

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

Michael Kahnert^{1,2} and Emma Andersson²

¹Research Department, Swedish Meteorological and Hydrological Institute, Folkborgsvägen 17, SE-601 76 Norrköping, Sweden

²Department of Earth and Space Science, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

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

1 Abstract.

2 We theoretically and numerically investigate the problem of assimilating multiwavelength lidar observations of extinction
3 and backscattering coefficients of aerosols into a chemical transport model. More specifically, we consider the inverse problem
4 of determining the chemical composition of aerosols from these observations. The main questions are how much information
5 the observations contain to determine the particles' chemical composition, and how one can optimise a chemical data assimila-
6 tion system to make maximum use of the available information. We first quantify the information content of the measurements
7 by computing the singular values of the scaled observation operator. From the singular values we can compute the number of
8 signal degrees of freedom, N_s , and the reduction in Shannon entropy, H . As expected, the information content as expressed by
9 either N_s or H grows as one increases the number of observational parameters and/or wavelengths. However, the information
10 content is strongly sensitive to the observation error. The larger the observation error variance, the lower the growth rate of N_s
11 or H with increasing number of observations. The right singular vectors of the scaled observation operator can be employed
12 to transform the model variables into a new basis in which the components of the state vector can be partitioned into signal-
13 related and noise-related components. We incorporate these results in a chemical data assimilation algorithm by introducing
14 weak constraints that restrict the assimilation algorithm to acting on the signal-related model variables only. This ensures that
15 the information contained in the measurements is fully exploited, but not over-used. Numerical tests show that the constrained
16 data assimilation algorithm provides a solution to the inverse problem that is considerably less noisy than the corresponding
17 unconstrained algorithm. This suggests that the restriction of the algorithm to the signal-related model variables suppresses the
18 assimilation of noise in the observations.

19 1 Introduction

20 Atmospheric aerosols have a substantial, yet highly uncertain impact on climate, they can cause respiratory health problems,
21 degrade visibility, and even compromise air-traffic safety. The physical and chemical properties of aerosols play a key role in
22 understanding these effects. The aerosol properties are determined by a complex interplay of different chemical, microphysi-

23 cal, and meteorological processes. These processes are investigated in environmental modelling by use of chemical transport
24 models (CTMs). However, modelling aerosol processes is plagued by substantial biases and errors (McKeen et al., 2007). It is,
25 therefore, fundamentally important to evaluate and constrain CTMs by use of measurements.

26 Measurements from satellite instruments provide consistent long-term data sets with global coverage. However, it is notori-
27 ously difficult to compare measured radiances to modelled aerosol concentrations. An alternative to using radiances is to make
28 use of satellite retrieval products. For instance, one of the products of the CALIPSO lidar instrument (Cloud-Aerosol Lidar
29 and Infrared Pathfinder Satellite Observations) is a rough classification of the aerosol types (i.e. dust, smoke, clean/polluted
30 continental, and clean/polluted marine). This retrieval product is based on lidar depolarisation measurements (Omar et al.,
31 2009). For the evaluation of aerosol transport models this provides us with a qualitative check for the chemical composition of
32 aerosols. However, this is of limited practical use, since what we really need is quantitative information on the particles' chem-
33 ical composition (which can be size-dependent). The most popular approach in evaluating and constraining aerosol transport
34 models is the use of retrieved optical properties, such as aerosol optical depth, or extinction and backscattering coefficients.
35 Yet another idea is to provide the particles' refractive index as a retrieval product (e.g. Müller et al., 1999; Veselovskii et al.,
36 2002). However, the use of such retrieval products still leaves us with the challenge of solving an ill-posed inverse problem,
37 namely, of determining the particles chemical composition from their retrieved optical or dielectric properties.

38 A systematic class of statistical methods for solving this inverse problem is known as data assimilation. Recent studies
39 have applied data assimilation to aerosol models with varying degrees of sophistication, ranging from simple dust mod-
40 els (Khade et al., 2013) and mass transport models (Zhang et al., 2014) to microphysical aerosol models based on modal
41 (Rubin and Collins, 2014) or sectional descriptions (Sandu et al., 2005; Saide et al., 2013) of the aerosol size distribution. The
42 assimilation techniques that have been used comprise variational methods, such as 2D (Zhang et al., 2014), 3D (Kahnert,
43 2008; Liu et al., 2011), and 4D variational methods (Benedetti et al., 2009), as well as ensemble approaches (Sekiyama et al.,
44 2010). Assimilation of satellite products for trace gases is relatively straightforward, since observed and modelled trace gas
45 concentrations are almost directly comparable. However, aerosol optical properties observed from satellites are not directly
46 comparable to the modelled size distribution and chemical composition of the aerosols. Solving this problem amounts to reg-
47 ularising a severely under-constrained inverse problem. Previous aerosol assimilation attempts have been mainly based on
48 educated guesses about the information content of the observations. For instance, there have been studies on the assimilation of
49 aerosol optical depth (AOD) in which all chemical aerosol components in all size classes and at all model layers were used as
50 independent control variables (Liu et al., 2011). This approach largely disregards the problems involved in inverse modelling.
51 By contrast, it has been proposed to only allow for the total aerosol mass concentration to be corrected by data assimilation
52 of AOD (Benedetti et al., 2009; Wang et al., 2014). This is a more prudent approach based on the plausible assumption that a
53 single optical variable only contains enough information to control a single model variable. There have also been intermediate
54 approaches in which the total aerosol mass per size bin have been used as control variables (Saide et al., 2013).

55 In all such approaches the choice of control variables is based on ad hoc assumptions. Numerical assimilation experiments by
56 Kahnert (2009) suggest that observations of several aerosol optical properties at multiple wavelengths may allow us to constrain
57 more than just the total mass concentration, but certainly not *all* aerosol parameters. However, it is still an unsolved mystery how

58 much information a given set of observations actually contains about the size distribution and chemical composition of aerosols,
59 and exactly which model variables are related to the observed signals, and which ones are related to noise. Thus a prerequisite
60 for assimilating remote sensing observations into aerosol transport models is to thoroughly understand the information content
61 of the observations as well as the relation between the model variables and the signal degrees of freedom.

62 In numerical weather prediction (NWP) modelling, several studies have discussed the information content of satellite obser-
63 vations for meteorological variables. For instance, Joiner and da Silva (1998) applied a singular-value decomposition (SVD)
64 approach in order to reduce the effect of prior information in the analysis, so that the retrieval and forecast errors can be as-
65 sumed to be uncorrelated. Rabier et al. (2002) considered assimilation of IR sounders, which typically provide a large number
66 of different channels. They applied methods of information and retrieval theory in order to decide which channels contain most
67 information about the vertical variation of temperature and humidity. Cardinali et al. (2004) employed the influence matrix to
68 compute diagnostics of the impact of observations in a global NWP data assimilation system. Johnson et al. (2005a, b) investi-
69 gated filtering and interpolation aspects in a 4DVAR assimilation system by use of an SVD approach. They also used Tikhonov
70 regularisation theory to optimise the signal-to-noise regularisation parameter in order to maximise the information that can be
71 extracted from observations. Xu (2006) compared different metrics, namely, the relative entropy and the Shannon-entropy dif-
72 ference, to measure information contents of radar observations assimilated into a coupled atmosphere-ocean model. Bocquet
73 (2009) used methods of information theory to address the question how to determine an optimum spatial resolution of the
74 discretised space of control variables in geophysical data assimilation.

75 Burton et al. (2016) have recently investigated the information content of “ $3\beta + 2\alpha$ ” lidar measurements, i.e., observations of
76 backscattering at three wavelengths and extinction at two wavelengths, where the information content was analysed with regard
77 to the refractive index and number distribution of the aerosol particles. Veselovskii et al. (2004, 2005) have performed similar
78 analyses of the information content of multiwavelength Raman lidar measurements with regard to the complex refractive index
79 and the effective radius of the aerosol particles. As mentioned earlier, the refractive index is a very useful retrieval product of
80 remote sensing observations. However, from the point of view of chemical transport modelling, the main quantities of interest
81 are the concentrations of the different chemical species of which the aerosol particles are composed. Although the chemical
82 composition determines the refractive index, the inversion of this relationship is still under-determined, hence an ill-posed
83 problem. In the present paper, we want to investigate the inverse problem that goes all the way from optical properties to the
84 chemical composition of particles.

85 The two main goals of this paper are (i) to apply a systematic method for analysing the information content of aerosol optical
86 properties with regard to the particles’ chemical composition, and (ii) to test an algorithm for making an automatic choice
87 of control variables in chemical data assimilation such that all control variables are signal-related, while the noise-related
88 variables remain unchanged by the assimilation procedure. The main hypothesis is that by constraining the data assimilation
89 algorithm to acting on the signal-related variables only, the output will be less noisy than in an unconstrained assimilation.
90 The focus of our study will be on spectral observations of extinction and backscattering coefficients, which can be retrieved

91 from lidar observations.¹ We will not restrict this analysis to any fixed choice of wavelengths, such as $3\beta + 2\alpha$. Instead, we will
92 investigate the information content for varying combinations of the three main wavelengths of the commonly used neodymium-
93 doped yttrium aluminium garnet (Nd:YAG) laser. However, it should be mentioned that extinction measurements at the lowest
94 harmonic of 1064 nm can be difficult and plagued by high errors; in practice, this will affect the observation error, resulting in
95 a low information content of this particular measurement.

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

101 **2 Methods**

102 This study consists of two parts. In the first part we quantify the information content of extinction and backscattering coeffi-
103 cients at multiple wavelengths. In the second part we perform a numerical test to investigate to what extent the concentrations
104 of different chemical aerosol components can be constrained by observations of extinction and backscattering coefficients.
105 The modelling tools required for this study are (i) a chemical transport model; (ii) an aerosol optics model; and (iii) a data
106 assimilation system.

107 **2.1 Multiple scale Atmospheric Transport and CHemistry modelling system (MATCH)**

108 We employ the chemical transport model MATCH, which is an off-line Eulerian CTM with flexible model domain. It has been
109 previously used from regional to hemispheric scales. Here we use a model version that contains a photochemistry module with
110 64 chemical species, among them four secondary inorganic aerosol (SIA), namely, ammonium sulphate, ammonium nitrate,
111 other sulphates, and other nitrates. It also contains a module with 16 primary aerosol variables, namely, sea salt, elemental
112 carbon (EC), organic carbon (OC), and dust particles, each emitted in four different size bins. Thus, the model contains 20
113 different aerosol variables. The size ranges of the four bins are as follows.

114 Size bin 1: 10–50 nm

115 Size bin 2: 50–500 nm

116 Size bin 3: 500–1250 nm

¹In 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 polarised components. The CATS instrument on-board the International Space Station has been operational since January 2015; It measures backscattering at 355 nm, 532 nm, and 1064 nm, were the latter two have two orthogonal polarisation 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).

