Regionalizing crop types to enhance global ecosystem modeling of maize production

Yi Yang1,∗, Stephen Ogle1,2, Stephen Del Grosso3, Nathaniel Mueller2,3, Shannon Spencer1 and Deepak Ray4

1 Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523, United States of America
2 Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO 80523, United States of America
3 Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523, United States of America
4 Institute on the Environment, University of Minnesota, Saint Paul, MN 55108, United States of America
5 Agricultural Research Service, USDA, Fort Collins, CO 80526, United States of America

∗Author to whom any correspondence should be addressed.
E-mail: yi.yang@colostate.edu

Keywords: model calibration, DayCent, parameter estimation, global maize production
Supplementary material for this article is available online

Abstract
Improving the prediction of crop production is critical for strategy development associated with global food security, particularly as the climate continues to change. Process-based ecosystem models are increasingly used for simulating global agricultural production. However, such simulations often use a single crop variety in global assessments, implying that major crops are identical across all regions of the world. To address this limitation, we applied a Bayesian approach to calibrate regional types of maize (Zea mays L), capturing the aggregated traits of local varieties, for DayCent ecosystem model simulations, using global crop production data from 2001 to 2013. We selected major cropping regions from the FAO Global Agro-Environmental Stratification as a basis for the regionalization and identified the most important model parameters through a global sensitivity analysis. We calibrated DayCent using the sampling importance resampling algorithm and found significant improvement in DayCent simulations of maize yields with the calibrated regional varieties. Compared to a single type of maize for the world, the regionalization of maize leads to reductions in root mean squared error of 11%, 31%, 27%, 30%, 19%, and 27% and reductions in bias of 59%, 59%, 50%, 81%, 32%, and 56% for Africa, East Asia, Europe, North America, South America, and South and Southeast Asia, respectively. We also found the optimum parameter values of radiation use efficiency are positively correlated with the income level of different regions, which indicates that breeding has enhanced the photosynthetic efficiency of maize in developed countries. There may also be opportunities for expanding crop breeding programs in developing countries to enhance photosynthesis efficiency and reduce the yield gap in these regions. This study highlights the importance of representing regional variation in crop types for achieving accurate predictions of crop yields.

1. Introduction
Climate change is arguably the greatest global environmental challenge facing society, influencing both natural and managed ecosystems, and agriculture is no exception. Climate change has a complex impact on global agriculture and food production (Rosenzweig and Parry 1994, Schmidhuber and Tubiello 2007, Wheeler and von Braun 2013, Fujimori et al 2019). For instance, higher temperatures in temperate regions may be of benefit to food production due to expansion of areas suitable for warm-season cropping, whereas increased evapotranspiration and reduced soil moisture can reduce the crop yields in arid regions (Schmidhuber and Tubiello 2007). Therefore, understanding this complexity is critical for decision-making to ensure food security (Wheeler and von Braun 2013). In addition,
Crop breeding has led to shifts in crop characteristics and production historically (Zaidi et al 2019), and this underlying variation will influence the responses of crops to changes in temperature and moisture regimes. Therefore, predicting future crop yields to support food security strategy will depend on understanding the complexity in climate change impacts as well as the characteristics of crop varieties available to the farming community.

Process-based ecosystem models are increasingly used for simulating global agricultural production to facilitate crop management, climate change impact and adaptation assessments (Ewert et al 2015). However, recent studies have shown that there is wide variability in crop production predictions among models (Martre et al 2015, Confalonieri et al 2016). Reducing the uncertainty of model predictions on large scales is important because it can help identify the factors driving crop growth in different regions and predict future yields with climate change. Further, improving the simulations of crop yields can also improve our estimation of greenhouse gas emissions in the agricultural sector, because crop growth is a key driver of C inputs to soils and formation of soil organic matter, as well as N uptake from soils by the crops to support growth, which affect soil C sequestration, and N2O and other trace gas emissions, and soil nutrient dynamics (Folberth et al 2019).

