General, specific and minor comments of the reviewer in Italic characters.
Replies of the authors in Normal characters.

Reviewer #3

SUMMARY

The paper describes methods of combining data retrieved from two limb-viewing instruments, MIPAS-STR and MARSCHALS, carried on the same aircraft. The first method (MSS) separates the measurements and constraints from each retrieval, combines just the measurement components and adds a new constraint, in this case Tikhonov regularization to produce stable output profiles. The second method is sequential: the MIPAS-STR profiles are first retrieved using regularisation, and the resulting profiles are then combined using optimal estimation with the MARSCHALS data. In principle, the results ought to be the same but significant differences are noted due to the implementation.

A second aspect of the paper is the demonstration of synergy between instruments using the infrared and millimetre spectral ranges, showing that under conditions of thin cloud, the MARSCHALS instrument can make a significant contribution. However, this is hardly surprising.

GENERAL COMMENTS

Any retrieval can be considered the sum of measurements and constraints (a priori estimates or regularisation) weighted by their inverse covariances. In principle, the results should be the same whether two sets of measurements are combined simultaneously with the constraint (the MSS method), or if one set of measurements is combined with the constraint, and the second set of measurements then added using optimal estimation (the L1+L2 method). The fact that the results are different are therefore simply due to differences in the details of the implementation, primarily the grid representation. It would have been better to have a consistent retrieval throughout.

The vertical grid on which MIPAS and MARSCHALS profiles are represented is the same in the two data fusion methods. A cause of the differences between the profiles obtained with the two methods could be the different implementation of the MIPAS systematic errors. In the (L1+L2) method, the MIPAS systematic errors are provided in the L2 outputs as errors on the retrieved profiles and we have added them to the random errors before to perform the optimal estimation of the MARSCHAL measurements. In the MSS method, coherently with what done for MARSCHALS systematic errors, the MIPAS systematic errors are taken into account including them in the covariance matrix of the radiances.
Another cause of the differences between the profiles obtained with the two methods could be due to the different linearization points that the two methods adopt. This difference is intrinsic to the two methods.
The values of the quantifiers are necessarily different for the two methods, because in the L1+L2 method the MIPAS measurements are considered as a constraint, while in the MSS methods they are considered as measurements.

A second criticism is the rather loose definition of what constitutes a synergistic retrieval. If, as the title suggests, the aim is to demonstrate synergy between the two types of instrument, beyond the obvious fact that microwave instruments are less sensitive to clouds that infrared instruments, I was left unconvinced. Probably the most important issue when combining data from two different types of observation is whether the absolute accuracy of the product is improved, and that was not discussed. The introduction of unecessary and unhelpful new diagnostics (RID, MQQ, SF – detailed below) did not help matters. Overall, this paper demonstrates a successful practical application of two different methods of combining profiles with different retrieval characteristics but in terms of showing the 'Synergy', as in the title of the paper, adds very little to what is immediately obvious: infrared is generally better than millimetre, but not when it comes to cloud.

From authors' point of view, the main goal of the paper is neither to demonstrate the well-known synergy between infrared and millimeter nor to compare the retrieval performances of the two spectral bands. The main goal of the paper is to quantify the synergistic properties of the two spectral bands. For this reason, we introduced the diagnostic quantifiers reported in the paper. The total retrieval error defined in the paper contains both the random and the systematic component, thus it is the absolute accuracy of the product. The total retrieval error has been discussed in the paper as quantifier of the synergy and used to evaluate the improvement in the quality of the product.

SPECIFIC COMMENTS

1) P11676, L15 - P11676, L11677: This list of pros and cons of IR/MW seems to be applicable to nadir sounding rather than limb sounding (eg negative lapse rates, surface emissivity) and comments about the 4.3um (2300cm-1) CO2 band seem irrelevant if MIPAS-STR only measures up to 2100cm-1. I suggest this list is reviewed and amended to make only the points relevant for this experiment. But it should perhaps also be noted that there are molecules, most notably O2, which have significant spectral features in the MW but not in the IR.
We accepted the suggestion of reviewer #3 and we removed the statements not directly applicable to the limb sounding measurements of MIPAS-STR and MARSCHALS in the revised version of the paper.

