Simulating N_2O Emission from Fertilized Mesocosm Using Knowledge Guided Machine Learning

Licheng LIU[1], Shaoming Xu[1], Zhenong Jin[2], Jinyun Tang[3], Timothy Griffis[4], Kaiyu Guan[5], Alexander Frie[6], Taegon Kim[1], Bin Peng[5], Yufeng Yang[1], Wang Zhou[5], and Vipin Kumar[1]

[1]University of Minnesota Twin Cities
[2]University of Minnesota-Twin Cities
[3]Lawrence Berkeley Natl Lab
[4]Univ Minnesota
[5]University of Illinois at Urbana Champaign
[6]University of California Riverside

November 24, 2022

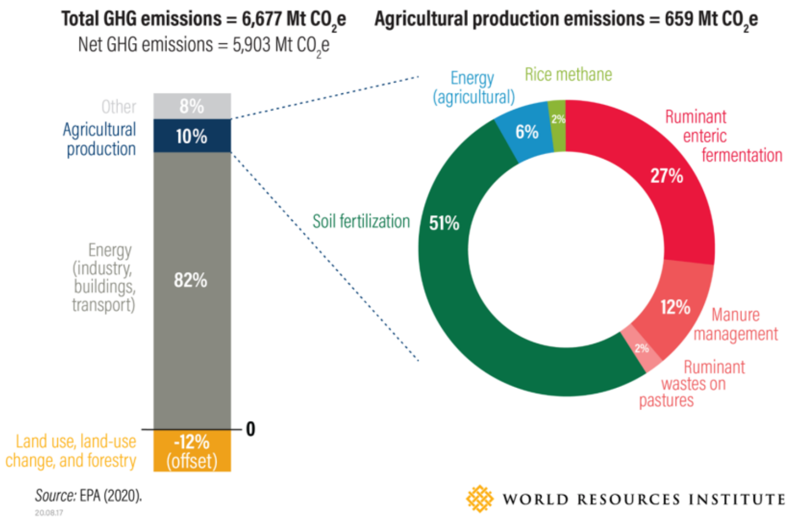
Abstract

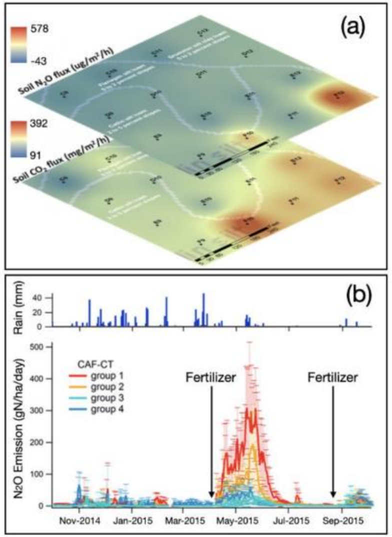
Nitrous oxide (N_2O) is one of the important greenhouse gases (GHGs), with its global warming potential 265 times greater than that of carbon dioxide (CO_2). About 60% of the anthropogenic N_2O emission is from agriculture production. To date, estimating N_2O emissions from cropland remains a challenging task because the related microbial origin processes (e.g. incomplete nitrification and denitrification) are controlled by a diverse factors of climate, soil, plant and human activities. In this study, we developed a ML model with physical/biogeochemical domain knowledge, namely knowledge guided machine learning (KGML), for simulating daily N_2O fluxes from the agriculture ecosystem. The Gated Recurrent Unit (GRU) was used as the basis to build the model structure. A range of ideas have been implemented to optimize the model performance, including 1) hierarchical structure based on variable causal relations, 2) intermediate variable (IMV) prediction and transfer, 3) inputting IMV initials for constraints, 4) model pretrain/retrain, and 5) multitask learning. The developed KGML was pre-trained by millions of synthetic data generated by an advanced PB model, *ecosys*, and then re-trained by observations from six mesocosm chambers during three growing seasons. Six other pure ML models were developed using the same data from mesocosm chambers to serve as the benchmark for the KGML model. The results show that KGML can always outperform the PB model in efficiency and ML models in prediction accuracy of capturing N_2O flux magnitude and dynamics. Besides, the reasonable predictions of IMVs increase the interpretability of KGML. We believe the footprint of KGML development in this study will stimulate a new body of research on interpretable machine learning for biogeochemistry and other related geoscience processes.

