Agricultural and Forest Meteorology 236 (2017) 145–161

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models

Élodie Blanc

Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 14 April 2016
Received in revised form
15 December 2016
Accepted 26 December 2016
Available online 26 January 2017

Keywords:
Crop yields
Crop model
Statistical model
Climate change

ABSTRACT

This study provides statistical emulators of crop yields based on global gridded crop model simulations from the Inter-Sectoral Impact Model Intercomparison Project Fast Track project. The ensemble of simulations is used to build a panel of annual crop yields from five crop models and corresponding monthly summer weather variables for over a century at the grid cell level globally. This dataset is then used to estimate, for each crop and gridded crop model, the statistical relationship between yields, temperature, precipitation and carbon dioxide. This study considers a new functional form to better capture the non-linear response of yields to weather, especially for extreme temperature and precipitation events, and now accounts for the effect of soil type. In- and out-of-sample validations show that the statistical emulators are able to replicate spatial patterns of yields crop levels and changes overtime projected by crop models reasonably well, although the accuracy of the emulators varies by model and by region. This study therefore provides a reliable and accessible alternative to global gridded crop yield models. By emulating crop yields for several models using parsimonious equations, the tools provide a computationally efficient method to account for uncertainty in climate change impact assessments.

1. Introduction

The vulnerability of crops to weather is well known and numerous studies have attempted to estimate the impact of climate change on yields (Challinor et al., 2014). These studies generally rely on either process-based crop models (e.g. Alexandrov and Hoogenboom, 2000; Butt et al., 2005; Deryng et al., 2014; Parry et al., 1999; Rosenzweig and Parry, 1994) or statistical techniques (e.g. Blanc, 2012; Blanc and Strobl, 2013; Haim et al., 2007; Lobell and Field, 2007; Schlenker and Roberts, 2009). While process-based crop models are able to capture the effect of weather and other environmental conditions on crop growth and yields at the grid cell or site level, they are computationally demanding and sometimes proprietary, which limits their accessibility. On the other hand, statistical models are more easily applicable but depend on the availability of observations to estimate the impact of average weather conditions on crop yields while controlling for other factors. To benefit from the capabilities of processed-based models while preserving the application simplicity of statistical models, Blanc and Sultan (2015) provide an ensemble of statistical tools emulating maize yields from process-based crop models at the grid cell level globally using a simple set of weather variables. They employ the 'perfect model' approach, consisting of training a statistical model on the output of a process-based crop model, based on the assumption that these output are 'true'. This method has been used by Holzkämper et al. (2012) and Lobell and Burke (2010) with the purpose of evaluating the ability of statistical models to predict crop yields out-of-sample. These studies find that statistical models are capable of replicating the out-of-sample outcomes of process-based crop models reasonably well. Oyebamiji et al. (2015) expand on these studies by estimating a crop yield emulator at the global level for five different crops but, as in previous studies, only consider one process-based crop model. As 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), Blanc and Sultan (2015) expanded the scope and applicability of statistical emulators by considering five different crop models. These emulators are based on simulations data from the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) Fast Track experiment dataset of global gridded crop models (GGCM) simulations. This project, coordinated by the Agricultural Model Intercomparison and Improvement Project (AgMIP) (Rosenzweig et al., 2013) as part of ISI-MIP (Warszawski et al., 2014), was tailored specifically to compare crop models. Therefore, all GGCMs simulations were driven by bias-corrected climate change projections

E-mail address: eblanc@mit.edu

http://dx.doi.org/10.1016/j.agrformet.2016.12.022

derived from the Coupled Model Intercomparison Project, phase 5 (CMIP5) archive (Hempel et al., 2013; Taylor et al., 2012). The statistical emulators produced by Blanc and Sultan (2015) provide an accessible tool to estimate the impact of climate change on crop yields while accounting for crop modeling uncertainty by allowing users to emulate yields projections from five different GGCMs. However, the crop yield emulators from Blanc and Sultan (2015) are only available for maize. This study proposes to expand the scope of these emulators to three additional crops: rice, soybean and wheat.

