School of Mathematical
and Physical Sciences

University of Reading

Department of Mathematics and Statistics

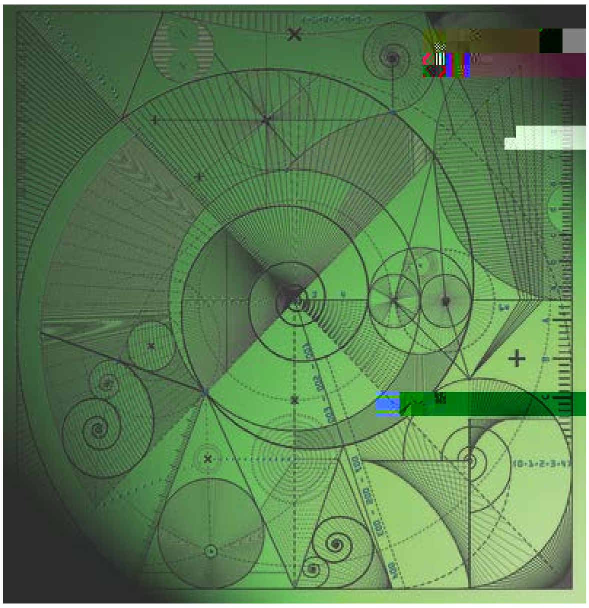
Preprint MPS-2015-04

2 March 2015

Theoretical insight into diagnosing observation error correlations using background and analysis innovation statistics

by

J.A. Waller, **S.L. Dance** and **N.K. Nichols**



Theoretical insight into diagnosing observation error correlations using background and analysis innovation statistics

J. A. Waller[1] , S. L. Dance[1], and N. K. Nichols[1]

1. School of Mathematical and Physical Sciences, University of Reading, Reading, Berkshire, RG6 6BB, United Kingdom

Abstract

To improve the quantity and impact of observations used in data assimilation it is
popular method is a diagnostic that make
and analysis innovations. The accuracy
diagnostic is sensitive to the di erence be

1 Introduction

Data assimilation techniques combine model states, known as forecasts or backgrounds, in simple model experiments in both variational [Stewart, 2010] and ensemble [Li et al., 2009, Miyoshi et al., 2013, Waller et al., 2014a] data assimilation systems and to estimate time varying observation errors [Waller et al., 2014a]. The diagnostic has also been applied to operational NWP observation types such as ATOVS, AIRS and IASI to calculate inter-channel error covariances [Stewart et al., 2009, 2014, Bormann and Bauer, 2010, Bormann et al., 2010, Weston et al., 2014]. When the correlated errors calculated using the diagnostic matrix variance increases as assumed observation error variance increases.

The power in the largest length scales can be obtained by considering the eigenstructure of the estimated matrix and provides some insight into the behaviour of the estimated correlation length scale. These results provide an understanding of the diagnostic that can aid the interpretation of results when the diagnostic is used to estimate spatial correlations in an operational setting e.g. Waller et al. [2014c]. We note that in the cases presented here, the statistical nature of the estimation is not considered since results are calculated is the difference between the observation y and the mapping of the forecast vector, x^b, into observation space by the observation operator H. The analysis innovations,

$$d^o_a = y - H(x^a); \qquad (2)$$

are similar to the background innovations, but with the forecast vector replaced by the analysis vector x^a. Desroziers et al. [2005] assume that the analysis is determined using,

$$x^a = x^b + \tilde{B}H^T(H\tilde{B}H^T + \tilde{R})^{-1} d^o_b; \qquad (3)$$

where H is the observation operator linearised about the current state and $\tilde{R}$ and $\tilde{B}$ are the assumed observation and background error statistics used to weight the observations and background in the assimilation. Taking the statistical expectation of the outer product of the analysis and background innovations and assuming that the forecast and observation errors are uncorrelated results in

$$E[d^o_a d^{oT}_b] = \tilde{R}(H\tilde{B}H^T + \tilde{R})^{-1}(HBH^T + R) = R^e; \qquad (4)$$

where R^e is the estimated observation error covariance matrix and B and R are the exact background and observation covariance ctionoeI-2.2219(nf)3.5[(o0(e)-307.594(t)-0.649399(h)1.56146

