Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-28-RC1, 2019

Open Access Earth System Science Data Discussions

Interactive comment on "Monthly Gridded Data Product of Northern Wetland Methane Emissions Based on Upscaling Eddy Covariance Observations" *by* Olli Peltola et al.

Anonymous Referee #1

Received and published: 11 April 2019

The authors use a machine learning technique (random forest-RF) to develop a monthly gridded data product of northern (>= 45°C) wetland CH4 emissions. Three CH4 emissions products (2013-2014) are derived from RF based on different wetland maps. Annual total CH4 emissions from these three products are comparable to previous studies and two process-based models, however, the areal extent and spatial-temporal patterns of the CH4 emissions are largely subject to the wetland map. These products can potentially become a good benchmark for both top-down and bottom-up models. Overall, the manuscript is well structured and written. The methods and results follow well the objectives. The figures clearly illustrate the results. However, I think there are still some aspects this study should spend more efforts to justify before

Printer-friendly version

Discussion paper

the manuscript can be accepted in the final publication of Earth System Science Data:

(1) Evaluation of the products is too simplistic. It would be nice to illustrate the statistics for model-observation comparison spatially, which means to make a plot to see the spatial patterns of R2, NSE, RE and BE.

There are two hot spots for CH4 emission across the circumpolar region (Hudson Bay lowlands and western Siberian lowlands). But unfortunately, there is only one site from each hot spot area. Why don't the authors select these two sites for the time series comparison, like what is shown in Figure 5? I notice that Figure 4 indicates regression lines between model and observations for these two sites have large deviations from the 1:1 line.

(2) Why does the study only select two years of observations? As most observations in the Arctic have only been collected during growing seasons, there are not many monthly data points left within two years. Therefore, it is difficult to say how significant the regression relationships between model and observation in each site is.

(3) Why do the authors compare RF-GLWD with LPJ-Bern and WetCHARTs? I don't see the point of making such a comparison, because they are not based on the same wetland map.

(4) I noticed the input variables listed in table 1 comprise of some categorical variables. Can the authors describe more details about how RF use such information? Can the authors consider making an RF model based on individual wetland type or biome type? Moreover, why don't these input variables include water table position, which is important drivers for CH4 emissions, particularly to explain spatial heterogeneity?

Specific comments:

Page 2

Line 11: utilize random forest (RF) -> utilize a random forest (RF)

Line 16: What does “confidence interval” refer to? a random forest ensemble? Please specify it.

Line 27: constraint -> constrain

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-28, 2019.

Printer-friendly version

Discussion paper