LETTER

Boosting climate change mitigation potential of perennial lignocellulosic crops grown on marginal lands

R A Martinez-Feria, B Basso and S Kim

1 Department of Earth and Environmental Science, Michigan State University, East Lansing, MI, 48823, United States of America
2 W. K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI, 49060, United States of America
3 Department of Chemical Engineering and Materials, Michigan State University, East Lansing, MI, 48823, United States of America

E-mail: basso@msu.edu

Keywords: emulators, nitrogen, marginal lands, biofuels, switchgrass, ensemble models

Abstract

Nitrogen fertilizer ($N_F$) is a major uncertainty surrounding the greenhouse gas (GHG) emissions of lignocellulosic biofuels. $N_F$ enhances agronomic yields and soil C inputs via plant litters, but results in soil organic carbon (SOC) decomposition, soil N$_2$O fluxes, and a large fossil energy footprint. Thus, whether $N_F$ is beneficial or detrimental to the GHG mitigation of biofuels is unknown. Here, we show the potential GHG mitigation of fertilizing switchgrass (*Panicum virgatum*) at the $N_F$ rate that minimizes net GHG emissions across 7.1 million ha of marginal lands in the Midwest US, with long-term production advantages surpassing emitted GHG by 0.66 Mg CO$_2$e ha$^{-1}$ yr$^{-1}$ on the aggregate. Marginal lands limited by poor N fertility could see a much greater benefit, but not SOC-rich lands, limited by low precipitation, or short growing seasons. The objectives of maximizing yield and minimizing GHG overlap only in a few environments, suggesting that maximum yield will reduce the climate benefit of cellulosic biofuels.

1. Introduction

Renewable lignocellulosic biofuels are one of the few mature technologies with significant potential to reduce fossil fuel consumption and greenhouse gas (GHG) emissions in the transportation sector (Gelfand et al 2020). Supplied sufficient lignocellulose feedstock to satisfy mandates for low-C biofuels will require massive land-use change, possibly involving the conversion of millions of hectares of existing cropland and marginally productive sites into perennial biomass crops (Gelfand et al 2013). Land-use change in such a scale, however, could have significant impacts on food security, biodiversity, and GHG emissions if bioenergy production directly or indirectly displaces food crops or sensitive habitats (Fargione et al 2008, Melillo et al 2009, El Akkari et al 2018). Furthermore, the potential carbon neutrality of lignocellulose-derived biofuel hinges on the ability of the production system to offset upstream and downstream supply-chain GHG emissions with the sequestration of atmospheric CO$_2$ in soil organic carbon (SOC) pools (Robertson et al 2011), which is a process highly dependent on the pedo-climatic conditions (Agostini et al 2015, Paustian et al 2016). Thus, careful selection of sites and management practices used for feedstock cultivation has become the core of strategies seeking the sustainable and efficient deployment of lignocellulose-based bioenergy systems (Somer ville et al 2010, Gelfand et al 2013, 2020, Field et al 2018, O’Neill and Maravelias 2021).

Nitrogen fertilizer ($N_F$) stands as one of the major uncertainties in the production of perennial bioenergy crops. Application of $N_F$ can increase plant growth and agronomic yields (Mclaughlin and Kszos 2005, Ruan et al 2016), and the additional biomass may contribute to climate mitigation by increasing availability of biofuel feedstocks and improving efficiency of biofuel supply chains (O’Neill and Maravelias 2021). However, the responses of perennial bioenergy crops to $N_F$ additions are quite variable and poorly understood (Fike et al 2017, Monti et al 2019). Enhancing plant productivity with fertilizer additions also generally means increased C inputs to
soils via plant litters (Miesel et al 2017), which may result in greater rates of SOC sequestration (Poffenbarger et al 2017). Yet, evidence suggests that in environments with low soil N availability to plants and microbes, N fertilizer additions can accelerate the rates of SOC decomposition and reduce soil C stocks, particularly in the mineral-associated fractions (Valdez et al 2017), with large uncertainty of whether N fertilizer will result in higher or lower SOC sequestration at a given candidate site. Moreover, N fertilizer also promotes microbial nitrification and denitrification processes, resulting in the release of nitrous oxide (N2O), a potent GHG with a global warming potential ~265 times that of CO2 (IPCC 2014). These emissions are on top of those incurred by the fossil fuel energy used for manufacturing, transportation, and application of the N fertilizer itself, which on the aggregate could negate the GHG mitigation benefit from any additional feedstock produced (Crutzen et al 2008).

From a climate mitigation perspective, the question of whether bioenergy crops ought to receive N fertilizer—and at what rate—needs to balance the benefits and costs of the practice in terms of GHG emissions. The objective here is to minimize the net GHG contribution of N fertilization to the life-cycle greenhouse warming intensity (GWIs, Mg CO₂e ha⁻¹ yr⁻¹) of the producing lignocellulosic biofuel. That can be accounted as the differences in GWI between the fertilized (GWIs) and unfertilized (GWIf) systems, as shown in equation (1):

\[
\text{GHG}_{\text{Net}} = \text{GWIs} - \text{GWIf}. \tag{1}
\]

