Author response regarding the manuscript 'How much information do extinction and backscattering measurements contain about the chemical composition of atmospheric aerosol?'

M. Kahnert[1,2] and E. Andersson[1]

[1]Swedish Meteorological and Hydrological institute (SMHI), SE-60176 Norrköping, Sweden
[1]Chalmers University of Technology, Department of Earth and Space Sciences, SE-41296 Göteborg, Sweden

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

Dear Matthias,

we were very happy to receive comments from reviewers who seemed to have many different backgrounds, ranging from data assimilation to lidar instrumentation and remote sensing. It was our hope and intention to write this manuscript in such a way that it would be interesting to a broad readership, which has not been quite easy. But the discussion showed that it is not impossible, and we received many good suggestions to better accomodate the expectations of each of these communities. Although most of the reviewer comments were rather straight forward to answer and implement, the sum of them (5 reviews plus one extra comment) amounted to quite substantial changes in the structure and content of the manuscript. We changed all figures, added two new ones, and removed one of the old figures. This also resulted in some changes in the abstract and conclusion section. Below we answer the comments by the reviewers and describe our changes in the manuscript. A manuscript version with the tracked changes is appended at the end of this document. Since this looks a bit messy, we also submitted (in an extra file) a clean version of the revised manuscript without any markings of the changes we did.

Below the reviewer comments are marked in blue, our response is marked in black.

1 Reviewer 1

The ACPD paper by Kahnert and Andersson deals with the assimilation of lidar observations into a chemical transport model. They investigate how much information about the chemical composition can be extracted from backscatter and extinction measurements and how this information is best assimilated into a chemical transport model.

Overall the paper is very well written and should be published as it is an interesting and important contribution to aerosol research. I only have a few minor comments which the authors may consider for their final paper. I have to say that my experience lies more on the lidar and aerosol optics side than on the information theory / mathematical side, thus I was not really able to review all theory details described in the appendix.

We very much appreciate receiving comments from the lidar and aerosol optics community. The parts that deal with theoretical developments and chemical data assimilation have been very well covered by reviewers 2 and 4. We thank the reviewer for his supportive review and helpful comments!

Comments:

1. It may be beneficial to say a few words about the refractive index and the size bins of the individual species of MATCH. I suggest to add a table with the refractive index of these species at the lidar wavelengths.
   We added a new table to Sect. 2 providing the refractive indices, and we added an itemised list of the size bins and the corresponding size ranges.

2. Line 105-109: The description of the MATCH aerosol microphysics module could be shortened as it is not used in this paper.
   It is difficult to shorten these 5 lines. We could only remove them. Then again, we would like the reader to understand that there do exist more realistic optics models, but they are not so straight forward to test, owing to their nonlinearity. Thus the present study is meant as a first step in a larger project, which we will, hopefully, be able to follow up with an investigation of information content based on a more sophisticated description of aerosol optics.

3. Line 118: What about the emissions of the other species? Are they also from EMEP?
   Yes. But EMEP does not deliver gridded emission for black carbon and elemental carbon, only for total primary particulate matter. The sentence in question was meant to explain how we converted these into gridded emission data for black carbon and elemental carbon. We reformulated this to make it clear that the emissions of *all* aerosol species are taken from EMEP.

4. Line 134: "an" -> "and"
   Yes.

5. Line 147 "we constrain to better than observation error": It is not clear to me what this means.
   This formulation was also criticised by another reviewer. We reformulated this part as follows: “Suppose we have an n dimensional model space. Given m observations (e.g., m_1 different parameters at m_2 different wavelengths, so that $m_1 \cdot m_2 = m$), how many independent model variables $N \leq n$ can we constrain with the observations?”

6. Line 151: Remove "the".
   OK. However, we reformulated this entire section to accommodate the comments by reviewer 2.

7. Line 177: "To be specific" could be removed.
   Agreed.

8. Line 177: Do the results (N_s and H) presented in this section depend on the order of the parameters? If yes, are the changes significant?
   We do not quite understand this question, especially not what the reviewer means by “order”. Is the reviewer inquiring about the *ordering* and *grouping*, or about the *magnitude*? In the latter case, the answer is no, because N_s and H

are computed from the scaled Jacobian of the observation operator, which does not depend on the magnitude of the parameters. In the former case, the results do depend on which parameters are being measured, but, of course, not on the ordering.

9. Line 185: "around 7.4 for a single wavelength to around 10-12 for two wavelengths" would be more precise.
OK. This text has changed significantly in the revised version, owing to comments by reviewer 2, who asked us to consider different and technically more realistic combinations of observables in table 1.

10. Line 203: I was not aware about the difference between "observation error" and "measurement error". Is this generally accepted terminology? Maybe you can add a reference here so that the reader not familiar with this terminology can see that is used also elsewhere or was introduced by someone (maybe Rodgers?).
We added the formal definition for the observation error as $\epsilon_o = \epsilon_f + \epsilon_m$ and a reference to Rabier et al. (2002). They use the same terminology as we do, and they also denote the forward model error by ϵ_f. However, they use the symbol ϵ_o for the measurement error, which is potentially confusing. We find it less confusing to denote the measurement error by ϵ_m, and to reserve the symbol ϵ_o for the observation error. We also mentioned that there can be other contributions to the observation error, such as representativity error. These concepts are well understood both in the data assimilation and in the satellite remote sensing/retrieval community, but not necessarily among instrument developers, who tend to identify ϵ_o with ϵ_m, while forgetting about ϵ_f. This can be a serious mistake in cases where $\epsilon_f \gg \epsilon_m$, as is the case, e.g., in lidar depolarisation measurements. We find this point sufficiently significant to repeat it, in rephrased form, in appendix B.

11. Fig. 1: The difference between the middle and the right sub-plot is hardly visible. Perhaps you find a better way to visualize it.
We removed this figure. The regional model is merely used to generate a test case, but we do not address questions of regional modelling or horizontal information spreading in 3DVAR. Therefore this figure conveys no useful information for this study.

12. Line 229 (and at other places): You use β_{sca} and β_{bak} for the backscatter coefficient. Please use only a single symbol throughout the paper.
Yes, we corrected this and consistently use β_{sca}.

13. Line 241 "the secondary inorganic aerosol (SIA) species are almost completely restored by the 3DVAR": Is it understandable why exactly SIA is restored? Because of the refractive index? Or does it have something to do with the order (index number) of the species in the model?
This question has also been brought up by other reviewers. We added a new figure to Sect. 3 in which we show the linear coefficients in the transformation of the control variables in Eq. (C16) — see the new Fig. 5. Based on this extra figure we added a discussion of the question which aerosol components in model space make the dominant contribution to the signal-related variables in the transformed space. This facilitates the interpretation of the analysis results.

14. Line 274 "there appeared ...": This was not really shown in the paper, so you might remove this sentence or write it in a different way.

    OK, we removed this sentence.

15. Fig. 3: In this figure the difference between "observations" and "analysis" is much smaller than 10 % (the assumed "observation error"). As this is somewhat unexpected (but understandable as an optimization is applied) you may add a brief discussion about the effect a "measurement error" (noise) would have. Because of the assumed linearity this probably is not very difficult to explain.

    OK, we added the following text. “In fact, the difference between the observation-equivalent analysis and the observations deviate by even less than 10 %. However, our tests confirmed that an increase in the observation error eventually results in analysis results of which the observation-equivalent increasingly deviates from the observations (not shown).”

16. Fig. 4: Could it be of interest to see which aerosol species (size bins) the individual variables represent? What would be the effect of changing the order of the species?

    We added an extra figure (new Fig. 1) that shows a selection of aerosol species in specific size bins. In response to reviewer 2 and 4, we even show a comparison with an unconstrained analysis. This makes it clearer that the constrained analysis reduces the noisiness of the analysis, since it is being constrained to assimilating signal rather than measurement noise.

    We do not understand the last question about changing the “order of the species”.

17. Line 277: "to be sure" could be removed.

    Agreed, but we re-wrote the whole sentence.

18. Line 314: I think some aerosol species exist for which assuming externally mixed spheres is not that wrong.

    It is unclear what kind of species the reviewer refers to. Certainly not dust or black carbon (BC). Sea salt is either mixed with water, or else it is nonspherical. Organic carbon (OC) and secondary inorganic aerosols (SIA) are rarely found in pure form. They are often mixed with each other, with water, NaCl, and even BC and dust. Even nucleation-mode particles are often the result of at least binary nucleation involving more than one species. In our more realistic aerosol microphysics model there is not a single size bin in which liquid-phase (i.e., spherical) aerosols consist of a single compound. We therefore prefer to keep the text in its present form.

2 Reviewer 2

2.1 General comments

1. This paper details an interesting way to assess the information content in lidar measurements of aerosol backscatter and extinction with respect to model assimilation. It also demonstrates how this knowledge may be used to optimize the incorporation of lidar measurements in the model. This is a very interesting and relevant topic. Assimilation of lidar

data into models is a field that is still developing rapidly, with a few different groups using very different techniques; therefore, well designed research into how best to use lidar data is very valuable. It is also potentially informative to the lidar community, since work must begin soon to design the next satellite lidar instruments if the lidar record is to continue. The choice of which measurements and which wavelengths to include has a large bearing on cost and technological difficulty, so having quantitative information about which measurements are most useful for improving models is critical. To that end, I would like to suggest some additional cases for Table 1, please see the specific comments below.

We thank the reviewer for the considerably thorough and supportive review, which very much helped us to improve various aspects of the manuscript. Our detailed response to the review comments follows.

2. The paper is well written with very nice clarity. However, the overall organization is somewhat difficult. The current organization consists of a very streamlined and easy-to-read main text with five very technically dense appendices. While the main text is pleasantly easy to read on the first pass, there is too much information missing. While it's appropriate to include extra, more detailed information in appendices, the main text still needs to be able to stand on its own, and in my opinion, it doesn't quite. I would suggest that the main equations and brief explanations should also be included in the main text, including all the equations that a reader would need to apply to calculate the kinds of results presented in this work. The appendices also include a lot of pedagogical development; this is the kind of information that I think rightly belongs in the appendix for readers who want more details. Since the appendices are 5 different topics, I also suggest that each appendix should be exist as a separate entity, with all variables defined, so that a reader can read Appendix D to learn about the application of constraints or Appendix E for the "practical aspects" without a close reading of Appendix A,B, and C, to find the definitions of the variables.

   The organisation of the paper is indeed a delicate issue that was also brought up by other reviewers. Our main goal is to make this paper accessible to a broad community, including lidar instrument developers, remote sensing groups, and data assimilation researchers. For this reason, we prefer to include most of the theoretical developments in the appendix. However, we agree that this creates a significant problem by removing essential information from the main body of the paper. In the revised paper we followed the reviewer's suggestion and re-state the most essential theoretical results from the appendix in the main text. These changes were done mainly in Sect. 2. This makes the paper more readable and self-contained, while avoiding the risk of making it too technical, which could narrow down the readership of this work.

3. The results and conclusions are also a little too abbreviated. Some key aspects are missing, like how was the specific weighting chosen and how do we know this is the best weighting? Also, as pointed out by another reviewer, the assessment (section 3.2) is really more of a demonstration. That is, although the theoretical development is compelling, the application/assessment section isn't sufficient to convince readers that this is a better way to assimilate lidar data than another way. This paper clearly reflects a lot of research on the part of the authors and I think the missing information probably exists but was left out in the effort to streamline the manuscript. I think adding this additional information should be fairly straightforward and would improve the usefulness of this research for the modeling and lidar communi-

ties without adding too much complication to the nice flow of the paper.

This is also an important point, which was brought up by several reviewers. We performed additional computations using the unconstrained assimilation algorithm and compared the constrained to the unconstrained analysis. The hypothesis is that the constrained analysis is less noisy, because the unconstrained analysis is at risk of assimilating noise. The results of this comparison, which are shown in the new Figs. 1, 2, and 4, are consistent with the hypothesis. Also, we eliminated all instances of "numerical experiment" and replace it by a more appropriate term, e.g., "numerical test", "demonstration", or "illustration". Further, we added more explanations to Sect. 2.4 about the construction of the covariance matrix in the constraint term. Finally, we amended the conclusion section and the abstract to incorporate the results of the comparison of the unconstrained and the constrained analysis algorithm.

2.2 Specific comments

1. Lines 151-158: Here is an example where I think some important things are missing from the main text which only appear in the appendices. These eight short lines are the methodology section for the key calculations that are the novel part of your research and are critically important for a reader to understand. I suggest that a way to decide what should also be included here would be to target the subset of equations that a reader would need to apply to calculate results like yours, but without their derivations. Also include enough supporting explanation to describe what the equations say and how to use them.

   We agree, and we made changes following the more detailed suggestions given in the following comments.

2. L152-153: Specifically here, Eq C6 and C16 should be included in the text, since they are required to understand the meaning of the sentence. Later, at L159-160 where readers are directed to the appendix for more background information, I think that's fine.

   OK, we revised the text and included the equations (with explanations) for the observation operator, the observability matrix, and the singular-value decomposition thereof. The changes pertain to Sect. 2.

3. L155-157: The equations for signal degrees of freedom and Shannon information content should also be included in the text.

   OK, this has been added with accompanying text to Sect. 2.

4. L165: "a numerical experiment". In fact, it's more of a demonstration than an experiment. It's useful as a demonstration of the results of the technique, but there's nothing in the demonstration that addresses a hypotheses. Sharing more of the background work would make the paper more compelling. For example, as another reviewer suggested, comparing to a control experiment would be necessary for convincing readers that this technique is useful. For another example, a pair of runs with different weightings in the assimilation would help answer the question of why the weighting that was ultimately chosen was the best one.

   We replaced "numerical experiment" everywhere in the paper, as mentioned previously. Next, we showed a control run with the unconstrained assimilation system (new Figs. 1, 2, and 4). The hypothesis is that the constrained analysis

should be less noisy than the unconstrained analysis. We revised the Figures and show both the unconstrained and the constrained analysis. Also, we added an extra figure (new Fig. 1) to show both analysis results for different aerosol species in different size bins, as these are even more sensitive than size-integrated total mass mixing ratios. Finally, the case we picked in the original manuscript was not particularly challenging, since the background state was fairly close to the reference state. In the revised paper, we picked a more challenging case in order to make the differences between both analysis runs as clear as possible. As for the different weightings, our tests, so far, indicate that the different approaches result in rather similar analysis results. So, the constrained analysis is not as strongly dependent on the weighting as one may expect. We clarified this point in a discussion added at the end of Sect. 2.4.

5. L177: Depolarization is not included in the studied parameters, yet lidar studies have shown that depolarization measurements contain some information about aerosol composition (for example, Omar et al. 2009 as referenced in the introduction, but there are many others). Do the authors have any comment on depolarization and why it isn't included in this study?

   There are two major problems. The obvious practical problem is that the forward model would need to be based on nonspherical particles (as spherical particles do not depolarise). However, our simpler optics model is entirely based on spherical particles, while our newer optics model only accounts for the nonsphericity of bare black carbon, but not for that of mineral dust or dry sea salt. Thus our capabilities of modelling depolarisation are presently limited. The second problem is that the observation error for depolarisation may be very high, even though the measurement error is very low. This is because the forward-model error is likely to be quite high, since even slight variations in particle geometry (e.g. Kahnert et al. (2012)) or inhomogeneity (e.g. Kahnert (2015)) can result in large variations in the depolarisation ratio. If the forward-model error is, indeed, high, then the prospects of using depolarisation for constraining CTM model results are likely to be low. However, this question is open and will be investigated in future studies. But in order to do so, one would first need to obtain estimates of the forward-model error (e.g, by computing depolarisation ratios while varying particle morphology).

