Tellus
SERIES B
CHEMICAL
AND PHYSICAL
METEOROLOGY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2008), 60B, 753–770
Printed in Singapore. All rights reserved

TELLUS

Variational data analysis of aerosol species in a regional CTM: background error covariance constraint and aerosol optical observation operators

By MICHAEL KAHNERT*, *Swedish Meteorological and Hydrological Institute, 601 76 Norrköping, Sweden*

(Manuscript received 6 May 2008; in final form 19 August 2008)

ABSTRACT

A multivariate variational data assimilation scheme for the Multiple-scale Atmospheric Transport and CHemistry (MATCH) model is presented and tested. A spectral, non-separable approach is chosen for modelling the background error constraints. Three different methods are employed for estimating background error covariances, and their analysis performances are compared. Observation operators for aerosol optical parameters are presented for externally mixed particles. The assimilation algorithm is tested in conjunction with different background error covariance matrices by analysing lidar observations of aerosol backscattering coefficient. The assimilation algorithm is shown to produce analysis increments that are consistent with the applied background error statistics. Secondary aerosol species show no signs of chemical relaxation processes in sequential assimilation of lidar observations, thus indicating that the data analysis result is well balanced. However, both primary and secondary aerosol species display emission- and advection-induced relaxations.

1. Introduction

Data assimilation methods can be employed for constraining the initial state of a model, for interpolating between observations sparsely distributed in time or space and for solving the problem of inverting indirect observations of the system. The basic idea is to join information provided by observations and by some background estimate of the state of the system, for example, a climatological field or a model result. This is performed by use of additional information about the error variances and covariances of the observations and the background information, thus ensuring that the analysed state represents the most likely estimate of the true state of the system. Information from observations of quantities that are indirectly related to the variables in the configuration space of the system are propagated to the state variables via the observation operator and the background error covariances.

One of the most widely used assimilation approaches is the variational method (Talagrand and Courtier, 1987). In numerical weather prediction (NWP) models, it has been employed operationally both in its three-dimensional (3DVAR) (Courtier et al., 1998; Gustafsson et al., 2001; Lindskog et al., 2001; Polavarapu et al., 2005) and in its more advanced four-dimensional (4DVAR) formulation (Saunders et al., 1998). The 3DVAR approach is a sequential data assimilation method in which observations at discrete analysis times are assimilated. The result of the analysis at the analysis time is fed into the model as a new initial state for the continued time-integration of the model. The 4DVAR approach, by contrast, is a continuous assimilation method, in which a whole time-series of observations within a certain time-window is simultaneously assimilated. The governing equations of the model constitute a strong constraint in 4DVAR.

Application of data assimilation to a chemical transport model (CTM) is complicated by the fact that the number of free variables in a CTM is usually much larger than in an NWP. Several applications of the variational method in CTMs have been reported (Robertson and Langner, 1992; Fisher and Lary, 1995; Elbern et al., 1997; Elbern and Schmidt, 1999; Elbern et al., 2000; Elbern and Schmidt, 2001; Wang et al., 2001; Elbern et al., 2007), of which most were focused on assimilating reactive gases. However, 4DVAR applications to aerosol mixing ratios (Benedetti and Fisher, 2007) and adjoint modelling of aerosol dynamics has also been reported (Sandu et al., 2005).

Another approach for solving the data assimilation problem is based on the Kalman filter technique (Lorenc, 1986), which is a sequential approach. By contrast to the variational method, the Kalman filter algorithm includes computation of time-dependent

*Correspondence.
e-mail: michael.kahnert@smhi.se
DOI: 10.1111/j.1600-0889.2008.00377.x

error covariance forecasts. In practice, the high degree of complexity of the method requires certain simplifications.

In conjunction with CTMs, methods based on the Kalman filter have been employed to assimilate both gas species (Levelt et al., 1998; Khattatov et al., 2000; Ménard and Chang, 2000) and aerosols (Rasch et al., 2001; Collins et al., 2001). A particularly versatile simplified assimilation algorithm based on this method is the Ensemble-Kalman filter (EnKF) (Evensen, 2003). Applications of the EnKF to chemical data assimilation have, for example, been reported by van Loon et al. (2000) and Constantinescu et al. (2007a,b).

Hybrids of the variational and the EnKF methods have also been reported (Hamill and Snyder, 2000). A detailed discussion of advantages and drawbacks of both approaches can be found in the recent literature (Gustafsson, 2007; Kalnay et al., 2007a,b).

Assimilation of chemical species in the aerosol phase has not quite reached the same degree of maturity as assimilation of gas components. A description of aerosols needs to address additional aspects that are of no concern for gas components, such as the mixing state (external or internal) and the size distribution of the chemical species. Information about the mixing state and the shape of the particles is usually not provided by the model and often difficult to extract from the observations. However, this information is needed for construction of aerosol optical observation operators. Thus, the formulation of the observation operators often needs to rely on assumptions that can introduce additional uncertainties and errors.

In this paper, we mainly address the formulation of the background error constraints and of the observation operators for aerosol species in a variational data analysis algorithm. The CTM we employ is the Multiple-scale Atmospheric Transport and CHemistry (MATCH) model (Robertson et al., 1999; Foltescu et al., 2005; Andersson et al., 2007). We adopt a spectral formulation of the variational method (Berre, 2000) that largely follows the corresponding formulation in the HIgh Resolution Limited-Area Model (HIRLAM) weather forecasting system (Gustafsson et al., 2001). For the background error covariance modelling, we choose an ansatz of error correlations that are non-separable in the vertical and horizontal coordinates (Derber and Bouttier, 1999; Berre, 2000). Error correlations are assumed to be horizontally homogeneous and isotropic, but error covariances are not required to have these symmetry properties. Aerosol optical observation operators are presented for external mixtures of aerosol species. Three different methods for modelling background error covariances are tested, and the differences in the resulting background error statistics are discussed. The properties of the data analysis algorithm in conjunction with the three different background error covariance models are studied by an assimilation experiment, in which we analyse lidar observations of the aerosol backscattering coefficient. Further, the properties of sequential 3DVAR assimilation in an aerosol chemistry model are investigated.

2. Spectral formulation of three-dimensional variational data analysis

Producing an accurate estimate of the chemical state of the atmosphere by use of observations is usually an under-determined problem, since observational data are typically sparse in space–time. Further, the observations are often indirectly related to the model state variables, such as in case of radiative measurements. As a result, the problem of retrieving the chemical state of the atmosphere from observations is often an ill-posed inverse problem. To make the problem well-posed, one can supplement observations by background information, which one can obtain from a model estimate.

Let $\mathbf{x} \in I\!R^n$ denote a vector of state variables in a discretized model space, let $\mathbf{x}_\mathrm{b} \in I\!R^n$ be a background estimate of the chemical state of the atmosphere in that space, and let $\mathbf{y} \in I\!R^m$ represent a vector of observations. If the errors $\epsilon_\mathrm{b} \in I\!R^n$ and $\epsilon_\mathrm{o} \in I\!R^m$ of the background estimate and the observations are independent, then their probability density function P is separable, that is,

$$P(\epsilon_\mathrm{b}, \epsilon_\mathrm{o}) = P_\mathrm{b}(\epsilon_\mathrm{b}) P_\mathrm{o}(\epsilon_\mathrm{o}) = \exp\left[\log P_\mathrm{b}(\epsilon_\mathrm{b}) + \log P_\mathrm{o}(\epsilon_\mathrm{o})\right]. \quad (1)$$

Assuming that the best estimate of the true chemical state of the atmosphere is given by a maximum likelihood solution, the solution of the data analysis problem is found by minimizing the quantity

$$J = -\log P_\mathrm{b} - \log P_\mathrm{o}. \quad (2)$$

J is called the cost function. If the background and observation errors are normally distributed, then one can write

$$P_\mathrm{b}(\mathbf{x}) = (2\pi\,|\mathbf{B}|)^{-1/2} \times \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}_\mathrm{b})^T \cdot \mathbf{B}^{-1} \cdot (\mathbf{x}-\mathbf{x}_b)\right], \quad (3)$$

$$P_\mathrm{o}(\mathbf{x}) = (2\pi\,|\mathbf{R}|)^{-1/2} \times \exp\left[-\frac{1}{2}(\hat{H}(\mathbf{x})-\mathbf{y})^T \cdot \mathbf{R}^{-1} \cdot (\hat{H}(\mathbf{x})-\mathbf{y})\right], \quad (4)$$

where $\mathbf{B}$ and $\mathbf{R}$ denote the background and observation error covariance matrices, $|\mathbf{B}|$ and $|\mathbf{R}|$ denote their determinants, and where $\hat{H}: I\!R^n \longrightarrow I\!R^m$ represents an operator mapping from model to observation space. Substitution of eqs. (3) and (4) into (2) yields

$$\begin{aligned} J[\mathbf{x}] &= J_\mathrm{b} + J_\mathrm{o} \\ &= \frac{1}{2}(\mathbf{x}-\mathbf{x}_\mathrm{b})^T \cdot \mathbf{B}^{-1} \cdot (\mathbf{x}-\mathbf{x}_\mathrm{b}) \\ &\quad + \frac{1}{2}(\hat{H}(\mathbf{x})-\mathbf{y})^T \cdot \mathbf{R}^{-1} \cdot (\hat{H}(\mathbf{x})-\mathbf{y}), \end{aligned} \quad (5)$$

where we have neglected constant terms that are irrelevant for the minimization procedure. One can see that the assumption of normally distributed background errors implies that the background term J_b in the cost function is a quadratic form. The

observation term J_o is a quadratic form only if the operator $\hat{H}$ is linear.

We adopt an incremental formulation (Courtier et al., 1994) based on the analysis increment $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$. Assuming that the norm of $\delta\mathbf{x}$ is small, and assuming that $\hat{H}$ is not a highly non-linear operator, we can use a first-order approximation of the $\hat{H}$-operator according to

$$J[\delta\mathbf{x}] = \frac{1}{2}\delta\mathbf{x}^T \cdot \mathbf{B}^{-1} \cdot \delta\mathbf{x} + \frac{1}{2}(\hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x} - \mathbf{y})^T \cdot \mathbf{R}^{-1} \times (\hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x} - \mathbf{y}), \quad (6)$$

where $\mathbf{H}$ represents the Jacobian of $\hat{H}$. By further assuming that $\delta\mathbf{x}$ is smoothly varying with horizontal position, it is usually possible to perform the variational data analysis at a lower horizontal resolution than the model calculations.

Given an ensemble of background errors ϵ_b, one can compute the background error covariance matrix according to

$$\mathbf{B} = \left\langle (\epsilon_b - \langle\epsilon_b\rangle) \circ (\epsilon_b - \langle\epsilon_b\rangle)^T \right\rangle, \quad (7)$$

where $\mathbf{a} \circ \mathbf{b}$ denotes the dyadic product of $\mathbf{a}$ and $\mathbf{b}$, and where $\langle \cdots \rangle$ denotes an ensemble-average. For a CTM with N_x longitudinal and N_y latitudinal gridpoints, N_z vertical levels, and N_c chemical compounds, the background error covariance matrix is typically of dimension $N_x N_y N_z N_c \sim 10^7$. This makes the minimization of the cost function given in eq. (6) prohibitively complex, not just in terms of computation time, but also with regard to numerical stability. Note that computation of the cost function requires inversion of the covariance matrices. One, therefore, needs to introduce simplifying assumptions to reduce the dimensionality of the problem. This can be achieved by making certain symmetry assumptions, such as homogeneity and isotropy in the horizontal dimensions. If the Lie group corresponding to the assumed symmetry properties is a simple or semi-simple Lie group of rank l, then there exist l invariant operators, the eigenvectors of which form a basis which subdivides the vector space into invariant, irreducible subspaces (e.g. Greiner and Müller, 1994). The idea behind such a choice of basis functions is to block-diagonalize the background error covariance matrix, and to handle each block matrix separately in each of the irreducible subspaces, which have a much lower dimension than the entire vector space. In case of homogeneity and isotropy of space, the corresponding Lie group is the translation–rotation group, which is neither simple nor semi-simple. However, this group contains both the translation and the rotation groups as invariant subgroups, both of which are semi-simple. Under the assumption of homogeneity and isotropy of space, one can therefore choose whether to use the appropriate eigenfunctions of the invariant operators of the translation or of the rotation group as an expansion basis. The former are harmonic functions (sines and cosines), the latter are spherical harmonics. For a global model, it is advantageous to use spherical harmonics, whereas for a limited-area model, a decomposition into harmonic functions (Fourier transformation) is best adapted to the geometry of the model domain. For our limited-area model, we shall therefore base our background error covariance modelling on a spectral (Fourier) formulation with respect to the horizontal coordinates.

