



# How much information do extinction and backscattering measurements contain about the chemical composition of atmospheric aerosol?

Michael Kahnert<sup>1,2</sup> and Emma Andersson<sup>2</sup>

<sup>1</sup>Research Department, Swedish Meteorological and Hydrological Institute, Folkborgsvägen 17, 601 76 Norrköping, Sweden

<sup>2</sup>Department of Earth and Space Science, Chalmers University of Technology, 412 96 Gothenburg, Sweden

Correspondence to: Michael Kahnert (michael.kahnert@smhi.se)

Received: 10 October 2016 – Discussion started: 10 November 2016

Revised: 9 February 2017 – Accepted: 11 February 2017 – Published: 9 March 2017

**Abstract.** We theoretically and numerically investigate the problem of assimilating multiwavelength lidar observations of extinction and backscattering coefficients of aerosols into a chemical transport model. More specifically, we consider the inverse problem of determining the chemical composition of aerosols from these observations. The main questions are how much information the observations contain to determine the particles' chemical composition, and how one can optimize a chemical data assimilation system to make maximum use of the available information. We first quantify the information content of the measurements by computing the singular values of the scaled observation operator. From the singular values we can compute the number of signal degrees of freedom,  $N_s$ , and the reduction in Shannon entropy,  $H$ . As expected, the information content as expressed by either  $N_s$  or  $H$  grows as one increases the number of observational parameters and/or wavelengths. However, the information content is strongly sensitive to the observation error. The larger the observation error variance, the lower the growth rate of  $N_s$  or  $H$  with increasing number of observations. The right singular vectors of the scaled observation operator can be employed to transform the model variables into a new basis in which the components of the state vector can be partitioned into signal-related and noise-related components. We incorporate these results in a chemical data assimilation algorithm by introducing weak constraints that restrict the assimilation algorithm to acting on the signal-related model variables only. This ensures that the information contained in the measurements is fully exploited, but not overused. Numerical tests show that the constrained data assimilation algorithm

provides a solution to the inverse problem that is considerably less noisy than the corresponding unconstrained algorithm. This suggests that the restriction of the algorithm to the signal-related model variables suppresses the assimilation of noise in the observations.

## 1 Introduction

Atmospheric aerosols have a substantial, yet highly uncertain impact on climate, they can cause respiratory health problems, degrade visibility, and even compromise air-traffic safety. The physical and chemical properties of aerosols play a key role in understanding these effects. The aerosol properties are determined by a complex interplay of different chemical, microphysical, and meteorological processes. These processes are investigated in environmental modelling by use of chemical transport models (CTMs). However, modelling aerosol processes is plagued by substantial biases and errors (McKeen et al., 2007). It is, therefore, fundamentally important to evaluate and constrain CTMs by use of measurements.

Measurements from satellite instruments provide consistent long-term data sets with global coverage. However, it is notoriously difficult to compare measured radiances to modelled aerosol concentrations. An alternative to using radiances is to make use of satellite retrieval products. For instance, one of the products of the CALIPSO lidar instrument (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) is a rough classification of the aerosol types (i.e.

dust, smoke, clean/polluted continental, and clean/polluted marine). This retrieval product is based on lidar depolarization measurements (Omar et al., 2009). For the evaluation of aerosol transport models this provides us with a qualitative check for the chemical composition of aerosols. However, this is of limited practical use, since what we really need is quantitative information on the particles' chemical composition (which can be size-dependent). The most popular approach in evaluating and constraining aerosol transport models is the use of retrieved optical properties, such as aerosol optical depth, or extinction and backscattering coefficients. Yet another idea is to provide the particles' refractive index as a retrieval product (e.g. Müller et al., 1999; Veselovskii et al., 2002). However, the use of such retrieval products still leaves us with the challenge of solving an ill-posed inverse problem, namely, of determining the particles' chemical composition from their retrieved optical or dielectric properties.

A systematic class of statistical methods for solving this inverse problem is known as data assimilation. Recent studies have applied data assimilation to aerosol models with varying degrees of sophistication, ranging from simple dust models (Khade et al., 2013) and mass transport models (Zhang et al., 2014) to microphysical aerosol models based on modal (Rubin and Collins, 2014) or sectional descriptions (Sandu et al., 2005; Saide et al., 2013) of the aerosol size distribution. The assimilation techniques that have been used comprise variational methods, such as 2-D (Zhang et al., 2014), 3-D (Kahnert, 2008; Liu et al., 2011), and 4-D variational methods (Benedetti et al., 2009), as well as ensemble approaches (Sekiyama et al., 2010). Assimilation of satellite products for trace gases is relatively straightforward, since observed and modelled trace gas concentrations are almost directly comparable. However, aerosol optical properties observed from satellites are not directly comparable to the modelled size distribution and chemical composition of the aerosols. Solving this problem amounts to regularizing a severely under-constrained inverse problem. Previous aerosol assimilation attempts have been mainly based on educated guesses about the information content of the observations. For instance, there have been studies on the assimilation of aerosol optical depth (AOD) in which all chemical aerosol components in all size classes and at all model layers were used as independent control variables (Liu et al., 2011). This approach largely disregards the problems involved in inverse modelling. By contrast, it has been proposed to only allow for the total aerosol mass concentration to be corrected by data assimilation of AOD (Benedetti et al., 2009; Wang et al., 2014). This is a more prudent approach based on the plausible assumption that a single optical variable only contains enough information to control a single model variable. There have also been intermediate approaches in which the total aerosol mass per size bin have been used as control variables (Saide et al., 2013).

In all such approaches the choice of control variables is based on ad hoc assumptions. Numerical assimilation experi-

ments by Kahnert (2009) suggest that observations of several aerosol optical properties at multiple wavelengths may allow us to constrain more than just the total mass concentration, but certainly not *all* aerosol parameters. However, it is still an unsolved mystery how much information a given set of observations actually contains about the size distribution and chemical composition of aerosols, and exactly which model variables are related to the observed signals, and which ones are related to noise. Thus a prerequisite for assimilating remote sensing observations into aerosol transport models is to thoroughly understand the information content of the observations as well as the relation between the model variables and the signal degrees of freedom.

In numerical weather prediction (NWP) modelling, several studies have discussed the information content of satellite observations for meteorological variables. For instance, Joiner and da Silva (1998) applied a singular-value decomposition (SVD) approach in order to reduce the effect of prior information in the analysis, so that the retrieval and forecast errors can be assumed to be uncorrelated. Rabier et al. (2002) considered assimilation of IR sounders, which typically provide a large number of different channels. They applied methods of information and retrieval theory in order to decide which channels contain most information about the vertical variation of temperature and humidity. Cardinali et al. (2004) employed the influence matrix to compute diagnostics of the impact of observations in a global NWP data assimilation system. Johnson et al. (2005a, b) investigated filtering and interpolation aspects in a 4DVAR assimilation system by use of an SVD approach. They also used Tikhonov regularization theory to optimize the signal-to-noise regularization parameter in order to maximize the information that can be extracted from observations. Xu (2006) compared different metrics, namely, the relative entropy and the Shannon-entropy difference, to measure information contents of radar observations assimilated into a coupled atmosphere–ocean model. Bocquet (2009) used methods of information theory to address the question how to determine an optimum spatial resolution of the discretized space of control variables in geophysical data assimilation.

Burton et al. (2016) have recently investigated the information content of “ $3\beta + 2\alpha$ ” lidar measurements, i.e. observations of backscattering at three wavelengths and extinction at two wavelengths, where the information content was analysed with regard to the refractive index and number distribution of the aerosol particles. Veselovskii et al. (2004, 2005) have performed similar analyses of the information content of multiwavelength Raman lidar measurements with regard to the complex refractive index and the effective radius of the aerosol particles. As mentioned earlier, the refractive index is a very useful retrieval product of remote sensing observations. However, from the point of view of chemical transport modelling, the main quantities of interest are the concentrations of the different chemical species of which the aerosol particles are composed. Although the chemical composition

determines the refractive index, the inversion of this relationship is still under-determined, hence an ill-posed problem. In the present paper, we want to investigate the inverse problem that goes all the way from optical properties to the chemical composition of particles.

The two main goals of this paper are (i) to apply a systematic method for analysing the information content of aerosol optical properties with regard to the particles' chemical composition, and (ii) to test an algorithm for making an automatic choice of control variables in chemical data assimilation such that all control variables are signal related, while the noise-related variables remain unchanged by the assimilation procedure. The main hypothesis is that by constraining the data assimilation algorithm to acting on the signal-related variables only, the output will be less noisy than in an unconstrained assimilation. The focus of our study will be on spectral observations of extinction and backscattering coefficients, which can be retrieved from lidar observations.<sup>1</sup> We will not restrict this analysis to any fixed choice of wavelengths, such as  $3\beta + 2\alpha$ . Instead, we will investigate the information content for varying combinations of the three main wavelengths of the commonly used neodymium-doped yttrium aluminium garnet (Nd:YAG) laser. However, it should be mentioned that extinction measurements at the lowest harmonic of 1064 nm can be difficult and plagued by high errors; in practice, this will affect the observation error, resulting in a low information content of this particular measurement.

The paper is organized as follows. Section 2 gives a rather concise introduction of the modelling tools and of the numerical approach employed to studying the information content of extinction and backscattering observations. Section 3 presents the main results of this study, and Sect. 4 offers concluding remarks. To make this paper self-contained, we include an appendix that gives a brief introduction to some essential concepts of data assimilation, and a detailed explanation of the methods we used for quantifying the information content of aerosol optical observables.

## 2 Methods

This study consists of two parts. In the first part we quantify the information content of extinction and backscattering

<sup>1</sup>In addition to lidar measurements from ground-based and aircraft-carried instruments (e.g. Burton et al., 2015), there are currently two space-borne lidar instruments in orbit. The CALIOP instrument on board the CALIPSO satellite has been launched in April 2006; it has three receiver channels – one at 1064 nm and two channels at 532 nm to measure orthogonally polarized components. The CATS instrument on board the International Space Station has been operational since January 2015. It measures backscattering at 355, 532, and 1064 nm, where the latter two have two orthogonal polarization channels. It is also capable of performing high-spectral-resolution measurements at 532 nm. A third instrument is planned to be launched in 2018 (ATLID on board EarthCARE).

coefficients at multiple wavelengths. In the second part we perform a numerical test to investigate to what extent the concentrations of different chemical aerosol components can be constrained by observations of extinction and backscattering coefficients. The modelling tools required for this study are (i) a chemical transport model, (ii) an aerosol optics model, and (iii) a data assimilation system.

### 2.1 Multiple scale Atmospheric Transport and CHemistry modelling system (MATCH)

We employ the chemical transport model MATCH, which is an off-line Eulerian CTM with flexible model domain. It has been previously used from regional to hemispheric scales. Here we use a model version that contains a photochemistry module with 64 chemical species, among them four secondary inorganic aerosols (SIAs) – namely ammonium sulfate, ammonium nitrate, other sulfates, and other nitrates. It also contains a module with 16 primary aerosol variables – namely sea salt, elemental carbon (EC), organic carbon (OC), and dust particles, each emitted in four different size bins. Thus, the model contains 20 different aerosol variables. The particle-radius ranges of the four bins are as follows:

- size bin 1: 10–50 nm;
- size bin 2: 50–500 nm;
- size bin 3: 500–1250 nm;
- size bin 4: 1250–5000 nm.

The model reads in emission data, meteorological data, and land use data and computes transport processes, chemical transformation, and dry and wet deposition of the various trace gases and aerosols. As output, it provides concentration fields of gases and aerosols, the deposition of these chemical species to land and water-covered areas, as well as the temporal evolution of these variables.

We mention that there exists another model version that includes aerosol microphysical processes, such as nucleation, condensational growth, and coagulation. In that model version the aerosol size distribution evolves dynamically. The model has 20 size bins and seven chemical species (EC, OC, dust, sea salt, particulate sulfate (PSOX), particulate nitrate (PNOX), and particulate ammonium (PNHX)), although not all species are encountered in all size bins. The total number of model variables currently in that version is 82.

More complete information about the mass transport model can be found in Andersson et al. (2007). The sea salt module is discussed in Foltescu et al. (2005). The aerosol microphysics module is described in Andersson et al. (2015).

For the sake of simplicity we here use the mass transport model without aerosol microphysical processes (see next section). The model is set up over Europe covering  $33^\circ$  in the longitudinal and  $42^\circ$  in the latitudinal direction in a rotated lat-long grid with  $0.4^\circ \times 0.4^\circ$  horizontal resolution. In

the vertical direction the model domain extends up to 13 hPa, using 40 terrain-following coordinates. The meteorological input data are taken from the numerical weather prediction model HIRLAM (Undén et al., 2002). For the emissions of all aerosol components we used EMEP data for the year 2007, where EC and OC emissions were computed from total primary particle emissions based on the data in Kupiainen and Klimont (2004, 2007).

## 2.2 Aerosol optics model

We have two different optics models coupled to MATCH: one to the mass transport module, and another to the aerosol microphysics module. The former assumes that all aerosol species are homogeneous spheres, and that each chemical species is contained in separate particles. Under these assumptions the optics model is linear, i.e. the optical properties are linear functions of the concentrations of the chemical aerosol species. The latter model accounts for the fact that in reality different chemical species can be internally mixed, i.e. they can be contained in one and the same particle. That model also accounts for the inhomogeneous internal structure of black carbon mixed with other aerosol components, and for the irregular fractal aggregate morphology of bare black carbon particles (Kahnert et al., 2012a, 2013). Under these assumptions the optics model becomes nonlinear, which introduces additional complications in the inverse-modelling problem. This is the main reason why we chose to use the simpler mass transport optics model in this study. Much of the theory explained in the Appendix B relies on the assumption that the optics model is either linear, or that it is only mildly nonlinear, so that it can be linearized – see Eq. (B6).

Table 1 lists the refractive indices in the mass-transport optics model at the three lidar wavelengths considered in this study. More information about the aerosol optics models implemented in MATCH can be found in Andersson and Kahnert (2016).

