Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2017-3-RC3, 2017

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Tiling soil textures for terrestrial ecosystem modelling via clustering analysis: a case study with CLASS-CTEM (version 2.1)" *by* Joe R. Melton et al.

Anonymous Referee #3

Received and published: 22 March 2017

General comments:

The authors examine a clustering algorithm to represent sub-grid heterogeneity of soil texture for a coupled land-surface-ecosystem model (CLASS-CTEM). A newer clustering algorithm is applied to the CLASS-CTEM model soil texture inputs to define the sub-grid tiles in the model. The algorithm has not been previously applied to land-surface/hydrology/ecosystem models for sub-grid tiling and has the potential advantage of determining the number of clusters from the data; no a priori number of clusters needs to be defined. The CLASS-CTEM model shows sensitivity to tiling of soil texture in a synthetic experiment and two grid point simulations. An equilibrium is reached at around 7-8 tiles in the synthetic experiment. They then run global offline simulations.

Printer-friendly version

Discussion paper

Overall, clustering by soil texture using the algorithmic parameters discussed in the paper results in roughly 3-4 tiles per grid cell across the globe, with some cells containing up to 7-8+ clusters. The model does not show much sensitivity to ecosystem parameters (e.g. gross or net productivity) across the global uncoupled simulations, except in some arid/semi-arid regions along the margins of deserts.

The paper is very easy to read and logically organized. I think the application of the new clustering algorithm is useful to the community and may have broader applicability; the experiments give insight into this particular clustering method-model combination. However, I have some major concerns. If the editor believes the authors can address them in the allotted time, then it should be acceptable for publication in GMD after revisions, otherwise I would suggest rejection with resubmission.

Major Comments:

1) Lines 24-25 on page 2 (Introduction) mentions water balance partitioning (runoff vs ET) and then lines 17-18 on page 3 mention the offline run conclusions should hold for the coupled runs because grid-mean fluxes are passed back to the atmospheric model. However, there is no mention or presentation of any latent or sensible heat fluxes anywhere in the paper. It is very likely that the grid-mean values may be different in the regions of large net ecosystem productivity changes, which would then affect coupled simulations.

I am not expecting fully coupled runs here, but presentation and discussion of flux partitioning changes (or lack thereof) since the primary function of the land-surface model is flux partitioning.

2) Section 3.1: This section needs to be discussed in more detail. What was the vegetation distribution for this synthetic test? Does the vegetation distribution influence the sensitivity? If it does, perhaps another test would be useful. Does this then impact the equilibrium number of clusters?

3) The density-based approach of this clustering algorithm could be problematic. If only 57% of the high-resolution data are being included in clusters on average, you are effectively letting the algorithm determine that 43% of the data may not matter. It is particularly worrisome to see many grid points with clusters that contain only 10-20% of the high-resolution data. Just because the grid-mean texture values and these cluster values agree is not a good enough reason to say the clustering algorithm is successful. The "outliers" or lower-density regions may be important to ecology and/or hydrology. Seyfried et al. (2009) discuss hydrologic hotspots, portions of a catchment that are small but have disproportionate impacts on runoff. The same is seen here in Figure 9 for gross primary production (GPP).

If the clustering is not considering areas of a grid cell that may be small individually (yet add up to a very significant portion of the grid-cell total in many cases), it is possible that some sub-grid variability is being missed (or possibly over-emphasized). For example, what percent of the large grid-cell in Figure 9 is classified on the high-resolution grid? How much of the high-resolution data is similar to (or higher) in sand content than cluster E? If the total area of the large grid cell has more(less) than 10% of the high-resolution points similar to cluster E (from expert evaluation), the grid-cell mean impacts of tiling may be under(over)-estimated.

4) Why is vegetation considered constant across clusters (lines 1-2, page 6)? The GLC2000 dataset could have been mapped to the clusters to have variable PFTs with each tile. Soil texture and vegetation likely co-vary in some or many locations; it would be good to capture that relationship if present.

Minor Comments:

1) Page 5, line 27: Why 10%? Also, this is a 10% change relative to the grid-mean value?

References:

Seyfried, M. S., L. E. Grant, D. Marks, A. Winstral, and J. McNamara, 2009: Simulated soil water storage effects on streamflow generation in a mountainous snowmelt environment, Idaho, USA. Hydrol. Processes, 23, 858–873, doi:10.1002/hyp.7211.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2017-3, 2017.