Tellus (1998), 50A, 42–57

TELLUS
ISSN 0280–6495

A direct way of specifying flow-dependent background error correlations for meteorological analysis systems

By LARS PETER RIISHØJGAARD*, *Joint Center for Earth Systems Technology, University of Maryland Baltimore County, USA*

(Manuscript received 2 May 1997; in final form 2 September 1997)

ABSTRACT

Identifying a viable strategy for specifying the background error covariance remains an important problem in meteorological data assimilation. From a formal point of view the number of independent parameters needed for this is $n^2/2$ where n is the dimension of the model state space. In most analysis systems used in operational mode at the present time, the error covariance is modeled using assumptions about homogeneity and isotropy, and the resulting background error covariance matrix thus does not depend on the state of the atmosphere. In this paper, we propose a simple and inexpensive method for specifying univariate background error correlations in terms of the background field itself. We illustrate the positive impact of the implementation of this model by a simple example in which we reconstruct a total ozone field from a sparse set of observations. We finally discuss the generalization of the basic idea involved to univariate correlations for general meteorological models.

1. Introduction

A substantial amount of progress has been made over the last few years in devising new methods and in improving already existing ones for atmospheric data assimilation. Most of these methods can (under certain assumptions about the statistical nature of the observation and background errors) be shown to be equivalent to either directly or approximately solving the following equation (Cohn, 1997):

$$x^a = x^f + K(y - Hx^f), \tag{1}$$

where x^f (a vector of length n) is the background estimate of the analyzed state given by a short range forecast initiated at the previous analysis time, x^a the analysis, and y the observations (a vector of length p). H is the so-called observation operator that projects the model state vector onto the observation space, and Hx^f thus provides a model estimate of the observations y. The $n \times p$ gain matrix K specifies how a given innovation (the difference between the forecast projected onto the observation space and the observations themselves) is to be interpreted back in the state space of the model. Assuming that both the forecast and the observations are unbiased, and that forecast errors are uncorrelated with observation errors, the specific gain matrix that leads to the best estimate of x in a minimum variance sense is (Cohn, 1997):

$$K = P^f H^T (HP^f H^T + R)^{-1}. \tag{2}$$

In this equation, P^f is the forecast error covariance matrix and R is the observation error covariance matrix. As already noted, the background estimate is invariably given by a short range forecast in present operational meteorological analysis systems, and in the remainder of this article we shall thus use the terms forecast error

* Corresponding address: Data Assimilation Office, Goddard Space Flight Center (Code 910.3), Greenbelt, MD 20771, USA.

covariance and background error covariance interchangeably. Apart from designing computationally efficient algorithms for solving this rather unwieldy equation, arguably the most important problem in data assimilation at the present time is that of specifying the background error covariance matrix. As is evident from eq. (2), this matrix directly controls the outcome of the assimilation procedure. According to its definition, P^f is found by taking the expectation value of differences ε between the true state of the atmosphere x^t and an ensemble of corresponding background estimates of the field:

$$P^f_{ij} = \langle \varepsilon_i \varepsilon_j \rangle = \langle (x^t_i - x^f_i)(x^t_j - x^f_j) \rangle \,. \tag{3}$$

For convenience, we also introduce the background error correlation matrix C, obtained by normalizing each element of P by the product of standard deviations of the errors at the points i and j.

$$C^f_{ij} = P^f_{ij}(\sigma_i \sigma_j) = \frac{\langle (x^t_i - x^f_i)(x^t_i - x^f_i) \rangle}{\sqrt{\langle (x^t_i - x^f_i)^2 \rangle}\sqrt{\langle (x^t_j - x^f_j)^2 \rangle}}. \tag{4}$$

The diagonal elements of C are all equal to one, whereas the diagonal of P is given by the forecast error variances, $P_{ii} = \sigma_i^2$. In this article we focus on the correlation part of the problem, and we assume that the forecast error variance is known. Specifying it correctly is in fact an important problem not only for the sake of data assimilation, but also because of its application as a measure of confidence in a given forecast.

The size of P^f is $n \times n$, and for present numerical weather prediction models it thus contains upwards of 10^{12} elements, even allowing for symmetry. Apart from the obvious problems this creates for the memory of present computers, there is no known strategy for generating this vast amount of information. The actual values of the elements P^f_{ij} are supposed to reflect the statistics of an ensemble of forecast errors of which we do not know even a single member. Real implementations are therefore generally based on radical simplifications with respect to the definition given in eq. (3).

At least four different basic approaches to the problem of specifying P^f can be identified in the literature. The first is climatological, the second is based on ensemble forecasting, the third involves solving a prognostic equation for the covariance evolution, and the fourth is direct specification of flow dependent correlation functions. Below we give a brief overview of the methodologies. However, we point out that until now only the first of them has been widely used in operational mode.

In the 1st approach, P^f is modeled assuming ergodicity of the forecasting system. Under this assumption the ensemble averaging of eq. (3) can be substituted by an average over a time sequence of innovations, which then leads to a climatological description of the forecast error. Further simplification is often achieved through assumptions of homogeneity and isotropy of the error correlations. This reduces the number of free parameters in the covariance model to a level where some form of tuning procedure can be used for fitting it to the data (Dee, 1995). The main disadvantage of the climatological approach is that the background error covariance is treated as being decoupled from the actual flow state of the atmosphere. There appears to be little justification for the additional assumptions of homogeneity and isotropy.

The 2nd approach is known as Monte Carlo filtering or ensemble Kalman filtering (e.g. Evensen, 1994). This is the method that most closely reflects the statistical definition of P^f. Here an ensemble of forecasts is generated by perturbing an initial state randomly, and the desired statistics of the ensemble are then calculated explicitly. A basic problem with the method is how to define a minimal set of initial perturbations that will allow for a reasonably complete reconstruction of the forecast error covariance.

The 3rd approach is represented by the Kalman filter and its various generalizations. Here a prognostic equation for the time evolution of P^f is solved explicitly (Kalman, 1960; for a discussion about its use in meteorology, see, e.g., Cohn (1997)). However, there are a number of problems associated with the Kalman filter in meteorological applications: (i) it remains prohibitively expensive for extensive three-dimensional work; (ii) the simplest and most widely used generalization of the Kalman filter to non-linear systems, the so-called Extended Kalman Filter, is inconsistent (Cohn, 1993); (iii) in the absence of inclusion of a model error term the filter will suffer from a systematic decrease of analysis error variance; this will lead to rejection of observations (filter divergence); (iv) for partially observed systems the background error covariance will depend on its (unknown) initial conditions (Cohn and Dee, 1988).

The 4th approach is to specificy the correlations directly in terms of the flow, assuming that a meaningful functional form can be identified and tuned. Since what we are about to propose falls in this category, we now discuss this fourth approach in general and some earlier works in particular in a slightly more detailed manner.

