We thank the reviewers for their positive reception of our work, insightful comments, and valuable suggestions that helped to improve this paper.

The replies to both reviews are given below. We do not discuss small technical or typesetting remarks and typos spotted by reviewers here, those were simply applied as recommended. The original referee comments are indented, excerpts from the revised version of the paper are given in italic.

1 Reply to Referee 1

1.1 High level comments

> My only high-level concern is that the end results, based on simulations for GLORIA, do not necessarily show the instrument or technique off at its best. In part, as I see it, this is because of the tight nature of the a priori constraints imposed, specifically the 0.645 K a priori precision estimate for temperature. As an aside, although the choice of the associated correlation lengths is well described in the manuscript, I was unable to find discussion of how the 0.645K number was arrived at (apologies if I missed it). A consequence of this very tight constraint is that the instrument has to "work very hard" to add useful information, as witness the similarity between the magnitude of the fields in figures 6 and 4 to those in the bottom row of figure 3. In other words, the instrument hasn't been "allowed" to tell you that much more than you already knew.

The following explanation was added to section 4.3: *Our choice of σ value for temperature, in particular, may seem rather low. This is related to the choice of a priori: smoothed ECMWF data was used instead of climatologies that would have been more typical in this case. Such an a priori can be expected to match the large scale structures of real temperature more closely, thus reducing the forward model errors in poorly resolved areas and improving the results. It is particularly useful for resolving fine structures, such as gravity waves, which were the main scientific interest of the measurment flight used as a basis for test retrievals [Krisch et al., 2017]. A likely closer match between a priori and retrieval thus requires a lower σ value, that, in this setup, represents expected strength of gravity wave disturbances, rather than full thermal variability of the atmospheric region in question.*

It must also be noted that the σ value is not an hard limit for deviations from a priori. A gravity wave with an amplitude of 3-3.5 K, which is considered to be among 1% of strongest gravity wave events expected in the measured region, was succesfully retrieved with this setup [Krisch et al., 2017].

1.2 Details

> Another common thread in my more minor comments is the author's use of the word "accuracy" to describe quantities that I believe are more commonly referred to as "precision".

This was corrected in several places, as suggested.

> Finally, the figures, while arguably numbered logically are not introduced in the text in that order. The current order has 3 first then, 6, 1a, 4, 5, then 2.

The figures and references were rearranged in the revised paper, now order of figure references in the text matches their numerical order. The old Figure 3 hence became Figure 1, bringing it closer to its first reference, and the old Figures 1 and 2, that were strongly related, were joined into a single figure. The color scale of the old Figure 3 was also changed, as suggested.

> Page 2, Line 3: "quality" is rather a loose term. Consider clarifying it further (e.g., information content, effective resolution, faithfulness to "true" atmosphere, etc.).

Adding detail to this, admittedly vague, statement would repeat the ideas at the end of the paragraph (or those in the previous paragraph). Hence the sentence was simply removed.

> Page 2, Line 6: Is the "classic Tikhonov" approach necessarily first order? I know of some teams using second-order Tikhonov routinely.

This statement was made more specific in the revised paper: *In this paper, we introduce an improved, more physically and statistically motivated approach to regularisation, that requires less tuning. It relies on the calculation of both the first spatial derivative and Laplacian of atmospheric quantities, which provide more complete information about smoothness and feasibility of a particular atmospheric state.* We are not aware of anyone else in the atmospheric community using a combination of first order derivatives and Laplacian for 2D or 3D problems.

> Page 3, Line 27: I'm not sure that characterizing the alpha terms as "unphysical" is that fair. One could easily scale them with some suitable length term to cast them into constraints on spatial variability (which is more or less what you yourselves are doing). Also, such scaling would arguably render them less "grid dependent".

One could, of course, use a physical length instead of each α parameter. That would also allow to, for example, change the grid constant of a regular grid and retain the regularisation strength. Such an approach would, however, result in a large number of tunable parameters (especially if derivatives of higher order than one are to be used). Estimating all of them from model or observation data would be difficult without establishing some relations between them. Our approach provides one way to obtain such relations. Also, a regularisation scheme derived in the form of an integral, as in equation (13) of the revised paper, is relatively easy to adopt to different kinds of grids (e.g. irregular). This gives much more "grid independence" than a mere rescaling that one could do while preserving the form of Tikhonov matrices.

> Page 4, Line 19: The term "covariance kernel" was new to me. Is it widely used? Is there some reference for it?

Covariance kernel, also known as covariance function (https://en.wikipedia.org/wiki/Covariance_function) is just a real function that describes a spatial covariance of the (continuum) random variable. In general, if one defines an operator X by means of an integral of a real (or complex) function over the whole space in question, it is common in applied mathematics to call this function an "X kernel". This is why we chose this term here. It was also used in [Lim and Teo, 2009]. A remark about covariance kernel being the same as the covariance function was added to the revised paper.

> Page 4 Equation 5: Doesn't C^{-1} need a _k subscript?

No. The scalar product (and hence the norm) associated with the covariance operator is induced by volume integral of the inverse of the operator, not the inverse of it's kernel. If one would represent a scalar field $\phi(\mathbf{r})$ by a state vector $\mathbf{x}_\phi$ on some grid, and similarly for $\varphi(\mathbf{r})$ and $H\mathbf{x}_\varphi$, then the discrete equivalent to the integral in equation (5) would be $\mathbf{x}_\phi^T \mathbf{S}_a^{-1} \mathbf{x}_\varphi$. Hopefully the more explicit statement about the equivalence between continuous and discrete formulations (equation (6) in the revised manuscript) will make things clearer for the readers.

> Page 5, Line 23 and 25: Not sure how, if L_h is constant (line 23) you can have "different correlation lengths...". Are you implying L_h only has to be horizontally homogeneous, not vertically constant?

The equation (12) (10 in the original manuscript) holds if, as it was stated in lines 20-21, the correlation lengths are functions of altitude only: $L_h = L_h(z)$, $L_v = L_v(z)$. If, in addition, L_h is constant (i.e. $L_h = \text{const.}$, $L_v = L_v(z)$), then the algebra is further simplified, as was described in lines 23-24.

> Page 10, Line 28: I like the term "Precision matrix" because, as with the word itself, "more" implies "better". However, is it in common use?

Precision, the reciprocal of variance, and precision matrix, the inverse of the covariance matrix, seem to be standard statistical terms (https://en.wikipedia.org/wiki/Precision_(statistics)).

> Discussion starting around line 11, and through the remainder of the section: It's not clear to me how the matrix M in the lines 11-16 relates to the matrix A in equation 22 and beyond. Also, I don't believe A is the Rodgers-style "Averaging kernel" matrix in this context is it (as

Rodgers' A is not s.p.d.)? Again, I'd suggest using a different letter than A. If is intended to be a measure of S_a, as the discussion on Line 30 (page 11) seems to imply, then why not use S_a itself, perhaps with some suitable diacritic (a tilde, or hat or something)? Apologies if I'm missing a critical point here.

The notation and organisation of section 3 was found to be confusing by both reviewers. Therefore, the section was significantly reorganised and notation changed. Please refer to the section 3 in the revised manuscript for details.

2   Reply to Referee 2

2.1   General comments

It should be clarified to what extent the new approaches are relevant for the more standard 1D and 2D inversions. Most importantly, can 1D and 2D regularisation matrices be constructed in the same manner?

Applicability to 1D and 2D retrievals is discussed in the newly added Appendix B.

The new way to construct the regularisation matrix is presented as an extension of Tikhonov regularisation, but I rather see it as a way to approximate the precision matrix of Equation 1? In any case, the approach bridges the gap between Bayesian and Tikhonov regularisation. This is important and should be stressed (depending a bit on if your approach works for 1D and 2D). To be clear, you argue that the simplest way to set the Tikhonov regularisation matrix is to consider statistics of the atmosphere. You use correlation structures, but that approach is essentially identical to the Bayesian approach. That is, you basically argue that the Bayesian approach is to prefer. I am trying to provoke here, on purpose, to encourage you to extend the discussion and clarify the nice link you provide between Bayesian and Tikhonov regularisation.

A remark about the similarity to Bayesian approach was added, and some statements in section 2.2 clarified.

Title: Should be changed. The present title is too generic, especially considering that the manuscript doesn't involve any real observations. The manuscript deals only with technical improvements of the retrieval step. On the other hand, I don't see any reason to make a restriction to "limb sounding" and "airborne", the new methods are relevant for any 3D observations (also 1D and 2D?).

The title was changed to *3D tomographic limb sounder retrieval techniques: irregular grids and Laplacian regularisation.*

Abstract: I find the abstract vague. See comment about Delaunay triangulation above. Some hard facts would be nice. For example, what reduction in the number of grid points was achieved?

Explicit mention of Delaunay triangulation and the achieved reductions in computational cost were added to the abstract. The new version reads: *Multiple limb sounder measurements of the same atmospheric region taken from different directions can be combined in a 3D tomographic retrieval. Mathematically, this is a computationally expensive inverse modelling problem. It typically requires an introduction of some general knowledge of the atmosphere (regularisation) due to its underdetermined nature. This paper introduces a consistent, physically motivated (no ad-hoc parameters) variant of the Tikhonov regularisation scheme based on spatial derivatives of first order and Laplacian. As shown by a case study with synthetic data, this scheme, combined with irregular grid retrieval methods employing Delaunay triangulation, improves both upon the quality and the computational cost of 3D tomography. It also eliminates grid dependence and the need to tune parameters for each use case. The few physical parameters required can be derived from in situ measurements and model data. Tests show that 82% reduction in the number of grid points and 50% reduction in total computation time, compared to*

previous methods, could be achieved without compromising results. An efficient Monte Carlo technique was also adopted for accuracy estimation of the new retrievals.

> As a concrete example, despite Section 2.2 is detailed it is not yet clear to me, after several readings, how the result of Equation 11 shall be used to actually construct the regularisation matrix. Maybe I miss something obvious but I don't see how the result of Eq 11 shall be used to generate S−1a.

Explanations of how to use the equation (11) for obtaining the precision matrix have been added. Refer to equations (6) and (14) in the revised manuscript and the paragraphs above them.

> The nomenclature should also be revised. For example, the symbol used on the left hand side of Equation 11 is not defined. Or rather, I assume it's should the same as in Equation 9. Further, it's unlucky to use y in Equation 14, as y represents the observation in Equation 1.

Φ instead of ϕ in equation 11 was indeed a typo. The notation in most of the section 2.4 was revised not to cause confusion with equation (1).

> With respect to Eq.2 and related text: I don't know how Tikhonov formulated the approach (and it was introduced independently by others), and your reference may be correct. On the other hand, I don't think it is fair to say that Tikhonov regularisation is today restricted to consider the first derivative. For example, the description of Tikhonov regularisation in "Numerical recipes" says "... measures of smoothness that derive from first or higher derivatives." An example from the atmospheric field where the second derivative was considered: Steck, Tilman. "Methods for determining regularization for atmospheric retrieval problems." Applied Optics 41.9 (2002): 1788-1797. The text can be interpreted as that considering the second derivative is novel, which is not true.

This statement was made more specific in the revised paper: *In this paper, we introduce an improved, more physically and statistically motivated approach to regularisation, that requires less tuning. It relies on the calculation of both the first spatial derivative and Laplacian of atmospheric quantities, which provide more complete information about smoothness and feasibility of a particular atmospheric state.* We are not aware of anyone else in the atmospheric community using a combination of first order derivatives and Laplacian for 2D or 3D retrievals.

> Section 3: This section has similar problems as Sec 2.

The notation and organisation of section 3 was found to be confusing by both reviewers. Therefore, the section was significantly reorganised and notation changed. Please refer to the section 3 in the revised manuscript for details.

> Section 4.1: Title and content do not agree. The section deals also with the forward model.

The title was changed to *The GLORIA instrument and data processing*

> Section 4.4: A demonstration of the new features is, of course, nice to see, and maybe even a demand. However, showing results for a single retrieval case does not prove much. Statistics of an ensemble of retrievals are required to judge if one retrieval is better than another one. For a single case, specifics of the case can make the poorer method to look better. Further, it is also very unclear how "optimal" the regularisation weights used in A actually are? Anyhow, can really optimal weights be found by manually tuning? There are in fact objective methods for setting the weights.

One of the aims of this paper is to progress towards a robust and standardised regularisation technique, that could be employed to perform 3D retrievals in various atmospheric conditions without much manual effort. The authors are aware of some of the objective methods for setting the weights for the classic Tikhonov approach, but they, so far, did not prove sufficient to achieve this goal. Therefore, adequately testing both regularisation methods on a large and representative set of atmospheric states and performing appropriate statistical analysis on the results was deemed to be beyond the scope

of this paper. A well-understood test case was used as an example instead. Performing statistical analysis of in situ or model data to determine correlation lengths and standard deviations of various atmospheric quantities, as well as further testing of retrieval methods on larger data sets is an option for further work.

> Some of the new features can be tested in a more direct manner, compared to doing full retrievals. For example, a basic demand when selecting a grid is that discretisation errors are kept at a sufficiently small level. That is, for me, the first test when introducing a new grid scheme (here Delaunay) is to simply compare forward model simulations and check that results only change in a tolerable way. This test is most critical for D. By the way, is the same simulated measurement inverted in A to D (presumably based on A)? If a new simulated measurement is done for each case, then a possible discretisation errors are swept under the carpet.

The authors are very grateful for pointing this out, this important part of test retrieval description was indeed overlooked. The following was added to section 4.2: *The same set of simulated measurements was used for all test retrievals described in this paper. These measurements were obtained by running the forward model on a very dense grid (about twice as dense in each dimension as those of the densest test retrievals). This was done to ensure that the discretisation errors in the simulated measurements would be minimal and would not give any one retrieval an advantage (as it could happen, if they were generated on the same grid as this test retrieval).* Also, a new appendix (Appendix C) with a comparison of forward model output in the case of each retrieval was added to the revised paper.

> In my mind, the most interesting question in the manuscript is how well the calculations actually manage to estimate the exponential covariance assumed (Equation 7)? Is it possible to derive/estimate the Sa implied by the derived S−1 a , and check how well the obtained Sa follows the start assumptions? Either for a sub-volume or a smaller test case. My interpretation of Section 2.4 is that you ensure S−1a to be positive definite and it should then be invertible (for a reasonable large case).

We have come up with two ways to test whether the precision matrix constructed as in equation (14) indeed represents the exponential covariance assumption (8). One way is to construct a covariance matrix directly for a relatively small grid, invert it, and compare the result with the directly obtained precision matrix. Matrix norms are poorly suited for such a comparison, since small test cases are prone to significant edge effects (finite differences perform poorly on grid boundaries, among other issues). Therefore, we simply constructed some test "atmospheric states" – Gaussian wave packets – for the small grid and compared their norms based on the two precision matrices obtained in different ways. This comparison was added to the revised paper as an Appendix A.

An alternative approach would be to construct the precision matrix, invert it, and see if the matrix elements satisfy the exponential relation. These results are harder to interpret: covariance matrix elements do not directly enter any retrieval calculations, it is hard to quantify what errors are acceptable in this case. We will therefore just provide such a comparison here.

We use 20x20x20 regular grid with grid constant equal to 1 in each direction. We set $\sigma = 1$ and consider two cases: $L_h = L_v = L = 2$ and $L = 3$. For each case, a precision matrix is obtained as a discretisation of equation (13) for rectilinear grid. Then the precision matrix is inverted to obtain a covariance matrix. One point near the centre of the grid, the reference point, is selected, and then covariances between every grid point and the reference point are read from the covariance matrix and plotted, as a function of distance between grid point and reference point, in Figure 1 of this document. The plots show the theoretical covariance value as a blue line. Also, the points with distance from reference point lower or equal to 7 were used to perform a least squares fit for covariance function of the form $C_k(\mathbf{r}) = \sigma^2 \exp\left(-\|\mathbf{r}\|/L\right)$ with respect to the parameters σ^2 and $1/L$. The fits are shown as green lines in Figure 2.1. Best fit was achieved with $\hat{\sigma} = 1.05$, $\hat{L} = 1.89$ in the $L = 2$ case and $\hat{\sigma} = 1.04$, $\hat{L} = 2.66$ in the $L = 3$ case. Hence, in these cases, the precision matrix constructed with our method is in good agreement with theoretical assumptions. One can see from the figures, however, that agreement is worse near the edges of the grid. This is not a surprise, since derivatives, upon which the precision matrix construction is based, cannot be properly evaluated at grid boundaries. The results would also look worse for large L values, those would need a larger grid to properly test.
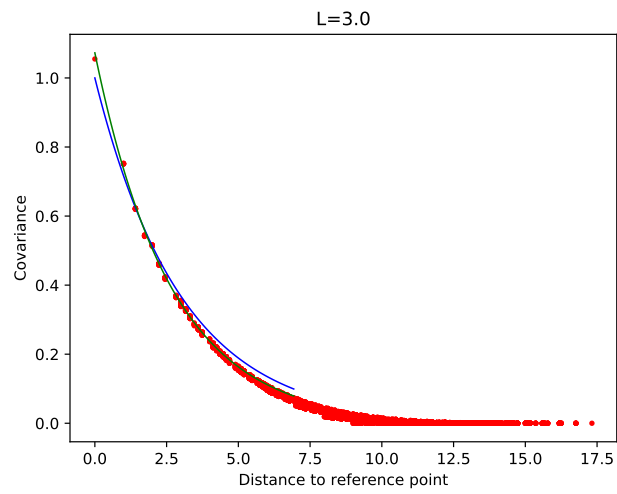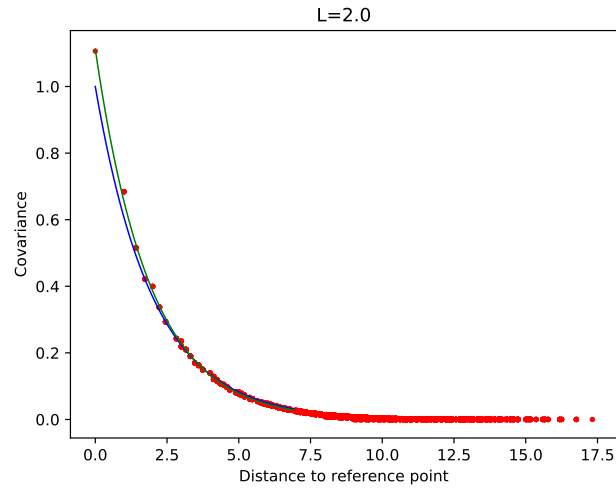
Figure 1: Covariance to the reference point as a function of distance to the reference point for correlation lengths $L = 2$ and $L = 3$. Blue line – theoretical covariance value, green line – least squares fit for exponential covariance.

