ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

Quantifying nitrogen loss hotspots and mitigation potential for individual fields in the US Corn Belt with a metamodeling approach

IOP Publishing Environ. Res. Lett. 16 (2021) 075008 https://doi.org/10.1088/1748-9326/ac0d21

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
22 December 2020

REVISED
15 June 2021

ACCEPTED FOR PUBLICATION
21 June 2021

PUBLISHED
9 July 2021

LETTER

Quantifying nitrogen loss hotspots and mitigation potential for individual fields in the US Corn Belt with a metamodeling approach

Taegon Kim[1], Zhenong Jin[1,2,*], Timothy M Smith[1,2], Licheng Liu[1], Yufeng Yang[1], Yi Yang[3], Bin Peng[4,5], Kathryn Phillips[1], Kaiyu Guan[4,5], Luyi C Hunter[1] and Wang Zhou[4]

1 Department of Bioproducts and Biosystems Engineering, University of Minnesota, St. Paul, MN, United States of America
2 Institute on the Environment, University of Minnesota, St. Paul, MN, United States of America
3 Key Laboratory of the Three Gorges Reservoir Region's Eco-Environment, Ministry of Education, Chongqing University, Chongqing 400045, People's Republic of China
4 College of Agricultural, Consumer and Environmental Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, United States of America
5 National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL, United States of America
* Author to whom any correspondence should be addressed.

E-mail: jinzn@umn.edu

Keywords: nitrogen loss, mitigation potential, metamodel, sustainable agriculture, the Corn Belt

Supplementary material for this article is available online

Abstract

The high productivity in the US Corn Belt is largely enabled by the consumption of millions of tons of manufactured fertilizer. Excessive application of nitrogen (N) fertilizer has been pervasive in this region, and the unrecovered N eventually escaped from croplands in forms of nitrous oxide (N_2O) emission and N leaching. Mitigating these negative impacts is hindered by a lack of practical information on where to focus and how much mitigation potential to expect. At a large scale, process-based crop models are the primary tools for predicting variables required by decision making, but their applications are prohibited by expensive computational and data storage costs. To overcome these challenges, we built a series of metamodels to learn the key mechanisms regarding the carbon (C) and N cycle from a well-validated process-based biogeochemical model, *ecosys*. The trained metamodel captures over 98% of the variability of the *ecosys* simulated outputs for 99 randomly selected counties in Iowa, Illinois, and Indiana. To identify hotspots with high mitigation potential, we introduce net societal benefit (NSB) as an indicator for synthesizing the loss in yield and social benefits through emissions and pollutants avoided. Our results show that reducing N fertilizer by 10% leads to 9.8% less N_2O emissions and 9.6% less N leaching at the cost of 4.9% more SOC depletion and 0.6% yield reduction over the study region. The estimated total annual NSB is \$395 M (uncertainty ranges from \$114 M to \$1271 M), including \$334 from social benefits (uncertainty ranges from \$46 M to \$1076 M), \$100 M from saving fertilizer (uncertainty ranges from \$13 M to \$455 M), and −\$40 M due to yield changes (uncertainty ranges from −\$261 M to \$69 M). For the median scenario, we noted that 20% of the study area accounts for nearly 50% of the NSB, and thus represent hotspot locations for targeted mitigation. Although the uncertainty range suggests that developing such a high-resolution framework is not yet settled and the scenario based estimations are not appropriate to inform the management practices for individual farmers, our efforts shed light on the new generation of analytical tools for life cycle assessment.

1. Introduction

The United States produces about 1/3 of the world's corn for food, feed, and biofuel (USDA-FAS 2020), largely in the US Midwest, also known as the Corn Belt. This high productivity was enabled by artificially draining the seasonally saturated soils of the Midwestern landscape and catalyzed by application of millions of tons of manufactured fertilizer (USDA 2019). It is estimated that 40%–80% of these applied fertilizers is lost from soils, through drainage tiles to water bodies as reactive nitrate (Chen *et al* 2016, Zhao *et al* 2016) and into the atmosphere as N_2O (Turner *et al* 2015, Zhang *et al* 2020). With the continued need to satisfy growing global demand for agricultural products, ensuring the sustainability of food production systems has become a major challenge in the US Midwestern region (Foley *et al* 2011, Tilman *et al* 2011).

To address the socio-economic expectations of different stakeholders, thorough understanding of the relationship between agricultural inputs and outputs is necessary. Although previous research on this topic is abundant, many studies rely on aggregated data of regional and national cropping systems (Johnston *et al* 2015, Marshall *et al* 2015, Burke and Emerick 2016, Cho and McCarl 2017, Kent *et al* 2017). Aggregated estimates from these studies can be easily compared with government-published benchmarks, but have diluted the spatial variability at a finer scale that are more relevant to growers' practices and supply chain management (Smith *et al* 2017). Meanwhile, many others have designed experiments on individual farms or fields to investigate the input and output flows of corn production (Kwon *et al* 2017, Wienhold *et al* 2018, McNunn *et al* 2020). Assessments obtained from the field approach provide valuable first-hand results, but generalizing these local findings for regional decision making is questionable (Basso and Liu 2019). Limitations of the two approaches call for an alternative pathway that can cover broad geographical regions while maintaining fine granularity of farm-field information.

Process-based crop models that simulate the complex interactions between soil, weather and management practices have been widely used to inform field management (Keating and Thorburn 2018, Morris *et al* 2018). At regional scale, crop models are often used to examine alternative management strategies through gridded simulations (e.g. 10 km) of baselines and proposed scenarios (Lu *et al* 2018, Peng *et al* 2020, Mandrini *et al* 2021). While instrumental to informing land management decisions, these models process each 'location' individually, creating, to date, unmanageable computational challenges for their application across larger landscapes or complex sourcing scenarios incorporating multiple nonadjacent locations. Even by implementing parallel or cloud computing, it is unlikely that the settings could accommodate the millions of croplands to be processed and the many different ways farmers might manage their fields, not to mention interactions with soil and weather conditions (Shahhosseini *et al* 2019).

Metamodeling, or making a 'model of a model', is the process of generating a statistical or machine learning model to approximate the process-based crop model; thus, it can overcome the above-mentioned computational challenges (Villa-Vialaneix *et al* 2012). Additionally, metamodels provide flexible applicability across different temporal and spatial scales (Britz and Leip 2009, Nolan *et al* 2018). Even though a metamodel is a simplification of a process-based model, a trained metamodel may even show better estimation with fine resolution data than the estimation from the original process-based model with low resolution data (Perlman *et al* 2014). In recent years, machine learning based metamodeling approaches have been widely tested (Villa-Vialaneix *et al* 2012, Mekonnen *et al* 2015) and have demonstrated early success in simulating corn yield and N losses across the US Midwest (Shahhosseini *et al* 2019).

Estimating crop productivity and its associated environmental performance at the field scale has also attracted surging interest among practitioners and civil society. Countless multi-stakeholder initiatives have emerged over the past three decades, from the organic movement to cooperative supply chain platforms aimed at assessing agriculture's 'fieldprint'. In addition, many food manufacturing corporations have committed to reducing their environmental impacts by optimizing their supply chain. Recent studies on subnational corn and soy mobility that highlight the spatial variability of GHGs and water footprints of upstream suppliers and downstream processors illustrate the consolidation of agricultural impacts within a handful of large firms (Smith *et al* 2017, Brauman *et al* 2020) and opportunities for more sustainable management practices through technical support and financial assistance (Eagle *et al* 2020). Yet, for these initiatives to achieve their desired effects mitigation strategies will likely require more targeted application and more prospective modeling at high spatial resolution (Groffman *et al* 2009). The recent finding that high-emission hotspot locations are largely unrelated to high-production areas (Carlson *et al* 2017) emphasizes that interventions will need to be characterized not only in terms of emissions reductions but also productivity. To the best of our knowledge, this high-resolution information is not currently available in the US Corn Belt.

In this study, we built a series of metamodels to learn the key mechanisms regarding the C and N cycle from a well-calibrated process-based biogeochemical model, *ecosys* (Grant *et al* 2016, Zhou *et al* 2021). Using these metamodels, we generated millions of scenario simulations and investigated two fundamental questions to the US Midwest sustainability: where are the mitigation hotspots? How much mitigation can we expect under different management scenarios? We synthesized four simulated indicators of agroecosystem sustainability (i.e. yield, N_2O emissions, N leaching, and changes in SOC (ΔSOC)) into net societal benefits as the basis for identifying hotspots and infeasible land for mitigation for every corn field across the study area. Given the inevitable uncertainty in the analysis due to data limitations, our results are not intended to inform the management practices for individual farmers, but instead to provide a realistic estimate of mitigation potentials in this corn production system.