Solving this expression (methods) reveals that GHGNet can be quantified if we know how much additional GHG is emitted due to N fertilizer synthesis and distribution (GHG_{FertSyn}), field application (GHG_{FertApp}), the changes in soil N₂O emissions (ΔGHG_{N₂O}), SOC sequestration (ΔGHG_{ΔSOC}), and the negative emissions due to the additional replacement of gasoline by the biofuel (ΔGHG_{BioFuel}):

\[
\text{GG}_{\text{Net}} = \text{GHG}_{\text{FertSyn}} + \text{GHG}_{\text{FertApp}} + \Delta\text{GHG}_{\text{N}2\text{O}} \\
+ \Delta\text{GHG}_{\Delta\text{SOC}} + \Delta\text{GHG}_{\text{BioFuel}} \tag{2}
\]

here we examine the potential GHG mitigation benefit of fertilizing perennial bioenergy crops at the N fertilizer rate that minimizes GHGNet, at the US Midwest regional scale. Our analysis focuses on the production of upland ecotypes of switchgrass (Panicum virgatum), the leading candidate crop for deployment in the region (McLaughlin and Kszos 2005), across 7.1 million ha of non-irrigated marginal lands identified in the Midwest. Understanding the magnitude of the potential benefits at each candidate production site, as well as their context-specific drivers, can aid in the design of targeted agronomic recommendations and policy incentives for the effective use of N fertilizer in bioenergy feedstock production.

2. Methods

2.1. Predictive model development

Given that data on upland switchgrass productivity and ΔSOC is available only for a few experimental locations in the Midwest, we project its N fertilizer response to heterogeneous soils, climate and management using a mixed-modeling approach that combines the strengths of process-based crop simulation models (CSMs) and machine learning (ML; figure 1(a)). To do this, we trained ML algorithms to replicate the ensemble mean of four CSMs using weather, soil, and management data summaries readily computed from available public databases. We selected the gradient boosting machine (GBM) among several algorithms examined because of its superior performance in prediction with this simulated dataset (supplementary figure 1 available online at stacks.iop.org/ERL/17/044004/mmedia). The CSM ensemble consisted of four process-based models (APSIM (Ojeda et al 2017), DAYCENT (Field et al 2016), EPIC (Jones et al 2018) and SALUS (Martinez-Feria and Basso 2020), all of which have been previously used to simulate upland ecotypes of switchgrass in the Midwest, and represent diverse model structures. We tested each of these CSMs and their ensemble against a dataset with plant and SOC measurements that encompass three decades of experimentation in the region and a range of N fertilizer rates (figures 1(b), (c) and supplementary table 1). Further details about the development of the predictive models for switchgrass yield and ΔSOC are included in the supplemental information.

2.2. In-silico nitrogen fertilizer rate experiments on marginal lands

We applied the developed predictive models to estimate geographical variations in switchgrass agronomic yield and ΔSOC sequestration on Midwestern marginal lands across a gradient of N fertilizer additions. Growing bioenergy crops on marginal lands has been proposed as a strategy (Gelfand et al 2013) to avoid indirect emissions occur when biofuels production displaces agricultural production and causes additional land-use change that leads to an increase in GHG emissions elsewhere (Melillo et al 2009). Here, we consider marginal lands that could be good candidates for switchgrass establishment as those with soils under recent or current agricultural use, but with very severe limitations to row-crop production. Similar to previous studies (Gelfand et al 2013, Liu and Basso 2017), we identified these prospective sites by selecting fields that were planted with corn at least once during the period between 2008 and 2018 according to the crop frequency layer in the CropScape database (Han et al 2012). Then, we subset these soils to only those identified with land capability class (LCC) IV–VIII in the SSURGO database (Soil Survey Staff) (see supplementary table 2 for detailed
Figure 1. Approach used to project yields and ΔSOC of switchgrass upland varieties across marginal lands in the US Midwest. (a)–(f) Graphical summary of the approach. Four process-based CSMs were tested against measurements (b) from 28 experimental sites (c) and (d), and then used to generate simulations across a sample of 5000 agricultural sites. The CSM ensemble (e) was used to train a ML algorithm, GBM (f). The ML emulation technique reduced the computational burden of the in-silico experiments across a large set of marginal-land locations, weather-years, and management, allowing us to efficiently scale up the bottom-up estimates derived from the CSM ensembles without compromising accuracy. Color shading in (c) shows concentration of the marginal lands identified (methods), aggregated at the 0.01° spatial resolution. Line-range in (e) indicate the spread in the predictions among the CSM models.

LCC descriptions). Because we assume that marginal lands are not typically irrigated, we use the dominant capability class for non-irrigated conditions for each soil map unit (variable name: niccdcd). In total, area used for this experiment consists of 7.08 million ha distributed across the 12-state region (figure 1 and supplementary figure 2). The marginal land identified was layered with gridded NLDAS2 dataset (Xia et al 2012) (0.125° × 0.125° resolution), which resulted in 175 130 unique ‘experimental’ units, each representing an area of up to 6235 ha, or 40.4 ha on average. For each of these units, we conducted experiments in which we vary $N_F$ application rates from 0 to 200 kg N ha$^{-1}$, in 10 kg N ha$^{-1}$ increments, while randomly sampling other management inputs from the same distributions as for the crop model simulations (see supplementary text and supplementary figure 3). The prediction time windows were 20 randomly selected, contiguous weather years from 1980 to 2019. Because we are interested in characterizing responses of established stands, we discarded the first 2 years of the prediction. In total this amounted to more than 63 million individual predictions.