Conclusions: Should be extended a bit. Are all problems solved? Or something lacking to attack real observations? Are the new methods applicable in other cases

The following was added to conclusion: *At the time of writing, the methods developed in this paper were already in use to process GLORIA limb sounder data.* Better approximations for regularisation parameters could help to improve the retrievals in the future.

2.2 Details

Page 1, line 8, and elsewhere: It should be considered how the word "accuracy" is used. Accuracy equals systematic error or at least includes this term.

This was corrected in several places.

Page 1, lines 22-23: This is not a specific 3D issue.

We were not trying to claim that it is. To make this more explicit, the word "typically" was replaced with: *as it often happens with remote sensing retrievals.*

Page 2, line 3: The choice of regularisation constraint does affect the output of the retrieval, but I don't agree that it changes the quality. The regularisation methods are mathematical tools, and, assuming that there is no numerical issues or similar problems, they simply optimize what you have told them to do. That is, if you change regularisation constraint, you select to optimize another metrics, and the result will differ. But can it be claimed generally that one metrics is better than another one? I would say that it depends on the application.

The (admittedly vague) sentence about retrieval quality was removed.

Sec 2.3: I assume you are using some kind of external library to derive the Delaunay triangulation. Which one should be specified? Any other libraries that should be mentioned?

Delaunay triangulation was generated using CGAL (the Computational Geometry Algorithms Library). This statement was added to section 2.3. CGAL was the only major library that needed to be added to JURASSIC2 code to implement the new methods described in this paper. For more information about JURASSIC2 and other libraries used by this software, refer to [Hoffmann et al., 2008, Ungermann et al., 2011, Ungermann, 2013].

Page 9, line 26: I don't agree that this is in general a difficult problem. As you point out, Rodgers (2000) explains how it should be done. Basically, all retrievals come with an error estimation, so it is, in general, a feasible task.

The beginning of the paragraph was condensed to clarify the main point: theoretical derivations for detailed error estimates do exist, but are usually prohibitively computationally expensive for large problems. In these cases, one needs to resort to other means, such as Monte Carlo. The revised formulation: *Estimating the precision of remote sensing data products generated by means of inverse modelling is essential for the users of the final data and also valuable for evaluation and optimisation of the inverse modelling techniques. Detailed quantitative descriptions of data accuracy can be derived in theory (see validation in section 4.6 and Rodgers (2000)) but they are, in case of large retrievals, too numerically expensive to calculate in practice.*

Page 13, line 15: The figures shall be introduced in order. You start with Figure 3.

The figures and references were rearranged in the revised paper, now order of figure references in the text matches their numerical order. The old Figure 3 hence became Figure 1, bringing it closer to its first reference, and the old Figures 1 and 2, that were strongly related, were joined into a single figure.

Page 16, lines 3 and 13: Join these two comments about weights, to more clearly describe what has been done.

Descriptions of the regularisations for retrievals A and B were extended slightly. For the case of A: *[Regularisation weights] are typically tuned ad-hoc (adjusted by trial-and-error until optimal retrieval results can be achieved), validated against model data and, when using real observations, in situ measurements.* For the case of B: *The grid and interpolation methods were identical to retrieval A, but the regularisation was replaced with the second order scheme from equation (12) and the correlation lengths derived in section 4.3 from in situ observations. No subsequent tuning of these parameters was performed.*

      Page 18, lines 30-33: I don't follow the explanation. Why can't you compare apples with apples? If this is not possible, there is little value in the exercise.

A Monte Carlo run that can be compared "apples to apples" with the retrieval results shown has been added to the revised manuscript (Figure 5 Ds). This description of the old and newly added Monte Carlo results should clarify what has been done: *Results for retrievals B and D are presented in Figure 5 B, D. They show the total error estimate for temperature based on the sensitivity of temperature value to (simulated) instrument noise and all other retrieved atmospheric quantities as well as expected error of non-retrieved quantities (e.g. pressure). These results tell us what temperature errors we could expect if the retrievals in question were performed with real data. The panel Ds shows an error estimate based on (simulated) instrument noise and uncertainties of retrieved parameters only. In the case of a synthetic retrieval, unretrieved parameters, upon which the simulated measurements are based, are known exactly. Therefore, panel Ds is a direct Monte Carlo prediction of the temperature deviation we see in Figure 3D, and the magnitude of this deviation is indeed similar.*

References

[Hoffmann et al., 2008] Hoffmann, L., Kaufmann, M., Spang, R., Müller, R., Remedios, J. J., Moore, D. P., Volk, C. M., von Clarmann, T., and Riese, M. (2008). Envisat mipas measurements of cfc-11: retrieval, validation, and climatology. *Atmospheric Chemistry and Physics*, 8(13):3671–3688.

[Krisch et al., 2017] Krisch, I., Preusse, P., Ungermann, J., Dörnbrack, A., Eckermann, S. D., Ern, M., Friedl-Vallon, F., Kaufmann, M., Oelhaf, H., Rapp, M., Strube, C., and Riese, M. (2017). First tomographic observations of gravity waves by the infrared limb imager gloria. *Atmospheric Chemistry and Physics*, 17(24):14937–14953.

[Lim and Teo, 2009] Lim, S. and Teo, L. (2009). Generalized whittle–matérn random field as a model of correlated fluctuations. *Journal of Physics A: Mathematical and Theoretical*, 42(10):105202.

[Ungermann et al., 2011] Ungermann, J., , J., Lotz, J., Leppkes, K., Hoffmann, L., Guggenmoser, T., Kaufmann, M., Preusse, P., Naumann, U., and Riese, M. (2011). A 3-D tomographic retrieval approach with advection compensation for the air-borne limb-imager GLORIA. *Atmos. Meas. Tech.*, 4(11):2509–2529.

[Ungermann, 2013] Ungermann, J. (2013). Improving retrieval quality for airborne limb sounders by horizontal regularisation. *Atmospheric Measurement Techniques*, 6(1):15–32.

~~Tomographic airborne~~ 3D tomographic limb sounder ~~retrievals on retrieval techniques:~~ irregular ~~grid with second order~~ grids and Laplacian regularisation

Lukas Krasauskas[1], Jörn Ungermann[1], Stefan Ensmann[1], Isabell Krisch[1], Erik Kretschmer[2], Peter Preusse[1], and Martin Riese[1]

[1]Institute of Energy and Climate Research – Stratosphere (IEK-7), Research centre Jülich GmbH, Jülich, Germany
[2]The Institute of Meteorology and Climate Research (IMK), Karlsruhe Institute of Technology, Karlsruhe, Germany

Correspondence to: L. Krasauskas (l.krasauskas@fz-juelich.de)

Abstract. Multiple limb sounder measurements of the same atmospheric region taken from different directions can be combined in a 3D tomographic retrieval. Mathematically, this is a computationally expensive inverse modelling problem. It typically requires an introduction of some general knowledge of the atmosphere (regularisation) due to its underdetermined nature.

This paper introduces a consistent, physically motivated (no ad-hoc ~~unphysical~~ parameters) variant of the Tikhonov regu-
5 larisation scheme based on spatial derivatives of first ~~and second order~~ order and Laplacian. As shown by a case study with synthetic data, this scheme, combined with irregular grid retrieval methods employing Delaunay triangulation, improves both upon the quality and the computational cost of 3D tomography~~, while eliminating~~. It also eliminates grid dependence and the need to tune parameters for each use case. The few physical parameters required can be derived from in situ measurements and model data. Tests show that 82% reduction in the number of grid points and 50% reduction in total computation time, compared
10 to previous methods, could be achieved without compromising results. An efficient Monte Carlo technique was ~~adopted for retrieval accuracy estimation~~ also adopted for accuracy estimation of the new retrievals.

1 Introduction

Dynamics and mixing processes in the upper troposphere and lower stratosphere (UTLS) are of great interest. They control the exchange between these layers (Gettelman et al., 2011) and have a strong influence on the composition and thereby on
15 radiative forcing (Forster and Shine, 1997; Riese et al., 2012). High spatial resolution observations are required to understand the low length-scale processes typical of this region.

Infrared limb-sounding is an important tool for measuring the temperature and volume mixing ratios of trace gases in UTLS with high resolution, especially in the vertical direction, which is critical for resolving typical structures there (e.g. Birner, 2006; Hegglin et al., 2009). Modern limb sounders are capable of high frequency operation resulting in many measurements
20 taken close ~~from~~ to one another, which are best exploited if the data from these measurements can be combined. Producing a 2D curtain of an atmosphere along the flight path of the instrument by combining the observations close to each other is already

an established technique (e.g. Steck et al., 2005; Worden et al., 2004). With suitable measurement geometries, this technique can be extended to obtain 3D data.

A tomographic data retrieval uses multiple measurements of the same air mass, taken from different directions, to obtain high resolution 3D temperature and trace gas concentration data. There is, as it often happens with remote sensing retrievals, a large number of atmospheric states that would agree with the given observations within their expected precision. A regularisation algorithm is employed to pick the solution that is in best agreement with our prior knowledge of the atmospheric state and general understanding of the physics involved. The mathematical framework of the tomographic retrieval is outlined in section 2.1.

The classic Tikhonov regularisation scheme (Tikhonov and Arsenin, 1977) is an established technique used for infrared limb sounding instruments such as the Michelson Interferometer for Passive Atmospheric Sounding (Steck et al., 2005; Carlotti et al., 2001)), GLORIA (Ungermann et al., 2010; Ungermann et al., 2011). It quantifies the spatial continuity of an atmospheric state by evaluating first order spatial derivatives of the retrieved quantities. This technique serves the purpose of ruling out very pathological, oscillatory retrieval results, but requires fine tuning of several unphysical ad-hoc parameters and subsequent validation. In this paper, we introduce an improved, more physically and statistically motivated approach to regularisation, that requires less tuning. It relies on the calculation of both the first spatial derivative and Laplacian of atmospheric quantities, which provide more complete information about smoothness and feasibility of a particular atmospheric state. The regularisation parameters we use are physical, in the sense that they can be estimated from in situ measurements or model data, and less grid dependant: their values, once established, can be more readily used for new retrievals.

Data retrievals from limb sounders are typically performed on a rectilinear grid. In this paper, we define rectilinear grid by taking a set of longitudes, a set of latitudes and a set of altitudes and placing a grid point at each of the possible combinations of these coordinates. Such grids can be a limiting factor for efficiency of numerical calculations. Due to the exponential nature of atmosphere density distribution with altitude, most of the radiance along any line of sight of a limb imager comes from the vicinity of the lowest altitude point on a given line of sight (called *tangent point*). The resolution of the retrieved data depends on the density of tangent points in the area. For airborne observations this density is highly inhomogeneous. The densely measured regions are limited in size, located below flight altitude only and rarely rectangular. The large, poorly resolved areas with little or no tangent points still need to be included in the grid as long as lines of sight of any measurements pass through them. A rectilinear grid retrieval tends to either underresolve the well-measured area, or waste memory and computation time for regions with few measurements. To avoid this problem, a tomographic retrieval on irregular grid with Delaunay triangulation was developed (section 2). It allows for significant computational cost improvements without compromising retrieval quality.

Most retrieval error estimation techniques would be unreasonably computationally expensive in the case of a 3D tomographic retrieval. Monte Carlo methods allow a relatively quick error estimation. In order to apply Monte Carlo for a 3D tomographic retrieval on an irregular grid and using our newly developed regularisation, dedicated algorithms have to be developed. These are presented in section 3.

The new methods described in sections 2 and 3 were tested with synthetic measurement data to compare their results and computational costs. The tests, and their results, are described in detail in section 4.

2 Regularisation improvements and irregular grids

2.1 The retrieval

5 The main focus of this paper is the algorithm for retrieving atmospheric quantities, such as temperature or trace gas mixing ratios, from limb sounder measurements in 3D. In this section, we outline the general inverse modelling approach to this problem and identify some aspects we aim to improve upon.

Let an atmospheric state vector $\boldsymbol{x} \in \mathbb{R}^n$ represent the values of some atmospheric quantities on a finite grid and let $\boldsymbol{y} \in \mathbb{R}^m$ be a set of m remote measurements taken within the region in question. The physics of the measurement process itself has to

10 be understood for the instrument to be practical, so one can usually build a theoretical model of the measurement (the forward model) $F : \mathbb{R}^n \to \mathbb{R}^m$. The inverse problem of determining $\boldsymbol{x}$ given $\boldsymbol{y}$ is then typically both underdetermined and ill-posed: many atmospheric states would result in the same measurement, but due to instrument and model errors no states would yield the exact measurement result $\boldsymbol{y}$. One can solve this problem by finding an atmospheric state $\boldsymbol{x}$ that minimises the following quadratic cost function

$$15 \quad J(\boldsymbol{x}) = (\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{y})^T \mathbf{S}_\epsilon^{-1} (\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{y}) + (\boldsymbol{x} - \boldsymbol{x}_\mathrm{a})^T \mathbf{S}_\mathrm{a}^{-1} (\boldsymbol{x} - \boldsymbol{x}_\mathrm{a}) \tag{1}$$

The first term of (1) quantifies the difference between the actual measurement and a simulated one (given the atmospheric state is $\boldsymbol{x}$). The matrix $\mathbf{S}_\epsilon^{-1} \in \mathbb{R}^m \times \mathbb{R}^m$ represents the expected instrument errors. The second term introduces regularisation, i.e. it is larger for $\boldsymbol{x}$ that are unlikely given our general, measurement unrelated, knowledge about the atmosphere. In practice this is achieved using an a priori state $\boldsymbol{x}_\mathrm{a}$, usually derived from climatologies. It is meant for introducing the best of our prior

20 knowledge into the retrieval process without imposing any features of the kind to be measured. The precision matrix $\mathbf{S}_\mathrm{a}^{-1}$ (the inverse of a covariance matrix for atmospheric states $\mathbf{S}_\mathrm{a}$) contains the prior knowledge about the probability distribution of atmospheric states. For a more detailed discussion of this approach refer to Rodgers (2000).

Finding the precision matrix $\mathbf{S}_\mathrm{a}^{-1}$ that would correctly represent the physical and statistical properties of the atmosphere is challenging. The Tikhonov regularisation approach (Tikhonov and Arsenin, 1977)

$$25 \quad \mathbf{S}_\mathrm{a}^{-1} = \alpha_0^2 \mathbf{L}_0 \mathbf{L}_0^T + \alpha_v^2 \mathbf{L}_z \mathbf{L}_z^T + \alpha_h^2 \mathbf{L}_x \mathbf{L}_x^T + \alpha_h^2 \mathbf{L}_y \mathbf{L}_y^T \tag{2}$$

is widely used for this purpose. Here the matrix $\mathbf{L}_0$ is diagonal and contains reciprocals of standard deviations of atmospheric quantities, and $\mathbf{L}_x$, $\mathbf{L}_y$, $\mathbf{L}_z$ represent spatial derivatives in the respective directions on regular rectangular grid, estimated by forward differences. The positive constants α_0, α_h, α_v are chosen ad-hoc. This is a convenient way to construct $\mathbf{S}_\mathrm{a}^{-1}$, and it serves the purpose of ruling out very pathological, oscillatory retrievals. The constants α_0, α_h, α_v are, however, unphysical, grid

30 dependent, and hence have to be estimated (typically by trial-and-error) for each use case. They are not related to the properties

3

of the atmosphere in any simple fashion. Also, the usage of first order derivatives only is justified mostly by simplicity and the need to keep the computations cheap.

