Dear Editor Ito and reviewers,

Thank you for reviewing our manuscript "Evaluating alternative ebullition models for predicting peatland methane emission and its pathways via data-model fusion". We are very appreciative of the time and effort that you have spent in improving our manuscript.

We have carefully gone through reviewers' comments and have responded to their concerns and suggestions, which we believe have greatly improved our paper. Please find our detailed responses below in blue underlined fonts.

Please let us know if you need any further information or clarifications.

Yours truly,

Shuang Ma, Ph.D.
Postdoctoral Scholar
Jet Propulsion Laboratory at Caltech
Telephone: +1 8183542113
Email: shuang.ma@jpl.nasa.gov

Reviewer 1

This study compared two algorithms of peatland methane ebullitive fluxes and examined the impact of data availability on simulation reliability. The EBG algorithm was found to be more accurate than the ECT in simulating both methane fluxes and pore water methane concentrations. Using both methane fluxes and pore methane concentrations in parameter calibration led to better constrained parameter values. Overall, the paper is well-written, and the experiments are solid. The conclusions can be potentially helpful for methane modelers. However, there remain doubts in the motivation and aim of the study, and more information and clarifications are needed regarding the methods and the data analysis.

We appreciate your suggestions, please find replies after each paragraph in blue underlined fonts.

Major comments

How available are pore water methane concentration measurements in wetland sites generally? More observations would always help to better constrain parameter values and obtain more accurate simulation results, but if such variables are quite rare or hard to be measured, how will the finding inform other modelers on how to improve their model performances?

Thanks for the question. And we realized the importance of explicitly pointing out the conveniency of conducting porewater methane concentration measurement, as our paper advocate the importance of this dataset to modelers. We explain here and will add it in our discussion.
Porewater methane concentrations are relatively easy to measure by methods that have been around for decades. The measurements are now fairly widespread and have been used across different wetland environments including: various peatlands from Arctic to tropical (Chanton et al., 2008; Hines et al., 2008; Holmes et al., 2015; Hodgkins et al., 2015; Wilson et al., 2021); agricultural rice fields (Kruger et al., 2002); marshes (Happell et al., 1993; Keller et al., 2009); Riparian wetlands (Itoh et al., fe2008), and arctic lakes (Walter et al., 2008).

L31: This sentence could be misleading since the paper didn't actually investigate the relative contributions of each pathway but only presented the modeled results which were not validated against measurements.

We agree, we will rephrase it as:

'Notwithstanding their importance, the relative contributions of the emission pathways are highly uncertain.'

Table 1: Why did you select growing season only? How long did each measurement last?

What does the 'CH4 fluxes' include here (from ebullition only)? Since ebullition is quite spatially heterogeneous, how many observation sites were deployed at the same time?

Thanks for the question. We should be more precise in describing months of measurement. We stop taking measurement of fluxes after October, in some years until December, well pass the growing season. We stop taking total CH4 emission measurements after getting very small values for that month, and as shown in figure 3, the fluxes at shoulder seasons are minimum. Another reason for missing those winter months is the cold, harsh weather with heavy snow cover at the experiment site, making it hard to deploy such measurements.

Below is the exact months we took measurement of CH4 fluxes, and we will add this information into Table 1.
• August, September, and October 2011,
• May through November 2012,
• July, September, and October 2013,
• June through December 2014,
• April through November 2015,
• March through December 2016

We measure total CH4 emission including all emission pathways. The fluxes are measured using the static chamber method (Hanson et al., 2014; Hanson et al., 2016a; Hanson et al., 2017b). Open-path CO2 x H2O (LI-7500A; LiCor Inc., Lincoln, NB) and CH4 (LI-7700) infrared sensors were used to measure change of CH4 concentration. A typical measurement interval lasts approximately 400 seconds. Total CH4 emission data are collected 2-6 times a day and we take the mean to represent the emission of the day for model validation. Also see lines 174-177 for data processing method.

Two plots were measured in the same day. We agree ebullition can be both temporally sporadic and spatially heterogeneous. It thus hinders quantification of CH4 emission pathways. The porewater CH4 concentration profile, however, is quite static (see figure 4), as it represents the amount of CH4 in each soil layer instead of the dynamic fluxes. Our study offered an alternative method toward estimating ebulliated, diffused, and plant-mediated CH4 fluxes by fusing vertically resolved mechanistic model with porewater CH4 concentration profile data. As individual pathways are not measured in this study, and realistically can't be measured with an ideal precision with current stage of techniques, we can not directly validate modeled emission pathways with observed ones. But we did make a case here that rather than using total CH4 emission data alone, CH4 concentration profile further constrain model structures and parameters, that potentially improve modeled emission pathways.