Another simplification can be achieved by assuming separability of horizontal and vertical correlations, see for example, Constantinescu et al. (2007b). This assumption implies that the spatial correlations can be decomposed into a product of height-independent horizontal correlation functions and vertical correlations functions independent of horizontal position. We shall in the following 'not' invoke this assumption, but instead model non-separable background error covariances. Non-separability of error covariances have, for example, been discussed by Berre (2000), Derber and Bouttier (1999).

To be more specific, the general strategy outlined above can be implemented by the following steps.

(1) As a first step, one obtains an ensemble of model states, for example, a time-series $x_{i,j,l,k}(t)$, where $i = 1, \ldots, N_x$, $j = 1, \ldots, N_y$ $l = 1, \ldots, N_z$, $k = 1, \ldots, N_c$ denote longitudinal, latitudinal, vertical and chemical indices, respectively. Next, one obtains a perturbed ensemble of model states $x'_{i,j,l,k}(t)$. Different perturbation approaches will be discussed in Section 4.

A bi-Fourier transformation in the horizontal coordinates is applied. This application, however, requires the input field to be bi-periodic in the horizontal coordinates. When enforcing the bi-periodicity one wants to ensure that the field and its partial derivatives are smoothly varying functions. To this end a horizontal extension zone is introduced, that is, the x and y components of the field are expanded to $N_x^{ex} > N_x$ and $N_y^{ex} > N_y$ gridpoints. The fields $\mathbf{x}$ and $\mathbf{x}'$ are interpolated into the extension zone by use of a two-dimensional cubic spline by enforcing periodic boundary conditions for the field and its partial derivatives. A critical issue is the size of the extended domain. When later performing an analysis with observations close to the domain boundary, there is a risk that the periodic boundary conditions result in spreading of information out of one domain boundary and into the opposite boundary. To prevent such artefacts in the analysis, the width of the extension zone needs to be larger than the characteristic horizontal length scale of the structure functions.

(2) Next step is to compute errors

$$\epsilon_{i,j,l,k}(t) = x_{i,j,l,k}(t) - x'_{i,j,l,k}(t) \quad (8)$$

and biases

$$\langle\epsilon_{i,j,l,k}\rangle = \langle x_{i,j,l,k} - x'_{i,j,l,k}\rangle. \quad (9)$$

Note that most chemical species have a pronounced diurnal cycle. For this reason it is usually a good idea to compute biases for each time of the day separately, i.e.

$$\langle\epsilon_{i,j,l,k}(t_0)\rangle = \frac{1}{N_{t_0}} \sum_{t \in M_{t_0}} (x_{i,j,l,k}(t) - x'_{i,j,l,k}(t)), \quad (10)$$

where

$$M_{t_0} = \{t \,||t - t_0| = n \cdot 24h, \quad n \in \mathbb{N}_0\} \tag{11}$$

is the equivalence class of all times in the ensemble belonging to the same time of the day, and where $N_{t_0} = |M_{t_0}|$ is the number of elements in that class.

(3) In the next step, one removes the biases from the errors and computes error variances, that is,

$$\tilde{\epsilon}_{i,j,l,k}(t) = \epsilon_{i,j,l,k}(t) - \langle \epsilon_{i,j,l,k}(t_0) \rangle, \qquad t \in M_{t_0}, \tag{12}$$

$$\sigma^2_{i,j,l,k} = \langle \tilde{\epsilon}^2_{i,j,l,k}(t) \rangle_t. \tag{13}$$

(4) Note that the assumption of horizontal homogeneity and isotropy is usually more reasonable for error correlations than for error covariances. Therefore, it is advisable to normalize the background errors with the standard deviations, before invoking the symmetry assumptions and computing error correlations. Thus one defines

$$\eta_{i,j,l,k}(t) = \frac{\tilde{\epsilon}_{i,j,l,k}(t)}{\sigma_{i,j,l,k}}. \tag{14}$$

(5) The next step is to perform a bi-Fourier transformation in the horizontal coordinates from gridpoint to spectral space

$$\hat{\eta}_{m,n,l,k}(t) = \frac{1}{N_x^{\mathrm{ex}} N_y^{\mathrm{ex}}} \times \sum_{i=1}^{N_x^{\mathrm{ex}}} \sum_{j=1}^{N_y^{\mathrm{ex}}} \eta_{i,j,l,k}(t) \exp\left[-2\pi \mathrm{i} \left(\frac{mi}{N_x^{\mathrm{ex}}} + \frac{nj}{N_y^{\mathrm{ex}}}\right)\right]. \tag{15}$$

The indices $m = -K_x, \ldots, K_x$ and $n = -K_y, \ldots, K_y$ denote indices in spectral space conjugate to the horizontal indices i and j in gridpoint space. Note that we denote the imaginary unit by roman face i to avoid confusion with the horizontal index i.

Following Berre (2000), we use an elliptic truncation $(m/N_x^{\mathrm{ex}})^2 + (n/N_y^{\mathrm{ex}})^2 \leq 1$ to ensure a homogeneous and isotropic resolution over the model domain. For brevity, this truncation will be omitted in the notation.

(6) Next step is to compute spectral covariances. As derived in detail by Berre (2000), the assumptions that the covariances are horizontally homogeneous and independent of the orientation of the separation vector in physical space implies the following form for the correlations $\hat{C}_{m,n,l,k;m',n',l',k'}$ in spectral space

$$\begin{aligned}\hat{C}_{m,n,l,k;m',n',l',k'} &= \langle \hat{\eta}_{m,n,l,k}(t)\, \hat{\eta}_{m',n',l',k'}(t) \rangle_t \\ &= \delta_{m,m'} \delta_{n,n'} \frac{1}{N_{t_0}} \sum_{t \in M_{t_0}} [\mathrm{Re}\, \hat{\eta}_{m,n,l,k}(t) \mathrm{Re}\, \hat{\eta}_{m,n,l',k'}(t) \\ &\quad + \mathrm{Im}\, \hat{\eta}_{m,n,l,k}(t) \mathrm{Im}\, \hat{\eta}_{m,n,l',k'}(t)].\end{aligned} \tag{16}$$

Thus, the covariance matrix is diagonal in the spectral indices, that is, the matrix as a whole is block-diagonal.

(7) As mentioned earlier, it was not possible to simultaneously exploit homogeneity and isotropy in the choice of expansion functions, which had to do with the algebraic properties of the translation–rotation group. However, we can make use of the isotropy of space by noting that the error correlations in spectral space only depend on the norm of the wave vector $\mathbf{k}^* = N_s(m/K_x, n/K_y)$ (where N_s is an arbitrary scale factor, which we set to max $\{N_x^{\mathrm{ex}}, N_y^{\mathrm{ex}}\}$). The norm of $\mathbf{k}^*$ is given by the wavenumber $k^* = N_s\sqrt{(m/K_x)^2 + (n/K_y)^2}$. Thus, we can define angularly averaged spectral error correlations

$$\hat{D}_{l,k;l',k'}(k^*) = \frac{1}{2\pi} \int_0^{2\pi} \hat{C}_{m,n,l,k;m,n,l',k'} \mathrm{d}\theta, \tag{17}$$

such that $N_s\sqrt{(m/K_x)^2 + (n/K_y)^2} \in [k^* - \mathrm{d}k^*, k^* + \mathrm{d}k^*]$ and $\tan\theta = [(n/K_y)/(m/K_x)]$, where we used a resolution of $\mathrm{d}k^* = 1.5$ in our calculations. The main benefit of the angular integration in eq. (17), resulting from the isotropy assumption, is a better sampling of the correlation estimates. Note that the above definition of the wavenumber, which is due to Berre (2000), implies that the wavenumber is unitless.

(8) Following Gustafsson et al. (2001), we normalize the $\hat{D}$ matrix by the spectral density of the horizontal error correlations

$$\gamma_{l,k}(k^*) = \frac{\hat{D}_{l,k;l,k}(k^*)}{\sum_{k_0^*=0}^{K_{\max}^*} 2\pi k_0^* \hat{D}_{l,k;l,k}(k_0^*)}, \tag{18}$$

thus defining

$$\hat{B}_{l,k;l',k'}(k^*) = \frac{\hat{D}_{l,k;l',k'}(k^*)}{\sqrt{\gamma_{l,k}(k^*)\gamma_{l',k'}(k^*)}}. \tag{19}$$

(9) Finally, the $\hat{B}$ matrix has to be diagonalized. We emphasize once more that the correlation matrix is block-diagonal, that is,

$$\hat{\mathbf{B}} = \begin{pmatrix} \hat{\mathbf{B}}\left(k_1^*\right) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{B}}\left(k_2^*\right) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{B}}\left(K_{\max}^*\right). \end{pmatrix}. \tag{20}$$

Each block matrix is of dimension $N_z N_c \sim 10^2$–10^3, which is considerably smaller than the dimension of the full background error covariance matrix ($\sim 10^7$). This greatly reduces the CPU time requirements involved in numerical matrix operations. Equally importantly, this substantially alleviates numerical instability problems in the inversion or, equivalently, diagonalization of this matrix as required in the computation of the cost function. The diagonalization of the $\hat{B}$ matrix is achieved by diagonalizing each block matrix separately, that is,

$$\hat{\mathbf{B}}(k^*) = \mathbf{X}(k^*) \cdot \mathbf{\Lambda}(k^*) \cdot \mathbf{X}^T(k^*), \tag{21}$$

where $\mathbf{X}$ (k^*) is an orthogonal matrix containing the eigenvectors of $\hat{\mathbf{B}}(k^*)$, and $\mathbf{\Lambda}(k^*)$ is a diagonal matrix containing the corresponding eigenvalues. Note that the inverse of each block matrix $\hat{\mathbf{B}}(k^*)$ is now readily obtained by $\hat{\mathbf{B}}^{-1}(k^*) = \mathbf{X}(k^*) \cdot \mathbf{\Lambda}^{-1}(k^*) \cdot \mathbf{X}^T(k^*)$. In our calculations, we employed a reduced eigenvalue diagonalization routine. It was ensured that all eigenvalues were positive, which is necessary and sufficient for the correlation matrix to be positive definite.

(10) In summary, we have computed

$$\mathbf{X}\cdot\mathbf{\Lambda}\cdot\mathbf{X}^T=\mathbf{L}\cdot\mathbf{F}\cdot\mathbf{S}^{-1}\cdot\langle\tilde{\epsilon}\circ\tilde{\epsilon}^T\rangle\cdot\mathbf{S}^{-T}\cdot\mathbf{F}^{\dagger}\cdot\mathbf{L}^T, \tag{22}$$

where

$$\mathbf{B}=\langle\tilde{\epsilon}\circ\tilde{\epsilon}^T\rangle \tag{23}$$

is the background error covariance matrix in gridpoint space. The diagonal matrix

$$\mathbf{S}^{-1}=\mathrm{diag}(\sigma_{i,j,l,k}^{-1}) \tag{24}$$

containing the inverse values of the standard deviations [cf. eq. (13)] transforms the covariance matrix into a correlation matrix. The linear transformation $\mathbf{F}$ represents the bi-Fourier transform (see eq. 15). The matrix

$$\mathbf{L}=\mathrm{diag}(\gamma_{l,k}^{-1/2}(k^*)) \tag{25}$$

performs the normalization with the spectral density of the horizontal error correlations (eqs. 18 and 19). The angular averaging (eq. 17) has been omitted from the notation in eq. (22). The matrix $\mathbf{X}$ is a block-diagonal matrix containing all spectral eigenvector matrices $\mathbf{X}(k^*)$. Similarly, the matrix $\mathbf{\Lambda}$ comprises all spectral eigenvalues $\mathbf{\Lambda}(k^*)$.

