Estimating the Autotrophic and Heterotrophic Respiration in the US Crop Fields using Knowledge Guided Machine Learning

Licheng LIU[1], Wang Zhou[2], Zhenong Jin[3], Jinyun Tang[4], Xiaowei Jia[5], Chongya Jiang[2], Kaiyu Guan[2], Bin Peng[2], Shaoming Xu[1], Yufeng Yang[1], and Vipin Kumar[1]

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

November 24, 2022

Abstract

Improving the estimation of CO_2 exchange between the atmosphere and terrestrial ecosystems is critical to reducing the large uncertainty in the global carbon budget. Large amounts of the atmospheric CO_2 assimilated by plants return to the atmosphere by ecosystem respiration (Reco), including plant autotrophic respiration (Ra) and soil microbial heterotrophic respiration (Rh). However, Ra and Rh are challenging to be estimated at large regional scales because of the limited understanding of the complex interactions among physical, chemical, and biological processes and the resulting high spatio-temporal dynamics. Traditional approaches for estimating Reco including process-based (PB) models are limited by human knowledge resulting in limited accuracy and efficiency. Accumulation of the *in situ* observation of net ecosystem exchange (NEE), weather, and soil, and satellite data of GPP, LAI and soil moisture make it possible for applying data driven machine learning (ML) approaches. But the ML model approach has disadvantages of omission of domain knowledge and lack of interpretability. Here we propose a novel knowledge guided machine learning (KGML) method for predicting daily Ra and Rh in the US crop fields. With Gated Recurrent Unit (GRU) as the basis, we develop the KGML models constructing the hierarchical structure of ML with a mass balance constraint. The KGML models were pre-trained using synthetic data generated by an advanced agroecosystem model, ecosys, and re-trained with real-world FLUXNET observation data. We extrapolate the best KGML model to crop fields over the US with the help of satellite data, reanalysis climate forcings, and soil database to reveal the spatio-temporal variations and key controlling factors. We believe this study advances the interpretable machine learning concept for carbon cycle estimation and will shed light on many other process-based biogeochemistry research.

Knowledge-based artificial intelligence for agroecosystem carbon budget and crop yield estimation

Licheng Liu#, Wang Zhou#, Kaiyu Guan*, Bin Peng, Chongya Jiang, Jinyun Tang, Sheng Wang, Robert Grant, Symon Mezbahuddin, Xiaowei Jia, Shaoming Xu, Vipin Kumar, Zhenong Jin*

Department of Bioproducts and Biosystems Engineering, University of Minnesota; College of Agricultural, Consumer and Environmental Sciences, University of Illinois at Urbana Champaign

#These authors contributed equally to this work; *Corresponding to: Zhenong Jin (jinzn@umn.edu), Kaiyu Guan (kaiyug@illinois.edu)

1. Introduction

- Accurate estimation of carbon budgets is vital to assessing the climate change mitigation potentials of terrestrial ecosystems. Cropland carbon budgets play an important role in regional carbon budgets over cropland dominates the landscape such as the U.S. Midwest.
- However, there is still no reliable product on cropland carbon budget with high spatial and temporal resolutions over the U.S. Midwest.
- Empirical studies use flux tower observations to quantify different components of cropland carbon budget at local scale, such as net ecosystem exchange (NEE), but it is difficult to scale local observations up to regional scales.
- Process-based models can simulate individual components of cropland carbon budget, but are lacking effective constraints from observations.
- Although there is an increasing interest in leveraging recent advances in machine learning, capturing this opportunity requires going beyond the ML limitations, including limited generalizability to out-of-sample scenarios, demand for massive training data, and low interpretability due to the "black-box" use of ML
- To fill this gap, we used the knowledge-based artificial intelligence to integrated the advanced ecosystem model, *ecosys*, with a new remotely-sensed daily ecosystem gross primary production (GPP) observations to estimate the crop yield, ecosystem respiration (Reco), and NEE at field scale in the U.S. Midwestern cropland.

