Review of "Simulating the variations of carbon dioxide in the global atmosphere on the hexagonal grid of DYNAMICO coupled with the LMDZ6 model" by Lloret et al.

This paper presents a comparison of two general gcms, LMDZORINCA (REG) which is an established setup, and ICO which is the novel version presented in this manuscript. The authors test the effects on simulated CO_2. Their main conclusion is that a similar performance is found between the two set-ups, but that ICO has a 20 percent less cpu cost by utilizing a smaller/reduced grid.

Although this comparison in itself can be an interesting step in improving the transport model performance, the current manuscript is not convincing in showing that this development is in fact an improvement of the REG setup, and it would require substantial revision before it could be considered for publication. Below are several general and specific comments that could help to improve the manuscript.

General comments

The manuscript seems to contain relatively little innovation. The abstract suggests that ICO is the innovation that is being presented. Section 2.1.2 seems to introduce the ICO concept for the first time, but no explanation is given on how this coupler works. For example, what time stepping is used, and why. How are these different in REG? Currently it is not clear which innovation is exactly documented by this manuscript. Either the innovation was already described in previous studies (shown by several references to previous work, e.g. on line 82), or if this is the first manuscript presenting, the description would need to be more elaborate. The methodology section would need to be rewritten so that it is clear which information is new, and only include information from previous work if it is necessary to understand this manuscript.

In the manuscript, several problems are identified which may explain the cases where ICO performs worse than REG in terms of simulating CO_2. However, these are either not solved, (mass conservation in Section 3.1, see also major concerns below), or indicated to be beyond the scope of this study (Section 3.3 temperature profiles, and their potential impact on the results presented in Section 3.5). It is therefore not clear if and why the authors would recommend using ICO instead of REG? Reading the manuscript as is, it seems that REG shows better performance, and yet the authors suggest continuing the development of ICO, and seem to indicate that ICO is in fact showing equal performance compared to REG. How are future developments going to impact the results? If there are major updates in the pipeline, would it not be better to merge those with the current manuscript, so that it does not lose its relevance after those updates?

Also, a discussion of the performance is missing, and the manuscript would benefit from having a discussion section included. In such section, the authors could go more in depth into the outcomes of the comparison and the implications. Currently, there is no comparison to the GPU version mentioned in line 40. It would be highly recommended to also include a comparison to that version, especially also in terms of computational efficiency. It is not clear where these two innovations stand with respect to each other. Also, the position compared to other models could be discussed, since the new transport scheme benefits mainly from reducing the number of computations that are required, which is comparable to reduced grids that other models use (see e.g. Petersen 1998).

Currently, it is not clear what the reason is for the violation of mass conservation. I would agree that the amount of mass being lost is indeed small, and not directly relevant for CO_2 inversion studies. But for a technical evaluation study such as presented in this manuscript, it is vital to know the origin and whether the irrelevance can be safely extrapolated e.g. when moving towards a higher resolution.

The technique used to evaluate the seasonal cycle is not explained. The curve-fitting procedure that was used is not specified (line 169). It is therefore not clear how conclusions can be drawn based on the seasonal cycle. What is the used criterion for a good fit and a bad fit? What are the residuals; do they still contain a trend or a seasonal cycle? What about the year-to-year variability? I think it would be relevant here to show the outcomes of the fitting routine to both setups. Based on this you may wish to introduce a rejection criterion for your final conclusions: if the original observed data is not fitted well enough, which I hypothesise to be the case for GIC/UTDBK and also CPT (as shown in figure 6 in the manuscript), it is not informative to see how the model performs there, it is just showing that the curve-fitting routine that is applied is not sufficient to yield a good answer (in other words, for those sites, I don't think we are looking at the REG/ICO's failure to reproduce the seasonal cycle, but to the capacity of the curve-fitting to perform a good fit). It is also good to check whether the quality of the seasonal cycle is due to feedback within the model between meteorology and the land surface model or is it related to the transport dynamics?

Specific comments

It would be useful to include tables, to help to better understand the parallelization schemes and how they lead to comparable setup, and which runs are performed for what results.

Figures 5, 7, and 8: Since the main difference between the models is the resolution towards the poles, it would be useful to see the differences by latitude. I would suggest ordering the stations by latitude. Also, aggregated statistics by latitude bands would be useful too.

L. 54: What does coupled configurations mean? What is coupled to what? Maybe a block diagram of both coupled configurations could illustrate the differences between the setups?

L. 61: the abbreviations are too long to be easy to read. Could you add dashes like in the full name in line 66?

L. 79-80: It is unclear how the "mix of finite difference and finite volume" affects either the primitive equations or the transport equations. Do both transport and the primitive equation use a mixed approach? Or does the primitive equation use a finite difference and the transport finite volume, and is the total model therefore considered a mix? Is this relevant e.g. in relation to mass-conservation?

L. 86: Is the explanation of parallelization required here?

L. 90: “our resolution” is unclear, but this is specified in lines 134-137, and could be moved up.

L. 102: What does “coarser” mean? How much?

L. 110: Here, further details on the parallelization scheme would be useful.

L. 120: I would rephrase this to a simpler sentence: "In both configurations the large-scale atmospheric circulation was nudged to the 6-hourly ERA5 reanalysis for wind." And why are the other parameters, like temperature not nudged? Could that (partially) solve the stratosphere and troposphere bias?