From eq. (22), we obtain the inverse background error covariance matrix

$$\begin{aligned}\mathbf{B}^{-1}&=\mathbf{S}^{-T}\cdot\mathbf{F}^{\dagger}\cdot\mathbf{L}^T\cdot\mathbf{X}\cdot\mathbf{\Lambda}^{-1/2}\cdot\mathbf{\Lambda}^{-1/2}\cdot\mathbf{X}^T\cdot\mathbf{L}\cdot\mathbf{F}\cdot\mathbf{S}^{-1}\\&=\mathbf{U}^{\dagger}\cdot\mathbf{U},\end{aligned} \tag{26}$$

with the abbreviation

$$\mathbf{U}=\mathbf{\Lambda}^{-1/2}\cdot\mathbf{X}^T\cdot\mathbf{L}\cdot\mathbf{F}\cdot\mathbf{S}^{-1}. \tag{27}$$

Inserting eq. (26) into the expression of the cost function given in eq. (6) and making the variable transformation

$$\chi=\mathbf{U}\cdot\delta\mathbf{x}, \tag{28}$$

one obtains

$$\begin{aligned}J[\chi]=&\frac{1}{2}\chi^{\dagger}\cdot\chi\\&+\frac{1}{2}\left[\hat{H}(\mathbf{x}_b)+\mathbf{H}\cdot\mathbf{U}^{-1}\cdot\chi-\mathbf{y}\right]^T\cdot\mathbf{R}^{-1}\\&\times\left[\hat{H}(\mathbf{x}_\mathrm{b})+\mathbf{H}\cdot\mathbf{U}^{-1}\cdot\chi-\mathbf{y}\right].\end{aligned} \tag{29}$$

Thus with the preconditioning given in eq. (28), the covariance matrix of the background term in the cost function becomes a unit matrix. The minimization of the cost function is performed with a descent algorithm making use of the gradient

$$\nabla_\chi J[\chi]=\chi+\mathbf{U}^{-\dagger}\cdot\mathbf{H}^T\cdot\mathbf{R}^{-1}\cdot\left[\hat{H}(\mathbf{x}_\mathrm{b})+\mathbf{H}\cdot\mathbf{U}^{-1}\cdot\chi-\mathbf{y}\right]. \tag{30}$$

As one can see in eqs. (29) and (30), the variable transformation $\mathbf{U}$ is never applied in its forward form, but only in its inverse form

$$\mathbf{U}^{-1}=\mathbf{S}\cdot\mathbf{F}^{-1}\cdot\mathbf{L}^{-1}\cdot\mathbf{X}\cdot\mathbf{\Lambda}^{1/2} \tag{31}$$

and in its inverse adjoint form

$$\mathbf{U}^{-\dagger}=\mathbf{\Lambda}^{1/2}\cdot\mathbf{X}^T\cdot\mathbf{L}^{-1}\cdot\mathbf{F}^{-\dagger}\cdot\mathbf{S}, \tag{32}$$

where we used the fact that a diagonal matrix is equal to its transpose.

Eqs. (29)–(32) describe the basic three-dimensional, non-separable spectral variational data analysis algorithm employed in conjunction with our CTM. The pre-conditioning method used here closely follows that employed in the NWP model HIRLAM (Gustafsson et al., 2001). The main difference is that HIRLAM contains additional features specific to NWP modelling, such as balancing of the wind fields and multilinear regressions in the formulation of cross-covariances (Berre, 2000).

3. Sequential analysis of unobserved species in spectral space

We now consider the frequently occurring case that the observation operator only acts on a subset of the chemical species in the model. For instance, for observations of the aerosol backscattering coefficient, the corresponding observation operator only acts on the aerosol species, not on the gas species in the model. Let $\mathbf{z}=(\mathbf{x},\mathbf{u})^T\in I\!R^{n_x+n_u}$ denote a state vector in configuration space, let $\mathbf{x}\in I\!R^{n_x}$ represent a vector of observed species and $\mathbf{u}\in I\!R^{n_u}$ a vector of unobserved species. The observation operator is now a map $\hat{H}:I\!R^{n_x}\longrightarrow I\!R^m$ that only operates on $\mathbf{x}$. The background error covariance matrix can be split up accordingly, that is,

$$\mathbf{B}=\begin{pmatrix}\mathbf{B}_{xx}&\mathbf{B}_{xu}\\\mathbf{B}_{ux}&\mathbf{B}_{uu}\end{pmatrix}. \tag{33}$$

It has been shown by Ménard et al. (2004) that the analysis of observed and unobserved species can be done sequentially. The former is done in the usual way by minimizing the cost function

$$\begin{aligned}J[\delta\mathbf{x}]=&\frac{1}{2}\delta\mathbf{x}^T\cdot\mathbf{B}_{xx}^{-1}\cdot\delta\mathbf{x}\\&+\frac{1}{2}(\hat{H}(\mathbf{x}_\mathrm{b})+\mathbf{H}\cdot\delta\mathbf{x}-\mathbf{y})^T\cdot\mathbf{R}^{-1}\\&\times(\hat{H}(\mathbf{x}_\mathrm{b})+\mathbf{H}\cdot\delta\mathbf{x}-\mathbf{y}),\end{aligned} \tag{34}$$

whereas the analysis increment of the unobserved species is given by

$$\delta\mathbf{u}=\mathbf{B}_{ux}\cdot\mathbf{B}_{xx}^{-1}\cdot\delta\mathbf{x}. \tag{35}$$

Analogous to eqs. (26) and (28), we now have

$$\mathbf{B}_{xx}^{-1}=\mathbf{U}^{\dagger}\cdot\mathbf{U}, \tag{36}$$

$$\delta\mathbf{x}=\mathbf{U}^{-1}\cdot\chi, \tag{37}$$

where, according to eq. (27),

$$\mathbf{U}=\mathbf{\Lambda}^{-1/2}\cdot\mathbf{X}^T\cdot\mathbf{L}_x\cdot\mathbf{F}_x\cdot\mathbf{S}_x^{-1}. \tag{38}$$

Here, $\mathbf{\Lambda}$ and $\mathbf{X}$ denote the eigenvalue and eigenvector matrices of $\mathbf{B}_{xx}$, and $\mathbf{S}_x$, $\mathbf{L}_x$ and $\mathbf{F}_x$ are matrices operating on $I\!R^{n_x}$. Substitution of eqs. (36)–(38) into eq. (35) yields

$$\delta\mathbf{u} = \mathbf{B}_{ux} \cdot \mathbf{S}_x^{-1} \cdot \mathbf{F}_x^{\dagger} \cdot \mathbf{L}_x \cdot \mathbf{X} \cdot \mathbf{\Lambda}^{-1/2} \cdot \chi. \tag{39}$$

Although the matrix $\mathbf{B}_{ux}$ does not need to be inverted, it is still advantageous to compute it in spectral space, since this greatly reduces its dimension, thus reducing memory requirements and the number of algebraic operations involved in the matrix multiplications in eq. (39). To this end, we now compute spectral error vectors $\hat{\eta}^{(x)}$ and $\hat{\eta}^{(u)}$ for observed and unobserved species following steps 1–5 in the previous section. With the vector of the spectral errors of the observed species $\hat{\eta}^{(x)}$, one proceeds according to steps 6–10. For instance, in step 6 one computes spectral error correlations

$$\begin{aligned}\hat{C}^{(xx)}_{m,n,l,k;m,n,l',k'} = \frac{1}{N_{t_0}} \sum_{t\in M_{t_0}} \big[&\mathrm{Re}\,\hat{\eta}^{(x)}_{m,n,l,k}(t)\mathrm{Re}\,\hat{\eta}^{(x)}_{m,n,l',k'}(t) \\ &+ \mathrm{Im}\,\hat{\eta}^{(x)}_{m,n,l,k}(t)\mathrm{Im}\,\hat{\eta}^{(x)}_{m,n,l',k'}(t)\big].\end{aligned} \tag{40}$$

Since $\mathbf{B}_{ux}$ does not need to be diagonalized, it is sufficient to proceed with steps 6–7 for the unobserved–observed cross-correlations. Thus one computes in step 6

$$\begin{aligned}\hat{C}^{(ux)}_{m,n,l,k;m,n,l',k'} = \frac{1}{N_{t_0}} \sum_{t\in M_{t_0}} \big[&\mathrm{Re}\,\hat{\eta}^{(u)}_{m,n,l,k}(t)\mathrm{Re}\,\hat{\eta}^{(x)}_{m,n,l',k'}(t) \\ &+ \mathrm{Im}\,\hat{\eta}^{(u)}_{m,n,l,k}(t)\mathrm{Im}\,\hat{\eta}^{(x)}_{m,n,l',k'}(t)\big],\end{aligned} \tag{41}$$

and in step 7

$$\hat{D}^{(ux)}_{l,k;l',k'}(k^*) = \frac{1}{2\pi}\int_0^{2\pi} \hat{C}^{(ux)}_{m,n,l,k;m,n,l',k'}\,\mathrm{d}\theta. \tag{42}$$

In compact matrix notation, we can write

$$\hat{\mathbf{D}}^{(ux)}(k^*) = \mathbf{F}_u \cdot \mathbf{S}_u^{-1} \cdot \left\langle \tilde{\epsilon}^{(u)} \circ \tilde{\epsilon}^{(x)} \right\rangle \cdot \mathbf{S}_x^{-T} \cdot \mathbf{F}_x^{\dagger}, \tag{43}$$

where $\langle \tilde{\epsilon}^{(u)} \circ \tilde{\epsilon}^{(x)} \rangle = \mathbf{B}_{ux}$, and where $\mathbf{F}_u$ and $\mathbf{S}_u$ operate on $I\!R^{n_u}$. Again, the angular averaging in eq. (42) has been omitted from the notation.

Solving eq. (43) for $\mathbf{B}_{ux}$ and substituting the result into eq. (39) finally yields

$$\delta\mathbf{u} = \mathbf{S}_u \cdot \mathbf{F}_u^{-1} \cdot \hat{\mathbf{D}}^{(ux)} \cdot \mathbf{L}_x \cdot \mathbf{X} \cdot \mathbf{\Lambda}^{-1/2} \cdot \chi, \tag{44}$$

where

$$\hat{\mathbf{D}}^{(ux)} = \begin{pmatrix} \hat{\mathbf{D}}^{(ux)}(k_1^*) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{D}}^{(ux)}(k_2^*) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{D}}^{(ux)}(K^*_{\max}) \end{pmatrix}. \tag{45}$$

Equation (44) is the spectral formulation of the analysis of unobserved species. Thus, the analysis of the observed species is performed first by iteratively minimizing the cost function given in eq. (34). Subsequently, the analysis result $\boldsymbol{\chi}$ can be directly substituted into eq. (44) to obtain the analysis increment for unobserved species.

4. Modelling of non-separable background error covariances

The errors ϵ_{b} in the definition of the B matrix in eq. (7) are defined by $\epsilon_{\mathrm{b}} = \mathbf{x}_{\mathrm{b}} - \mathbf{x}_{\mathrm{t}}$, where $\mathbf{x}_{\mathrm{b}}$ denotes the background estimate and $\mathbf{x}_{\mathrm{t}}$ denotes the true state of the system. However, since the true state is unknown one needs to employ appropriate methods for estimating the B matrix. In assimilation systems, one often employs the NMC method (Parrish and Derber, 1992), which is based on computing correlations between forecasts of different forecast lengths. The ensembles-of-analyses method (Fisher, 2003) is a different approach based on analysing the covariances of an ensemble-of-analysis runs with perturbed observations.

A pre-requisite for applying these methods is an existing assimilation system that has to be run with some first estimate of the background error covariance matrix. One relatively simple way of obtaining a first estimate of the background error covariances is based on investigating innovation statistics (Hollingsworth and Lönnberg, 1986). Drawbacks of this method are that one obtains background error statistics in observation space, that the statistics are not global and do not cover all model levels, that one needs a relatively dense, uniform observation network and that the statistics tend to be biased towards data-dense areas. Polavarapu et al. (2005) proposed and applied two alternative methods for estimating background error statistics based on investigating the statistical properties of the model. The starting point is to compute model errors by use of eq. (8) with a suitably defined perturbation $\mathbf{x}'$ of a reference model state $\mathbf{x}$. The first method proposed by Polavarapu et al. (2005) compares the state vector at different times $\mathbf{x}(t)$ to a 'climatological' field $\langle \mathbf{x} \rangle_t$ obtained by averaging over the entire time-series. Thus the model errors in eq. (8) become

$$\epsilon_{i,j,l,k}(t) = x_{i,j,l,k}(t) - \langle x_{i,j,l,k} \rangle_t. \tag{46}$$

The second method by Polavarapu et al. (2005) compares fields at different times, thus modelling spatial error correlations by use of temporal correlations. Hence, the background errors are modelled according to

$$\epsilon_{i,j,l,k}(t) = x_{i,j,l,k}(t + \Delta t) - x_{i,j,l,k}(t), \tag{47}$$

where Δt was set to 6 h. We emphasize once more that the bias removal according to eqs. (10)–(12) is essential to avoid artefacts in the error correlations related to the diurnal variability of chemical species.

