The FORUM End-to-End Simulator project: architecture and results. Answers to the referees

Luca Sgheri[1], Claudio Belotti[2], Maya Ben-Yami[3], Giovanni Bianchini[2], Bernardo Carnicero Dominguez[3], Ugo Cortesi[4], William Cossich[5], Samuele Del Bianco[4], Gianluca Di Natale[2], Tomás Guardabrazo[6], Dulce Lajas[3], Tiziano Maestri[5], Davide Magurno[5], Hilke Oetjen[3], Piera Raspollini[4], and Cristina Sgattoni[1]

[1]IAC – CNR, Via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy
[2]INO – CNR, Via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy
[3]ESA – ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
[4]IFAC – CNR, Via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy
[5]Physics and Astronomy Department, University of Bologna, Bologna, Italy
[6]DEIMOS Space S.L.U., Ronda de Pte. 19, 28760 Tres Cantos – Madrid, Spain

1 Answer to the referees

We thank both referees for reading a quite long paper and suggesting improvements. The required additional data to be shown will make the paper even longer, but we try anyway to adhere to all requests.

1.1 First Referee

This paper describes a complete End-to-End Simulator (E2ES) for the FORUM mission, with the purpose of estimating the performances of the FORUM Sounding Instrument, which will observe the Earth from the orbit in the Far Infrared (FIR) for the first time. The authors showcase different simulations in clear and cloudy sky, and perform retrievals on synthetic data, reporting accuracy and sensitivity to atmospheric and surface parameters, comparing the E2ES performances to other codes (KLIMA and SACR). Applications to real scenes from MODIS data are also presented as an example to show how the E2ES deals with real data.

Overall the manuscript presents the fundamental aspects of the E2ES, and gives interesting elements to evaluate the impact of including the FIR in the observed spectral interval when a simultaneous retrieval of many parameters is performed. Therefore the content of this work is certain to be of interest for the future activities involving the FORUM mission preparation. However, there are several revisions that need to be implemented to this manuscript before it can be considered for final publication in AMT, as the text suffers particularly from lack of clarity on several aspects.

1.1.1 Major comments:

1) The work does not provide any specific detail about the computational performances of the E2ES, and in particular of the SGM and L2M_I modules, which are heavily involved in the operational use. While the accuracy is assessed (and also based on a long series of studies and on the fact that the E2ES is based on well-established codes such as LBLRTM), the work should provide many more details about: 1) the computational time required to generate a scene with the SGM, and 2) the average time required to perform retrievals in each one of the cases explored in this work. Based on this, the work should also contain a discussion about eventual strategies to improve performances.

DONE, added a full section.

2) While the work contains a lot of details about retrieved parameters, there is very little shown about retrieved vs. modeled radiances. Those are especially important as the work deals with the FIR, for which no hyperspectral observations from orbit exist. More specifically, the work should show: 1) spectra for the case 1.1 to 7.1 (or at least a sample of those) with simulated observations, best fit from L2M_I, KLIMA, and residuals; 2) spectra and residuals related to the results in Fig. 16 and 17; 3) spectra and residuals related to the results in Fig. 18 and 19; 4) the spectra for the MODIS scene case.

ANSWER: We added spectra and residuals from L2M_I for the seven clear cases in Figure 5. Residuals are well comparable with the given noise, and residual from KLIMA and L2M are almost identical, as can be seen from the chi2 reported in Table 6. Thus, difference in residuals and difference between simulated (by KLIMA and SGM) and fitted spectra are not really appreciable. Only residuals between simulated observations by the E2E and L2M_I best fit spectra were reported and, to highlight small residuals, we had to use a double scale in this and the subsequent similar figures. We added spectra and residuals for the cloud contamination cases in Figure 18, for the heterogeneous surface cases in Figure 21. Finally, MODIS spectra are reported in Figure 25.

3) Regarding the MODIS case, because the spectra are not shown, it is unclear - at a first glance - whether actual MODIS data are fitted (obviously adapting the Radiative Transfer to the spectral response of MODIS) or SGM scenes are generated using the MODIS L2 products and the described databases. It seems like the latter is the case, and this should be properly conveyed by the text, as the title of the section mentioning "MODIS data" is misleading.

ANSWER: The scenes are generated using MODIS L2 products. The title of the section is changed to “Tests based on MODIS L2 products”. The text is improved to better clarify that MODIS L2 data are mapped into the FEI grid and used as inputs to compute the spectral radiances for each grid point. Radiances are calculated outside the SGM, by using a dedicated subroutine, and stored into the FE2ES as ancillary data.

The modifications of the text are here reported for your convenience:

The surface, atmospheric and cloud parameters used to build the scene are derived from the following databases:

- Surface emissivity: the Global Land Cover Map (GlobCover 2009, http://due.esrin.esa.int/page_globcover.php) is an ESA’s composite product based on MERIS (Medium Resolution Imaging Spectrometer Instrument) data collected at 300 m spatial resolution. The GlobCover map provides 22 land cover classes, defined with the United Nations land cover classification system, that are associated to the 11 surface types from Huang et al. (2016) used in the SGM surface definition;

– Atmosphere: the MODIS Atmospheric Profile product (07_L2, https://modis-images.gsfc.nasa.gov/products.html) defines the temperature and water vapor profiles, and the surface height at 5 km resolution;

– Cloud properties and surface temperature: the MODIS Cloud Product (06_L2) are used to define the cloud parameters (particles phase and effective radius, cloud top height and optical thickness) and the surface temperature, with spatial resolution of 1 km.