2.3. Estimating net emissions associated with nitrogen fertilizer additions

The impact of fertilizer additions on the life-cycle emissions of biofuel production was calculated as the net difference between the total GWI under a given $N_F$ rate minus the GWI under unfertilized
management ($N_0$) as shown in equation (1). The GWI of switchgrass production can be accounted as the sum of the GHG emitted by fossil fuel used for field operations ($GHG_{Fuel}$), seed production ($GHG_{Seed}$), agrochemicals other than $N_F$ ($GHG_{AgCh}$) and $N_F$ synthesis ($GHG_{FertSyn}$), $N_F$ application operations ($GHG_{FertApp}$), the biogenic direct and indirect $N_2O$ emissions ($GHG_{N2O}$) and changes in soil C sequestration ($GHG_{\Delta SOC}$) associated with fertilizer management, and downstream emissions related to conversion of the biomass feedstock into biofuel ($GHG_{Feedstock}$):

$$GHG_{Net} = [GHG_{FertApp} + GHG_{FertSyn} + GHG_{N2O} + GHG_{\Delta SOC} + GHG_{BioFuel}]$$

$$= GHG_{FertApp} + GHG_{FertSyn} + [GHG_{N2O} - GHG_{N2O}] + [GHG_{\Delta SOC} - GHG_{\Delta SOC}]$$

$$+ [GHG_{BioFuel} - GHG_{BioFuel}].$$

$GHG_{FertSyn}$ was assumed to be 4.5 kg CO$_2$ per kg of N, and $GHG_{FertApp}$ as 26 kg CO$_2$ e ha$^{-1}$. Direct and indirect $N_2O$ emissions were followed using the tier 1 method in the IPCC Guidelines for National Greenhouse Gas Inventories (IPCC 2019), which include $N_2O$ from nitrogen fertilizer, dead biomass residues and SOC loss (supplemental table 3). $GHG_{\Delta SOC}$ were the predicted yearly changes in SOC times 44/12 to convert to CO$_2$ e. Finally, for a given $N_F$, $GHG_{BioFuel}$ is calculated as the product of the predicted agronomic biomass yield ($Yield_{Ag}$, Mg dm ha$^{-1}$) and the net GHG offset ($GHG_{offset}$, Mg CO$_2$ e Mg$^{-1}$ dm), that is, the emissions avoided from gasoline combustion:

$$GHG_{offset} = \frac{GHG_{Tranp}}{1 - LT} + \frac{GHG_{Storage} + Yield_{BioFuel} \times NCV_{BF} \times (GHG_{BioRef} + GHG_{Dist} + GHG_{Combust} - GWI_{Gasoline})}{(1 - LT) \times (1 - LS)}$$

where $LT$ and $LS$ (unitless) are the portion of biomass lost during transportation and storage, respectively; $Yield_{BioFuel}$ is the yield from biomass-to-ethanol conversion, $NCV_{BF}$ is the net caloric value of bioethanol and $GW_{Gasoline}$ is the life-cycle GHG emissions intensity of gasoline. Values for each of the parameters in equation (6) were populated from values reported in the GREET (Greenhouse gases, Regulated Emissions, and Energy use in Transportation) model database (Argonne National Laboratory 2020) and studies shown on supplementary table 3.

3. Results

3.1. Projected switchgrass response to nitrogen fertilizer additions on marginal lands

Consistent with previous CSM intercomparison studies (Rosenzweig et al 2013, Bassu et al 2014, Martre...
et al. 2015, Basso et al. 2018), the ensemble means of the CSMs proved superior to the prediction of any single model, reducing root mean square error of the prediction by 16% and 12% for yield and ΔSOC, respectively (supplementary figure 4). Loss of prediction accuracy by the GBM emulator was negligible (figure 1(f)), as demonstrated by the performance against the measured data which was similar to that of the CSM ensemble (supplementary figure 5).

Projected across all the marginal lands identified in the Midwest and 20 years of weather (figure 2), site-mean agronomic yields without N$_F$ are expected to range between 2.1 and 11.9 Mg ha$^{-1}$, averaging at 5.6 Mg ha$^{-1}$ (figure 2(a)). Our predictions, however, show substantial spatial variability due to differences in soil characteristics and climates (figure 2(b)). Highest yields of unfertilized switchgrass are predicted mainly in the upper Midwest, eastern areas of the Northern plains, and scattered pockets of Indiana and Ohio. Unfertilized yields are projected to be relatively low in the drier parts of the central and northern plains (west of the 99th meridian) and western Illinois. Further, most marginal land cultivated with upland switchgrass (~95%) is expected to accumulate SOC under unfertilized conditions, with ΔSOC projected across all marginal land averaging 0.58 Mg C ha$^{-1}$ yr$^{-1}$ (figure 2(d)). Negative ΔSOC under unfertilized conditions is predicted in some areas in North Dakota and South Dakota, and isolated areas in Minnesota, Wisconsin, Ohio and Michigan (figure 2(e)).

Analysis of emulator predictions reveals that N$_F$ rate is among the most important determinants of agronomic yields, ranking 3rd most important among all 40 variables included, and accounting for 17% of the total variance (supplementary figure 6). With annual applied N$_F$ (up to 200 kg N ha$^{-1}$) upland switchgrass is modeled to respond positively in most marginal sites following diminishing-returns pattern. With full N$_F$, the projected long-term mean yield boost averages 3.1 Mg ha$^{-1}$ (or 54%), although yield increases are generally predicted to be small beyond 100 Kg N ha$^{-1}$ rate (12% on average, figure 2(c)). This result is supported by findings from literature syntheses that suggests that 100–180 kg N ha$^{-1}$ are
generally needed to maximize yields (Parrish and Fike 2005, Chen et al 2019, Monti et al 2019).

