Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2021-317-RC1, 2021

Comment on gmd-2021-317

Anonymous Referee #1

Referee comment on "KGML-ag: a modeling framework of knowledge-guided machine learning to simulate agroecosystems: a case study of estimating N_2O emission using data from mesocosm experiments" by Licheng Liu et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-317-RC1, 2021

Liu et al. presented a promising predictive framework that combined a process-based model (physical knowledge and pre-train dataset) and a machine learning model for agroecosystem N2O emission estimate. The modeling framework is robust and thoroughly validated. This work will be an important milestone towards a better understanding, monitoring, and predicting agroecosystem greenhouse gas emissions.

The paper is well organized and written. Below are some of my comments that may help elucidate the strength and limitations of the proposed KGML-ag framework.

- Robustness of physical (prior) knowledge

ecosys model plays a central role in guiding the ML model in terms of structure and providing a pre-train dataset. It will be important to discuss the structure uncertainty in ecosys N2O module, including e.g., underlying theories, major processes, difference/similarity to the classic leaky pipe type model (Davidson et al., 2000), and so on.

Again ecosys provides pretrain dataset, which has its own uncertainty and biases. It's worthwhile to at least show some ecosys model performance across various different conditions at agroecosystems. For example, does ecosys pick up the high-frequency signals (fluctuation) of CO2/N2O flux that are observed in the chambers data? If not, is that the reason why PGML-ag could not capture the high fluctuation of CO2/N2O emissions in the field?

- It’s not obvious which variables are used as inputs or intermediate variables and how that relates to the feature importance ranking. It will be better to show each variable in Figure 1. For example, W will be temperature and precipitation. Furthermore, feature importance analysis highlight NH3, H2, N2, O2, CH4, ET, CO2 are important variables that drive N2O emission (~ L230). It’s not clear in the main text, how this feature importance ranking helps the design of PGML-ag. What can we get out of this feature importance analysis?
- There is a lack of discussion on uncertainty in PGML-ag, which is fundamentally important for predictive modeling. Also, what about chamber measurements uncertainty?

L254 based on the structure of process representation in ecosys

Reference:

Davidson, E. A., Keller, M., Erickson, H. E., Verchot, L. V., & Veldkamp, E. (2000). Testing a conceptual model of soil emissions of nitrous and nitric oxides: using two functions based on soil nitrogen availability and soil water content, the hole-in-the-pipe model characterizes a large fraction of the observed variation of nitric oxide and nitrous oxide emissions from soils. Bioscience, 50(8), 667-680.