2. Method

2.1. Build a metamodel to approximate the process-based model

To provide the synthetic data for training metamodels, the process-based model employed should not rely too much on empirical relationships that may vary from one place to another. *Ecosys* is one of the few agroecosystem models that meet this restrictive criterion since it is constructed from basic principles of physics and biochemistry using parameters that may be determined independently of the model itself (Grant 2001); therefore, it is widely applicable to different soils, climates, and managements. Previous work using *ecosys* has fully demonstrated its capabilities in simulating carbon and nitrogen cycling (Grant 2001) and the impacts of different management practices (Grant *et al* 2006, 2016). Uniquely, *ecosys* is one of the few models that has resolved the biogeochemical coupling between different crops and microbial population dynamics (including nitrogen fixation, nitrification, and denitrification) and has been evaluated extensively for soil C and N_2O modeling in agroecosystems (Grant *et al* 2001, 2006, 2016, Zhou *et al* 2021). Validation results showed that *ecosys* was able to reasonably estimate hourly N_2O flux at the field level ($R^2 = 0.46$) (Metivier *et al* 2009) as well as the total flux aggregated over the measured period (relative error between 0% and 16.5%) (Grant and Pattey 2003). Relative error in a long-term simulation of ΔSOC with no-fertilizer can be as low as 5.1% (Grant *et al* 2001). A recent model evaluation study by Zhou *et al* (2021) using data from seven agricultural eddy-covariance flux tower sites shows that with moderate calibration, *ecosys* is able to capture the daily time series of gross primary production (GPP), net ecosystem exchange (NEE), ecosystem respiration (Reco), and leaf area index (LAI) with R^2 equal to 0.92, 0.87, 0.87, and 0.78, respectively. For regional scale simulations, *ecosys* reproduced the spatial distribution of USDA NASS county-level yield statistics with R^2 equal 0.83 and 0.80 for corn and soybean, respectively. From a mass balance perspective, these results further justified using *ecosys* to simulate annual SOC change since ΔSOC can be approximately estimated as GPP minus Reco and harvested yield.

We generated synthetic data (i.e. yield, N_2O emissions, N leaching, and ΔSOC from 2001 to 2018) for training metamodels by running *ecosys* (the same version as was used in Zhou *et al* 2021) for 99 randomly selected counties in Iowa, Illinois, and Indiana (figure S1 available online at stacks.iop.org/ERL/16/075008/mmedia). Hourly meteorological inputs, including net radiation, air temperature, precipitation, relative humidity, and wind speed, for *ecosys* are collected from the phase 2 of North American Land Data Assimilation System (NLDAS-2) with 0.125 degree resolution (Xia *et al* 2012). A total of 21 variables describing soil physics (e.g. bulk density, sand content, silt content), hydraulics (e.g. field capacity and wilting point), and soil biogeochemistry (e.g. pH, SOC, organic N and P) were derived from the SSURGO database (Soil Survey Staff 2020) for each soil map unit and resampled into 12 layers with fixed layer depths to set up *ecosys* soil inputs. Crop management in the baseline simulation was configured as rain-fed, no-till, and with corn-soybean rotation. To investigate the impacts of varying N fertilizer rate, the simulation experiments were tested at 20 different N rates (i.e. 0, 40, 80, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 240, 260, 280, 300 lb N/acre). More details about the input data of *ecosys* regional simulations are summarized in table S1.

We built four separate models for yield, N_2O, N leaching, and ΔSOC for 0–30 cm depth. Among many choices in machine learning models, random forest (Breiman 2001) and XGBoost (Chen and Guestrin 2016) have shown good performance in metamodeling studies for agroecosystems (Perlman *et al* 2014, Shahhosseini *et al* 2019). During the model selection stage, we noticed XGBoost outperformed RF in the testing set, although both methods performed well with the training set (table S3). Therefore, we used XGBoost for generating results for subsequent analysis. The proposed metamodels were trained by mapping the dependent variables to a range of variables characterizing key soil properties, aggregated weather and fertilizer rates (table S2), all derived from the inputs and outputs of *ecosys* simulations. The synthetic dataset was partitioned into 70% of the training metamodels and 30% for validation using random sampling. We used the xgboost package (Chen and Guestrin 2016) in python 3.6.10. The default hyperparameters were used since tuning their values did not lead to meaningful improvement during the preliminary analysis. The final metamodels were applied to every SSURGO soil map unit within all cornfields. The USDA crop data layer (CDL) (Johnson 2013) was used to identify cornfields in each year.

2.2. Net societal benefit from farm management practices

The difference in societally adjusted net operating revenue of production between the benchmark scenario and various scenarios of management practices is defined as net societal benefit (NSB) and measured with equation (1):

$$\text{NSB} = \Delta\text{yield} \times C_C + \Delta F \times C_F + \text{SB} \qquad (1)$$

where NSB is the sum of farm revenue change and social benefits from avoiding environmental cost ($ ha^{-1}), Δyield is the change in yield (t ha^{-1}), C_C is the corn price ($ t^{-1}), ΔF is the reduced N fertilizer amount (kg N ha^{-1}), C_F is the fertilizer cost ($ N^{-1} kg^{-1}), and SB is the social benefit ($ ha^{-1}). SB is estimated with equation (2)

$$\mathrm{SB} = (\mathrm{GHG}_{\Delta \mathrm{N2O}} + \mathrm{GHG}_{\Delta \mathrm{fert}} + \mathrm{GHG}_{\Delta \mathrm{SOC}}) \times C_{\mathrm{GHG}} + \Delta N_{\mathrm{leaching}} \times C_{\mathrm{leaching}} \quad (2)$$

where $\mathrm{GHG}_{\Delta \mathrm{N2O}}$, and $\mathrm{GHG}_{\Delta \mathrm{SOC}}$ is the reduced N_2O emission, reduced emission associated with fertilizer manufacturing and transport, and reduced ΔSOC converted to CO_2 equivalent (CO_2e) greenhouse gas (GHG) emissions (t CO_2e ha^{-1}), respectively; C_{GHG} is social cost of GHG emissions ($/t CO_2e), $\Delta N_{\mathrm{leaching}}$ is reduced N leaching (kg N ha^{-1}), and C_{leaching} is the social cost for N leaching. $\mathrm{GHG}_{\Delta \mathrm{fert}}$ is derived from the life cycle inventory, a database that provides the embedded GHG of a product based on resource extraction to the factory gate (cradle-to-gate) assessments GHG (PE International 2014).

The cost for different items adjusted as dollars in 2014 is listed in table S4. With a discount rate of 2.6%, corn and fertilizer prices are set to 150 $/t and 1.15 $ N^{-1} kg^{-1} based on the average market value for 2001–2018 (figure S8). Social cost consists of two categories: GHG emission and water pollution. Cost of N_2O emissions is estimated by converting into CO_2e based on the fact that N_2O is 265times more powerful than CO_2 regarding global warming potential for a 100 year time horizon (IPCC 2014). CO_2e of fertilizer is estimated based on the assumption that farmers apply one-unit ammonium nitrate and two units of URAN that produce emission of 9.1 and 6.1 kg CO_2e kg^{-1} N^{-1}, respectively (PE International 2014). We assumed a price of $50/t for CO_2, which is the Interagency Working Group's central estimate (IWG 2016, Revesz *et al* 2017). Cost for N leaching, largely due to eutrophication, is estimated as 18.54 $ kg^{-1} N^{-1} (8.88 and 31.58 $ kg^{-1} N^{-1} for lowest and highest estimates, respectively) based on Sobota *et al* (2015). All metamodels output density variables as quantity per ha. Aggregated estimations for county- or state-level were obtained by multiplying corn production acres from the USDA census data in 2017 (USDA NASS 2019).

2.3. Hotspots and mitigation potentials

We calculated profit changes under various N reduction scenarios to determine where cornfields can break even or make more expected profits through avoiding social cost. Since no data is available at the field scale to provide the current N application rate, we used the maximum return to nitrogen (MRTN) calculator (Sawyer *et al* 2006) to derive the baseline scenario of N management. MRTN is a well-known method to estimate the economically optimal N rate based on state-specific yield to N response curves that are derived from multiple years and sites (Sawyer *et al* 2006). The MRTN calculator is widely used by University Extension specialists, farm consultants and fertilizer vendors to provide the recommended N fertilizer rate (Morris *et al* 2018), therefore is a reasonable approximation to farmers' practices. More details to derive MRTN rate for each map unit are provided in Jin *et al* (2019). The N mitigation scenarios were configured as 10% and 20% N application reductions from the optimal applied N rate over the study region. For each map unit from 2001 to 2018, the annual NSB was calculated by comparing multiple scenario simulations by the metamodel to the benchmark. Based on these simulations, we identified map units with NSB ($ ha^{-1}) greater than 0 as feasible spots for GHG mitigation, and those with NSB greater than $40 ha^{-1} in more than 75% of the years during 2001–2018 as hotspots (figure S2). Here, the threshold of $40 ha^{-1} was selected because it is roughly at the 75 percentiles of the attainable NSB.