## 2.3 Three-dimensional variational data assimilation (3DVAR)

Data assimilation is a class of statistical methods for combining model results and observations. The algorithm weighs these two pieces of information according to their respective error variances and covariances. As output the assimilation returns a result in model space of which the error variances are smaller than those of the original model estimate. In our case the model variables are the mass mixing ratios of aerosol components in a three-dimensional discretized model domain. These model variables are summarized in a vector  $\mathbf{x}$ . The model provides us with a background (or first guess) estimate  $\mathbf{x}_b$  (with an error  $\epsilon_b$ ). The observations, summarized in a vector  $\mathbf{y}$ , are related to the model state  $\mathbf{x}$  by

$$\mathbf{y} = \hat{H}(\mathbf{x}) + \epsilon_o, \quad (1)$$

where  $\hat{H}$  is known as the observation operator, and  $\epsilon_o$  denotes the vector of observation errors. The problem is to determine the most likely state vector  $\mathbf{x}_a$  given  $\mathbf{x}_b$  and  $\mathbf{y}$ , and given the background error covariance matrix  $\mathbf{B} = \langle \epsilon_b \cdot \epsilon_b^T \rangle$ , and the observation error covariance matrix  $\mathbf{R} = \langle \epsilon_o \cdot \epsilon_o^T \rangle$ . Here  $\langle \dots \rangle$  denotes the expectation value. In the three-dimensional variational method (3DVAR), the maximum-likelihood solution is found by numerically minimizing the cost function

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

Data assimilation is commonly employed for constraining model results by use of observations. However, one can also employ data assimilation as an inverse-modelling tool, i.e. for retrieving a model state from measurements. A summary of the theoretical basis of variational data assimilation is given in Appendices B–D.<sup>2</sup>

The MATCH model contains a 3DVAR data assimilation module. This model uses a spectral method, i.e. the model state vector is Fourier transformed in the two horizontal coordinates. All error correlations in the horizontal direction are assumed to be homogeneous and isotropic. The background error covariance matrix is modelled with a method that follows similar principles to the NMC method (Parrish and Derber, 1992). A more complete description of our 3DVAR program can be found in Kahnert (2008).

## 2.4 Analysis of the information content of aerosol optical parameters

The questions we ask are these:

1. Suppose we have an  $n$  dimensional model space. Given  $m$  observations (e.g.  $m_1$  different parameters at  $m_2$  different wavelengths, so that  $m_1 \cdot m_2 = m$ ), how many independent model variables  $N \leq n$  can we constrain with the observations? Obviously, the best we can achieve would be  $N = \min\{m, n\}$ , but often we will have  $N < \min\{m, n\}$ .
2. Which are the  $N$  model variables (or linear combinations of model variables) that can be constrained by the measurements?

<sup>2</sup>Many authors distinguish between data *assimilation* and data *analysis*. In data analysis one merely post-processes a model results by incorporating the information provided by observations. In data assimilation, the data analysis process is part of the time integration of the CTM. Thus, in each time step the result of the analysis becomes the new initial state for the next model forecast. Our 3DVAR code can be used in either analysis or assimilation mode. However, in this study we only perform numerical tests at a fixed point in time. Thus we use the 3DVAR code as a data analysis tool.

**Table 1.** Refractive indices at the three harmonics of the Nd:YAG laser assumed in the MATCH mass-transport optics model.

Wavelength ( $\mu\text{m}$ )	0.355	0.532	1.064
SIA	$1.53 + 5.0e - 3i$	$1.53 + 5.6e - 3i$	$1.52 + 1.6e - 2i$
Dust	$1.53 + 1.7e - 2i$	$1.53 + 6.3e - 3i$	$1.53 + 4.3e - 3i$
NaCl	$1.51 + 2.9e - 7i$	$1.50 + 1.0e - 8i$	$1.47 + 2.0e - 4i$
OC	$1.53 + 5.0e - 3i$	$1.53 + 5.6e - 3i$	$1.52 + 1.6e - 2i$
EC	$1.66 + 7.2e - 1i$	$1.73 + 6.0e - 1i$	$1.82 + 5.9e - 1i$

Here we only give a summary of the most essential theoretical tools for answering these questions. A more thorough explanation of these concepts is given in Appendix C.

First we want to explain what we mean by *signal degrees of freedom* and *noise degrees of freedom*, closely following an example in Rodgers (2000) (p. 29f). Suppose we have a direct measurement  $y$  of a scalar variable  $x$  with error  $\epsilon_0$ , i.e.

$$y = x + \epsilon_0. \quad (3)$$

Suppose further that we have a background estimate  $x_b$  with background error variance  $\sigma_b^2$ , and that the error  $\epsilon_0$  has variance  $\sigma_0^2$ . The prior variance of  $y$  is given by  $\sigma_y^2 = \sigma_b^2 + \sigma_0^2$ , assuming that background and observation errors are uncorrelated. One can show that the best estimate  $x_a$  of  $x$  will be

$$x_a = \frac{\sigma_b^2 y + \sigma_0^2 x_b}{\sigma_b^2 + \sigma_0^2}. \quad (4)$$

Hence, if  $\sigma_b^2 \gg \sigma_0^2$ , then the measurement  $y$  will provide information for estimating  $x_a$ , i.e. the measurement provides a *degree of freedom for signal*. However, if  $\sigma_b^2 \ll \sigma_0^2$ , then  $x_a$  will be close to  $x_b$ , and  $y$  provides little information to estimating  $x_a$ . The measurement mostly contains information on  $\epsilon_0$ , i.e. it provides a *degree of freedom for noise*.

In a more general case we have to consider a state vector  $\mathbf{x}$  and a set of measurements  $\mathbf{y}$  with errors  $\boldsymbol{\epsilon}_0$ . The number  $N_s$  of signal degrees of freedom is a measure for the information content of the set of measurements. It provides us with an estimate of the number  $N$  of model variables that can be controlled by assimilating measurements.

The mapping from model space to observation space given in Eq. (1) can be Taylor expanded to first order according to

$$\mathbf{y} = \hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x} + \boldsymbol{\epsilon}_0, \quad (5)$$

where  $\hat{H}$  is the observation operator,  $\mathbf{H}$  denotes its Jacobian, and  $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$ . The background or prior estimate  $\mathbf{x}_b$  is often obtained from a model run. The (in general non-square) matrix  $\mathbf{H}$  is the main quantity we need to investigate in order to address the questions formulated at the beginning of this section. It is transformed to the so-called observability matrix  $\tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2}$ , where  $\mathbf{R}$  is the observation error covariance matrix, and  $\mathbf{B}$  denotes the error covariance matrix of the background estimate. Subsequently, one performs

a singular-value decomposition (SVD)

$$\mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2} = \mathbf{V}_L \cdot \mathbf{W} \cdot \mathbf{V}_R^T, \quad (6)$$

where the matrices  $\mathbf{V}_L$  and  $\mathbf{V}_R$  contain the left and right singular vectors, respectively, and  $\mathbf{W}$  is a matrix that contains the singular values along the main diagonal, while all other matrix elements are zero. It turns out that the singular values  $w_i$  can be employed to compute the number of signal degrees of freedom  $N_s$  according to

$$N_s = \sum_{i=1}^{\min\{n,m\}} w_i^2 / (1 + w_i^2). \quad (7)$$

Another useful measure is obtained by expressing our incomplete knowledge of the atmospheric aerosol state by use of the Shannon entropy. The use of measurement information reduces the entropy, and this entropy reduction  $H$  can be expressed in terms of the singular values:

$$H = \frac{1}{2} \sum_{i=1}^{\min\{n,m\}} \log_2(1 + w_i^2). \quad (8)$$

Both  $N_s$  or  $H$  allow us to quantify the information content of a set of measurements. More detailed explanations of these concepts are given in Appendix C. A comprehensive discussion of information aspects and inverse methods for atmospheric sounding can be found in Rodgers (2000).

By performing the transformation

$$\delta\mathbf{x}' = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta\mathbf{x} \quad (9)$$

we go from our physical model space to an abstract phase space – see Eq. (C16) in Appendix C. In this phase space the components of  $\delta\mathbf{x}'$  can be separated into signal-related and noise-related variables. The signal-related components can be controlled by the measurements, the noise-related components cannot. We therefore introduce constraints into our 3DVAR program such that only the  $N_s$  signal-related components of  $\delta\mathbf{x}'$  are allowed to be adjusted in the data analysis procedure, while the noise-related components are not altered. This is accomplished by adding an extra term  $J_G$  to the cost function in Eq. (2), where

$$J_G = \frac{1}{2} \delta\mathbf{x}'^T \cdot \mathbf{B}^{-1/2} \cdot \mathbf{V}_R \cdot \mathbf{B}_G^{-1} \cdot \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta\mathbf{x}, \quad (10)$$

and where  $\mathbf{B}_G$  is a diagonal matrix which we assume to have the form

$$\mathbf{B}_G = \sigma_G \text{diag}(w_1, w_2, \dots, w_K, c, \dots, c). \quad (11)$$

Here  $K = \min\{n, m\}$ , and the number  $c$  is assumed to be much smaller than the smallest singular value. We note that the formulation of the constraint term in Eq. (11) is by no means unique. Other possible choices of the matrix  $\mathbf{B}_G$  are discussed in Appendix D3. However, we performed preliminary tests which indicate that the constrained 3DVAR approach is not very sensitive to exactly how one chooses to formulate the matrix  $\mathbf{B}_G$ , as long as it behaves in such a way that the noise-related phase-space variables are tightly constrained, while the signal-related variables can be varied relatively freely by the analysis. The free parameters  $\sigma_G$  and  $c$  should be tuned in such a way that the constraints are neither too hard nor too soft. In the former case, the analysis will stay too close to the background estimate. In the latter case, it will not differ much from the unconstrained analysis.

## 2.5 Numerical test of the constrained assimilation algorithm

We study the performance of the 3DVAR system by performing a numerical test. To this end, we first perform a reference run by driving the MATCH model with analysed meteorological data. These reference results are taken as the “true” chemical state of the atmosphere. We apply the optics model to the model output to generate synthetic “observations”, i.e. a vertical profile at a selected observation point of extinction and backscattering coefficients at three typical lidar wavelengths. Next we run the MATCH model again, this time driven with 48 h forecast meteorological data. The results are taken as a proxy for a background model-estimate that is impaired by uncertainties. Finally, we perform a 3DVAR-analysis of the “observations” and the background estimate in an attempt to restore the reference results. In this numerical test we have perfect knowledge of the true state, and we assume that our optics model is nearly perfect, thus providing nearly perfect observations (we assumed that the observation error standard deviation is 10 % of the measurement value). The only factor that may prevent us from fully restoring the reference state is a lack of information in the observed parameters. Thus, comparison of the retrieval and reference results gives us an indication of how strongly different model variables can be controlled by the information contained in the observations.

We perform this test (i) with the unconstrained 3DVAR algorithm and (ii) with the constrained 3DVAR algorithm. We compare both runs in order to make a first assessment of the impact of the constraints. In particular, we are interested in the prospect of reducing the risk of assimilating noise in such a highly under-constrained inverse problem.

## 3 Results

### 3.1 Analysis of the information content of aerosol optical parameters

We consider the set of parameters  $\{k_{\text{ext}}(\lambda_1), k_{\text{ext}}(\lambda_2), \beta_{\text{sca}}(\lambda_1), \beta_{\text{sca}}(\lambda_2), \beta_{\text{sca}}(\lambda_3)\}$ , where  $k_{\text{ext}}$  and  $\beta_{\text{sca}}$  denote the extinction and backscattering coefficients, respectively, and the wavelengths  $\lambda_1 = 1064$  nm,  $\lambda_2 = 532$  nm, and  $\lambda_3 = 355$  nm denote the first three Nd:YAG harmonics. Hereafter, we will abbreviate these parameters by  $k_{\text{ext}}(\lambda_i) = k_i$ ,  $\beta_{\text{sca}}(\lambda_j) = \beta_j$ ,  $i = 1, 2$ ,  $j = 1, 2, 3$ . Out of this five-parameter set we pick different subsets and analyse the singular values of the corresponding observability matrices. From those we compute the number of signal degrees of freedom as well as the change in Shannon entropy for each subset of measurements. We will focus on those parameter subsets that are technically relevant in practical lidar applications.

Table 2 shows the number of signal degrees of freedom  $N_s$  and the reduction in Shannon entropy  $H$  for different values of the observation standard deviation  $\sigma_o$ . For low values of  $\sigma_o$ , the number of signal degrees of freedom is identical to the number of observational parameters. However, as we increase  $\sigma_o$  we observe a decrease in  $N_s$ . For instance, for  $\sigma_o = 100\%$  the five parameters  $\beta_1 + \beta_2 + \beta_3 + k_2 + k_3$  (last row) only provide roughly  $N_s = 3$  signal degrees of freedom. The reduction in Shannon entropy  $H$  displays an analogous behaviour. For instance, for  $\sigma_o = 1\%$  we see that  $H$  consistently increases as one increases the number of observational parameters. This is much less pronounced for  $\sigma_o = 100\%$ . In that case,  $H$  does increase as one goes from a single parameter to two parameters (compare the first to the second and fourth rows). However, as one adds more parameters, the increase in  $H$  slows down considerably. For five parameters (last row),  $H$  is only about twice as high as for a single parameter (first row).

This illustrates the pivotal importance of the observation error for the amount of information that can be obtained from measurements. It is important to understand that the observation error  $\epsilon_o$  is not the same as the measurement error  $\epsilon_m$ . Rather, in our case we have  $\epsilon_o = \epsilon_m + \epsilon_f$ , where  $\epsilon_f$  denotes the forward-model error (see, e.g., Eq. 1 and accompanying text in Rabier et al., 2002). Any simplifying assumptions in the optics model or incomplete knowledge of the particle size distribution, morphology, chemical composition, or dielectric properties can contribute to  $\epsilon_f$ . Such assumptions enter into our relatively simple optics model.<sup>3</sup> Note also that in operational applications there may be other terms contributing to  $\epsilon_o$ . For instance, if a point measurement is taken at a location that does not provide a good representation of the

<sup>3</sup>A more realistic optics model, such as the one investigated in Andersson and Kahnert (2016) would help to reduce the observation standard deviation. For future studies, such a model should be linearized and investigated in a similar way.

**Table 2.** Number of signal degrees of freedom  $N_s$  and reduction in entropy  $H$  as a function of observation standard deviation, taken from the lowest model layer (closest to the surface). Results are shown for different subsets of  $k_1, k_2, \beta_1, \beta_2, \beta_3$ , where  $k_i$  and  $\beta_i$  represent the extinction and backscattering coefficient, respectively, at the wavelengths  $\lambda_1 = 1064$  nm,  $\lambda_2 = 532$  nm, and  $\lambda_3 = 355$  nm.

No.	Obs. SD (%) Parameters	1		5		10		50		100	
		$N_s$	$H$	$N_s$	$H$	$N_s$	$H$	$N_s$	$H$	$N_s$	$H$
1.	$\beta_3$	1.00	10.9	1.00	8.58	1.00	7.58	1.00	5.26	1.00	4.26
2.	$\beta_1 + \beta_2$	2.00	20.6	2.00	15.99	2.00	13.98	1.97	9.36	1.90	7.42
3.	$\beta_1 + \beta_2 + \beta_3$	3.00	27.3	3.00	20.3	2.99	17.3	2.72	10.5	2.33	8.00
4.	$\beta_3 + k_3$	2.00	19.4	2.00	14.8	2.00	12.8	1.92	8.21	1.74	6.37
5.	$\beta_1 + \beta_2 + k_2$	3.00	28.0	3.00	21.0	2.99	18.0	2.77	11.2	2.42	8.63
6.	$\beta_1 + \beta_2 + \beta_3 + k_2 + k_3$	5.00	40.0	4.97	28.4	4.91	23.5	3.89	12.9	2.97	9.49

grid-cell average, then one would have to add a representativity error  $\epsilon_r$  to the observation error.