117 Size bin 4: 1250–5000 nm.

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

122 We mention that there exists another model version that includes aerosol microphysical processes, such as nucleation, con-
123 densational growth, and coagulation. In that model version the aerosol size distribution evolves dynamically. The model has
124 20 size bins and seven chemical species (EC, OC, dust, sea salt, particulate sulphate (PSOX), particulate nitrate (PNOX), and
125 particulate ammonium (PNHX)), although not all species are encountered in all size bins. The total number of model variables
126 currently in that version is 82.

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

129 For the sake of simplicity we here use the mass transport model without aerosol microphysical processes (see next sec-
130 tion). The model is set up over Europe covering 33° in the longitudinal and 42° in the latitudinal direction in a rotated lat-
131 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
132 terrain-following coordinates. The meteorological input data are taken from the numerical weather prediction model HIRLAM
133 (Undén et al., 2002). For the emissions of all aerosol components we used EMEP data for the year 2007, where EC and OC
134 emissions were computed from total primary particle emissions based on the data in Kupiainen and Klimont (2004, 2007).

135 **2.2 Aerosol optics model**

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

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

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

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

149 2.3 Three-dimensional variational data assimilation (3DVAR)

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

$$156 \mathbf{y} = \hat{H}(\mathbf{x}) + \epsilon_o, \quad (1)$$

157 where \hat{H} is known as the observation operator, and ϵ_o denotes the vector of observation errors. The problem is to determine the
 158 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
 159 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
 160 (3DVAR), the maximum-likelihood solution is found by numerically minimising the cost function

$$161 J = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \cdot \mathbf{B}^{-1} \cdot (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}[\hat{H}(\mathbf{x}) - \mathbf{y}]^T \cdot \mathbf{R}^{-1} \cdot [\hat{H}(\mathbf{x}) - \mathbf{y}]. \quad (2)$$

162 Data assimilation is commonly employed for constraining model results by use of observations. However, one can also
 163 employ data assimilation as an inverse-modelling tool, i.e. for retrieving a model state from measurements. A summary of the
 164 theoretical basis of variational data assimilation is given in the appendix.²

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

²Many 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.

170 2.4 Analysis of the information content of aerosol optical parameters

171 The questions we ask are these.

172 1. Suppose we have an n dimensional model space. Given m observations (e.g., m_1 different parameters at m_2 different
173 wavelengths, so that $m_1 \cdot m_2 = m$), how many independent model variables $N \leq n$ can we constrain with the observa-
174 tions? Obviously, the best we can achieve would be $N = \min\{m, n\}$; but often we will have $N < \min\{m, n\}$.

175 2. Which are the N model variables (or linear combinations of model variables) that can be constrained by the measure-
176 ments?

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

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

$$181 \quad y = x + \epsilon_o. \quad (3)$$

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

$$185 \quad x_a = \frac{\sigma_b^2 y + \sigma_o^2 x_b}{\sigma_b^2 + \sigma_o^2}. \quad (4)$$

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

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

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

$$193 \quad \mathbf{y} = \hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x} + \boldsymbol{\epsilon}_o, \quad (5)$$

194 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
195 obtained from a model run. The (in general non-square) matrix \mathbf{H} is the main quantity we need to investigate in order to
196 address the questions formulated at the beginning of this subsection. It is transformed to the so-called observability matrix
197 $\tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2}$, where \mathbf{R} is the observation error covariance matrix, and \mathbf{B} denotes the error covariance matrix of the
198 background estimate. Subsequently, one performs a singular-value decomposition (SVD)

$$199 \quad \mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2} = \mathbf{V}_L \cdot \mathbf{W} \cdot \mathbf{V}_R^T, \quad (6)$$

200 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
 201 singular values along the main diagonal, while all other matrix elements are zero. It turns out that the singular values w_i can
 202 be employed to compute the number of signal degrees of freedom N_s according to

$$203 \quad N_s = \sum_{i=1}^{\min\{n,m\}} w_i^2 / (1 + w_i^2). \quad (7)$$

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

$$207 \quad H = \frac{1}{2} \sum_{i=1}^{\min\{n,m\}} \log_2(1 + w_i^2). \quad (8)$$

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

211 By performing the transformation

$$212 \quad \delta \mathbf{x}' = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \mathbf{x} \quad (9)$$

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

$$219 \quad J_G = \frac{1}{2} \delta \mathbf{x}^T \cdot \mathbf{B}^{-1/2} \cdot \mathbf{V}_R \cdot \mathbf{B}_G^{-1} \cdot \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \mathbf{x}, \quad (10)$$

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

$$221 \quad \mathbf{B}_G = \sigma_G \text{diag}(w_1, w_2, \dots, w_K, c, \dots, c). \quad (11)$$

222 Here $K = \min\{n, m\}$, and the number c is assumed to be much smaller than the smallest singular value. We note that the
 223 formulation of the constraint term in Eq. (11) is by no means unique. Other possible choices of the matrix \mathbf{B}_G are discussed
 224 in appendix D3. However, we performed preliminary tests which indicate that the constrained 3DVAR approach is not very
 225 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
 226 phase-space variables are tightly constrained, while the signal-related variables can be varied relatively freely by the analysis.
 227 The free parameters σ_G and c should be tuned in such a way that the constrains are neither too hard nor too soft. In the former
 228 case, the analysis will stay too close to the background estimate. In the latter case, it will not differ much from the unconstrained
 229 analysis.

230 2.5 Numerical test of the constrained assimilation algorithm

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

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

246 3 Results

247 3.1 Analysis of the information content of aerosol optical parameters

248 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_{ext} and β_{sca} denote the extinction
249 and backscattering coefficients, respectively, and the wavelengths $\lambda_1 = 1064$ nm, $\lambda_2 = 532$ nm, and $\lambda_3 = 355$ nm denote the
250 first three Nd:YAG harmonics. Hereafter, we will abbreviate these parameters by $k_{\text{ext}}(\lambda_i) = k_i$, $\beta_{\text{sca}}(\lambda_j) = \beta_j$, $i = 1, 2$, $j =$
251 $1, 2, 3$. Out of this five-parameter set we pick different subsets and analyse the singular values of the corresponding observability
252 matrices. From those we compute the number of signal degrees of freedom as well as the change in Shannon entropy for each
253 subset of measurements. We will focus on those parameter subsets that are technically relevant in practical lidar applications.

254 Table 2 shows the number of signal degrees of freedom N_s and the reduction in Shannon entropy H for different values of
255 the observation standard deviation σ_o . For low values of σ_o , the number of signal degrees of freedom is identical to the number
256 of observational parameters. However, as we increase σ_o we observe a decrease in N_s . For instance, for $\sigma_o = 100$ % the five
257 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
258 entropy H displays an analogous behaviour. For instance, for $\sigma_o = 1$ % we see that H consistently increases as one increases
259 the number of observational parameters. This is much less pronounced for $\sigma_o = 100$ %. In that case, H does increase as one
260 goes from a single parameter to two parameters (compare the first to the second and fourth rows). However, as one adds more

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 β_i represents the extinction and backscattering coefficient, respectively, at the wavelengths $\lambda_1 = 1064$ nm, $\lambda_2 = 532$ nm, and $\lambda_3 = 355$ nm.

No.	Obs. Std. dev. [%] Parameters	1		5		10		50		100	
		N_s	H	N_s	H	N_s	H	N_s	H	N_s	H
1.	β_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.	β_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

261 parameters, the increase in H slows down considerably. For five parameters (last row), H is only about twice as high as for a
 262 single parameter (first row).

263 This illustrates the pivotal importance of the observation error for the amount of information that can be obtained from mea-
 264 surements. It is important to understand that the observation error ϵ_o is not the same as the measurement error ϵ_m . Rather, in our
 265 case we have $\epsilon_o = \epsilon_m + \epsilon_f$, where ϵ_f denotes the forward-model error [see, e.g., Eq. (1) and accompanying text in Rabier et al.
 266 (2002)]. Any simplifying assumptions in the optics model or incomplete knowledge of the particle size distribution, morphol-
 267 ogy, chemical composition, or dielectric properties can contribute to ϵ_f . Such assumptions enter into our relatively simple
 268 optics model.³ Note also that in operational applications there may be other terms contributing to ϵ_o . For instance, if a point
 269 measurement is taken at a location that does not provide a good representation of the grid-cell average, then one would have to
 270 add a representativity error ϵ_r to the observation error.

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

272 1. In order to make the forward-model error ϵ_f as small as possible, it is essential to develop accurate and realistic aerosol
 273 optics models. The most accurate measurements may intrinsically contain a wealth of information on aerosol properties.
 274 But we can only make use of this information to the extent that our observation operator is able to accurately describe
 275 the relation between the physical and chemical particle characteristics and their optical properties.

276 2. It is equally essential to accurately estimate the contribution of the uncertainties in the aerosol optics model, i.e., to
 277 estimate the forward-model error ϵ_f . If we underestimate this error, we will rely too much on the measurements than we
 278 should, thus assimilating noise. If we overestimate this error, we will waste information contained in the observations.
 279 In practice, one way to estimate ϵ_f is to compute optical properties while varying the particles' size, morphology, and

³A 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 linearised and investigated in a similar way.

280 dielectric properties within typical ranges. The resulting variation in the optical properties then allows us to estimate ϵ_f .
281 (For a review of aerosol optics modelling see Kahnert et al. (2014, 2016) and references therein).

282 In Table 2 we sorted the results for N_s and H by different values of the observation standard deviation. However, it is
283 important to realise that the results also depend on the background error standard deviation, or, more precisely, on how large
284 the background error standard deviations are compared to the observation error standard deviations. Johnson et al. (2005a)
285 made this point very explicit. They discussed an idealised case with diagonal background error covariance matrix $\mathbf{B} = \sigma_b^2 \mathbf{1}$ and
286 observation error covariance matrix $\mathbf{R} = \sigma_o^2 \mathbf{1}$. They considered the case of direct measurements, i.e., the model variables and
287 the observed parameters are the same type of variables. Under such idealised conditions, they showed that one can maximise
288 the amount of information that can be obtained from the observations by optimising the regularisation parameter σ_b/σ_o (or,
289 equivalently, the regularisation parameter σ_o^2/σ_b^2). In our more general case, instead of σ_b we need to consider the full matrix
290 $\mathbf{B}^{1/2}$, instead of σ_o^{-1} we need to consider $\mathbf{R}^{-1/2}$, and in order to compare the two matrices we need to first transform $\mathbf{B}^{1/2}$
291 from model to observation space according to $\mathbf{H} \cdot \mathbf{B}^{1/2}$. So in place of σ_b/σ_o we need to consider the more general quantity
292 $\mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2}$, and we need to diagonalise it by a singular value decomposition according to Eq. (6). Thus the singular
293 values w_i generalise the parameter σ_b/σ_o . The latter applies to the case of direct observations and error covariance matrices
294 that are proportional to unit matrices. The former apply to the general case of non-diagonal error covariance matrices and
295 indirect observations.

