The authors thank both reviewers for their careful reading of the manuscript and the helpful suggestions which have added clarity to the paper.

Review 1

Comment: *Overview*
The paper draws attention to a hitherto overlooked problem with the application of averaging kernel matrices, specifically that the AK matrix itself, A, has some dependence on the retrieved state, x. Hence, when using averaged data, ¡x¿ (e.g., a monthly mean), the appropriate averaging kernel ¡Ax¿ is not simply ¡A¿¡x¿ constructed from an average AK matrix ¡A¿.

Main Comments:

The paper is concise, well-argued and includes a suitable illustrative example, although I have some suggestions below as to how it might be further clarified so as to reach a wider audience...

Reply: We thank the reviewer for the appreciation of our work.

Action: See actions in reply to the specific comments.

Comment: *... While the main recommendation of the paper is that the data providers should also provide a correction term for the mean averaging kernel, it does seem more practical if, instead, the data providers themselves use the guidance in this paper to produce a suitable averaging kernel to accompany the averaged products. We all know that it is a struggle to get data user to understand and apply an averaging kernel, so we should avoid making their task any more complicated.*

Reply: The adequacy of the correction vector can be deduced with mathematical rigour. Since the correction vector is additive while a modified averaging kernel would modify the mean profile in a multiplicative manner, we doubt that it is possible to find a modified averaging kernel which does in effect the same as our suggested additive correction. We think that work with scientific data is best done in cooperation with the data providers, who will then help the data users to apply the averaging kernel and the correction term correctly.

Action: None

Comment: *...I doubt if this paper will be the last word in the matter – there are a number of open issues which require a little more thought, such as logarithmic retrievals, retrievals of temperature/pointing/pressure, non-constant a priori data, averaging kernel matrices which are not square. However, this paper is a good starting point for the conversation.*

Reply: We do agree that this paper can only be a first step.

Action: None

Comment: *Minor Comments*

1) The authors frequently resort to Latin. Personally I find it a welcome change from the usual stock phrases, although I expect some readers may not be quite so appreciative.

Reply: We have gone through the manuscript and checked where it seemed appropriate to replace the latin terms by English ones.

Action:
a priori (various places): This is so common that it is not even printed in italic font in Copernicus journals; thus we have decided not to change it.
viz. (p1 l18): has been replaced with 'namely'.
mutatis mutandis (p3 l11): has been replaced with 'in an analogue way'.
cum grano salis (p3 l17): has been replaced with "with some qualification..."
a fortiori (p8 l3): has been replaved with 'even more'.

Comment: *2) Abstract (and elsewhere): reference to 'covariance profile' although the suggested correction is a matrix rather than a profile.*

Reply: Here we respectfully disagree. The correction is a column vector. It results from a product of an $n \times n$ square matrix with a profile (see, e.g., Eq. 7).

Action: None

Comment: *3) P1, L17: 'this is does not make sense here.*

Reply: Agreed.

Action: Has been replaced with 'that is to say'

Comment: *4) P1, L20: I don't think the use of monthly means requires any references, although no doubt Hegglin and Tegtmeier will appreciate being selected for multiple citations from among many, many such users.*

Reply: To our knowledge, the SPARC data initiative was the first large-scale international activity where the work was entirely monthly zonal means. And, beyond this, it was during the SPARC Data Initaive discussions where the title paper of the problem emerged. We agree that the original manuscript includes an over-exaggerated number of related references.

Action: The number of references has been reduced to a single one.

Comment: *5) P1, L20: suggest 'their rather than her.*

Reply: agreed.

Action: Changed as suggested.

Comment: *6) P2, L13: The casual reader may interpret this comment as suggesting none of this applies to non-linear, iterative retrievals, so I suggest rewording to emphasise that it still does.*

Reply: We agree.

Action: A clarifying sentence has been added.

Comment: *7) P2, Eq 2: It could be pointed out that the main dependence on $\vec{x}$ in the AK matrix comes from the Jacobian matrix,* $\mathbf{K}$ *(although possibly also from* $\mathbf{R}$ *if some form of adaptive regularization is used), so whether or not there is any dependence of* $\mathbf{A}$ *on $\vec{x}$ is usually a consequence of whether or not* $\mathbf{K}$ *depends on $\vec{x}$.*

Reply: Agreed.

Action: A statement has been added after Eq (2) and the following statement has been slightly modified because otherwise it might no longer be clear what 'this' refers to.

Comment: *8) P2, L25: An extra equation, $\vec{y} - \mathbf{F}(\vec{x}_a) = \mathbf{K}(\vec{x} - \vec{x}_a)$ would help the reader get from eq (2) to eq (3).*

Reply: Since $\mathbf{F}(\vec{x})$ might be nonlinear, this extra equation does not necessarily hold in a general sense, although we see its didactic value.

Action: We have added: With the averaging kernel matrix introduced above, and using the linearization

$$\vec{y} - \mathbf{F}(\vec{x}_\mathrm{a}) \approx \mathbf{K}(\vec{x} - \vec{x}_\mathrm{a}), \quad (1)$$

Comment: *9) P2, L28 onwards. This is confusing. Elsewhere averaging kernels are discussed as a characteristic of the lower-resolution (satellite) retrievals, but in this example (Eq 4) the averaging kernel seems to be on the grid of the higher resolution 'original' retrieval. Despite the similarity of Eq 4 and Eq 3, these seem to be two quite different things.*

Reply: We assume that the averaging kernel of the coarser resolved data set has been evaluated on a grid fine enough to do this transformation. For example, MIPAS averaging kernels are evaluated on a 1-km altitude grid although the MIPAS altitude resolution is typically only about 3 km. This allows application of Eq. 4 to, e.g., model data sampled on a 1-km grid. We realize that our terminology in the paper might lead the reader astray: Our $x_{original}$ is any high-resolved profile to be degraded. It is NOT our original retrieval. In the application described above, $x_{original}$ was meant to be the high-resolution model data to be degraded.

Action: We have changed our terminology and use the subscript "high-resolution" to avoid this kind of misunderstanding. We have also added some prose for clarification.

Comment: *10) P3, Eq 6 and elsewhere: if this is prepared with LaTeX, I suggest using nlangle and nrangle rather than* $<$ *and* $>$ *for the angle-brackets.*

Reply: ok, thanks!

Action: done as suggested.

