Review of "Evaluation of radiation schemes in the CMA-MESO model using high time-resolution radiation measurements in China: I. Long-wave radiation" by Yang et al. submitted to Geoscientific Model Development

This study compares the downward longwave irradiance predicted by the China Meteorological Administration mesoscale model (CMA-MESO) with high-resolution long-term measurements from 42 sites in China. The authors conclude that the model generally agrees with the observations. The correlation between the prediction error and multiple factors (including geolocation, hour of day, cloud amount, etc.) is analyzed. Generally speaking, the dependence of model error on clouds is important and interesting to me. However, the results and the findings presented in current form may not be sufficiently insightful and novel enough to serve as direct feedback to model improvements. In other words, we know that clouds are among the largest contributors to the model uncertainty. Emphasizing it again is not valuable for future model development. It is still necessary to figure out what physical processes specifically contribute to model biases. Approximations made in the radiation scheme? Errors in cloud microphysics? I also have other major concerns regarding the title and the methodology of this study, notably the use of the MODTRAN model. Overall, I recommend a major revision or even a rejection of this version of the manuscript.

Please see below for specific major concerns and minor suggestions.

Major Concerns

1. Title: The title of the manuscript is "**Evaluation of radiation schemes** in the CMA-MESO model using high time-resolution radiation measurements in China: I. **Long-wave radiation**". The title is indeed very general and large. Upon reading the title, I would expect the authors will comprehensively compare the model outputs to various radiation measurements (e.g., radiation measured at the top of the atmosphere and at the surface, upwards and downwards). Yet, this study narrows down to only compare the model outputs against the downward longwave irradiance, an important variable though but just one variable in the radiation budget. I would say the title is very misleading, and the authors should polish it to make it more specific and accurate.

In addition, this manuscript seems to be the first part of a multi-part article with the subtitle "I. Long-wave radiation". The authors mentioned in the final section that an evaluation of the shortwave radiation scheme may follow, which is okay. However, for the first part of this article, I would expect an overarching introduction to this model, like what are the highlights of the CMA-MESO model and what technically makes it different from other mesoscale models. What I read in the second paragraph in the introduction is generally the model adopts the RRTM scheme and what the RRTM scheme is. I don't have a picture of what improvements have been made based on this existing radiation scheme, which motivates an evaluation of the model. The authors may also need to include a sentence in the introduction section to inform the readers what should be expected in the next part of the article series.

2. Essentially, this study compares the predictions from the CMA-MESO model to the field measurements. While I understand the difficulty in comparing the gridded model outputs to the sparse and uneven radiation measurements, the authors also introduced a few other changing factors that make the analysis more complicated. Based on my understanding, the authors used the concatenated time series of model data from mixed versions of the model. I am not very sure whether the change is big or small, physical or cosmetic, in each version of the model. However, I do think that such comparison using different versions of model outputs can be challenging and unfair. Different versions of models are evaluated against measurements in different time periods, possibly with variable meteorological conditions. If I were the authors, I would do a "retrospective prediction" using a single version of the model. Say, I could predict a certain time frame in the past records based on the input data before that point. Given that, different versions of the model can be compared to each other, and it would be interesting to see whether one version is better than the others.

Without this complexity, to make it an evaluation of the model, it is important but still difficult to attribute the model error by simply analyzing the statistical correlation between the model bias and the physical variables. For example, the authors claim that the RRTM scheme in the CMA-MESO model struggles with cold and dry conditions. There are so many confounding factors that may lead to a different conclusion. What if those cold and dry samples all come from the plateau region where the model indeed struggles with the topography? An interesting question to answer could be what should be blamed for the difference between the model and observations. The structural error in the radiation scheme? Or the prediction error in clouds/aerosols/temperature/humidity? At the end of this manuscript, the authors mentioned the possibility of checking the model prediction of those meteorological variables with the measurements, which I think should be done in this manuscript to make the argument stronger.

3. Section 3.2.2 tries to elucidate the underlying physical mechanisms that control the downward longwave irradiance using the spectral fluxes from the MODTRAN radiation model. Personally, I think it is basically a radiation 101 material that teaches the readers what factors can affect downward longwave irradiance. It is not relevant to the model evaluation and does not directly explain the model bias. It would be great if the measurements contained spectral fluxes, and then we would be able to diagnose what could be wrong with the model predictions and link it to physical mechanisms. Even the model does not output spectral fluxes that could be used to compare against the MODTRAN simulations. As far as I am concerned, this subsection does not contribute to the main theme of this manuscript and should be entirely removed.

Minor Suggestions

L22: "but" I don't think the sentences before and after present contrasting ideas and require a turning point. It might be appropriate to make them separate sentences.

L39-40: "[…] due to the radiation emitted by the instrument body is comparable to that being measured in wavelength […]". First grammatical error "due to" => "because". Second I don't quite understand what this sentence actually means. Please consider rephrasing it.

L42: "Brunt et al., 1932" => "Brunt, 1932". This is a single-author paper.

L57-58: If talking about the longwave fluxes and cooling rates in general, the authors should also consider ground emission. I suggest a simpler version of this sentence: "[…], which are governed by the absorption and emission of the infrared radiation from both the atmosphere and the ground surface (Shen et al., 2004)."