The strong impact of the observation errors on the information content of measurements suggests two conclusions.

1. In order to make the forward-model error  $\epsilon_f$  as small as possible, it is essential to develop accurate and realistic aerosol optics models. The most accurate measurements may intrinsically contain a wealth of information on aerosol properties. But we can only make use of this information to the extent that our observation operator is able to accurately describe the relation between the physical and chemical particle characteristics and their optical properties.
2. It is equally essential to accurately estimate the contribution of the uncertainties in the aerosol optics model, i.e. to estimate the forward-model error  $\epsilon_f$ . If we underestimate this error, we will rely too much on the measurements than we should, thus assimilating noise. If we overestimate this error, we will waste information contained in the observations. In practice, one way to estimate  $\epsilon_f$  is to compute optical properties while varying the particles' size, morphology, and dielectric properties within typical ranges. The resulting variation in the optical properties then allows us to estimate  $\epsilon_f$ . (For a review of aerosol optics modelling see Kahnert et al., 2014, 2016, and references therein).

In Table 2 we sorted the results for  $N_s$  and  $H$  by different values of the observation standard deviation. However, it is important to realize that the results also depend on the background error standard deviation, or, more precisely, on how large the background error standard deviations are compared to the observation error standard deviations. Johnson et al. (2005a) made this point very explicit. They discussed an idealized case with diagonal background error covariance matrix  $\mathbf{B} = \sigma_b^2 \mathbf{1}$  and observation error covariance matrix  $\mathbf{R} = \sigma_o^2 \mathbf{1}$ . They considered the case of direct measurements, i.e. the model variables and the observed parameters are the same type of variables. Under such idealized conditions, they showed that one can maximize the amount of

information that can be obtained from the observations by optimizing the regularization parameter  $\sigma_b/\sigma_o$  (or, equivalently, the regularization parameter  $\sigma_o^2/\sigma_b^2$ ). In our more general case, instead of  $\sigma_b$  we need to consider the full matrix  $\mathbf{B}^{1/2}$ , instead of  $\sigma_o^{-1}$  we need to consider  $\mathbf{R}^{-1/2}$ , and in order to compare the two matrices we need to first transform  $\mathbf{B}^{1/2}$  from model to observation space according to  $\mathbf{H} \cdot \mathbf{B}^{1/2}$ . Thus, in place of  $\sigma_b/\sigma_o$  we need to consider the more general quantity  $\mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2}$ , and we need to diagonalize it by a singular value decomposition according to Eq. (6). Thus the singular values  $w_i$  generalize the parameter  $\sigma_b/\sigma_o$ . The latter applies to the case of direct observations and error covariance matrices that are proportional to unit matrices. The former apply to the general case of non-diagonal error covariance matrices and indirect observations.

From this we learn that the singular values  $w_i$  provide us with a (however abstract) means to quantify how the background standard deviations compare to the observation standard deviations. We pick one of the columns in Table 2, namely the one for  $\sigma_o = 50\%$ , and expand it in Table 3. We show the singular values  $w_i$ , as well as their contributions  $N_s^i = w_i^2/(1 + w_i^2)$  and  $H_i = 0.5 \log_2(1 + w_i^2)$  to the sums in Eqs. (7) and (8), respectively. The results reveal that the singular values  $w_i$  can decrease quite rapidly from the largest to the smallest value (see, e.g., case no. 6 in the table). However, the corresponding contribution  $N_s^i$  to the number of signal degrees of freedom changes rather smoothly. Even those singular values that are only slightly larger than 1 make contributions  $N_s^i$  that lie close to 1 (see, e.g.,  $i = 4$  in case no. 6). However, once  $w_i$  falls below 1, the corresponding contribution  $N_s^i$  becomes much smaller than 1 (see  $i = 5$  in case no. 6).

Let us now compare the different subsets of parameters in Tables 2 and 3. In case no. 1 we observe a single parameter that provides a single degree of freedom. In cases no. 2 and 4 we observe two parameters, which nearly doubles  $N_s$ . Comparison of these two cases shows that it does not make a significant difference whether we observe backscattering coefficients at different wavelengths, or both extinction and backscattering coefficients each at a single wavelength. In either case the measurements provide roughly the same amount

**Table 3.** Signal degrees of freedom  $N_s$  and change in entropy  $H$  for the lowest model layer (closest to the surface). Also shown are the singular values  $w_i$  and their contributions  $N_s^i$  and  $H_i$  to  $N_s$  and  $H$ , respectively. The results have been obtained by assuming an observation standard deviation of 50 %.

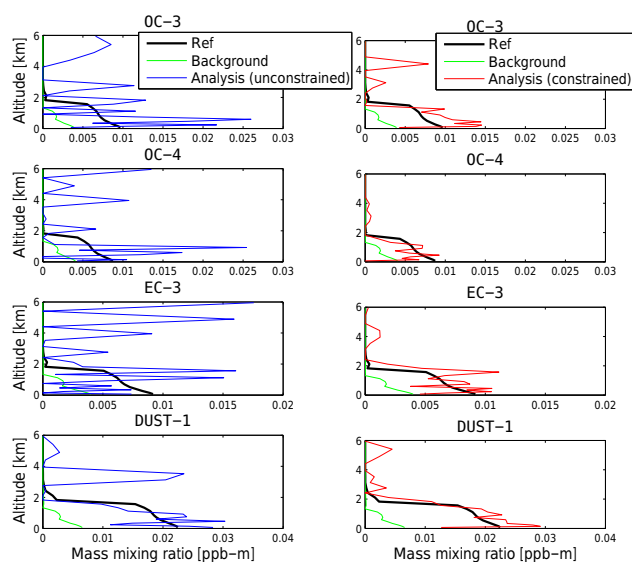
No.	Parameters	$i$	$w_i$	$N_s^i$	$H_i$	$N_s$	$H$
1.	$\beta_3$	1	38.2	1.00	5.26	1.00	5.26
2.	$\beta_1, \beta_2$	1	108	1.00	6.76	1.97	9.36
		2	6.00	0.97	2.61		
3.	$\beta_1, \beta_2, \beta_3$	1	115	1.00	6.84	2.71	10.5
		2	6.54	0.98	2.73		
		3	1.68	0.74	0.97		
4.	$\beta_3, k_3$	1	83.3	1.00	6.38	1.92	8.22
		2	3.43	0.92	1.84		
5.	$\beta_1, \beta_2, k_2$	1	128	1.00	7.00	2.77	11.24
		2	8.71	0.99	3.13		
		3	1.90	0.78	1.10		
6.	$\beta_1, \beta_2, \beta_3, k_2, k_3$	1	153	1.00	7.26	3.89	12.9
		2	9.52	0.99	3.26		
		3	1.94	0.79	1.13		
		4	1.63	0.73	0.93		
		5	0.79	0.38	0.35		

of information (in terms of  $N_s$  or  $H$ ). The same is true when considering three observational parameters (compare case nos. 3 and 5). The  $3\beta + 2\alpha$  case (no. 6) clearly provides the largest amount of information in comparison to the other cases. However, as we saw in Table 2, observation errors that are large in comparison to the background errors can significantly reduce the effective information that can be assimilated into a model.

### 3.2 Numerical inverse-modelling test

We integrated the findings of Sect. 3.1 into our 3DVAR program by constraining the algorithm to varying only the signal-related model variables. To illustrate the method we conduct a numerical test as described in Sect. 2.5. We perform a 3DVAR analysis by assimilating “ $3\beta + 2\alpha$ ” profiles, i.e. synthetic lidar measurements of  $\beta_{\text{sca}}$  at the three wavelengths 1064, 532, and 355 nm together with  $k_{\text{ext}}$  at the two wavelengths 532 and 355 nm. Thus in our case the number of singular values in each vertical layer is  $K = 5$ . We assume an idealized situation in which the observation standard deviation is only 10 %. As we see in Table 2 (case no. 6), the number of signal degrees of freedom is  $N_s = 4.9$  in this case. Thus, we roughly have as many signal degrees of freedom as we have measurements.

Figure 1 shows vertical profiles of selected aerosol components, namely (from top to bottom): organic carbon (OC) in the third size bin (OC-3), OC in the fourth size bin (OC-4), elemental carbon (EC) in the third size bin (EC-3), and mineral dust in the first size bin (DUST-1). The reference and

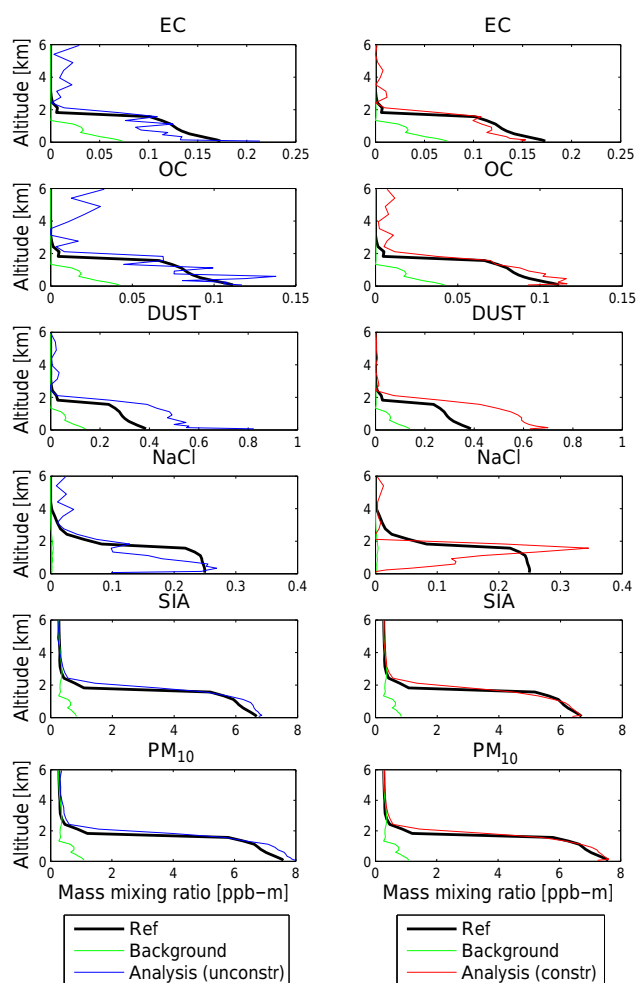


**Figure 1.** Vertical profiles of selected aerosol components in different size bins. From top to bottom: organic carbon in the third size bin (OC-3), OC in the fourth size bin (OC-4), elemental carbon in the third size bin (EC-3), and dust in the first size bin (DUST-1). The reference results are shown in black, and the background (first guess) estimate is shown in green. The unconstrained 3DVAR analysis results are presented in the left panels in blue, the constrained 3DVAR analysis results are shown in the right panels in red.

background mixing ratios are shown in black and green, respectively. The 3DVAR analysis was first performed without any constraints; the results are shown in the left column by the blue line. Then the 3DVAR analysis was repeated with the constraints in Eqs. (10) and (11); the results are represented in the right column by the red line. Clearly, the unconstrained analysis (blue lines in the left panels) yields results that oscillate quite erratically in the vertical direction. Also, the unconstrained analysis can yield conspicuously high values at higher altitudes, even though the reference and background values are both close to zero. By contrast, the constrained analysis (red lines in the right panels) yields results that better agree with the reference results. The noisiness in the vertical direction is significantly reduced, and the results at higher altitudes are generally lower than those obtained with the unconstrained analysis.

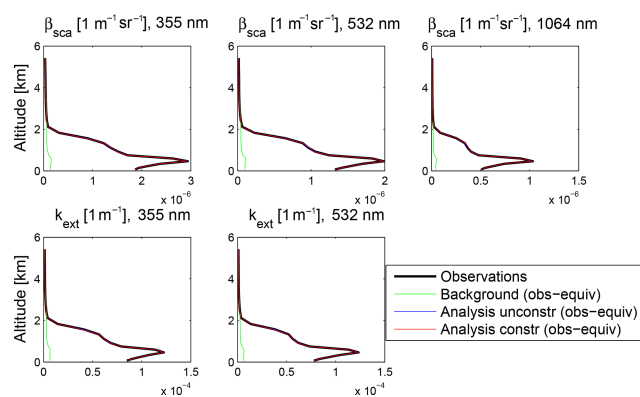
Figure 2 shows analogous results for the mass mixing ratios of different aerosol components, each summed over all size bins. The aerosol components are (from top to bottom): elemental carbon (EC), organic carbon (OC), mineral dust (DUST), sea salt (NaCl), secondary inorganic aerosols (SIA, i.e. the sum over all sulfate, nitrate, and ammonium species), and PM<sub>10</sub> (i.e. the sum over all aerosol components). Clearly, the constrained analysis faithfully retrieves both PM<sub>10</sub> and SIA. The unconstrained analysis performs almost equally well for these two variables. Sea salt and mineral dust are not well retrieved from the measurements in either the con-





**Figure 2.** As Fig. 1, but for the total mass mixing ratio (summed over all size bins). The components are (from top to bottom): EC, OC, mineral dust, sea salt, secondary inorganic aerosols (sum of all sulfate, nitrate, and ammonium species), and  $PM_{10}$  (sum of all aerosol components).

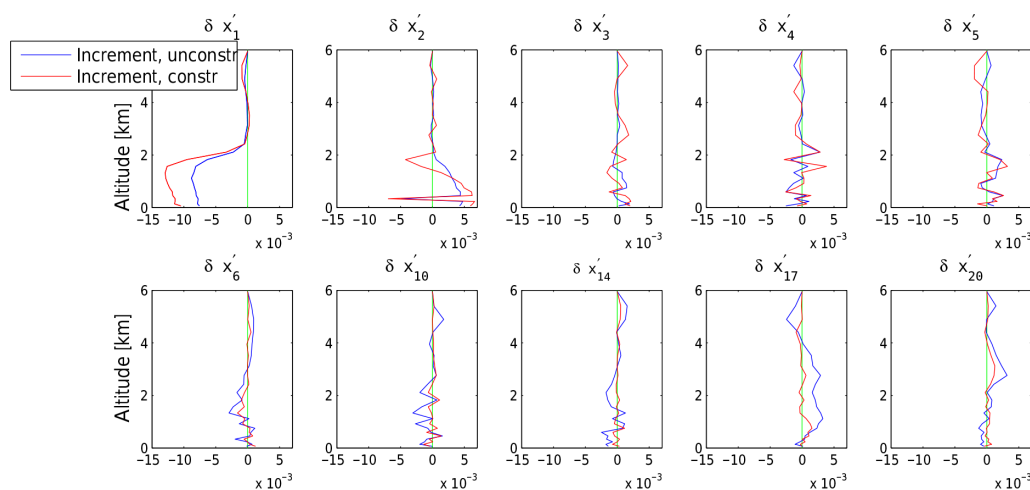
strained or unconstrained approach. EC and OC are very well retrieved by the constrained analysis. For these components, the unconstrained analysis has a very small bias compared to the reference results, but it is considerably more noisy (i.e. oscillating in the vertical direction) than the constrained analysis. We also see, again, that the mixing ratios at higher altitudes obtained with the unconstrained analysis can be unreasonably high. This is especially pronounced for OC. In general, however, the problems we encounter in the unconstrained analysis are less pronounced in Fig. 2 than in Fig. 1. A possible explanation is that SIA may be most strongly related to the measurement signal, and SIA is dominating the aerosol mass in this case. We will return to this point shortly. Another possible factor is that the noise in the analysis can be damped by summing up results over several size bins.



