Tellus (1997), 49A, 54–73
Printed in UK – all rights reserved

Copyright © Munksgaard, 1997
TELLUS
ISSN 0280–6495

Discretization error and signal/error correlation in atmospheric data assimilation

(II). The effect of unresolved scales

By HERSCHEL L. MITCHELL* and ROGER DALEY[1], *Direction de la Recherche en Météorologie, Atmospheric Environment Service, 2121 route Transcanadienne, 5e étage, Dorval, Québec, Canada H9P 1J3*

(Manuscript received 28 August 1995; in final form 5 July 1996)

ABSTRACT

This paper is the 2nd of a 2-part study of the effects of forward interpolation error and model discretization error on atmospheric data assimilation. Utilizing the generalized Kalman filter developed in part (I), that investigation is here extended to the case where not all scales in the signal are resolvable by the analysis mesh (although they may all be sampled by the observation network). The error in the unresolved scales is thus able to affect the forward interpolation and model discretization errors and the signal/error correlations considered in part (I). In some circumstances, the results are found to be qualitatively similar to those observed when all scales were resolved, but at a much higher error level. This strengthens the corresponding results of part (I) by extending them to the situation where there are unresolved scales. The most important findings of this type are the following. (a) Models, however imperfect, reduce analysis error. (b) The neglect of signal/error correlation can result in a serious underestimate of analysis error. (c) While model truncation error has a blue spectrum, proper inclusion of signal/error correlation has the effect of reddening the spectrum. Other important findings of this paper are the following. (i) The error in the unresolved scales cannot be eliminated by data assimilation. Moreover, the existence of unresolved scales not only increases the magnitude of the forward interpolation and model discretization errors due to errors in these scales, but also produces error in the resolved scales. (ii) While simple models can provide useful information about the mechanisms of analysis error growth and decay, the coupling of these models with very special observation networks (e.g., coincident with the analysis mesh) *may* yield results which are misleading with respect to more realistic models and networks. (iii) For imperfect models, both the magnitude of the analysis error and the forecast-error correlation structure are much more sensitive to model resolution than to observation network density.

1. Introduction

For observations to be assimilated optimally, the effects of unresolved scales on both the observation error and model error must be taken into account. With regard to the observation error, this is often done by adding an error-of-representativeness component (Daley, 1991) to the observation error. Lorenc (1986) has identified this component with forward interpolation error. The effect of unresolved scales on model error, and in particular on model discretization error, has not been considered.

In the first part of this study (Mitchell and Daley, 1997; hereinafter referred to as Part I), the effects on data assimilation of forward interpolation error and model discretization error were examined in the context of the one-dimensional

* Corresponding author.
[1] Current affiliation: Naval Research Laboratory, Monterey, California, USA.

linear advection equation. Explicit expressions, previously obtained by Daley (1993) and Cohn and Dee (1988), for the forward interpolation and model discretization errors respectively, were utilized. In Part I, it was assumed that the analysis mesh could resolve all scales in the signal. This simplifying assumption is dispensed with in the present paper; more realistic analysis meshes, unable to resolve all scales in the signal, are considered.

The set of second-moment equations derived in Part I applies also when there are unresolved scales. Aside from the relatively simple instrument error covariance, this set of equations is forced by the forward interpolation and model discretization error covariances. Both of these are functions of the signal. As discussed in Part I, this gives rise to cross-terms (involving these errors and signal/error correlations) not present in the standard Kalman filter (KF) equations.

Both the forward interpolation and model discretization errors are also functions of a (Fourier) transform from spectral space to the analysis mesh and therefore depend on the resolution of the analysis mesh. Consequently, the forward interpolation and model discretization error covariances and the cross-terms in the second-moment equations are affected when the analysis mesh cannot resolve all scales in the signal. In this paper, we examine the effect of these unresolved scales. The organization of the paper parallels that of Part I.

2. Theory

The task here is to generalize the theory developed in Section 2 of Part I to the case when the analysis mesh cannot resolve all scales in the signal. We consider the same one-dimensional periodic domain, the same Fourier series representation of the signal (truncated at wavenumber K), and the same uniform analysis mesh, as described by (2.1)–(2.3) of Part I. (Hereafter we will adopt the convention of referring to equations in Part I using the prefix "I", so that these equations, for example, will be referred to as I.2.1–I.2.3.) The following discussion will deal only with those aspects of Section 2 (Part I) that must be generalized.

As in Part I, we take J, the number of analysis meshpoints, to be odd, but it is no longer assumed that J equals $2K+1$. Since the analysis mesh cannot necessarily resolve all scales in the signal, it is assumed only that $J \leqslant 2K+1$.

In numerical models there is a duality in the definition of gridpoint values of the dependent variables. Thus in the dynamic portions of the model, each gridpoint value is usually (but not always) assumed to be the actual point value of the variable at the gridpoint. On the other hand, in model physical parametrizations, gridpoint values are often assumed to represent some sort of spatial average over the box surrounding the point. Frequently, this duality is not spelled out explicitly. For the present study, ambiguity in the definition of gridpoint values leads to ambiguity in the experimental results. For this reason, we here define the gridpoint values precisely and unambiguously. We assume that *the gridpoint values represent all those, and only those, scales that can be explicitly resolved by the uniform mesh.* That is, a grid with J points resolves the J spectral components with largest scale. Thus, define

$$\bar{h}(x_j, t_n) = \frac{1}{2}\hat{h}_0^c(t_n)c_0^j + \sum_{k=1}^{K}[\hat{h}_k^c(t_n)c_k^j + \hat{h}_k^s(t_n)s_k^j], \tag{2.1}$$

where

$$\begin{aligned} &c_k^j = \cos(kx_j/a), && 0 \leqslant k \leqslant (J-1)/2, \\ &c_k^j = 0 && \text{otherwise}, \\ &s_k^j = \sin(kx_j/a), && 1 \leqslant k \leqslant (J-1)/2, \\ &s_k^j = 0 && \text{otherwise}. \end{aligned} \tag{2.2}$$

Thus, $\bar{h}(x_j, t_n)$ is a representation of the variable h at the gridpoints x_j in which all Fourier components with wavenumbers smaller than $(J-1)/2$ have been filtered. The overbar can be thought of as indicating the spatial averaging (Fourier filtering). Formally we can write (2.2) in the same form as (I.2.4), i.e., $\bar{\boldsymbol{h}}_n = \boldsymbol{F}\hat{\boldsymbol{h}}_n$, where $\hat{\boldsymbol{h}}_n$ is defined in (I.2.2), $\boldsymbol{F}$ is the $J \times (2K+1)$ Fourier transform matrix with elements given in (2.2) above and $\bar{\boldsymbol{h}}_n$ is the vector of length J of Fourier filtered values at the gridpoints x_j. Unlike the situation in Part I, here the elements of $\boldsymbol{F}$ for $k > (J-1)/2$ are equal to zero. Definition of gridpoint values other than (2.1)–(2.2) are possible; some consequences of this particular choice are discussed at the end of this section.

As in Part I, the signal is assumed to satisfy the

one-dimensional linear advection equation, (I.2.5), and we consider data assimilation with an assimilating model, $\boldsymbol{M}$. The model $\boldsymbol{M}$ is a $J \times J$ matrix (as in I.2.10) which approximates the actual advection of the signal. Thus, interpreting $\bar{\boldsymbol{h}}_n^a$ and $\bar{\boldsymbol{h}}_n^f$ as the *Fourier filtered* analyzed and forecast values of h on the grid x_j at time t_n, eqs. (I.2.5)–(I.2.13) need no modification to permit unresolved scales.

As in Part I, we consider a time-independent observation network x_i, $1 \leqslant i \leqslant I$, which is assumed to consist of in situ *point* observations, e.g., radiosondes. *Point* observations include all wavenumbers $\leqslant K$. For this reason $\boldsymbol{G}$, the $I \times (2K+1)$ Fourier transform matrix between the Fourier coefficients and the observation locations (defined in I.2.15), is a *point* Fourier transform matrix, defined as a sum over all K wavenumbers using (I.2.1) for an observation location x_i. $\boldsymbol{G}$ is *not* a Fourier filtered expression as in (2.1)–(2.2). As in Part I, we assume that the instrument error is stationary, unbiased and neither spatially nor serially correlated. Thus, (I.2.14)–(I.2.18) also remain unaltered when there are unresolved scales.

We consider the three choices for the gain matrix $\boldsymbol{K}_n$ given by eqs. (I.2.19)–(I.2.21). It should be noted that the first of these, the generalized KF gain matrix, *minimizes the analysis error in the resolved scales only* ($k \leqslant (J-1)/2$).

As in the case where all scales were resolved, (I.2.10) and (I.2.14) can be integrated to predict the (Fourier filtered) values of $\boldsymbol{h}$ at the gridpoints. The necessary analysis weights are obtained from (I.2.19), (I.2.20) or (I.2.21), which in general requires $\bar{\boldsymbol{P}}_n^f$ and $\boldsymbol{Y}_n$. These are obtained by integrating the second-moment equations (I.2.12), (I.2.13), (I.2.17) and (I.2.18), which also yield $\bar{\boldsymbol{P}}_n^a$, the analysis error covariance on the grid of the assimilating model. We would like to examine $\bar{\boldsymbol{P}}_n^a$ and $\bar{\boldsymbol{P}}_n^f$ to determine the effect of a given model resolution on the assimilation process.

There is a difficulty, however: $\bar{\boldsymbol{P}}_n^f$ and $\bar{\boldsymbol{P}}_n^a$ are the error covariances corresponding to the scales resolved by the (possibly very coarse) grid of the assimilating model. If the grid is very coarse, then it will only be able to resolve the largest spatial scales of the signal or any error in the signal. Scales too small to be resolved by the mesh will not make any contribution to the errors $\bar{\boldsymbol{P}}_n^f$ and $\bar{\boldsymbol{P}}_n^a$. Thus, it is also necessary to consider the errors in the unresolved scales. To do this we introduce another Fourier transform, which is a generalization of that defined by (I.2.22) and (I.2.23). Consider an arbitrary vector $\bar{\boldsymbol{z}}_n$ of elements $\bar{z}(x_j, t_n)$, $1 \leqslant j \leqslant J$. Consistent with (2.1), we define a discrete Fourier transform by

$$\hat{z}_k^c(t) = \frac{2}{J}\sum_{j=1}^{J} \bar{z}(x_j, t) c_k^j,$$
$$\hat{z}_k^s(t) = \frac{2}{J}\sum_{j=1}^{J} \bar{z}(x_j, t) s_k^j, \tag{2.3}$$

where c_k^j and s_k^j are defined in (2.2). The first relation is valid for $k \geqslant 0$ and the second for $k \geqslant 1$. Note that for $k > (J-1)/2$, $\hat{z}_k^c(t) = \hat{z}_k^s(t) = 0$. In matrix form, we can write (2.3) as,

$$\hat{\boldsymbol{z}}_n = \tilde{\boldsymbol{F}}\bar{\boldsymbol{z}}_n, \tag{2.4}$$

where, following (I.2.23), $\hat{\boldsymbol{z}}_n^T = [\hat{z}_0^c(t_n)\ \hat{z}_1^c(t_n)\ \hat{z}_1^s(t_n) \ldots\ \hat{z}_K^s(t_n)]$ and $\tilde{\boldsymbol{F}}$ is the $(2K+1) \times J$ Fourier transform matrix from the grid x_j to Fourier coefficients whose elements are given in (2.3). The last $(2K+1-J)$ rows of $\tilde{\boldsymbol{F}}$ are all zero and no longer is $\tilde{\boldsymbol{F}}$ the matrix inverse of $\boldsymbol{F}$, as it was in Part I. Instead, $\tilde{\boldsymbol{F}}$ is a generalized inverse of $\boldsymbol{F}$ as discussed in Appendix A.

Now, left multiply the analysis error vector $\bar{\boldsymbol{e}}_n^a$ ($= \bar{\boldsymbol{h}}_n^a - \boldsymbol{F}\hat{\boldsymbol{h}}_n$) by $\tilde{\boldsymbol{F}}$ to obtain,

$$\tilde{\boldsymbol{F}}\bar{\boldsymbol{h}}_n^a = \tilde{\boldsymbol{F}}\bar{\boldsymbol{e}}_n^a + \tilde{\boldsymbol{F}}\boldsymbol{F}\hat{\boldsymbol{h}}_n. \tag{2.5}$$

We now introduce $\hat{\boldsymbol{e}}_n^a = \tilde{\boldsymbol{F}}\bar{\boldsymbol{h}}_n^a - \hat{\boldsymbol{h}}_n$, which is the vector of length $2K+1$ of spectral components of the analysis error. Substitution from (2.5) gives,

$$\hat{\boldsymbol{e}}_n^a = \tilde{\boldsymbol{F}}\bar{\boldsymbol{e}}_n^a + [\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]\hat{\boldsymbol{h}}_n. \tag{2.6}$$

While the first right-hand-side term represents the analysis error in the scales resolved by the assimilation model, the second term represents the analysis error in the unresolved scales. Right multiplication of (2.6) by its transpose and application of the expectation operator, yields,

$$\hat{\boldsymbol{P}}_n^a = \tilde{\boldsymbol{F}}\bar{\boldsymbol{P}}_n^a\tilde{\boldsymbol{F}}^T + [\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]\hat{\boldsymbol{S}}[\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]^T + \tilde{\boldsymbol{F}}\boldsymbol{X}_n[\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]^T + [\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]\boldsymbol{X}_n^T\tilde{\boldsymbol{F}}^T, \tag{2.7}$$

where $\hat{\boldsymbol{P}}_n^a = \langle \hat{\boldsymbol{e}}_n^a (\hat{\boldsymbol{e}}_n^a)^T \rangle$ is the spectral analysis error covariance matrix and $\boldsymbol{X}_n = \langle \bar{\boldsymbol{e}}_n^a \hat{\boldsymbol{h}}_n^T \rangle$ is the covariance between the gridpoint values of the analysis error and the spectral form of the signal. Expressions corresponding to (2.6) and (2.7) for the forecast error are,

$$\hat{\boldsymbol{e}}_n^f = \tilde{\boldsymbol{F}}\bar{\boldsymbol{e}}_n^f + [\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]\hat{\boldsymbol{h}}_n, \tag{2.8}$$

$$\hat{\boldsymbol{P}}_n^f = \tilde{\boldsymbol{F}}\bar{\boldsymbol{P}}_n^f\tilde{\boldsymbol{F}}^T + [\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]\hat{\boldsymbol{S}}[\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]^T + \tilde{\boldsymbol{F}}\boldsymbol{Y}_n[\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]^T + [\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]\boldsymbol{Y}_n^T\tilde{\boldsymbol{F}}^T, \tag{2.9}$$

where $\hat{\boldsymbol{P}}_n^{\mathrm{f}} = \langle \hat{\boldsymbol{e}}_n^{\mathrm{f}} (\hat{\boldsymbol{e}}_n^{\mathrm{f}})^{\mathrm{T}} \rangle$, $\hat{\boldsymbol{e}}_n^{\mathrm{f}} = \tilde{\boldsymbol{F}} \bar{\boldsymbol{h}}_n^{\mathrm{f}} - \hat{\boldsymbol{h}}_n$ is the (spectral) forecast error and $\boldsymbol{Y}_n = \langle \bar{\boldsymbol{e}}_n^{\mathrm{f}} \hat{\boldsymbol{h}}_n^{\mathrm{T}} \rangle$.