Uncertainties surrounding parameter values for process-based models can lead to errors in extending simulations across space and time (Seidel et al 2018). Most models are initially constructed based on field studies with parameters representing crop physiology and environmental factors. To predict realistic crop yields, model parameters should be calibrated, and model output evaluated, especially when extending model simulations to regional, national and global scales. To improve the accuracy in predicting future crop yields under climate change, the parameters can be calibrated with yield data from the historical period. Common practices for optimizing parameters includes manual calibration and automated calibration with statistical methods, such as Bayesian parameter estimation (Seidel et al 2018). The manual calibration method involves trial and error by manually shifting individual parameter values, generally one-at-a-time, until the model predictions are more similar to observations, and this approach has been widely applied among modeling studies. A survey of 211 modeling researchers reported that nearly half the respondents were still using the trial and error method to find the most influential parameters and their best-fit values (Seidel et al 2018). However, manual parameter calibration is not only inefficient and time-consuming but also unable to provide uncertainty ranges for each parameter that can be used to derive uncertainty in model predictions. Parameter interactions are also often ignored in manual calibration. In contrast, Bayesian calibration methods can search ranges of values for key parameters in an automated manner and determine which parameter values are likely realistic given the observed crop yields. This produces parameter value distributions that can be used to derive uncertainty in model predictions and accounts for parameter interactions.

In the last century, crop breeding made significant progress in improving crop growth and production (van de Wouw et al 2010). Breeders have been developing new crop varieties that have higher yields and are adapted to various environments and farming systems (van de Wouw et al 2010). In recent years, crop varieties have been introduced even more rapidly due to the application of genomic sequencing and genotyping (Furbank et al 2019). To simulate crop yields on large-scales, regionalizing crop types is essential by representing variation in characteristics that influence parameter values, and in turn, improve crop yield simulation accuracy. Further, identifying crop region-specific characteristics can also provide insights for breeders to determine key traits that could enhance yields. However, regionalizing crop types on a large-scale with manual calibration is nearly impossible because process-based models tend to have numerous parameters for crop growth. Many previous studies ignored spatial variability of crops by applying default parameters in all regions or used parameters calibrated in industrialized countries with high-input agriculture to regions with low crop yield (Stehfest et al 2007, Kamali et al 2018). Therefore, an automatic parameter calibration method that can account for temporal fluctuation and spatial variation of crop growth should be applied when using models to simulate large-scale crop production.

Maize (Zea mays L) is the one of the most widely produced grain crops in the world, and is used for human consumption, livestock feed and bioenergy (Ranum et al 2014). New varieties of maize are constantly under development to improve production, pest/disease resilience and nutrient quality (Duvick 2005, Meissle et al 2010, Cairns and Prasanna 2018). Incorporating regional types would significantly improve the accuracy of maize yield simulation at large scales. However, information on the spatial and temporal distribution of specific maize varieties across the world and the measurements of their physiological and related traits are difficult to obtain. Instead, we created regional maize types by calibrating the DayCent ecosystem model with the reported maize yield data (Ray et al 2019) and environmental factors as the model inputs. The regional types represent the aggregated traits of local maize varieties. To achieve this, we first applied a global sensitivity analysis (GSA) to identify the most influential parameters driving crop production in DayCent model simulations, followed by a Bayesian parameter estimation method to identify the likely ranges of the parameter values given the observed crop yields. With
this automated calibration procedure, we expected to (a) improve the accuracy of the simulations of global maize production by DayCent, (b) identify the spatial variability and uncertainty of the parameters that control crop growth, and (c) understand differences among regional crop types based on the variation in the resulting parameters from the calibration analysis.

2. Method

2.1. DayCent model

DayCent is a process-based ecosystem model that operates on a daily time step and simulates biogeochemical and water dynamics in croplands, forests, grasslands, and savanna (Parton et al 1998, Del Grosso et al 2010). The model uses daily weather, soil properties, land management, and plant types as primary inputs. Model outputs include crop yield, soil organic carbon, and daily trace gas fluxes (N\textsubscript{2}O, NO\textsubscript{x}, NH\textsubscript{3}, CH\textsubscript{4}). The vegetation production module was recently improved by using green leaf weight ratio (fraction of green leaf biomass to aboveground biomass) in a function based on growing degree days (GDDs) to simulate crop growth and canopy development (Zhang et al 2020). The recent version of DayCent also incorporates N losses from ammonium volatilization in the soil (Gurung et al 2020).

2.2. Calibration zones and grid cells selection

We calibrated the DayCent model for maize yield at a 0.5° grid resolution between 2001 to 2013 for the agricultural regions based on the Global Agro-Environmental Stratification (GAES) developed by Mücher et al (2016), and maize yield and harvest area that were reported by Ray et al (2019). The GAES dataset was produced by stratifying the agricultural production zones according to the region’s agro-environmental characteristics. The strata contain four spatial levels, which were further characterized by factors such as phenology, management practices, field size, and other key characteristics. We used the third level of GAES for maize yield parameter calibration, which consists of 77 zones. The major maize-growing regions were determined by selecting regions containing more than 100 grid cells with maize production, which resulted in a dataset that included 95% of the total world maize production. We applied systematic sampling to select training and validation sets of grid cells for parameter calibration and evaluation of the results, respectively (figure 1). In addition, there were ten grid cells for each sample that were selected in regions with less than 100 grid cells of maize production. We used the ratio of mean annual precipitation to potential evapotranspiration (MAP:PET) to determine the irrigation management in each grid cell, which is based on the 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories (IPCC 2019).