This study also improves the response functions estimated by Blanc and Sultan (2015) by estimating more precisely the response of crop yields to weather. The effect of weather on crop yields is non-linear and is therefore usually modeled in regression analyses by including a quadratic term in the specification (e.g. Blanc, 2012; Grassini et al., 2013; Schlenker and Lobell, 2010). However, the symmetrical concave relationship imposed by this functional form might be too restrictive. Blanc and Sultan (2015) find that a fifth order polynomial transformation is well suited to represent the nonlinear relationship between weather and crop yields. However, the polynomial form exhibits behaviors difficult to explain for extreme values of temperature and precipitation. As an alternative, this study applies the fractional polynomial method from Royston and Altman (1994). This approach provides the flexibility and improved fit of a non-parametric model, but with the simplicity of a parametric model.

Data and methods used to statistically estimate relationship between yields and weather variables are presented in Section 2. Results are presented and discussed in Section 3. The models are validated in Section 4. Section 5 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. Weather and CO_2

Bias-corrected weather data used as input into each crop model are obtained from the CMIP5 climate data simulations. Daily weather data generated by three CMIP5 climate models, or General Circulation Models (GCMs): HadGEM2-ES, NorESM1-M, and GFDL-ESM2 M. These GCMs are selected to be representative of respectively, high, medium and low levels of global warming (Warszawski et al., 2014).

GCM simulations are provided for the 'historical' period of 1975–2005 and the 'future' period of 2006–2099. For the 'future' period, one Representative Concentration Pathway (RCP) consistent with the highest level of global warming compared to historical conditions, RCP 8.5, and the corresponding CO_2 concentrations data (Riahi et al., 2007) is considered.[2] Combined with the large range of climate change patterns represented by the three GCMs, this study considers the broadest plausible range of future climate change.

Using daily precipitation, and minimum and maximum temperature produced by each GCM and used as inputs by GGCMs, monthly averages of precipitation (*Pr*) and temperature (*Tmean*) are calculated for each summer month.[3] For ease of reference, in this study numbers suffixes are used to represent each summer month, so _1, _2, and _3 refer to, respectively, June, July and August in the Northern Hemisphere and December, January and February in the Southern Hemisphere.

2.1.2. Crop yields

Crop yields are obtained from GGCMs members of the ISI-MIP Fast Track experiment. Due to data limitations, simulations from five crop models are selected: 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 Agrotechnology 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). For each of these GGCMs, model simulations considering the effect of CO_2 concentrations are selected in order to account for the CO_2 fertilization effect, which plays an important role on biomass production. In this study, only simulations assuming no irrigation are considered in order to capture the effect of precipitation on crop yields.

All GGCMs estimate annual crop yields in metric tons per hectare (t/ha) at a $0.5 \times 0.5^{\circ}$ resolution (about 50 km^2). And although they differ in their representation of crop phenology, leaf area development, yield formation, root expansion and nutrient assimilation, they all account for the effect of water, heat stress and CO_2 fertilization, and assume no technological change. A more detailed description of each model's processes is provided by Rosenzweig et al. (2014). As mentioned in Blanc and Sultan (2015), caveats are associated with each model leading to divergences and GGCM-specific periodic patterns of yield projections.[4]

Crop models simulate yields from 1975 to 2005 for the 'historical' period and 2006–2099 for the 'future' period. As only one RCP scenario is selected for each GCM, the panel is constructed over the consecutive period 1975–2099 without distinction (i.e. one historical scenario and one future scenario for each GCM). In the final sample, grid cells for which there are less than 10 yield observations after data cleaning are omitted.

2.1.3. Soil orders

Soil orders at the $0.5 \times 0.5^{\circ}$ resolution are extracted from the FAO-UNESCO (2005) Soil Map of the World using the USDA soil taxonomy (Soil Survey Staff, 1999) which classifies soils on the basis of soil physical and chemical properties observed in situ (e.g. soil horizons,[5] structure, texture, color) and inferred from environmental conditions (e.g., soil temperature and moisture regimes). As shown in Fig. 1, soils are grouped into 12 main soil orders: Gelisols which are permanently frozen or contain permafrost near the soil surface and are found in the Arctic, Antarctic and extremely high elevations; Histosols which are composed mainly of decomposing organic matters and forming in areas of poor drainage; Spodosols which develop under coniferous vegetation where litter contributes to acid accumulations in the soil; Andisols which

[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] The data are available at http://tntcat.iiasa.ac.at/RcpDb/dsd?Action=htmlpage&page=welcome

[3] Mean temperature is calculated as $Tmean = (Tmin + Tmax)/2$, Tmin and Tmax are, respectively, the minimum and maximum daily temperatures.

[4] 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

[5] Soil order data are available for download at https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/?cid=nrcs142p2_054013