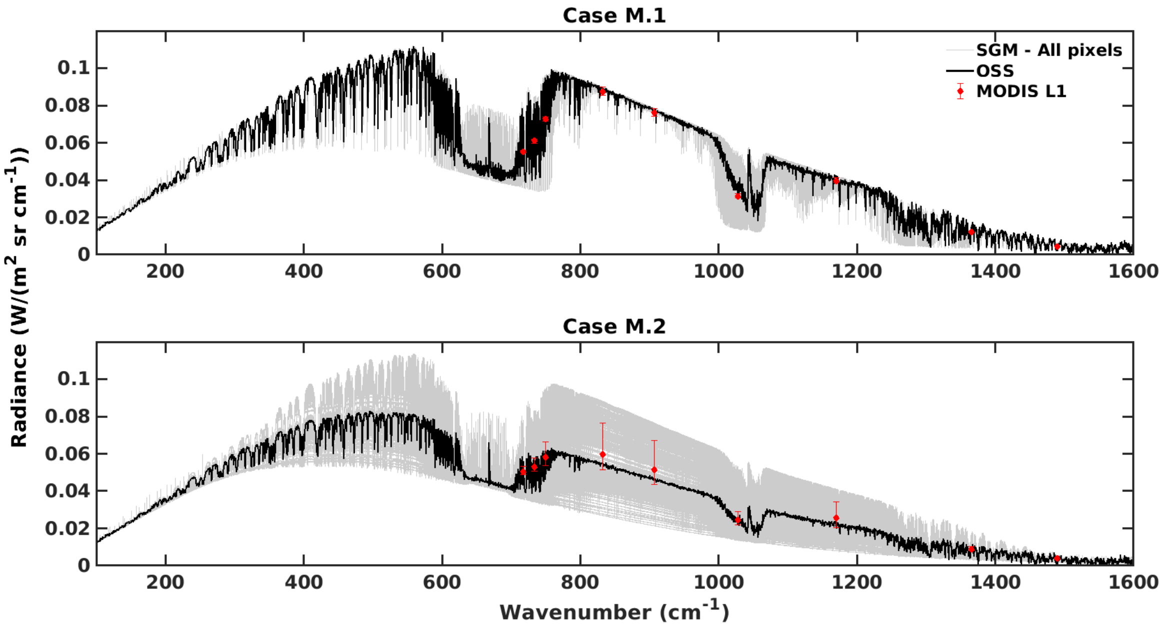
The above products are characterized by different grids and spatial resolutions; thus, they are remapped into the FEI grid. Once the cloud, atmospheric and surface information are available at the same grid, a dedicated subroutine, mimicking the SGM, is applied to compute the spectral radiance for each pixel. Due to the complexity of the process the subroutine is kept external to the E2ES chain. Nevertheless, following the above procedure, multiple scenarios can be built from satellite products for complex atmospheric conditions. The radiances are stored in the E2ES as ancillary data and used for testing the FORUM E2ES performances on realistic conditions.

Figure 24 shows examples of the surface and cloud parameters remapped into the FEI grids. The scene observed in the M.1 clear sky case (top panel) has irregularly sparse urban areas (brown) surrounded by different kinds of vegetation (light and dark green) and few wet areas (blue, probably lakes or ponds), creating a much more complex scenario than an ideal homogeneous case. Similarly, clear sky areas alternate to high and low level clouds, with unevenly distributed optical and microphysical parameters, in the cloudy case M.2 (bottom panels). Multiple spectral radiances are computed by the SGM for all the FEI grid points within the FSI FoV (the black circle in the panels of Figure 24), according to the different properties of the scene. Within the E2ES, the pre-computed radiances are ingested by the FSI-OSS module to produce a single spectral radiance. In figure 25 the spectral radiances computed for each FEI pixel inside of the FSI FoV (by using the dedicated subroutine mimicking the SGM) are plotted together with the FSI-OSS output radiance for the cases M.1 and M.2. The figure also reports radiance measurements in multiple MODIS channels (red dots) with their variability within the FSI FoV. Note that the FSI-OSS spectral radiance based on MODIS products and the mean MODIS radiance at specific bands are consistent, but a perfect match is not expected due to the non-linearity between L2 products and the radiance.

Finally, the L2M is used to retrieve the characterizing parameters of the complex scenes, in particular: temperature and water vapour profiles, and surface skin temperature and emissivity for the case M.1, correctly classified by CIC as a clear sky scene; cloud properties for the cases M.2 and M.3, classified by CIC as a cloudy scene."

1.1.2 Other/minor comments:

1) Lines 78-79: This sentence is not very clear: more than "routines", the FORUM data will likely provide feedback about the foundations of radiative transfer in the FIR in the Earth's atmosphere itself. Unless this is a quote from the document ESA (2019), it should be slightly clarified.



Spectral radiances computed by using the dedicated subroutine mimicking the SGM (grey lines) for each FEI pixel within the FSI field of view are plotted with radiances obtained from the FSI-OSS (black lines) for the cases M.1 (top panel) and M.2 (bottom panel). The MODIS radiance at specific bands are also reported in red dots. The vertical red bars indicate the MODIS radiance variability within FSI FoV.

DONE: We agree that this phrase is confusing. We changed the sentence to: 'the dataset aims at providing a benchmark for the validation of radiative transfer models and with this also for radiative routines used in climate models through the use of satellite simulators (ESA, 2019).'

2) Line 81: A more recent work about continuum in the FIR, which goes at wavelengths as long as 200 um, is the following, and it is recommended to cite it: https://www.sciencedirect.com/science/article/pii/S0022407320310141?via%3Dihub

DONE

3) Line 82: point ii) needs to be more specific, as in the present form seems redundant w.r.t. the other elements mentioned here. Is the text referring to, e.g., the way in which aerosol scattering is handled? Test how high opacities by water are treated?

DONE: We agree with the reviewer. Actually the water vapour continuum belongs into the category of the 2nd bullet point and not in the 1st that refers to spectroscopy. We have updated the text accordingly.

4) Line 88: among the works regarding parametrization of cirrus clouds, perhaps it is also appropriate to add the most recent reference to the work by Martinazzo et al. (2021): https://www.sciencedirect.com/science/article/pii/S0022407321002326

DONE

5) Line 113: can the text be more specific about the requirement of having a spatial sampling of 0.6 km for FSI to be sensitive to clouds?

