INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2008; **56**:1521–1527
Published online 5 December 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.1636

Correlated observation errors in data assimilation

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

The Department of Mathematics, The University of Reading, P.O. Box 220, Reading RG6 6AX, U.K.

SUMMARY

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 idealized data set, the information content under each simplified assumption is compared with that under correct correlation specification. Treating the errors as uncorrelated results in a significant loss of information. However, retention of an approximated correlation gives clear benefits. Copyright © 2007 John Wiley & Sons, Ltd.

Received 29 April 2007; Revised 28 August 2007; Accepted 28 August 2007

KEY WORDS: observation error correlation; Shannon Information Content; degrees of freedom; data assimilation

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; hence, 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 (where phenomena observed by a sensitive instrument

*Correspondence to: L. M. Stewart, The Department of Mathematics, The University of Reading, P.O. Box 220, Reading RG6 6AX, U.K.
†E-mail: l.m.stewart@reading.ac.uk

Contract/grant sponsor: NERC
Contract/grant sponsor: Met Office

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 [1]. 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, and as the number of observations is of order 10^6 [2], 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, variance enlargement is constrained by the need for a physically accurate error estimate [3].

The assumption of zero correlations is often used in conjunction with data thinning methods such as superobbing [4]. 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 the analysis accuracy [5]. Although discarding available information may be appropriate when the spatial resolution of the observations is denser than the model grid, recent technological advances have challenged the practicality of such methods. 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 [7] 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's 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 are given in Sections 3 and 4, respectively.

2. METHODS AND DATA

2.1. Data assimilation

The main aim of variational data assimilation methods is to minimize a cost function that measures the distance of the solution to the background and the observations, weighted by the inverse of their respective error covariances

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

DOI: 10.1002/fld

where $\mathbf{x}$ is the model state vector, $\mathbf{x}^{\mathrm{b}}$ 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, $\mathbf{H}x \approx \mathbf{h}(x)$. The vector $\mathbf{y}$ is the observation vector, whose relationship to the model state vector, under the assumption of linearity, is given by $\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}^{\mathrm{o}}$, where $\boldsymbol{\varepsilon}^{\mathrm{o}}$ 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 $\mathbf{x}^{\mathrm{b}}$, 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, $\mathbf{x}^{\mathrm{a}}$, of the model state $\mathbf{x}$ that minimizes the cost function J:

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{b}}), \quad \mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \tag{2}$$

Known as the *analysis*, $\mathbf{x}^{\mathrm{a}}$ is used for the initialization of the model variables in a forecast.

2.2. Information theory

In ignoring observation error correlations, we overlook a portion of the available data, and so information that could be utilized is lost. In this context, the amount of information provided by a set of observations is a measure of how much they reduce uncertainty in our analysis. To evaluate numerically the information lost when using simplified observation error correlations, we use Shannon Information Content (SIC) and the number of degrees of freedom of signal ($\mathrm{dof_S}$) [8].

2.2.1. Shannon Information Content. The SIC is a measure of the reduction of entropy [8]. 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 then

$$\mathrm{SIC} = \tfrac{1}{2} \ln |\mathbf{S}_{\mathrm{a}}^{-1}\mathbf{B}| \tag{3}$$

where $\mathbf{S}_{\mathrm{a}}$ is the analysis error covariance matrix given by (i) $\mathbf{S}_{\mathrm{a}}^{(1)} = (\mathbf{H}^{\mathrm{T}}\mathbf{R}_{\mathrm{t}}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}$ if the correct covariance matrix $\mathbf{R}_{\mathrm{t}}$ is used; (ii) $\mathbf{S}_{\mathrm{a}}^{(2)} = \mathbf{S}_{\mathrm{a}}^{(1)} + \mathbf{K}(\mathbf{R}_{\mathrm{t}} - \mathbf{R}_{\mathrm{f}})\mathbf{K}^{\mathrm{T}}$ if an approximation, $\mathbf{R}_{\mathrm{f}}$, to $\mathbf{R}_{\mathrm{t}}$ is used. An alternative philosophy in which $\mathbf{S}_{\mathrm{a}}^{(1)}$ is used in all cases is discussed in [9]. The larger the SIC, the greater the reduction in uncertainty in our analysis.