6. Table 1 and related discussion: From a lidar standpoint, some combinations of channels are more technologically affordable than others, so the discussion of which channels add significant information content is very interesting. However, the utility for the lidar community would be maximized if the combinations were ordered such that they roughly increase in technological difficulty. Also, some combinations don't really make sense from a technological standpoint. There is no lidar that measures extinction but not backscatter at the same channel (although modelers may use only the extinction). On the other hand, backscatter (actually attenuated backscatter) without a direct measurement of extinction is common. Also, since CALIPSO, CATS, EarthCARE and the $3\beta + 2\alpha$ combination of airborne HSRL2 are mentioned in the introduction and motivation sections, it would be useful if the combinations relevant to those instruments were included. CALIPSO = CATS =$\beta(\lambda 1) + \beta(\lambda 2)$. EarthCARE = $\beta(\lambda 3) + k(\lambda 3)$. HSRL2 =$\beta(\lambda 1) + \beta(\lambda 2) + \beta(\lambda 3) + k(\lambda 2) + k(\lambda 3)$. I would suggest these combinations of backscatter and extinction would be most interesting and useful to the lidar community: $\beta(\lambda 3)$

$\beta(\lambda 1) + \beta(\lambda 2)$

$\beta(\lambda 1) + \beta(\lambda 2) + \beta(\lambda 3)$

$\beta(\lambda 3) + k(\lambda 3)$

$\beta(\lambda 1) + \beta(\lambda 2) + k(\lambda 2)$

$\beta(\lambda 1) + \beta(\lambda 2) + \beta(\lambda 3) + k(\lambda 2) + k(\lambda 3)$

For these experiments, it appears that the observation error was always assumed to be the same in every channel. I think it's a reasonable assumption, to first approximation, that the measurement error would be similar in every channel, but as pointed out at L78-79, some lidar retrievals include additional non-random errors that can be much larger. This could and should affect the choice of channels to assimilate. For example, the Raman, HSRL, and transmittance techniques are fairly direct measures of extinction, but techniques that require an inferred lidar ratio to convert backscatter to extinction have relatively little additional measurement information content in the extinction.

We welcome the reviewer's suggestion to take technical realisations of lidar systems into account, and we revised Tables 1 and 2 (i.e., Tables 2 and 3 in the revised manuscript) according to the reviewer's specific suggestions. We also added a comment on the observation errors of lidar measurements, specifically on the fact that the observation errors may be different for different channels/parameters.

7. L 197-201. Here also the discussion of incorporating soft constraints and the specifics of the three weighting schemes should be in the main text of the paper and not just the appendix, since it is discussed here in the results section. This section is not understandable without the equations from the appendix and most of section D3.

   We removed this discussion here. Instead, we briefly discussed the construction of the constraint covariance matrix in Sect. 2.4.

8. L 203-204. Discussion of observation error vs. measurement error. This is interesting and useful, but could be clarified as to whether the forward model error (due to poor assumptions) is considered part of the observation error or is another separate source of error. If it is part of the observation error, how are the forward model errors represented and how are they transformed into the space of the measurement vector?

   We extended the text to clarify that the observation error is given by $\boldsymbol{\epsilon}_o = \boldsymbol{\epsilon}_m + \boldsymbol{\epsilon}_f$, where $\boldsymbol{\epsilon}_f$ denotes the forward-model error. We also added a citation to the paper by Rabier et al. (2002) with a hint to their Eq. (1), which explains this terminology. A way to determine the forward-model errors theoretically is to perform light-scattering calculations while varying various parameters, such as particle morphology, refractive index, and size distribution within typical uncertainty ranges. This can provide us with an estimate of $\boldsymbol{\epsilon}_f$. To the best of our knowledge, it would be very difficult to determine $\boldsymbol{\epsilon}_f$ with experimental methods.

   We are not sure if we understand the last question. $\boldsymbol{\epsilon}_f$ enters into the definition of the observation error covariance matrix, i.e. $\mathbf{R} = \langle \boldsymbol{\epsilon}_o \cdot \boldsymbol{\epsilon}_o^T \rangle$, which is a matrix in the space of the measurement vector. No further transformation is necessary.

9. L 207 While there may be retrieval errors in the lidar backscatter and extinction due to assumptions, assumptions on particle shape and size distribution are not among the assumptions used in lidar retrievals. These examples belong only

to the optics model (forward model). So, perhaps delete "also". Poor assumptions in the optics model or in lidar retrievals would presumably lead to bias errors, whereas measurement errors would more typically be random. Does this make a difference in the analysis?

OK, we deleted "also". We would generally not be sure if assumptions in the optics model necessarily (mainly) lead to biases. For instance, model errors may be dependent on size and morphology of the actual particles. The errors would, correspondingly, fluctuate over time as the aerosol size and composition changes over time. The amplitude of this fluctuation may well be larger than any possible biases. However, in case that the forward-model does introduce a large bias, this would, indeed, be a problem, since analysis algorithms are typically based on the assumption that the errors are unbiased.

10. L 219. I strongly agree that estimating the uncertainties in the optics model is very important. Some discussion here seems warranted about how that can be done. Later I see that this is discussed in the summary (L281 – 292) but I think it would be better if it comes up first here in the discussion section.

Agreed. We added an explanation here, but we also mentioned it again in the conclusion section.

11. L 256 and caption to Fig 4. In both places, it would be kind to remind readers that the delta notation in $\delta x'$ means this is the difference between the value and the background value.

It is not so simple. δx in physical space is the difference between the value and the background, while $\delta x'$ is obtained from δx by applying the transformation $\delta \boldsymbol{x}' = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \boldsymbol{x}$. We repeated this definition in the text with a reference to the definition (which is now found both in the main text and the appendix), and we added a reference to the defining equation both at this point in the text and in the caption to the figure. But we think it would be a bit overdone to repeat the equation in the figure caption.

12. L 259-263. The choice of D21 with its sharp drop-off in weighting appears to mean that only one transformed variable is allowed to change in a meaningful way, although the measurement scenario chosen has nearly the maximum amount of information content available, close to DOF=4. Why was D21 chosen instead of D18, which would allow the measurements to play a bigger role? The only discussion of this choice is the rather vague comment in the Appendix "it is a matter of experience to test different approaches and select the one that proves to be most suited". How and why was this approach determined to be the most suited?

We have done some additional tests and found, in fact, that the analysis is less sensitive to the choice of weighting than we expected. We explain this in the revised paper in Sect. 2.4. Also, we did the following changes to Fig. 4. First, we show $\delta x'$ for both the constrained and the unconstrained analysis. Thus the whole discussion of the figure shifts from a mere description of the behaviour of the constrained analysis to a comparative discussion. This makes it much clearer what kind of effects the weak constraints have on the analysis increments. Second, following a suggestion by reviewer 4, we show not all 20 panels, but only a subset of panels sufficient to illustrate the different behaviour of signal- and noise-related (phase-space) model variables. Third, as mentioned earlier, we picked a more challenging case in which the

reference and background results differ more strongly than in the case we originally picked. So this figure has changed considerably, and the accompanying discussion has become a lot more informative.

13. Comparison of Figure 3 and Figure 2, if I understand right, underscores the fact that there is a significant null space, not controlled by the measurements, since essentially the same measurements in Fig 3 correspond to both the black and red lines in Fig 2. What is not clear to me is what happens in a standard assimilation to the variables that are not well controlled by the measurements? Do they remain close to the background values, or do they vary wildly and arbitrarily? If the former, then the exercise of determining the singular values wouldn't help the assimilation very much (but would still be useful in terms of building knowledge about what we can and can't actually measure). On the other hand, if a standard assimilation arbitrarily varies state variables in the null space, then this is a very important motivation for this technique (and maybe that motivation could be emphasized a little bit more in the introduction and conclusions). Not being very familiar with the field of model assimilation, I guess but don't actually know that there must be other "regularization" techniques in use to prevent an assimilation from arbitrarily varying parameters that are mostly in the null space of the observations, although I imagine existing techniques may be more ad hoc than the method presented here. Can you comment on other methods and demonstrate how this method performs better than other methods?

The reviewer's comment about the null space and the behaviour of the unconstrained (standard) assimilation raises an important issue. As mentioned earlier, we have now run an additional unconstrained assimilation, and we show a comparison of both methods. Figure 2 has been replaced by two figures. The new Fig. 2, similarly to the old figure 2, shows the total mass concentration of different aerosol species, but now for both the constrained and the unconstrained analysis. The new Fig. 1 shows a similar comparison of a selection of aerosol species in specific size bins. This comparison illustrates that the unconstrained analysis yields more erratically varying vertical profiles (i.e., results that vary more wildly in the null-space).

As for ad hoc methods, we did review previously reported approaches in the introduction, such as the one by Benedetti et al. (2009) (L 53-54) based on constraining the total aerosol mass mixing ratio, and the one by Saide et al. (2013) (L 55-56) based on constraining the mass mixing ratio per size bin. One obvious disadvantage is that these approaches are quite inflexible. The number of constraints is fixed in these methods, so one cannot easily adapt the number of constraints to the number of independent measurements to be assimilated, as we can in our approach. (In fact, our method automatises this process.) Also, the available information may not be optimally exploited by these methods (L 57-59). We have not tested such methods, so we cannot comment on their performance. However, we also believe that the burden of proof for such a demonstration does not lie with us. We are employing a mathematically well-founded approach based on information theory. If other groups choose to not follow us, but continue to use ad hoc methods (which, admittedly, may be quite attractive owing to their simplicity), then it is up to them to demonstrate that such ad hoc methods yield sufficiently accurate results while exploiting the available measurement information. Owing to the ad hoc nature of these methods, such a demonstration would have to be repeated for any new set of measurements to be assimilated. Our method can serve as a reference for such tests.

14. L 298-299. “It also appeared”. This result is disappointingly empirical for such a well-founded theoretical study. This observation that SIC was most faithfully retrieved was made in a single case– would you expect this result to be general for all cases, and why? Answering the question is complicated since the singular variables are defined only in the transformed space and therefore the information about what variables are or are not constrained by the measurements is only in this transformed space, not the state space. Yet this statement highlights that it’s desirable to have information about which chemical species and size bins are constrained by the measurements. Is there any way to provide information about this quantitatively? For example, since each state variable is a linear combination of the transformed variables, would showing the linear coefficients in a table make it more obvious which state variables are most closely related to the most significant transformed variables? Perhaps there is a way to use the coefficients to calculate a “fractional significance” that would indicate that x% of the variability in a given state parameter is orthogonal with significant transformed variables while (1-x)% is orthogonal with insignificant variables?

    This is a very good suggestion. We added an extra figure (new Fig. 5) with accompanying discussion and show the magnitude of the linear coefficients for the signal-related control variables. However, the coefficients depend on the B- and R-matrices, which vary spatially. So, we cannot draw very general conclusions from a single test case. But we do think that this discussion helps the reader to understand why the analysis behaves the way it does in our specific case.

2.3 Minor comments

1. L37: Muller et al. 1999 and Veselovskii et al. 2002 and related papers (there are many) would be more relevant references here since they detail retrievals of refractive index, etc., from lidar. (Mishchenko et al. 2007 is an introduction to the Glory satellite and was about retrievals from a polarimeter.)

    Müller, D., U. Wandinger, and A. Ansmann (1999), Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: theory, Appl Optics, 38(12), 2346-2357, doi: 10.1364/AO.38.002346.

    Veselovskii, I., A. Kolgotin, V. Griaznov, D. Müller, U. Wandinger, and D. N. White- man (2002), Inversion with regularization for the retrieval of tropospheric aerosol parameters from multiwavelength lidar sounding, Appl Optics, 41(18), 3685-3699, doi: 10.1364/AO.41.003685.

    Agreed. The references have been replaced.

2. L99: I infer that the ratios in the different size bins are fixed, or else there would be much more than 20 total variables. Is there a way to concisely clarify this in the sentence?

    We are not sure what the reviewer means by the “ratios in the different size bins”. The concentration ratios are certainly not fixed; they can change from one grid cell to the next. The size ranges are fixed. The latter point should now be clear, since we explicitly list the size rages in Sect. 2.1 of the revised manuscript.

3. L109: maybe replace “in the present setup” with “currently in that version”. “The present setup” seems to refer to “the setup used in the present study” but that is misleading, since the present study uses the 20-variable version of the model.

    Agreed.

4. L134: “an” should be “and”

   Yes.

5. L142: “Error correlations ::: are not assumed to be separable”. I’m not sure what this means. What is (or is not) separable from what?

   Vertical and horizontal correlations are often assumed to be separable. We do not make such assumptions, because vertical correlations are often stronger on larger horizontal length scales. In our spectral model (where the horizontal correlations are Fourier-transformed) this means that vertical correlations are larger for smaller horizontal wavenumbers. Since this is not so essential in the context of this study (and potentially confusing), we removed this text in L 142.

6. L153: “see Eq. D16”. Should this be C16?

   Yes. However, following earlier suggestions by the reviewer, this text has been revised and supplied with the main equations from the appendix. So the text in its present form has been replaced.

7. L162-164: Should this sentence perhaps be part of section 2.4, as part of the description of the new technique? The rest of this paragraph (L164-174) is more about the demonstration of the new technique and so seems like a somewhat distinct topic.

   Agreed, we have moved this text.

8. Figure 1: The caption says “note the nonlinear colour scale” Actually, the scale is hardly visible. Please expand the axis labels so they are a similar text size to the caption text.

   Actually, we think that this figure is not particularly relevant in the context of our study, since we do not consider aspects of regional modelling or horizontal information spreading in the analysis. It merely shows one out of many model variables in a single model layer, which does not convey much useful information. Also, since we consider a single profile, the analysis impacts the mass mixing ratio only at and around the observation site, which is difficult to see in a regional plot. We therefore removed this figure in the revised manuscript.

9. Figure 2: The axis labels’ and inset box labels’ font size should also be increased here.

   OK, we increased the font size in all figures wherever it was necessary and possible.

10. L 391. The variable n is not defined. Possibly this is the only case, but I would also request that variables be re-defined frequently when used in key equations. If a reader is directed from another part of the paper to Equation D18 or C12, for example, then it would be nice if all the information relevant to understanding that equation is given immediately after that equation, rather than having to scroll through 8 or 10 pages to relocate the definitions of key variables.

    Agreed, we have added the definition of n. Also, the problem with directing the reader to equations in the appendix is now significantly alleviated in the revised versions, since we re-stated the key equations in the main body of the paper (see our response to an earlier comment).

11. L563. The symbol lambda is used for wavelength elsewhere in the text. You might consider using a different symbol here.

    OK, we have replace it by mu.

3 Reviewer 3

The line-number references of the reviewer seem to be offset relative to those given in the online pdf of our manuscript. But we think that we figured out each point in the text the reviewer referred to.

1. It is well known, that the problem of inversion of standard $3\beta + 2\alpha$ lidar measurements to the particle microphysics is undetermined and to constrain it, numerous techniques were considered. The authors suggest an interesting approach to assimilation of lidar measurements into chemical transport model. It looks like a promising concept to extract the information about particle parameters from lidar measurements. Paper is very well written and should be published.

   We thank the reviewer for his positive evaluation of our manuscript and for his helpful comments.

2. The structure may be questionable, because a half of material is put in appendices. These appendices are clearly written and are definitely useful for unprepared reader. I personally, had no problems with material structure.

   We agree that the structure was not optimal for all types of readers. We found that the compromise suggested by reviewer 2 adequately addresses these concerns. We refer to our detailed response to reviewer 2, which explains the changes we implemented in the revisions.

3. Additional references to the previous studies of lidar data inversion would be desirable , and other Referees have already suggested several.

   Agreed. We added a paragraph in the introduction with a brief discussion of other studies, also from numerical weather prediction data assimilation.

4. Stability of retrieval strongly depends on aerosol type. It is more challenging for aerosols with dominant coarse mode and for particles with strong absorption. The authors consider only one example (not the most challenging) in their simulation, so it is not very clear how the approach will work for other aerosol types. But this may be a subject of separate study.

   Yes. Although it is not the subject of this paper to comprehensively test all sorts of mixed aerosol populations, we do agree that the case we picked was a little bit too easy. This is mostly because the background and reference cases were very close to each other. In such a case one does not see very clear differences between a constrained and unconstrained analysis. In the revised paper we have picked a more challenging case, and we now show our test results for both the constrained and the unconstrained 3DVAR algorithm. This helps to better illustrate what practical significance the constraints can have.

4 Reviewer 4

Summary

The authors consider the case of assimilation of remote-sensing data (specifically aerosol extinction and backscattering coefficients) applied to aerosols fields within a chemical transport model. They describe how an additional term can be added to the 3D-var cost-function so that the assimilation adjusts only those components (in a transformed space) for which the observations provide information. The additional term relies on the singular value decomposition of the scaled observation operator. In this way, the assimilation automates the choice of control variables in an otherwise highly under-constrained inverse problem.

