Knowledge Guided Machine Learning for Simulating Agricultural N2O Emission

Licheng Liu[1], Shaoming Xu[2], Zhenong Jin[1,3], Jinyun Tang[4], Timothy J Griffis[5], Kaiyu Guan[6,7], AlexanderLee Frie[5], Taegon Kim[1], Bin Peng[6,7], Xiaowei Jia[8], Yufeng Yang[1], Wang Zhou[6] and Vipin Kumar[2]

[1]Department of Bioproducts & Biosystems Engineering, University of Minnesota
[2]Department of Computer Science, University of Minnesota
[3]Institute on the Environment, University of Minnesota
[4]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory
[5]Department of Soil, Water & Climate, University of Minnesota
[6]Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana-Champaign
[7]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign
[8]Department of Computer Science, University of Pittsburgh

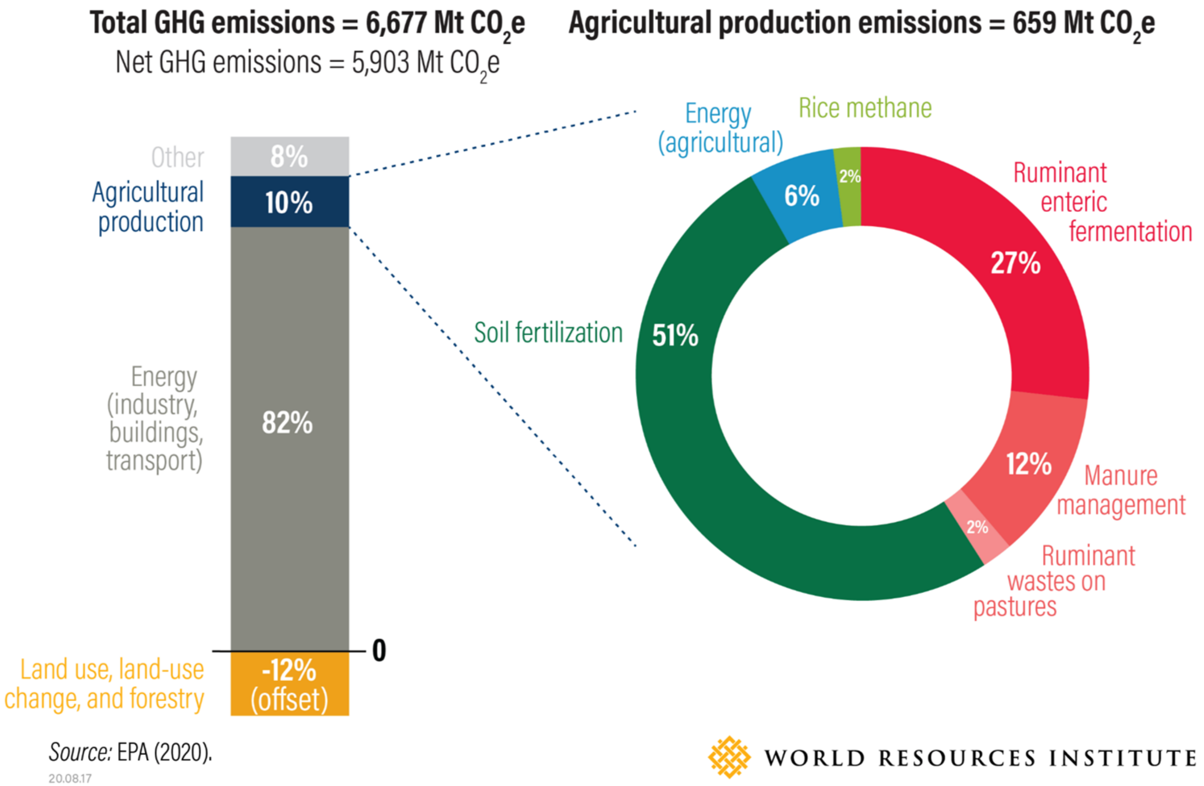
Contact: lichengl@umn.edu

AGU Fall meeting 2021

UNIVERSITY OF MINNESOTA
Driven to Discover®

Agriculture and Climate Change

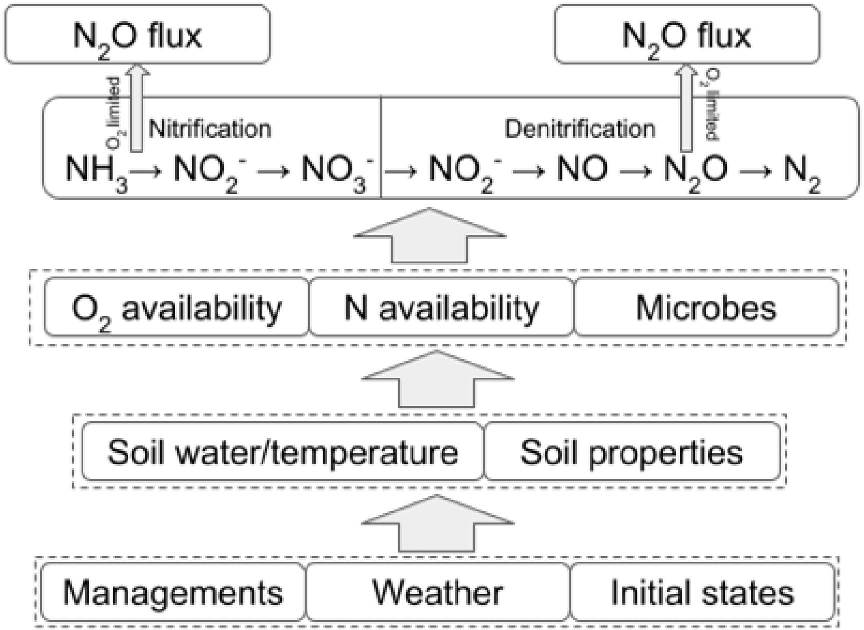
Agriculture contributes a quarter of global greenhouse gas (GHG) emissions that are causing climate change: **~14%** directly from agricultural activities and **~10%** through land use change.



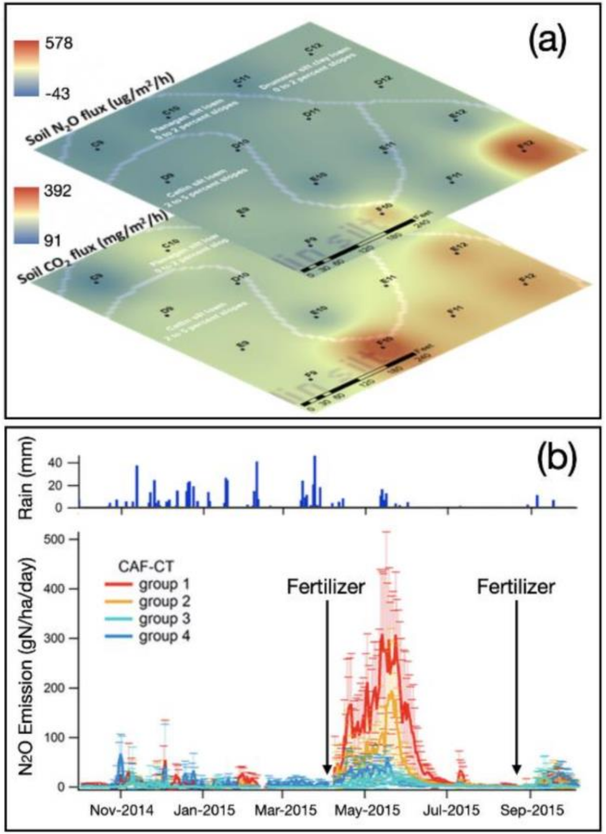
- ❑ Agricultural production accounts for **10%** of U.S. GHG emissions in 2018
- ❑ Fertilizer use accounts for 51% of the ag emissions, largely in forms of **N_2O, 265x more powerful than CO_2** as a GHG
- ❑ Over applying fertilizer also causes **water** & **air** pollution and **land** degradation

Why estimating N_2O is so hard?