Knowledge Guided Machine Learning for Simulating Agricultural N_2O Emission

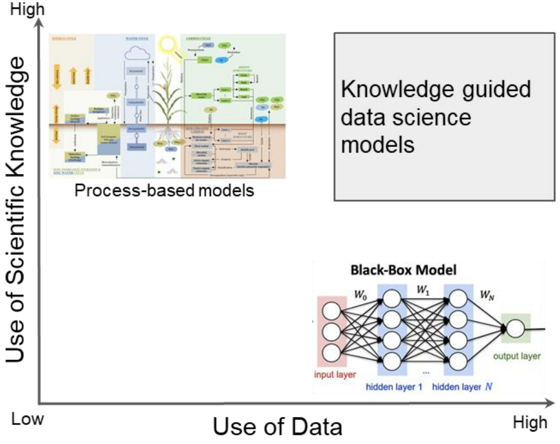
Presenter: Licheng Liu; lichengl@umn.edu;

Motivation:





Waldo et al. 2019



- ❑ Fertilizer use accounts for 51% of the ag emissions, largely in forms of **N_2O, 265x more powerful than CO_2** as a GHG
- ❑ Hard to estimate due to **hot spots, hot moment** of N_2O fluxes
- ❑ KGML model can take full advantage of data without ignoring the treasure of accumulated scientific knowledge

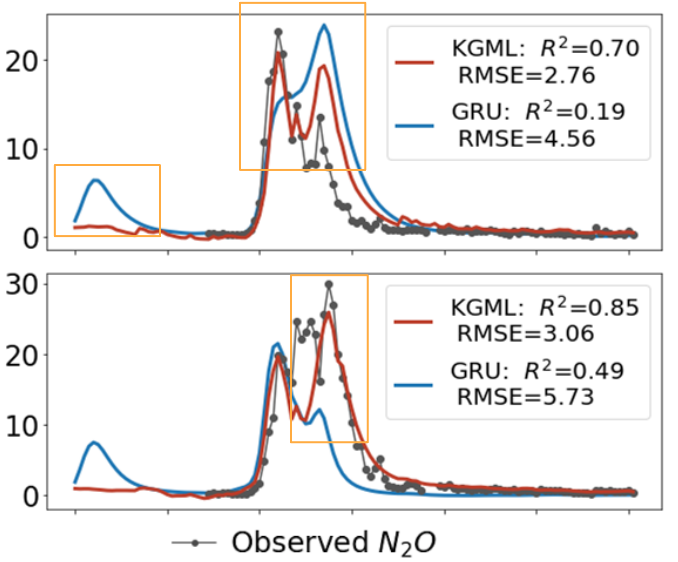
AGU FALL MEETING

UNIVERSITY OF MINNESOTA Driven to Discover®

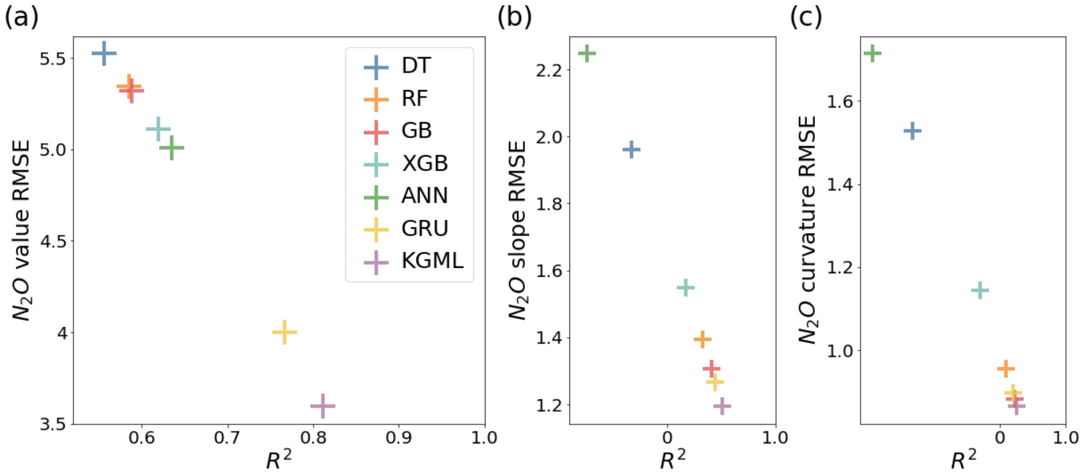
Knowledge Guided Machine Learning for Simulating Agricultural N_2O Emission

Presenter: Licheng Liu; lichengl@umn.edu;

Key results:



One example of KGML model comparing to GRU model in mesocosm experiment data



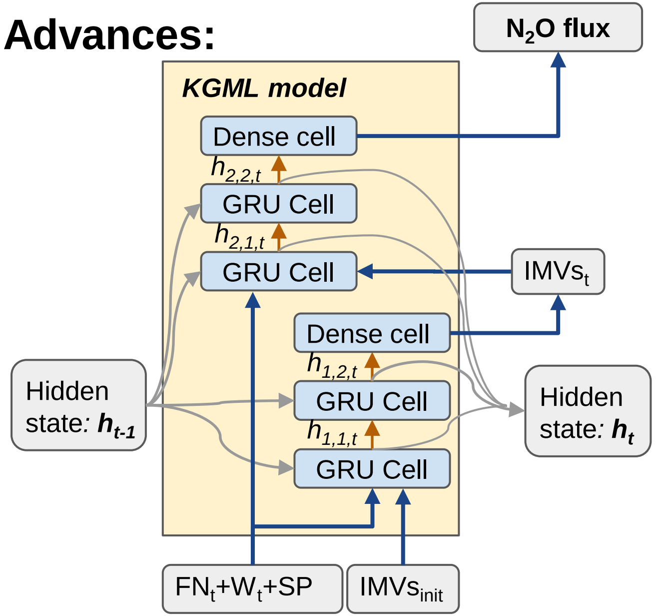
- ❑ KGML (purple) outperform all other ML models
- ❑ This is mainly because (1) pre-training using synthetic data, (2) knowledge guided architecture, (3) knowledge guided initial values

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.

