Review of « A test of the ability of current bulk optical models to represent the radiative properties of cirrus cloud across the mid- and far-infrared », by Richard J. Bantges et al.

General comments

This paper investigates the capability of a cloud single scattering properties (SSPs) database (taken from Baum et al., 2014) to explain downlooking airborne radiance observations above cirri throughout the mid- and far-infrared taken with the ARIES and TAFTS instruments during the CIRCCREX experiment in March 2015. Short flight periods are selected for the quality of the radiation data, and are completed by extinction profiles below the aircraft provided by a lidar. Atmospheric profiles are taken from dropsondes, and completed with ERA-Interim information. Radiative transfer simulations are performed with the LBLDIS model in an attempt to perform a radiative closure with observations. Clouds are assumed to have vertically homogeneous ice habit and effective radius, and several single scattering properties from SSPs database are used to represent cirri. It is found that none of the database used allows to match the observations in both the MIR and FIR within instrument uncertainties. It also shows that retrieving the best properties based only on MIR observations results in large residuals once applied to the FIR, and vice versa. Interestingly, any optimal set of cloud properties is characterized by an error compensation when it comes to broadband fluxes, with the MIR overestimated by the model and the FIR underestimated. The study demonstrates with unique observations the limitations of the currently used SSPs database in the FIR and point to the need for further investigation to achieve consistency between spectral regions.

The paper perfectly fits in "Atmospheric Chemistry and Physics" as it reports new findings relevant to the atmosphere energy budget and remote sensing based on a combination of novel observations and state-of-the-art radiative transfer modeling. The paper is well-written, relatively easy to follow except for a few points detailed further. While there is no doubt that the conclusions of the paper are worth being shared among the atmospheric community, I believe the paper could be significantly strengthened by extending the investigation to other SSPs models and by providing more quantitative and physical insight into the reasons for the failure of the considered database. Some leads for improvement are quickly proposed and ruled out as quickly, so that the reader, including experts in scattering, are let with few hints to how the next generation databases could perform better than the current ones. Such modifications are probably minor in terms of amount of work, but major in terms of clarification. They would make the paper more convincing and valuable to a wider audience.

Specific comments

1) A major issue of the paper, which might however be ruled out in a few sentences or with complementary simulations, is the treatment of the atmosphere above the aircraft. Indeed the paper clearly details how the atmosphere is prescribed below the aircraft, but nothing is said about the presence of an atmosphere above the aircraft, suggesting that deep space is considered starting at 9.3 km. At the same time the authors point in the introduction the fact that scattering is important in the FIR (l.46), much larger than in the MIR. This means that any downward flux coming from above the aircraft will be partly reflected in the FIR, hence will contribute to the observed upward radiance. The presence of a cloud above the aircraft or the tiny residual of water vapor at this altitude would certainly be visible. As a consequence the absence of any cloud should be verified and stated as long as possible, and a water vapor profile used for the whole atmosphere (for instance taking a co-located ERA-I profile). A sensitivity study could be performed to ensure that what happens above the aircraft cannot be the reason for the residuals in the FIR. Note that the additional scattering from the cloud would tend to enhance the simulated FIR radiance, which is currently underestimated.

2) The objective of the paper is to demonstrate that current SSPs databases don't work throughout the MIR and FIR spectral ranges. However to demonstrate this only one database is used, that of Baum et al. (2014). Why weren't more extensive databases used, in particular those of Yang et al. (2013) including a larger variety of habits and the effect of roughness, which is mentioned in the introduction (l.30) but not further. Also, could the database of Baran et al. (2014) be used as well? Consider also that of van Diedenhoven and Cairns (2020) to be exhaustive. If none of those databases (which probably cover all the available databases) manage to reconcile the MIR and the FIR, then the conclusion of the paper would be much stronger. At least, it should be specified to which extent the presently used database is representative of all those available in the literature.

3) The authors mention an exhaustive set of optical probes, many of them providing detailed information about the ice crystals habits and size distributions. Although it is clear that taking these information as a raw input to the simulations wouldn't work, at least because of the elapsed time between the radiative and microphysical observations, these rich observations are not mentioned at all. Maybe the complexity of the habits, the singularity of the PSDs would point to possible deficiencies of the SSPs databases. Also this could provide useful information regarding the vertical structure of the clouds, which is currently too quickly ruled out as a potential explanation for the inadequacy observed and would deserve more investigation and a dedicated sensitivity study.

4) More generally, the paper would greatly benefit from physical insight about the limitations of the databases. In which direction should experts work ? What's the next step ? Could you inform whether the temperature dependence of the refractive index may solve something. To do so it should depend on the spectral range, does it ? What about surface roughness etc. ? Such discussion could of course be very exploratory but would have the merit to provide meaningful leads for improvement.

5) What are the practical consequences of the paper for energy budgets or for ice cloud retrievals? How is cirrus radiative effect erroneous in climate simulations, how does it matter? How comes that radiative closures have been satisfying in the MIR if adding FIR channels would have resulted in different parameters? Do FIR channels provide significantly different retrievals, or do they narrow the range of possible values (hence uncertainties)?

