Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2020-297-RC2, 2021

Comment on gmd-2020-297

Robin Hogan (Referee)

Referee comment on "Towards an improved treatment of cloud–radiation interaction in weather and climate models: exploring the potential of the Tripleclouds method for various cloud types using libRadtran 2.0.4" by Nina Črnivec and Bernhard Mayer, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-297-RC2, 2021

This paper presents an interesting evaluation of the Tripleclouds (TC) method for representing cloud structure in a radiation scheme using benchmark Monte Carlo calculations and three contrasting, realistic 3D cloud scenes. While the McICA method is a more common way to treat cloud structure in operational models, Tripleclouds is attractive because it is free from stochastic noise. The paper goes into some depth in testing and optimizing the various ways in which Tripleclouds can be configured to reduce the errors. However, I have some major concerns that should be addressed before this paper can be published. The paper is generally well written otherwise.

MAJOR COMMENTS

1. The use of maximum-random overlap is well behind the state-of-the-art and will have led to appreciable biases in your GCM and TC schemes. The cirrus case is the most obvious: Hogan and Kew (2005, Table 2) found that for this exact case, a GCM-type radiation scheme with maximum-random overlap underestimated the TOA cloud radiative forcing by 42% in the longwave and 48% in the shortwave, compared to the same GCM-type scheme but with the "true" overlap. I expect significant overlap biases to be present in the other two scenes, and indeed I can only explain your TC flux biases with respect to ICA as being due to overlap being too maximal. To test this, simply compute the total cloud cover predicted from the cloud-fraction profile and your maximum-random overlap assumption, and compare it to the actual total cloud cover of the scene. Note that even the overcast stratocumulus case can have an overlap bias because of the overlapping of sub-grid cloud structures, which ought to be represented by non-maximum overlap of the high-LWC and low-LWC cloudy regions (some schemes use a vertical decorrelation scale for cloud heterogeneities of half that used for cloud boundaries, i.e. around 1 km). I don't believe it would be too complicated to incorporate realistic overlap: this can be done using overlap matrices introduced by Shonk et al. (2010), and an example implementation is in the open-source ecRad package - see the radiation_overlap.F90 source file on GitHub. The capability has been in older TC implementations for longer - for example, Shonk and Hogan (2010) evaluated the impact of horizontal heterogeneity and vertical overlap using

an implementation of Tripleclouds incorporating exponential-random overlap in the Met Office radiation scheme, and stressed the importance of making both horizontal heterogeneity and vertical overlap as realistic as possible.

2. When comparing TC results to the benchmark, a number of errors are compounded and there is insufficient effort to separate them out for the reader: (1) the overlap parameterization is biased (see above); (2) the split percentile of 50% cannot adequately represent large FSDs; (3) a value of FSD=0.75 is used at various points which is different (often much lower) than the "truth"; and (4) TC is benchmarked against at 3D model when it makes no attempt to represent 3D effects. This leads to tuning of one parameterization to fix a problem caused by another (see comments 18 and 19). Wouldn't it be more satisfactory to evaluate TC against ICA when all the inputs are correct, in order to identify the intrinsic error in TC, then the impact of not knowing the exact overlap or FSD, or not representing 3D effects, can be quantified separately? Indeed, this is the approach taken by Hogan et al. (2019): their Fig. 7a-c shows that when the correct overlap and FSD is used, the shortwave bias against ICA over 65 scenes is less than 3 W m-2. This was then a firm foundation for them to look at the representation of 3D effects, whereas in your case it would be a firm foundation to look primarily at errors due to parameterizing FSD (although you also have some interesting 3D results).

3. The proposed solution to the high-FSD problem in section 4 should be compared to the solution to the same problem presented by Hogan et al. (2019, appendix). Their solution is derived from theoretical distributions, rather than by trying to minimize an error against a benchmark calculation in which several other errors (notably in overlap) are also present. The structure of the present manuscript is a little frustrating - section 3 contains some puzzling errors that are only addressed, or even properly mentioned, when the reader gets to section 4. Why not flag up the need to address the problem sooner in the paper?

SPECIFIC COMMENTS

4. Abstract line 6: while the optically thicker part could be used to represent convection, most clouds are not convection and the use of a thicker and thinner parts are simply a first-order approximation to the horizontal distribution of optical depth that is found in stratiform clouds.

5. Figure 1: Caption should stress that this is a schematic; the vertical resolution shown here is coarser than any operational model.

6. Introduction: This seems unnecessarily long. The primary purpose of the paper concerns testing the Tripleclouds scheme for representing horizontal cloud heterogeneity in a 1D radiative transfer context, for which the appropriate benchmark is the Independent Column Approximation. Many of the references and discussion concern 3D radiative effects, which seems not so central to the topic of this paper; a little shortening

would therefore seem in order. The introduction should mention the McICA scheme of Pincus et al. (2003), which is much more commonly used to represent cloud structure than Tripleclouds in current global models. Figure 1, panel 2, could just as easily be used to illustrate the McICA scheme as a cloud-resolving model.

7. Line 120: Define FSD here, saying particularly that it is the standard deviation divided by the mean, in both cases considering only the non-zero water content values in the horizontal LWC distribution.