2. Method and Data

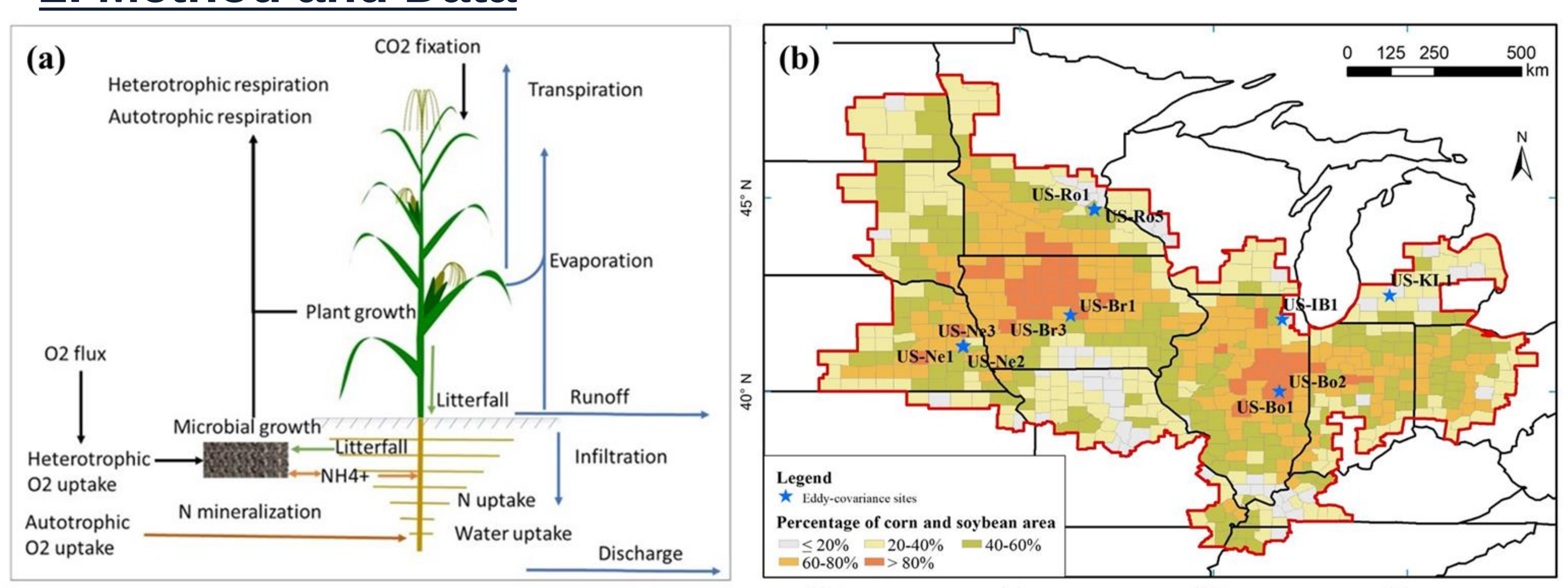

Fig. 1. (a) Major processes represented in the *ecosys* model (revised from (Grant, 2004; Zhou, 2021)), and (b) locations of the flux towers and the counties selected in this study.