Soil nitrous oxide (N_2O) emissions are highly variable in space and time due to dynamic controls by a range of biotic and abiotic factors.

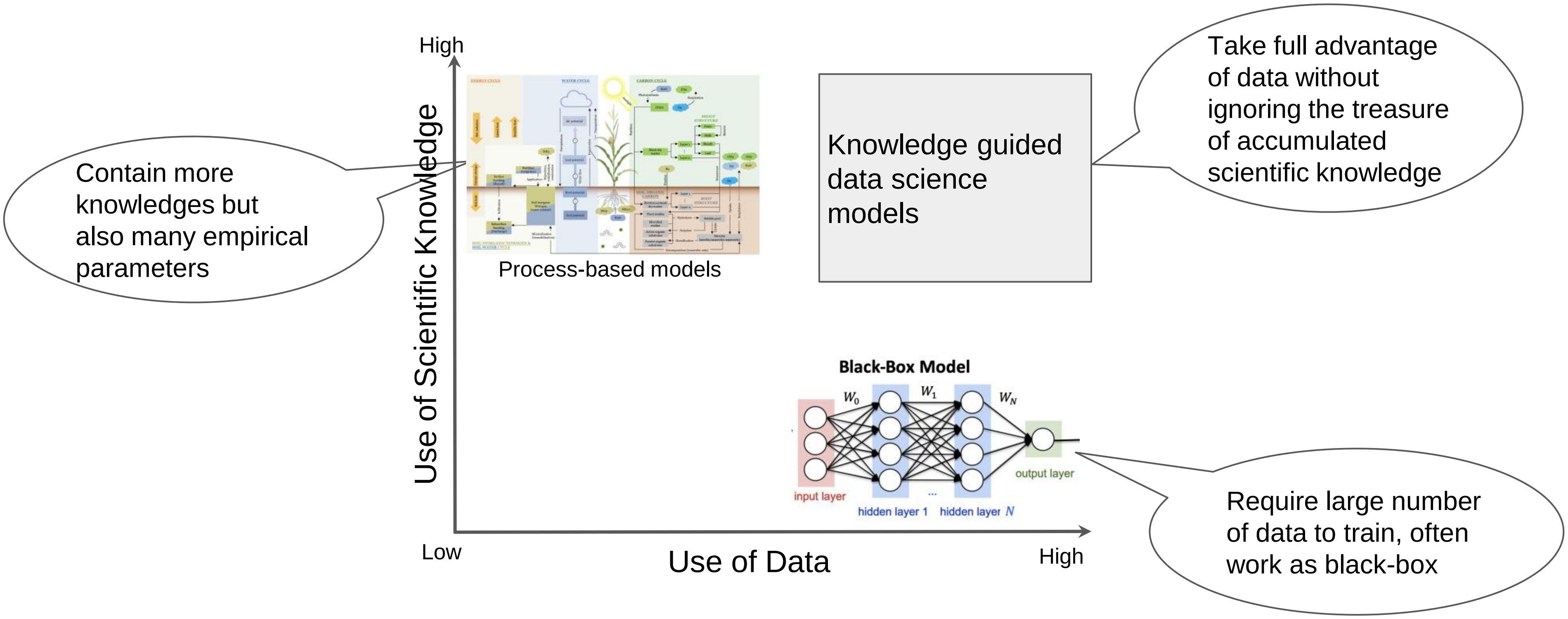


Hot spots, hot moment of N_2O fluxes

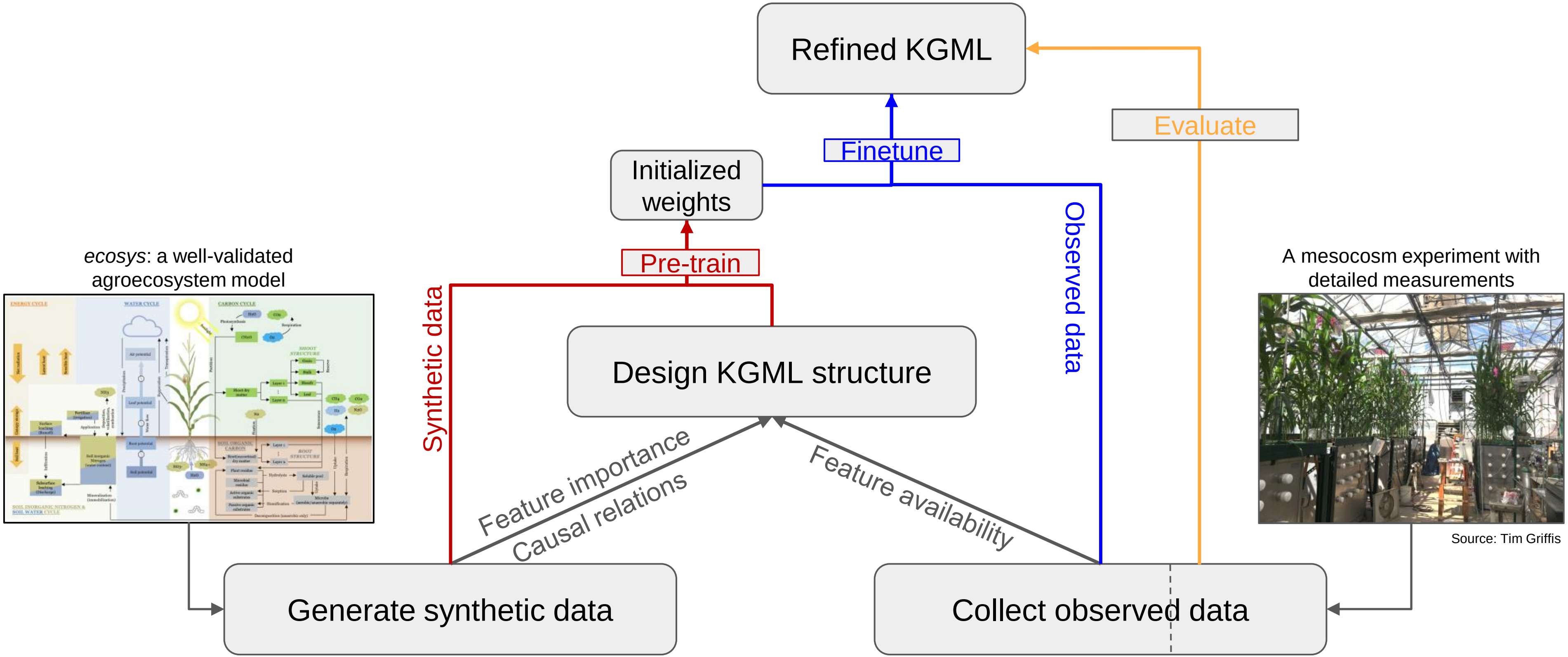


