Data assimilation with correlated observation errors: analysis accuracy with approximate error covariance matrices

L.M. Stewart[*], **S.L. Dance**[†], **N.K. Nichols**
School of Mathematical and Physical Sciences,
University of Reading,

In this paper we carry out numerical experiments using incremental 4D-Var (section 2) to address the following questions: Is it better to model observation error correlation structure approximately than not at all? and Is it computationally feasible to model observation error correlations? We use identical twin experiments so that a 'truth' trajectory is available and we are able to consider analysis errors explicitly. We specify two 'true' correlated observation error covariance matrices that we use to simulate synthetic observation errors, described in section 3. Howe

3.3 Markov matrix

The second approximate form of matrix that we employ is a Mark in Figure 2. The plots show that the eigenvalue size declines sharply as the eigenvalue number increases. The condition number (ratio of largest to smallest eigenvalue) for the Markov matrix is 400 and for the SOAR matrix is $4\,8 \times 10^5$. After 100 eigenvalues 80% and 99% of the overall uncertainty is represented for the Markov and SOAR matrix respectively. (Uncertainty percentages are calculated using (sum of eigenvalues used)/(sum of all eigenvalues or trace of matrix) $\times$ 100%.) Therefore we use 100 eigenpairs as an empirical upper limit to the number of eigenpairs used in the assimilation.

The number of eigenpairs we use in our approximations are = 10, = 20, = 50, and = 100. This represents 1%, 2%, 5% and 10% of the total number of eigenpairs. An ED approximation using the full number of eigenpairs = 1001 is equivalent to using the true error correlation matrix in the system. Obviously using all the eigenpairs is an expensive procedure and would not be attempted operationally. However in these smaller dimensioned experiments, knowing the performance of the

The height of the obstacle in the fluid is given by

$$
{}_{o}(\ {}_{D}) = \begin{cases} {}_{C}\ 1 - \frac{\ {}^{2}_{D}}{\ {}^{2}} & 0 \leq |\ {}_{D}| \leq \\ 0 & |\ {}_{D}| \quad 0 \text{ or } |\ {}_{D}| \end{cases}
$$

where ${}_{C}$ is the maximum height of the obstacle and is half the length over which the base of the obstacle extends. The values of and ${}_{C}$ are set as: $= 40\Delta\ {}_{D} = 0\ 4$m, ${}_{C} = 0\ 05$m. The temporal domain is 100 timesteps with step size $= 9\ 2 \times 10^{-3}$s. At $= 0$ the:

(a) **Error 1 (E1):** The norm of the analysis error in the true solution

$$\left\| \bar{\ }_{R_f} - {}^{*} \right\|_2 \tag{18}$$

where * is the true solution of the original model run from which the observations are sampled, and $\bar{\ }_{R_f}$ is the converged solution to the assimilation problem when the approximation ${}_{f}$ is used in calculating the cost function, but the observation errors themselves are sampled using the true error covariance ${}_{t}$;

(b) **Error 2 (E2):** The percentage norm of the analysis error in the converged solution relative to the norm of the true converged solution

8 Experiment 2: SOAR error correlation structure

In this section we consider the effect of our choice of the true observation error correlation structure. In Experiment 1 the true error correlation matrix was generated from a Markov distribution. We now change the correlation matrix to represent a SOAR distribution with length scale L_R = 0 1m. The matrix representations used to approximate this correlation structure are the same as those used in Experiment 1. Using a SOAR matrix will allow us to determine whether the Markov approximation also minimises analysis error when the true correlation structure is not in Markov form, and how well the ED and diagonal approximations perform in comparison.

The analysis errors E1 and E2 at = 0 for the different approximations to the SOAR error covariance matrix are given in Tables 3 and 4. Comparing the results to Table 1 and 2, we observe that the qualitative nature of the errors is very similar. For example, using the true error covariance matrix structure results in the smallest errors and diagonal approximations result in the largest errors. The approximations resulting in the smallest analysis errors are a Markov matrix with length scale L_R = 0 2m and an ED matrix using 100 eigenpairs. It is intuitive that a Markov matrix with a longer length scale is preferable, because of the longer tails in a SOAR function. The E2 error in the field is also small for Markov approximations with length scale between L_R = 0 2m and L_R = 0 05m, compared to a 9 4% error when a 4$\times$ diagonal approximation is used. Inflated diagonal approximations perform slightly worse than a simple diagonal approximation; this is in line with the information content results in Stewart (2010), when the background errors were uncorrelated.

It is also expected that an ED matrix using 100 eigenpairs results in a very small analysis error relative to the converged solution, because as we observed in section 3.4, 100 eigenpairs represent 99% of the overall uncertainty in the matrix. It is encouraging that an ED approximation using even fewer eigenpairs also results in an improved E2 error relative to a diagonal approximation; using 5% of the available eigenpairs results in an E2 error in the field of 2 3% compared to 5 3% when a diagonal approximation is used. The E1 errors in using an ED approxima

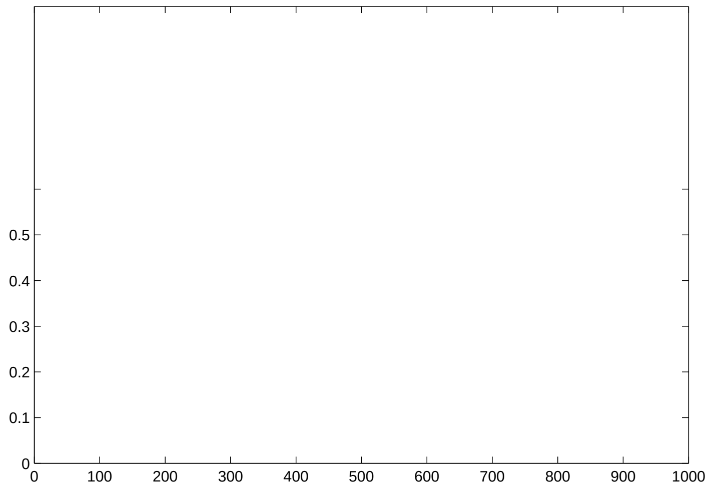
A.D. Collard. On the choice of observation errors for the assimilation of AIRS brightness temperatures: A theoretical study. ECMWF Technical Memoranda, AC/90, 2004.

