

The GGCMI Phase II experiment: simulating and emulating global crop yield responses to changes in carbon dioxide, temperature, water, and nitrogen levels

James Franke^{a,b,*}, Joshua Elliott^{b,c}, Christoph Müller^d, Alexander Ruane^e, Abigail Snyder^f, Jonas Jägermeyr^{c,b,d,e}, Juraj Balkovic^{g,h}, Philippe Ciais^{i,j}, Marie Dury^k, Pete Falloon^l, Christian Folberth^g, Louis François^k, Tobias Hank^m, Munir Hoffmannⁿ, Cesar Izaurralde^{o,p}, Ingrid Jacquemin^k, Curtis Jones^o, Nikolay Khabarov^g, Marian Kochⁿ, Michelle Li^{b,l}, Wenfeng Liu^{r,i}, Stefan Olin^s, Meridel Phillips^{e,t}, Thomas Pugh^{u,v}, Ashwan Reddy^o, Xuhui Wang^{i,j}, Karina Williams^l, Florian Zabel^m, Elisabeth Moyer^{a,b}

^aDepartment of the Geophysical Sciences, University of Chicago, Chicago, IL, USA

^bCenter for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA

^cDepartment of Computer Science, University of Chicago, Chicago, IL, USA

^dPotsdam Institute for Climate Impact Research, Leibniz Association (Member), Potsdam, Germany

^eNASA Goddard Institute for Space Studies, New York, NY, United States

^fJoint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA

^gEcosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria

^hDepartment of Soil Science, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovak Republic

ⁱLaboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France

^jSino-French Institute of Earth System Sciences, College of Urban and Environmental Sciences, Peking University, Beijing, China

^kUnité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d'Astrophysique et de Géophysique, University of Liège, Belgium

^lMet Office Hadley Centre, Exeter, United Kingdom

^mDepartment of Geography, Ludwig-Maximilians-Universität, Munich, Germany

ⁿGeorg-August-University Göttingen, Tropical Plant Production and Agricultural Systems Modelling, Göttingen, Germany

^oDepartment of Geographical Sciences, University of Maryland, College Park, MD, USA

^pTexas AgriLife Research and Extension, Texas A&M University, Temple, TX, USA

^qDepartment of Statistics, University of Chicago, Chicago, IL, USA

^rEAWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

^sDepartment of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

^tEarth Institute Center for Climate Systems Research, Columbia University, New York, NY, USA

^uKarlsruhe Institute of Technology, IMK-IFU, 82467 Garmisch-Partenkirchen, Germany.

^vSchool of Geography, Earth and Environmental Science, University of Birmingham, Birmingham, UK.

Abstract

Concerns about food security under climate change have motivated efforts to better understand the future changes in yields by using detailed process-based models in agronomic sciences. Process-based crop models differ on many details affecting yields and considerable uncertainty remains in future yield projections. Phase II of the Global Gridded Crop Model Intercomparison (GGCMI), an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP), consists of a large simulation set with perturbations in atmospheric CO₂ concentrations, temperature, precipitation, and applied nitrogen inputs and constitutes a data-rich basis of projected yield changes across twelve models and five crops (maize, soy, rice, spring wheat, and winter wheat) using global gridded simulations. In this paper we present the simulation output database from Phase II of the GGCMI effort, a targeted experiment aimed at understanding the sensitivity to and interaction between multiple climate variables (as well as management) on yields, and illustrate some initial summary results from the model intercomparison project. We also present the construction of a simple “emulator or statistical representation of the simulated 30-year mean climatological output in each location for each crop and model. The emulator captures the mean-climatological response of the process-based models in a lightweight, computationally tractable form that facilitates model comparison as well as potential applications in subsequent modeling efforts such as integrated assessment.

Keywords: climate change, food security, model emulation, AgMIP, crop model

1. Introduction

Understanding crop yield response to a changing climate is critically important, especially as the global food production system will face pressure from increased demand over the next century. Climate-related reductions in supply could therefore have severe socioeconomic consequences. Multiple studies using different crop or climate models concur in predicting sharp yield reductions on currently cultivated cropland under business-as-usual climate scenarios, although their yield projections show considerable spread (e.g. Porter et al. (IPCC), 2014, Rosenzweig et al., 2014, Schauberger et al., 2017, and references therein). Modeling crop responses continues to be challenging, as crop growth is a function of complex interactions between climate inputs and management practices. Intercomparison projects targeting model responses to important drivers are critical to improve future projections.

Computational models have been used to project crop yields since the 1950's, beginning with statistical models that attempt to capture the relationship between input factors and resultant yields (e.g. Heady, 1957, Heady & Dillon, 1961). These statistical models were typically developed on a small scale for locations with extensive histories of yield data. The emergence of electronic computers allowed development of numerical models that simulate the process of photosynthesis and the biology and phenology of individual crops (first proposed by de Wit (1957) and Duncan et al. (1967) and attempted by Duncan (1972); for a history of crop model development see Rosenzweig et al. (2014)). A half-century of improvement in both models and computing resources means that researchers can now run crop simulations for many years at high spatial resolution on the global scale.

Both types of models continue to be used, and compara-

tive studies have concluded that when done carefully, both approaches can provide similar yield estimates (e.g. Lobell & Burke, 2010, Moore et al., 2017, Roberts et al., 2017, Zhao et al., 2017). Models tend to agree broadly in major response patterns, including a reasonable representation of the spatial pattern in historical yields of major crops (e.g. Elliott et al., 2015, Müller et al., 2017) and projections of decreases in yield under future climate scenarios.

Process-based models do continue to struggle with some important details, including reproducing historical year-to-year variability (e.g. Müller et al., 2017), reproducing historical yields when driven by reanalysis weather (e.g. Glotter et al., 2014), and low sensitivity to extreme events (e.g. Glotter et al., 2015). These issues are driven in part by the diversity of new cultivars and genetic variants, which outstrips the ability of academic modeling groups to capture them (e.g. Jones et al., 2017). Models also do not simulate many additional factors affecting production, including pests, diseases, and weeds. For these reasons, individual studies must generally re-calibrate models to ensure that short-term predictions reflect current cultivar mixes, and long-term projections retain considerable uncertainty (Wolf & Oijen, 2002, Jagtap & Jones, 2002, Iizumi et al., 2010, Angulo et al., 2013, Asseng et al., 2013, 2015). Inter-model discrepancies can also be high in areas not yet cultivated (e.g. Challinor et al., 2014, White et al., 2011). Finally, process-based models present additional difficulties for high-resolution global studies because of their complexity and computational requirements. For economic impacts assessments, it is often impossible to integrate a set of process-based crop models directly into an integrated assessment model to estimate the potential cost of climate change to the agricultural sector.

Nevertheless, process-based models are necessary for understanding the global future yield impacts of climate change for many reasons. First, cultivation may shift to new areas, where

*Corresponding author at: 4734 S Ellis, Chicago, IL 60637, United States.
email: jfranke@uchicago.edu

no yield data are currently available and therefore statistical models cannot apply. Yield data are also often limited in the developing world, where future climate impacts may be the most critical. Finally, only process-based models can capture the growth response to novel conditions and practices that are not represented in historical data (e.g. Pugh et al., 2016, Roberts et al., 2017). These novel changes can include the direct fertilization effect of elevated CO₂, and changes in management practices that may ameliorate climate-induced damages.

Interest has been rising in statistical emulation, which allows combining advantageous features of both statistical and process-based models. The approach involves constructing a statistical representation or “surrogate model” of complicated numerical simulations by using simulation output as the training data for a statistical model (e.g. O’Hagan, 2006, Conti et al., 2009). Emulation is particularly useful in cases where simulations are complex and output data volumes are large, and has been used in a variety of fields, including hydrology (e.g. Razavi et al., 2012), engineering (e.g. Storlie et al., 2009), environmental sciences (e.g. Ratto et al., 2012), and climate (e.g. Castruccio et al., 2014, Holden et al., 2014). For agricultural impacts studies, emulation of process-based models allows capturing key relationships between input variables in a lightweight, flexible form that is compatible with economic studies.

