Reply to reviewer 1

September 5, 2026

A deep learning-driven emission estimator utilizing a mixture of experts for local wind speed situations applied to high-resolution methane imagery

Plewa et al.

The authors would like to sincerely thank the reviewer for their expertise and time in reviewing the first version of the manuscript and their helpful comments. We considered every comment carefully to improve the revised version of the manuscript. In the following we respond to the comments and suggestions of reviewer 1. The reviewer comments are marked as "RC1" and appear in boxes. The authors' responses are indicated with "AC".

Reviewer 1

RC1:

> This study demonstrated that the proposed MoE model outperformed previous end-to-end approaches based on metrics such as MPE and MAPE. While these metrics are useful for evaluating accuracy and bias, they do not provide information on predictive uncertainty. In particular, the MoE model introduces an intermediate expert-selection step, which may affect uncertainty propagation, so it would be better to include some metrics on uncertainty evaluation. Ideally the authors can add a table including accuracy and uncertainty comparison between the MoE model, the Plewa et al. (2025) model, and the Jongaramrungruang et al. (2022) model.

AC: We would like to thank the reviewer for pointing out that a direct comparison of the uncertainty evaluation between the different versions of the model was not present. We added a statement comparing the overall performance of the MoE network to its previous end-to-end trained version, stating that the previous version without any filtering provided better uncertainty estimates overall (page 20 line 396 to 399). We also added an additional statement regarding the performance on individual wind speed ensembles, where after filtering the quality of the uncertainty estimates is improved compared to the end-to-end trained version (page 20 line 408 to 412). Further, we included a statement in the context of future work that provides the possibility to obtain end-to-end trained uncertainty estimates without giving up on any of the advantages of the MoE design. We also would like to point out that we discussed negative effects of the uncertainty prediction of the MoE architecture on page 9 in line 200 and throughout the manuscript while evaluating the standardized residuals. We excluded Jongaramrungruang et al. [2022] from the comparison as there were no predicted uncertainties and did not include a table as we did not find a suitable summary statistic that would add to and/or summarize the current analysis.

RC1:

> The current Results and Summary sections would benefit from a more explicit discussion of future work. For example, on page 19 Line 361-371, the manuscript could include a brief discussion on how to improve model robustness under high wind-speed conditions. In addition, have the authors explored extending the proposed approach to other observational platforms, particularly satellite platforms? While the model is conceptually transferable, even a limited quantitative assessment of cross-platform performance would strengthen the evidence for its broader applicability, especially given that we now have a lot of available satellite observations.

AC: We included a statement at the end of the conclusion (page 21 line 433 to 441) that includes an explicit discussion of future work. This includes a statement on the extension of the approach to satellite platforms, which we had in mind during the development of this method. For further increases of the performance in high wind speed situations we do not have any concrete conceptual plans beyond the now mentioned improvement of the used base network architecture. It might just require more data, but could also be a systematic limit, as discussed in the paragraph in question.

RC1:

> Page 3 Line 80: "explicit" − > "explicitly"

AC: We were referring to make the assignment of the wind speed categories explicit, instead of them being implicit, as in our previous work.

RC1:

> Page 5 Table 1: Please capitalize the first letter of the table title and column/row headers

AC: We thank the reviewer for pointing out these mistakes, we capitalized the title and the headers of the table.

RC1:

> Page 6 Figure 1: The presentation of plume images is not immediately clear to me. Particularly, it is not obvious to me that the four images in the bottom-right panel all correspond to resulting scenes. I'd suggest adding colored or bordered panels for each image group (simulated plumes, background, resulting scenes with plumes) to highlight visual separation.

AC: We added colored frames around the different image groups and hope that this way the visual separation and relation between the groups is improved.

RC1:

> Page 8 Figure 2: Please add the full name of FFN and explanations of the nodes with the plus sign and times sign in the caption.

AC: We adapted the Figure and added a description including the now more precise fully connected (FC) layer of the network and included a description of the signs and their meaning. We hope that the figure is more clear now.

RC1:

> Page 8 Line 180 & Line 192: Why use a wind speed value range (+- 1m/s) instead of a percentage range? For example, I'd assume that it is much more likely for a 10 m/s wind speed estimate to be assigned to 8 m/s (-20%) than a 1 m/s to be assigned to 3 m/s (+200%).

AC: We would like to thank the reviewer for this question and we agree that a percentage range should be more suitable. The selection of the ± 1 m/s instead of a percentage error is in parts a technical limitation as our data is currently clustered by integer geostrophic wind speed forcings. However, we would also like to point out that every wind speed cluster contains a variety of different local realizations, over which we strive to have more information in future iterations. The router of the network agrees with the statement that a 10 m/s scene is more likely to be classified as a 8 m/s scene and we would in general too, but for application we would argue that the distribution of the expected external wind speed source would be the preferred source to determine such parameters.