Eqs. (2.6)–(2.9) include the error due to unresolved scales. When $J = 2K + 1$, then $\tilde{\boldsymbol{F}}\boldsymbol{F} = \boldsymbol{I}$, and therefore they reduce to (I.2.24)–(I.2.25). $\hat{\boldsymbol{P}}_n^{\mathrm{f}}$ and $\hat{\boldsymbol{P}}_n^{\mathrm{a}}$ can be transformed to spatial form on any grid as described in Appendix A of Part I. (Note that the Fourier transform matrix between the Fourier coefficients and the high-resolution output mesh, is a *point* Fourier transform matrix, like $\boldsymbol{G}$, and *not* a Fourier filtered expression as in (2.1)–(2.2).) The resulting error will be referred to as the total error, in contrast to $\bar{\boldsymbol{P}}_n^{\mathrm{f}}$ and $\bar{\boldsymbol{P}}_n^{\mathrm{a}}$ which contain error in the resolved scales only.

$\hat{\boldsymbol{P}}_n^{\mathrm{f}}$ and $\hat{\boldsymbol{P}}_n^{\mathrm{a}}$ are $(2K+1) \times (2K+1)$ matrices, which are, in general, full. Consistent with (2.6) and (2.8), each of these spectral covariance matrices can be considered as a 2×2 block matrix, i.e., letting $\hat{\boldsymbol{P}}$ denote $\hat{\boldsymbol{P}}_n^{\mathrm{f}}$ or $\hat{\boldsymbol{P}}_n^{\mathrm{a}}$, we can write

$$\hat{\boldsymbol{P}} = \begin{bmatrix} \hat{\boldsymbol{P}}_{\mathrm{rr}} & \hat{\boldsymbol{P}}_{\mathrm{ru}} \\ \hat{\boldsymbol{P}}_{\mathrm{ur}} & \hat{\boldsymbol{P}}_{\mathrm{uu}} \end{bmatrix}, \tag{2.10}$$

where $\hat{\boldsymbol{P}}_{\mathrm{rr}}$ the $J \times J$ upper diagonal block represents the spectral covariances in the resolved part of the spectrum, $\hat{\boldsymbol{P}}_{\mathrm{uu}}$ the $(2K+1-J) \times (2K+1-J)$ lower diagonal block represents the covariances in the unresolved part of the spectrum, and the off-diagonal blocks $\hat{\boldsymbol{P}}_{\mathrm{ru}}$ and $\hat{\boldsymbol{P}}_{\mathrm{ur}}$ represent interactions between these two parts of the spectrum. (Note that the "r" in the subscript denotes "resolved" and should not be confused with the instrument error introduced in (I.2.15).)

$\hat{\boldsymbol{S}}$ and other spectral covariance matrices can be partitioned following (2.10). The rectangular matrices $\boldsymbol{F}$, $\tilde{\boldsymbol{F}}$ and $\boldsymbol{G}$ can also be partitioned in a similar way. Thus, for example, the $J \times (2K+1)$ matrix $\boldsymbol{F}$ and the $(2K+1) \times J$ matrix $\tilde{\boldsymbol{F}}$ can be written

$$\boldsymbol{F} = [\boldsymbol{F}_{\mathrm{rr}} \;\; \boldsymbol{F}_{\mathrm{ru}}] \qquad \text{and} \qquad \tilde{\boldsymbol{F}} = \begin{bmatrix} \tilde{\boldsymbol{F}}_{\mathrm{rr}} \\ \tilde{\boldsymbol{F}}_{\mathrm{ur}} \end{bmatrix}, \tag{2.11}$$

where $\boldsymbol{F}_{\mathrm{rr}}$ and $\tilde{\boldsymbol{F}}_{\mathrm{rr}}$ are $J \times J$, $\boldsymbol{F}_{\mathrm{ru}}$ is $J \times (2K+1-J)$ and $\tilde{\boldsymbol{F}}_{\mathrm{ur}}$ is $(2K+1-J) \times J$. For these particular matrices, it follows from (2.2) and (2.3) that $\boldsymbol{F}_{\mathrm{ru}} = 0$ and $\tilde{\boldsymbol{F}}_{\mathrm{ur}} = 0$. Partitioning the various matrices in the second-moment equations in this way allows an examination of the interaction between resolved and unresolved scales. This is pursued for (I.2.17) and (I.2.18) in Subsection 4.1 and for (I.2.12) and (I.2.13) in Subsection 5.1.

Before continuing, it is appropriate to say a few more words about the transforms $\boldsymbol{F}$ and $\tilde{\boldsymbol{F}}$. It is certainly possible to define the gridpoint values using the *point* Fourier transform (I.2.1) (including all wavenumbers $\leqslant K$) rather than the *Fourier filtered* transform of (2.2). In this case there would be no requirement for eqs. (2.3)–(2.9). However, this simpler approach has at least two problems, which led to our choice of the Fourier-filtered transforms.

Firstly, with the present Fourier-filtered definition of the gridpoint values, the forward interpolation error $[\boldsymbol{G} - \boldsymbol{H}\boldsymbol{F}]\hat{\boldsymbol{h}}_n$ of (I.2.16) is a weak function of the spatial co-ordinate x. However, if point Fourier transforms had been used instead, the forward interpolation error would have been a function of the exact spatial relationship between analysis grid and observation stations. Thus, if an analysis gridpoint and an observation station were collocated, this term would be zero, while it would be large if the observation point were halfway between two gridpoints. This would lead to a very inhomogeneous forward interpolation error covariance $[\boldsymbol{G} - \boldsymbol{H}\boldsymbol{F}]\hat{\boldsymbol{S}}[\boldsymbol{G} - \boldsymbol{H}\boldsymbol{F}]^{\mathrm{T}}$.

Secondly, point Fourier transforms will indeed give estimates of the analysis and forecast variables and their errors at the gridpoints x_j of the analysis grid. These estimates include all scales $k \leqslant K$, not merely the scales resolved by the grid, which would seem to be an advantage of the point Fourier method. However, in the point Fourier transform case, there only exist estimates at the gridpoints, there are no estimates (and no way of obtaining estimates without making further assumptions) between gridpoints. Thus, for a very coarse grid, one might obtain analyzed values and corresponding expected errors at the few scattered gridpoints, but would know nothing about the state variable anywhere else in the domain.

Eqs. (I.2.12), (I.2.13), (I.2.17) and (I.2.18), plus one of the gain matrix eqs. (I.2.19)–(I.2.21), together with (2.7) and (2.9) above constitute a complete set of second-moment equations. This set allows the effects of forward interpolation error, model discretization error and signal/error correlation to be studied, irrespective of whether or not all scales are resolved by the analysis mesh.

3. The signal and the assimilation system

The signal, the various types of observation networks and the free parameters are defined exactly as in Section 3 of Part I. That is $a = 1$, $K = 32$ and $E^r = 10$ m. The two assimilating models and the two forward interpolation matrices are defined in Part I, but since these depend on the analysis mesh, they will be affected if the analysis mesh cannot resolve all scales. As we shall see in this section, these effects stem from the fact that, when $J < 2K + 1$, $\boldsymbol{F}$ and $\tilde{\boldsymbol{F}}$ are no longer square matrices which are inverses of each other. Unless indicated to the contrary in this section, all of the properties of the assimilating models and the forward interpolation matrices discussed in Section 3 of Part I remain valid despite the presence of unresolved scales.

As before, $\boldsymbol{M}_1$, the exact Fourier model, is defined by (I.3.3) but now with the Fourier truncated definitions of $\boldsymbol{F}$ and $\tilde{\boldsymbol{F}}$ (i.e., (2.1)–(2.2) and (2.3)–(2.4), respectively). Thus $\boldsymbol{M}_1$ is now truncated at wavenumber $(J-1)/2$. $\boldsymbol{M}_2$, the cubic spline interpolation model, is defined as in Part I and consists of interpolating the values of $h(x - U\Delta t)$ from the values of h at the gridpoints of the mesh, as in a semi-Lagrangian time marching algorithm. As before, $\boldsymbol{M}_2$ can be expressed by (I.3.4), i.e., $\boldsymbol{M}_2 = \boldsymbol{F}\hat{\boldsymbol{M}}\hat{\boldsymbol{D}}\tilde{\boldsymbol{F}} = \boldsymbol{F}\hat{\boldsymbol{D}}\hat{\boldsymbol{M}}\tilde{\boldsymbol{F}}$. Partitioning the various matrices, as in (2.10) and (2.11), yields the alternate expressions

$$\boldsymbol{M}_2 = \boldsymbol{F}_{rr}\hat{\boldsymbol{M}}_{rr}\hat{\boldsymbol{D}}_{rr}\tilde{\boldsymbol{F}}_{rr} = \boldsymbol{F}_{rr}\hat{\boldsymbol{D}}_{rr}\hat{\boldsymbol{M}}_{rr}\tilde{\boldsymbol{F}}_{rr}. \tag{3.1}$$

Note that multiplication by $\hat{\boldsymbol{D}}_{rr}$, the $J \times J$ block diagonal matrix with upper left-hand corner element equal to 1 and remaining 2×2 blocks given by (B.2) of Part I, represents the effect of the spurious damping and phase shift due to cubic spline interpolation. The other blocks of the $(2K+1) \times (2K+1)$ matrix $\hat{\boldsymbol{D}}$, i.e., $\hat{\boldsymbol{D}}_{ru}$, $\hat{\boldsymbol{D}}_{ur}$ and $\hat{\boldsymbol{D}}_{uu}$, have no effect on the calculations and can be set to zero.

Proceeding in a similar way, we find that the $J \times (2K+1)$ matrix

$$\begin{aligned}\boldsymbol{MF} - \boldsymbol{F}\hat{\boldsymbol{M}} &= [\boldsymbol{F}_{rr}\hat{\boldsymbol{M}}_{rr}(\hat{\boldsymbol{D}} - \boldsymbol{I})_{rr} \quad 0] \\ &= [\boldsymbol{MF}_{rr} - \boldsymbol{F}_{rr}\hat{\boldsymbol{M}}_{rr} \quad 0].\end{aligned} \tag{3.2}$$

This is valid for both $\boldsymbol{M} = \boldsymbol{M}_1$ (in which case $\hat{\boldsymbol{D}}_{rr} = \boldsymbol{I}$) and $\boldsymbol{M} = \boldsymbol{M}_2$. Thus, while model $\boldsymbol{M}_1$ is perfect in the resolved scales, model $\boldsymbol{M}_2$ is not.

Again the model Courant number is defined as $C_M = U\Delta tJ/2\pi a$. Since $J \leqslant 2K + 1$, $C_M \leqslant C_s$ (the signal Courant number). When C_M has been chosen to be an integer, $\hat{\boldsymbol{D}}_{rr} = \boldsymbol{I}$ and $\boldsymbol{M}_1 = \boldsymbol{M}_2$ takes on a particularly simple form, as in Part I. When C_M is not an integer (i.e., the point of departure is *not* a gridpoint) then the cubic spline method damps and phase shifts the spatial scales near the truncation limit, i.e., $\hat{\boldsymbol{D}}_{rr} \neq \boldsymbol{I}$, as discussed in Appendix B of Part I. As in the case $J = 2K + 1$, this differs from the Fourier method in which there is no damping for any Courant number. Therefore, for $\boldsymbol{M}_1$, there is no model discretization error and $\boldsymbol{Q}$, the $J \times J$ model-error covariance matrix defined in (I.2.12), is zero. On the other hand, for $\boldsymbol{M}_2$, there is model discretization error in general (from 3.2), and $\boldsymbol{Q}$ is non-zero.