In the past decade, multiple studies have developed emulators of process-based crop simulations. Early studies proposing or describing potential crop yield emulators include Howden & Crimp (2005), Räisänen & Ruokolainen (2006), Lobell & Burke (2010), and Ferrise et al. (2011), who used a machine learning approach to predict Mediterranean wheat yields. Studies developing single-model emulators include Holzkämper et al. (2012) for the CropSyst model, Ruane et al. (2013) for the CERES wheat model, and Oyebamiji et al. (2015) for the

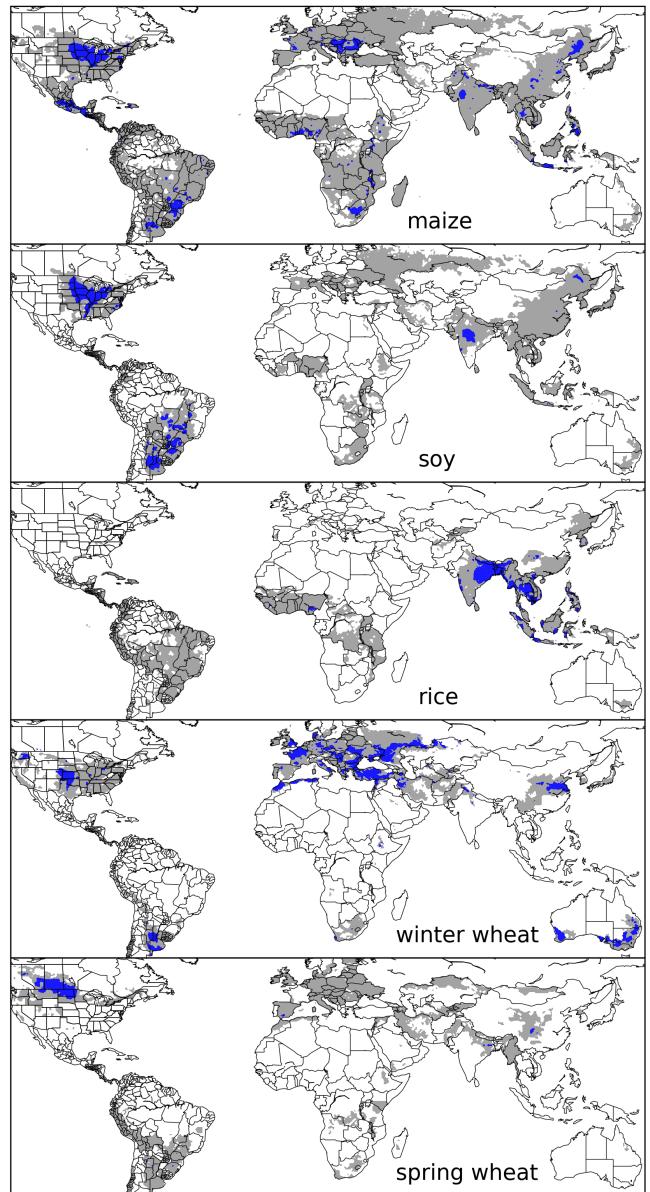


Figure 1: Presently cultivated area for rain-fed crops. Blue indicates grid cells with more than 20,000 hectares (~10% of the equatorial grid cells) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated areas for maize, rice, and soy are taken from the MIRCA2000 (“monthly irrigated and rain-fed crop areas around the year 2000”) dataset (Portmann et al., 2010). Areas for winter and spring wheat areas are adapted from MIRCA2000 data and sorted by growing season. For analogous figure of irrigated crops, see Figure S1.

LPJmL model (for multiple crops, using multiple scenarios as a training set). More recently, emulators have begun to be used in the context of multi-model intercomparisons, with Blanc & Sultan (2015), Blanc (2017), Ostberg et al. (2018) and Misra et al. (2017) using them to analyze the five crop models of the Inter-Sectoral Impacts Model Intercomparison Project

(ISIMIP) (Warszawski et al., 2014), which simulated yields for maize, soy, wheat, and rice. Choices differ: Blanc & Sultan (2015) and Blanc (2017) base their emulation on historical simulations and a single future climate/emissions scenario (RCP8.5), and use local weather variables and yields in their regression but then aggregate across broad regions; Ostberg et al. (2018) consider multiple future climate scenarios, using global mean temperature change (and CO₂) as regressors but then pattern-scale to emulate local yields; while Mistry et al. (2017) compare emulated and observed historical yields, using local weather data and a historical crop simulation. These efforts do share important common features: all emulate annual crop yields across the entire scenario or scenarios, and when future scenarios are considered, they are non-stationary, i.e. their input climate parameters evolve over time.

An alternative approach is to construct a training set of multiple stationary scenarios in which parameters are systematically varied. Such a “parameter sweep” offers several advantages for emulation over scenarios in which climate evolves over time. First, it allows separating the effects of different variables that impact yields but that are highly correlated in realistic future scenarios (e.g. CO₂ and temperature). Second, it allows making a distinction between year-over-year yield variations and climatological changes, which may involve different responses to the particular climate regressors used (e.g. Ruane et al., 2016). For example, if year-over-year yield variations are driven predominantly by variations in the distribution of temperatures throughout the growing season, and long-term climate changes are driven predominantly by shifts in means, then regressing on the mean growing season temperature will produce different yield responses at annual vs. climatological timescales.

Systematic parameter sweeps have begun to be used in crop model evaluation and emulation, with early efforts in 2015 (Makowski et al., 2015, Pirttioja et al., 2015), and several re-

cent studies in 2018 (Fronzek et al., 2018, Snyder et al., 2018, Ruiz-Ramos et al., 2018). All three studies sample multiple perturbations to temperature and precipitation (with Snyder et al. (2018) and Ruiz-Ramos et al. (2018) adding CO₂ as well), in 132, 99 and 220 different combinations, respectively, and take advantage of the structured training set to construct emulators (“response surfaces”) of climatological mean yields, omitting year-over-year variations. All are limited in some respects and focus on a limited number of sites. Fronzek et al. (2018) and Ruiz-Ramos et al. (2018) simulate only wheat (over many models) and Snyder et al. (2018) analyzes four crops (maize, wheat, rice, soy) for agricultural impacts experiments with the GCAM (Calvin et al., 2019) model.

In this paper we describe a new comprehensive dataset designed to expand the parameter sweep approach still further. The Global Gridded Crop Model Intercomparison (GGCMI) Phase II experiment involves running a suite of process-based crop models across historical conditions perturbed by a set of discrete steps in different input parameters, including an applied nitrogen dimension. The experimental protocol involves over 700 different parameter combinations for each model and crop, with simulations providing near-global coverage at a half degree spatial resolution. The experiment was conducted as part of the Agricultural Model Intercomparison and Improvement Project (AgMIP) (Rosenzweig et al., 2013, 2014), an international effort conducted under a framework similar to the Climate Model Intercomparison Project (CMIP) (Taylor et al., 2012, Eyring et al., 2016). The GGCMI protocol builds on the AgMIP Coordinated Climate-Crop Modeling Project (C3MP) (Ruane et al., 2014, McDermid et al., 2015) and will contribute to the AgMIP Coordinated Global and Regional Assessments (CGRA) (Ruane et al., 2018, Rosenzweig et al., 2018). GGCMI Phase II is designed to allow addressing goals such as understanding where highest-yield regions may shift

under climate change; exploring future adaptive management
 176 strategies; understanding how interacting input drivers affect
 crop yield; quantifying uncertainties across models and major
 178 drivers; and testing strategies for producing lightweight em-
 ulators of process-based models. In this paper, we describe
 180 the GGCMI Phase II experiments, present initial results, and
 demonstrate that it is tractable to emulation.