Background error statistics were computed in our CTM by use of these two methods, to which we shall, for brevity, refer to as the 'climatological' and the 'Δt' method, respectively. In addition, a third method was employed in which the model was run with two sets of meteorological input data, namely, analysed

meteorological data and 48 h forecasts. The model states obtained from these two model runs were compared at equal times. Thus, background errors were modelled according to

$$\epsilon_{i,j,l,k}(t) = x_{i,j,l,k}(t; 0\,\mathrm{h}) - x_{i,j,l,k}(t; 48\,\mathrm{h}), \tag{48}$$

where the first argument indicates the time of the model results, and the second argument represents the forecast length of the input meteorological data. This approach is inspired by the NMC method, in which forecast runs of an assimilation system are compared at the same time with 24 and 48 h forecast lengths. In our case, we are only comparing model runs with perturbed input meteorological data, and not forecast results of an assimilation system. We chose to use a rather long forecast length difference to have a sufficiently strong perturbation of the model fields. For brevity, this method shall be referred to as the modified NMC method. One may be concerned that this modified NMC method will result in error correlations between chemical species that mainly reflect the statistics of the meteorological input data. However, this is not confirmed by the results, which show statistical properties for different chemical species that are distinctly different from one another and from meteorological variables.

In the following, we shall focus on chemical species in the aerosol phase. The MATCH model (Robertson et al., 1999; Andersson et al., 2007) employed in this study has a photochemistry module with 64 chemical species, a sea salt module (Foltescu et al., 2005), and a primary particulate matter (PPM) module. Particle mass mixing ratios were computed for four different size classes containing aerosols, with diameters ranging from 20 to 100 nm, from 100 nm to 1 μm, from 1 to 2.5 μm, and from 2.5 to 10 μm, respectively. Emission data for the photochemistry and PPM module were taken from EMEP. Based on the recent work by Kupiainen and Klimont (2004, 2007), the PPM emission data were split up into gridded emission data for elemental carbon (EC), primary organic carbon (OC) and other primary particles. Meteorological input data were taken from HIRLAM. The MATCH model was run on a European scale with a horizontal resolution of 0.4°, whereas 3DVAR computations were performed with 0.8° resolution.

Model results obtained with MATCH and stored every 3 h for the period of July 2006 were employed for performing the background error covariance modelling. For the modified NMC method, the model run with the 48 h forecast meteorological data was started on 25 June, to allow for a 6-d spin-up period during which the forecast run had enough time to diverge from the analysis run. The six spin-up days were not included in the error statistics calculations. In this way, it was ensured that the results were not influenced by the error statistics of the input meteorological data.

The results of the error statistics analyses can be presented in different ways. The power spectrum

$$P_{l,k}(k^*) = \frac{2\pi k^*}{N_s^2}\gamma_{l,k}(k^*) \tag{49}$$

is a measure for the contribution of each horizontal wavenumber k^* (corresponding to a horizontal length scale) to the total variance at level l and for species k. Figure 1 presents power spectra for different species, at different levels, and obtained with different methods.

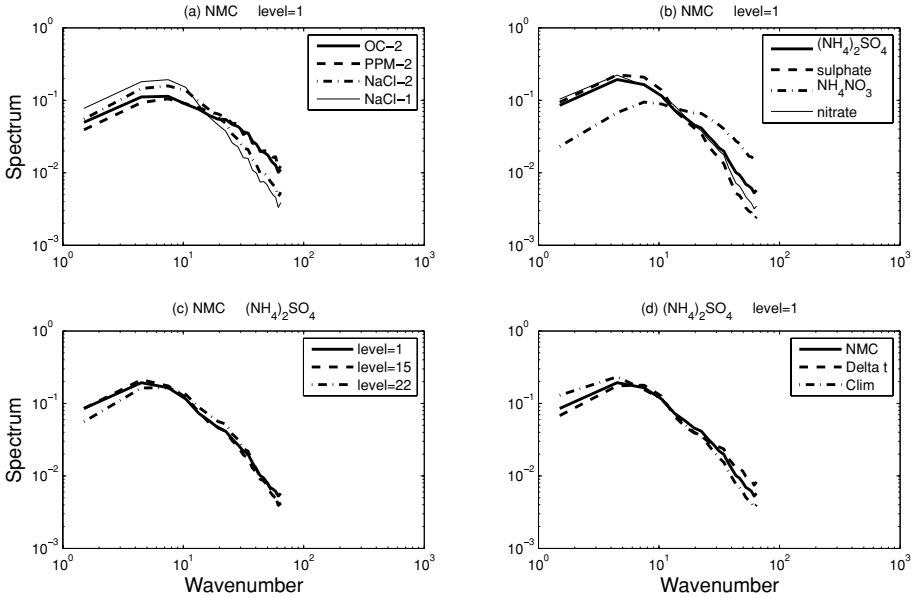

Fig. 1. (a) Power spectra of four different chemically inert aerosol species in the lowest model layer, obtained with the modified NMC method; (b) similarly for secondary inorganic species; (c) similarly for ammonium sulphate at three different vertical layers; (d) power spectra of ammonium sulphate in the lowest model layer obtained with the modified NMC, Δt, and climatological methods.

Figure 1a depicts the power spectra in the lowest model layer obtained with the modified NMC method for OC particles in the second size class, that is, with diameters between 100 nm and 1 μm (OC-2, thick solid line), inorganic primary particles in the same size class (PPM-2, dashed line), sodium chloride aerosols in the same size class (NaCl-2, dash–dotted line) and sodium chloride particles in the first size class (NaCl-1, thin solid line). All curves show a maximum at meso-scale wavenumbers. Sodium chloride aerosols are shifted to smaller wavenumbers, thus to larger horizontal scales as compared to primary particles. The reason is that emissions of sea salt over the ocean are usually generated by large-scale wind fields, whereas primary organic aerosols and other dust particles are mainly emitted from sources that vary on smaller spatial scales. Smaller aerosols can be advected over larger distances before being deposited, that is why the power spectrum of the smallest NaCl particles is shifted towards the smallest wavenumbers.

Figure 1b shows corresponding results obtained with the photochemistry module for ammonium sulphate (thick solid line), other sulphates including sulphuric acid in liquid phase (dashed line), ammonium nitrate (dash–dotted line) and other nitrates including nitric acid in liquid phase (thin solid line). Ammonium sulphate results from the reaction of sulphuric acid with ammonia, whereas ammonium nitrate is formed from nitric acid and ammonia. If the amount of ammonia in the atmosphere is limited, then the former reaction dominates and scavenges the available ammonia, thus leaving most oxidized nitrogen in the form of nitric acid. In the model, the ammonium nitrate concentration is basically instantaneously adapted in each time step to the amount of ammonia that is left after the formation of ammonium sulphate. Consequently, ammonium nitrate mixing ratios can fluctuate on considerably shorter temporal and spatial scales than the other long-range transported secondary aerosol species, resulting in a power spectrum peaking at much larger wavenumbers.

Figure 1c shows the power spectrum of ammonium sulphate at different model layers. In the hybrid vertical coordinates employed in MATCH, the same model layer can correspond to different altitudes at different horizontal gridpoints and different times. For instance, for the time and location shown in Fig. 7, model layers 15 and 22 have mid-altitudes of ca. 2600 and 5600 m, respectively. The figure shows similar power spectra of ammonium sulphate in the lowest layer and in layer no. 15 (solid and dashed lines), and a shift to larger wavenumbers in layer no. 22 (dash–dotted line).

Figure 1d shows the power spectra of ammonium sulphate in the lowest layer, computed with the modified NMC (solid line), Δt (dashed line) and climatological methods (dash–dotted line), respectively. As one may have expected, the climatological method gives rise to large-scale variations of the variance. The Δt method, which, following Polavarapu et al. (2005), we ran with a rather short time difference of 6 h, is shifted towards smaller horizontal length scales.

The horizontal length scale of the error correlation functions behaves similarly as the power spectra of the variance. Following Daley (1991), we define the length scale such that its squared value is -2 times the inverse Laplacian of the correlation function at the origin, which in spectral space can be written as (Berre, 2000)

$$L_{l,k}^2 = \left(\frac{N_\mathrm{s}}{2\pi}\right)^2 \frac{2\sum_{m,n} \hat{D}_{l,k;l,k}(k^*(m,n))}{\sum_{m,n}(k^*(m,n))^2 \hat{D}_{l,k;l,k}(k^*(m,n))}. \quad (50)$$

As noted by (Berre, 2000), this parameter mainly reflects short-distance correlations and is more sensitive to larger wavenumbers. Therefore, the results shown in Fig. 2 are mainly useful for a qualitative comparisons rather than quantitative assessments of correlation length scales. One can see that L varies with altitude, which is a consequence of our assumption of non-separable horizontal and vertical error correlations. Figure 2a presents the horizontal correlation length scale obtained with the modified NMC method for the first (PPM-1, solid line), second (PPM-2, dashed line) and fourth size class (PPM-4, dash–dotted line) of primary dust particles. The vertical dependence is qualitatively similar for the different size classes. The smaller particles, which are advected over larger distances and less effectively deposited, have somewhat larger length scales.

Figure 2b shows corresponding results for EC (EC-4, thick solid line) and OC (OC-4, dashed line), both in the fourth size class, as well as NaCl in the first (NaCl-1, dash–dotted line) and fourth size classes (NaCl-4, thin solid line). Due to the large scale, homogeneous distribution of emissions, sea salt particles have larger horizontal correlation length scales than the primary particle components, which are emitted from more inhomogeneously distributed sources.

Figure 2c presents similar results for the secondary inorganic aerosol (SIA) components. Ammonium nitrate (dash–dotted line) stands out with a considerably smaller length scale than the other compounds. As in the case of the variance power spectra, this small horizontal scale is related to the fast timescale on which this species is formed in the model.

Figure 2d depicts the vertical dependence of the horizontal correlation length scale of ammonium sulphate modelled with the modified NMC (solid line), Δt (dashed line) and climatological methods (dash–dotted line). In complete analogy to the behaviour of the variance power spectra, the horizontal correlation length scale determined with the climatological method is largest, and that obtained with the Δt method is smallest.

Figure 3 shows spectral variances and covariances $\hat{D}_{l,k;l,k'}(k^*)$ between aerosol components k and k' modelled with the modified NMC method for different vertical levels and wavenumbers. Inspecting the colour scales of the different panels, we see that the covariances are generally largest at continental scales ($k^* = 1.5$, upper row), about an order of magnitude smaller for typical meso-scales ($k^* = 7.5$, middle row) and three orders of magnitude smaller at local scales ($k^* = 64.5$, lower row). We further observe that at the larger horizontal scales (upper and middle

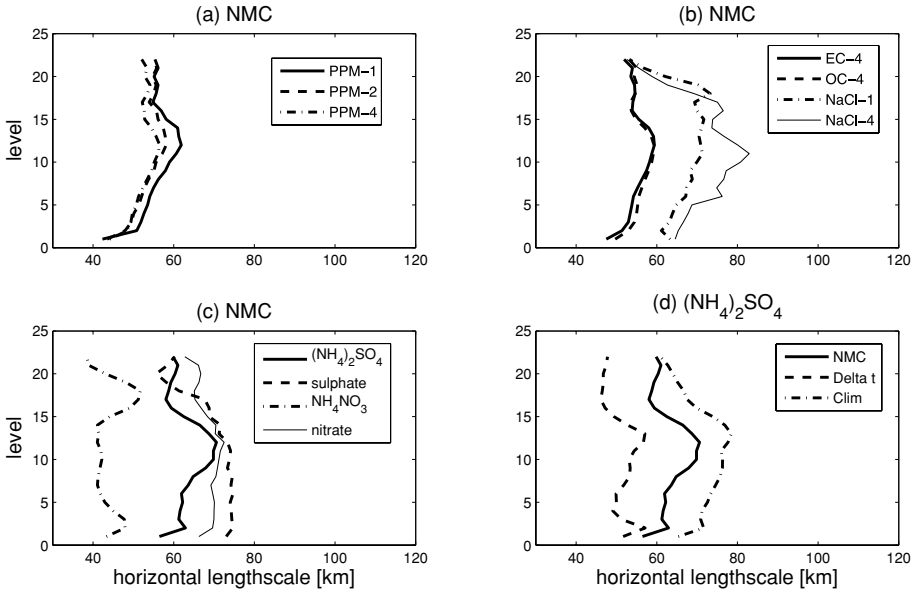

Fig. 2. (a) Horizontal length scale as a function of vertical layer obtained with the modified NMC method for different size classes of dust particles; (b) similarly for EC, OC and NaCl of different sizes; (c) similarly for secondary inorganic aerosols; (d) length scale of ammonium sulphate obtained with the modified NMC, Δt, and climatological methods.

