

## Review on Monteil et al., egusphere-2024-3122

### General comments

This paper studies Nordic wetland CH<sub>4</sub> emissions from the LPJ-GUESS process model and the LUMIA atmospheric inverse model. In the example year (2019), they found good improvements with atmospheric CH<sub>4</sub> observations over Europe, modelled by FLEXPART Lagrangian particle dispersion model, after optimizations using GRaB-AM flux data assimilation framework, LUMIA, and the combination of both data assimilation systems. Although the authors did not fully couple the flux and atmospheric data assimilation systems, the results look promising. I found that the way they examined the process model estimates also using atmospheric transport models is valuable, which has not been a very common approach. Ideally, additional data assimilation (i.e. in addition to EC data, also assimilating atmospheric data) would improve regional emission estimates, providing improved fit to data to both EC and atmospheric data. This was not yet archived in this study, and possibly because the coupling was not done fully. I anyway look forward to see their development to fulfil the goal. In addition, improving uncertainty estimates from the process models and using such information in the atmospheric inverse models are advanced approach, which should be explored more. This will also improve source attribution of emission estimates from atmospheric inverse models, which is a hot-topic now. Therefore, this paper presents an important work that contributes significantly to the carbon cycle studies.

In general, I have a few points which could be discussed more.

- The assimilated data are different, not only the quantity but also regarding number of sites/observations. How much of the differences between the optimization approaches are due to this? What if you only assimilate atmospheric data from the stations that are close to the EC sites, do you think you can still archive same improvements?
- I understand that it is difficult to examine which processes contribute to the over- or underestimation of emissions from the atmospheric inverse model. However, now you have attempted to combine the information from the process model and its parameter estimation, could you further discuss what is the problem that causes the process model based estimates to be over- or underestimated even after assimilating the EC data?
- As you mentioned briefly, validation of this model seems to be challenging. After assimilating both EC and atmospheric data, there are not much validation data left. But what data are available and would be the optimal or good way to evaluate the results? How can you say that end results are (or will be) better than only assimilating one or the other?
- Based on your results, would you argue that we need both EC and atmospheric data to estimate regional wetland CH<sub>4</sub> emissions, or only one or the other is enough? Can you identify an optimal way of designing observation network – should we have EC and atmospheric measurements at the same locations, or mostly in different locations?

Please also revise following specific and technical points before being accepted for publication.

### Specific comments

Abstract: I am not sure if you need all the references. Could you try to write without references? If some spaces are left after revision, could you possibly add a bit more about results?

L3: “poor spatial and process resolution” and L413: “process resolution”

I am not sure what you mean by “process resolution”. Could you rephrase a bit?

L5: “leading to more realistic emission estimates.”

Would it be better to say e.g. “realistic emission distribution”? You said earlier that inversion approach can provide robust total emissions.

L11-13: “In this study, we used the LUMIA regional atmospheric inversion system (Monteil and Scholze, 2021) to confront wetland emissions from the GRaB-AM approach to atmospheric CH<sub>4</sub> measurements in Europe. We then perform inversions using the information from GRaB-AM as prior.”

Does the second sentence repeating what is said in the first sentence? Could you consider to merge these sentences?

L23: “for the global methane budget”

Do you perhaps mean global natural methane budget? On overall, I think anthropogenic emissions contribute more than tropical wetlands.

Section 2.1.2

Please provide information about EC sites in LUMIA domain. Although a full list can be found in Kallingal et al. (2024), it is reader friendly to repeat the information also in this paper.

L108-109: Why do you use different ranges for prior and posterior ensembles? How are these percentages decided?

L160-161: CAMS provides two inversion estimates, one using surface air-samples, and another using also satellite data. Which product did you use?

L171-172: “All the emissions were regridded from their original resolution to the 0.25°, daily resolution of the FLEXPART footprints.”

What method did you use in case original resolution is coarser than 0.25° or daily?

L175: “influence on observations”

I guess here you mean atmospheric observations. This is true, but would it be better if you say e.g. “contribution to seasonality of total emissions” so that it is connected to emissions rather than atmospheric states.

L178: “the fact that the observation network is relatively dense in Northern Europe”

It is probably not only about how the dense the observation network is, but also locations and surroundings of the sites. If all the sites are close to anthropogenic sources, the constraint on wetland sources will be weak. Please clarify this point a bit more.

L187: “they are set proportional to the absolute value of the emissions”

Can you specify how much?