6) Here is a suggestion to illustrate the differences in FIR and MIR retrievals for one selected case. On a 2D LUT with r_{eff} and τ as axes (assuming a fixed habit), highlight the regions corresponding to MIR and FIR matching (for the different methods). This would help understand the minimisation procedure and indicate in which direction FIR channels tend to drive the retrievals (for instance).

<u>Technical corrections</u>

l.15 : single-scattering is probably more detailed than "optical" so should not be in parentheses

l.18 : state whether those fluxes are broadband or spectral

l.19 : "strong" is not quantitative, is it $\pm$ 2 or $\pm$ 10 W m^{-2} ? Not clear how there can be a compensation between something that is within the residuals and something that is not.

l.22 : "cloud properties" is not defined, and the link to retrieval is not that straightforward.

l.23 : "guidance" is probably not sufficient currently for the practical development of new databases

l.26 : an additional sentence to present the SW (thin so often limited albedo) and LW (cold so large greenhouse effect) effects of cirrus clouds may be useful before talking about net effect.

l.27 : "geographical position" is not very clear. How does it impact the radiative effect? Do you mean temperature contrast with the local surface and atmosphere ? This last point should not overlap with the first two characteristics pointed out. Also, given the subsequent definition of the microphysical properties, I feel like optical thickness or particle number concentration is lacking here, unless it is included in the PSD (at its zeroth moment)

l.35 : I tend to write *in situ* as it is a Latin phrase. Holds elsewhere

l.38 : do the authors mean that all ice clouds are cirrus clouds or that they focus on cirri only ? Ice clouds could be tackled more broadly.

l.42 : no lower wavenumber limit given for the FIR ? Can be misleading

l.45 : the formulation "sensitivie to radiation" is unclear. Do you mean that the optical properties are highly variable across the FIR ? That the broadband properties are sensitive to what happens in the FIR ?

l.49 : maybe state that this holds for narrowband channels, not necessarily for hyperspectral observations

l.51 : "spectrally-resolved" has not been properly defined. Maybe give a hint to which spectral resolution this refers, because depending whether the reader is a climate modeler or a spectroscopist the expectations might differ.

l.94 : could you detail if relevant what those probes measure : PSD, scattering properties, habit ? Are all these instruments used in the paper? Are they to some extent redundant? Only relevant data should be presented.

l.96 : how are cloud phase and total amount of ice measured ?

l.98 : "information" is vague, do you mean geometric thickness here, as extinction follows ?

l.100 : here more details on the assumptions to convert backscatter profiles into extinction profiles are needed because (as discussed later on) this is key for the consistency of the synergistic radiative closure.

l.108: knowledge of the atmospheric profile above the aircraft is key as well because of scattering (including backscattering from the clouds). In particular, the absence of clouds above the aircraft is critical.

l.117: the acquisition time of a TAFTS spectrum is lacking to understand why and how 3 sets of radiance can be taken in 1 min 12 s. Please also clarify the ARIES acquisition time.

l.120: what does this "two second period" refer to? It is not clear

l.121: what's the reason for converting radiance spectra into brightness temperature (BT)? Is it practical when it comes to including instrumental error, which is more uniform in radiance than in BT?

l.125: what "variations in the cirrus properties" do you refer to? Do you simply mean the presence of cirrus ?

l.127: why is that "useful"? Is this used further in the study? Is it original, unexpected, instructive?

l.131: "a frequency dependent sensitivity to cirrus properties" is unclear. Sensitivity of what?

l.145: are these really two radiative codes, or does LBLDIS merge the LBLRTM model for gas optical thickness and DISORT for the radiative transfer equation solver?

l.148: this should be more explicit that most parameterizations try to express the single scattering properties in terms of the effective radius. Note also that it differs from the approach of Baran et al. (2014) who use temperature and ice water content to estimate single scattering properties.

l.152: this match is surprisingly low

l.153: does this emissivity model spectrally extend into the FIR?

l.154: again, no information about what the atmosphere above the aircraft looks like, although this may be critical

l.169: how many streams were used?

l.171: why only focusing on these 3 databases while Yang et al. (2013) proposes much more? In particular the effect of roughness could be investigated as a solution to overcome the current deficiencies.

l.197: separated into → composed of, split into, discretised into ?

l.218: The Baum model was already mentioned

l.227: how wide are these spectral regions?

l.231: is this "τ" referring to a single cloud layer, or to the whole cloud? Is the profile still scaled on the lidar profile? Also precise whether habit and r_{eff} are assumed vertically homogeneous.

l.241: could you explain what is the physical meaning of weighting by the error? What differences do you expect in comparison with the second approach? Why duplicating similar approaches?

l.242: are the Rs in the formula spectra? In which case how is the absolute difference defined? Unless one wavenumber region is actually a single channel? This should be clarified

l.244: is the minimisation performed through interpolation (or selection) of the LUT, or using a dedicated algorithm?

l.257: can such an inconsistency really explain 45% differences?

l.258: Details are needed to clarify the lidar estimate of extinction

l.260: why using two methods which are so close, in particular if the consistency is not surprising (l. 274)?