Comment: *11) P3, Eq 7: it would be helpful to further simplify this here, giving* $\mathrm{cov}(\mathbf{A}, \tilde{\mathrm{x}}) = \langle \mathbf{A}\tilde{\mathrm{x}} \rangle - \langle \mathbf{A} \rangle \langle \tilde{\mathrm{x}} \rangle$ *which makes Eq 6 clearer.*

Reply: Agreed.

Action: Done as suggested.

Comment: *12) It does not help that many of these equations are split over two lines, but that may not be the choice of the authors.*

Reply: This indeed looks odd in the one-column format of the discussion paper. But the final paper will be formatted in two columns, and to avoid errors in the equations when the manuscript will be transformed from a one-colum format to a two-column format, where many of these equations have to be split over two lines, we have chosen a format for the equations which will be compatible also with the two-column style file.

Action: None.

Comment: *13) P3, L22: There seems to be more to be said than this simple phrase 'an individual prior x_a'. For example, an individual but* **almost** *constant a priori could be used for each profile, in which case Eq (9) applies rather than Eq (11). The key is obviously what sort of 'individual prior' leads to the two covariances being approximately equal.*

Reply: We had two different applications in mind:

1. The retrieval scientist might use climatologies as a priori. This is often done, and Rodgers' account of the smoothing error seems to suggest that this kind of priori information can be used. In this case Eq(9) applies.
2. In some other cases, the best available prior information for each individual retrieval can be used. For example, for MIPAS we use meteorological analyses as a priori for temperature retrievals (we constrain only the shape, not the values, but that is another story...). Eq (10) has been tailored for this kind of applications.

We agree that the entire 'grey scale' between these extremes deserves some discussion.

Action: Some text has been added.

Comment: *14) P4, Eq (11): In this case I think the extra equation confuses (especially when split over multiple lines) rather than clarifies. Perhaps better to refer back to Eq (6) and simply state the simplified result.*

Reply: The purpose of Eq (11) is to make clear which of the terms cancel out with the given assumptions in force. This is not easily seen in Eq (6).

Action: None.

Comment: *15) P4, L26: nbar r seems to be introduced in the wrong place in this sentence, presumably it should be after 'normalized covariance term'*

Reply: yes, indeed, thanks for spotting.

Action: corrected.

Comment: *16) P4, L14: I'm surprised that this produces stable results, eg for HCN at higher altitudes, where the ¡x¿ in the denominator would tend to zero. Covariance terms, as in the Pearson correlation, are usually scaled by the square root of the variance, so don't have this problem.*

Reply: Well, rather by the product of both standard deviations involved than by the variance, but we see your point. Our normalized covariance can be interpreted as a relative error. It is admittedly unstable, similarly as relative or percentage errors in the case of small measured values (and in the latter case nobody complains). A normalized quantity calculated along the idea of Pearson's correlation coefficient would be more stable but it would require an entirely different interpretation. Since we are interested in the relative error related with ignoring the covariance effect, we think that our variant is more adequate.

Action: None.

Comment: *17) P5, L17: 'recommend' (spelling).-*

Reply: Thanks for spotting.

Action: Corrected.

Review 2

Comment: *This manuscript discusses an important and often ignored issue involving the application of averaging kernels to mean profiles. A solution to the problem is presented where the covariance between the averaging kernel and the atmospheric state is calculated. Examples are shown applying the method to MIPAS, and recommendations are given to data producers of monthly zonal mean data.*
The manuscript is well written and suitable for publication in AMT after a few comments are taken into account.

Reply: We thank the reviewer for this positive evaluation.

Action: See actions in reply to the specific comments.

Comment: *General Comments*
The discussion and conclusion (including the recommendations) of the paper focuses on the ideal case where the data producer actually calculates (and stores) an averaging kernel for each individual profile. It is somewhat common to only produce representative averaging kernels and perhaps use them as a metric for retrieval performance in a validation/retrieval paper or data quality document. Would a possible recommendation of this work be that a few of these covariance terms should be calculated and included as an assessment of the data quality?

Reply: These covariance terms refer to mean averaging kernels. We have no idea to which degree they are applicable to representative individual averaging kernels. Since the covariance terms depend on the ensemble over which it is averaged, and since it may not be known in advance what kind of averages the data user wants, it will not be easily possible to produce useful covariance profiles in advance. Instead one might consider to calculate the averaging kernels for each retrieval (which is not so much additional effort) and to calculate zonal mean profiles and averaging kernels immediately as the last step of each individual retrieval. In this case one would have the mean averaging kernels and the covariances without storing each averaging kernel. They can be deleted immediately after their consideration for the mean values.

Action: None

Comment: *Related to the above point, I have to wonder, is the covariance profile useful beyond a correction when applying the mean averaging kernel? My (perhaps wrong) interpretation is that when the covariance profile is 0, the mean of the retrieved profile is a smoothed version of the true mean atmospheric state. I suppose what I am asking is that if the covariance profile is not 0, is it wrong to interpret the retrieved mean as a smoothed version of the true atmospheric mean? If so, I would like to see a discussion of this included in the manuscript.*

Reply: We are not aware of any other interpretation beyond the one offered. The interpretation that zero covariance means that the mean of the retrieved profiles is a smoothed version of the true mean atmospheric state is not true, at least not in a general sense. Assume a case with infinite noise, i.e., no measurement information. The retrieval will then be identical to the a priori. Assume further that a constant (e.g. climatological) a priori has been chosen. The result will not vary at all. Thus also the covariance will be zero although the mean result is fully determined by the a priori, in shape and values.

Action: None

Comment: *Minor Comments*
p.1 l.9: ". . . on a given altitude grid . . . "
Here and throughout this section it is written that altitude is the vertical coordinate, however all of the arguments should equally apply to any vertical coordinate.

Reply: With "altitude" we mean any vertical coordinate, not only geometrical altitude.

Action: We have added a footnote explaining that we mean altitude in a more general sense, not limited to geometrical altitude.

Comment: *p.2 l.18: "For a constrained retrieval of the type" The way this is presented the reader may assume that what follows only applies to retrievals applying a (possibly iterative) form of eq. 1, when the concepts here are more general.*

Reply: We agree, but this type or retrieval is the only one for which averaging kernels are reported at all. But you are right, in principle our arguments hold for any other type of retrievals if averaging kernels are made available. These can be evaluated, e.g., by perturbation studies.

Action: We have added "...or any equivalent formulation of it."
Comment: *p.2 l.29: eq. 4: Somewhere here I would like to see a brief mention that $x_o riginal$ needs to be converted to the same grid and representation*

(vmr/number density and altitude/pressure) as the retrieval.

