Agricultural and Forest Meteorology 214–215 (2015) 134–147

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Emulating maize yields from global gridded crop models using statistical estimates

Elodie Blanc[a,*], Benjamin Sultan[b]

[a] Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA
[b] Laboratoire d'Océanographie et de Climatologie par Expérimentation et Approche Numérique, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris, France

ARTICLE INFO

Article history:
Received 6 February 2015
Received in revised form 12 July 2015
Accepted 16 August 2015

Keywords:
Crop yields
Crop model
Statistical model
Climate change
Maize

ABSTRACT

This study estimates statistical models emulating maize yield responses to changes in temperature and precipitation simulated by global gridded crop models. We use the unique and newly released *Inter-Sectoral Impact Model Intercomparison Project Fast Track* ensemble of global gridded crop model simulations to build a panel of annual maize yields simulations from five crop models and corresponding monthly weather variables for over a century. This dataset is then used to estimate statistical relationship between yields and weather variables for each crop model. The statistical models are able to closely replicate both in- and out-of-sample maize yields projected by the crop models. This study therefore provides simple tools to predict gridded changes in maize yields due to climate change at the global level. By emulating crop yields for several models, the tools will be useful for climate change impact assessments and facilitate evaluation of crop model uncertainty.

1. Introduction

The impact of climate change on crop yields has been extensively studied. To estimate these impacts, two approaches are usually taken: (i) process-based crop models, which represent mechanistically or functionally the effect of weather, soil conditions, management practices and abiotic stresses on crop growth and yields; or (ii) statistical techniques that empirically estimate the effect of weather on crop yields while controlling for other factors based on historical observations.

Process-based crop models are able to consider the detailed effect of weather and climate change on crop yields at the global level or at the site level by considering monthly, daily, or even hourly weather information (Basso et al., 2013). Some models can also capture other factors, such as pest damages, soil properties, fertilizer application, planting dates, and the carbon dioxide (CO_2) fertilization effect. These models are either calibrated at the field scale (Elliott et al., 2013; Izaurralde et al., 2006; Jones et al., 2003), the national level (Bondeau et al., 2007) or the grid cell level across the globe (Deryng et al., 2011). These models can simulate a wide range of weather and environmental conditions, but are computationally demanding and sometimes proprietary, which limits their accessibility.

Statistical models, usually in the form of regression analysis, on the other hand, use observed data to estimate the impact of weather on crop yields and are usually based on data aggregated by month (Carter and Zhang, 1998), growth stage (Dixon et al., 1994) or year (Blanc, 2012; Schlenker and Lobell, 2010). Regression analyses usually consider the effect of temperature and precipitation on crop yields (Corobov, 2002; Lobell and Field, 2007; Nicholls, 1997) and its derived composites, such as growing degree days (GDD) (Lobell et al., 2011), evapotranspiration (Blanc, 2012), and drought indices (Blanc, 2012; Carter and Zhang, 1998; Lobell et al., 2014). Some studies control for alternative effects, such as cloud cover (You et al., 2009); sources of water availability such as proximity to streams (Blanc and Strobl, 2014) and dams (Blanc and Strobl, 2013; Strobl and Strobl, 2010); management strategies, such as fertilizer application (Cuculeanu et al., 1999) or changes in planting dates (Alexandrov and Hoogenboom, 2000); and technological trends (Lobell and Field, 2007). The ability of these models to provide large-scale yields estimates is limited by data availability, and they are thus generally based on crop yield data averaged globally (Lobell and Field, 2007), at the country level (Blanc, 2012; Schlenker and Lobell, 2010), or at the county level (Lobell and Asner, 2003).

The out-of-sample predictive ability of statistical models is a concern when estimating impacts for scenarios of climate change not previously observed. This issue has been considered in recent

---

* Corresponding author.
E-mail address: eblanc@mit.edu (E. Blanc).

http://dx.doi.org/10.1016/j.agrformet.2015.08.256
0168-1923/© 2015 Elsevier B.V. All rights reserved.

studies by Holzkämper et al. (2012) and Lobell and Burke (2010) using the so-called 'perfect model' approach, which consists of training a statistical model on the output of a process-based crop model, assuming that this output is 'true'. The main aim of these studies is to evaluate the ability of statistical models to provide predictions out-of-sample. They find that statistical models are capable of replicating the outcomes of process-based crop models reasonably well. The spatial and temporal scope of these studies is, however, fairly small. Oyebamiji et al. (2015) expand on these studies and estimates an empirical crop yield emulator at the global level for five different crops but, as in previous studies, they only consider one process-based crop model. This is a concern because the choice of crop model is an important source of uncertainty in climate change impact assessments on crop yields (e.g. Bassu et al., 2014; Mearns et al., 1999). Therefore, having access to a tool capable of replicating yields from a wide ensemble of crop models would facilitate the analysis of crop model uncertainty in climate change impact assessments.