296 From this we learn that the singular values w_i provide us with a (however abstract) means to quantify how the background
297 standard deviations compare to the observation standard deviations. We pick one of the columns in Tab. 2, namely, the one
298 for $\sigma_o = 50\%$, and expand it in Tab. 3. We show the singular values w_i , as well as their contributions $N_s^i = w_i^2/(1 + w_i^2)$ and
299 $H_i = 0.5 \log_2(1 + w_i^2)$ to the sums in Eqs. (7) and (8), respectively. The results reveal that the singular values w_i can decrease
300 quite rapidly from the largest to the smallest value (see, e.g., case No. 6 in the table). However, the corresponding contribution
301 N_s^i to the number of signal degrees of freedom changes rather smoothly. Even those singular values that are only slightly
302 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
303 corresponding contribution N_s^i becomes much smaller than 1 (see $i = 5$ in case No. 6).

304 Let us now compare the different subsets of parameters in Tab. 2 and 3. In case No. 1 we observe a single parameter that
305 provides a single degree of freedom. In cases No. 2 and 4 we observe two parameters, which nearly doubles N_s . Comparison of
306 these two cases shows that it does not make a significant difference whether we observe backscattering coefficients at different
307 wavelengths, or both extinction and backscattering coefficients each at a single wavelengths. In either case the measurements
308 provide roughly the same amount of information (in terms of N_s or H). The same is true when considering three observational
309 parameters (compare cases No. 3 and 5). The $3\beta + 2\alpha$ case (No. 6) clearly provides the largest amount of information in
310 comparison to the other cases. However, as we saw in Tab. 2, observation errors that are large in comparison to the background
311 errors can significantly reduce the effective information that can be assimilated into a model.

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.	β_3	1	38.2	1.00	5.26	1.00	5.26
2.	β_1, β_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.	β_3, k_3	1	83.3	1.00	6.38	1.92	8.22
		2	3.43	0.92	1.84		
5.	β_1, β_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		

312 3.2 Numerical inverse-modelling test

313 We integrated the findings of 3.1 into our 3DVAR program by constraining the algorithm to varying only the signal-related
314 model variables. To illustrate the method we conduct a numerical test as described in Sect. 2.5. We perform a 3DVAR analysis
315 by assimilating “ $3\beta + 2\alpha$ ” profiles, i.e., synthetic lidar measurements of β_{sca} at the three wavelengths 1064, 532, and 355 nm
316 together with k_{ext} at the two wavelengths 532 and 355 nm. Thus in our case the number of singular values in each vertical
317 layer is $K = 5$. We assume an idealised situation in which the observation standard deviation is only 10 %. As we see in Table
318 2 (case No. 6), the number of signal degrees of freedom is $N_s = 4.9$ in this case. So we roughly have as many signal degrees
319 of freedom as we have measurements.

320 Figure 1 shows vertical profiles of selected aerosol components, namely (from top to bottom): organic carbon (OC) in the 3rd
321 size bin (OC-3), OC in the 4th size bin (OC-4), elemental carbon (EC) in the 3rd size bin (EC-3), and mineral dust in the 1st size
322 bin (DUST-1). The reference and background mixing ratios are shown in black and green, respectively. The 3DVAR analysis
323 was first performed without any constraints; the results are shown in the left column by the blue line. Then the 3DVAR analysis

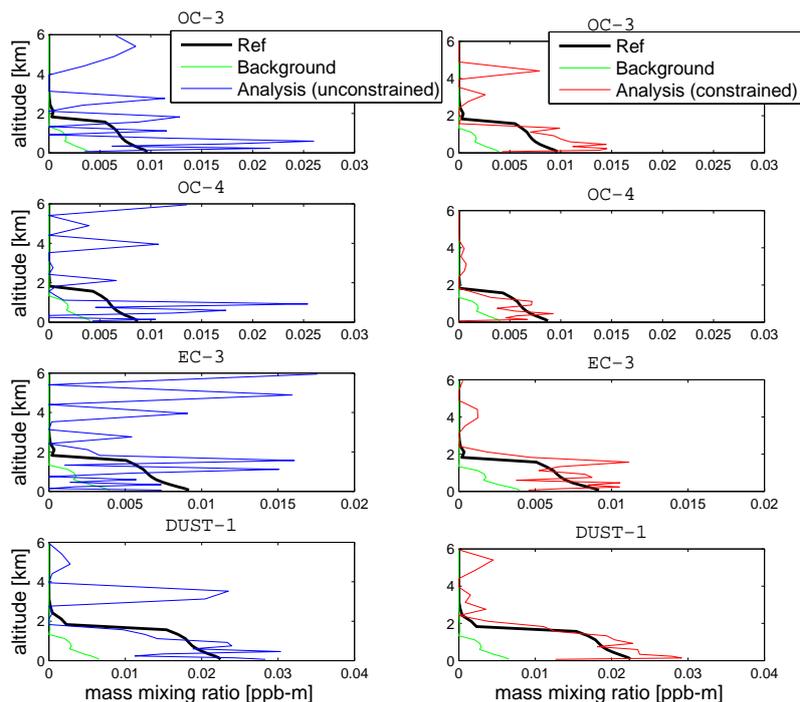


Figure 1. Vertical profiles of selected aerosol components in different size bins. From top to bottom: organic carbon in the 3rd size bin (OC-3), OC in the 4th size bin (OC-4), elemental carbon in the 3rd size bin (EC-3), and dust in the 1st size bin (DUST-1). The reference results are shown in black, 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.

324 was repeated with the constraints in Eq. (10) and (11); the results are represented in the right column by the red line. Clearly,
 325 the unconstrained analysis (blue lines in the left panels) yields results that oscillate quite erratically in the vertical direction.
 326 Also, the unconstrained analysis can yield conspicuously high values at higher altitudes, even though both the reference and
 327 background values are both close to zero. By contrast, the constrained analysis (red lines in the right panels) yields results that
 328 better agree with the reference results. The noisiness in the vertical direction is significantly reduced, and the results at higher
 329 altitudes are generally lower than those obtained with the unconstrained analysis.

330 Figure 2 shows analogous results for the mass mixing ratios of different aerosol components, each summed over all size bins.
 331 The aerosol components are (from top to bottom): elemental carbon (EC), organic carbon (OC), mineral dust (DUST), sea salt
 332 (NaCl), secondary inorganic aerosols (SIA, i.e., the sum over all sulphate, nitrate, and ammonium species), and PM10 (i.e., the
 333 sum over all aerosol components). Clearly, the constrained analysis faithfully retrieves both PM10 and SIA. The unconstrained
 334 analysis performs almost equally well for these two variables. Sea salt and mineral dust are not well retrieved from the mea-
 335 surements in either the constrained or unconstrained approach. EC and OC are very well retrieved by the constrained analysis.

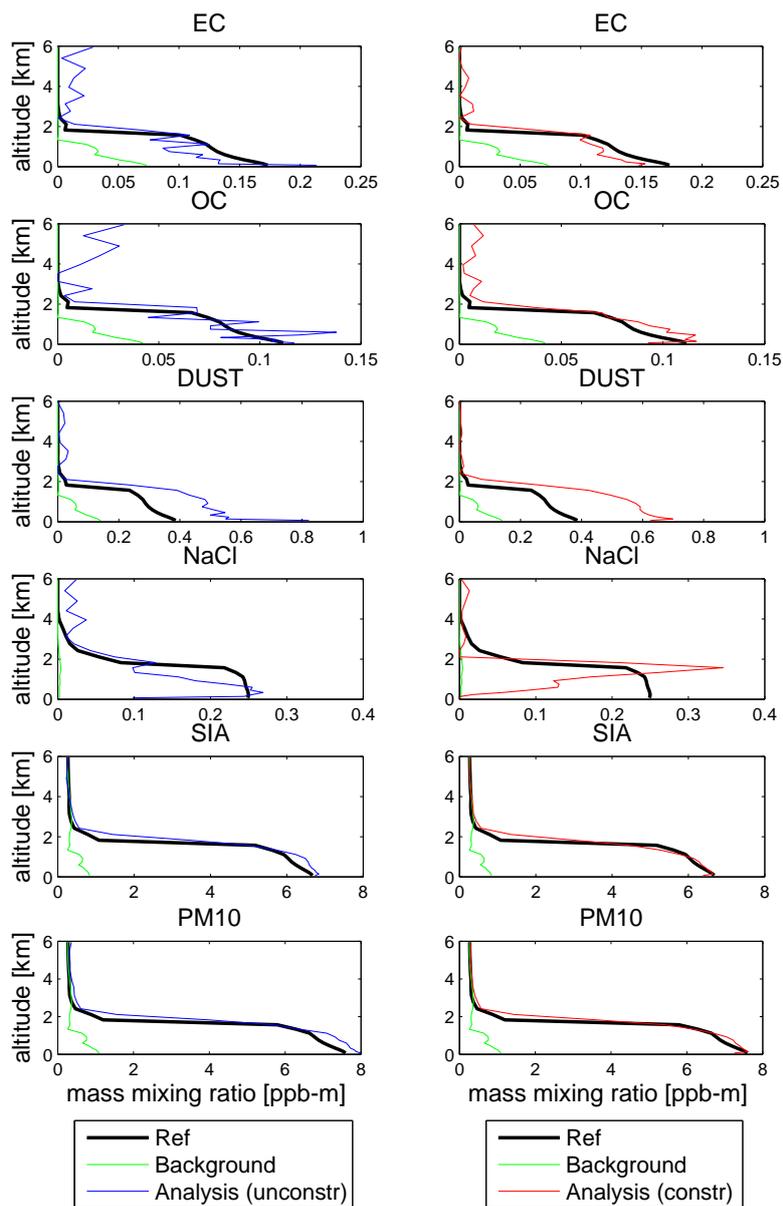


Figure 2. As Fig. 1, but for the total mass mixing ratio (summed over all size bins). The components are (from top to bottom): EC, OC, mineral dust, sea salt, secondary inorganic aerosols (sum of all sulphate, nitrate, and ammonium species), and PM10 (sum of all aerosol components).