Reply: agreed.

Action: we have added "It goes without saying that the high-resolved profile has to be resampled on the grid on which the application of the averaging kernel is performed, and, if applicable, transformed to the same units (volume mixing ratio, number density, etc."

Comment: *p.3 l.5: "Calculation of zonal averages over L profiles . . . " Why restrict to zonal?*

Reply: This is meant as an example.

Action: We have added 'e.g.'.

Comment: *p.3 l.12: "For a retrieval with $\vec{x}_a = 0$ . . . " This is a nitpick and I dont necessarily think it should be changed, but the same would be true with $\vec{x}_a$=constant and a Tiknonov regularized retrieval. I guess the general condition would be if $\vec{x}_a$ is in the null space of* **R**.

Reply: Ok, indeed for an altitude-constant prior and an **A** with unity measurement response in all altitudes, we have $\langle \mathbf{A} \rangle \langle \vec{x}_\mathrm{a} \rangle = \langle \vec{x}_\mathrm{a} \rangle$, which cancels with the first $\langle \vec{x}_\mathrm{a} \rangle$ term. And a covariance involving a time-constant $\vec{x}_\mathrm{a}$ will also be zero.

Action: We have added "The same is true if for all retrievals the same altitude-constant prior is used in combination with an averaging kernel with unity row sums as associated with purely smoothing constraints.".

Comment: *p.3 l.22: "For a retrieval where an individual prior $\vec{x}|texta$ is used for each profile . . . " I suppose this assumes that the prior used is a good representation of the true atmospheric state/variability.*

Reply: Yes, this is indeed a precondition for Eq 10.

Action: We have added "...i.e., that the prior information is a good representation of the true atmospheric state and variability."

Comment: *p.3 l.15: "cov($\mathbf{A}; \vec{x}$) and be approximated by cov($\mathbf{A}; \hat{\vec{x}}$)" I have a hard time intuitively understanding the implications of this approximation. I think that there are two things going on here, the first is the switch from the true state to the smoothed state, which I don't expect to have a large effect. But since the intention is to use this to compare two measurements, are we also assuming that both instruments have approximately equal sampling within whatever bin is being averaged?*

Reply: Yes, we assume either identical or representative sampling. If the sampling is representative with respect to mean and variance, then different sampling should not be an issue.

Action: None

Comment: *p.4 l.8: "For retrievals performed in the log-space, all this becomes slightly more complicated . . . " It is fine to ignore the issues with log retrievals, since, as stated, averaging may have its own issues, but I have to wonder is this not a more general representation issue? Presumably if our goal was to compare a high resolution and a low resolution retrieval that both operated in log space, it would be possible using this framework if the averaging was done in log space.*

Reply: Averaging in the log-space typically does not solve related problems; particularly it will not remove biases introduced by the retrieval in the log space (see, Funke and von Clarmann, 2012).

Action: None.

Comment: *p.4 l.10: eq. 12 Perhaps related to above, but this equation is hard to interpret when the x's do not represent the same thing (some are in linear space some are logarithmic). Or maybe all the x's are intended to be in linear space and the logarithm being applied to xoriginal is missing?*

Reply: The latter is the case; it should read $Ax_{original}$. Thanks for spotting!

Action: Corrected.

Comment: *p.7 l.12: "The covariance effects can exceed 10% and thus need to be considered when mean profiles are used for quantitative analysis and mean averaging kernels are applied." This statement had me wondering about the implications of this effect beyond comparisons of two measurements. Say a data user is using zonally averaged MIPAS HCN data, but not actually applying any mean averaging kernel. Would having knowledge of the magnitude of this covariance term guide them in their analysis, similar to the way having a measure of vertical resolution from the averaging kernel would?*

Reply: We do not have any idea how to use the covariance term for other purposes than that described in the paper.

Action: None.

Comment: *Technical Comments*
p.4 l.18: eq. 13 Equation has extra equal signs.

Reply: Thanks for spotting

Action: Corrected.

Comment: *p.7 l.3: "consistes" consistes ! consists*

Reply: Thanks for spotting!

Action: corrected.

Comment: *p.7 l.17: "we recommed" recommed ! recommend*
Reply: *Thanks for spotting.*

Action: *Corrected.*

The application of mean averaging kernels to mean trace gas distributions

Thomas von Clarmann and Norbert Glatthor

Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research,
Karlsruhe, Germany

Correspondence: thomas.clarmann@kit.edu

Abstract. To avoid unnecessary data traffic it is sometimes desirable to apply mean averaging kernels to mean profiles of atmospheric state variables. Unfortunately, application of averaging kernels and averaging are not commutative in cases when averaging kernels and state variables are correlated. That is to say, the application of individual averaging kernels to individual profiles and subsequent averaging will, in general, lead to different results than averaging of the original profiles prior to the application of the mean averaging kernels unless profiles and averaging kernels are fully independent. The resulting error, however, can be corrected by subtraction of the covariance between the averaging kernel and the vertical profile. Thus it is recommended to calculate the covariance profile along with the mean profile and the mean averaging kernel.

1 Introduction

More often than not satellite data retrievals are constrained because the unconstrained profile retrieval on a given altitude[1] grid would lead to an ill-posed inverse problem. The constrained retrieval is more robust but the price to pay typically is, among other effects, a certain loss in vertical resolution. The effect of the constraint is characterized by the averaging kernel matrix (Rodgers, 2000).

Many applications of remotely sensed data involve comparison with independent model or independent measurement data. If these comparison data are better resolved than the remotely sensed data, the averaging kernel of the latter has to be applied to the former to make the comparison meaningful (Connor et al., 1994). Otherwise differences caused by the different altitude resolution would mask scientifically significant differences. Unfortunately, for a vertical profile of n values of an atmospheric state variable, the related averaging kernel matrix is of the size $n \times n$, ~~this is~~ that is to say, the data traffic is dominated by the averaging kernel data while the data product of interest, ~~viz.~~ namely the profile, could be communicated with much less effort. Often the data users are not interested in the individual measurements but prefer to work, e.g., with monthly zonal mean profiles ~~(Hegglin and Tegtmeier, 2011; von Clarmann et al., 2012; Hegglin et al., 2013; Hegglin and Tegtmeier, 2017; Tegtmeier et al., 2013, 2016~~ In this case, it would be convenient if the data user could simply apply monthly zonal mean averaging kernels to ~~her~~ their better resolved monthly zonal mean data to make them comparable to the coarser resolved zonal monthly mean measurements.

