Table of Contents

We thank both reviewers for their time spent reviewing the manuscript and are grateful for their recommendations. Specific answers to comments are provided in the separate Reviewer 1 and 2 sections below, but before that we provide a response to the two main issues raised by both reviewers. In all cases the reviewers' comments are provided in black font and our responses in green.

Firstly, the aim of this paper was not to show the improvement of using observations in the far-infrared with respect to the mid-infrared, as this has been done in other work on both simulated cases, and observations from aircraft level. We simply aim to perform a sanity check on the extension to this scheme in comparison to its ability when applied to IASI as this was the original instrument that the retrieval scheme was built for. To further emphasise this, we have rephrased comparisons in the abstract, section 4, and conclusion.

Secondly, we agree that the measurement covariance built for the FORUM Sounding Instrument should include both an estimation of the calibration uncertainty, and off-diagonals. The measurement covariance for FORUM is now the combination of two components – the apodised target noise (AP NOISE) and the target absolute radiometric accuracy (ARA) that is provided in brightness temperature. The latter is converted into radiance units for each individual spectrum and converted into a fully spectrally correlated covariance matrix. The two components are then combined, and examples are shown in Figure A1 below, alongside the IASI measurement covariance.

Figure A2 shows the noise, calibration, and final measurement covariance used in their 2D form. Most of the contributions towards the covariance are coming from close to the diagonal, or between 100-200cm-1 and 1300-1600cm-1. This has also been included in the supplementary information.

Fig A1. The diagonals and first three off-diagonals for the matrices associated with the original IASI measurement covariance, the FSI apodised (AP) noise covariance, and the FSI target ARA. For the FSI target ARA, we assume the matrix is fully correlated.