Eq 1: Why were fstp, fph, fred not calibrated? Did you do any sensitivity tests to determine what parameters to calibrate?

Yes! Sensitivity test was done in Ma et al. (2017) to determine what parameters to calibrate. fstp is a soil temperature scalar and CH4 production is very sensitive to this parameter value. The equation for fstp is in equation 9 in Ma et al. (2017) and shown below. We used Q10pro in data model fusion (parameter calibration).

$$f_{stp}(t) = \begin{cases} 0 & \text{if } T_{soil} < 0 \\ Q_{10pro}^{\dfrac{T_{soil}(t) - T_{optpro}}{10}} & \text{if } 0 \leq T_{soil} \leq T_{max} \text{ ,} \\ 0 & \text{if } T_{soil} > T_{max} \end{cases}$$

where $T_{soil(t)}$ is the hourly soil temperature and T_{optpro} is the optimum temperature for CH4 production, which varies across ecosystems. In this study we chose a value of 20°C since this was the maximum temperature for which methane production was examined in incubations of peat from this site (Wilson et al., 2016).

However, the factors fpH and fred are nominally set to a constant value of 1.0 due to the model sensitivity (Meng et al., 2012; Riley et al., 2011) and uncertainty in characterizing these two parameters (Le Mer & Roger, 2001; Wania et al., 2010; Whalen, 2005). Citing the corresponding paragraph from Ma et al 2017:

In the CLM4Me model, the effect of pH and redox potential on net fluxes were tested in the sensitivity analysis and resulted in less than a 20% change in net CH4 emission at high latitudes (Riley et al., 2011). Redox potential does not have substantial impacts on methane emissions at seven wetland sites including one adjacent to the Marcell Experimental Forest in north central Minnesota (Meng et al., 2012; Shurpali & Verma, 1998). Wania et al. (2010) argued that the pH and redox factors are so poorly characterized that they should be excluded. Many of the current process-based methane models use a single value for the pH scaler calculated from the soil property that does not change with time and depth. In many process-based methane models a step function is used for calculating the redox potential scaler (Fiedler & Sommer, 2000; Segers & Kengen, 1998; Zhang et al., 2002), which is decided by root distribution, fraction of water-filled pore space, the water table position, and several other constant parameters with a single value across different ecosystems such as change rate of soil redox potential under saturated conditions, cross-sectional area of a typical fine root, and fine root length density. In our model, the potential ratio of anaerobically mineralized C released as CH4 can reflect some of the information on the effects of pH and redox potential to methane production. We kept fpH, and fred in equation 8 because as more information become available, we might be able to improve their calculation in our later versions.

To help reading, we will add in our manuscript after introducing Equation 1:

'A sensitivity test was done to decide which parameters need to be optimized by data-model fusion (Ma et al. 2017), and more detailed descriptions on the CH4 module can be found in Ma et al. (2017).'

Method: More information about the modeling approach is needed. How was the model structured horizontally? For how long and with what time step were the models run? What was the spin-up period? How did you deal with data with different frequency and time span for model calibration?

C cycle is only vertically resolved like is done in the other state-of-the-art models (Tang et al., 2010; Riley et al., 2011; Wania et al., 2010; Peltola et al., 2018).

The model was run at hourly time step. We did not need spin up period because we know the initial values for carbon states from measurement (Table 1). In fact, this northern peatland site is not at a 'steady state' due to two strip cuts 5 years apart (1969 and 1974, Verry et al., 2011).

Data-model fusion does not require zero-gap data, and is flexible for the time span of data. For example, 1) if we had hourly CH4 emissions data, we would be able to compare modeled-observed CH4 at each time step where we have observations, add up the model-observation difference to be accounted for in the cost function; 2) we have daily CH4 emission data in this study, so we add up modeled CH4 emission of that day and compare it with measured fluxes.

To clarify our method, we will add this statement in section 2.3:

'Both the observed data and simulated results were rescaled to a daily emission unit for comparison.'

Table 4: Is there a quantitative criterion of well and poorly-constrained parameters? Why do you choose mean+-sd instead of median and q1-q3? For Figure 2 e,f, the posterior distributions of EBG-F are skewed so mean+-sd may reflect a wide range but the values are still relatively concentrated. However, you claim the parameter values of EBG-F in these two figures to be poorly constrained, which doesn't seem to be a fair judgement.

Thanks for bringing up this question.