2.4. Uncertainty analysis

Similar to other model-based studies, uncertainty is a necessary component in our estimates. While quantifying the uncertainty from all sources is beyond the scope of this study, here we particularly focused on three major categories, including baseline N rate, pricing for corn yield and environmental benefits, and management practices. Since our baseline N rate was derived from the MRTN calculator (which by itself carries some uncertainty), we tested the impacts of varying the recommended MRTN rate by $\pm$10% on the estimated regional NSB. To account for uncertainty caused by pricing, we examined the extent to which mitigation potentials can vary by different cost combinations at the benchmark N rate. Detailed pricing combinations for changes in yield, N_2O, N leaching and ΔSOC are listed in table S6. We also evaluated the impacts of adopting different management practices compared with the baseline scenario of corn-soybean rotation and no-till for every cornfield. Modifiers to simulated yield, N_2O emission, N leaching under continuous corn and conventional tillage are extracted from several recent meta-analysis studies and review papers (see table S7). For example, Eagle *et al* (2020) suggested N_2O emission from continuous corn fields is on average 43% more than corn-soybean rotation field based on a sample of more than 600 site-year observations, therefore a modifier of 1.43 will be multiplied to the baseline N_2O estimation.

3. Results

3.1. Metamodel performance

The metamodels based on the XGBoost method can fully capture the variability of *ecosys* simulations, with testing R^2 for all four variables (i.e. yield, ΔSOC, N_2O and N leaching) greater than 0.98 and relative RMSE smaller than 5% (table S3 and figure S2). The scatter plots of metamodel versus *ecosys* simulations for testing dataset are clustered closely around the 1:1 ratio line (figure S2), suggesting metamodels were able to provide unbiased approximations of *ecosys* for the variables of interest. We did not further validate metamodel's performance on representing the spatial pattern in ΔSOC, N2O and N leaching due to lack of field measurements. At the regional scale, the USDA NASS county-level yield statistics is the only available dataset for validation, thus it was used to evaluate our metamodels. We detrend survey data using linear regression to eliminate the effects associated with technology improvements, which is not considered by the metamodel. Results showed that the metamodel could capture about 30% of the variability in detrended yields when all counties are evaluated (R^2 of 0.304 and RMSE of 1.96 t ha^{-1}), as well as the magnitude and interannual variability of detrended yields in most states (figure S3). Errors were larger in 2012, suggesting the metamodel did not adequately capture some extreme climatic events such as a severe drought.

Analysis on feature importance suggested that yield is largely determined by temperature, especially max temperature in July (tmax07) and minimum temperature in June (tmin06); the N_2O emission is determined by tmax07, fertilized N rate (N_rate) and precipitation in June (prec06); N leaching is primarily controlled by cation exchange capacity (cec7_r); and ΔSOC is mainly determined by total organic C (om_r), minimum temperature in July (tmin07), maximum temperature in July (tmax07) (figure S4). These findings were in line with the key predictive variables identified by a few recent studies focusing on yield (Peng *et al* 2018, Shahhosseini *et al* 2019), N_2O (Perlman *et al* 2014) and N leaching (Villa-Vialaneix *et al* 2012).

3.2. Baseline simulation

The baseline simulations for corn yield, N_2O emission, N leaching and annual ΔSOC are illustrated in figure 1. For corn yield, the 18 year averaged map showed a similar spatial pattern to a recent satellite-based estimation for the same region (Lobell *et al* 2020), except in Ohio. Low yielding areas in the north are coincident with high N leaching, low N_2O emissions and high ΔSOC (figure 1). The estimated mean annual N_2O emission is 2.59 kg N ha^{-1} over the study region, which falls in the range of 1.0–5.3 kg N ha^{-1} (table S5) as reported by several other meta-analysis studies for this region (Ingraham and Salas 2019, Pelton 2019, Chatterjee 2020, Eagle *et al* 2020). Estimated N_2O emissions are higher in the southern Corn Belt than in northern areas (figure 1(b)). Similar spatial patterns, although with slightly lower emission intensity, have been reported by a multi-model ensemble study (Tian *et al* 2019), by the EPA's GHG emissions report (EPA 2020), as well as by an inventory-based emission estimation (Janssens-Maenhout *et al* 2019). The estimated regional average N leaching is 23.8 kg N ha^{-1}, which is in line with the estimates of 5.8–58.4 kg N ha^{-1} (table S5) by other modeling and meta-analysis studies (Ingraham and Salas 2019, Jin *et al* 2019, Jungers *et al* 2019, Wang *et al* 2019, Chatterjee 2020, Eagle *et al* 2020). The high N leaching in the North can be explained by the often-saturated soil and hence higher drainage water flow in these areas. Estimated ΔSOC shows more loss in the south (figure 1(d)), likely due to the warmer climate and hence faster SOC decomposition. The estimated regional average ΔSOC is 0.18 t C ha^{-1} yr^{-1}, which is within the range of −0.16–1.27 t C ha^{-1} yr^{-1} (table S5) by other modeling and meta-analysis studies (Poffenbarger *et al* 2017, Zomer *et al* 2017, Horwath and Kuzyakov 2018, Xu *et al* 2019).

3.3. Hotspots for mitigating GHG emission and leaching

Two mitigation scenarios, 10% and 20% reductions of N-fertilizer application, were compared with the baseline simulation to quantify the mitigation potential. For the 10% reduction scenario, despite a slight decrease in crop yield (−0.61%) and in ΔSOC (−4.91%), the Corn Belt could benefit from significantly reduced N_2O emission (−9.78%) and N leaching (−9.62%) (table 1). Changes in yield and GHG emissions varied among states (tables 1 and 2). Among those, Michigan showed the highest reduction of N_2O emission (−17.43%), while Minnesota and Illinois observed 12.89% and 10.44% less N leaching, respectively. High yield reductions were observed in Illinois and Missouri, indicating greater yield to N responses in these states. When synthesizing the four components as net societal benefits, only 19.9% of the total area were identified as hotspots, whereas mitigation through N reduction was infeasible in more than 27% of the study region (figure 2). Such a high interannual variability poses an extra challenge to the planning of mitigation strategies. Social benefits from these hotspots was $334 million USD in addition to a $60 million USD increase in farm revenue, which collectively account for 50% of the total NSB over the study area (figure 3(a)).

Mitigation hotspots were largely identified in Illinois, Indiana, Iowa, Michigan, and Minnesota (table 2, figure 3). The contribution of each component to the NSB was different across the region (figure S5). Annual NSB showed high-interannual variability in many places except for northern Minnesota (figure S6), suggesting the dominant component of GHG emissions were more sensitive to changes in weather conditions. Due to the high pricing for water pollution by reactive N, fields with the highest mitigation potential were mostly found in northern Minnesota, northern Indiana, and southern Michigan, where N leaching reduced

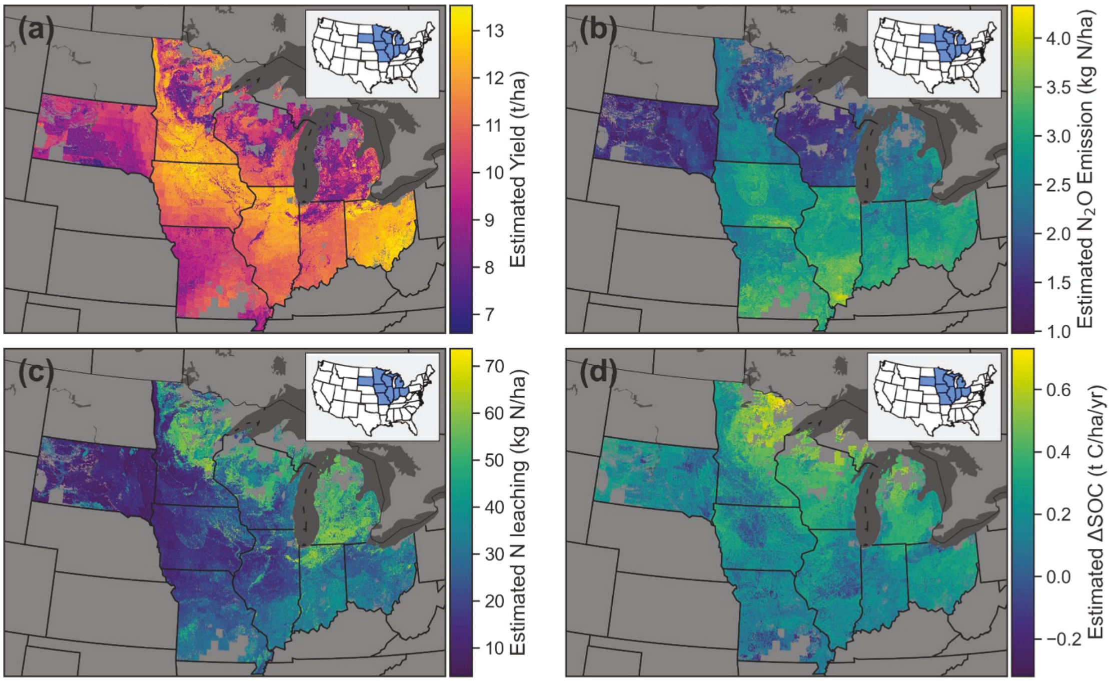

Figure 1. Baseline estimation of yields (a), N_2O emissions (b), N leaching (c), and annual changes of SOC (ΔSOC) (d) during 2001–2018 using the metamodels. N application rates are based on the state-level MRTN estimation.