[1] We use the term altitude in a generic sense, not limited to geometric altitude but for any vertical coordinate. Alternative vertical coordinates can be, e.g., pressure or potential temperature.

Unfortunately averaging and application of the averaging kernel are not commutative. As soon as the data and the averaging kernels covary, the application of the mean averaging kernel to mean profiles gives a different result than the application of individual averaging kernels prior to averaging. We solve this problem by providing statistically inferred covariance terms which can be used to correct the related error. In the next section we describe the theoretical framework used. As a case study, covariances applicable to trace gas profiles retrieved from MIPAS (Michelson Interferometer for Passive Atmospheric Sounding, Fischer et al. 2008) measurements are inferred in Section 3. The varying importance of the covariance effect is illustrated in Section 4. Section 5 is an interlude where we investigate into pitfalls regarding the applicability of averaging kernels to comparison data, before a critical discussion of the applicability of our suggested appproach concludes the paper (Section 6).

2 The formal concept

We borrow the formal concept of retrieval theory from Rodgers (2000). The intended application of our study is, at worst, moderately nonlinear retrievals. That is to say, linear theory is assumed to be adequate for the characterization of the retrieval in terms of error estimation, assessment of vertical resolution, and so forth. Thus, we ignore all complication which may arise from non-linearity and thus do not discuss the retrievals in an iterative setting. Within the framework of moderately nonlinear problems our results are still applicable to the results of iterative retrievals.

The vertical resolution of a profile of an atmospheric state variable, e.g., temperature or the volume mixing ratio of a trace gas, with n gridpoints, is usually characterized by the averaging kernel matrix $\mathbf{A}$ of size $n \times n$. Its elements are the partial derivatives $\frac{\partial \hat{x}_i}{\partial x_j}$ of the estimated state variables $\hat{x}_i$ with respect to the true state variable x_j. While the indices i and j typically run over altitude levels of one vertical profile, the concept as such has a much wider range of applicability, e.g., horizontal averaging kernels (von Clarmann et al., 2009a), or characterization of cross-dependence of multiple species. In this study, we restrict ourselves to averaging kernels of vertical profiles of single species. For a constrained retrieval of the type

$$\hat{\boldsymbol{x}} = \boldsymbol{x}_{\mathrm{a}} + \left(\mathbf{K}^T \mathbf{S}_{\mathrm{y}}^{-1} \mathbf{K} + \mathbf{R}\right)^{-1} \mathbf{K}^T \mathbf{S}_{\mathrm{y}}^{-1} \left(\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x}_{\mathrm{a}})\right), \tag{1}$$

or any equivalent formulation of it, where the $\hat{\boldsymbol{x}}$ vector represents the estimated profile, $\boldsymbol{x}_{\mathrm{a}}$ is an a priori profile, $\mathbf{K}$ is the Jacobian matrix $\frac{\partial y_i}{\partial x_j}$, ^{T} indicates a transposed matrix, $\mathbf{S}_{\mathrm{y}}$ is the measurement error covariance matrix, $\mathbf{R}$ is a regularization matrix, $\boldsymbol{F}$ is the radiative transfer function, and $\boldsymbol{y}$ is the vector of measurements (von Clarmann et al. 2003a, building largely upon Rodgers 2000), the averaging kernel matrix is

$$\mathbf{A} = \left(\mathbf{K}^T \mathbf{S}_{\mathrm{y}}^{-1} \mathbf{K} + \mathbf{R}\right)^{-1} \mathbf{K}^T \mathbf{S}_{\mathrm{y}}^{-1} \mathbf{K}. \tag{2}$$

~~With this,~~ The state dependence of the averaging kernel is largely due to the state dependence of the Jacobian $\mathbf{K}$. With the averaging kernel matrix introduced above, and using the linearization

$$\boldsymbol{y} - \mathbf{F}(\boldsymbol{x}_{\mathrm{a}}) \approx \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_{\mathrm{a}}), \tag{3}$$

Eq. 1 can be rewritten as

$$\hat{\boldsymbol{x}} = (\mathbf{I} - \mathbf{A})\boldsymbol{x}_{\mathrm{a}} + \mathbf{A}\boldsymbol{x} \tag{4}$$

The most common application of the averaging kernel matrix is the degradation of high-resolved vertical profiles to make them comparable to poorer resolved profiles (Connor et al., 1994). by application of the averaging kernel matrix of the poorer resolved profile to the high-resolved profile (Connor et al., 1994).

$$\boldsymbol{x}_{\mathrm{degraded}degraded} = (\mathbf{I} - \mathbf{A})\boldsymbol{x}_{\mathrm{a}} + \mathbf{A}\boldsymbol{x}_{\mathrm{original}\,\mathrm{high-resolved}} \tag{5}$$

Here $\mathbf{A}$ and $\boldsymbol{x}_{\mathrm{a}}$ refer to the poorer resolved profile. It goes without saying that the high-resolved profile has to be resampled on the grid on which the application of the averaging kernel is performed, and, if applicable, transformed to the same units (volume mixing ratio, number density, etc.).

For applications where the a priori profiles are all zero, as being the case for most trace gas profiles retrieved from MIPAS (von Clarmann et al., 2009b), which often is appropriate if a smoothing regularization (Steck and von Clarmann 2001, building on Tikhonov 1963) is used instead of an inverse a priori covariance matrix as suggested by Rodgers (1976, 2000), this reduces to

$$\boldsymbol{x}_{\mathrm{degraded}} = \mathbf{A}\boldsymbol{x}_{\mathrm{original}\,\mathrm{high-resolved}}. \tag{6}$$

The same is true if for all retrievals the same altitude-constant prior is used in combination with an averaging kernel with unity row sums as associated with purely smoothing constraints.