Proceeding as in Section 3 of Part I and using (2.10)–(2.11), it can be shown that

$$\boldsymbol{Q} = \boldsymbol{F}_{rr}(\boldsymbol{I} - \hat{\boldsymbol{D}})_{rr}\hat{\boldsymbol{S}}_{rr}(\boldsymbol{I} - \hat{\boldsymbol{D}})_{rr}^{T}\boldsymbol{F}_{rr}^{T}, \tag{3.3}$$

and therefore that

$$\hat{\boldsymbol{Q}} = \tilde{\boldsymbol{F}}\boldsymbol{Q}\tilde{\boldsymbol{F}}^{T} = \begin{bmatrix} (\boldsymbol{I} - \hat{\boldsymbol{D}})_{rr}\hat{\boldsymbol{S}}_{rr}(\boldsymbol{I} - \hat{\boldsymbol{D}})_{rr}^{T} & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.4}$$

Note that $\hat{\boldsymbol{Q}}$ is diagonal (from I.3.6) and hence $\boldsymbol{Q}$ is homogeneous.

As shown in Appendix B of Part I, the term $(\boldsymbol{I} - \hat{\boldsymbol{D}})_{rr}$ becomes large near the truncation limit. In view of the exponential form of the signal variance (see Fig. 1 of Part I), reducing the truncation limit (i.e., reducing J) increases the magnitude of $\hat{\boldsymbol{Q}}$ near the truncation limit dramatically (from

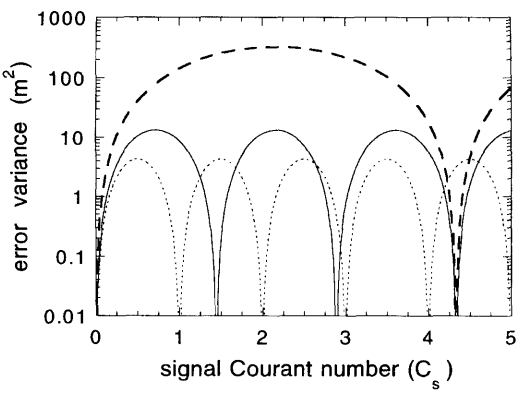

Fig. 1. The model discretization error variance for model $\boldsymbol{M}_2$ as a function of C_s, the signal Courant number, for analysis grids with $J = 65$ (dotted curve), $J = 45$ (solid curve) and $J = 15$ (dashed curve).

3.4). This produces a corresponding increase in $\boldsymbol{Q}$, as illustrated in Fig. 1, where the model discretization error variance for model $\boldsymbol{M}_2$ is shown as a function of C_s for three different resolutions. In the case $J = 65$, there are no unresolved scales, so the dotted curve corresponds to the heavy dashed curve in Fig. 6 of Part I. In all cases, $\boldsymbol{Q} = 0$ when the corresponding C_M is an integer, as noted above.

As before, $\boldsymbol{H}_1$, the exact Fourier forward interpolation matrix is defined by (I.3.8) but now with $\tilde{\boldsymbol{F}}$ defined by (2.3)–(2.4) above. Thus $\boldsymbol{H}_1$, like $\boldsymbol{M}_1$, is now truncated at wavenumber $(J-1)/2$. We note that $[\boldsymbol{G} - \boldsymbol{H}_1\boldsymbol{F}]$ is *equal to zero for the resolvable scales* (but not the unresolvable). $\boldsymbol{H}_2$, the second forward interpolation matrix, uses periodic cubic spline interpolation, as in Part I. For a uniform observation network which coincides with the analysis mesh, $\boldsymbol{H}_1 = \boldsymbol{H}_2 = \boldsymbol{I}$, so that $[\boldsymbol{G} - \boldsymbol{H}\boldsymbol{F}] = [\boldsymbol{G} - \boldsymbol{F}]$. In summary, while $\boldsymbol{H}_2$ has error in both resolved and unresolved scales, $\boldsymbol{H}_1$ has error in the unresolved scales only.

In the case of a uniform observation network with $I = J$, eqs. (I.3.9)–(I.3.12) and the related discussion in Part I remain valid, except that $\hat{\boldsymbol{R}}^i$, the spectral form of $\boldsymbol{R}^i$, now has upper left-hand corner element equal to $4(E^r)^2/J$, the following $(J-1)$ diagonal elements equal to $2(E^r)^2/J$, and all remaining elements equal to zero (from Appendix A). Note that while the integral of instrument error over all resolved scales is independent of J, the amount of error in any given wavenumber depends on J.

We are interested in the stationary statistics and proceed as in Part I by integrating the second-moment equations until stationarity is achieved. A hierarchy of cases essentially parallel to the hierarchy of Part I is considered, which allows the effect of the different errors to be examined in a controlled manner. Thus, Section 4 examines the case where there is no assimilating model and so only the forward interpolation error and error in the unresolved scales come into play, while the effect of an assimilating model is considered in Section 5.

4. No prior estimate (no model)

In this case, as in Part I, $\boldsymbol{K}_n = \boldsymbol{K} = [\boldsymbol{H}^T\boldsymbol{H}]^{-1}\boldsymbol{H}^T$, for $I \geqslant J$. Eqs. (I.4.1)–(I.4.3) and the related discussion remain valid. Thus $\bar{\boldsymbol{P}}_n^a$, $\boldsymbol{X}_n$ and (by (2.7)) $\hat{\boldsymbol{P}}_n^a$ are still time-independent, and when $I = J$ and $\boldsymbol{H} = \boldsymbol{H}_1$, the analysis passes through the observations.

As in Part I, the first case considered is that of uniform observation networks with the same number of points as the analysis mesh. This is extended to homogeneous networks in Subsection 4.2 and to the infinite observation-density limit in Subsection 4.3.

4.1. *Uniform observation network with $I = J \leqslant 2K + 1$*

The existence of unresolved scales gives rise to analysis error in *both* unresolved *and* resolved scales. As shown in Appendix B, the matrix product

$$\begin{bmatrix} \boldsymbol{I}_{rr} & \boldsymbol{G}_{rr}^{-1}\boldsymbol{G}_{ru} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \hat{\boldsymbol{S}}_{uu} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{rr} & 0 \\ \boldsymbol{G}_{ru}^T(\boldsymbol{G}_{rr}^{-1})^T & 0 \end{bmatrix}, \tag{4.1}$$

which arises in the forward interpolation error component of the analysis error, plays a key role. The effect of this matrix product is to "fold back" the elements of $\hat{\boldsymbol{S}}_{uu}$ around the Nyquist scale from the unresolved to the resolved part of the spectrum. This produces a "foldback" error component, $\hat{\boldsymbol{B}}$, *in the resolved scales* as large as the error due to unresolved scales.

$\hat{\boldsymbol{B}}$ is independent of the displacement of the observation network with respect to the analysis grid and, once K and $\hat{\boldsymbol{S}}$ are specified, depends only on I, the number of observations in the uniform network. Fig. 2 shows the spectra of $b^2(k)$ for $I = J = 45$ and 35, which clearly illustrate this "fold back" of covariance error. Also shown in Fig. 2 are the corresponding spectra for $I = J = 25$ and 15. In these latter cases, since $2K + 1 - I > I$, we see the error from more than one unresolved wave being aliased onto the same resolved wave. The error in the resolved scales due to the misrepresentation of scales which the observation network is unable to resolve (given, from (B.7), by $\hat{\boldsymbol{D}}_{rr}^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{D}}_{rr}^{-T}$) is a type of observation error commonly referred to as the "error of representativeness" (e.g., Daley, 1991).

As in Subsection 4.1 of Part I, (B.7) can be decomposed into $(J-1)/2$ sets of 2×2 matrix equations, one for each wavenumber $k \leqslant (J-1)/2$, following Daley and Ménard (1993). This yields

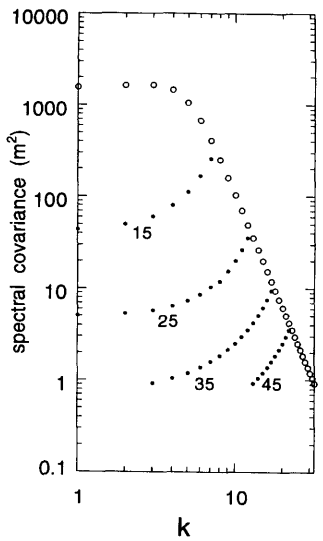

Fig. 2. The spectrum of the signal, as in Fig. 1a of Part I but in discrete form, is represented by the open circles. The spectra of the forward interpolation error covariance for $I = 15$, 25, 35 and 45 are represented by the dots.

a diagonal matrix with diagonal elements given by

$$a^2(k) = d^{-2}(k)\{r^2(k) + s^2(k) \times [1 + d^2(k) - 2d(k)\cos\phi(k)] + b^2(k)\}, \tag{4.2}$$

where for a given k, $a^2(k)$, $b^2(k)$, $r^2(k)$, and $s^2(k)$ denote the diagonal elements of $[\tilde{\boldsymbol{F}}\bar{\boldsymbol{P}}^{\mathrm{a}}\tilde{\boldsymbol{F}}^{\mathrm{T}}]_{\mathrm{rr}}$, $\hat{\boldsymbol{B}}$, $\hat{\boldsymbol{R}}^{\mathrm{i}}_{\mathrm{rr}}$, and $\hat{\boldsymbol{S}}_{\mathrm{rr}}$, respectively, and $d(k)$ and $\phi(k)$ are the damping and phase shift due to cubic spline interpolation (Appendix B, Part I). Since $[\tilde{\boldsymbol{F}}\bar{\boldsymbol{P}}^{\mathrm{a}}\tilde{\boldsymbol{F}}^{\mathrm{T}}]_{\mathrm{rr}}$, the only non-zero block of $\tilde{\boldsymbol{F}}\bar{\boldsymbol{P}}^{\mathrm{a}}\tilde{\boldsymbol{F}}^{\mathrm{T}}$, is diagonal, $\bar{\boldsymbol{P}}^{\mathrm{a}}$ is homogeneous. Like (I.4.6), (4.2) is valid for any forward interpolation operator, with the appropriately specified damping, $d(k)$, and phase shift, $\phi(k)$.

Unlike the situation in Part I, however, it follows from (2.7) that the unresolved scales give rise to further terms in addition to the homogeneous, resolved error of (4.2). The combined effect of forward interpolation error and unresolved scales will now be illustrated for a very special network.

4.1.1. Uniform observation network coincides with analysis mesh ($I = J \leqslant 2K + 1$). In this case, $\boldsymbol{F}_{\mathrm{rr}} = \boldsymbol{G}_{\mathrm{rr}}$ and, whether obtained by the Fourier procedure or the cubic spline interpolation, $\boldsymbol{H} = \boldsymbol{G}\tilde{\boldsymbol{F}} = \boldsymbol{I}$ so $\hat{\boldsymbol{D}} = \boldsymbol{I}$. It follows that $\boldsymbol{K} = \boldsymbol{H}^{-1} = \boldsymbol{I}$, i.e., the analyzed value is set equal to the observed value at each analysis point. Although there is no damping, the unresolved scales will produce an error of representativeness in the resolved scales.

Using (B.1), eqs. (I.4.1) and (I.4.2) can be written:

$$\bar{\boldsymbol{P}}^{\mathrm{a}} = \boldsymbol{G}\{\hat{\boldsymbol{R}}^{\mathrm{i}} + [\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]\hat{\boldsymbol{S}}[\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]^{\mathrm{T}}\}\boldsymbol{G}^{\mathrm{T}}, \tag{4.3}$$

$$\boldsymbol{X} = \boldsymbol{G}[\boldsymbol{I} - \tilde{\boldsymbol{F}}\boldsymbol{F}]\hat{\boldsymbol{S}}. \tag{4.4}$$

It follows then from (2.7) that

$$\hat{\boldsymbol{P}}^{\mathrm{a}} = \tilde{\boldsymbol{F}}\bar{\boldsymbol{P}}^{\mathrm{a}}\tilde{\boldsymbol{F}}^{\mathrm{T}} + [\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]\hat{\boldsymbol{S}}[\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]^{\mathrm{T}} - \tilde{\boldsymbol{F}}\boldsymbol{G}[\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]\hat{\boldsymbol{S}}[\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]^{\mathrm{T}} - [\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]\hat{\boldsymbol{S}}[\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\tilde{\boldsymbol{F}}^{\mathrm{T}}. \tag{4.5}$$

The various terms on the right-hand side of (4.5) will now be considered.

From (B.7), the sum of the first two terms on the right-hand side of (4.5) is the diagonal matrix

$$\begin{bmatrix} \hat{\boldsymbol{R}}^{\mathrm{i}}_{\mathrm{rr}} + \hat{\boldsymbol{B}} & 0 \\ 0 & \hat{\boldsymbol{S}}_{\mathrm{uu}} \end{bmatrix},$$

where $\hat{\boldsymbol{R}}^{\mathrm{i}}_{\mathrm{rr}}$, $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{S}}_{\mathrm{uu}}$ are all in spectral form and represent: the instrument error (in the resolved scales), the representativeness (foldback) error, and the unresolved portion of the signal, respectively. Since each of these are diagonal, they give rise to homogeneous analysis errors when transformed to physical space. Moreover from (B.9), since $\hat{\boldsymbol{B}}$ is a "folded back" image of the error in the unresolved scales, $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{S}}_{\mathrm{uu}}$ contribute equally to the total analysis error variance in spatial form. *Thus the error due to unresolved scales contributes twice to the homogeneous part of the analysis error.* Returning to (4.3), we see that $\bar{\boldsymbol{P}}^{\mathrm{a}}$, the analysis error in the scales resolved by the analysis mesh, consists of instrument error and error of representativeness, i.e., foldback error, only. Eq. (4.2) reduces to $a^2(k) = \{r^2 + b^2(k)\}$.