Table 1. State-specific average changes in yields, N_2O emissions, annual changes of soil organic carbon (ΔSOC), and N leaching across hotspot areas under 10% N fertilizer reduction during 2001–2018, compared with the baseline simulation.

		Changes in environmental release (%)			Hotspot areas	
State	Yield changes (%)	N_2O	ΔSOC[a]	N leaching	Acreage (10^6 ha)	(% of state corn acreage)
Illinois	−1.24	−8.94	−6.20	−10.44	1.65	36.9
Indiana	−0.78	−7.85	−4.95	−10.03	0.64	29.2
Iowa	−0.83	−12.77	−4.85	−6.67	0.56	10.6
Michigan	1.40	−17.43	−9.38	−8.16	0.44	51.3
Minnesota	0.49	−7.58	−2.21	−12.89	0.51	16.2
Missouri	−1.29	−9.80	−0.90	−10.31	0.21	15.3
Ohio	−0.55	−10.85	−5.80	−6.89	0.10	7.4
South Dakota	−0.56	−6.67	5.17	−6.18	0.02	0.8
Wisconsin	0.55	−2.16	−1.39	−6.82	0.24	19.9
Corn Belt	−0.61	−9.78	−4.91	−9.62	4.36	19.9

[a] Positive number of ΔSOC means SOC accumulation and negative number means SOC depletion.

significantly in response to 10% N reduction. Mitigation of N_2O emission and fertilizer embedded GHG were the dominant contributors to NSB in Iowa and Illinois (table 2). Overall reduced N_2O emission contributed less to the NSB than reduced N leaching, managing N_2O fluxes has greater potential in reducing N loss in Illinois, Iowa, and Michigan (table 2).

Admittedly, not all of the four components are of interest to stakeholders who care about sustainability. For example, a water quality scientist may only care about N leaching, and would prefer to know priority locations with a particular focus on mitigating N loss to water bodies. We thus masked figure 3(a) to reflect priority regions for reducing N_2O emission, and N leaching independently (see figure S7 for thresholds for masking). Figure 3(c) showed that by masking areas currently with high N_2O emissions, an NSB of $141 million USD can be expected from reducing 10% N fertilizer. Similarly, the expected social benefits are $305 million USD if the targets are N leaching (figure 3(e)).

Mitigation potential varied among soil types (table 3). Sandy soil, including loam sand, sand, sandy clay loam, and sandy loam, is the most sensitive soil type in response to N reduction, with over 80% of the sandy soil fields belonging to hotspots. This can be explained by the fact that high N leaching was observed among the sandy soil profiles. Yield increased in sandy soils while decreased in other soil

Table 2. Changes in GHG emissions, N leaching, profit, and social benefit in hotspot areas across nine Corn Belt states under 10% N fertilizer reduction from 2001–2018.

	Changes in GHG emissions (MMT CO_2e) [a]				Changes in net societal benefit ($ millions)			
State	N_2O	ΔSOC[b]	Fertilizer embedded GHG	N leaching(TMT N) [a]	Corn production	Fertilizer savings	Social benefit	Total
Illinois	−0.195	0.057	−0.259	−5.2	−34.0	41.9	115.4	123.3
Indiana	−0.056	0.022	−0.084	−3.0	−7.3	13.6	60.7	67.0
Iowa	−0.082	0.018	−0.084	−1.1	−7.1	13.5	27.8	34.2
Michigan	−0.079	0.049	−0.055	−1.9	8.1	8.9	38.9	55.8
Minnesota	−0.034	0.015	−0.067	−2.7	3.6	10.8	55.0	69.4
Missouri	−0.024	0.001	−0.027	−0.6	−3.8	4.4	14.0	14.6
Ohio	−0.013	0.002	−0.015	−0.2	−1.0	2.4	5.1	6.5
South Dakota	−0.001	−0.001	−0.002	0.0	−0.1	0.3	0.6	0.8
Wisconsin	−0.004	0.005	−0.028	−0.8	1.7	4.5	16.8	23.0
Corn Belt	−0.487	0.169	−0.620	−15.5	−40.1	100.4	334.5	394.7

[a] MMT: million metric tons; TMT: thousand metric tons.

[b] Positive number of ΔSOC means SOC accumulation and negative number means SOC depletion.

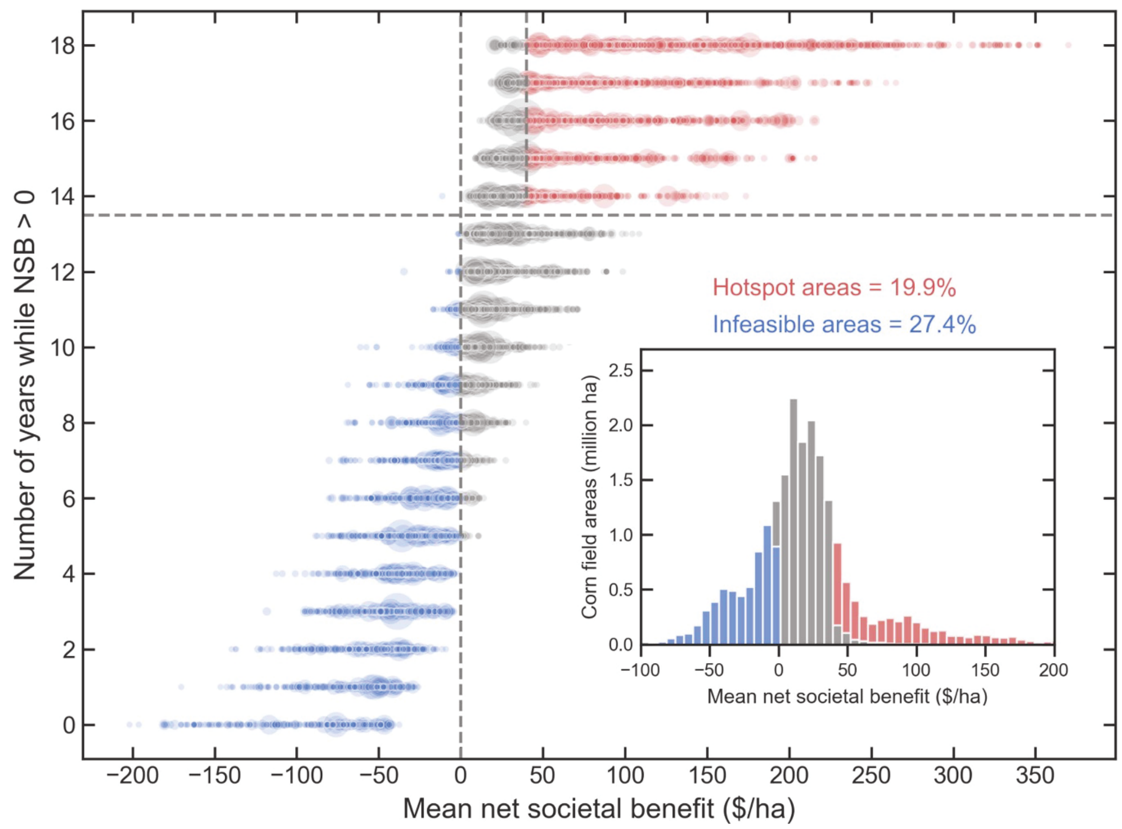

Figure 2. Hotspot areas (red bubble), feasible areas (gray bubble), and infeasible areas (blue bubble) classified by mean net societal benefit (NSB) and number of years with profitable NSB (>0) during the 18 year period. Each bubble represents one soil map unit while bubble size represents corn field areas in each map unit. The NSB is estimated based on comparing 10% N reduction and the baseline simulation.

types, likely because sandy soils have a lower fertility in general. Mean NSB for hotspots is 90.5 \$ ha^{-1}, with sandy soil areas having NSB above the average. Although the clay soil area has the largest yield increase and NSB ha^{-1}, only very small portions of the area are hotspots. In the corn belt, the most common soil type is silt loam, silt clay loam, and loam, covering together over 80% of cornfields. The expected NSB from areas with those soil types are 71.5, 48.3, 89.8 \$ ha^{-1}, respectively.