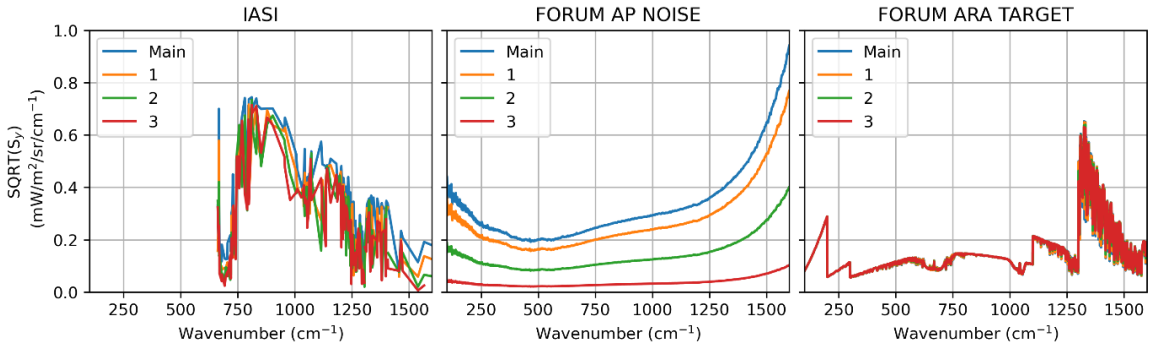
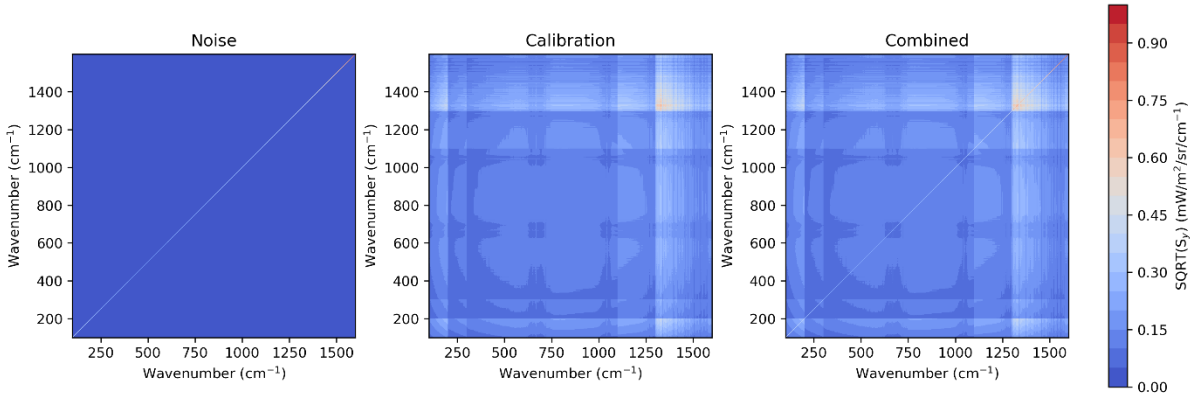
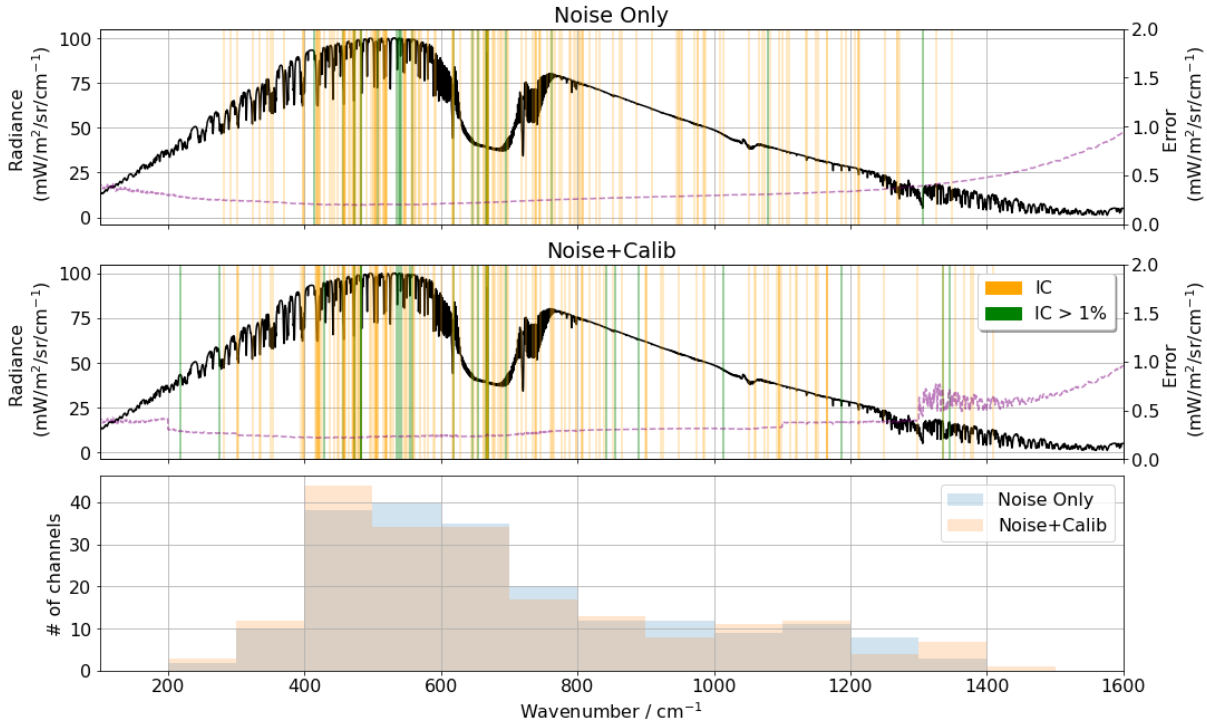


Fig A2. Components of the measurement covariance developed for the FSI. (a) Contributions from the apodised target NESR to the final measurement covariance. (b) The average contribution from the target absolute radiometric accuracy (ARA) when applied to the ECMWF 83-profiles as an example. (c) An example measurement covariance as the combination of these two components.



This slightly impacts the channel selection (see Figure A3 below). The top panel shows the 200 channels that were selected in the original manuscript submission using the diagonal noise Sy only, while the middle panel shows the 200 channels selected when the calibration uncertainty and off-diagonals are included. However, as noted in the revised paper on line 200, this only slightly changes the total information content contained in the retrieval. The bottom panel shows the difference in the distribution of selected channels.

Fig A3. The top 200 channels selected for the FSI configuration using the method outlined in Section 3.2. The top plot shows the original selection, where the measurement covariance was a diagonal matrix composed of only the apodised noise. The middle plot shows the new selection, where the measurement covariance includes off-diagonals and an estimate of the calibration uncertainty. The bottom plot summarises the distributions of both sets of channels.