336 For these components, the unconstrained analysis has a very small bias compared to the reference results, but it is considerably
 337 more noisy (i.e., oscillating in the vertical direction) than the constrained analysis. We also see, again, that the mixing ratios

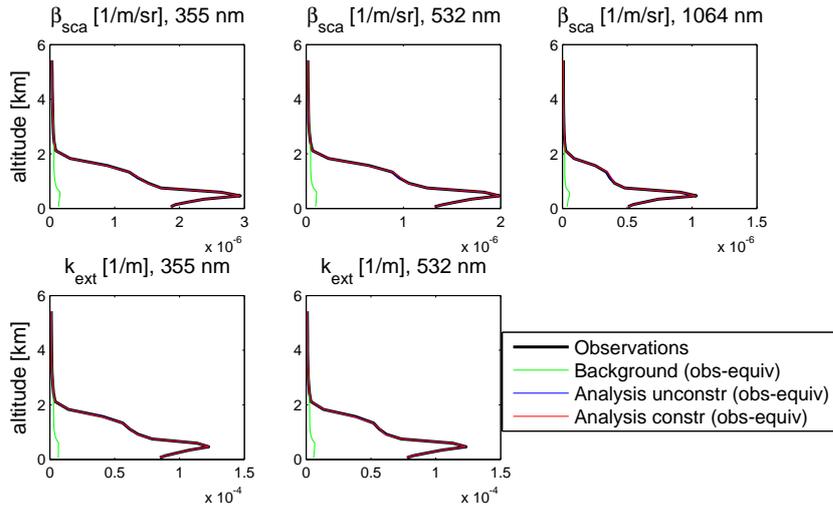


Figure 3. Observations (black solid line), and observation-equivalents of the background estimate (green), and of the unconstrained (blue) and constrained (red) 3DVAR analysis. The optical parameters and wavelengths are indicated above each panel.

338 at higher altitudes obtained with the unconstrained analysis can be unreasonably high. This is especially pronounced for OC.
 339 In general, however, the problems we encounter in the unconstrained analysis are less pronounced in Fig. 2 than in Fig. 1. A
 340 possible explanation is that SIA may be most strongly related to the measurement signal, and SIA is dominating the aerosol
 341 mass in this case. We will return to this point shortly. Another possible factor is that the noise in the analysis can be damped
 342 by summing up results over several size bins.

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

349 We have seen that the analysis provides a reasonable, but, as expected, not a perfect answer to the inverse problem. We have
 350 further seen that at the observation site it relies more on the observations than on the background estimate. Most importantly,
 351 we have seen that the constraints introduced in the 3DVAR algorithm suppress noise in the analysis, especially in EC and OC.
 352 However, the previous figures do not provide us with any direct insight of how exactly the constraints accomplish this. To learn
 353 more about that we need to inspect the analysis in the abstract phase space of the transformed model variables $\delta\mathbf{x}'$. (Recall that
 354 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,
 355 20 variables $\delta x'_i$. The background estimate corresponds to $\delta x'_i = 0$ and is represented by the green line. The unconstrained
 356 3DVAR analysis increment is represented by the blue line, the constrained 3DVAR analysis increment is shown by the red

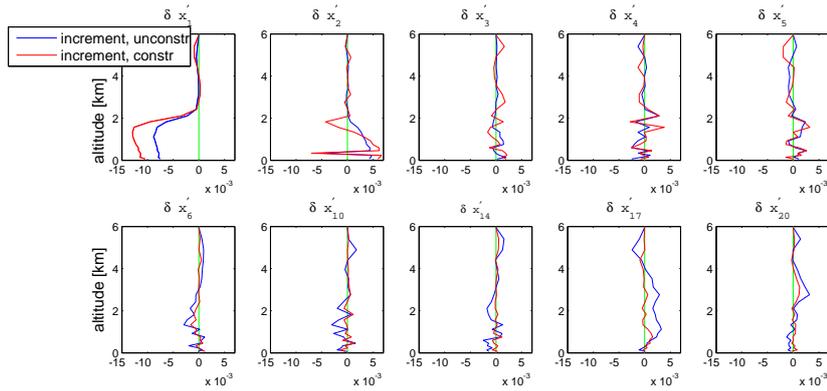


Figure 4. Vertical profiles of the transformed model variables $\delta x'$, defined in Eq. (9). The figure shows results obtained with the constrained (red) and unconstrained (blue) 3DVAR analysis.

357 line. The first five phase space elements in the top row are the signal-related control variables. Generally, the magnitude of the
 358 constrained increments (red) is larger than that of the unconstrained increments (blue). The noise-related phase space elements,
 359 five of which are shown in the bottom row, display the opposite behaviour. The constrained increments are close to zero, as
 360 they should. The unconstrained elements consistently show higher magnitudes than the constrained elements. However, we
 361 also see that the unconstrained analysis does produce increments that are largest for the two elements $\delta x'_1$ and $\delta x'_2$, which
 362 most strongly relate to the measurement signal. Based on our single test case we cannot say if this is a lucky coincidence
 363 or a consistent property. If the latter, it may indicate that we are using rather reasonable background error statistics, so that
 364 the analysis increment in observation space is distributed to the different variables in model space in a sensible way. If the
 365 former, it could be the case that the success of the unconstrained analysis is largely dependent on whether or not those aerosol
 366 components dominate the total aerosol mass that most strongly relate to the signal degrees of freedom. (In our case the total
 367 mass is dominated by SIA, which is very well retrieved by the analysis).

368 Finally, we want to obtain a better understanding of how the aerosol components x in model space, or their increments δx , are
 369 linked with the signal-related phase-space elements $\delta x'$. To this end we inspect the first five row vectors of the transformation
 370 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
 371 of δx 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
 372 for $j = 1, \dots, 20$, where 5 is the number of signal-related phase-space elements, and 20 is the number of aerosol components
 373 in model space. Results are shown for model layers 2 (left column) and 22 (right column), which correspond to altitudes of
 374 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
 375 four SIA components, i.e., sulphates (SOX) other than $(\text{NH}_4)_2\text{SO}_4$, ammonium sulphate (AS), ammonium nitrate (AN), and
 376 nitrates (NOX) other than NH_4NO_3 .

377 Comparison of the two columns clearly demonstrates that the elements of the transformation matrix can vary considerably
 378 with vertical layer (or, more generally, with location). This is because the error covariance matrix \mathbf{B} varies with location, and

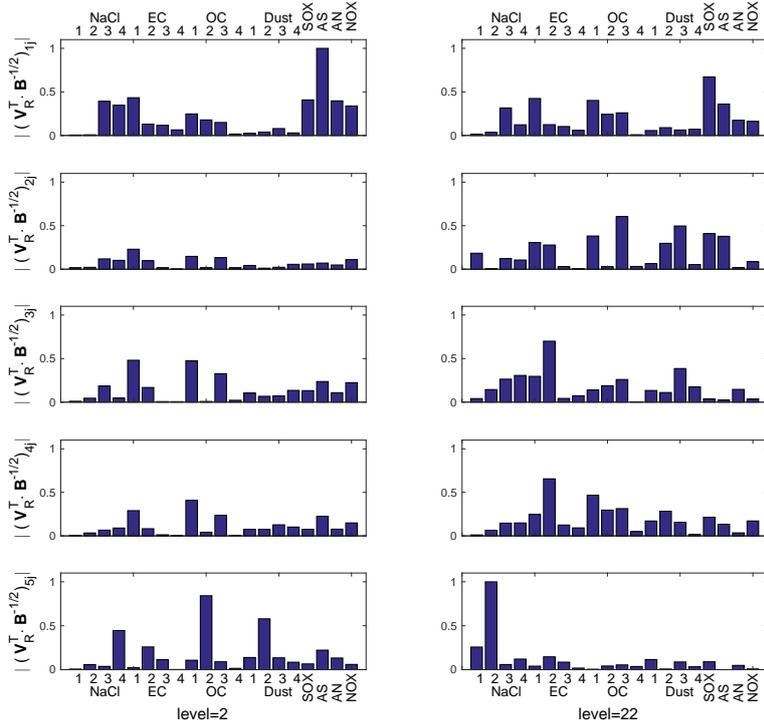


Figure 5. The first five rows (from top to bottom) of the matrix $\mathbf{V}_R^T \cdot \mathbf{B}^{-1/2}$ at the observation site, and for model layers 2 (left) and 22 (right). The y-values are normalised by dividing them by the maximum element. The x-axis indicates the aerosol components in model space 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: sulphates (SOX) other than $(\text{NH}_4)_2\text{SO}_4$, ammonium sulphate (AS), ammonium nitrate (AN), and nitrates (NOX) other than NH_4NO_3 .

379 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
 380 also dependent on location — see Eq. (6). Consequently, it is very difficult to draw general conclusions about which aerosol
 381 components make a dominant contribution to the signal-related phase-space variables; this can vary with location, and it can
 382 vary for different data sets.

383 However, in our case the SIA components consistently make a strong contribution to the first signal-related element $\delta x'_1$.
 384 Since SIA is dominating the aerosol mass mixing ratio in this test case, the analysis was able to retrieve PM10. We also see that
 385 the dust components make only a weak contribution to most of the signal-related elements $\delta x'_i$, especially to the first one. This
 386 is a likely explanation for the difficulties encountered in retrieving the dust mass mixing ratio. Sea salt is more complicated.
 387 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
 388 not make a significant contribution to most of the five signal-related control variables. In our case the sea salt mass is strongly
 389 dominated by the second size bin (not shown). This explains the difficulties we encountered in the retrieval of sea salt.

390 4 Summary and conclusions

391 We have quantified the information content of multiwavelength lidar measurements with regard to the chemical composition
392 of aerosol particles. Different combinations of extinction and backscattering observations at several wavelengths have been
393 investigated by determining the singular values of the scaled observation operator, by computing the number of signal degrees
394 of freedom N_s , and by calculating the reduction in Shannon entropy H caused by taking measurements. We first quantified
395 N_s and H as a function of observation standard deviation σ_o . The information content of the observations, as expressed by N_s
396 and H , decreased as σ_o was increased. This became the more pronounced the larger the number of simultaneously observed
397 parameters was.

398 The observation error depends not only on the measurement error, but also on the forward-model error. The latter de-
399 pends on the uncertainties in the aerosol-optics model. This highlights the importance of developing accurate aerosol optics
400 models and of obtaining an accurate estimate of the observation error, especially of the uncertainty in the aerosol optics
401 model. This is a prerequisite for extracting as much information as possible from the measurements, while avoiding to ex-
402 tract noise rather than signal. More often than not, computational limitations and lack of knowledge force us to introduce
403 simplifying assumptions about the particles' morphologies. However, we know that aerosol optical properties can be highly
404 sensitive to the shape (Mishchenko et al. (1997); Kahnert (2004)), small-scale surface roughness (Kahnert et al., 2012b), inho-
405 mogeneity (Mishchenko et al., 2014; Kahnert, 2015), aggregation (Fuller and Mackowski, 2000; Liu and Mishchenko, 2007;
406 Kahnert and Devasthale, 2011), irregularity (Muinonen, 2000; Bi et al., 2010), porosity (Vilaplana et al., 2006; Lindqvist et al.,
407 2011; Kylling et al., 2014), and combinations thereof (Lindqvist et al., 2009; Kahnert et al., 2013; Lindqvist et al., 2014). We
408 need to know how much these sources of uncertainty contribute to the observation standard deviation. One way of estimating
409 this is to compare aerosol optical properties computed with simple shape models to either measurements or to computations
410 based on more realistic particle shape models — see Kahnert et al. (2014) for a recent review and a more detailed discussion.