2.2.2. Degrees of freedom of signal. The number of $\mathrm{dof_S}$ indicates the number of quantities deemed measured by the observations; the closer the $\mathrm{dof_S}$ is to the total number of dof, the more information the observations have provided.

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

We take a non-singular square matrix $\mathbf{L}$, as in [10], such that $\mathbf{L}\mathbf{B}\mathbf{L}^{\mathrm{T}} = \mathbf{I}$ and $\mathbf{L}\mathbf{S}_{\mathrm{a}}\mathbf{L}^{\mathrm{T}} = \hat{\mathbf{S}}_{\mathrm{a}}$. This transformation is not unique as we can replace $\mathbf{L}$ by $\mathbf{X}^{\mathrm{T}}\mathbf{L}$ where $\mathbf{X}$ is an orthogonal matrix. Now if we take $\mathbf{X}$ to be the matrix of the eigenvectors of $\hat{\mathbf{S}}_{\mathrm{a}}$, then we simultaneously reduce $\mathbf{B}$ to the identity matrix and $\hat{\mathbf{S}}_{\mathrm{a}}$ to a diagonal matrix of its eigenvalues, $\boldsymbol{\Lambda}$;

$$\mathbf{X}^{\mathrm{T}}\mathbf{L}\mathbf{B}\mathbf{L}^{\mathrm{T}}\mathbf{X} = \mathbf{X}^{\mathrm{T}}\mathbf{X} = \mathbf{I}, \quad \mathbf{X}^{\mathrm{T}}\mathbf{L}\mathbf{S}_{\mathrm{a}}\mathbf{L}^{\mathrm{T}}\mathbf{X} = \mathbf{X}^{\mathrm{T}}\hat{\mathbf{S}}_{\mathrm{a}}\mathbf{X} = \boldsymbol{\Lambda} \tag{4}$$

DOI: 10.1002/fld

After this transformation, the diagonal elements (eigenvalues) of the transformed matrix $\mathbf{LBL}^{\mathrm{T}}$ are found to be unity and each corresponds to an individual dof. The eigenvalues of $\hat{\mathbf{S}}_{\mathrm{a}}$ may therefore be interpreted as the relative reduction of variance in each of the independent directions. Hence, if N is the total number of dof, then $\mathrm{dof}_{\mathrm{S}}$ is given by

$$\mathrm{dof}_{\mathrm{S}} = N - \mathrm{trace}(\mathbf{\Lambda}) = N - \mathrm{trace}(\hat{\mathbf{S}}_{\mathrm{a}}) = N - \mathrm{trace}(\mathbf{B}^{-1}\mathbf{S}_{\mathrm{a}}) \tag{5}$$

Equations (3) and (5) describe the SIC and $\mathrm{dof}_{\mathrm{S}}$ in terms of the scaled analysis error variances; therefore, information inferred from one measure can be directly related to the other.

2.3. Idealized data set

To evaluate information content quantitatively under different treatments of error correlations, we investigate a scalar quantity on an idealized data set. Consider observations on a regular flat $n \times n$ grid, with a 200 km 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 $B(i, j) = \exp(-r_{ij}^2/2L^2)$, where r_{ij} is the Euclidean distance between points i and j and $L = 190$ is the length scale.

The test error covariance matrix $\mathbf{R}_{\mathrm{t}}$ is calculated using empirically derived error variances [2], and isotropic correlations described by $C(i, j) = (1 + r_{ij}/L)\exp(-r_{ij}/L)$. This produces a correlation matrix $\mathbf{C}$, with components $C(i, j)$, which is used to describe $\mathbf{R}_{\mathrm{t}}$: $\mathbf{R}_{\mathrm{t}} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2}$, where $\mathbf{D}$ is the diagonal matrix of error variances. The variances are obtained from the analysis of pairs of collocations between atmospheric motion vectors 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. The analysis error covariance matrix is $\mathbf{S}_{\mathrm{a}}^{(1)}$ for A1 and $\mathbf{S}_{\mathrm{a}}^{(2)}$ for A2–A4.