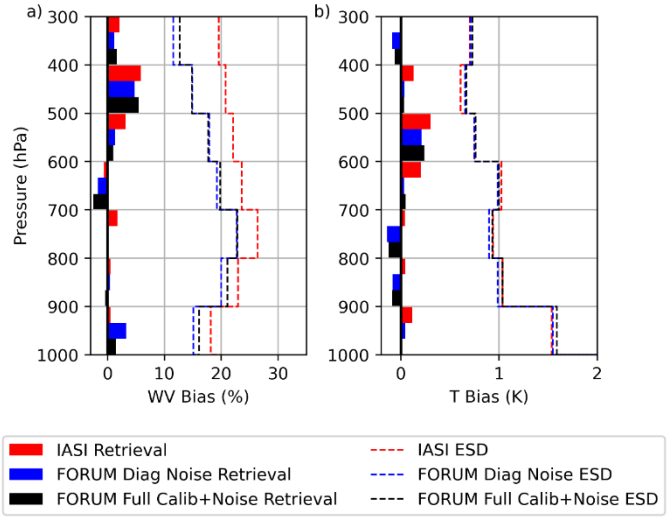


Using this new channel selection and measurement covariance for the FSI configuration, the retrievals from FSI simulations (Section 4) have been redone. Figure A4 below shows the impact on the median temperature and water vapour bias associated with the retrievals, in comparison to the original FSI configuration.

Introducing the calibration uncertainty and off-diagonals for the FORUM configuration (blue vs black) slightly increases the ESD but does not significantly affect the median retrieved profiles. The difference between the IASI and new FORUM retrievals is slightly reduced.

The red and black curves and bars are what are included in Figure 5 in the revised paper (with the CDOFS recalculated for Figure 4 in the revised paper).

Fig A4: The median (a) water vapour (WV) and (b) temperature (T) bias of the retrieved profile to the true state across the testing cases for the optimised IASI (red) and previous FORUM (blue) and updated FORUM (black) configurations. Biases are evaluated between 300 and 1000 hPa in 100 hPa bins as in Trent et al. (2023). The water vapour percentage bias is calculated from absolute values in ppmv. Dashed lines show the median estimated standard deviation (ESD) for a single retrieval for the test cases in each bin. For water vapour this is plotted with an offset of - 10% for clarity.



Reviewer 1 Comments

"This work by Panditharatne et al. reports on the adaption of the RAL Infrared Microwave Sounding retrieval processor to the evaluation of data from the far-infrared spectral range. The scheme is tested twofold: first by simulating FORUM-like observations and performing test retrievals of water vapour and temperature in comparison to IASI simulations. Second, simultaneous aircraft nadir measurements from two instruments covering the far- and mid-infrared are used as the input to the retrieval scheme. Results are discussed against vertical profiles of temperature and water vapour from dropsondes during the same flight.

This analysis clearly shows the applicability of the adapted retrieval scheme for the analysis of observations in the far-infrared spectrum. A second aim of the paper (I guess) was to show the improvement of using observations in the far-infrared with respect to pure mid-infrared retrievals (e.g. from IASI). However, I am not really convinced about the outcome of the study in this respect. E.g. assumptions on the measurement uncertainty of the two instruments (FORUM and IASI) are rather different with including all kind of instrumental errors for IASI while using only the spectral noise in case of FORUM. Also, the very different horizontal

sampling/resolution between the two instruments has not been taken into consideration.

Still, I would support the publication of the manuscript in AMT after revision on basis of the comments and suggestions below."

Specific comments:

Retrieval Framework

l.91-99: Please explain a bit more about the species included in LBLRTM as well as the spectroscopy used. Has e.g. the N2-continuum at around 100 cm-1 been simulated? Which water-vapour continuum has been applied?

An additional sentence has been added to line 100 outlining the continuum and number of species, as well as on line 344 related to the updated version of LBLRTM used in the FORUM-aircraft retrievals.

Line 100*: The LBLRTM transmittances include the effects of 28 gas species, alongside continua (MT_CKDv3.2) due to water vapour (self- and foreign broadening), ozone, carbon dioxide, nitrogen, and oxygen (Saunders et al., 2017).*

Line 344: *This version of LBLRTM uses the MT_CKDv3.5 continuum model which, most notably for this study, contains an update to the water vapour continuum, that increases water vapour absorption within the far-infrared (Mlawer et al., 2019)*

l.91-99: It would also be important to know about the uncertainty in the following retrievals introduced by the spectroscopic errors (including errors in the continua). Some discussion about this should be included in the discussions.

There is a brief discussion of the impact of spectroscopic uncertainty on line 450 in the context of the retrievals of upper tropospheric water vapour from the FORUM-aircraft observations.

Line 450*: Figure 13a shows the channels with the highest sensitivity between 5 and 7 km tend to be clustered between 100 and 400 cm^{-1} and above 1400 cm^{-1}, within the water vapour rotation and vibrational-rotation bands. Between 100 and 400 cm^{-1} water vapour spectroscopy, including line and continuum contributions, has been continuously updated over the last couple of decades but is still subject to relatively large uncertainty (Mlawer et al., 2019, 2023), which will contribute to forward model error.*

l.73, 'Retrieval Framework':

Please add (e.g. in a Table) all elements and dimensions of the retrieval state vector, as well as their units. Additionally show (e.g. as supplement) the results and

covariances of all parameters of the state vector (surface-temperature, surface emissivity).