182 2. Simulation – Methods

GGCMI Phase II is the continuation of a multi-model com-
 184 parison exercise begun in 2014. The initial Phase I compared
 harmonized yields of 21 models for 19 crops over a 31-year
 186 historical (1980-2010) scenario with a primary goal of model
 evaluation (Elliott et al., 2015, Müller et al., 2017). Phase II
 188 compares simulations of 12 models for 5 crops (maize, rice,
 soybean, spring wheat, and winter wheat) over the same histor-
 190 ical time series (1980-2010) used in Phase I, but with individ-
 ual climate or management inputs adjusted from their historical
 192 values. The reduced set of crops includes the three major global
 cereals and the major legume and accounts for over 50% of hu-
 194 man calories (in 2016, nearly 3.5 billion tons or 32% of total
 global crop production by weight (Food and Agriculture Orga-
 196 nization of the United Nations, 2018).

The guiding scientific rationale of GGCMI Phase II is to pro-
 198 vide a comprehensive, systematic evaluation of the response
 of process-based crop models to different values for carbon
 200 dioxide, temperature, water, and applied nitrogen (collectively

known as “CTWN”). The dataset is designed to allow re-
 202 searchers to:

- Enhance understanding of how models work by character-
 izing their sensitivity to input climate and nitrogen drivers.
 204
- Study the interactions between climate variables and nitro-
 gen inputs in driving modeled yield impacts.
 206
- Explore differences in crop response to warming across the
 Earth’s climate regions.
 208
- Provide a dataset that allows statistical emulation of crop
 model responses for downstream modelers.
 210
- Illustrate differences in potential adaptation via growing
 season changes.
 212

The experimental protocol consists of 9 levels for precipita-
 214 tion perturbations, 7 for temperature, 4 for CO₂, and 3 for ap-
 plied nitrogen, for a total of 672 simulations for rain-fed agri-
 culture and an additional 84 for irrigated (Table 1). For irri-
 216 gated simulations, soil water is held at either field capacity or,
 for those models that include water-log damage, at maximum
 beneficial level. Temperature perturbations are applied as ab-
 solute offsets from the daily mean, minimum, and maximum
 218 temperature time series for each grid cell used as inputs. Pre-
 cipitation perturbations are applied as fractional changes at the
 grid cell level, and carbon dioxide and nitrogen levels are spec-
 ified as discrete values applied uniformly over all grid cells.
 Note that CO₂ changes are applied independently of changes
 220 in climate variables, so that higher CO₂ is not associated with
 222

Input variable	Abbr.	Tested range	Unit
CO ₂	C	360, 510, 660, 810	ppm
Temperature	T	-1, 0, 1, 2, 3, 4, 5*, 6	°C
Precipitation	W	-50, -30, -20, -10, 0, 10, 20, 30, (and W _{inf})	%
Applied nitrogen	N	10, 60, 200	kg ha ⁻¹

Table 1: GGCMI Phase II input variable test levels. Temperature and precipitation values indicate the perturbations from the historical, climatology and are selected to represent reasonable ranges for potential climate changes in the medium term. * Only simulated by one model. W-percentage does not apply to the irrigated (W_{inf}) simulations, which are all simulated at the maximum beneficial levels of water.

Model (Key Citations)	Maize	Soy	Rice	Winter Wheat	Spring Wheat	N Dim.	Simulations per Crop
APSIM-UGOE , Keating et al. (2003), Holzworth et al. (2014)	X	X	X	–	X	Yes	37
CARAIB , Dury et al. (2011), Pirttioja et al. (2015)	X	X	X	X	X	No	224
EPIC-IIASA , Balkovi et al. (2014)	X	X	X	X	X	Yes	35
EPIC-TAMU , Izaurrealde et al. (2006)	X	X	X	X	X	Yes	672
JULES* , Osborne et al. (2015), Williams & Falloon (2015), Williams et al. (2017)	X	X	X	–	X	No	224
GEPIC , Liu et al. (2007), Folberth et al. (2012)	X	X	X	X	X	Yes	384
LPJ-GUESS , Lindeskog et al. (2013), Olin et al. (2015)	X	–	–	X	X	Yes	672
LPJmL , von Bloh et al. (2018)	X	X	X	X	X	Yes	672
ORCHIDEE-crop , Valade et al. (2014)	X	–	X	–	X	Yes	33
pDSSAT , Elliott et al. (2014), Jones et al. (2003)	X	X	X	X	X	Yes	672
PEPIC , Liu et al. (2016a,b)	X	X	X	X	X	Yes	130
PROMET*† , Mauser & Bach (2015), Hank et al. (2015), Mauser et al. (2009)	X	X	X	X	X	Yes†	239
Totals	12	10	11	9	12	–	3993 (maize)

Table 2: Models included in GGCMI Phase II and the number of C, T, W, and N simulations that each performs for rain-fed crops (“Sims per Crop”), with 672 as the maximum. “N-Dim.” indicates whether the simulations include varying nitrogen levels. Two models provide only one nitrogen level, and †PROMET provides only two of the three nitrogen levels (and so is not emulated across the nitrogen dimension). All models provide the same set of simulations across all modeled crops, but some omit individual crops. (For example, APSIM does not simulate winter wheat.) Irrigated simulations are provided at the level of the other covariates for each model, i.e. an additional 84 simulations for fully-sampled models. Simulations are nearly global but their geographic extent can vary, even for different crops in an individual model, since some simulations omit regions far outside the currently cultivated area. In most cases, historical daily climate inputs are taken from the 0.5 degree NASA AgMERRA daily gridded re-analysis product specifically designed for agricultural modeling, with satellite-corrected precipitation (Ruane et al., 2015), but two models (marked with *) require sub-daily input data and use alternative sources. See Elliott et al. (2015) for additional details.

higher temperatures. An additional, identical set of scenarios (at the same C, T, W, and N levels) not shown or analyzed here simulate adaptive agronomy under climate change by varying the growing season for crop production. The resulting GGCMI Phase II dataset captures a distribution of crop responses over the potential space of future climate conditions.

The 12 models included in GGCMI Phase II are all mechanistic process-based crop models that are widely used in impacts assessments (Table 2). Although some models share a common base (e.g. the LPJ family or the EPIC family of models), they have subsequently developed independently. (For more details on model genealogy, see Figure S1 in Rosenzweig et al. (2014).) Differences in model structure mean that several key factors are not standardized across the experiment, including secondary soil nutrients, carry-over effects across growing years including residue management and soil moisture, and the extent of simulated area for different crops. Growing seasons are standardized across models (with assumptions based

on Sacks et al. (2010) and Portmann et al. (2008, 2010)), but vary by crop and by location on the globe. For example, maize is sown in March in Spain, in July in Indonesia, and in December in Namibia. All stresses are disabled other than factors related to nitrogen, temperature, and water (e.g. alkalinity and salinity). No additional nitrogen inputs, such as atmospheric deposition, are considered, but some model treatments of soil organic matter may allow additional nitrogen release through mineralization. See Rosenzweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017) for further details on models and underlying assumptions.

The participating modeling groups provide simulations at any of four initially specified levels of participation, so the number of simulations varies by model, with some sampling only a part of the experiment variable space. Most modeling groups simulate all five crops in the protocol, but some omitted one or more. Table 2 provides details of coverage for each model. Note that the three models that provide less than 50 simulations

are excluded from the emulator analysis.