Conversely, ∆SOC is projected to have a decreasing trend with increasing N_F rate. This is consistent with the theory that N_F in switchgrass has an overall dampening effect on the sequestration potential of in soil C pools via enhanced SOC decomposition (Valdez et al 2017). Yet, the mean effect is minimal, decreasing from 0.58 Mg C ha\(^{-1}\) yr\(^{-1}\) with no N_F to 0.46 Mg C ha\(^{-1}\) yr\(^{-1}\) with full N_F (figure 2(f)). In fact, N_F rate seems to have much less influence in the emulator for predicting ∆SOC, ranking 9th in variable importance and accounting for less than 2% of the variance (supplementary figure 6). Nevertheless, both predicted responses to N_F exhibit substantial spatial variation, explained in part by fundamental differences in soil characteristics and climates. For example, the yield response to N_F is expected to be much more marked in sites with high precipitation, coarser texture and low SOC, whereas the negative ∆SOC response is more pronounced with increasing stand age, in sites with low precipitation and in soils with mesic levels of SOC (∼1%–2% C; supplementary figure 7).

3.2. Optimized nitrogen fertilizer rate based on minimizing net GHG emissions

Next, we contextualize these projected site-specific N_F responses in terms of how they could impact the GHG emissions of the biofuel production system as defined in equation (2). In this calculation, we include the projected changes in yields and SOC sequestration, and the biogenic N_2O emissions due to N_F application using the IPCC tier 1 methodology (IPCC 2019). In addition, we apply a life cycle assessment (LCA) methodology to account for the upstream emissions due to fertilizer synthesis, transportation and application, and changes in the downstream emissions related to the manufacturing of the biofuel, its combustion, and the value replacement of gasoline fuel (methods).

A graphical example of the calculated net GHG balance as a function of N_F rates is shown on figure 3. Here we display the mean GHG balances across all marginal land sites. In figure 3(a) we show both GHG\(_{\text{FertSyn}}\) and GHG\(_{\text{N2O}}\) increasing monotonically with N_F applied, whereas GHG\(_{\text{BioFuel}}\) (i.e. the emissions offset credit due to the additional biofuel produced) becomes more negative mirroring the yield response (figure 2(c)). Furthermore, GHG\(_{\Delta C} \) is counted as net positive emissions, given the small dampening effect of N_F on the SOC sequestration rate predicted by the model (figure 2(f)). Including all GHG sources and credits indicates net GHG emissions mitigation (i.e. negative GHG\(_{\text{Net}}\)) for a wide range of N_F levels up to 140 kg N ha\(^{-1}\).

Focusing solely on the average response, however, would masks important variation. Indeed, further examination of a gradient of SOC content levels reveals that yield increases in sites with greater than 2% SOC in the top 25 cm are, on average, predicted to be insufficient to overcome the additional life-cycle emissions incurred by applying fertilizer (figure 3(b)). Thus, from a GHG mitigation perspective, these C-rich soils should not generally receive N_F. Yet, these sites represent a small fraction of marginal lands in our dataset (<5%). Optimal N rates are projected to increase monotonically with decreasing topsoil SOC content, up to ∼120 kg N ha\(^{-1}\) in soils with very low topsoil SOC content (figure 3(b)).

![Figure 3](https://example.com/figure3.png)

Figure 3. Example of N fertilization rate optimization based on minimization of net GHG emissions. (a) Net balance between the avoided gasoline due to the additional biofuel produce, and the additional emissions associated to nitrogen fertilizer synthesis, changes in SOC sequestration, and release of soil N_2O from soils. Responses are averaged across all marginal lands. (b) Change in the net GHG across fertilizer levels for a gradient of SOC content in the topsoil. Line indicates changes in the optimum predicted for each level of SOC content. Histogram shows the distribution of binned SOC content levels across the marginal lands examined.
Across all marginal lands, the modeled optimum N<sub>F</sub> rate averages at 74.8 kg N ha<sup>−1</sup>, although its distribution is quite wide and spatially variable as shown in figure 4(a). Areas with low optimal rates (<50 kg N ha<sup>−1</sup>) are mainly projected in western states, and large areas in Minnesota, southern Iowa and northern Missouri, and isolated areas in the eastern states. These low optimal rates occur mostly in sites with low precipitation, SOC-rich soils, or a combination of both as is the case of areas of North and South Dakota. Indeed, these conditions in many marginal lands in the Great Plains, and to a lesser extend areas of Missouri and northern Wisconsin, lead to environments where the optimum N<sub>F</sub> rate is much less likely to be greater than zero (figure 4(b)). On the other hand, high optimal N<sub>F</sub> rate (>100 kg N ha<sup>−1</sup>) is projected mainly for the eastern states, particularly in the southern tip of Indiana, Illinois and Missouri (figure 4(a)).