A1. Use the test error covariance matrix $\mathbf{R}_{\mathrm{t}}$.
A2. Set the correlations to zero in $\mathbf{R}_{\mathrm{t}}$ to obtain $\mathbf{R}_{\mathrm{f}}$.
A3. Set the correlations to zero in $\mathbf{R}_{\mathrm{t}}$ and inflate the error variances to obtain $\mathbf{R}_{\mathrm{f}}$.
A4. Approximate $\mathbf{R}_{\mathrm{t}}$ by a truncated eigendecomposition [7]

$$\mathbf{R}_{\mathrm{f}} = \mathbf{D}^{1/2}\left(\alpha\mathbf{I} + \sum_{m=1}^{M}(\lambda_m - \alpha)\mathbf{v}_m\mathbf{v}_m^{\mathrm{T}}\right)\mathbf{D}^{1/2} = \mathbf{D}^{1/2}\tilde{\mathbf{C}}\mathbf{D}^{1/2} \tag{6}$$

Here $(\lambda_m, \mathbf{v}_m)$ is an (eigenvalue, eigenvector) pair of $\mathbf{C}$, M is the number of leading eigenpairs used in the approximation, and α is chosen such that $\mathrm{trace}(\mathbf{R}_{\mathrm{f}}) = \mathrm{trace}(\mathbf{D})$, i.e. so that there is no misapproximation of the total error variance.

3. RESULTS

Studying the results obtained using a diagonal approximation of $\mathbf{R}_{\mathrm{t}}$ (A2 and A3), we conclude that if we neglect observation error correlations, information is lost. As scaled representations of the total analysis error variance, both the SIC and $\mathrm{dof}_{\mathrm{S}}$ are directly proportional to the number of

DOI: 10.1002/fld

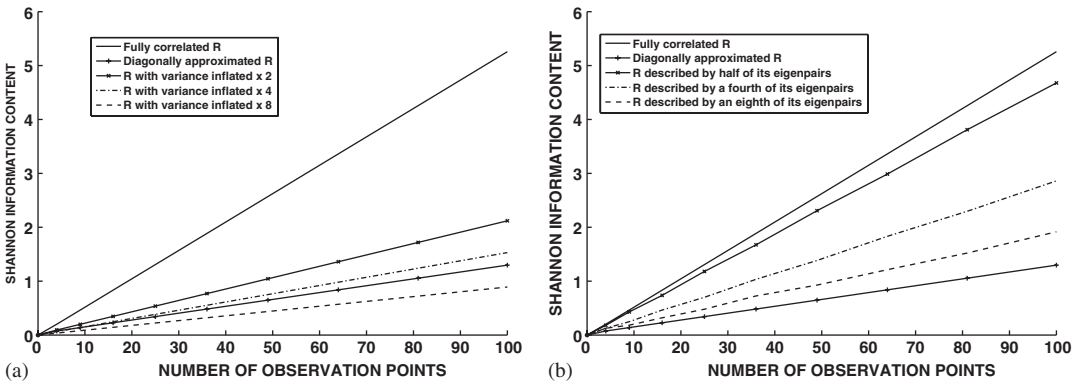

Figure 1. (a) The SIC for different scales of variance enlargement and (b) the SIC for a correlated ($\mathbf{R}_t$), uncorrelated and eigenpair described $\mathbf{R}$.

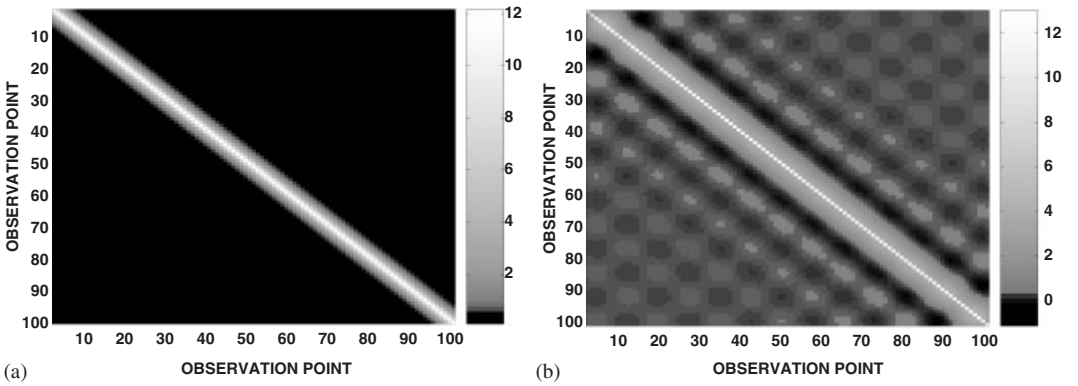

Figure 2. Spatial structure of correlations for (a) $\mathbf{R}_t$ and (b) $\mathbf{R}_t$ described by its leading 10 eigenpairs.