It is useful at this point to temporarily abandon the statistical point of view when contemplating the background error covariance matrix P^f, and instead focus on its impact on the outcome of the analysis. Assume that the observation vector y consists of a single observation of the quantity represented by the element x_k of the model state vector. In this case both y and R in eq. (2) are scalars, and H is thus a $1 \times n$ matrix with $h_{1,k}=1$ being the only element different from zero. By doing the matrix multiplications on the right-hand side of eq. (2) it is easy to verify that the $n \times 1$ gain matrix K distributing the innovation y back in the model state space simply becomes proportional to the kth column of P^f

$$K = P^f_{|,k}\,(P^f_{k,k}+R)^{-1}\,. \tag{5}$$

Assuming that the variance field is constant ($\sigma_i = \sigma_j$ for all i,j) the analysis increment due to the single observation at the point k will look like a map of the background error correlation between that single point and the rest of the model domain. It follows that the increment due to any such set of uncorrelated observations can be written as a sum of column vectors of P^f. These vectors are therefore often also referred to as structure functions. The structure functions of an isotropic background error covariance matrix have concentric circles as isolines.

One of the goals of an analysis scheme is to calculate an increment that lets an observation at a given point influence the analysis not only at that point itself, but also at surrounding points for which the observation is considered representative. However, it should not change the field at points (even close to the observation) where the dynamical regime obviously differs from that of the observation point. A classical example is that of wind or temperature measurements obtained near a front. It is intuitively clear that the correlation length in the direction along the front should be longer than in the direction perpendicular to it. It is also likely that it should be different in the direction towards the front than in the direction away from it.

This general idea appears first to have been pursued by Atkins (1974). She used the gradient of the background relative humidity field to modulate the weights assigned to individual observations in a simple objective analysis procedure in which the field was expressed by a weighted sum of a background field and observations. The weight w_{ik} assigned to the observation at the point i in the analysis at the point k was given by

$$w_{ik} = \frac{1}{1 + a|\boldsymbol{r}_{ik}|^n + b\,(\nabla R_{bi}\cdot\boldsymbol{r}_{ik})^2}\,. \tag{6}$$

where R_b is the background field, and $\boldsymbol{r}_{ik}$ is the position vector of point i with respect to point k, and a, b, and n are free parameters. Although the anisotropy was found to have a positive impact on the forecast of rainfall, this particular approach does not appear to have been used in other contexts. We note that in the modern terminology of statistical data assimilation the weight w_{ik} would be given by the background error correlation between the points i and k. With this interpretation of w_{ik}, eq. (6) is invalid, since it is asymmetric with respect to a switch in positions. In other words it does not meet the criterion

$$w_{ik} = w_{ki}\,. \tag{7}$$

In other studies anisotropic weighting functions have been formulated by seeking to express a given desired shape in physical space in terms of flow parameters. Benjamin and Seaman (1985) provide an example of this approach, and references in their paper point to earlier efforts. They expressed the weights in terms of one of three weighting functions: circular, elliptical, or curved elliptical ("banana-shaped"). The choice of the function to be used as well as the parameter settings for the appropriate anisotropic formula was set to depend on the local flow at the observation point. The transition from one weighting function to another as one moves between different domains of the flow thus introduces some additional complexity in the analysis procedure, and the actual anisotropic weighting functions proposed in the article violate the symmetry criterion (7). The elliptical weighting functions were later implemented in a forecast/analysis system that used potential temperature as the vertical coordinate (Benjamin et al., 1991).

Correlations that are isotropic in one coordinate system may be anisotropic in another coordinate system. This concept was illustrated by Benjamin (1989) who showed that isotropic vertical correlations in an isentropic coordinate system lead to baroclinically tilting correlation structures when plotted in isobaric coordinates, even though the horizontal correlations remained isotropic in the separable formulation that he used.

Desroziers and Lafore (1993) proposed another way of achieving anisotropic correlations through a change of coordinate. They transformed wind and temperature fields into semi-geostrophic coordinates defined locally in terms of filtered wind observations and found by an iterative procedure. They then analyzed these fields assuming isotropic error correlations in the transformed coordinates. Among the problems associated with the method is the specification of the amount of filtering needed in determining the new coordinates, and the fact that the transformation is ill-defined in the tropics. A recently proposed generalization of this concept to a spectral three-dimensional variational analysis system (Desroziers, 1997) has not yet been implemented. In the variational implementation special care needs to be taken in accounting for the state dependence of the coordinate transformation.

In Section 2, we introduce a new and conceptually simple method for specifying the background error correlations in terms of the field itself. The underlying idea is that the deformation history of the flow remains embedded in the distribution of the field itself, and that this should guide the shape of the error correlations. We then illustrate the impact of the anisotropic covariances in a simple two-dimensional example. Before we summarize the ideas and results of this paper, we discuss the possible application of the basic concept to more general analysis systems.

2. A simple anisotropic background error correlation model

In the previous section we gave an overview of earlier efforts towards a direct flow-dependent specification of background error correlations. Although the studies differed both in terms of meteorological context, and in terms of the proposed solutions, they all identified and tried to cure basically the same problem: Isolines of isotropic structure functions that randomly cross isolines of the background field. That this occurs is at odds with the notion that observations should modify the field at points within the same air mass, and that the background estimate contains most of what we know about the air mass distribution of the field.

We now therefore seek to mathematically constrain the structure functions to follow the background field to some degree. Note that the basic assumption underlying the use of isotropic correlation models is extremely simple: "Background errors at nearby points are similar". A simple anisotropic formulation may be obtained by slightly modifying this assumption so that it reads "Background errors at nearby points *that have similar values of the field* are similar".

This assumption naturally leads to a correlation model in which the standard isotropic correlation between two points is multiplied by a function that decreases with the norm of the difference between their respective values of the field for which the correlation is being calculated.

Let $x^{\mathrm{b}}=\{x_1^{\mathrm{b}}, =\ldots, x_n^{\mathrm{b}}\}$ be a background estimate of the state variable x defined on the space M. Assume that the error covariance of x^{b} is given by an isotropic single-level covariance matrix P (e.g., DAO, 1996):

$$P_{ij}=\sigma_i\sigma_j\rho(r_{ij}), \tag{8}$$

where σ_i is the forecast error standard deviation at point i, and r_{ij} is the Euclidean distance between points i and j. The real-valued function ρ is assumed to be chosen such that P is a covariance matrix on M (see Gaspari and Cohn, 1996, for a detailed discussion on covariance modeling in two and three dimensions). This implies that P will be at least positive semidefinite. For certain choices of ρ, it will be positive definite.

Let furthermore the matrix D be defined as follows:

$$D_{ij}=\nu(\|x_i^{\mathrm{b}}-x_j^{\mathrm{b}}\|). \tag{9}$$

Here $x_i^{\mathrm{b}}-x_j^{\mathrm{b}}$ is the difference between the background estimate of the field at points i and j. If ν is a correlation function on R, it can be shown that D will be a positive semi-definite correlation matrix on the model state space. Positive definiteness cannot be guaranteed, since several grid-points may have the same value of x.

A simple flow-dependent covariance matrix B is obtained by letting B_{ij} be the product of the covariance given by eq. (8) and the correlation given by eq. (9):

$$B_{ij} = P_{ij} D_{ij} = \sigma_i \sigma_j \rho(r_{ij})\, v(\|x_i - x_j\|)\,. \tag{10}$$

Since B is the Hadamard product of P and D, it can be shown that if P is positive definite and D is a positive semi-definite correlation matrix, B will be positive definite. Since the product ρv is a correlation function on $M \times R$, B is thus a legitimate background error covariance matrix.