With optimal N<sub>F</sub> rate, agronomic yields across all marginal lands are expected to increase by an average of 1.99 Mg ha<sup>−1</sup> (figure 4(c)), which aggregated at the regional scale would amount to 14.1 million tons of additional lignocellulosic feedstock. Accounting for transportation and storage biomass losses and assuming a conversion yield of 355 l per Mg of biomass at the biorefinery (Argonne National Laboratory 2020) (supplementary table 3), the boost in productivity would amount to nearly 4.6 million m<sup>3</sup> of additional bioethanol. This represents about 12% of the total statutory cellulosic biofuel production goal of 39 million m<sup>3</sup> for 2020 in the US (National Research Council 2012). The optimal N<sub>F</sub> is also projected to have a more mixed effect on SOC sequestration, with only a few marginal lands seeing increase SOC sequestration, particularly in the Great Lakes states (Michigan, Minnesota, Wisconsin), areas of Iowa, and the southern tip of Indiana, Illinois and Missouri. Overall, however, the effect for the entire region is projected to be neutral, with a mean change in soil C sequestration rate of 0.006 Mg C ha<sup>−1</sup> yr<sup>−1</sup>.

Given these responses, optimum N<sub>F</sub> additions would mitigate emissions by 0.66 Mg CO<sub>2</sub>e ha<sup>−1</sup> yr<sup>−1</sup> on average across all marginal lands (figure 5(a)). Areas surrounding the Great Lakes are projected to have the greatest N<sub>F</sub>-related GHG mitigation potential (up to 2.8 Mg CO<sub>2</sub>e ha<sup>−1</sup> yr<sup>−1</sup>), likely a result of concomitant gains in SOC sequestration and a large yield response to N<sub>F</sub> (figure 4). On the aggregate, optimal NF additions could provide up to
4.63 Tg CO₂e yr⁻¹ of GHG mitigation compared to unfertilized switchgrass production across all marginal lands examined.

Overall, these results reveal the significant contributions that optimal Nₑ could provide to the GHG mitigation of future lignocellulosic biofuel production in the Midwest. However, these projections should be considered as a hypothetical ceiling given that on-the-ground implementation of site-specific optimal Nₑ management could face important challenges due to the uncertainty in production costs (Fike et al 2017). Therefore, we further examine how optimum Nₑ would compare to a more simplistic blanket recommendation of 75 kg N ha⁻¹. This value not only represents the average optimum Nₑ rate across all marginal lands, but also approximates the long-term average N removal of the harvested biomass (Monti et al 2019).

The blanket-rate strategy is projected to result in a net GHG mitigation of 0.35 CO₂e ha⁻¹ yr⁻¹ on average, or 2.48 Tg CO₂e yr⁻¹ on the aggregate, when compared with the unfertilized scenario (figures 5(a) and (b)). This GHG mitigation is about 54% of that projected to be achievable with the optimal rate. Both of these scenarios compare favorably to a situation where Nₑ is applied at the rate that maximizes long-term yields (~160–180 kg N ha⁻¹; supplementary figure 8 and Monti et al 2019), resulting in a slight increase in the GHG emissions over the unfertilized scenario (0.036 CO₂e ha⁻¹ yr⁻¹ on average or 0.26 Tg CO₂e yr⁻¹ aggregated; figures 5(a) and (b)). It should be noted that this latter scenario exhibits the strongest spatial divergence, with areas of large positive net GHG in the Great Plains states and southern Iowa, and large negative net GHG in the Great Lakes states and areas of Illinois and Indiana (figure 5(b)).

4. Discussion and conclusion

Our findings seem to challenge the prevalent belief that Nₑ application is generally detrimental to bioenergy GHG balances and that its use should be
avoided in the production of lignocellulosic feedstock (Erisman et al 2009, Smith et al 2010). On the contrary, here projections suggest that the long-term yield advantages of relatively low amounts of Nf in marginal lands would surpass the GHG emissions stemming from its use. Even a blanket recommendation of 75 kg N ha\(^{-1}\) that would forego nearly half of the potential GHG mitigation benefit, is predicted to be a net positive on the aggregate for Midwestern marginal lands (figure 5(c)). The empirical evidence, though limited, seems to concur. Indeed, a field study in Michigan (Ruan et al 2016) that examined field-level GHG measurements found that Nf additions could support GHG mitigation of switchgrass production up to 140 kg N ha\(^{-1}\), although GHG\(_{\text{Net}}\) was minimized with a Nf application of 56 kg N ha\(^{-1}\). Importantly, our projections seem to imply that the benefit could be much greater in marginal lands that are limited by poor N fertility as is the case of much of Michigan, Wisconsin, and isolated areas of the eastern Corn Belt (figure 4). Meanwhile, areas with SOC-rich soils (Iowa, Minnesota) and/or limited by low precipitation (Nebraska, Kansas) or short growing seasons (North and South Dakota) are predicted to be less likely to respond to Nf. Thus, the potential GHG mitigation benefit in these marginal lands is minimal. Lastly, it should be noted that the objectives of maximizing yield and minimizing GHG\(_{\text{Net}}\) do not seem to overlap in all but a few environments (figure 5(c)). As such, widespread implementation of fertilizing strategies aiming to maximize yield will likely result in the overall reduction of the climate benefit of cellulosic biofuels (figures 5(b) and (c)). (We include a map classifying Midwestern marginal lands in four levels of optimal N rates in the supplementary figure 9 to facilitate discussions with stakeholders).

Of course, it is likely that the optimal Nf rates estimated here, which are based on minimizing GHG emissions, are different from those Nf rates that are most economically efficient (i.e. where Nf marginal revenue equals marginal cost). Indeed, a recent study (Fike et al 2017) that examined production sites in five states concluded that typical Nf recommendations (56–112 kg N ha\(^{-1}\)) added between $37 and $74 ha\(^{-1}\) to production costs and required switchgrass biomass prices in excess of $70 Mg\(^{-1}\) to be economically justifiable under the observed yield responses. Whether the bioenergy industry or ancillary carbon markets can support such pricing is yet to be determined and lies beyond the scope of our analysis.