Waldo et al. 2019

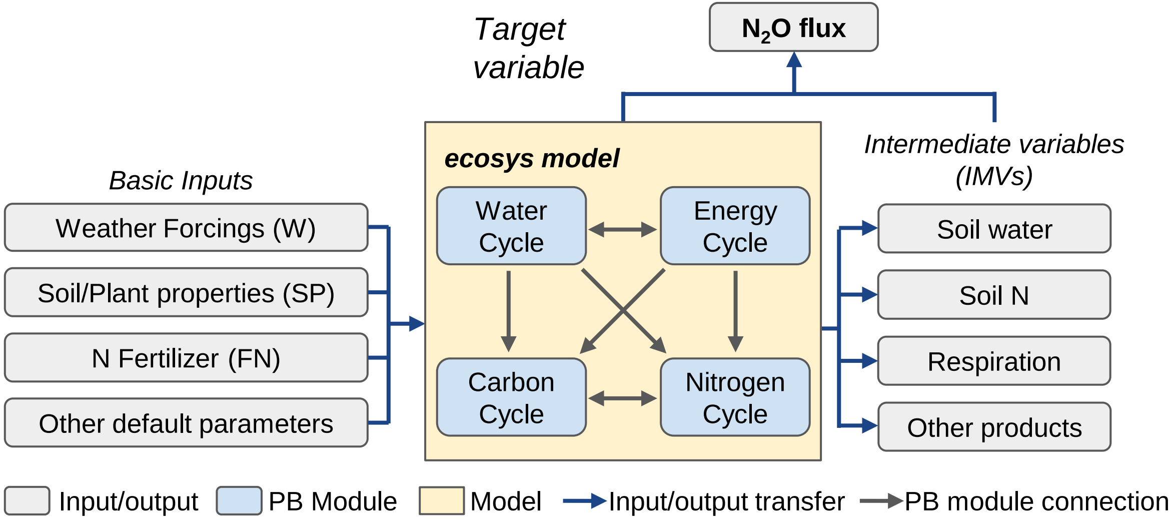
Opportunities with knowledge guided machine learning



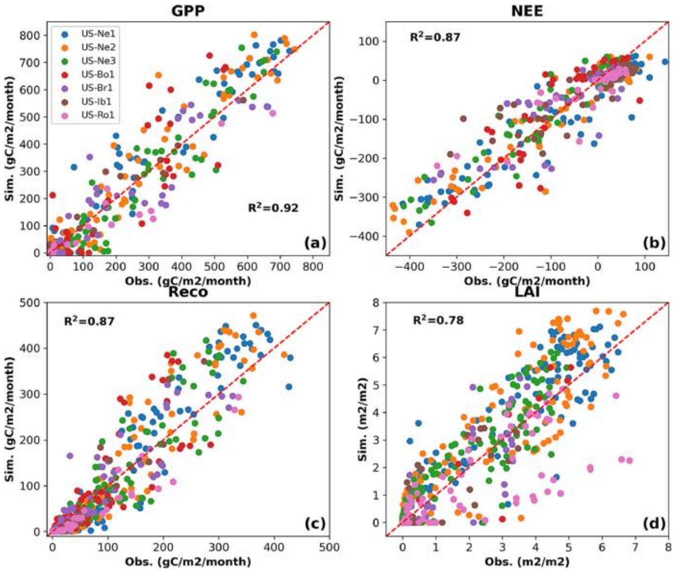
Experiments overview



Generate synthetic data from an advanced agroecosystem model, *Ecosys*



Validation at 7 US FluxNet sites for ag



Zhou et al. (2021)

<u>Ecosys simulation:</u>

- 99 random sites from Illinois, Indiana and Iowa
- 18 years simulations
- Over 4 million synthetic data samples

Observed N_2O fluxes from mesocosm experiments

Experiment setup:

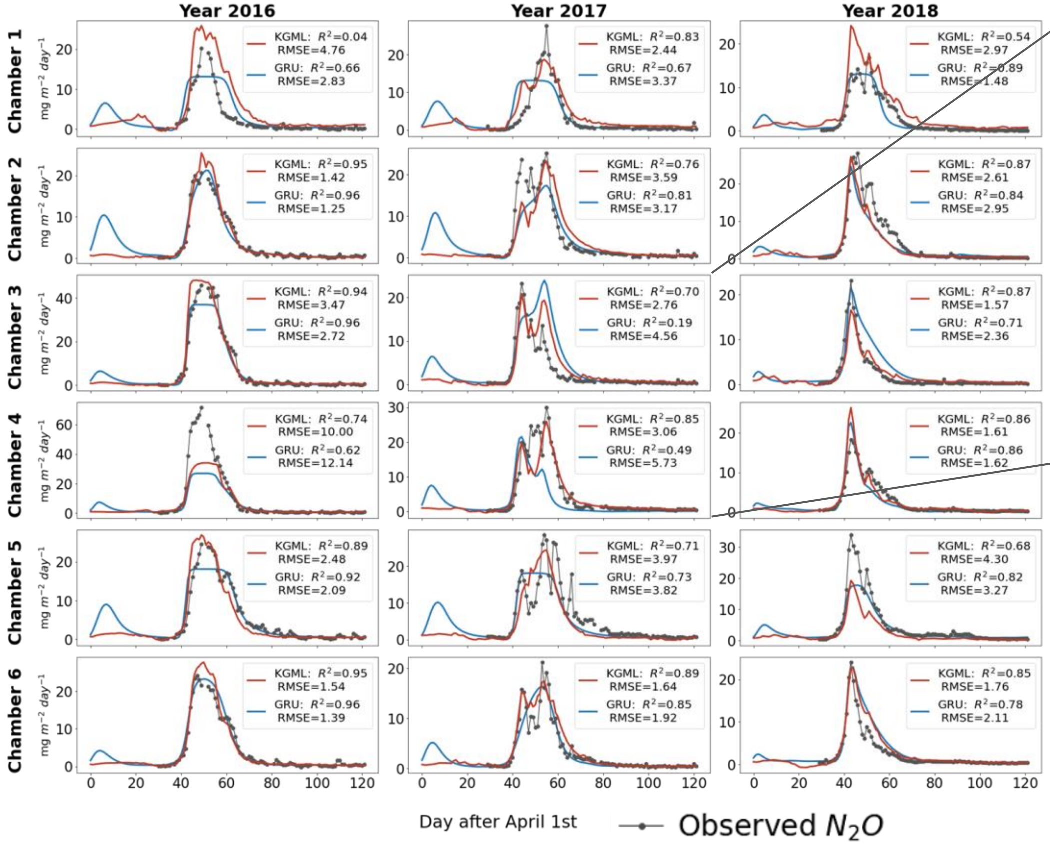
- Growing seasons during 2016-2018
- 6 chambers with different precipitation treatment
- N_2O flux was measured by Teledyne M320EU Analyzer in automatic chamber

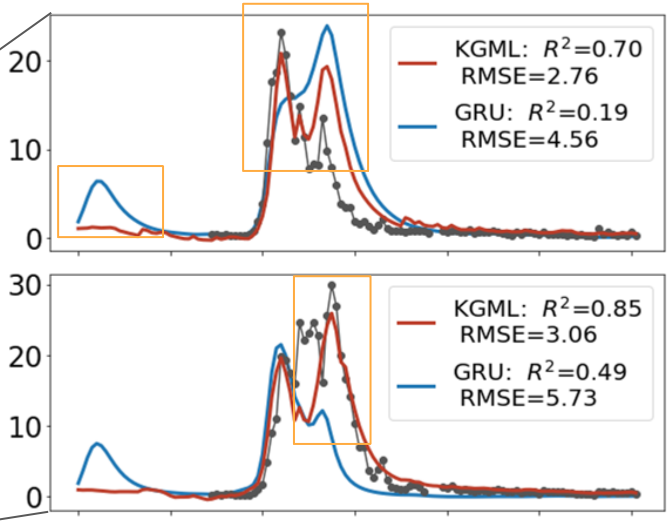
Available variables:

- Controlled weather conditions
- Hourly N_2O fluxes, CO_2 fluxes, and soil moisture at 15 cm depth
- Weekly soil [NO_3-], [NH_4+] at 15cm depth
- Management info: planting/harvesting dates, fertilizer application timing and rate