8. Figure 2: the linear colour scale is not really suitable for the cirrus cloud since it is entirely white for optical depths up to 5, yet a significant fraction of the radiative impact of this cloud in the longwave will be from optical depths less than 5.

9. Line 155: Please say how the distributions were fitted (least-squares or fitting three of the moments of the distribution?) and I'm not sure what use a Gaussian distribution is, except as an excuse to use the 16th percentile, since it is unbounded on the lower end and so negative water contents are predicted. I'm also curious as to whether you can say that either lognormal or gamma are really better; Hogan and Illingworth (JAS 2003, Figs. 4-5) found that there was little to choose between them when comparing to real data, and often the gamma and lognormal were much closer to each other than either were to the noisy distributions of individual scenes.

10. Line 202: \bar{LWC} as defined here is not the layer-mean LWC, but the in-cloud mean LWC, i.e. ignoring the clear region.

11. Line 234: a short further discussion is required - like the bottom row of your Fig. 3, Hogan et al. (2019, Fig. 10a-d) also found much larger FSD values than reported by Shonk et al. (2010). This is not possible to represent if the two cloudy regions are assumed to have the same area, but in the appendix to that paper they showed how an improved representation of large-FSD distributions could be achieved by making the optically thinner region occupy a larger area.

12. Eq. 8: surely to be a bias, x and y should be averaged over more than one event? Otherwise it is an instantaneous error. Also it should be clarified whether x and y represent horizontal averages (e.g. of heating rate)... but are they also vertical averages of heating rate through the cloud layer? "Cloud-layer RMSE" is ambiguous - does the "layer" refer to model layer or the entire layer of cloud?

13. Table 1 and text: I understand that the TC(FSD) method uses FSD=0.75, but it is unclear from either this table or the text whether the TC(LP) method uses the 16th percentile of the *true* in-cloud LWC distribution for each cases, or for an idealized (e.g. lognormal/gamma) distribution with an FSD of 0.75. If the former, then surely the main

difference between the two methods is whether the true distribution is used or not, information that is not stated (at least not clearly). It would seem much more satisfactory in all cases to use the observed FSD in order that we are evaluating the intrinsic TC method, not the rather old and simple parameterization of Shonk et al. (2010).

14. Table 1: "Conventional GCM" is not an appropriate label since many/most operational GCMs now use the McICA method. You could call it "Homogeneous cloud"? For this table to be most easily understood, the acronyms LP and FSD should be defined in the table.

15. Line 310: Say *why* the "GCM" scheme has his bias: by homogenizing the cloud, the probability of radiation being scattered or emitted is greater. This is even more clearly evident in the cirrus case.

16. Line 350 and lower panels of Fig. 6: In a fractional sense both the "GCM" and TC errors are very large, especially in the infrared. However, this is not a fair evaluation of the TC method because the fractional standard deviation is extreme in this case (around 3) whereas you are feeding it with a value of 0.75, or using a split percentile of 50% which cannot capture large FSDs. It would be better to discuss your solution to the problem earler.

17. Figure 8, left panels (and also discussion at lines 431, 437 and elsewhere in section 3.2): It is not the net surface flux that should be shown here, but the cloud radiative effect. This way the true fractional error of the various methods can be worked out. For example, the cirrus case at 60 degrees SZA shows the "GCM" method has a solar bias of -25 W m-2, but the net flux is around 300 W m-2, implying that this is less than a 10% bias. However, it should really be compared to the cloud radiative forcing which Hogan and Kew (2005) estimated to be -39 W m-2 for this case. Thus the error is more like 64%.

18. Section 3.2: I find the discussion of the biases in Fig. 8 rather unsatisfactory because there is inadequate discussion of the role of heterogeneity and overlap, or the problem of compensating errors (although the 3D effect is discussed at length). My interpretation would be as follows:

18a. Stratocumulus: GCM is biased low because it overly homogenizes the cloud, but then the question is why TC is biased high even though the homogeneity is about right. This raises the question as to whether the assumption of maximum overlap of the in-cloud heteorogeneities explains the excessive transmission to the surface. The vertical decorrelation length of in-cloud heterogeneities is typically assumed to be half that of cloud boundaries, so about 1 km, and if this was implemented it would block more of the solar radiation and reduce the positive surface-flux bias. Just an idea.

18b. Cirrus: GCM/TC models all significantly underestimate surface transmission because they all fail to capture the strong cloud heterogeneity. Only later does the reader find Fig. 10 in which a "fix" is presented, but it would be simpler to present the fix at the same time as the original problem.

18c. Cumulonimbus: There is a strong 3D effect for this case so if TC is closer to the 3D benchmark than it is to ICA then it is for the wrong reason: surely you should aim for TC to agree with ICA, then use some other scheme to try to capture the 3D effects on top of TC? The fact that all the TC calculations have a positive bias with respect to ICA is probably due to the incorrect maximum-random overlap assumption in this deep cloud system.

19. Line 493 and bottom-left panel of Fig. 10: I don't think you can conclude much from this analysis because you have a large error from the maximum-random overlap assumption, so tuning the treatment of cloud horizontal heterogeneity is probably leading to the wrong conclusions.