Calculation of Using

$$\mathrm{cov}(\mathbf{a}, \boldsymbol{x}) = \langle \mathbf{A}\boldsymbol{x} \rangle - \langle \mathbf{A} \rangle \langle \boldsymbol{x} \rangle, \tag{7}$$

calculation of, e.g., zonal averages over L profiles renders[2]

$$\begin{aligned} \leq\langle \hat{\boldsymbol{x}} \geq\rangle &= \leq\langle (\mathbf{I} - \mathbf{A})\boldsymbol{x}_{\mathrm{a}} + \mathbf{A}\boldsymbol{x} \geq\rangle \\ &= \leq\langle \boldsymbol{x}_{\mathrm{a}} \geq\rangle - \leq\langle \mathbf{A} \geq\leq\rangle\langle \boldsymbol{x}_a \geq\rangle - \\ &\quad \mathrm{cov}(\mathbf{A}, \boldsymbol{x}_{\mathrm{a}}) + \leq\langle \mathbf{A} \geq\leq\rangle\langle \boldsymbol{x} \geq\rangle + \mathrm{cov}(\mathbf{A}, \boldsymbol{x}), \end{aligned} \tag{8}$$

where

$$\mathrm{cov}(\mathbf{A}, \boldsymbol{x}) = \frac{1}{L}\sum_{l=1}^{L}(\mathbf{A}_l - \leq\langle \mathbf{A} \geq\rangle)(\boldsymbol{x}_l - \leq\langle \boldsymbol{x} \geq\rangle), \tag{9}$$

and mutatis mutandis in an analogue way for $\mathrm{cov}(\mathbf{A}, \boldsymbol{x}_{\mathrm{a}})$.

[2] Here a caveat is in order. The average of profiles which are 'optimal' in the sense of maximum a posteriori information and where the a priori information is the same for all averaged profiles is not the optimal average. This is, because the weight of the a priori information will be too large in the average. A more thorough discussion of this issue, however, is beyond the scope of this paper.

For a retrieval with $\boldsymbol{x}_\mathrm{a} = \mathbf{0}$ (or $\boldsymbol{x}_\mathrm{a}$ constant with altitude and a purely smoothing constraint) this simplifies to

$$\begin{aligned} \underline{<}\langle\hat{\boldsymbol{x}}\underline{>}\rangle &= \underline{<}\langle\mathbf{A}\boldsymbol{x}\underline{>}\rangle \\ &= \underline{<}\langle\mathbf{A}\underline{>}\underline{<}\rangle\langle\boldsymbol{x}\underline{>}\rangle + \mathrm{cov}(\mathbf{A},\boldsymbol{x}). \end{aligned} \tag{10}$$

$\mathrm{cov}(\mathbf{A},\boldsymbol{x})$ can be approximated by $\mathrm{cov}(\mathbf{A},\hat{\boldsymbol{x}})$ which can easily be evaluated statistically from the available results and distributed to the data user along with the mean averaging kernel ~~$<\mathbf{A}>$~~ $\langle\mathbf{A}\rangle$ and the mean profile ~~$<\boldsymbol{x}>$~~ $\langle\boldsymbol{x}\rangle$ and used to correct profiles of averaged comparison data. All this is valid only ~~cum grano salis~~with some qualification. Related problems will be discussed in Section 6.

For a retrieval with constant climatological $\boldsymbol{x}_\mathrm{a}$ for the entire sample of profiles we get

$$\begin{aligned} \underline{<}\langle\hat{\boldsymbol{x}}\underline{>}\rangle &= \underline{<}\langle(\mathbf{I}-\mathbf{A})\boldsymbol{x}_\mathrm{a} + \mathbf{A}\boldsymbol{x}\underline{>}\rangle \\ &= \boldsymbol{x}_\mathrm{a} - \underline{<}\langle\mathbf{A}\underline{>}\rangle\boldsymbol{x}_a + \\ &\quad + \underline{<}\langle\mathbf{A}\underline{>}\underline{<}\rangle\langle\boldsymbol{x}\underline{>}\rangle + \mathrm{cov}(\mathbf{A},\boldsymbol{x}). \end{aligned} \tag{11}$$

For a retrieval where an individual prior $\boldsymbol{x}_a$ is used for each profile retrieval, i.e., a prior which represents the best available information on the current state not in a climatological sense, but, e.g., from independent measurements specific to each measurement of the ensemble, it may also be adequate to assume

$$\mathrm{cov}(\mathbf{A},\boldsymbol{x}) \approx \mathrm{cov}(\mathbf{A},\boldsymbol{x}_\mathrm{a}) \tag{12}$$

~~and~~ , i.e., that the prior information is a good representation of the true atmospheric state and variability. In this case the correction by the covariance terms becomes approximately obsolete, because

$$\begin{aligned} \underline{<}\langle(\mathbf{I}-\mathbf{A})\boldsymbol{x}_\mathrm{a} + \mathbf{A}\boldsymbol{x}\underline{>}\rangle &= \\ \underline{<}\langle\boldsymbol{x}_\mathrm{a}\underline{>}\rangle - \underline{<}\langle\mathbf{A}\underline{>}\underline{<}\rangle\langle\boldsymbol{x}_a\underline{>}\rangle - \mathrm{cov}(\mathbf{A},\boldsymbol{x}_a) & \\ + \underline{<}\langle\mathbf{A}\underline{>}\underline{<}\rangle\langle\boldsymbol{x}\underline{>}\rangle + \mathrm{cov}(\mathbf{A},\boldsymbol{x}) &\approx \\ \underline{<}\langle\boldsymbol{x}_\mathrm{a}\underline{>}\rangle - \underline{<}\langle\mathbf{A}\underline{>}\underline{<}\rangle\langle\boldsymbol{x}_a\underline{>}\rangle - \mathrm{cov}(\mathbf{A},\boldsymbol{x}_a) + & \\ \underline{<}\langle\mathbf{A}\underline{>}\underline{<}\rangle\langle\boldsymbol{x}\underline{>}\rangle + \mathrm{cov}(\mathbf{A},\boldsymbol{x}_\mathrm{a}) &= \\ \underline{<}\langle\boldsymbol{x}_\mathrm{a}\underline{>}\rangle - \underline{<}\langle\mathbf{A}\underline{>}\underline{<}\rangle\langle\boldsymbol{x}_a\underline{>}\rangle + \underline{<}\langle\mathbf{A}\underline{>}\underline{<}\rangle\langle\boldsymbol{x}\underline{>}\rangle & \end{aligned} \tag{13}$$

For retrievals performed in the log-space, all this becomes slightly more complicated (e.g., Stiller et al., 2012). Equation (5) then reads

$$\boldsymbol{x}_\mathrm{degraded} = \exp\left((\mathbf{I}-\mathbf{A})\ln\boldsymbol{x}_\mathrm{a} + \mathbf{A}\ln\boldsymbol{x}_{\text{original}\,\text{high-resolved}}\right), \tag{14}$$

where $\mathbf{A}$ is $\frac{\ln\hat{\boldsymbol{x}}_i}{\ln\boldsymbol{x}_j}$. For log retrievals there is no obvious way to correct for the averaging artefacts as long as the averaging is performed linearly in the vmr-space. Since averaging of logarithmic rerievals in the logarithmic domain has its own problems (Funke and von Clarmann, 2012), we do not pursue this option any further.