We chose a slightly different, more physically and statistically motivated approach to regularisation, which is introduced in the following section.

2.2 Covariance

This section describes the theoretical background for the regularisation algorithm used for our retrieval, i.e. the motivation for $\mathbf{S}_a^{-1}$ in equation (1).

If the atmospheric state x has the mean x_{apr} and covariance matrix $\mathbf{S}_a$ then its entropy is maximised if x has multivariate normal distribution ~~(proved e.g. by Tarantola, 2013)~~(proved e.g. by **?**). Hence by treating x as a random vector with this distribution we impose the smallest restriction possible while still introducing a priori information from $\mathbf{S}_a$ into our retrieval. Also, the probability density of x is then

$$f_x(\hat{x}) \sim \exp\left(-(\hat{x} - x_{apr})^T \mathbf{S}_a^{-1}(\hat{x} - x_{apr})\right) \tag{3}$$

This immediately justifies the second term of equation (1), provided that $\mathbf{S}_a^{-1}$ represents, to some extent, the actual statistics of the atmosphere. This is often referred to as the Bayesian approach to regularisation (Rodgers, 2000). The precision matrix from equation (2), is, however, ~~just~~ a mathematical device not meant to represent the physical world, as discussed in the previous section. Furthermore, we would like to derive a coordinate independent expression for the regularisation term of cost function (also unlike equation (2)), as this would be useful for work with irregular grids and allow us to use the same regularisation on completely different grid geometries. To achieve these goals, we ~~go back to the theoretical foundations of the classic Tikhonov approach, but do not limit ourselves to just minimising first-order spatial derivatives of the result.~~ use continuous covariance operators and their associated norms and only discretise them as a last step, hence preserving the grid independence and their statistical interpretation. The resulting discrete regularisation is a variant of Tikhonov in its final numerical form, but can be interpreted as a realisation of general continuous covariance relations. The following paragraph introduces some of the required formalism (Lim and Teo, 2009; Tarantola, 2013). Let us denote the departure of the atmospheric quantity $f(r)$ from the a priori by $\phi(r) = f(r) - f_{apr}(r)$ for some $r \in \mathbb{R}^3$. In some finite volume $V \subset \mathbb{R}^3$, we define the covariance operator C by

$$C\phi(r) = \int_{r' \in V} \phi(r') C_k(r, r') \, dV \tag{4}$$

where $C_k : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ is the covariance kernel (also known as covariance function). Then we can treat the scalar fields $\phi(r)$ as elements of Hilbert space with the product

$$\langle \phi, \varphi \rangle = \int_V \phi(x) \, C^{-1} \varphi(x) \, dV \tag{5}$$

~~It~~ which induces the norm $\|\phi\|^2 = \langle \phi, \phi \rangle$~~, which is represented, in our discrete numerical approach, by the second term of (1).~~ . Once the explicit expression for the norm is found it can be discretised, representing the scalar field $\phi(r)$ by the atmospheric

state vector $\boldsymbol{x} - \boldsymbol{x}_{\mathrm{apr}}$. Then we can write

$$\|\phi\|^2 = (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{apr}})^T \mathbf{S}_{\mathrm{a}}^{-1} (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{apr}}) \tag{6}$$

It now remains to find an appropriate covariance kernel C_k. Let us first consider an atmosphere that is isotropic and has the same physical and statistical properties everywhere. Then one would expect $C_k(\boldsymbol{r}, \boldsymbol{r}') = g(\|\boldsymbol{r} - \boldsymbol{r}'\|)$ with some monotonously

5 decreasing function $g : [0, \infty) \to [0, \infty)$. The most common kernel used in literature for fluid dynamics, meteorology and similar applications is the parametric Matérn covariance kernel (introduced, e.g. by Lim and Teo (2009))

$$C_\nu(\boldsymbol{r}, \boldsymbol{r}') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{d}{L} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{d}{L} \right) \tag{7}$$

where $d = \|\boldsymbol{r} - \boldsymbol{r}'\|$, σ and L are the standard deviation and typical length scale of structures (correlation length), respectively, of the atmospheric quantity in question. K_ν is the modified Bessel function of the second kind, $\Gamma(\nu)$ is the Gamma function.

10 Exponential and Gaussian covariance kernels are special cases of the kernel (7), with $\nu = 0.5$ and $\nu = 1$ respectively. We choose the exponential covariance

$$C_k(\boldsymbol{r}, \boldsymbol{r}') = \sigma^2 \exp\left(-\frac{\|\boldsymbol{r} - \boldsymbol{r}'\|}{L} \right) \tag{8}$$

as it closely resembles the Matérn with ν values that are typically used for fluid problems and allows for analytic derivation of the subsequently required quantities. Also, having a significant number of parameters to estimate theoretically, we could not

15 make a good use of the flexibility provided by the free parameter of the Matérn covariance in any case.

It can be shown (Tarantola, 2013) that the norm associated with the covariance (8) is

$$\|\phi\|^2 = \frac{1}{8\pi\sigma^2} \int_{\boldsymbol{r} \in V} \left[\frac{\phi^2}{L^3} + \frac{2\|\nabla\phi\|^2}{L} + L(\Delta\phi)^2 \right] \mathrm{d}V \tag{9}$$

Neglecting some boundary terms. In a more realistic picture of the atmosphere, the correlation length L depends on altitude and strongly depends on direction: correlation between vertically separated air parcels is much weaker than between air parcels

20 separated by the same distance horizontally. We propose to deal with anisotropic or variable L by performing a coordinate transformation such that L would be isotropic and constant in resulting coordinates. In particular, let $U, V \subset \mathbb{R}^3$ and consider a bijective map $\boldsymbol{\xi} : V \to U$ such that both $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^{-1}$ are twice differentiable on their respective domains and have non-zero first and second derivatives everywhere. Then, using integration by substitution and basic vector calculus identities, equation (9) can be written as

$$\|\phi\|^2 = \int_{\boldsymbol{u} \in \boldsymbol{\xi}^{-1}(V)} \frac{|\det(\mathrm{D}\boldsymbol{\xi})|}{8\pi\sigma^2} \left(\frac{(\phi(\boldsymbol{\xi}))^2}{L^3} + \frac{2\|\boldsymbol{\delta}(\phi(\boldsymbol{\xi}))\|^2}{L} + L\operatorname{Tr}^2 \left[(\mathrm{D}\boldsymbol{\xi})^{-T} [\mathrm{D}^2\phi(\boldsymbol{\xi}) - \boldsymbol{\delta}(\phi(\boldsymbol{\xi})) \cdot \mathrm{D}^2\boldsymbol{\xi}] (\mathrm{D}\boldsymbol{\xi})^{-1} \right] \right) \mathrm{d}U$$

25

$$\tag{10}$$

$$\boldsymbol{\delta}(\phi(\boldsymbol{\xi})) = (\mathrm{D}\boldsymbol{\xi})^{-T} \nabla(\phi(\boldsymbol{\xi})) \tag{11}$$

Here $(\mathrm{D}\boldsymbol{\xi})_{ij} = \partial\xi_i/\partial u_j$ is the Jacobian, and we define the matrix $(\mathrm{D}^2 f)_{ij} = (\partial^2 f)/(\partial u_i \partial u_j)$. $\boldsymbol{\xi}$ can be chosen so that all its spatial derivatives could be computed analytically, and $\nabla(\phi(\boldsymbol{\xi}))$, $\mathrm{D}^2(\phi(\boldsymbol{\xi}))$ are the derivatives of ϕ in the transformed space, so numerical evaluation of (10) is similar in complexity to that of (9), the only notable difference being the need to compute the mixed derivatives $\partial^2\phi/(\partial u_i \partial u_j)$, $i \neq j$ if $\mathrm{D}\boldsymbol{\xi}$ is not diagonal. A large and relevant class of suitable maps $\boldsymbol{\xi}$ can be defined, for

5   example, by assuming that the correlation lengths in vertical and horizontal direction are smooth functions of altitude $L_v(z)$, $L_h(z)$ (we use Cartesian coordinates (x, y, z) where z axis is vertical). Then the map

$$\boldsymbol{\xi}^{-1}(x, y, z) = \frac{1}{L}\left(x L_h(z),\, y L_h(z),\, \int\limits_0^z L_v(z')\mathrm{d}z'\right) \tag{12}$$

defines a corresponding transformation. If, in addition, L_h is constant, mixed derivatives are not required and we can simply evaluate $\nabla(\phi(\boldsymbol{\xi}))$, $\Delta(\phi(\boldsymbol{\xi}))$ in transformed space. Transformations of this type could be used, for example, to perform regu-

10   larisation in tropopause-based coordinates and use different correlation lengths for troposphere and stratosphere. For the study described in this paper, we only introduce anisotropy (i.e. L_v and L_h are different, but do not depend on altitude) and equation (10) becomes

$$\left\|\underline{\Phi\phi}\right\|^2 = \frac{1}{8\pi\sigma^2}\left[\int\limits_V \frac{\phi^2 \mathrm{d}V}{L_h^2 L_v} + \frac{2}{L_h}\int\limits_V \frac{L_h}{L_v}\left(\frac{\partial\phi}{\partial x}\right)^2 + \frac{L_h}{L_v}\left(\frac{\partial\phi}{\partial y}\right)^2 + \frac{L_v}{L_h}\left(\frac{\partial\phi}{\partial z}\right)^2 \mathrm{d}V + L_v\int\limits_V \left(\frac{L_h}{L_v}\frac{\partial^2\phi}{\partial x^2} + \frac{L_h}{L_v}\frac{\partial^2\phi}{\partial y^2} + \frac{L_v}{L_h}\frac{\partial^2\phi}{\partial z^2}\right)^2 \mathrm{d}V\right] \tag{13}$$

This general expression can now be discretised for use in the retrieval. We can represent the scalar field $\phi(\boldsymbol{r})$ by the

15   atmospheric state vector $\boldsymbol{x} - \boldsymbol{x}_{\mathrm{apr}}$ and then calculate a discrete approximation to the integral in (13). This can then be interpreted as $(\boldsymbol{x} - \boldsymbol{x}_{\mathrm{apr}})^T \mathbf{S}_{\mathrm{a}}^{-1}(\boldsymbol{x} - \boldsymbol{x}_{\mathrm{apr}})$ as seen from the equation (6). For example, if a matrix $\mathbf{L}_{\mathrm{z}}$ represents a finite difference scheme to calculate the derivative in z direction, i.e. $(\mathbf{L}_{\mathrm{z}}\boldsymbol{x})_j \approx \frac{\partial\phi}{\partial z}|_{\boldsymbol{r}=\boldsymbol{r}_j}$, and the diagonal matrix $\mathbf{V}$ represents the volumes of grid cells ($\mathbf{V}_{ii}$ being the volume of grid cell containing grid point $\boldsymbol{r}_i$), then

$$\sum_i (\mathbf{V}\mathbf{L}_{\mathrm{z}}\boldsymbol{x})_i \approx \int\limits_V \frac{\partial\phi}{\partial z}\mathrm{d}V \qquad \boldsymbol{x}^T \mathbf{L}_{\mathrm{z}}^T \mathbf{V}\mathbf{L}_{\mathrm{z}}\boldsymbol{x} \approx \int\limits_V \left(\frac{\partial\phi}{\partial z}\right)^2 \mathrm{d}V \tag{14}$$

20   Other terms of the integral in (13) can be computed similarly. To implement the calculation of this cost function on an irregular grid one needs to have a triangulation for that grid, an interpolation algorithm (both described in section 2.3), a method for calculating the spatial derivatives $\nabla\phi$ and $\Delta\phi$ (section 2.4), and a way to compute volume integrals (section 2.5). The estimation of the physical parameters in equation (13) is discussed in section 4.3. The errors introduced while constructing the precision matrix S_{a}^{-1} are estimated in Appendix A. Applicability of methods presented in this section to 1D and 2D

25   retrievals is briefly introduced in Appendix B.

2.3 Delaunay triangulation and interpolation

We aim to perform a retrieval on an arbitrary, finite grid. This section explains how vertically stretched Delaunay triangulation with linear interpolation is employed for that purpose.

To maintain generality and compatibility with arbitrary grids, we partition the retrieval volume into Euclidean simplexes (tetrahedrons in our 3D case) with vertices at grid points. Many such partitions exist, but it is beneficial to ensure that the number of very elongated tetrahedrons, that increase interpolation and volume integration errors, are kept to a minimum. The standard technique to achieve this is to use a Delaunay triangulation (Delaunay (1934), Boissonnat and Yvinec (1998)), which maximizes the minimum solid angle of any tetrahedron employed. Recall, also, from section 2.2 that atmospheric quantities in UTLS and the stratosphere above tend to have much more variation vertically than they do horizontally within similar length scales. This difference can be quantified by comparing horizontal and vertical correlation lengths L_h and L_v. Hence, stretching the space vertically by means of coordinate transformation $(x, y, z) \mapsto (x, y, \eta z)$ with $\eta \sim L_h/L_v$ before constructing a Delaunay triangulation ensures that, on average, the amount of variation of the retrieved quantity in vertical and horizontal directions is similar. This, in turn, allows us to make use of the aforementioned benefits of Delaunay partitioning.

In our implementation, we retrieve several quantities on the same grid, and L_h and L_v may differ for each of them. The cost function of equation (13) is evaluated separately for each quantity, so regularisation can still be calculated correctly as long as spatial structures of every quantity are adequately sampled by the grid. In practice, this means that as long as L_h/L_v are of the same order of magnitude for all quantities, they can all be retrieved on one grid. Fortunately this is usually the case (see section 4.3). For test retrievals in this paper $L_h/L_v = 200$ for trace gas concentrations and $L_h/L_v = 67$ for temperature, therefore we set $\eta = 100$. Retrieval results do not indicate systematic oversampling or undersampling in either horizontal or vertical directions and are not sensitive to a change in η.

In order to keep the computational costs low, we use a simple linear interpolation scheme: in each Delaunay cell we express an atmospheric quantity f at any point $\boldsymbol{x}$ as

$$f(\boldsymbol{r}) = f(\boldsymbol{r_0}) + \boldsymbol{k} \cdot (\boldsymbol{r} - \boldsymbol{r_0}), \quad \boldsymbol{k} = \text{const.} \tag{15}$$

If $\boldsymbol{r}_i$, $0 \leq i \leq 4$ are the four vertices of the Delaunay cell, the (unique) constant gradient $\boldsymbol{k}$ can obtained from a system of 3 linear equations

$$f(\boldsymbol{r}_i) = f(\boldsymbol{r_0}) + \boldsymbol{k} \cdot (\boldsymbol{r}_i - \boldsymbol{r}_0), \quad 1 \leq i \leq 3 \tag{16}$$

This method ensures that the interpolated quantity is continuous and consistent with the volume integration scheme described in section 2.5. Implementation is simple and fast, as any point inside a given Delaunay cell can be interpolated with data about that cell only. The gradient ∇f of the atmospheric quantity, is, however, generally discontinuous at cell boundaries making the interpolation unsuitable for direct use in spatial derivative evaluation. The Computational Geometry Algorithms Library (CGAL, https://www.cgal.org) was used to construct Delaunay triangulations for our grids.

2.4 Derivatives

In order to evaluate the cost function based on 3D exponential covariance (13), we need to estimate $\nabla\phi$ and $\Delta\phi$ at every grid point (as before, $\phi = f - f_{\mathrm{apr}}$ is the departure of atmospheric quantity f from the a priori). This section describes the algorithm we use to achieve that on irregular Delaunay grids.