Full details of the state vector dimensions are presented in (Siddans, 2019) with a summary published in (Siddans, 2017) and (Trent, 2023) and are unchanged in this work. The apriori covariance matrices for temperature, water, and ozone are shown in (Siddans, 2019) and are consistent with what was used in the retrievals from simulations in this work.

The retrieval targets and units have been added on line 104.

Line 104: *In this work, we perform simultaneous retrievals of temperature (K), water vapour (ppmv), surface skin temperature (K), and surface emissivity...*

The following has been included in the supplementary information:

- New covariance matrices for the FSI
- The tighter apriori covariance for temperature used in Section 7.2.
- The temperature and water vapour retrieval biases for the FORUM-aircraft configuration
- The retrieved surface skin temperature and emissivity from simulations for the IASI, FORUM, and FORUM-aircraft configurations.

In addition, information should be provided about the influence of other spectrally interfering trace gases (e.g. ozone) on the retrievals: e.g. which profiles are used and kept fixed (e.g. also for CO2, etc)? Have their uncertainty been considered for the selection of channels? What is the expected uncertainty introduced in the actual retrieval?

Information about which profiles are used and kept fixed as a baseline has been added to line 107 but further details are included in each section as they are different in each case.

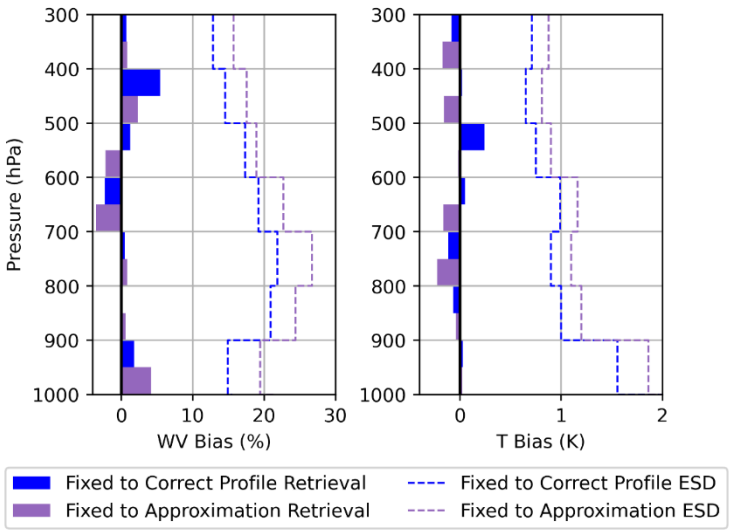
Line 107: *Vertical profiles for CO2, O3, CH4, N2, and other trace gases are fixed to the US 1976 Standard profile unless specified otherwise.*

For the retrievals from the FORUM-aircraft observations, the ozone and carbon dioxide are well constrained. The ozone profile was measured using a TECO 49 UV absorption ozone photometer, and the CO2 was reported by NOAA/ESRL on the day of the flight. Test simulations that used the CAMS reanalysis data closest to the flight produced no noticeable difference to simulated spectra. As a result, the expected uncertainty is minimal, and their uncertainty has not been considered for the selection of these channels.

As an approximation to the uncertainty on the retrieval for the FORUM configuration, the 240 retrievals were rerun with the O3 fixed to ERA5, and CO2, CH4, and N2O fixed to the respective CAMS profile while the 'true' profile in each case remained as specified in the paper.

The impact on the retrieval bias and estimated standard deviation (ESD) is shown below. As expected, the retrieval uncertainty generally increases, as does the associated ESD accounting for these incorrect gases.

Fig A5: The median (a) water vapour (WV) and (b) temperature (T) bias of the retrieved profile to the true state across the testing cases using the FORUM configuration for when the O3, CO2, CH4, and N2O profiles are (blue) fixed to their true values and (purple) when they are fixed to approximations. The O3 profile is fixed to ERA5, while CO2, CH4, and N2O are fixed to their respective CAMS profiles. Biases are evaluated between 300 and 1000 hPa in 100 hPa bins as in Trent et al. (2023). The water vapour percentage bias is calculated from absolute values in ppmv. Dashed lines show the median estimated standard deviation (ESD) for a single retrieval for the test cases in each bin. For water vapour this is plotted with an offset of - 10% for clarity.



l.108:Does the a-priori covariance of surface emissivity also contain correlations (off-diagonal elements) in spectral space?

This covariance matrix does include off-diagonal elements and the updated FSI compatible matrix has been included in the supplementary information.