264 Each model is run at 0.5 degree spatial resolution and cov-
 265 ers all currently cultivated areas and much of the uncultivated
 266 land area. (See Figure 1 for the present-day cultivated area of
 267 rain-fed crops, and Figure S1 in the Supplemental Material for
 268 irrigated crops.) Coverage extends considerably outside cur-
 269 rently cultivated areas because cultivation will likely shift under
 270 climate change. However, areas are not simulated if they are
 271 assumed to remain non-arable even under an extreme climate
 272 change; these regions include Greenland, far-northern Canada,
 273 Siberia, Antarctica, the Gobi and Sahara Deserts, and central
 274 Australia.

275 All models produce as output crop yields (tons ha^{-1} year $^{-1}$)
 276 for each 0.5 degree grid cell. Because both yields and yield
 277 changes vary substantially across models and across grid cells,
 278 we primarily analyze relative change from a baseline. We take

280 as the baseline the scenario with historical climatology (i.e. T
 281 and P changes of 0), C of 360 ppm, and applied N at 200 kg
 282 ha^{-1} . We show absolute yields in some cases to illustrate geo-
 283 graphic differences in yields.

284 The GGCMI Phase II simulations are designed for evaluat-
 285 ing changes in yield but not absolute yields, since they omit
 286 detailed calibrations. To provide some validation of the skill of
 287 the process-based models used, we repeat the validation exer-
 288 cises of Müller et al. (2017) for GGCMI Phase I. See Appendix
 289 A for details on simulation model validation.

3. Simulation – Results

290 Crop models in the GGCMI Phase II ensemble show broadly
 291 consistent responses to climate and management perturbations
 292 in most regions, with a strong negative impact of increased tem-
 293 perature in all but the coldest regions. We illustrate this re-

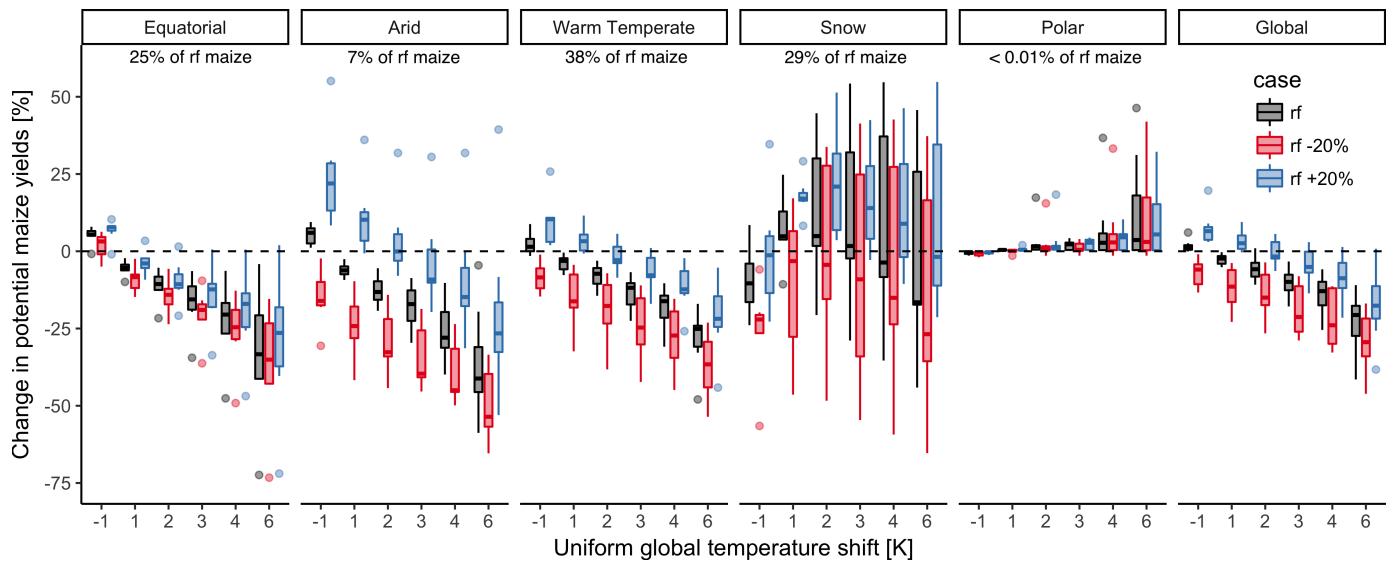


Figure 2: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek, 2010). Note that ‘Equatorial’ and ‘Snow’ regions are sometimes referred to as ‘tropical’ and ‘cold-continental’ respectively. We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation levels (rain-fed (rf) -20%, rain-fed (0), and rain-fed +20%), with CO₂ and nitrogen held constant at baseline values (360 ppm and 200 kg ha^{-1} yr^{-1}). Y-axis is fractional change in the regional average climatological potential yield relative to the baseline. The figure shows all modeled land area; see Figure S6 in the supplemental material for only currently-cultivated land. Panel text gives the percentage of rain-fed maize presently cultivated in each climate zone (data from Portmann et al., 2010). Box-and-whiskers plots show distribution across models, with median marked. Box edges are first and third quartiles, i.e. box height is the interquartile range (IQR). Whiskers extend to maximum and minimum of simulations but are limited at 1.5·IQR; otherwise the outlier is shown. Models generally agree in most climate regions (other than cold continental), with projected changes larger than inter-model variance. Outliers in the tropics (strong negative impact of temperature increases) are the pDSSAT model; outliers in the high-rainfall case (strong positive impact of precipitation increases) are the JULES model. Inter-model variance increases in the case where precipitation is reduced, suggesting uncertainty in model response to water limitation. The right panel (Global) shows yield responses to an globally uniform temperature shift; note that these results are not directly comparable to simulations of more realistic climate scenarios with the same global mean change.