**Figure 3.** Observations (black solid line), and observation-equivalents of the background estimate (green), and of the unconstrained (blue) and constrained (red) 3DVAR analysis. The optical parameters and wavelengths are indicated above each panel.

Figure 3 shows the observations (black) as well as the observation-equivalents of the background estimate (green) and the unconstrained (blue) and constrained (red) 3DVAR analysis for all five observations. We learn from this figure that the analysis follows the observations faithfully. The reason for this is that we assumed that the observations were highly accurate with an error standard deviation of only 10%. In fact, the difference between the observation-equivalent analysis and the observations deviate by even less than 10%. However, our tests confirmed that an increase in the observation error eventually results in analysis results of which the observation-equivalent increasingly deviates from the observations (not shown).

We have seen that the analysis provides a reasonable, but, as expected, not a perfect answer to the inverse problem. We have further seen that at the observation site it relies more on the observations than on the background estimate. Most importantly, we have seen that the constraints introduced in the 3DVAR algorithm suppress noise in the analysis, especially in EC and OC. However, the previous figures do not provide us with any direct insight of how exactly the constraints accomplish this. To learn more about that we need to inspect the analysis in the abstract phase space of the transformed model variables  $\delta\mathbf{x}'$ . (Recall that we defined this variable in Eq. 9 as  $\delta\mathbf{x}' = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot (\mathbf{x} - \mathbf{x}_b)$ ). Figure 4 shows vertical profiles of a selection of the, in total, 20 variables  $\delta x'_i$ . The background estimate corresponds to  $\delta x'_i = 0$  and is represented by the green line. The unconstrained 3DVAR analysis increment is represented by the blue line, the constrained 3DVAR analysis increment is shown by the red line. The first five phase space elements in the top row are the signal-related control variables. Generally, the magnitude of the constrained increments (red) is larger than that of the unconstrained increments (blue). The noise-related phase space elements, five of which are shown in the bottom row, display the opposite behaviour. The constrained increments are



**Figure 4.** Vertical profiles of the transformed model variables  $\delta x'_i$ , defined in Eq. (9). The figure shows results obtained with the constrained (red) and unconstrained (blue) 3DVAR analysis.

close to zero, as they should be. The unconstrained elements consistently show higher magnitudes than the constrained elements. However, we also see that the unconstrained analysis does produce increments that are largest for the two elements  $\delta x'_1$  and  $\delta x'_2$ , which most strongly relate to the measurement signal. Based on our single test case we cannot say if this is a lucky coincidence or a consistent property. If the latter, it may indicate that we are using rather reasonable background error statistics, so that the analysis increment in observation space is distributed to the different variables in model space in a sensible way. If the former, it could be the case that the success of the unconstrained analysis is largely dependent on whether or not those aerosol components dominate the total aerosol mass that most strongly relate to the signal degrees of freedom. (In our case the total mass is dominated by SIA, which is very well retrieved by the analysis).

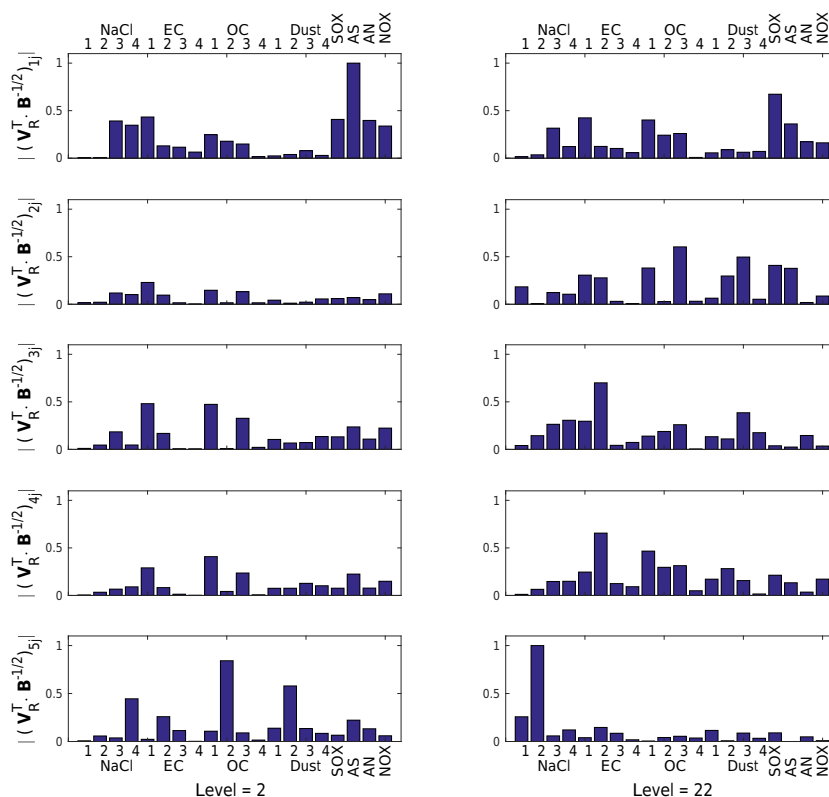
Finally, we want to obtain a better understanding of how the aerosol components  $x$  in model space, or their increments  $\delta x$ , are linked with the signal-related phase-space elements  $\delta x'$ . To this end we inspect the first five row vectors of the transformation matrix  $\mathbf{V}_R^T \cdot \mathbf{B}^{-1/2}$  in Eq. (9). The magnitude of these elements can be taken as a measure for how much each aerosol component of  $\delta x$  in model space contributes to the signal-related elements of  $\delta x'$ . Figure 5 shows  $|(\mathbf{V}_R^T \cdot \mathbf{B}^{-1/2})_{ij}|$  for  $i = 1, \dots, 5$ , and for  $j = 1, \dots, 20$ , where 5 is the number of signal-related phase-space elements, and 20 is the number of aerosol components in model space. Results are shown for model layers 2 (left column) and 22 (right column), which correspond to altitudes of about 100 m and 6 km, respectively. The  $x$  axis shows sea salt (NaCl), EC, OC, and dust, each in four size bins, as well as the four SIA components, i.e. sulfates ( $\text{SO}_x$ ) other than  $(\text{NH}_4)_2\text{SO}_4$ , ammonium sulfate (AS), ammonium nitrate (AN), and nitrates ( $\text{NO}_x$ ) other than  $\text{NH}_4\text{NO}_3$ .

Comparison of the two columns clearly demonstrates that the elements of the transformation matrix can vary considerably with vertical layer (or, more generally, with location). This is because the error covariance matrix  $\mathbf{B}$  varies with location, and the matrix  $\mathbf{R}$  varies from one observation site to another (in our case, from one altitude to another). Hence the matrix  $\mathbf{V}_R$  is also dependent on location – see Eq. (6). Consequently, it is very difficult to draw general conclusions about which aerosol components make a dominant contribution to the signal-related phase-space variables; this can vary with location, and it can vary for different data sets.

However, in our case the SIA components consistently make a strong contribution to the first signal-related element  $\delta x'_1$ . Since SIA is dominating the aerosol mass mixing ratio in this test case, the analysis was able to retrieve  $\text{PM}_{10}$ . We also see that the dust components make only a weak contribution to most of the signal-related elements  $\delta x'_i$ , especially to the first one. This is a likely explanation for the difficulties encountered in retrieving the dust mass mixing ratio. Sea salt is more complicated. Size bins 3 and 4 do contribute considerably to  $\delta x'_1$ , and also to some of the other four increments, while size bins 1 and 2 do not make a significant contribution to most of the five signal-related control variables. In our case the sea salt mass is strongly dominated by the second size bin (not shown). This explains the difficulties we encountered in the retrieval of sea salt.

#### 4 Summary and conclusions

We have quantified the information content of multiwavelength lidar measurements with regard to the chemical composition of aerosol particles. Different combinations of extinction and backscattering observations at several wavelengths have been investigated by determining the singular values of the scaled observation operator, by computing the



**Figure 5.** The first five rows (from top to bottom) of the matrix  $\mathbf{V}_R^T \cdot \mathbf{B}^{-1/2}$  at the observation site, and for model layers 2 (left) and 22 (right). The y values are normalized by dividing them by the maximum element. The x axis indicates the aerosol components in model space to which the elements of the row vectors correspond, namely, sea salt (NaCl), EC, OC, and dust, each in four size bins, as well as the four SIA components: sulfates ( $\text{SO}_x$ ) other than  $(\text{NH}_4)_2\text{SO}_4$ , ammonium sulfate (AS), ammonium nitrate (AN), and nitrates ( $\text{NO}_x$ ) other than  $\text{NH}_4\text{NO}_3$ .

number of signal degrees of freedom  $N_s$ , and by calculating the reduction in Shannon entropy  $H$  caused by taking measurements. We first quantified  $N_s$  and  $H$  as a function of observation standard deviation  $\sigma_o$ . The information content of the observations, as expressed by  $N_s$  and  $H$ , decreased as  $\sigma_o$  was increased. This became the more pronounced the larger the number of simultaneously observed parameters was.

The observation error depends not only on the measurement error, but also on the forward-model error. The latter depends on the uncertainties in the aerosol optics model. This highlights the importance of developing accurate aerosol optics models and of obtaining an accurate estimate of the observation error, especially of the uncertainty in the aerosol optics model. This is a prerequisite for extracting as much information as possible from the measurements, while avoiding to extract noise rather than signal. More often than not, computational limitations and lack of knowledge force us to introduce simplifying assumptions about the particles' morphologies. However, we know that aerosol optical properties can be highly sensitive to the shape (Mishchenko et al., 1997; Kahnert, 2004), small-scale surface roughness (Kahnert et al., 2012b), inhomogeneity (Mishchenko et al., 2014;

Kahnert, 2015), aggregation (Fuller and Mackowski, 2000; Liu and Mishchenko, 2007; Kahnert and Devasthale, 2011), irregularity (Muinonen, 2000; Bi et al., 2010), porosity (Vilaplana et al., 2006; Lindqvist et al., 2011; Kylling et al., 2014), and combinations thereof (Lindqvist et al., 2009, 2014; Kahnert et al., 2013). We need to know how much these sources of uncertainty contribute to the observation standard deviation. One way of estimating this is to compare aerosol optical properties computed with simple shape models to either measurements or to computations based on more realistic particle shape models – see Kahnert et al. (2014) for a recent review and a more detailed discussion.

The singular values of the scaled observation operator provide us with an abstract measure to compare the standard deviations of the background (prior) estimate to those of the observations. The reason why this is a rather abstract measure is because background and observation errors are, in general, in different spaces and cannot be directly compared. However, we constructed a mapping that transforms the state vector in physical (model) space to an abstract phase space in which the components of the state vector can be partitioned into signal-related and noise-related components. The singu-

lar values indicate to what extent the signal-related phase-space variables can be constrained by the measurements. We exploited this fact by constructing weak constraints in a 3DVAR data assimilation code, which limited the assimilation algorithm to acting on the signal-related phase-space variables only (hereafter referred to as the *constrained analysis*). The idea was to maximize the use of information, while avoiding the risk of assimilating noise by overusing the measurements. Thus, our main hypothesis was that the constrained analysis will yield less noisy results than the unconstrained analysis. Numerical tests confirmed this hypothesis. Notably in the case of elemental carbon (EC) and organic carbon (OC) the unconstrained analysis gave mixing ratios that oscillated considerably in the vertical direction. The constrained analysis results were considerably less noisy.

When mapped into observation space, the analysis result closely reproduced the measurements. When viewed in the abstract phase space, we found that the constrained analysis did, indeed, yield noise-related components that were close to zero, as they should be. This was not so in the unconstrained analysis. Also, the magnitude of the signal-related phase-space components was generally larger in the constrained analysis than in the unconstrained analysis. This confirms that the constraints we introduced work as intended.

In our specific test case secondary inorganic aerosol components were most faithfully retrieved by the inverse modelling solution, followed by organic and black carbon. Dust and sea salt mass mixing ratios were more challenging to retrieve. We could explain this by inspecting the linear coefficients in the transformation from physical space to the abstract phase space. We found that those aerosol components that had the largest weight in the transformation were most faithfully retrieved by the analysis. However, these linear coefficients depend on the background error covariances (which can change with location), and on the observation error variances. Therefore, it is difficult to draw general conclusions about which aerosol components are most easily retrieved by a given set of measurements.

The results presented here suggest further questions for future studies. We have performed this investigation with a mass transport model, thus focusing on the information content of optical measurements with respect to the chemical composition of aerosols. When we include aerosol microphysical processes, then the model delivers the aerosols' size distribution, as well as their size-resolved chemical composition. This makes the problem quite different from that we investigated here. First, the dimension of the model space is considerably larger for an aerosol microphysics transport model. Constraining such a model with limited information from measurements becomes even more challenging than in the case of a mass transport model. On the other hand, an aerosol microphysics model delivers information on the particles' size distribution and mixing state. Therefore, this would require us to make fewer assumptions in the aerosol optics model, which may reduce the observation error. The

present study could be extended to investigate the information contained in extinction and backscattering measurements for simultaneously constraining the chemical composition and the size of aerosol particles.

Another important issue concerns the choice of the aerosol optics model. In the present study we employed a simple homogeneous-sphere model in which all chemical components were assumed to be externally mixed. There is little one can put forward in defence of this model other than pure convenience. (Regarding the applicability of simplified model particles in atmospheric optics see the review by Kahnert et al., 2014). As a result of the external-mixture assumption, the observation operator is linear, which is a prerequisite for much of the theoretical foundations of this study – see Appendices B–D for details. However, it has been demonstrated that drastically simplifying assumptions, such as the external-mixture approximation, can give model results for aerosol optical properties that differ substantially from those obtained with more realistic nonlinear optics models (Andersson and Kahnert, 2016). It would therefore be important to extend the present study to include more accurate and realistic optics models. A first step could be to analyse the degree of nonlinearity of optics models that account for internal mixing of different aerosol species. If they turn out to be only mildly nonlinear, then one can linearize them and work with the Jacobian of the nonlinear observation operator. Otherwise the theoretical methods employed in this paper would have to be extended in order to accommodate nonlinear observation operators.

*Data availability.* The data used in this study are included in the Supplement.