Grid cells with MAP:PET < 1 are arid or semi-arid regions, where irrigation management is applied in our analysis. The MAP and PET data were provided by the Climate Research Unit (CRU) climate datasets (Harris et al 2020). We simulated maize monoculture in each grid cell due to the lack of global datasets with crop rotation information.

2.3. Parameter calibration process

We applied the Bayesian model analysis framework developed by Gurung et al (2020) for the regional calibration of model parameters to improve maize yields predictions. Process-based models usually use dozens to hundreds of parameters to control crop growth, requiring an immense number of computational resources and time to fully implement calibration routines. Therefore, we applied a GSA to identify the most influential parameters that drive the variation in model outputs before calibration. We then used the sampling importance resampling (SIR) method for calibration based on applying the DayCent model with randomly selected parameter sets for the influential parameters on multi-core computer cluster to accelerate the calibration process while fully exploring parameter spaces (Tang and Zhuang 2009, Gurung et al 2020).

Before calibration, key state variables were initialized to reach steady-state condition, particularly SOM levels, by simulating native vegetation, historical weather, and soil properties for several 1000 years, followed by simulating cultivation histories before the calibration period (2001–2013). The steady-state simulations used the weather data between 1901 and 2000 by simulating the same 100 years for several 1000 years until the state variables steady-state. The exact number of years for the steady-state simulations varied among the grid cells. Model inputs, such as land management, irrigation, planting dates, and soil characteristics were adopted from the Environmental Protection Agency (EPA) report on global mitigation of non-CO\textsubscript{2} greenhouse gases: 2010–2030 (EPA 2013); the weather data, including daily precipitation, maximum and minimum air temperatures, were based on the Inter-Sectoral Impact Model Intercomparison Project’s GSWP3-WSES data (Lange 2019); and the native vegetation before cultivation was based on Cramer et al (1999). We used the same model inputs data and global dataset of maize yield for both calibration and validation.

We performed the GSA with the Sobol method (Sobol 2001, Zhang et al 2015), which is based on variance decomposition techniques to provide a quantitative measure of the contributions of the input to the output variance. First, we created prior distribution for each of the 64 parameters that control crop growth in DayCent. All priors had uniform distributions, except for the parameters that were associated with the temperature effect on the plant growth.
(PPDF1, PPDF2, PPDF3, and PPDF4). The priors of these four parameters produce a distribution that describes the influence of temperature on maize production and so the initial parameters were created collectively as a joint distribution based on curves that produced growth within expected ranges (no growth at or below 0 °C, minimal growth at 10 °C and no growth at or above 47 °C). We considered parameters with total sensitivity indices over 0.025 as influential.

For the Bayesian parameterization, we used the SIR algorithm (Rubin 1988) and followed the general procedure outlined by Gurung et al (2020). We derived 100,000 parameter sets with Latin Hypercube Sampling from the prior distributions of the influential parameters. We simulated crop production with DayCent using these parameter sets for randomly selected grid cells (training set) in each region from 2001 and 2013, while fixing the non-influential parameters to their default values. We computed log-likelihoods based on the goodness-of-fit for the modeled yields to observed yields for each GAES region and calculated the standardized importance weights. Finally, we resampled 500 parameter sets without replacement from the initial 100,000 parameter sets based on their importance weights and used them to construct the marginal joint posterior distributions of the parameters. Gurung et al (2020) provides more information about this Bayesian calibration method. We used the out-of-samples to validate the calibration, by calculating root mean squared errors (RMSEs) and model biases were calculated with the following equation:

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^{n} (M_i - O_i),$$  \hspace{1cm} (2)

where \( n \) is the number of crop yield values, \( M_i \) are the modeled values and \( O_i \) are measured values.

After the calibration, we compared the parameter values in each region with environmental and economic factors to explore the regional differences of the parameters. The environmental variable data, such as mean annual temperature, rainfall, and cloud fraction were from the GAES strata. The gross domestic production (GDP) per capita of each maize growing zone were based on the means between 2001 and 2013 in each country from World Bank. For zones that include two or more countries, we calculated the mean GDP per capita in each zone weighted by area in each country.