L67: From the perspective of spectral resolution, LBLRTM can be very accurate. But it doesn't sound right to call it the most accurate radiative transfer model. I would say it could be regarded as the baseline model to be compared to.

L77: What is the temporal resolution for those in-situ DnLWI measurements?

Tables 1 & 2: It could be informative to include a map to show the locations of the sites and the instruments each site uses.

L144-145: "There are 2501×1671 grid points in the **north-south** and **east-west** directions, respectively." I suspect there should be 2501 grid points in the east-west direction ($\frac{(145-70)°}{0.03°} + 1$) and 1671 in the north-south direction. Please correct the order.

L149-150: Grammatical error. I don't know which is the main sentence and which is the clause. Please consider rephrasing it.

L169-171: Why should we average ten 1-min DnLWI measurements centered at the punctual hour of the prediction and compare this quantity to hourly instantaneous prediction? What's the temporal step for each integration of the model?

L188: "Talor, 2001" Typo in the author's name.

L195: What does "PDLR" actually stand for? In lines 154-155, the authors explain it as the Prediction Deviation. But where is "LR"? I don't think this is a standard abbreviation and discourage the use of such a confusing term.

L256-257: "The RRTM scheme was remarkably stable when the atmosphere was wet." As I stated in my major concern, there are many confounding factors, and it is hard to argue that humidity directly affects model performance without more sounding proof. Also note the difference in sample size between the bins (18,967 in dry, 330,576 in semi-humid, and 51,738 in humid).

L289-290: "The DnLWIs observed were generally lower than those predicted […] under the cloudless and partly cloudy conditions, but higher than those under the overcast conditions." This is inconsistent with what Fig. 2b-d tells me. Observations (black) are generally higher than predictions (red) except under the overcast conditions.

L297: "[…] mostly come from the insufficient consideration of the PBL in the CMA-MESO model". This argument does not have sufficient supporting evidence, especially considering that much evidence points towards the high and medium clouds, which are located beyond the boundary layer.

L300: "[…] insufficient prediction ability of the RRTM scheme under extremely cold and dry conditions". Again, cold and dry may not be the culprit of the worse model performance.

Figure 2: The y-axis scale is different across different panels. At least for each category (e.g., humidity, season, cloud cover, etc.), the scale should be unified.

L337-339: I don't quite observe the reduced amplitude of PLDR with increasing temperature and surface pressure in Figure 3.

L332: "**All** correlation coefficients". Do they include the negative ones discussed in the previous paragraph?

L355: One caveat here is that the visibility is defined in the shortwave, visible spectrum.

L361: "small correlation coefficient (0.094) between the DnLWI and Vis". The correlation is between PDLR (i.e., the model error in DnLWI) and Vis, not the absolute value of DnLWI and Vis. Also, the sample size and confounding factors matter here.

L384-396: It is valuable to compare predicted meteorological variables to the reanalysis, which can help explain the contribution of model biases. Such should be applied to other variables as well.

L389-391: "CCDs less than –10%". Is it worse than or better than an underestimation of 10%? It could be better phrased as "CCDs are underestimated by greater/smaller than 10%".

L395: "was may relate to" grammatical error.

L403-405: Why T/Q profiles at those vertical layers are filled with values from the standard atmospheres? Are these layers missing in the CMA-MESO model? Then how is radiation treated beyond the 100 hPa level in the CMA-MESO model? It could induce inconsistencies between the CMA-MESO model and the MODTRAN simulations.

L407: "MDDTRAN" typo.

L409-411: Does it mean that only the vertically integrated cloud amount is input to the MODTRAN simulations for the three standard types of clouds? If so, the standard cloud profile defined in MODTRAN for low, medium, and high cloud is scaled and used, which can be very different from the real cloud profiles in the CMA-MESO model. Is my understanding correct?

L412: "anisotropy" This word means the radiation may be stronger in one direction than in others. Do you mean spectral dependency rather than directional dependency?

L439-441: "[…] the reliability of the DnLWIs predicted by the CMA-MESO model may be affected by the inappropriate input of the cloud types, especially the high and medium clouds." Given the input cloud profile described in previous paragraphs, do you mean that the standard cloud profiles defined in the MODTRAN profile are better than the simulated profiles in the CMA-MESO model?

L467-468: "The reason why **uses** […] is **to facilitate detect** the […]." Grammatical errors.

L471: "[…] discrepancies between the RRTM scheme and the radiative transfer model." In my previous comments, I mentioned that the input profiles of both models are not exactly the same. Thus, the comparison of the outputs may not purely present structural differences between the two models.

L490: Delete "e.g."

L491: "[…] they **[are]** close~~d~~ to zero when the atmosphere altitude **[is]** greater than 10 km." Grammatical errors.

L491: "Not" => "Note"

L493: "[…] the CO_2 profiles decreased exponentially with increasing height." It might be worth mentioning the unit of CO_2 concentration in the figure.

L574: "[…] due to the strong thermal emission of the clouds". Strong thermal emission of the clouds only explains larger DnLWI under the overcast conditions. It is the model overestimating the longwave cloud radiative effect (LWCRE) that explains the change of model bias from negative to positive.

L589: "uncertain" => "uncertainty".

L590: "are looking forward to" => "are expected to".