ANSWER: We reported the references where we found this assertion. Actually, we had no control on this value. We suspect the FEI has a standard 60x60 matrix, the 0.6 km permits to have both a good coverage of the FSI FoV with ample margin on the adjacent zones and a good resolution to identify heterogeneities.

6) Lines 136-137: This is not necessarily true, as the chosen approach is not equivalent to knowing how "radiative transfer works". Rather, it is useful to "making sure that the radiative transfer introduces no bias in the retrieval w.r.t. simulated observations". The text should be revised to be more precise in this sense.

DONE

7) Line 179: can the authors specify the 11 gases considered in the radiative transfer?

DONE: The 11 gases considered are: CO2, O3, N2O, CO, CH4, O2, NO, SO2, NO2, NH3 and HNO3. This information was added in the text:

The atmospheric profiles of temperature, pressure, water vapor, and other 11 gases mixing ratios, interpolated at fixed altitude levels, are derived from the ERA 5 reanalysis and the IG2 climatological data (Remedios et al., 2007). The gases included in the simulation are $\mathrm{H_2O}$, $\mathrm{CO_2}$, $\mathrm{O_3}$, $\mathrm{N_2O}$, CO, $\mathrm{CH_4}$, $\mathrm{O_2}$, NO, $\mathrm{SO_2}$, $\mathrm{NO_2}$, $\mathrm{NH_3}$, and $\mathrm{HNO_3}$. Cross section data using predefined values for the $\mathrm{CCl_4}$, $\mathrm{CFC_{11}}$ and $\mathrm{CFC_{12}}$ molecules were also added.

8) Line 188: the text is not sufficiently clear about how the aerosol properties are derived over the whole FORUM spectral range, as only the ones at 900 cm-1 are mentioned.

DONE: New text is inserted to better clarify how the aerosol properties are derived. The modifications of the text are here reported for your convenience:

The cloud optical depth (OD) at 900 $\mathrm{cm^{-1}}$ is derived, layer by layer, from information about the IWC (in case of ice clouds) or LWC (in case of liquid water clouds) and the effective radius of the particle size distributions (PSD). The equation defining the OD at 900 $\mathrm{cm^{-1}}$ for vertically homogeneous clouds of geometrical thickness Δz is

$$\mathrm{OD} = \mathrm{IWC}\beta(R_{eff}, 900)\Delta z \tag{1}$$

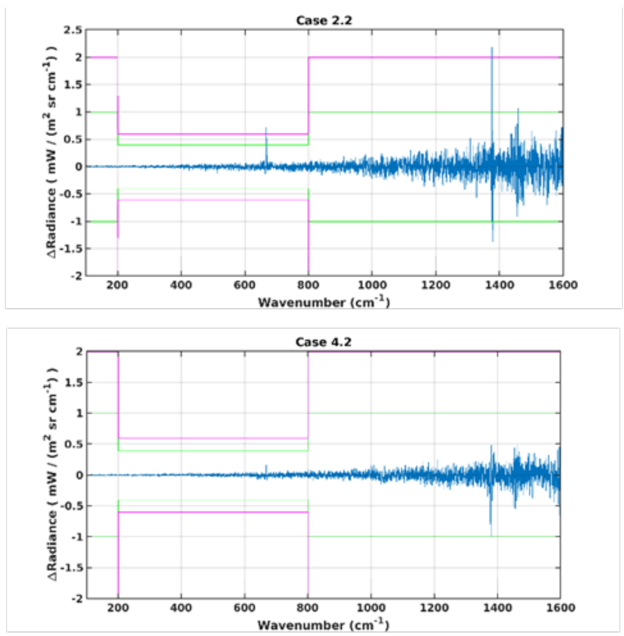
where $\beta(R_{eff}, 900)$ is the extinction coefficient at 900 $\mathrm{cm^{-1}}$ of the PSD corresponding to a specific effective radius R_{eff}, normalized with respect to a unit IWC (or LWC for liquid water clouds).

Water clouds are simulated as PSDs of liquid water spheres while ice clouds are assumed as composed of 8 columns aggregates crystals. Single scattering properties for single particles (SSSP) are combined to compute the optical properties (extinction, scattering and absorption coefficients and the phase function) for a set of PSDs over the whole FORUM spectral range. The assumed PSDs are modified gamma distributions (Hansen, 1971) representatives of effective radii from 2 to 200 micrometers in case of ice clouds and from 1 to 30 micrometers in case of liquid water clouds. The SSSP are derived from

Yang et al. (2013) in case of ice crystals and computed by using a Mie scattering code (Peña and Pal, 2009) when liquid water spherical particles are assumed.

9) Line 193: the text should be explicit about the reasons why two different resolutions are adopted for cloudy and clear cases. This is customary in other codes.

ANSWER: The default spectral resolution for clear sky conditions is $10^{-3}\ \mathrm{cm}^{-1}$. The typo is corrected in the paper. Tests have been performed to evaluate the effect of using a lower resolution ($10^{-2}\ \mathrm{cm}^{-1}$) in case of cloudy conditions instead of the higher resolution adopted in clear sky ($10^{-3}\ \mathrm{cm}^{-1}$). Results are reported here below for your convenience as difference between FSI-OSS radiances. The tests refer to the conditions of cases 2 and 4 and show that the adoption of a lower resolution implies differences with respect to the adoption of the higher resolution that are within the FORUM FSI goal measurement error (NESR).



Differences between FSI-OSS radiances, computed from a lower ($10^{-2}\ \mathrm{cm}^{-1}$) and a higher ($10^{-3}\ \mathrm{cm}^{-1}$) spectral resolution, for the cases 2.2 (top panel) and 4.2 (bottom panel).