A two-way ANOVA was used to examine the relationship between parameter values and income levels in all cropping zones. Tukey’s test was used for the post-hoc analysis. The normality of the residuals was tested with Q–Q plot. All statistical analyses were performed using R version 4.0.5 (R Core Team 2021). The following publicly available R packages were also used in the analyses: ‘tidyverse’ (Wickham et al 2019) was used to import data, transform data sets and construct figures; ‘sf’ (Pebesma 2018) and ‘raster’ (Hijmans 2020) were used to analyze spatial data; ‘tmap’ (Tennekes 2018) was used to construct maps; ‘sensitivity’ (Iooss et al 2021) was used to perform GSA; ‘lhs’ (Carnell 2020) was used to perform latin hypercube sampling; ‘car’ (Fox and Weisberg 2018) was used to perform ANOVA analyses and ‘emmeans’
Table 1. Description and prior ranges of all parameters that control crop growth in DayCent.

| Parameter | Description                                                                 | Default values | Prior ranges   |
|-----------|-----------------------------------------------------------------------------|----------------|----------------|
| RUETB     | Radiation use efficiency for total biomass production                        | 3.89           | 3.5–7          |
| PPDF1     | Optimum temperature for production for parameterization of a Poisson density function curve to simulate temperature effect on growth | 27             | 25–33          |
| PPDF2     | Maximum temperature for production for parameterization of a Poisson density function curve to simulate temperature effect on growth | 45             | 40–47          |
| PPDF3     | Left curve shape for parameterization of a Poisson density function curve to simulate temperature effect on growth | 1              | 0.5–1.5        |
| PPDF4     | Right curve shape for parameterization of a Poisson density function curve to simulate temperature effect on growth | 2.5            | 2–4            |
| PLTMRF    | Initial total biomass at emergence                                           | 0.3            | 0.05–0.5       |
| FRTC1     | Fraction of C allocated to roots at planting, with no water or nutrient stress | 0.4            | 0.1–0.6        |
| DDBASE    | Number of degree days required to reach anthesis for a grain filling annual crop | 800            | 500–900        |
| KLIGHT    | Extinction coefficient of Beer’s Law for light interception of crop canopy    | 0.55           | 0.4–0.8        |
| SLA       | Specific leaf area (m$^2$ g$^{-1}$)                                          | 0.02           | 0.015–0.03     |

Figure 2. Total sensitivity index for parameters selected by the GSA. There are 64 parameters in DayCent that drive crop growth and production. Here we show the most influential parameters for modeled crop yield. PPDF1–4 are the parameters controlling the temperature effect on vegetation growth; DDBASE is the number of GGDs required to reach the flowering stage in the annual crop’s life cycle; RUETB is the radiation use efficiency of the crop; KLIGHT is light interception of the crop canopy; SLA is the specific leaf area; PLTMRF is the initial total biomass at emergence and FRTC1 is the fraction of C allocated to roots at planting. See table 1 for the detailed parameter definitions and supp table 1 for the parameters that were set to default values for Bayesian calibration.

(Russell 2019) was used to perform post-hoc tests for the ANOVA.

3. Results

3.1. GSA

There are 64 parameters in DayCent that drive crop production (table 1) and ten parameters are most influential (figure 2), including (a) parameters that control the temperature effect on vegetation growth (parameters: PPDF1, PPDF2, PPDF3 and PPDF4), (b) the number of GDDs required to reach the flowering stage in the annual crop’s life cycle (parameter: DDBASE), (c) radiation use efficiency (RUE) of the crop (parameter: RUETB), (d) light interception of the crop canopy (parameter: KLIGHT), (e) specific leaf area (SLA) (parameter: SLA), (f) initial total biomass at emergence (parameter: PLTMRF), and (g) fraction of C allocated to roots at planting (parameter: FRTC1).
3.2. Bayesian calibration
The Bayesian calibration was performed in the 77 GAES zones, and here we show the posterior and prior distributions of the zone with the largest amount of maize production in North America (figure 3) while the distributions for South America, Europe, Africa, East Asia, and South and Southeast Asia are provided in supplementary material (supp figure 2 available online at stacks.iop.org/ERL/17/014013/mmedia). The SIR results show that ranges of the posterior distributions for parameters are reduced over their priors, which indicates that there was sufficient information in the crop yield data to calibrate the model. There is regional variation, however, with some parameters only having moderate or little difference between the posterior and prior probability distributions (e.g. FRTC1 in zone 450 of East Asia, supp figure 1(B)). In addition, some parameters have bimodal posterior distributions, such as RUE in zone 1547 of North America, indicating that there may be different varieties with distinct characteristics within a region (RUE in this case), and further disaggregation of these zones for calibration could reduce uncertainty in crop yield predictions.