411 The singular values of the scaled observation operator provide us with an abstract measure to compare the standard de-
412 viations of the background (prior) estimate to those of the observations. The reason why this is a rather abstract measure is
413 because background and observation errors are, in general, in different spaces and cannot be directly compared. However,
414 we constructed a mapping that transforms the state vector in physical (model) space to an abstract phase space in which the
415 components of the state vector can be partitioned into signal-related and noise-related components. The singular values indi-
416 cate to what extent the signal-related phase-space variables can be constrained by the measurements. We exploited this fact
417 by constructing weak constraints in a 3DVAR data assimilation code, which limited the assimilation algorithm to acting on
418 the signal-related phase-space variables only (hereafter referred to as the *constrained analysis*). The idea was to maximise the
419 use of information, while avoiding the risk of assimilating noise by over-using the measurements. Thus, our main hypothesis
420 was that the constrained analysis will yield less noisy results than the unconstrained analysis. Numerical tests confirmed this
421 hypothesis. Notably in the case of elemental carbon (EC) and organic carbon (OC) the unconstrained analysis gave mixing
422 ratios that oscillated considerably in the vertical direction. The constrained analysis results were considerably less noisy.

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

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

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

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

457 Appendix A: Inverse problems

458 Suppose we have a system described by a set of variables x_1, \dots, x_n , summarised in a vector \mathbf{x} . Suppose also that we have an
459 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 , summarised in a vector \mathbf{y} . To
460 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
461 optical properties, and \hat{H} as an aerosol optics model. The operator \hat{H} maps from model space into observation space, which
462 allows us to compare model output and observations. We consider the following two problems:

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

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

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

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

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

474 Appendix B: Three-dimensional variational data assimilation

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

478 In a CTM we discretise the geographic domain of interest into a three-dimensional grid. In each grid cell, the aerosol
479 particles are characterised by the mass mixing ratio of each chemical component in the aerosol phase, such as sulphate, nitrate,
480 ammonium, mineral dust, black carbon, organic carbon, and sea salt. Suppose we summarise all these mass mixing ratios
481 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,
482 known as a *background estimate* \mathbf{x}_b .⁴ Suppose also that we have m observations, which we summarise in a vector $\mathbf{y} \in \mathbb{R}^m$. We
483 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

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

484 space⁵. 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
 485 $\boldsymbol{\epsilon}_o = \hat{H}(\mathbf{x}_t) - \mathbf{y}$ the observation error.⁶ The background and observation errors are assumed to be unbiased and uncorrelated
 486 with each other. Then their joint probability distribution becomes separable, i.e.

$$487 \quad P(\boldsymbol{\epsilon}_b, \boldsymbol{\epsilon}_o) = P_b(\boldsymbol{\epsilon}_b)P_o(\boldsymbol{\epsilon}_o). \quad (\text{B1})$$

488 The true state of the atmosphere is, of course, unknown. Therefore, our definition of the errors and their probability distri-
 489 bution is only of conceptual use, but not of any practical value. However, we can reinterpret the probability distributions by
 490 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
 491 that both the background and the observation errors are normally distributed. Thus we may write

$$492 \quad P_b(\mathbf{x}) = (2\pi |\mathbf{B}|)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \cdot \mathbf{B}^{-1} \cdot (\mathbf{x} - \mathbf{x}_b)\right) \quad (\text{B2})$$

$$493 \quad P_o(\mathbf{x}) = (2\pi |\mathbf{R}|)^{-1/2} \exp\left(-\frac{1}{2}(\hat{H}(\mathbf{x}) - \mathbf{y})^T \cdot \mathbf{R}^{-1} \cdot (\hat{H}(\mathbf{x}) - \mathbf{y})\right). \quad (\text{B3})$$

494 Here \mathbf{B} and \mathbf{R} denote the covariance matrices of the background and observation errors, respectively, and $|\cdot|$ denotes the
 495 matrix determinant. In this form, $P_b(\mathbf{x})$ represents the probability that the atmospheric aerosol particles are found in state \mathbf{x} ,
 496 given a background estimate \mathbf{x}_b with error covariance matrix \mathbf{B} . Similarly, $P_o(\mathbf{x})$ is the probability that the system is found in
 497 state \mathbf{x} , given measurements \mathbf{y} with error covariances \mathbf{R} .⁷

498 Equations (B1)–(B3) can be summarised in the form

$$499 \quad P(\mathbf{x}) = \frac{1}{2\pi(|\mathbf{B}| \cdot |\mathbf{R}|)^{1/2}} \exp(-J(\mathbf{x})) \quad (\text{B4})$$

$$500 \quad J(\mathbf{x}) = \frac{1}{2} \left[(\mathbf{x} - \mathbf{x}_b)^T \cdot \mathbf{B}^{-1} \cdot (\mathbf{x} - \mathbf{x}_b) + (\hat{H}(\mathbf{x}) - \mathbf{y})^T \cdot \mathbf{R}^{-1} \cdot (\hat{H}(\mathbf{x}) - \mathbf{y}) \right], \quad (\text{B5})$$

501

502 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
 503 simultaneously deviate from the background estimate and the measurements within the permitted error bounds. The deviations
 504 are weighted with the inverse error covariance matrices. For instance, this means that for measurements with a small error
 505 variance, a deviation $\hat{H}(\mathbf{x}) - \mathbf{y}$ becomes “more costly”.

⁵The 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.

⁶We 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 contains also 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 $\boldsymbol{\epsilon}_o$.

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

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

510 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
 511 operator,

$$512 \hat{H}(\mathbf{x}) = \hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x}, \quad (\text{B6})$$

513 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 non-linear, and if the components of $\delta\mathbf{x}$
 514 are sufficiently small, then we can substitute this first-order approximation into Eq. (B5), which yields

$$515 J = J_b + J_o \quad (\text{B7})$$

$$516 J_b(\delta\mathbf{x}) = \frac{1}{2} \delta\mathbf{x}^T \cdot \mathbf{B}^{-1} \cdot \delta\mathbf{x} \quad (\text{B8})$$

$$517 J_o(\delta\mathbf{x}) = \frac{1}{2} \left(\hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x} - \mathbf{y} \right)^T \cdot \mathbf{R}^{-1} \cdot \left(\hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta\mathbf{x} - \mathbf{y} \right) \quad (\text{B9})$$

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

520 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 -
 521 dimensional model space); we input the observations \mathbf{y} into the algorithm, and as output we obtain a result in model space that
 522 is consistent with the measurements (within the given error bounds).⁸ What if the measurements contain insufficient informa-
 523 tion about the state \mathbf{x} ? The algorithm will still provide an answer to the inverse problem, but the missing information will be
 524 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
 525 respective error covariance matrices. Thus data assimilation is a statistical approach, which can be expected to give good results
 526 *on average*, but not in every single time-step of the model run. This can become highly problematic if we only have very few
 527 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
 528 by the assimilation algorithm in such a severely under-constrained case, then the algorithm may just assimilate noise from
 529 the measurements rather than signal, resulting in unreasonable solutions to the inverse problem (e.g. Kahnert, 2009). To avoid
 530 such problems, one needs to systematically analyse the information content of the observations and constrain the assimilation
 531 algorithm to only operate on the signal degrees of freedom.

⁸By 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.

532 Appendix C: Information content of measurements

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

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

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

$$542 \mathbf{y} = \hat{H}(\mathbf{x}_b) + \mathbf{H} \cdot \delta \mathbf{x} + \epsilon_o. \quad (\text{C1})$$

543 The error covariance matrices are given by the expectation values $\mathbf{B} = \langle \delta \mathbf{x} \cdot \delta \mathbf{x}^T \rangle$, and $\mathbf{R} = \langle \epsilon_o \cdot \epsilon_o^T \rangle$, where the dot denotes a
 544 dyadic product.⁹ From Eq. (C1) we see that the covariance matrix of $\delta \mathbf{y} = \mathbf{y} - \hat{H}(\mathbf{x}_b)$ is given by $\langle \delta \mathbf{y} \cdot \delta \mathbf{y}^T \rangle = \mathbf{H} \cdot \mathbf{B} \cdot \mathbf{H}^T + \mathbf{R}$,
 545 where we assumed that background and observation errors are uncorrelated. This last equation suggests that we can compare
 546 model and observation errors in the same space by transforming the background error covariance matrix from the space of
 547 $(n \times n)$ matrices to the space of $(m \times m)$ matrices viz. $\mathbf{H} \cdot \mathbf{B} \cdot \mathbf{H}^T$.

548 To address the first problem, we diagonalise the covariance matrices by making the following change of variables

$$549 \delta \tilde{\mathbf{x}} = \mathbf{B}^{-1/2} \cdot \delta \mathbf{x} \quad (\text{C2})$$

$$550 \delta \tilde{\mathbf{y}} = \mathbf{R}^{-1/2} \cdot (\mathbf{y} - \hat{H}(\mathbf{x}_b)) \quad (\text{C3})$$

$$551 \tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2}. \quad (\text{C4})$$

552 Here $\mathbf{B}^{1/2}$ denotes the positive square root¹⁰ of the matrix \mathbf{B} , and $\mathbf{B}^{-1/2}$ denotes its inverse. The scaled observation operator
 553 $\tilde{\mathbf{H}}$ is sometimes referred to as the observability matrix. In the new basis, the cost function in (B7)–(B9) becomes

$$554 J = \frac{1}{2} \delta \tilde{\mathbf{x}}^T \cdot \delta \tilde{\mathbf{x}} + \frac{1}{2} \left(\tilde{\mathbf{H}} \cdot \delta \tilde{\mathbf{x}} - \delta \tilde{\mathbf{y}} \right)^T \cdot \left(\tilde{\mathbf{H}} \cdot \delta \tilde{\mathbf{x}} - \delta \tilde{\mathbf{y}} \right). \quad (\text{C5})$$

555 The covariance matrices are now unit matrices. This can also be seen by considering the transformed errors, e.g. $\tilde{\epsilon}_o = \mathbf{R}^{-1/2} \cdot \epsilon_o$
 556 and computing $\langle \tilde{\epsilon}_o \cdot \tilde{\epsilon}_o^T \rangle = \mathbf{R}^{-1/2} \cdot \langle \epsilon_o \cdot \epsilon_o^T \rangle \cdot \mathbf{R}^{-1/2} = \mathbf{1}_{m \times m}$, since $\langle \epsilon_o \cdot \epsilon_o^T \rangle = \mathbf{R}$. (Here, $\mathbf{1}_{m \times m}$ denotes the unit matrix in m -
 557 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
 558 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
 559 space, while the second term (the unit matrix) is the diagonalised observation error covariance matrix.

⁹The 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 $\langle a \rangle = \sum_{i=1}^n p_i a_i$.

¹⁰A 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.

