ENVIRONMENTAL RESEARCH LETTERS

LETTER

Observational constraint of process crop models suggests higher risks for global maize yield under climate change

Xiaomeng Yin\(^1\*) and Guoyong Leng\(^{2,\*}\)

1 Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, People’s Republic of China
2 University of Chinese Academy of Sciences, Beijing 100049, People’s Republic of China

* Author to whom any correspondence should be addressed.
E-mail: lenggy@igsnrr.ac.cn

Abstract

Projecting future changes in crop yield usually relies on process-based crop models, but the associated uncertainties (i.e. the range between models) are often high. In this study, a Machine Learning (i.e. Random Forest, RF) based observational constraining approach is proposed for reducing the uncertainties of future maize yield projections by seven process-based crop models. Based on the observationally constrained crop models, future changes in yield average and yield variability for the period 2080–2099 are investigated for the globe and top ten producing countries. Results show that the uncertainties of crop models for projecting future changes in yield average and yield variability can be largely reduced by 62% and 52% by the RF-based constraint, respectively, while only 4% and 16% of uncertainty reduction is achieved by traditional linear regression-based constraint. Compared to the raw simulations of future change in yield average (−5.13 ± 18.19%) and yield variability (−0.24 ± 1.47%), the constrained crop models project a much higher yield loss (−34.58 ± 6.93%) and an increase in yield variability (3.15 ± 0.71%) for the globe. Regionally, the constrained models show the largest increase in yield loss magnitude in Brazil, India and Indonesia. Our results suggest more agricultural risks under climate change than previously expected after observationally constraining crop models. The results obtained in this study point to the importance for observationally constraining process crop models for robust yield projections, and highlight the added value of using Machine Learning for reducing the associated uncertainties.

1. Introduction

Understanding how crop yield would change under global warming is critical for adaptation and mitigation. Through representing the key physiological and phenological processes governing crop growth and yield, process-based crop models have been widely used for large-scale projections of crop yield under various climate change scenarios (Asseng et al. 2014, Rosenzweig et al. 2014, Li et al. 2015, Martre et al. 2015, Franke et al. 2020). However, uncertainties arising from crop model structure, parameters and input data are often high (Asseng et al. 2013, Bassu et al. 2014, Rosenzweig et al. 2014, Mueller et al. 2017, Tao et al. 2018, Ruane et al. 2021), limiting the effective design of adaptation and mitigation strategies. A recent systematic review highlights the need for better understanding and quantifying uncertainty to improve the confidence of crop yield projection under climate change (Chapagain et al. 2022). Therefore, how to estimate and narrow down the uncertainty of yield projections has attracted huge attentions in the crop modeling community (Waha et al. 2015, Ramírez-Villegas et al. 2017, Wang et al. 2017, Toreti et al. 2020, Muller et al. 2021, Ringeval et al. 2021).

To date, various approaches have been proposed to minimize such uncertainties through model development, parameter calibration and parameterization improvement (Iizumi et al. 2009, Maiorano et al. 2017, Makowski 2017, Wang et al. 2017, Brown et al. 2018, Jägermeyr and Frieler 2018). The uncertainties of crop models can also be reduced by using harmonized
meteorological forcing, soil dataset and parameters (Chen and Gournède 2014, Elliott et al 2015, Folberth et al 2016, 2019). Many studies showed that crop model ensemble exhibit better performances than a single or a limited number of models (Bassu et al 2014, Martre et al 2015), especially when fed with harmonized inputs and parameters (Mueller et al 2017). However, such conclusions could be biased, because crop models are not completely independent as some ensemble members share common structure and parameterizations (e.g. EPIC, GEPIEC). Therefore, how to better utilize the information of crop model ensemble for improving future yield predictions remain an open question and deserves more investigations.

Theoretically, if there exists a robust relationship across models between the predictand of interest and predictors that can be observed, then the derived relation function combined with observations of predictors can be used to constrain the future projection of the variable of interest. Such idea is often referred to as ‘emergent constraint’ and has been successfully applied in the climate community for reducing the uncertainty of climate models (Boé et al 2009, Bracegirdle and Stephenson 2012, Cox et al 2013, Klein and Hall 2015, Tsushima et al 2015, Li et al 2017, Eyring et al 2019). Recently, a few studies have tested its efficiency in crop modeling community (Zhao et al 2016, Wang et al 2020). For example, Wang et al (2020) showed that the emergent constraint can reduce the uncertainties of crop models in simulating global yield sensitivity to temperature rise by 12%–54% for maize, rice, soybean and wheat. However, how the constrained models perform in projecting future changes in the year-to-year variability of crop yield remains unknown at the global scale, despite its great implications for food security.

What’s more, though the approach of emergent constraint has been demonstrated useful, its key limitations should be noted. That is, the relationship function fitted between the predictand and predictor across models is often assumed to be linear. For example, Li et al (2017) built a linear relationship between historical western Pacific precipitation and future Indian summer monsoon rainfall across 24 coupled model intercomparison project phase 5 (CMIP5) models, which is used for constraining the future projection of rainfall change. Similarly, Zhao et al (2016) fitted a linear regression between the historical and future sensitivity of rice yield to warming across 17 models, based on which future rice yield change is constrained. However, we hypothesize that the constraint of future yield changes conditioned on historical observations may exhibit non-linear features, given the complex and nonlinear interactions between crops and the environment.