The issues discussed in this Section have to be considered if mean averaging kernels are to be applied to mean profiles in the spirit of Eq. 5, in order to make mean profiles of different sources comparable.

3 Covariances

The covariances between the averaging kernel matrices and the state vectors are calculated as

$$\mathrm{cov}(\mathbf{A},\boldsymbol{x}) = \tag{15}$$

$$\underline{=}\frac{1}{L}\left(\sum_{l=1}^{L}(\mathbf{A}_l - \underline{<}\langle\mathbf{A}\underline{>}\rangle)(\boldsymbol{x}_l - \underline{<}\langle\boldsymbol{x}\underline{>}\rangle)\right) =$$

$$\underline{=}\frac{1}{L}\left(\sum_{l=1}^{L}\mathbf{A}_l\boldsymbol{x}_l - \frac{1}{L}\sum_{l=1}^{L}\mathbf{A}_l\sum_{l=1}^{L}\boldsymbol{x}_l\right),$$

where L denotes the sample size; we divide by L instead of $L-1$ because the latter would entail an inconsistency with Eq. (8) and Eqs. (10–13). The formulation in the lowermost line of Eq. (15) is computationally more efficient. For our case study, averaging kernel matrices and state vectors retrieved from limb emission spectra measured by the MIPAS are used. The general processing scheme is described by von Clarmann et al. (2003b, 2009b). We study covariances for MIPAS O_3 and HCN profiles (Laeng et al. 2018 and Glatthor et al. 2015, respectively).

To illustrate the relevance ~~$\tilde{r}$~~ of the correction terms, we also present the normalized covariance term $\tilde{r}$ for each profile element:

$$\tilde{r}_n = \mathrm{cov}(\mathbf{A},\boldsymbol{x})_\mathrm{n}/(\underline{<}\langle\mathbf{A}\underline{><}\rangle\langle\boldsymbol{x}\underline{>}\rangle)_\mathrm{n} \tag{16}$$

where index n runs over the profile elements. The ˜ symbol shall avoid confusion with the product moment correlation coefficient established by Pearson (1895) for which r is often used as a symbol and which is widely used for normalization of covariances but which causes some headache when applied to correlations of matrices with vectors. For simplicity, we still call the normalized covariance 'correlation', however without claiming equivalence with its scalar counterpart.

4 Results

Case studies have been performed using ozone and HCN vertical profiles retrieved from MIPAS measurements of February 10, 2009. The test data set ~~consistes~~ consists of 1385 geolocations. This day was characterized by a significantly disturbed Arctic vortex. Figure 1 shows the covariances between the profiles and the averaging kernel matrices of ozone globally (black solid line) and for various latitude bands of different size (dashed and dotted lines). In general the values are largest at the extreme ends of the profiles, where the effect of the constraint on the retrieved profile is typically largest.

These results suggest that for MIPAS ozone in the middle and upper stratosphere the effect studied here can be safely ignored. Problems are limited to the upper troposphere and lower stratosphere and the mesosphere. The relevance of this effect can better be judged on the basis of the correlation profiles (Fig. 2). From 20 to about 60 km the effect is negligibly small for all latitude bands investigated in this case study. Only at the uppermost and lowermost altitudes the effect becomes relevant. The large effects at lower altitudes are simply caused by normalization of the original covariances by low ozone mixing ratios.

To study HCN is particularly interesting in the tropical upper troposphere and lower stratosphere. This is because HCN has tropospheric sources, and its pathway into the stratosphere is a particular research issue. The covariance effects can exceed

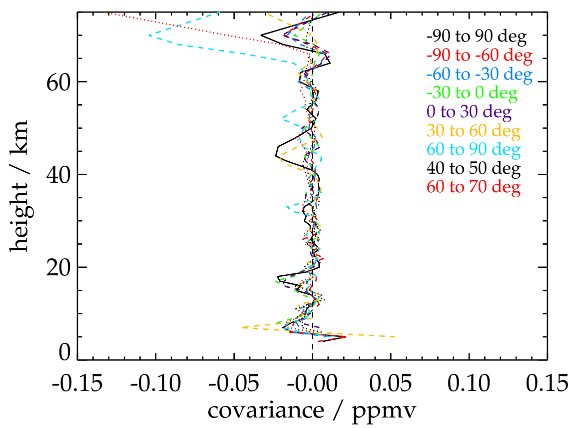

Figure 1. : Covariance of the averaging kernel and ozone mixing ratio for various latitude bands. The black solid line refers to global data. The dashed lines refer to 30-degrees latitude bands, and the dotted lines to 10-degrees latitude bands.

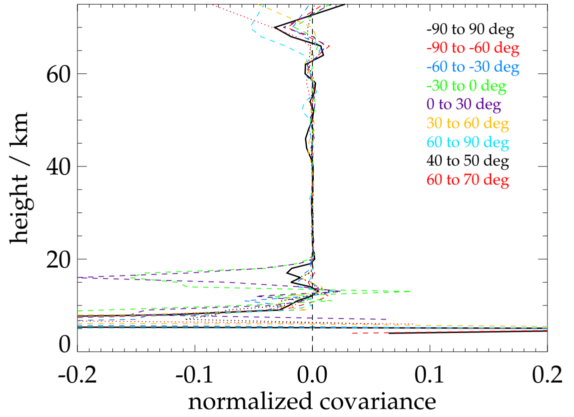

Figure 2. : Correlation of the averaging kernel and ozone mixing ratio for various latitude bands.

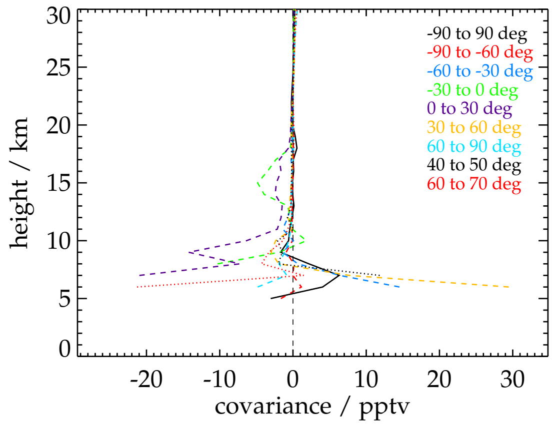

Figure 3. : As Fig 1, but for HCN.