Menard et al. [2009] also show, again in the scalar case, that if both variances are iterated concurrently then the diagnostics converge in one iteratio observation operator so long as HBH^T and $H\overset{e}{B}H^T$

In our case we are considering the eigenvalues m of n correlation matrices with ones on the diagonal, and therefore the trace of such a matrix and hence the sum of the eigenvalues will be n. This allows the estimated error variance to be written as,

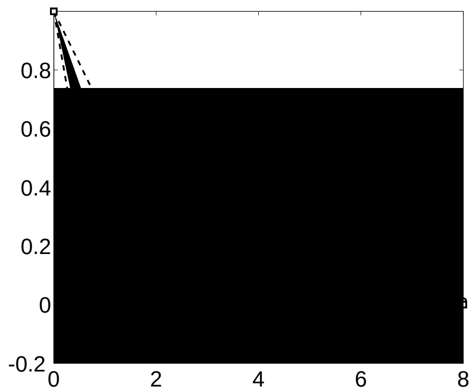
$$_e = \frac{1}{n}\sum_k \frac{_k + \quad _k}{1 + (\tilde{\ } = \tilde{\ })\tilde{\ }}$$

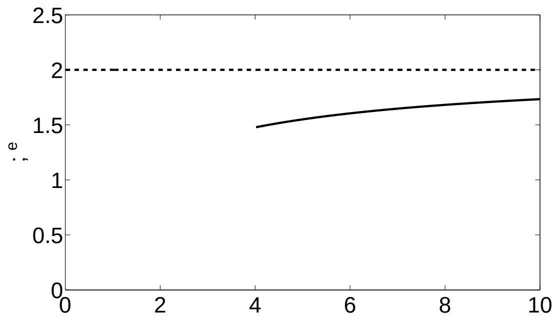
where s_k = $_k$ +

The misspecified length scale in $\hat{B}$ results in correlations in the estimated observation

Table 1: Estimated observation error variances when lengthscales (defined using the SOAR function in equation (22)) and variances in $\tilde{R}$ and $\tilde{B}$ used in the assimilation are incorrect. The exact observation and background error variances are set to = = 1 and length scales to $L = 2$ and $L = 5$ respectively. The matrix $\tilde{R}$ used in the assimilation is always diagonal.

Exp. Label	~	~	$\tilde{B}$ Length scale (L)	e
Control	1	1	5	0.94
0:5	0.5	1	5	0.68
1:1	1.1	1	5	0.98
2	2	1	5	1.22
10	10	1	5	1.73
0:5	1	0.5	5	1.22
0:75	1	0.75	5	1.06
0:99	1	0.99	5	0.94
1:5	1	1.5	5	0.78
2	1	2	5	0.68
L3	1	1	3	0.91
L4	1	1	4	0.92
L6	1	1	6	0.97
L7	1	1	7	1.00
1:5L6	1	1.5	6	0.82
2L6	1	2	6	0.73
1:5L7	1	1.5	7	0.87
2L7	1	2	7	0.77
2 1:5L6	2	1.5	6	1.08
2 2L6	2	2	6	0.97
2 1:5L7	2	1.5	7	1.10
2 2L7	2	2	7	1.00