To address the above gaps, we propose a novel emergent constraint approach based on machine learning for global yield projections. The proposed method advances previous studies in that it can handle non-linear dependences between variables. Besides yield averages, we focus on the changes in the inter-annual variability of yields which has great implications for food security given its relevance to yield stability and crop price. Specifically, the following scientific questions are addressed in this study: (a) How much uncertainties of crop models can be reduced after being observationally constrained using machine learning? (b) How will yield average and variability change by the end of 21st century at the global and country scales? (c) How does the constrained crop models compare with the unconstrained ensemble for future yield projections?

2. Materials and methods

2.1. Maize yield from census data and crop model simulations

Historical census maize yield data for the globe and top ten producing countries are obtained from the Food and Agriculture Organization of the United Nations (FAO) Statistical Database food and agriculture organization corporate statistical database (FAOSTAT) (www.fao.org/faostat/en/#data/QC) for the period 1986–2005. Seven global gridded crop models (GGCMs), including EPIC, GEPIEC, IMAGE, LPJ-GUESS, LPJmL, pDSSAT and PEGASUS, driven by five general circulation models, i.e. GFDL-ESM2M, LPSL-CM5A-LR, MIROC-ESM-CHEM, NorESM1-M and HadGEM2-ES, are used to simulate historical (1986–2005) and future (2080–2099) global maize yield. These models differ in their structure, assumptions and parameterizations, and their key information is summarized in table 1. The model simulation outputs (seven GGCMs and two time periods) are downloaded from the Inter-Sector Impact Model Intercomparison Project (ISI-MIP) Fast Track (Rosenzweig et al 2014), which are publicly available from the ISI-MIP Repository at https://data.isimip.org/. These simulations have been widely used for assessing future climate impacts on crop yield, and previous studies showed that uncertainties are several times larger in the highest RCP8.5 than other scenarios (Rosenzweig et al 2014). Therefore, like previous studies (e.g. Schauberger et al 2017), we select RCP8.5 as an example to demonstrate the effectiveness of our approach in reducing the uncertainty of future yield projections. The historical period 1986–2005 is selected to maximize the number of simulations, while the future period 2080–2099 is selected because uncertainties are much greater in the later decades of the 21st century than early periods (Rosenzweig et al 2014). Since these crop models are run at the grid scale driven with gridded data of climate, soil and management, the intra-country variability of environments/crop management are accounted for. These models have been well evaluated and validated separately by each model developing
| Model | Type | Soil | Root distribution | Key citations |
|-------|------|------|-------------------|--------------|
| EPIC  | Site-based | ISRIC-WISE | Automatic N, P, K | Williams (1995), Zaugraudere et al (2006) |
| GEPIC | Site-based | ISRIC-WISE | NP, dynamic application | Site-specific and global H, F, I, O, T, A, N, P, K, BD, AL |
| IMAGE | Global | H, T, A, N, P, K | No nutrient limitation | Global, H, F, I, O, T, A, N, P, K, BD, AL |
| GDD, 2 | Site-specific | ND | GDD, 2 | Global, H, F, I, O, T, A, N, P, K, BD, AL |
| Crop cultivars | | | | Site-specific and global H, F, I, O, T, A, N, P, K, BD, AL |
| Calibration | | | | Site-specific and global H, F, I, O, T, A, N, P, K, BD, AL |
| Notes | | | | Notes: (NA where not applicable). |

**Stresses:**

- **W**: water stress;
- **T**: temperature stress;
- **H**: specific-heat stress;
- **A**: oxygen stress;
- **N**: nitrogen stress;
- **P**: phosphorus stress;
- **K**: potassium stress;
- **BD**: bulk density;
- **AL**: aluminum stress (based on pH and base saturation).

**Calibration:**

- **GDD, 2**: Global GDD requirement (no adaptation); clim. adapt: dynamic GDD requirement (adaptation to climate change).

**Crop cultivars:**

- **GDD**: Growing Degree Days; Number of cultivars; GDD + V: GDD requirements and vernalization requirements; BT base temperature (fixed); clim. adapt: dynamic GDD requirement (adaptation to climate change).

**Notes:**

- **NA**: not applicable.
- **H, T, A, N, P, K**: nitrogen, phosphorus, potassium.
- **GDD**: Growing Degree Days; Site-specific and global H, F, I, O, T, A, N, P, K, BD, AL.
- **HI**: harvest index; LAImax: maximum LAI; HI: harvest index; α: leaf-level photosynthesis scaling factor; β: radiation-use efficiency factor; D: fertilizer application rate; T: temperature stress; A: oxygen stress; N: nitrogen stress; P: phosphorus stress; K: potassium stress; BD: bulk density; AL: aluminum stress (based on pH and base saturation).

