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 $\mathrm{H}\mathbf{x}_\varphi$, then the discrete equivalent to the integral in equation (5) would be $\mathbf{x}_\phi^T \mathbf{S}_\mathrm{a}^{-1} \mathbf{x}_\varphi$. 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.

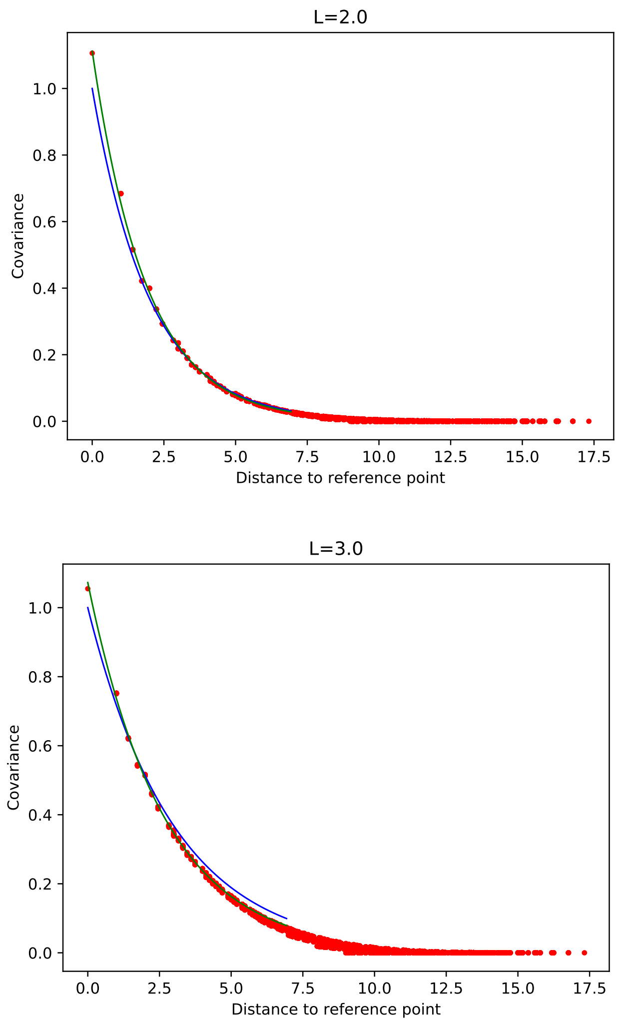

Figure 1: Covariance to the reference point as a function of distance to the reference point for correlation lengths $L = 2$ and $L = 3$. Blue line – theoretical covariance value, green line – least squares fit for exponential covariance.

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.