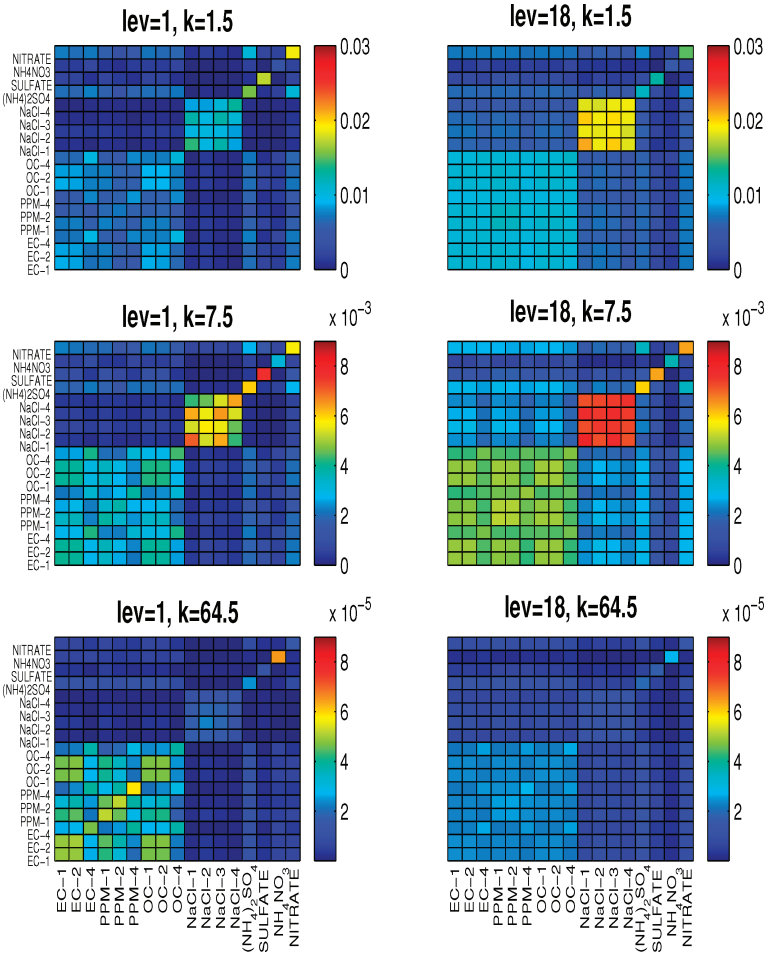

Fig. 3. Spectral covariances between different aerosol components obtained with the modified NMC method for wavenumbers $k^* = 1.5$ (upper row), 7.5 (middle row), 64.5 (lower row) and computed at the lowest model layer (left-hand column) and at layer no. 18 (right-hand column).

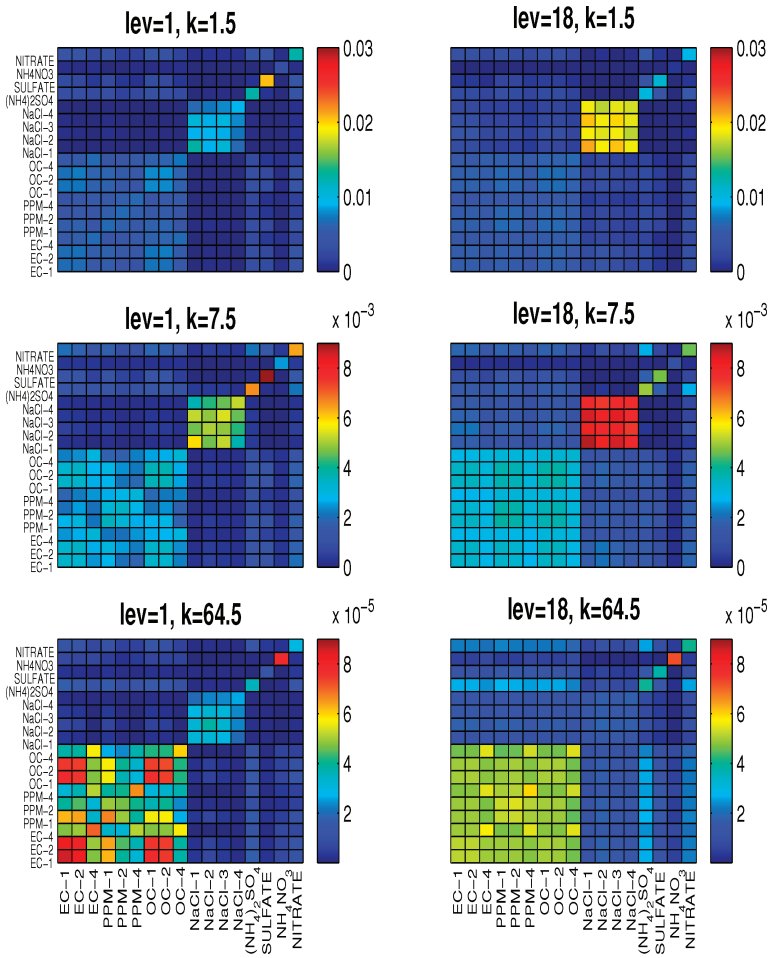

Fig. 4. As Fig. 3, but computed with the Δt method.

rows), the covariances are larger in the free atmosphere (right-hand column) than near the surface (left-hand column), which is due to the larger horizontal advection scales. Near the surface, however, the covariances are larger for smaller horizontal length scales, thus for larger wavenumbers (compare the lower left and lower right panels). Finally, we see that the covariances are rather small for the SIA components as compared to sea salt and PPM components. This indicates that statistical cross-correlations between different aerosol components are mainly caused by correlating emission sources, not so much by interconnections caused by chemical reactions. This can, of course, be quite different for gas components with fast chemical reaction rates.

Figure 4 presents corresponding results obtained with the Δt method. At larger horizontal scales (upper and middle row), the covariances are rather similar. At local scales, the Δt method predicts larger covariances for primary emitted particles than the modified NMC method.

By contrast, one can see in Fig. 5 that the climatological method predicts almost zero covariances at local scales (bottom row), and in general more pronounced covariances at larger scales (upper and middle rows), especially in the free atmosphere (right-hand column).

The differences in statistical properties obtained with the three different methods conform with our intuitive expectations. When defining model errors as differences between instantaneous field values and a 'climatological', that is, long-term average, small-scale correlations between species almost vanish, whereas long-scale correlations dominate. By contrast, defining model errors in terms of 6-h differences results in much more pronounced small-scale correlations. For performing data analysis with a climatological background estimate, the climatological B matrix modelling method is probably the most appropriate. However, when using model forecasts as background values, the NMC method appears to be a more reasonable choice. The Δt method may underestimate the horizontal length scale of the error correlations.

Figure 6 presents vertical spectral error covariances $\hat{D}_{l,k;l',k}(k^*)$ for $l = 18$, k representing ammonium nitrate. The vertical level index l' is plotted on the vertical axis, and the wavenumber k^* is plotted along the horizontal axis. We can clearly see how the vertical error covariances are depending on

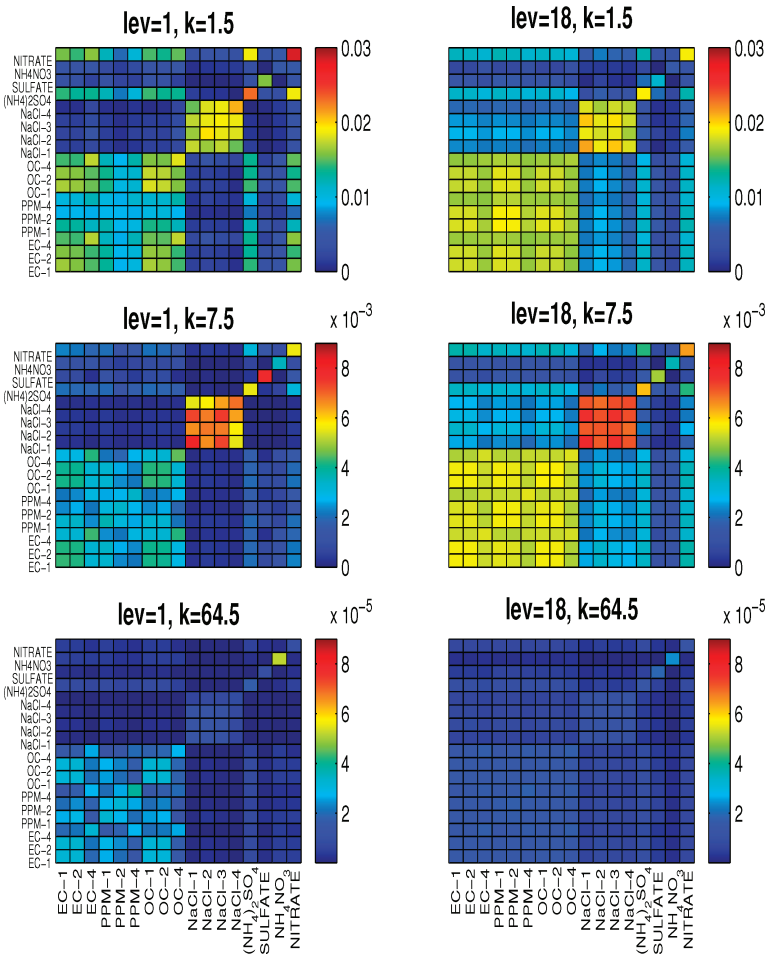

Fig. 5. As Fig. 3, but computed with the climatological method.

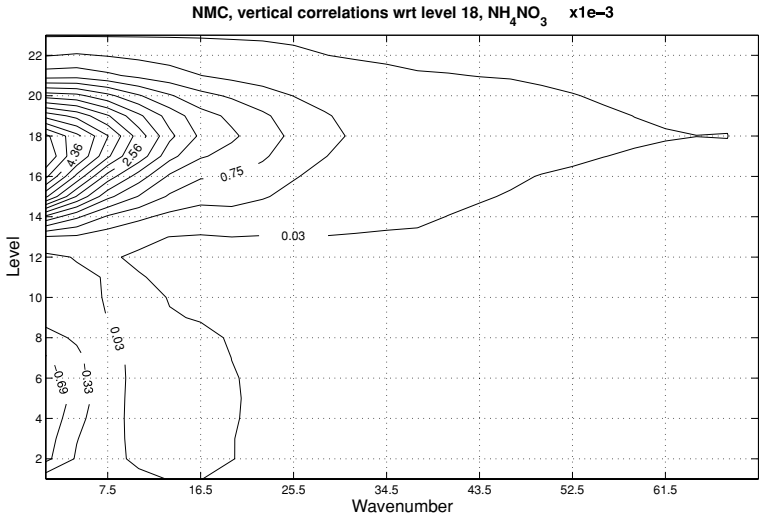

Fig. 6. Vertical spectral covariances with respect to model layer 18 for ammonium nitrate, obtained with the modified NMC method.

the wavenumber k^*, which is a consequence of our non-separable background error covariance ansatz. The vertical covariances are most pronounced at small wavenumbers, that is, at large horizontal length scales. We shall see shortly the consequences of the non-separability of horizontal and vertical error covariances in a specific application of the 3DVAR analysis code. As a preparation, we will have to briefly discuss observation operators for aerosol optical parameters.

5. Aerosol optical observation operators

Data assimilation can be employed as a method for inverting remote sensing observations. Potential weaknesses in other inversion techniques are circumvented. For instance, even if the set of observed parameters by itself is not complete enough to allow for unambiguously retrieving the physical and chemical properties of the aerosols, data assimilation always provides an answer to the inverse problem, because the information missing in the observations is supplemented by the background estimate. Similarly, in cases in which data are too sparse in space and/or time, the observation gaps are filled by the background estimate. The use of error variance and covariance constraints ensures that the information from background estimates and observations is merged by the assimilation in such a way as to yield the most probable chemical state of the atmosphere.

