The paper “On the use of IASI spectrally resolved radiances to test the EC-Earth climate model (v3.3.3) in clear-sky conditions” by Della Fera et al. explores the use of top-of-Atmosphere out going spectral radiance measurements to evaluate biases in climate models and linked them to key variables of the Earth system. The method is applied to the Earth-system model EC-Earth discussing biases in the temperature and water vapour content. The paper is generally well organised and written and I propose it for publication after the following points have been addressed by the authors.

Major comments:

The selection of the clear-sky areas to analyse goes through a series of steps and eventually a set of clear sky spectra are obtained from the dataset of simulated radiances. Would it be possible to show a map with the regional density of pixels surviving the selection, to help identify which regions are most contributing to the analysis? For example, are the pixels used to create figure 3 distributed homogeneously or with a higher density somewhere? Or the analysis is robust enough to any regional bias?

Jacobians in Figure 2 are shown as normalized to the relative maximum. I understand that this is to show qualitatively the spectral sensitivity of the radiance to various parameters, but would it make sense to also show the absolute values somehow (for example in Figure 6)? It would be interesting for example to estimate if the biases in temperature and water vapour concentration shown in figure 8 would indeed reasonably map into the observed radiance/brightness temperature biases discussed for figure 6 and figure 5c.

Figure 5 panel B shows that EC-Earth has a positive temperature bias in the region 5-15km in the 30S-30N area as seen in the 700-710 cm-1 spectral region. Yet in the comparison against ERA5 that area shows generally a negative bias up to 15km (positive biases are seen north and south of 30 degrees, above the tropopause). This seems in contradiction to what it is said at line 360 Ch. 3.4

In the conclusion I’d like to see a clearer discussion of the limitations of the presented method, namely the difficulty of making sure that a meaningful comparison between simulated and measured radiances can be obtained (this is briefly touched in chapter 3.4 regarding sampling biases). The analysis presented in this work does indeed show the advantage of having spectrally resolved OLR computed from the model fields overcoming some of the limitations of retrieval products, but on the other hand, assumptions and parametrisations are still needed to perform accurate radiative transfer computations. Restricting the analysis to clear-sky ocean areas circumvents the difficulties of a correct simulation of surface radiative properties or the representation of cloud microphysical characteristics whose uncertainties might create larger differences than the model biases one attempts to investigate. Can these shortcomings be somehow mitigated or is the method really only suited for a limited set of cases?

Minor comments:

Sect. 1, line 39: “and, thus, to a specific parameter”: what is it meant here by parameter? An active model variable such as temperature or concentration of a particular atmospheric constituent or rather a model parameter governing a particular process/parametrisation in the model? I think that it is difficult that a single model parameter, if the latter is intended, would be able to explain a bias in a particular spectral region. Indeed, many of the parameters tuned in EC-Earth to achieve satisfactory radiative fluxes at TOA, affect cloud parametrisations. Please, clarify.

Sect. 1, line 43: “1970s, but only starting from the 2000s long term….”

Sect. 1, line 56: “In anticipation of FORUM measurements….”. This is repeated few time in the document: does it mean here something along the lines of “To demonstrate the potential of the future FORUM measurements”?

Sect. 1, line 65: “MIR” has not been defined while FIR yes.

Sect 1, line 79: “paves the way for direct assimilation” Not clear here: why the implementation of an online simulator would help in that direction?

Sect. 2.1.1, line 101: CLOUDSAT is an active radar. ISCCP is not a sensor, if it means the International Satellite Cloud Climatology Project. References and full description of the acronym would help here, as done for RTTOV mentioned later, for example.

Sec 2.2.1, line 137: is there a reference for the AVHRR cloud mask?

Sec 2.2.2: I think there should be references to the CERES dataset used

Sec 3.1, line 197: “partial derivatives of radiance with respect to the most relevant atmospheric parameter”. Perhaps a more general definition like “with respect to any relevant atmospheric variable. Here we show the most relevant…”

Sec 3.3, line 298: The example discussed in this section shows that spectral OLR are indeed useful to provide a complete view of model biases, but it also shows the difficulty to select meaningful area to compare with measurement, thus limiting somehow the possibility to map directly a bias to a parameter in a global model. Perhaps the sentence can be rephrased to reflect this limitation.

Sec 3.3, line 330: “witnessed by the negative sign of the water Jacobian” -> “as represented by the negative values of the water vapour Jacobian”

Figure 8: The representation of the differences in specific humidity as percentage tends to highlight the stratosphere due to the very low water vapour concentration there. Absolute differences could perhaps be better related to the absolute value of the Jacobian that could be shown e.g. in Figure 6?

Sec 4, line 392: The results are only shown for ocean areas, land biases are not really discussed in depth in the text. Is it worth it reporting it in the conclusions?