

## ***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 #2**

Received and published: 17 April 2019

The authors have produced a gridded product of monthly CH<sub>4</sub> fluxes from Northern (>=45°) wetlands using a random forest (machine learning) algorithm. The random forest model was trained against a number of predictor variables such as nighttime Land Surface Temperature, absence or presence of permafrost, potential sunshine, sedge cover and biome class. Eddy covariance measurements of CH<sub>4</sub> from 25 sites was used to optimize model performance during the training. The fluxes were then upscaled using three different wetland maps and compared against previous studies and two process models. There results are comparable to these.

The manuscript is well written and explanatory, highlights uncertainties and discuss

C1

data in a good way. The product have potential to become valuable to process modellers as a benchmarking product further on, which is also the aim of the study.

I have a few thoughts that could be considered before the final report is submitted.

1) The most important predictor variable is the nighttime land surface temperature (LST<sub>n</sub>). While this is interesting and maybe not completely unexpected I lack the discussion around this. Is there a mechanism behind the nighttime temperature or is it an outcome of the machine learning algorithm? For instance, I would assume that most process based models would use daily temperature as that would be better correlated with ecosystem productivity.

2) In the upscaling process you initially calculate the flux density per grid cell and then use the wetland maps to produce the amount of CH<sub>4</sub> per grid cell. As the maps used contribute highly to the uncertainties in these amounts I would consider to add a map of the flux densities. Such a map could also be used as a benchmark product by bottom-up models.

3) The comparison to the process models contains uncertainty both in the wetland map used as pointed out by referee 1, but also in the climate forcing data of the models. This should maybe at least be clarified if a comparison should be made.

4) In the introduction the contribution of CH<sub>4</sub> emissions from natural wetland is discussed (page 2, line 30-33). Although it might be implicitly understood, I lack a clarification that these are global estimates.

Technical corrections: 1) I cannot find where the abbreviation of SC is defined. I assume that it stands for sedge cover?

2) Page 14, line 11 th an -> than

3) Page 14 line 12. Consider not using . . . for timespans. Use instead, e.g., years 2008 - 2011

C2

4) Page 14, line 15, same as 3)

5) Also why not use subscript for ecosystem respiration Reco ->  $R_{eco}$ ?

---

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