The Bayesian calibration significantly improved the model prediction of maize yield in all large regions compared to model prediction with default parameter values. In the validation set of grid cells, modeled crop yield after calibration is positively associated with observed yield data in all regions (figures 4 and 5). After calibration, the RMSE was reduced by 11%, 31%, 27%, 30%, 19%, and 27% and bias was reduced by 59%, 59%, 50%, 81%, 32%, and 56% in Africa, East Asia, Europe, North America, South America, and South and Southeast Asia, respectively. There is good agreement in crop yield between calibrated model prediction and observed data in South Africa, East Asia, Europe, and North America in terms of median crop yields and associated variations (figure 6). However, despite the significant reduction of RMSE and bias after calibration, the model predictions in South America and South and Southeast Asia had large biases compared to the observed crop yield.

3.3. Regional parameter evaluation
Each GAES zone has distinct parameter posterior distributions based on our analysis that we evaluated based on environmental and socio-economic variables. First, we evaluated the role of climate factors on the selection of maize varieties in each region. A correlation analysis between environmental factors and calibrated parameter values showed that the optimum growing temperature (PPDF1) and GDDs (DDBASE) of maize are positively correlated with mean annual temperature, which suggests that there is a variation in selection of maize varieties on a gradient from cool to warmer regions (table 2). The parameter PPDF1 is positively associated with irrigation and the initial biomass at emergence (PLTMRF) is also positively associated with temperature. Second, we compared the parameter values with GDP per capita to evaluate the influence of socioeconomic factors on crop breeding. We found that maize varieties in...
4. Discussion

Global food security requires maintaining and potentially increasing global crop yields in the future (Schmidhuber and Tubiello 2007, Fischer et al 2014, Fujimori et al 2019). Therefore, there is a need to improve modeling to identify factors and management strategies that can sustain yields in high production systems, while enhancing production and reducing the yield gap in regions with low production agricultural systems. Regionalizing crop types in process-based models can significantly improve crop yield prediction on a global scale (Folberth et al 2012). However, manual calibration methods by trial and error are inefficient and time-consuming for

developed regions tend to have higher RUE, but lower SLA and light interception (figure 7).
large-scale simulations (Seidel et al 2018). In this study, we used the SIR method in a Bayesian analysis framework to calibrate the DayCent ecosystem model and simulated global maize yields. Facilitated by the GSA that limited the number of parameters needed for calibration, the automated Bayesian method streamlined the process of calibration, identified regional variation in parameter values, and improved the model predictions of maize yields in all regions.

4.1. Calibrated model predictions
The predictions of maize yields have good agreement with observed yields in temperate regions, such as Southern Africa, East Asia, Europe and North America (figure 6). The maize production in these regions accounts for about 84% of the world maize production, which means the regionalization of maize crop types significantly improved DayCent simulations of crop yield in major maize production regions. The calibrated crop yields were generally underestimated in tropical regions such as South America, East Africa, West Africa, South Asia, and Southeast Asia, despite the RMSE and bias statistics demonstrating that the model results were more accurate following calibration. Water stress is unlikely the reason for the underestimation of maize yields in these regions because we modeled irrigated maize growth in grid cells where the MAP: PET is less than one. Therefore, simulated maize production was generally not under water stress.

DayCent does not capture the effect of pest and disease outbreaks on crop yields. However, similar to limited water stress, it is unlikely that the lack of pest/disease outbreaks in the model simulations would lead to an underestimation of yields. Furthermore, we calibrated the parameters against the reported yield data, which represent crop production under

![Figure 6](image_url). The differences between observed and modeled maize yields. The crop yield unit is gC m$^{-2}$.
local environmental conditions. If pests and disease outbreaks are common in these tropical regions and leading to lower yields, then the model parameter values may capture some of those effects by adjusting the crop physiological trait parameters. For example, the calibrated radiation use efficiency model parameter (RUETB) of maize would be lower if pest/disease outbreaks are leading to an overall reduction in yield throughout the area in the region. Nonetheless, explicitly representing these processes in the future version of DayCent could improve overall accuracy of crop simulations.