To be able to compare the background estimate with the observations, one needs to have an operator that maps from model to observation space. Thus, data assimilation requires, in principle, only a forward model, no inverse model. The first step is to compute aerosol optical properties, such as the extinction and scattering cross-sections $C_{\rm ext}$, $C_{\rm sca}$ and the phase function $p(\Theta)$, where Θ denotes the scattering angle. If the particles are approximated by homogeneous spheres, one can employ the Lorenz–Mie solution for the electromagnetic scattering computations (Mie, 1908). Each of the optical properties is a function of wavelength λ, particle size and refractive index m. We precomputed size-averaged optical properties for different λ and m and for each of the four size classes, assuming for each size class a log-normal size distribution with a fixed variance σ, where we used $\sigma_1 = \sigma_3 = \sigma_4 = 1.8, \sigma_2 = 1.5$ (Neusüß et al., 2002). The assumption of a log-normal size distribution allows for converting aerosol mass density M_k to aerosol number density N_k for each component k. Note that k is a super-index that includes both different size classes and chemical components.

Assuming that the aerosol components are externally mixed with one another and internally mixed with water, the refractive index in the Mie computations $m = m(k, RH(i, j, l))$ depends on the chemical component and on the relative humidity RH, which in turn determines how much water is adsorbed by the particles. We computed the effective refractive index of the internal aerosol–water mixture with effective medium theory (Chýlek et al., 2000), where we employed the Maxwell-Garnett rule (Maxwell-Garnett, 1904) if the water volume ratio exceeds 80% in the aerosols, the inverse Maxwell-Garnett rule if it lies below 20 and the Bruggeman rule (Bruggeman, 1935) in the intermediate regime. Making the independent scattering approximation, ensemble-averaged optical properties are computed for each wavelength and in each gridpoint (i, j, l) according to

$$\langle C_{\rm ext}\rangle(\lambda; i, j, l) = \frac{1}{N_{i,j,l}} \sum_k N_{i,j,l,k} \times C_{\rm ext}(\lambda, m(k, RH(i, j, l))) \tag{51}$$

and likewise for $C_{\rm sca}$, and

$$\langle p\rangle(\lambda; i, j, l; \Theta) = \frac{1}{N_{i,j,l}\langle C_{\rm sca}\rangle(\lambda; i, j, l)} \times \sum_k N_{i,j,l,k} C_{\rm sca}(\lambda, m(k, RH(i, j, l))) \times p(\lambda, m(k, RH(i, j, l)); \Theta), \tag{52}$$

$$N_{i,j,l} = \sum_k N_{i,j,l,k}. \tag{53}$$

Observable quantities of interest are, for example, the extinction optical depth

$$\tau_{\rm ext}(\lambda; i, j) = \sum_l N_{i,j,l}\langle C_{\rm ext}\rangle(\lambda; i, j, l)\Delta z_{i,j,l}, \tag{54}$$

(where Δz denotes the layer thickness), the extinction coefficient $\epsilon = {\rm d}\tau_{\rm ext}/{\rm d}z$, or the backscattering coefficient

$$\beta_{\rm sca}(\lambda; i, j, l) = \frac{1}{4\pi} N_{i,j,l}\langle C_{\rm sca}\rangle(\lambda; i, j, l)\langle p\rangle(\lambda; i, j, l; 180^\circ). \tag{55}$$

Thus, we can compute aerosol optical parameters from the mixing ratios $x_{i',j',l',k'}$ by multiplication with observation operators $\mathbf{H}^{\tau_{\rm ext}}$, $\mathbf{H}^{\epsilon}$ and $\mathbf{H}^{\beta_{\rm sca}}$ with components

$$H^{\tau_{\rm ext}}_{i,j;i',j',l',k'}(\lambda) = \delta_{i,i'}\delta_{j,j'}\Delta z_{i,j,l'} \frac{\rho^{\rm air}_{i,j,l'}}{\frac{4}{3}\pi R^3_{k'}\rho_{k'}} \times C_{\rm ext}(\lambda; i, j, l', k'), \tag{56}$$

$$H^{\epsilon}_{i,j,l;i',j',l',k'}(\lambda) = \delta_{l,l'} H^{\tau_{\rm ext}}_{i,j;i',j',l',k'}/\Delta z_{i,j,l'}, \tag{57}$$

$$H^{\beta_{\rm sca}}_{i,j,l;i',j',l',k'}(\lambda) = \delta_{i,i'}\delta_{j,j'}\delta_{l,l'} \frac{1}{4\pi} \frac{\rho^{\rm air}_{i,j,l}}{\frac{4}{3}\pi R^3_k \rho_{k'}} \times C_{\rm sca}(\lambda; i, j, l, k')p(\lambda; i, j, l, k'; 180^\circ). \tag{58}$$

The air density $\rho^{\rm air}$ converts mass mixing ratios to mass concentrations, whereas the factor $4\pi R^3_k\rho_k/3$ in the denominator converts mass densities to number densities, where ρ_k is the average mass density of species/size class k, and where particles of radius R_k make the average mass contribution in the respective size class. It is obtained by computing zeroth and third moments of the log-normal size distribution in each size interval.

The observation operators given in (56)–(58) are obtained from the Lorenz–Mie computations in conjunction with some meteorological variables (relative humidity, layer thickness). The observable quantities are obtained according to, for example,

$$\beta_{\rm sca}(\lambda; i, j, l) = \sum_{i',j',l',k'} H^{\beta_{\rm sca}}_{i,j,l;i',j',l',k'}(\lambda)\, x_{i',j',l',k'}. \tag{59}$$

The full observation operators need to be supplemented with a linear interpolation from the gridpoints to the observation points. For observations of vertical profiles, this includes a mapping from the hybrid vertical coordinates to altitudes in metres.

The main point we wish to stress in this section is that the observation operators in eqs. (56)–(58) are linear. This is a direct consequence of the assumption that the aerosol components are externally mixed and that the independent scattering approximation applies, so that the ensemble-averaged optical properties can be computed by linear superpositions of individual particle optical properties according to eqs. (51) and (52). A more realistic treatment would assume that the aerosols are partially internally and partially externally mixed, where external mixtures are more common among freshly formed or emitted, smaller particles, and internal mixtures are more typical for larger and aged aerosols. In such a treatment, however, the linearity of the observation operators would be lost, because both the effective refractive indices in the effective medium theory computations and the Mie coefficients that enter into the computation of the optical cross-sections and the phase function would be non-linear functions of the mixing ratios. To employ the incremental formulation of the cost function given in eq. (6), suitable numerical simplifications would be necessary. For internally mixed aerosols, inhomogeneity will affect the aerosol optical properties so that the use of the homogeneous sphere approximation may no longer be justified. Also, for dry aerosols non-sphericity may have a similarly strong impact. In such cases, Lorenz–Mie simulations would have to be replaced by more sophisticated electromagnetic scattering methods (Mishchenko et al., 2002; Kahnert, 2003).

Note that the optical properties are rather sensitive to the aerosol size distribution. A more detailed dynamical description of the size distribution may therefore improve our current treatment of the observation operators. In such case, a bin representation of the aerosol size distribution would be preferable to a modal description. The modal approach would require online computations of ensemble-averaged optical properties. By contrast, a sectional description would allow for pre-computation of size-averaged optical properties for each size section, which would be a considerable computational advantage.

6. An illustrative example for a 3DVAR assimilation of lidar observations of aerosol backscattering coefficients

In a recent global 4DVAR assimilation experiment of aerosol optical depth (Benedetti and Fisher, 2007), background error covariances were modelled by the NMC method. The total aerosol mass mixing ratio was chosen as the control variable. Thus, the relative contribution of each chemical species and size class to the total aerosol mass mixing ratio was kept constant over the 6-h assimilation window. This choice was motivated by the high computational costs associated with the 4DVAR minimization procedure. The 3DVAR, on the other hand, is conceptually less sophisticated than 4DVAR, but computationally less demanding. It does not require an adjoint model and an iterative integration of the model and its adjoint forward and backward in time. Rather, it merely involves minimization of the cost function in eq. (29) at each analysis time by using the gradient in eq. (30) in a decent algorithm. As a result, the considerably lower computational requirements in 3DVAR allow us to employ a more detailed description of the aerosol phase in the assimilation scheme. Thus, we employ all 17 aerosol components mentioned in Section 4 (see Figs. 3–5) as control variables.

Note that the single-observation experiment presented in this section provides a rather weak constraint for an aerosol field characterized by 17 different variables. Thus, in operational applications it would be highly beneficial to assimilate observations at several different wavelengths and/or observations of different aerosol optical properties. For our purposes, however, assimilation of vertically resolved observations provides an ideal test case for studying some properties of the non-separable background error covariance scheme employed in this study.

Figure 7 shows a vertical profile of the aerosol backscattering coefficient observed with the lidar of the Swedish Defense Research Agency (Gustafsson et al., 2006) at mid-night on 24 June 2006 in Linköping (thin solid line). The lidar wavelength is 355 nm. The thick line shows computational results obtained with MATCH in conjunction with eqs. (58) and (59). The circles represent the lidar observations reduced to the vertical resolution of the MATCH model. The boundary layer is usually not well-defined at that time of the day. However, the turbulent kinetic energy displays a sharp decrease above 1500 m at that particular location and time. Thus within the lowest well-mixed, layer the model underpredicts the observations. The high aerosol concentrations near the surface are likely to be caused by local emissions within the city that are not resolved by our regional model. It is also possible that part of the backscattering signal in

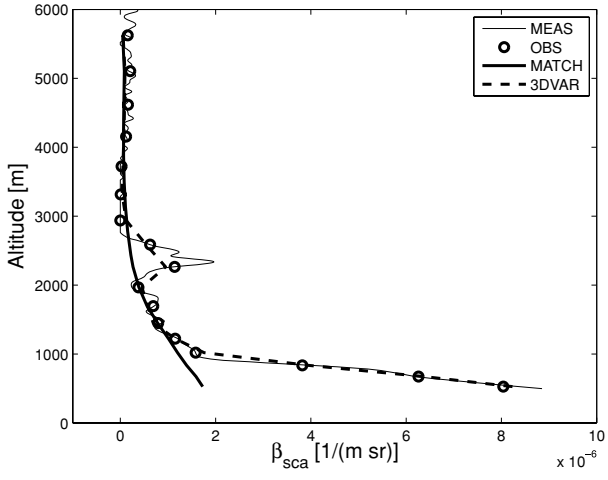

Fig. 7. Vertical profile of aerosol backscattering coefficient on 24 June 2006 at mid-night in Linköping, Sweden. The lidar wavelength is 355 nm. The figure shows the measurements (thin solid line), the measurements reduced to the model's vertical resolution (circles), the MATCH results in observation space (thick solid line) and the 3DVAR analysis in observation space (dashed line).

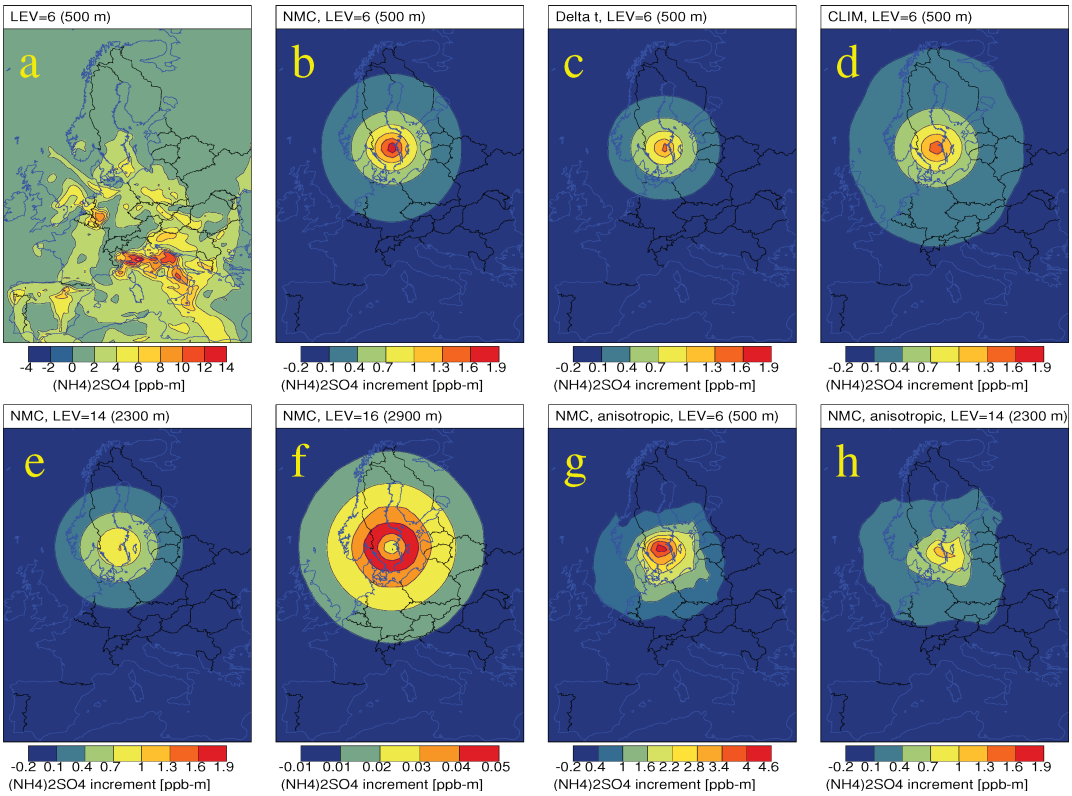

Fig. 8. (a) Background field of ammonium sulphate mixing ratio on 24 July 2006 at mid-night in model layer 6 (500 m), computed with the MATCH photochemistry model; (b) 3DVAR analysis increment, obtained with the modified NMC covariance matrix; (c) as (b), but obtained with the Δt covariance matrix; (d) as (b), but obtained with the climatological covariance matrix; (e) as (b), but for model layer 14 (2300); (f) as (b), but for model layer 16 (2900 m); (g) as (b), but obtained with an inhomogeneous, anisotropic covariance matrix; (h) as (g), but for model layer 14 (2300 m).

the boundary layer is caused by haze droplets. At altitudes larger than about 1300 m, the model generally agrees well with the observations apart from an aerosol layer around 2300 m altitude, which the model does not reproduce.