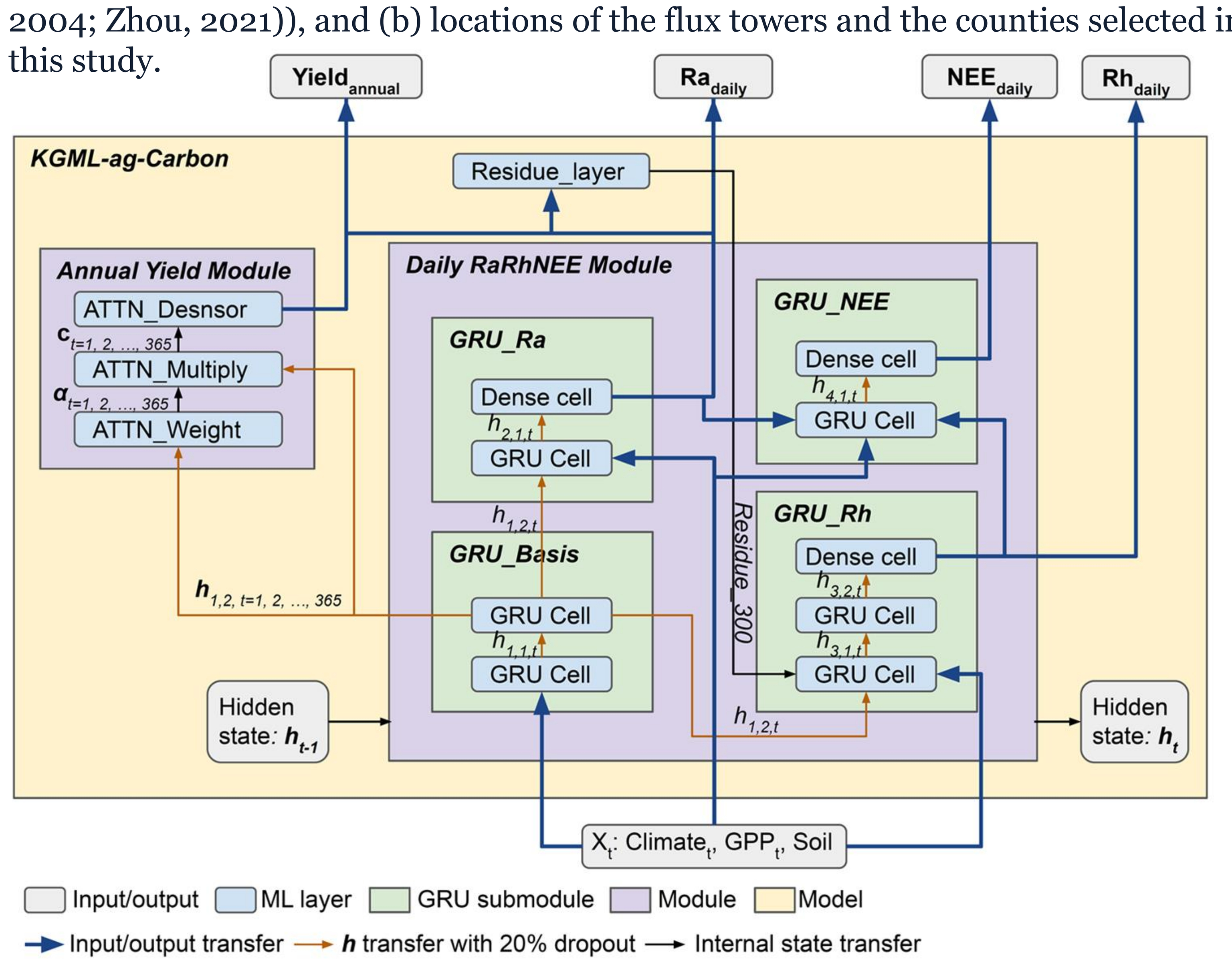

Fig. 2. The structural of knowledge-based machine learning for agricultural carbon budget estimation (KGML-ag-Carbon) developed in this study. KGML-ag-Carbon model would be carefully trained by including knowledge constraints before regional extrapolation.