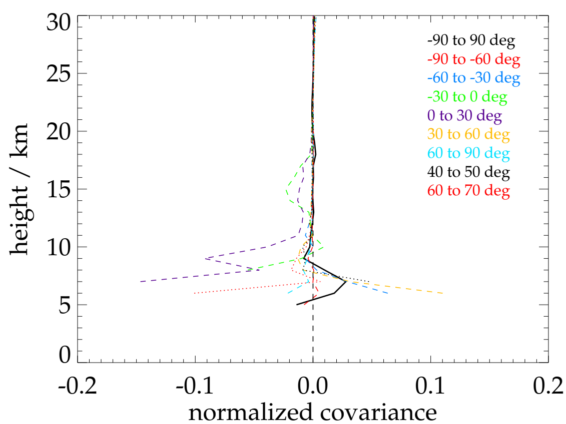

Figure 4. : As Fig 2, but for HCN.

10% (violet and yellow dashed lines) and thus need to be considered when mean profiles are used for quatitative analysis and mean averaging kernels are applied.

These case studies are not meant to be representative for other gases or other instruments. Instead, they are shown to give a flavour about the order of magnitude this kind of effect can reach. Unless $\mathrm{cov}(\mathbf{A}, \boldsymbol{x})$ can be shown to be small, we recommed recommend to use this covariance term for an additive correction when mean averaging kernels are applied to averaged comparison data.

5 An Important Side Remark

The issue of the limited applicability of averaging kernels to independent comparison data deserves awareness. When averaging kernels of a measurement are applied to better resolved comparison data, it is almost always tacitly assumed that the atmospheric state represented by the measurement is the same as that of the comparison data and that thus the averaging kernel of the measurement can be safely applied to the comparison data. However, since averaging kernels are in general state-dependent, a caveat is in order.

Application of the formalism of Connor et al. (1994) (our Eq. 5) has its own specific problems which fully apply to our proposed scheme. The application of the averaging kernel matrix of a poorly resolved profile $\boldsymbol{x}_{\mathrm{coarse}}$ to a better resolved profile $\boldsymbol{x}_{\mathrm{fine}}$ is only adequate if both data sets describe approximately the same atmospheric state, i.e., if the one profile is in the linear domain of the other. That is to say, if the same Jacobians apply to both profiles. Otherwise it would be necessary to construct an averaging kernel using the Jacobian $\mathbf{K}$ evaluated for the atmospheric state represented by the profile $\boldsymbol{x}_{\mathrm{fine}}$ but with the measurement covariance matrix $\mathbf{S}_{\mathrm{y}}$ and the regularization matrix $\mathbf{R}$ corresponding the the retrieval producing $\boldsymbol{x}_{\mathrm{coarse}}$. Within linear theory, the state-dependence of the Jacobian $\mathbf{K}$ and with this the state dependence of the averaging kernel matrix $\mathbf{A}$ are often ignored. To do so is justifiable as long as the profiles to be intercompared are sufficiently similar. In this case the comparison will show reasonable agreement.

If, in turn, the profiles are very different, two components contribute to the disagreement seen after application of the Connor-method. First the genuine difference of the profiles, and second any artefact caused by the inadequate averaging kernels. Thus, in the logic of a testing scheme, good apparent agreement hints at good genuine agreement a fortiorieven more, because it is extremely unlikely that genuine differences which would survive the application of the Connor-method with the correct averaging kernel are 'convolved away'[3] with the averaging kernel evaluated for the wrong atmosphere.

6 Discussion and Conclusion

We have identified the problem that it is not generally allowable to apply mean averaging kernels to mean atmospheric profiles in situations where the averaging kernels and the profiles covary. The relevance of this effect, however, depends on the instrument, the species, the latitude band and the altitude under investigation. To solve this problem, we have proposed a statistical

[3]We put this term in quotes to highlight that this method is, mathematically speaking, not a convolution. It is not even a numerical approximation to a convolution.

correction scheme which involves the covariance between the averaging kernel and the profile. With this correction in place, the scheme suggested by Connor et al. (1994) to make better resolved vertical profiles of atmospheric state variables comparable to coarser resolved ones can be applied also to averaged profiles.

For data producers who distribute, besides their original retrievals, also zonal mean data or similar data products, we recommend the following: Along with the generation of zonal mean data and averaging kernels, the correlation profiles should be calculated. Compared to averaging kernels and covariance matrices they need negligible storage and cause negligible data traffic. In cases when zonal mean data have already been generated but when mean covariance matrices and covariance profiles are not available, the huge I/O load associated with reading all individual averaging kernels may be prohibitive. In these cases one might consider to estimate the mean averaging kernel and the covariance profile on the basis of a limited random sample out of the measurements which went into the zonal mean.

Data availability. MIPAS data used for this case study are available via https://www.imk-asf.kit.edu/english/308.php

Author contributions. TvC identified the problem, suggested the research method, proposed the solution and wrote the paper. NG critically reviewed the methodology, implemented the suggested solution as a computer program, performed the calculations and visualized the results. Both authors discussed and evaluated the results and elaborated the recommendation.

Competing interests. TvC is associate editor of AMT but has not been involved in the evaluation of this paper.

Acknowledgements. NG acknowledges support by BMBF project SEREMISA (50EE1547). SPARC provided funding for the meetings of the activity "Towards Unified Error Reporting (TUNER)", the framework in which this research took place. The International Space Science Institute (ISSI) provided support and facilities for two TUNER International Team meetings.

References

Connor, B. J., Siskind, D. E., Tsou, J. J., Parrish, A., and Remsberg, E. E.: Ground-based microwave observations of ozone in the upper stratosphere and mesosphere, J. Geophys. Res., 99, 16,757–16,770, https://doi.org/10.1029/94JD01153, 1994.

Fischer, H., Birk, M., Blom, C., Carli, B., Carlotti, M., von Clarmann, T., Delbouille, L., Dudhia, A., Ehhalt, D., Endemann, M., Flaud, J. M., Gessner, R., Kleinert, A., Koopmann, R., Langen, J., López-Puertas, M., Mosner, P., Nett, H., Oelhaf, H., Perron, G., Remedios, J., Ridolfi, M., Stiller, G., and Zander, R.: MIPAS: an instrument for atmospheric and climate research, Atmos. Chem. Phys., 8, 2151–2188, https://doi.org/10.5194/acp-8-2151-2008, 2008.