Consider the remaining terms of (4.5). Since in the present case $\boldsymbol{F}_{\mathrm{rr}} = \boldsymbol{G}_{\mathrm{rr}}$, $\tilde{\boldsymbol{F}}_{\mathrm{rr}} = \boldsymbol{G}^{-1}_{\mathrm{rr}}$, it follows from (B.4) that with the inclusion of these terms, (4.5) can be written

$$\hat{\boldsymbol{P}}^{\mathrm{a}} = \begin{bmatrix} \hat{\boldsymbol{R}}^{\mathrm{i}}_{\mathrm{rr}} + \hat{\boldsymbol{B}} & \tilde{\boldsymbol{F}}_{\mathrm{rr}}\boldsymbol{G}_{\mathrm{ru}}\hat{\boldsymbol{S}}_{\mathrm{uu}} \\ \hat{\boldsymbol{S}}_{\mathrm{uu}}\boldsymbol{G}^{\mathrm{T}}_{\mathrm{ru}}\tilde{\boldsymbol{F}}^{\mathrm{T}}_{\mathrm{rr}} & \hat{\boldsymbol{S}}_{\mathrm{uu}} \end{bmatrix}. \tag{4.6}$$

While $\hat{\boldsymbol{R}}^{\mathrm{i}}_{\mathrm{rr}}$, $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{S}}_{\mathrm{uu}}$ are diagonal, the two off-diagonal contributions give rise to inhomogeneous

error covariances when transformed to spatial form.

The contributions of the various terms of (4.3) and (4.5) are illustrated in Fig. 3 for the cases $I = J = 45$ and $I = J = 15$. The resolved analysis error variance (the dots) has values of 119 m^2 and 892 m^2 for the two cases illustrated here, 19 m^2 and 792 m^2 being the variance in the unresolved scales at any point in physical space for these values of $I = J$ (as can be roughly confirmed from the spectrum of the signal plotted in Fig. 2) and 100 m^2 being the instrument error variance. The dashed line in each panel represents the homogeneous portion of the total analysis error variance. Since $\hat{\boldsymbol{R}}^{\mathrm{i}}_{\mathrm{rr}}$, $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{S}}_{\mathrm{uu}}$ are diagonal, this can be obtained by transforming these terms to the high-resolution output grid (Appendix A, Part I). The homogeneous portion of the total analysis error variance has a value of 138 $(= 100 + 2 \times 19)$ m^2 for $I = J = 45$ and 1684 $(= 100 + 2 \times 792)$ m^2 for $I = J = 15$. The total analysis error variance on the high-resolution output grid (the solid curve) does not pass through the dots, as it did in Part I, due to the unresolved scales. The inhomogeneous terms produce an oscillation with amplitude approximately equal to twice the variance in the unresolved scales, i.e., 38 and 1584 m^2. This oscillation reduces the total analysis error variance to the instrument error variance at each point of the analysis grid, consistent with the fact that the analysis passes through the observations and these coincide with the analysis mesh. However the analysis error now reaches a maximum value equal to the instrument error plus roughly *four* times the variance in the unresolved scales midway between the points of the analysis grid.

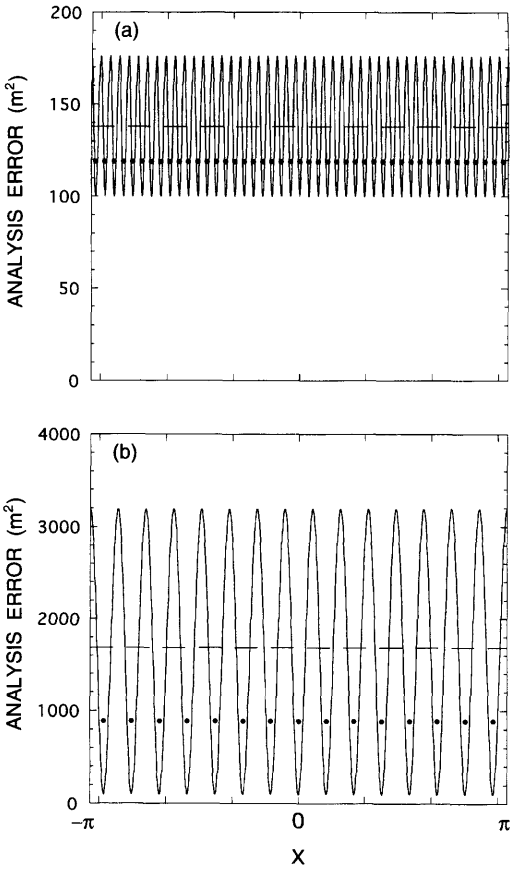

Fig. 3. The analysis error variance as a function of the spatial coordinate, x, in the case of coincident observation network and analysis mesh with 45 points (panel (a)) and 15 points (panel (b)). No model is used. The dots represent the analysis error variance resolved by the analysis grid (i.e., the diagonal elements of $\bar{\boldsymbol{P}}^{\mathrm{a}}$). The solid curve represents the total analysis error variance, (i.e., from $\hat{\boldsymbol{P}}^{\mathrm{a}}$ transformed to spatial form on a high-resolution output grid), while the dashed line is the homogeneous portion of the total analysis error variance.

The unresolved scales also produce oscillations in the analysis error/signal covariance (not shown). This has a rather simple structure for the very special network considered here. Calculated on the analysis mesh, as $\boldsymbol{F}\hat{\boldsymbol{X}}\boldsymbol{F}^{\mathrm{T}}$, this is zero (as it was with $\boldsymbol{H}_1$ in Section 4 of Part I), while on the high-resolution output mesh, it oscillates about zero.

The effect of unresolved scales together with a non-zero displacement between the observation network and analysis mesh will now be examined.

4.1.2. General case of a uniform observation network with $I = J \leqslant 2K + 1$. When $\boldsymbol{KH} = \boldsymbol{I}$, $\bar{\boldsymbol{P}}^{\mathrm{a}}$ is a sum of two terms, one due to instrument error and the other due to forward interpolation error (see (I.4.1) and the discussion following (I.2.17)). For $\boldsymbol{H}_1$, the Fourier forward interpolation, $d(k) = 1$ and $\phi(k) = 0$, so (4.2) reduces to $a^2(k) = r^2(k) + b^2(k)$, i.e., the forward interpolation error reduces to the foldback term. Fig. 4a shows the analysis error variance in the resolved scales as a function of the number of points $(I = J)$ and the displacement between the observation network and analysis grid. In this case, the contribution due to instrument error is 100 m^2 independent of

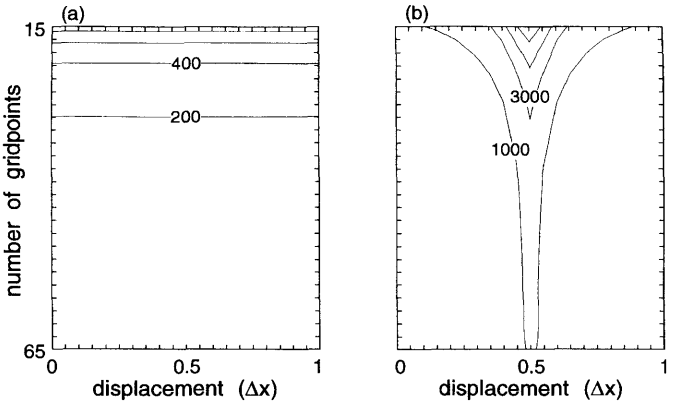

Fig. 4. The resolved analysis error variance (m^2) as a function of the number of analysis points (vertical axis) and the displacement (in gridlengths) between the uniform observation network and the analysis mesh (horizontal axis). The number of observations equals the number of analysis points. Panels (a) and (b) show the analysis error variance for $\boldsymbol{H} = \boldsymbol{H}_1$ (contour interval 200) and $\boldsymbol{H} = \boldsymbol{H}_2$ (contour interval 2000), respectively.

the number of points or the displacement, while the foldback error, independent of the displacement only, gives rise to horizontal contours. The contour gradient increases rapidly as the number of points becomes small, consistent with the spectrum of the signal.

As in Part I, the behaviour with $\boldsymbol{H}_2$, the cubic spline forward interpolation, is rather different. Both components of the analysis error variance, and therefore also their sum shown in Fig. 4b, are now strong functions of the displacement. (Note the difference in contour interval between the two panels of Fig. 4.) The maximum error occurs for a displacement of half a gridlength, due to the variation of $d(k)$ and $\phi(k)$ with displacement, as in Fig. 3 of Part I.

Fig. 4 implies that the effect of displacing the observation network of Fig. 3 with respect to the analysis grid is very different for the two forward interpolation operators. If $\boldsymbol{H} = \boldsymbol{H}_1$, only the phase of the inhomogeneous terms changes: as in Fig. 3, the minima of the total analysis error occur at the observation locations, but these are now displaced with respect to the analysis mesh. If $\boldsymbol{H} = \boldsymbol{H}_2$ there can be a substantial increase in the magnitude of the analysis error in addition to the change in phase of the inhomogeneous terms. When $\boldsymbol{H} = \boldsymbol{H}_2$, the analysis error/signal covariance (not shown) oscillates about a positive mean value, consistent with the results of Section 4, Part I.

4.2. *Homogeneous (nonuniform) observation networks*

An example of the analysis error for this type of network is presented in Fig. 5. Here $I = J = 15$ and each randomly situated observation is constrained to lie not more than one-third of a gridlength from the corresponding point of the analysis mesh. The result for $\boldsymbol{H} = \boldsymbol{H}_1$ is shown by the dashed curve and crosses. Comparison with Fig. 3b reveals the increase in the average analysis error due to the use of a nonuniform observation network. The corresponding analysis error when $\boldsymbol{H}_2$ is used in place of $\boldsymbol{H}_1$ is indicated by the solid curve and dots. As in Fig. 4 of Part I, use of $\boldsymbol{H}_2$ results in a damping of the largest oscillations.

The effects of displacing the observation network by half a gridlength, and therefore constraining the observations to lie close to the midpoints of the mesh, were consistent with Fig. 4: for $\boldsymbol{H} = \boldsymbol{H}_1$, this displacement caused a half-gridlength shift in the total analysis error; while for $\boldsymbol{H} = \boldsymbol{H}_2$, it resulted in a much larger total analysis error.

4.3. *The infinite observation-density limit*

For a fixed value of J and either forward interpolation operator, the analysis error decreases as the number of observations, I, increases. It was found (results not shown) that substantially larger

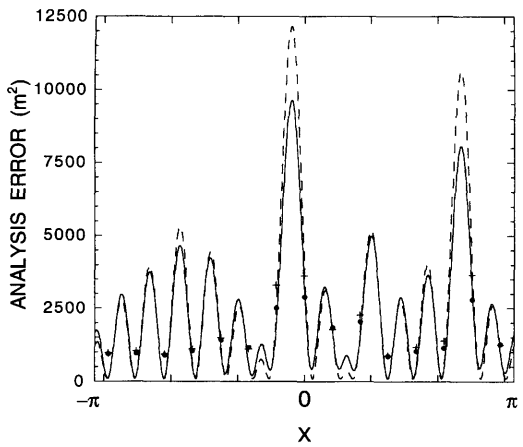

Fig. 5. The analysis error variance as a function of the spatial coordinate, x, in the case where the analysis mesh and the homogeneous observation network both have 15 points. Each randomly situated observation is constrained to lie not more than one-third of a gridlength from the corresponding point of the analysis mesh. No model is used. When $\boldsymbol{H}_2$ is used for the forward interpolation, the dots and the solid curve apply and have the same significance as in Fig. 3. When $\boldsymbol{H} = \boldsymbol{H}_1$, the corresponding quantities are given by the crosses and the dashed curve.

reductions occurred with $\boldsymbol{H}_1$ than with $\boldsymbol{H}_2$. Also with $\boldsymbol{H}_1$ the oscillation in the total analysis error had as many maxima as I and became smaller as I increased, while with $\boldsymbol{H}_2$ the oscillation was related to J and eventually became independent of I. These differences between $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ can be elucidated by letting $I \to \infty$ (Appendix C). The two limiting cases for $J = 15$ are shown in Fig. 6: for $\boldsymbol{H}_1$ (panel (a)), $\bar{\boldsymbol{P}}^{\mathrm{a}} \to 0$ and the total analysis error consists only of the error in the unresolved scales; for $\boldsymbol{H}_2$ (panel (b)), $\bar{\boldsymbol{P}}^{\mathrm{a}} \neq 0$ and the interaction between error in the resolved and unresolved scales produces a large amplitude oscillation.

We now introduce the models $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ into the assimilation system.

5. The effect of an assimilating model

As in Part I, the introduction of an assimilating model makes the situation more complex; the full set of second-moment equations ((I.2.12), (I.2.13), (I.2.17), (I.2.18)) must now be included. We examine a hierarchy of increasingly complex cases, beginning with the case where there are no observations.

5.1. *The no-observation case*

In this case $\bar{\boldsymbol{P}}^{\mathrm{a}}_n = \bar{\boldsymbol{P}}^{\mathrm{f}}_n$ and $\boldsymbol{X}_n = \boldsymbol{Y}_n$, as in Section 5.1 of Part I.

For the perfect model $\boldsymbol{M}_1$, (I.2.12)–(I.2.13) reduce to (I.5.1)–(I.5.2), i.e., $\bar{\boldsymbol{P}}^{\mathrm{a}}_{n+1} = \boldsymbol{M}\bar{\boldsymbol{P}}^{\mathrm{a}}_n\boldsymbol{M}^{\mathrm{T}}$ and $\boldsymbol{X}_{n+1} = \boldsymbol{M}\boldsymbol{X}_n\hat{\boldsymbol{M}}^{\mathrm{T}}$. The first of these equations involves resolved scales only; while partitioning the matrices in the second equation, following (2.10)–(2.11), shows that there is no interaction between resolved and unresolved scales. Thus, the resolved scales are not affected by the unresolved scales and the conclusions obtained in Part I remain unaltered, i.e., the initial covariance is simply advected around the domain and therefore remains unchanged, if initially homogeneous. Therefore, for the perfect model with no observations, the analysis error and the signal/error correlation at any time are the same as they were at initial time, apart from a possible displacement.