## Appendix A: Inverse problems

Suppose we have a system described by a set of variables  $x_1, \dots, x_n$ , summarized in a vector  $\mathbf{x}$ . Suppose also that we have an operator  $\hat{H}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\mathbf{x} \mapsto \mathbf{y} = \hat{H}(\mathbf{x})$  that allows us to compute a set of variables  $y_1, \dots, y_m$ , summarized in a vector  $\mathbf{y}$ . To take a specific example, we may think of  $\mathbf{x}$  as a vector of mass mixing ratios of chemical aerosol species,  $\mathbf{y}$  as a set of aerosol optical properties, and  $\hat{H}$  as an aerosol optics model. The operator  $\hat{H}$  maps from model space into observation space, which allows us to compare model output and observations. We consider the following two problems:

1. *Direct problem*: given  $\mathbf{x}$  and  $\hat{H}$ , calculate  $\mathbf{y} = \hat{H}(\mathbf{x})$ .
2. *Inverse problem*: given  $\mathbf{y}$  and  $\hat{H}$ , solve  $\mathbf{y} = \hat{H}(\mathbf{x})$  for  $\mathbf{x}$ .

A pair of such problems is inverse to each other; it is, therefore, somewhat arbitrary which problem we choose to call the direct problem, and which one we call the inverse problem. However, one of the problems is usually *well posed*, while the other one is *ill-posed*. Such is also the case in aerosol optics modelling. It is customary to call the well-posed problem the *direct problem*, and the ill-posed one the *inverse problem*.

An equation  $\mathbf{y} = \hat{H}(\mathbf{x})$  is called *well posed* if it has the following properties:

1. *Existence*: for every  $\mathbf{y} \in \mathbb{R}^m$ , there is at least one  $\mathbf{x} \in \mathbb{R}^n$  for which  $\mathbf{y} = \hat{H}(\mathbf{x})$ .
2. *Uniqueness*: for every  $\mathbf{y} \in \mathbb{R}^m$ , there is at most one  $\mathbf{x} \in \mathbb{R}^n$  for which  $\mathbf{y} = \hat{H}(\mathbf{x})$ .
3. *Stability*: the solution  $\mathbf{x}$  depends continuously on  $\mathbf{y}$ .

If any of these properties is not fulfilled, then the problem is called *ill-posed*.

## Appendix B: Three-dimensional variational data assimilation

Data assimilation is usually employed for constraining models by use of measurements, but it can also be used to solve inverse problems. Here we focus on one specific data assimilation method known as three-dimensional variational data assimilation, or 3DVAR.

In a CTM we discretize the geographic domain of interest into a three-dimensional grid. In each grid cell, the aerosol particles are characterized by the mass mixing ratio of each chemical component in the aerosol phase, such as sulfate, nitrate, ammonium, mineral dust, black carbon, organic carbon, and sea salt. Suppose we summarize all these mass mixing ratios from all grid cells into one large vector  $\mathbf{x} \in \mathbb{R}^n$ . The model provides us with a first guess of the atmospheric

aerosol state, known as a *background estimate*  $\mathbf{x}_b$ .<sup>4</sup> Suppose also that we have  $m$  observations, which we summarize in a vector  $\mathbf{y} \in \mathbb{R}^m$ . We further have an observation operator  $\hat{H}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\mathbf{x} \mapsto \hat{H}(\mathbf{x})$  that maps the state vector  $\mathbf{x}$  from model space to observation space.<sup>5</sup> We further denote by  $\mathbf{x}_t$  the true state of the atmosphere, by  $\boldsymbol{\epsilon}_b = \mathbf{x}_t - \mathbf{x}_b$  the error of the background estimate, and by  $\boldsymbol{\epsilon}_o = \hat{H}(\mathbf{x}_t) - \mathbf{y}$  the observation error.<sup>6</sup> The background and observation errors are assumed to be unbiased and uncorrelated with each other. Then their joint probability distribution becomes separable, i.e.

$$P(\boldsymbol{\epsilon}_b, \boldsymbol{\epsilon}_o) = P_b(\boldsymbol{\epsilon}_b)P_o(\boldsymbol{\epsilon}_o). \quad (\text{B1})$$

The true state of the atmosphere is, of course, unknown. Therefore, our definition of the errors and their probability distribution is only of conceptual use, but not of any practical value. However, we can reinterpret the probability distributions by replacing  $\boldsymbol{\epsilon}_b$  in the argument of  $P_b$  with  $\mathbf{x} - \mathbf{x}_b$ , and by replacing  $\boldsymbol{\epsilon}_o$  in the argument of  $P_o$  with  $\hat{H}(\mathbf{x}) - \mathbf{y}$ . We further assume that both the background and the observation errors are normally distributed. Thus we may write

$$P_b(\mathbf{x}) = (2\pi |\mathbf{B}|)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \cdot \mathbf{B}^{-1} \cdot (\mathbf{x} - \mathbf{x}_b)\right) \quad (\text{B2})$$

$$P_o(\mathbf{x}) = (2\pi |\mathbf{R}|)^{-1/2} \exp\left(-\frac{1}{2}(\hat{H}(\mathbf{x}) - \mathbf{y})^T \cdot \mathbf{R}^{-1} \cdot (\hat{H}(\mathbf{x}) - \mathbf{y})\right). \quad (\text{B3})$$

Here  $\mathbf{B}$  and  $\mathbf{R}$  denote the covariance matrices of the background and observation errors, respectively, and  $|\cdot|$  denotes the matrix determinant. In this form,  $P_b(\mathbf{x})$  represents the probability that the atmospheric aerosol particles are found in state  $\mathbf{x}$ , given a background estimate  $\mathbf{x}_b$  with error covariance matrix  $\mathbf{B}$ . Similarly,  $P_o(\mathbf{x})$  is the probability that the

<sup>4</sup>In the remote sensing and inverse modelling community, the background estimate is more commonly referred to as the a priori estimate.

<sup>5</sup>The optics model  $\hat{H}$  usually has to invoke assumptions about physical aerosol properties that are relevant for the optical properties, but not provided by the CTM output, e.g. assumptions about the morphology of the particles. If the CTM is a simple mass-transport model without aerosol microphysics, then it is also necessary to invoke assumptions about the size distribution of the aerosols.

<sup>6</sup>We stress, once more, that the observation error must not be confused with the measurement error  $\boldsymbol{\epsilon}_m$ . The latter contributes to the former, but the observation error also contains other sources of error. For instance, if we deal with morphologically complex particles, but our lack of knowledge forces us to make assumptions and invoke approximations about the particle shapes, then this forward-model error  $\boldsymbol{\epsilon}_f$  contributes to the observation error. The same is the case if we lack information about the particles' size distribution. In operational applications the representativity error  $\boldsymbol{\epsilon}_r$  can also make a substantial contribution to  $\boldsymbol{\epsilon}_o$ .

system is found in state  $\mathbf{x}$ , given measurements  $\mathbf{y}$  with error covariances  $\mathbf{R}$ .<sup>7</sup>

Equations (B1)–(B3) can be summarized in the form

$$P(\mathbf{x}) = \frac{1}{2\pi(|\mathbf{B}| \cdot |\mathbf{R}|)^{1/2}} \exp(-J(\mathbf{x})) \quad (\text{B4})$$

$$J(\mathbf{x}) = \frac{1}{2} \left[ (\mathbf{x} - \mathbf{x}_b)^T \cdot \mathbf{B}^{-1} \cdot (\mathbf{x} - \mathbf{x}_b) + (\hat{H}(\mathbf{x}) - \mathbf{y})^T \cdot \mathbf{R}^{-1} \cdot (\hat{H}(\mathbf{x}) - \mathbf{y}) \right], \quad (\text{B5})$$

where  $J$  is suggestively called the cost function, since it can be interpreted as a measure for how “costly” it is for a state  $\mathbf{x}$  to simultaneously deviate from the background estimate and the measurements within the permitted error bounds. The deviations are weighted with the inverse error covariance matrices. For instance, this means that for measurements with a small error variance, a deviation  $\hat{H}(\mathbf{x}) - \mathbf{y}$  becomes “more costly”.

We are interested in the most probable aerosol state of the atmosphere, i.e. in that state  $\mathbf{x}_a$  for which the probability distribution attains its maximum. This is obviously the case when the argument of the exponential in Eq. (B4) assumes a minimum. Thus we seek to minimize the cost function  $J$ . The variational method is based on computing the gradient of the cost function,  $\nabla_{\mathbf{x}} J$ , and to use this in a descent algorithm to iteratively search for the minimum of  $J$ .

In practice it is common to introduce the variable  $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$ , and use the first-order Taylor expansion of the observation operator,

$$\hat{H}(\mathbf{x}) = \hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x}, \quad (\text{B6})$$

where the ( $m \times n$ ) matrix  $\mathbf{H}$  denotes the Jacobian of  $\hat{H}$  at  $\mathbf{x} = \mathbf{x}_b$ . If  $\hat{H}$  is only mildly nonlinear, and if the components of  $\delta\mathbf{x}$  are sufficiently small, then we can substitute this first-order approximation into Eq. (B5), which yields

$$J = J_b + J_o \quad (\text{B7})$$

$$J_b(\delta\mathbf{x}) = \frac{1}{2} \delta\mathbf{x}^T \cdot \mathbf{B}^{-1} \cdot \delta\mathbf{x} \quad (\text{B8})$$

$$J_o(\delta\mathbf{x}) = \frac{1}{2} \left( \hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x} - \mathbf{y} \right)^T \cdot \mathbf{R}^{-1} \cdot \left( \hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x} - \mathbf{y} \right). \quad (\text{B9})$$

The components of the vector  $\delta\mathbf{x}$  are the *control variables* that are iteratively varied by the algorithm until the minimum of the cost function is found.

The solution to the equation  $\nabla_{\mathbf{x}} J = \mathbf{0}_n$  is a solution to the inverse problem (where  $\mathbf{0}_n$  denotes the null vector in  $n$ -dimensional model space); we input the observations  $\mathbf{y}$  into the algorithm, and as output we obtain a result in model space

<sup>7</sup>The observation errors are often assumed to be uncorrelated (this is not always true). In such a case the matrix  $\mathbf{R}$  is diagonal, where the diagonal elements are the observation error variances.

that is consistent with the measurements (within the given error bounds).<sup>8</sup> What if the measurements contain insufficient information about the state  $\mathbf{x}$ ? The algorithm will still provide an answer to the inverse problem, but the missing information will be supplemented by the background estimate  $\mathbf{x}_b$ . The weighting of the two pieces of information,  $\mathbf{x}_b$  and  $\mathbf{y}$ , is controlled by the respective error covariance matrices. Thus data assimilation is a statistical approach, which can be expected to give good results *on average*, but not in every single time step of the model run. This can become highly problematic if we only have very few observations, i.e.  $m \ll n$ , where  $n$  is the dimension of the model space. If we allow all model variables to be freely adjusted by the assimilation algorithm in such a severely under-constrained case, then the algorithm may just assimilate noise from the measurements rather than signal, resulting in unreasonable solutions to the inverse problem (e.g. Kahnert, 2009). To avoid such problems, one needs to systematically analyse the information content of the observations and constrain the assimilation algorithm to only operate on the signal degrees of freedom.

### Appendix C: Information content of measurements

Our ultimate goal is to formulate the data assimilation problem in such a way that the information contained in the measurements is fully exploited, but not overused. To this end, we first need to know how many independent quantities can be determined from a specific set of measurements. We investigate this question by borrowing ideas from retrieval and information theory – see Rodgers (2000) for more detailed explanations.

The main idea is to compare the variances of the model variables to those of the observations. Only those model variables whose variance is larger than those of the observations can be constrained by measurements. However, to actually make such a comparison poses two problems. The first problem is that one cannot readily compare error covariance *matrices*. The second problem is that model variables and measurements are in different spaces. We first address the second problem.

When we account for observation errors  $\epsilon_o$ , then the basic relation between model variables and observations is, to first order

$$\mathbf{y} = \hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x} + \epsilon_o. \quad (\text{C1})$$

<sup>8</sup>By solving the equation  $\nabla J|_{\mathbf{x}=\mathbf{x}_a} = \mathbf{0}_n$  for the analysed state  $\mathbf{x}_a$  it can be shown that the solution to the inverse problem is given by  $\mathbf{x}_a = \mathbf{x}_b + \mathbf{K} \cdot (\mathbf{y} - \hat{H}(\mathbf{x}_b))$ , where  $\mathbf{K} = \mathbf{B} \cdot \mathbf{H}^T \cdot (\mathbf{H} \cdot \mathbf{B} \cdot \mathbf{H}^T + \mathbf{R})^{-1}$  is known as the gain matrix. This illustrates that the analysis updates the background estimate  $\mathbf{x}_b$  by mapping the increment  $(\mathbf{y} - \hat{H}(\mathbf{x}_b))$  from observation space to model space by use of the gain matrix. The correlations among the model variables enter into the gain matrix through the matrix  $\mathbf{B}$ . In our case the vertical correlations are rather weak in comparison to correlations among different aerosol species.

The error covariance matrices are given by the expectation values  $\mathbf{B} = \langle \delta \mathbf{x} \cdot \delta \mathbf{x}^T \rangle$ , and  $\mathbf{R} = \langle \boldsymbol{\epsilon}_o \cdot \boldsymbol{\epsilon}_o^T \rangle$ , where the dot denotes a dyadic product.<sup>9</sup> From Eq. (C1) we see that the covariance matrix of  $\delta \mathbf{y} = \mathbf{y} - \hat{H}(\mathbf{x}_b)$  is given by  $\langle \delta \mathbf{y} \cdot \delta \mathbf{y}^T \rangle = \mathbf{H} \cdot \mathbf{B} \cdot \mathbf{H}^T + \mathbf{R}$ , where we assumed that background and observation errors are uncorrelated. This last equation suggests that we can compare model and observation errors in the same space by transforming the background error covariance matrix from the space of  $(n \times n)$  matrices to the space of  $(m \times m)$  matrices, namely  $\mathbf{H} \cdot \mathbf{B} \cdot \mathbf{H}^T$ .