L. 122: mixing ratios is not the correct term to use, this should be replaced by mole fractions (also in other lines in the text).

L. 123-124: Could you specify the prior fluxes used? In line 226 it is mentioned that the prior fluxes are relevant, but they are not known to the reader. For example, the prior fluxes might partially explain the seasonalities that are presented.

L. 142: Have you assessed the impact of the used sampling scheme? There can be substantial horizontal and vertical gradients in the simulated CO_2 mole fractions. If you want to assess how these two simulations perform, especially at the surface. The two grids are very different, so the sampling scheme can be relevant.

L. 144: Since you are using the observations from this obspack product elaborately in the manuscript, it would be appropriate to contact the PIs of the datasets to discuss how these should be acknowledged. Currently, only Jungfraujoch and Aircore are specifically mentioned in the acknowledgements, but JFJ is not shown explicitly, while other stations are, e.g. in Figure 6.

L. 152: The explanation of the AirCore technique is not fully correct. It does not take “many successive samples of the ambient air when descending”. It rather takes a single sample while descending and utilizes the length of the tube to preserve the vertical gradients as much as possible (affected by diffusion) to obtain a vertical profile. Either rephrase or consider if the explanation is needed at all. What matters for the manuscript is that it provides a vertical profile.

L. 171: typo in 19080.

L. 176: This section could benefit from describing the calculations with a couple of equations.

L. 178: It is better to consistently use mole fractions rather than concentrations (nor mixing ratios).

L. 214: This mass conservation issue at first depends on the advection scheme used, and to what extend mass balancing is applied. Secondly, it will depend on the accuracy of floatingpoint arithmetic. Could you elaborate on how these two components affect either the ICO simulation or the REG simulation? This information is relevant for the discussion. I agree that the loss could be acceptable, and at this scale would not significantly influence inversion-based estimates of fluxes, however at least the process behind should be known for such a model evaluation study. For example, what would happen when you would go to a higher model resolution with ICO?

L. 219: Several of the abbreviations in the equations are not explained in the text. What is the superscript ‘e’? What is ‘emi’? Please check these abbreviations. Also, the notations are not standardized, e.g. chemical elements should not be italic. Mass is recommended to be written with small m, not capital. See IUPAC Green Book.

L. 226: As mentioned before, these surface fluxes are not described in Section 2.2.

L. 247-248: I would be interested to know the comparison to the GPU version of Chevallier et al. 2023.

L. 257: This difference of 10 K that is not shown in the figure: which is better, the REG or ICO setup?

L. 289: Considering the RMSE and looking at figure 4, I would not say the ICO has a slightly lower overall bias, since the difference in bias is probably insignificant considering the total spread of the data. They are both significantly different from the 1:1 correspondence line, but not from each other. Checking whether this difference is significant could be done using a student's t-test.

Figure 5: How is the data on x-axis ordered? And why is it not ordered geographically? I would suggest sorting by latitude, so that it is easier to see whether larger differences start to occur closer to the poles. See also general comment above.

L. 297: Before you can conclude that ICO better captures the gradients (growth rate?), it is necessary to add uncertainty estimates to the 1.43 and 1.3 values. Are they significantly different? If that is larger than 0.15, they are essentially the same.

L. 300-305: It would be very useful to see a latitude-bias plot, also as for example averaged data per 5- or 10-degree latitude bins, over zonal bands, to show that the bias doesn't increase significantly with latitude.

L. 307: I would use different selection criteria here for the stations shown. First, for certain sites, 8 harmonics may either be overkill, or not enough. I think it would be justified to ignore the sites where the curve fit on observations fails to represent the data. Then, the main point of the reduced grid in the ICO has coarser resolution at the poles. If you select 1 station for every 20-degree latitude bin, you can clearly illustrate that the seasonal cycle is well captured. It is already present in this figure through BRW.

L. 350: What do you mean with “variations”?

L. 375-384: What do you mean by “effective”? I am not sure if I agree with the general message in the conclusion that “it did not worsen either” or “comparable” vertical profiles. From the results section, I would say that the ICO setup gives slightly worse results, and that the only advantage is that it is somewhat faster. If computational efficiency is the most important aspect, and the slightly worse results are not an issue, one could go for ICO, but I would not see a reason to do this, since the gain is marginal. Also, as mentioned before, there is the GPU version of the same setup, and the comparison to that is missing. I wonder if with the GPU you still need the ICO configuration, since the main gain is already solved in another way.

References:

Petersen, A. C., E. J. Spee, H. vanDop, and W. Hundsdorfer (1998), An evaluation and intercomparison of four new advection schemes for use in global chemistry models, *J. Geophys. Res.*, 103(D15), 19253–19269, doi:10.1029/98JD01380.

Quantities, Units, and Symbols in Physical Chemistry, IUPAC Green Book, 3rd edition, prepared for publication by E.R. Cohen, T. Cvitas, J.G Frey, B. Holmstrom, K. Kuchitsu, R. Marquardt, I. Mills, F. Pavese, M. Quack, J. Stohner, H. Strauss, M. Takami, and A.J. Thor, RSC Publishing, 2007 [ISBN 0 85404 433 7; ISBN-13 978 0 85404 433 7];