Geosci. Model Dev. Discuss., 6, C562–C565, 2013
www.geosci-model-dev-discuss.net/6/C562/2013/

Geoscientific Model Development Discussions

Interactive comment on "Correction of approximation errors with Random Forests applied to modelling of aerosol first indirect effect" *by* A. Lipponen et al.

S. J. Ghan (Referee)

steve.ghan@pnnl.gov

Received and published: 14 May 2013

This manuscript describes the application of a statistical method of reducing errors in a parameterization. The statistical method is trained on a large dataset of more accurate solutions, and then applied to an independent set of solutions. Although the method reduces mean squared error by an order of magnitude, the cost is a large fraction of the cost of the accurate solution.

Major Comments.

I do not find the choice of the sectional ARG scheme for both the accurate and approximate models to be optimal. The modal ARG scheme is used much more extensively

in global and regional models, and hence would be of greater interest to the modeling community. Of greater concern is the use of the ARG scheme for the accurate model. Because it relies on many of the same assumptions as the modal ARG scheme, it is not necessarily more accurate, except perhaps because of truncation errors in the sectional scheme when the number of sections is small. If you really want to reduce the number of sections, why not use a modal scheme. A much more valuable test of the methodology would be to use a detailed numerical model of aerosol activation to provide the reference solutions. I can provide such a numerical model, which is orders of magnitude slower than the sectional parameterization, but also much more robust, complete, and flexible.

The other major comment is that I would like to see an explanation for why the Random Forest method reduces errors. Presumably it is because it brings training data to the scheme, but perhaps a statement about how the training data provides more information about parameter dependencies that the approximate model misses.

Minor Comments

Page 2553, Line 13. Replace "is increasing" with "has increased".

Page 2553, Line 16. Remove comma.

Page 2553, Line 17-20. I know of only one climate model that uses a sectional activation parameterization. Modal aerosol schemes, i.e., Abdul-Razzak and Ghan (2000) and Fountoukis and Nenes (2005) are used much more extensively. Note that since the latter scheme cited is for modal, if you want to cite sectional schemes you should cite Abdul-Razzak and Ghan (2002) and Nenes and Seinfeld (2003). But why does your analysis focus on sectional parameterizations? Why not focus on the modal schemes? I see no reason why not, and strongly urge you to repeat your analysis with a modal scheme.

Page 2554, line 3. New paragraph beginning with "The main".

Page 2554, line 25. Remove comma.

Page 2555, line 11. Remove “of”.

Page 2555, lines 26-28. The approximation errors for the parameterization are not just caused by the limited number of sections. The key challenge of all activation schemes is determining the maximum supersaturation. If that is not diagnosed accurately the number of sections makes little difference. Why is this application expressed in terms of number of sections? In is full numerical model the number of sections is the only remaining approximation, but that is not true for activation parameterizations. Why not just say that the parameterization produces an approximate estimate with errors due to a number of assumptions and approximations?

Page 2558 line 9. Add “a” after “as”.

Pager 2560. A motivation for the distinction between Algorithms 2 and 3 is needed. I had to reread the text to find that Algorithm 2 is for training, and Algorithm 3 is for application.

Page 2562, line 8. For completeness list the modal schemes Abdul-Razzak and Ghan (2000) and Fountoukis and Nenes (2005). I strongly urge you to focus your analysis on a modal scheme, as the sectional schemes are not used in climate models (I know that sectional models are used in some global aerosol models, but the computational cost of sectional models is so high that they are never used in climate simulations, which are run for one hundred years or more.

Page 2562, line 9 – page 2563, line 17. Now I see why the sectional parameterization Abdul-Razzak and Ghan (2002) was chosen for the analysis.

Page 2563, lines 20-25. Why is a many-bin version of the Abdul-Razzak and Ghan (2002) sectional model chosen to be the reference model? It is still a parameterization that relies upon many assumptions to determine the maximum supersaturation. Is there any evidence that it is more accurate with 70 bins than with 7? On what basis

do you claim that the 70-bin ARG parameterization is “sufficiently accurate”. A more accurate reference model is needed here, to really put your correction methodology to the test. I can provide you with a numerical model that solves the time-dependent Kohler equations with a large number of bins. I urge you to use it or a comparably accurate model.

Page 2565, line 17. Both the “accurate” and “approximate” models use the ARG parameterization. The choice of words is only appropriate if the “accurate” model is based on numerical simulations rather than the ARG parameterization.

Page 2568, lines 5-9. These reductions are not impressive to me, particularly since the accurate solution is not necessarily that accurate. I suspect the results would be much more impressive if a full numerical model is used for the accurate solution. It is much slower, so the speedup would be considerable. It remains to be seen how much more accurate the RF model would be.

Steve Ghan

Interactive comment on Geosci. Model Dev. Discuss., 6, 2551, 2013.