With further testing, we found the modeled crop yield in the tropical regions was constrained by N availability, which is one of the most critical drivers of crop growth in DayCent (Stehfest et al. 2007). This is consistent with previous research in tropical regions, such as Sub-Saharan Africa, which found that N stress was the major growth-limiting factor when simulating maize production (Folberth et al. 2012). In our study, for example, we found maize yields in the Pampas region of Argentina were underestimated after calibration, and this may be partly related to N fertilization rates. The N fertilizer data in this study are derived from (Conant et al. 2013), which are largely based on data published by the Food and Agricultural Organization of the United Nations (FAO). The FAO fertilizer report data in the Pampas region is about 20 kg N ha\(^{-1}\) yr\(^{-1}\), whereas the observed N fertilization rates can range from 20 to 80 kg N ha\(^{-1}\) yr\(^{-1}\) and have been increasing over time (Lavado and Taboada 2009). We performed a test by manually increasing the fertilization rate in the Pampas region to 80 kg N ha\(^{-1}\) yr\(^{-1}\) and found significantly increased maize yields that were more consistent with the observed yields, implying that the N fertilization rates may be larger than reported in the FAO statistics. In addition, due to the lack of global datasets of crop rotation, we simulated maize monocultures in all grid cells, which could exaggerate the N limitation in regions that have maize rotated with N fixing legumes. These results suggest that quality of input data, such as N fertilization rates and crop rotation, are a key factor for accurate yield predictions.

The accuracy of the reported maize yields can also impact the calibration. The maize yield data in this study were acquired from the census bureau or agricultural statistics from different political units, and the data availability and resolution varied among regions (Ray et al. 2019). For example, data from the United States were reported at the county level, whereas data from regions in Southeast Asia are mostly at the national level. In addition, the inaccurate data of crop production and harvested area could lead to unrealistically high or low yields in some regions. For instance, regions with lower quality crop yield data tend to have less accurate results from the model calibration. Finally, the DayCent model was mostly developed and tested with data from temperate regions, and it is possible that certain processes relevant to tropical cultivation are not well-represented.

4.2. Regional parameter patterns and crop breeding

The calibration identified regional variation in the parameters that are driving maize production in the
DayCent model simulations and provide insight into reducing yield gaps through crop breeding. Over the past decades, selective breeding has improved many physiological traits of maize, such as ones related to grain production, pest and disease resistance and stress tolerance. Among these traits, harvest index has become a major morphological driver of increased maize yield (Fischer et al 2014). However, as breeding is reaching the biological limits of cereal crops, improving harvest index has become increasingly difficult in recent years (Fischer et al 2014). Instead, breeders are currently focusing on improving crop biomass accumulation efficiency, which involves traits related to photosynthesis (Furbank et al 2019).

Our GSA results are consistent with the current crop breeding trends in which parameters related to photosynthesis (RUE, SLA, and light interception) are among the most influential parameters, while harvest index was found to be less sensitive given the observed yields. Among the influential parameters, RUE is a measure of the conversion of light intercepted in relation to crop biomass. Higher RUE would lead to higher biomass accumulation across the crop growing season. SLA is the ratio of leaf area to the leaf dry mass. A lower SLA is associated with crop that have smaller but thicker leaves, which have a lower amount of light interception, but an increased amount of photosynthetic machinery per unit leaf area (Richards 2000). Both results are consistent with the goals of crop breeding programs.

When comparing spatial variability of the parameters, we found RUE is positively correlated with GDP per capita while SLA and light interception are negatively correlated with GDP per capita. The more developed regions also have higher maize yields. These findings suggest that advanced breeding techniques are applied mostly in developed regions, which is consistent with larger budgets for research and development that lead to improvements in traits related to photosynthesis (Ribaut et al 2010, Qaim 2020). Screening for genetic variation in photosynthetic traits at the leaf level is time-consuming, and scaling leaf-level data to the canopy adds even more complexity to crop breeding programs. In recent years, new technology such as LiDAR, thermal imaging, canopy spectral reflectance as well as machine learning have been applied in developed countries to accelerate crop breeding programs for improving photosynthetic traits (Furbank et al 2019). Nevertheless, there are great opportunities for developing countries to improve maize yield as the new generation of breeding technology matures. For example, the RUE values in East Asia are significantly lower compared with North America and Europe. The relatively high maize yield in East Asia is facilitated by large fertilizer applications, but advanced breeding technology could further enhance crop yields and reduce fertilizer usage. Furthermore, less fertilization would also be beneficial for reducing greenhouse gas emissions of N₂O (Tian et al 2019) and other N losses that contribute to air and water pollution. Regardless, advancements in crop breeding that enhance crop traits related to RUE and other model parameters may have an even larger impact on crop yields and food security in regions with large yield gaps, such as Southeast Asia, Latin America and Africa (Lobell et al 2009).