Line 113: *The a-priori covariance for temperature, water vapour and skin temperature is a 2-dimensional matrix derived from the differences between the zonal mean of ERA-5 profiles for three days (17 April, 17 July, and 17 October 2013). The a-priori covariance for the surface emissivity is derived from surface emissivities from the IREMIS atlas for*

the same time period. Both of these covariance matrices contain correlations manifested as non-zero off-diagonal elements.

l.111: 'yi':I think i should be a subscript to gamma. Is gamma_i decreasing during the iterations?

Yes, you're absolutely correct and this has been changed.

l.130: G, K_x, \hat{x} have not been defined. Please do so.

G and $\hat{x}$ are defined on line 140:

Line 140: *This is, in practice, calculated using the gain matrix,* ***G*** *which is the sensitivity of the retrieval to the measurement and its error, and can be used to smooth a true state,* $\hat{x}$*, , to the resolution of the retrieval:*

K_x is synonymous with K so the subscript has been removed on line 130.

l.135: as you report for the RMSE the logarithm of specific humidity, has this quantity also been used in the retrieval state vector? Otherwise, why not reporting the RSME directly of what is in the state vector?

Water vapour is stored in ppmv within the state vector. As water vapour in ppmv is several orders larger at lower altitudes than at higher altitudes, the log of the specific humidity is used to allow comparisons between different altitudes. This is discussed on line 147.

3 Extension of IMS for FORUM

l.155: Please explain more in detail which instrumental errors are considered in the Sy of FSI as well as of IASI and how these have been produced? Also mention the horizonal resolution on ground as well as measurement time related to the FSI versus IASI measurements those Sy values are valid for.

Additional detail has been added to the measurement covariance for the FSI on line 168. The IASI measurement covariance details are introduced on line 91, however they are further explored here.

Line 168*: The measurement covariance was built using the apodised target noise-equivalent-spectral-radiances (NESR) and target absolute radiometric accuracy (ARA) for the FSI. These two components have been combined to produce a correlated covariance matrix. The leading diagonal of the matrix is shown in Figure 1b. It has been separated into the NESR and ARA components in Figure 1c with the full matrix included in Figure S2*

in the supplementary information. The apodised NESR includes four sets of non-zero off-diagonals due to the thin instrument line shape and is the dominant contributor to the measurement covariance below 1300 cm^{-1} and above 1450 cm^{-1}. The target ARA is 0.1 K between 300 and 1100 cm^{-1}, 0.2 K between 200 and 300 cm^{-1}, and 1100 and 1300 cm^{-1}, and 1K elsewhere. As the ARA is defined in brightness temperature, it is converted into radiance units for each individual spectrum, and the average is shown in Figure 1c for the ECMWF 83-profiles. We assume that the ARA covariance is fully correlated, and so it has contributions off the leading diagonal of up to 0.5 [mW m^{-2} sr^{-1}/cm^{-1}]^2 below 100 cm^{-1} and above 1300 cm^{-1}.

This measurement covariance relates to the FSI 15km footprint and acquisition time of 8s. The leading diagonal of the IASI measurement covariance outlined in Section 2 is shown in Figure 1a. This corresponds to the IASI 12 km footprint which also has an acquisition time of 8s. As can be deduced from Figure 1, the FSI measurement covariance is between 0.08 and 0.37 [mW m^{-2} sr^{-1}/cm^{-1}]^2 smaller than its IASI counterpart between 750 and 1200 cm^{-1}. At wavenumbers above 1300 cm^{-1} the FSI measurement covariance increases, reaching a difference of 0.9 [mW m^{-2} sr^{-1}/cm^{-1}]^2 at 1600 cm^{-1}.

l.167: Matrices of dimension (mxm) and (nxn) cannot be directly multiplied, please correct/modify.

The equation has been corrected to $H = \frac{1}{2}\ln |I - A|$ as in (Rodgers et al, 2000) where the dependence of the information content on the measurement and state covariance is contained within the averaging kernel such that the dimensions of the matrices are now compatible.

Figure 5: Are the water vapour biases plotted as the biases of the log()-values or the real wv-values in the retrieval state vector? If only the log(WV) is shown, please show also the 'native' ones.

The water vapour biases show real wv – values. Additional notes have been added in the caption to figure 5

***Fig 5 Caption**: Biases are evaluated between 300 and 1000 hPa in 100 hPa bins as in Trent et al. (2023). The water vapour percentage bias is calculated from absolute values in ppmv.*

4.1 Retrievals from Simulations

l.227: As the retrievals are performed with noise-free simulated 'measured' spectral radiances, can you explain more clearly what the median retrieval biases are significant for as a retrieval diagnostic?