We begin by establishing some requirements that our derivative-estimation algorithm will have to meet. Firstly, recall that our inverse model assembles the precision matrix $\mathbf{S}_a^{-1}$ so that $~~(\boldsymbol{y}-\boldsymbol{y}_{\mathrm{apr}})^T\mathbf{S}_a^{-1}(\boldsymbol{y}-\boldsymbol{y}_{\mathrm{apr}})~~$ $(\boldsymbol{x}-\boldsymbol{x}_{\mathrm{apr}})^T\mathbf{S}_a^{-1}(\boldsymbol{x}-\boldsymbol{x}_{\mathrm{apr}})$ would be a numerical representation of (13) (or, more generally, (10)). Let us write $\mathbf{S}_a^{-1} = \mathbf{A}+\mathbf{B}$, where $\mathbf{A}$ is a diagonal matrix representing the first integral of (13) (or first term of (10)), and $\mathbf{B}$ represents the remaining terms. If $\mathbf{B}$ were an exact representation, it would be positive definite by construction, but this may not hold with $\nabla\phi$ and $\Delta\phi$ obtained numerically for a finite grid. If indeed an atmospheric state vector $\boldsymbol{q}$ exists so that $\boldsymbol{q}^T\mathbf{B}\boldsymbol{q} = 0$, then for every atmospheric state $~~\boldsymbol{y}~~$ $\boldsymbol{x}$

$$(\boldsymbol{x}\pm\boldsymbol{q}-\boldsymbol{x}_{\mathrm{apr}})^T\mathbf{B}(\boldsymbol{x}\pm\boldsymbol{q}-\boldsymbol{x}_{\mathrm{apr}}) = (\boldsymbol{x}-\boldsymbol{x}_{\mathrm{apr}})^T\mathbf{B}(\boldsymbol{x}-\boldsymbol{x}_{\mathrm{apr}})\pm 2(\boldsymbol{x}-\boldsymbol{x}_{\mathrm{apr}})^T\mathbf{B}\boldsymbol{q} \tag{17}$$

hence one of the states $~~\boldsymbol{y}\pm\boldsymbol{q}~~$ $\boldsymbol{x}\pm\boldsymbol{q}$ would be favoured (have a smaller cost function) over $~~\boldsymbol{y}~~$ $\boldsymbol{x}$ by the inverse modelling algorithm. Therefore, a scaled version of $\boldsymbol{q}$ would appear on any retrieval as noise. This particular type of noise would only be suppressed by the cost function term from $~~(\boldsymbol{y}-\boldsymbol{y}_{\mathrm{apr}})^T\mathbf{A}(\boldsymbol{y}-\boldsymbol{y}_{\mathrm{apr}})~~$ $(\boldsymbol{x}-\boldsymbol{x}_{\mathrm{apr}})^T\mathbf{A}(\boldsymbol{x}-\boldsymbol{x}_{\mathrm{apr}})$. This guarantees only that $~~\boldsymbol{y}~~$ $\boldsymbol{x}$ is not very far from a priori, but does nothing to ensure that it is even continuous. This often results in retrievals with unphysical periodic structures (noise) and is unacceptable.

We deal with this problem by explicitly ensuring that $\mathbf{B}$ is positive definite. Consider the estimation of derivatives at a single grid point $\boldsymbol{a}\in\mathbb{R}^3$. $\nabla\phi(\boldsymbol{a})$ and $\Delta\phi(\boldsymbol{a})$ are numerically estimated from the values of ϕ in some (say m) grid points near $\boldsymbol{a}$. We can write this as a map

$$D_{\boldsymbol{a}}(\phi):\mathbb{R}^m\to\mathbb{R}^6, \quad \{\phi(\boldsymbol{r}_i)-\phi(\boldsymbol{a}):1\le i\le m\}\mapsto\left\{\frac{\partial\phi}{\partial x},\frac{\partial\phi}{\partial y},\frac{\partial\phi}{\partial z},\frac{\partial^2\phi}{\partial x^2},\frac{\partial^2\phi}{\partial y^2},\frac{\partial^2\phi}{\partial z^2}\right\} \tag{18}$$

Now $\boldsymbol{q}^T\mathbf{B}\boldsymbol{q} = 0$, $\boldsymbol{q}\ne\boldsymbol{0}$ implies that at every grid point $\boldsymbol{a}\in\mathbb{R}^3$ we have $\nabla\phi(\boldsymbol{a}) = 0$, $\Delta\phi(\boldsymbol{a}) = 0$ for the $\phi(\boldsymbol{a})$ corresponding to state $\boldsymbol{q}$. Hence such $\boldsymbol{q}$ will not exist if we require that $~~D_a~~$ $D_{\boldsymbol{a}}$ have trivial null space for all grid points $\boldsymbol{a}$ (i.e. $~~D_a(\phi) = \{0,\dots,0\}~~$ ~~only if $\phi(\boldsymbol{x}_i) = \phi(\boldsymbol{a})$~~ $D_{\boldsymbol{a}}(\phi) = \{0,\dots,0\}$ only if $\phi(\boldsymbol{r}_i) = \phi(\boldsymbol{a})$, $1\le i\le m$). The converse is not always true, so the trivial null space requirement is a stronger condition than absolutely necessary. It is, however, advantageous because it restricts the derivative-estimation algorithm at each grid point independently and hence can be implemented without computationally expensive operations on $\mathbf{S}_a^{-1}$ as a whole.

Since (18) will be used directly to construct $\mathbf{S}_a^{-1}$, which does not depend on atmospheric state $~~\boldsymbol{y}~~$ $\boldsymbol{x}$ and hence on $\phi(\boldsymbol{a})$, $D_{\boldsymbol{a}}$ must be a linear map, so null space is a vector space of dimension $\max\{0,m-6\}$, hence we need $m\le 6$. We also want to avoid the derivative estimation to be underdetermined ($m\ge 6$), so we choose $m = 6$, i.e. we aim to construct a linear map (18) that estimates derivatives at grid point $\boldsymbol{a}$ using atmospheric quantity values at 6 points around $\boldsymbol{a}$. This prevents us from using interpolation for derivative calculations, since each interpolated value depends on atmospheric quantity values at multiple (in our case 4) grid points and consequently, even if the same grid point is reused for several interpolated values, the total number of grid points employed exceeds 6 in any interpolation-based scheme we could come up with.

A solution satisfying the above criterion and able to deal with grid points with neighbours at irregular positions is based on polynomial fitting. For 6 grid points around a grid point $\boldsymbol{a} = (x_a, y_a, z_a)$ denoted as $\boldsymbol{r}_i = (x_a + x_i, x_a + y_i, x_a + z_i)$, $1 \leq i \leq 6$ we write

$$\phi(\boldsymbol{r}_i) - \phi(\boldsymbol{a}) = \phi_x x_i + \phi_y y_i + \phi_z z_i + \frac{\phi_{xx}}{2} x_i^2 + \frac{\phi_{yy}}{2} y_i^2 + \frac{\phi_{zz}}{2} z_i^2, \quad 1 \leq i \leq 6 \tag{19}$$

5 We solve this as a linear system for the unknowns $\phi_x, \phi_y, \phi_z, \phi_{xx}, \phi_{yy}, \phi_{zz}$ which can then be directly identified with the derivatives we seek. Note that, as we required, the system can be solved by inverting a 6-by-6 matrix that only depends on $\boldsymbol{r}_i$ and not $\phi(\boldsymbol{r}_i)$, so we can use this method to precompute a matrix that would act on the atmospheric state vector to return the value of the volume integral over the derivative (13).

The only remaining issue is selecting a suitable set of grid points $\boldsymbol{r}_i$ so that the resulting matrix that needs to be inverted
10 would not be singular or have a very high condition number. We found that this is much easier to achieve by imposing simple geometric criteria for point selection rather than algebraic conditions.

Let $\boldsymbol{a}$ be a point on Delaunay grid. The main idea of the algorithm we propose is to pick, for each spatial dimension d, a pair of points r_1, r_2 suitable for estimating the derivative in that direction. Points of such a pair should be sufficiently separated in d direction (i.e. projections of the vectors $\boldsymbol{r}_i - \boldsymbol{a}$ to d direction sufficiently different), be on the opposite sides of $\boldsymbol{a}$ (unless $\boldsymbol{a}$ is
15 at the boundary of the grid and this is not possible), and be reasonably close to $\boldsymbol{a}$. We obtain the required 6 points by choosing one pair of points for each of the 3 dimensions. This can be implemented as follows.

Let $\boldsymbol{a} = (x_a, y_a, z_a)$, and let us write the coordinates of other points on this grid as $\boldsymbol{r}_i = (x_a + x_i, y_a + y_i, z_a + z_i)$ with $r_i = \|\boldsymbol{r}_i - \boldsymbol{a}\|$. Let $0 < \beta < 1$ and $\gamma > 1$ be constants, suitable numeric values will be discussed further on. Then:

1. For each Delaunay grid neighbour $\boldsymbol{r}_i$ of $\boldsymbol{a}$, compute $\alpha_{ix} = x_i/r_i$, $\alpha_{iy} = y_i/r_i$, $\alpha_{iz} = z_i/r_i$.

20 2. If i, j such that $|\alpha_{ix}| > \beta$, $|\alpha_{jx}| > \beta$, $\alpha_{ix}\alpha_{jx} < 0$ exist, then pick ~~x_i, x_j~~ r_i, r_j for derivative calculation. Else if k, l such that $|\alpha_{kx}| > |\alpha_{lx}| > \beta$, $\alpha_{kx}/\alpha_{lx} > \gamma$ exist, then pick ~~x_l, x_k~~ r_l, r_k for derivative calculation.

3. Repeat step 2 for y and z dimensions. Each point can be only picked for one of the dimensions.

4. If neither condition of step 2 is satisfied for at least 1 dimension, repeat steps 1-3 with not only direct Delaunay grid neighbours of $\boldsymbol{a}$, but also second neighbours (i.e. neighbours of neighbours). If this fails as well, do not calculate deriva-
25 tives at the point $\boldsymbol{a}$ and simply assume them to be zero.

We found that the polynomial interpolation method is robust and a rather low value of $\beta = 0.3$ was sufficient to produce suitable sets of neighbouring points (i.e. almost all 6-by-6 matrices resulting from such choice of points can be reliably inverted for polynomial fitting). In practice this means that the algorithm can find a suitable pair of points for dimension d unless all the Delaunay neighbours of the point $\boldsymbol{a}$ are located very close to the other two axes w.r.t. the point $\boldsymbol{a}$.

30 ## 2.5 Volume integration

Computing (13) requires a way to estimate volume integrals on Delaunay grid, which is described in this section.

Let a Delaunay cell have vertices $\boldsymbol{r}_i$ and face areas S_i for $0 \leq i \leq 3$, and volume V. Then integrating an atmospheric quantity f as interpolated in (15) over V we get

$$\int_V f(\boldsymbol{r})\,\mathrm{d}V = \frac{V}{4}\sum_{i=0}^{3} f(\boldsymbol{r}_i) \tag{20}$$

One can prove this by writing (15) with $\boldsymbol{r}_i$, $0 \leq i \leq 3$ instead of $\boldsymbol{r}_0$ and adding up the resulting 4 equations to obtain

$$5 \quad f(\boldsymbol{r}) = \frac{1}{4}\sum_{i=0}^{3} f(\boldsymbol{r}_i) + \boldsymbol{k} \cdot \left(\boldsymbol{r} - \frac{1}{4}\sum_{i=0}^{3} \boldsymbol{r}_i\right) \tag{21}$$

Now choose Cartesian coordinates (x', y', z') where $\boldsymbol{r_0} = (H, 0, 0)$ and $\boldsymbol{r}_i$, $1 \leq i \leq 3$ lie in the plane $x' = 0$ and consider the x' component of the integral $\boldsymbol{I} = \int_V \left(\boldsymbol{r} - 1/4\sum_{i=0}^{3}\boldsymbol{r}_i\right)\mathrm{d}V$

$$I_{x'} = \int_V \left(x' - \frac{H}{4}\right)\mathrm{d}V = \int_0^H x'S_0\left(\frac{H-x'}{H}\right)^2\mathrm{d}x' - \frac{HV}{4} = \frac{H^2 S_0}{12} - \frac{HV}{4} = 0 \tag{22}$$

As we can choose such coordinates for any vertex instead of $\boldsymbol{r}_0$, the integral $\boldsymbol{I}$ has zero component in 3 linearly independent
10 directions, so $\boldsymbol{I} = 0$. Then integrating (21) over V yields (20) and completes the proof.

We can now see that (20) gives an exact value to the volume integral of f provided that f behaves exactly as prescribed by our interpolation scheme. Since the interpolation is linear we may expect the values of volume integrals to be correct to the first order in cell dimensions. Since, generally, all data fed into the inverse modelling problem is interpolated beforehand, there is little reason to expect that a more elaborate and accurate volume integration scheme would improve the accuracy of retrievals
15 on its own. A method based on Voronoi cells was also attempted and did not produce meaningfully different results.

3 Monte Carlo diagnostics

Estimating the ~~accuracy~~ precision of remote sensing data products generated by means of inverse modelling is ~~, in general, rather difficult due to the indirect manner in which they are obtained from the actual measurements. These estimates are, however, very valuable not only~~ essential for the users of the final data ~~, but also~~ and also valuable for evaluation and optimi-
20 sation of the inverse modelling techniques. Detailed quantitative descriptions of data accuracy can be derived in theory (see validation in section 4.6 and Rodgers (2000)) but they are, in case of large retrievals, too numerically expensive to calculate in practice. The accuracy calculation typically requires $O\left(n^3\right)$ operations for an atmospheric state vector of n elements (for 3D tomography n might be of the order 10^6). One way of reducing computational cost is is employing the Monte Carlo technique to generate a set of random (Gaussian) atmospheric state vectors with the required covariance. One can then add these random
25 vectors to the result of the retrieval and run the forward model on the perturbed atmospheric states. The magnitude of perturbations that is required to produce a variation of forward model output similar to the expected error of measurements would then be an estimate of the retrieval error (Ungermann et al., 2011).

In order to implement this technique, one needs a way to generate a random vector ~~$\boldsymbol{y}$~~ $\boldsymbol{x}$ with the required covariance, represented by a known, symmetric, positive definite covariance matrix $\mathbf{S}_\mathrm{a}$. This can be done by generating a vector ~~$\boldsymbol{x}$~~ $\boldsymbol{u}$ of independent standard normal random variables and finding a square matrix $\tilde{\mathbf{L}}$ such that $\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T = \mathbf{S}_\mathrm{a}$. Then we can set ~~$\boldsymbol{y} = \tilde{\mathbf{L}}\boldsymbol{x}$~~ $\boldsymbol{x} = \tilde{\mathbf{L}}\boldsymbol{u}$, and indeed ~~$\langle \boldsymbol{yy}^T \rangle = \langle \tilde{\mathbf{L}}\boldsymbol{xx}^T\tilde{\mathbf{L}}^T \rangle = \tilde{\mathbf{L}}\langle \boldsymbol{xx}^T \rangle \tilde{\mathbf{L}}^T = \mathbf{S}_\mathrm{a}$~~ $\langle \boldsymbol{xx}^T \rangle = \langle \tilde{\mathbf{L}}\boldsymbol{uu}^T\tilde{\mathbf{L}}^T \rangle = \tilde{\mathbf{L}}\langle \boldsymbol{uu}^T \rangle \tilde{\mathbf{L}}^T = \mathbf{S}_\mathrm{a}$, i.e. the vector ~~$\boldsymbol{y}$~~ $\boldsymbol{x}$ has the required covariance (the angle brackets in the expression above denote the expected value).

A widely used technique for obtaining the matrix $\mathbf{L}$ is the Cholesky decomposition. Given the covariance matrix $\mathbf{S}_\mathrm{a} = \{s_{ij}\}$ it explicitly provides a lower triangular matrix root $\tilde{\mathbf{L}} = \{l_{ij}\}$ satisfying $\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T = \mathbf{S}_\mathrm{a}$ as shown in (23).

$$
l_{ij} = \begin{cases}
\sqrt{s_{ij} - \sum_{k=1}^{i-1} l_{ik}^2}, & i = j \\
l_{jj}^{-1}\left(s_{ij} - \sum_{k=1}^{j-1} l_{ik}l_{jk}\right), & i > j \\
0, & i < j
\end{cases}
\tag{23}
$$

In practice we do not usually assemble the covariance matrix $\mathbf{S}_\mathrm{a}$, but rather its inverse: the precision matrix $\mathbf{S}_\mathrm{a}^{-1}$, because the latter is sparse. This is not an issue, since $\mathbf{S}_\mathrm{a}^{-1}$ is then also symmetric positive definite, so one can compute its root $\mathbf{L}$ such that $\mathbf{L}\mathbf{L}^T = \mathbf{S}_\mathrm{a}^{-1}$ in the same way as above, and then obtain ~~$\boldsymbol{y}$~~ $\boldsymbol{x}$ from the linear system ~~$\mathbf{S}_\mathrm{a}^{-1}\boldsymbol{y} = \mathbf{L}\boldsymbol{x}$~~ $\mathbf{S}_\mathrm{a}^{-1}\boldsymbol{x} = \mathbf{L}\boldsymbol{u}$. We will refer to the components of precision matrix by $\mathbf{S}_\mathrm{a}^{-1} = \{a_{ij}\}$.

Cholesky decomposition does not preserve the sparse structure of $\mathbf{S}_\mathrm{a}^{-1}$, i.e. the $\mathbf{L}$ obtained in this way will typically have many more non-zero entries than $\mathbf{S}_\mathrm{a}^{-1}$. It follows from (23), however, that if $\mathbf{S}_\mathrm{a}^{-1}$ has lower half-bandwidth w (definition in equation (24) below), then so has $\mathbf{L}$

$$
w = \min\{k \geq 0 : i - j > k \Rightarrow a_{ij} = 0\}
\tag{24}
$$

In practice, this means that if $\mathbf{S}_\mathrm{a}^{-1}$ is a sparse $N \times N$ matrix with half-bandwidth $w \ll N$, $\mathbf{L}$ will have approximately $N(w+1)$ non zero entries. Cholesky decomposition is hence well suited for computing diagnostics of 1D retrieval: assuming that the value of the retrieved atmospheric parameter at one point only directly correlates with its w nearest neighbouring points in each direction, one gets a precision matrix with half-bandwidth w and can compute $\mathbf{L}$ cheaply with $O(Nw^2)$ operations.