The computational times for the two cases presented are:

	resolution=$10^{-2}\ \mathrm{cm}^{-1}$	**resolution=**$10^{-3}\ \mathrm{cm}^{-1}$
case 2.2	12 min	5 hours
case 4.2	2 hours	26 hours

The text is modified as follows:

These radiative transfer routines compute the spectral radiances for the FORUM imager and sounder. The default output spectral resolution is set to $10^{-2}\ \mathrm{cm}^{-1}$ and $10^{-3}\ \mathrm{cm}^{-1}$ for cloudy and clear sky simulations respectively. The different resolution is adopted for the two cases as a compromise between computational speed and accuracy. In fact, for a resolution of $10^{-2}\ \mathrm{cm}^{-1}$ the computational times in cloudy conditions decrease by a factor ranging from about 10 to about 20 with respect to using a $10^{-3}\ \mathrm{cm}^{-1}$ resolution. Nevertheless, the simulation accuracy is maintained since OSS-FSI radiances in cloudy conditions computed using 10^{-2} or $10^{-3}\ \mathrm{cm}^{-1}$ differ less than FSI goal NESR (Figure not shown).

10) Lines 240-241: the "level 1 data analysis code" component of the OSS is never cited before and if it is part of the OSS seems redundant to mention it here. Is this the same "level 1 module" as the one mentioned in line 273?

- DONE The level 1 code is the part of the OSS code which is included in the three level 1 submodules, and it is mentioned immediately at the beginning of the OSS section in order to stress the main subdivision in the OSS code, i.e. instrument simulator and level 1 analysis. The following text adds more detail introducing the code modularity and the internal subdivision of the level 1 code, so we would like to maintain the redundancy for clarity.

11) Line 281: How is the resampling performed? Can the text provide some more details on this?

DONE the resampling is done through FFT interpolation and zero padding in case of oversampling. We clarify this in the text.

Citing the ATBD:

the resampling of the output on an arbitrary wavenumber grid will be performed by using an FFT/IFFT algorithm in which if the final sampling step is smaller than the original one, zero padding is applied. In the specific case, a sampling step equal to the instrumental resolution (0.5/MOPD) will be applied, obtaining in this way a final output spectrum which is optimally sampled, i.e. all the spectral channels are uncorrelated and carry independent information.

12) Lines 355-356: the text does not explain why the choices made in the simulations "do not have an impact on the assessment of the performances of the FEES". Those performances (as shown later in the paper) relate indeed also to the degree of inhomogeneity of the FoV.

ANSWER: Here we are not talking about inhomogeneity of the FoV, but only about atmospheric profiles. Clarified also according to the remark of the other referee.

13) Lines 384-390: The nomenclature of some of the variables used in the retrieval process is confusing. Is one case the background term in the OE equations and the other one the first guess? The authors should clarify the nomenclature of the different terms ("initial condition" and "a-priori" vs "first guess" and "background") and there should be no ambiguity between the two.

DONE. We removed in the paper the two references to background and first guess, so now only a-priori and initial guess are used.

14) Lines 391-392: This is not entirely accurate, In cases where the surface T and emissivity are retrieved simultaneously, the regularization of the two simultaneously could easily bring the retrieval to a local minimum if not properly regularized. The best choice of the first guess value for emissivity is indeed the background in real-data applications (as it comes from climatology), however the regularization of the combined retrieval of emissivity and surface temperature is a separate problem.

DONE Correct, we reworded the sentence mentioning the issue that is discussed in the paper.

15) Lines 394-400: as it is evident from this description and from the references contained, the KLIMA FM is heavily based on the same linelist and continuum used by LBLRTM, and it works with the same line-by-line approach. The SGM itself uses LBLRTM. Since this part of the work describes the comparison between the KLIMA and L2M performances, it would be useful to also state what are the key differences between KLIMA FM and the way LBLRTM is used in the SGM, as this gives the full perspective about which components of the two codes are effectively independent.

DONE: We pointed out that the difference between KLIMA and LBLRTM FM lies in the line shape. While KLIMA uses the Voigt line-shape, LBLRTM uses a combination of simpler functions, which grants it a faster execution time.

16) Lines 433-434: the retrieved emissivity seems to have a lot of unphysical structures (on its sampling of 5 cm-1) which is problematic as it could, in principle, bias the retrieval of atmospheric parameters. The text should discuss if such bias have been observed and if they are significant, and ideally include this aspect among the directions to further improve retrievals in sight of FORUM operational retrievals.

ANSWER: Actually this paragraph is more devoted to showing differences (if any) between the two codes. In the following section (Discussion of the results) we showed how to cure the bias that we (both L2M and KLIMA) observed in one case. We revised and extended the part according to your suggestion.

17) Lines 503-506: This part of the text is contradictory, as it says that the objective is "to perform a realistic validation" and then it says "we cannot truly speak of validation". Also, the text does not fully explain why the atmospheric parameters are perturbed and not retrieved and why should this grant a realistic validation.

DONE. The text was not clear. While an advanced version of the inversion module is indeed able to retrieve simultaneously the cloud and atmospheric parameters, the aim of the project was to show that, even with a perturbed atmosphere, the cloud parameters could be detected within reasonable bounds. We clarify this in the text.

18) Line 523: In an Optimal Estimation approach, it is customary to internally normalize the matrices and arrays such that the value of each parameter is comparable to the others. Therefore this point is not necessarily granted.

CORRECTED. We only did this partially (i.e. the possibility of retrieving the log of the water vapour profile), and it is surely a suggestion that we will keep into consideration. However, in the specific case, the ranges of the cloud variables are comparable. Our statement was not only linked to different ranges, but also different behaviors due to different dependence in the forward model. We modified the sentence accordingly.

19) Line 524: to prove this point it is necessary (and very informative to the readers, given the novelty of the spectra interval as well) to show a full a-posterior covariance matrix resulting from the analysis for this case.

ANSWER: We added a sentence to report the correlations for the cloudy case. We also added a figure for the correlation marices in two clear cases.

20) Lines 562-563: This is not shown anywhere in the section, and it should be rephrased or not mentioned.

DONE. Rephrased, according to results of figure 13.

21) Figure 16: it would be very useful to indicate the boundaries of the cloud in the plots (e.g., with a gray patch or horizontal lines).

DONE

22) Lines 645-646: While this is certainly true, the emissivity exhibits much lower variability than other cases (e.g. a scene with sand + canopy). A much more challenging test would include a scenario like this and would give valuable information on the validity of the scheme. Have the authors considered to include this in the analysis?