The aim of performing retrievals on these simulations is to present an average result of the retrieval in this configuration. If a large number of noisy simulations

were used, the random component of this uncertainty would be minimal in the final average. Using the median retrieval bias represents this average without skewing from large/small outliers in the dataset.

The median absolute deviation on Figure 5 has been removed following comments from reviewer 2 as given the simulations are noise free it does not provide much information as a retrieval diagnostic.

l.235, 'the partial derivatives of the forward model output against components of the state vector also contributes to the retrieval uncertainty':

Why not writing 'the Jacobian matrix K'? But why is this mentioned here at all? Is K strongly different between the instruments apart from the channel selection and the different spectral regions?

Following your recommendation this sentence has been removed. K isn't strongly different bar the channel selection and spectral regions.

5 Observations for Testing

l.274: It would be good to mention here also the maximum optical path difference of the interferometer for the measurements used here.

This has been included on line 299. ARIES has a maximum optical path difference of 1.01 cm.

Figure 7, 'taken from ERA5 around the path of each dropsonde':

Please quantify what 'around' means (delta-time, delta-km)?

This has been added into the caption for Figure 7.

***Figure 7 Caption**: ±30 mins, ±48 km)*

6 Development for Aircraft-Level Retrievals

l.345: Can you specify which single errors are included in the instrumental 'Uncertainty' curve in Fig. 9? Especially outline also possible errors due to spectral as well as radiometric calibration?

More information about the 'uncertainty' components has been added in the caption to Figure 9.

***Figure 9 Caption**: The uncertainty used for FORUM-aircraft configurations is overlaid on an example simulated upwelling spectrum at the altitude of the higher SLR. This is composed of the instrument uncertainty and the uncertainty that arises from the FORUM apodisation process discussed in Section 6.1. This uncertainty squared is used as the main diagonal of **Sy**. The instrument uncertainty is the radiometric calibration and random noise for each instrument combined in quadrature for a single TAFTS scan and six ARIES scans. For both instruments, this uncertainty is dominated by the calibration, particularly for ARIES, where the random noise is reduced due to multiple scans. The impact of the FORUM apodisation process can be seen most clearly in the 15 micron CO2 band and between 1200-1400 cm^{-1}. The 200 channels selected for the aircraft retrievals are also shown.*

l.362-368: It would be good to see these results as figures (e.g. in the appendix).

Now included in the supplementary information.

l.364, 'looser': Clearer to say 'smaller' or do you mean 'larger' values?

Changed from looser to larger (now on line 385)

7 Retrievals from Aircraft Observations

l.385, 'AK-treated':

Please provide the formula used (including averaging kernel and a-priori profile).

The equation has been added on line 399.

***Line 399:** This is referred to as AK-treated dropsonde measurements, x_{AK} and is calculated using the following equation:*

$$x_{\mathrm{AK}} = x_a + \boldsymbol{A}(x - x_a)$$

where the averaging kernel used is calculated for each individual retrieval.

Figure 12a:

Please also show the deviations of WV also as percentage as well as absolute (e.g. in a supplement) differences of WV (ppmv) from the 'Drops_AK'. With always showing the log-scale plots, it is very difficult to follow the argumentation.

Now included in the supplementary information.

l.398, 'This tightening of the a-priori temperature covariance also leads to an increase in the water vapour AKs':

Can you make it plausible why this is the case?

Information about the temperature and water vapour profiles significantly overlaps, and so as you tighten the apriori temperature covariance (i.e increase confidence in the temperature profile) you increase the available information about the water vapour profile.

l.436, 'however this remains within the measurement covariance':

Has the measurement error also been divided by sqrt(n) where n is the number of samples? Otherwise a direct comparison to the measurement error of one observation might be misleading.

The measurement error contains the calibration and noise added in quadrature for each instrument combined with the uncertainty associated with the apodisation process.

The mean residual in each wavenumber bin (shown by individually coloured stars) is calculated for a single FORUM-aircraft observation (i.e a single sample (n=1)). The measurement error in each bin has then been averaged across each channel within the bin.

Chapter 8 Conclusions

Please discuss here (or before, e.g. in the introduction) the general differences in the method as well as the results between your analysis of the aircraft measurements and the work by Warwick et al., 2022 (https://doi.org/10.1029/2020JD034229).