To address the limitations of simulations based on processed-based models and to consider crop model uncertainty, we design an ensemble of simple statistical models able to accurately replicate the outcomes of process-based crop models at the grid cell level over the globe using only a limited set of weather variables. To this end, we use the recently released *Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) Fast Track* experiment dataset of global gridded crop models (GGCM) simulations. This project was coordinated by the Agricultural Model Intercomparison and Improvement Project (AgMIP) (Rosenzweig et al., 2013) as part of ISI-MIP (Warszawski et al., 2014). To enable comparison across models, all GGCMs are driven with consistent bias-corrected climate change projections derived from the Coupled Model Intercomparison Project, phase 5 (CMIP5) archive (Hempel et al., 2013; Taylor et al., 2012). Our statistical models are trained on the crop yields simulated by these process-based crop models and are subject to the widest range of climate conditions estimated in CMIP5. The statistical models are then used to predict the spatial responses of maize yields to weather. Differences between predictions from the process-based and statistical models are then assessed in order to measure how well statistical models can capture yield responses to weather variations driven by climate change.

Based on the evaluation of a large set of weather variables, non-linear transformations and interactions effects, we show that a simple specification including temperature and precipitation in polynomial form and interaction terms performs relatively well. Various validation exercises show that out-of-sample maize yield predictions are reasonably accurate, especially with respect to long-term trends. Robustness analyses considering either transformed dependent variable, more precise representations of the growing season, or region-specific estimates support the overall preferability of the parsimonious specification for global climate change projections.

This paper has five further sections. Section 2 presents the data and methods used to statically estimate relationship between yields and weather variables. Results are presented and discussed in Section 3. The models are validated in Section 4 and sensitivity analyses are performed in Section 5. Section 6 concludes.

2. Material and methods

2.1. Data

Data used in this study are sourced from the *ISI-MIP Fast Track* experiment, an inter-comparison exercise of global gridded process-based crop models using the CMIP5 climate simulations.[1] In this exercise, several modeling groups provided results from global gridded process-based crop models run under the same set of weather and CO_2 concentration inputs.

2.1.1. Crop yields and growing seasons

Crop yields and growing season information are obtained from GGCMs members of the *ISI-MIP Fast Track* experiment. Based on data availability, we consider five crop models: the Geographic Information System (GIS)-based Environmental Policy Integrated Climate (GEPIC) model (Liu et al., 2007; Williams, 1995), the Lund Potsdam-Jena managed Land (LPJmL) dynamic global vegetation and water balance model (Bondeau et al., 2007; Waha et al., 2012), the Lund-Potsdam-Jena General Ecosystem Simulator (LPJ-GUESS) with managed land model (Bondeau et al., 2007; Lindeskog et al., 2013; Smith et al., 2001), the parallel Decision Support System for Agro-technology Transfer (pDSSAT) model (Elliott et al., 2013; Jones et al., 2003), and the Predicting Ecosystem Goods And Services Using Scenarios (PEGASUS) model (Deryng et al., 2011).

Each GGCM simulation provides estimates of annual maize yields in metric tons (t) per hectare (ha), as well as planting and maturity dates, at a 0.5×0.5 degree resolution (about 50 km^2). For each of these models, we select model simulations considering the effect of CO_2 concentration in order to account for CO_2 fertilization effect, which plays an important role in biomass production. Also, we consider simulations assuming no irrigation in order to capture the effect of precipitation on crop yields.

GGCMs differ in their representation of crop phenology, leaf area development, yield formation, root expansion and nutrient assimilation. However, they all account for the effect of water, heat stress and CO_2 fertilization. None of the models considered assume technological change. A more detailed description of each model's processes is provided by Rosenzweig et al. (2014). Some caveats are associated with each model.[2] For instance, the LPJ-GUESS model estimates potential yields (yield non-limited by nutrient or management constraints) rather than actual yield and therefore only relative change should be considered when assessing the impact of climate change on crop yield using this model. Also, the GEPIC model accounts for soil fertility erosion, which requires the simulations to be run independently for each decade, while the pDSSAT model only updates CO_2 inputs every 30 years, which results in a periodic step in yield projections. As a result, these GGCM simulations are more suited to assess long-term trends in yields rather than inter-annual yield variability.

2.1.2. Weather

Bias-corrected weather data used as input into each crop model are obtained from the CMIP5 climate data simulations. This study uses daily weather data for three of the five climate models, or General Circulation Models (GCMs) included in CMIP5: HadGEM2-ES, NorESM1-M, and GFDL-ESM2M. As summarized in Warszawski et al. (2014), these GCMs project, respectively, high, medium and low level of global warming.

GCM simulations are available for an 'historical' period of 1975–2005 and a 'future' period of 2006 to 2099. For the 'future' period, each GCM is run under four Representative Concentration Pathways (RCPs), each representative of different level of radiative forcing (RCP 2.6, RCP 4.5, RCP 6.0 and RCP 8.5). We selected

---

[1] The data are available for download at https://www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/research/rd2-cross-cutting-activities/isi-mip/data-archive/fast-track-data-archive.

[2] These caveats are discussed at https://www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/research/rd2-cross-cutting-activities/isi-mip/data-archive/fast-track-data-archive/data-caveats.