The constraint of the posterior distribution is relatively speaking (Luo et al., 2016; Liang et al., 2018; Xu et al., 2006). For a uniform prior distribution, eg. in our study, a 'flat' posterior distribution would be regarded as not well-constrained. In general, we can focus on how much the posterior distribution differ from the prior assumption. Mean, median, variance, and quantiles are all taken into consideration.

We agree for a skewed distribution, mean (sd) reflect a wider range than median (q1q3). And definitely, for figure 2e and f, it is inappropriate to say parameter values of EBG-F is poorly constrained.

To avoid confusion, we will change our wording to be descriptive ('bell-shaped') instead of intuitive ('well-constrained') in Table 4 and rephrase the corresponding paragraphs:

'Of the three ebullition-related parameters used only in the EBG approach, when assimilating only the CH4 emission flux data, $V_{maxfraction}$ (maximum fraction of volume occupied by bubbles) was **constrained with a unimodal shaped posterior distribution** (Fig. 2g), f (CH4 mixing ratio in bubbles) **was edge hitting with a marginal distribution downward** (Fig. 2e), and bubprob (probability that a bubble will get trapped at a certain soil layer) was **constrained with a wide, slightly domed distribution** (Fig. 2f).'

Figure 4: How does Tveg affect the uncertainty range of ECT?

Thanks for the question. The effect of parameter Tveg is minor to the uncertainty range of ECT, mainly because at this experiment site (and typically the case in northern peatlands), more than 97% percent of root is distributed in less than 30cm depth. While the deeper layer CH4 concentration is the hardest component for ECT to resolve.

L440: As mentioned in Introduction, the ECT is assumed to overestimate ebullition, which is the opposite of the finding here. This needs more explanation.

Thank you to pointing it out! That sentence in the introduction was ambiguous and need some more context.

Originally, we said:

'However, this approach ignores the possibility of a CH_4 bubble moving into a less saturated layer where it can subsequently dissolve and possibly be oxidized, potentially overestimating ebullition.'

Instead, we will revise as:

'However, this approach ignores the possibility of a CH_4 bubble moving into a less saturated layer where it can subsequently dissolve and possibly be oxidized, potentially overestimating ebullition **from deep, saturated layers**.'

Minor comments

Figure 4: the legend for observations should be dots with ranges.

Thank you! We will match the legend with dots and ranges in the revised manuscript.

Reviewer 2

In their manuscript "Evaluating alternative ebullition models for predicting peatland methane emission and its pathways via data-model fusion", Shuang Ma and co-authors present their evaluation of two different formulations of the ebullition process in methane components for Land Surface Models, the Ebullition Bubble Growth approach and the Ebullition Concentration Threshold approach. They evaluate these approaches against methane flux and concentration measurements from the SPRCUE site using an MCMC approach and find that the Ebullition Bubble Growth model is far superior.

The authors make a very convincing argument why the Ebullition Bubble Growth approach is superior, which solves one of the many issues that need to be addressed in making methane emission models more reliable.

I am very much impressed by the manuscript. It is superbly written and about ready for publication as it is, though I have found a (very) few minor details. I really must congratulate the authors, as I've never before seen a manuscript that was as good as this during the first round of reviews.

We appreciate your suggestions, please find replies after each paragraph in blue underlined fonts.

As mentioned above, the manuscript is nearly ready for publication, but there are a few minor things to be sorted out:

- Lines 160-162: Double "in this study" -- please remove one.

Thank you! Will remove the latter one:

In this study, however, all observed data we used were only from ambient plots (no infrastructures and no warming treatment) for our research goals and we did not explore the warming effects on CH_4 processes.

- Lines 181-183: Sentence starting with "Samples shallower than..." sounds slightly awkward, I suggest rephrasing it

Indeed! Will rephrase as:

'A perforated stainless-steel tube was inserted into the peat to collect samples within 0-25cm depth.'

- Line 357: I don't see any blue shaded areas in the plot, I only see blue lines. Reformulate.

Thank you for catching this mistake. Will change in the manuscript as:

'The blue lines are the parameter posterior distributions (PPDs) from the ECT model structure trained with CH_4 emission data (ECT_F). The green lines are the PPDs from the EBG model

structure trained with CH$_4$ emission data (EBG_F). The purple lines are the PPDs from EBG trained with both CH$_4$ emission and CH$_4$ concentration data (EBG_FC).'

- Line 452: The model is called ELM_SPRUCE, I assume, not ELS-SRUCE, as is written?