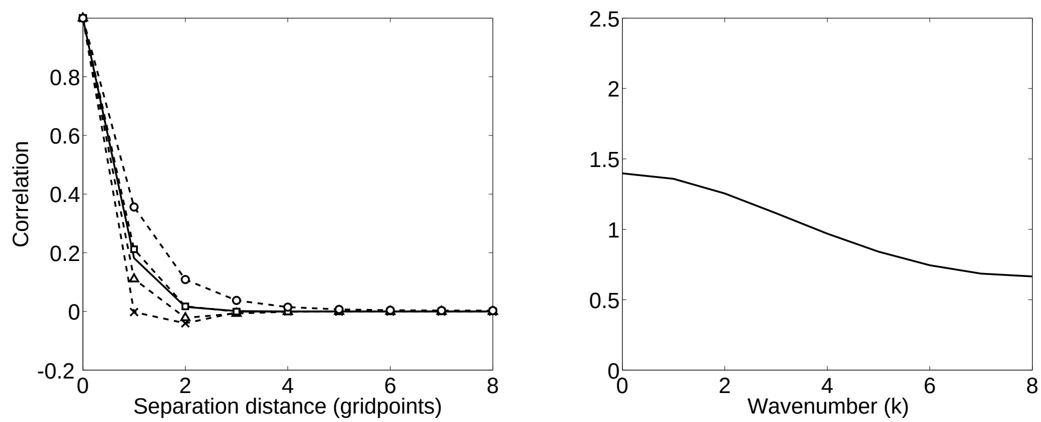

0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8



2.5
2
1.5
1
0.5
0
0
2
4
6
8
10

estimated first eigenvalue, and hence the power in the lowest wave numbers, to increase as a function of ~.

We plot the estimated correlation function and corresponding eigenvalues in Figure 5. From



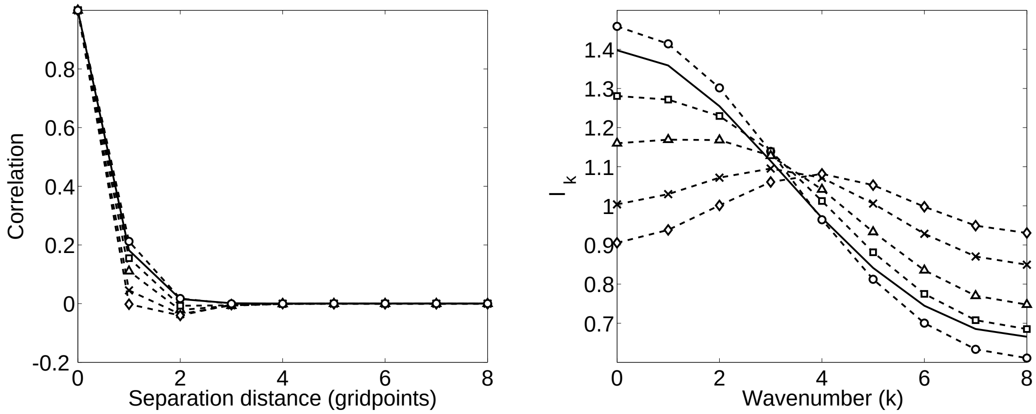


Figure 7: Estimated observation error correlations and corresponding eigenvalues for Experiments 0.5 ($\tilde{}$ = 0.5, dashed line circles), 0.75 ($\tilde{}$ = 0.75 , dashed line squares), Control ($\tilde{}$ = 1.0, dashed line triangles), 1.5 ($\tilde{}$ = 1.5, dashed line crosses) and 2 ($\tilde{}$ = 2.0, dashed line diamonds) when the variance in the background error covariance matrix is misspecified in the assimilation. The exact observation correlation function C_r (solid line) is plotted for comparison.

background error variance is largest. The observation error correlation length scale remains underestimated as the background error variance decreases. The observation error correlation length scale is only overestimated when the assumed background error variance is half the value of the actual background error variance or less. Considering the eigenvalues of R^e we see that unless the assumed background error variance is much smaller than the true background error variance, the power in the low wave numbers (large scales) will be underestimated and the power in the high wave numbers (small scales) will be overestimated. This is consistent with the theoretical result that the first eigenvalue decreases as $\tilde{}$ increases.

In summary, in the case of misspecified background error variances we are able to show that:

- As the assumed background error variance increases the estimated observation error variance decreases.
- As the assumed background error variance increases the estimated power in the largest scales decreases.
- In the multi-dimensional case, where observation errors are neglected, it does not hold that an assumed observation error variance that is too small (large) will result in an estimated observation error variance that is overestimated (underestimated).