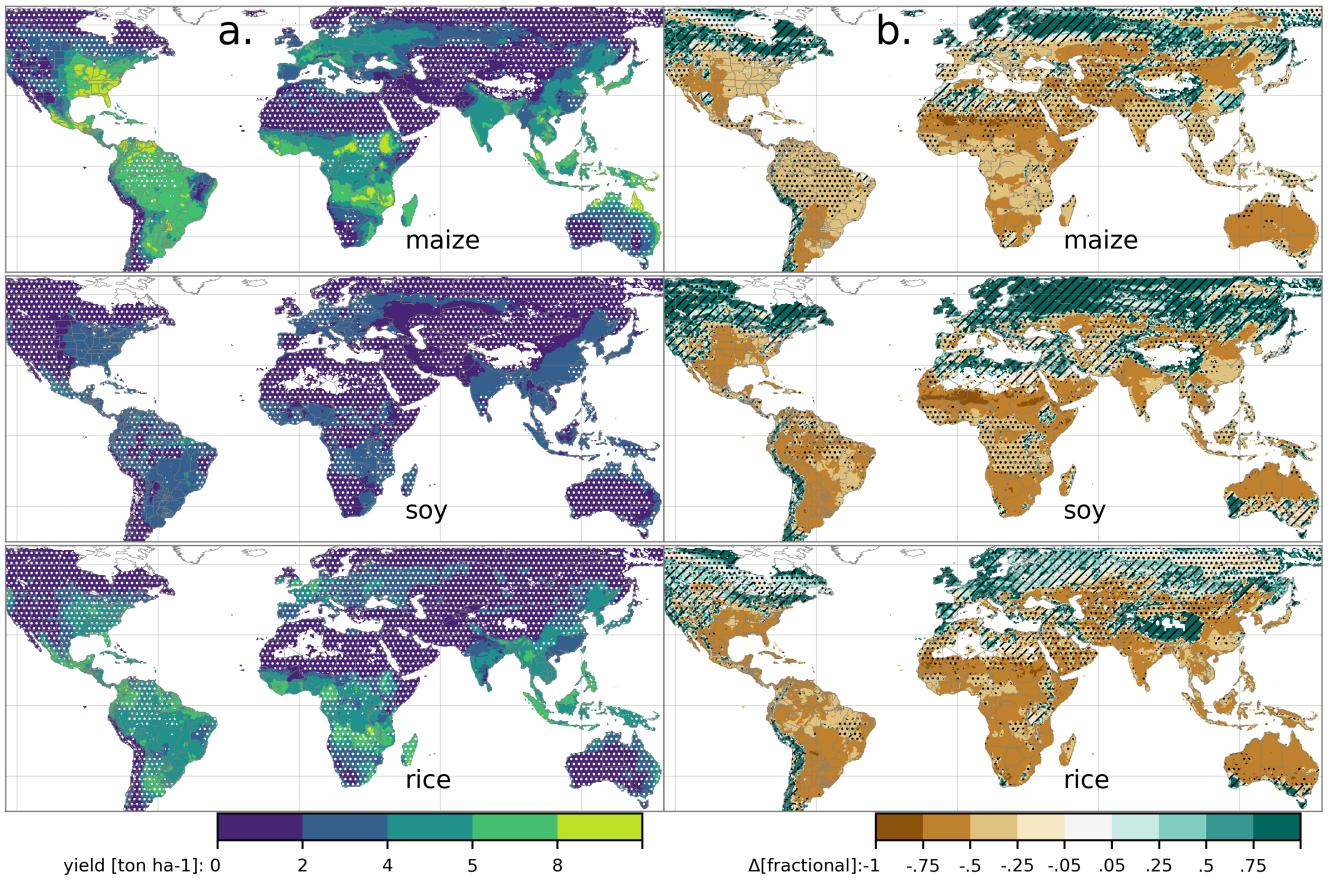


Figure 3: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (Wheat shows a qualitatively similar response, see Figure S11 in the supplemental material.) White stippling indicates areas where these crops are not currently cultivated. Absence of cultivation aligns well with the lowest yield contour ($0\text{--}2 \text{ ton ha}^{-1}$). Right column (b) shows the multi-model mean fractional yield change in the extreme $T + 4 \text{ }^{\circ}\text{C}$ scenario (with other inputs at baseline values). Areas without hatching or stippling are those where confidence in projections is high: the multi-model mean fractional change exceeds two standard deviations of the ensemble. ($\Delta > 2\sigma$). Hatching indicates areas of low confidence ($\Delta < 1\sigma$), and stippling areas of medium confidence ($1\sigma < \Delta < 2\sigma$). Crop model results in cold areas, where yield impacts are on average positive, also have the highest uncertainty.

sult for rain-fed maize in Figure 2, which shows yields across all grid cells for the primary Köppen-Geiger climate regions (Rubel & Kottek, 2010). In warming scenarios with precipitation held constant, all models show decreases in maize yield in the ‘warm temperate’, ‘equatorial’(tropical), and arid regions that account for nearly three-quarters of global maize production. These impacts are robust for even moderate climate perturbations. In the ‘warm temperate’ zone, even a 1 degree temperature rise with other variables held fixed leads to a median yield reduction that outweighs the variance across models. A 6 degree temperature rise results in median loss of $\sim 25\%$ of yields with a signal to noise ratio of nearly three to one. A notable exception is the ‘snow’ (‘cold-continental’) region, where

models disagree strongly, extending even to the sign of impacts. Other crops show similar responses to warming, with robust yield losses in warmer locations and high inter-model variance in the cold continental regions (Figure S7).

The effects of rainfall changes on maize yields shown in Figure 2 are also as expected and are consistent across models. Increased rainfall mitigates the negative effect of higher temperatures by counteracting the increased evapo-transpiration to some degree, most strongly in arid regions. Decreased rainfall amplifies yield losses and also increases inter-model variance more strongly, suggesting that models have difficulty representing crop response to water stress or increased evapo-transpiration due to warmer temperatures. We show only rain-

fed maize here; see Figure S5 for the irrigated case. As expected, irrigated crops are more resilient to temperature increases in all regions, especially so where water is limiting.

Mapping the distribution of baseline yields and yield changes shows the geographic dependencies that underlie these results. Figure 3 shows baseline and changes in the T+4 scenario for rain-fed maize, soy, and rice in the multi-model ensemble mean, with locations of model agreement marked. Absolute yield potentials show strong spatial variation, with much of the Earth's surface area unsuitable for any of these crops. In general, models agree most on yield response in regions where yield potentials are currently high and therefore where crops are currently grown. Models show robust decreases in yields at low latitudes, and highly uncertain median increases at most high latitudes. For wheat crops see Figure S11; wheat projections are more uncertain, possible because calibration is especially important for wheat (e.g. Asseng et al., 2013).

4. Emulation – Methods

As part of our demonstration of the properties of the GGCMI Phase II dataset, we construct an emulator of 30-year climatological mean yields. This approach is made possible by the structured set of simulations involving systematic perturbations. In the GGCMI Phase II dataset, the year-over-year responses are generally quantitatively distinct from (and larger than) climatological mean responses. In the example of Figure 4, responses to year-over-year temperature variations are 100% larger than those to long-term perturbations in the baseline case, and larger still under warmer conditions, rising to nearly 200% more in the T+6 case. The stronger year-over-year response under warmer conditions also manifests as a wider distribution of yields (Figure 5). As discussed previously, year-over-year and climatological responses can differ for many reasons including memory in the crop model, lurking covariants, and

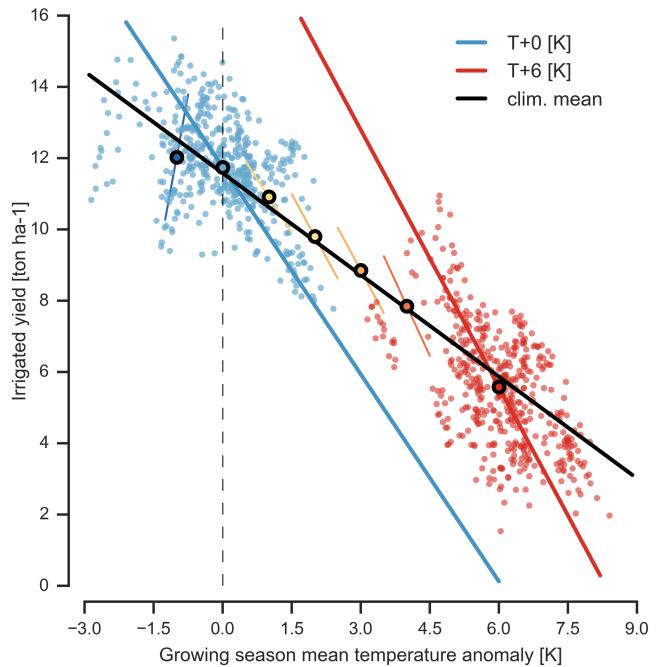


Figure 4: Example showing distinction between crop yield responses to year-to-year and climatological mean temperature shifts. Figure shows irrigated maize for a representative high-yield region (nine adjacent grid cells in northern Iowa) from the pDSSAT model, for the baseline 1981-2010 historical climate (blue) and for the scenario of maximum temperature change (+6 K, red). Other variables are held at baseline values, and the choice of irrigated yields means that precipitation is not a factor. Open black circles mark climatological mean yield values for all six temperature scenarios (T-1, +0, +1, +2, +3, +4, +6). Colored lines show total least squares linear regressions of year-over-year variations in each scenario. Black line shows the fit through the climatological mean values. Responses to year-over-year temperature variations (colored lines) are 100–200% larger than those to long-term climate perturbations, rising under warmer conditions.

differing associated distributions of daily growing-season daily weather (e.g. Ruane et al., 2016). Note that the GGCMI Phase II datasets do not capture one climatological factor, potential future distributional shifts, because all simulations are run with fixed offsets from the historical climatology. Prior work has suggested that mean changes are the dominant drivers of climatological crop yield shifts in non-arid regions (e.g. Glotter et al., 2014).