Yes. ELM_SRUCE is the name of model. Also, to be consistent with our previous word choice, we will replace 'Earth system model' with 'Land Surface Model (LSM)'

- Line 455: "...concentrations ... was still not...". Noun is plural, verb is singular. Better use "were" instead of "was".

Thanks! Will correct it.

References

Krüger, M., Eller, G., Conrad, R., & Frenzel, P. (2002). Seasonal variation in pathways of CH4 production and in CH4 oxidation in rice fields determined by stable carbon isotopes and specific inhibitors. *Global Change Biology*, *8*(3), 265–280. https://doi.org/https://doi.org/10.1046/j.1365-2486.2002.00476.x

Happell, J. D., Chanton, J. P., Whiting, G. J., & Showers, W. J. (1993). Stable isotopes as tracers of methane dynamics in Everglades marshes with and without active populations of methane oxidizing bacteria. *Journal of Geophysical Research: Atmospheres*, *98*(D8), 14771–14782. https://doi.org/https://doi.org/10.1029/93JD00765

Hodgkins, S. B., Chanton, J. P., Langford, L. C., Mccalley, C. K., Saleska, S. R., Rich, V. I., et al. (2015). Soil incubations reproduce field methane dynamics in a subarctic wetland. *Biogeochemistry*, *126*(1), 241–249. https://doi.org/10.1007/s10533-015-0142-z

Wilson, R. M., Tfaily, M. M., Kolton, M., Johnston, E. R., Petro, C., Zalman, C. A., et al. (2021). Soil metabolome response to whole-ecosystem warming at the Spruce and Peatland Responses under Changing Environments experiment, *118*(25), 1–11. https://doi.org/10.1073/pnas.2004192118

Holmes, M. E., Chanton, J. P., Tfaily, M. M., & Ogram, A. (2014). Global Biogeochemical Cycles. https://doi.org/10.1002/2014GB004951.Received

Hines, M. E., Duddleston, K. N., Rooney-Varga, J. N., Fields, D., & Chanton, J. P. (2008). Uncoupling of acetate degradation from methane formation in Alaskan wetlands: Connections to vegetation distribution. *Global Biogeochemical Cycles*, *22*(2), 1–12. https://doi.org/10.1029/2006GB002903

Walter, K. M., Chanton, J. P., Chapin, F. S., Schuur, E. A. G., & Zimov, S. A. (2008). Methane production and bubble emissions from arctic lakes: Isotopic implications for source

pathways and ages. *Journal of Geophysical Research: Biogeosciences*, *113*(3). https://doi.org/10.1029/2007JG000569

Chanton, J. P., Glaser, P. H., Chasar, L. S., Burdige, D. J., Hines, M. E., Siegel, D. I., et al. (2008). Radiocarbon evidence for the importance of surface vegetation on fermentation and methanogenesis in contrasting types of boreal peatlands. *Global Biogeochemical Cycles*, *22*(4), 1–11. https://doi.org/10.1029/2008GB003274

Keller, J. K., Weisenhorn, P. B., & Megonigal, J. P. (2009). Humic acids as electron acceptors in wetland decomposition. *Soil Biology and Biochemistry*, *41*(7), 1518–1522. https://doi.org/10.1016/j.soilbio.2009.04.008

Itoh, M., Ohte, N., Koba, K., Sugimoto, A., & Tani, M. (2008). Analysis of methane production pathways in a riparian wetland of a temperate forest catchment, using δ13C of pore water CH 4 and CO2. *Journal of Geophysical Research: Biogeosciences*, *113*(3), 1–16. https://doi.org/10.1029/2007JG000647

Verry, E. S., Brooks, K. N., Nichols, D. S., Ferris, D. R. & Sebestyen, S. D. Watershed Hydrology, in Peatland Biogeochemistry and Watershed Hydrology at the Marcell Experimental Forest (eds Kolka, R. K. et al.) 193–212 (CRC Press, Boca Raton, FL, 2011).

Liang, J., Xia, J., Shi, Z., Jiang, L., Ma, S., Lu, X., et al. (2018). Biotic responses buffer warming-induced soil organic carbon loss in Arctic tundra. *Global Change Biology*, *24*(10), 4946–4959. https://doi.org/doi:10.1111/gcb.14325

Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E. A., Finzi, A., et al. (2016), Toward more realistic projections of soil carbon dynamics by Earth system models, *Global Biogeochem. Cycles*, 30, 40– 56, doi:10.1002/2015GB005239.

Xu, T., White, L., Hui, D., & Luo, Y. (2006). Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochemical Cycles*, *20*(2), 1–15. https://doi.org/10.1029/2005GB002468