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

$$563 \quad \tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \cdot \mathbf{H} \cdot \mathbf{B}^{1/2} = \mathbf{V}_L \cdot \mathbf{W} \cdot \mathbf{V}_R^T. \quad (\text{C6})$$

564 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-
 565 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 block of \mathbf{W} is a
 566 $(m \times m)$ -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
 567 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
 568 diagonal, while the lower block is a $((m - n) \times n)$ -null matrix.

569 We now make another change of variables:

$$570 \quad \delta \mathbf{x}' = \mathbf{V}_R^T \cdot \delta \tilde{\mathbf{x}} \quad (\text{C7})$$

$$571 \quad \delta \mathbf{y}' = \mathbf{V}_L^T \cdot \delta \tilde{\mathbf{y}} \quad (\text{C8})$$

$$572 \quad \mathbf{H}' = \mathbf{V}_L^T \cdot \tilde{\mathbf{H}} \cdot \mathbf{V}_R. \quad (\text{C9})$$

573 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 (C7)–(C9) into
 574 (C5) yields

$$575 \quad J = \frac{1}{2} \delta \mathbf{x}'^T \cdot \delta \mathbf{x}' + \frac{1}{2} (\mathbf{H}' \cdot \delta \mathbf{x}' - \delta \mathbf{y}')^T \cdot (\mathbf{H}' \cdot \delta \mathbf{x}' - \delta \mathbf{y}'). \quad (\text{C10})$$

576 Evidently, the transformation given in (C7)–(C9) preserves the diagonality of the background and observation error covariance
 577 matrices. What about the covariance matrix $\langle \delta \mathbf{y}' \cdot \delta \mathbf{y}'^T \rangle$ in the new basis? Using $\epsilon'_o = \mathbf{V}_L^T \cdot \tilde{\epsilon}_o = \mathbf{V}_L^T \cdot \mathbf{R}^{-1/2} \cdot \epsilon_o$, as well as
 578 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
 579 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),
 580 which yields

$$581 \quad \mathbf{H}' \cdot \mathbf{H}'^T = \mathbf{W} \cdot \mathbf{W}^T, \quad (\text{C11})$$

582 which is a $(m \times m)$ diagonal matrix. Thus in this coordinate system we can readily compare the diagonal elements of the
 583 transformed background error covariance matrix $\mathbf{H}' \cdot \mathbf{H}'^T$ to the diagonal (unit) elements of the observation error covariance
 584 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
 585 $\delta x'_i$ that can be controlled by the measurements. Those singular values smaller than unity correspond to model variables that
 586 are only related to noise.

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

591 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
 592 be expressed in terms of the singular values w_i of the observability matrix:

$$593 \quad N_s = \sum_{i=1}^{\min\{m,n\}} w_i^2 / (1 + w_i^2), \quad (\text{C12})$$

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

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

$$597 \quad S(P) = - \int P(x) \log_2 \left(\frac{P(x)}{P_0(x)} \right) dx, \quad (\text{C13})$$

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

$$601 \quad H = S(P_i) - S(P_f). \quad (\text{C14})$$

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

$$603 \quad H = \frac{1}{2} \sum_{i=1}^{\min\{m,n\}} \log_2(1 + w_i^2). \quad (\text{C15})$$

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

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

$$608 \quad \delta \mathbf{x}' = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \mathbf{x}. \quad (\text{C16})$$

609 Then we want to formulate the minimisation of the cost function in such a way that only those components of $\delta \mathbf{x}'$ are adjusted
 610 by the assimilation algorithm that correspond to the largest singular values of the matrix \mathbf{W} in (C6). All other elements of
 611 $\delta \mathbf{x}'$ should be left alone. In other words, we want to constrain the minimisation of the cost function to the subspace of the
 612 signal degrees of freedom of the state vector. Thus, in order to implement this idea, we first need to discuss how to incorporate
 613 constraints into the theory.

614 **Appendix D: Minimisation of the cost function with constraints**

615 In the minimisation of the cost function all elements of the control vector $\delta \mathbf{x}$ are independently adjusted until the minimum
 616 of J is found. This may not be a prudent approach if the information contained in the observations is insufficient to constrain
 617 all model variables. In such case one should introduce constraints that reduce the number of independent control variables.

618 However, this needs to be done in a clever way; the goal is to neither under-use the measurements (thus wasting available
619 information), nor to over-use them (thus assimilating noise).

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

622 D1 Minimisation of the cost function with strong constraints

623 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
624 constraints g_i is the method of Lagrange multipliers. More specifically, one introduces k Lagrange multipliers $\lambda_1, \dots, \lambda_k$ and
625 defines the function

$$626 \quad L(\delta x_1, \dots, \delta x_n, \lambda_1, \dots, \lambda_k) = J(\delta x_1, \dots, \delta x_n) + \sum_{i=1}^k \lambda_i g_i(\delta x_1, \dots, \delta x_n); \quad (D1)$$

627 then one solves the minimisation problem

$$628 \quad \nabla L(\delta x_1, \dots, \delta x_n, \lambda_1, \dots, \lambda_k) = \mathbf{0}_{n+k}, \quad (D2)$$

629 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
630 $(n+k)$ -dimensional space. Note that in this general formulation of the problem the constraints can even be nonlinear. We are
631 specifically interested in linear constraints, which can be expressed in the form $\mathbf{G} \cdot \delta\mathbf{x} = \mathbf{0}_k$. Then the constrained minimisation
632 problem becomes

$$633 \quad L(\delta\mathbf{x}, \boldsymbol{\lambda}) = J(\delta\mathbf{x}) + \boldsymbol{\lambda}^T \cdot \mathbf{G} \cdot \delta\mathbf{x} \quad (D3)$$

$$634 \quad \nabla_{\delta\mathbf{x}, \boldsymbol{\lambda}} L(\delta\mathbf{x}, \boldsymbol{\lambda}) = \begin{pmatrix} \nabla_{\delta\mathbf{x}} J(\delta\mathbf{x}) + \boldsymbol{\lambda}^T \cdot \mathbf{G} \\ \mathbf{G} \cdot \delta\mathbf{x} \end{pmatrix} = \mathbf{0}_{n+k}. \quad (D4)$$

635 Compared to the unconstrained minimisation problem, the introduction of k constraints has increased the dimension of the
636 problem from n to $n+k$. Naively, one may have expected that the dimension would, on the contrary, be reduced to $n-k$. This
637 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,
638 let us first write those equations more concisely in the form

$$639 \quad J = \frac{1}{2} (\delta\mathbf{x}^T \cdot \mathbf{Q}_1 \cdot \delta\mathbf{x} + \mathbf{Q}_2^T \cdot \delta\mathbf{x} + \delta\mathbf{x}^T \cdot \mathbf{Q}_2 + \mathbf{Q}_3) \quad (D5)$$

$$640 \quad \mathbf{Q}_1 = \mathbf{B}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \quad (D6)$$

$$641 \quad \mathbf{Q}_2 = \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot (\hat{H}(\mathbf{x}_b) - \mathbf{y}) \quad (D7)$$

$$642 \quad \mathbf{Q}_3 = (\hat{H}(\mathbf{x}_b) - \mathbf{y})^T \cdot \mathbf{R}^{-1} \cdot (\hat{H}(\mathbf{x}_b) - \mathbf{y}). \quad (D8)$$

643 Note that the covariance matrices and their inverses are symmetric (i.e. $\mathbf{R}^T = \mathbf{R}$, etc.) The unconstrained minimisation problem
644 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 minimise the cost function subject to the the
645 linear constraints

$$646 \quad \mathbf{G} \cdot \delta\mathbf{x} = \mathbf{0}_k, \quad (D9)$$

647 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 kernel¹¹ of \mathbf{G} by $\ker(\mathbf{G})$. Let
648 further $\mathbf{z}_1, \dots, \mathbf{z}_{n-k}$ denote a basis of $\ker(\mathbf{G})$. We define the $(n \times (n - k))$ -matrix

$$649 \quad \mathbf{Z} = \begin{pmatrix} \mathbf{z}_1 & \cdots & \mathbf{z}_{n-k} \end{pmatrix} \quad (\text{D10})$$

650 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
651 $((k \times (n - k))$ -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
652 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 minimisation
653 problem by substitution of $\delta \mathbf{x} = \mathbf{Z} \cdot \boldsymbol{\xi}$ into Eq. (D5), which yields

$$654 \quad J = \frac{1}{2} \left(\boldsymbol{\xi}^T \cdot \mathbf{Z}^T \cdot \mathbf{Q}_1 \cdot \mathbf{Z} \cdot \boldsymbol{\xi} + \mathbf{Q}_2^T \cdot \mathbf{Z} \cdot \boldsymbol{\xi} + \boldsymbol{\xi}^T \cdot \mathbf{Z}^T \cdot \mathbf{Q}_2 + \mathbf{Q}_3 \right) \quad (\text{D11})$$

$$655 \quad \mathbf{0}_k = \nabla_{\boldsymbol{\xi}} J = \mathbf{Z}^T \cdot \mathbf{Q}_1 \cdot \mathbf{Z} \cdot \boldsymbol{\xi} + \mathbf{Z}^T \cdot \mathbf{Q}_2. \quad (\text{D12})$$

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

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

662 D2 Minimisation of the cost function with weak constraints

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

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

$$668 \quad J = J_b + J_o + J_G \quad (\text{D13})$$

$$669 \quad J_G = \frac{1}{2} \delta \mathbf{x}^T \cdot \mathbf{G}^T \cdot \mathbf{B}_G^{-1} \cdot \mathbf{G} \cdot \delta \mathbf{x}, \quad (\text{D14})$$

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

$$671 \quad \nabla_{\delta \mathbf{x}} J_G = \mathbf{G}^T \cdot \mathbf{B}_G^{-1} \cdot \mathbf{G} \cdot \delta \mathbf{x}. \quad (\text{D15})$$

672 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”
673 σ_i^G along the diagonal of \mathbf{B}_G allow us to fine-tune the influence of each constraint on the solution. If σ_i^G is small, then the i th

¹¹The *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 $\dim \ker(\mathbf{G}) = n - k$.

674 constraint is relatively strong, and vice versa. Typically, if the σ_i^G are made too large, then there is a risk that the minimisation
675 algorithm ignores the constraints all together. In that case the solution will be very similar to the unconstrained solution. On the
676 other hand, if the σ_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
677 minimisation routine largely ignores the observations and returns a solution that lies quite close to the background estimate.

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

679 We now want to incorporate the results of Section C into the variational data assimilation method. More specifically, we want
680 to formulate weak constraints, Eq. (D14), based on the singular values of the observation operator in Eq. (C6). To this end,
681 we make the change of variables given in Eq. (C16). We assume, without loss of generality, that the first ℓ singular values are
682 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
683 3DVAR algorithm, while the remaining components remain unchanged, at least approximately, within specified error bounds.
684 If we were to formulate this requirement as a strong constraint, as in Eq. (D9), then it would take the form

$$685 \quad \delta \mathbf{x}' = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \mathbf{x} = \begin{pmatrix} \delta x'_1 \\ \vdots \\ \delta x'_\ell \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (\text{D16})$$

686 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.