5. Conclusions

Bayesian calibration of the DayCent model significantly improved the accuracy of global maize yield predictions by regionalizing parameters values. Spatial variation in parameter values also indicates that enhancing photosynthetic efficiency through crop breeding may reduce the yield gap in developing countries. The major limitations of this study appear to be related to lack of accurate fertilization data and observed maize yields in tropical regions. In addition, the DayCent model was originally developed for simulation of ecosystem dynamics in temperate regions, so the model may be missing a critical driver(s) of crop production in tropical climates, or the prior ranges for parameters may not be representative of the parameter distributions in these regions. Therefore, further improvements in model accuracy are likely possible with more accurate fertilizer data and maize yield observations, and further model development in the tropical agricultural systems.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This research was funded by the US Environmental Protection Agency and USDA Forest Service (Agreement No. 18-CR-11242305-109). We thank Yao Zhang, Bill Parton, Stephen Williams, Ram Gurung, and Melannie Hartman for support in conducting the Bayesian analysis and application of the DayCent model. We also thank Jared Creason and Shaun Ragnauth for insightful discussions about the analysis framework.

ORCID iDs

Yi Yang https://orcid.org/0000-0002-5548-8811
Stephen Ogle https://orcid.org/0000-0003-1899-7446
Stephen Del Grosso https://orcid.org/0000-0001-7486-3958
Nathaniel Mueller https://orcid.org/0000-0003-1857-5104
Deepak Ray https://orcid.org/0000-0002-2856-9608