ANSWER: We agree emissivity retrieval needs definitely more attention (see e.g. the paper under discussion in AMT by Maya Ben Yami et al.: https://amt.copernicus.org/preprints/amt-2021-232). Our paper reports the first results in many areas, so there are surely many additional tests that can be performed, but time and space limit our effort.

23) Lines 647-648: is the surface temperature retrieved? Can it be added to Figure 19 (e.g. as a marker on the bottom x axis)? Also, are the emissivity fractions assumed as a-priori input in the retrieval?

DONE: Surface Temperature is reported in table 11 (which tex decided to put a bit far from the text referring to it). We also added the used a-priori for emissivity which is 1 (actually snow emissivity + 1 sd, constrained to 1) and tskin.

24) Figure 18: Using a light grey for the errorbars would significantly improve the readability of the figure.

DONE

25) Figure 20: at which wavelength is the OD represented? (perhaps 900 cm-1, but it should be specified)

DONE, in the figure caption.

26) Line 680: the reason for this limitation and how this will be dealt with in an operational context should be specified.

DONE. The SGM is the module that builds the scene. The computer time needed to build the scene would be too large if all the pixels in the FoV of the FEI and of the FOV should be reconstructed separately. Specified in the text, but check also the new timing section added. Therefore these calculations needs to be done offline. In an operational context normally you don't want to simulate data.

27) Figure 21: like in the previous cases, the difference between true and retrieved profiles compared to the errorbars should be shown, as it is difficult to see here.

ANSWER: The problem here is that we do not know what is the true profile because each pixel of the FOV has a distinctive profile. We checked different alternatives, we hope the chosen solution (showing the t and h2o profile range along with the retrieved profile and retrieval error) is informative and easy enough to read.

28) Line 735-736: as said in one of my previous comments, the text should be more detailed on how to improve retrievals, perhaps gathering the main criticities found during this work in a bullet point list.

DONE

1.1.3 Technical comments:

1) There is a series of acronyms not spelled out in the text: OLR (line 74), FEES (line 346, perhaps FORUM End-to-End Simulator?), GN (line 390), NLSF (line 410), and maybe other for which the authors should carefully check the text. Also, the use of "E2E simulator" in a couple of instances should be replaced by "E2ES".

DONE, thanks for pointing these out.

2) Line 38: "absorption" instead of "coefficients".

DONE

3) Line 128-130: This passage is a little repetitive, the text can be shortened.

DONE

4) Table 5: Is this the total OD or the OD per km?

DONE, Total inserted in table caption.

5) Table 6: The Table could be improved if reformatted with two columns, one for KILMA and one for L2M_I.

DONE

6) Line 464: the "substantial features" are usually the Quartz Reststrahlen bands, and this should be explicited.

DONE

7) Line 587: The FSI is the instrument. Perhaps the authors mean the OSS?

DONE

1.2 Second Referee

This paper from Sgheri et al. describes the End-to-End Simulator (E2ES) developed to fully investigate the potentiality of the future Earth Explorer 9 FORUM (Far-infrared Outgoing Radiation Understanding and Monitoring) mission that will be launched in the next years. The work explains in detail the architecture of the E2ES and then investigate the sensitivity of the acquired signal in different scenarios to atmospheric and surface geophysical variables through retrieval tests performed by the L2M_I module inverting the synthetic radiances. The E2ES retrieval module has been validated against the KLIMA and SACR independent codes, finally it has been tested for inhomogeneous scenes and in a real scenario built up from MODIS data.

The work is overall well written and structured, and it will represent reasonably a benchmark for the FORUM preparatory phase during the next years. However, some revisions have to be implemented on the manuscript before its final publication on the AMT journal.

1.2.1 Major comments:

The comparison of the retrieved products from L2M_I and KLIMA codes is fully investigated in Section 5.1, however I agree with point 2) of major comments and suggestions from Referee 1. Furthermore, the comparison between L2M_I and SACR codes has been poorly inspected, since relevant differences in the retrieved parameters reported in Table 9 have not been discussed. The cloud parameters retrieved by L2M_I are generally in agreement with true values, but L2M_I and SACR are generally not consistent (e.g. case 2.2, where difference on cloud top amount to 5.6 km). It is not clear at all if the SACR code tends to converge to local minima, however it should be clarified studying the retrieval quality through the chi-square reduction for - at least - two or three critical cases. Therefore, the authors should assess in Section 5.2.2 if this problem is affecting the SACR retrieval, if this may affect the L2M_I retrieval in cloudy sky conditions and how this drawback could be prevented.

DONE, see also answer to reviewer 1. We redid all the cloudy tests and comparisons. One of the reason for the discrepancies was that the SACR code started from the L2M_CIC estimates for the cloud parameters. On the other hand the L2M_I includes a cloud preprocessor that tries to get better estimates for the cloud parameters. This is important especially when the cloud is a cirrus, as explained in the text.

Now the SACR starts from the output of the L2M_I preprocessor, and the values are definitely closer. We also used a revised version of the L2M_I processor. The previous version was assuming during the GN iterations (but not in the final one), that the cloud top and bottom coincide with the boundary of a layer, moving them to the closest values. Since the SGM simulate clouds on full layers, it was a sort of aid to the L2M_I. The new version adds temporary layers to the radiative transfer to cope for top and bottom that do not coincide with a level, so also the results of the L2M_I are somewhat different.

There is still some difference between SACR and L2M_I, mainly for case 6.2 (very thick cloud), but, as explained in the text, the sensibility to cloud parameters in these case is low.

The code used for the FE2ES project was only able to assess the sensitivity to the cloud parameters with a fixed, if perturbed, atmosphere. The code is still under development, the new version allows the simultaneous retrieval of atmospheric and cloud parameters which is of course the final aim.

1.2.2 Minor comments:

LINE 92: when presenting FEI, it would be better to explain in detail its role in the detection of the FOV inhomogeneities, that’s not quite clear in this Section DONE

LINE 100: is there any references confirming the 96.7% of the OLR being included in the range from 100 to 1600 cm-1?