Additional information has been added on line 66 (introduction) and line 423 (Section 7.1)

Line 66: *The second builds on the earlier work of Warwick et al. (2022) by adapting aircraft-based observations to mimic the expected FSI instrumental characteristics, developing what we term FORUM-aircraft radiances. We also include a greater number of radiometric observations from the flight and optimise the IMS channel selection for the FORUM-aircraft configuration.*

Line 423: *This method enables the joint retrieval of temperature, water vapour, surface emissivity, and surface skin temperature that was not possible in Warwick et al., (2022) due to the limitations outlined for the ARIES spectra.*

l.484, 'The performance of IMS was first assessed…':

It is not clear to me what the aim of the paper is. Is it the assessment of the retrieval code or the assessment of the FORUM vs. IASI performance or the analysis of the aircraft observations? On l.63 it is written: 'To validate this scheme in the far-infrared..'. However, I'm not convinced how retrieval test, with simulated as well as measured data 'validates' the processor? Please try to make this a bit clearer in the introduction and even in the title where the retrieval scheme validation is not mentioned.

The aim of this paper is to present the extension of this retrieval scheme as a tool that will be ready for use on the FORUM Sounding Instrument. We expect that following the FSI launch, this retrieval scheme will need further updates, however this preliminary setup is ready to be applied to observations. In an ideal scenario, this scheme would be tested on TOA observations to assess the likely quality of its products from the FORUM satellite. However, as these are not currently available, we have tested it on simulations to cover a variety of atmospheric profiles, and then on aircraft observations to account for potential spectroscopic differences.

l.491, 'Because FORUM has not flown yet these aspects are not possible to quantify':

However, there are requirement targets/threshold on the instrumental performance (e.g. calibration, etc) which could have gone into this analysis. Please explain why only the noise was used here.

Please see main comment where this has been corrected.

Reviewer 2 Comments

The paper reports the adaptation of a retrieval code developed for IASI to the analysis of FORUM-like measurements. It is an interesting paper that can go ahead to be published.

However, I have some concern on the validity of the reached results and their application to the forthcoming FORUM instrument.

- The forward model used to simulate the FORUM measurements uses a 0.3 cm-1 sampling, but no mention is made to the spectral resolution. Is it the same as the spectral sampling? If so, the resolution of FORUM reported in ESA documentation is about 0.5 cm-1 and not 0.3 cm-1, and if the the spectral resolution used in this paper is 0.3 then this difference affects the conclusion reached in the paper, that cannot be directly applicable to FORUM.

  The spectral resolution of FORUM within the forward model is 0.6 cm-1 and this has been added on line 155 for further clarity.

***Line 155**: The RTTOVv12 FSI radiances cover the full FSI spectral range at a sampling of 0.3 cm^{-1} and have the strong Norton-Beer apodisation applied. This results in a total of 5000 channels and a spectral resolution of 0.6 cm^{-1}.*

The simulated spectra are produced with a Norton Beer apodization, that makes the simulated channels signal spectrally correlated. However, in the retrieval, you use a diagonal covariance matrix, that neglects these correlations. This will affect the results of your retrievals. Moreover, since the covariance matrix is used to select the channels used in the retrieval, the whole process is not completely robust.

See main comment. We now include the off-diagonals in the measurement covariance.

- The error used for the FSI includes only the NESR, while the one for IASI includes a lot of other components. Some of these could be deduced from the FORUM documentation (i.e. the required radiometric accuracy that can be used to deduce the expected calibration error) therefore the results presented here are biased in favor of the FORUM instrument

  See main comment. Following your suggestions we now use the target radiometric accuracy to deduce the expected calibration error.

- The simulated retrievals are performed using the MERRA2 data. This is fine, but I think that, since the diverse profile dataset used in the development of the RTTOV code are relative to 2008, the use of any atmosphere of the more recent years could have been safely used for these tests.
- You do not retrieve ozone. No mention on which profile of ozone you use in the retrievals of the simulations is made in the text, while you correctly report the value for the retrievals of real observations

The ozone profile is fixed to the 'true' MERRA-2 profile that is used in the initial simulation. Clarification has been added on line 221.

***Line 221**: Ozone is not a retrieval target in this work and is fixed to the 'true' profile in each retrieval, but it has been chosen to vary across the test cases given its strong absorption in the mid-infrared.*

- You report all the equations of the optimal estimation (that are very well known) but you do not report any description of the used quantifiers (measurement cost per channel, State cost, cumulative degrees of freedom, hPa/DOF etc. I suggest adding few sentences describing what they represent

Information about the measurement/state cost has been added on line 124. Further clarification about the RMSE value has been added on line 147. Information about the cumulative DOFS has been added on line 246.