**Key citations:**

- Williams (1995), Zaugraudere et al (2006).
- Site-specific and global H, F, I, O, T, A, N, P, K, BD, AL.
- Site-specific and global H, F, I, O, T, A, N, P, K, BD, AL.
- Global, H, F, I, O, T, A, N, P, K, BD, AL.
- Site-specific and global H, F, I, O, T, A, N, P, K, BD, AL.
- Global, H, F, I, O, T, A, N, P, K, BD, AL.
- Site-specific and global H, F, I, O, T, A, N, P, K, BD, AL.
- Global, H, F, I, O, T, A, N, P, K, BD, AL.
- Site-specific and global H, F, I, O, T, A, N, P, K, BD, AL.
team and collectively by the ISI-MIP community (e.g. Rosenzweig et al 2014, Schauerger et al 2017, Leng and Hall 2020, Yin et al 2022). More information on the model setup and simulation protocol can be found in the literature (Rosenzweig et al 2014, Warszawski et al 2014). To be consistent with the scale of FAO census yield, the simulated maize yields at a spatial resolution of 0.5° × 0.5° are aggregated to the global and country scales, with weights assigned according to the gridded harvest area map from the MIRCA 2000 (Portmann et al 2010).

2.2. Machine-learning based emergent constraint

Unlike calibrations aiming for matching historical simulations closer with observations in terms of distribution and mean, we propose a machine-learning based approach of emergent constraint for constraining or narrowing the uncertainty of crop model simulations for the future period. More specifically, the basic concept of emergent constraint approach is that when we analyze a large ensemble of crop models, a clear relationship may emerge between a variable \(X\) in models’ simulations in the historical period and a variable \(Y\) in models’ projections in the future period. When such emergent relationship is identified, we can then plug in actual observations of variable \(X\) to narrow the uncertainty of projections for variable \(Y\). Similar ideas have been successfully implemented in the climate modeling community (e.g. Cox et al 2018, Thackeray and Hall 2019, Tokarska et al 2020, Shigama et al 2022). Like previous studies (e.g. Mueller et al 2017, Schewe et al 2019, Franke et al 2020, Heinicke et al 2022, Yin et al 2022), the FAO yield data is treated as observations and used for constraining GGCMs to reduce their uncertainty in future yield projections.

In our study, uncertainty refers to the range between crop models, calculated as one standard deviation of 35 model simulations, similar to previous studies (Zhao et al 2016, Tokarska et al 2020, Shigama et al 2022). For each model simulation, two indicators are calculated for analysis, i.e. long-term yield average and interannual yield variability over a 20 year period. Here, the interannual yield variability is measured by the metric of coefficient of variance, which is calculated by dividing the standard deviation of a 20 year yield time series by its mean value.

Previous studies often assume a linear emergent relation across models (Zhao et al 2016, Hall et al 2019, Wang et al 2020), which we hypothesize may not hold in the future under a changing environment. To test our hypothesis, we derive the emergent relation based on Random Forest (RF) which can handle non-linear relations without prior determination of the relationship function (Jeong et al 2016, Hoffman et al 2018, Feng et al 2019). Specifically, the emergent relationship for yield average and yield variability are derived using equation (1)

\[
y_{f,m} = RF(y_{h,m})
\]

where RF denotes the RF, \(y_{h,m}\) and \(y_{f,m}\) are the historical and future indicators (i.e. yield average and yield variability) for model \(m\), respectively. To demonstrate the added value of Machine-learning, the traditional linear regression is adopted for comparison with RF-based constraint:

\[
y_{f,m} = a_0 + a_1 \cdot y_{h,m} + \epsilon_{a,m}
\]

where \(a_0\) and \(a_1\) are the intercept and coefficient, respectively; \(\epsilon_{a,m}\) is the error term.

Following Zhao et al (2016), the least squares model error \(s^2\) is calculated as

\[
s^2 = \frac{1}{n-2} \sum_{m=1}^{n} (y_{f,m} - \hat{y}_{f,m})
\]

and when given \(y_{h,m}\), the predicted error for \(\hat{y}_{f,m}\) is

\[
\sigma(\hat{y}_{f,m}|y_{h,m}) = \sqrt{\frac{1}{n} \cdot \frac{(y_{h,m} - y_h)^2}{\sigma^2 + \frac{\sigma^2 y_h}{n}}}
\]

where \(\sigma^2\) is the variance of \(y_h\); \(\sigma_{y_h}^2 = \frac{\sum_{m=1}^{n} (y_{h,m} - y_h)^2}{n}\), \(y_{h,m}\) is the historical indicator for model \(m\), \(y_h\) is the average of \(y_{h,m}\) for all models, \(n\) is total number of models, \(y_{f,m}\) is the simulated future indicator and \(\hat{y}_{f,m}\) is the predicted future indicator based on the emergent ‘historical-future’ relationship for model \(m\).