Verdict

The paper is very well written and is surprisingly clear, given the subject matter. The manuscript introduces a potentially very powerful concept for variable selection into the field of aerosol data assimilation. The authors have probed the idea in a minimal test case, which assists in understanding the effects. I found that the shortcomings of the paper were relatively minor. I felt there was insufficient discussion of the literature of related treatment. I was unsure about whether the organisation of the material was optimal (see the "Main comments"). Finally, a counter-experiment without the addition of the new constraint in the 4D-var cost function was, in my opinion, lacking. All in all, I believe that the paper should be published, pending the minor revisions suggested below.

We are grateful for this encouraging assessment of our work, as well as for the insightful comments and suggestions. It is obvious that the reviewer has devoted considerable time into studying the manuscript and providing constructive criticism on various aspects of the content and organisation of the paper. Our detailed response to these comments follows.

4.1 Main comments

1. There was little or no discussion of literature on related treatments. I have not the time to read all of these myself, however I have included a list at the end of articles that may be relevant, for example those that deal with information content of observations in data assimilation or those that refer to the singular value decomposition of the observability matrix.

   We have, indeed, only cited studies on aerosol data assimilation. Most of the studies cited by the reviewer are concerned with numerical weather prediction (NWP). We have added a paragraph to the introduction to discuss related NWP studies and include the citations suggested by the reviewer.

2. I believe that a small counter-experiment was lacking. In the results presented in section 3.2, I would suggest also presenting results for the assimilation experiment which did not include the additional constraint in the 3D-var cost-function.

   This is a very valid point that was also brought up by other reviewers. We have included these results and revised the figures and discussion accordingly. In particular, we replaced Fig. 2 by two new figures. The new Fig. 2 shows, similarly to the old Fig. 2, the total mass concentration of different aerosol species, but now for both the constrained and the unconstrained analysis. The new Fig. 1 shows a similar comparison of a selection of aerosol species in specific size bins.

This comparison illustrates that the unconstrained analysis yields more erratically varying vertical profiles (i.e., results that vary more wildly in the null-space).

3. I was unsure whether the organisation of the material was optimal - I highlight this as an issue that the editor may wish to take up. The introduction concludes by urging the reader to read the Appendix before proceeding onto the rest of the methods and results section. Much of the interesting methodology is contained within the Appendix, and we agree that it would be difficult to make sense of the main part of the paper without a good understanding of the contents of the Appendix. As such, I would suggest incorporating the Appendix into the main body of the text. At one level, this is really a matter of taste, and thus I leave it to the editor.

   This is a tricky point. We put some thought into this before writing the paper, and we concluded that the appendix is, indeed, most interesting for readers who are mainly interested in data assimilation methodology, and for those who are very eager to learn something about it. But other readers, e.g. lidar instrument developers, will most likely be deterred from reading the paper if we merge the entire appendix with the main body of the paper. However, the reviewer's criticism is very valid, and it has been brought up by several reviewers. We believe that reviewer 2 has suggested a very good compromise, namely, to state and explain the main results (equations) from the appendix in the methodology section of the paper, while retaining the derivations and more detailed explanations in the appendix. This alleviates the problem that parts of the main text are hard to understand without the information given in the appendix. At the same time, we avoid the risk of making the paper inaccessible (or just too boring) for those who do not mainly work with data assimilation methodology.

   We therefore followed the suggestions of reviewer 2 in this point. It seems to us that this also adequately addresses the main point of criticism brought up by reviewer 4.

4.2 Minor comments

1. When describing observation errors, there was no reference to the component from "representativity errors" (i.e. measurements are made at a point, or over a small area in the case of remote sensing, while model grid-boxes are typically in the order of kilometres across in the horizontal dimensions). All of the discussion about observation errors was in terms of the measurement error and errors in the observation operator, both of which are relevant. However the representativity component is not insignificant in many contexts.

   Agreed. We have added a discussion of the representativity error in the text accompanying table 2, where we made it clear that in this numerical test we have neglected this source of error.

2. Observation standard deviation was reported in percentage, but it was unclear what this was a percentage of. Please clarify.

   It is a percentage of the observed backscattering coefficient or extinction coefficient. We changed the text in Sect. 2.5 from "We assumed an observation error standard deviation of 10 %" to "We assumed that the observation error standard deviation is 10 % of the measurement value."

3. I would suggest replacing all instances of the term “costfunction” with “cost function” (or “cost-function”). The latter is about 15 times more common (on the web, at least). Similarly, I believe that the compound word “nullmatrix” is used in German (capitalised, that is) whereas it is “null matrix” (or “zero matrix”) in English.
   Agreed.

4. I could not find a definition to the term “signal degrees of freedom”. Please include this somewhere (preferably at first usage, or in the Appendix).
   We have added a detailed explanation of the terminology to Sect. 2.4. Following the suggestions of reviewer 2, we have also provided key equations of the appendix with explanations in the main text. This also applies to Eq. C12, which is now provided in the methodology section. Thus the explanation and definition of the term “signal degrees of freedom” now appears much earlier in the revised paper. (Note that in the remote sensing community the number of signal degrees of freedom is also known as the “effective rank” of the problem.)

5. Line 48: Please replace “This is a rather bold approach that largely disregards ...” with “This is approach largely disregards ...” – please use argument rather than rhetoric to explain what is wrong with the work of others.
   Agreed.

6. Line 54: The reference to Kahnert (2009) is used to show that several optical properties at multiple wavelengths may allow constraining more than just the total mass concentration. Surely other authors have looked into this. If so, please summarise other work done. If not, please say so.
   We did cite the study by Burton et al. (2016) (L64), although we did so in the introduction. We now have added two more references that analyse the information content of lidar observations, namely, the papers by Veselovskii et al. (2004) and Veselovskii et al. (2005). However, these papers analyse the information content with respect to particle size and refractive index, not with respect to chemical composition. Therefore, we put these citations into the introduction.

7. Line 98: “... using 40 eta-layers with variable thickness depending on the under- lying topography” – do you just mean that this is a terrain-following coordinate? Or is there something more sophisticated about this?
   OK, we have replaced this with “using 40 terrain-following coordinates”.

8. Line 125: “The background error covariance matrix of the model a priori is modelled with the NMC method ...” . I checked the reference (Kahnert, 2008), in which I believe this is described. If the implementation is the same here as in the 2008 article, then I believe that it is best to say that it “follows similar principles to the NMC method” or “is inspired by the NMC method”. If it is indeed the NMC method, the authors should clarify the difference to methodology laid out in Kahnert (2008).
   OK, we have replaced this with “follows similar principles to the NMC method”.

9. Line 129: I would suggest replacing “Given m observations of, e.g., m1 different parameters at m2 different wavelengths, so that m1 m2=m, how many...” with “Given m observations (e.g., m1 different parameters at m2 different wavelengths,

so that m1 m2=m), how many..."

Agreed.

10. Line 130: "... we can constrain to better than observation error" – do you mean "model error"? If not, please explain that the transformation makes the (rescaled) observation errors and (rescaled) model variables comparable.

    This was a bit confusing. We have reformulated this sentence, and we have addd a more detailed explanation of the terminology *signal degrees of freedom*.

11. Line 134: Please replace "... a singular value decomposition of the Jacobian of the observation operator ..." with "... a singular value decomposition of the Jacobian of the scaled observation operator ..." or something similar. By the way, this scaled observation operator appears to have a name: "the observability matrix"

    Yes. Actually, this text has been extended with a lot more explanations, and it now provided the main equations from the appendix. We have followed the reviewer's suggestion and introduced the term *observability matrix* for the scaled Jacobian.

12. Footnote 2, page 5: I found this distinction a bit cryptic. Please consider rephrasing.

    There seem to be two fractions in the community. One that uses *data analysis* and *data assimilation* almost interchangeably, and another that insist on keeping these two concepts apart. We are mostly guilty of belonging to the first one, but we do not want to make a big deal out of mere questions of terminology (which is why we put this into a footnote rather than into the main text). However, we did our best and clarified the text as best as we could.

13. Line 150: I realise that this is something that is clarified later on, but I would suggest saying a few words at this point about the synthetic observations; namely, what kind of observations they were and how many observation points there were.

    Agreed; we have added this information in the revised manuscript.

14. Line 154: I would suggest the following change "thus providing nearly perfect observations. (We assumed an observation error standard deviation of 10 %) The only ..." becomes "thus providing nearly perfect observations (we assumed an observation error standard deviation of 10 %). The only...". See also my comment about about describing the units for the observation error standard deviation.

    Agreed (replacing "observations. (We" by "observations (we". In addition, in response to an earlier request to be more specific what me mean by "10 %" (percent of what?), we have replaced the text in parenthesis with "(we assumed that the observation error standard deviation is 10 % of the measurement value)".

15. Line 162: What is "Nd:YAG"? Please clarify. I suspect that this is some error with the bibliography manager.

    It is no error. "Nd:YAG" is the standard abbreviation for "neodymium-doped yttrium aluminium garnet" laser, one of the most commonly used solid-state lasers in remote sensing. We have now added this information at the first instance, which is in the introduction section.

16. Line 168: I would suggest the following change: "... two wavelengths. (Compare, e.g., cases 1., 2., and 3. to cases 4., 5., and 6.) Hence ..." becomes "... two wavelengths (compare, e.g., cases 1., 2., and 3. to cases 4., 5., and 6.). Hence ..."
Agreed. However, reviewer 2 has suggested to replace the cases considered in Table 1 with different cases that are more closely associated to combinations of wavelengths and parameters that are technologically feasible and common. Thus the text accompanying Table 1 (now Table 2 in the revised manuscript) has changed considerably.

17. Line 171: a missing full stop after the right parenthesis.
Agreed.

18. Table 1, caption: the "Nd:YAG" term appears again.
See our earlier response.

19. Line 181: I believe that "weak constrains" should be "weak constraints".
Yes.

20. Line 189: See my comment above about the representativity component to the observation error.
Agreed, see our earlier response.

21. Figure 1: I think it would be interesting to see the increment as an additional panel in this figure.
Figure 1 has been criticised by several reviewers. In fact, this figure is not particularly useful in the context of our paper. We are not discussing any aspects of regional modelling or horizontal information spreading in the assimilation algorithm. The model merely serves us to provide us with a test case. So, we have removed this figure in the revised manuscript (see also our response to reviewer 2).

22. Figure 1: The text on the scale is a bit too small. I would suggest having one scale, rather than three, and enlarging the scale so that the labels can be read.
See the previous item.

23. Figure 2: The units appear to be "mixing ratio [ppb-m]". Do you mean mass mixing ratio? Please clarify.
Yes. This has been corrected.

24. Figure 4: Do we need all panels? Why not just show the first three or four, and then a selection of the remaining terms.
Agreed, we now show 10 instead of 20 panels. Following the comment by reviewer 2, we have run a $3\alpha + 2\beta$ test case, in which case we have 5 signal degrees of freedom. Thus we now show the first 5 signal-related transformed increments, and 5 out of the 15 noise-related increments (Fig. 4).

25. Line 262: I would suggest the following change " ... dramatic decrease in both the entropy and signal degrees of freedom ..." becomes " ... dramatic decrease in both the entropy-change and signal degrees of freedom ..."
Agreed; however, owing to the changes in Sect. 3 this part of the text in the conclusions has now also changed.

26. Line 282: "It also appeared that among the original model variables, secondary inorganic aerosol components were most faithfully retrieved by the inverse modelling solution" – why is this? why SIA? Do they have specific optical properties to make them more observable by such LIDAR pseudo-observations?

This question has been brought up by several reviewers. We follow the suggestion of reviewer 2 and add an analysis of the linear coefficients that transform the elements in model space to the signal-related control variables. We have added a new figure (Fig. 5) and a discussion — see our detailed response to reviewer 2.

27. Line 293: I would suggest the following change: "The present study should be extended..." becomes "The present study could be extended..."

Agreed.

28. Line 295: I believe that the expression "highly underrated" is somewhat dramatic and relatively colloquial, and does not fit with the tone in the rest of the paper. The authors are encouraged to use argument rather than rhetoric to make their point.

OK, we have replaced the text with "Another important issue concerns the choice of ...".

29. Line 297: Regarding the statement "There is little one can put forward in defence of this model other than pure convenience". Some justification is required (e.g. some references) to demonstrate why this model is untenable. There's a saying (attributed to George Box) "All models are wrong, some models are useful". Does this model give significantly worse results than representations, or is it just inaccurate in its assumptions?

Worst of all, this model is rather unpredictable, since its accuracy depends on the size, refractive index, and shape of the aerosols. Also, it may, in some cases, give reasonable results at one wavelength and for one specific parameter, and fail at other wavelength or for other optical parameters.

There is a large body of work concerned with aerosol optics and the shortcomings of simplified model particles. Some of these studies focus on specific types of aerosols, others on specific morphological properties, such as non-sphericity, inhomogeneity, surface roughness, or chemical heterogeneity. It is difficult to pick just a few of such studies as representative citations. So, we found that the best solution was to cite a recent review paper on aerosol optics modelling that discusses the strengths and shortcomings of various morphological models (Kahnert et al., 2014).

30. Paragraph beginning at line 307: It may be worth making it clear that y is not observed, but a model equivalent of the observations

We have inserted the following sentence: "The operator $\hat{H}$ maps from model space into observation space, which allows us to compare model output and observations."

31. Lines 324 and 326: I would suggest replacing all instances of "3-dimensional" with "three-dimensional"

Agreed.

32. Paragraph beginning 336: I would suggest mentioning that the assumption of unbiased background and observation errors

    Agreed.

33. Footnote 6, page 15: See my comments above about the representativity component of the observation error.

    OK; see our earlier response.

34. Footnote 7, page 16: I would suggest the following change: "The observation errors are often uncorrelated" becomes "The observation errors are often assumed to be uncorrelated (this is not always true)"

    Agreed.

35. Paragraph beginning at line 368: Please comment on the role of spatial and inter- species correlations, particularly in light of the comment "if we allow all model variables to be freely adjusted" (line 374).

    OK. We have added the following footnote (after "(within the given error bounds).": By solving the equation $\nabla J|_{\boldsymbol{x}=\boldsymbol{x}_a} = \mathbf{0}$ for the analysed state $\boldsymbol{x}_a$ it can be shown that the solution to the inverse problem is given by $\boldsymbol{x}_a = \boldsymbol{x}_b + \mathbf{K} \cdot (\boldsymbol{y} - \hat{H}(\boldsymbol{x}_b))$, 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 $\boldsymbol{x}_b$ by mapping the increment $(\boldsymbol{y} - \hat{H}(\boldsymbol{x}_b))$ 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.

36. Line 369: It might be worth noting that δx is not constrained to ensure that all components of x remain positive in the analysis.

    There is no such constraint in the minimisation process itself, but we do post-process the results for δx such that negative concentrations would be set to zero. In practice, this rarely ever happens.

37. Line 386: The phrase "rather tricky" strikes me as somewhat colloquial. I would suggest the following change: "However, to actually make such a comparison israther tricky" becomes "However, to actually make such a comparison poses two problems."

    Agreed.

38. Paragraph beginning at line 390: Please introduce the meaning of the angle- bracket notation. I believe that this is common in physics, but other disciplines (e.g. statistics) often use different notation for the expectation.

    OK, we have added a formal definition of the expectation value for discrete variables in a footnote.

39. Footnote 8, page 17: Should $A \cdot A = B$ be $A^T \cdot A = B$?

    Yes!

40. Line 425: Should "(C7)-(C9)" not be "(C6)-(C9)"? As far as I can see, Eq. (C6) is required here.

    Yes.

41. Line 434-435: Please state which particular sections/chapters of Rodger (2000) the reader is referred to.
Agreed.

42. Equations C12, C15: I would suggest showing the range of the summation to indicate that it is a summation over observations (i.e. i ranges from 1 to m)
This is not generally true. The summation goes from 1 to $\min\{m,n\}$, where n is the dimension of model space, and m is the dimension of observation space. We have added these summation limits to the sums.