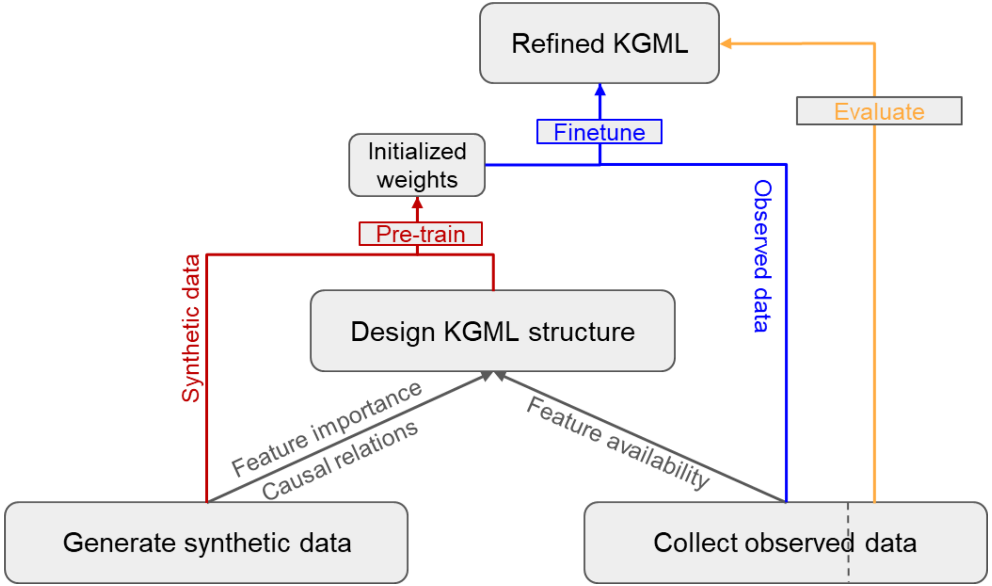
Knowledge Guided Machine Learning for Simulating Agricultural N_2O Emission

Presenter: Licheng Liu; lichengl@umn.edu;

Advances:



KGML model structure



KGML model development workflow

- ❑ High performance
- ❑ Low data demand
- ❑ Flexible structure
- ❑ Structure and workflow can be easily transfer to other similar geoscience tasks!

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.

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]

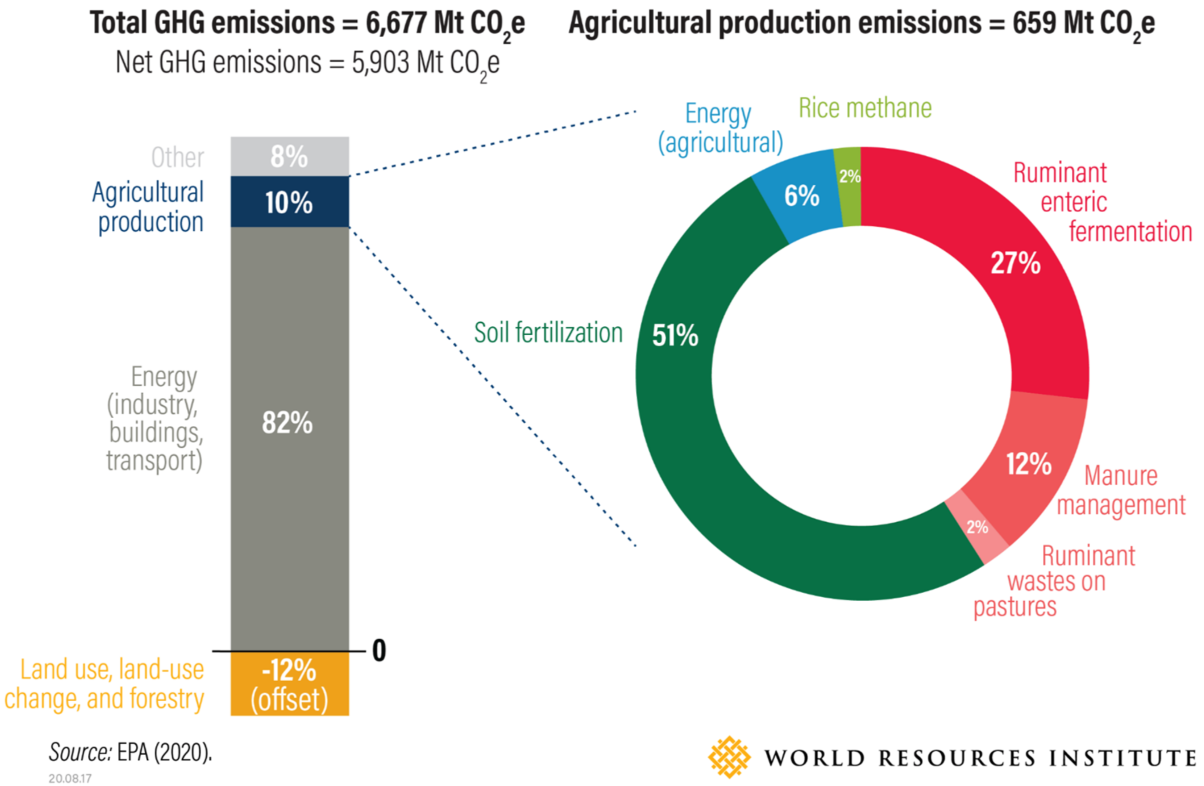
[1]Department of Bioproducts & Biosystems Engineering, University of Minnesota
[2]Department of Computer Science, University of Minnesota
[3]Institute on the Environment, University of Minnesota
[4]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory
[5]Department of Soil, Water & 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