3. Performance of KGML-ag-Carbon in crop yield estimation

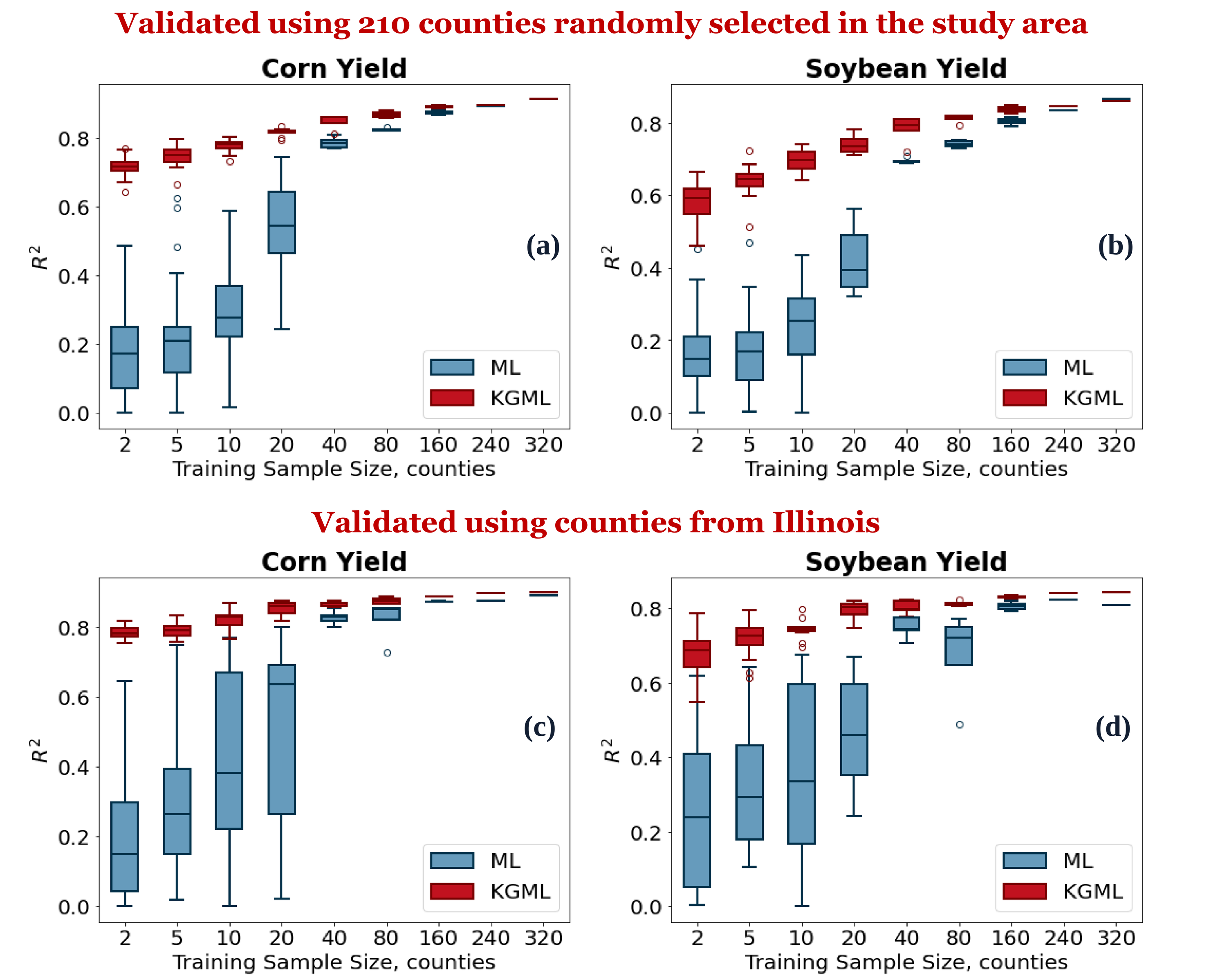

Fig. 3. The performance of KGML-ag-Carbon and pure machine learning (ML) in crop yield estimation. (a) and (b) were the model performance validated using 210 counties randomly selected in the study area for corn and soybean, respectively, and (c) and (d) were the model performance validated using the counties from Illinois for corn and soybean, respectively.