The situation is very different in the higher dimensions. We will use some results of graph theory to show that Cholesky decomposition is not practical in those cases. The n-dimensional lattice graph P_k^n consists of an n-dimensional rectangular grid of size k in each direction, with edge between any two grid neighbours. Let us label the vertices of P_k^n with integers: $P_k^n = \{v_i : 1 \leq i \leq k^n\}$, and let q be the maximum difference of indexes of neigbouring vertices ($q = \max\{|i - j| : v_i, v_j \text{adjacent}\}$). Then the *bandwidth* $\varphi(P_k^n)$ of P_k^n is defined as the minimum possible value of q among all possible ways to label the vertices. Now say we have some physical quantity defined on each vertex of this grid. Then any reasonable precision matrix $\mathbf{S}_\mathrm{a}^{-1}$ for this grid would at least give non-zero correlations between grid neighbours, i.e. $\left(\mathbf{S}_\mathrm{a}^{-1}\right)_{ij} \neq 0$ if v_i and v_j are neighbours. Then, by comparing the definitions of $\varphi(P_k^n)$ and lower-half bandwidth w of the precision matrix from the last paragraph, one can see that $2w + 1 \geq \varphi(P_k^n)$. FitzGerald (1974) showed that $\varphi\left(P_k^2\right) = k$ and $\varphi\left(P_k^3\right) = \lfloor 3k^2/4 + k/2 \rfloor$. Therefore, the narrowest possible half-bandwidths of $\mathbf{S}_\mathrm{a}^{-1}$ are $w = O(k)$ in 2D and $w = O\left(k^2\right)$ in 3D. It follows that if the grid contains a total of N

points, the computational cost of Cholesky decomposition would be $O(N^2)$ in 2D and $O\left(N^{7/3}\right)$ in 3D, which is unsatisfactory for large retrievals.

For these higher dimensions, we need to use sparse matrix iterative techniques to reduce computational cost and memory storage requirements, such as Krylov subspace methods. ~~These allow to solve a linear system $\mathbf{M}\boldsymbol{x} = \boldsymbol{b}$, where $\mathbf{M}$ is a linear~~

5 ~~map, for $\boldsymbol{x}$ iteratively. The cost of each iteration is proportional to the number of non-zero elements in $\mathbf{M}$.~~ In general, it is rather difficult to find a simple iteration scheme that would compute a square root of a matrix and converge reasonably fast. Here we will follow an algorithm proposed by Allen et al. (2000). Consider a system of linear ordinary differential equations (ODEs) with initial condition

$$
\begin{cases}
\mathrm{d}\boldsymbol{v}/\mathrm{d}t = -\frac{1}{2}\left(\mathbf{S}_{\mathrm{a}}^{-1}t - (1-t)\,I\right)^{-1}\left(\mathbf{I} - \mathbf{S}_{\mathrm{a}}^{-1}\right)\boldsymbol{v}(t) \\
\boldsymbol{v}(0) = \boldsymbol{u}
\end{cases}
\tag{25}
$$

10 where ~~$\boldsymbol{x}(t)$~~ $\boldsymbol{v}(t)$ is a column vector of size N, $\mathbf{I}$ is the identity matrix and ~~$\mathbf{A}$~~ $\mathbf{S}_{\mathrm{a}}^{-1}$ is a $N \times N$ symmetric positive definite (s. p. d. ~~matrix, $\|\mathbf{A}\|_\infty < 1$ (so that $\mathbf{A} - \mathbf{I}$)~~ matrix, prescaled so that $\|\mathbf{S}_{\mathrm{a}}^{-1}\|_\infty < 1$ (i. e. $\mathbf{S}_{\mathrm{a}}^{-1} - \mathbf{I}$ is non-singular). ~~$(\mathbf{A} - \mathbf{I})$ and~~ ~~$(\mathbf{A}t + (1-t)\mathbf{I})^{-1}$~~ $(\mathbf{S}_{\mathrm{a}}^{-1} - \mathbf{I})$ and $(\mathbf{S}_{\mathrm{a}}^{-1}t + (1-t)\mathbf{I})^{-1}$ commute for $0 \le t \le 1$, ~~because both of them have the same set of~~ ~~eigenvectors as $\mathbf{A}$. This~~ which makes it easy to verify that

$$
\boldsymbol{v}(t) = \left(\underline{\mathbf{A}t}\,\mathbf{S}_{\mathrm{a}}^{-1}t - (1-t)\,\mathbf{I}\right)^{1/2}\boldsymbol{u}
\tag{26}
$$

15 is the solution of (25). Hence we can obtain ~~$\boldsymbol{x}(1) = \mathbf{A}^{1/2}\boldsymbol{c}$~~ $\boldsymbol{v}(1) = \mathbf{S}_{\mathrm{a}}^{-1/2}\boldsymbol{u}$ by solving (25) numerically~~. Note~~, and then solve the linear system $\mathbf{S}_{\mathrm{a}}^{-1}\boldsymbol{x} = \boldsymbol{v}(1)$ for the vector $\boldsymbol{x}$, so that $\boldsymbol{x} = \mathbf{S}_{\mathrm{a}}\mathbf{S}_{\mathrm{a}}^{-1/2}\boldsymbol{u}$. The matrix $\mathbf{S}_{\mathrm{a}}^{-1/2}$ above is symmetric by construction (sum of products of symmetric matrices), hence

$$
\langle \boldsymbol{x}\boldsymbol{x}^T \rangle = \mathbf{S}_{\mathrm{a}}\mathbf{S}_{\mathrm{a}}^{-1/2}\langle \boldsymbol{u}\boldsymbol{u}^T \rangle \mathbf{S}_{\mathrm{a}}^{-1/2}\mathbf{S}_{\mathrm{a}} = \mathbf{S}_{\mathrm{a}}
\tag{27}
$$

i.e. $\boldsymbol{x}$ is indeed the vector required for Monte Carlo simulations. Note also that one does not need to explicitly calculate

20 ~~inverses of any of the matrices involved, since $\mathrm{d}\boldsymbol{x}/\mathrm{d}t$ can be found by employing iterative methods to solve~~ matrix inverses while solving the ODE, since

$$
\left(\mathbf{S}_{\mathrm{a}}^{-1}t - (1-t)\,\mathbf{I}\right)\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = -\frac{1}{2}\left(\mathbf{I} - \mathbf{S}_{\mathrm{a}}^{-1}\right)\boldsymbol{v}(t)
\tag{28}
$$

is a linear system

$$
\left(\mathbf{A}t - (1-t)\,\mathbf{I}\right)\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -\frac{1}{2}\left(\mathbf{I} - \mathbf{A}\right)\boldsymbol{x}(t)
$$

25 ~~Furthermore, such a numerical solution can then be formally written as~~

$$
\boldsymbol{x}(1) = \sum_i \prod_j \left[\xi_{ij}\left(\mathbf{A}t_{ij} - (1-t_{ij})\,\mathbf{I}\right)^{-1}(\mathbf{I} - \mathbf{A})\right]\boldsymbol{c}
$$

12

~~for some real constants ξ_{ij}, t_{ij}, which implies that $\mathbf{A}^{1/2}$ is sum of the products of commuting symmetric matrices, therefore $\mathbf{A}^{1/2}$ is symmetric. This means $\left(\mathbf{A}^{1/2}\right)^T \mathbf{A}^{1/2} = \mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$. Hence, as shown before, if c is a vector of independent random variables, each with standard normal distribution, $x(1)$ has covariance matrix $\mathbf{A}$.~~ that can be solved for $d\boldsymbol{v}/dt$.

We chose a classical approach – a Runge-Kutta method, to solve the linear ODE system numerically. Using constant step size proved to be inefficient, so adaptive step size control was introduced (in particular, a fifth order Runge-Kutta-Fehlberg method by Fehlberg (1985)). Conjugate gradients method (Saad, 2003) was employed for all the linear equation systems involved. If we make the same assumption about the structure of the inverse of the correlation matrix as before, i.e. that atmospheric quantities are only correlated between nearest grid neighbours, the said matrix will have a constant number of non-zero entries in each row for any amount of rows (grid points) N. More explicitly, under this assumption the covariance matrix for 1D retrieval would be tridiagonal, and for a 3D retrieval on a regular rectangular grid it would contain 6 non-zero entries in each row. Therefore, each iteration of the conjugate gradients algorithm will cost $O(N)$ operations. It is difficult, in general, to estimate the dependence of required number of iterations on N, but in our case large matrices seem to converge similarly to smaller ones and the whole computational cost of generating random vectors remains close to $O(N)$, which is a huge improvement over the Cholesky decomposition. The algorithm does not, though, explicitly yield the root of the covariance matrix and only gives one random vector per run. Hence, unlike the Cholesky decomposition, it must be executed again for each new random vector required, so it is only feasible if the number of random vectors required is a lot smaller than N. Fortunately, this is very much the case in practice, since the number of random vectors required is typically only of the order of 100 (Ungermann, 2013).

Although this root-finding algorithm does not provide a way to directly multiply the matrix root with itself, there is an inexpensive way to verify that the matrix root was computed correctly. Let ~~$\mathbf{A}$~~ $\mathbf{S}$ be a $n \times n$ s. p. d. matrix, let $\boldsymbol{e}_i$ be the i'th unit vector ($1 \leq i \leq n$), and compute ~~$\boldsymbol{v}_i = \mathbf{A}^{1/2}\boldsymbol{e}_i$~~ $\boldsymbol{v}_i = \mathbf{S}^{1/2}\boldsymbol{e}_i$ using our algorithm. Then for any i,j we have ~~$\boldsymbol{v}_i^T\boldsymbol{v}_j = \boldsymbol{e}_i^T\left(\mathbf{A}^{1/2}\right)^T\mathbf{A}^{1/2}\boldsymbol{e}_i = \boldsymbol{e}_i^T\mathbf{A}\boldsymbol{e}_i = \mathbf{A}_{ij}$~~ $\boldsymbol{v}_i^T\boldsymbol{v}_j = \boldsymbol{e}_i^T\left(\mathbf{S}^{1/2}\right)^T\mathbf{S}^{1/2}\boldsymbol{e}_i = \boldsymbol{e}_i^T\mathbf{S}\boldsymbol{e}_i = \mathbf{S}_{jj}$, an element of ~~$\mathbf{A}$~~ $\mathbf{S}$. Hence the quality of the matrix root can be evaluated by comparing $\boldsymbol{v}_i^T\boldsymbol{v}_j$ with ~~$\mathbf{A}_{ij}$~~ $\mathbf{S}_{jj}$. The error tolerances of conjugate gradients and Runge-Kutta were adjusted so that these values would differ by at most 10^{-4} when factoring a matrix prescaled so that the largest eigenvalue is approximately 0.95.

4 Performance study

4.1 The GLORIA instrument and data processing

The Gimballed Limb Observer of Air Radiance in the Atmosphere (GLORIA) is an aircraft-based infrared limb imager. It is a Michelson interferometer with a 2D detector array, spectral coverage from 770 to 1400 cm^{-1} and spectral sampling of up to 0.0625 cm^{-1} (Riese et al. (2014); Friedl-Vallon et al. (2014)). It is carried aboard the Russian M55 Geophysica or German HALO research aircraft, and is generally intended to look to the side of the aircraft, but can pivot horizontally so that the angle between aircraft heading and observation direction can be changed from 45° to 135°. The limb observation setup naturally allows high vertical resolution (up to approximately 200 m) and, with a proper trade-off between spectral resolution and measurement time, a high density of vertical profiles sampled along-track. The pivoting allows to look at the same air mass

from several points along the flight path and use tomography to improve resolution in the horizontal direction perpendicular to the flight path. Best results with tomographic retrievals are achieved, however, when the aircraft flies in a path close to circular (hexagonal flight paths are used in practice, so that the aircraft could fly straight most of the time) of around 400 km in diameter, and GLORIA observes the air masses in the middle from many directions (Ungermann et al., 2011). A horizontal resolution down to 25 km in both directions can be achieved this way. As any limb observer, GLORIA can only provide tangent point data about air masses at and below aircraft flight altitude, which is limited to, approximately 20 km for the M55 Geophysica and 15 km for HALO. For more detail on GLORIA, refer to e.g. Friedl-Vallon et al. (2014).

The implementations of the algorithms described in the sections 2 and 3 were integrated into the Jülich Rapid Spectral Simulation Code Version 2 (JURASSIC2). The forward model of atmospheric radiance used in this code employs a forward model for atmospheric radiation based on radiances obtained from the emissivity growth approximation method (Weinreb and Neuendorffer (1973), Gordley and Russell (1981)) and the Curtis–Godson approximation (Curtis (1952), Godson (1953)). For more information about JURASSIC2, refer to Hoffmann et al. (2008); Ungermann et al. (2011); Ungermann (2013). The performance of the new algorithms, both in terms of output quality and computational cost, is evaluated in the rest of this section.

4.2 Retrieval setup and test data

A hexagonal flight pattern intended for tomography was realised on 25th of January 2016, as part of flight 10 of the POL-STRACC measurement campaign. The flight path of the HALO research aircraft contained a regular hexagon with diameter (distance between the opposite vertices) of around 500 km over Iceland, which allowed for a high spacial resolution retrieval in the central part of the aforementioned hexagon (Figure 1). The flight altitude throughout the hexagonal flight segment remained close to 14 km. For detailed information about this flight refer to Krisch et al. (2017).

The test case used in this paper is based on the actual aircraft path, measurement locations, spectral lines used for retrieval and meteorological situation during this flight. Synthetic measurement data was used instead of real GLORIA observations: the forward model of JURASSIC2 was employed to simulate the observed radiances in the atmospheric state given by the ECWMF temperature and WACCM model data for trace gases, simulated instrument noise was subsequently added to those radiances. This setup allows us to use model data as a reference (the "true" atmospheric state) for evaluation of retrieval quality. The same set of simulated measurements was used for all test retrievals described in this paper. These measurements were obtained by running the forward model on a very dense grid (about twice as dense in each dimension as those of the densest test retrievals). This was done to ensure that the discretisation errors in the simulated measurements would be minimal and would not favour any retrieval (as it may could happen, if they were generated on the same grid). An evaluation of forward model errors on different grids is presented in Appendix C.

The retrieval derives temperature and volume mixing ratios of O_3, HNO_3, CCl_4, $ClONO_2$. Temperature and CCl_4 concentration are derived in the altitude range from 3 km to 20 km; O_3, HNO_3, $ClONO_2$ are retrieved from 3 km to 64 km. The latter three trace gases have larger volume mixing ratios above flight level and hence significant contributions to measured radiances from these high altitudes. Interpolated WACCM data was used as a priori for all trace gases. A priori data for tem-
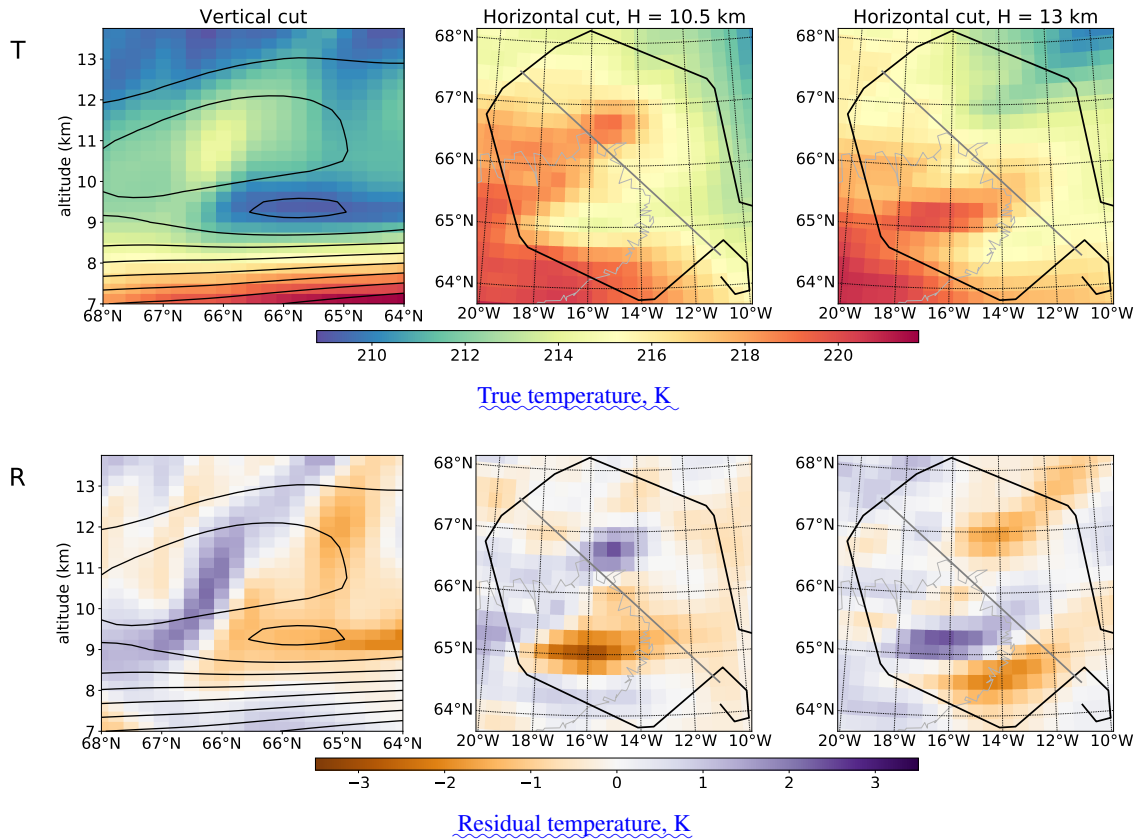
Figure 1. The first row (T) shows the ECMWF temperature data used to generate the synthetic measurements (true temperature). Second row (R) shows the temperature residual: the difference between the true temperature and the smoothed temperature field used as an a priori. In the first column, black lines represent contours of a priori temperature value. In columns 2 and 3: black line – flight path, thick grey line – position of cut shown in the first column, thin grey line – topography (Icelandic coastline visible).

perature was obtained by applying polynomial smoothing to ECMWF data. The retrieval is hence not supplied with the full temperature structure that was used to simulate the measurement. Therefore, the agreement between the retrieval result and the "true" atmospheric state, upon which the simulated measurements are based, cannot be achieved by overregularisation.