The predicted \(\hat{y}_{f,m}\) is fit into a Gaussian probability density function (PDF), and its standard deviation of \(\sigma(\hat{y}_{f,m}|y_{h,m})\): \(p(\hat{y}_{f,m}|y_{h,m}) \sim N(\hat{y}_{f,m}, \sigma(\hat{y}_{f,m}|y_{h,m})^2)\) is used to derive the constrained PDF. Based on the emergent relationship, the FAO observed yield average and yield variability are used to constrain future projections of maize yield average and yield variability for the globe and top ten producing countries. As for the unconstrained PDF, \(p(\epsilon_m) \sim N(c_m, \sigma_{c_m}^2)\) is defined, where \(c_m\) is the change of \(y_{f,m}\) relative to \(y_{h,m}\) for model \(m\), \(c_m\) and \(\sigma_{c_m}\) are the mean and standard deviation of \(c_m\), respectively. Based on that, a Gaussian probability distribution is fitted for future changes in yield average and variability in the multi-model ensemble.

3. Results

3.1. Uncertainties of process crop model simulations

Figure 1 shows the boxplots of crop model simulations for the historical and future yield average and variability for the globe and top ten producing countries. It is evident that a large discrepancy exists across crop models for yield simulations both regionally and globally. For the globe as a whole, a large range of yield average from 3.29 t ha\(^{-1}\) to 9.22 t ha\(^{-1}\) is found...
across models for the historical period, and such discrepancy is further amplified to 2.73–10.04 t ha$^{-1}$ for the future period. Regionally, the largest discrepancy is observed for India with the range between 0.42 and 11.70 t ha$^{-1}$, followed by South Africa and Argentina. These revealed ranges across crop model projections point to the importance of reducing the uncertainties for robust yield projections. In general,
the ensemble median of crop models tends to match observation more closely than an individual model, consistent with previous studies (Rosenzweig et al 2013, Bassu et al 2014, Li et al 2015, Martre et al 2015, Wang et al 2016). However, such conclusion does not hold at the country scale, and process models show a significant overestimation of yield average in Brazil, Mexico, India, Indonesia and South Africa. Relative to the historical period (1986–2005), the ensemble median of process models simulates a large decrease up to −8.62% for global maize yield during the period 2080–2099. Regionally, negative effects of global warming are projected in low latitude regions while positive effects are simulated in high-latitude regions, consistent with previous studies (Rosenzweig et al 2014). For example, the largest yield decrease is found in USA with a yield loss up to 23.18%, while an increase by 4.99% is projected for Canada. Such regional differences reflect the spatial heterogeneity of cropping system and yield sensitivity to climate change as represented in GGCMs. Indeed, the baseline climate in most of low latitude regions is already quite close to the high-temperature thresholds for suitable production, and thus rising temperatures are detrimental there and vice versa for high latitude region (e.g. Canada).

Similar to yield average, the ensemble median of process model exhibits a promising performance for reproducing historical year-to-year variability of global maize yield. However, in contrast, process models show an underestimation of yield variability in most of major producing countries, including China, Brazil, Mexico, South Africa and Canada. Relative to the historical period, a slight decrease of 0.27% is projected by model ensemble median for future yield variability. Regionally, a mixed pattern of yield variability change is observed across the top ten countries, with a projected increase in USA, France and Canada but a decrease in the remaining countries. However, the uncertainties are also high at the global and country level, and the largest uncertainty is found in South Africa with the range of values between 1.79% and 33.68%. Overall, the revealed large uncertainties highlight the need for observationally constraining crop models for robust projection of future changes in yield average and variability.

3.2. The emergent relationship

Figure 2 shows the scatterplot of projected changes in yield average against historical simulations for the globe and top ten producing countries. In general, models which simulate higher yield in the historical period is also likely to project higher yield in the future period. Such an emergent relation can be well described by both traditional linear regression and RF, with a $R^2$ of 0.81 and 0.97 for the globe, respectively. Similar to global-scale results, robust emergent relationships are detected for all the top ten countries. Compared to linear regression, RF exhibits a higher performance in reproducing the emergent relation. Since the main difference between linear regression and RF is that the latter can capture non-linear relations while the former cannot, the difference in their performances indicates the role of non-linear effects for the emergent constraint. Indeed, a non-linear curve is fitted by RF (red line in figure 2), which outperforms the linear regression (blue line in figure 2).

Similarly, models showing larger yield variability tend to predict more instable yields in the future (figure 3). For the globe as a whole, linear regression and RF exhibit a high $R^2$ of 0.70 and 0.91 for fitting the emergent relation, respectively. Such a high performance is also found for most countries, with the highest $R^2$ of 0.72 and 0.93 by linear regression and RF for China, respectively. In Brazil, however, linear regression fails to capture the emergent relation with a $R^2$ of 0.10, while RF exhibit a high performance with $R^2$ 0.86. Overall, our results demonstrate that RF is better than linear regression in describing the emergent relations at the global and country levels (supplementary table S1), which allows for observationally constraining the projections of future changes in yield average and yield variability.