3.4. Uncertainty in estimates

We repeated the NSB estimates under three alternative management scenarios compared with the baseline scenario, i.e. adopt continuous corn instead of corn-soybean rotation, adopt conventional tillage or conservation tillage instead of no-till. We noted that the distribution of our baseline and alternative scenario simulations for all four indicators were comparable to literature reported values (figure 4), although high uncertainty was observed in both empirical and simulated estimates. Results show that 10% N-fertilizer reduction will avoid more GHG emissions and N leaching, although at a cost of more yield loss, under continuous corn or conventional tillage practices (table 4), but estimates only change slightly under conservation tillage practice. Due to a lack of spatial information, we were not able to estimate the NSB as the area weighted sum of each possible practice. The wide range of uncertainty in estimating NSB (from \$392 to \$562 million USD) highlighted the importance of mapping crop rotation and tillage intensity.

Estimated NSB is also sensitive to assumptions of corn and fertilizer price as well as the pricing for social benefits. We thus assessed 27 combinations of pricing scenarios based on table S4, with full analysis results given in table S6. Briefly, the most beneficial scenario is 'LHH' (low corn price, high fertilizer cost, and high social benefit), which yields \$1271 M of NSB; Low fertilizer cost and low social benefit scenarios (LLL, MLL, HLL) yield similar lowest NSBs (\$115 M, \$114 M, \$128 M). Assuming the same pricing for social benefits, increasing fertilizer price significantly increased the extent of hotspots as well as the total NSB, especially in the lower states (table S6 and figure S9). Varying the MRTN rate also changed the NSB, and the impacts were in general larger with lower-than-optimal rates in a few states (figure S10).

A 20% N-fertilizer reduction scenario was also evaluated to assess potential benefits under this stricter N policy. Overall, NSB under the 20% reduction scenario is higher than that under the 10% reduction scenario. The density curves of mitigation effects

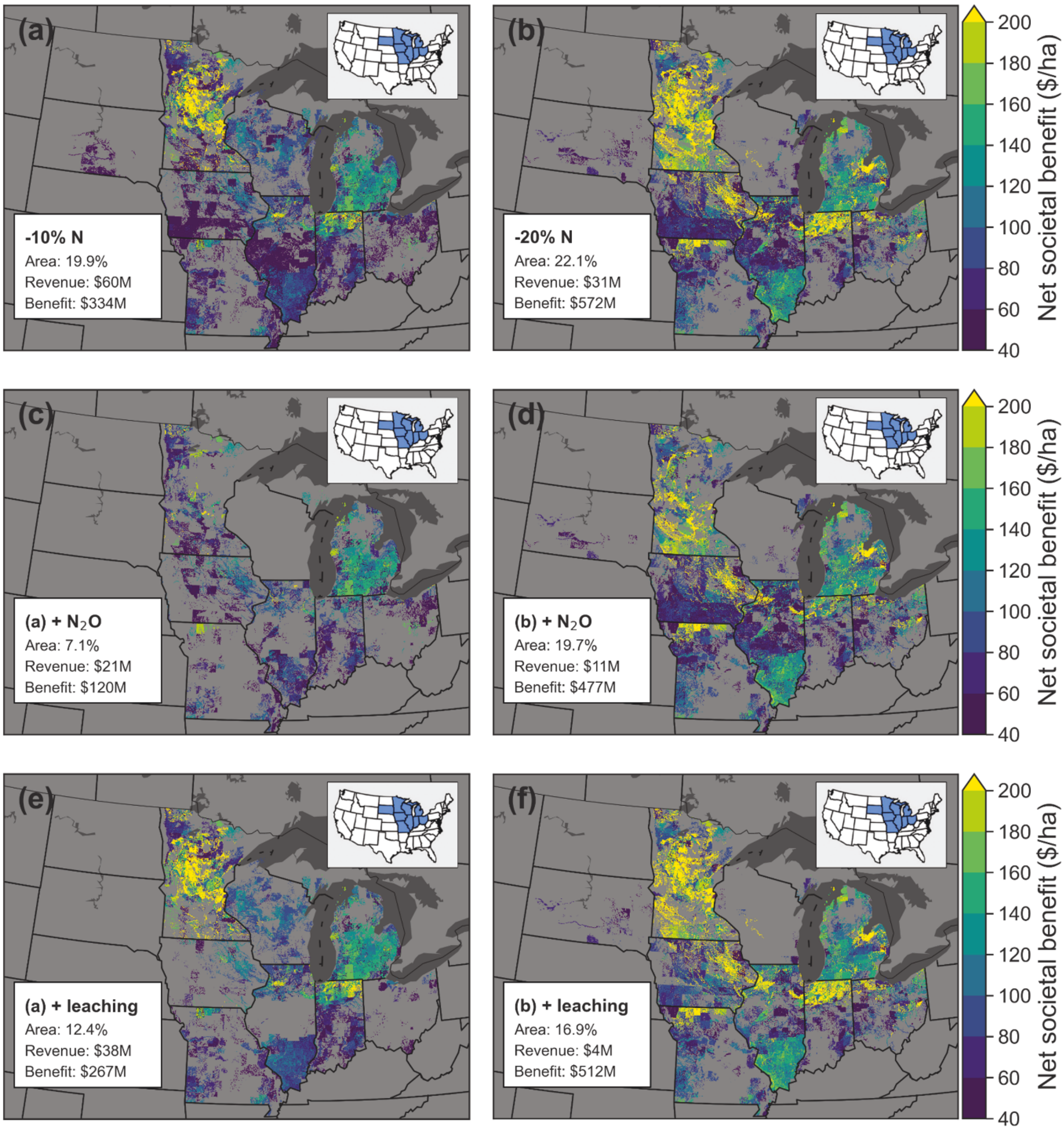

Figure 3. Mitigation potential at hotspot locations under (a) 10% N fertilizer reduction and (b) 20% N fertilizer reduction scenarios. The following analysis demonstrates the hotspots where contributing over 50% of total N_2O emissions ((c), (d)), N leaching ((e), (f)), respectively. 'Area' shows the percentage of hotspots in the Corn Belt region. 'Revenue' shows the farmers' direct revenue changes from related hotspots in each scenario. 'Benefit' shows the social benefit from related hotspots in each scenario. NSB equal to the sum of 'Benefit' and 'Revenue'. Hotspots are defined as areas where NSB is greater than \$40 ha^{-1} in more than 75% of the years during 2001–2018 are feasible (NSB > 0).

under the 20% reduction scenario were flattened compared to that under the 10% reduction scenario (figure S11), which suggested hotspots could deliver more benefits through N mitigation whereas infeasible areas became even more challenging to mitigate.

4. Discussion

In this study, we demonstrated the feasibility of upscaling *ecosys* with machine learning based metamodels and its applications to identifying nitrogen loss hotspots and GHG mitigation potential in the US Corn Belt. Although environmental life cycle assessment has been applied widely to evaluate the environmental burdens associated with food production, existing studies often estimate impacts based on highly aggregated data that fall short of predictive capability (Smith *et al* 2017). These challenges are particularly amplified for systems that are subject to heterogeneous inputs and/or outputs across supply landscapes and time, further complicated by the interactions between soil, weather, cropping history and management practices (McNunn *et al* 2020). The fine-granular estimation of societal costs and benefits by this study, as well as the proposed pipeline to derive so, thus can inform decision-makers on the implications of alternative policies and interventions.

Compared to the original *ecosys* model, our metamodels drastically reduced the computational time and memory requirement. To generate a 20 year

Table 3. Changes in yields, GHG emissions, N leaching for different soil textures in hotspot areas across nine Corn Belt states under 10% N fertilizer reduction from 2001–2018, compared with the baseline simulation.

	Components changes (%)			Hotspot areas		Net societal benefit	
Soil texture	Yield	GHGs[a]	N leaching	(%[b])	Acreage (10^3 ha)	(\$ ha^{-1})	(\$ millions)
Clay	5.88	−10.62	−3.26	2.56	2.0	132.4	0.3
Clay loam	−1.41	−9.77	−9.75	3.69	76.4	54.7	4.2
Loam	−0.21	−15.52	−9.79	16.47	663.2	89.8	59.6
Loam sand	1.25	−14.34	−7.60	84.87	347.3	124.8	43.3
Sand	0.87	−10.47	−6.58	79.77	47.6	108.1	5.1
Sandy clay loam	3.00	−12.78	−3.91	81.31	4.2	96.9	0.4
Sandy loam	1.75	−14.71	−8.33	71.34	938.6	136.8	128.4
Silt clay	−0.35	−9.36	−7.12	5.23	20.1	61.1	1.2
Silt clay loam	−1.15	−9.10	−8.57	8.52	401.6	48.3	19.4
Silt loam	−1.76	−11.30	−11.81	21.06	1,858.5	71.5	132.8
All	−0.61	−12.12	−9.62	19.92	4,359.5	90.5	394.7

[a] Greenhouse gases (GHGs) is aggregated N_2O emission, fertilizer embedded GHG, and SOC depletion converted to GHG.

[b] % of corn acreage in each soil texture.

