Geosci. Model Dev. Discuss., 6, C319–C324, 2013
www.geosci-model-dev-discuss.net/6/C319/2013/

Geoscientific Model Development Discussions

GMDD

6, C319–C324, 2013

Interactive Comment

Interactive comment on "The Subgrid Importance Latin Hypercube Sampler (SILHS): a multivariate subcolumn generator" *by* V. E. Larson and D. P. Schanen

P. Räisänen (Referee)

petri.raisanen@fmi.fi

Received and published: 16 April 2013

GENERAL COMMENTS

Cloud and precipitation related microphysical variables feature a lot of variability within areas corresponding to grid cells of large-scale atmospheric models. Generally, both cloud microphysical processes and radiative transfer depend non-linearly on these microphysical properties, which implies that if mean values over GCM grid cells are used as the sole input to the microphysical calculations, the corresponding grid-cell mean process rates will be biased. To account for the variations in microphysical properties, upscaling methods have been devised. One (and in most cases most straightforward) approach is Monte Carlo integration, in which the process rates are computed for subcolumns and then averaged over the grid cell.

This manuscript introduces a new method (SILHS) for generating subcolumns inside a large-scale model grid column, and reports its tests for a couple of case studies in the single-column model framework. A key feature of this method is that it generates multivariate samples, and can thus properly account for correlations between different variables, as well as vertical overlap. The approach is described in a logical manner, although not in such detail that the model could be reproduced based on this manuscript only. However, I do not consider this a problem, since the authors' model is freely available for downloading.

The SILHS approach is potentially very useful, and the paper is mostly well written. In addition to several minor comments mainly directed towards improving the clarity, I have one suggestion for extending the paper (which is, in part, inspired by Peter Caldwell's minor comment 17: "... I expect that the massive noise injection from this would outweigh the benefit of subgrid sampling"). This comment is listed first below, while other comments follow it in the order they appear in the manuscript.

SPECIFIC COMMENTS

1. As noted above, the main benefit of the proposed approach is that it can handle subgrid-scale variability of several variables and their correlations, along with vertical overlap. On the other hand, the disadvantage of the approach (like all approaches based on Monte Carlo sampling) is that the results contain random sampling noise, which is quite substantial for individual time steps, especially for the (computationally most attractive) lh2_int approach.

For comparison, typical GCM microphysics parameterizations distinguish between the cloudy and cloud-free part of each layer, but assume both to be homogeneous. Thus, in-cloud mean values of cloud water and other microphysical variables are used, with the potential for biases in microphysical process rates, due to their non-linear nature.

GMDD

6, C319–C324, 2013

On the other hand, there is no sampling noise.

Therefore, I suggest that the random variations be evaluated a bit more comprehensively, and their effect be put in the perspective by contrasting them with errors associated with the neglect of cloud subgrid-scale variability (i.e., the typical GCM approach).

First, this could be done in an offline mode: use the analytic (exact) solution for driving the SCM, and perform diagnostic computations for (i) full SILHS (several, e.g. 10 realizations for quantifying the variability); doing this for the lh_int2 version might be sufficient; (ii) a GCM-type approach where subgrid-scale variations are neglected (for ease of interpretation, you might still use the same overlap assumptions).

The most useful quantity for the analysis might be surface precipitation rate: How does the random error in precipation (e.g., the std. dev.) compare with the systematic error associated with neglecting subgrid-scale variability. For instantaneous values, and for integrals over time?

Second, the test could be repeated "on-line", so that each approach is used to drive the model. If the outcome is like I would expect, for short periods SILHS's random errors would dominate over the bias due to neglecting subgrid variations in the GCM-type approch, but when model integration is extended, the situation ** might ** be reversed (e.g., for accumulated precipitation) – which is to be tested here. Here, in addition to precipitation, LWP would be another interesting quantity.

2. p. 2138, line 9: please explain what you actually mean by "non-intrusive".

