Review of "A Novel Model Hierarchy Isolates the Effect of Temperature-dependent Cloud Optics on Infrared Radiation" by Gilbert *et al.* submitted to *Geoscientific Model Development*

This study investigated the effect of "temperature-dependent cloud optics" on infrared radiation, with a specific focus on the Arctic region. The analysis is done through a combination of a simple mathematical model for two-stream radiative transfer, a single-column atmospheric model, an atmospheric model, and a wind-nudged atmospheric model. The results suggest that the impact of "temperature-dependent cloud optics" is less significant compared to the internal variability in the Arctic region. When model winds are nudged towards reanalysis, the internal variability is partially constrained, and the effect of temperature-dependent cloud optics becomes more prominent.

This study has the potential to update our understanding of the impact of temperature-dependent cloud optics on climate simulations. However, there are a few major issues in this manuscript which I list below. The authors may need to perform additional experiments and data analyses. And based on that, I would recommend major revision.

1. It is not correct to claim that the designed model simulations study the effect of "temperature-dependent cloud optics". The authors simply switched the cloud optics at 298 K in the original model to the cloud optics at other temperatures. It is essentially cloud optics at a constant temperature (or temperature-independent cloud optics). While it is OK to simply do this in idealized single-column model experiments, because the cloud temperature can be set at any value to quantify the flux changes in the extreme cases, it is not appropriate to do this in the full atmospheric model simulations. Although the authors mentioned in the discussion section that this will be part of future work, "temperature-dependent" is still a confusing term to describe the current approach. I recommend the authors rephrasing it or implementing the physics to the atmospheric model.
2. Based on what has been presented in this manuscript, I don't think the analyses are sufficiently thorough, and the power of model hierarchy on understanding the impact of physical assumptions in climate models is not fully realized in this study. For example, the change of surface downward longwave radiative flux due to the use of temperature-dependent cloud optics is not well quantified. Only spatial pattern of differences between model runs are shown (Figures 7~10). The ranges given in the manuscript are mostly approximate (e.g., 1~2 W/m^2, 1~3 W/m^2, 1~7 W/m^2, etc.). Also, only surface downward longwave flux changes are quantified here, but the impact on OLR is also important from the perspective of the TOA radiation budget. I suggest that the authors should start from analyzing the global mean and regional mean time series of OLR and surface downward longwave flux, providing an estimate of flux differences, and then go further to analyze the spatial pattern of flux changes.
3. For the two-stream radiative transfer model described in section 2.3, the authors chose to use a very simple mathematical model to do the calculation. This does not take into account the atmospheric absorption, while it is an important factor that may mask the effect of cloud optics change. The authors may use a more developed two-stream radiative transfer model. For example, RRTMG_LW provides a single-column version that users can specify any profile to

test. Using this model, the authors can calculate the flux differences in broad cases and even plot the sensitivity of flux difference to the meteorological factors and cloud properties.

4. For the single-column atmospheric model, what variables are prescribed by the observations? My understanding is that clouds are not constrained by the observations. For most observational period in Figure 6, the flux difference is very close to 0. Are they cloud-free scenarios? I would suggest filtering out the clear-sky cases and focus on the cloudy scene.

