

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2021-317-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-317

Anonymous Referee #2

Referee comment on "KGML-ag: a modeling framework of knowledge-guided machine learning to simulate agroecosystems: a case study of estimating N₂O emission using data from mesocosm experiments" by Licheng Liu et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-317-RC2>, 2021

General comments

This manuscript presents a new method for estimating N₂O flux from cropland. The inputs to the method are known fertilization rate, weather forcings, soil and crop properties. The method also requires initial concentrations of nitrate ions, ammonium ions, and water in the soil, and optionally CO₂ flux. The method employs gated recurrent networks organized in a hierarchical structure to mirror the time-dependence and causality present in the process. A process-based model provides pre-training data, and fine-tuning is done using observations from mesocosm experiments. The trained neural network models outperform the process-based model and many basic machine learning approaches.

The methodology employed is both novel and sound. The use of GRUs in hierarchical structures is well-justified and appropriate to the problem. The models have been well-validated, and various alternate choices for model architecture have been explored. I believe this work represents a substantive advance in modelling science. Below I list specific comments which I hope will serve to improve the manuscript.

Specific comments

- The use of the term "initials" confuses me. Upon first reading I thought it referred to the acronyms for various intermediate variables. I think it actually refers to the initial values of a sequence. Is this usage standard? If not, I recommend a different phrase such as "initial values" in place of the word "initials." Alternatively, clarify the meaning of the term in the manuscript.
- Another possible explanation for why KGML-ag2 better predicts IMVs but does not predict N₂O as well is that KGML-ag1 may learn to use the IMVs as a kind of extra hidden layer, encoding information relevant to N₂O predictions in them.

- Why not include KGML-ag2 in Figure 4? I can see simplifying the comparison by choosing only the best-performing model.
- Many standard deep learning models were included for comparison, but an LSTM was not among them. I would expect the LSTM to perform similarly to the GRU. I don't think it is crucial that an LSTM be included in this comparison. However, if the GRU outperforms an LSTM, it could provide further justification for choosing to use a GRU instead of an LSTM. Again, I could understand simplifying the comparison by including only one recurrent neural network.
- You tested two input combinations, IMVcb1 and IMVcb2, but it is not clear how that test informed the model development.
- The reason for evaluating slope and curvature in addition to N2O value could be stated more clearly.
- I recommend that the paragraph starting at line 194 be rewritten for clarity. First, data augmentation is a class of methods, not a single method. Second, Meyer et al. use copula-based models in particular to augment datasets. Do you use copula-based methods? The way this reference is cited suggests that you follow their approach. Third, do you randomly sample observed data, or synthetically generated data, or both? Do you randomly sample only the data which are hourly, e.g., air temperature, net radiation, N2O, CO2, and VWC? How is the daily value calculated from the sampled data? I did not find the answers to these questions to be clear from the text.
- How well does the model perform out-of-sample? Out-of-sample performance is mentioned in the introduction, but the discussion does not address it.

Technical corrections

- At line 239, Sec. 4.4 does not exist.
- At line 240, I believe this should refer to Fig. 1c and 1d, not 1b and 1c.
- Tables 1 and 2 have identical captions but different contents.
- Sections 4.1 and 4.2 are both entitled "Interpretability of KGML-ag."