687 The weak constraint approach is, arguably, more suitable in our case. We have, in the preceding text, frequently used the
688 terms *signal degrees of freedom* and *noise degrees of freedom*. Although it was conceptually useful to make this distinction, it
689 is important to stress that there is no sharp boundary between the two. Rather, there is a smooth transition from singular values
690 $w_1 > w_2 > \dots > w_\ell \geq 1$ to singular values $1 > w_{\ell+1} > w_{\ell+2} > \dots > w_K$ ($K = \min\{n, m\}$). For this reason we choose to
691 formulate the constraints as weak constraints. This allows us to make a smooth transition from free to constrained control
692 variables, where the transition from one regime to the other can be controlled by the singular values.

693 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),
694 which yields

$$695 \quad J_G = \frac{1}{2} \delta \mathbf{x}^T \cdot \mathbf{B}^{-1/2} \cdot \mathbf{V}_R \cdot \mathbf{B}_G^{-1} \cdot \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \mathbf{x}, \quad (\text{D17})$$

696 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
697 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
698 the matrix \mathbf{B}_G would be

$$699 \quad \mathbf{B}_G = \sigma_G \text{diag}(w_1, w_2, \dots, w_\ell, \dots, w_k, c, \dots, c), \quad (\text{D18})$$

700 where σ_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
 701 smaller than the smallest singular value w_k .

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

$$703 \mathbf{B}_G = \sigma_G \text{diag}(w_1^p, w_2^p, \dots, w_\ell^p, \dots, w_k^p, c, \dots, c), \quad (\text{D19})$$

704 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
 705 from unconstrained to constrained control variables takes place. Yet another choice would be

$$706 \mathbf{B}_G = \sigma_G \cdot \text{diag}(\mu_1, \mu_2, \dots, \mu_\ell, \dots, \mu_k, c, \dots, c), \quad (\text{D20})$$

$$707 \mu_i = w_i^2 / (1 + w_i^2), \quad (\text{D21})$$

708 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
 709 contribution to the number of signal degrees of freedom. We tested all three approaches (the one in Eq. (D19) for $p = 2$). These
 710 tests showed that the different approaches often yield analysis results that are quite similar. However, in each approach the
 711 free parameters σ_G and c are tuned to different values. If they are not well tuned, then the analysis tends either toward the
 712 background estimate or toward the unconstrained analysis, as explained earlier in the text following Eq. (D15).

713 Appendix E: Practical aspects of the implementation

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

715 One of the main practical problems is the dimension n of the model space. The grid-size is typically on the order $N_x \times$
 716 $N_y \times N_z \sim 100 \times 100 \times 10$, and the number of aerosol components is on the order of $N_c \sim 10-100$. Hence the dimension of the
 717 model space is $n \sim 10^6-10^7$. In our case, the matrix $\tilde{\mathbf{H}}$ in (C6) is a $(m \times n)$ matrix. To numerically perform a singular value
 718 decomposition of such a large matrix would be a formidable task.

719 In variational data assimilation we encounter a similar problem in the inversion of the matrix \mathbf{B} . In our 3DVAR code this
 720 problem is alleviated by using a so-called spectral formulation. The idea is to make a Fourier-transformation in the horizontal
 721 coordinates and to assume that all horizontal error correlations are homogeneous and isotropic. Under these assumptions
 722 one obtains one background error covariance matrix for each horizontal wavenumber; each of these matrices has dimension
 723 $N_z \times N_c \sim 10^3-10^4$. This can further be reduced to about 10^2 by making a reduced eigenvalue diagonalisation. The details are
 724 explained in Kahnert (2008).

725 In our case we are primarily interested in constraining the aerosol components. Therefore, we formulate our weak con-
 726 straints in a suitable subspace of the physical space. Suppose, for simplicity, that we have reduced all data to the verti-
 727 cal resolution of our model. Let $\nu_l = 1, \dots, m_l$ label all measurements that lie within model layer l . Suppose further than
 728 (i_α, j_α) is the horizontal grid point belonging to observation ν_l (so that the index α depends on the layer l and the observa-
 729 tion ν_l). Consider the reduced background error covariance matrix with elements $B_{k,k'}^{(\alpha,l)} = B_{i_\alpha, j_\alpha l k, i_\alpha, j_\alpha l k'}$, $k, k' = 1, \dots, N_c$,
 730 and N_c is the number of aerosol components. Consider further the reduced observability matrix with elements $\tilde{H}_{\nu_l, k}^{(l)} =$

731 $\sum_{k'=1}^{N_c} R_{\nu_l, \nu_l}^{-1/2} H_{m, i_\alpha, j_\alpha l k'} \{(B^{(\alpha, l)})^{1/2}\}_{k', k}$, where $m = m(l, \nu_l)$ labels the ν_l th observation in model layer l . Analogous to
 732 Eq. (C6), we now perform a singular value decomposition in the reduced space

$$733 \quad \tilde{H}_{\nu_l, k}^{(l)} = \sum_{s=1}^{\min\{m_l, N_c\}} (V_L^{(l)})_{\nu_l, s} w_s^{(l)} (V_R^{(l)})_{k, s}. \quad (\text{E1})$$

734 The dimension of this SVD-problem is now considerably reduced. The number of singular values is equal to $K = \min\{N_c, m_l\}$.

735 The constraint matrix $\mathbf{G} = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2}$ reduces to

$$736 \quad G_{s, k} = \sum_{k'=1}^{N_c} (V_R^{(l)})_{k', s} \{(B^{(\alpha, l)})^{-1/2}\}_{k', k}. \quad (\text{E2})$$

737 We now invoke the assumption that the constraints computed at the observation site are also valid at neighbouring points, i.e.,

738 we apply the constraint matrix given in Eq. (E2) in Eq. (D17) according to

$$739 \quad J_G = \frac{1}{2} \sum_{ijklk's} \delta x_{ijklk'} G_{s, k'} (B_G^{-1})_s G_{s, k} \delta x_{ijklk}, \quad (\text{E3})$$

740 where $(B_G)_s$ denotes the diagonal elements of the matrix given in (D18).¹²

741 Another aspect concerns the positive square root of the background error covariance matrix, which appears in essential
 742 parts of the theory, namely, in Eqs. (C6) and (D16). In theoretical developments it is, arguably, didactically expedient to work
 743 with the matrix $\mathbf{B}^{1/2}$. But in practice there are numerically more efficient formulations. One such approach is discussed in
 744 Kahnert (2008) in the context of a spectral formulation of the variational method. The spectral formulation is applied to the full
 745 B-matrix in order to reduce the dimension of the problem of diagonalising this matrix. This method is our method of choice
 746 in the formulation of the background and observation terms in the cost function given in Eqs. (B8) and (B9), respectively.
 747 However, in the formulation of the constraint term given in Eq. (D17) we can substantially reduce the dimension of the matrix
 748 \mathbf{B} by working in the reduced space in which only the covariances $\mathbf{B}^{(\alpha, l)}$ among aerosol components are considered. One could
 749 compute the matrix $(\mathbf{B}^{(\alpha, l)})^{-1/2}$ in Eq. (D17) by diagonalising the matrix $\mathbf{B}^{(\alpha, l)}$. However, a numerically much more efficient
 750 approach is to not work with positive square root, but with the so-called Cholesky decomposition¹³ of the B-matrix,

$$751 \quad \mathbf{B}^{(\alpha, l)} = \mathbf{C}_u^T \cdot \mathbf{C}_u, \quad (\text{E4})$$

752 where \mathbf{C}_u is an upper triangular matrix. Thus the actual algorithm we used for formulating the constrained minimisation of the
 753 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,
 754 similarly, by replacing the inverse matrix $\mathbf{B}^{-1/2}$ by the inverse of the Cholesky factor, \mathbf{C}_u^{-T}).

¹²For 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 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 ∇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.

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

755 *Author contributions.* MK worked with the theoretical developments and and numerical implementation, EA performed the testing of the
756 method.

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

758 References

- 759 Andersson, C., Langner, J., and Bergström, R.: Interannual variation and trends in air pollution over Europe due to climate variability during
760 1958-2001 simulated with a regional CTM coupled to the ERA40 reanalysis, *Tellus*, 59B, 77–98, 2007.
- 761 Andersson, C., Bergström, R., Bennet, C., Robertson, L., Thomas, M., Korhonen, H., Lehtinen, K. H. J., and Kokkola, H.: MATCH-SALSA
762 — Multi-scale Atmospheric Transport and CHemistry model coupled to the SALSA aerosol microphysics model — Part 1: Model de-
763 scription and evaluation, *Geosci. Model Dev.*, 8, 171–189, 2015.
- 764 Andersson, E. and Kahnert, M.: Coupling aerosol optics to the MATCH (v5. 5.0) chemical transport model and the SALSA (v1) aerosol
765 microphysics module, *Geosci. Model Dev.*, 9, 1803–1826, 2016.
- 766 Benedetti, A., Morcrette, M. J.-J., Boucher, O., Dethof, A., Engelen, R. J., Huneeus, M. F. H. F. N., Jones, L., and S. Kinne, J. W. K.,
767 Mangold, A., Razinger, M., Simmons, A. J., and Suttie, M.: Aerosol analysis and forecast in the European Centre for Medium-Range
768 Weather Forecasts Integrated Forecast System: 2. Data assimilation, *J. Geophys. Res.*, 114, D13 205, 2009.
- 769 Bi, L., Yang, P., Kattawar, G., and Kahn, R.: Modeling optical properties of mineral aerosol particles by using nonsymmetric hexahedra,
770 *Appl. Opt.*, 49, 334–342, 2010.
- 771 Bocquet, M.: Toward optimal choices of control space representation for geophysical data assimilation, *Mon. Wea. Rev.*, 137, 2331–2348,
772 2009.
- 773 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,
774 J. E., Fenn, M. A., and Rogers, R. R.: Observations of the spectral dependence of particle depolarization ratio of aerosols using NASA
775 Langley airborne High Spectral Resolution Lidar., *Atmos. Chem. Phys.*, 15, 13 453–13 473, 2015.
- 776 Burton, S. P., Chemyakin, E., Liu, X., Knobelspiesse, K., Stamnes, S., Sawamura, P., Moore, R. H., Hostetler, C. A., and Ferrare, R. A.: In-
777 formation content and sensitivity of the $3\beta+2\alpha$ lidar measurement system for aerosol microphysical retrievals, *Atmos. Meas. Techniques*,
778 9, 5555–5574, 2016.
- 779 Cardinali, C., Pezzulli, S., and Andersson, E.: Influence-matrix diagnostic of a data assimilation system, *Q. J. R. Meteorol. Soc.*, 130, 2767–
780 2786, 2004.
- 781 Foltescu, V., Pryor, S. C., and Bennet, C.: Sea salt generation, dispersion and removal on the regional scale, *Atmos. Environ.*, 39, 2123–2133,
782 2005.
- 783 Fuller, K. A. and Mackowski, D. W.: Electromagnetic scattering by compounded spherical particles, in: *Light scattering by nonspherical*
784 *particles*, edited by Mishchenko, M. I., Hovenier, J. W., and Travis, L. D., pp. 226–273, Academic Press, San Diego, 2000.
- 785 Johnson, C., Hoskins, B. J., and Nichols, N. K.: Very large inverse problems in atmosphere and ocean modelling, *Int. J. Numer. Meth. Fluids*,
786 47, 759–771, 2005a.
- 787 Johnson, C., Nichols, N. K., and Hoskins, B. J.: A singular vector perspective of 4D-Var: Filtering and interpolation, *Q. J. R. Meteorol. Soc.*,
788 131, 1–19, 2005b.
- 789 Joiner, J. and da Silva, A. M.: Efficient methods to assimilate remotely sensed data based on information content, *Q. J. R. Meteorol. Soc.*,
790 124, 1669–1694, 1998.
- 791 Kahnert, F. M.: Reproducing the optical properties of fine desert dust aerosols using ensembles of simple model particles, *J. Quant. Spectrosc.*
792 *Radiat. Transfer*, 85, 231–249, 2004.
- 793 Kahnert, M.: Variational data analysis of aerosol species in a regional CTM: background error covariance constraint and aerosol optical
794 observation operators, *Tellus*, 60B, 753–770, 2008.