Agriculture and Climate Change

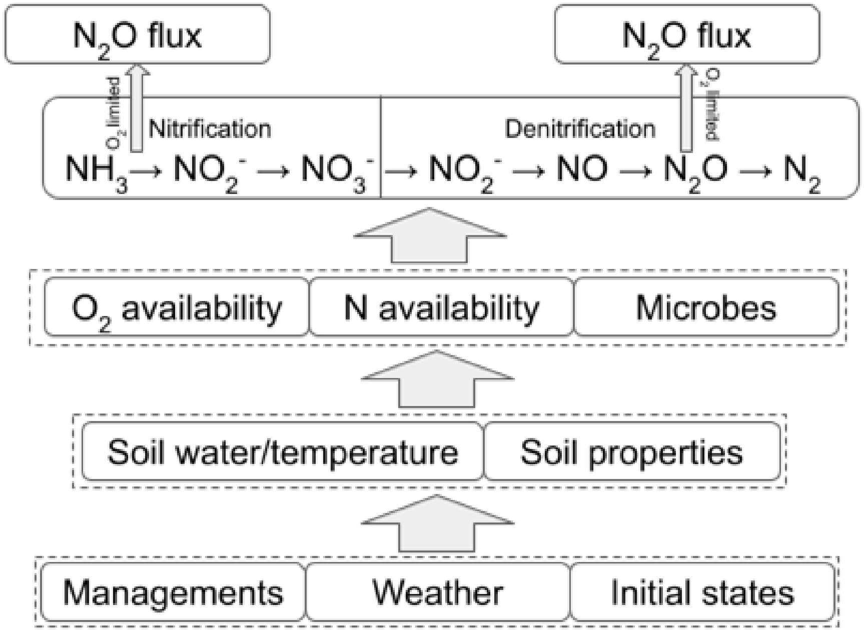
Agriculture contributes a quarter of global greenhouse gas (GHG) emissions that are causing climate change: **~14%** directly from agricultural activities and **~10%** through land use change.



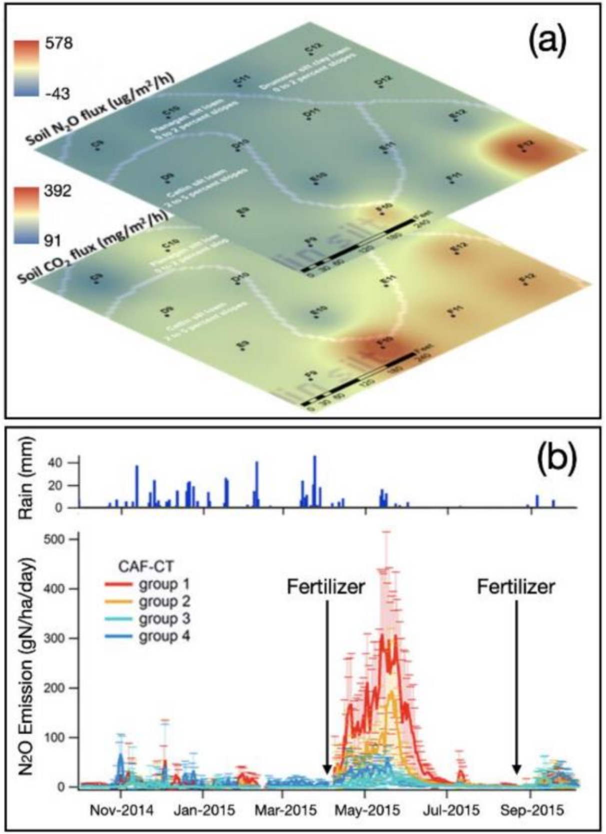
- ❑ Agricultural production accounts for **10%** of U.S. GHG emissions in 2018
- ❑ Fertilizer use accounts for 51% of the ag emissions, largely in forms of **N_2O, 265x more powerful than CO_2** as a GHG
- ❑ Over applying fertilizer also causes **water** & **air** pollution and **land** degradation

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.



