The University of Reading

Correlated observation errors in data assimilation

L.M. Stewart, S.L. Dance and N.K. Nichols

NUMERICAL ANALYSIS REPORT 3/2007

Presented at the ICFD International Conference on Numerical Methods for Fluid Dynamics, Reading, March 2007

Department of Mathematics
The University of Reading
Whiteknights, PO Box 220
Reading
Berkshire RG6 6AX

Department of Mathematics

Abstract

Data assimilation provides techniques for combining observations and prior model forecasts to create initial conditions for numerical weather prediction (NWP). The relative weighting assigned to each observation in the analysis is determined by its associated error. Remote sensing data usually has correlated errors, but the correlations are typically ignored in NWP. Here we describe three approaches to the treatment of observation error correlations. For an idealised data set, the information content under each simplified assumption is compared to that under the correct correlation specification. Treating the errors as uncorrelated results in a significant loss of information. However, retention of an approximated correlation gives clear benefits.

Contents

1 Introduction

In Numerical Weather Prediction, an accurate, high resolution representation of the current state of the atmosphere is needed as an initial condition for the propagation of a weather forecast. Data assimilation combines observations of atmospheric variables with *a priori* knowledge of the atmosphere to obtain a consistent representation. The weighted importance of each is determined by the size of their associated errors, so it is crucial to the accuracy of the analysis that these errors be specified correctly.

Satellite instruments are regularly calibrated, so that instrument errors are usually uncorrelated. However, observation error correlations will arise from observation pre-processing and errors in the forward model, including representativity (when phenomena observed by a sensitive instrument cannot be resolved by the model). The inclusion of such correlations results in a lower weighting of the observations in the analysis, when compared with assimilating the same observations with the same error variances and no correlations. The correlations increase the accuracy of gradients of the observed field represented in the analysis, but make a smaller impact on the accuracy of the observed field itself [9]. They also act in conjunction with the prior error covariance to specify how observation information should be smoothed.

Unfortunately, such errors are not easily measured. Even when good estimates can be made, the number of observations is of order 10^6 [2], and so the storage and subsequent computation using observation error correlations is infeasible. Hence operationally, observations are usually assumed uncorrelated. In most cases, to compensate for the lack of correlation, the observation error variances are inflated so that the observations have the correct lower weighting in the analysis. However, results have shown [3] that error variances can be made at most 2-4 times larger before the analysis field becomes degraded through excessive error amplification. So variance enlargement is constrained by the need for a physically accurate error estimate.

The assumption of zero correlations is often used in conjunction with data thinning methods such as superobbing [1]. This reduces the density of data by averaging

the properties of observations in a region, and assigning this average as a single observation value. Under such assumptions, increasing observation density beyond some threshold value has been shown to yield very little or no improvement in analysis accuracy [7]. Clearly these methods discard much of the available information. Although these may be appropriate when the spatial resolution of the observations is denser than the model grid, recent technological advances have challenged their practicality. For example, as high resolution models are used in forecasting convective storms, there is a requirement to retain all the available data to provide detail on the appropriate scales. Such shortcomings suggest that an alternative approach to dealing with observation error correlations is needed.

Approximating observation error correlation is a relatively new direction of research but progress has been made. Healy and White [6] have used circulant matrices to approximate symmetric Toeplitz observation error covariance matrices. Results indicated that assuming uncorrelated observation errors gave misleading estimates of information content. Fisher [5] proposes giving the observation error covariance matrix a block diagonal structure, with (uncorrelated) blocks corresponding to different instruments or channels; individual block matrices are approximated by a truncated eigendecomposition. On a simple domain, spurious long range correlations have been observed.

In this paper, we expand on the work of the above and quantify the loss in information content when ignoring error correlations, using simplified diagonal matrix structures, and using Fisher proposed structures. We further extend Fisher's work and investigate long range correlations on larger domains. The question of whether information loss is significant enough to warrant a change in operational treatment is addressed. In Section 2 we give a brief overview of data assimilation and information theory for this problem, and the structure of the experiment. Results and subsequent conclusions will be given in Section 3 and 4, respectively.

2 Methods and data

2.1 Data assimilation

The main aim of variational data assimilation methods is to minimise a cost function which measures the distance of the solution to the background and the observations, weighted by the inverse of their respective errors,

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x_B})^{\mathbf{T}}\mathbf{B^{-1}}(\mathbf{x} - \mathbf{x_B}) + (\mathbf{y} - \mathbf{h}(\mathbf{x}))^{\mathbf{T}}\mathbf{R^{-1}}(\mathbf{y} - \mathbf{h}(\mathbf{x})), \tag{1}$$

where $\mathbf{x}$ is the model state vector, $\mathbf{x_B}$ is the background state, and $\mathbf{h}$ is the observation operator (known as the forward model). For simplicity in our analysis we use a linear approximation to the forward model, $\mathbf{H} \approx \mathbf{h}$. The vector $\mathbf{y}$ is the observation vector, whose relationship to the model state vector is given by $\mathbf{y} = \mathbf{Hx} + \epsilon^{\mathbf{o}}$, where $\epsilon^{\mathbf{o}}$ is the measurement error. Matrices $\mathbf{B}$ and $\mathbf{R}$ are the background and observation error covariance matrices, respectively. $B(i, j)$ describes the error covariance between components i and j of $\mathbf{x_B}$, and $R(i, j)$ describes the error covariance between components i and j of $\mathbf{y}$.