4.3 Correlation length estimation

5 The standard deviation σ and correlation the lengths L_h and L_v in equation (13) can be obtained from statistical analysis of in situ measurement or model data. Rigorous derivation of these parameters for each retrieved trace gas and temperature would require detailed analysis of in situ measurement databases and is, therefore, out of scope of this paper. However, the following rough approximation proved sufficient to improve upon retrieval results compared to the old regularisation scheme.
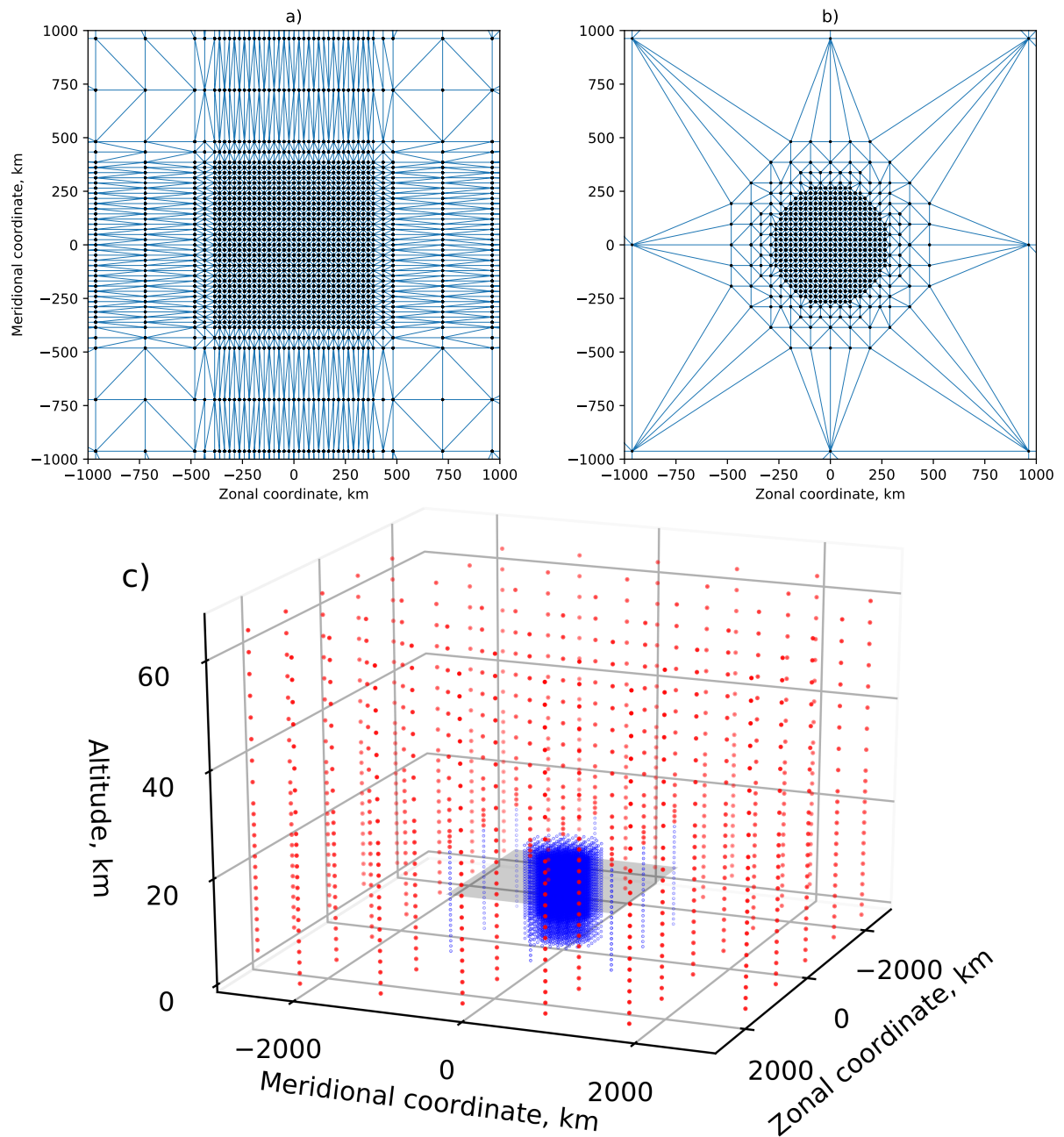
Figure 2. ~~Horizontal~~ Panels (a) and (b): horizontal cuts of retrieval grids. Dots show grid points, blue lines indicate Delaunay cell boundaries. ~~Panel~~ (a) shows the full initial grid for retrievals A, B, C (all horizontal layers of this grid are identical), ~~panel~~ (b) shows the thinned out grid for retrieval D (cut at 11 km). Panel (c): grid for retrieval D. Radial distance is defined as distance to the vertical line at the centre of the hexagonal flight path (66°N, 15°W). Points within the inner core (1000 km $\times$ 1000 km $\times$ 20 km) are shown in blue, other points – in red, shaded rectangle shows the position of horizontal cut in panel (b).

~~Grid for retrieval D. Radial distance is defined as distance to the vertical line at the centre of the hexagonal flight path (66°N, 15°W). Points within the inner core (1000 × 1000 × 20) are shown in blue, other points – in red, shaded rectangle shows the position of horizontal cut in Figure ??b.~~

Table 1. Retrieval parameters

Entity	α_0	α_h	σ
Temperature	10^{-3}	10^{-1}	0.645 K
O_3	10^{-5}	10^{-3}	70.7 ppbv
HNO_3	10^{-5}	10^{-3}	902 pptv
CCl_4	10^{-5}	10^{-3}	2.13 pptv
$ClONO_2$	10^{-5}	10^{-3}	76.7 pptv

Regularisation weights α_0, α_h, (as in (2)) are used for retrieval A, the standard deviation σ (see (13)) used for retrievals B-D

3D tomographic retrievals are most useful for those trace gases that have high vertical gradients and complex spatial structure. The spatial correlation lengths of concentrations of such gases are mostly determined by stirring, mixing and other dynamical processes and are therefore similar. Furthermore, Haynes and Anglade (1997) and Charney (1971) estimate of the dynamically determined aspect ratio $\alpha \approx 200-250$ in lower stratosphere, and we expect less than that in the upper troposphere. Hence we will assume $L_h/L_v = 200$ for all trace gases.

An analysis of spatial variability of some airborne in situ measurements was performed by Sparling et al. (2006). For ozone concentration $\chi(s, z)$, where s is the horizontal location and z is the altitude, they define fractional difference parameters

$$\Delta_{r,h} = \frac{2\,|\chi(s+r, z+h) - \chi(s, z)|}{[\chi(s+r, z+h) + \chi(s, z)]} \tag{29}$$

$$\Delta_r = \lim_{h \to 0} \Delta_{r,h} \tag{30}$$

and provide the dependence of the standard deviation $\sigma(\Delta_r)$ on the horizontal separation r. If the typical spatial variation in ozone concentration is significantly smaller than the mean ozone concentration $\langle\chi\rangle$, one can approximate $\Delta_r \approx |\chi(s+r) - \chi(s)|/\langle\chi\rangle$. Then, using the covariance relation (8) and various properties of the normal distribution, we get $\sigma^2(\Delta_r) = 2\,(\sigma/\langle\chi\rangle)^2\,(1 - 2/\pi)\,(1 - \exp(-r/L_h))$. By comparing this to experimental results of Sparling et al. (2006), we estimate $L_h = 200\,\text{km}$. Using the aspect ratio argument above, we also set $L_v = 1\,\text{km}$.

The spatial structure of retrieved temperature differs from that of the trace gases. It is determined not only ~~my~~ by mixing and stirring, but also radiative processes and gravity waves. Radiative transfer tends to erase some of the fine vertical structure determined by isentropic transport, hence one should use a ~~higher~~ greater vertical correlation length than in the case of trace gases. Gravity waves have a variety length scales, but due to the finite resolution of the instrument not all of them can be retrieved. Following the gravity wave observational filter study for the GLORIA instrument in Krisch et al. (2018) we can estimate lowest observable gravity wave horizontal wavelength to be $\lambda_h \sim 200\,\text{km}$ and lowest vertical wavelength $\lambda_v \sim 3\,\text{km}$. λ_h coincides with the dynamically determined correlation length $L_h = 200\,\text{km}$, so this value should be appropriate for temperature as well. The λ_v limit is not only determined by instrument resolution, but is also related the fundamental properties of

Table 2. Grid densities in different regions for retrieval D in kilometers

Radius	<300	300-400	400-550	550-1500	>1500
Altitude		Horizontal grid separation			
1-5	100	100	100	1000	1000
5-12.5	25	50	100	1000	1000
12.5-14.5	25	25	100	1000	1000
14.5-17	25	50	100	1000	1000
17-20	100	100	100	1000	1000
20-64	500	500	500	1000	1000
Altitude		Vertical grid separation			
1-5	1	1	1	1	2
5-12.5	0.125	0.25	1	1	2
12.5-14.5	0.125	0.125	1	1	2
14.5-18	0.125	0.25	1	1	2
18-24	2	2	2	2	2
24-64	4	4	4	4	4

gravity waves. Preusse et al. (2008) gives the following expression for the gravity wave saturation amplitude

$$\hat{T}_{\max} = \frac{\lambda_v \bar{T} N^2}{2\pi g} \tag{31}$$

where $\bar{T}$ is the mean temperature and N is the Brunt–Väisälä frequency. In typical lower stratosphere conditions and $\lambda_v = 3\,\mathrm{km}$ the equation (31) gives $\hat{T}_{\max} \approx 4\mathrm{K}$. Since gravity waves are usually not saturated and considering typical measurement error
5 (see ~~Figure 5~~4.6), waves of significantly ~~lower~~ shorter wavelengths would probably not be observed. We hence set $L_v = 3\,\mathrm{km}$. Time series of measurements at a fixed point in atmosphere are then used to approximate the standard deviations σ for all retrieved quantities (Table 1).

Our choice of σ value for temperature, in particular, may seem rather low. This is related to the choice of a priori: smoothed ECMWF data was used instead of climatologies that would have been more typical in this case. Such an a priori can be expected
10 to match the large scale structures of real temperature more closely, thus reducing the forward model errors in poorly resolved areas and improving the results. It is particularly useful for resolving fine structures, such as gravity waves, which were the main scientific interest of the measurment flight used as a basis for test retrievals Krisch et al. (2017). A likely closer match between a priori and retrieval thus requires a lower σ value, that, in this setup, represents expected strength of gravity wave disturbances, rather than full thermal variability of the atmospheric region in question.
15 ~~True temperature, K~~

~~Residual temperature, K The first row (T) shows the ECMWF temperature data used to generate the synthetic measurements (true temperature). Second row (R) shows the temperature residual: the difference between the true temperature and the~~

18

4.4 Test ~~runs~~retrievals

The test data described in the previous section was processed with several different regularisation setups and retrieval grids. Here we present the results for four different runs.

Firstly, as a reference, the latest version of JURRASIC2 without the implementations of any new algorithms described in this paper was used (retrieval A). It uses a rectangular grid that covers altitudes from 1 km to 64 km. The vertical spacing of the grid is 1 km between the altitudes of 1 km and 5 km, 250 m between altitudes of 5 km and 20 km, 1 km between altitudes of 20 km and 25 km and 4 km between 28 km and 64 km. As required for the rectilinear grid, the number and distribution of points is the same in each altitude (Figure ~~??~~2a). For this reference run, a first order regularisation (1) with trilinear interpolation (Bai and Wang, 2010) was used. The regularisation weights, as defined in (2), are given in Table 1. These are typically tuned ad-hoc (adjusted by trial-and-error until optimal retrieval results can be achieved), validated against model data and, when using real observations, in situ measurements.

Then, to evaluate the performance of second order regularisation, retrieval B was performed. The grid and interpolation methods were identical to retrieval A, but the regularisation was replaced with the second order scheme from equation (13) ~~.~~ and correlation lengths derived in section 4.3 from in situ observations. No subsequent tuning of these parameters was performed.

We compare the quality of different retrievals by inspecting the differences between the retrieved temperature and temperature used to generate the synthetic measurements ("true" temperature). These differences are shown on horizontal and vertical slices of the observed atmospheric volume in Figure 3. Figure 4 shows the variation of the retrieved and true temperatures on horizontal lines in order to visualise the response of retrievals to different spatial structures in true temperature.

Comparison of Figure 3 rows A and B shows the effect of the new regularisation (13). The new algorithm (retrieval B) demonstrates better agreement with the true temperature. It shows markedly less retrieval noise in the central area, as well as better results (less effect of a priori) just outside it, at the edges of the vertical cut shown in Figure 3. Considering also that, unlike A, the regularisation strength of B was not manually tuned for this test case in particular, we can conclude that the new regularisation performed better in this test case.