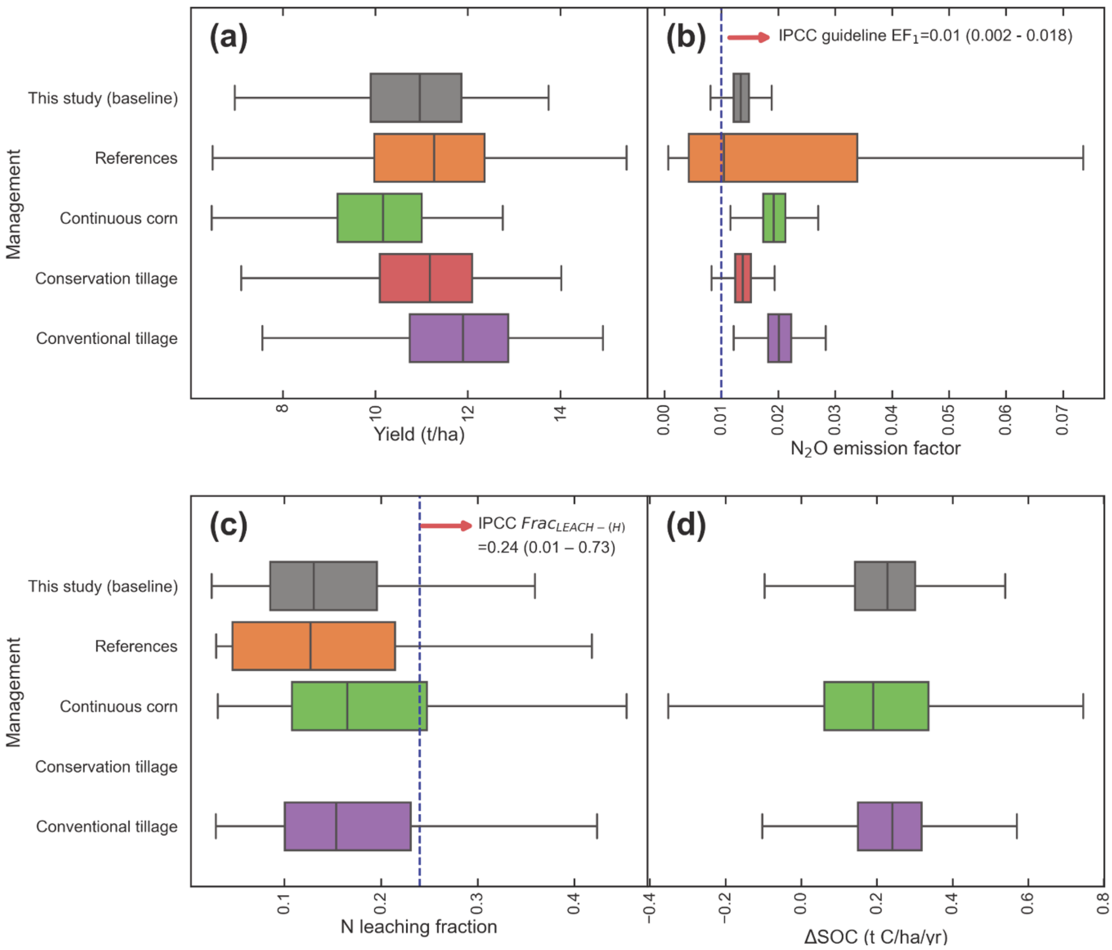

Figure 4. Uncertainty of management effects on (a) yield, (b) N_2O emissions factor, (c) N leaching fraction, and (d) ΔSOC. The reference for yield is county-level corn yield in 2018 from USDA survey. The reference data for N_2O emission factor is collected from the studies in the US Corn Belt ($n = 329$) from Eagle *et al* (2020) and the N leaching fraction is harvested from the studies in the US Corn Belt ($n = 272$) from Chatterjee (2020). N_2O factor and N leaching fraction is calculated using N_2O emission and N leaching divided by N application rate. Both impact factors are compared with IPCC (2019) Refinements (IPCC 2019). Continuous corn effects for ΔSOC were estimated by metamodel in this study. Due to limitation of reliable data, the reference for ΔSOC and conservation tillage effects for N leaching and ΔSOC were not considered.

Table 4. Annual average impacts of 10% and 20% N reduction from the baseline scenario under alternative management practices in hotspot areas of 2001–2018.

Management scenarios	Changes in GHG emissions (MMT CO_2e) [a]			N leaching(TMT N) [a]	Net societal benefit ($ millions)			
	N_2O	ΔSOC[b]	Fertilizer embedded GHG		Corn production	Fertilizer savings	Social benefit	Total
Baseline (10% N reduction)	−0.487	0.169	−0.620	−15.5	−40.1	100.4	334.5	394.7
-Continuous corn	−1.028	0.466	−0.878	−24.1	−98.5	142.2	518.2	561.8
-Conservation tillage	−0.491	0.167	−0.611	−15.4	−38.8	99.0	331.6	391.9
-Conventional tillage	−0.910	0.209	−0.761	−20.4	−76.6	123.2	450.8	497.5
Baseline (20% N reduction)	−1.146	0.458	−1.405	−25.2	−196.2	227.5	571.9	603.3
-Continuous corn	−2.203	1.232	−1.873	−38.7	−319.2	303.4	860.1	844.3
-Conservation tillage	−1.148	0.449	−1.371	−24.9	−191.7	222.1	564.3	594.8
-Conventional tillage	−1.872	0.523	−1.522	−31.3	−251.4	246.6	724.1	719.3

[a] MMT: million metric tons; TMT: thousand metric tons.

[b] Positive number of ΔSOC means SOC accumulation and negative number means SOC depletion.

simulation for one site, *ecosys* required almost 1 h, including approximately 2 min for simulating each year and a spin-up period of 20 min to allow the complex process-based model to reach an equilibrium status. Thus, it took us about 2400 CPU hours (Intel Haswell E5-2680v3) to simply generate the synthetic training data for building the metamodels (i.e. 99 sites over 18 years and 20 different N fertilizer rates in this study). In contrast, once trained, the metamodel can finish 3.2 million simulations (i.e. combinations of about 45 000 unique soil map units, 18 years of weather conditions, and four indicators to calculate the social benefits) within 10 s. More importantly, our metamodels showed high performance in reproducing the aggregated annual C and N fluxes simulated by *ecosys*. This capability provides an excellent opportunity for scenario-based assessments in which millions or even billions of combinations need to be simulated (Banger *et al* 2017, McNunn *et al* 2020).

However, this is not to say metamodel can replace any process-based models, especially when the goal is to understand physiological responses to the changing climate. In fact, the task of emulating annual fluxes from cropland is relatively easy compared to simulating daily fluxes. For example, CO_2 and N_2O fluxes within a field are characterized as 'hot-moments' over the growing season (Waldo *et al* 2019), making it extremely challenging to model with simple machine learning based metamodeling methods. To capture those temporal patterns, more advanced deep learning methods that can handle pulses in time series data such as attention-based hierarchical recurrent neural networks (Qin *et al* 2017) should be considered. Moreover, these data-driven metamodels are not suitable for making predictions under future climate change if climate scenarios are outside the range in training data. To make out-of-sample predictions on yield, ΔSOC and N losses, more physiological and biogeochemical mechanisms should be added to the 'black-box' training of metamodels. This can be potentially achieved by a new method called physics-guided machine learning that fully integrates scientific knowledge embodied in process-based models as well as information from data (Read *et al* 2019), which will be addressed in our future studies.

One major uncertainty source in our estimates is the configuration of field-level management practices. Our baseline and N scenario simulations assumed every field is rainfed, adopts corn-soybean rotation but does not apply tillage. Constrained by data availability, uncertainty was roughly estimated by introducing literature-based modifiers to account for changes in rotation and tillage (table S7). Similarly, due to a lack of spatially explicit information or conclusive experimental results, we were not able to analyze a few other management practices that also affect yields and emissions. For example, adding cover crops has been suggested by several recent studies as a viable mitigation solution (Kaye and Quemada 2017). The impacts of cover crop on crop yield (Alvarez *et al* 2017), N leaching (Thapa *et al* 2018), SOC and N_2O emission (Muhammad *et al* 2019) vary by the type of cover crop, management practices and soil properties. In addition, applying manure instead of synthetic fertilizers can help build soil organic matter and hence SOC (Maillard and Angers 2014), but may increase emissions by 40% (Decock 2014). Last but not least, irrigation is ignored in this study because only 3.8% of cornfields use irrigation in the Corn Belt (Perlman *et al* 2014), but will be an important factor to consider if expanding the analysis to the entire US. Estimating these management practices at field scale over the Corn Belt thus can help reduce the uncertainty in evaluating mitigation strategies. Some recent preliminary success in mapping field-level information on tillage, cover crop and irrigation using high-resolution satellite imagery has shed light on this direction (Peng *et al* 2020). Future integration of these practices may yield new insights.