The matrix B is obtained by multiplying each element of the original isotropic covariance matrix P by the term (9) that we shall call the field term. It is of a similar form to the spatial correlation term ρ, apart from the fact that it decreases not with the physical distance between two points, but with the norm of the difference between their field values. The effect of the field term is that the originally concentric circular isolines of the correlation function are distorted towards following the isolines of the background field to some extent. The combined correlation will fall off rapidly in directions in which the field is changing rapidly, while the isotropic correlation will dominate in directions in which the rate of change of the field is slow. If both components (8) and (9) are given by one-parameter correlation functions, the combined anisotropic correlation will involve two tunable parameters. If L_1 is the length scale in physical space and L_2 is the length scale in field space, we see that in the limit $L_2 \to \infty$, the anisotropic covariance (10) reduces to the isotropic case (8). Both the horizontal extent of the correlation and the degree of anisotropy (i.e., similarity to the background field), can thus be controlled at will or subjected to rigorous tuning procedures (Dee, 1995).

From a computational point of view eq. (10) should not be much more expensive than the corresponding isotropic formulation. In real analysis systems based on isotropic error correlation models the full P^{f} matrix is never stored. Only the elements needed at any given time are even calculated. Similarly for the anisotropic formulation proposed here, the elements of P^{f} can be calculated on a "need to know" basis as indicated by the two terms $P^{\mathrm{f}} H^{\mathrm{T}}$ and $HP^{\mathrm{f}} H^{\mathrm{T}}$ of eq. (2). H is a $p \times n$ matrix, and we normally have $p \ll n$. Assuming that the observation operator H is local in physical space, only a limited number of the elements of P^{f} would thus be needed for calculating K at any given time. Even though the proposed model is fully flow-dependent, it differs in this respect from approaches that aim at calculating the entire error covariance matrix. For the Kalman filter in particular, the full covariance matrix is needed at a given time in order to calculate its value at the next filter time. In our approach the covariance propagation is presumed to be taken care of directly by the flow itself, and we are thus free to extract only the parts of P^{f} that we need and when we need them.

Ideally one would want the analysis increment to resemble the background error as much as possible. By expressing the error covariance as proposed in eq. (10), we are effectively assuming that the shape of the background error correlation is similar to that of the background field itself. While by definition we cannot know the background error exactly, the assumption that its correlation structure follows the background field itself can to a certain extent be justified as follows, if the field under consideration is conserved following the flow: Air parcels that are near each other at a given time t_0 will presumably have a high correlation between their errors at that time. In the absence of external error growth the error correlation between two parcels is conserved in a Lagrangian sense. Because of the deformation of the field due to the flow, the parcels may be far from each other at a later time t_1. Their field values are also conserved, however, so isolines of the error correlation should locally be similar to the isolines of the field. In the absence of a rigorous derivation of the error correlation, these remarks will have to serve as justification of our model, along with the case study to be presented in the following section.

3. The impact of anisotropic covariances in a two-dimensional constituent assimilation problem

In this section, we present results from a series of experiments designed to test the impact of the background error covariance in the assimilation of nadir observations of a passive tracer from a polar orbiting satellite. The physical framework is that of reconstructing a perfectly known total ozone field from a sparse and noisy set of observations under various assumptions about the background error covariance.

While the problem itself and the specific background estimate that we use both have links to an actual topic in stratospheric research, we emphasize that we are not here proposing a particular strategy for assimilating ozone observations. For one thing total ozone is not a conserved tracer, and the Lagrangian reasoning used for justifying the correlation model therefore strictly speaking does not apply. On the other hand it has been shown that total ozone is well correlated with a wide range of dynamical quantities in the lower stratosphere (Stanford and Ziemke, 1996, and numerous references therein), and over short spans of time Levelt et al. (1996) have successfully advected total ozone using single level winds for data assimilation purposes. We therefore use the total ozone estimation problem here as a convenient framework for illustrating some features of the state-dependent correlation model.

The state variable x to be analyzed is the vertically integrated ozone column, and the model domain M is a $2^\circ \times 2.5^\circ$ latitude/longitude grid covering the area from 60 W to the Greenwich meridian, and from 30 N to 60 N. In Fig. 1 the total ozone field over the model domain is shown, as seen by the Nimbus-7 TOMS instrument (McPeters et al., 1996). The data were obtained on 2 February, 1992, and while they are not strictly speaking synoptic, the difference in time over this limited region is small enough for us to consider them all valid at 12 UTC. TOMS is a nadir-scanning instrument, and the mean inherent resolution in the data is about $1^\circ \times 1^\circ$. The raw level 2 satellite retrievals have been resampled to the model grid by simple averaging of observations within a grid cell, and we define this resampled field to represent the true state x^t.

In addition to the field itself we have marked by asterisks the grid-cells in which the nadir observation is taken for each of the scans used in building the field shown in Fig. 1. For each such "observation grid-cell" we have calculated the average of the nadir TOMS measurements obtained within it, and added pseudo-random Gaussian noise with zero mean and a spread of

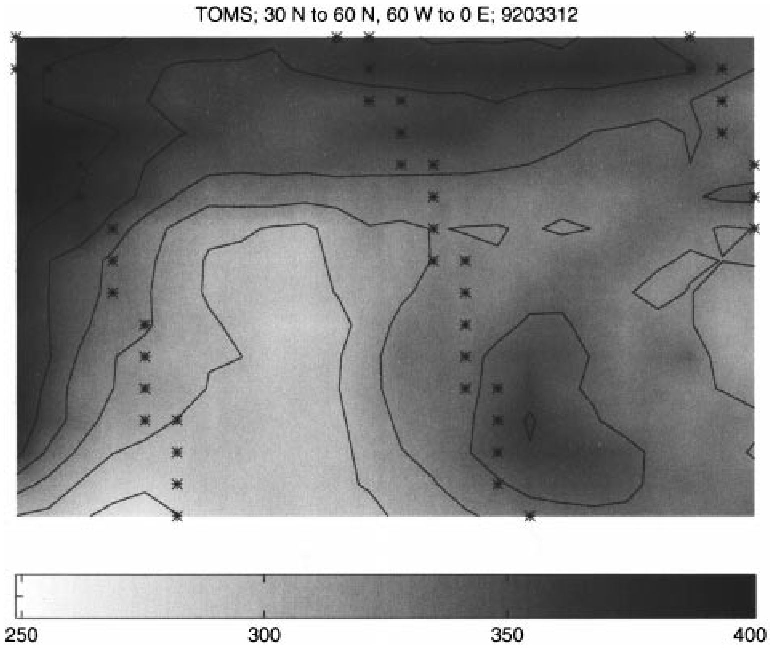

Fig. 1. TOMS total ozone (in Dobson Units) over model domain on 2 February 1992; this field represents true state of the model; asterisks indicate location of pseudo-observations.

3% of the averaged value. The 3% random error represents a conservative estimate of the actual measurement error, and the observation error standard deviation is thus 10.3 Dobson Units (DU). This sparse and noise-contaminated dataset forms the set of observations from which we attempt to reconstruct the full field.