To address the first problem, we diagonalize the covariance matrices by making the following change of variables:

$$\delta \tilde{\mathbf{x}} = \mathbf{B}^{-1/2} \cdot \delta \mathbf{x} \quad (\text{C2})$$

$$\delta \tilde{\mathbf{y}} = \mathbf{R}^{-1/2} \cdot (\mathbf{y} - \hat{H}(\mathbf{x}_b)) \quad (\text{C3})$$

$$\tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2}. \quad (\text{C4})$$

Here  $\mathbf{B}^{1/2}$  denotes the positive square root<sup>10</sup> of the matrix  $\mathbf{B}$ , and  $\mathbf{B}^{-1/2}$  denotes its inverse. The scaled observation operator  $\tilde{\mathbf{H}}$  is sometimes referred to as the observability matrix. In the new basis, the cost function in Eqs. (B7)–(B9) becomes

$$J = \frac{1}{2} \delta \tilde{\mathbf{x}}^T \cdot \delta \tilde{\mathbf{x}} + \frac{1}{2} (\tilde{\mathbf{H}} \cdot \delta \tilde{\mathbf{x}} - \delta \tilde{\mathbf{y}})^T \cdot (\tilde{\mathbf{H}} \cdot \delta \tilde{\mathbf{x}} - \delta \tilde{\mathbf{y}}). \quad (\text{C5})$$

The covariance matrices are now unit matrices. This can also be seen by considering the transformed errors, e.g.  $\tilde{\boldsymbol{\epsilon}}_o = \mathbf{R}^{-1/2} \cdot \boldsymbol{\epsilon}_o$  and computing  $\langle \tilde{\boldsymbol{\epsilon}}_o \cdot \tilde{\boldsymbol{\epsilon}}_o^T \rangle = \mathbf{R}^{-1/2} \cdot \langle \boldsymbol{\epsilon}_o \cdot \boldsymbol{\epsilon}_o^T \rangle \cdot \mathbf{R}^{-1/2} = \mathbf{1}_{m \times m}$ , since  $\langle \boldsymbol{\epsilon}_o \cdot \boldsymbol{\epsilon}_o^T \rangle = \mathbf{R}$ . (Here,  $\mathbf{1}_{m \times m}$  denotes the unit matrix in  $m$ -dimensional observation space.) Similarly, we find  $\langle \delta \tilde{\mathbf{x}} \cdot \delta \tilde{\mathbf{x}}^T \rangle = \mathbf{1}_{n \times n}$ . The covariance matrix of the transformed measurement vector  $\delta \tilde{\mathbf{y}}$  is given by  $\langle \delta \tilde{\mathbf{y}} \cdot \delta \tilde{\mathbf{y}}^T \rangle = \tilde{\mathbf{H}} \cdot \tilde{\mathbf{H}}^T + \mathbf{1}_{m \times m}$ . The first term is the model error covariance term transformed into observation space, while the second term (the unit matrix) is the diagonalized observation error covariance matrix.

We are still not in a position to make a meaningful comparison of model and observation errors, since the first term,  $\tilde{\mathbf{H}} \cdot \tilde{\mathbf{H}}^T$ , is still not diagonal. To make it so we need to perform one more transformation. To this end, we consider the singular value decomposition of the matrix  $\tilde{\mathbf{H}}$ :

$$\tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2} = \mathbf{V}_L \cdot \mathbf{W} \cdot \mathbf{V}_R^T. \quad (\text{C6})$$

Here  $\tilde{\mathbf{H}}$  is a  $(m \times n)$  matrix, the matrix of the left-singular vectors  $\mathbf{V}_L$  is a  $(m \times m)$  matrix, the matrix  $\mathbf{V}_R$  containing the right-singular vectors is a  $(n \times n)$  matrix, and the  $(m \times n)$  matrix  $\mathbf{W}$  consists of two blocks. If  $m < n$ , then the left

<sup>9</sup>The expectation value of a discrete variable  $a$  that assumes values  $a_1, a_2, \dots, a_n$  with corresponding probabilities  $p_1, p_2, \dots, p_n$  is given by  $\langle a \rangle = \sum_{i=1}^n p_i a_i$ .

<sup>10</sup>A matrix  $\mathbf{A}$  is called a square root of a matrix  $\mathbf{B}$  if  $\mathbf{A}^T \cdot \mathbf{A} = \mathbf{B}$ . The positive square root of  $\mathbf{B}$ , which is denoted by  $\mathbf{B}^{1/2}$ , has the property  $\mathbf{x}^T \cdot \mathbf{B}^{1/2} \cdot \mathbf{x} \geq 0$  for all  $\mathbf{x}$ . If  $\mathbf{B}$  is itself positive and symmetric, as is the case for covariance matrices, then the positive square root exists and is unique.

block of  $\mathbf{W}$  is a  $(m \times m)$ -diagonal matrix containing the  $m$  singular values  $w_1, \dots, w_m$  on the diagonal; the right block is a  $(m \times (n - m))$ -null matrix. Similarly, if  $m > n$ , then the upper block of  $\mathbf{W}$  is a  $(n \times n)$ -diagonal matrix containing the  $n$  singular values on the diagonal, while the lower block is a  $((m - n) \times n)$ -null matrix.

We now make another change of variables:

$$\delta \mathbf{x}' = \mathbf{V}_R^T \cdot \delta \tilde{\mathbf{x}} \quad (\text{C7})$$

$$\delta \mathbf{y}' = \mathbf{V}_L^T \cdot \delta \tilde{\mathbf{y}} \quad (\text{C8})$$

$$\mathbf{H}' = \mathbf{V}_L^T \cdot \tilde{\mathbf{H}} \cdot \mathbf{V}_R. \quad (\text{C9})$$

The matrices  $\mathbf{V}_L$  and  $\mathbf{V}_R$  are orthogonal, i.e.  $\mathbf{V}_L^T \cdot \mathbf{V}_L = \mathbf{1}_{m \times m}$ , and similarly for  $\mathbf{V}_R$ . Thus, substitution of Eqs. (C7)–(C9) into (C5) yields

$$J = \frac{1}{2} \delta \mathbf{x}'^T \cdot \delta \mathbf{x}' + \frac{1}{2} (\mathbf{H}' \cdot \delta \mathbf{x}' - \delta \mathbf{y}')^T \cdot (\mathbf{H}' \cdot \delta \mathbf{x}' - \delta \mathbf{y}'). \quad (\text{C10})$$

Evidently, the transformation given in Eqs. (C7)–(C9) preserves the diagonality of the background and observation error covariance matrices. What about the covariance matrix  $\langle \delta \mathbf{y}' \cdot \delta \mathbf{y}'^T \rangle$  in the new basis? Using  $\boldsymbol{\epsilon}'_o = \mathbf{V}_L^T \cdot \tilde{\boldsymbol{\epsilon}}_o = \mathbf{V}_L^T \cdot \mathbf{R}^{-1/2} \cdot \boldsymbol{\epsilon}_o$ , as well as Eqs. (C1), (C2)–(C4), and (C6)–(C9), we obtain  $\langle \delta \mathbf{y}' \cdot \delta \mathbf{y}'^T \rangle = \mathbf{H}' \cdot \mathbf{H}'^T + \mathbf{1}_{m \times m}$ . The contribution of the background error covariances in this coordinate system is  $\mathbf{H}' \cdot \mathbf{H}'^T$ , which is a diagonal matrix. This becomes clear from Eqs. (C6) and (C9), which yields

$$\mathbf{H}' \cdot \mathbf{H}'^T = \mathbf{W} \cdot \mathbf{W}^T, \quad (\text{C11})$$

which is a  $(m \times m)$  diagonal matrix. Thus in this coordinate system we can readily compare the diagonal elements of the transformed background error covariance matrix  $\mathbf{H}' \cdot \mathbf{H}'^T$  to the diagonal (unit) elements of the observation error covariance matrix  $\mathbf{1}_{m \times m}$ . Roughly, those singular values  $w_i$  on the diagonal of  $\mathbf{W}$  that are larger than unity correspond to model variables  $\delta x'_i$  that can be controlled by the measurements. Those singular values smaller than unity correspond to model variables that are only related to noise.

In the above discussion we relied on plausibility arguments. We mention that there are more systematic ways of approaching the problem. Here we merely state some key results without going into details. The interested reader is referred to chapter 2 in Rodgers (2000). However, in all approaches the main quantities of interest are always the singular values of the observability matrix  $\mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2}$ .

One can compute the number of signal degrees of freedom  $N_s$  from the expectation value of  $J_b$  in Eq. (B8). The result can be expressed in terms of the singular values  $w_i$  of the observability matrix:

$$N_s = \sum_{i=1}^{\min\{m,n\}} w_i^2 / (1 + w_i^2), \quad (\text{C12})$$

where  $n$  is the dimension of model space, and  $m$  is the dimension of observation space.

Another approach is based on information theory. Given a system described by a probability distribution function  $P(x)$ , one defines the Shannon entropy

$$S(P) = - \int P(x) \log_2 \left( \frac{P(x)}{P_0(x)} \right) dx, \quad (\text{C13})$$

where  $P_0$  is a normalization factor needed to make the argument of the logarithm dimensionless. A decrease in entropy expresses an increase in our knowledge of the system. For instance, if we initially describe the system by  $P_i(x)$ , and, after taking measurements, by  $P_f(x)$ , then the measurement process has changed the entropy by an amount

$$H = S(P_i) - S(P_f). \quad (\text{C14})$$

In our case, we assume that all errors are normally distributed. In that case, one can show that

$$H = \frac{1}{2} \sum_{i=1}^{\min\{m,n\}} \log_2(1 + w_i^2). \quad (\text{C15})$$

$H$  can be interpreted as a measure for the information content of a set of measurements.

Our findings so far suggest a general strategy for how to optimize the amount of information that can be extracted from measurements. First, we need to compute the singular value decomposition in Eq. (C6), as well as the transformation given in Eqs. (C2) and (C7), which we can summarize as

$$\delta \mathbf{x}' = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \mathbf{x}. \quad (\text{C16})$$

Then we want to formulate the minimization of the cost function in such a way that only those components of  $\delta \mathbf{x}'$  are adjusted by the assimilation algorithm that correspond to the largest singular values of the matrix  $\mathbf{W}$  in Eq. (C6). All other elements of  $\delta \mathbf{x}'$  should be left alone. In other words, we want to constrain the minimization of the cost function to the subspace of the signal degrees of freedom of the state vector. Thus, in order to implement this idea, we first need to discuss how to incorporate constraints into the theory.

#### Appendix D: Minimization of the cost function with constraints

In the minimization of the cost function all elements of the control vector  $\delta \mathbf{x}$  are independently adjusted until the minimum of  $J$  is found. This may not be a prudent approach if the information contained in the observations is insufficient to constrain all model variables. In such a case one should introduce constraints that reduce the number of independent control variables. However, this needs to be done in a clever way; the goal is to neither underuse the measurements (thus wasting available information), nor to overuse them (thus assimilating noise).

For reasons we will explain later we formulate the constraints as weak conditions. However, for didactic reasons as well as for the sake of completeness, we will also mention how to formulate constraints as strong conditions.

#### D1 Minimization of the cost function with strong constraints

Given  $k$  constraints in the form  $g_i(\delta \mathbf{x}) = 0$ ,  $i = 1, \dots, k$ , the most general way of finding the minimum of  $J(\delta \mathbf{x})$  under the constraints  $g_i$  is the method of Lagrange multipliers. More specifically, one introduces  $k$  Lagrange multipliers  $\lambda_1, \dots, \lambda_k$  and defines the function

$$L(\delta x_1, \dots, \delta x_n, \lambda_1, \dots, \lambda_k) = J(\delta x_1, \dots, \delta x_n) + \sum_{i=1}^k \lambda_i g_i(\delta x_1, \dots, \delta x_n); \quad (\text{D1})$$

then one solves the minimization problem

$$\nabla L(\delta x_1, \dots, \delta x_n, \lambda_1, \dots, \lambda_k) = \mathbf{0}_{n+k}, \quad (\text{D2})$$

where  $\nabla = \nabla_{\delta x_1, \dots, \delta x_n, \lambda_1, \dots, \lambda_k}$  is now a  $(n+k)$ -dimensional gradient operator, and where  $\mathbf{0}_{n+k}$  denotes the null vector in an  $(n+k)$ -dimensional space. Note that in this general formulation of the problem the constraints can even be nonlinear. We are specifically interested in linear constraints, which can be expressed in the form  $\mathbf{G} \cdot \delta \mathbf{x} = \mathbf{0}_k$ . Then the constrained minimization problem becomes

$$L(\delta \mathbf{x}, \boldsymbol{\lambda}) = J(\delta \mathbf{x}) + \boldsymbol{\lambda}^T \cdot \mathbf{G} \cdot \delta \mathbf{x} \quad (\text{D3})$$

$$\nabla_{\delta \mathbf{x}, \boldsymbol{\lambda}} L(\delta \mathbf{x}, \boldsymbol{\lambda}) = \begin{pmatrix} \nabla_{\delta \mathbf{x}} J(\delta \mathbf{x}) + \boldsymbol{\lambda}^T \cdot \mathbf{G} \\ \mathbf{G} \cdot \delta \mathbf{x} \end{pmatrix} = \mathbf{0}_{n+k}. \quad (\text{D4})$$

Compared to the unconstrained minimization problem, the introduction of  $k$  constraints has increased the dimension of the problem from  $n$  to  $n+k$ . Naively, one may have expected that the dimension would, on the contrary, be reduced to  $n-k$ . This is indeed the case if the constraints are linear, and if the function  $J$  is quadratic, as is the case in Eqs. (B7)–(B9). To see this, let us first write those equations more concisely in the form

$$J = \frac{1}{2} (\delta \mathbf{x}^T \cdot \mathbf{Q}_1 \cdot \delta \mathbf{x} + \mathbf{Q}_2^T \cdot \delta \mathbf{x} + \delta \mathbf{x}^T \cdot \mathbf{Q}_2 + \mathbf{Q}_3) \quad (\text{D5})$$

$$\mathbf{Q}_1 = \mathbf{B}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \quad (\text{D6})$$

$$\mathbf{Q}_2 = \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot (\hat{H}(\mathbf{x}_b) - \mathbf{y}) \quad (\text{D7})$$

$$\mathbf{Q}_3 = (\hat{H}(\mathbf{x}_b) - \mathbf{y})^T \cdot \mathbf{R}^{-1} \cdot (\hat{H}(\mathbf{x}_b) - \mathbf{y}). \quad (\text{D8})$$

Note that the covariance matrices and their inverses are symmetric (i.e.  $\mathbf{R}^T = \mathbf{R}$ , etc.) The unconstrained minimization problem requires us to solve the equation  $\nabla_{\delta \mathbf{x}} J = \mathbf{Q}_1 \cdot \delta \mathbf{x} + \mathbf{Q}_2 = \mathbf{0}_n$ . Now we want to minimize the cost function subject to the linear constraints

$$\mathbf{G} \cdot \delta \mathbf{x} = \mathbf{0}_k, \quad (\text{D9})$$



where  $\mathbf{G}$  is a  $(k \times n)$  matrix,  $\delta\mathbf{x}$  is an  $n$ -vector, and  $\mathbf{0}_k$  is the null-vector in  $\mathbb{R}^k$ . Let us denote the kernel<sup>11</sup> of  $\mathbf{G}$  by  $\ker(\mathbf{G})$ . Let further  $\mathbf{z}_1, \dots, \mathbf{z}_{n-k}$  denote a basis of  $\ker(\mathbf{G})$ . We define the  $(n \times (n-k))$  matrix

$$\mathbf{Z} = (\mathbf{z}_1 \quad \dots \quad \mathbf{z}_{n-k}) \quad (\text{D10})$$

the column vectors of which are just the basis vectors of  $\ker(\mathbf{G})$ . Obviously,  $\mathbf{G} \cdot \mathbf{Z} = \mathbf{0}_{k \times (n-k)}$ , where  $\mathbf{0}_{k \times (n-k)}$  denotes the  $((k \times (n-k))$ -null matrix. If  $\delta\mathbf{x}$  is a vector in  $\mathbb{R}^n$  for which there exists a vector  $\boldsymbol{\xi} \in \mathbb{R}^{n-k}$  such that  $\mathbf{Z} \cdot \boldsymbol{\xi} = \delta\mathbf{x}$ , then we automatically have  $\mathbf{G} \cdot \delta\mathbf{x} = \mathbf{0}_k$ , i.e.  $\delta\mathbf{x}$  satisfies the linear constraints. Thus we can formulate the constrained minimization problem by substitution of  $\delta\mathbf{x} = \mathbf{Z} \cdot \boldsymbol{\xi}$  into Eq. (D5), which yields

$$J = \frac{1}{2} (\boldsymbol{\xi}^T \cdot \mathbf{Z}^T \cdot \mathbf{Q}_1 \cdot \mathbf{Z} \cdot \boldsymbol{\xi} + \mathbf{Q}_2^T \cdot \mathbf{Z} \cdot \boldsymbol{\xi} + \boldsymbol{\xi}^T \cdot \mathbf{Z}^T \cdot \mathbf{Q}_2 + \mathbf{Q}_3) \quad (\text{D11})$$

$$\mathbf{0}_k = \nabla_{\boldsymbol{\xi}} J = \mathbf{Z}^T \cdot \mathbf{Q}_1 \cdot \mathbf{Z} \cdot \boldsymbol{\xi} + \mathbf{Z}^T \cdot \mathbf{Q}_2. \quad (\text{D12})$$

Thus we have reduced the  $(n+k)$ -dimensional constrained minimization problem given in Eq. (D4) to a problem consisting of the following two steps:

1. Determine a basis of the null space  $\ker(\mathbf{G})$ ; this yields the matrix  $\mathbf{Z}$ .
2. Solve the unconstrained  $(n-k)$ -dimensional optimization problem given in Eq. (D12). From the  $(n-k)$ -vector  $\boldsymbol{\xi}$  that minimizes the cost function in Eq. (D11), we then obtain the solution  $\delta\mathbf{x} = \mathbf{Z} \cdot \boldsymbol{\xi}$  that minimizes the cost function in Eq. (D5) subject to the constraint Eq. (D9).

