

## 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 convenience 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 'CH<sub>4</sub> 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 CH<sub>4</sub> 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 CH<sub>4</sub> 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 CH<sub>4</sub> 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 CO<sub>2</sub> x H<sub>2</sub>O (LI-7500A; LiCor Inc., Lincoln, NB) and CH<sub>4</sub> (LI-7700) infrared sensors were used to measure change of CH<sub>4</sub> concentration. A typical measurement interval lasts approximately 400 seconds. Total CH<sub>4</sub> 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 CH<sub>4</sub> emission pathways. The porewater CH<sub>4</sub> concentration profile, however, is quite static (see figure 4), as it represents the amount of CH<sub>4</sub> in each soil layer instead of the dynamic fluxes. Our study offered an alternative method toward estimating ebulliated, diffused, and plant-mediated CH<sub>4</sub> fluxes by fusing vertically resolved mechanistic model with porewater CH<sub>4</sub> 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 CH<sub>4</sub> emission data alone, CH<sub>4</sub> concentration profile further constrain model structures and parameters, that potentially improve modeled emission pathways.

Eq 1: Why were f<sub>stp</sub>, f<sub>ph</sub>, f<sub>red</sub> 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. f<sub>stp</sub> is a soil temperature scalar and CH<sub>4</sub> production is very sensitive to this parameter value. The equation for f<sub>stp</sub> is in equation 9 in Ma et al. (2017) and shown below. We used Q10pro in data model fusion (parameter calibration).

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

where  $T_{\text{soil}(t)}$  is the hourly soil temperature and  $T_{\text{optpro}}$  is the optimum temperature for CH<sub>4</sub> 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  $f_{\text{pH}}$  and  $f_{\text{red}}$  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 CH<sub>4</sub> 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 CH<sub>4</sub> can reflect some of the information on the effects of pH and redox potential to methane production. We kept  $f_{\text{pH}}$ , and  $f_{\text{red}}$  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 CH<sub>4</sub> 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 CH<sub>4</sub> emissions data, we would be able to compare modeled-observed CH<sub>4</sub> 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 CH<sub>4</sub> emission data in this study, so we add up modeled CH<sub>4</sub> 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 CH<sub>4</sub> emission flux data,  $V_{\max}$ fraction (maximum fraction of volume occupied by bubbles) was **constrained with a unimodal shaped posterior distribution** (Fig. 2g), f (CH<sub>4</sub> 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<sub>4</sub> 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<sub>4</sub> 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.