ANSWER: we changed the percentage we stated to "95% or more" and added a reference to (ESA, 2019, p. 43) we can also note that it is possible to calculate the flux over the FORUM spectral band, spectrally integrating the Planck function (Widgerand Woodall (1976)) assuming Earth's equivalent blackbody temperature $T_E = 255K$ (e.g. Wallace and Hobbs (2006), Liou(2002)) for an isotropic radiation $F = \pi I_\lambda$). The ratio of the flux over the FORUM spectral band and the total flux from Stefan-Boltzmann Law is about 97%.

LINE 179: Looking at the AER database, the gaseous species having absorption lines in the spectral range of FORUM are much more than 12. Which trace gases have been excluded from simulations? Did the authors perform any test to check if their spectral signatures on TOA radiances are negligible?

ANSWER: Yes, we have performed some tests to evaluate each gas impact on the TOA radiances. In the figure below we report the differences between the TOA radiances, processed using OSS-FSI, for one atmospheric case in which all gases are accounted for and all gases but without one specific gas. The gases and thermodynamics vertical profiles used are those from case 2.1, as well as the surface temperature. The surface emissivity is set equal to 1 in the entire FORUM spectral range for this simulation. The figure also report the goal and threshold of the FSI NESR (green and magenta line, respectively). The AER gases are ordinated following the "lines" of the figure, i.e., the first gas is H_2O, the second one isCO_2, and the last one is N_2. We observe that after the gas number 12 (HNO_3), all the other radiance differences are lower than FSI NESR. Besides water vapor, the other 11 gases considered are: CO_2, O_3, N_2O, CO, CH_4, O_2, NO, SO_2, NO_2, NH_3, and HNO_3. This information was added in the text:

"The atmospheric profiles of temperature, pressure, water vapor, and other 11 gases mixing ratios, interpolated at fixed altitude levels, are derived from the ERA 5 reanalysis and the IG2 climatological data. The gases included in the simulation are H_2O, CO_2, O_3, N_2O, CO, CH_4, O_2, NO, SO_2, NO_2, NH_3, and HNO_3. Cross section data using predefined values for the CCl_4, CFC_{11} and CFC_{12} molecules were also added."

LINE 193: Simulations in cloudy sky are generally more time consuming, so I guess that this choice was made to preserve computational efficiency. Did the authors test which is the effect on the high resolution cloudy spectra of a degradation of the spectral resolution from 10-4 to 10-2 cm-1?

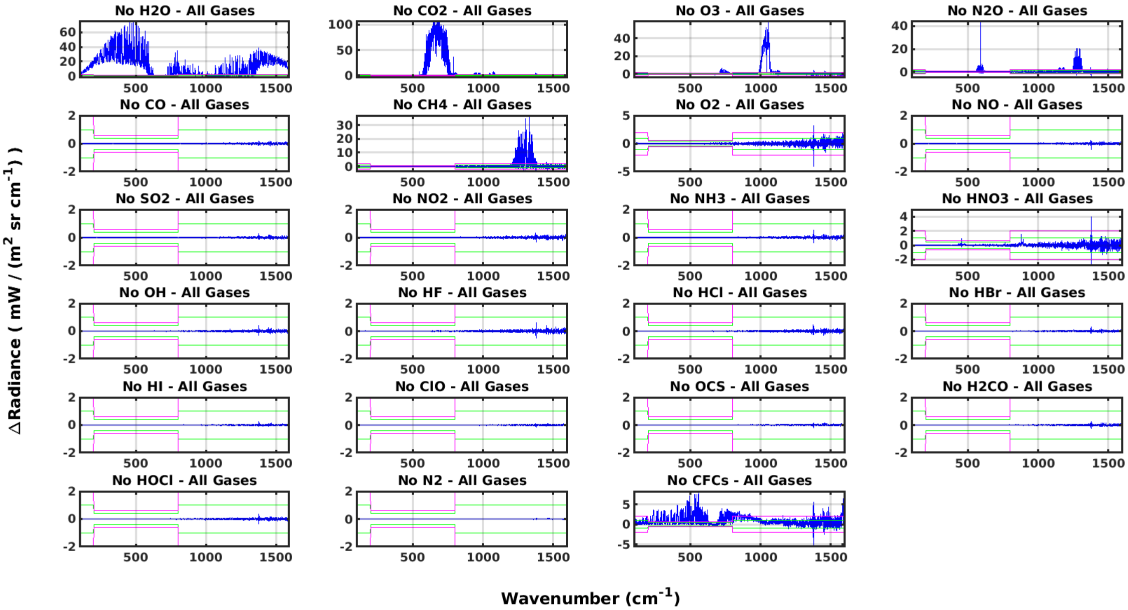
ANSWER: The default spectral resolution for clear sky conditions is 10-3 cm-1. The typo is corrected in the paper.

Tests have been performed to evaluate the effect of using a lower resolution (10-2 cm-1) in case of cloudy conditions instead of the higher resolution adopted in clear sky (10-3 cm-1).

Results are reported here below for your convenience as difference between FSI-OSS radiances. The tests refer to the conditions of cases 2 and 4 and show that the adoption of a lower resolution implies differences with respect to the adoption of the higher resolution that are within the FORUM FSI goal measurement error (NESR)

The computational times for the two cases presented are:

The text is modified as follows:



Differences between the TOA radiances, processed using OSS-FSI, for one atmospheric case in which all gases are accounted for and all gases but without one specific gas.