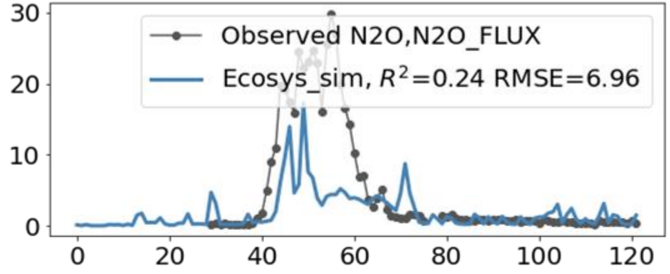
Source: Tim Griffis

The KGML model outperformed the pure ML model

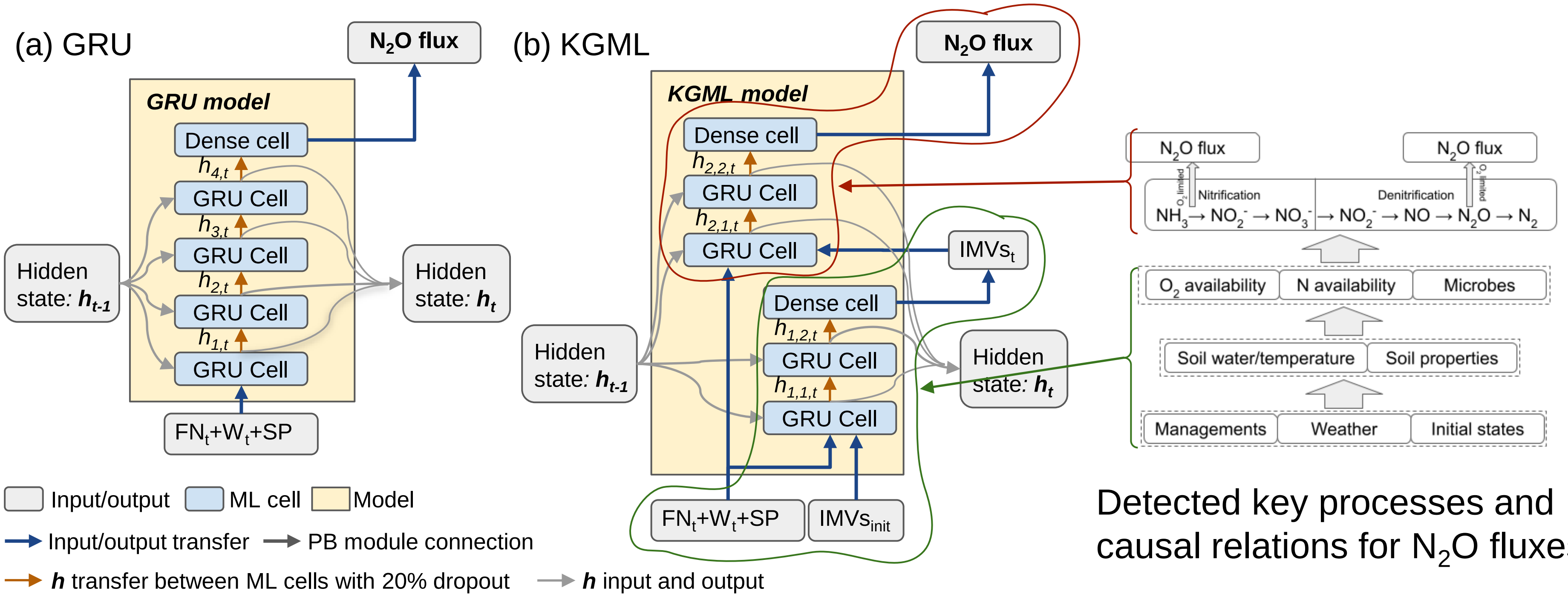




Same chamber-year Ecosys performance



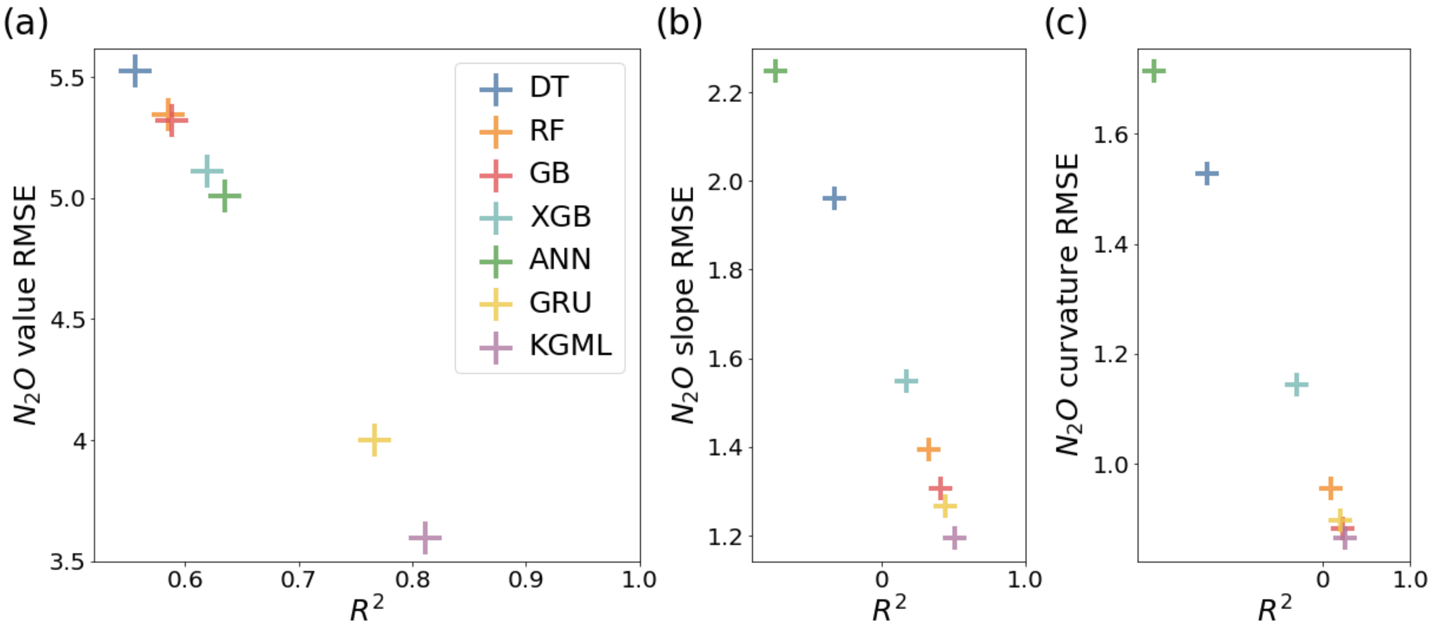
Develop KGML model based on causal relations and feature importance



Detected key processes and causal relations for N_2O fluxes

- GRU outperformed LSTM with its simpler structure in N_2O simulation
- GRU model was used to do feature importance tests
- Knowledge guided initialization and architecture constraints were applied

The KGML model outperformed the pure ML model



- ❑ KGML (purple) outperform all other ML models
- ❑ This is mainly because (1) pre-training using synthetic data, (2) knowledge guided architecture, (3) knowledge guided initial values

Liu, L., Xu, S., Jin, Z., Tang, J., Guan, K., Griffis, T. J., ... & Kumar, V. (2021). 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. *Geoscientific Model Development Discussions*, 1-29.

Conclusion

- ❑ We used 1) knowledge-guided initialization, 2) hierarchical architecture, and 3) initial values of intermediate variebles to develop the KGML-ag strucure for N_2O prediction
- ❑ The KGML-ag model has been tested on mesocosm experiment observations and can outperform all other pure ML models
- ❑ KGML reduced data demand significantly comparing to ML
- ❑ More N_2O flux data are needed for further improvement of KGML
- ❑ The structures are flexible and can be easily revised or transferred to other data/study (Another study of KGML-ag for Carbon is presented in AGU poster named: *Estimating the Autotrophic and Heterotrophic Respiration in the US Crop Fields using Knowledge Guided Machine Learning*)

Thank you so much for your interest!

If you have any questions, please feel free to contact
Licheng Liu: lichengl@umn.edu