We advocate caution when interpreting these results given the various sources of uncertainty contained in our projections. Foremost are the uncertainties associated with our characterization of the responses to Nf additions across the pedo-climatic conditions of Midwestern marginal lands. Although the process-based CSMs here used are reasonably capable of capturing the variation in biomass production and SOC changes and their dynamic interaction with agricultural management (supplementary figure 4 (Surendran Nair et al 2012, Field et al 2016, Jiang et al 2017, Liu and Basso 2017, Ojeda et al 2017, Jones et al 2018, Jarecki et al 2020, Martinez-Feria and Basso 2020)), their complex structure, large sets of parameters, and statistical misspecification (Wallach 2011) generally means that prediction uncertainty is large when CSMs are applied outside the calibrated domain. These uncertainties are partially addressed by using the ensemble mean of independent CSMs, which has been shown to provide more accurate predictions outside of sample populations (Rosenzweig et al 2013, Bassu et al 2014, Martre et al 2015, Basso et al 2018). The emulation step unlikely added significant uncertainty at this scale because of the independence between CSM and ML errors, which would tend to cancel out when aggregating across soil types and geographical locations.

Nevertheless, there are a few known weaknesses shared by all CSMs that could bias our projections. Chief among them is the representation of associative N fixation (ANF) by free-living diazotrophs, which is increasingly recognized as a potentially important source of N to switchgrass (40–60 kg ha\(^{-1}\) yr\(^{-1}\)), especially in N-limiting environments (Roley et al 2018, Bahulikar et al 2021). Yet, in contrast to biological N fixation (BNF) in legumes, ANF is largely episodic and irregularly detectable (Roley et al 2019), and our current understanding of the environmental factors controlling ANF rates is rather incomplete (Bahulikar et al 2021). Thus, current switchgrass CSMs, either do not account for this pathway (EPIC) or use instead BNF routines as a surrogate (APSIM, SALUS). Only the DAYCENT model has an explicit routine that, in the absence of soil phosphorus data, calculates free-living N fixation rates as a linear function of annual precipitation (Necpálova et al 2015). We are not aware of any work verifying the robustness of these routines for capturing ANF environmental variation in switchgrass, or whether it is even possible at this stage given the paucity of data available. The likely effect of underestimating non-Nf sources under certain environments would be the overestimation of the yield response, which would lead to an overestimation of the optimal N rate. Although we found no widespread evidence of overestimation of the site-specific yield response in the experiments (supplementary figure 10), a conservative interpretation of our data may still be warranted until further examination of this mechanism within the models can be performed.

An additional source of potential bias in our optimal Nf rate estimates is the use of the offline calculation of direct and indirect soil N\(_2\)O emissions following the IPCC tier 1 methodology. We opted not to use simulated N\(_2\)O because not all the CSMs we selected contain an explicit nitrification-denitrification routine, or have not been comprehensively validated...
in production of perennial bioenergy grasses. On the other hand, the uncertainties of the IPCC methodologies are more thoroughly understood. Namely, that the response of annual N$_2$O fluxes to N$_F$ applied tends to be exponential rather than linear (Philibert et al 2012). Thus, the use of a constant emission factor as required by IPCC tier 1, tends to overestimate N$_2$O fluxes at lower N$_F$ rates, and underestimate them at higher rates. This pattern seems to hold in switchgrass production (Ruan et al 2016). Using a non-constant emissions factor that increases with N$_F$ rates to calculate N$_2$O fluxes (as shown by Ruan et al 2016) results in overall lower optimum N$_F$ rates with a much narrower distribution (supplementary figure 11). Nevertheless, it is not known how well this exponential relationship, which is derived from a few experiments, would represent emissions at the regional scale. Thus, more work is needed to improve the characterization of soil N$_2$O emissions when estimating GHG balances of bioenergy.

A final, possibly large source of uncertainty are the assumptions made for calculating the biofuel GHG offset, which essentially scales the projected yield and ΔSOC responses into GHG emissions (supplementary table 3). Here, we used a purely static LCA approach, which assumes constant values derived from the GREET model (methods) notwithstanding the spatial distribution of future infrastructure (e.g. depots, biorefineries, etc) or biomass end use (e.g. pelletization, direct combustion, energy cogeneration). These would result in changes in the efficiency of utilization of the feedstock supply (O’Neill and Maravelias 2021), particularly because lower switchgrass yields would require larger collection radius, and therefore greater GHG emissions of switchgrass transport (Gelfand et al 2013). Yet, the uncertainty associated with these parameters may be substantial. A sensitivity analysis where each of these values are randomly varied within ±20% of the original value reveals a potentially large departure from the mean GHG response (supplementary figure 12), though the computed optima tend to cluster around the mean found in our main analysis (75 kg N ha$^{-1}$) suggesting low bias. Yet, two parameters the yield from biomass-to-ethanol conversion (Yield$_{bioFuel}$) and the life-cycle GHG emissions intensity of gasoline (GWI$_{Gasoline}$), resulted to have the greatest impact in the calculation of the biofuel offset (supplementary figure 10). Thus, much of the uncertainty can be reduced by finding best values for those parameters.