43. Line 479: "Naively, one may have expected that the dimension would, on the contrary, be reduced to $n-k$" – why? is this because the number of unknowns remains the same but the number of equations to be solved has increased by k?
In physics one usually learns about holomorphic constraints in theoretical mechanics, often by considering a point mass moving on a hypersurface. So, this is often the mental picture one invokes when dealing with constrained problems. For instance, a point mass in three-dimensional Euclidean space with a single holomorphic (i.e. strong) constraint can be pictured as moving on a two-dimensional surface. Thus this constraint reduces the dimension of the manifold on which the the point mass can move from three to two. One would therefore *naively* expect that one is now dealing with a two dimensional problem. The reason why this is naive is because a nonlinear constraint will correspond to a *curved* manifold. To characterise this manifold requires additional equations. Only if we have *linear* constraints, then the hypersurface is simply a tilted plane, which, by a suitable rotation-translation, can be brought into coincidence with, e.g., the xy plane. In such cases, and only in such cases, can the dimension of the problem actually be reduced, as one would naively have expected.

44. Line 486: I would suggest the following change: "(Note that the covariance matrices and their inverses are symmetric, i.e., $R^T = R$, etc.)" becomes "Note that the covariance matrices and their inverses are symmetric (i.e. $R^T = R$, etc.)."
Agreed.

45. Appendix: For all unit and zero matrices (and vectors), I would suggest indicating the dimension as a sub-script.
Agreed. We have changed this throughout the manuscript.

46. Line 498: I would suggest adding a subscript to clarify with respect to what the differentiation refers (i.e. replace ∇ with ∇_ξ).
Agreed.

47. Paragraph beginning line 515: how was this tuning done in practice?
As it is explained in the text. When the error variance is too large, one can see that the analysis is close to the unconstrained one. When it is too small, the analysis lies very close to the background estimate. One varies the variance until one obtains an analysis that departs from the background without drifting over to the (often noisy) unconstrained analysis.

48. Line 549: "It turns out that Eq. (D18) gives a relatively sharp transition from unconstrained to constrained model variables, while Eq. (D19) gives a very gentle transition" – this can be seen from the equations. I would suggest replacing the sentence with "It can be seen that Eq. (D18) gives a relatively sharp transition from unconstrained to constrained model variables, while Eq. (D19) gives a very gentle transition"

    We have replaced the text with "We tested all three approaches . These tests showed that the different approaches often yield analysis results that are quite similar. However, in each approach the free parameters σ_G and c are tuned to different values. If they are not well tuned, then the analysis tends either toward the background estimate or toward the unconstrained analysis, as explained earlier in the text following Eq. (D15)." Our tests, so far, showed that the differences between these approaches are not quite as dramatic as we expected.

49. Paragraph beginning line 567: I found that this went too fast and skipped a bit too much detail, after what was otherwise a very well-written paper that included a fair bit of theory. In particular, can you please explain in further detail the reduced matrices. The phrase "we are primarily interested in constraining the chemical components" was surprising, since I thought the authors were mainly interested in the aerosol components. What does it mean to "restrict ourselves to the chemical subspace"?

    This seems to be a misunderstanding. What we mean by "chemical components" is "chemical components in the aerosol phase". Since our paper is exclusively concerned with aerosols, we thought that there was no risk of misunderstanding. Thus, by "chemical subspace" we mean "subspace of aerosol components". We have revised the text accordingly and replaced all instances of "chemical components" by "aerosol components", and similarly for "chemical subspace". Also, we have revised the text in response to point 51 (see below).

50. Line 570: Full stop missing after N_c.

    OK.

51. Paragraph beginning 574: similar to the above comment, I found that this skipped over too much detail. Please add further explanation. The authors state that in their present study, they use a Cholesky decomposition of the B-matrix. Is this what was used in Kahnert (2008), or is this described as the "spectral formulation"? If it is different, it may be relevant to understand why the Cholesky decomposition was preferable to the author's previously presented methodology. This is mainly to understand the requirements and limitations of the proposed methodology.

    We *are* using the spectral formulation for the minimisation of the cost function. However, we formulate the weak constraints in a subspace of physical space, as explained above. The Cholesky decomposition is only applied to the reduced B-matrix in the formulation of the weak constraints. We do not go into the details of spectral data assimilation, since these questions are rather specific to our particular implementation, while the paper is not restricted to spectral methods. However, we have rewritten this entire subsection and explained the reduced subspace approach in much more detail. We have also added a short footnote on how to incorporate this into the spectral formulation.

4.3 Minor formatting issues

1. References with parentheses inside parentheses: lines 33, 268, 269, 376

   This has been corrected.

2. Some of the in-line equations appeared to be missing spaces on one or both sides of the equals sign – this only appeared in the appendix. See lines: 391, 403, 404, 425, 515. I might just be imagining it. The paper was otherwise very well laid out.

   Our latex program seems to insert spaces when using the eqnarray environment, but not when using the equation environment. We trust that the copy editor will take care of this problem.

4.4 References the authors may wish to consider

- Qin, X. Measuring information content from observations for data assimilation: relative entropy versus shannon entropy difference. Tellus: Series A. 59, 2, 198- 209, 2007.
- J. Joiner, A. M. da Silva. Efficient methods to assimilate remotely sensed data based on information content. Q. J. R. Meteorol, SOC. (1998), 124, pp. 1669- 1694
- C Cardinali, S Pezzulli, E Andersson. Influence-matrix diagnostic of a data as- similation system. Q. J. R. Meteorol. Soc. (2004), 130, pp. 2767-2786. doi: 10.1256/qj.03.205
- C. Johnson, N. K. Nichols; B. J. Hoskins. Very large inverse problems in atmo- sphere and ocean modelling. Int. J. Numer. Meth. Fluids 2005; 47:759-771.
- M Bocquet, 2009: Toward Optimal Choices of Control Space Representation for Geophysical Data Assimilation. Mon. Wea. Rev., 137, 2331-2348, doi: 10.1175/2009MWR2789.1.
- F Rabier, N Fourrie, D Chafai, P Prunet. Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. Q. J. R. Meteorol. Soc. (2002), 128, pp. 1011-1027
- C Johnson, B. J. Hoskins, N. K. Nichols. A singular vector perspective of 4D-Var: Filtering and interpolation. Q. J. R. Meteorol. Soc. (2005), 131, pp. 1-19 doi: 10.1256/qj.03.231

Agreed; these have been added to and discussed in the introduction.

5 Reviewer 5

Summary

The inversion of aerosol optical properties into the aerosol chemical composition is a ill posed problem. The authors use information theory techniques to estimate the amount of information contained in LIDAR observations. They present different methods to make use of it as contains in a 3DVAR algorithm. This is meant to avoid assimilating noise inherent to the observations. To evaluate their constrain methods, they create synthetic observations from CTM simulations and assimilate them back

into the CTM.

Recommendation

The paper is well written and should be published. The methodology proposed is novel and can be applied to different observations within the variational assimilation framework.

We thank the reviewer for this positive evaluation of our paper.

Main comments

The authors choose to place all equations and their derivations into different appendixes. This hindered slightly the reading of sections 2.4, 3.1 and 3.2. However, the overall readability of the manuscript is improved by the focus on the description and evaluation of the method in the main text.

We agree. This point has been brought up by the other reviewers as well. We have followed the recommendations given by reviewer 2 and included the key equations with explanations in the main text, while providing the more detailed derivations in the appendix. This is a good compromise that keeps the paper accessible to non-theorists, while providing all the necessary details in the appendix for the interested readers.

Minor comments

Figure 1 is hard to read, specially the colour bar. Otherwise, the previous Referees have a number of valid suggestions for improvement, and I have nothing to add.

We have removed this figure in the revised manuscript. Since the paper is not concerned with those aspects specific to regional modelling, this regional plot conveys no useful information in the context of this paper.

6 Comment by P. Chazette

May be you can check the papers of Wang at al. in ACP (2013, 2014a and b), where lidar assimilation is tested.

We thank Patrick Chazette for bringing these three papers, which he co-authored, to our attention. The results reported in these articles are very interesting. The paper by Wang, Sartelet, Bocquet, and Chazette (2013) is particularly impressive. It investigated assimilation of lidar and ground observations of PM10 and performed an observing system simulation experiment. The results demonstrate that a relatively small lidar network can give analyses and forecasts of similar, and in some cases even higher accuracy than corresponding results obtained with an extensive network of ground stations, such as AirBase. This clearly demonstrates the potential of lidar observations. However, this study is only marginally relevant in the context of our paper, because it considers assimilation of lidar measurements for determining PM10, not for determining the concentrations of each aerosol component. It does not discuss the question of how to constrain the assimilation algorithm in order not to assimilate noise. For this reason, we did not feel compelled to add a citation to this article.

The paper by Wang, Sartelet, Bocquet, and Chazette (2014) presents a comparison of modelled and measured backscattering profiles, where the measurements were taken by a mobile lidar in the vicinity of Paris. The results of this comparison are highly encouraging. They also describe their assimilation methodology. If we understand it correctly, they set up the assimilation to

correct PM10, and they distribute the analysis increment back to the various aerosol components in each size class according to the a priori distribution. In the context of our study, this is the most relevant fact in this paper, since it describes an ad hoc method for specifying constraints. Essentially, this approach seems to be based on the same idea as that described in Benedetti et al. (2009). However, we found that the explanations in the paper by Wang et al. (2014) were more detailed than in the paper by Benedetti et al. (2009). For this reason, we have added a citation to this paper.

Finally, the paper by Wang, Sartelet, Bocquet, Chazette, et al. (2014) presents a very impressive and comprehensive evaluation work of the potential of assimilating lidar measurements from the EARLINET network into an aerosol transport model. Since it is an application rather than methodology paper, we did not cite it here; but we will be sure to cite it when we have come that far and submit a paper on the operational evaluation of our lidar assimilation system.

References

Benedetti, A., Morcrette, M. J.-J., Boucher, O., Dethof, A., Engelen, R. J., Huneeus, M. F. H. F. N., Jones, L., andS. Kinne, J. W. K., Mangold, A., Razinger, M., Simmons, A. J., and Suttie, M.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, D13 205, 2009.

Burton, S. P., Chemyakin, E., Liu, X., Knobelspiesse, K., Stamnes, S., Sawamura, P., Moore, R. H., Hostetler, C. A., and Ferrare, R. A.: Information content and sensitivity of the $3\beta+2\alpha$ lidar measurement system for aerosol microphysical retrievals, Atmos. Meas. Techniques, 9, 5555–5574, 2016.

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

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

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

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

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 for a size-resolved sectional model: impacts of observationally constrained, multi-wavelength and fine mode retrievals on regional scale analysis and forecasts, Atmos. Chem. Phys., 13, 10 425–10 444, 2013.

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

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

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

Michael Kahnert[1,2] and Emma Andersson[2]

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

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

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

Abstract.

We theoretically and numerically investigate the problem of assimilating multiwavelength lidar observations of extinction and backscattering coefficients of aerosols into a chemical transport model. More specifically, we consider the inverse problem of determining the chemical composition of aerosols from these observations. The main questions are how much information the observations contain to ~~constrain~~ determine the particles' chemical composition, and how one can optimise a chemical data assimilation system to make maximum use of the available information. We first quantify the information content of the measurements by computing the singular values of the scaled observation operator. From the singular values we can compute the number of signal degrees of freedom, N_s, and the reduction in Shannon entropy~~. For an observation standard deviation of 10 %, it is found that simultaneous measurements of extinction and backscattering allows us to constrain twice as many model variables as extinction measurements alone. The same holds for measurements at two wavelengths compared to measurements at a single wavelength. However, when we extend the set of measurements from two to three wavelengths then we observe only a small increase in the number of signal degrees of freedom, anda minor change in the Shannon entropy. The~~ , H. As expected, the information content as expressed by either N_s or H grows as one increases the number of observational parameters and/or wavelengths. However, the information content is strongly sensitive to the observation error~~; both the number of signal degrees of freedom and the reduction in Shannon entropy steeply decrease as the observation standard deviation increases in the range between 1 and 100 %~~. The larger the observation error variance, the lower the growth rate of N_s or H with increasing number of observations. The right singular vectors of the scaled observation operator can be employed to transform the model variables into a new basis in which the components of the state vector can be ~~divided~~ partitioned into signal-related and noise-related components. We incorporate these results in a chemical data assimilation algorithm by introducing weak constraints that restrict the assimilation algorithm to acting on the signal-related model variables only. This ensures that the information contained in the measurements is fully exploited, but not over-used. Numerical ~~experiments confirm~~ tests show that the constrained data assimilation algorithm ~~solves~~ provides a solution to the inverse problem ~~in a way that automatises the choice of control variables, and that restricts the minimisation of the costfunction~~ that is considerably less noisy than the

corresponding unconstrained algorithm. This suggests that the restriction of the algorithm to the signal-related model variables suppresses the assimilation of noise in the observations.

1 Introduction

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

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

A systematic class of statistical methods for solving this inverse problem is known as data assimilation. Recent studies have applied data assimilation to aerosol models with varying degrees of sophistication, ranging from simple dust models (Khade et al., 2013) and mass transport models (Zhang et al., 2014) to microphysical aerosol models based on modal (Rubin and Collins, 2014) or sectional descriptions (Sandu et al., 2005; Saide et al., 2013) of the aerosol size distribution. The assimilation techniques that have been used comprise variational methods, such as 2D (Zhang et al., 2014), 3D (Kahnert, 2008; Liu et al., 2011), and 4D variational methods (Benedetti et al., 2009), as well as ensemble approaches (Sekiyama et al., 2010). Assimilation of satellite products for trace gases is relatively straightforward, since observed and modelled trace gas concentrations are almost directly comparable. However, aerosol optical properties observed from satellites are not directly comparable to the modelled size distribution and chemical composition of the aerosols. Solving this problem amounts to regularising a severely under-constrained inverse problem. Previous aerosol assimilation attempts have been mainly based on educated guesses about the information content of the observations. For instance, there have been studies on the assimilation of aerosol optical depth (AOD) in which all chemical aerosol components in all size classes and at all model layers were used as indepen-

dent control variables (Liu et al., 2011). This ~~is a rather bold approach that~~ approach largely disregards the problems involved in inverse modelling. By contrast, it has been proposed to only allow for the total aerosol mass concentration to be corrected by data assimilation of AOD ~~(Benedetti et al., 2009)~~(Benedetti et al., 2009; Wang et al., 2014). This is a more prudent approach based on the plausible assumption that a single optical variable only contains enough information to control a single model variable. There have also been intermediate approaches in which the total aerosol mass per size bin have been used as control variables (Saide et al., 2013).

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

In ~~a recent study, Burton et al. (2016) have~~ numerical weather prediction (NWP) modelling, several studies have discussed the information content of satellite observations for meteorological variables. For instance, Joiner and da Silva (1998) applied a singular-value decomposition (SVD) approach in order to reduce the effect of prior information in the analysis, so that the retrieval and forecast errors can be assumed to be uncorrelated. Rabier et al. (2002) considered assimilation of IR sounders, which typically provide a large number of different channels. They applied methods of information and retrieval theory in order to decide which channels contain most information about the vertical variation of temperature and humidity. Cardinali et al. (2004) employe the influence matrix to compute diagnostics of the impact of observations in a global NWP data assimilation system. Johnson et al. (2005a, b filtering and interpolation aspects in a 4DVAR assimilation system by use of an SVD approach. They also used Tikhonov regularisation theory to optimise the signal-to-noise regularisation parameter in order to maximise the information that can be extracted from observations. Xu (2006) compared different metrics, namely, the relative entropy and the Shannon-entropy difference, to measure information contents of radar observations assimilated into a coupled atmosphere-ocean model. Bocquet (2009) used methods of information theory to address the question how to determine an optimum spatial resolution of the discretised space of control variables in geophysical data assimilation.

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

problem. In ~~this~~ the present paper, we want to investigate the inverse problem that goes all the way from optical properties to the chemical composition of particles.

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

The paper is organised as follows. Section 2 gives a rather concise introduction of the modelling tools and of the numerical approach employed to studying the information content of extinction and backscattering observations. Section 3 presents the main results of this study, and Sect. 4 offers concluding remarks. To make this paper self-contained, we included an appendix that gives a brief introduction to some essential concepts of data assimilation, and a detailed explanation of the methods we used for quantifying the information content of aerosol optical observables. ~~It is advisable to read the appendix first before reading the body of the paper. Readers that are not interested in the theory may leave out the appendix at the risk of missing some of the discussions in the main body of the paper.~~

2 Methods

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

[1] In addition to lidar measurements from ground-based and aircraft-carried instruments (e.g. Burton et al., 2015), there are currently two space-borne lidar instruments in orbit. The CALIOP instrument on-board the CALIPSO satellite has been launched in April 2006; it has three receiver channels, one at 1064 nm, and two channels at 532 nm to measure orthogonally 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).