3.3. Constraint on future changes in yield average and yield variability

Figure 4 shows the probability density function (PDF) of the constrained and unconstrained (raw) projections of future changes in yield average for the globe and top ten producing countries. For the globe as a whole, a large uncertainty of 18.19% is found by the unconstrained crop models. After observationally constraining crop models, such uncertainty is reduced to 6.93% by RF, while a slight uncertainty reduction is achieved by linear regression. The significant uncertainty reduction achieved with RF in our study is also higher than most of previous studies using emergent constraint (Cox et al 2013, Wenzel et al 2016, Zhao et al 2016, Eyring et al 2019, Wang et al 2020, Xu et al 2020), demonstrating the added value of using RF for the constraint. Importantly, the constrained models based on linear regression and RF both project a more severe yield loss of −31.01% and −34.58% for the globe, respectively, compared to −5.13% in the unconstrained models. This indicates that previous yield change projections based on raw crop models may have largely underestimated yield loss magnitude (Rosenzweig et al 2014, Mueller et al 2017).

Regionally, larger uncertainties of unconstrained models up to 20.93%–31.22% are found for the top ten producing countries (supplementary table S2). When constrained against observations by RF, the uncertainty of projected yield changes is largely reduced by 31%–81%, while a much smaller uncertainty reduction is obtained by linear regression (−6%–38%). Overall, compared to the raw projections, the constrained models based on RF
and linear regression both suggest more yield loss magnitude for most countries, except for France and Canada. For example, maize yield in USA is projected to decrease by $-24.70\%$ based on RF, which is five times larger than the loss in the raw crop models ($-4.64\%$). In Brazil, the projected yield loss could even increase from $-13.63\%$ in raw models to $-82.22\%$ in the RF-based constrained models. Compared to the projections based on linear regression, the RF-based constrained models simulate more yield loss magnitude in USA, China, Brazil, India and Indonesia, but smaller decrease in Mexico.
Argentina and South Africa. Such differences between the constrained projections between linear regression and RF indicate the role of non-linear effects in the constraint.

As for yield variability (figure 5), the unconstrained models show an uncertainty range of 1.47% for the globe, which is six times larger than the ensemble mean change value (−0.24%). After applying the emergent constraint by RF, the uncertainty is reduced by half to 0.71%, which triples the magnitude of uncertainty reduction achieved by linear regression. In contrast to the projected decrease in yield...
variability by the raw models, the constrained models based on RF shows a large increase of yield variability of 3.15%, implying more unstable yield in the future. Such a disagreement in the change direction of yield variability cannot be obtained, however, when using the linear regression for the constraint.

For the top ten producing countries, the unconstrained models show diverse behaviors for projecting
future changes in yield variability, which ranges from 1.54% to 9.55% across countries (supplementary table S3). The RF-based constraint can well reduce the uncertainty by 41%–80%, much larger than that based on linear regression (−3%–51%). In China, the linear regression-based constraint even leads to an amplification of the uncertainty. Similar to global-scale results, the RF-based constrained models project a future increase in yield variability in most countries. For example, maize yield variability in USA is projected to increase by 0.22% based on RF, while a decrease is simulated by the raw crop models (−0.19%). Given

Figure 5. Same as figure 4, but for yield variability (CV).
the promising performance of RF in describing the emergent relation as shown in figure 3, the projected increase in yield variability suggests more risks for maize production than previously expected based on the raw crop models.

4. Discussions

It is well known that GGCMs differ in their approaches, assumptions, and structures, and often show large uncertainty in future yield projections (Asseng et al. 2013, Bassu et al. 2014, Rosenzweig et al. 2014, Tao et al. 2018). Calibration in a detailed manner would help reduce the uncertainties, but is challenging for seven crop models across the globe. In addition, calibration outcome is sensitive to the choices of parameters to calibrate, data used for calibration, calibration method (e.g. Ordinal Least Square, Markov chain Monte Carlo, Bayesian and Generalized Likelihood Uncertainty Estimation) and etc (Angulo et al. 2013, He et al. 2017, Seidel et al. 2018). Therefore, instead of further recalibrating the seven GGCMs globally, we propose an alternative approach (i.e. emergent constraint) which is efficient to implement and also effective in reducing the uncertainty in future yield projections. Specifically, our approach is based on the fitting of emergent relations that would emerge across the ensemble of crop models. Besides traditional linear regression, we introduce machine-learning (i.e. RF) to demonstrate its added value for the constraint. Our results indicate that both regression and RF can well capture the emergent relations, with RF outperforming regression in the fitting. Importantly, we show that observationally constraining ISI-MIP crop models can largely reduce their uncertainties in future yield projections.

For better interpretation of our findings, however, there are some limitations that should be acknowledged. We also identify several aspects of future research needs as below.

First, we constrain future changes in yield average and yield variability at the country and global scale, ignoring the spatial heterogeneity. This is because FAO is the only publicly available crop yield data, and only global and country-level values are provided. When aggregating the gridded simulations of crop models to global and country scales, grid-specific weights are assigned according to its harvest area, which is hold constant due to the lack of time-varying harvest area maps. Leng and Huang (2017) showed that US maize yield response depends on the spatial distribution of crops. Sloat et al. (2020) further demonstrated the importance of crop migration in mitigating the adverse climate impacts. However, how the changes in cropping area would affect the constraint of yield changes requires further investigation, and is not within the goal of our study as process models adopted in our study does not account for such dynamics.