Fig. 2 shows an independent background estimate, x^b, of the total ozone field from Fig. 1. The estimate is calculated using a hemispheric least squares regression of vertically integrated potential vorticity on total ozone. This method for estimating total ozone from meteorological analysis fields was first introduced by Allaart et al. (1993), and the data shown here are taken from Riishøjgaard et al. (1996).

This way of generating a background estimate shares some characteristic aspects with the general meteorological analysis problem: The background field is produced by an incomplete model of the physical system acting on stochastic input, and the result tends to contain the main features of the true state, while erring in a spatially non-white way in the details. Comparing Figs. 1 and 2 it is thus evident that the background field captures most of the main features of the true state, even though there are some fairly large differences locally. In Fig. 3 the background error $\varepsilon^b \equiv x^t - x^b$ is shown. We note that the total ozone field is underestimated by the background field in a broad band extending from west to east near the northern boundary of the domain, whereas it is overestimated both in a tongue originating at the southeastern corner and in a coherent region in the extreme northeastern part of the domain. The RMS of the background error is 20.9 DU, i.e., it is about twice as big as the observation error standard deviation.

The analysis problem now consists of reconstructing the true field (Fig. 1) from the background field (Fig. 2) and the observations, and the method is that of directly solving the analysis eq. (1). This is done for a range of different gain matrices that are all calculated using eq. (2), but differing in the way the forecast error covariance matrix P^f is specified. The background error variance field is set to be constant, and the value is the square of the known RMS background error (20.9 DU). We first show the increment due to a single observation using an isotropic covariance

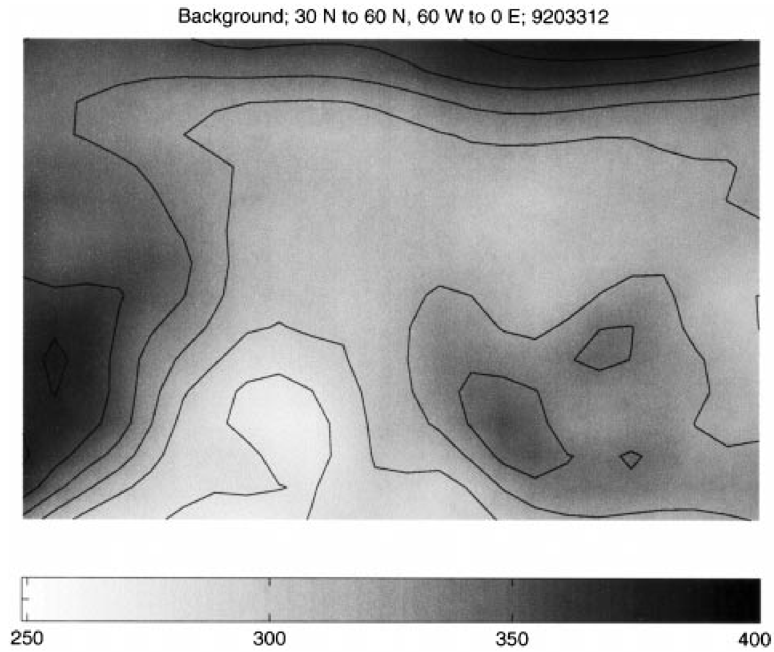

Fig. 2. Background estimate of total ozone field shown in Fig. 1, obtained by regression on vertically integrated PV (see text).

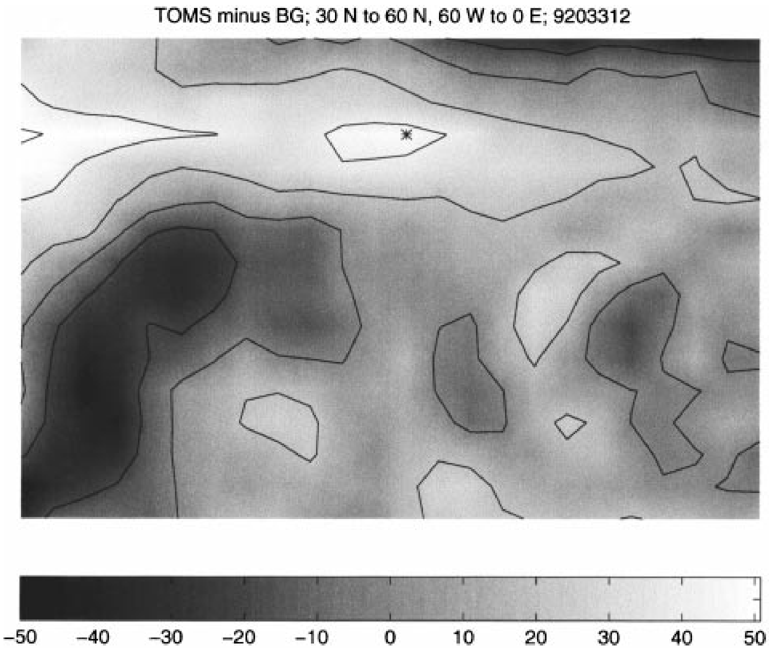

Fig. 3. Background error field (field from Fig. 1 minus field from Fig. 2); asterisk indicate location of single pseudo-observation.

model (eq. (8)) with the horizontal correlations expressed by a power-law function (e.g. DAO, 1996):

$$\rho(r_{ij}) = \left[1 + \frac{1}{2} \left(\frac{r_{ij}}{L} \right)^2 \right]^{-1}. \tag{11}$$

In this equation, L is the length scale as defined in Daley (1991) governing the horizontal extent of the correlation.

In the first example, the observation vector y consists of a single observation taken at (54 N, 330 E), indicated by the asterisk in Fig. 3. Taking advantage of our knowledge of the true state, we can explore the parameter space of L in order to find the value that yields the best analysis based on this particular observation. In Fig. 4, the RMS analysis error is shown as a function of the length scale L (in units of analysis grid cells) in eq. (11). We see that the optimal correlation length for this case is about 2 grid cells, or 450 km. This seems physically reasonable, if somewhat on the low side for a total ozone field. The minimum RMS analysis error is 18.8 DU.

In Fig. 5 we show the analysis increment obtained using a correlation length of 2 grid cells. Comparing with fig. 3, we can verify that the increment has roughly the correct value in the neighborhood of the observation point. On the

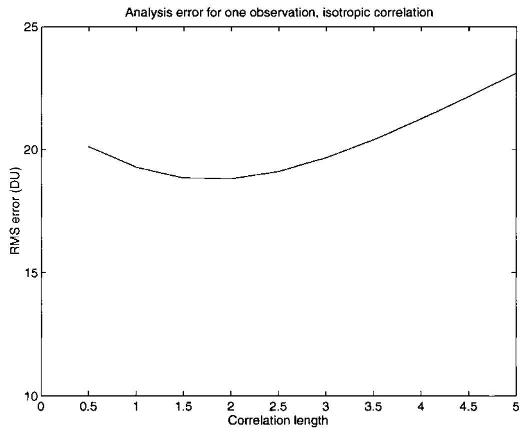

Fig. 4. RMS error for analysis based on a single observation using isotropic background error correlation; the error is shown as function of the correlation length scale.