RC1:

> Page 10 Line 248: "the performance" $->$ "the performance of wind speed classification", since there are two types of performance in discussion, it'd be better to clarify explicitly. I'd also recommend carefully reviewing the Results section to change any unambiguous terminology.

AC: Thank you for catching this ambiguous formulation from our side. We tried to keep the two performances separated and reviewed the manuscript again for other instances where we were unclear.

RC1:

> Page 10 Line 248-257: Back to my previous point, wouldn't a percentage range of surrounding wind speeds help mitigate the bias? For example, by providing a +-20% range you are telling the model that a 10 m/s wind speed estimate could possibly be 8 m/s in reality.

AC: Yes, it should be possible to reduce the bias, assuming you meant the emission estimation bias, by adopting the distribution of the trainings data, however this also impacts the overall performance of the expert and makes the performance kind of more specific for the input distribution. This would make sense if one had a well characterized external input source of wind speed information and is expected to rely on it heavily. However, in our application scenario it seemed favorable to rely on the router for all higher emission scenarios and only fall back on external information for lower emission scenarios. This way the performance was very good and it would still be possible to achieve an improvement at low emissions, where the router starts to fail, by using external wind speed information. For other applications it might be beneficial for the overall performance to widen the distribution due to a large expected uncertainty at high wind speed situations and a poor router performance. This is a tuneable hyperparameter for training that depends on the expected uncertainty of the wind speed information, the performance of the individual expert and the performance of the router. Therefore, the choice of the distribution design during training is quite complicated and will be largely determined by the observed performance on real data, which we now address in the outlook at the end of the manuscript (page 21 line 433 to 441).

RC1:

> Page 11 Figure 3: The red 1:1 line is not mentioned in the caption. It would also be useful to add a fitted line with R2 marked in the plot.

AC: We added the information to the figure caption, we decided not to add a fitted line as it is essentially the $1:1$ line and it just makes the plot harder to read.

RC1:

> Page 13 Line 282: I think here the author meant MAPE rather than MPE? It's claimed that MAPE represents "the spread of our data"; and in Table 2 only MAPE in the filtered model is lower than the unfiltered one for all the three emission rates.

AC: We are sorry for the confusion, we were referring to change of the MPE, which is smaller when the low emission scenarios are filtered or corrected. The revised manuscript now refers to the change of the MPE.

RC1:

> Page 16 Line 303: A difference of 1 m/s between the correct wind speed and the external wind speed seems too optimistic to me. How was it determined?

AC: It was selected to illustrate the capabilities and the limitations of the experts and thus reflected the prior assumption included into the training of the experts. Therefore, it is more a technical choice than a realistic choice which would most likely be a percentage of the wind speed as you mentioned in your previous comments.

RC1:

> Page 17 Line 325: It would be helpful to explicitly clarify that, the comparison between the unfiltered and filtered models is intended as a diagnostic analysis to separate errors from wind misclassification from those related to emission estimation under different scene conditions, rather than as an indication that the filtered model would be preferable in operational settings, since we never know the true wind speed in practice. The same clarification applies to the "correct" and "corrected" configurations.

AC: We are sorry for not being precise enough in this statement. For the most part it shows the basic functionalities of the network design and confirms their potential functionality. We added a statement to emphasize this point: "..., and thus shows that the design properties of the network transition well into application.". Previously in line 306 we illustrated how the experts perform when they get scenes from the wrong wind speed classes. In our case we limited this study to the selected ± 1 m/s as it is the chosen stability of the experts, defined by the spread of wind speed forcings in the training data, and thus a further design goal of our network.

We obviously agree on the fact that we will never know the true wind speed, however, we would like to point out that this fact can be accounted for in the training of the experts. Thus, while it has a negative impact on the performance to use the wrong wind speed information this impact should be dampened by the selected training data distribution for the expert. For an operational application this would in fact mean that filtering the scenes should be the best performing option to optimize the sheer quality of the output, but this is in practice a consequence of the fact that most difficult scenes are filtered out, as we stated later in line 330. For low emission scenarios, where the accuracy of the router is very low, one is left with either filtering scenes that show deviations in the wind speed estimates or using the external information. While the latter should be more reliable at this point, leading to the best possible estimate for these cases, it results in a worse average performance overall compared to filtering these scenes out completely. However, for operational applications these decisions are complicated and their impact cannot be gauged at this point. Therefore, we would argue that it is beyond the scope of this paper that tries to first establish the method by presenting its general functionality in a controlled synthetic environment.

References

Siraput Jongaramrungruang, Andrew Thorpe, Georgios Matheou, and Christian Frankenberg. MethaNet – an AI-driven approach to quantifying methane point-source emission from high-resolution 2-D plume imagery. *Remote Sensing of Environment*, 269:112809, 02 2022. doi: 10.1016/j.rse.2021.112809.