Despite the uncertainties, our study represents an important step forward, as it is the first to examine prospective responses of switchgrass at such scale and resolution, and the emulator model developed could provide the groundwork for building dynamic tools to guide future decisions in agronomic management, technoeconomic analysis, and policy formulation (Ruan et al 2016). Only through this kind of high-resolution predictive tools, biofuel supply-chain network design can move toward integrating temporally and spatially explicit data to reflect the reality of heterogeneity in biomass productivity and SOC sequestration capacity across the landscape, and to allow consideration of land allocation and crop management decisions in addition to infrastructure and operational decisions (O’Neill and Maravelias 2021). This is becoming more critical given the interest in targeted integration of perennial bioenergy crops within the margins of row-crop fields or other marginal subfield areas (Brandes et al 2018, Basso and Antle 2020), which could not only avoid GHG through the production of biofuels, but also by replacing low yielding, highly polluting croplands (Basso et al 2019). Therefore, further research is needed that comprehensively evaluates the modeling uncertainties here outlined to advance prospects for efficient site-specific N$_F$ management in perennial bioenergy crops, and to maximize the GHG mitigation potential of this technology.

**Data availability statements**

The data that support the findings are within this study in the cited references. R code and the database needed to run the developed GBM emulators has been made public at the following repository: https://doi.org/10.5281/zenodo.5116736. The data that support the findings of this study are available upon reasonable request from the authors.

**Acknowledgments**

This research was supported in part by the Great Lakes Bioenergy Research Center, US Department of Energy, Office of Science, Office of Biological and Environmental Research (Awards DESC0018409 and DE-FC02-07ER64494), the USDA National Institute of Food and Agriculture (Awards: 2015-68007-23133, 2018-67003-27406, 2019-67012-29595) and Michigan State University AgBioResearch. Authors would like to thank Max de Antoni Migliorati (DAYCENT model support), Sasmita Sahoo (EPIC model support), the APSIM initiative, and Lydia Price (spatial data support).

**Conflict of interest**

The authors declare no competing interests.

**Author contributions**

R M F and B B conceived and designed the research; R M F assembled the datasets and developed predictive models. S K defined LCA values. R M F analyzed the predicted datasets and visualization of the results. All authors contributed to interpreting the results and writing the manuscript.
References

Agostini F, Gregory A S and Richter G M 2015 Carbon sequestration by perennial energy crops: is the jury still out? Bioenergy Res. 8 1057–80

Argonne National Laboratory 2020 Summary of expansions and updates in GREET 2020—ANL Energy System Division

Bahulikar R A, Chaluvadi S R, Torres-Jerez I, Mosalj I, Bennetzen J L and Udvardi M 2021 Nitrogen fertilization reduces nitrogen fixation activity of diverse diazotrophs in switchgrass roots Physiologiae J. S 80–87

Basso B et al 2018 Soil organic carbon and nitrogen feedbacks on crop yields under climate change Agric. Environ. Lett. 3 180026

Basso B and Antle J 2020 Digital agriculture to design sustainable agricultural systems Nat. Sustain. 3 254–6

Basso B, Shuai G, Zhang J and Robertson G P 2019 Yield stability of various maize crop models vary in their responses to climate change factors Glob. Change Biol. 20 2301–20

Brandes E, McNunn G S, Schulte L A, Muth D J, VanLoocke A and Bassu S Bassu B, Shuai G, Zhang J and Robertson G P 2019 Yield stability analysis reveals sources of large-scale nitrogen loss from the US Midwest Sci. Rep. 9 5774

Bassu S et al 2014 How do various maize crop models vary in their responses to climate change factors? Glob. Change Biol. 20 333–46

Crutzen P J, Mosier A R, Smith K A and Winiswarter W 2008 N2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels Atmos. Chem. Phys. 8 389–95

El Akkari M, Réchauchere O, Bispo A, Gabrielle B and Makowski D 2018 A meta-analysis of the greenhouse gas abatement of bioenergy factoring in land use changes Sci. Rep. 8 8563

Erisman J W, van Grinsven H, Leip A, Mosier A and Bleeker A 2009 Nitrogen and biofuels; an overview of the current state of knowledge Nutr. Cycl. Agroecosyst. 86 211–23

Fargione J, Hill J, Tilman D, Polasky S and Hawthorne P 2008 Land clearing and the biofuel carbon debt Science 319 1235–8

Field J L, Evans S G, Marx E, Easter M, Adler P R, Dinh T, Willson B and Paustian K 2018 High-resolution technob-ecological modelling of a bioenergy landscape to identify climate mitigation opportunities in cellulosic ethanol production Nat. Energy 3 211–9

Field J L, Marx E, Easter M, Adler P R and Paustian K 2016 Ecosystem model parameterization and adaptation for sustainable cellulosic biofuel landscape design GCB Bioenergy 8 1106–23

Fike J H, Pease J W, Owens V N, Farris R L, Hansen J L, Heaton E A, Hong C O, Mayton H S, Mitchell R B and Viands D R 2017 Switchgrass nitrogen response and estimated production costs on diverse sites GCB Bioenergy 9 1526–42

Gelfand I, Hamilton S K, Kravchenko A N, Jackson R D, Thelen K D and Robertson G P 2020 Empirical evidence for the potential climate benefits of decarbonizing light vehicle transport in the US with bioenergy from purpose-grown biomass with and without BECCS Environ. Sci. Technol. 54 2961–74