795 Kahnert, M.: On the observability of chemical and physical aerosol properties by optical observations: Inverse modelling with variational
796 data assimilation, *Tellus*, 61B, 747–755, 2009.

797 Kahnert, M.: Modelling radiometric properties of inhomogeneous mineral dust particles: Applicability and limitations of effective medium
798 theories, *J. Quant. Spectrosc. Radiat. Transfer*, 152, 16–27, 2015.

799 Kahnert, M. and Devasthale, A.: Black carbon fractal morphology and short-wave radiative impact: a modelling study, *Atmos. Chem. Phys.*,
800 11, 11 745–11 759, 2011.

801 Kahnert, M., Nousiainen, T., Lindqvist, H., and Ebert, M.: Optical properties of light absorbing carbon aggregates mixed with sulfate:
802 assessment of different model geometries for climate forcing calculations, *Opt. Express*, 20, 10 042–10 058, 2012a.

803 Kahnert, M., Nousiainen, T., Thomas, M. A., and Tyynelä, J.: Light scattering by particles with small-scale surface roughness: comparison
804 of four classes of model geometries, *J. Quant. Spectrosc. Radiat. Transfer*, 113, 2356–2367, 2012b.

805 Kahnert, M., Nousiainen, T., and Lindqvist, H.: Models for integrated and differential scattering optical properties of encapsulated light
806 absorbing carbon aggregates, *Opt. Express*, 21, 7974–7992, 2013.

807 Kahnert, M., Nousiainen, T., and Lindqvist, H.: Review: Model particles in atmospheric optics, *J. Quant. Spectrosc. Radiat. Transfer*, 146,
808 41–58, 2014.

809 Kahnert, M., Nousiainen, T., and Markkanen, J.: Morphological models for inhomogeneous particles: light scattering by aerosols, cometary
810 dust, and living cells, in: *Light Scattering Reviews 11*, edited by Kokhanovsky, A., Springer, Berlin, p. 299–339, 2016.

811 Khade, V. M., Hansen, J. A., Reid, J. S., and Westphal, D. L.: Ensemble filter based estimation of spatially distributed parameters in a
812 mesoscale dust model: experiments with simulated and real data, *Atmos. Chem. and Phys.*, 13, 3481–3500, 2013.

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

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

816 Kylling, A., Kahnert, M., Lindqvist, H., and Nousiainen, T.: Volcanic ash infrared signature: porous non-spherical ash particle shapes com-
817 pared to homogeneous spherical ash particles, *Atmos. Meas. Tech.*, 7, 919–929, 2014.

818 Lindqvist, H., Muinonen, K., and Nousiainen, T.: Light scattering by coated Gaussian and aggregate particles, *J. Quant. Spectrosc. Radiat.*
819 *Transfer*, 110, 1398–1410, doi:10.1016/j.jqsrt.2009.01.015, 2009.

820 Lindqvist, H., Nousiainen, T., Zubko, E., and Muñoz, O.: Optical modeling of vesicular volcanic ash particles, *J. Quant. Spectrosc. Radiat.*
821 *Transfer*, 112, 1871–1880, 2011.

822 Lindqvist, H., Jokinen, O., Kandler, K., Scheuven, D., and Nousiainen, T.: Single scattering by realistic, inhomogeneous mineral dust
823 particles with stereogrammetric shapes, *Atmos. Chem. Phys.*, 14, 143–157, doi:10.5194/acp-14-143-2014, 2014.

824 Liu, L. and Mishchenko, M. I.: Scattering and radiative properties of complex soot and soot-containing aggregate particles, *J. Quant. Spec-*
825 *trosc. Radiat. Transfer*, 106, 262–273, 2007.

826 Liu, Z., Liu, Q., Lin, H.-C., Schwartz, C. S., Lee, Y.-H., and Wang, T.: Three-dimensional variational assimilation of MODIS aerosol optical
827 depth: Implementation and application to a dust storm over East Asia, *J. Geophys. Res.*, 116, D23 206, 2011.

828 McKeen, S., Chung, S. H., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Gong, W., Bouchet, V., Moffet, R., Tang, Y., Carmichael, G. R.,
829 Mathur, R., and Yu, S.: Evaluation of several PM_{2.5} forecast models using data collected during the ICARTT/NEAQS 2004 field study, *J.*
830 *of Geophys. Res.*, 112, d10S20, 2007.

831 Mishchenko, M. I., Travis, L. D., Kahn, R. A., and West, R. A.: Modeling phase functions for dustlike tropospheric aerosols using a shape
832 mixture of randomly oriented polydisperse spheroids, *J. Geophys. Res.*, 102, 16,831–16,847, 1997.

833 Mishchenko, M. I., Dlugach, Z. M., and Zakharova, N. T.: Direct demonstration of the concept of unrestricted effective-medium approxima-
834 tion, *Opt. Lett.*, 39, 3935–3938, 2014.

835 Muinonen, K.: Light scattering by stochastically shaped particles, in: *Light scattering by nonspherical particles*, edited by Mishchenko, M. I.,
836 Hovenier, J. W., and Travis, L. D., pp. 323–354, Academic Press, San Diego, 2000.

837 Müller, D., Wandinger, U., and Ansmann, A.: Microphysical particle parameters from extinction and backscatter lidar data by inversion with
838 regularization: theory, *Appl. Opt.*, 38, 2346–2357, 1999.

839 Omar, A. H., Winker, D. M., Vaughan, M. A., Hu, Y., Trepte, C. R., Ferrare, R. A., Lee, K.-P., Hostetler, C. A., Kittaka, C., Rogers, R. R.,
840 Kuehn, R. E., and Liu, Z.: The CALIPSO automated aerosol classification and lidar ratio selection algorithm, *J. Atmos. and Ocean.*
841 *Technol.*, 26, 1994–2014, 2009.

842 Parrish, D. F. and Derber, J. C.: The National Meteorological Centre’s spectral statistical interpolation analysis system, *Mon. Wea. Rev.*, 120,
843 1747–1763, 1992.

844 Rabier, F., Fourrié, N., Chafaï, D., and Prunet, P.: Channel selection methods for infrared atmospheric sounding interferometer radiances, *Q.*
845 *J. R. Meteorol. Soc.*, 128, 1011–1027, 2002.

846 Rodgers, C. D.: *Inverse methods for atmospheric sounding*, World Scientific, Singapore, 2000.

847 Rubin, J. I. and Collins, W. D.: Global simulations of aerosol amount and size using MODIS observations assimilated with an Ensemble
848 Kalman Filter, *J. Geophys. Res.*, 119, 12,780–12,806, 2014.

849 Saide, P. E., Charmichael, G. R., Liu, Z., Schwartz, C. S., Lin, H. C., da Silva, A. M., and Hyer, E.: Aerosol optical depth assimilation
850 for a size-resolved sectional model: impacts of observationally constrained, multi-wavelength and fine mode retrievals on regional scale
851 analysis and forecasts, *Atmos. Chem. Phys.*, 13, 10 425–10 444, 2013.

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

854 Sekiyama, T. T., Tanaka, T. Y., Shimizu, A., and Miyoshi, T.: Data assimilation of CALIPSO aerosol observations, *Atmos. Chem. Phys.*, 10,
855 39–49, 2010.

856 Undén, P., Rontu, L., Järvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., Garcia-Moya, J. A., Jones, C., Lender-
857 link, G., McDonald, A., McGrath, R., Navascues, B., Nielsen, N. W., Ødegaard, V., Rodriguez, E., Rummukainen, M., Rööm, R., Sattler,
858 K., Sass, B. H., Savijärvi, H., Schreur, B. W., Sigg, R., The, H., and Tijm, A.: HIRLAM-5 Scientific Documentation, <http://www.hirlam.org>,
859 2002.

860 Veselovskii, I., Kolgotin, A., Griaznov, V., Müller, D., Wandinger, U., and Whiteman, D. N.: Inversion with regularization for the retrieval of
861 tropospheric aerosol pa rameters from multiwavelength lidar sounding, *Appl. Opt.*, 41, 3685–3699, 2002.

862 Veselovskii, I., Kolgotin, A., Griaznov, V., Müller, D., Franke, K., and Whiteman, D. N.: Inversion of multiwavelength Raman lidar data for
863 retrieval of bimodal aerosol size distribution, *Appl. Opt.*, 43, 1180–1195, 2004.

864 Veselovskii, I., Kolgotin, A., Müller, D., and Whiteman, D. N.: Information content of multiwavelength lidar data with respect to microphys-
865 ical particle properties derived from eigenvalue analysis, *Appl. Opt.*, 44, 5292–5303, 2005.

866 Vilaplana, R., Moreno, F., and Molina, A.: Study of the sensitivity of size-averaged scattering matrix elements of nonspherical particles
867 to changes in shape, porosity and refractive index, *J. Quant. Spectrosc. Radiat. Transfer*, 100, 415–428, doi:10.1016/j.jqsrt.2005.11.068,
868 2006.

869 Wang, Y., Sartelet, K. N., Bocquet, M., and Chazette, P.: Modelling and assimilation of lidar signals over Greater Paris during the MEGAPOLI
870 summer campaign, *Atmos. Chem. Phys.*, 14, 3511–3532, 2014.

- 871 Xu, Q.: Measureing information content from observations for data assimilation: relative entropy versus shannon entropy difference, *Tellus*,
872 59A, 198–209, 2006.
- 873 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
874 on a global aerosol particle transport model, *J. Geophys. Res.*, 119, 4674–4689, 2014.