To achieve sustainable management in the food system, communication among researchers, stakeholders, farmers, and policymakers is important (Reimer *et al* 2017, Zhang *et al* 2020). Through linking food retailers and manufacturers to farmers and suppliers in food supply chains, companies' sustainability efforts can be quantified (Smith *et al* 2017), and initiatives can be designed for farmers to reduce N losses (Eagle *et al* 2020). Our metamodels, by providing key indicators related to these activities, can serve as a useful and usable tool for practitioners. For example, these metamodels could be integrated into FoodS^3 (Smith *et al* 2017), which would enable companies to quantify the emissions during upstream productions and distinguish mitigation options for setting their sustainability goals. To promote mitigation strategies, policymakers and stakeholders need to encourage farmers' incentives via social and economic actions, such as subsidies or carbon offset programs (Sykes *et al* 2020). However, planning the right mitigation strategy is challenging since the ability for multiple proposed practices to sequester additional carbon or avoid emissions varies substantially over the landscape. In this case, the proposed metamodel can support decision making for individual farmers by providing evaluations of N mitigation cost with small computational resources; meanwhile, it could be an effective tool to assist policymakers to balance tradeoffs between social benefits and potential impacts on farm productivity.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgement

The information, data, or work presented herein was funded in part by the National Science Foundation SitS program (Award #: 2034385) and the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0001227 and DE-AR0001382. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORCID iDs

Taegon Kim https://orcid.org/0000-0002-7931-6627
Zhenong Jin https://orcid.org/0000-0002-1252-2514
Licheng Liu https://orcid.org/0000-0002-9649-1056
Yi Yang https://orcid.org/0000-0002-1131-6196
Bin Peng https://orcid.org/0000-0002-7284-3010
Kaiyu Guan https://orcid.org/0000-0002-3499-6382

References

Alvarez R, Steinbach H S and De Paepe J L 2017 Cover crop effects on soils and subsequent crops in the pampas: a meta-analysis *Soil Till. Res.* **170** 53–65
Banger K, Yuan M, Wang J, Nafziger E D and Pittelkow C M 2017 A vision for incorporating environmental effects into nitrogen management decision support tools for U.S. Maize production *Front. Plant Sci.* **8** 1270
Basso B and Liu L 2019 Chapter four—seasonal crop yield forecast: methods, applications, and accuracies *Adv. Agron.* **154** 201–55
Brauman K A, Goodkind A L, Kim T, Pelton R E O, Schmitt J and Smith T M 2020 Unique water scarcity footprints and water risks in US meat and ethanol supply chains identified via subnational commodity flows *Environ. Res. Lett.* **15** 105018
Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
Britz W and Leip A 2009 Development of marginal emission factors for N losses from agricultural soils with the DNDC-CAPRI metamodel *Agric. Ecosyst. Environ.* **133** 267–79
Burke M and Emerick K 2016 Adaptation to climate change: evidence from US agriculture *Am. Econ. J. Econ. Policy* **8** 106–40
Carlson K M *et al* 2017 Greenhouse gas emissions intensity of global croplands *Nat. Clim. Change* **7** 63–8
Chatterjee A 2020 Extent and variation of nitrogen losses from non-legume field crops of conterminous United States *Nitrogen* **1** 34–51
Chen T and Guestrin C 2016 XGBoost: a scalable tree boosting system *Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (*San Francisco, CA, USA, 13–17 August 2016*) (ACM) pp 785–94
Chen Z *et al* 2016 Partitioning N_2O emissions within the U.S. Corn Belt using an inverse modeling approach *Glob. Biogeochem. Cycle* **30** 1192–205
Cho S J and McCarl B A 2017 Climate change influences on crop mix shifts in the United States *Sci. Rep.* **7** 40845
Decock C 2014 Mitigating nitrous oxide emissions from corn cropping systems in the Midwestern U.S.: potential and data gaps *Environ. Sci. Technol.* **48** 4247–56
Eagle A J *et al* 2020 Quantifying on-farm nitrous oxide emission reductions in food supply chains *Earth's Future* **8** e2020EF001504
EPA 2020 Inventory of U.S. Greenhouse gas emissions and sinks: 1990-2018 (available at: www.epa.gov/sites/production/files/2020-04/documents/us-ghg-inventory-2020-main-text.pdf) (Accessed 18 December 2020)
Foley J A *et al* 2011 Solutions for a cultivated planet *Nature* **478** 337–42
Grant R F 2001 A review of the Canadian ecosystem model ecosys *Modeling Carbon and Nitrogen Dynamics for Soil Management* (Boca Raton, FL: CRC Press) pp 173–264
Grant R F, Juma N G, Robertson J A, Izaurralde R C and McGill W B 2001 Long-term changes in soil carbon under different fertilizer, manure, and rotation: testing the mathematical model ecosys with data from the breton plots *Soil Sci. Soc. Am. J.* **65** 205–14
Grant R F, Neftel A and Calanca P 2016 Ecological controls on N_2O emission in surface litter and near-surface soil of a managed grassland: modelling and measurements *Biogeosciences* **13** 3549–71
Grant R F, Pattey E, Goddard T W, Kryzanowski L M and Puurveen H 2006 Modeling the effects of fertilizer application rate on nitrous oxide emissions *Soil Sci. Soc. Am. J.* **70** 235–48
Grant R and Pattey E 2003 Modelling variability in N2O emissions from fertilized agricultural fields. *Soil Biol Biochem* **35** 225–43
Groffman P M, Butterbach-Bahl K, Fulweiler R W, Gold A J, Morse J L, Stander E K, Tague C, Tonitto C and Vidon P 2009 Challenges to incorporating spatially and temporally explicit phenomena (hotspots and hot moments) in denitrification models *Biogeochemistry* **93** 49–77
Horwath W R and Kuzyakov Y 2018 Chapter three—the potential for soils to mitigate climate change through carbon sequestration *Dev. Soil Sci.* **35** 61–92
Ingraham P A and Salas W A 2019 Assessing nitrous oxide and nitrate leaching mitigation potential in US corn crop systems using the DNDC model *Agric. Syst.* **175** 79–87
Interagency Working Group 2016 Technical update on the social cost of carbon for regulatory impact analysis-under executive order 12866 (Washington, DC: Office of Management and Budget)
IPCC 2019 *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories* ed E Calvo Buendia, K Tanabe, A Kranjc, J Baasansuren, M Fukuda, S Ngarize, A Osako, Y Pyrozhenko, P Shermanau and S Federici (Switzerland: IPCC) Published
IPCC 2014 *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* ed IPCC, R K Pachauri and L A Meyer (Geneva: IPCC)
Janssens-Maenhout G, Crippa M, Guizzardi D, Muntean M, Schaaf E, Dentener F, Bergamaschi P, Pagliari V, Olivier J G and Peters J A 2019 EDGAR v4.3.2 global atlas of the three major greenhouse gas emissions for the period 1970–2012 *Earth Syst. Sci. Data* **11** 959–1002
Jin Z, Archontoulis S V and Lobell D B 2019 How much will precision nitrogen management pay off? An evaluation based on simulating thousands of corn fields over the US Corn-Belt *Field Crop. Res.* **240** 12–22
Johnson D M 2013 A 2010 map estimate of annually tilled cropland within the con-terminous United States *Agric. Syst.* **114** 95–105
Johnston R Z, Sandefur H N, Bandekar P, Matlock M D, Haggard B E and Thoma G 2015 Predicting changes in yield and water use in the production of corn in the United States under climate change scenarios *Ecol. Eng.* **82** 555–65
Jungers J M, DeHaan L H, Mulla D J, Sheaffer C C and Wyse D L 2019 Reduced nitrate leaching in a perennial grain crop compared to maize in the Upper Midwest, USA *Agric. Ecosyst. Environ.* **272** 63–73
Kaye J P and Quemada M 2017 Using cover crops to mitigate and adapt to climate change. A review *Agron. Sustain. Dev.* **37** 4

Keating B A and Thorburn P J 2018 Modelling crops and cropping systems—evolving purpose, practice and prospects *Eur. J. Agron.* **100** 163–76

Kent C, Pope E, Thompson V, Lewis K, Scaife A A and Dunstone N 2017 Using climate model simulations to assess the current climate risk to maize production *Environ. Res. Lett.* **12** 054012

Kwon H, Ugarte C M, Ogle S M, Williams S A and Wander M M 2017 Use of inverse modeling to evaluate CENTURY-predictions for soil carbon sequestration in US rain-fed corn production systems *PLoS One* **12** e0172861

Lobell D B, Deines J M and Di Tommaso S 2020 Changes in the drought sensitivity of US maize yields *Nat. Food* **1** 729–35

Lu C, Yu Z, Tian H, Hennessy D A, Feng H, Al-Kaisi M, Zhou Y, Sauer T and Arritt R 2018 Increasing carbon footprint of grain crop production in the US Western Corn Belt *Environ. Res. Lett.* **13** 124007