2) Section 3.1 The MSS solution is described at considerable length, and rather confusingly, in section 3.1, as if it were some complicated transform. However, looking at Eq(1), it appears to be simply the standard unconstrained least squares fit solution and, to most readers, that would be a much shorter and simpler description. Substituting the standard LSF covariance (KˆT Syˆ-1 K)-1 (the inverse of the Fisher information matrix defined in Eq8) for the S_MSS in Eqs (1)-(7) seems to give the standard equations for regularised inverses.

The MSS is the component of the profile in the measurement space and is represented using an orthonormal basis of the measurement space, so that the components of the profile in this basis are uncorrelated with each other. The MSS coincides with the standard unconstrained least squares fit solution only when the measurement space coincides with the space in which the profile is represented. In general the representation of the profile on a fine grid makes the measurement space a subspace of the space in which the profile is represented with the consequence that infinity unconstrained least squares fit solutions exist (because the problem is ill-posed), while the MSS is always well defined.
So probably, the description of the reviewer of the MSS is shorter and simpler, but it is not correct in all cases.

3) P11690 - discussion of Total Retrieval Error I found this very difficult to follow.
The Total Retrieval Error definition has been re-arranged in a clearer manner in the revised version of the paper.

4) P11692 - Relative information distribution. I don't understand why this RID diagnostic is introduced here at all. With equal profile spacing it is basically just the inverse of the square of the retrieval error expressed as a fraction of the profile value, thus there is a 1:1 relationship between RID and %retrieval error. So the RID panels in Figs 5-7 show exactly the same information as the error panels above, but with different and obscure contour values. I have no idea what a RID value of 4000 means, but I do understand a retrieval error of 0.5% or 0.5K.

The RID is not just the inverse of the square of the retrieval error expressed as a fraction of the profile value. Indeed the square of the retrieval errors are given by the diagonal elements of the covariance matrix. As we can see from Eq. (6), these errors depend on the constraint we used. On the other hand, the RID defined in Eq. (10) is defined using the diagonal elements of the Fisher information matrix, which does not depend on the constraint.
This propriety makes the RID a quantifier preferable with respect to the retrieval errors, because it quantifies in an absolute way the information that the observations have on the retrieved profile independently from the constraint applied to the retrieval.

5) Synergy Factor The original use by Aires was to investigate combinations of nadir sounding instruments where the error could be measured with respect to a defined 'truth', and some combinations could clearly be worse than single instruments. It seems a misuse to apply the same statistic to these results where the errors are now just the predicted retrieval errors, which can, by definition, never get worse. Thus everything appears to prove 'synergistic' to some degree. If you are going to use this definition of synergy, then the results should be compared to some independent truth.

In Aires et al. (2011), where the synergy factor is introduced, the errors are estimated by Eq (3) (of Aires et al. (2011)) and they correspond to the predicted retrieval errors. We do not think that calculating the synergy factor with the predicted retrieval errors is a misuse. This is confirmed by our Fig. 9, where there are a lot of points where the synergy factor is one and, therefore, in that cases we can conclude that there is not synergy between the two measurements.

*6) The Shannon information gain is, I think, too easily dismissed as an alternative. Firstly, it can be defined wrt to *any* 'a priori' covariance, not necessarily the one actually used (if any) within the retrieval. As such it can provide a sensible weighting of each profile retrieval error in the context of an expected uncertainty in the quantity. For example 10% accuracy in tropospheric ozone is much more useful than 5% accuracy in stratospheric ozone but the RID would weight them the other way around. Secondly, the Shannon information can also be defined level-by-level (replacing the covariance matrices with variances) so can provide altitude discrimination. It also has the useful property of being additive, so in my opinion would provide a better measure of the relative contributions of the two instruments than the Synergy Factor or RID.*