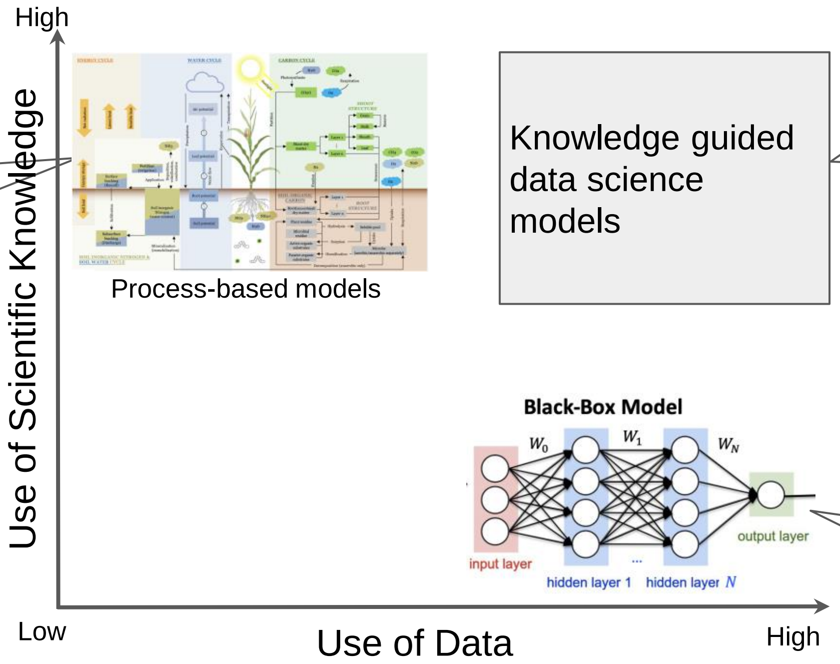
Hot spots, hot moment of N_2O fluxes



Waldo et al. 2019

Opportunities with knowledge guided machine learning

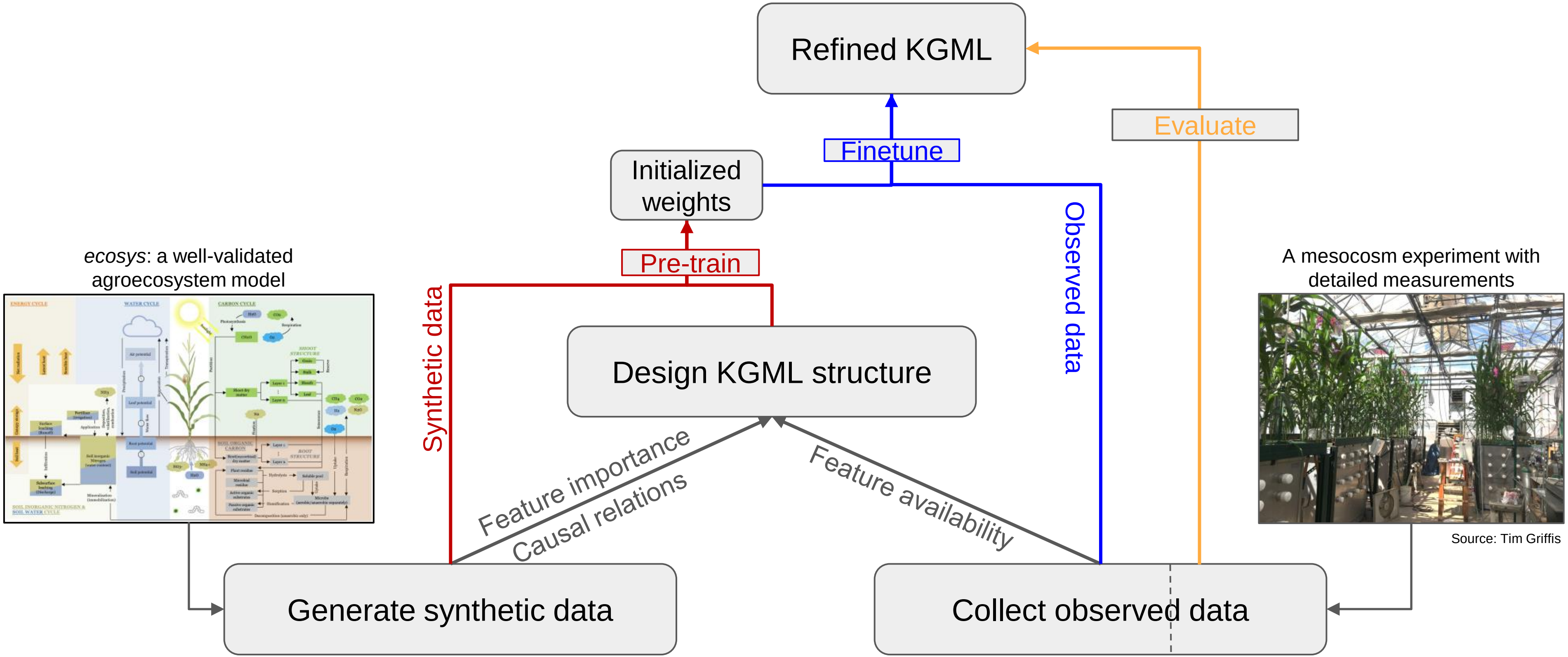
Contain more knowledges but also many empirical parameters