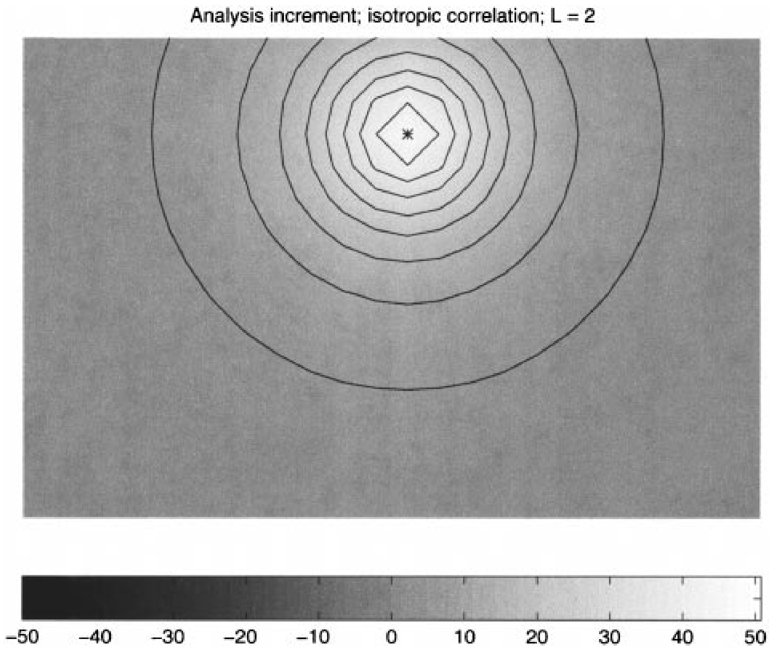

Fig. 5. Isotropic analysis increment (in Dobson Units) corresponding to the minimum RMS analysis error of Fig. 4; compare with error shown in Fig. 3.

other hand it is clear that the presence of nearby regions with background errors of the opposite sign both north and south of the observation point tends to keep the increment more localized than might be desired in the east-west direction. We now solve exactly the same problem using the simple anisotropic covariance model defined by eq. (10). As in the example above we use a power-law function for the horizontal correlation (eq. (8)), and we use a Gaussian for the field term (9):

$$v(x_i, x_j) = \exp - \left(\frac{x_i - x_j}{T} \right)^2 \qquad (12)$$

In this equation T is a length scale in the space of the field ($T/\sqrt{2}$ in eq. (12) corresponds to the correlation length L in eq. (11)), and it thus controls the degree of anisotropy. It is worth pointing out here that the choice of a power-law function for the horizontal correlation and a Gaussian for the field term is arbitrary. Apart from computational efficiency and a possible fit to data (if data are indeed available) the only real restriction that applies to these functions is that they should define a valid covariance matrix on the particular model domain on which they are evaluated. The forms chosen here meet this requirement.

As for the isotropic case we scan the parameter space of the correlation model in order to find the set of (L, T) that yields the best analysis. In Fig. 6 we show the RMS analysis error as a function of L and T. The minimum value is found to be 18.0 DU for $L = 2.5$ and $T = 40$ DU. Comparing with the background error of 20.9 DU, and the isotropic analysis error of 18.8 DU we see that the anisotropic correlation yields an improvement over the isotropic correlation in terms of error reduction of almost 40% for this particular case.

We emphasize that the optimal set of parameters will depend on the field that is analyzed, the functional forms of ρ and v used in eq. (10), and on the physical distribution of the observation network. Not surprisingly the optimal length scale for the physical space component ρ (in the anisotropic case L is no longer strictly speaking a correlation length) is somewhat longer here than for the isotropic case. We noted above that the

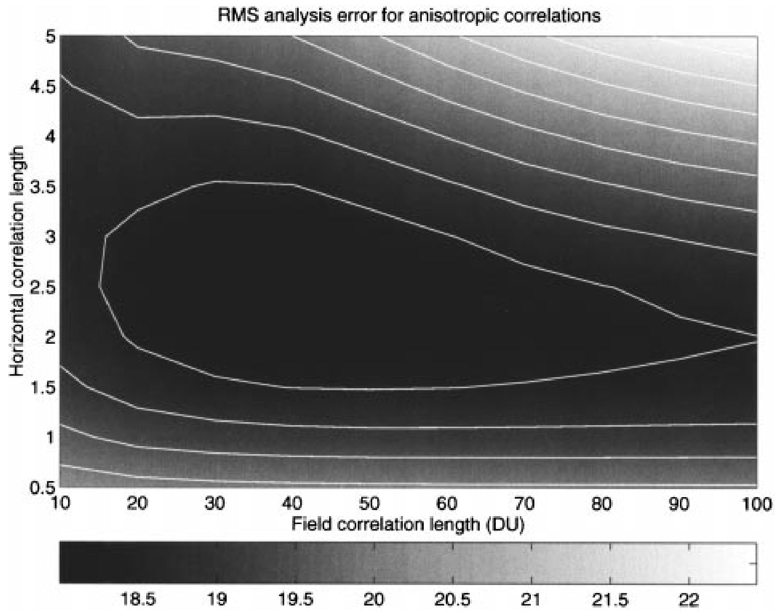

Fig. 6. RMS error (in Dobson Units) for analysis based on a single observation using anisotropic background error correlation; the error is shown as function of the two correlation length scales L (vertical axis) and T (horizontal axis).

isotropic correlation length results as a compromise between a broad correlation in the zonal direction and a shorter correlation in the meridional direction, particularly north of the observation. In Fig. 7, which shows the increment calculated from the optimal set of parameters for the anisotropic correlation, we see that the background field successfully keeps the increment out of regions in which the error has the opposite sign, while allowing it to spread out zonally as it should according to the background error (Fig. 3).

As a second example we show the analysis of the ozone field from the full set of pseudo-observations (indicated by asterisks in Fig. 1).

As for the single observation case discussed above, we start by showing the analysis error as a function of isotropic correlation length in Fig. 8. We see that the optimal correlation length is about 4 grid cells, and that the minimum RMS analysis error is 12.8 DU. This correlation length is supposedly more representative than the somewhat low value we found for a single observation. At any rate, it is closer to what one would intuitively expect to be optimal for this particular observation network, simply from expecting the impact of the observations to be felt to some extent throughout the data void between the orbits.

The corresponding analysis increment is shown in Fig. 9. Because of the increased number of observations the fact that this field is just a linear combination of circular increments like the one shown in Fig. 5 is now less evident. Note, however, that the SSW/NNE orientation of the patch of negative background error emanating from the lower left corner of Fig. 3, as well as the E/W oriented patch of negative error at the upper rim of the plot are both difficult to discern in the increment. This is due to the fact that neither of these features is aligned with the track of the satellite orbit defining the sampling pattern. This particular observation network thus has a highly anisotropic horizontal resolution, and in the direction in which the resolution is low the information is not effectively used under the assumption of isotropic correlation.