The third retrieval (C) was performed on the same set of grid points as B and with the same input parameters, but the grid was treated as irregular, i.e. a Delaunay triangulation was found, the algorithms described in section 2 were used for regularisation. A horizontal cut of the Delaunay triangulation for this grid is shown in Figure ~~??~~2a. By comparing B and C one can evaluate the agreement between the old interpolation, derivative calculation and volume integration techniques for rectilinear grids and their newly developed irregular-grid-capable alternatives. The temperature fields from retrievals B and C are very similar (Figure 3), which confirms that the new Delaunay triangulation based code gives consistent results when run on rectilinear grid.
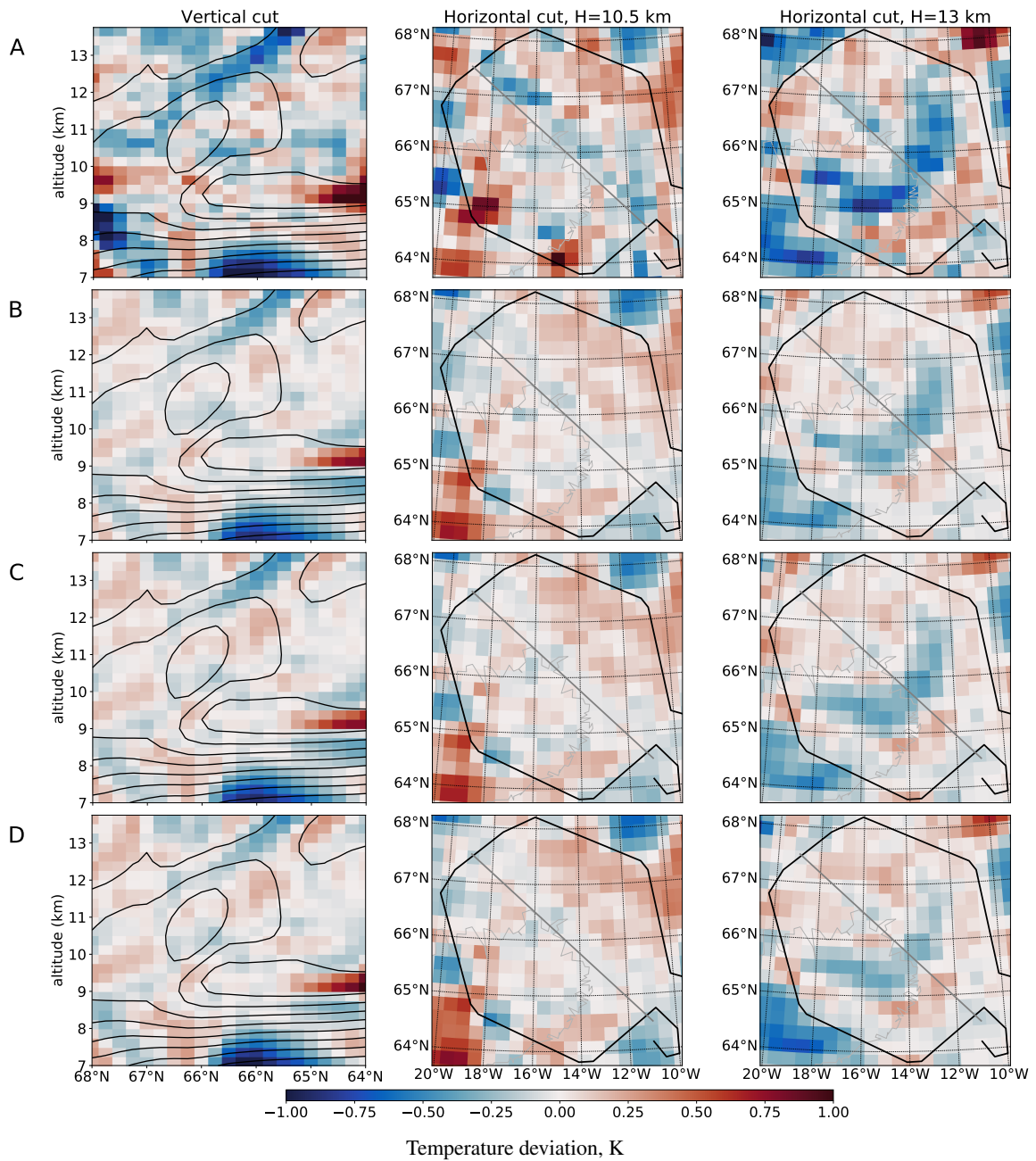
Figure 3. Panels show difference between retrieved temperature and true temperature of the simulated atmosphere. Rows A-D show results of respective retrievals. In the first column, black lines represent contours of a priori temperature value. In columns 2 and 3: black line – flight path, thick grey line – position of cut shown in the first column, thin grey line – topography (Icelandic coastline visible).
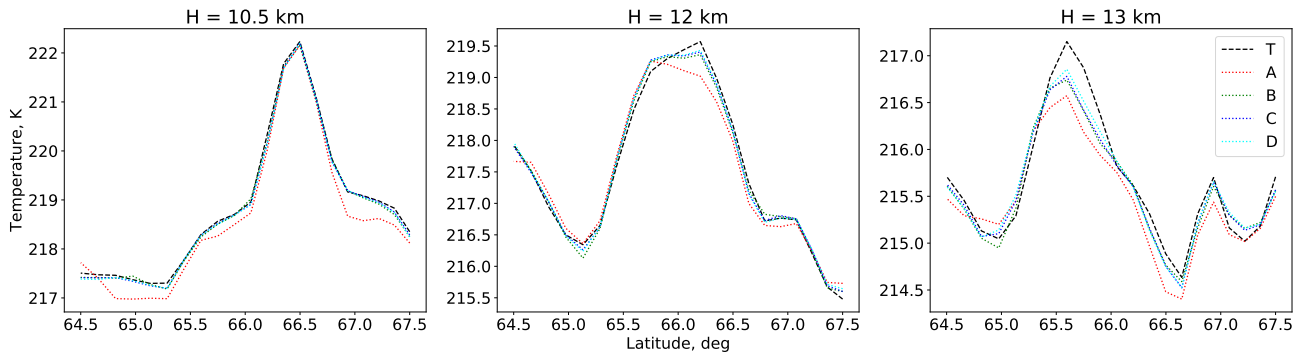
Figure 4. Retrieved temperature at different altitudes in the vertical cut shown in Figures 1 and 3. Line T represents the true temperature of the simulated atmosphere, A-D: respective retrievals.

Finally, retrieval D was performed with the same input parameters and methods as C, but using only the subset (22 %) of the original grid points to reduce computational cost. These points were chosen so that grid density would be highest in the volumes best resolved by GLORIA and sparser where little measurement data is available. In particular, a set of axially symmetric regions around the center of original retrieval (i.e. symmetric w.r.t. vertical axis at 66°N, 15°W) was defined as shown in Table 2. Points were removed from the original grid to achieve the specified vertical and horizontal resolutions for each region, resulting in the grid shown in Figure ~~??~~2. Results changed little compared to retrieval C. Most of the minor differences between C and D occur near the edges of the hexagonal flight path (Figure 3). Even though the grid spacing for retrieval D is still the same as C in some these areas, larger grid cells nearby have some effect on their neighbours. Note also that C and D retrievals tend to differ in the areas where the agreement between each of them and the true temperature is relatively poor and diagnostics predict lower accuracy (Figure 5), so D can be considered to be as good as C in well-resolved areas. Unphysical temperature structures far outside the measured area like the one partially seen at 64°N, 20°W are not unique to this retrieval, they just occur a little closer to the flight path in this case due to the coarser grid outside the hexagon.

4.5 Computational costs

JURASSIC2 finds an atmospheric state minimising (1) iteratively. In each iteration it first computes a Jacobian matrix of the forward model, hence obtains a Jacobian of the cost function (1) and then minimises the cost function using the Levenberg-Marquardt (Marquardt, 1963) algorithm and employing conjugate gradients (CG) (Saad (2003), Hanke (1995)) to solve the linear systems involved. The forward model is then run to simulate the measurements for the new iteration of atmospheric state.

All calculations were performed on the Jülich Research on Exascale Cluster Architectures (JURECA) supercomputer operated by the Jülich Supercomputing Centre (Jülich Supercomputing Centre, 2016). Each retrieval was executed in parallel on 4 computation nodes, each with 24 CPU cores (twin Intel Xeon E5-2680 v3 CPU's) and 128 GB of system main memory.
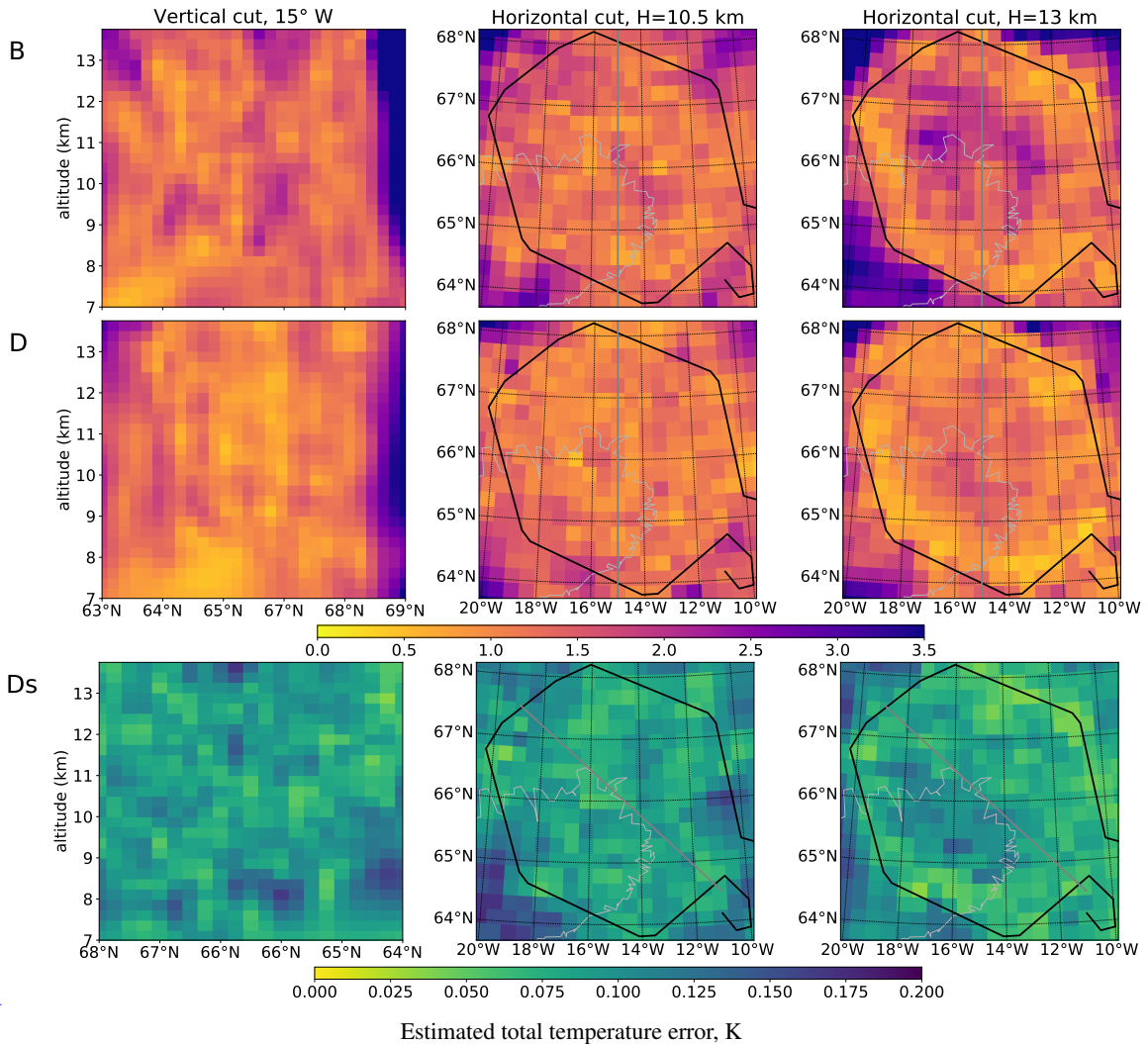
Figure 5. Panels B and D show the Monte Carlo estimate of the total temperature error ~~for~~ the retrievals B and D would have, if the measurements were real. Panel Ds shows an error estimate for the synthetic retrieval D (this assumes that unretrieved quantities, e. g. pressure, are known exactly). In columns 2 and 3: black line – flight path, thick grey line – position of cut shown in the first column, thin grey line – topography.

Table 3 shows the time required for different parts of calculation averaged over 3 identical runs (although < 1% variation was observed). The "Other calculations" field mostly includes time required by input and output, but also preparatory steps, such as Delaunay triangulation in the case of retrievals C and D.

The results in Table 3 reveal the impact of the new methods on computation times. The new, more complex regularisation results in ~~significantly~~ slower convergence in the conjugate gradient calculation, making retrieval B about 6% more expensive

than A. The Delaunay triangulation, however, decreases Jacobian calculation costs by roughly 25%, even the one used on the same grid as the old approach. This has to do with the fact that interpolation based on Delaunay triangulation only requires 4 values of the interpolated quantity at neighbouring points, and linear interpolation in three directions requires 8, thus resulting in more non-zero terms in the Jacobian. The time required for initialisation of Delaunay interpolation and derivatives, includ-

5   ing construction of the matrix $\mathbf{S}_a^{-1}$, increases rather fast with grid size with current implementation, contributing to "other calculations" of retrieval C. Further optimisation of this step is possible and will be implemented in the future. In total, the new retrieval (C) is 9% faster then the old approach (A) while using the same grid.

Table 3. Computational cost of retrievals (calculation times given in minutes)

Retrieval	A	B	C	D
Number of grid points	302526	302526	302526	54580
Jacobian calculation	134.4	133.2	101.2	61.3
Minimisation (CG)	4.2	14.9	12.8	6.0
Forward model	2.5	2.5	2.6	2.6
Other calculations	5.6	4.4	17	3.0
Total elapsed time	146.7	155.0	133.6	72.9

Retrieval D gives a further 45% reduction of computation time by removing 82% of points in the grid. A speedup of such order was expected. Only a minority of grid points contribute to a forward model (and hence Jacobian) calculations, and

10   JURASSIC2 code is heavily optimised for sparse matrix operations. Most of the grid points contributing to forward model are in the areas well resolved by the instrument and thus cannot be removed from the grid without compromising data products. Therefore, the actual savings in computation time mostly come from removing points above flight level and choosing a more appropriate non-rectangular shape for the densest part of the grid when measurement geometry requires that. Some of the points above flight path must be included since infrared radiation from there is measured by the instrument, but very little

15   information about atmosphere far above flight level can be retrieved, so very low grid densities are sufficient (Table 2).

4.6   Diagnostics

We follow Ungermann et al. (2011) and use Monte Carlo approach to obtain diagnostics, but replace Cholesky decomposition with the method developed in section 3. Results for retrievals B and D are presented in Figure 5 B, D. They show the total error estimate for temperature based on the sensitivity of temperature value to (simulated) instrument noise and all other

20   retrieved atmospheric quantities as well as expected error of non-retrieved quantities (e.g. pressure). These results tell us what temperature errors we could expect if the retrievals in question were performed with real data. The panel Ds shows an error estimate based on (simulated) instrument noise and uncertainties of retrieved parameters only. In the case of a synthetic retrieval, unretrieved parameters, upon which the simulated measurements are based, are known exactly. Therefore, panel Ds

is a direct Monte Carlo prediction of the temperature deviation we see in Figure 3D, and the magnitude of this deviation is indeed similar.

These results confirm that temperature can only be reliably retrieved within the flight hexagon or very close to the actual path. Also, the error estimates are higher in the central area of hexagon close to flight altitude, in agreement with test results (compare column 3 of Figures 3 and 5).

Monte Carlo retrieval results provided here were verified by estimating temperature error at one grid point using a more direct approach. Rodgers (2000) proves that retrieved atmospheric state x can be written in terms of the a priori state x_a, the (unknown) true state x_t and measurement error ϵ as

$$x = \mathbf{A}x_t + (\mathbf{I} - \mathbf{A})x_a + \mathbf{G}\epsilon \tag{32}$$

where

$$\mathbf{G} = \mathbf{M}^{-1}(\mathrm{D}\mathbf{F}(x))^T \mathbf{S}_\epsilon^{-1} \qquad \mathbf{A} = \mathbf{G}\,\mathrm{D}\mathbf{F}(x) \qquad \mathbf{M} = \mathbf{S}_a^{-1} + [\mathrm{D}\mathbf{F}(x)]^T \mathbf{S}_\epsilon^{-1}\mathrm{D}\mathbf{F}(x) \tag{33}$$

Here D denotes ~~Jacobian~~ the Jacobian operator and all other notation as in (1). The so-called *gain matrix* $\mathbf{G}$ fully describes the effect of measurement uncertainties ϵ on the retrieved state x. Computing $\mathbf{G}$ would be prohibitively expensive due to the required matrix inversion, but note that i'th element of x only depends on i'th line of the matrix $\mathbf{G}$, so one can use Krylov subspace methods (e.g. conjugate gradients) to solve $\mathbf{M}r_i = e_i$ for the i'th row r_i^T of the inverse matrix $\mathbf{M}^{-1}$. The i'th row of $\mathbf{G}$ can then be obtained as $r_i^T (\mathrm{D}\mathbf{F}(x))^T \mathbf{S}_\epsilon^{-1}$.

This approach allows for error estimation for a small number of grid points in reasonable time. We chose one point in the centre of the hexagon ($66°$ N, $15°$ W, 11.75 km altitude) for retrieval B. The total temperature error was found to be 1.27 K and agrees with Monte Carlo result (1.25 K) at ~~those points~~ this point within 2%, which is an excellent accuracy for an error estimate.

~~The results of Monte Carlo runs for retrievals B and D are presented in Figure 5. They clearly confirm that temperature can only be reliably retrieved within the flight hexagon or very close to the actual path. Also, the error estimates are higher in the central area of hexagon close to flight altitude, in agreement with test results (compare column 3 of Figures 3 and 5). Note that the actual deviations of retrieved temperatures (Figure 3 A-D) from the temperature values used to generate the synthetic measurements (depicted in Figure 1 T) tend to be significantly lower than the Monte Carlo estimate. This happens because Monte Carlo provides a rather cautious approximation for work with real data, which includes the uncertainties of unretrieved quantities taken from the models (e.g. pressure), but those quantities are treated as exact when generating the simulated atmosphere, and only measurement noise is introduced. The major spatial features of the Monte Carlo estimate, however, do agree well with the errors of the simulated retrieval results.~~

5 Conclusions

A new regularisation algorithm for 3D tomographic limb sounder retrievals employing second spatial derivatives (based on equation 13) was developed and implemented. The ~~unphysical~~ ad-hoc parameters (regularisation weights) of the previous first

spatial derivative based approach were replaced with physical quantities – retrieved quantity standard deviation and correlation lengths in vertical and horizontal directions. Unlike the regularisation weights, these quantities do not require as manual tuning for each retrieval and can be estimated from in situ measurements, model data and theoretical considerations. Tests with synthetic data showed slightly superior performance of the new algorithm, even compared to the old approach with tuned

5 parameters. A rather general technique (equation (10)) of adapting the tomographic retrieval for atmospheric regions of high anisotropy and spatial variability was also proposed.