Equation (1) can be solved to determine the value, $\mathbf{x_A}$, of the model state $\mathbf{x}$ that minimises J; $\mathbf{x_A}$ is used for the initialisation of the model variables in a forecast,

$$\mathbf{x_A} \;=\; \mathbf{x_B} + \mathbf{BH^T}(\mathbf{HBH^T} + \mathbf{R})^{-\mathbf{1}}(\mathbf{y} - \mathbf{H}(\mathbf{x_B})). \tag{2}$$

2.2 Information theory

In ignoring observation error correlation, we overlook a portion of the available data, and so information that could be utilised is lost. In this context, the amount of information provided by a set of observations is a measure of how far they go to reduce uncertainty in our analysis. To numerically evaluate the information lost when using simplified observation error correlations, we use Shannon Information Content and the number of degrees of freedom of signal.

2.2.1 Shannon Information Content

The Shannon Information Content (SIC) is a measure of the reduction of entropy. Entropy physically corresponds to the volume of state space occupied by the probability density function (pdf) describing the knowledge of the state. Assuming all pdfs are Gaussian, [8],

$$SIC = \frac{1}{2}\left|\mathbf{S_A^{-1}B}\right|, \tag{3}$$

$$\mathbf{S_A^{-1}} = \mathbf{H^T R^{-1} H + B^{-1}}, \tag{4}$$

where $\mathbf{S_A}$ is the analysis error covariance matrix ($S_A(i,j)$ describes the error covariance between components i and j of $\mathbf{x_A}$). The larger the SIC, the greater the reduction in uncertainty in our analysis.

2.2.2 Degrees of freedom of signal

The number of degrees of freedom of signal (dof_S) indicate the number of quantities deemed measured by the observations; the closer dof_S is to the total number of degrees of freedom (dof), the more information the observations have provided.

We have an initial covariance matrix $\mathbf{B}$, and performing an analysis to minimise the variance in observed directions gives us a posterior matrix $\mathbf{S_A}$. The size of the eigenvalues in each matrix represent the size of the uncertainty in the direction of the associated eigenvector; in comparing the eigenvalues of the two, we can determine the reduction in uncertainty.

Take a non-singular square matrix $\mathbf{L}$, as in [4], such that $\mathbf{LBL^T} = \mathbf{I}$ and $\mathbf{LS_AL^T} = \hat{\mathbf{S}}_\mathbf{A}$. This transformation is not unique as we can replace $\mathbf{L}$ by $\mathbf{X^T L}$ where $\mathbf{X}$ is an orthogonal matrix. Now if we take $\mathbf{X}$ to be the matrix of the eigenvectors of $\hat{\mathbf{S}}_\mathbf{A}$, then we simultaneously reduce $\mathbf{B}$ to the identity matrix and $\mathbf{S_A}$ to a diagonal matrix of its eigenvalues, $\mathbf{\Lambda}$;

$$\begin{aligned}\mathbf{X^T LBL^T X} &= \mathbf{X^T X = I},\\ \mathbf{X^T LS_AL^T X} &= \mathbf{X^T \hat{S}_A X = \Lambda}.\end{aligned}$$

After this transformation, the diagonal elements (eigenvalues) of the transformed matrix $\mathbf{B}$ are unity, and each corresponds to an individual dof. The eigenvalues of $\hat{\mathbf{S}}_{\mathbf{A}}$ may therefore be interpreted as the relative reduction of variance in each of the independent directions. So the dof_S are given by

$$dof_S = N - \text{trace}(\mathbf{\Lambda}). \tag{5}$$

2.3 Idealised data set

To quantitatively evaluate information content under different treatments of error correlations, we investigate a scalar quantity on an idealised data set. Consider observations on a regular flat $N \times N$ grid, with a 200km spacing between observation points. Assume that every observation is taken directly, $\mathbf{H} = \mathbf{I}$, and the background errors are uniform and described by the correlation function $B_{ij} = \exp\left(-r_{ij}^2/2L\right)$, where r_{ij} is the Euclidean distance between point i and j and $L = 190$ is the length scale.