In Fig. 10, we show the RMS analysis error as a function of L and T using the anisotropic correlation model. The minimum value of 10.7 DU is found for $L=6$ and $T=140$ DU. The relative

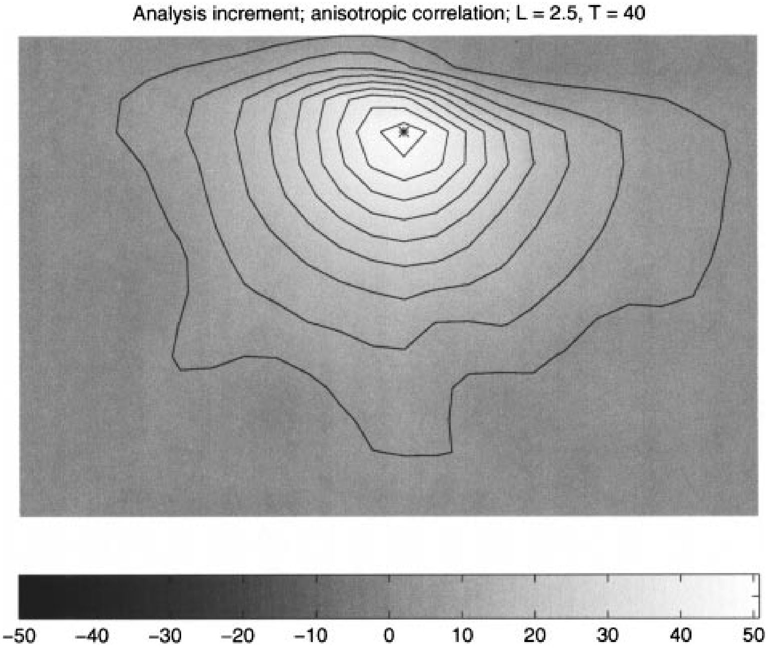

Fig. 7. Anisotropic analysis increment (in Dobson Units) corresponding to the minimum RMS analysis error of Fig. 6; compare with error shown in Fig. 3.

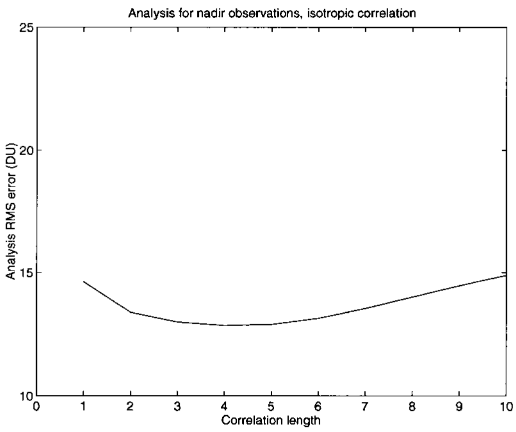

Fig. 8. As Fig. 4, but for the full set of nadir observations indicated by asterisks in Fig. 1.

performance gain with respect to the isotropic analysis is thus less impressive here than for the single observation case. On the other hand the analysis error is now close to the RMS observation error (10.3 DU), and bearing in mind that the field is undersampled by the observations, the performance of the analysis may be nearly optimal in this case.

The corresponding increment is shown in Fig. 11. Comparing with Figs. 3 and 9, we see a much better coherence of the features that cross the direction defined by the orbits as well as a better orientation of the tongue of negative values in the SE corner of the domain. The effect of the anisotropy is clearly seen even though the length scale of the field component almost equals the range of field values found within the domain. It is noteworthy that even such a moderate amount of anisotropy allows us to increase the length scale of ρ by 50%, while the same time yielding the distinct qualitative improvement seen when comparing Figs. 9 and 11. The anisotropy in the error correlation thus in a sense transforms the effective resolution of the satellite sampling pattern into being nearly isotropic.

4. Anisotropic covariances for non-linear three-dimensional models

In this section we discuss possible ways of generalizing the simple method for generating

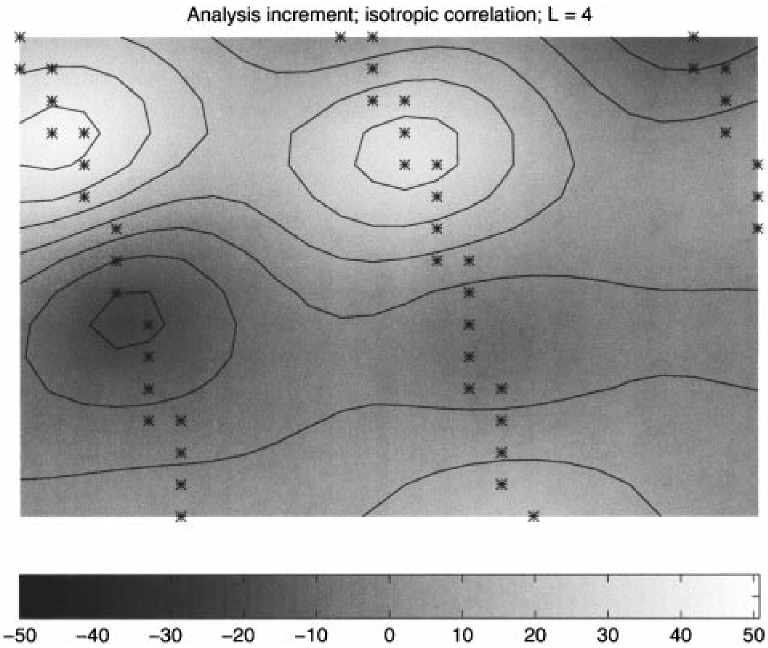

Fig. 9. As Fig. 5, but for the full set of nadir observations indicated by asterisks in Fig. 1.

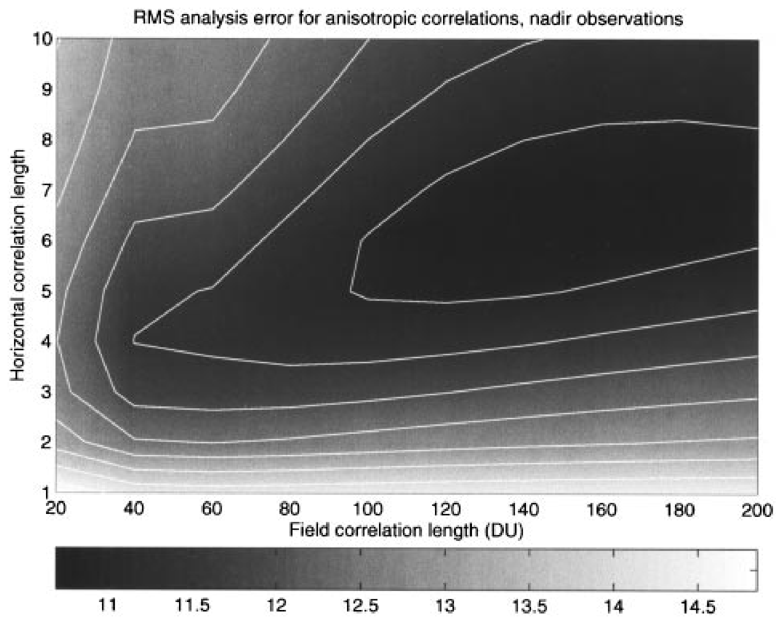

Fig. 10. As Fig. 6, but for the full set of nadir observations indicated by asterisks in Fig. 1.