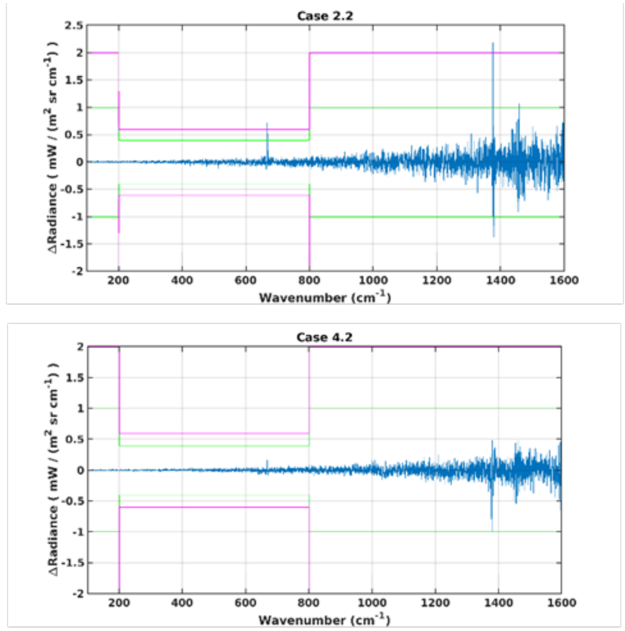
	resolution=10^{-2} cm^{-1}	**resolution**10^{-3} cm^{-1}
case 2.2	12 min	5 hours
case 4.2	2 hours	26 hours

These radiative transfer routines compute the spectral radiances for the FORUM imager and sounder. The default output spectral resolution is set to 10^{-2} cm^{-1} and 10^{-3} cm^{-1} for cloudy and clear sky simulations respectively. The different resolution is adopted for the two cases as a compromise between computational speed and accuracy. In fact, for a resolution of 10^{-2} cm^{-1} the computational times in cloudy conditions decrease by a factor ranging from about 10 to about 20 with respect to using a 10^{-3} cm^{-1} resolution. Nevertheless, the simulation accuracy is maintained since OSS-FSI radiances in cloudy conditions computed using 10^{-2} or 10^{-3} cm^{-1} differ less than FSI goal NESR (Figure not shown).

LINE 200: can the authors specify from which database the 11 pre-defined surface types were taken?

DONE: The surface emissivity database is derived from Huang, 2016. The text is modified as follows:

The SGM allows the user to define surface temperature and emissivity, or to select specific emissivities from a list of 11 pre-defined surface types derived from Huang (2016), e.g., water, desert, snow, deciduous vegetation.



Differences between FSI-OSS radiances, computed from a lower (10^{-2} cm^{-1}) and a higher (10^{-3} cm^{-1}) spectral resolution, for the cases 2.2 (top panel) and 4.2 (bottom panel).

LINES 242-246: This is a little bit redundant, maybe it would be enough to explain that the safety margin on the FOV was considered to model the self-apodization effect

ANSWER: Actually, the integration of the different lines of sight across the FOV accounts for the self-apodization, while the safety margin is used in case a LOS offset is configured, to model pointing errors. We clarified this in the text. Also, we would like to maintain the explanation because we believe not every reader is an expert in Fourier spectroscopy.

LINE 282: can the authors explain why is the resampling of the average spectrum required for the L2M_I? Which are the spectral resolutions of Level 1b and 1c respectively?

DONE: The level 1b resolution is internal and is used to satisfy some technical aspects such as changing of pointing during the scan. Thus the level 1b spectrum needs to be resampled, and the best choice from the level 2 point of view is that adjacent measurements are statistically independent. We added some explanation in the paper.

LINE 339: the meaning of FWHM vertical resolution and how it should be calculated is not quite clear here

ANSWER: The FWHM of each Averaging Kernel is more properly called FWHH (Full Width at Half Width) according to Rodgers. It is calculated assuming that the AK can be approximated with a triangle, so that the FWHH can be computed as (Area subtended by the AK curve) / AK(i,i), where AK(i,i) represents the altitude of the triangle. We modified the definition and inserted the reference.

LINE 342: In my opinion, the first sentence of Section 4 is redundant, it could be removed

DONE

LINE 355: I suggest to re-write sentence as “we depart from real data using this approximation, however it does not have an impact on the assessment of the FE2ES performances"

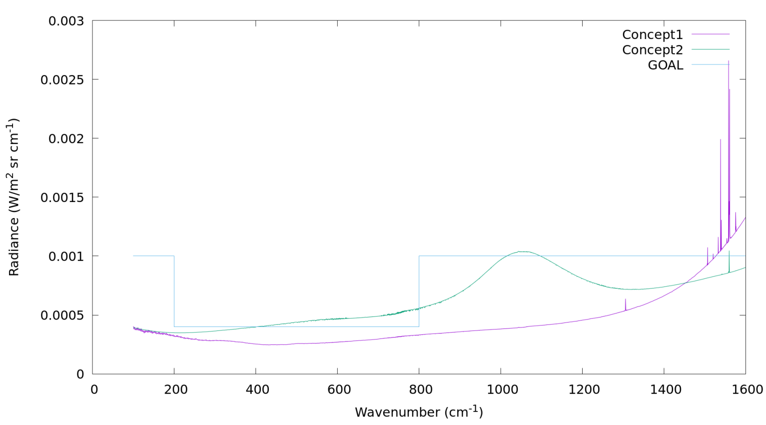
DONE: See also comment from author 1. Text modified accordingly to both comments.

LINE 376: The vertical retrieval grid is considerably different from that used by Ridolfi et al. (2020) for FORUM. Can the authors justify the choice of this grid? Have any tests been performed to get an optimized grid?

DONE: Yes, we started with an "optimized" grid for the retrieval. However, since the profiles were generated on a finer grid, the retrieval was not able to reproduce the finer features, so that the smoothing error introduced was considerably larger than the bias due to the OE term. Added a sentence in the paper.

LINE 378: The authors could consider whether compare more in detail the information provided by the industrial consortia to the goal parameters (e.g. insert figure for comparison between G and industrial consortia noise, to show that this last is compliant to the goal)

ANSWER: We have a figure, we show it here:



FORUM requirements and instrument concepts noise

However, maybe for an excess of prudence, we prefer not to show it in the paper to avoid disclosure of industrial data. Of course, the noise shown in all the figures is derived from instrument concept 2, the one we used for all the tests.