Funke, B. and von Clarmann, T.: How to average logarithmic retrievals?, Atmos. Meas. Tech., 5, 831–841, https://doi.org/10.5194/amt-5-831-2012, 2012.

Glatthor, N., Höpfner, M., Stiller, G. P., von Clarmann, T., Funke, B., Lossow, S., Eckert, E., Grabowski, U., Kellmann, S., Linden, A., Walker, K. A., and Wiegele, A.: Seasonal and interannual variations in HCN amounts in the upper troposphere and lower stratosphere observed by MIPAS, Atmos. Chem. Phys., 15, 563–582, https://doi.org/10.5194/acp-15-563-2015, 2015.

Hegglin, M. and Tegtmeier, S.: The SPARC Data Initiative, SPARC Newsletter, 36, 22–23, 2011.

Hegglin, M. I. and Tegtmeier, S., eds.: The SPARC Data Initiative: Assessment of stratospheric trace gas and aerosol climatologies from satellite limb sounders, SPARC Report No. 8, WCRP-5/2017, SPARC, https://doi.org/10.3929/ethz-a-010863911, 2017.

Hegglin, M. I., Tegtmeier, S., Anderson, J., Froidevaux, L., Fuller, R., Funke, B., Jones, A., Lingenfelser, G., Lumpe, J., Pendlebury, D., Remsberg, E., Rozanov, A., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., Wang, R., and Weigel, K.: SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders, J. Geophys. Res. Atmos., 118, 11,824–11,846, https://doi.org/10.1002/jgrd.50752, 2013.

Laeng, A., Eckert, E., von Clarmann, T., Kiefer, M., Hubert, D., Stiller, G., Glatthor, N., López-Puertas, M., Funke, B., Grabowski, U., Plieninger, J., Kellmann, S., Linden, A., Lossow, S., Babenhauserheide, A., Froidevaux, L., and Walker, K.: On the improved stability of the version 7 MIPAS ozone record, Atmos. Meas. Tech., 11, 4693–4705, https://doi.org/10.5194/amt-11-4693-2018, https://doi.org/10.5194/amt-11-4693-2018, 2018.

Pearson, K.: VII. Notes on regression and inheritance in the case of two parents, Proc. R. Soc. Lond., 58, 240–242, https://doi.org/10.1098/rspl.1895.0041, 1895.

Rodgers, C. D.: Retrieval of Atmospheric Temperature and Composition From Remote Measurements of Thermal Radiation, Rev. Geophys. Space Phys., 14, 609–624, 1976.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, vol. 2 of *Series on Atmospheric, Oceanic and Planetary Physics, F. W. Taylor, ed.*, World Scientific, Singapore, New Jersey, London, Hong Kong, 2000.

Steck, T. and von Clarmann, T.: Constrained profile retrieval applied to the observation mode of the Michelson Interferometer for Passive Atmospheric Sounding, Appl. Opt., 40, 3559–3571, 2001.

Stiller, G. P., von Clarmann, T., Haenel, F., Funke, B., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., Lossow, S., and López-Puertas, M.: Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period, Atmos. Chem. Phys., 12, 3311–3331, https://doi.org/10.5194/acp-12-3311-2012, 2012.

Tegtmeier, S., Hegglin, M. I., Anderson, J., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Krüger, Y. K. K., Kyrölä, E., Lingenfelser, G., Lumpe, J., Nardi, B., Neu, J., Pendlebury, D., Remsberg, E., Rozanov, A., Smith, L., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R. H. J.: SPARC Data Initiative: A comparison of ozone climatologies from international satellite limb sounders, J. Geophys. Res. Atmos., 118, 12,229–12,247, https://doi.org/10.1002/2013JD019877, 2013.

Tegtmeier, S., Hegglin, M. I., Anderson, J., Funke, B., Gille, J., Jones, A., Smith, L., von Clarmann, T., and Walker, K. A.: The SPARC Data Initiative: comparisons of CFC-11, CFC-12, HF and SF_6 climatologies from international satellite limb sounders, Earth Syst. Sci. Data, 8, 61–78, https://doi.org/10.5194/essd-8-61-2016, 2016.

Tikhonov, A.: On the solution of incorrectly stated problems and method of regularization, Dokl. Akad. Nauk. SSSR, 151, 501–504, 1963.

von Clarmann, T., Ceccherini, S., Doicu, A., Dudhia, A., Funke, B., Grabowski, U., Hilgers, S., Jay, V., Linden, A., López-Puertas, M., Martín-Torres, F.-J., Payne, V., Reburn, J., Ridolfi, M., Schreier, F., Schwarz, G., Siddans, R., and Steck, T.: A blind test retrieval experiment for infrared limb emission spectrometry, J. Geophys. Res., 108, 4746, https://doi.org/10.1029/2003JD003835, 2003a.

von Clarmann, T., Glatthor, N., Grabowski, U., Höpfner, M., Kellmann, S., Kiefer, M., Linden, A., Mengistu Tsidu, G., Milz, M., Steck, T., Stiller, G. P., Wang, D. Y., Fischer, H., Funke, B., Gil-López, S., and López-Puertas, M.: Retrieval of temperature and tangent altitude pointing from limb emission spectra recorded from space by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), J. Geophys. Res., 108, 4736, https://doi.org/10.1029/2003JD003602, 2003b.

von Clarmann, T., De Clercq, C., Ridolfi, M., Höpfner, M., and Lambert, J.-C.: The horizontal resolution of MIPAS, Atmos. Meas. Techn., 2, 47–54, https://doi.org/10.5194/amt-2-47-2009, 2009a.

von Clarmann, T., Höpfner, M., Kellmann, S., Linden, A., Chauhan, S., Funke, B., Grabowski, U., Glatthor, N., Kiefer, M., Schieferdecker, T., Stiller, G. P., and Versick, S.: Retrieval of temperature, H_2O, O_3, HNO_3, CH_4, N_2O, $ClONO_2$ and ClO from MIPAS reduced resolution nominal mode limb emission measurements, Atmos. Meas. Techn., 2, 159–175, https://doi.org/10.5194/amt-2-159-2009, 2009b.

von Clarmann, T., Funke, B., Glatthor, N., Kellmann, S., Kiefer, M., Kirner, O., Sinnhuber, B.-M., and Stiller, G. P.: The MIPAS HOCl climatology, Atmos. Chem. Phys., 12, 1965–1977, https://doi.org/10.5194/acp-12-1965-2012, 2012.