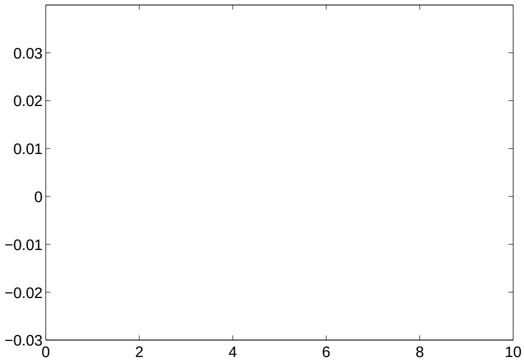
A.D. Collard. Selection of IASI channels for use in Numerical Weather Prediction. Q.J.R.Meteorol.Soc., 133:1977–1991, 2007.

P. Courtier, J.-N. Thépaut, and A.Holingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. Q.J.R.Meteorol.Soc., 120:1367–1387, 1994.

M.L. Dando, A.J. Thorpe, and J.R. Eyre. The optimal density of atmospheric sou8(o)0.433745(5.6167733(l)-0.3332r85.

sphe, nnmaeaten42y fmeyRRrat mu8(o)0.43302(o)0.4337449(p)-28.5809(e)-0.46074316r9422706(r)-1.06599(a)0.433749(t)-0..0194o9(8)0.43


0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
400
500
600
700
800
900
1000


0.03
0.02
0.01
0
−0.01
−0.02
−0.03
0
2
4
6
8
10

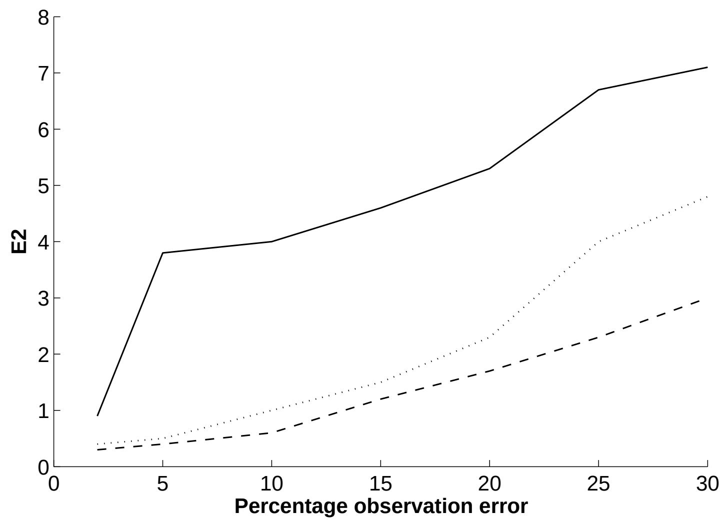


Figure 6: As in Figure 5 but for field.

Approximation	E1: $\|\bar{}_{R_f} - {}^*\|_2$	$\|\bar{}_{R_f} - \bar{}_{R_t}\|_2$	E2 (%)
Truth	0.20	0	0
Diagonal	0.30	0.23	7.2
2 $\times$ Diagonal	0.31	0.23	7.2
4 $\times$ Diagonal	0.31	0.24	7.5
Markov (L_R = 0 2)	0.21	0.06	1.9
Markov (L_R = 0 1)	0.20	0	0
Markov (L_R = 0 05)	0.21	0.05	1.6
Markov (L_R = 0 01)	0.27	0.18	5.6
ED (= 10)	0.28	0.19	5.9
ED (= 20)	0.28	0.19	5.9
ED (= 50)	0.25	0.15	4.7
ED (= 100)	0.23	0.10	3.1

Table 1: Analysis errors in field at = 0 for different approximations to a Markov error covariance matrix ($\|\bar{}_R\|_2$ = 3 20)

Approximation	E1: $\|\bar{}_R$

Approximation	E1: $\|\bar{\ }_{R_f} - {}^{*}\|_2$	$\|\bar{\ }_{R_f} - \bar{\ }_{R_t}\|_2$	E2 (%)
Truth	0.57	0	0
Diagonal	3.36	3.32	5.3
2 $\times$ Diagonal	3.59	3.55	5.7
4 $\times$ Diagonal	3.99	3.95	6.3
Markov (L_R = 0 2)	0.81	0.63	1.0
Markov(L_R = 0 1)	1.18	1.06	1.7
Markov (L_R = 0 05)	1.69	1.60	2.6
Markov (L_R = 0 01)	2.89	2.84	4.5
ED (= 10)	3.90	3.87	6.2
ED (= 20)	3.71	3.67	5.9
ED (= 50)	1.56	1.45	2.3
ED (= 100)	1.06	0.85	1.4

Table 4: Analysis errors in field at = 0 for different diagonal approximations to a SOAR error covariance matrix ($\|\bar{\ }_R\|_2$ = 62 54)