5.3.4 Impact of misspecifying the background error correlation length scale

We now consider what happens when the background error variance is correctly specified but the correlation length scale is misspecified. We give the assumed background correlation function length scales and estimated observation error variances in Table 1, Experiments L3, L4, L6 and L7 and we plot the estimated observation error correlation functions and corresponding eigenvalues in Figure 8. Again we plot the result from the control experiment for comparison.

From the table and figure we see that:

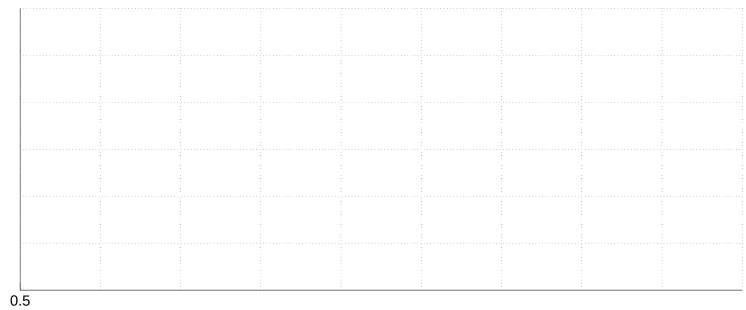
- As the assumed background error length scale increases the estimated observation error variance increases.
- as the assumed background error length scale increases the estimated correlation length scale and leading eigenvalues decrease.

However, in all but the case of the largest length scale the observation error variances are underestimated. We see that when the assumed background correlation length scale is too

5.3.6 Impact of misspecifying the background error varianc e and correlation length scale

2L7 and F

we nd that:


0.5

5.3.8 Impact of misspecifying all assumed error variances a nd length scales

are also able to prove that:

- Estimated observation error variance increases as assumed observation error variance increases.
- Estimated observation error variance decreases as assumed background error variance increases.

We are able to verify this through our simple experiments and show that the bounds on the variance are respected in the experimental cases. Under the additional assumption of the exact and assumed correlation matrices having non-negative coefficients, we are also able to prove results relating to the first eigenvalue, which provide some information about the estimated correlation length scale. We prove that:

- The power in the large scales of the estimated observation error correlation matrix increases as assumed observation error variance increases
- The power in the large scales of the estimated observation error correlation matrix decreases as assumed background error variance increases.

This provides some insight into the behaviour of the estimated correlation length scale. In general we are able to show that if observation error correlations are neglected in the assumed observation error covariance matrix, then it is likely that the diagnostic will underestimate the strength of the correlations, though the result from the diagnostic will be better a estimate of $\mathbf{R}$ than one that is assumed diagonal.

The theoretical results are more complex when the background error length scales are misspecified, so to aid our understanding we considered the results of some simple experiments. A more detailed knowledge of the exact and assumed spectra are required to predict whether the variance will increase or decrease as the assumed background error length scales are increased. It does appear, however, in the case of the SOAR function that an increase in the assumed background error length scale causes a reduction in the estimated observation error length scales.

Another important conclusion drawn from the illustrative examples is that if the observation error covariance matrix is assumed diagonal in the assimilation, then the observation error correlation matrix calculated by the diagnostic is li

G. Desroziers and S. Ivanov. Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. Quarterly Journal of the Royal Meteorological Society, 127: 1433{1452, 2001. doi: 10.1002/qj.49712757417.

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.

G. Desroziers, L. Berre, and B. Chapnik. Objective validation of data assimilation systems: diagnosing sub-optimality. In Proceedings of ECMWF Workshop on diagnostics of data assimilation system performance, 15-17 June 2009, 2009.

G. H. Golub and C. F. Van Loan. Matrix Computations, chapter 7, page 348. Baltimore: The Johns Hopkins University Press, fourth edition, 2013.

R. Gray. Toeplitz and Circulant Matrices: A Review Now Publishers, 2006.

S. B. Healy and A. A. White. Use of discrete Fourier transform in the 1D-Var retrieval

L. M. Stewart. Correlated observation errors in data assimilation. PhD thesis, University of Reading, 2010. http://www.reading.ac.uk/maths-and-stats/research/theses/maths-phdtheses.aspx.