Emulation involves fitting individual regression models for each crop, simulation model, and 0.5 degree geographic pixel from the GGCMI Phase II dataset; the regressors are the applied constant perturbations in CO₂, temperature, water, and nitrogen (C, T, W, N). We regress 30-year climatological mean yields against a third-order polynomial in C, T, W, and N with interac-

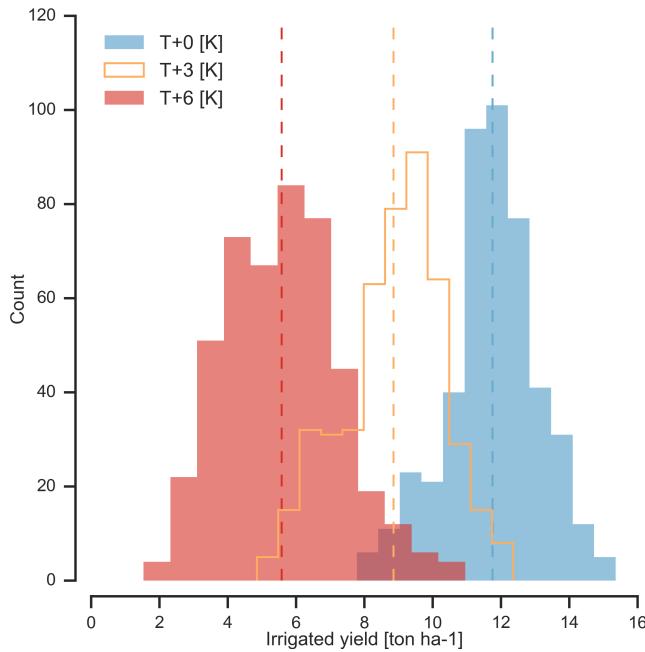


Figure 5: Example showing climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows irrigated maize for nine adjacent high-yield grid cells of Figure 4 from the pDSSAT model, for the baseline 1981-2010 historical climate (blue) and for scenarios with temperature shifted by T+3 (orange) and T+6 K (red), with other variables held at baseline values. The stronger year-over-year temperature response with higher temperatures seen in Figure 4 is manifested here as larger variance in annual yields even though the variance in climate drivers is identical. In this work we emulate not the year-over-year distributions but the climatological mean response (dashed vertical lines).

tion terms. (We aggregate the entire 30-year run in each case to improve signal to noise ration in our model.) The higher-order terms are necessary to capture any nonlinear responses, which are well-documented in observations for temperature and water perturbations (e.g. Schlenker & Roberts (2009) for T and He et al. (2016) for W). We include interaction terms (both linear and higher-order) because past studies have shown them to be significant effects. For example, Lobell & Field (2007) and Tebaldi & Lobell (2008) showed that in real-world yields, the joint distribution in T and W is needed to explain observed yield variance. (C and N are fixed in these data.) Other observation-based studies have shown the importance of the interaction between water and nitrogen (e.g. Aulakh & Malhi, 2005), and between nitrogen and carbon dioxide (Osaki et al., 1992, Nakamura et al., 1997). To avoid over-fitting or unstable parameter estimation, we apply a feature selection procedure (described

below) that reduces the potential 34-term polynomial (for the rain-fed case) to 23 terms.

We do not focus on comparing different functional forms in this study, and instead choose a relatively simple parametrization that allows for some interpretation of coefficients. Some prior studies have used more complex functional forms and larger numbers of parameters, e.g. 39 in Blanc & Sultan (2015) and Blanc (2017), who borrow information across space by fitting grid points simultaneously across a large region in a panel regression. The simple functional form used here allows emulation at the grid cell level. The emulation therefore indirectly includes any yield response to geographically distributed factors such as soil type, insolation, and the baseline climate itself. We hold the statistical specification constant across all crops and models to facilitate parameter by parameter simulation model comparison.

4.1. Feature selection procedure

Although the GGCMI Phase II sampled variable space is large, it is still sufficiently limited that use of the full polynomial expression described above can be problematic. We therefore reduce the number of terms through a feature selection cross-validation process in which terms in the polynomial are tested for importance. In this procedure higher-order and interaction terms are added successively to the regression model; we then follow the reduction of the the aggregate mean squared error with increasing terms and eliminate those terms that do not contribute significant reductions. See supplemental documents for more details. We select terms by applying the feature selection process to the three models that provided the complete set of 672 rain-fed simulations (pDSSAT, EPIC-TAMU, and LPJmL); the resulting choice of terms is then applied for all emulators.

Feature importance is remarkably consistent across all three models and across all crops (see Figure S4 in the supplemental material). The feature selection process results in a final poly-

nomial in 23 terms, with 11 terms eliminated. We omit the N³ term, which cannot be fitted because we sample only three nitrogen levels. We eliminate many of the C terms: the cubic, the CT, CTN, and CWN interaction terms, and all higher order interaction terms in C. Finally, we eliminate two 2nd-order interaction terms in T and one in W. Implication of this choice include that nitrogen interactions are complex and important, and that water interaction effects are more nonlinear than those in temperature. The resulting statistical model (Equation 1) is used for all grid cells, models, and rain-fed crops. (The regressions for irrigated crops do not contain the W terms and the models that do not sample the nitrogen levels omit the N terms).

$$\begin{aligned} Y &= K_1 \\ &+ K_2C + K_3T + K_4W + K_5N \\ &+ K_6C^2 + K_7T^2 + K_8W^2 + K_9N^2 \\ &+ K_{10}CW + K_{11}CN + K_{12}TW + K_{13}TN + K_{14}WN \\ &+ K_{15}T^3 + K_{16}W^3 + K_{17}TWN \\ &+ K_{18}T^2W + K_{19}W^2T + K_{20}W^2N \\ &+ K_{21}N^2C + K_{22}N^2T + K_{23}N^2W \end{aligned} \quad (1)$$

To fit the parameters K , we use a Bayesian Ridge probabilistic estimator (MacKay, 1991), which reduces volatility in parameter estimates when the sampling is sparse, by weighting parameter estimates towards zero. The Bayesian Ridge method is necessary to maintain a consistent functional form across all models and locations. We use the implementation of the Bayesian Ridge estimator from the scikit-learn package in Python (Pedregosa et al., 2011). In the GGCMI Phase II experiment, the most problematic fits are those for models that provided a limited number of cases or for low-yield geographic regions where some modeling groups did not run all scenarios. We do not attempt to emulate models that provided less than 50 simulations. The lowest number of simulations emu-

lated across the full parameter space is then 130 (for the PEPIC model). The resulting parameter matrices for all crop model emulators are available on request [give location?](#), as are the raw simulation data and a Python application to emulate yields. The yield output for a single GGCMI Phase II model that simulates all scenarios and all five crops is ~12.5 GB; the emulator is ~100 MB, a reduction by over two orders of magnitude.

5. Emulation – Results

Emulation provides not only a computational tool but a means of understanding and interpreting crop yield response across the parameter space. Emulation is only possible when crop yield responses are sufficiently smooth and continuous to allow fitting with a relatively simple functional form, but this condition largely holds in the GGCMI Phase II simulations. Responses are quite diverse across locations, crops, and models, but in most cases local responses are regular enough to permit emulation. We show illustrations of emulation fidelity in this section; for more detailed discussion see Appendix B.