Maillard É and Angers D A 2014 Animal manure application and soil organic carbon stocks: a meta-analysis *Glob. Change Biol.* **20** 666–79

Mandrini G, Bullock D S and Martin N F 2021 Modeling the economic and environmental effects of corn nitrogen management strategies in Illinois *Field Crop. Res.* **26** 108000

Marshall E, Aillery M, Malcolm S and Williams R 2015 *Climate Change, Water Scarcity, and Adaptation in the US Field Crop Sector* (United States Department of Agriculture, Economic Research Service)

McNunn G, Karlen D L, Salas W, Rice C W, Mueller S, Muth D and Seale J W 2020 Climate smart agriculture opportunities for mitigating soil greenhouse gas emissions across the U.S. Corn-Belt *J. Clean Prod.* **268** 122240

Mekonnen B A, Nazemi A, Mazurek K A, Elshorbagy A and Putz G 2015 Hybrid modelling approach to prairie hydrology: fusing data-driven and process-based hydrological models *Hydrol. Sci. J.* **60** 1473–89

Metivier K A, Pattey E and Grant R F 2009 Using the ecosys mathematical model to simulate temporal variability of nitrous oxide emissions from a fertilized agricultural soil *Soil Biol. Biochem.* **41** 2370–86

Morris T F *et al* 2018 Strengths and limitations of nitrogen rate recommendations for corn and opportunities for improvement *Agron. J.* **110** 1–37

Muhammad I, Sainju U M, Zhao F, Khan A, Ghimire R, Fu X and Wang J 2019 Regulation of soil CO_2 and N_2O emissions by cover crops: a meta-analysis *Soil Till. Res.* **192** 103–12

Nolan B T, Green C T, Juckem P F, Liao L and Reddy J E 2018 Metamodeling and mapping of nitrate flux in the unsaturated zone and groundwater, Wisconsin, USA *J. Hydrol.* **559** 428–41

PE International 2014 GaBi LCA software (available at: www.gabi-software.com/merica/index/)

Pelton R 2019 Spatial greenhouse gas emissions from US county corn production *Int. J. Life Cycle Assess.* **24** 12–25

Peng B *et al* 2020 Towards a multiscale crop modelling framework for climate change adaptation assessment *Nat. Plants* **6** 338–48

Peng B, Guan K, Chen M, Lawrence D M, Pokhrel Y, Suyker A, Arkebauer T and Lu Y 2018 Improving maize growth processes in the community land model: implementation and evaluation *Agric. For. Meteorol.* **250–251** 64–89

Perlman J, Hijmans R J and Horwath W R 2014 A metamodelling approach to estimate global N_2O emissions from agricultural soils *Glob. Ecol. Biogeogr.* **23** 912–24

Poffenbarger H J, Barker D W, Helmers M J, Miguez F E, Olk D C, Sawyer J E, Six J and Castellano M J 2017 Maximum soil organic carbon storage in Midwest US cropping systems when crops are optimally nitrogen-fertilized *PLoS One* **12** e0172293

Qin Y, Song D, Chen H, Cheng W, Jiang G and Cottrell G W 2017 A dual-stage attention-based recurrent neural network for time series prediction *Proc. of the Twenty-Sixth Int. Joint Conf. on Artificial Intelligence*

Read J S *et al* 2019 Process-guided deep learning predictions of lake water temperature *Water Resour. Res.* **55** 9173–90

Reimer A, Doll J E, Basso B, Marquart-Pyatt S T, Robertson G P, Stuart D and Zhao J 2017 Moving toward sustainable farming systems: insights from private and public sector dialogues on nitrogen management *J. Soil Water Conserv.* **72** 5A–9A

Revesz R, Greenstone M, Hanemann M, Livermore M, Sterner T, Grab D, Howard P and Schwartz J 2017 Best cost estimate of greenhouse gases *Science* **357** 655

Sawyer J, Nafziger E, Randall G, Bundy L, Rehm G and Joern B 2006 Concepts and rationale for regional nitrogen rate guidelines for corn (Iowa: Iowa State University Extension) (available at: http://publications.iowa.gov/3847/1/PM2015.pdf) (Accessed 01 October 2020)

Shahhosseini M, Martinez-Feria R, Hu G and Archontoulis S V 2019 Maize yield and nitrate loss prediction with machine learning algorithms *Environ. Res. Lett.* **14** 124026

Smith T M, Goodkind A L, Kim T, Pelton R E O, Suh K and Schmitt J 2017 Subnational mobility and consumption-based environmental accounting of US corn in animal protein and ethanol supply chains *Proc. Natl Acad. Sci. USA* **114** E7891–9

Sobota D J, Compton J E, McCrackin M L and Singh S 2015 Cost of reactive nitrogen release from human activities to the environment in the United States *Environ. Res. Lett.* **10** 025006

Soil Survey Staff 2020 Gridded soil survey geographic (gSSURGO) database for the United States of America and the Territories, Commonwealths, and Island Nations served by the USDA-NRCS United States Department of Agriculture, Natural Resources Conservation Service (available at: https://gdg.sc.egov.usda.gov/) (Accessed 15 June 2020)

Sykes A J *et al* 2020 Characterising the biophysical, economic and social impacts of soil carbon sequestration as a greenhouse gas removal technology *Glob. Change Biol.* **26** 1085–108

Thapa R, Mirsky S B and Tully K L 2018 Cover crops reduce nitrate leaching in agroecosystems: a global meta-analysis *J. Environ. Qual.* **47** 1400–11

Tian H *et al* 2019 Global soil nitrous oxide emissions since the preindustrial era estimated by an ensemble of terrestrial biosphere models: magnitude, attribution, and uncertainty *Glob. Change Biol.* **25** 640–59

Tilman D, Balzer C, Hill J and Befort B L 2011 Global food demand and the sustainable intensification of agriculture *Proc. Natl Acad. Sci. USA* **108** 20260–4

Turner P A, Griffis T J, Lee X, Baker J M, Venterea R T and Wood J D 2015 Indirect nitrous oxide emissions from streams within the US Corn Belt scale with stream order *Proc. Natl Acad. Sci. USA* **112** 9839–43

USDA National Agricultural Statistics Service (NASS) 2019 2017 census of agriculture Complete data (available at: www.nass.usda.gov/AgCensus) (Accessed 29 October 2019)

USDA-FAS 2020 World agricultural production (Washington, DC: USDA FAS) (available at: www.fas.usda.gov/data/world-agricultural-production)

Villa-Vialaneix N, Follador M, Ratto M and Leip A 2012 A comparison of eight metamodeling techniques for the simulation of N_2O fluxes and N leaching from corn crops *Environ. Modell. Softw.* **34** 51–66

Waldo S, Russell E S, Kostyanovsky K, Pressley S N, O'Keeffe P T, Huggins D R, Stöckle C O, Pan W L and Lamb B K 2019 N_2O Emissions from two agroecosystems: high spatial variability and long pulses observed using static chambers and the flux-gradient technique *J. Geophys. Res. Biogeosci.* **124** 1887–904

Wang Y, Ying H, Yin Y, Zheng H and Cui Z 2019 Estimating soil nitrate leaching of nitrogen fertilizer from global meta-analysis *Sci. Total Environ.* **657** 96–102

Wienhold B J, Jin V L, Schmer M R and Varvel G E 2018 Soil carbon response to projected climate change in the US Western Corn Belt *J. Environ. Qual.* **47** 704–9

Xia Y *et al* 2012 Continental-scale water and energy flux analysis and validation for the North American land data assimilation system project phase 2 (NLDAS-2): 1. Intercomparison and application of model products *J. Geophys. Res.* **117** D03109
Xu H, Sieverding H, Kwon H, Clay D, Stewart C, Johnson J M F, Qin Z, Karlen D L and Wang M 2019 A global meta-analysis of soil organic carbon response to corn stover removal *GCB Bioenergy* **11** 1215–33
Zhang X, Davidson E A, Zou T, Lassaletta L, Quan Z, Li T and Zhang W 2020 Quantifying nutrient budgets for sustainable nutrient management *Glob. Biogeochem. Cycle* **34** e2018GB006060
Zhao X, Christianson L E, Harmel D and Pittelkow C M 2016 Assessment of drainage nitrogen losses on a yield-scaled basis *Field Crops Res.* **199** 156–66
Zhou W, Guan K, Peng B, Tang J, Jin Z, Jiang C, Grant R and Mezbahuddin S 2021 Quantifying carbon budget, crop yields and their responses to environmental variability using the ecosys model for U.S. Midwestern agroecosystems *Agric. For. Meteorol.* 108521 (accepted)
Zomer R J, Bossio D A, Sommer R and Verchot L V 2017 Global sequestration potential of increased organic carbon in cropland soils *Sci. Rep.* **7** 15554