For the imperfect model $\boldsymbol{M}_2$ (which has discretization error) and no observations, (I.2.12)–(I.2.13) do not simplify to the same extent. In view of (3.1)–(3.3), partitioning (I.2.12) following (2.10)–(2.11) shows that this equation involves only resolved scales. On the other hand, partitioning (I.2.13) at equilibrium yields (with $\boldsymbol{X} = \boldsymbol{Y}$)

$$\boldsymbol{X}_{\mathrm{rr}} = \boldsymbol{M}\boldsymbol{X}_{\mathrm{rr}}\hat{\boldsymbol{M}}^{\mathrm{T}}_{\mathrm{rr}} + [\boldsymbol{M}\boldsymbol{F}_{\mathrm{rr}} - \boldsymbol{F}_{\mathrm{rr}}\hat{\boldsymbol{M}}_{\mathrm{rr}}]\hat{\boldsymbol{S}}_{\mathrm{rr}}\hat{\boldsymbol{M}}^{\mathrm{T}}_{\mathrm{rr}}, \qquad (5.1)$$

$$\boldsymbol{X}_{\mathrm{ru}} = \boldsymbol{M}\boldsymbol{X}_{\mathrm{ru}}\hat{\boldsymbol{M}}^{\mathrm{T}}_{\mathrm{uu}}. \qquad (5.2)$$

It is then straightforward to show that

$$\bar{\boldsymbol{P}}^{\mathrm{a}} = \boldsymbol{F}_{\mathrm{rr}}\hat{\boldsymbol{S}}_{\mathrm{rr}}\boldsymbol{F}^{\mathrm{T}}_{\mathrm{rr}} \quad \text{and} \quad \hat{\boldsymbol{X}}_{\mathrm{rr}} = \tilde{\boldsymbol{F}}_{\mathrm{rr}}\boldsymbol{X}_{\mathrm{rr}} = -\hat{\boldsymbol{S}}_{\mathrm{rr}} \qquad (5.3)$$

are solutions of (I.2.12) and (5.1), respectively. Substitution into (2.7) yields

$$\hat{\boldsymbol{P}}^{\mathrm{a}} = \begin{bmatrix} \hat{\boldsymbol{S}}_{\mathrm{rr}} & -\boldsymbol{F}_{\mathrm{rr}}\boldsymbol{X}_{\mathrm{ru}} \\ -\boldsymbol{X}^{\mathrm{T}}_{\mathrm{ru}}\boldsymbol{F}^{\mathrm{T}}_{\mathrm{rr}} & \hat{\boldsymbol{S}}_{\mathrm{uu}} \end{bmatrix}, \qquad (5.4)$$

which is a generalization of the corresponding result obtained in Part I. Note that $\boldsymbol{X}_{\mathrm{ru}}$ is not forced (from 5.2) and therefore if it is zero at any time it will remain zero. In that case $\hat{\boldsymbol{P}}^{\mathrm{a}} = \hat{\boldsymbol{S}}$, as in Part I.

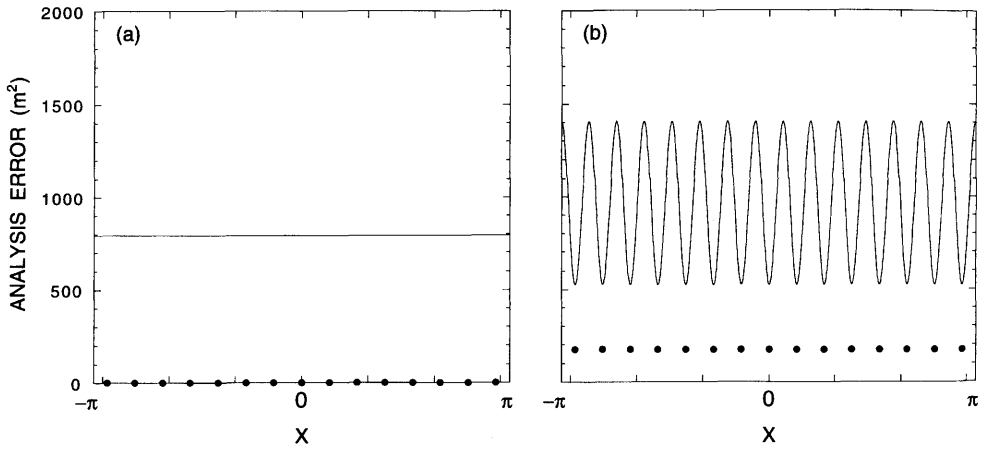

Fig. 6. The analysis error variance as a function of the spatial coordinate, x, for an analysis mesh having 15 points in the limit as the number of observations, $I \to \infty$. $\boldsymbol{H}_1$ is used for the forward interpolation in panel (a), while $\boldsymbol{H}_2$ is used in panel (b). The dots and the solid curve have the same significance as in Fig. 3b.

5.2. *Perfect model ($\boldsymbol{M} = \boldsymbol{M}_1$) with any observation network*

In this case, as discussed in Section 3 above, $[\boldsymbol{MF} - \boldsymbol{F}\hat{\boldsymbol{M}}] = 0$ as in Part I. Therefore (I.5.3)–(I.5.4) and (I.2.17)–(I.2.18) apply, despite the presence of unresolved scales. It follows that the conclusions of Section 5.2 of Part I largely apply in the present situation. In particular, $\bar{\boldsymbol{P}}^{\mathrm{f}} = \bar{\boldsymbol{P}}^{\mathrm{a}} = \boldsymbol{X} = \boldsymbol{Y} = \boldsymbol{K} = 0$ is a stationary solution, as before. Also, filter performance, i.e., its ability to reduce the analysis error substantially and quickly, is generally enhanced by: (i) use of the Fourier forward interpolation ($\boldsymbol{H}_1$) rather than the cubic spline forward interpolation ($\boldsymbol{H}_2$); (ii) use of the generalized KF gain matrix (I.2.19) rather than the conventional one (I.2.20); and (iii) use of homogeneous (non-uniform) observation networks rather than uniform ones, in particular when $I = J$.

The presence of unresolved scales does, however, give rise to differences as compared to the corresponding situation in Part I. In particular, since $[\boldsymbol{G} - \boldsymbol{H}_1\boldsymbol{F}]$ is equal to zero for the resolvable scales *only*, qualitative behaviour with $\boldsymbol{H}_1$ changes and can assume aspects previously observed only with $\boldsymbol{H}_2$. Specifically, the gain matrices defined by (I.2.19) and (I.2.20) are no longer identical for $\boldsymbol{H} = \boldsymbol{H}_1$. Also, $[\bar{\boldsymbol{P}}_n^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{Y}_n(\boldsymbol{G} - \boldsymbol{HF})^{\mathrm{T}}]$, and hence $\boldsymbol{K}_n$, can become zero, halting further reduction of the analysis error, with $\boldsymbol{H}_1$ as well as $\boldsymbol{H}_2$. In fact, this behaviour was observed for integer values of C_{M} even when the observation network coincided with the analysis mesh. In this situation there is not a unique stationary solution, which is independent of the initial conditions.

Even when $\bar{\boldsymbol{P}}^{\mathrm{a}} = \bar{\boldsymbol{P}}^{\mathrm{f}} = \boldsymbol{X} = \boldsymbol{Y} = 0$, we still have the error in the unresolved scales, as with $\boldsymbol{H}_1$ in the limit $I \to \infty$ (Fig. 6a). That is, from (2.7), $\hat{\boldsymbol{P}}^{\mathrm{a}} = \hat{\boldsymbol{P}}^{\mathrm{f}} = [\tilde{\boldsymbol{F}}\tilde{\boldsymbol{F}} - \boldsymbol{I}]\hat{\boldsymbol{S}}[\tilde{\boldsymbol{F}}\tilde{\boldsymbol{F}} - \boldsymbol{I}]^{\mathrm{T}}$, the diagonal matrix, identical to $\hat{\boldsymbol{S}}$ in the unresolved part of the spectrum and zero elsewhere. Thus, even if the error in the resolved scales is eliminated, the error in the unresolved scales remains.

The effects of model discretization error will now be considered. We first consider the case where the observation network and analysis mesh coincide.

5.3. *Coincident observation network and analysis mesh with imperfect model ($\boldsymbol{M}_2$)*

In this case $\boldsymbol{H} = \boldsymbol{I}$. Since $[\boldsymbol{G} - \boldsymbol{HF}] = [\boldsymbol{G} - \boldsymbol{F}]$ is equal to zero for the resolved scales only, (I.2.17)–(I.2.18) do not simplify as they did in Appendix C of Part I, and, with regard to the definition of the gain matrix, no longer does (I.2.19) reduce to (I.2.20). We use the generalized KF gain matrix and examine results for the low-resolution analysis mesh, $I = J = 15$.

From Fig. 1, the (homogeneous) model error is

a maximum when $C_M = 0.5$ and equals 322 m^2 in this case. The corresponding analysis error is shown in Fig. 7. (The forecast error was qualitatively similar, but about twice as large.) It can be seen that the perceived analysis error, represented by the dots, is now very much larger than the maximum error when all scales were resolved (see Fig. 6 of Part I). The total analysis error (solid curve in Fig. 7) is in the mean larger by an amount equal to the variance in the unresolved scales (792 m^2), oscillating about this value as in Sect. 4. (Use of the conventional gain matrix (I.2.20), in place of (I.2.19), produced only a 2% increase in the analysis error.) A comparison of Fig. 7 with Fig. 3b, where no model was used, shows the substantial reduction in analysis error that results from even the use of an imperfect model.

For $C_M = 0.5$ there is no phase shift due to the model M_2, and therefore at stationarity, from Appendices B and C of Part I, $\hat{X}_{rr}$ and $\hat{Y}_{rr}$, like $\hat{D}_{rr}$, become diagonal. This facilitates the examination of these quantities as a function of wavenumber, k, shown in Fig. 8. In each of Figs. 8a (analysis error) and 8b (corresponding forecast error), the heavy solid and dashed curves show the error when the effect of the signal/error correlations is included and neglected, respectively. In the unresolved scales, i.e., for $k > (J-1)/2$, both of these errors are identical to the signal, as in (4.6), while the instrument and model errors are zero. In the resolved scales, the instrument error (light dotted line) is constant, while the model error (heavy dotted line) rises sharply to reach a maximum at the truncation limit. In these scales, this behaviour of the model error is seen to be reflected both in the analysis error, panel (a), and the forecast error, panel (b). The effect of the signal/error correlations is seen to broaden the spectrum of the analysis and forecast errors, increasing the error in the longer waves, as noted in Fig. 7 of Part I where all scales were resolved. Also shown in panels (a) and (b) are the diagonal elements of $-\hat{X}$ and $-\hat{Y}$, respectively. $\hat{X}_{rr}$ and $\hat{Y}_{rr}$ are diagonal with spectra that are seen to be similar to that shown in Fig. 7 of Part I, while $\hat{X}_{uu} = \hat{Y}_{uu} = 0$. However, $\hat{X}$ and $\hat{Y}$ are not diagonal, since $\hat{X}_{ru}$ and $\hat{Y}_{ru}$ are non-zero.

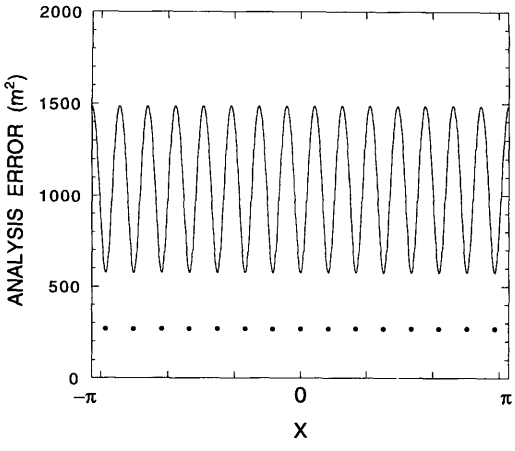

Fig. 7. As in Fig. 3b but for $C_M = 0.5$ using model M_2.

Fig. 9a shows how the analysis error in the resolved part of the spectrum (solid curve) varies with model Courant number (C_M) for the case $I = J = 15$. (Fig. 9b will be discussed in Subsection 5.5 below.) Also shown are the model error variance (see also Fig. 1), the signal/analysis error covariance and the analysis error variance when the signal/error correlations are neglected. As was the case for the corresponding figure when all scales were resolved (Fig. 6, Part I), the curves vary periodically with C_M for C_M outside of the interval $[0, 1]$ and, within this interval, are symmetric about $C_M = 0.5$.