Crop yield responses are geographically diverse, even in high-yield and high-cultivation areas. Figure 6 illustrates geographic diversity for a single crop and model (rain-fed maize in pDSSAT); this heterogeneity supports the choice of emulating at the grid cell level. Each panel in Figure 6 shows simulated yield output from scenarios varying only along a single dimension (CO₂, temperature, precipitation, or nitrogen addition), with other inputs held fixed at baseline levels, compared to the full 4D emulation across the parameter space. Yields evolve smoothly across the space sampled, and the polynomial fit captures the climatological response to perturbations. Crop yield responses generally follow similar functional forms across models, though with a large spread in magnitude likely due to the lack of calibration. Figure 7 illustrates inter-model diversity for a single crop and location (rain-fed maize in northern Iowa,

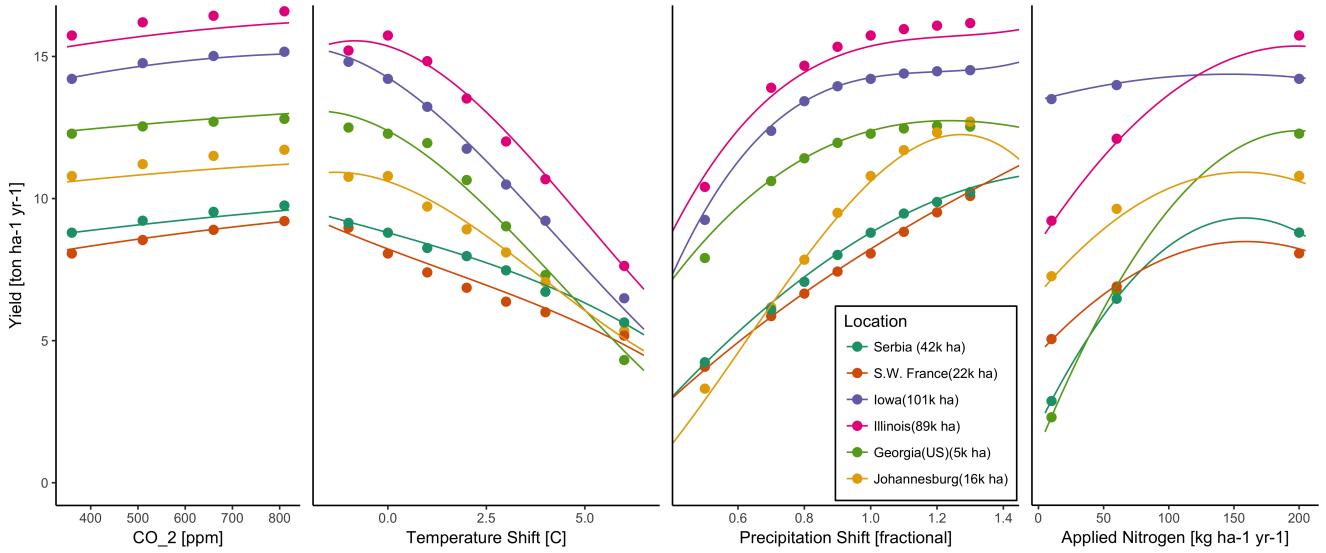


Figure 6: Illustration of spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation along a single variable, with others held at baseline values. Dots show climatological mean yields and lines the results of the full 4D emulator of Equation 1. In general the climatological response surface is sufficiently smooth that it can be represented with the sampled variable space by the simple polynomial used in this work. Extrapolation can however produce misleading results. Nitrogen fits may not be realistic at intermediate values given limited sampling. For more detailed emulator assessment, see Appendix B.

also shown in Figure 6). Differences in response shape can lead to differences in the fidelity of emulation, though comparison here is complicated by the different sampling regimes across models. Note that models are most similar in their responses to temperature perturbations.

While the nitrogen dimension is important, it is also the most problematic to emulate in this work because of its limited sampling. The GGCMI Phase II protocol specified only three nitrogen levels (10, 60 and 200 kg N y⁻¹ ha⁻¹), so a third-order fit would be over-determined but a second-order fit can result

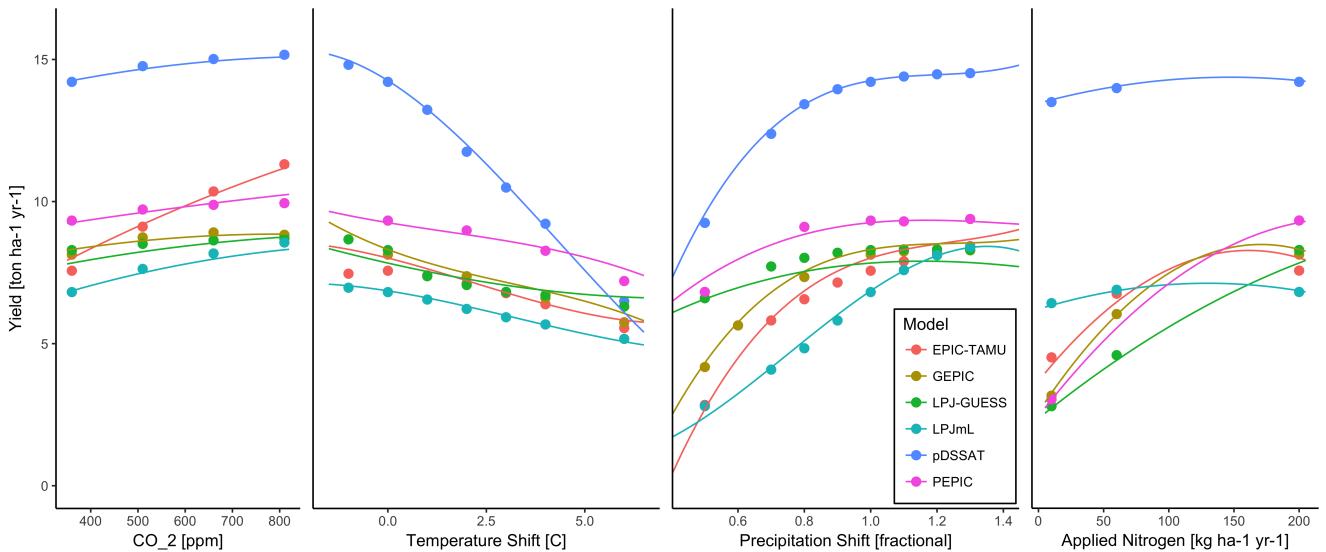


Figure 7: Illustration of across-model variations in yield response. Figures shows simulations and emulations from six models for rain-fed maize in the same Iowa grid cell shown in Figure 6, with the same plot conventions. Models that do not simulate the nitrogen dimension are omitted for clarity. Note that models are uncalibrated, increasing spread in absolute yields. While most model responses can readily emulated with a simple polynomial, some response surfaces are more complicated (e.g. LPJ-GUESS here) and lead to emulation error, though error generally remains small relative to inter-model uncertainty. For more detailed emulator assessment, see Appendix A. As in Figure 6, extrapolation out of the sample space is problematic.