Specific Comments

1. L19-21: A reference may be necessary to support the statement that “All else being equal, clouds with small particle sizes also scatter more shortwave and emit more downwelling longwave than clouds with large particle sizes.
2. L39-40: “Specifically, temperature-dependent liquid water optics are not used in RRTMG.” Related to the first major issue, this sentence is very confusing as the authors did not implement the full temperature-dependent liquid water optics in the model, either. The authors may be more specific on what specific cloud optics RRTMG has used (e.g., at 298 K), and point out that this may not reflect the truth in the supercooled liquid cloud regime.
3. L39: Also cite Clough et al. (2005; https://doi.org/10.1016/j.jqsrt.2004.05.058)
4. L45-47: This long sentence is a bit confusing. “supercooled liquid clouds frequently occur in both observations […] and the climate model […] and where the atmosphere is typically cold and dry.” These three are not in parallel. Consider this alternative: “supercooled liquid clouds frequently occur in the cold and dry region, as evidenced by observations and climate model simulations.”
5. L92: For surface, “albedo” is specific for solar radiation. A better term could be “reflectivity”.
6. Figure 2: “reflected ground emission” is ambiguous. A better alternative is “ground emission scattered by clouds”
7. Figure 2: In longwave radiative transfer, it better aligns with the convention to use emissivity rather than reflectivity.
8. Figure 3: For panels (c) and (d), it could be better to visualize the difference between 263 K optics and CESM optics.
9. Table 1: Do these model runs include model spin-up period? It takes time for the model to adjust to the new state.
10. Table 1: Why is the 263 K run missing in the F1850 experiment? Especially consider that Figure 3 highlights the comparison between 263 K optics and CESM optics, and also the 263 K run appears in all other experiments.
11. L141: “the next time step”. Note that 6-hourly ERA-Interim reanalysis is used here while the model step is 30 minutes by default. According to the referred literature, this is indeed the next available analysis time, not the next model time. Please be more specific and clear.
12. Figure 4: In panel (a), I noticed that there is a smoothing gradient at the boundaries of the latitudinal band. The previous study cited by the authors explicitly mentioned that they

applied smoothing (by setting α to a value between 0 and 1 in some region). Did the authors also apply the same technique? Also, in panel (b), a solid line is connected between $\alpha = 0$ and $\alpha = 1$ at around 800 hPa. Is the smoothing technique also applied here? To make it clear, instead of using line plot, the authors may choose scatter plot instead to visualize the exact α values at each discrete layer.

13. L164-165: "the downwelling irradiance and flux was higher for temperature-dependent optics than temperature-independent optics" This is confusing. It would be better to state that the downwelling irradiance and flux was higher for cloud optics at X temperature than the optics at Y temperature.
14. L165: "The thinnest clouds […] showed the largest difference." This statement is not supported by Figure 5, as no results are presented for clouds at different thickness.
15. L167: What is the meaning of "all cloud temperatures"? Rephrase this sentence.
16. L168-169: "However, as cloud thickness increased from 100 to 500 m […]" This is not shown in any figure.
17. L170~171: "but our model was meant to be a proof of concept and not realistic". Why not use a realistic model, given that a quantitative estimate of the effect is provided above (0.35 W/m^2)?
18. L177-179: The authors mentioned that when cloud optics at different temperatures are used, the cloud fraction and cloud phase in the simulations are different. I assume that the authors do not prescribe the model simulations with observed clouds. What are the differences in cloud fraction and properties exactly? Having these differences, I don't think this is an apple-to-appple comparison to show the net effect of cloud optics at different temperatures since cloud variability has played a role.
19. Figure 7: I don't see stippling in the figure, so it is better to say that no significance in the figure caption.
20. Figure 8: What's the regional mean difference in these plots? The average can be performed over 50ºN~90ºN, consistent with the given latitudinal band in Table 3, and the values can be added to the panel title.
21. L208: I suggest adding "at 5% significance level" to be more accurate and specific.
22. L209-210: "because the wind nudging reduced the variability in the annual mean flux between the ensemble members" A figure may be necessary to show this. If there are too many figures, consider combining the information in one figure. For instance, Figures 7~10 show similar information and can be merged into one figure.
23. L218~220: "no flux differences […] were statistically significant" Instead of setting some threshold, I suggest providing a p-value so that we can understand how far it is from the significance threshold.
24. L232-236: Given that the authors simply change the cloud optics at another temperature, the effect on mean 2-m air temperature difference should be more prominent than the effect on 2-m air temperature trend, since the temperature-cloud property feedback is muted. Also,

considering that no greenhouse gas and aerosol forcings are included in the simulations, it makes no sense to compare to the ERA-I 2-m air temperature trend.

25. L246-247: “Whereas for the global cliate model, an effect of a few W m^{-2} is within climate variability and thus relatively small.” Note that the historical change in effective radiative forcing from 1750 and 2019 is also a few W m^{-2}.
26. Table 3: The values in the “Effect of optics” column should be the regional mean values as defined in the “Spatial scale” column.