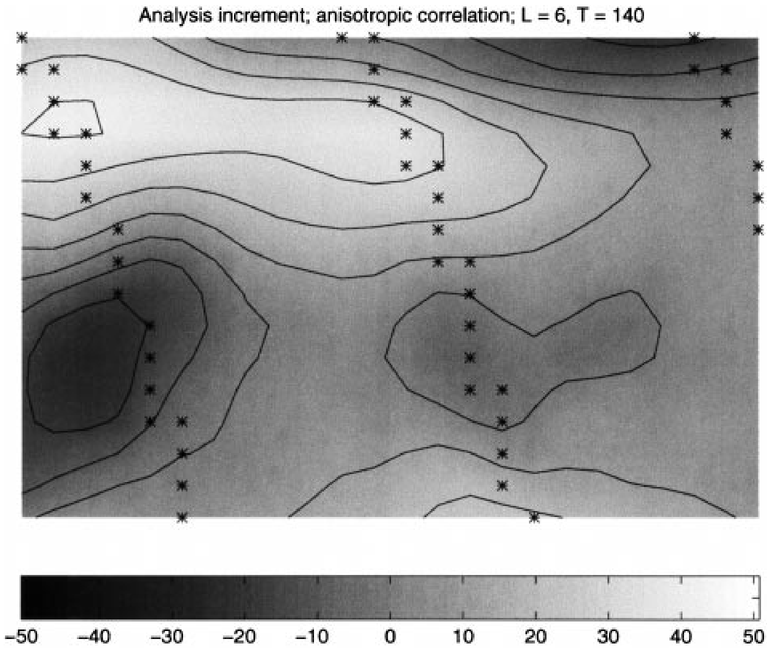

Fig. 11. As Fig.7, but for the full set of nadir observations indicated by asterisks in Fig. 1.

anisotropic covariances introduced in Section 1, so that it may be used for three-dimensional non-linear models. Consider a general meteorological model where the state variable x is a multivariate function (wind, temperature, humidity, surface pressure) of space and time, and the forward operator in time is non-linear. We again apply the notion that the background estimate of a conserved quantity represents most of what we know about the deformation history of the flow, and this leads us to explore whether it is possible to model the error covariance of x using (10).

For passive variables this may be feasible by analogy with the ozone example described above. Because of the Lagrangian character of the approach, however, errors in the dynamical variables may be more appropriately expressed in terms of some conservative function of the state rather than directly in terms of the state itself. Note that by allowing a (possibly multivariate) function of the field variable to appear in the last term of (10) rather than just the field itself, we get the more general form:

$$B_{ij}=\sigma_i\sigma_j\rho(r_{ij})v(\|f(x_i^{\rm b})-f(x_j^{\rm b})\|)\,. \tag{13}$$

B is still a legitimate background error covariance matrix provided that v is a correlation function on R^k, where k is the number of components in f. Potential vorticity (PV) and potential temperature (θ) are both simple functions of x that are conserved under certain assumptions about the flow on time-scales of up to about a week throughout much of the domain, and either one (or both simultaneously) are thus candidate quantities for being used in a covariance model based on (13). Before primitive equation models became economically tractable, barotropic vorticity equation models were widely used for forecasting the large-scale flow in the free troposphere, and in this altitude region (where potential temperature and potential vorticity are of limited use) the vorticity field might therefore provide a useful constraint on the background error covariance.

Let us now consider the problem of three-dimensional error covariances. It turns out that there is an interesting parallel between the covariance model of eq. (10) and the separable isotropic model commonly used for describing three-dimensional univariate error covariances in present meteorological analysis schemes (DAO, 1996):

$$B_{ij}^{(mn)}=\sigma_i^m\sigma_j^n\rho(r_{ij})v^{(mn)}\,. \tag{14}$$

Here m and n indicate the vertical levels, and $v^{(mn)}$ are the vertical correlation coefficients. It is easy to see that if the field x is set to be the vertical coordinate, eqs. (10) and (14) are of the same form, except for the fact that (10) is continuous in the third coordinate, whereas (14) is discrete.

A multi-level univariate non-separable covariance model which is more general than (14) is given by (DAO, 1996):

$$B_{ij}^{(mn)} = \sigma_i^m \sigma_j^n \rho^{(mn)}(r_{ij}) v^{(mn)}, \tag{15}$$

where m and n are level indices, $\rho^{(mn)}(r_{ij})$ are cross-correlation functions, and $v^{(mn)}$ are vertical correlation coefficients. If the horizontal cross-correlations are the same at all levels it reduces to the separable model (eq. (14)). This equation is cumbersome to implement because of the amount of input information that is needed in terms of the cross-correlation functions and vertical correlation coefficients. Although it does eliminate the assumption of separability, even eq. (15) is not completely general, since the correlation between a pair of points (i, m) and (j, n) will be identical to the correlation between (i', m) and (j', n), provided that $r_{ij} = r_{i'j'}$.

It may be useful at this point to examine why it is common practice in covariance modeling to distinguish between the vertical and the horizontal directions as in eqs. (15) and (14). One reason obviously has to do with the desire to reduce the mathematical and computational complexity of the problem. This is indeed a major motivation for the separable model. Another important reason, however, is that the large aspect ratio (10^2–10^3) typically found in atmospheric flows makes a fully general three-dimensional approach impractical. In building a separable model we thus implicitly assume the aspect ratio to be constant, but this, on the other hand, seems excessively restrictive. Even the more general form (15) contains unphysical limitations, as pointed out above. In reality the aspect ratio is a local attribute of the flow, and the error covariance model should respect this fact as well as the horizontal anisotropy.

We now point out a remarkable consequence of the analogy between our two-dimensional anisotropic model (13) and the three-dimensional separable model (eq. (14)). If f in (13) is a monotonic function of height or pressure, this equation can actually be used for constructing three-dimensional flow-dependent error covariances without the explicit use of vertical correlation coefficients. As an illustration, consider the following expression for the covariance between a given quantity at horizontal grid-point i, level m and horizontal grid-point j, level n, in which potential temperature has been substituted for f:

$$B_{ij}^{(mn)} = \sigma_i^{(m)} \sigma_j^{(n)} \rho(r_{ij}^{mn}) v(\| \theta(x_i^{b,(m)}) - \theta(x_j^{b,(n)}) \|). \tag{16}$$

The distance r_{ij}^{mn} is here taken to be the Euclidean (rather than merely the horizontal) distance, and this formulation therefore does not assume separability between the horizontal and vertical directions. Instead, control of the aspect ratio of the correlations is left to the potential temperature field that also controls the horizontal anisotropy. This is a natural choice since the aspect ratio of the flow has a very particular value in isentropic coordinates: It is infinite under adiabatic conditions. In other words there is no motion along the third coordinate, and the amount of vertical exchange within the flow (which at least intuitively is what we would want the vertical correlations to reflect) is instead tied to the vertical gradient of θ. A model of the form of (16) is conceptually less complicated than (15). It is fully flow-dependent, since it allows the increments, or correlations, to extend in all three dimensions without any restrictions on their shape other than what is imposed on them by the flow itself.