4. Carbon fluxes estimation over the U.S. Midwestern cropland

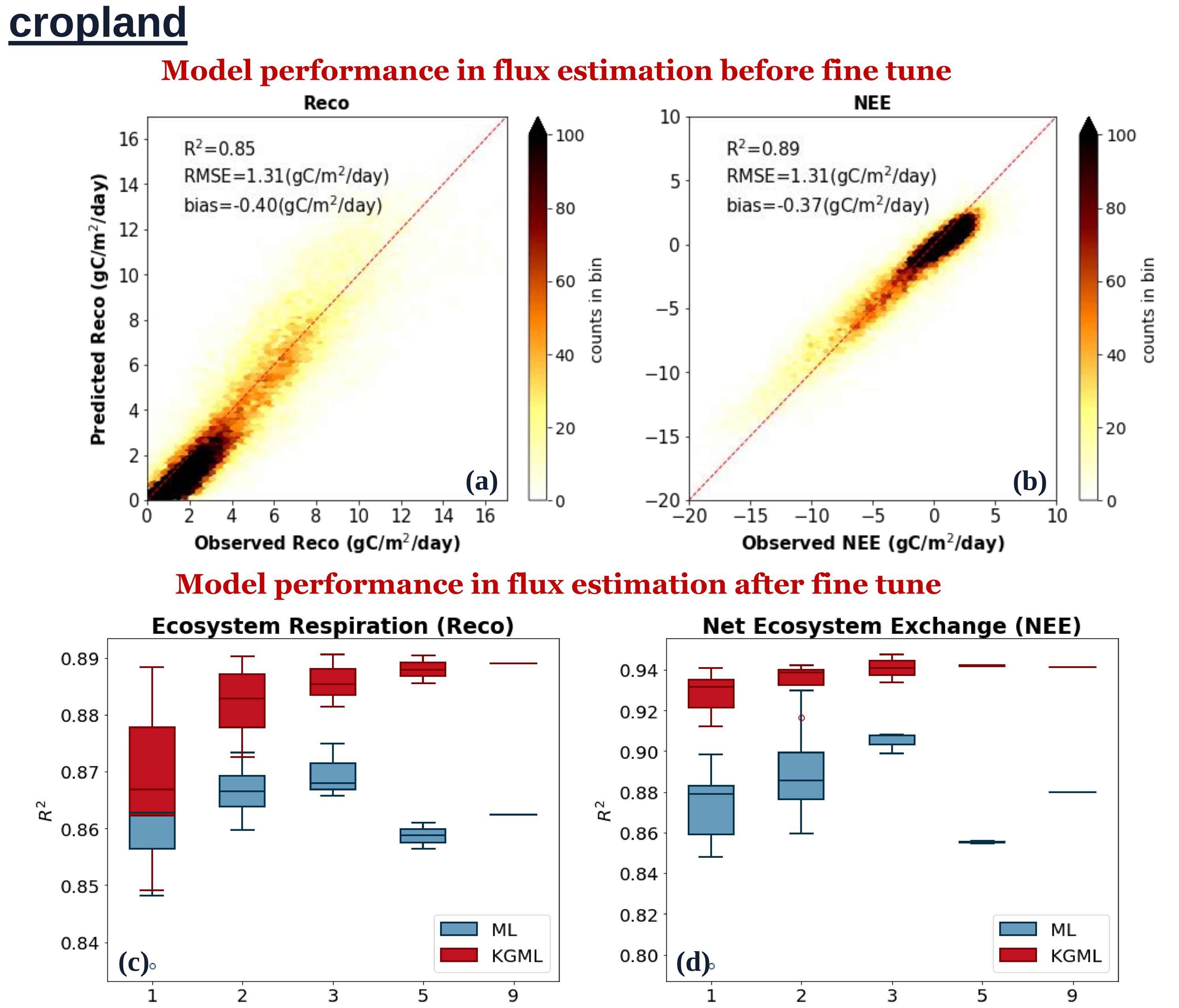

Fig. 4. The performance of KGML-ag-Carbon and pure machine learning (ML) in carbon fluxes (Reco and NEE) estimations. (a) and (b) were the KGML-ag-Carbon performance at 11 fluxtower sites before fine tuning, and (c) and (d) were the model performance for KGML and ML at US-Br1 and US-Br3 with different number of sites for fine tuning (or training), respectively.