In practice, one would not want to assimilate local air pollution effects or aerosol signals perturbed by haze droplets into a regional model. One would therefore assume a relatively high representativeness error in the lowest model layers, thus inducing the assimilation algorithm to place very little weight on such observations. However, for illustration purposes, we shall set the observation standard deviations to a fairly low value of $10^{-7}\,\mathrm{m}^{-1}\,\mathrm{sr}^{-1}$ independent of altitude. In such case, the 3DVAR analysis follows the observations faithfully. The dashed line in Fig. 7 shows the analysis result in observation space. Note that the observations are uncorrelated, which implies that the observation error covariance matrix is diagonal.

We shall now turn to the 3DVAR analysis result in model space. As an example, we focus on analysis increments $\delta\mathbf{x}$ of ammonium sulphate mixing ratios. Figure 8a shows the background estimate obtained from the MATCH photochemistry model. The colour scale extends up to $14\,\mu\mathrm{g\,kg}^{-1}$. For simplicity, we first run the 3DVAR analysis with a homogeneous and isotropic background error covariance matrix. This can be achieved by dropping the horizontal dependence in the normalization of the covariance matrix with the standard deviations, thus replacing $\sigma_{i,j,l,k}$ in eq. (24) by the horizontally averaged values

$$\bar{\sigma}_{l,k} = \sqrt{\frac{1}{N_x^{\mathrm{ex}} N_y^{\mathrm{ex}}} \sum_{i=1}^{N_x^{\mathrm{ex}}} \sum_{j=1}^{N_y^{\mathrm{ex}}} \sigma_{i,j,l,k}^2}. \tag{60}$$

Figure 8b shows the analysis increment for model level 6 (the lowest level, for which lidar data are available, corresponding to an altitude of about 500 m). The 3DVAR analysis was performed by using the modified NMC background error covariance matrix. The colour scale extends up to $1.9\,\mu\mathrm{g\,kg}^{-1}$. Figures 8c and d show corresponding 3DVAR results obtained with the Δt and the climatological error covariance matrices, respectively. The analysis increment is positive, since the model underestimates the observations at this altitude. As expected, the horizontal length scale of the propagation of information is lowest for the analysis with the Δt covariance matrix and largest for the climatological covariance matrix (compare with Fig. 2d).

At altitudes around 1500 m, the analysis increment becomes nearly zero (not shown) due to the good agreement between model results and observations. At about 2300 m altitude, the

analysis increment becomes positive again, as can be seen in Fig. 8e for the modified NMC analysis. This is caused by the aerosol layer at this altitude that the model was not able to reproduce.

At about 2900 m altitude (Fig. 8f), the analysis increment is rather small since the magnitude of the innovation $\hat{H}(\mathbf{x}) - \mathbf{y}$ is almost zero. For this reason, the colour scale has now been reduced to the range from -0.01 to $0.05\ \mu g\,kg^{-1}$ to make the qualitative behaviour of the analysis increment clearer. Even though the innovation is negative here, one obtains a slightly positive increment at the observation site. Even more remarkable, the analysis increment is increasing with distance from the observation site and reaches a maximum at a horizontal distance of few hundred kilometres. At large distances from the observation site, the covariances and thus the analysis increment decrease to zero. The reason for the increase of the analysis increment as one moves away from the observation site is related to the vertical spreading of information. The positive innovation within the aerosol layer at around 2300 m gives rise to a positive analysis increment that is vertically spread and that counteracts the effect of the negative innovation at the observation site at 2900 m. However, as we saw in Fig. 6, vertical spreading is more efficient on large horizontal length scales. This is why the analysis increment in Fig. 8f is increasing with horizontal distance from the observation site. We see here an example for the practical consequences of our ansatz of non-separable background error correlations.

We shall now drop the assumption of horizontally homogeneous and isotropic background error covariances, and only assume homogeneity and isotropy of the background error correlations. This is achieved by using the horizontally dependent standard deviations $\sigma_{i,j,l,k}$ in eq. (24) instead of their horizontal averages given in (60). Figure 8g shows the NMC analysis increment at an altitude of 500 m, which should be compared to Fig. 8b. The anisotropic spreading of information is immediately apparent. Furthermore, the analysis increment is larger in Fig. 8g than in Fig. 8b. The colour scale had to be extended up to $4.6\ \mu g\,kg^{-1}$. This is an indication of the inhomogeneity of the background error covariances.

Figure 8h shows corresponding NMC results at 2300 m altitude, which should be compared to Fig. 8e. Note that the colour scale in Figs. 8e and h are identical, and that the analysis increments are approximately of the same magnitude in both cases. This indicates that the inhomogeneous structure of the background error covariances is altitude-dependent. A comparison of Figs. 8g and h reveals further that the anisotropic structure of the background error covariances is not uniform with altitude. This feature is again a consequence of the coupling between horizontal and vertical background error covariances in our non-separable scheme.

Finally, we consider what happens when performing a sequential 3DVAR data assimilation, thus using the analysis result as initial state for continuing the time integration of the model. Figure 9 shows the time-series of ammonium sulphate (upper panel) and EC (lower panel) at ground level in Linköping obtained from a pure model run (thick line), and obtained by assimilating the lidar observations (thin line). One clearly observes that 24–48 h after the analysis the model has basically relaxed to the reference run. This is caused by advective transport from surrounding regions and the forcing effect of the emissions. Elbern et al. (2007) have emphasized the importance of optimizing not only initial values but also emission rates in chemical data assimilation. They performed four-dimensional variational data assimilation (4DVAR) with an additional term in the cost function accounting for the emission errors and achieved significant improvements in the forecast of various reactive gas species with their approach. Constantinescu et al. (2007a) came to analogous conclusions by assimilating ozone observations using an EnKF scheme.

Whereas NWP models are mainly sensitive to initial states, CTMs are strongly driven by the emissions. For this reason,

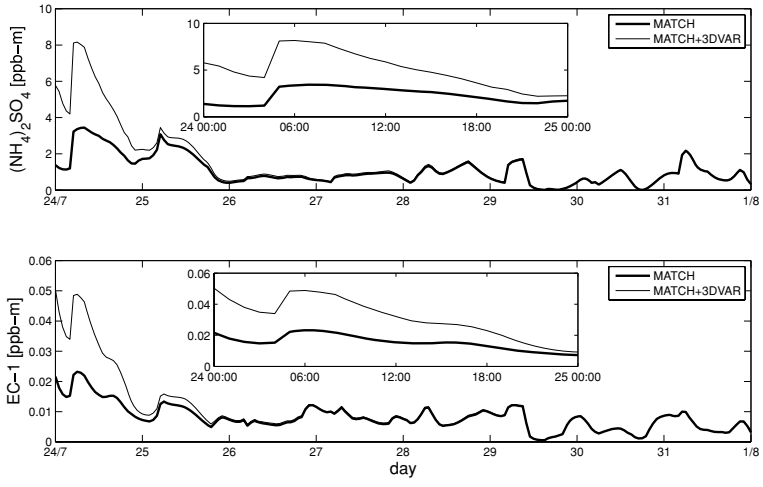

Fig. 9. Upper panel: Time-series of ammonium nitrate mixing ratio for the period 24 July-1 August 2006. The thin and thick lines show model results with and without data assimilation, respectively. Lower panel: as upper panel, but for elemental carbon (lowest size class).

chemical data assimilation schemes that only focus on adjusting mixing ratios are likely to improve only short-term forecasting abilities. In such case, one needs to have sufficiently frequent observations to counteract the relaxation to the model state dictated by the input emissions. Also one needs to have a reasonably good spatial coverage to avoid relaxation of the analysis result, driven by advection from regions that lie outside the radius of the analysis increments.

An assimilation scheme in which the governing equations do not enter as a strong constraint is potentially problematic. In chemical data assimilation, an analysis result that is in conflict with the model equations can give rise to spurious relaxations to chemical equilibrium following the analysis. However, in our test case we do not find these suspicions confirmed. By comparing the chemically inert EC to the chemically produced ammonium sulphate, we can see that their time evolutions are similar. For both components, the assimilation runs display the same oscillations as the reference runs, and for both components the relaxation of the assimilation runs to the reference runs occurs on the same timescale. This indicates that the analysis result is well balanced in our case. However, it is conceivable that attaining a well-balanced analysis within the 3DVAR scheme is more difficult for highly reactive gas components.

7. Summary and conclusions

We have employed a non-separable spectral formulation of variational data analysis and discussed three simple methods for obtaining a first estimate of the background error statistics based on model results. The climatological method yields rather large-scale horizontal and chemical cross-correlations, whereas the Δt method tends towards small-scale spatial and chemical cross-correlations. The modified NMC method yields background error statistics that lies between these two extremes. The NMC results appear to provide a reasonable first estimate for the background error covariance matrix that can be employed for modelling the background error statistics of an assimilation system, for example, with the regular NMC method (Parrish and Derber, 1992) or with the ensembles-of-analyses method (Fisher, 2003).

We found that statistical cross-correlations between chemical aerosol components and spatial horizontal correlations are mainly influenced by interrelations between emission sources, and to a much lesser degree by interrelations due to chemical reactions forming secondary aerosol species. Sea salt, for example, is emitted rather uniformly over extended source regions, thus giving rise to large horizontal correlation length scales and high spectral covariances at low wavenumbers between sea salt aerosols belonging to different size classes. The non-separability of horizontal and vertical correlations in our scheme manifests itself in spectral space as vertical covariances that increase with decreasing horizontal wavenumber, and as horizontal correlation length scales that are dependent on the vertical coordinate.

Observation operators for aerosol optical depth, extinction coefficient and backscattering coefficient have been implemented into the 3DVAR scheme. The basic assumptions in the construction of these operators were that aerosols can be represented by homogeneous spheres, that different aerosol components are externally mixed and that the aerosol density is sufficiently low for the independent scattering approximation to be valid. The first two assumptions constitute rather crude simplifications, whereas the third assumption is well justified. The first assumption makes the computations of aerosol optical parameters particularly fast and simple. However, since the aerosol optical properties are precomputed off-line, it would be feasible to use more sophisticated model particles, such as homogeneous spheroids (Dubovik et al., 2002) or inhomogeneous, coated spheres (Toon and Ackermann, 1981). The last two assumptions result in linear observation operators, which greatly reduces the computational burden in the iterative minimization of the cost function.

A 3DVAR single-observation analysis of lidar observations of backscattering coefficient illustrated the practical implications of the various statistical properties of the modelled error covariance matrices. More specifically, the different correlation length scales obtained with the three different error covariance modelling methods were reflected by the horizontal extent of the analysis increments. The non-separability of horizontal and vertical background error covariances resulted in scale-dependent vertical spreading of information in the analysis algorithm. The altitude-dependent structure of inhomogeneous and anisotropic analysis increments was another manifestation of the non-separability in our scheme.