3. p. 2140-2141: While the literature review in the Introduction appears comprehensive, the copula approach (Norris et al., Quart. J. Roy. Meteor. Soc.,134, 1843-1864) should be mentioned. Like SILHS, this approach can handle correlations between two or more variables, although it appears relatively complex.

4. p. 2143:, line 9: "horizontal correlations" was not immediately clear to me. Horizontal correlation between the variates?

GMDD

6, C319–C324, 2013

5. p. 2146, line 3: For completeness, show also the formula for the normal mixture marginals.

6. p. 2148, line 5: "..an uncorrelated multivariate sample." I found this puzzling when I first read the manuscript. It would help to add a note like "the correlations between the different variates are handled below (point 4)".

7. p. 2147, line 9: suggestion "the following five tasks".

8. p. 2150, line 27: How are the sample points chosen when the cloud fraction is larger than 0.5?

9. p. 2150, Eq.(3): The variable name "vert_corr_coef" is slightly misleading, because with increasing "vert_corr_coef", "vert_corr" becomes smaller. Why not e.g. "vert_decorr_coeff"?

10. A follow-up comment: It could be worth stating explicitly that based on Eq. (3), the vertical correlations are the same for all variates. (This is not an unreasonable approximation as such, but not always true: e.g., precipitating hyrdrometeors tend to be closer to maximum overlap than cloud droplets).

11. p. 2151, lines 12-15: This is rather confusing. Try to reformulate? E.g., "Note that SILHS imposes (influences?) directly only the vertical correlation of uniformly distributed points. Although the vertical correlations for variates in the normal-mixture/lognormal distributions are related to those of uniformly distributed points, they are generally not equal".

12. p. 2155, lines 11-13: You use 2-64 samples per grid box and time step, but do you also use optimization of the set of samples over time steps? Larson et al. (2005) generated n_t samples, and used n of them at each time step, where n could be less than n_t. Are n and n_t the same (or different) here?

13. A follow-up on the previous one: how would the random noise for standard Monte Carlo sampling compare with the Latin Hypercube sampling? It was demonstrated by

Larson et al. (2005) that LHS reduces the sampling noise for time-averaged quantities, when compared to standard MC sampling (their Table 1) but not for individual time steps. That test, however, considered a simpler system (autoconversion only, overlap not considered).

This question could be addressed by adding one more case to the tests suggested in the first comment: standard Monte Carlo sampling without LHS (although, you might still keep the importance sampling).

14. p. 2156, line 9: Why is the convergence of rain water mixing ratio so slow for the RICO case? Is this because rain is generated mainly in the upper part of the cloud (above 1 km) with very small cloud fraction, while your "importance sampling" presumably targets the layer with more cloud at ~600-700m?

15. p. 2157, line 20: You should state explicitly that the timing estimates in this paragraph and Table 1 refer to the most "computer-friendly" case, with only two subcolumns. For a larger number of subcolumns, the cost is higher.

16. p. 2160, lines 25-. This is a good point to mention, and I think this argument has been made before in connection to McICA. Still, while the motivation for adding noise in ensemble forecasting is pragmatic (and the methods sometimes ad-hoc), its physical justification is that in the real world, tendencies of atmospheric variables are not determined solely by the grid-mean values, as in physical parameterizations. I think it would be just to say that there is no particular reason to believe that the nature of sampling noise introduced by SILHS would properly represent the uncertainty associated with the physical parameterization.

17. Figs 1-2. Consider adding graphs of accumulated rainfall at surface (i.e., integrals from t=0 to each point in time). It would be very interesting to see how this compares between the different experiments.

18. Caption of Fig. 7: Trivially, sudden jumps also occur for random overlap!

TECHNICAL CORRECTIONS:

1. Figs. 1-4. Consider using thinner lines. Now, the curves cannot be distinguished properly (especially the analytic solution is invisible in most of the plots).

2. Fig. 6: The second panel on the right (and the font in its title) is visibly smaller than the others.

Interactive comment on Geosci. Model Dev. Discuss., 6, 2137, 2013.