The Shannon information gain is defined in Eq. (11), so we can put for S_a any a priori covariance matrix, but the covariance matrix S depends on the a priori covariance matrix (or any constraint) used in the retrieval (see Eq. (6)). Therefore, to use S_a different from that used in the retrieval provides a quantifier with a not clear significance.
The dependence of the Shannon information gain on the covariance matrix determines the dependence of the Shannon information gain on the constraint used in the retrieval, and for this reason, we prefer to use a quantifier as the RID, which does not depend on the constraint, as already explained at the reply to the specific comment 4).
The definition of the Shannon information level-by-level given by the reviewer neglects the correlations between errors at different altitudes. Since not only the errors but also the correlations contribute to the information content of the measurement, the definition given by the reviewer provides a quantifier with a partial significance.
Finally, as stated in Ceccherini at al (2012) and demonstrated in Ceccherini et al. (2010), the Shannon information gain does not satisfy the additivity property for data fusion, while the quantifiers that are derived from the Fisher information matrix as the RID satisfy this property.

MINOR COMMENTS

a) 'Inverse Processing' there are a number of occurrences of this phrase eg P11674, L24, Section header 2.1.2. This confused me at first since I thought it might refer to simulating L1 radiances from L2 products (ie the 'inverse' of the usual 'processing'). I would suggest replacing with 'Inversion', but that I s just my preference.
Done

b) P11679, Section 2.0. Given the equal weight here applied to all three flights in this section, I was expecting results from all three to be shown rather than just the 2010 campaign. Please adjust the text accordingly to make it clear at this point which of the various datasets are to be investigated in the remainder of this paper.
In section 2, we added the statement: "The results reported in this paper are based on data acquired during thr PemierEx flight on March 10, 2010.".

c) P11680, L4-5: I suggest 'In the upper troposphere, at millimetre wavelengths, problems can be posed by ...'. However I didn't understand the point that was being made in the following sentence about 'residual absorption'. Is this a reference to the MARSCHALS field-of-view extending to the mid-troposphere where water vapour concentration is even higher?
The point about residual absorption is related to the fact that strong absorption features of water vapor are present in the far-infrared/sub-millimeter range and in presence of high water vapour concentration their pressure broadening might have a significant impact also on the wings at longer (i.e., millimeter) wavelengths. In the revised version of the paper, we modified the sentence to clarify this aspect.

d) P11683, L21 (also P11695, L7) 'along flight track'. Are the measurements along the actual flight track or parallel to it? I assume that both instruments look sideways to the flight direction although that hasn't been stated so far (in fact, only evident in Figs 2 and 3, and not mentioned in the text at all). Horizontal gradients have more of an impact if the viewing direction is orthogonal to the flight direction, so it is an important distinction.
P11683, L21 Modified the text: " … spectrometer. Both instruments performed limb-sounding observations with side-viewing geometry on the starboard side of the aircraft during the flights of …".

e) P11684, section 2.1.2/2.2.2. It would be better to have an equivalents to Table 1 and 2 to describe the MIPAS-STR instrument and error characteristics.
Since the characteristics of the MIPAS-STR infrared observations are different from the MARSCHALS mm-wave observations and the errors are estimated using different approaches, a direct comparison is difficult. Therefore, we provide the most relevant aspects of MIPAS-STR and the applied error budget in Sect. 2.2.1 and 2.2.2 and refer to Woiwode et al. (2012) for more detailed information.

f) P11694 L1-9: Synergy Factor It is not obvious here, but Aires only defines synergy factor for a scalar quantity, ie a profile level rather than complete profiles. Otherwise there is the question of how to compute a synergy factor for vector quantities.
As in Aires et al. (2011) and Aires et al. (2012), the synergy factor is calculated for each atmospheric layer and represented as a vertical profile.

g) P11696 L5 'These included ....' Not clear if this is the list of all the targets retrieved from both instruments, or just the subset of common targets studied here, or just some examples of common targets.
The list includes a set of common targets retrieved from simultaneous and collocated observations by the two instruments during the flight on March 10, 2010. We modified the text to clarify this point.

h) P11697 & Figs 5-8 I am unclear on what is being shown on the DOF panels for the MARSCHALS data. Does 'B and B' refer to band B data only? And for the B+C and B+D data are data from successive scans being averaged together to retrieve a single profile?
There was a misunderstanding. In the figures, we reported DOFs separated with respect to MARSCHALS bands. In the figures, we reported 'Band B' and not 'B and B'. We will change the label from 'Band B' to 'band B' to avoid this confusion (the same for band C and D).