The implicit assumption behind eq. (16) is that error correlations are closer to being isotropic in isentropic coordinates than in height or pressure coordinates. While it remains unproven, Benjamin (1989) cites a number of studies that seem to support this hypothesis. Note that although eq. (16) is given in terms of the potential temperature, its use is not limited to analysis on isentropic levels along the lines of Benjamin (1989). It can be applied in any vertical coordinate system, and using it for analyzing on, say, isobaric levels would provide a way of reaping most of the benefits of an isentropic formulation of the analysis system without changing the vertical coordinate.

5. Summary and conclusions

In this article we have presented a new method for specifying the correlation part of the background error covariances needed for atmospheric data analysis. In our model the correlations are expressed in terms of horizontal distances and differences between field values. The model should be easy to implement in most physical space analysis systems (Cohn et al., 1997), and it is inexpensive to evaluate. In the limit where the correlation length goes to infinity in the space of the field, the model reduces to the now commonly used isotropic model, and it should therefore be possible to implement it in an operational system gradually without causing major shocks to the system. Although the model is fully flow-dependent, it has very few free parameters once an exact functional form has been chosen, and it is therefore well suited for rigorous tuning methods.

Some aspects of the potential impact of the anisotropic correlation model were illustrated by means of a simple example, in which we analyzed the total ozone field over a limited domain using sparse observations. We first demonstrated the fundamental difference in character between isotropic and anisotropic correlations by calculating the increment due to a single observation under the two assumptions. The anisotropic increment yielded a better analysis by objective skill, and it eliminated some features that were deemed unphysical in the isotropic increment. Using data from a network corresponding to nadir observations from a satellite in a polar orbit it was shown that the RMS error of the analysis approached that of the observations when the anisotropic formulation was used. For this particular observation network that level of skill could not be reached when isotropic correlations were used.

We then proceeded to discuss the implications of the new correlation model for general meteorological analysis systems. It was shown that the model is easy to generalize to incorporate non-trivial functions of the flow field, and that for a two-dimensional field it is formally equivalent to a three-dimensional separable model. We finally showed that for some particular choices of the functional form used in the field term of the correlation, the model leads to a potentially useful class of three-dimensionally flow-dependent structure functions at no additional cost in terms of mathematical or conceptual complexity.

To a large extent the new correlation model has been developed and is justified in terms of physical reasoning rather than mathematical rigor. Where possible we have sought to specify the assumptions invoked in constructing it. There remains, however, a significant amount of work to be done in examining and reinterpreting the model proposed here in terms of other known approaches to covariance modeling. This will hopefully provide some guidance both with respect to the choice of the functional forms, and at a later stage with respect to the actual tuning of the model. In addition to the simple analysis example reported here, work is underway to test the model in a more realistic meteorological context, and results of these tests will be reported in due course.

Finally, we note that the univariate correlation model proposed here addresses but one aspect of the problem of covariance modeling for data assimilation. Models for both the error variances and the multivariate correlations for the linear transport problem are currently under development, and we hope that generalizations to multivariate correlations for non-linear models can be found.

6. Acknowledgements

The author wishes to thank his colleagues at the Goddard Space Flight Center for numerous presentations and informal discussions that lead to the work presented here. In particular he wishes to thank Dick Dee for sharing his thoughts on covariance modeling and for constructive criticism of earlier versions of the manuscript for this article. Greg Gaspari provided insight on the mathematical definition of covariance models for different model domains, and Ivanka Stajner supplied the initial proof that the model proposed here fits the standard definition. Steve Cohn and Ricky Rood both provided useful comments on parts of the manuscript. Roger Daley of the Naval Research Laboratory in Monterey is thanked for his commentary, including the suggestion of the Benjamin (1989) reference on isentropic analysis. Financial support for this work was obtained through the Cooperative Agreement (NCC 592) between University Maryland Baltimore County and NASA Goddard Space Flight Center.

REFERENCES

Allaart, M. F., Kelder, H. and Heijboer, L. C. 1993. On the relation between ozone and potential vorticity, *Geophys. Res. Lett.* **20**, 811–814.

Atkins, M. J. 1974. The objective analysis of relative humidity. *Tellus* **26**, 663–671.

Benjamin, S. G. 1989. An isentropic mesoα-scale analysis system and its sensitivity to aircraft and surface observations. *Mon. Wea. Rev.* **117**, 1586–1603.

Benjamin, S. G. and Seaman, N. L. 1985. A simple scheme for objective analysis in curved flow. *Mon. Wea. Rev.* **113**, 1184–1198.

Benjamin, S. G., Brewster, K. A., Brümmer, R., Jewett, B. F., Schlatter, T. W., Smith, T. L. and Stamus, P. A. 1991. An isentropic three-hourly data assimilation system using ACARS aircraft observations. *Mon. Wea. Rev.* **119**, 888–906.

Cohn, S. E. 1993. Dynamics of short-term univariate forecast error covariances. *Mon. Wea. Rev.* **121**, 3123–3149.

Cohn, S. E. 1997. An introduction to estimation theory. *J. Meteorol. Soc. Jap.* **1B**, 257–288.

Cohn, S. E. and Dee, D. 1988. Observability of discretized partial differential equations. *SIAM J. Numer. Anal.* **25**, 586–616.

Cohn, S. E., Da Silva, A., Guo, J., Sienkiewicz, M. and Lamich, D. 1997. Assessing the effects of data selection with the DAO physical-space statistical analysis system. DAO Office Note 97–08. Data Assimilation Office, NASA Goddard Space Flight Center.

DAO, 1996. Algorithm theoretical baseline document, version 1.01. Data Assimilation Office, NASA Goddard Space Flight Center.

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

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

Desroziers, G. 1997. A coordinate change for data assimilation in spherical geometry of frontal structures. *Mon. Wea. Rev.* **125**, 582–600.

Desroziers, G. and Lafore, J.-P. 1993. A coordinate transformation for objective frontal analysis. *Mon. Wea. Rev.* **121**, 1531–1553.

Evensen, G. 1994. Inverse methods and data assimilation in nonlinear ocean models. *Physica D* **77**, 108–129.

Gaspari, G. and Cohn, S. E. *Construction of correlation functions in two and three dimensions.* DAO Office Note 96–03, Data Assimilation Office, NASA Goddard Space Flight Center.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45.

Levelt, P., Allaart, M. A. F. and Kelder, H. 1996. On the assimilation of total-ozone satellite data. Ann. Geophys., **14**, 1111–1118.

McPeters et al. 1996. *Nimbus-7 total ozone mapping spectrometer (TOMS) data products user's guide.* NASA Reference Publications 1384.

Riishøjgaard, L. P., Lou, G. P., Menard, R. and Cohn, S. E. 1997. A study of the relationship between GEOS-DAS PV and TOMS total ozone from February 1992. *Proceedings of the XVIII Quadrennial Ozone Symposium.*

Stanford, J. L. and Ziemke, J. R. 1996. A practical method for predicting midlatitude total column ozone from operational forecast temperature fields. *J. Geophys. Res.* **101**, 28 769–28 773.