References

Cairns J E and Prasanna R M 2018 Developing and deploying climate-resilient maize varieties in the developing world Curr. Opin. Plant Biol. 45 226–30
Carmel R 2020 Ibs Latin hypercube samples R package version 1.1.1. (available at: https://CRAN.R-project.org/package=lhs)
Conant R T, Berdanier A B and Grace P R 2013 Patterns and trends in nitrogen use and nitrogen recovery efficiency in world agriculture Glob. Biogeochem. Cycles 27 558–66
Confalonieri R et al 2016 Uncertainty in crop model predictions: what is the role of users? Environ. Model. Softw. 81 165–73
Cramer W, Kicklighter D W, Bondeau A, Ili B M, Churkina G, Nemry B, Ruiny A and Schloss A L Intercomparison, T P O T P N M 1999 Comparing global models of terrestrial net primary productivity (NPP): overview and key results Glob. Change Biol. 5 1–15
Del Grosso S J, Ogle S M, Parton W J and Breidt F J 2010 Estimating uncertainty in N₂O emissions from US cropland soils Glob. Biogeochem. Cycles 24 Gb1009
Duvick D N 2005 The contribution of breeding to yield advances in maize (Zea mays L.). Advances in Agronomy ed D L Sparks (Cambridge, MA: Academic Press) Vol 86 pp 83–145
EPA 2013 V.1. Non-Rice Croplands Global Mitigation of non-CO₂ Greenhouse Gases: 2010–2030 (Washington, DC: United States Environmental Protection Agency) (https://www.epa.gov/sites/default/files/2016-06/documents/mac_report_2013.pdf)
Ewert F et al 2015 Crop modelling for integrated assessment of risk to food production from climate change Environ. Model. Softw. 72 287–303
Fischer T, Byerlee D and Edmeades G 2014 Crop yields and global food security: will yield increase continue to feed the world? vol 158 pp 1–634
Folberth C et al 2019 Parameterization-induced uncertainties and impacts of crop management harmonization in a global gridded crop model ensemble PLoS One 14 e0221862
Folberth C, Gaizer T, Abbaspour K C, Schulin R and Yang H 2012 Regionalization of a large-scale crop growth model for sub-Saharan Africa: model setup, evaluation, and estimation of maize yields Agric. Ecosyst. Environ. 151 21–33
Fox J and Weisberg S 2018 An R Companion to Applied Regression (Sage publications)
Fujimori S et al 2019 A multi-model assessment of food security implications of climate change mitigation Nat. Sustain. 2 386–96
Furbanck R T, Jimenez-Berni J A, George-Jaegli B, Potgieter A B and Deery D M 2019 Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops New Phytol. 223 174–27
Gurung R B, Ogle S M, Breidt F J, Williams S A and Parton W J 2020 Bayesian calibration of the DayCent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty Geoderma 376 114529
Harris I, Osborn T J, Jones P and Lister D 2020 Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset Sci. Data 7 109
Hijmans R J 2020 Geographical data analysis and modeling [R package raster version 3.4-6]
Iooss B et al 2021 sensitivity: global sensitivity analysis of model outputs R package version 1.25.0 (available at: https://CRAN.R-project.org/package=sensitivity)
IPCC 2019 2019 refinement to the 2006 IPCC guidelines for national greenhouse gas inventories Agriculture, Forestry and Other Land Use vol 4, Chapter 5: Cropland
Kamali B, Abbaspour K C, Lehmann A, Wehrli B and Yang H 2018 Uncertainty-based auto-calibration for crop yield—the EPIC + procedure for a case study in sub-Saharan Africa Eur. J. Agron. 93 57–72
Lange S 2019 Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BSD (v1.0) Geosci. Model Dev. 12 3055–70
Lavado R S and Taboada M A 2009 The Argentinean Pamplas: a key region with a negative nutrient balance and soil degradation needs better nutrition management and conservation programs to sustain its future viability as a world agroresource J. Soil Water Conserv. 64 150a–153a
Lobell D B, Cassman K G and Field C B 2009 Crop yield gaps: their importance, magnitudes, and causes Annu. Rev. Environ. Resour. 34 179–204
Marthe P et al 2013 Multimodel ensembles of wheat growth: many models are better than one Glob. Change Biol. 21 911–25
Meisle M et al 2010 Pests, pesticide use and alternative options in European maize production: current status and future prospects J. Appl. Entomol. 134 357–75
Mücher S, De Simone L, Kramer H, de Wit A, Roupioz L, Hazeu G, Boogaard H, Schuingl R, Fritz S and Latham J 2016 A New Global Agro-environmental Stratification (GAES) pp 1566–7197
Parton W J, Hartman M, Ojima D and Schimel D 1998 DAILYCENT and its land surface submodel: description and testing Glob. Planet. Change 19 35–48
Pebesma E J 2018 Simple features for R: standardized support for spatial vector data R J. 10 439
Qaim M 2020 Role of new plant breeding technologies for food security and sustainable agricultural development Appl. Econ. Perspect. Policy 42 129–50
R Core Team 2021 R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing (Vienna, Austria) (available at: https://www.R-project.org/)
Ranum P, Peña-Rosas J P and Garcia-Casal M N 2014 Global maize production, utilization, and consumption Ann. New York Acad. Sci. 1312 105–12
Ray D K, West P C, Clark M, Gerber J S, Prischepov A V and Chatterjee S 2019 Climate change has likely already affected global food production PLoS One 14 e0217148
Ribaut J M, de Vicente M C and Delannay X 2010 Molecular breeding in developing countries: challenges and perspectives Curr. Opin. Plant Biol. 13 213–8
Richards R A 2000 Selectable traits to increase crop photosynthesis and their importance, magnitudes, and causes Glob. Change Biol. 6 226–38
Rubin D B 1988 Using the SIR algorithm to simulate posterior distributions Bayesian Stat. 3 395–402
Russell L I 2019 Emmeans: estimated marginal means, aka least-squares means (available at: https://CRAN.R-project.org/package=emmeans)
Schmidhuber J and Tubiello F N 2007 Global food security under climate change Proc. Natl Acad. Sci. USA 104 19703–8
Seidel S J, Palosuo T, Thorburn P and Wallack D 2018 Towards improved calibration of crop models—where are we now and where should we go? Eur. J. Agron. 94 25–35
Sobol M I 2001 Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates Math. Comput. Simul. 55 271–80
Stehfest E, Heistermann M, Pruess J A, Ojima D S and Alcamo J 2007 Simulation of global crop production with the ecosystem model DayCENT (Article) Ecol. Model. 209 203–19
Tang J and Zhuang Q 2009 A global sensitivity analysis and Bayesian inference framework for improving the parameter estimation and prediction of a process-based terrestrial ecosystem model J. Geophys. Res. 114 D15
Tennenes M 2018 map: thematic maps in R J. Stat. Softw. 84 1–39
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
van de Wouw M, van Hintum T, Kik C, van Treuren R and Visser B 2010 Genetic diversity trends in twentieth century crop cultivars: a meta analysis Theor. Appl. Genet. 120 1241–52

Wheeler T and von Braun J 2013 Climate change impacts on global food security Science 341 508–13

Wickham H et al 2019 Welcome to the Tidyverse J. Open Source Softw. 4 1686

Zaidi S S E A, Vanderschuren H, Qaim M, Mahfouz M M, Kohli A, Mansoor S and Tester M 2019 New plant breeding technologies for food security Science 363 1390–1

Zhang X Y, Trame M N, Lesko L J and Schmidt S 2015 Sobol sensitivity analysis: a tool to guide the development and evaluation of systems pharmacology models CPT Pharmacometrics Syst. Pharmacol. 4 69–79

Zhang Y, Gurung R, Marx E, Williams S, Ogle S M and Paustian K 2020 DayCent model predictions of NPP and grain yields for agricultural lands in the contiguous US J. Geophys. Res. Biogeo. 125