Second, the uncertainties related to the effects of climate change on maize yield have not been analyzed. As shown in previous studies (Zhao et al. 2016, Wang et al. 2020), the uncertainties of crop yield responses to temperature increase have been successfully investigated with the emergent constraint approach. Besides, there are large changes of climate factors and their interactions in the future scenarios, including temperature, precipitation and CO2, which would disturb the estimation of individual factor (Rosenzweig et al. 2014). Therefore, we focus on the future changes of yield and yield variability and using the emergent constraint to reduce derived uncertainties.

Third, the uncertainty from observed data is not completely sampled. As demonstrated in the literature (Cox et al. 2013, Zhao et al. 2016, Eyring et al. 2019), emergent constraint is conducted in two steps: (a) first, develop the emergent-relationship across models; (b) second, use observations to constrain models for future projections. So, there are two sources of uncertainties that could affect the performance of emergent constraint. Zhao et al. (2016) collected observed temperature sensitivities of rice yield from 83 field experiments based on which the uncertainties from the observations can be derived. Due to lack of such information for field-level yield observations worldwide, we focus on the uncertainty in emergent relationship. To reduce such uncertainty, we need to explore a robust relationship. To address this, we use RF for deriving the emergent relationship, which lead to smaller uncertainties than using traditional linear regression.

Fourth, it is well-known that crop yield depends on many secondary traits and bio-physical processes. Therefore, it is important to constrain its primary drivers, processes and secondary traits to advance our scientific understandings of future changes in crop yield. Here, we constrain crop yield only, because crop yield is the priority variable requested to be outputted by each crop modeling team within ISI-MIP, while other variables related to crop growth, phenology and other secondary traits are not routinely archived. What’s more, observations on crop yield are available worldwide from the FAO database, while observations of other secondary traits are not readily available worldwide. Future efforts should be devoted to constraining both gain yield and its secondary traits, once both simulations and observations are available. Fifth, climate change impacts on crop yield could be further modulated by various adaptation strategies. For example, cultivar choice has been shown to be effective in reducing the negative effects of future warming (Liu et al. 2013, Waha et al. 2013, Huang et al. 2020). Besides cultivar improvement, other adaptation strategies such as irrigation, fertilization, soil mulching and crop migration could further modulate crop yield response to climate change.
(Karlen et al. 2013, Qin et al. 2015, Seifert and Lobell 2015, Troy et al. 2015, Sloat et al. 2020). To account for all possible factors influencing crop yield is not within the scope of this study. Rather, this study aims to propose an efficient and effective approach for reducing the uncertainty of crop models in future yield projections. In our study, we use the simulations from ISI-MIP, which have been widely used in the community for assessing future climate impacts on crop yield (Rosenzweig et al. 2014, Warszawski et al. 2014, Schaubberger et al. 2017). Unfortunately, the simulation protocol is not designed for evaluating adaptation effects. Therefore, the output of our approach would depend on the specific environmental, crop management and genotypic factors represented in these specific simulations.

Despite the above limitations, our study contributes to literature by demonstrating the added value of using Machine Learning for observationally constraining process crop models for projecting future changes in yield average and variability. Indeed, combining machine learning and process models lead to a large reduction of uncertainties for future yield projections, with significant influence on the changes in yields not only for the long-term average but also for the year-to-year variability. Our proposed approach is demonstrated to be efficient and effective, and can therefore be easily extended to other crops, variables, periods, regions and environments of interest, once ensemble simulations and observations are both available.

5. Conclusions

Process-based crop models have been widely used for projecting future yield changes, but often show substantial uncertainties. In this study, we adopt the approach of emergent constraint for observationally constraining the projections of 35 crop model simulations for future changes in global maize yield average and yield variability. Instead of assuming a linear emergent relationship as pursued in previous studies, we propose a new approach based on Machine-Learning, which can handle non-linear relations between the predicant and predictors. Compared with the traditional approach, the RF shows a much better performance in modeling the emergent relations, leading to a large reduction of the uncertainty of crop model projections of both yield average and yield variability at the global and country scales. Specifically, the uncertainty range of constrained crop models is reduced by 62% and 52% for projecting future changes in global maize yield and yield variability, respectively, which more than triple that achieved with traditional linear regression. Based on the observationally constrained models by RF, a more severe yield loss (−34.58 ± 6.93%) and larger year-to-year variability (3.15 ± 0.71%) are simulated than based on the raw models by the end of 21st century. Regionally, the constrained models show the largest increase in yield loss magnitude in Brazil, India and Indonesia compared to the raw models. Our results suggest that previous studies based on the raw crop models may have largely underestimated the agricultural risks under climate change. This study highlights the added value of using machine learning for observationally constraining process crop models for more robust projections of future yield changes.

