Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2018-1117-RC2, 2019

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Calibration of a multi-physics ensemble for greenhouse gas atmospheric transport model uncertainty estimation" *by* Liza I. Díaz-Isaac et al.

Bruhwiler (Referee)

lori.bruhwiler@noaa.gov

Received and published: 11 January 2019

This is a very interesting study that seems to make some progress on an important issue for atmospheric inversions - how can we estimate atmospheric transport uncertainties? Most of us just use educated guesses, so it's really nice to be shown a potential way to do better even if it appears to be a lot of work and computational expense. I think the paper should be useful to the community of "flux inversers". One slightly disappointing thing is that CO2 BC errors cannot be distinguished from transport errors making me look forward to trying this with a global model.

I have mainly minor comments, and there are a few things I didn't follow and would like

to better understand.

Abstract, L19 - I think "observations" should be added to the beginning of this sentence.

P2, L1- On what basis do these studies rule out spatial scale as a factor in inversion differences? Some of these studies use results from models with different spatial resolutions.

P2, L30 - The measurement people would object to the use of "ppmv" rather than "ppm" here because CO2 deviates from being an ideal gas. ppmv also appears elsewhere.

P5, L25 and throughout - The v in the for the virtual potential temperature gradient should be subscript to avoid confusion with a product.

P5, L26 - How robust is this definition for the PBLH? Is there a reference discussing this?

P6, L11- There's an extra "s" after rank.

P6, L20-25 - Would it be better to describe an under-dispersive ensemble as a distribution that is sharply peaked and shows less variability than observed? This would match up with the description of over-dispersive as having too much variability. Just a minor point though, I had tor read the sentence a couple of times, but then understood it.

P6, Eqn 1 - Is N the number of ensemble members and is this the same as "the number of models"? Also, it could be noted that the expectation is obs. evenly distributed over bins.

P7, L1 - Where does our statistical expectation of how well the ensemble matches the observed variability come from? Suppose that rj = r(bar) in equation 1, then it seems that the model is getting the observed variability right, but what helps us to decide that this is overconfidence and not an extremely successful model?

P7, L4 - Does "samples" in this sentence refer to ensemble members or observations?

If covariances are underestimated, would this mean that there is nonindependent data and over-representation in a certain bin?

P7, L9- I think “mismatches” should not be plural here.

P7, L17-19 - Check the grammar here, “These” appears twice.

P8, L26 - The “flatness score” is the rank histogram score? Should stick with same terminology if possible.

P8, L27 - Is this N the same as the N that was used previously (e.g. the number of ensemble members)? I think this must be a different N that is something less than the previous one.

P8, L28 - It seems like a new symbol is being used for the rank histogram score here (it is delta in eqn 1). Is this because it’s going to be optimized by the SA/GA procedures and so a cost function will be defined?

P9, L9-21 - I have a few questions about this description. First, isn’t the deviation of delta from 1 what is being optimized here? I don’t see how this is explicit in the notation. The other question I have is about the size of the sub-ensemble. Can the procedure test sub-ensemble sizes all of the way to N-1 and all of the way down to some minimum number, maybe 2?

P9, L30-31 - Is mutation a separate step here? Or is is considered part of “crossover”?

P10, L103 - I have the same question that I had for the SA, are the sizes of sub-ensembles allowed to vary?

P12, Section 3.2 - Does this answer my question about exploring the sizes of the sub-ensembles? One uses the largest frequency from the rank histogram and since this happens to be the first box, then than one gets used? Why are 5-member ensembles used?

P16, L7-9 - I’m struggling with the implication of this statement. It means that even

though the sub-ensemble has the right spread it doesn’t mean the simulation will encompass the true values? What about bias? If the model is biased one could get this situation, right?

P16, Section 4.4 - I’m not sure I follow this argument. I see from Fig 15 that the spatial correlations of CO2 get closer to 1 or -1, but I’m not sure why this happens with fewer ensemble members. It’s stated that this is because of sample size (i.e. number of realizations?) but why should this result in a more intense correlation pattern? I would like to understand this.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-1117, 2018.