The test error covariance matrix $\mathbf{R_t}$ is calculated using empirically derived error variances [2], and isotropic correlations described by $C_{ij} = \left(1 + \frac{r_{ij}}{L}\right) \exp\left(-r_{ij}/L\right)$, $L = 190$. This produces a correlation matrix $\mathbf{C}$, with components C_{ij}, which can be used to alternatively describe $\mathbf{R_t}$: $\mathbf{R_t} = \mathbf{D^{1/2}CD^{1/2}}$ where $\mathbf{D}$ is the diagonal matrix of error variances. The values are obtained from the analysis of pairs of collocations between Atmospheric Motion Vectors (AMVs) and radiosonde observations. It is this matrix against which we measure information loss.

2.4 Observation error correlation matrix structures

Using the above experimental structure, we compare four different approaches to observation error correlation:

1. Use the test error covariance matrix $\mathbf{R_t}$;

2. Set the correlations to zero in $\mathbf{R_t}$;

3. Set the correlations to zero in $\mathbf{R_t}$ and inflate the error variances;

4. Describe $\mathbf{R}$ by a truncated eigendecomposition [5];

$$\tilde{\mathbf{R}} = \mathbf{D^{1/2}}(\alpha\mathbf{I} + \sum_{\mathbf{k=1}}^{\mathbf{K}} \left(\lambda_\mathbf{k} - \alpha\right)\mathbf{v_k}\mathbf{v_k^T}) \, \mathbf{D^{1/2}} = \mathbf{D^{1/2}}\tilde{\mathbf{C}}\mathbf{D^{1/2}} \tag{6}$$

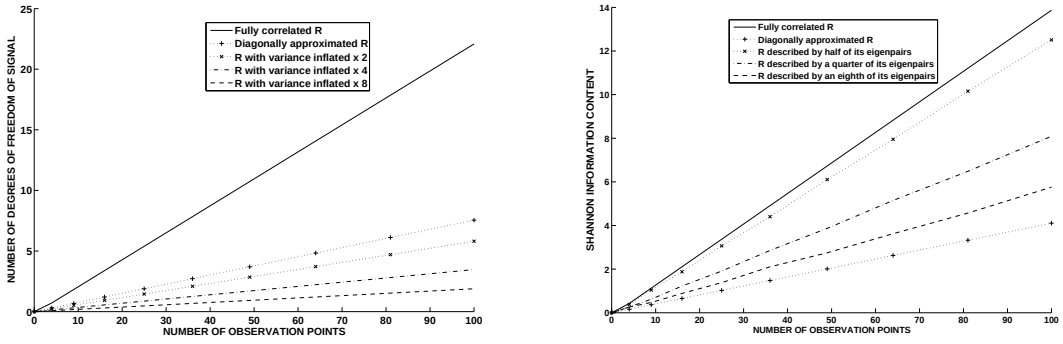
$(\lambda_k, \mathbf{v_k})$ is an eigenvalue, eigenvector pair of $\mathbf{C}$, K is the number of leading eigenpairs used in the approximation, and α is chosen such that trace($\tilde{\mathbf{R}}$)=trace($\mathbf{D}$), i.e, so that there is no mis-approximation of the total error variance.

3 Results

First we present the results obtained when using a diagonal approximation of $\mathbf{R_t}$: Approach 2 and Approach 3. As anticipated, under simplified assumptions of observation error correlations, information is lost. Both the SIC and dof_S are directly proportional to the number of observation points; so increasing our grid size provides access to more information. But as the number of observation points increases, the greater the difference in information between $\mathbf{R_t}$ and the diagonal approximation used in Approach 2. For a grid with 100 points, both the SIC and dof_S decrease by 75% between $\mathbf{R_t}$ and the diagonal approximation (Figure 1(a)).

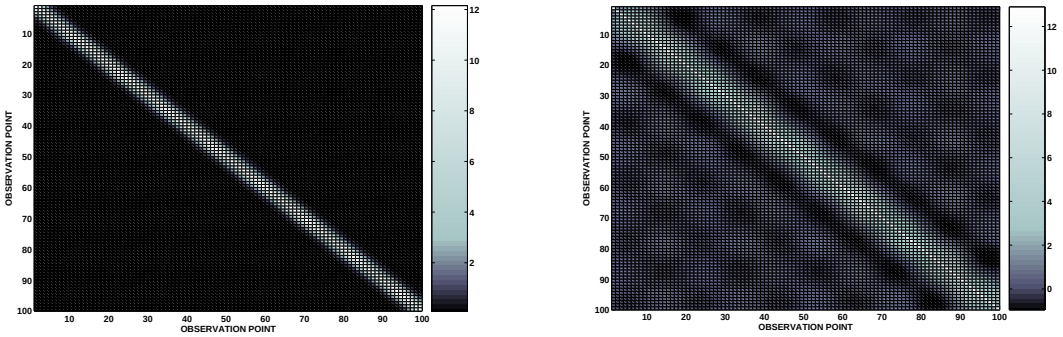
The depletion in information increases with the scale of variance enlargement. Variance enlargement (Approach 3) is shown to have a detrimental effect on the information provided by a set of observations; more so than a simple diagonal approximation. For a grid with 100 points, the decrease in information is up to 93% between $\mathbf{R_t}$ and a diagonal approximation with an 8$\times$ variance enlargement (Figure 1(a)). If, as Collard [3] suggests, we are limited to a variance enlargement of between 2-4 times, then we still lose between 74% and 84% of the information available.