L198-199: Did you use all the data available, or did some filtering based on observation quality, quantified by flags? From which dataset did you collect these from?

L201-202: ICOS data provides various measurement uncertainty. What variable did you use? 20 ppb measurement uncertainty sounds rather high. What is the rationale for the choice?

Section 3.1: Please consider modifying “for each emission category” to “for wetland emission category” in L227 if you wish to describe only about wetland emission uncertainties in Section 3.1. Please anyway add somewhere in text about description of how you calculated uncertainties for non-wetland emission category.

L214-215: From the description here, I assumed  $\epsilon_{mod}^i$  changes in time as  $n_{obs}$  vary weekly, but in Table 2, only a single “model error” value is given for each site. Could you explain how you end up with one value per site?

L321-323: How much uncertainties did you put to the background concentrations? What makes SMR and BIR special that  $y_{bg}$  explains 100% of observed mixing ratios? If it only matters on several days in the end of year, can it explain the large reduction that starts already in the beginning of November?

L350: “the way they did.”

I do not quite understand this last part. Maybe the sentence can be slightly revised?

L360-363: Did you take into account wetland area (i.e. how large are wetlands) within the grid cell? If so, did you take the wetland area information from LPJ-GUESS?

L370: I see the largest reduction in RMSE in LPJ-GUESS-opt (site) rather than LUMIA-Lpost.

L371: Zarnekow has observed emission peak in August, where models failed to capture. This is probably the main reasons why RMSE did not improve. What is this associated with, and why models failed to capture the peak?

L345: “overestimating CH<sub>4</sub> emissions” → “overestimating regional CH<sub>4</sub> emissions”?

L434-439: Is it only due to lack of emissions from wet mineral soils? How about overestimation of wetland emissions in regions where EC data are not available, e.g. due to wrong representation of wetland area, or biases in climate forcing data?

L443-444: “lower temporal resolution of the methane emissions”

I do not fully agree. Do you mean temporal variations of methane emissions are weaker in the sense that it does not have strong diurnal cycle? Please rephrase.

L476-479: How are these estimates compared to other studies? Is it a common feature that emissions are reduced from process-based model estimates after optimizations (either by assimilating EC or atmospheric data)? If so, does it mean that some significant improvement is needed in the process models related to descriptions of processes? You may combine with last paragraph of 4.3 and expand discussion a bit more? I would also be interested to know about seasonality – how is it compared to other models/estimates?

L480-481: I guess ideally, you’d have even better fit to EC data with atmospheric inversion. Why do we fail to do so?

L501-505: Could you kindly elaborate more on what you mean by the complexity of the vegetation model (i.e. complexity in what sense) and “diagnostic models” (i.e. what are you going to diagnose)?

L506-508: As you showed in this study, anthropogenic emissions play an important role in estimation of regional CH<sub>4</sub> fluxes in Europe as their magnitude is larger than wetland emissions. So this development will be crucial. What are the uncertainties in anthropogenic CH<sub>4</sub> emissions, and how large/small are they compared to the wetland emission uncertainties? If you implement the information, what would be the consequent results? Is there any atmospheric inversion studies (possibly of CH<sub>4</sub>) that employ different uncertainty estimates for different sources?

Figure 1.

- Please add latitude/longitude coordinate labels of the grids.
- Please check coordinates of the EC sites. I wonder what is a site on the Baltic sea...?
- I do not see Siikaneva with green star. Is it just overlapped with red star? Please make it clear (also for Zarnekow).
- Please add/rephrasing the caption a bit to make it clear what type of observations you are talking about:
  - LUMIA inversion domain (cyan grid), the position of the **atmospheric observation sites** used in the LUMIA inversions... The red dots mark the position of the **eddy-covariance** sites used in Section 3.3.1.
- Labels: I would prefer to use “sites” than “obs”, and modify “CH<sub>4</sub> obs” to e.g. “Atmospheric sites” to make it clear (EC also measures CH<sub>4</sub>).

Figure 2:

- Are these regional emissions for the LUMIA or Nordic domain?
- Unit in the y-axis should be Tg CH<sub>4</sub> day<sup>-1</sup>, with 4 being subscript.
- What is bmb?
- Could you kindly add wetland emissions used in LUMIA-Lpost inversion?

Figure 3:

- Is this annual averages? Please specify in caption.
- The spatial resolutions in the panels seem to be different. Please consider using the same resolution in all.
- Why the maximum colour scale for wetland sources does not have arrow as in others? Is it on purpose?