observation points; hence, increasing our grid size provides access to more information. However, as the number of observation points increases, the greater the difference in information between $\mathbf{R}_t$ and the diagonal approximation used in A2. For a grid with 100 points, both the SIC (Figure 1) and dof_S (not shown) decrease by 75% between $\mathbf{R}_t$ and the diagonal approximation. Further results for dof_S can be found in [9].

A depletion in information also occurs when a rescaled diagonal approximation, as in A3, is made; the scale of variance enlargement determines how detrimental the diagonal assumption is. It is observed that for a diagonal approximation with a two or four times variance enlargement, more information is retained than under the standard diagonal approximation, but less for an eight times variance enlargement (Figure 1(a)). This supports Collard's conclusions that we are limited to a variance enlargement of between 2 and 4 times [3].

In A4 we acknowledge error correlations by forming a truncated eigendecomposition of $\mathbf{R}_t$. Results show that the more eigenpairs used in the decomposition, the smaller the difference in information between $\mathbf{R}_t$ and the eigenpair approximation. The eigenpair approximation retains a higher percentage of the information available than the diagonal approximations if more than a

DOI: 10.1002/fld

quarter of the eigenpairs are used. Describing $\mathbf{R}_t$ by an eighth of its eigenpairs and by a diagonal approximation with a four times variance enlargement produce similar values of SIC (Figure 1). When $\mathbf{R}_t$ is described by half of its eigenpairs, SIC only decreases by 11% compared with 60% under A3 with a two times variance enlargement (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 [7]. 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 of the observation error variances, as in A2, is overly detrimental to the information content, but a thoughtfully chosen rescaled diagonal approximation can retain more of the information. An approach in which $\mathbf{R}_t$ is approximated by a truncated eigendecomposition retains much of the information available if a sufficient number of eigenpairs are used. But, addressing Fisher's concerns [7], 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 relatively 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.

Calculations of information content are additionally dependent on both $\mathbf{B}$ and the idealized observation operator $\mathbf{H}$. In future work, different approaches to incorporating correlation structures in the observation error covariance matrix $\mathbf{R}$ will be conducted under a more realistic specification of the two. The operational feasibility of including correlated observation errors in the data assimilation algorithm must also be addressed.

ACKNOWLEDGEMENTS

The authors would like to thank John Eyre for his involvement and suggestions. This work is partly funded by the NERC and the Met Office.

REFERENCES

1. Seaman R. Absolute and differential accuracy of analyses achievable with specified observation network characteristics. *Monthly Weather Review* 1977; **105**:1211–1222.
2. Bormann N, Saarinen S, Kelly G, Thepaut J-N. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Monthly Weather Review* 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*, *AC/90*, 2004.
4. Berger H, Forsythe M. *Satellite Wind Superobbing*. Forecasting Research Technical Report, vol. 451. Met Office, 2004.
5. Liu Z-Q, Rabier F. The interaction between model resolution, observation resolution and observation density in data assimilation. *Quarterly Journal of the Royal Astronomical Society* 2002; **128**:1367–1386.
6. Healy S, White A. Use of discrete Fourier transforms in the 1D-Var retrieval problem. *Quarterly Journal of the Royal Astronomical Society* 2005; **131**:63–72.
7. Fisher M. Accounting for correlated observation error in the ECMWF analysis. *ECMWF Technical Memoranda*, *MF/05106*, 2005.

DOI: 10.1002/fld

8. Rodgers C. *Inverse Methods of Atmospheric Sounding*, *Theory and Practice*. Atmospheric, Oceanic and Planetary Physics, vol. 2. World Scientific: Singapore, 2000.
9. Stewart L, Dance S, Nichols N. Correlated observation errors in data assimilation. *Numerical Analysis Report 3/05*, University of Reading, 2005. Available at: http://www.maths.rdg.ac.uk/research/publications/Na_reports.asp.
10. Fisher M. Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. *ECMWF Technical Memoranda*, *397*, 2003.

DOI: 10.1002/fld