4. Carbon fluxes estimation over the U.S. Midwestern cropland

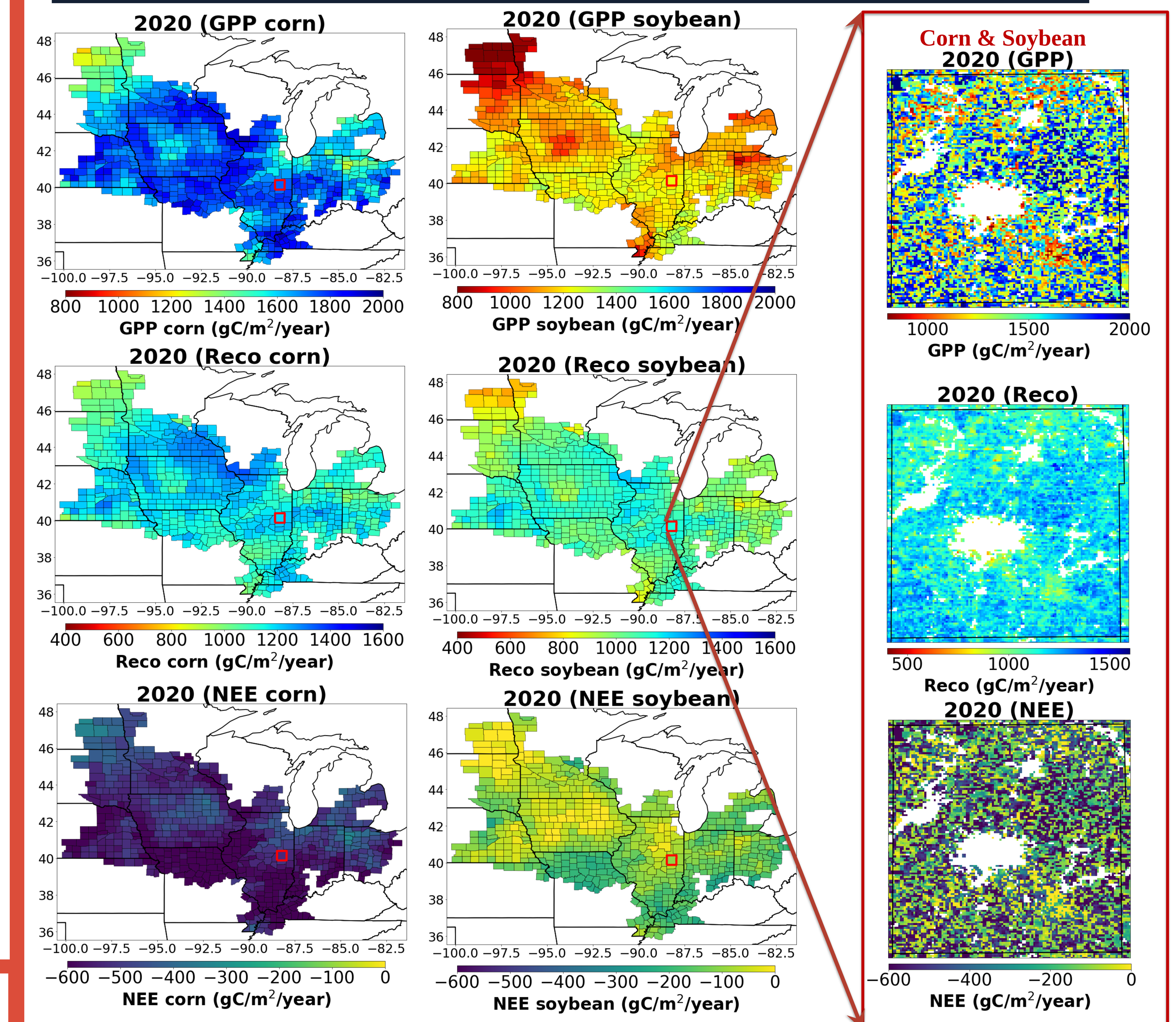

Fig. 5. An example of NIRv-based GPP and KGML-ag-Carbon estimated carbon fluxes (Reco and NEE) in 2020.

5. Conclusion

- Built an AI-based field scale carbon budget (Yield, Reco, NEE) estimation framework (KGML-ag-Carbon) by integrating knowledge-based artificial intelligence, advanced ecosystem model, and remotely-sensed GPP observations.
- Validated the performance of KGML-ag-Carbon in corn and soybean yield and carbon fluxes estimations using the crop yield from USDA NASS at county scale, and Reco and NEE at 11 cropland eddy-covariance sites in the U.S. Midwest.
- An efficient and reliable tool to estimate crop carbon and detailed components in high spatial and temporal resolution
- It has potential to evaluate soil, climate and management influences on carbon credit in field level and therefore guide farmers and policy makers to make right decisions
- The final product can feedback to process-based model for calibration and testing

6. References and Acknowledgements

[1] Grant (2004). Modeling topographic effects on net ecosystem productivity of boreal black spruce forests. *Tree Physiology*, 24, 1-18.

[2] Jiang, et al. (2021). A daily, 250 m and real-time gross primary productivity product (2000–present) covering the contiguous United States. *Earth System Science Data*, 13, 281–298.

[3] Liu, et al. (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.*

[4] Zhou, et al. (2021). Quantifying carbon fluxes, crop yields, and their responses to environmental variability using the *ecosys* model for U.S. Midwestern agroecosystems. *Agricultural and Forest Meteorology*, 307, 108521.

We acknowledge the support from DOE Advanced Research Projects Agency-Energy (ARPA-E), NASA Carbon Monitoring System Program, National Science Foundation Career Award, USDA National Institute of Food and Agriculture (NIFA) Program, Illinois Nutrient Research & Education Council, and Foundation for Food and Agriculture Research. We also acknowledge the following AmeriFlux sites for their data records: US-Ne1, US-Ne2, US-Ne3, US- Bo1, US- Bo2, US-Br1, US-Br3, US-Ib1, US-Ro1, US-Ro5, and US-KL1.

ILLINOIS | UNIVERSITY OF MINNESOTA Driven to Discover | arpa-e CHANGING WHAT'S POSSIBLE | NASA | FFAR