Data availability statement

The data that support the findings of this study are openly available. Census data on annual maize yields of the globe and top ten producing countries are obtained from the Food and Agriculture Organization of the United Nations (FAO) FAOSTAT database (www.fao.org/faostat/en/#data/QC). The simulated yields by process-based models are from the Inter-Sector Impact Model Intercomparison Project (ISI-MIP) Fast Track project at www.isimip.org. The crop harvest area map is from www.uni-frankfurt.de/45218023/MIRCA.

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

Acknowledgments

This research was funded by the National Natural Science Foundation of China (No. 42077420) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA28060100).

ORCID iD

Xiaomeng Yin @ https://orcid.org/0000-0003-3990-1124

References

Angulo C, Rötter R, Lock R, Enders A, Fronzek S and Ewert F 2013 Implication of crop model calibration strategies for assessing regional impacts of climate change in Europe Agric. For. Meteorol. 170 32–46

Asseng S et al. 2013 Uncertainty in simulating wheat yields under climate change Nat. Clim. Change 3 827–32

Asseng S et al. 2014 Rising temperatures reduce global wheat production Nat. Clim. Change 5 143–7

Bassu S et al. 2014 How do various maize crop models vary in their responses to climate change factors? Glob. Change Biol. 20 2301–20

Boé J, Hall A and Xu Q 2009 September sea-ice cover in the Arctic Ocean projected to vanish by 2100 Nat. Geosci. 2 341–3

Bondeau A et al. 2007 Modelling the role of agriculture for the 20th century global terrestrial carbon balance Glob. Change Biol. 13 679–706

Bouwman A F, Kram T, Klein T and Goldewijk K (eds) 2006 Integrated Modelling of Global Environmental Change an Overview of IMAGE 2.4 (The Hague: PBL Netherlands Environmental Assessment Agency)

Bracegirdle T J and Stephenson D B 2012 On the robustness of emergent constraints used in multimodel climate change projections of Arctic warming J. Clim. 26 669–78

[12]
Brown H, Guth N and Holzworth D 2018 Crop model improvement in APSIM: using wheat as a case study Eur. J. Agron. 100 141–50
Chapagain R, Remenyi T A, Harris R M B, Mohammed C L, Guth N, Wallach D, Rezaei E E and Ojeda J J 2022 Decomposing crop model uncertainty: a systematic review Field Crops Res. 279 108448
Chen Y and Courtois D H 2014 Data Assimilation to Reduce Uncertainty of Crop Model Prediction with Convolution Particle Filtering Ecol. Modelling 290 165–77
Cox P M, Huntingford C and Williamson M S 2018 Emergent constraint on equilibrium climate sensitivity from global temperature variability Nature 553 319–22
Deryng D, Conway D, Ramankutty N, Price J and Warren R 2014 Global crop yield response to extreme heat stress under multiple climate change futures Environ. Res. Lett. 9 034011
Deryng D, Sacks W J, Barford C C and Ramankutty N 2011 Simulating the effects of climate and agricultural management practices on global crop yield Glob. Biogeochem. Cycles 25 GB2006
Elliott J et al 2015 The global gridded crop model intercomparison: data and modeling protocols for phase 1 (v1.0) Geosci. Model Dev. 8 261–77
Elliott J, Kelly D, Chrysanthiscopoulos J, Glotter M, Hoffman S, Jones J W, Hoogenboom G, Porter C H, Boote K J, Batchelor W D, Hunt L A, Wilkes P W, Singh U, Gisjesman A J and Ritchie J T 2003 The DSSAT cropping system model Eur. J. Agron. 18 235–65
Karlen D L, Kovar J L, Cambardella C A and Colvin T S 2013 Thirty-year tillage effects on crop yield and soil fertility indicators Soil Tillage Res. 130 24–41
Klein S A and Hall A 2015 Emergent constraints for cloud feedbacks Curr. Clim. Change Rep. 1 276–87
Leemans R and Solomon A M 1993 Modeling the potential change in yield and distribution of the earth's crops under a warmed climate Clim. Res. 3 79–96
Leng G and Hall J W 2020 Predicting spatial and temporal variability in crop yields: an inter-comparison of machine learning, regression and process-based models Environ. Res. Lett. 15 044027
Leng G and Huang M 2017 Crop yield response to climate change varies with crop spatial distribution pattern Sci. Rep. 7 14463
Li G, Xie S-P, He C and Chen Z 2015 Western Pacific emergent constraint lowers projected increase in Indian summer monsoon rainfall Nat. Clim. Change 7 708–12
Li T et al 2015 Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions Glob. Change Biol. 21 1326–41
Linderskog M, Arneth A, Bondeau A, Waha K, Sequeijt J, Olin S and Smith B 2013 Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa Earth Syst. Dyn. 4 385–407
Liu J, Williams J R, Zehnder A J B and Yang H 2007 GEPICH—modelling wheat yield and crop water productivity with high resolution on a global scale Agric. Syst. 94 478–93
Liu Z, Hubbard K G, Lin X and Yang X 2013 Negative effects of climate warming on maize yield are reversed by the changing of sowing date and cultivar selection in Northeast China Glob. Change Biol. 19 3481–92
Maiorano A et al 2017 Crop model improvement reduces the uncertainty of the response to temperature of multi-model ensembles Field Crops Res. 202 5–20
Malowski D 2017 A simple Bayesian method for adjusting ensemble of crop model outputs to yield observations Eur. J. Agron. 58 476–83
Martre P et al 2015 Multimodel ensembles of wheat growth: many models are better than one Glob. Change Biol. 21 911–25
Mueller C et al 2017 Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications Geosci. Model Dev. 10 1403–22
Muller C et al 2021 Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios Environ. Res. Lett. 16 034009
Portmann F T, Siebert S and Döll P 2010 MIRCA2000-global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling Glob. Biogeochem. Cycles 24 GB1011
Quin W, Hu C and Oenema O 2015 Soil mulching significantly enhances yields and water and nitrogen use efficiencies of maize and wheat: a meta-analysis Sci. Rep. 5 16210
Ramireir-Villegas M, Koehler A K and Challinor A J 2017 Assessing uncertainty and complexity in regional-scale crop model simulations Eur. J. Agron. 88 84–95
Ringeval B, Müller C, Pugh T A M, Mueller N D, Ciais P, Folberth C, Liu W, Deheke P and Pellerin S 2021 Potential yield simulated by global gridded crop models: using a
process-based emulator to explain their differences Geosci. Model Dev. 14 1639–56
Rosenzweig C et al 2013 The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies Agric. For. Meteorol. 170 166–82
Rosenzweig C et al 2014 Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison Proc. Natl Acad. Sci. USA 111 3268–73
Ruane A C et al 2021 Strong regional influence of climatic forcing datasets on global crop model ensembles Agric. For. Meteorol. 300 108313
Schauberger B et al 2017 Consistent negative response of US crops to high temperatures in observations and crop models Nat. Commun. 8 15931
Schewe J et al 2019 State-of-the-art global models underestimate impacts from climate extremes Nat. Commun. 10 1003
Seidel S J, Palosuo T, Thorburn P and Wallach D 2018 Towards improved calibration of crop models—where are we now and where should we go! Eur. J. Agron. 94 25–35
Seifert C A and Lobell D B 2015 Response of double cropping to warmer temperature in observations and crop models Proc. Natl Acad. Sci. USA 112 2479–509
Yin X and G Leng 2022 Disentangling the separate and confounding effects of temperature and precipitation on global maize yield using machine learning, statistical and process crop models Environ. Res. Lett. 17 044036
Zhao C et al 2016 Plausible rice yield losses under future climate warming Nat. Plants 3 16202