In Approach 4 we acknowledge error correlations by forming a truncated eigendecompostion of $\mathbf{R_t}$. Results show that the more eigenpairs used in the decomposition, the smaller the difference in the information between $\mathbf{R_t}$ and the eigenpair approximation. The eigenpair approximation retains a higher percentage of the information available than the diagonal approximations. For a grid size with 100 points, even when $\mathbf{R_t}$ is described by an eighth of its eigenpairs, the resultant loss of information is considerably less than for any diagonal approximation. When $\mathbf{R_t}$ is described by



(a) (b)

Figure 1: (a) The dof_S for different scales of variance enlargement; (b) The SIC for a correlated ($\mathbf{R_t}$), uncorrelated and eigenpair described $\mathbf{R}$.



(a) (b)

Figure 2: Spatial structure of correlations for (a) $\mathbf{R_t}$ and (b) $\mathbf{R_t}$ described by its leading 10 eigenpairs.

half of its eigenpairs, the SIC only decreases by 5% (Figure 1(b)). In describing $\mathbf{R_t}$ by its eigenpairs, using too few will lead to spurious error correlations (Figure 2), as suggested by Fisher [5]. Under this set up the correlations are not large enough to discount the approach, but care must be taken for larger problems.

4 Conclusions

We have evaluated the loss of information under three different treatments of correlated observation errors. Approximating $\mathbf{R_t}$ with a diagonal matrix and ignoring error correlations, has been shown as so detrimental to the information provided by a set of observations that an alternative approach must be sought. One such ap-

proach is the approximation of $\mathbf{R_t}$ through its leading eigenpairs; this retains much of the information available even with less than half the available eigenpairs. But, addressing Fisher's concerns [5], we find that spurious long range correlations are present even for larger observation sets.

Although creating a truncated decomposition of $\mathbf{R_t}$ is more costly than the traditional operational approach, it includes some of the correlation structure of $\mathbf{R_t}$ and is still realtively easy to invert. If the computational cost involved in this is not too extensive, then it may be possible to include correlations operationally, leading to a more accurate forecast.

The above results are only currently applicable to the idealised framework in which they have been obtained. Although we have used empirically derived observation errors, the background errors are not realistic. Since the calculations of information are dependent on both $\mathbf{B}$ and the idealised observation operator $\mathbf{H}$, a more realistic specification of these would produce more general results. Also, this paper has not addressed the operational feasibility of including correlated observation errors in the data assimilation algorithm. In future work, comparisons will be made using more realistic models, and using different approaches to incorporating correlation structures in the observation error covariance matrix.

References

[1] Berger H, Forsythe M. *Satellite Wind Superobbing.* Forecasting Research Technical Report. Met Office, 2004; **451**.

[2] Bormann N, Saarinen S, Kelly G, Thepaut J-N. *The spatial structure of observation errors in Atmospheric Motion Vectors from geostationary satellite data.* Mon. Wea. Rev., 2003. **131**:706–718.

[3] Collard A. *On the choice of observation errors for the Assimilation of AIRS brightness temperatures: A theoretical study.* ECMWF report, 2004. **AC/90**.

[4] Fisher M. *Estimation of Entropy Reduction and Degrees of Freedom for Signal for Large Variational Analysis Systems*. ECMWF Technical Memoranda, 2003. **397**.

[5] Fisher M. *Accounting for Correlated Observation Error in the ECMWF Analysis*. ECMWF Technical Memoranda, 2005. **MF/05106**

[6] Healy S, White A. *Use of discrete Fourier transforms in the 1D-Var retrieval problem*. Quart.J.Roy.Met.Soc., 2005. **131**:63–72.

[7] Liu Z-Q, Rabier F. *The interaction between model resolution, observation resolution and observation density in data assimilation*. Quart.J.Roy.Met.Soc., 2002. **128**:1367–1386.

[8] Rodgers C. *Inverse Methods of Atmospheric Sounding, Theory and Practice, volume 2 of Atmospheric, Oceanic and Planetary Physics*. World Scientific, 2000.

[9] Seaman R. *Absolute and differential accuracy of analyses acheivable with specified observation network characteristics*. Mon. Wea. Rev., 1977; **105**:1211–1222.