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

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

Size bin 1: 10–50 nm

Size bin 2: 50–500 nm

Size bin 3: 500–1250 nm

Size bin 4: 1250–5000 nm.

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

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

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

For the sake of simplicity we here use the mass transport model without aerosol microphysical processes (see next section). The model is set up over Europe covering 33° in the longitudinal and 42° in the latitudinal direction in a rotated lat-long grid with $0.4° \times 0.4°$ horizontal resolution. In the vertical direction the model domain extends up to 13 hPa, using 40 ~~η-layers with variable thickness depending on the underlying topography~~terrain-following coordinates. The meteorological input data are taken from the numerical weather prediction model HIRLAM (Undén et al., 2002). For the emissions of all aerosol components we used EMEP data for the year 2007, where EC and OC emissions were computed from total primary particle emissions based on the data in Kupiainen and Klimont (2004, 2007).

2.2 Aerosol optics model

We have two different optics models coupled to MATCH, one to the mass transport module, and another to the aerosol microphysics module. The former assumes that all aerosol species are homogeneous spheres, and that each chemical species

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+5.0e-3 i	1.53+5.6e-3 i	1.52+1.6e-2 i
Dust	1.53+1.7e-2 i	1.53+6.3e-3 i	1.53+4.3e-3 i
NaCl	1.51+2.9e-7 i	1.50+1.0e-8 i	1.47+2.0e-4 i
OC	1.53+5.0e-3 i	1.53+5.6e-3 i	1.52+1.6e-2 i
EC	1.66+7.2e-1 i	1.73+6.0e-1 i	1.82+5.9e-1 i

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

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

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

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

$$\boldsymbol{y} = \hat{H}(\boldsymbol{x}) + \boldsymbol{\epsilon}_o, \tag{1}$$

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

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

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

The MATCH model contains a 3DVAR data assimilation module. This model uses a spectral method, i.e., the model state vector is Fourier-transformed in the two horizontal coordinates. All error correlations in the horizontal direction are assumed to be homogeneous and isotropic. ~~Error correlations in the vertical direction and among different chemical species are *not* assumed to be separable.~~ The background error covariance matrix ~~of the model a priori~~ is modelled with a method that follows similar principles to the NMC method (Parrish and Derber, 1992). A more complete description of our 3DVAR program can be found in Kahnert (2008).

2.4 Analysis of the information content of aerosol optical parameters

The questions we ask are these.

1. Suppose we have an n dimensional model space. Given m observations ~~of,~~ (e.g., m_1 different parameters at m_2 different wavelengths, so that $m_1 \cdot m_2 = m$), how many independent model variables ~~ℓ~~ $N \leq n$ can we constrain ~~to better than observation error~~with the observations? Obviously, the best we can achieve would be ~~$\ell = m$; but in general,~~ $N = \min\{m,n\}$; but often we will have ~~$\ell \leq m$~~ $N < \min\{m,n\}$.

2. Which are the ~~ℓ~~ N model variables (or linear combinations of model variables) that can be constrained by the measurements?

~~It turns out the the answer to these questionsare found by performing a singular value decomposition of the Jacobian of the observation operator — see Eq. (C6)in the appendix. The right singular vectors can be used to construct a transformation of the model state vector — see~~ Here we only give a summary of the most essential theoretical tools for answering these questions. A more thorough explanation of these concepts is given in the appendix.

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

$$y = x + \epsilon_o. \tag{3}$$

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 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

[2] ~~One should actually~~ Many authors distinguish between data *assimilation* and data *analysis*. ~~The latter refers to post-processing the~~ In data analysis one merely post-processes a model ~~output by statistically weighing model~~ results ~~and~~ by incorporating the information provided by observations. ~~The former refers to a process in which~~ In data assimilation, the data analysis process is ~~incorporated into~~ 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 ~~experiments~~ tests at a fixed point in time. Thus we use the 3DVAR code as a data analysis tool.

can show that the best estimate x_a of x will be

$$x_a = \frac{\sigma_b^2 y + \sigma_o^2 x_b}{\sigma_b^2 + \sigma_o^2}. \tag{4}$$

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 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 measurement mostly contains information on ϵ_o, i.e., it provides a *degree of freedom for noise*.

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

The mapping from model space to observation space given in Eq. (~~D16) in the appendix. The transformed vector components fall into two categories, namely, the signal-related components, which can be constrained by the measurements,~~ 1) can be Taylor-expanded to first order according to

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

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

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

where the matrices $\mathbf{V}_L$ and $\mathbf{V}_R$ contain the left and right singular vectors, respectively, and $\mathbf{W}$ is a matrix that contains the singular values along the main diagonal, while all other matrix elements are zero. It turns out that the ~~noise-related components, which cannot be constrained by the measurements. From the~~ singular values ~~we can~~ w_i can be employed to compute the number of signal degrees of freedom ~~, i.e., the number of model variables that can be constrained to better than observation error. We can further compute how much a set of measurements reduces the Shannon-entropy of the model state. This is a quantitative measure for~~ N_s according to

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

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

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

Both N_s or H allow us to quantify the information content of ~~the measurements.~~

~~Readers who are unfamiliar with~~ a set of measurements. More detailed explanations of these concepts are ~~urged to read the brief introduction~~ given in the appendix. A ~~more complete~~ comprehensive discussion of information aspects and inverse methods for atmospheric sounding can be found in Rodgers (2000).

2.5 ~~Numerical assimilation experiments~~

~~We use the results of this analysis to modify our 3DVAR program. More specifically, we implement weak constraints into~~ By performing the transformation

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

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

$$J_G = \frac{1}{2} \delta \boldsymbol{x}^T \cdot \mathbf{B}^{-1/2} \cdot \mathbf{V}_R \cdot \mathbf{B}_G^{-1} \cdot \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \boldsymbol{x}, \tag{10}$$

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

$$\mathbf{B}_G = \sigma_G \, \mathrm{diag}(w_1, w_2, \ldots, \ldots, w_K, c, \ldots, c). \tag{11}$$

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

2.5 Numerical test of the constrained assimilation algorithm

We study the performance of the 3DVAR system by performing a numerical ~~experiment~~test. To this end, we first perform a reference run by driving the MATCH model with analysed meteorological data. These reference results are taken as the "true" chemical state of the atmosphere. We apply the optics model to the model output to generate synthetic "observations"~~.~~, i.e., a

vertical profile at a selected observation point of extinction and backscattering coefficients at three typical lidar wavelengths. Next we run the MATCH model again, this time driven with 48 hour-forecast meteorological data. The results are taken as a proxy for a background model-estimate that is impaired by uncertainties. Finally, we perform a 3DVAR-analysis of the "observations" and the background estimate in an attempt to restore the reference results. In this numerical ~~experiment~~ test we have perfect knowledge of the true state, and we assume that our optics model is nearly perfect, thus providing nearly perfect observations ~~. (We assumed an~~ (we assumed that the observation error standard deviation ~~of~~ is 10 % of the measurement value).The only factor that may prevent us from fully restoring the reference state is a lack of information in the observed parameters~~on the chemical composition of the aerosol particles~~. Thus, comparison of the retrieval and reference results gives us an indication of how strongly different model variables can be controlled by the information contained in the observations.

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

3 Results

3.1 Analysis of the information content of aerosol optical parameters

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

~~When we increase~~ Table 2 shows the number of ~~wavelengths from one to two, then the number of~~ signal degrees of freedom N_s ~~shows a corresponding increase from 1 to around 1.95–2.00. The change in~~ and the reduction in Shannon entropy H ~~indicates a similar trend; it increases from around 7 for a single wavelength to up to 12 for two wavelengths. (Compare, e.g., cases 1., 2., and 3. to cases 4., 5., and 6.) Hence we almost double the information contained in the measurements. When we increase the number of wavelengths further from two to three, then N_s only increases from around 2.0 to around 2.7 (compare cases 4., 5., and 6. to case 7, or case 9. to case 10.) This is also reflected in H; it only increases from around 12 to 13. This indicates that in our particular case there is little extra information to be gained by extending~~ for different values of the observation standard deviation σ_o. For low values of σ_o, the number of ~~spectral measurements beyond 2–3 wavelengths. Supplementing extinction with backscattering measurements results in a significant increase~~ signal degrees of freedom is identical to the number of observational parameters. However, as we increase σ_o we observe a decrease in N_s ~~and H. This can be seen by comparing, e.g., cases 5. and 11. By adding $\beta_{\rm sca}$ observations to $k_{\rm ext}$ observations the number of~~ . For instance, for

Table 2. ~~Signal~~ Number of signal degrees of freedom N_s and ~~change~~ reduction in entropy H ~~for~~ as a function of observation standard deviation, taken from the lowest model layer (closest to the surface)~~and~~. Results are shown for different subsets of ~~$\{k_{\rm ext}(\lambda_1), k_{\rm ext}(\lambda_2), k_{\rm ext}(\lambda_3), \beta_{\rm sca}(\lambda_1), \beta_{\rm sca}(\lambda_2), \beta_{\rm sca}(\lambda_3)\}$~~$k_1$, ~~where $k_{\rm ext}$ denotes the extinction coefficient~~k_2, ~~$\beta_{\rm sca}$~~ β_1, β_2, β_3, where k_i and β_i represents the extinction and backscattering coefficient, ~~and~~ respectively, at the wavelengths $\lambda_1 = 1064$ nm, $\lambda_2 = 532$ nm, and $\lambda_3 = 355$ nm~~denote the first three Nd:YAG harmonics~~. ~~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 10 %.~~

~~No.~~

~~Parameters~~ Obs. Std. dev. [%]		~~Wavelengths~~		
No.	Parameters	N_s		H ~~1.~~
~~4.~~ height1.	~~$k_{\rm ext}$~~ β_3	λ_2, λ_3 ~~2 4.57 0.95 2.23~~ 1.95 10.1 ~~1 253~~ 1.00	~~7.98~~ ~~5. $k_{\rm ext}$~~ λ_1, λ_3 ~~2 15.0 1.00 3.91~~	2.00 11
~~9.~~ 2.	~~$\beta_{\rm sca}$~~ β_1+β_2	λ_1, λ_3 2.00		~~2~~ 20.6
~~10.~~ 3.	~~$\beta_{\rm sca}$~~ β_1+β_2+β_3	~~λ_1, λ_2, λ_3~~ 3.00		~~2~~ 27.3
4.	β_3+k$_3$	2.00		~~2~~ 19.4
5.	β_1+β_2+k$_2$	3.00		~~2~~ 28.0
~~12.~~ 6.	~~$k_{\rm ext}$, $\beta_{\rm sca}$~~ β_1+β_2+β_3+k$_2$+k$_3$	~~λ_1, λ_2, λ_3~~ 5.00		~~4~~ 40.0

~~6 0.53 0.22 0.18~~ height

$\sigma_o = 100$ % the five parameters β_1+β_2+β_3+k$_2$+k$_3$ (last row) only provide roughly $N_s = 3$ signal degrees of freedom~~increases from 2 to 3.9, so it almost doubles, while~~. The reduction in Shannon entropy H ~~increases from 12 to 18. Case 12 clearly reveals the limitations of extending the set of observed parameters; N_s is only 4.6, significantly lower than the number of observed parameters~~displays an analogous behaviour. For instance, for $\sigma_o = 1$ % we see that H consistently increases as one increases the number of observational parameters. This is much less pronounced for $\sigma_o = 100$ %. In that case, H does increase as one goes from a single parameter to two parameters (compare the first to the second and fourth rows). However, as one adds more parameters, the increase in H slows down considerably. For five parameters (last row), ~~m=6. An inspection of w_i shows that the singular values often display quite a dramatic decrease from the largest to the smallest value. By contrast, the contribution N_s^i to the total number of signal degrees of freedom decreases more gently. This fact is interesting in relation to the choice of the covariance matrix of the weak constrains — see Eqs. (D18), (D19) and (D20)in the appendix. In view of our findings here, we conclude that (D18) would yield a very sharp transition from unconstrained to constrained model variables, Eq. (D19)would give a smooth transition, and Eq. (D20)would give a moderately sharp transition.~~ H is only about twice as high as for a single parameter (first row).

~~We performed a sensitivity study on how the observation error affects the information content in the analysis~~This illustrates the pivotal importance of the observation error for the amount of information that can be obtained from measurements. It is important to understand that the observation error $\boldsymbol{\epsilon}_o$ is not the same as the measurement error $\boldsymbol{\epsilon}_m$. Rather, in our case we have $\boldsymbol{\epsilon}_o = \boldsymbol{\epsilon}_m + \boldsymbol{\epsilon}_f$, where $\boldsymbol{\epsilon}_f$ denotes the forward-model error [see, e.g., Eq. (1) and accompanying text in Rabier et al. (2002)]. Any simplifying assumptions in the optics model or incomplete knowledge of the particle size distribution, morphology, chemical composition, or dielectric properties can contribute to $\boldsymbol{\epsilon}_f$. ~~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 source of error contributes to the observation error. The same is the case if we lack information about the particles' size distribution.~~ Such assumptions ~~also~~ enter into our relatively simple optics model~~, so our previous assumption of an observation standard deviation of 10 % represents, most likely, a highly idealised case~~. [3] Note also that in operational applications there may be other terms contributing to $\boldsymbol{\epsilon}_o$. For instance, if a point measurement is taken at a location that does not provide a good representation of the grid-cell average, then one would have to add a representativity error $\boldsymbol{\epsilon}_r$ to the observation error.

~~To get an idea about the significance of the observation error on the amount of information we can extract from measurements, we consider case 12 in table 2, and we varied the observation standard deviation from 1% to 100%. Table 3 shows how the total entropy and signal degrees of freedom vary with the observation standard deviation. The larger the standard deviation, the less information can be obtained from the observations. Both the total entropy H and the signal degrees of freedom N_s decrease with increasing standard deviation. For a standard deviation of 100 %, we only have two signal degrees of freedom contained in the six observed optical parameters. This demonstrates two important things.~~ The strong impact of the observation errors on the information content of measurements suggests two conclusions.

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

2. It is equally essential to accurately estimate the contribution of the uncertainties in the aerosol optics model~~to the observation error~~, i.e., to estimate the forward-model error $\boldsymbol{\epsilon}_f$. If we underestimate this error, we will rely too much on the measurements than we should, thus assimilating noise. If we overestimate this error, we will waste information contained in the observations. In practice, one way to estimate $\boldsymbol{\epsilon}_f$ is to compute optical properties while varying the particles' size, morphology, and dielectric properties within typical ranges. The resulting variation in the optical properties then allows us to estimate $\boldsymbol{\epsilon}_f$. (For a review of aerosol optics modelling see Kahnert et al. (2014, 2016) and references therein).

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

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

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

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

3.2 ~~Numerical inverse-modelling experiment~~

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

~~We integrated the findings of 3.1 into our~~

3.2 Numerical inverse-modelling test

We integrated the findings of 3.1 into our 3DVAR ~~program by constraining the algorithm to varying only the~~ program by constraining the algorithm to varying only the signal-related model variables. ~~We employed the weak-constraint approach described in the appendix. More specifically, the constraints are formulated by use of the ansatz given in Eq. (D20), where we set the constant $c'' = \min\{w_k, 0.1\}$, and where w_k is the smallest singular value of the scaled observation operator — see Eq. (C6).~~

To illustrate the method we conduct a numerical ~~experiment~~ test as described in Sect. 2.5. We perform a 3DVAR analysis ~~of the background field by assimilating four different vertical profilesof optical properties, namely, the backscattering coefficient,~~

~~Signal degrees of freedom N_s and change in entropy H as a function of observation standard deviation, taken from the first model layer and case 12 in table 2.~~

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 %.