Gelfand I, Sahajpal R, Zhang X, Izaurralde R C, Gross K L and Robertson G P 2013 Sustainable bioenergy production from marginal lands in the US Midwest Nature 493 14–517

Han W, Yang Z, Di L and Mueller R 2012 CropScape: a web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support Comput. Electron. Agric. 84 111–23

IPCC 2019 N2O emissions from managed soils, and CO2 emissions from lime and urea application 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories

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 Core Writing Team, R K Pachauri and L A Meyer (Geneva)

Jarecki M, Kariyapperuma K, Deen B, Graham J, Bazarag A B, Vijayakumari S, Thimmanagari M, Gordon A, Voroney P and Thevathasan N 2020 The potential of switchgrass and miscanthus to enhance soil organic carbon sequestration—predicted by Daycent model Land 9 509

Jiang R, Wang T T, Jin S H, Sheng G U, Wei Z H, Yu Y J, Chen S L and Hatano R 2017 Modeling the biomass of energy crops: descriptions, strengths and prospective J. Integr. Agric. 16 1197–210

Jones C D, Oates I G, Robertson G P and Izaurralde R C 2018 Perennialization and cover cropping mitigate soil carbon loss from residue harvesting J. Environ. Qual. 47 710–7

Liu L and Basso B 2017 Spatial evaluation of switchgrass productivity under historical and future climate scenarios in Michigan GCB Bioenergy 9 1320–32

Martínez-Feria R and Basso B 2020 Predicting soil carbon changes in switchgrass grown on marginal lands under climate change and adaptation strategies GCB Bioenergy 12 742–53

Martre P et al 2015 Multimodel ensembles of wheat growth: many models are better than one Glob. Change Biol. 21 911–23

McLaughlin S B and Kszos L A 2005 Development of switchgrass (Panicum virgatum) as a bioenergy feedstock in the United States Biomass Bioenergy 28 515–35

Melillo J J, Reilly J M, Kicklighter D W, Gurgel A C, Cronin T W, Paltsev S, Felzer B S, Wang X, Sokolov A P and Adam Schlosser C 2009 Indirect emissions from biofuels: how important? Science 326 1397–9

Miesel J R, Jach-Smith L C, Renz M J and Jackson R D 2017 Distribution of switchgrass (Panicum virgatum L) aboveground biomass in response to nitrogen addition and across harvest dates Biomass Bioenergy 100 74–83

Monti A, Zegada-Lizarazu W, Zanetti F and Casler M 2019 Nitrogen fertilization management of switchgrass, miscanthus and giant reed: a review Adv. Agron. 153 87–119

National Research Council 2012 Renewable fuel standard: Potential economic and environmental effects of US biofuel policy

Nepolivá M, Anex R P, Fienien M N, del Grosso S J, Castellano M J, Sawyer E J, Iqbal J, Pantoja J L and Barker D W 2015 Understanding the DayCent model: calibration, sensitivity, and identifiability through inverse modeling Environ. Model. Softw. 66 110–50

O’Neill E G and Maravelias C T 2021 Towards integrated landscape design and biofuel supply chain optimization Curr. Opin. Chem. Eng. 31 100666

Ojeda J J, Volene J J, Brouder S M, Caviglia O P and Thevathasan N 2017 Modeling the biomass of energy crops: descriptions, strengths and prospective J. Integr. Agric. 16 1197–210
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

Robertson G P, Hamilton S K, del Grosso S J and Parton W J 2011 The biogeochemistry of bioenergy landscapes: carbon, nitrogen, and water considerations *Ecol. Appl.* 21 1055–67

Roley S S, Duncan D S, Liang D, Garoutte A, Jackson R D, Tiedje J M and Robertson G P 2018 Associative nitrogen fixation (ANF) across a nitrogen input gradient pp 1–37

Roley S S, Xue C, Hamilton S K, Tiedje J M and Robertson G P 2019 Isotopic evidence for episodic nitrogen fixation in switchgrass (*Panicum virgatum* L.) *Soil Biol. Biochem.* 129 90–98

Rosenzweig C et al 2013 The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies *Agric. For. Meteorol.* 170 166–82

Ruan L, Bhardwaj A K, Hamilton S K and Robertson G P 2016 Nitrogen fertilization challenges the climate benefit of cellulosic biofuels *Environ. Res. Lett.* 11 064007

Smith A et al 2010 Measuring sustainable intensification in smallholder agroecosystems: a review *J. Fam. Consum. Sci.* 130 95–109

Soil Survey Staff Soil Survey Geographic (SSURGO) database (available at: https://sdmdataaccess.sc.egov.usda.gov)

Somerville C, Youngs H, Taylor C, Davis S C and Long S P 2010 Feedstocks for lignocellulosic biofuels *Science* 329 790–2

Surendran Nair S, Kang S, Zhang X, Miguez F E, Izaurralde R C, Post W M, Dietze M C, Lynd L R and Wallishegler S D 2012 Bioenergy crop models: descriptions, data requirements, and future challenges *GCB Bioenergy* 4 620–33

Valdez Z P, Hockaday W C, Masiello C A, Gallagher M E and Philip Robertson G 2017 Soil carbon and nitrogen responses to nitrogen fertilizer and harvesting rates in switchgrass cropping systems *Bioenergy Res.* 10 456–64

Wallach D 2011 Crop model calibration: a statistical perspective *Agron. J.* 103 1144–51

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. Atmos.* 117