Geosci. Model Dev. Discuss., 6, C266–C269, 2013
www.geosci-model-dev-discuss.net/6/C266/2013/

Geoscientific Model Development Discussions

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

P. M. Caldwell (Referee)

caldwell19@llnl.gov

Received and published: 10 April 2013

This paper describes the implementation of a new subcolumn generator which can include correlations between state variables and between columns in the vertical. I believe there is great need for the kind of tool developed here and I think the paper does a great job of describing a complex topic. I have several suggestions for clarity and a couple questions about methodology, so I suggest acceptance with minor revisions.

Major Comments: 1. I don't like your methodology for computing vertical correlations. While the methodology of sampling from a uniform distribution of width 2*vert_corr centered on the sample drawn for the cell above gives the right behavior when vert_corr

is 0 or 1, it isn't clear to me that values for the 2 cells will have correlation = vert_corr for intermediate values of vert_corr. Further, I don't think "folding back" preserves the desired correlation structure. Here's an example: suppose vert_corr=0.95. If X_u(K) = 1, then X_u(k+1) $\sim$ unif(0.95,1.05) so half of the time samples will be >1 and will be folded back to 2 – X_u(k+1) => values will be between 0 and 0.05. If we were to plot X_u(k) vs X_u(k+1) on a scatter plot, this would correspond to a value very far away from the 1:1 line expected for a good correlation.

Minor Comments:

1. p. 2138 L 24: I suggest rewriting as "parameterizations are usually based on understanding of physical processes at a point in space, while models are typically formulated in terms of grid-box means. "

2. P. 2141 L 10: I think the biggest benefit of subcolumns is commonly thought to be the ability to handle any kind of vertical overlap. You should mention this!

3. P. 2143 L1: The 2-component normal mixture terminology threw me for a loop. Is this typical terminology in some other community? If not, why not just call it a double Gaussian?

4. P. 2143 L 27: Maybe change to "using a weighted mean if needed (described further below)". . . I was initially confused why you'd ever want to use a weighted mean.

5. P. 2148 L 4: Can't you say "we start by creating a vector of independent samples drawn from a uniform PDF on (0,1), where each element of the vector corresponds to a variable needed by the routines using the subcolumn."

6. P. 2148 L 7: It took me a really long time to understand this paragraph (partly because of confusion about the normal mixture terminology). I think what you're saying is that drawing a sample from a double-Gaussian is equivalent to drawing a sample from one or the other of the component standard Gaussians, with the probability of choosing one standard Gaussian or the other given by the weighting parameter of the

double Gaussian. So you select 2 random #s to represent the 2 standard Gaussians, then choose a 3rd random # to choose between the choices? If so, then the paragraph on P. 2151 ∼L20 saying that all sample points are later converted to normal mixture distributions doesn't make sense to me.

7. A follow-on question to the above: Do you always sort the initial pair of random variables in increasing order so the 3rd # chooses between big and small values? Otherwise, wouldn't you get the same distribution by just choosing a single random variable for position in the PDF and using another more random variable + double-Gaussian weighting parameter to say how to convert the first value to a Gaussian w/ the correct mean and sigma?

8. A general question/comment is that if you're using this sampling for microphysics, you'd want to only sample in the cloudy or precipitating region, since microphysical rates are 0 outside cloud/precip. Thus your inclusion of clear-sky samples confused me for most of the paper. I think you do so in order to keep your sampler general, so it could also be used for radiation (for example). You should clarify this.

9. P. 2149 L 1: I think sample weighting for in-cloud points would be cldfrac/(0.5*ncols) and clear-sky points would get weight (1-cldfrac)/(0.5*ncols). If so, you might want to say this. If not. . . well, then I don't understand.

10. P. 2152 L 13: Why don't you just say "For each hydrometeor x, we assume a normal PDF for y=ln(x)"?

11. P. 2152 L 23: I think you mean "all variates" instead of "all hydrometeors"?

12. P. 2153 L 11: change "in the prior step we have" to "we previously" to avoid confusion about whether "prior step" means prior timestep.

13. P. 2153 L 24-P2154 L2: I'm confused – these 2 sentences seem to contradict each other.

14. P. 2155 L 18: why the square brackets?

15. P. 2155 L 22-27: I suggest replacing these sentences with something like "Rain is created entirely within SILHS subcolumns, while subcolumn microphysics is only one of several terms contributing to LWP. As a result, subcolumn noise is felt more strongly by rain than LWP."

16. P2158 L 10: This section made me uncomfortable because 12 hr averaging time is long enough to destroy the diurnal cycle. Will this have a big effect on the simulations? I suspect not and I guess you could get around this by compositing over similar times for many days, but the 12 hr suggestion without worrying about diurnal sampling seemed kind of cavalier.

17. P. 2158 L 16: I think most climate models already break grid cells into clear and cloudy portions and compute radiation and microphysics separately for each, then do a cloud-fraction weighted average of the result. In this context, 2-sample SILHS differs from the default treatment only by choosing properties randomly instead of by using layer averages. I expect the massive noise injection from this would outweigh the benefit of subgrid sampling. . . So I need more convincing that 2 sample SILHS is a good idea. Additionally, isn't there a good probability for cldfrac>0.5 that all SILHS samples will be from cloudy conditions, so clear-sky conditions wouldn't be sampled at all? This seems distinctly worse for both radiation and microphysics than what's currently done.

18. P 2160 L 12: "of PDF of": something's wrong here.

19. P 2160 L 19: "may be a suitable problem" -> "is a suitable problem": subcolumns are perfectly parallel so there's no question that we can speed them up w/ more processors (or with GPUs!)

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