~~Obs. std. dev. (%)~~ No.	Parameters	i	w_i	N_s^i	H_i	N_s	H
1.	β_3	1	~~5.96~~ 38.2	~~37.6~~ 5 1.00	~~5.29~~ 5.26	~~24.3~~ 1.00	5.26
~~10~~ 2.	~~4.61~~ β_1, β_2	1	108	1.00	6.76	1.97	9.36
		~~19.3~~ 2	6.00	0.97	2.61		
~~50~~ 3.	~~3.00~~ $\beta_1, \beta_2, \beta_3$	1	115	1.00	6.84	2.71	10.5
		2	6.54	0.98	2.73		
~~100~~	~~2.33~~	~~7.8~~ 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		

β_{bak}~~, and the extinction coefficient,~~ by assimilating "$3\beta + 2\alpha$" profiles, i.e., synthetic lidar measurements of β_{sca} at the three wavelengths 1064, 532, and 355 nm together with k_{ext} ~~, each at a wavelength of~~ at the two wavelengths 532 and 355 ~~and 1064~~ nm. Thus in our case the number of singular values in each vertical layer is ~~$k = 4$~~ $K = 5$. We assume an idealised situation in which the observation standard deviation is only 10 %. As we see in Table 2 (~~row 11~~case No. 6), the number of ~~singular values larger than unity is $\ell = k = 4$, and the number of~~ signal degrees of freedom is ~~$N_s = 3.9$~~ $N_s = 4.9$ in this case. So we roughly have as many signal degrees of freedom as we have measurements.

~~As an example, Fig. 1 shows the ammonium sulphate mixing ratio in the lowest model layer (closest to the surface) computed for the reference run (left), the background estimate (centre~~ Figure 1 shows vertical profiles of selected aerosol components, namely (from top to bottom): organic carbon (OC) in the 3rd size bin (OC-3), OC in the 4th size bin (OC-4), elemental carbon (EC) in the 3rd size bin (EC-3), and ~~the~~ mineral dust in the 1st size bin (DUST-1). The reference and background mixing ratios

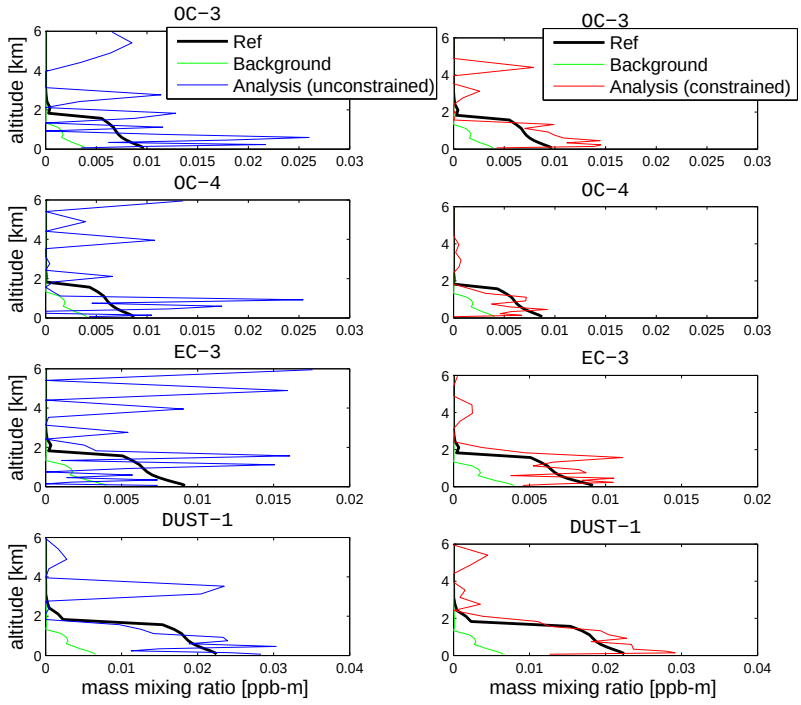


Figure 1. Vertical profiles of selected aerosol components in different size bins. From top to bottom: organic carbon in the 3rd size bin (OC-3), OC in the 4th size bin (OC-4), elemental carbon in the 3rd size bin (EC-3), and dust in the 1st size bin (DUST-1). The reference results are shown in black,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

are shown in black and green, respectively. The 3DVAR ~~analysis (right)[4]. Clearly,~~ analysis was first performed without any constraints; the results are shown in the left column by the blue line. Then the 3DVAR analysis was repeated with the constraints in Eq. (10) and (11); the results are represented in the right column by the red line. Clearly, the unconstrained analysis (blue lines in the ~~background field underestimates the reference field in most areas. We picked a location in Central Sweden (60°N, 15°E) to compute backscattering and extinction profiles from the reference results , which were then 3DVAR-analysed in conjunction with the background field. The analysis (right) restores the reference mixing ratios at and near the observation site. So, at least for ammonium sulphate mixing ratios in the lowest model layer we seem to obtain a satisfying solution to the inverse modelling problem.~~ left panels) yields results that oscillate quite erratically in the vertical direction. Also, the unconstrained analysis can yield conspicuously high values at higher altitudes, even though both the reference and background values are both close to zero. By contrast, the constrained analysis (red lines in the right panels) yields results that better agree with the

[4] ~~The approach for generating a reference and background model-run has been explained in Sect. 2.5~~

reference results. The noisiness in the vertical direction is significantly reduced, and the results at higher altitudes are generally lower than those obtained with the unconstrained analysis.

~~A closer inspection of the analysis performance is given in Fig. 2 . Each panel shows vertical profiles of mixing ratios at the observation site. We compare the analysis results (red solid line) to both the background estimate (blue dashed line)and the reference results (black solid line). The reference results of the secondary inorganic aerosol species~~ Figure 2 shows analogous results for the mass mixing ratios of different aerosol components, each summed over all size bins. The aerosol components are (from top to bottom): elemental carbon (EC), organic carbon (OC), mineral dust (DUST), sea salt (NaCl), secondary inorganic aerosols (SIA, ~~which is the sum of~~ i.e., the sum over all sulphate, nitrate, and ammonium ~~mixing ratios)are almost completely restored by the 3DVAR analysisat all altitudes. For elemental carbon (EC), organic carbon (OC), dust, and, even more so, sodium chloride (NaCl) the analysis overestimates~~ species), and PM10 (i.e., the sum over all aerosol components). Clearly, the constrained analysis faithfully retrieves both PM10 and SIA. The unconstrained analysis performs almost equally well for these two variables. Sea salt and mineral dust are not well retrieved from the measurements in either the constrained or unconstrained approach. EC and OC are very well retrieved by the constrained analysis. For these components, the unconstrained analysis has a very small bias compared to the reference results~~at altitudes between 0–2 km, while above 2 km the reference results are at least partially restored by the analysis. When we compare the different scales on the x-axes, we see that SIA makes the dominant contribution to the aerosol mixing ratio. Accordingly, the total aerosol mass mixing ratio (PM_{10}) is almost equally well restored by the analysis as the SIA mixing ratio. Vertical profiles of elemental carbon (EC), organic carbon (OC), dust, secondary inorganic aerosols (SIA), sea salt, and total aerosol mass mixing ratio (PM_{10}). Each panel shows the reference results (black solid line), background estimate (dashed blue line), and the 3DVAR analysis (red solid line).~~ , but it is considerably more noisy (i.e., oscillating in the vertical direction) than the constrained analysis. We also see, again, that the mixing ratios at higher altitudes obtained with the unconstrained analysis can be unreasonably high. This is especially pronounced for OC. In general, however, the problems we encounter in the unconstrained analysis are less pronounced in Fig. 2 than in Fig. 1. A possible explanation is that SIA may be most strongly related to the measurement signal, and SIA is dominating the aerosol mass in this case. We will return to this point shortly. Another possible factor is that the noise in the analysis can be damped by summing up results over several size bins.

Figure 3 shows the observations (black~~solid line~~) as well as the observation-equivalents of the background estimate (~~blue dashed line~~green) and the unconstrained (blue) and constrained (red) 3DVAR ~~analysis (red solid line) for all four observations, namely, β_{bak} at 355 wavelength (top left), β_{bak} at 1064 (top right), k_{ext} at 355 (bottom left), and k_{ext} at 1064 (bottom right)~~analysis for all five observations. We learn from this figure that the analysis follows the observations faithfully. The reason for this is that we assumed that the observations were highly accurate with an error standard deviation of only 10 %. In fact, the difference between the observation-equivalent analysis and the observations deviate by even less than 10 %. However, our tests confirmed that an increase in the observation error eventually results in analysis results of which the observation-equivalent increasingly deviates from the observations (not shown).

We have seen that the analysis provides a reasonable, but, as expected, not a perfect answer to the inverse problem. We have further seen that at ~~(and near)~~ the observation site it relies more on the observations than on the background estimate. Most

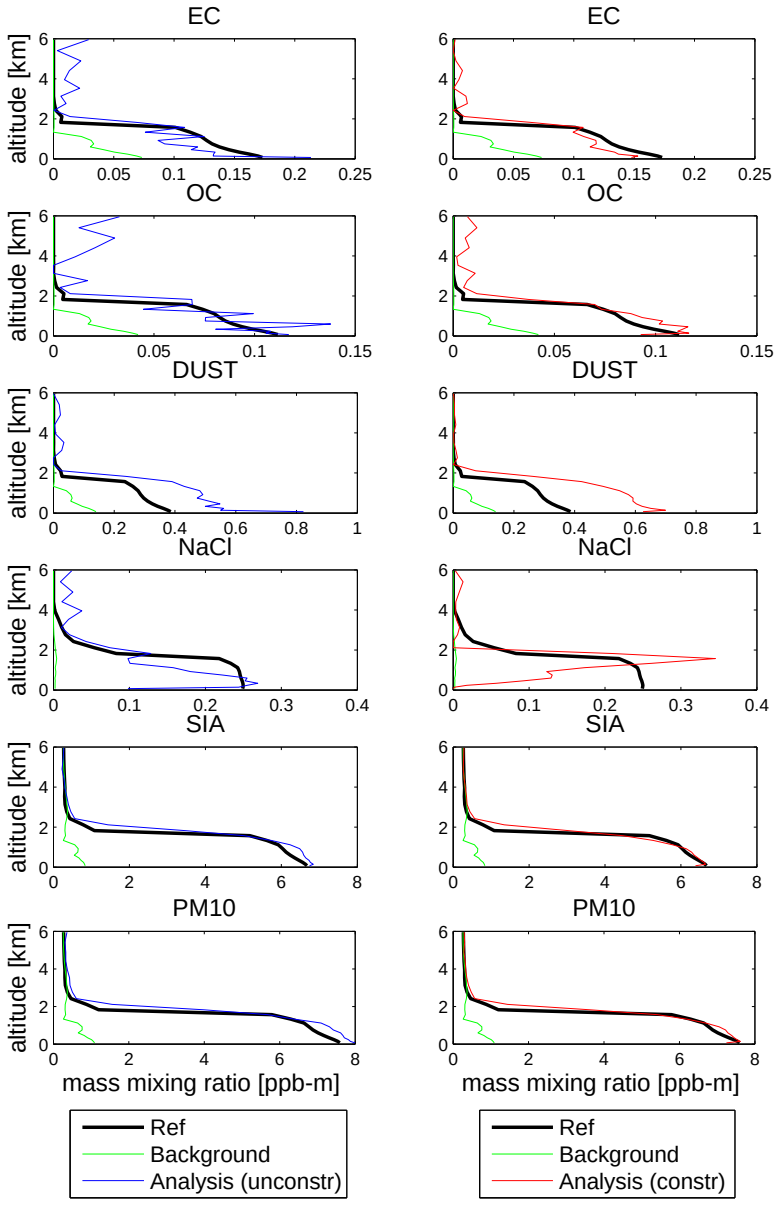


Figure 2. ~~Ammonium sulphate~~ As Fig. 1, but for the total mass mixing ratio (summed over ~~Europe~~all size bins). ~~Left~~The components are (from top to bottom): ~~reference field~~EC, ~~centre: background field~~OC, ~~right: 3DVAR analysis~~mineral dust, sea salt, secondary inorganic aerosols (sum of all sulphate, nitrate, and ammonium species), and PM10 (sum of all aerosol components).~~The observation site is indicated by the white circle. Note the nonlinear colour scale!~~

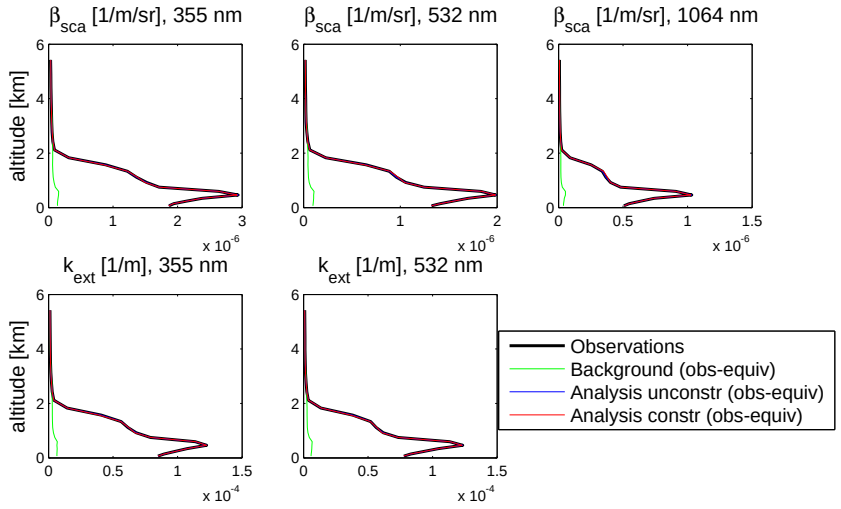
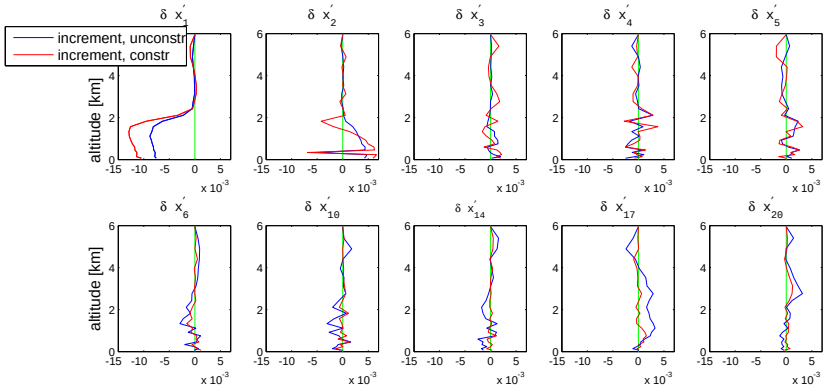


Figure 3. Observations (black solid line), and observation-equivalents of the background estimate (~~dashed blue line~~green), and of the ~~3DVAR analysis~~ unconstrained (blue) and constrained (red~~solid line~~) 3DVAR analysis. The optical parameters and wavelengths are indicated above each panel.

importantly, we have seen that the constraints introduced in the 3DVAR algorithm suppress noise in the analysis, especially in EC and OC. However, the previous figures ~~tell us little about the effect of the constraints we introduced~~do not provide us with any direct insight of how exactly the constraints accomplish this. To learn more about that we need to inspect the analysis in the abstract phase space of the transformed model variables $\delta\boldsymbol{x}'$ ~~given~~ . (Recall that we defined this variable in Eq. (~~C16)~~ 9) as $\delta\boldsymbol{x}' = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot (\boldsymbol{x} - \boldsymbol{x}_b)$). Figure 4 shows vertical profiles of ~~all~~ a selection of the, in total, 20 variables $\delta x_i'$. The ~~error variance within which each of these variables is allowed to vary in the analysis is given by the diagonal elements of the matrix $\mathbf{B}_C$ in Eq. (D20) . The first four of these are the singular values given in Table 2 (row11). The remaining 16 variances are set to 0.1. Thus the first element $\delta x_1'$ has by far the largest freedom to be adjusted by the~~ background estimate corresponds to $\delta x_i' = 0$ and is represented by the green line. The unconstrained 3DVAR analysis increment is represented by the blue line, the constrained 3DVAR ~~algorithm. The error variance of the second element,~~ analysis increment is shown by the red line. The first five phase space elements in the top row are the signal-related control variables. Generally, the magnitude of the constrained increments (red) is larger than that of the unconstrained increments (blue). The noise-related phase space elements, five of which are shown in the bottom row, display the opposite behaviour. The constrained increments are close to zero, as they should. The unconstrained elements consistently show higher magnitudes than the constrained elements. However, we also see that the unconstrained analysis does produce increments that are largest for the two elements $\delta x_1'$ and $\delta x_2'$~~, is smaller by roughly one order of magnitude. The error variances of $\delta x_3'$ and $\delta x_4'$ are of comparable magnitude, and each one is about one order of magnitude smaller than that of $\delta x_2'$. Finally,~~ , which most strongly relate to the ~~error variances of the remaining 16 elements are about one order of magnitude smaller than that of $\delta x_4'$. Accordingly, the first element, $\delta x_1'$, is the one that deviates most~~



~~Vertical profiles of the transformed model variables $\delta \boldsymbol{x}'$.~~

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

~~strongly from zero. The elements $\delta x_2'$, $\delta x_3'$, and $\delta x_4'$ are varied much less in comparison. If we had imposed strong constraints, then the remaining elements would be exactly zero. However, our weak-constraint formulation allows even the other elements to deviate from zero within relatively tight limits. But several of them are, in fact, very close to zero, notably the elements $\delta x_i'$ for $i = 7$–9, 11–14, 16, and 19.~~ measurement signal. Based on our single test case we cannot say if this is a lucky coincidence or a consistent property. If the latter, it may indicate that we are using rather reasonable background error statistics, so that the analysis increment in observation space is distributed to the different variables in model space in a sensible way. If the former, it could be the case that the success of the unconstrained analysis is largely dependent on whether or not those aerosol components dominate the total aerosol mass that most strongly relate to the signal degrees of freedom. (In our case the total mass is dominated by SIA, which is very well retrieved by the analysis).