## D2 Minimization of the cost function with weak constraints

In the approach described in the previous section the solution satisfies the constraints exactly. Therefore, this approach is known as the minimization of the cost function with *strong constraints*. In the *weak-constraint* approach the constraints only need to be satisfied within specified error bounds.

The formulation of the weak-constraint approach is conceptually quite simple. One incorporates the constraints by adding an extra term to the cost function Eq. (B7), i.e.

$$J = J_b + J_o + J_G \quad (\text{D13})$$

$$J_G = \frac{1}{2} \delta\mathbf{x}^T \cdot \mathbf{G}^T \cdot \mathbf{B}_G^{-1} \cdot \mathbf{G} \cdot \delta\mathbf{x}, \quad (\text{D14})$$

which also gives an extra term in the gradient of the cost function,

$$\nabla_{\delta\mathbf{x}} J_G = \mathbf{G}^T \cdot \mathbf{B}_G^{-1} \cdot \mathbf{G} \cdot \delta\mathbf{x}. \quad (\text{D15})$$

<sup>11</sup>The *kernel* or *null space* of a matrix is the set of all vectors  $\mathbf{z}$  such that  $\mathbf{G} \cdot \mathbf{z} = \mathbf{0}$ . The kernel is a subspace of the full vector space  $\mathbb{R}^n$  with  $\dim \ker(\mathbf{G}) = n - k$ .

We will assume that the matrix  $\mathbf{B}_G = \text{diag}(\sigma_1^G, \dots, \sigma_k^G)$  is diagonal, where  $k$  is the number of constraints. The “error variances”  $\sigma_i^G$  along the diagonal of  $\mathbf{B}_G$  allow us to fine-tune the influence of each constraint on the solution. If  $\sigma_i^G$  is small, then the  $i$ th constraint is relatively strong, and vice versa. Typically, if the  $\sigma_i^G$  are made too large, then there is a risk that the minimization algorithm ignores the constraints all together. In that case the solution will be very similar to the unconstrained solution. On the other hand, if the  $\sigma_i^G$  are made too small, then  $J_G$  can make the dominant contribution to  $J$ . In that case, there is a risk that the minimization routine largely ignores the observations and returns a solution that lies quite close to the background estimate.

## D3 Constraints designed for making optimum use of the information contained in the observations

We now want to incorporate the results of Appendix C into the variational data assimilation method. More specifically, we want to formulate weak constraints, Eq. (D14), based on the singular values of the observation operator in Eq. (C6). To this end, we make the change of variables given in Eq. (C16). We assume, without loss of generality, that the first  $\ell$  singular values are greater than unity. Thus we only want to use the corresponding components  $\delta x'_1, \dots, \delta x'_\ell$  as independent control variables in the 3DVAR algorithm, while the remaining components remain unchanged, at least approximately, within specified error bounds. If we were to formulate this requirement as a strong constraint, as in Eq. (D9), then it would take the form

$$\delta\mathbf{x}' = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta\mathbf{x} = \begin{pmatrix} \delta x'_1 \\ \vdots \\ \delta x'_\ell \\ 0 \\ \vdots \end{pmatrix}. \quad (\text{D16})$$

Thus the matrix expressing the constraints is given by  $\mathbf{G} = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2}$ , which is a  $(n \times n)$  matrix.

The weak constraint approach is, arguably, more suitable in our case. We have, in the preceding text, frequently used the terms *signal degrees of freedom* and *noise degrees of freedom*. Although it was conceptually useful to make this distinction, it is important to stress that there is no sharp boundary between the two. Rather, there is a smooth transition from singular values  $w_1 > w_2 > \dots > w_\ell \geq 1$  to singular values  $1 > w_{\ell+1} > w_{\ell+2} > \dots > w_K$  ( $K = \min\{n, m\}$ ). For this reason we choose to formulate the constraints as weak constraints. This allows us to make a smooth transition from free to constrained control variables, where the transition from one regime to the other can be controlled by the singular values.

In order to apply the weak-constraint approach, we need to substitute the constraint matrix  $\mathbf{G} = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2}$  into

Eq. (D14), which yields

$$J_G = \frac{1}{2} \delta \mathbf{x}^T \cdot \mathbf{B}^{-1/2} \cdot \mathbf{V}_R \cdot \mathbf{B}_G^{-1} \cdot \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \mathbf{x}, \quad (\text{D17})$$

where  $\mathbf{B}_G$  is a  $(n \times n)$  matrix. We want to set up this matrix in such a way that we obtain a smooth transition from freely adaptable control variables  $\delta x'_1, \dots, \delta x'_\ell$  to increasingly constrained variables  $\delta x'_{\ell+1}, \dots, \delta x'_k, \dots, \delta x'_n$ . One possible choice of the matrix  $\mathbf{B}_G$  would be

$$\mathbf{B}_G = \sigma_G \text{diag}(w_1, w_2, \dots, w_\ell, \dots, w_k, c, \dots, c), \quad (\text{D18})$$

where  $\sigma_G$  is a free scaling factor, and where the last  $n - k$  diagonal elements are equal to a constant  $c \ll w_k$  chosen to be much smaller than the smallest singular value  $w_k$ .

Clearly, how we set up the matrix  $\mathbf{B}_G$  is not unique. For instance, a more general choice would be

$$\mathbf{B}_G = \sigma_G \text{diag}(w_1^p, w_2^p, \dots, w_\ell^p, \dots, w_k^p, c, \dots, c), \quad (\text{D19})$$

where  $c \ll w_k^p$ , and where the exponent  $p$  would be another parameter that can be employed to tune how steeply the transition from unconstrained to constrained control variables takes place. Yet another choice would be

$$\mathbf{B}_G = \sigma_G \cdot \text{diag}(\mu_1, \mu_2, \dots, \mu_\ell, \dots, \mu_k, c, \dots, c), \quad (\text{D20})$$

$$\mu_i = w_i^2 / (1 + w_i^2), \quad (\text{D21})$$

where  $c \ll \mu_k$ . This ansatz is suggested by Eq. (C12), i.e. each of the elements  $\delta x'_1, \dots, \delta x'_k$  is weighted with its corresponding contribution to the number of signal degrees of freedom. We tested all three approaches (the one in Eq. D19 for  $p = 2$ ). These tests showed that the different approaches often yield analysis results that are quite similar. However, in each approach the free parameters  $\sigma_G$  and  $c$  are tuned to different values. If they are not well tuned, then the analysis tends either toward the background estimate or toward the unconstrained analysis, as explained earlier in the text following Eq. (D15).

## Appendix E: Practical aspects of the implementation

We will here discuss some practical aspects that are mainly interesting for model developers.

One of the main practical problems is the dimension  $n$  of the model space. The grid-size is typically on the order  $N_x \times N_y \times N_z \sim 100 \times 100 \times 10$ , and the number of aerosol components is of the order of  $N_c \sim 10$ – $100$ . Hence the dimension of the model space is  $n \sim 10^6$ – $10^7$ . In our case, the matrix  $\tilde{\mathbf{H}}$  in Eq. (C6) is a  $(m \times n)$  matrix. To numerically perform a singular value decomposition of such a large matrix would be a formidable task.

In variational data assimilation we encounter a similar problem in the inversion of the matrix  $\mathbf{B}$ . In our 3DVAR code

this problem is alleviated by using a so-called spectral formulation. The idea is to make a Fourier transformation in the horizontal coordinates and to assume that all horizontal error correlations are homogeneous and isotropic. Under these assumptions one obtains one background error covariance matrix for each horizontal wavenumber; each of these matrices has dimension  $N_z \times N_c \sim 10^3$ – $10^4$ . This can further be reduced to about  $10^2$  by making a reduced eigenvalue diagonalization. The details are explained in Kahnert (2008).

In our case we are primarily interested in constraining the aerosol components. Therefore, we formulate our weak constraints in a suitable subspace of the physical space. Suppose, for simplicity, that we have reduced all data to the vertical resolution of our model. Let  $v_l = 1, \dots, m_l$  label all measurements that lie within model layer  $l$ . Suppose further that  $(i_\alpha, j_\alpha)$  is the horizontal grid point belonging to observation  $v_l$  (so that the index  $\alpha$  depends on the layer  $l$  and the observation  $v_l$ ). Consider the reduced background error covariance matrix with elements  $B_{k,k'}^{(\alpha,l)} = B_{i_\alpha j_\alpha l k, i_\alpha j_\alpha l k'}$ ,  $k, k' = 1, \dots, N_c$ , and  $N_c$  is the number of aerosol components. Consider further the reduced observability matrix with elements  $\tilde{H}_{v_l, k}^{(l)} = \sum_{k'=1}^{N_c} R_{v_l, v_l}^{-1/2} H_{m, i_\alpha, j_\alpha l k'} \{ (B^{(\alpha,l)})^{1/2} \}_{k', k}$ , where  $m = m(l, v_l)$  labels the  $v_l$ th observation in model layer  $l$ . Analogous to Eq. (C6), we now perform a singular value decomposition in the reduced space

$$\tilde{H}_{v_l, k}^{(l)} = \sum_{s=1}^{\min\{m_l, N_c\}} (V_L^{(l)})_{v_l, s} w_s^{(l)} (V_R^{(l)})_{k, s}. \quad (\text{E1})$$

The dimension of this SVD problem is now considerably reduced. The number of singular values is equal to  $K = \min\{N_c, m_l\}$ . The constraint matrix  $\mathbf{G} = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2}$  reduces to

$$G_{s, k} = \sum_{k'=1}^{N_c} (V_R^{(l)})_{k', s} \{ (B^{(\alpha,l)})^{-1/2} \}_{k', k}. \quad (\text{E2})$$

We now invoke the assumption that the constraints computed at the observation site are also valid at neighbouring points, i.e. we apply the constraint matrix given in Eq. (E2) in Eq. (D17) according to

$$J_G = \frac{1}{2} \sum_{ijklk's} \delta x_{ijkl} G_{s, k'} (B_G^{-1})_s G_{s, k} \delta x_{ijkl}, \quad (\text{E3})$$

where  $(B_G)_s$  denotes the diagonal elements of the matrix given in Eq. (D18).<sup>12</sup>

<sup>12</sup>For those readers interested in spectral formulations of 3DVAR we refer to Eqs. (28)–(30) in Kahnert (2008). Expressed by the spectral control vector  $\chi = \mathbf{U} \cdot \delta \mathbf{x}$ , the weak constraint in the cost function takes the spectral form  $J_G = \frac{1}{2} \chi^\dagger \cdot \mathbf{U}^{-\dagger} \cdot \mathbf{G}^T \cdot \mathbf{B}_G^{-1} \cdot \mathbf{G} \cdot \mathbf{U}^{-1} \cdot \chi$ , and its contribution to the gradient of the cost function becomes  $\nabla_\chi J_G = \mathbf{U}^{-\dagger} \cdot \mathbf{G}^T \cdot \mathbf{B}_G^{-1} \cdot \mathbf{G} \cdot \mathbf{U}^{-1} \cdot \chi$ . We see that these expressions involve the computation of the variable  $\delta \mathbf{x} = \mathbf{U}^{-1} \cdot \chi$  in physical

Another aspect concerns the positive square root of the background error covariance matrix, which appears in essential parts of the theory, namely in Eqs. (C6) and (D16). In theoretical developments it is, arguably, didactically expedient to work with the matrix  $\mathbf{B}^{1/2}$ . But in practice there are numerically more efficient formulations. One such approach is discussed in Kahnert (2008) in the context of a spectral formulation of the variational method. The spectral formulation is applied to the full  $\mathbf{B}$  matrix in order to reduce the dimension of the problem of diagonalizing this matrix. This method is our method of choice in the formulation of the background and observation terms in the cost function given in Eqs. (B8) and (B9), respectively. However, in the formulation of the constraint term given in Eq. (D17) we can substantially reduce the dimension of the matrix  $\mathbf{B}$  by working in the reduced space in which only the covariances  $\mathbf{B}^{(\alpha,l)}$  among

aerosol components are considered. One could compute the matrix  $(\mathbf{B}^{(\alpha,l)})^{-1/2}$  in Eq. (D17) by diagonalizing the matrix  $\mathbf{B}^{(\alpha,l)}$ . However, a numerically much more efficient approach is to not work with positive square root, but with the so-called Cholesky decomposition<sup>13</sup> of the  $\mathbf{B}$  matrix,

$$\mathbf{B}^{(\alpha,l)} = \mathbf{C}_u^T \cdot \mathbf{C}_u, \quad (\text{E4})$$

where  $\mathbf{C}_u$  is an upper triangular matrix. Thus the actual algorithm we used for formulating the constrained minimization of the cost function is obtained by replacing in the preceding formulas all incidences of the matrix  $\mathbf{B}^{1/2}$  with the matrix  $\mathbf{C}_u^T$  (and, similarly, by replacing the inverse matrix  $\mathbf{B}^{-1/2}$  by the inverse of the Cholesky factor,  $\mathbf{C}_u^{-T}$ ).