l.282: do these 14 simulations refer to method 4? Otherwise it reads like they are those among the 739 that also match FIR observations, which is obviously not the case reading the following

sentence. More explicitly, matching has not been properly defined. It is generally completed by "within uncertainties", which is clear for MIR, but not for FIR. Maybe the difference between match in the selected channels and across the whole spectrum should also be more clearly explained.

l.289-290: this suggests that no combination works in the FIR? So I guess the 14 spectra mentioned previously were matching only for the selected channels.

l.291: this is not clear why retrieval would be more constrained. If no set of parameters works, then what to conclude? Something that is sure is that using different spectral regions for the retrieval gives different results, which is of course worth pointing. But speaking of retrieval quality sounds hazardous so far.

l.295: could there be a spectral signature of the angular signal? Maybe look at spectra at 3 different viewing angles to ensure that this approximation is acceptable.

l.297: if this compensation occurs within the uncertainty range of observations, can it be considered significant? Physically, does it mean that among the possible parameters after MIR matching, FIR selects the largest/smallest r_{eff} or optical thickness? This compensation should be further discussed because this provides physical insight about what individual spectral ranges would try to converge to (specific comment 6).

l.301: there is no more mention of the minimisation methods. Which method results in 2 W m^{-2} errors?

l.301: I think that at this stage the main conclusion should be that none of the optical models investigated allows to match observations, which points to the need for new models. It is a result and should be mentioned before the next paragraph.

l.305: state-of-the-art for sure, but encompassing all those available in the literature?

l.308: "tested here" suggests that other models could work, so makes the conclusions weaker

l.313: how can you be sure that this tighter constraint result in a better retrieval? I think a retrieval quality should be regarded through the uncertainty associated with this retrieval, not only through the absolute error of the optimal parameters. In that sense, how does adding FIR observations help reducing the retrieval possibilities is informative.

l.314: the habit was not much discussed for the retrieval. If it is forced, are the optimal r_{eff} and optical depths significantly different?

l.315: energy analysis is most meaningful at global scale. Could you provide hints to the expected global error given the distribution of cirrus (occurrence and optical depth). If 2 W m^{-2} is specific to the case studied here, it could have limited implications in a climate framework.

l.320: Could you, based on your simulations, provide a more quantitative (adding a figure for instance) discussion of this potential impact on the heating rates? This would bring the attention of the climate modelers. Maybe comparing the heating/cooling rates profiles of the 4 methods for the same case.

l.324. If temperature dependence is a potential venue, could you briefly explain why this may help reconcile MIR and FIR. For this, some different sensitivities should exist in this temperature

dependence between the FIR and MIR. Is that the case? The personal communication could be expanded.

l.326: this is indeed an important point, but not sufficiently detailed. How was this vertical heterogeneity investigated? What vertical gradients were used? How could cloud probes provide quantitative information about this vertical layering? So far, the short explanation lacks details to rule out the possibility that vertical layering associated with distinct penetration depth into the cloud of the MIR and FIR (because of scattering) could be a reason for the observed mismatch. Especially when looking at the sensitivity displayed in Figure 6a.

l.331: does this mean that "new" parameterizations were built as in Baum et al. (2014) based on these new PSDs? Alike the other leads investigated, this should be quantified more properly, in terms of error bars associated with this kind of assumption of the PSDs. Other theoretical PSDs (different shapes, different widths) could also be investigated.

l.339: how do you solve the issue of concomitant cloud microphysics observation in the spaceborne configuration? Accounting for the mismatch of spatial scales.

l.340: how long is the journey to the ultimate information? Again for the modelers, the paper would benefit from providing concrete leads towards improvement. Said differently, how should a climate modeler take these results?

Table 3: the retrieved habit for the method 3 differs from all the others. Would there be an explanation why including FIR observations tends to constrain the habit to GHM?

Table 4: none of the broadband fluxes differences reaches 2 W m^{-2}, which seems contradictory with the statement in the text (l.301).

References

Baran, A. J., Hill, P., Furtado, K., Field, P., & Manners, J. (2014). A coupled cloud physics–radiation parameterization of the bulk optical properties of cirrus and its impact on the Met Office Unified Model Global Atmosphere 5.0 configuration. *Journal of Climate*, *27*(20), 7725-7752.

Baum, B. A., Yang, P., Heymsfield, A. J., Bansemer, A., Cole, B. H., Merrelli, A., ... & Wang, C. (2014). Ice cloud single-scattering property models with the full phase matrix at wavelengths from 0.2 to 100 µm. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *146*, 123-139.

van Diedenhoven, B., & Cairns, B. (2020). A flexible parameterization for shortwave and longwave optical properties of ice crystals and derived bulk optical properties for climate models. *Journal of the Atmospheric Sciences*, *77*(4), 1245-1260.

Yang, P., Bi, L., Baum, B. A., Liou, K. N., Kattawar, G. W., Mishchenko, M. I., & Cole, B. (2013). Spectrally consistent scattering, absorption, and polarization properties of atmospheric ice crystals at wavelengths from 0.2 to 100 µ m. *Journal of the Atmospheric Sciences*, *70*(1), 330-347.