A comparison of Fig. 9a with Fig. 6 of Part I shows that, while the error levels are substantially larger here, the curves for the model error and the analysis error when signal/error correlations are neglected are very similar in the two figures. As before, these quantities assume their maximum values when the fractional part of $C_M = 0.5$ and fall to zero when C_M is an integer. By contrast, it can be seen that the other two curves, i.e., those which apply in the general case, are changed in form when there are unresolved scales. As in Fig. 6 (Part I), the signal/error covariance is seen to be a minimum when the fractional part of $C_M = 0.5$ and it is negative there, but this covariance now changes sign, becoming positive for C_M within about 0.15 of an integer value. The analysis error variance (solid curve) no longer has a global maximum when the fractional part of $C_M = 0.5$, but instead is relatively constant over the domain and actually increases as C_M approaches an integer value. These counter-intuitive results (i.e., analysis error increasing as the model becomes more accurate) are consistent with the fact that the signal/error correlation is negative (damping)

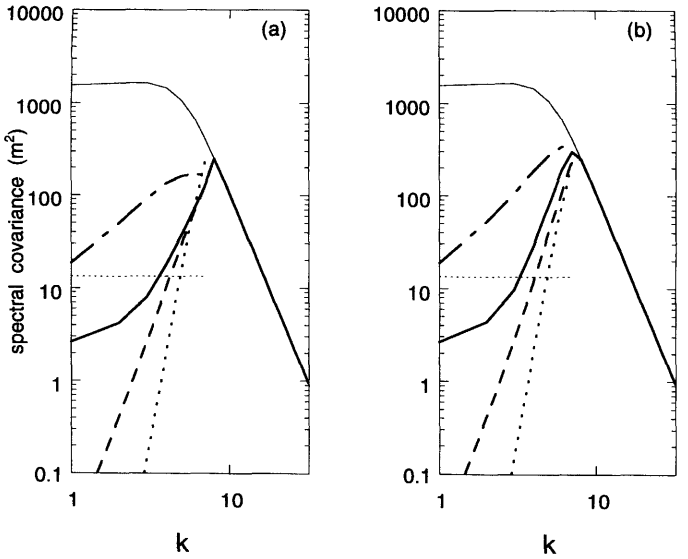

Fig. 8. The effect of model discretization error and unresolved scales as a function of wavenumber, k, for $C_M = 0.5$ and an observation network which coincides with the analysis mesh ($I = J = 15$). Panel (a) shows the analysis error variance (heavy solid curve) and $-x(k)$ (heavy dash-dotted curve). The analysis error variance when the serial correlations are neglected is shown as the heavy dashed curve. Panel (b) is the same as panel (a) but for the forecast error variance and $-y(k)$. Also shown for reference in both panels are: the model error variance (heavy dotted line), the instrument error variance (thin dotted line) and the signal (thin solid line).

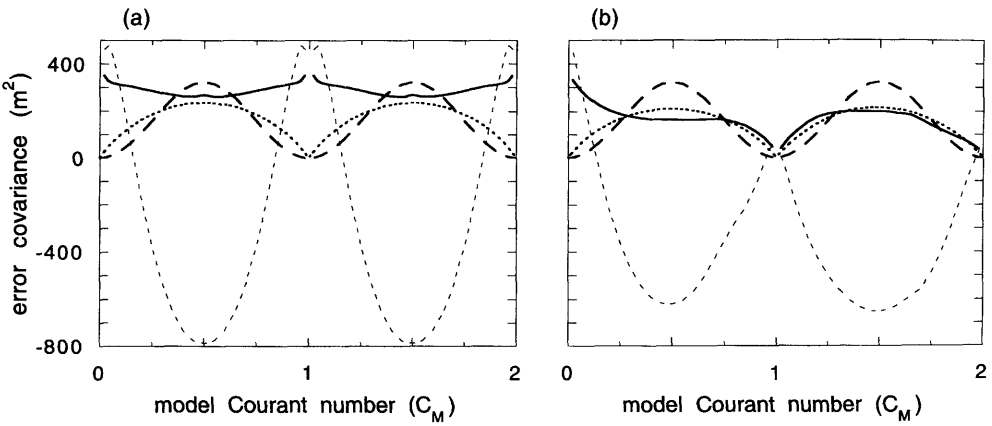

Fig. 9. Panel (a) shows the analysis (solid curve) and model (heavy dashed curve) error variance due to the discretization error of model M_2 as a function of C_M. The thin dashed curve is the signal/analysis error covariance, i.e., $\hat{X}$ transformed to spatial form. The dotted curve is the analysis error variance when signal/error correlations are neglected. In panel (a) the observation network coincides with the analysis mesh ($I = J = 15$). Panel (b) is the same as panel (a) except that the average analysis error is shown for the homogeneous observation network of Fig. 5.

when the model error is a maximum and positive (amplifying) when the model error is small. As will be discussed in Subsection 5.5, this counter-intuitive behaviour does not occur with more general observation networks. Rather, it seems to occur here because of the combination of the special observation network, the constant advecting velocity (I.2.5), and the periodic boundary conditions.

It may be noted in Fig. 9a that the signal/analysis error covariance (thin dashed curve) and the corresponding analysis error (solid curve) are not plotted for values of C_M within 0.02 of an integer value. For such values of C_M, i.e., when the model is nearly-perfect, the analysis error became larger and convergence towards a fixed point became much more problematic. Such behaviour could be expected from the results of the previous subsection.

We now turn to the minimal observation network, i.e., the case of a single observation.

5.4. *Single observation with imperfect model (M_2)*

The dots and solid curve in Fig. 10a show the resolved and total analysis error in the case $J = 15$ for a single observation located at $x = 0$. These results were obtained with the generalized KF gain matrix, i.e., (I.2.19). A comparison with Fig. 8a of Part I, in which all scales were resolved, shows that while the unresolved scales have greatly increased the magnitude of the analysis error, they have had little effect on its overall structure. The corresponding forecast error (not shown) was rather similar, and as in Fig. 8 of Part I, the negative of the corresponding signal/analysis error covariance, i.e., $-\hat{X}$ transformed to spatial form, (not shown) was also very similar in form to the total analysis error.

The crosses and the dashed curve in Fig. 10a show the corresponding resolved and total analysis error when $\boldsymbol{K}$ is defined by (I.2.20), instead of (I.2.19). Use of this conventional gain matrix is seen to result in only a slight change in the analysis error.

By contrast, the effect of dropping the signal/error correlations, shown in Fig. 10b, is seen to have a major impact, as in Subsection 5.4 of Part I. While some decrease in the resolved analysis error in the vicinity of the observation is evident, overall the analysis error is much smaller and much more homogeneous than before. Also, it is interesting to note that the marked asymmetry about $x = 0$ in Fig. 10a is no longer evident in Fig. 10b.

While Fig. 10a indicates that the choice of conventional or generalized gain matrix has only a small impact on the stationary component of the analysis error, an examination of the time-dependent behaviour shows important differences. This is illustrated in Fig. 11. The top two panels show the time-dependent behaviour of the analysis error at two analysis points, $x = -4\Delta x$ (where the analysis error is a maximum) and $x = 0$ (where it is a minimum), over the last 300 timesteps of a long integration of the second-moment equations with the generalized KF gain matrix, (I.2.19). There is evidence of a very small amplitude periodic oscillation having a period of 30 timesteps.

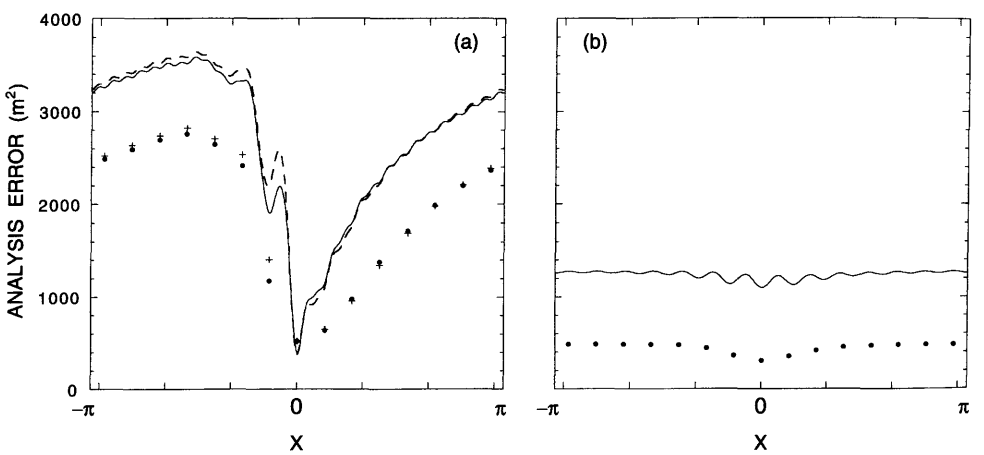

Fig. 10. As in Fig. 3b but there is only one observation (located at $x = 0$) and model M_2 is used. $C_M = 0.5$. The effect of the serial correlations is retained in panel (a) but neglected in panel (b). In panel (a) the dots and the solid curve indicate the resolved and total analysis error obtained using the generalized KF gain matrix, while the crosses and dashed curve show the corresponding quantities obtained with the conventional gain matrix.

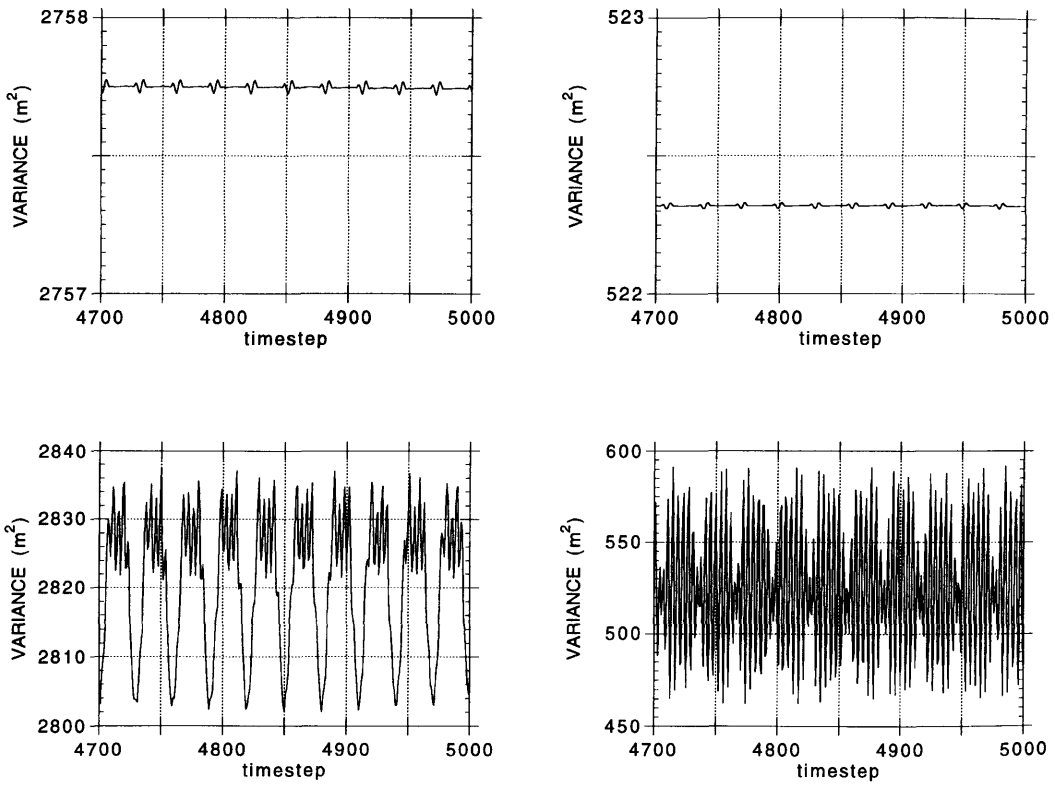

Fig. 11. The top two panels show the time-evolution of the analysis error at $x = -4\Delta x$ and $x = 0$ over an interval of 300 timesteps when the generalized KF gain matrix is used in the single-observation experiment. The corresponding quantities, except using the conventional gain matrix, are plotted in the bottom two panels.

Other experiments showed that this period is determined by the time taken for the effect of an observation to make a complete circuit of the domain, i.e., J/C_M (= 15/0.5 in the present case). Examination of the time-dependent behaviour at other times (not shown) indicated that the amplitude of the oscillation is decreasing with time. In any case, for all practical purposes, the solution shown in these panels is stationary. In contrast, the corresponding results with the conventional gain matrix (I.2.20), shown in the bottom two panels of Fig. 11, exhibit large-amplitude time-dependent behaviour. (Note the change in scale from upper to lower panels.) In addition to the evident 30-timestep period, there are also seen to be other important oscillations having shorter periods, all of which combine to produce a complex time-dependent behaviour. In view of these results, rather than showing the final solution in Fig. 10a, the solutions actually plotted were obtained by time-averaging over the 300 timesteps considered in Fig. 11.

The results of Fig. 11 (like those of Subsection 5.2) clearly demonstrate the superiority of the generalized KF gain matrix over the conventional one, when there are signal/error correlations. This, even though with the single observation located at a point of the analysis mesh, the effect of the additional terms in (I.2.19) is minimized, since $(\boldsymbol{G} - \boldsymbol{HF})_{rr} = 0$.

5.5. *Homogeneous (nonuniform) observation networks with imperfect model ($\boldsymbol{M}_2$)*

The general case of a homogeneous observation network with the imperfect model, $\boldsymbol{M}_2$, and imperfect forward interpolation matrix, $\boldsymbol{H}_2$, will now be considered. We will examine analysis error magni-

tude as a function of model Courant number (C_M) in the case of a homogeneous observation network and compare the effects of observation density and model resolution on analysis error magnitude and forecast error correlation structure. In view of the results above, all experiments which include signal/error correlation will be done with the generalized KF gain matrix.

