Referee report of revised BG manuscript:
„Evaluation of Wetland CH_4 in the JULES Land Surface Model Using Satellite Observations“
by Robert J. Parker et al.

In this manuscript it is presented an evaluation of modeled wetland CH_4 emissions using the JULES Land Surface Model against atmospheric CH_4 concentrations from GOSAT satellite observations. For this comparison, the surface CH_4 emissions produced with ensemble simulations by JULES, are used in the global chemistry transport model TOMCAT in order to remove the non-wetland methane emissions from the modeled methane emissions.
A total of 24 ensemble members were produced using JULES, and those ensembles are a combination of varying atmospheric forcing data, vegetation, temperature dependence of soil-producing methane and wetland extent configuration.

The simulation results are at global scale; however, this work focuses mostly in tropical areas. This work provides a step forward in the evaluation of modeled CH_4 emissions in tropical wetland areas, a region of great interest in the global methane budget and of challenge for models. The study also offers a methodological consistent approach (ensemble) that was followed to analyze the performance of a land surface model under different criteria, providing key findings for other model studies.

The performance of the modeled emissions by the land surface model is considerably improved by including the SWAMP wetland extent. This approach was used to improve/deal with wetland extent in land surface models and brings an additional step forward to a known issue in the CH_4-wetlands modeling community: modelling wetland extent. It would be interesting to evaluate such performance in the more challenging boreal regions.

The original manuscript was reviewed by two referees who provided comments that were well responded by the authors. After reading the author comments, I recommend this work to be published in Biogeosciences.

Below, I report a summary of the general points discussed in the revision of this manuscript and provide own minor comments in the hope they can further support the improvement of this manuscript during the peer review process:

1) The considerations taken for non-wetland CH_4 emissions in the TOMCAT transport model and associated errors in the wetland CH_4 emissions
This comment was done by the two reviewers. The authors responded that several mitigation steps were taken including using model versions that have been previously used in other published works and that were consistently evaluated well against priors in CH_4 inversion exercises. Additionally, the authors suggest that the uncertainties generated in the TOMCAT atmospheric transport and chemistry model are likely larger than those introduced by not considering non-wetland emissions in JULES, and that quantitatively the uncertainties in emissions from e.g., biomass burning have shown to be much smaller than the uncertainties in wetland methane emissions, hence not strongly interfering in the modeled emissions.

The authors added a paragraph to discuss the impact of the assumptions taken, and performed an additional analysis in three regions of the world that are not typically dominated by wetlands. A comparison of detrended methane seasonal cycles between the satellite and model data showed a good agreement between both data sets, thus evidencing that non-wetland emissions can also be indirectly well constrained by the model.

I found that this approach to respond this point was well taken and well evidenced.

2) On scaling annual emissions to the Global Methane Budget
The authors scaled the annual emissions of 2000 to 180 Tg CH_4 yr^{-1} to keep consistency with the reported value in Saunois et al., 2016 for wetland emissions. The authors argue that this step is necessary because the wetland area masked by SWAMPS in the JULES emissions is smaller giving as a result “unrealistically low methane emissions”. The authors included a paragraph to clarify this step requested by referee #1.

However, in this added paragraph by the authors they mention: “the time series of annual wetland emissions of **each** ensemble member is separately scaled …”. Wouldn’t this scaling procedure needed to be done **only** for the ensemble members that were masked with SWAMPS? And not in **each** ensemble member. This is unclear and needs to be indicated more precisely or the added paragraph need to be moved to section 2.3.4

Also, it will be good to add a quantity to the expression “unrealistically low methane emissions” to give further validity to this approach in this particular set up.

3) Comparison of GOSAT vs JULES
Referee #2 points out the challenges in comparing GOSAT data to model surface emissions due to the potential of the satellite data to account for other methane sources that are not necessarily located in the region where the satellite is passing by, and also in relation to the problems with the seasonal coverage for specific regions. The authors provided an acceptable response referring to the technique utilized in GOSAT which uses only shortwave infrared and hence is limited only to land areas that emit such wavelength (i.e. wetlands).

Regarding the comparison of the two data sets shown in Figure 5, it is shown the difference in the amplitude between wetland seasonal cycles from the model emissions to that of the GOSAT data. From this figure, it is clear that GOSAT generally provides higher CH_4 emissions (about by 15 ppb) than the JULES simulated values in all the presented regions.
The color bar indicates the correlation coefficient between these two data sets, but: how is it possible to obtain a higher correlation coefficient in areas where the difference between the data sets is larger and with higher uncertainty (given by the whiskers in the bars)?
For example, the Pantanal region shows a correlation coefficient higher than 0.8 and has the highest range of difference and uncertainty of all the regions, and the South Tropics and S.E. Australia shows a close to zero difference but a correlation coefficient below 0.2. Is this a numerical artifact because the amplitude of the seasonal cycle is small or because the apparent time shift (Fig. 4) between the seasonal cycles of JULES and GOSAT? This should be make clearer in the text.

Also, it should be indicated why this time shift between the two data sets is happening.

4) Regarding the atmospheric forcing data

The authors concluded that the WFDEI data performed better than ERA-Interim in terms of wetland extent. The WFDEI data is based on ERA-Interim but its precipitation field is biased corrected using observational data, hence it is expected this improvement especially in the tropical areas.

However, WFDEI data is only available until December 2016 (and ERA-Interim until August 2019, and suppressed by ERA5), hence limiting the period of simulations using that forcing data.

According to NCAR, WFDEI is only available until December 2016 (https://rda.ucar.edu/datasets/ds314.2/), and the simulations using JULES in this manuscript were done until 2017, how the authors obtained forcing year for 2017?

Also, if model simulations using WFDEI are limited until 2016 (or 2017), this should be mentioned in the manuscript and, what recommendations can the authors provide for further studies that require improved wetland extent in simulations for years 2017 onwards? A test with ERA5 which has an improved precipitation field compared to ERA-Interim might be a way forward.

Minor comments:

The minor comments from both referees were answered accordingly and changes were implemented in the revised manuscript.

Own few minor comments:

Figure 1 – This is rather a table and not a figure, and in row 3 replace the expression “q10” by “Q_{10}”

Section 2.3.3 - Will be good to explicitly know why 3.7 and 5.0 were used as Q_{10} values