Finally, we want to obtain a better understanding of how the aerosol components $\boldsymbol{x}$ in model space, or their increments $\delta \boldsymbol{x}$, are linked with the signal-related phase-space elements $\delta \boldsymbol{x}'$. To this end we inspect the first five row vectors of the transformation matrix $\mathbf{V}_R^T \cdot \mathbf{B}^{-1/2}$ in Eq. (9). The magnitude of these elements can be taken as a measure for how much each aerosol component of $\delta \boldsymbol{x}$ in model space contributes to the signal-related elements of $\delta \boldsymbol{x}'$. Figure 5 shows $|\,(\mathbf{V}_R^T \cdot \mathbf{B}^{-1/2})_{ij}\,|$ for $i = 1, \ldots, 5$, and for $j = 1, \ldots, 20$, where 5 is the number of signal-related phase-space elements, and 20 is the number of aerosol components in model space. Results are shown for model layers 2 (left column) and 22 (right column), which correspond to altitudes of about 100 m and 6 km, respectively. The x-axis shows sea salt (NaCl), EC, OC, and dust, each in four size bins, as well as the four SIA components, i.e., sulphates (SOX) other than $(NH_4)_2SO_4$, ammonium sulphate (AS), ammonium nitrate (AN), and nitrates (NOX) other than NH_4NO_3.

Comparison of the two columns clearly demonstrates that the elements of the transformation matrix can vary considerably with vertical layer (or, more generally, with location). This is because the error covariance matrix $\mathbf{B}$ varies with location, and the matrix $\mathbf{R}$ varies from one observation site to another (in our case, from one altitude to another). Hence the matrix $\mathbf{V}_R$ is

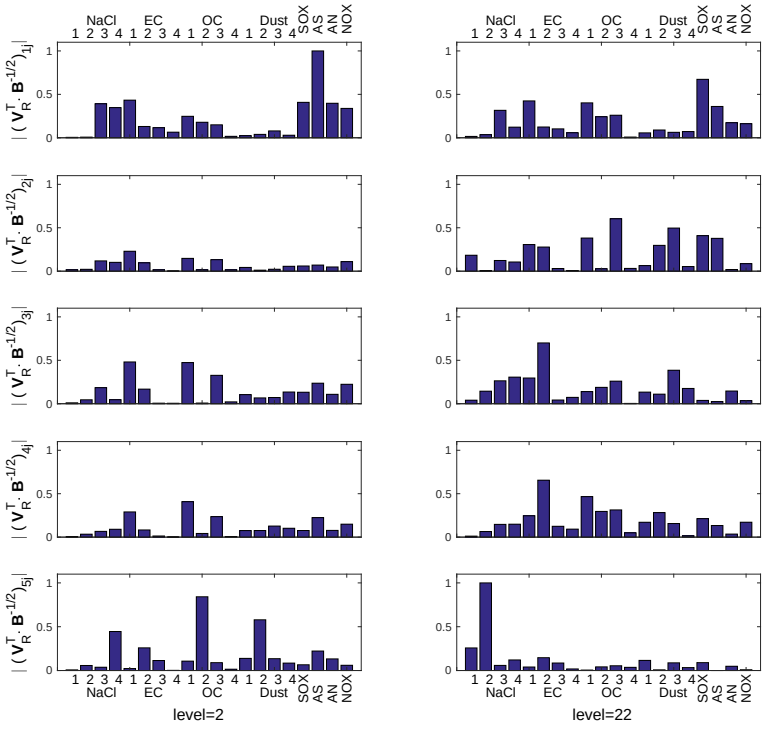


Figure 5. The first five rows (from top to bottom) of the matrix $\mathbf{V}_R^T \cdot \mathbf{B}^{-1/2}$ at the observation site, and for model layers 2 (left) and 22 (right). The y-values are normalised by dividing them by the maximum element. The x-axis indicates the aerosol components in model space 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 $(NH_4)_2SO_4$, ammonium sulphate (AS), ammonium nitrate (AN), and nitrates (NOX) other than NH_4NO_3.

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

However, in our case the SIA components consistently make a strong contriubution to the first signal-related element $\delta x_1'$. Since SIA is dominating the aerosol mass mixing ratio in this test case, the analysis was able to retrieve PM10. We also see that the dust components make only a weak contribution to most of the signal-related elements $\delta x_i'$, especially to the first one. This is a likely explanation for the difficulties encountered in retrieving the dust mass mixing ratio. Sea salt is more complicated. Size bins 3 and 4 do contribute considerably to $\delta x_1'$, and also to some of the other four increments, while size bins 1 and 2 do not make a significant contribution to most of the five signal-related control variables. In our case the sea salt mass is strongly dominated by the second size bin (not shown). This explains the difficulties we encountered in the retrieval of sea salt.

4 Summary and conclusions

We have quantified the information content of ~~extinction and backscattering~~ multiwavelength lidar measurements with regard to the chemical composition of aerosol particles. ~~This has been done~~ Different combinations of extinction and backscattering observations at several wavlengths have been investigated by determining the singular values of the scaled observation operator, by computing the number of signal degrees of freedom N_s, and by calculating the ~~change in Shannon-entropy~~ reduction in Shannon entropy H caused by taking measurements. We first ~~assumed a relatively low~~ quantified N_s and H as a function of observation standard deviation σ_o. The information content of ~~10 %. In that case, when adding measurements of β_{bak} to measurements of k_{ext}, the information content nearly doubles. The same is true when we increase the number of wavelengths from a single wavelength to two wavelengths. However, when we further increase the number of wavelengths from two to three, then the gain in information is rather modest; there appeared to be little use in increasing the number of optical wavelengths beyond three. When the full set of six observations(both optical parameters at three wavelengths) is considered, then the number of signal degrees of freedom is 4.6. Thus we can constrain, at most, 4–5 model variables with this set of observations.~~ the observations, as expressed by N_s and H, decreased as σ_o was increased. This became the more pronounced the larger the number of simultaneously observed parameters was.

~~These conclusions depend, to be sure, on the assumed observation standard deviation. We therefore performed a sensitivity study were we investigated how the observation standard deviation affects the information content. We observed a rather dramatic decrease in both the entropy and signal degrees of freedom with increasing observation standard deviation. Note that not only~~ The observation error depends not only on the measurement error, but also ~~the~~ on the forward-model error. The latter depends on the uncertainties in the ~~aerosol optics modelcontribute to the observation error~~aerosol-optics model. This highlights the importance of developing accurate aerosol optics models and of obtaining an accurate estimate of the observation error, especially of the uncertainty in the aerosol optics model. This is a prerequisite for extracting as much information as possible from the measurements, while avoiding to extract noise rather than signal. More often than not, computational limitations and lack of knowledge force us to introduce simplifying assumptions about the particles' morphologies. However, we know that aerosol optical properties can be highly sensitive to the shape (Mishchenko et al. (1997); Kahnert (2004)), small-scale surface roughness (Kahnert et al., 2012b), inhomogeneity (Mishchenko et al., 2014; Kahnert, 2015), aggregation (Fuller and Mackowski, 2000; Liu and Mishchenko, 2007; Kahnert and Devasthale, 2011), irregularity (Muinonen, 2000; Bi et al., 2010), porosity (Vilaplana et al., 2006; Lindqvist et al., 2011; Kylling et al., 2014), and combinations thereof (Lindqvist et al., 2009; Kahnert et al., 2013; Lindqvist et al., 2014). We need to know how much these sources of uncertainty contribute to the observation standard deviation. One way of estimating this is to compare aerosol optical properties computed with simple shape models to either measurements or to computations based on more realistic particle shape models — see ~~Kahnert et al. (2016)~~ Kahnert et al. (2014) for a recent review and a more detailed discussion.

~~We exploited our analysis of the information content and the number of signal degrees of freedom by formulating weak constraints in a 3DVAR algorithm. More specifically, we transformed the model variables into a new basis~~ The singular values of the scaled observation operator provide us with an abstract measure to compare the standard deviations of the background

(prior) estimate to those of the observations. The reason why this is a rather abstract measure is because background and observation errors are, in general, in different spaces and cannot be directly compared. However, we constructed a mapping that transforms the state vector in physical (model) space to an abstract phase space in which the components of the state vector can be ~~divided~~ partitioned into signal-related and noise-related components. ~~We then added weak constraints to~~ The singular values indicate to what extent the signal-related phase-space variables can be constrained by the measurements. We exploited this fact by constructing weak constraints in a 3DVAR data assimilation code, which limited the assimilation algorithm ~~in such a way that only~~ to acting on the signal-related ~~transformed model variables are varied by the 3DVAR analysis . Numerical experiments showed that the 3DVAR algorithm provided a reasonable solution to the inverse problem; when~~ phase-space variables only (hereafter referred to as the *constrained analysis*). The idea was to maximise the use of information, while avoiding the risk of assimilating noise by over-using the measurements. Thus, our main hypothesis was that the constrained analysis will yield less noisy results than the unconstrained analysis. Numerical tests confirmed this hypothesis. Notably in the case of elemental carbon (EC) and organic carbon (OC) the unconstrained analysis gave mixing ratios that oscillated considerably in the vertical direction. The constrained analysis results were considerably less noisy.

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

In our specific test case secondary inorganic aerosol components were most faithfully retrieved by the inverse modelling solution~~. Most importantly, it was demonstrated that the 3DVAR analysisfollows, indeed, the imposed constraints; the transformed model variables are adjusted within certain limits according to how strongly they relate to the signal degrees of freedom~~, followed by organic and black carbon. Dust and seasalt mass mixing ratios were more challenging to retrieve. We could explain this by inspecting the linear coefficients in the transformation from physical space to the abstract phase space. We found that those aerosol components that had the largest weight in the transformation were most faithfully retrieved by the analysis. However, these linear coefficients depend on the background error covariances (which can change with location), and on the observation error variances. Therefore, it is difficult to draw general conclusions about which aerosol components are most easily retrieved by a given set of measurements.

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

may reduce the observation error. The present study ~~should~~ could be extended to investigate the information contained in extinction and backscattering measurements for simultaneously constraining the chemical composition and the size of aerosol particles.

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

Appendix A: Inverse problems

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

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

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

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

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

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

Appendix B: ~~3-dimensional~~ Three-dimensional variational data assimilation

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

In a CTM we discretise the geographic domain of interest into a ~~3-dimensional~~ three-dimensional grid. In each grid cell, the aerosol particles are characterised by the mass ~~concentrations~~ mixing ratio of each chemical component in the aerosol phase, such as sulphate, nitrate, ammonium, mineral dust, black carbon, organic carbon, and sea salt. Suppose we summarise all these mass ~~concentrations~~ mixing ratios from all grid cells into one large vector $\boldsymbol{x} \in \mathbb{R}^n$. The model provides us with a first guess of the atmospheric aerosol state, known as a *background estimate* [4] $\boldsymbol{x}_b$.[4] Suppose also that we have m observations, which we summarise in a vector $\boldsymbol{y} \in \mathbb{R}^m$. We further have an observation operator $\hat{H} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{x} \mapsto \hat{H}(\boldsymbol{x})$ that maps the state vector $\boldsymbol{x}$ from model space to observation space[5]. We further denote by $\boldsymbol{x}_t$ the true state of the atmosphere, by $\boldsymbol{\epsilon}_b = \boldsymbol{x}_t - \boldsymbol{x}_b$ the error of the background estimate, and by $\boldsymbol{\epsilon}_o = \hat{H}(\boldsymbol{x}_t) - \boldsymbol{y}$ the observation error.[6] ~~If the background errors are not correlated with the observation errors, then~~ The background and observation errors are assumed to be unbiased and uncorrelated with each other. Then their joint probability distribution becomes separable, i.e.

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

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

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

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

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

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

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

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

[6] ~~The~~ We stress, once more, that the observation error must not be confused with the measurement error $\boldsymbol{\epsilon}_m$. The latter contributes to the former, but the observation error 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 ~~source of~~ 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 $\boldsymbol{\epsilon}_o$.

given a background estimate $\boldsymbol{x}_b$ with error covariance matrix $\mathbf{B}$. Similarly, $P_o(\boldsymbol{x})$ is the probability that the system is found in state $\boldsymbol{x}$, given measurements $\boldsymbol{y}$ with error covariances $\mathbf{R}$.[7]

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

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

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

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

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

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

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

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

$$J = J_b + J_o \tag{B7}$$

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

$$J_o(\delta\boldsymbol{x}) = \frac{1}{2}\Big(\hat{H}(\boldsymbol{x}_b)+\mathbf{H}\cdot\delta\boldsymbol{x}-\boldsymbol{y}\Big)^T\cdot\mathbf{R}^{-1}\cdot\Big(\hat{H}(\boldsymbol{x}_b)+\mathbf{H}\cdot\delta\boldsymbol{x}-\boldsymbol{y}\Big) \tag{B9}$$

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

The solution to the equation ~~$\nabla J=0$~~ $\nabla_{\boldsymbol{x}} J=\mathbf{0}_n$ is a solution to the inverse problem (where $\mathbf{0}_n$ denotes the null vector in n-dimensional model space); we input the observations $\boldsymbol{y}$ into the algorithm, and as output we obtain a result in model space that is consistent with the measurements (within the given error bounds).[8] What if the measurements contain insufficient

[7] The 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.

[8] By solving the equation $\nabla J|_{\boldsymbol{x}=\boldsymbol{x}_a}=\mathbf{0}_n$ for the analysed state $\boldsymbol{x}_a$ it can be shown that the solution to the inverse problem is given by $\boldsymbol{x}_a=\boldsymbol{x}_b+\mathbf{K}\cdot(\boldsymbol{y}-\hat{H}(\boldsymbol{x}_b))$, 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

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

Appendix C: Information content of measurements

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

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

When we account for observation errors $\boldsymbol{\epsilon}_o$, then the basic relation between model variables and observations is, to first order

$$\boldsymbol{y} = \hat{H}(\boldsymbol{x}_b) + \mathbf{H} \cdot \delta\boldsymbol{x} + \boldsymbol{\epsilon}_o. \tag{C1}$$

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

background estimate $\boldsymbol{x}_b$ by mapping the increment $(\boldsymbol{y} - \hat{H}(\boldsymbol{x}_b))$ 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.

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

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

$$\delta\tilde{\boldsymbol{x}} = \mathbf{B}^{-1/2}\cdot\delta\boldsymbol{x} \tag{C2}$$

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

$$\tilde{\mathbf{H}} = \mathbf{R}^{-1/2}\cdot\mathbf{H}\cdot\mathbf{B}^{1/2}. \tag{C4}$$

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

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

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

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

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

Here $\tilde{\mathbf{H}}$ is a $(m\times n)$-matrix, the matrix of the left-singular vectors $\mathbf{V}_L$ is a $(m\times m)$-matrix, the matrix $\mathbf{V}_R$ containing the right-singular vectors is a $(n\times n)$-matrix, and the $(m\times n)$-matrix $\mathbf{W}$ consists of two blocks. If $m<n$, then the left block of $\mathbf{W}$ is a $(m\times m)$-diagonal matrix containing the m singular values $w_1,\ldots,w_m$ on the diagonal; the right block is a $(m\times(n-m))$~~-nullmatrix~~-null matrix. Similarly, if $m>n$, then the upper block of $\mathbf{W}$ is a $(n\times n)$-diagonal matrix containing the n singular values on the diagonal, while the lower block is a $((m-n)\times n)$~~-nullmatrix~~-null matrix.