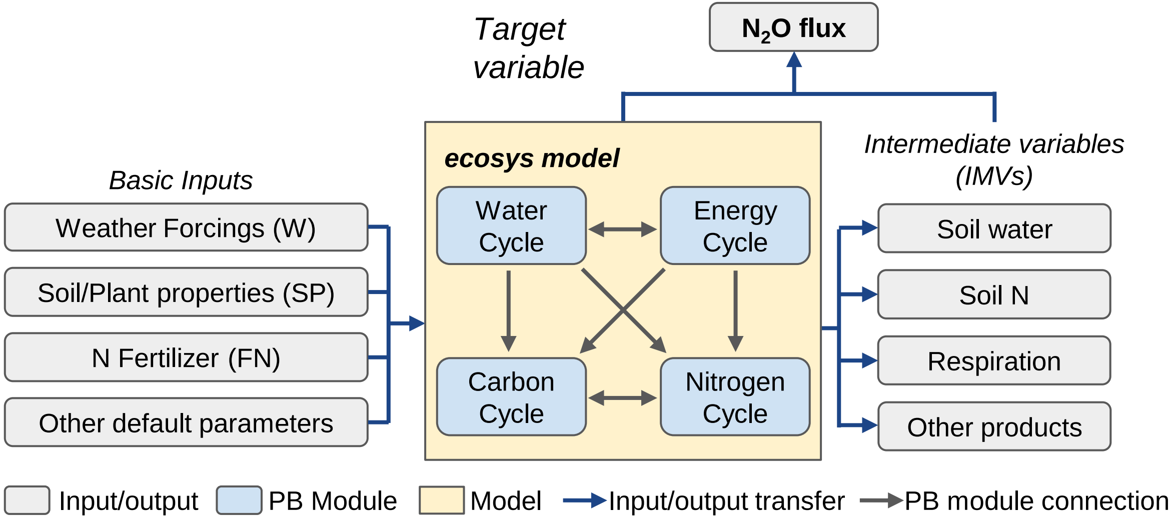
Take full advantage of data without ignoring the treasure of accumulated scientific knowledge

Require large number of data to train, often work as black-box

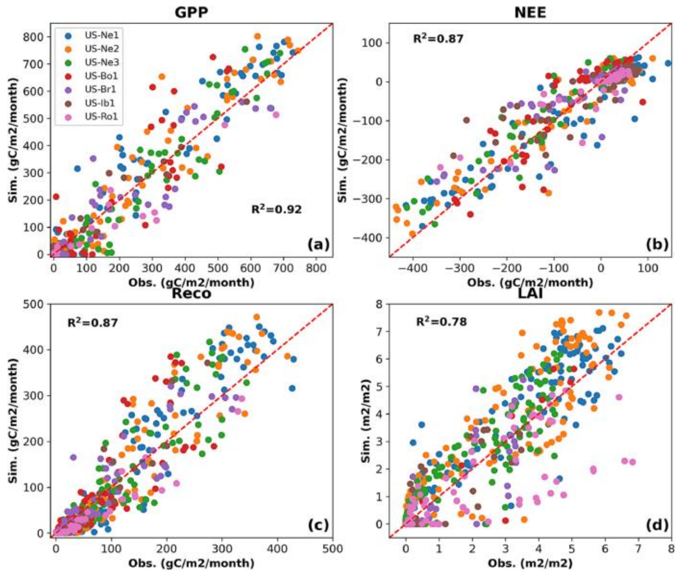
Experiments overview



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:

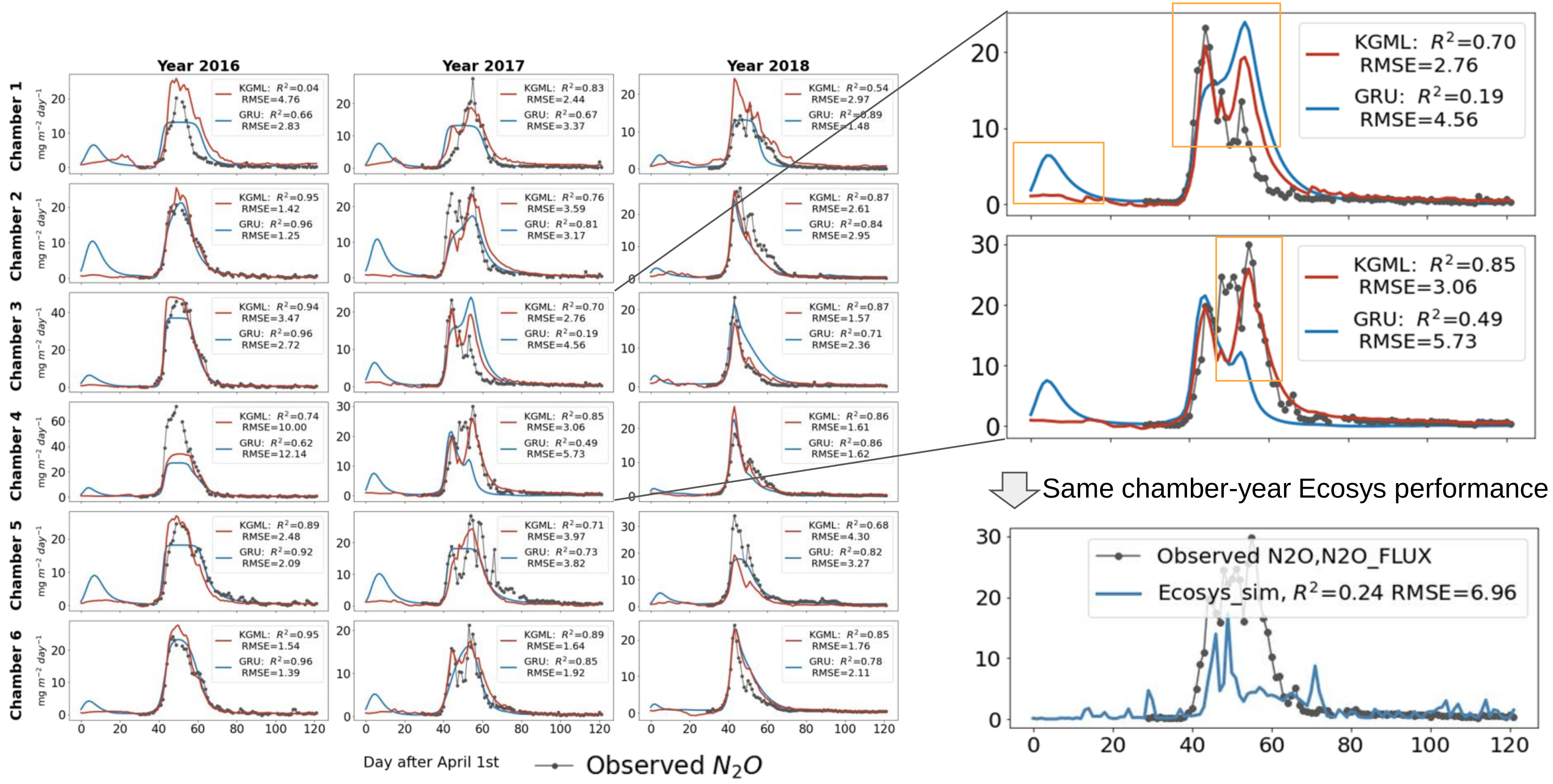
- Growing seasons during 2016-2018
- 6 chambers with different precipitation treatment
- N_2O flux was measured by Teledyne M320EU Analyzer in automatic chamber

Available variables:

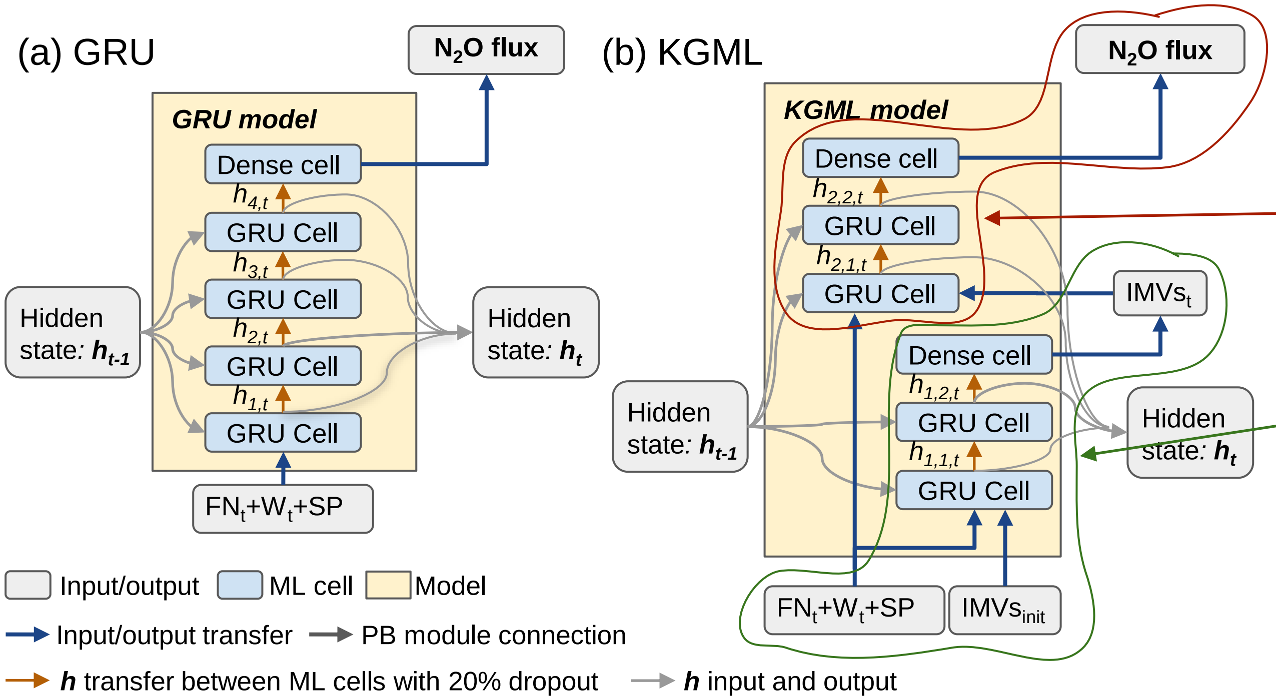
- Controlled weather conditions
- Hourly N_2O fluxes, CO_2 fluxes, and soil moisture at 15 cm depth
- Weekly soil [NO_3-], [NH_4+] at 15cm depth
- Management info: planting/harvesting dates, fertilizer application timing and rate