LINE 402-403: Are the spectroscopic database and continuum model in KLIMA the same of those used by the L2M_I module? If no, would the authors expect an even better consistency between the retrieved products from the two models if using the same spectroscopic data?

DONE: Yes, the spectroscopic databases and the continuum model are exactly the same. We reported this fact in the paper.

LINE 417: Can the authors quantify the term “negligible” pointing out a threshold for the difference between retrieval errors provided by KLIMA and L2M_I?

DONE: We reported the two explicit thresholds for the retrieval profiles and the retrieval error.

LINE 445: How is the sensitivity of the retrieval linked to the retrieval error? A reference to the information gain quantifier (Dinelli et al., 2009) would be useful to clarify this point

DONE: Rewritten the sentence in terms of the IGQ.

Figure 5 (bottom right panel): How can the authors explain the positive bias in the difference between retrieved and true water vapour above 100 hPa?

DONE: There is a low sensitivity (as shown by retrieval error and a-priori error being similar), so that the retrieval converges to the a-priori which is the true value plus 1 sd.

LINE 470: How many emissivity values have been retrieved for the full spectral range and the FIR region only? The authors could consider whether to include (e.g. with a Table) a comparison between the number of retrieved values for the emissivity and the obtained DOFs. It would be useful to clarify the information content on surface emissivity in the different analysed scenario.

ANSWER: The number of DOF for FIR and MIR was already present in the paper. We grouped this information in a table as suggested. As written in the paper, the numbers are directly comparable because the emissivity grid used in the retrieval is fixed.

LINE 503-506: It’s not quite clear from the text why the unretrieved atmospheric parameters are perturbed in this test

DONE, according also to comment of referee 1. The aim of the FE2ES was to show that, even with a perturbed atmosphere, the cloud parameters could indeed be retrieved. We clarified this in the text.

LINE 541: What’s the reduction factor on computational time obtained with the application of the single layer approximation?

DONE in the timing section. Quite stunning: the computation time with the single layer approximation is about 1/10 of that without the single layer approximation.

LINE 513: This sentence is contradictory with results shown in Table 9. For example, differences in the retrieved cloud top are even higher than 5 km. As just mentioned in the major comments, the authors should explain at this point why it may occur

DONE, see answer to major comment.

Figure 13: The figure should include also some case tests for which the cloud is under the limit of detectability (e.g.: OD from 0.01 to 0.2 for CASE 1.2, from 0.01 to 0.15 for CASE 2.2)

DONE

LINE 621: It is redundant to specify here the physical properties of the studied cloudy scenario, citing Table 5 would be enough

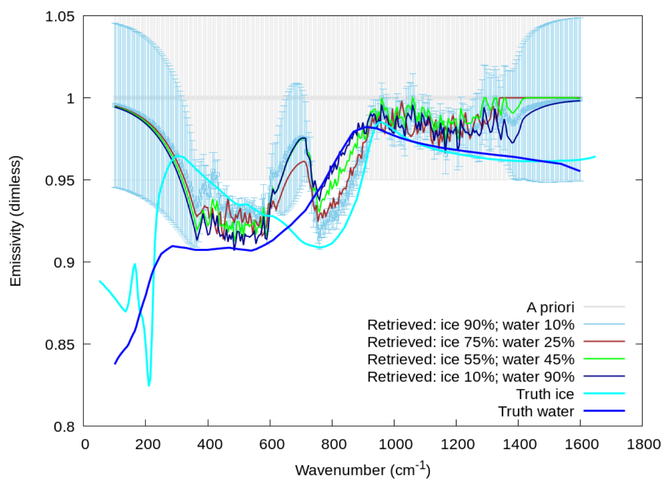
DONE

Figure 17: It's quite tricky to observe anti-correlation between retrieved emissivity and surface temperature from this graph, as left panel shows [retrieved-true] values and right panel shows [true-retrieved] values

DONE: We have changed the figure reporting the Retrieved - True profile also for the surface temperature. Check also the explicit plot in the right panel of the new figure 13.

LINE 654: How can the authors explain this counterintuitive result on surface emissivity? I would have predicted a retrieved emissivity moving from water toward snow type with a larger coverage of the latter, as it is for surface temperature

Actually the reported figure was relative to the test with a mix of ice and water surface, while the emissivity of snow was wrongly reported. Thanks for spotting this. The correct figure is reported below.



In the paper we have reported the figure relative to a mix of fine snow and water since their emissivities show the largest differences.

LINE 729: The authors should confirm in the conclusions that the CIC algorithm – in most cases - tends to prevent a clear sky classification even with low percentage of cloud contamination in the FSI FOV

DONE: The information was included in the text: Even small contamination of a cloud in the FoV induces errors in the retrieved atmospheric and surface quantities. The error increases with the increase of the percentage of the FoV affected by the contamination and of the optical depth of the cloud. However, it is important to underline that, by exploiting the information content of the FIR channels, the CIC algorithm is highly sensitive to the presence of a cloud in the FSI FoV, even with a low percentage of cloud contamination. Thus, in most cases, the cloud identification is performed and the performances of the L2M_I are preserved.

1.2.3 Technical revisions

LINE 40: Delete “respectively”. In the following sentence it’s better to shift “thus” at the beginning

LINE 41 and LINE 101: “interannual” instead of “longer term”

LINE 52: Please, specify the acronym of GCMs Section 4: Here is FEES acronym, I guess it should be FE2ES or E2ES

Table 4: The right nomenclature is “cumulonimbus”, not “cumulus nimbus”

LINE 401: I suggest “taking into account” instead of “takes”

LINE 405: It would be better to specify the AER acronym at line 402, where AER is cited for the first time

LINE 431: “Each column of the table” instead of “each row”

LINE 445: “sensitivity” instead of “sensibility”

LINE 456 “precision” instead of “performance”

LINE 489: “top of the cloud” instead of “altitude of the top of the cloud”

LINE 498: “and” is repeated twice

LINE 515: the parenthesis is redundant, it could be removed

ALL DONE, thanks for pointing these out.