P11698 and subsequently Discussion of 'factors' when applied to intrinsically dimensionless quantities such as DOFs is confusing. Better to refer to absolute increases. In any case, if the changes are associated with the change in the number of cloud free profile levels, it really is a change in the number of DOFs rather than a fractional increase.
Fractional increase values replaced with absolute increase values at P 11698 L8, P 11701 L1, P 11705 L19 and P 11706 L9.

j) Figs 5-8: Since the text refers to changes for different 'Legs' of the flight, these should also be marked on the figures (perhaps just the DOF panels).
As proposed by the reviewer 'Legs' will be included in the figures.

TYPOGRAPHICAL/GRAMMATICAL POINTS

Throughout: inconsistent spelling 'vapor' and 'vapour'
P11674 L6: change 'is focusing' to 'focuses' L20: change 'O2' to 'O3'
Done

P11675 L5: change 'is providing' to 'provides' L18: remove comma after 'sounders'. I think what the authors are trying to say here is: either using different spectral regions or using different observation geometries.
Done

P11678 L1: change 'embarking' eg to 'carrying' L2: change 'was composed by... and by...' to either 'comprised' or 'was composed of ... and ...'
Done (changed 'was composed by … and by …' to 'was composed of … and …'

P11680 L12-13: change 'an uniform' to 'a uniform'
Done

P11681 L10: No need for definition of IMU acronym if it does not appear elsewhere.
Acronym removed

P11681 L16: (pedantically) MARC retrieval code is a tautology since the 'RC' also means 'retrieval code'
Replaced 'MARC retrieval code' with 'MARC'

P11682 L9: ILS acronym apparently not repeated elsewhere
Acronym removed

P11682 L10: change 'uncertainties on' to 'uncertainties in' L12: change 'MARC algorithm provides also ' to 'The MARC algorithm also provides '.
Done

P11683 L13: change 'allows to observe' to 'allows the observation of '
Done

P11683 L15: suggest changing 'tropospheric altitudes' to 'tropospheric tangent altitudes' to clarify that the point refers to the target rather than the observing instrument.
Done

P11683 L21: change 'allow to resolve' to 'allow the resolution of'
Done

P11684 L8: missing ')' after '990 cm-1'
Done

P11685 L8: change 'inverted' to 'retrieved'
Done

P11686 L10: Section 3.1 MSS should be expanded here on its first use in the main part of the paper.
Done

P11690 L26: change 'we remind here' to 'we recall here' (remind is a transitive verb so requires a direct object, eg 'we remind the reader').
Done

P11693 L1: change 'independent on' to 'independent of' L8: change 'in alternative' to 'as an alternative' L24: change 'allows to evaluate in detail' to 'allows the detailed evaluation of'
Done

P11695 L7: change 'scenery' to 'scenes'
Done

P11696 L5: change 'o3' to 'O3' L12: change 'remind' to 'recall' L14: (pedantically) 'alternate' refers to just two cases, not three.
Done

P11697 L5: change 'cfr.' to 'cf.' L9: change 'fucntion' to 'function'
Done

P11703 L20: change 'capable to evaluate' to 'capable of evaluating'
Done

P11704 L2: change 'allow separating' to 'allow the separation of'
Done

P11715 (Table 1) 'Band A' listed as header for first column. Also, there should be some mention of noise characteristics.
Receiver noise temperature and Noise equivalent Brightness Temperature values in each band added to Table 1.

P11716 (Table2) change 'hundredth' to 'hundredths'
Done

P11718 (Figure 1) Since several MARSCHALS datasets are mentioned here, it should be made explicit in the figure caption the date/latitudes of this particular flight.
Replaced caption figure 1: 'The flight altitude and MARSCHALS scan position plotted vs. the acquisition time (UTC).'
With: 'The flight altitude and MARSCHALS tangent points plotted vs. the acquisition time (UTC) for the Arctic Geophysica flight on 10 March 2010.'

P11724 (Figure 7) change 'H3O' to 'H2O'
Done

P11728 (Figure 11) change 'Sinergy' to 'Synergy' on right axes.
Done