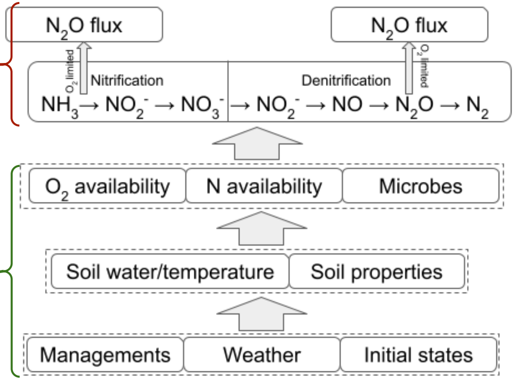
Source: Tim Griffis

The KGML model outperformed the pure ML model



Develop KGML model based on causal relations and feature importance

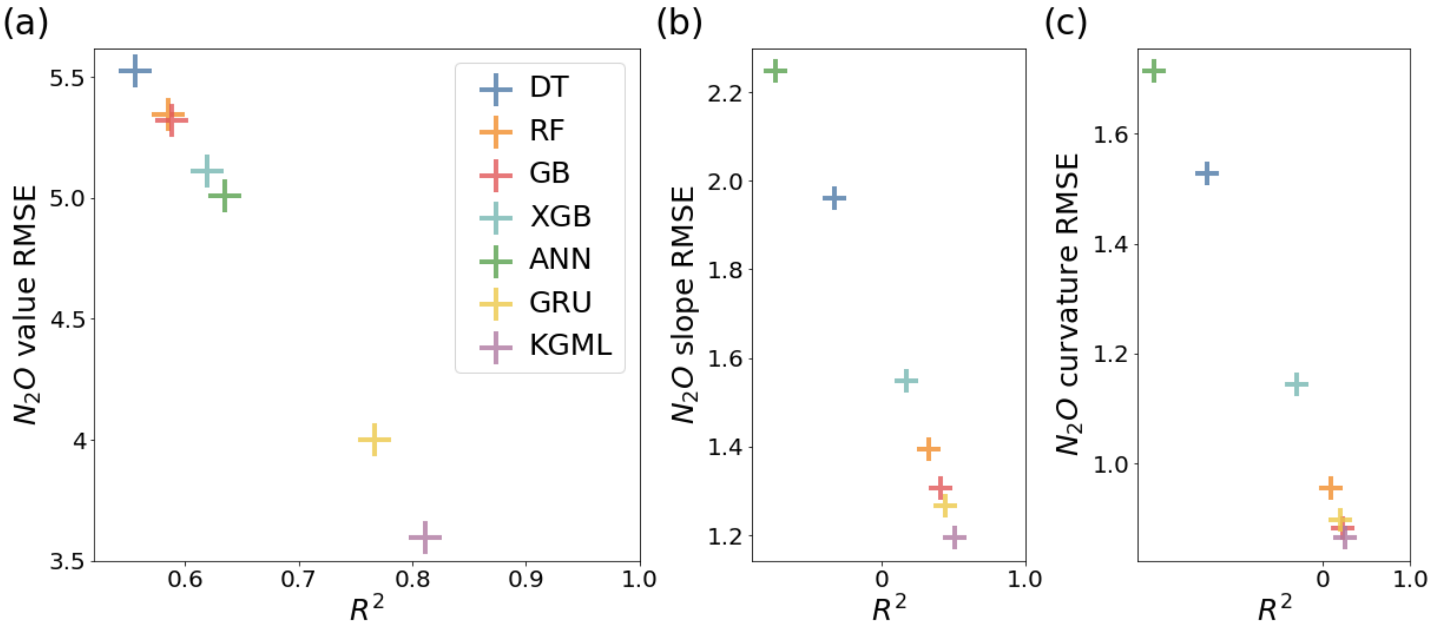




Detected key processes and causal relations for N_2O fluxes

- 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