Sequential data assimilation with the 3DVAR approach requires observations that are sufficiently dense in time and space. One problem is that the long-term time evolution of the model state is mainly forced by the emissions and, by contrast to an NWP model, less sensitive to the initial state. Thus, if observations are too sparse in time, the adjustment of the model state by the analysis has no lasting effect in the continuing time-integration of the model. A promising remedy is the approach by Elbern et al. (2007) and by Constantinescu et al. (2007a), which is based on including the emissions in the control vector, thus allowing the assimilation algorithm to adjust not only the model state but also the emissions within the bounds of their error variances to fit the observations. This approach has been tested by Elbern et al. (2007) within a 4DVAR scheme and by Constantinescu et al. (2007a) within an EnKF scheme.

The analysis result in the 3DVAR method does, in general, not lie within the manifold in configuration space that is defined by the model constraint. This can, in principle, result in spurious relaxations of mixing ratios of reactive species to chemical equilibrium following an analysis. However, in our test case we did not observe such relaxation processes, thus indicating that the analysis result of the secondary aerosol components was sufficiently well balanced.

Finally, it is noted that the spectral approach pursued in this study has two main limitations. First, as previously mentioned, it is limited to homogeneous error correlations, even though the error covariances are not required to have this symmetry property. Second, one needs to add a horizontal extension zone, which, depending on the horizontal error correlations, may have to be a few hundred kilometres broad. A promising new approach that circumvents both of these shortcomings is based on employing wavelet transforms (Selesnick et al., 2005) instead of Fourier transforms. First applications in meteorological data assimilation have been reported by Fisher (2003, 2006) and Pannekoucke et al. (2007).

8. Acknowledgments

Ove Gustafsson and Rolf Persson from the Swedish Defense Research Agency are gratefully acknowledged for having made their lidar measurements available. This work was supported by the Swedish Governmental Agency for Innovation Systems (VINNOVA) as part of the AMAGAL project (2006-02314).

References

Andersson, C., Langner, J. and Bergström, R. 2007. Interannual variation and trends in air pollution over Europe due to climate variability during 1958-2001 simulated with a regional CTM coupled to the ERA40 reanalysis. *Tellus* **59B**, 77–98.

Benedetti, A. and Fisher, M. 2007. Background error statistics for aerosols. *Q. J. R. Meteorol. Soc.* **133**, 391–405.

Berre, L. 2000. Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Mon. Wea. Rev.* **128**, 644–667.

Bruggeman, D. A. G. 1935. Berechnung verschiedener physikalischer Konstanten von heterogenen Substanzen. 1. Dielektrizitätskonstanten und vinnova Leitfähigkeiten der Mischkörper aus isotropen Substanzen. *Ann. Phys.* **24**, 636–664.

Chýlek, P., Videen, G., Geldart, D. J. W., Dobbie, J. S. and Tso, H. C. W. 2000. Effective medium approximations for heterogeneous particles. In: *Light Scattering by Nonspherical Particles* (eds. M. I. Mishchenko, J. W. Hovenier and L. D. Travis, Academic Press, San Diego , 274–308.

Collins, W. D., Rasch, P. J., Eaton, B. E., Khattatov, B. V. and Lamarque, J.-F. 2001. Simulating aerosols using a chemical transport methodology for INDOEX. *J. Geophys. Res.* **106**, 7313–7336.

Constantinescu, E. M., Sandu, A., Chai, T. and Carmichael, G. R. 2007a. Ensemble-based chemical data assimilation. I: ceneral approach. *Q. J. R. Meteorol. Soc.* **133**, 1229–1243.

Constantinescu, E. M., Sandu, A., Chai, T. and Carmichael, G. R. 2007b. Ensemble-based chemical data assimilation. II: Covariance localization. *Q. J. R. Meteorol. Soc.* **133**, 1245–1256.

Courtier, P., Thépaut, J.-N. and Hollingsworth, A. 1994. A strategy for operational implementation of 4d-var using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1388.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljević, D., and co-authors. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: formulation. *Q. J. R. Meteorol. Soc.* **124**, 1783–1807.

Daley, R. 1991. *Atmospheric Data Analysis*. Cambridge University Press, Cambridge, UK .

Derber, J. and Bouttier, F. 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus* **51A**, 195–221.

Dubovik, O., Holben, B. N., Lapyonok, T., Sinyuk, A., Mishchenko, M. I., and co-authors. 2002. Non-spherical aerosol retrieval method employing light scattering by spheroids. *Geophys. Res. Lett.* **29**, doi: 10.1029/2001GL014506.

Elbern, H. and Schmidt, H. 1999. A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling. *J. Geophys. Res.* **104**, 18 583–18 598.

Elbern, H. and Schmidt, H. 2001. Ozone episode analysis by four-dimensional variational chemistry data assimilation. *J. Geophys. Res.* **106**, 3569–3590.

Elbern, H., Schmidt, H. and Ebel, A. 1997. Variational data assimilation for tropospheric chemistry modeling. *J. Geophys. Res.* **102**, 15 967–15 985.

Elbern, H., Schmidt, H., Talagrand, O. and Ebel, A. 2000. 4D-variational data assimilation with and adjoint air quality model for emission analysis. *Environ. Model. Software* **15**, 539–548.

Elbern, H., Strunk, A., Schmidt, H. and Talagrand, O. 2007. Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmos. Chem. Phys.* **7**, 3749–3769.

Evensen, G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.

Fisher, M. 2003. Background error covariance modelling. In: *Proceedings of the Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*, 8-12 September 2003, Reading, UK, pp. 45–64.

Fisher, M. 2006. Wavelet Jb—a new way to model the statistics of background errors. *ECMWF Newslett.*, ECMWF, Reading, UK, **106**, 23–28.

Fisher, M. and Lary, D. J. 1995. Lagrangian four-dimensional variational data assimilation of chemical species. *Q. J. R. Meteorol. Soc.* **121**, 1681–1704.

Foltescu, V., Pryor, S. C. and Bennet, C. 2005. Sea salt generation, dispersion and removal on the regional scale. *Atmos. Environ.* **39**, 2123–2133.

Greiner, W. and Müller, B. 1994. *Quantum Mechanics (Symmetries)*. Springer, Berlin .

Gustafsson, N. 2007. Discussion on '4-D-Var or ensemble Kalman filter?'. *Tellus* **59A**, 774–777.

Gustafsson, N., Berre, L., Hörnquist, S., Huang, X.-Y., Lindskog, M., and co-authors. 2001. Three-dimensional variational data assimilation for a limited area model part I: general formulation and the background error constraint. *Tellus* **53A**, 425–446.

Gustafsson, O. K. S., Persson, R. and Thorin, E. 2006. Use of lidar measurements of aerosol extinction and backscatter coefficients as a part of assessing data from meteorological forecast models and scattering calculations. In: *Lidar Technologies, Techniques, and Measurements for Atmospheric Remote Sensing II*. (ed. U. N. Singh), Proceeding of SPIE Vol. 6367, 636710, doi: 10.1117/12.689919 (2006).

Hamill, T. M. and Snyder, C. 2000. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.* **128**, 2905–2919.

Hollingsworth, A. and Lönnberg, P. 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. part I: the wind field. *Tellus* **38A**, 111–136.

Kahnert, F. M. 2003. Numerical methods in electromagnetic scattering theory. *J. Quant. Spectrosc. Radiat. Transfer* **79-80**, 775–824.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C. and Ballabrera-Poy, J. 2007a. 4-D-Var or ensemble Kalman filter? *Tellus* **59A**, 758–773.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C. and Ballabrera-Poy, J. 2007b. Response to the discussion on "4-D-Var or ensemble Kalman filter?" by Nils Gustafsson. *Tellus* **59A**, 778–780.

Khattatov, B. V., Lamarque, J.-F., Lyjak, L. V., Menard, R., Levelt, and co-authors. 2000. Assimilation of satellite observations of long-lived chemical species in global chemistry transport models. *J. Geophys. Res.* **105**, 29 135–29 144.

Kupiainen, K. and Klimont, Z. 2004. Primary emissions of submicron and carbonaceous particles in Europe and the potential for their control, Technical Report IR-04-079, IIASA, Laxenburg, Austria.

Kupiainen, K. and Klimont, Z. 2007. Primary emissions of fine carbonaceous particles in Europe. *Atmos. Environ.* **41**, 2156–2170.

Levelt, P. F., Khattatov, B. V., Gille, J. C., Brasseur, G. P. and Tie, X. X. 1998. Assimilation of MLS ozone measurements in the global three-dimensional transport model ROSE. *J. Geophys. Res.* **25**, 4493–4496.

Lindskog, M., Gustafsson, N., Navascués, B., Mogensen, K. S., Huang, X.-Y., and co-authors. 2001. Three-dimensional variational data assimilation for a limited area model part II: observation handling and assimilation experiments. *Tellus* **53A**, 447–468.

Lorenc, A. C. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**, 1177–1194.

Maxwell-Garnett, J. C. 1904. Colours in metal glasses and in metallic films. *Philos. Trans. R. Soc. A* **203**, 385–420.

Ménard, R. and Chang, L.-P. 2000. Assimilation of stratospheric chemical tracer observations using a Kalman filter. part II: χ^2-validated results and analysis of variance and correlation dynamics. *Mon. Wea. Rev.* **128**, 2672–2686.

Ménard, R., Yang, Y. and Polavarapu, S. 2004, Model error estimation: its application to chemical data assimilation, *Proceedings of Workshop on Modelling and Assimilation for the Stratosphere and Tropopause*, ECMWF/SPARC, pp. 137–145.

Mie, G. 1908. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. Phys.* **25**, 377–445.

Mishchenko, M. I., Travis, L. D. and co-authors.Lacis, A. A. 2002. *Scattering, Absorption, and Emission of Light by Small Particles*. Cambridge University Press, Cambridge.

Neusüß, C., Wex, H., Birmili, W., Wiedensohler, A., Koziar, C., and 2002. Characterization and parametrization of atmospheric particle number-, mass-, and chemical-size distributions in central Europe during LACE 98 and MINT. *J. Geophys. Res.* **107**, doi:10.1029/2001JD000514.

Pannekoucke, O., Berre, L. and Desroziers, G. 2007. Filtering properties of wavelets for local background-error correlations. *Q. J. R. Meteorol. Soc.* **133**, 363–379.

Parrish, D. F. and Derber, J. C. 1992. The National Meteorological Centre's spectral statistical interpolation analysis system. *Mon. Wea. Rev.* **120**, 1747–1763.

Polavarapu, S., Ren, S., Rochon, Y., Sankey, D., Ek, N., and co-authors. 2005. Data assimilation with the Canadian middle atmosphere model. *Atmos. Ocean* **43**, 77–100.

Rasch, P. J., Collins, W. D. and Eaton, B. E. 2001. Understanding the Indian Ocean experiment (INDOEX) aerosol distributions with an aerosol assimilation. *J. Geophys. Res.* **106**, 7337–7355.

Robertson, L. and Langner, J. 1992. Source function estimate by means of variational data assimilation applied to the ETEX-I tracer experiment. *Atmos. Environ.* **32**, 4219–4225.

Robertson, L., Langner, J. and Enghardt, M. 1999. An Eulerian limited-area atmospheric transport model. *J. Appl. Meteorol.* **38**, 190–210.

Sandu, A., Liao, W., adn D. K., Henze, G. R. C. and Seinfeld, J. H. 2005. Inverse modeling of aerosol dynamics using adjoints: theoretical and numerical considerations. *Aerosol. Sci. Technol.* **39**, 677–694.

Saunders, R. E., Andersson, E., Järvinen, H., Gérard, E., Rohn, M. and co-authors. 1998. Recent improvements to the ECMWF 4DVar data assimilation system. *ECMWF Newslett.* **81**, 2–7.

Selesnick, I. W., Baraniuk, R. G. and Kingsbury, N. G. 2005. The dual-tree complex wavelet transform. *IEEE Signal Process. Mag.* **22**, 123–151.

Talagrand, O. and Courtier, P. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: theory. *Q. J. R. Meteorol. Soc.* **113**, 1311–1328.

Toon, O. B. and Ackermann, T. P. 1981. Algorithms for the calculation of scattering by stratified spheres. *Appl. Opt.* **20**, 3657–3660.

van Loon, M., Builtjes, P. J. H. and Segers, A. J. 2000. Data assimilation of ozone in the atmospheric transport chemistry model LOTOS. *Environ. Model. Software* **15**, 603–609.

Wang, K.-Y., Lary, D. J., Shallcross, D. E., Hall, S. M. and Pyle, J. A. 2001. A review on the use of the adjoint method in four-dimensional atmospheric-chemistry data assimilation. *Q. J. R. Meteorol. Soc.* **127**, 2181–2204.