Troy T J, Kipgen C and Pal I 2015 The impact of climate extremes and irrigation on US crop yields Environ. Res. Lett. 10 054015
Tsuchiya Y et al 2015 Robustness, uncertainties, and emergent constraints in the radiative responses of stratuscumulus cloud regimes to future warming Clim. Dyn. 46 3023–39
Waha K, Huth N, Carberry P and Wang E 2015 How model and input uncertainty impact maize yield simulations in West Africa Environ. Res. Lett. 10 024017
Waha K, Müller C, Bonneau A, Dietrich J P, Kurukulasuriya P, Heinke J and Lotze-Campen H 2013 Adaptation to climate change through the choice of cropping system and sowing date in sub-Saharan Africa Glob. Environ. Change 23 130–43
Waha K, van Bussel L G J, Müller C and Bonneau A 2012 Climate-driven simulation of global crop sowing dates Glob. Ecol. Biogeogr. 21 247–59
Wang B, Liu D L, Macadam J, Alexander L V, Abramowitz G and Yu Q 2016 Multi-model ensemble projections of future extreme temperature change using a statistical downscaling method in south eastern Australia Clim. Change 138 85–98
Wang E et al 2017 The uncertainty of crop yield projections is reduced by improved temperature response functions Nat. Plants 3 17102
Wang X et al 2020 Emergent constraint on crop yield response to warmer temperature from field experiments Nat. Sustain. 3 908–16
Warszawski L, Fricker K, Huber V, Piontek F, Serdeczny O and Schewe J 2014 The inter-sectoral impact model intercomparison project (ISI-MIP): project framework Proc. Natl Acad. Sci. USA 111 3228–32
Wenzel S, Cox P M, Eyring V and Friedlingstein P 2016 Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO2 Nature 538 499–501
Williams J R 1995 The EPIC Computer Models of Watershed Hydrology ed V P Singh (Littleton, CO: Water Resources Publications) ch 25, pp 909–1000
Xu W et al 2020 Reducing uncertainties of future global soil carbon responses to climate and land use change with emergent constraints Glob. Biogeochem. Cycles 34
Yin X, Leng G and Yu I 2022 Disentangling the separate and confounding effects of temperature and precipitation on global maize yield using machine learning, statistical and process crop models Environ. Res. Lett. 17 044036
Zhao C et al 2016 Plausible rice yield losses under future climate warming Nat. Plants 3 16202