Figure 4:

- Unit in the figure on left is missing.
- I guess the colour label on the right figure should not be emissions, but emission uncertainties.

Figure 5: bottom panel: please use same labels as in the top panel

Figure 6: Is it possible add uncertainty ranges?

Figure 8: Do I understand it correctly that wetlands LUMIA-Lpost panel (left middle) show differences compared to LPJ-GUESS-unopt? How do the differences between LUMIA-Lpost and LPJ-GUESS-opt look like?

Figure 9: Could you use different colour for LIMIA-Lpost? It's a bit hard to distinguish it from LPJ-GUESS-unopt.

Figure 10:

- Are these observed values hourly data, and are they all assimilated in LUMIA?
- How about plotting observations also as weekly averages? It is now hard to compare between observations and modelled values as they have different temporal resolution.

Table 1:

- Please check resolution columns. Temporal/spatial resolution columns seem to be mixed up.

- Column on “climatological” can be filled with “no” also. I wonder if that matters here, though, as the study is only for one year.
- Why spatial(?) resolution of geology and termites are missing?

Table 2:

- Is the values “model error”  $\epsilon_{mod}^i$ , i.e. not the quadratic sum of measurement and model uncertainties, i.e. diagonals of R, but just model uncertainties? If so, could you also add e.g. range of measurement uncertainties as well?
- nobs seems to exceed number of observations per year (365 days x 7 hours (11:00-17:00 LT) = 2555). Could you explain (somewhere in text) why? Is the information about e.g. spin-up spin-down missing in the description of inversion setup?
- Please add units for latitude, longitude, elevation and inlet height.
- Lan et al. reference seems to be missing in Bibliography. Please also remember to add publication year for this reference.

Table 3:

- Do I understand it correctly that RMSE here means RMSE in differences between simulated and the EC measurements? Please make this clear. Could you also add the RMSE in LPJ-GUESS-unopt site simulation?
- Is it correct that negative values mean increase in RMSE?

Table 4: Could you also add biases?

### Technical comments

Unit “/year” should be written as year<sup>-1</sup>

In general, overuse of parentheses could be improved. Short ones are ok, but I would prefer to write them as proper sentences in many cases. Please consider revising.

Please check reference styles. For example, “(e.g. Rayner et al. (2005))” should be written as “(e.g. Rayner et al., 2005)”.

don’t / doesn’t / didn’t → do not / does not / did not

L109: “time” → “times”?

L147: Please provide references to FLEXPART.

L166: The term super-category is only used here. Can you just say categories?

L173: Please mention Figure 2 first.

L174: “that has exhibits”: there are two verbs?

L186:  $\sigma_w \rightarrow \sigma_w$  is not defined yet. do you mean  $\sigma_{xc}$ ?

L191: “in-situ observations from 43 European in-situ and flask measurement sites, ”  
“in-situ” is repeated. Should it be rephrased as e.g. “continuous and discrete observations from 43 European in-situ sites”?

L192: “ICOS network of in-situ measurements”

I would prefer to say “ICOS network of atmosphere stations.”

L194: “and the free ,”  
Some words missing?

L232-233: “were estimated by through two” → “estimated by” or “estimated through”

L275: Figure 5 → do you mean Figure 6?

L292-293:  
“roughly a halving” → “roughly a half of”  
“mainly during the” → “mainly due to reduction of emissions during”?

L350: “two optimizations system” → “two optimization systems”

L356: “CH<sub>4</sub> fluxes from wetlands” → “CH<sub>4</sub> flux measurements (*or data or observations*) from wetlands”

L364: “improved for most sites” → “improved for most sites after optimizations”

L395-409: I would prefer to have this in Introduction as this is a kind of general description.  
L410: “two data-informed, in principle complementary:” → “two data-informed, in principle complementary, approaches:

L413: Remove “s” after “(LUMIA)”

L434: “and/or” → “or”

L476-477: Please add simulation name that estimated 2.5 Tg CH<sub>4</sub> year<sup>-1</sup>.

L511-501: “based on the LPJ-GUESS model Kallingal et al. (2024)” → “based on the LPJ-GUESS model by Kallingal et al. (2024)” or “based on the LPJ-GUESS model (Kallingal et al., 2024)”

L517: overestimation of the **wetland** CH<sub>4</sub> emissions?

L525: “emissions processes” → “emission processes”

Acknowledgement: Please add acknowledgement for the ICOS data. <https://www.icos-cp.eu/how-to-cite>