We now make another change of variables:

$$\delta\boldsymbol{x}' = \mathbf{V}_R^T\cdot\delta\tilde{\boldsymbol{x}} \tag{C7}$$

$$\delta\boldsymbol{y}' = \mathbf{V}_L^T\cdot\delta\tilde{\boldsymbol{y}} \tag{C8}$$

$$\mathbf{H}' = \mathbf{V}_L^T\cdot\tilde{\mathbf{H}}\cdot\mathbf{V}_R. \tag{C9}$$

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

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

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

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

$$\mathbf{H}' \cdot \mathbf{H}'^T = \mathbf{W} \cdot \mathbf{W}^T, \tag{C11}$$

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

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

One can compute the number of signal degrees of freedom N_s from the expectation value of J_b in Eq. (B8). The result can be expressed in terms of the singular values w_i of the ~~transformed observation operator in Eq. (C6)~~observability matrix:

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

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

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

$$S(P) = -\int P(x) \ln \log_2 \left(\frac{P(x)}{P_0(x)} \right) \mathrm{d}x, \tag{C13}$$

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

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

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

$$H = \frac{1}{2} \sum_i \ln_{i=1}^{\min\{m,n\}} \log_2(1+w_i^2). \tag{C15}$$

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

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

$$\delta \boldsymbol{x}' = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \boldsymbol{x}. \tag{C16}$$

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

Appendix D: Minimisation of the ~~costfunction~~ cost function with constraints

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

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

D1 Minimisation of the ~~costfunction~~ cost function with strong constraints

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

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

then one solves the minimisation problem

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

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

minimisation problem becomes

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

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

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

$$J = \frac{1}{2}\left(\delta\boldsymbol{x}^T \cdot \mathbf{Q}_1 \cdot \delta\boldsymbol{x} + \mathbf{Q}_2^T \cdot \delta\boldsymbol{x} + \delta\boldsymbol{x}^T \cdot \mathbf{Q}_2 + \mathbf{Q}_3\right) \tag{D5}$$

$$\mathbf{Q}_1 = \mathbf{B}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \tag{D6}$$

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

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

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

$$\mathbf{G} \cdot \delta\boldsymbol{x} = \mathbf{0}_k, \tag{D9}$$

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

$$\mathbf{Z} = \begin{pmatrix} \boldsymbol{z}_1 & \cdots & \boldsymbol{z}_{n-k} \end{pmatrix} \tag{D10}$$

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

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

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

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

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

1. Determine a basis of the ~~nullspace~~ null space $\ker(\mathbf{G})$; this yields the matrix $\mathbf{Z}$.

2. Solve the unconstrained $(n-k)$-dimensional optimisation problem given in Eq. (D12). From the $(n-k)$-vector $\boldsymbol{\xi}$ that minimises the ~~costfunction~~ cost function in (D11), we then obtain the solution $\delta\boldsymbol{x} = \mathbf{Z}\cdot\boldsymbol{\xi}$ that minimises the ~~costfunction~~ cost function in (D5) subject to the constraint (D9).

D2 Minimisation of the ~~costfunction~~ cost function with weak constraints

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

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

$$J = J_b + J_o + J_G \tag{D13}$$

$$J_G = \frac{1}{2}\delta\boldsymbol{x}^T\cdot\mathbf{G}^T\cdot\mathbf{B}_G^{-1}\cdot\mathbf{G}\cdot\delta\boldsymbol{x}, \tag{D14}$$

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

$$\nabla_{\delta\boldsymbol{x}} J_G = \mathbf{G}^T\cdot\mathbf{B}_G^{-1}\cdot\mathbf{G}\cdot\delta\boldsymbol{x}. \tag{D15}$$

We will assume that the matrix $\mathbf{B}_G$=diag$(\sigma_1^G,\ldots,\sigma_k^G)$ is diagonal, where k is the number of constraints. The "error variances" 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 ith constraint is relatively strong, and vice versa. ~~The choice of these variances is a matter of experimenting and tuning.~~ Typically, if the σ_i^G are made too large, then there is a risk that the minimisation algorithm ignores the constraints all together. In that case the solution will be very similar to the unconstrained solution. On the other hand, if the 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 minimisation routine largely ignores the observations and returns a solution that lies quite close to the background estimate.

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

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

If we were to formulate this requirement as a strong constraint, as in Eq. (D9), then it would take the form

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

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

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

In order to apply the weak-constraint approach, we need to substitute the constraint-matrix $\mathbf{G} = \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2}$ into Eq. (D14), which yields

$$J_G = \frac{1}{2} \delta \boldsymbol{x}^T \cdot \mathbf{B}^{-1/2} \cdot \mathbf{V}_R \cdot \mathbf{B}_G^{-1} \cdot \mathbf{V}_R^T \cdot \mathbf{B}^{-1/2} \cdot \delta \boldsymbol{x}, \tag{D17}$$

where $\mathbf{B}_G$ is a $(n \times n)$ matrix. We want to set up this matrix in such a way that we obtain a smooth transition from freely adaptable control variables $\delta x'_1, \ldots \delta x'_\ell$ to increasingly constrained variables ~~$\delta x'_{\ell+1}, \ldots \delta x'_k, \ldots, \delta x_n$~~ $\delta x'_{\ell+1}, \ldots \delta x'_k, \ldots, \delta x'_n$. One possible choice of the matrix $\mathbf{B}_G$ ~~, which is suggested by Eq. (C11),~~ would be

$$\mathbf{B}_G = \sigma_G \, \mathrm{diag}(w_1^2, w_2^2, \ldots, w_\ell^2, \ldots, w_k^2, c, \ldots, c), \tag{D18}$$

where σ_G is a free scaling factor, and where the last $n-k$ diagonal elements are equal to a constant ~~c~~ $c \ll w_k$ chosen to be much smaller than ~~w_k^2. Another possible~~ the smallest singular value w_k.

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

$$\mathbf{B}_G = \sigma_G \cdot \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_\ell, \ldots, \lambda_k, c', \ldots, c'),$$

$$\lambda_i = w_i^2/(1+w_i^2),$$

$$\mathbf{B}_G = \sigma_G \, \mathrm{diag}(w_1^p, w_2^p, \ldots, w_\ell^p, \ldots, w_k^p, c, \ldots, c), \tag{D19}$$

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

$$\mathbf{B}_G \;=\; \sigma_G \cdot \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_\ell, \ldots, \mu_k, c, \ldots, c), \tag{D20}$$

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

where $c \ll \mu_k$. This ansatz is suggested by Eq. (C12), i.e., each of the elements $\delta x'_1, \ldots \delta x'_k$ is weighted with its corresponding contribution to the number of signal degrees of freedom. ~~It turns out that Eq. (D18) gives a relatively sharp transition from unconstrained to constrained model variables, while~~ We tested all three approaches (the one in Eq. (D19) ~~gives a very gentle transition. Another ansatz that lies in between these two extremes would be~~

$$\mathbf{B}_G = \sigma_G \,\mathrm{diag}(w_1, w_2, \ldots, w_\ell, \ldots, w_k, c'', \ldots, c''),$$

~~where $c'' \ll w_k$.~~

~~Despite the mathematical foundation of this approach we are left with some room for experimentation in~~ for $p = 2$). These tests showed that the ~~formulation of the matrix $\mathbf{B}_G$. It is a matter of experience to test different approaches and select the one that proves to be most suited.~~ different approaches often yield analysis results that are quite similar. However, in each approach the free parameters σ_G and c are tuned to different values. If they are not well tuned, then the analysis tends either toward the background estimate or toward the unconstrained analysis, as explained earlier in the text following Eq. (D15).

Appendix E: Practical aspects of the implementation

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

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

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

In our case we are primarily interested in constraining the ~~chemical~~ aerosol components. Therefore, we formulate our weak constraints in a suitable subspace of the physical space. Suppose, for simplicity, that we have reduced all data to the ~~best solution for us is to simply restrict ourselves to the chemical subspace. To this end, we select a grid-point~~[12] ~~(i,j,l) and consider the~~

[12] ~~Since the error correlations are assumed to be homogeneous in space any point in the horizontal direction will do.~~

vertical resolution of our model. Let $\nu_l = 1, \ldots, m_l$ label all measurements that lie within model layer l. Suppose further than (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 observation ν_l). Consider the reduced background error covariance matrix ~~$\mathbf{B}^0$ with components $B^0_{k;k'} = B_{i,j,l,k;i,j,l,k'}$, where~~ with elements $B^{(\alpha,l)}_{k,k'} = B_{i_\alpha j_\alpha l k i_\alpha j_\alpha l k'}$, $k, k' = 1, \ldots, N_c$~~Similarly, we consider reduced matrices $\mathbf{R}^0$, $\mathbf{H}^0$,~~, and $\tilde{\mathbf{H}}^0 = (\mathbf{R}^0)^{-1/2} \cdot \mathbf{H}^0 \cdot \mathbf{B}^0$, and we numerically compute the~~ N_c is the number of aerosol components. Consider further the reduced observability matrix with elements $\tilde{H}^{(l)}_{\nu_l,k} = \sum_{k'=1}^{N_c} R^{-1/2}_{\nu_l,\nu_l} H_{m i_\alpha j_\alpha l k'} \{(B^{(\alpha,l)})^{1/2}\}_{k',k}$, where $m = m(l, \nu_l)$ labels the ν_lth observation in model layer l. Analogous to Eq. (C6), we now perform a singular value decomposition ~~of the latter. From this we obtain, for each level l, a constraint matrix $(\mathbf{V}^0_R)^T \cdot (\mathbf{B}^0)^{-1/2}$ analogous to the one~~ in the reduced space

$$\tilde{H}^{(l)}_{\nu_l,k} = \sum_{s=1}^{\min\{m_l, N_c\}} (V^{(l)}_L)_{\nu_l,s} w^{(l)}_s (V^{(l)}_R)_{k,s}. \tag{E1}$$

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

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

We now invoke the assumption that the constraints computed at the observation site are also valid at neighbouring points, i.e., we apply the constraint matrix given in Eq. (~~D16) , but reduced to the chemical subspace. We then simply apply these chemical constraints throughout the horizontal domain (again, assuming horizontal homogeneity).~~E2) in Eq. (D17) according to

$$J_G = \frac{1}{2} \sum_{ijlkk's} \delta x_{ijlk'} G_{s,k'} (B^{-1}_G)_s G_{s,k} \delta x_{ijlk}, \tag{E3}$$

where $(B_G)_s$ denotes the diagnonal elements of the matrix given in (D18).[12].

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

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

reduce the dimension of the matrix $\mathbf{B}$ by working in the reduced space in which only the covariances $\mathbf{B}^{(\alpha,l)}$ among aerosol components are considered. One could compute the matrix $(\mathbf{B}^{(\alpha,l)})^{-1/2}$ in Eq. (D17) by diagonalising the matrix $\mathbf{B}^{(\alpha,l)}$. However, a numerically much more efficient approach is to not work with positive square root, but with the so-called Cholesky decomposition[13] of the B-matrix,

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

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

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

Acknowledgements. This work was funded by the Swedish National Space Board through ~~the OSCES project (no.~~ project nos. 100/16 (MK) and 101/13 (EA).

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

References

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

Andersson, C., Bergström, R., Bennet, C., Robertson, L., Thomas, M., Korhonen, H., Lehtinen, K. H. J., and Kokkola, H.: MATCH-SALSA — Multi-scale Atmospheric Transport and CHemistry model coupled to the SALSA aerosol microphysics model — Part 1: Model description and evaluation, Geosci. Model Dev., 8, 171–189, 2015.

Andersson, E. and Kahnert, M.: Coupling aerosol optics to the MATCH (v5. 5.0) chemical transport model and the SALSA (v1) aerosol microphysics module, Geosci. Model Dev., 9, 1803–1826, 2016.

Benedetti, A., Morcrette, M. J.-J., Boucher, O., Dethof, A., Engelen, R. J., Huneeus, M. F. H. F. N., Jones, L., andS. Kinne, J. W. K., Mangold, A., Razinger, M., Simmons, A. J., and Suttie, M.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, D13 205, 2009.

Bi, L., Yang, P., Kattawar, G., and Kahn, R.: Modeling optical properties of mineral aerosol particles by using nonsymmetric hexahedra, Appl. Opt., 49, 334–342, 2010.

Bocquet, M.: Toward optimal choices of control space representation for geophysical data assimilation, Mon. Wea. Rev., 137, 2331–2348, 2009.

Burton, S. P., Hair, J. W., Kahnert, M., Ferrare, R. A., Hostetler, C. A., Cook, A. L., Harper, D. B., Berkoff, T. A., Seaman, S. T., Collins, J. E., Fenn, M. A., and Rogers, R. R.: Observations of the spectral dependence of particle depolarization ratio of aerosols using NASA Langley airborne High Spectral Resolution Lidar., Atmos. Chem. Phys., 15, 13 453–13 473, 2015.

Burton, S. P., Chemyakin, E., Liu, X., Knobelspiesse, K., Stamnes, S., Sawamura, P., Moore, R. H., Hostetler, C. A., and Ferrare, R. A.: Information ~~Content and Sensitivity of the $3\alpha + 2\beta$ Lidar Measurement System for Aerosol Microphysical Retrievals~~content and sensitivity of the $3\beta+2\alpha$ lidar measurement system for aerosol microphysical retrievals, Atmos. Meas. ~~Tech.Discuss., 2016, ,~~ Techniques, 9, 5555–5574, 2016.

Cardinali, C., Pezzulli, S., and Andersson, E.: Influence-matrix diagnostic of a data assimilation system, Q. J. R. Meteorol. Soc., 130, 2767–2786, 2004.

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

Fuller, K. A. and Mackowski, D. W.: Electromagnetic scattering by compounded spherical particles, in: Light scattering by nonspherical particles, edited by Mishchenko, M. I., Hovenier, J. W., and Travis, L. D., pp. 226–273, Academic Press, San Diego, 2000.

Johnson, C., Hoskins, B. J., and Nichols, N. K.: Very large inverse problems in atmosphere and ocean modelling, Int. J. Numer. Meth. Fluids, 47, 759–771, 2005a.

Johnson, C., Nichols, N. K., and Hoskins, B. J.: A singular vector perspective of 4D-Var: Filtering and interpolation, Q. J. R. Meteorol. Soc., 131, 1–19, 2005b.

Joiner, J. and da Silva, A. M.: Efficient methods to assimilate remotely sensed data based on information content, Q. J. R. Meteorol, Soc., 124, 1669–1694, 1998.

Kahnert, F. M.: Reproducing the optical properties of fine desert dust aerosols using ensembles of simple model particles, J. Quant. Spectrosc. Radiat. Transfer, 85, 231–249, 2004.

Kahnert, M.: Variational data analysis of aerosol species in a regional CTM: background error covariance constraint and aerosol optical observation operators, Tellus, 60B, 753–770, 2008.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

McKeen, S., Chung, S. H., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Gong, W., Bouchet, V., Moffet, R., Tang, Y., Carmichael, G. R., Mathur, R., and Yu, S.: Evaluation of several PM2.5 forecast models using data collected during the ICARTT/NEAQS 2004 field study, J. of Geophys. Res., 112, d10S20, 2007.

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

Mishchenko, M. I., Cairns, B., Kopp, G., Schueler, C. F., Fafaul, B. A., Hansen, J. E., Hooker, R. J., Itchkawich, T., Maring, H. B., and Travis, L. D.: Accurate Monitoring of Terrestrial Aerosols and Total Solar Irradiance: Introducing the Glory Mission, Bull. Am. Met. Soc., 88, 677–691, 2007.

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

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

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

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., Kuehn, R. E., and Liu, Z.: The CALIPSO automated aerosol classification and lidar ratio selection algorithm, J. Atmos. and Ocean. Technol., 26, 1994–2014, 2009.

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

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

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

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

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 for a size-resolved sectional model: impacts of observationally constrained, multi-wavelength and fine mode retrievals on regional scale analysis and forecasts, Atmos. Chem. Phys., 13, 10 425–10 444, 2013.

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

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

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

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

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

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

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

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

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

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