A Delaunay triangulation based approach for retrievals on irregular grids was developed. By better adapting the grid to the retrieval geometry and thinning it out in the regions where lack of measurement data limits the resolution, the computation time of a tomographic retrieval was reduced by a total of 50% without significant deterioration of the results. Irregular grids

10 can be further employed for efficient tomographic retrievals in non-standard measurement geometries. At the time of writing, the methods developed in this paper were already in use to process GLORIA limb sounder data.

Monte Carlo Diagnostics were newly implemented for a 3D tomographic retrieval allowing for quick and reliable evaluation of measurement quality, identification of reliably resolved volumes and selection of optimal 3D tomography setups. Error estimates for retrieved atmospheric quantities can now be calculated at each grid point, while the previous approaches to

15 diagnostics only allowed to do that for a few selected points within similar time frame.

Appendix A: Regularisation tests

The equations (13) and (6) were used to construct precision matrices $\mathbf{S}_a^{-1}$ directly, since constructing a covariance matrix $\mathbf{S}_a$ explicitly and inverting it would be prohibitively computationally expensive. While the equation (13) was derived analytically, its discrete implementation cannot be exact, which raises the question whether our precision matrices still have required

20 properties, namely, whether they are still inverses of $\mathbf{S}_a$ that obeys equation (8). Clearly, this can only be verified directly for a small test case, where covariance matrices are small enough to be inverted. Also, employing matrix norms to compare matrices is not very useful in this case, as the directly computed $\mathbf{S}_a^{-1}$ is based on finite difference derivative estimates which do not work very well at grid boundaries. This is not a problem for the retrieval, which, by design, approaches a priori near grid boundaries, but it has strong effects on matrix norms. Instead, we compared the covariance matrices by using them to compute

25 the norms of several test vectors, that represent typical structures expected for that covariance.

Precision matrix $\mathbf{S}_a^{-1}$ was constructed for one entity on a regular 20x20x20 grid (grid constant 1 in each direction) using the same techniques as for retrieval B ($\sigma = 1$ and $L_v = L_h = L = 2$). Then a covariance matrix was explicitly constructed for the same grid using equation (8) and inverted to obtain the precision matrix $\hat{\mathbf{S}}_a^{-1}$. A set of vectors $\boldsymbol{x}_\phi$ – discrete representations of a "Gaussian wave packet" disturbance

30 $$\phi\left(\boldsymbol{r}\right) = \exp\left(-\frac{\|r\|^2}{d^2}\right)\cos\left(\boldsymbol{k}\cdot\boldsymbol{r}\right) \tag{A1}$$

was computed. The parameter d was chosen so that the amplitude of the wave would decay to at most 0.01 at the grid boundary, thus suppressing edge effects. Two different wavelengths, namely $\lambda = 15, 20$, in grid units, and were chosen and 50 $\boldsymbol{k}$ values

Table A1. Forward model tests

Measurement set	original + noise	A	C	D
$\Delta\left(\boldsymbol{y}, \boldsymbol{y}_{\mathrm{orig}}\right)$	0.0193	0.0035	0.0036	0.0047

Difference in forward model results, compared to original synthetic measurement set without noise. A-D refer to forward model results of respective retrievals

pointing at every possible direction were generated for each wavelength (i.e. $\|\boldsymbol{k}\| = 2\pi/\lambda$). Then, for each of the resulting 100 disturbances the norms $\|\boldsymbol{x}_\phi\|_p$ and $\|\boldsymbol{x}_\phi\|_t$, where $\|\boldsymbol{x}_\phi\|_p^2 = \boldsymbol{x}_\phi^T \mathbf{S}_a^{-1} \boldsymbol{x}_\phi$ and $\|\boldsymbol{x}_\phi\|_t^2 = \boldsymbol{x}_\phi^T \hat{\mathbf{S}}_a^{-1} \boldsymbol{x}_\phi$, were calculated, as well as their relative difference $\delta = 2|\|\boldsymbol{x}_\phi\|_p - \|\boldsymbol{x}_\phi\|_t|/(\|\boldsymbol{x}_\phi\|_p + \|\boldsymbol{x}_\phi\|_t)$. For the case $\lambda = 15$ we found that the mean δ value was 0.050, with standard deviation 0.0042, and for the case $\lambda = 20$ mean was 0.036, with standard deviation 0.0007. We believe that precision of this order is sufficient for regularisation purposes.

Appendix B: Applicability to 1D and 2D retrievals

The exponential covariance relation, given in equation (8), can be trivially extended to any dimension. It is used as a statistical model to many processes. The norm associated with this exponential covariance, given in equation (9), is, however, a 3D-specific result. Known 2D equivalents involve fractional powers of derivatives and therefore are not as easy to implement numerically. A norm associated with 1D exponential covariance can be shown (**?**) to have the simple form

$$\|\phi\|^2 = \frac{1}{2\sigma^2 L}\left(\int_{x_1}^{x_2}\phi^2\mathrm{d}x + L^2\int_{x_1}^{x_2}\left(\frac{\partial\phi}{\partial x}\right)^2\mathrm{d}x\right) \tag{B1}$$

neglecting some boundary terms. Here the scalar field $\phi(x)$ represents the difference between the quantity to be retrieved and a priori, and L is the correlation length. All the main theoretical results and numerical methods of this paper can therefore be reproduced for 1D inverse modelling problems. The usefulness of our methods for practical 1D applications ultimately depends on the size of the problem in question. 1D limb sounding retrievals, for example, are usually too small for the methods developed here to be useful. It would typically be possible, and indeed simpler, to explicitly construct and invert the covariance matrix of a 1D problem, rather than discretise the associated norm (B1). In the case of a very large 1D problem with strong covariance, however, one may find the methods presented in this paper more lucrative.

Appendix C: Forward model tests

As pointed out in section 4.2, the same measurement set, synthetically generated on dense grid, was used for retrievals A-D. This provides an easy way to verify if the forward model is performing consistently on different grids: run it on each grid, based on the "true" atmospheric state and compare the results with the synthetic measurements from the dense grid.

In our test case, the radiances are calculated for $n = 31$ spectral windows. If vector $\boldsymbol{y}$ represents a set of radiances, and $\boldsymbol{y}_i$ represents the radiances at i-th spectral window, we evaluate the difference between those vectors with the parameter $\Delta(\boldsymbol{y}, \boldsymbol{y}') = (1/n) \sum_i^n 2 \|\boldsymbol{y}_i - \boldsymbol{y}'_i\|_2 / (\|\boldsymbol{y}_i\|_2 + \|\boldsymbol{y}'_i\|_2)$ where $\|\cdot\|_2$ is the euclidean norm. The results of this comparison are presented in Table A1. Forward model for retrievals A and C used the same grid, but A used trilinear interpolation, and C used Delaunay interpolation. We can see that the change in interpolation does not introduce significant errors (compared to those introduced by the grid change). Thinned out grid and Delaunay interpolation was used for grid D. Note that the simulated measurements that were actually used also have simulated instrument noise added to them. The results clearly show that the effects of different discretisation are well below iNote that the simulated measurements that were actually used also have simulated instrument noise added to them instrument noise level and are therefore acceptable. Also, the error is lower than our generally assumed forward model error of 1-3% (Hoffmann, 2006).

References

Allen, E., Baglama, J., and Boyd, S.: Numerical approximation of the product of the square root of a matrix with a vector, Linear Algebra and its Applications, 310, 167–181, doi:10.1016/s0024-3795(00)00068-9, 2000.

Bai, Y. and Wang, D.: On the Comparison om Trilinear, Cubic Spline, and Fuzzy Interpolation Methods in the High-Accuracy Measurements, IEEE Transactions on Fuzzy Systems, 18, 1016–1022, doi:10.1109/TFUZZ.2010.2064170, 2010.

Birner, T.: Fine-scale structure of the extratropical tropopause region, J. Geophys. Res., 111, doi:10.1029/2005JD006301, 2006.

Boissonnat, J.-D. and Yvinec, M.: Algorithmic Geometry, Cambridge University Press, New York, NY, USA, 1998.

Carlotti, M., Dinelli, B. M., Raspollini, P., and Ridolfi, M.: Geo-fit approach to the analysis of limb-scanning satellite measurements, Appl. Optics, 40, 1872–1885, doi:10.1364/AO.40.001872, 2001.

Charney, J. G.: Geostrophic Turbulence, Journal of the Atmospheric Sciences, 28, 1087–1095, doi:10.1175/1520-0469(1971)028<1087:GT>2.0.CO;2, 1971.

Curtis, A. R.: Discussion of 'A statistical model for water vapour absorption' by R. M. Goody, Quart. J. Roy. Meteorol. Soc., 78, 638–640, 1952.

Delaunay, B.: Sur la sphere vide, Bull. Acad. Sci. USSR: Class. Sci. Math., 7, 793–800, 1934.

Fehlberg, E.: Some old and new Runge-Kutta formulas with stepsize control and their error coefficients, Computing, 34, 265–270, doi:10.1007/BF02253322, 1985.

FitzGerald, C.: Optimal indexing of the vertices of graphs, Mathematics of Computation, 28, 825–831, doi:10.2307/2005704, 1974.

Forster, P. M. d. F. and Shine, K. P.: Radiative forcing and temperature trends from stratospheric ozone changes, J. Geophys. Res., 102, 10 841–10 855, doi:10.1029/96JD03510, 1997.

Friedl-Vallon, F., Gulde, T., Hase, F., Kleinert, A., Kulessa, T., Maucher, G., Neubert, T., Olschewski, F., Piesch, C., Preusse, P., Rongen, H., Sartorius, C., Schneider, H., Schönfeld, A., Tan, V., Bayer, N., Blank, J., Dapp, R., Ebersoldt, A., Fischer, H., Graf, F., Guggenmoser, T., Höpfner, M., Kaufmann, M., Kretschmer, E., Latzko, T., Nordmeyer, H., Oelhaf, H., Orphal, J., Riese, M., Schardt, G., Schillings, J., Sha, M. K., Suminska-Ebersoldt, O., and Ungermann, J.: Instrument concept of the imaging Fourier transform spectrometer GLORIA, Atmos. Meas. Tech., 7, 3565–3577, doi:10.5194/amt-7-3565-2014, 2014.

Gettelman, A., Hoor, P., Pan, L. L., Randel, W. J., Hegglin, M. I., and Birner, T.: The extra tropical upper troposphere and lower stratosphere, Rev. Geophys., 49, RG3003, doi:10.1029/2011RG000355, 2011.

Godson, W. L.: The evaluation of infra-red radiative fluxes due to atmospheric water vapour, Quart. J. Roy. Meteorol. Soc., 79, 367–379, 1953.

Gordley, L. L. and Russell, J. M.: Rapid inversion of limb radiance data using an emissivity growth approximation, Appl. Optics, 20, 807–813, doi:10.1364/AO.20.000807, 1981.

Hanke, M.: Conjugate gradient type methods for ill-posed problems, John Wiley & Sons, 1995.

Haynes, P. and Anglade, J.: The Vertical-Scale Cascade in Atmospheric Tracers due to Large-Scale Differential Advection, J. Atmos. Sci., 54, 1997.

Hegglin, M. I., Boone, C. D., Manney, G. L., and Walker, K. A.: A global view of the extratropical tropopause transition layer from Atmospheric Chemistry Experiment Fourier Transform Spectrometer O_3, H_2O, and CO, J. Geophys. Res., 114, D00B11, doi:10.1029/2008JD009984, 2009.

Hoffmann, L.: Schnelle Spurengasretrieval für das Satellitenexperiment Envisat MIPAS, vol. 4207 of *Berichte des Forschungszentrums Jülich*, Forschungszentrum, Zentralbibliothek, Jülich, record converted from VDB: 12.11.2012; Wuppertal, Univ., Diss., 2006, 2006.

Hoffmann, L., Kaufmann, M., Spang, R., Müller, R., Remedios, J. J., Moore, D. P., Volk, C. M., von Clarmann, T., and Riese, M.: Envisat MIPAS measurements of CFC-11: retrieval, validation, and climatology, Atmospheric Chemistry and Physics, 8, 3671–3688, doi:10.5194/acp-8-3671-2008, https://www.atmos-chem-phys.net/8/3671/2008/, 2008.

Jülich Supercomputing Centre: JURECA: General-purpose supercomputer at Jülich Supercomputing Centre, Journal of large-scale research facilities, 2, doi:10.17815/jlsrf-2-121, http://dx.doi.org/10.17815/jlsrf-2-121, 2016.

Krisch, I., Preusse, P., Ungermann, J., Dörnbrack, A., Eckermann, S. D., Ern, M., Friedl-Vallon, F., Kaufmann, M., Oelhaf, H., Rapp, M., Strube, C., and Riese, M.: First tomographic observations of gravity waves by the infrared limb imager GLORIA, Atmospheric Chemistry and Physics, 17, 14 937–14 953, doi:10.5194/acp-17-14937-2017, https://www.atmos-chem-phys.net/17/14937/2017/, 2017.

Krisch, I., Ungermann, J., Preusse, P., Kretschmer, E., and Riese, M.: Limited angle tomography of mesoscale gravity waves by the infrared limb-sounder GLORIA, Atmospheric Measurement Techniques Discussions, 2018, 1–24, doi:10.5194/amt-2018-72, https://www.atmos-meas-tech-discuss.net/amt-2018-72/, 2018.

Lim, S. and Teo, L.: Generalized Whittle–Matérn random field as a model of correlated fluctuations, Journal of Physics A: Mathematical and Theoretical, 42, 105 202, 2009.

Marquardt, D. W.: An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial and Applied Mathematics, 11, 431–441, doi:10.1137/0111030, 1963.

Preusse, P., Eckermann, S. D., and Ern, M.: Transparency of the atmosphere to short horizontal wavelength gravity waves, J. Geophys. Res., 113, doi:10.1029/2007JD009682, 2008.

Riese, M., Ploeger, F., Rap, A., Vogel, B., Konopka, P., Dameris, M., , and Forster, P.: Impact of uncertainties in atmospheric mixing on simulated UTLS composition and related radiative effects, J. Geophys. Res., 117, D16 305, doi:10.1029/2012JD017751, 2012.

Riese, M., Oelhaf, H., Preusse, P., Blank, J., Ern, M., Friedl-Vallon, F., Fischer, H., Guggenmoser, T., Höpfner, M., Hoor, P., Kaufmann, M., Orphal, J., Plöger, F., Spang, R., Suminska-Ebersoldt, O., Ungermann, J., Vogel, B., and Woiwode, W.: Gimballed Limb Observer for Radiance Imaging of the Atmosphere (GLORIA) scientific objectives, Atmos. Meas. Tech., 7, 1915–1928, doi:10.5194/amt-7-1915-2014, 2014.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, vol. 2 of *Series on Atmospheric, Oceanic and Planetary Physics*, World Scientific, Singapore, 2000.

Saad, Y.: Iterative Methods for Sparse Linear Systems, Society for Industrial and Applied Mathematics, second edn., doi:10.1137/1.9780898718003, 2003.

Sparling, L. C., Wei, J. C., and Avallone, L. M.: Estimating the impact of small-scale variability in satellite measurement validation, Journal of Geophysical Research: Atmospheres, 111, doi:10.1029/2005JD006943, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006943, 2006.

Steck, T., Höpfner, M., von Clarmann, T., and Grabowski, U.: Tomographic retrieval of atmospheric parameters from infrared limb emission observations, Appl. Optics, 44, 3291–3301, doi:10.1364/AO.44.003291, 2005.

Tarantola, A.: Inverse Problem Theory Methods for Data Fitting and Model Parameter Estimation, Elsevier Science, 2013.

Tikhonov, A. N. and Arsenin, V. Y.: Solutions of ill-posed problems, Winston, Washington D.C., USA, 1977.

Ungermann, J.: Improving retrieval quality for airborne limb sounders by horizontal regularisation, Atmospheric Measurement Techniques, 6, 15–32, doi:10.5194/amt-6-15-2013, https://www.atmos-meas-tech.net/6/15/2013/, 2013.

Ungermann, J., Kaufmann, M., Hoffmann, L., Preusse, P., Oelhaf, H., Friedl-Vallon, F., and Riese, M.: Towards a 3-D tomographic retrieval for the air-borne limb-imager GLORIA, Atmos. Meas. Tech., 3, 1647–1665, doi:10.5194/amt-3-1647-2010, 2010.

Ungermann, J., , J., Lotz, J., Leppkes, K., Hoffmann, L., Guggenmoser, T., Kaufmann, M., Preusse, P., Naumann, U., and Riese, M.: A 3-D tomographic retrieval approach with advection compensation for the air-borne limb-imager GLORIA, Atmos. Meas. Tech., 4, 2509–2529, doi:10.5194/amt-4-2509-2011, 2011.

Weinreb, M. P. and Neuendorffer, A. C.: Method to Apply Homogeneous-path Transmittance Models to Inhomogeneous Atmospheres, J. Atmos. Sci., 30, 662–666, 1973.

Worden, J. R., Bowman, K. W., and Jones, D. B.: Two-dimensional characterization of atmospheric profile retrievals from limb sounding observations, Journal of Quantitative Spectroscopy and Radiative Transfer, 86, 45–71, 2004.

5