Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-152-RC2, 2017

Geoscientific Model Development Discussions
Open Access

Interactive comment on "A map of global peatland distribution created using machine learning for use in terrestrial ecosystem and earth system models" *by* Yuanqiao Wu et al.

Anonymous Referee #2

Received and published: 1 November 2017

Recognizing the importance of peatlands in global carbo cycling, and a lack of dedicated peatland extent maps, the authors present an attempt at broad-scale mapping of global peatland extent for use in terrestrial ecosystem and earth system models. I applaud the effort and agree that it is very important work. I also believe that the approach of combining relevant datasets in a machine learning framework is a good one. However, I see numerous problems with the method and choices of input, training and validation data. These are summarized as: (1) lack of definition of peatlands, (2) problematic choice of input variables, (3) inconsistent and auto-correlated training data and (4) the model is validated against data which is strongly correlated to the input data. See more details on the four points below. Given these issues it is impossible to

Printer-friendly version

Discussion paper

assess if the derived map has any advantages over e.g. using existing global soil maps (e.g. HWSD, WISE30sec or SoilGrids) to parameterize the extent of organic soils in models.

Given these strong limitations in the basic methods, I cannot recommend this work for publication.

1. One of the problems with harmonizing or reconciling different approaches to peatland extent mapping (or modelling) is that different definitions of what constitutes a peatland exist. My first concern is that the authors do not themselves provide a clear definition of how they have defined peatland in their study. This is problematic since your input data is based on three different definitions (see below). I recommend the authors look at e.g. Joosten and Clarke (2002, "Wise use of mires and peatlands") as a first guide in their choice. You must choose one definition that you find useful and then design your study based on that. Since your stated purpose is to improve broad-scale modelling of the carbon cycle, a definition that reflects the depth of peat accumulation seems sensible (it is also the most commonly used).

2. I also object to the choice of using the HWSD soil carbon maps as input data. Those variables in the HWSD are calculated from HWSD soil coverage. In essence, organic C% in the HWSD is so autocorrelated to peatland extent in the HWSD that you are effectively used (a derived variable of) a coarse peatland map as "independent environmental" data with training data from other peatland maps (in some cases the very same map) and finally, ground-truthing it against that same map again (because Yu et al., 2010 is largely based on the HWSD, see further below in point 4). Your model is completely dominated by HWSD organic soil carbon variables, with scattered influence of climatic variables (none of which exceed 4% explanatory power). Your model shows no strong response to topography or any of the climatic variables that are believed to influence peatland extent (temperature, precipitation etc.). Most likely, those strong signals are masked out under the driving force of the HWSD data, which implicitly already includes that information.

3. You train the model with regional peatland presence and global peatland absence. This is inconsistent and somewhat problematic. The three different datasets (a, b c below) all use different definitions of peatland. And they are mapped at very different spatial scales and have very different thematic and spatial accuracy. The regional peatland presence is from (a) thematic soil maps in Canada (the Canadian system for peatland classes is the same as the soil classification system) and from (b) thematic wetland maps in west Siberia. The latter map actually also includes non-peat forming wetland systems: quote from Peregon et al 2009: "One reason [for the higher coverage compared to earlier studies] is that our estimate comprises not only open peat-accumulating wetlands but also forested and grass-dominated wetlands with/or without peat deposits." The Global peatland absence is from (c) a threshold of topsoil organic carbon content in the HWSD. The choice of a topsoil carbon threshold is difficult to justify; carbon stocks in the HWSD are calculated based on mapped coverage of different soil types, including maps of Histosols/peatlands. How can the authors justify that the derived HWSD variable of topsoil soil carbon should be superior for peatland mapping purposes compared to the actual mapped peatland extent in that dataset? Note that this threshold presumably misses many areas where there is >13 kg C but still no peatlands (eg upland Tundra soils or Boreal soils frequently have more than 13 kg C in the topsoil without wetland or peatland conditions).

4. And as mentioned above, your environmental data, training data and validation data are all strongly correlated. Given this fact, it is of course entirely unsurprising that you model performs very well. Further the split of the dataset in random pixels for training and validation also gives a strong autocorrelation which boosts the performance metrics. This is already well covered in a comment by M. Bechtold. Note that the peatland distribution map of Yu et al. (2010) is mostly based on soil maps, and shows peatlands to exist in cases where >5% of the terrain is peat. This percentage is a rather low threshold, and may have led you to overestimate peat cover. Further the use of the map by Yu et al for assessing your model is deeply problematic, since it is in no way independent from your input training data. If you look into the supplemental materials of

Yu et al (2010) you will find that the source for peatland maps of Canada is Tarnocai et al., (2002) and for West Siberia Sheng et al., (2004). These datasets are very closely related (or near identical) to the input data used for Tarnocai et al (2011) and the base peat type map used by Peregon et al (2009). Much of the world's other peatlands in Yu et al 2010 are mapped from the HWSD Histosol (and some Gleysol) coverage. So, your validation data is almost the same data as your environmental variables and your training data.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-152, 2017.