Line 124: *The measurement cost indicates the fit of the simulated and observed spectra in relation to the Sy in the selected channels. A larger measurement cost per channel indicates a worse fit and suggests that the retrieved state vector is less likely to match the true state. The state cost indicates how far the retrieved state has deviated from the apriori in relation to Sa. A high state cost could be caused by either an apriori which is not similar to the true state, or a small Sa, which suggests the retrieval is tightly constrained.*

Line 147: *The RMSE values for water vapour are calculated using the units of the logarithm of specific humidity as concentrations can vary by several orders of magnitude throughout the vertical, and are used as a summary value to further assess the quality of the retrieval.*

Line 246: *This is shown in the median cumulative degrees of freedom for signal (CDOFS) in Figure 4 which is indicative of the vertical resolution of the retrieved profile.*

- You perform the simulated retrieval on noiseless simulated measurements. Why? No explanation for this is given in the paper

The aim of performing retrievals on these simulations is to present an average result of the retrieval in this configuration. If a large number of noisy simulations are used, the random component of the uncertainty would be minimal in the final average.

- The part of the paper describing the retrievals of the real observations (TAFT+ARIES retrievals) is extremely long and reports a lot of tests that are somehow redundant for the purpose of the paper. Moreover, the description of the tests is really confused, and it is very hard to follow. I suggest either to shorten it or try to report the results in a clearer way.

We agree that this paper is very long and so as well as removing some of the tests mentioned in Section 6 and 7, we have also aimed to tighten the entire paper to make it easier to read.

- When the dropsonde data are reported, you say AK treated drop-sonde. Which AK are you applying?

  The AK varies between each individual case and as such a different averaging kernel is applied to each dropsonde for each retrieval. Clarification has been added on line 407.

Line 399: *This is referred to as AK-treated dropsonde measurements,* $\boldsymbol{x}_{\mathrm{AK}}$ *and is calculated using the following equation:*

$$\boldsymbol{x}_{\mathrm{AK}} = \boldsymbol{x_a} + \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x_a})$$

where the averaging kernel used is calculated for each individual retrieval.

Specific comments

Retrieval Framework

Line 119: why the convergence is reached when the change in the chi-square is less than 1?

The convergence of the retrieval is dependent on three conditions: the reduction of chi-square, that chi-square is the lowest value so far, and as you stated that the change in chi-square is less than 1. These are variable settings within the retrieval framework, and these conditions were found based on iterative tests of the retrieval framework on IASI observations.

Line 155: 'FSI measurement covariance is smaller', please quantify this

Following the changes made to the FSI measurement covariance, this paragraph has been rewritten. Line 180 now includes a comparison between the IASI and FORUM measurement covariance

Line 180: *As can be deduced from Figure 1, the FSI measurement covariance is between 0.08 and 0.37 [mW m^{-2} sr^{-1}/cm^{-1}]^2 smaller than its IASI counterpart between 750 and 1200 cm^{-1}. At wavenumbers above 1300 cm^{-1}, the FSI measurement covariance increases, reaching a difference of 0.9 [mW m^{-2} sr^{-1}/cm^{-1}]^2 at 1600 cm^{-1}.*

4 Simulated Test Cases

Lines 205-208 was the surface emissivity in the retrievals the same of the simulations?

The surface emissivity was also retrieved. The median retrieved surface emissivity was comparable between the IASI and FORUM configurations in their overlapping spectral range (645-1600 *cm^{-1})* and the difference to the input surface emissivity was of the order of 10^{-2}. Plots have been added to the supplementary information.

Caption of Figure 5 '*Dotted lines show the median absolute deviation (MAD) in the median retrieval bias in each bin'* Since you use the optimal estimation this value is

not significative, since if you have low information the results stay close to the a-priori and the MAD is very low

Per your recommendation the MAD variable has been removed from Figure 5.

Lines 233-235. The sentence is really not clear, can you rephrase it?

'While this is in part due to the reduced measurement covariance for the FORUM configuration, the partial derivatives of the forward model output against 235 components of the state vector also contributes to the retrieval uncertainty (equation 3).'

Following comments from Reviewer 1 this sentence has been removed as it doesn't add useful information.

5 Observations for Testing

The title of section 5.1.1 should be changed

Changed from 'Radiation Instrumentation' to 'Radiometric Instrumentation'

Line 269 'is substantially less than' please give estimates

Added on Line 294.

Line 294: *is of the order ±0.5 of [mW m^{-2} sr^{-1}/cm^{-1}]*

6.1 Simulating and Constructing FORUM-aircraft observations

The title of section 6.1 suggests that you describe how you treat TAFT and ARIES measurements to resemble FORUM but the section describes the way RTTOV simulates the measurements in the retrievals. Please change it

This section is now entitled "Constructing the FORUM-aircraft observations" and has been shortened. Information about the RTTOV simulations have been summarised in the section above.