Fig. 9b shows the effect on the average analysis error of varying C_M in the case $I = J = 15$ with the homogeneous observation network of Fig. 5. A comparison of the two panels of Fig. 9 shows the effect of using a homogeneous observation network, rather than one which coincides with the analysis mesh. It can be seen that while changing the observation network has little effect when the signal/error correlation is neglected (dotted curve), there is a significant effect in the general case. The signal/error correlation, while exhibiting the same overall behaviour in the two panels, no longer becomes strongly positive near $C_M = 1$ and 2. In addition, it is only for $C_M < 0.25$ that the behaviour of the analysis error (solid curve) is similar in the two panels. For larger values of C_M, the analysis error in the general case is rather similar to the analysis error when the signal/error correlation is neglected. Note that it now assumes its maximum value close to where the fractional part of $C_M = 0.5$. An examination of the two panels also shows that, except for small values of C_M, the homogeneous observation network gives rise to smaller analysis errors than the coincident uniform network. This differs from the no-prior-estimate case of Section 4 but is consistent with the perfect-model results in Subsection 5.2 above and in Part I, where we saw that uniform networks impede filter performance. Here we find that this is true even of uniform networks which coincide with the analysis mesh.

The upper-left panel of Fig. 12 shows the analysis error for the case $C_M = 0.5$. This panel will be used as a baseline for examining the effect on analysis error magnitude of increasing observation density and/or model resolution, but first a number of remarks. Comparison with Fig. 5 (for which the same observation network and no model was used) again shows the very substantial benefit of using even an imperfect model. The use of $\boldsymbol{H}_2$ (dots and solid curve), in place of $\boldsymbol{H}_1$ (crosses and dashed curve), is seen to result in a damping of the largest oscillations, as in the no-model case. Comparison with Fig. 7 illustrates the discussion of the previous paragraph with a specific example: for the same C_M, smaller analysis errors are obtained with a homogeneous observation network than with a uniform observation network that coincides with the analysis mesh.

The effect of increasing the number of observations and/or the resolution of the analysis mesh is illustrated in the other panels of Fig. 12. (Note the change in scale from upper to lower panels.) The lower-left panel shows the analysis error for the same observation network but with $J = 45$. This increase in J is seen to substantially reduce the analysis error, in both the resolved and unresolved scales. Tripling the number of observations but retaining the coarse analysis mesh (upper-right panel) substantially reduces the analysis error in the resolved scales, but cannot eliminate the large error in the unresolved scales. (A similar effect was noted when the number of observations was increased in the absence of a model, see Subsection 4.3.) The error in the unresolved scales can only be reduced by increasing the resolution of the analysis mesh, as in the lower-right panel. Due to this error, the resolution of the analysis mesh is more important than that of the observation network in determining the total analysis error.

The shape of the correlation functions at equilibrium for each of the four cases of Fig. 12 is also of interest. Due to their utility in statistical interpolation and 3-d variational analysis, we show the forecast-error correlations in Fig. 13. (The analysis-error correlations do not damp out quite as quickly away from the origin, but otherwise are similar.) To average-out the variation in correlation structure over the domain, each of the correlations plotted in Fig. 13 is actually a spatial average of the correlations at all the points of the high-resolution output grid.

Fig. 13 again shows the importance of the analysis-mesh resolution: the correlation widths of the heavy curves, for which $J = 45$, are seen to be substantially narrower than those of the thin curves, for which $J = 15$. The results also show that for a given value of J, the dotted curves, for which $I = 45$, are slightly narrower than the solid curves, for which $I = 15$. Thus, as with the magnitude of the total analysis error (Fig. 12), both the resolution of the analysis mesh and the observation network influence forecast-error correlation width, but the effect of the former is substantial

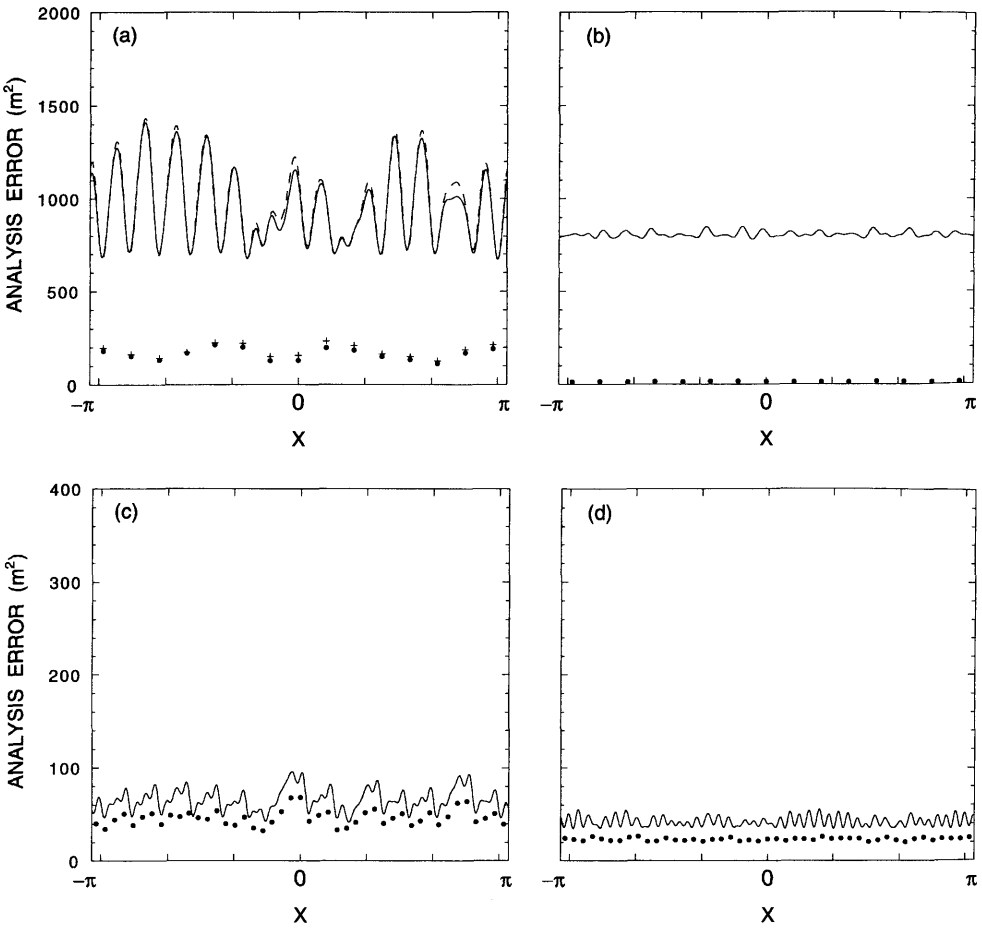

Fig. 12. The analysis error variance as a function of the spatial coordinate, x, in the case of a homogeneous observation network with 15 observations (left-hand panels) and 45 observations (right-hand panels). For the upper panels, the number of analysis points, J, is 15; while for the lower panels, $J = 45$. $C_M = 0.5$. Model $\boldsymbol{M}_2$ and forward interpolation matrix $\boldsymbol{H}_2$ are used in all cases except $\boldsymbol{H}_1$ is used for the dashed curve and crosses in the upper-left panel.

while the effect of the latter is slight. These results could depend to some extent on the specified observation errors and signal spectrum (Fig. 1a, Part I).

6. Summary and conclusions

Spatial scales which are unresolved by the analysis mesh can have major negative effects on atmospheric data assimilation. In fact, it is quite easy to seriously underestimate the analysis error when the analysis grid resolution is inadequate to resolve the signal. Unresolved scales have a primary effect on the unresolved scales and a secondary effect due to foldback (aliasing) of error into the resolved scales. The primary effect is always present; even when the errors in the resolved scales can be eliminated, the errors in the unresolved scales always remain. The secondary effects due to foldback of error into the resolved scales are more subtle.

When there is no model (Section 4), the effect of the foldback error on the analysis error is, on average, as large as the primary unresolved-scale errors. The true analysis errors may be much

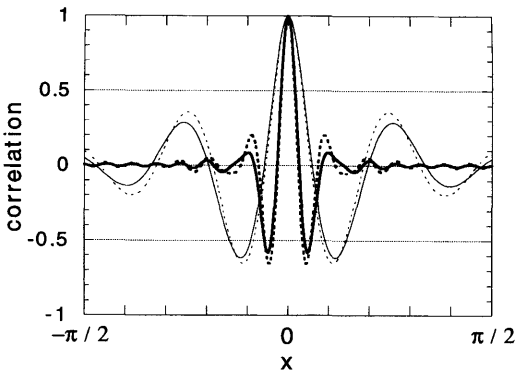

Fig. 13. The forecast-error correlation for different values of I and J. The heavy curves are for $J = 45$, while the thin curves are for $J = 15$. The dotted curves are for $I = 45$, while the solid curves are for $I = 15$.

larger than the perceived analysis errors on the analysis mesh, particularly between observation stations, depending upon the type of forward interpolation operator and the configuration of the observation network. In the asymptotic limit of an infinite observation density, there is no analysis error in the resolved scales for a perfect (Fourier) forward interpolation operator, while an imperfect (cubic spline) interpolator has analysis errors in both resolved and unresolved scales.

The primary effect of a model (Section 5) is to substantially reduce the error in the resolved scales, but there is no effect on the unresolved scales. For the Fourier semi-Lagrangian model (no error in resolved scales), the second-moment equations degenerate, as when all scales were resolved (Part I). However when there are unresolved scales, the asymptotic analysis error can depend on the initial conditions, not only with the cubic spline forward interpolator, but also with the Fourier forward interpolator, even though the latter is perfect in the resolved scales.

For the cubic spline semi-Lagrangian model (which has error in both resolved and unresolved scales) and either a uniform or non-uniform observation network, the analysis error tends to be minimized at the observations, but can get much larger between observations. For an observation network coincident with the analysis mesh, it was found that as the model Courant number (C_M) varies, the sign of the signal/error correlation changes and the magnitude of the asymptotic analysis error increases as the model error decreases. This counter-intuitive behaviour was not observed in Part I, nor is it observed here in the case of more general observation networks or when the signal/error correlation is neglected. Its occurence here seems to be due to a combination of the special (coincident) observation network, the constant advecting velocity and the periodic boundary conditions. Thus while such special networks can provide useful information about the mechanisms of error growth when analytic results can be obtained, these results *may* be misleading with respect to the properties of more realistic networks and models. For example, while it seems that the standard Kalman filter is always well-behaved in these circumstances, the introduction of signal/error correlation and/or unresolved scales may occasionally give rise to pathological results with these special networks.

For a cubic spline semi-Lagrangian model which does not resolve all scales, the model error variance has a minimum at large scales and a maximum at the smallest resolvable scales. The corresponding forecast and analysis error have much redder spectra, particularly when the signal/error correlation is properly accounted for. This result is consistent with that of Part I (except at a much higher error level) in demonstrating the importance of the signal/error correlation and its effect on reddening the error spectra. Thus, the relatively small (and blue) model discretization error can have a large effect on the large scale analysis through foldback from the unresolved scales coupled with signal/error correlation. The importance of signal/error correlation was also strikingly demonstrated in the case of the cubic-spline semi-Lagrangian model and a single observation, as in Part I but at a much higher error level.

Dee (1991, 1995) has argued that the real obstacle to a successful implementation of the Kalman filter is not the computational burden, but rather our inability to properly specify the model error. He concludes that the only feasible approach is to try to parameterize the model error. Our results indicate that parametrization of model error statistics must explicitly or implicitly account for correlation with the signal in some fashion. Indeed this study indicates how very complicated an attempt to specify the model error from first principles could become.

A final set of results showed that, for imperfect

models, the maximum reduction in analysis error occurs by increasing model resolution, rather than by increasing observation network density. The forecast error correlation structure (whose correct specification is so important in statistical interpolation and 3-d variational analysis algorithms), was found to be much more sensitive to model resolution than to observation network density.

7. Appendix A

Properties of the Fourier transforms $\boldsymbol{F}$ and $\tilde{\boldsymbol{F}}$

When there are unresolved scales, $\tilde{\boldsymbol{F}}$ is not the matrix inverse of $\boldsymbol{F}$, as it was in Part I. Instead, $\tilde{\boldsymbol{F}}$ is a generalized inverse of $\boldsymbol{F}$ in the sense that $\boldsymbol{F}\tilde{\boldsymbol{F}}$ is the $J \times J$ identity matrix, while $\tilde{\boldsymbol{F}}\boldsymbol{F}$ is the $(2K+1) \times (2K+1)$ matrix whose elements are all zero except for the first J diagonal elements which equal one. It follows that $[\tilde{\boldsymbol{F}}\boldsymbol{F} - \boldsymbol{I}]$, which appears in (2.6)–(2.9), is the $(2K+1) \times (2K+1)$ diagonal matrix whose first J diagonal elements are zero and whose remaining diagonal elements are -1.

$\tilde{\boldsymbol{F}}\tilde{\boldsymbol{F}}^{\mathrm{T}}$ and $\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}$ are diagonal $(2K+1) \times (2K+1)$ matrices having only the first J diagonal elements non-zero. (When all scales are resolved, as in Part I, all diagonal elements are non-zero.) In the case of $\tilde{\boldsymbol{F}}\tilde{\boldsymbol{F}}^{\mathrm{T}}$, the upper left-hand corner element equals $(4/J)$ and the following $J-1$ diagonal elements equal $(2/J)$, while in the case of $\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}$, the corresponding values are $(J/4)$ and $(J/2)$, respectively, consistent with the fact that $(\tilde{\boldsymbol{F}}\tilde{\boldsymbol{F}}^{\mathrm{T}})(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}) = \tilde{\boldsymbol{F}}(\boldsymbol{F}\tilde{\boldsymbol{F}})^{\mathrm{T}}\boldsymbol{F} = \tilde{\boldsymbol{F}}\boldsymbol{F}$.