---

space. Thus, even when using a spectral formulation of the 3DVAR method, one can still compute the constraints in physical space and add their contributions to  $J$  and  $\nabla J$ . The advantage of this is, as explained above, that the SVD of the observability matrix can be computed in the reduced subspace, which substantially reduces the dimension of the numerical SVD problem.

---

<sup>13</sup>The Cholesky decomposition is, essentially, a special case of a LU decomposition, which applies to symmetric real (or Hermitian complex), positive definite matrices.

**The Supplement related to this article is available online at doi:10.5194/acp-17-3423-2017-supplement.**

*Author contributions.* Michael Kahnert worked with the theoretical developments and numerical implementation, Emma Andersson performed the testing of the method.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* This work was funded by the Swedish National Space Board through project nos. 100/16 (MK) and 101/13 (EA).

Edited by: M. Tesche

Reviewed by: five anonymous referees

## References

- Andersson, C., Langner, J., and Bergström, R.: 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 B*, 59, 77–98, 2007.
- Andersson, C., Bergström, R., Bennet, C., Robertson, L., Thomas, M., Korhonen, H., Lehtinen, K. E. J., and Kokkola, H.: MATCH-SALSA – Multi-scale Atmospheric Transport and CHemistry model coupled to the SALSA aerosol microphysics model – Part 1: Model description and evaluation, *Geosci. Model Dev.*, 8, 171–189, doi:10.5194/gmd-8-171-2015, 2015.
- Andersson, E. and Kahnert, M.: Coupling aerosol optics to the MATCH (v5.5.0) chemical transport model and the SALSA (v1) aerosol microphysics module, *Geosci. Model Dev.*, 9, 1803–1826, doi:10.5194/gmd-9-1803-2016, 2016.
- Benedetti, A., Morcrette, M. J.-J., Boucher, O., Dethof, A., Engelen, R. J., Huneeus, M. F. H. F. N., Jones, L. S., Kinne, J. W. K., Mangold, A., Razinger, M., Simmons, A. J., and Suttie, M.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, *J. Geophys. Res.*, 114, D13205, doi:10.1029/2008JD011115, 2009.
- Bi, L., Yang, P., Kattawar, G., and Kahn, R.: Modeling optical properties of mineral aerosol particles by using nonsymmetric hexahedra, *Appl. Optics*, 49, 334–342, 2010.
- Bocquet, M.: Toward optimal choices of control space representation for geophysical data assimilation, *Mon. Weather Rev.*, 137, 2331–2348, 2009.
- Burton, S. P., Hair, J. W., Kahnert, M., Ferrare, R. A., Hostetler, C. A., Cook, A. L., Harper, D. B., Berkoff, T. A., Seaman, S. T., Collins, J. E., Fenn, M. A., and Rogers, R. R.: Observations of the spectral dependence of linear particle depolarization ratio of aerosols using NASA Langley airborne High Spectral Resolution Lidar, *Atmos. Chem. Phys.*, 15, 13453–13473, doi:10.5194/acp-15-13453-2015, 2015.
- Burton, S. P., Chemyakin, E., Liu, X., Knobelspiesse, K., Stammes, S., Sawamura, P., Moore, R. H., Hostetler, C. A., and Ferrare, R. A.: Information content and sensitivity of the  $3\beta + 2\alpha$  lidar measurement system for aerosol microphysical retrievals, *Atmos. Meas. Tech.*, 9, 5555–5574, doi:10.5194/amt-9-5555-2016, 2016.
- Cardinali, C., Pezzulli, S., and Andersson, E.: Influence-matrix diagnostic of a data assimilation system, *Q. J. Roy. Meteor. Soc.*, 130, 2767–2786, 2004.
- Foltescu, V., Pryor, S. C., and Bennet, C.: Sea salt generation, dispersion and removal on the regional scale, *Atmos. Environ.*, 39, 2123–2133, 2005.
- Fuller, K. A. and Mackowski, D. W.: Electromagnetic scattering by compounded spherical particles, in: *Light scattering by non-spherical particles*, edited by: Mishchenko, M. I., Hovenier, J. W., and Travis, L. D., 226–273, Academic Press, San Diego, 2000.
- Johnson, C., Hoskins, B. J., and Nichols, N. K.: Very large inverse problems in atmosphere and ocean modelling, *Int. J. Numer. Meth. Fl.*, 47, 759–771, 2005a.
- Johnson, C., Nichols, N. K., and Hoskins, B. J.: A singular vector perspective of 4D-Var: Filtering and interpolation, *Q. J. Roy. Meteor. Soc.*, 131, 1–19, 2005b.
- Joiner, J. and da Silva, A. M.: Efficient methods to assimilate remotely sensed data based on information content, *Q. J. Roy. Meteor. Soc.*, 124, 1669–1694, 1998.
- Kahnert, F. M.: Reproducing the optical properties of fine desert dust aerosols using ensembles of simple model particles, *J. Quant. Spectrosc. Ra.*, 85, 231–249, 2004.
- Kahnert, M.: Variational data analysis of aerosol species in a regional CTM: background error covariance constraint and aerosol optical observation operators, *Tellus B*, 60, 753–770, 2008.
- Kahnert, M.: On the observability of chemical and physical aerosol properties by optical observations: Inverse modelling with variational data assimilation, *Tellus B*, 61, 747–755, 2009.
- Kahnert, M.: Modelling radiometric properties of inhomogeneous mineral dust particles: Applicability and limitations of effective medium theories, *J. Quant. Spectrosc. Ra.*, 152, 16–27, 2015.
- Kahnert, M. and Devasthale, A.: Black carbon fractal morphology and short-wave radiative impact: a modelling study, *Atmos. Chem. Phys.*, 11, 11745–11759, doi:10.5194/acp-11-11745-2011, 2011.
- Kahnert, M., Nousiainen, T., Lindqvist, H., and Ebert, M.: Optical properties of light absorbing carbon aggregates mixed with sulfate: assessment of different model geometries for climate forcing calculations, *Opt. Express*, 20, 10042–10058, 2012a.
- Kahnert, M., Nousiainen, T., Thomas, M. A., and Tyynelä, J.: Light scattering by particles with small-scale surface roughness: comparison of four classes of model geometries, *J. Quant. Spectrosc. Ra.*, 113, 2356–2367, 2012b.
- Kahnert, M., Nousiainen, T., and Lindqvist, H.: Models for integrated and differential scattering optical properties of encapsulated light absorbing carbon aggregates, *Opt. Express*, 21, 7974–7992, 2013.
- Kahnert, M., Nousiainen, T., and Lindqvist, H.: Review: Model particles in atmospheric optics, *J. Quant. Spectrosc. Ra.*, 146, 41–58, 2014.
- Kahnert, M., Nousiainen, T., and Markkanen, J.: Morphological models for inhomogeneous particles: light scattering by aerosols, cometary dust, and living cells, in: *Light Scattering Reviews 11*, edited by: Kokhanovsky, A., Springer, Berlin, 299–339, 2016.
- Khade, V. M., Hansen, J. A., Reid, J. S., and Westphal, D. L.: Ensemble filter based estimation of spatially distributed parameters

- in a mesoscale dust model: experiments with simulated and real data, *Atmos. Chem. Phys.*, 13, 3481–3500, doi:10.5194/acp-13-3481-2013, 2013.
- Kupiainen, K. and Klimont, Z.: Primary emissions of submicron and carbonaceous particles in Europe and the potential for their control, Tech. Rep. IR-04-079, IIASA, Laxenburg, Austria, 2004.
- Kupiainen, K. and Klimont, Z.: Primary emissions of fine carbonaceous particles in Europe, *Atmos. Environ.*, 41, 2156–2170, 2007.
- Kylling, A., Kahnert, M., Lindqvist, H., and Nousiainen, T.: Volcanic ash infrared signature: porous non-spherical ash particle shapes compared to homogeneous spherical ash particles, *Atmos. Meas. Tech.*, 7, 919–929, doi:10.5194/amt-7-919-2014, 2014.
- Lindqvist, H., Muinonen, K., and Nousiainen, T.: Light scattering by coated Gaussian and aggregate particles, *J. Quant. Spectrosc. Ra.*, 110, 1398–1410, doi:10.1016/j.jqsrt.2009.01.015, 2009.
- Lindqvist, H., Nousiainen, T., Zubko, E., and Muñoz, O.: Optical modeling of vesicular volcanic ash particles, *J. Quant. Spectrosc. Ra.*, 112, 1871–1880, 2011.
- Lindqvist, H., Jokinen, O., Kandler, K., Scheuvsens, D., and Nousiainen, T.: Single scattering by realistic, inhomogeneous mineral dust particles with stereogrammetric shapes, *Atmos. Chem. Phys.*, 14, 143–157, doi:10.5194/acp-14-143-2014, 2014.
- Liu, L. and Mishchenko, M. I.: Scattering and radiative properties of complex soot and soot-containing aggregate particles, *J. Quant. Spectrosc. Ra.*, 106, 262–273, 2007.
- Liu, Z., Liu, Q., Lin, H.-C., Schwartz, C. S., Lee, Y.-H., and Wang, T.: Three-dimensional variational assimilation of MODIS aerosol optical depth: Implementation and application to a dust storm over East Asia, *J. Geophys. Res.*, 116, D23206, doi:10.1029/2011JD016159, 2011.
- McKeen, S., Chung, S. H., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Gong, W., Bouchet, V., Moffet, R., Tang, Y., Carmichael, G. R., Mathur, R., and Yu, S.: Evaluation of several PM<sub>2.5</sub> forecast models using data collected during the ICARTT/NEAQS 2004 field study, *J. Geophys. Res.*, 112, d10S20, 2007.
- Mishchenko, M. I., Travis, L. D., Kahn, R. A., and West, R. A.: Modeling phase functions for dustlike tropospheric aerosols using a shape mixture of randomly oriented polydisperse spheroids, *J. Geophys. Res.*, 102, 16831–16847, 1997.
- Mishchenko, M. I., Dlugach, Z. M., and Zakharova, N. T.: Direct demonstration of the concept of unrestricted effective-medium approximation, *Opt. Lett.*, 39, 3935–3938, 2014.
- Muinonen, K.: Light scattering by stochastically shaped particles, in: *Light scattering by nonspherical particles*, edited by: Mishchenko, M. I., Hovenier, J. W., and Travis, L. D., 323–354, Academic Press, San Diego, 2000.
- Müller, D., Wandinger, U., and Ansmann, A.: Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: theory, *Appl. Optics*, 38, 2346–2357, 1999.
- Omar, A. H., Winker, D. M., Vaughan, M. A., Hu, Y., Treppe, C. R., Ferrare, R. A., Lee, K.-P., Hostetler, C. A., Kittaka, C., Rogers, R. R., Kuehn, R. E., and Liu, Z.: The CALIPSO automated aerosol classification and lidar ratio selection algorithm, *J. Atmos. Ocean. Tech.*, 26, 1994–2014, 2009.
- Parrish, D. F. and Derber, J. C.: The National Meteorological Centre's spectral statistical interpolation analysis system, *Mon. Weather Rev.*, 120, 1747–1763, 1992.
- Rabier, F., Fourrié, N., Chafaï, D., and Prunet, P.: Channel selection methods for infrared atmospheric sounding interferometer radiances, *Q. J. Roy. Meteor. Soc.*, 128, 1011–1027, 2002.
- Rodgers, C. D.: *Inverse methods for atmospheric sounding*, World Scientific, Singapore, 2000.
- Rubin, J. I. and Collins, W. D.: Global simulations of aerosol amount and size using MODIS observations assimilated with an Ensemble Kalman Filter, *J. Geophys. Res.*, 119, 12780–12806, 2014.
- Saïde, P. E., Carmichael, G. R., Liu, Z., Schwartz, C. S., Lin, H. C., da Silva, A. M., and Hyer, E.: Aerosol optical depth assimilation for a size-resolved sectional model: impacts of observationally constrained, multi-wavelength and fine mode retrievals on regional scale analyses and forecasts, *Atmos. Chem. Phys.*, 13, 10425–10444, doi:10.5194/acp-13-10425-2013, 2013.
- Sandu, A., Liao, W., Carmichael, G. R., Henze, D. K., and Seinfeld, J. H.: Inverse modeling of aerosol dynamics using adjoints: Theoretical and numerical considerations, *Aerosol Sci. Tech.*, 39, 677–694, 2005.
- Sekiyama, T. T., Tanaka, T. Y., Shimizu, A., and Miyoshi, T.: Data assimilation of CALIPSO aerosol observations, *Atmos. Chem. Phys.*, 10, 39–49, doi:10.5194/acp-10-39-2010, 2010.
- Undén, P., Rontu, L., Järvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., Garcia-Moya, J. A., Jones, C., Lenderink, G., McDonald, A., McGrath, R., Navascues, B., Nielsen, N. W., Ødegaard, V., Rodriguez, E., Rummukainen, M., Rööm, R., Sattler, K., Sass, B. H., Savijärvi, H., Schreur, B. W., Sigg, R., The, H., and Tijm, A.: HIRLAM-5 Scientific Documentation, available at: <http://www.hirlam.org> (last access: 30 March 2009), 2002.
- Veselovskii, I., Kolgotin, A., Griaznov, V., Müller, D., Wandinger, U., and Whiteman, D. N.: Inversion with regularization for the retrieval of tropospheric aerosol parameters from multiwavelength lidar sounding, *Appl. Optics*, 41, 3685–3699, 2002.
- Veselovskii, I., Kolgotin, A., Griaznov, V., Müller, D., Franke, K., and Whiteman, D. N.: Inversion of multiwavelength Raman lidar data for retrieval of bimodal aerosol size distribution, *Appl. Optics*, 43, 1180–1195, 2004.
- Veselovskii, I., Kolgotin, A., Müller, D., and Whiteman, D. N.: Information content of multiwavelength lidar data with respect to microphysical particle properties derived from eigenvalue analysis, *Appl. Optics*, 44, 5292–5303, 2005.
- Vilaplana, R., Moreno, F., and Molina, A.: Study of the sensitivity of size-averaged scattering matrix elements of nonspherical particles to changes in shape, porosity and refractive index, *J. Quant. Spectrosc. Ra.*, 100, 415–428, doi:10.1016/j.jqsrt.2005.11.068, 2006.
- Wang, Y., Sartelet, K. N., Bocquet, M., and Chazette, P.: Modelling and assimilation of lidar signals over Greater Paris during the MEGAPOLI summer campaign, *Atmos. Chem. Phys.*, 14, 3511–3532, doi:10.5194/acp-14-3511-2014, 2014.

Xu, Q.: Measureing information content from observations for data assimilation: relative entropy versus shannon entropy difference, *Tellus A*, 59, 198–209, 2006.

Zhang, J., Campbell, J. R., Hyer, E. J., Reid, J. S., Westphal, D. L., and Johnson, R. S.: Evaluating the impact of multisensor data assimilation on a global aerosol particle transport model, *J. Geophys. Res.*, 119, 4674–4689, 2014.