8. Appendix B

No prior estimate, uniform observation network with $I = J \leqslant 2K+1$

In this case, $\boldsymbol{H}$ is a square matrix, $\boldsymbol{K} = \boldsymbol{H}^{-1}$ and we can write $\boldsymbol{H} = \boldsymbol{G}\hat{\boldsymbol{D}}\tilde{\boldsymbol{F}}$, where $\hat{\boldsymbol{D}} = \boldsymbol{I}$ for $\boldsymbol{H} = \boldsymbol{H}_1$ and $\hat{\boldsymbol{D}}$ is as defined in Appendix B (Part I) for $\boldsymbol{H} = \boldsymbol{H}_2$. Therefore

$$\boldsymbol{G} - \boldsymbol{H}\boldsymbol{F} = \boldsymbol{G}[\boldsymbol{I} - \hat{\boldsymbol{D}}\tilde{\boldsymbol{F}}\boldsymbol{F}]. \tag{B.1}$$

Substituting into (I.4.1)–(I.4.2) yields expressions for $\bar{\boldsymbol{P}}^{\mathrm{a}}$ and $\boldsymbol{X}$, from which we easily obtain

$$\tilde{\boldsymbol{F}}\bar{\boldsymbol{P}}^{\mathrm{a}}\tilde{\boldsymbol{F}}^{\mathrm{T}} = [\tilde{\boldsymbol{F}}\boldsymbol{H}^{-1}\boldsymbol{G}]\{\hat{\boldsymbol{R}}^{\mathrm{i}} + [\boldsymbol{I} - \hat{\boldsymbol{D}}\tilde{\boldsymbol{F}}\boldsymbol{F}]\hat{\boldsymbol{S}} \times [\boldsymbol{I} - \hat{\boldsymbol{D}}\tilde{\boldsymbol{F}}\boldsymbol{F}]^{\mathrm{T}}\}[\tilde{\boldsymbol{F}}\boldsymbol{H}^{-1}\boldsymbol{G}]^{\mathrm{T}}, \tag{B.2}$$

$$\hat{\boldsymbol{X}} = \tilde{\boldsymbol{F}}\boldsymbol{X} = [\tilde{\boldsymbol{F}}\boldsymbol{H}^{-1}\boldsymbol{G}][\boldsymbol{I} - \hat{\boldsymbol{D}}\tilde{\boldsymbol{F}}\boldsymbol{F}]\hat{\boldsymbol{S}}. \tag{B.3}$$

The matrix $\tilde{\boldsymbol{F}}\boldsymbol{H}^{-1}\boldsymbol{G}$ can be written

$$\begin{bmatrix} \tilde{\boldsymbol{F}}_{\mathrm{rr}} \\ \tilde{\boldsymbol{F}}_{\mathrm{ur}} \end{bmatrix} [\boldsymbol{H}^{-1}][\boldsymbol{G}_{\mathrm{rr}} \;\; \boldsymbol{G}_{\mathrm{ru}}].$$

Since $\tilde{\boldsymbol{F}}_{\mathrm{ur}} = 0$, $\boldsymbol{H} = \boldsymbol{G}\hat{\boldsymbol{D}}\tilde{\boldsymbol{F}} = \boldsymbol{G}_{\mathrm{rr}}\hat{\boldsymbol{D}}_{\mathrm{rr}}\tilde{\boldsymbol{F}}_{\mathrm{rr}}$. Therefore

$$\tilde{\boldsymbol{F}}\boldsymbol{H}^{-1}\boldsymbol{G} = \begin{bmatrix} \hat{\boldsymbol{D}}_{\mathrm{rr}}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{\mathrm{rr}} & \boldsymbol{G}_{\mathrm{rr}}^{-1}\boldsymbol{G}_{\mathrm{ru}} \\ 0 & 0 \end{bmatrix}, \tag{B.4}$$

where $\boldsymbol{I}_{\mathrm{rr}}$ is the $J \times J$ identity matrix. As we shall see presently, it is through the non-zero elements of $\boldsymbol{G}_{\mathrm{rr}}^{-1}\boldsymbol{G}_{\mathrm{ru}}$ that the scales that are not resolved affect those that are. We focus on (B.2), noting that (B.3) can be examined in a similar way.

Since

$$\boldsymbol{I} - \hat{\boldsymbol{D}}\tilde{\boldsymbol{F}}\boldsymbol{F} = \begin{bmatrix} \boldsymbol{I}_{\mathrm{rr}} - \hat{\boldsymbol{D}}_{\mathrm{rr}} & 0 \\ 0 & \boldsymbol{I}_{\mathrm{uu}} \end{bmatrix}, \tag{B.5}$$

it follows that

$$[\boldsymbol{I} - \hat{\boldsymbol{D}}\tilde{\boldsymbol{F}}\boldsymbol{F}]\hat{\boldsymbol{S}}[\boldsymbol{I} - \hat{\boldsymbol{D}}\tilde{\boldsymbol{F}}\boldsymbol{F}]^{\mathrm{T}} = \begin{bmatrix} [\boldsymbol{I}_{\mathrm{rr}} - \hat{\boldsymbol{D}}_{\mathrm{rr}}]\hat{\boldsymbol{S}}_{\mathrm{rr}}[\boldsymbol{I}_{\mathrm{rr}} - \hat{\boldsymbol{D}}_{\mathrm{rr}}]^{\mathrm{T}} & 0 \\ 0 & \hat{\boldsymbol{S}}_{\mathrm{uu}} \end{bmatrix}. \tag{B.6}$$

Moreover since $\hat{\boldsymbol{D}}$ is block diagonal with 2×2 blocks given by (B.2) of Part I, $[\boldsymbol{I}_{\mathrm{rr}} - \hat{\boldsymbol{D}}_{\mathrm{rr}}]\hat{\boldsymbol{S}}_{\mathrm{rr}}[\boldsymbol{I}_{\mathrm{rr}} - \hat{\boldsymbol{D}}_{\mathrm{rr}}]^{\mathrm{T}}$, the $J \times J$ upper diagonal block corresponding to the resolved part of the spectrum, satisfies (I.3.6) and is diagonal. Furthermore, as noted in Section 3, $\hat{\boldsymbol{R}}^{\mathrm{i}}$ is also diagonal with $\hat{\boldsymbol{R}}^{\mathrm{i}}_{\mathrm{uu}} = 0$.

Substitution into (B.2) shows that the only non-zero block of $\tilde{\boldsymbol{F}}\bar{\boldsymbol{P}}^{\mathrm{a}}\tilde{\boldsymbol{F}}^{\mathrm{T}}$ is

$$(\tilde{\boldsymbol{F}}\bar{\boldsymbol{P}}^{\mathrm{a}}\tilde{\boldsymbol{F}}^{\mathrm{T}})_{\mathrm{rr}} = \hat{\boldsymbol{D}}_{\mathrm{rr}}^{-1}\{\hat{\boldsymbol{R}}^{\mathrm{i}}_{\mathrm{rr}} + \hat{\boldsymbol{S}}_{\mathrm{rr}} \times [\boldsymbol{I}_{\mathrm{rr}} + \hat{\boldsymbol{D}}_{\mathrm{rr}}\hat{\boldsymbol{D}}_{\mathrm{rr}}^{\mathrm{T}} - (\hat{\boldsymbol{D}}_{\mathrm{rr}} + \hat{\boldsymbol{D}}_{\mathrm{rr}}^{\mathrm{T}})] + \hat{\boldsymbol{B}}\}\hat{\boldsymbol{D}}_{\mathrm{rr}}^{-\mathrm{T}}, \tag{B.7}$$

where $\hat{\boldsymbol{B}}$ is the $J \times J$ matrix defined by

$$\hat{\boldsymbol{B}} = \boldsymbol{G}_{\mathrm{rr}}^{-1}\boldsymbol{G}_{\mathrm{ru}}\hat{\boldsymbol{S}}_{\mathrm{uu}}\boldsymbol{G}_{\mathrm{ru}}^{\mathrm{T}}\boldsymbol{G}_{\mathrm{rr}}^{-\mathrm{T}}. \tag{B.8}$$

The first two terms in the braces on the right-hand side of (B.7) involve only the resolved scales, but in the term $\hat{\boldsymbol{B}}$ the resolved scales are directly affected by the unresolved scales. In fact, it can be shown that $\hat{\boldsymbol{B}}$ is diagonal and that

$$\text{Trace } \hat{\boldsymbol{B}} = \text{Trace } \hat{\boldsymbol{S}}_{\mathrm{uu}}, \tag{B.9}$$

where the upper left-hand corner element of $\hat{\boldsymbol{B}}$, which corresponds to wavenumber 0, must be multiplied by $\frac{1}{2}$, for consistency with (I.2.1).

9. Appendix C

No prior estimate, infinite observation density

The analysis error in the resolved scales, $\bar{\boldsymbol{P}}^{\mathrm{a}}$, is a sum of two terms, from (I.4.1): one due to instrument error and the other due to forward interpolation error. If $\boldsymbol{H}=\boldsymbol{H}_1$, the instrument error term reduces to $(E^{\mathrm{r}})^2(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}$, which equals $(E^{\mathrm{r}})^2(\tilde{\boldsymbol{F}}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\tilde{\boldsymbol{F}})^{-1}$, from (I.3.8). As $I\to\infty$, $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ becomes a diagonal matrix with diagonal elements $\approx(I/2)$, (except for the upper left-hand corner element which is half as large), while $\tilde{\boldsymbol{F}}^{\mathrm{T}}\tilde{\boldsymbol{F}}$, which is independent of I, does not change. It follows that, for large I, $\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{H}_1$ is proportional to I and therefore as $I\to\infty$, $(\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{H}_1)^{-1}\to 0$. $(\boldsymbol{H}_2^{\mathrm{T}}\boldsymbol{H}_2)^{-1}$ is observed to behave in the same way for large I, so the instrument error term becomes vanishingly small as $I\to\infty$ for both $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$.

For $\boldsymbol{H}=\boldsymbol{H}_1$, the term due to forward interpolation error can be expressed as

$$(\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{H}_1)^{-1}[\tilde{\boldsymbol{F}}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}(\tilde{\boldsymbol{F}}\boldsymbol{F}-\boldsymbol{I})\hat{\boldsymbol{S}} \times(\tilde{\boldsymbol{F}}\boldsymbol{F}-\boldsymbol{I})^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\tilde{\boldsymbol{F}}](\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{H}_1)^{-\mathrm{T}}.$$

As $I\to\infty$, this can be rewritten as

$$(\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{H}_1)^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}[\tilde{\boldsymbol{F}}^{\mathrm{T}}(\tilde{\boldsymbol{F}}\boldsymbol{F}-\boldsymbol{I})\hat{\boldsymbol{S}} \times(\tilde{\boldsymbol{F}}\boldsymbol{F}-\boldsymbol{I})^{\mathrm{T}}\tilde{\boldsymbol{F}}]\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{H}_1)^{-\mathrm{T}}.$$

Now $(\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{H}_1)^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ and its transpose remain finite as $I\to\infty$. However, the term in square brackets in the above expression vanishes since, from (2.11), $\tilde{\boldsymbol{F}}$ only has non-zero elements in $\tilde{\boldsymbol{F}}_{\mathrm{rr}}$, while the only non-zero elements of $(\tilde{\boldsymbol{F}}\boldsymbol{F}-\boldsymbol{I})\hat{\boldsymbol{S}}(\tilde{\boldsymbol{F}}\boldsymbol{F}-\boldsymbol{I})^{\mathrm{T}}$ are those in $\hat{\boldsymbol{S}}_{\mathrm{uu}}$. Thus for $\boldsymbol{H}=\boldsymbol{H}_1$, the contribution to the analysis error in the resolvable scales due to forward interpolation error vanishes as $I\to\infty$, as illustrated for $J=15$ in Fig. 6a. This is basically due to the fact that $[\boldsymbol{G}-\boldsymbol{H}_1\boldsymbol{F}]$ *is equal to zero for the resolvable scales*, as discussed in Section 3, and is true despite the fact that $(\boldsymbol{G}-\boldsymbol{H}_1\boldsymbol{F})\hat{\boldsymbol{S}}(\boldsymbol{G}-\boldsymbol{H}_1\boldsymbol{F})^{\mathrm{T}}$, the forward interpolation error covariance for $\boldsymbol{H}=\boldsymbol{H}_1$ is spatially correlated, as shown by Daley (1993).

On the other hand, $[\boldsymbol{G}-\boldsymbol{H}_2\boldsymbol{F}]$ *is not equal to zero for the resolvable scales.* Therefore, for $\boldsymbol{H}=\boldsymbol{H}_2$, the contribution to the analysis error in the resolvable scales due to forward interpolation error does not vanish as $I\to\infty$. Fig. 6b shows the analysis error in this case for $J=15$.

REFERENCES

Cohn, S. E. and Dee, D. P. 1988. Observability of discretized partial differential equations. *SIAM J. Numer. Anal.* **25**, 586–617.

Daley, R. 1991. *Atmospheric data analysis*, Cambridge University Press, Cambridge, 457 pp.

Daley, R. 1993. Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophysicae* **11**, 634–647.

Daley, R. and Ménard, R. 1993. Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev.* **121**, 1554–1565.

Dee, D. P. 1991. Simplification of the Kalman filter for meteorological data assimilation. *Quart. J. Roy. Meteor. Soc.* **117**, 365–384.

Dee, D. P. 1995. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.* **123**, 1128–1145.

Lorenc, A. C. 1986. Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.* **112**, 1177–1194.

Mitchell, H. L. and Daley, R. 1997. Discretization error and signal/error correlation in atmospheric data assimilation. Part I: All scales resolved. *Tellus*, this issue.