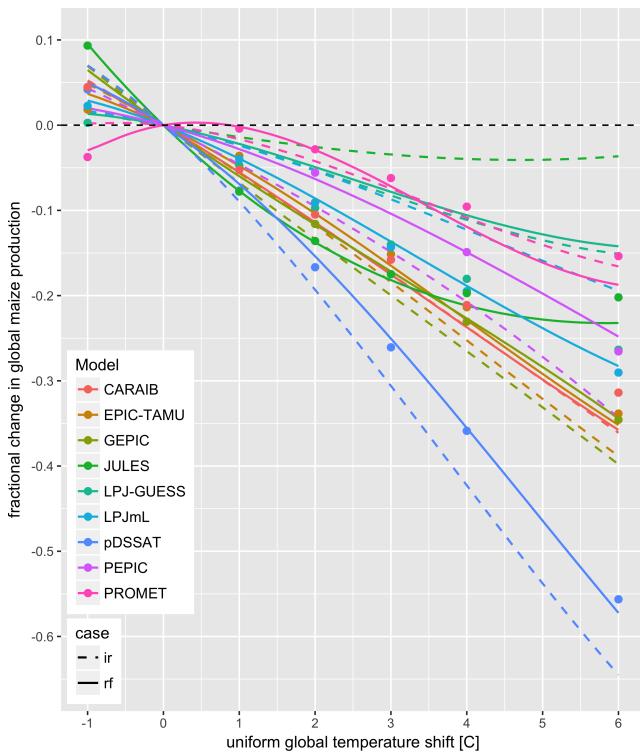


Figure 8: Global emulated damages for maize on currently cultivated lands for the GGCMI Phase II models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating up emulated values at the grid cell level, not from a regression of global mean yields.) Lines are emulations for rain-fed (solid) and irrigated (dashed) crops; for comparison, dots are the simulated values for the rain-fed case. For most models, irrigated crops show a sharper reduction than do rain-fed because of the locations of cultivated areas: irrigated crops tend to be grown in warmer areas where impacts are more severe for a given temperature shift. (The exceptions are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures S16- S19 in the supplemental material.

the emulator or “surrogate model” transforms the discrete simulation sample space into a continuous response surface at any geographic scale, it can be used for a variety of applications, including construction of continuous damage functions. As an example, we show a damage function constructed from the 4D emulation, aggregated to global yield, with simulated values shown for comparison (Figure 8, which shows maize on currently cultivated land; see Figures S16- S19 for other crops and dimensions). The emulated values closely match simulations even at this aggregation level. Note that these functions are presented only as examples and do not represent true global projections, because they are developed from simulation data with a uniform temperature shift while increases in global mean temperature should manifest non-uniformly. The global coverage of the GGCMI Phase II simulations allows impacts modelers to apply arbitrary geographically-varying climate projections, as well as arbitrary aggregation masks, to develop damage functions for any climate scenario and any geopolitical or geographic level.

6. Conclusions and Discussion

The GGCMI Phase II experiment provides a database targeted to allow detailed study of crop yields from process-based models under climate change. the systematic input parameter variations are designed to facilitate not only comparing the sensitivities of process-based crop yield models to changing climate and management inputs but also evaluating the complex interactions between driving factors (CO_2 , temperature, precipitation, and applied nitrogen). Its global nature also allows identifying geographic shifts in high yield potential locations. We expect that the simulations will yield multiple insights in future studies, and show here a selection of preliminary results to illustrate their potential uses.

First, the GGCMI Phase II simulations allow identifying ma-

in potentially unphysical results. Steep and nonlinear declines in yield with lower nitrogen levels mean that some regressions imply a peak in yield between the 100 and 200 $\text{kg N y}^{-1} \text{ha}^{-1}$ levels. While it is possible that over-application of nitrogen at the wrong time in the growing season could lead to reduced yields, these features are potentially an artifact of under sampling. In addition, the polynomial fit cannot capture the well-documented saturation effect of nitrogen application (e.g. Ingestad, 1977) as accurately as would be possible with a non-parametric model.

The emulation fidelity demonstrated here is sufficient to allow using emulated response surfaces to compare model responses and derive insight about impacts projections. Because

516 **jor areas of uncertainty. Inter-model uncertainty is qualitatively**
518 **similar across all four inputs tested at the globally aggregate**
520 **level with some notable exceptions (see Figure S20). For exam-**
522 **ple, uncertainty in the response to CO₂ and water dominates for**
wheat and soy is insensitive to nitrogen. Across geographic re-
gions, projections are most uncertain in the high latitudes where
yields may increase, and most robust in low latitudes where
yield impacts are largest.

524 Second, the GGCMI Phase II simulations allow understand-
526 ing the way that climate-driven changes and locations of cul-
528 tivated land combine to produce yield impacts. One coun-
530 terintuitive result immediate apparent is that irrigated maize
532 shows steeper yield reductions under warming than does rain-
534 fed maize when considered only over currently cultivated land
536 (Figure 8). The effect results from geographic differences in
538 cultivation. In any given location, irrigation (or additional rain-
540 fall increases) crop resiliency to temperature increase (partly by
542 reducing negative effects from increased evapo-transpiration),
but irrigated maize is grown in warmer locations where the im-
544 pacts of warming are more severe (Figures S5–S6). The same
behavior holds for rice and winter wheat, but not for soy or
546 spring wheat (Figures S8–S10). Irrigated wheat and maize are
548 also more sensitive to nitrogen fertilization levels than are anal-
ogous non-irrigated crops, presumably because those rain-fed
crops are limited by water as well as nitrogen availability (Fig-
ure S19). (Soy as an efficient atmospheric nitrogen-fixing is rel-
550 atively insensitive to nitrogen, and rice is not generally grown
in water-limited conditions).

554 Third, we show that even the relatively limited GGCMI
556 Phase II sampling space allows emulation of the climatological
558 response of crop models with a relatively simple reduced-form
statistical model. The systematic parameter sampling in the
560 GGCMI Phase II procedure provides information on the influ-
562 ence of multiple interacting factors in a way that single projec-
564

566 tions cannot, and emulating the resulting response surface then
568 produces a tool that can aid in both physical interpretation of
570 the process-based models and in assessment of agricultural im-
572 pacts under arbitrary climate scenarios. Emulating the climato-
574 logical response isolates long-term impacts from any confound-
576 ing factors that complicate year-over-year changes, and the use
578 of simple functional forms offer the possibility of physical in-
580 terpretation of parameter values. We anticipate that systematic
582 parameter sampling will become the norm in future crop model
584 intercomparison exercises.

586 While the GGCMI Phase II database should offer the foun-
588 dation for multiple future studies, several cautions need to be
590 noted. Because the simulation protocol was designed to focus
592 on change in yield under climate perturbations and not on repli-
594 cating real-world yields, the models are not formally calibrated
596 so cannot be used for impacts projections unless in used in con-
598 junction with historical data (or data products). Because the
600 GGCMI Phase II simulations apply uniform perturbations to
602 historical climate inputs, they do not sample changes in higher
604 order moments, and cannot address the additional crop yield
606 impacts of potential changes in climate variability. Although
608 distributional changes in model projections are fairly uncertain
610 at present, follow-on experiments may wish to consider them.
Several recent studies have described procedures for generat-
612 ing simulations that combine historical data with model pro-
614 jections of changes not only in temperature and precipitation
616 means but in their marginal distributions or temporal depen-
618 dence (e.g. Leeds et al. (2015), Poppick et al. (2016), Chang
et al. (2016) and Haugen et al. (2018)). The ranges for in-
620 put perturbations for the historical climatology were selected to
622 represent the range of potential future climatological changes.
Using the emulator to extrapolate beyond these ranges (Table
624 1) may lead to misinterpretations.

The GGCMI Phase II output dataset invites a broad range

584 of potential future avenues of analysis. A major target area of
research is studying the models themselves including: a de-
586tailed examination of interaction terms between the major in-
put drivers, a robust quantification of the sensitivity of differ-
588ent models to the input drivers, and comparisons with field-
level experimental data. The parameter space tested in GGCMI
590 Phase II will allow detailed investigations into yield variabil-
ity and response to extremes under changing management and
592 CO₂ levels and allow the study of geographic shifts in opti-
mal growing regions for different crops. The output dataset
594 also contains other runs and variables not analyzed or shown
here. Runs include several which allowed adaptation to climate
596 changes by altering growing seasons, and additional variables
include above ground biomass, LAI, and root biomass (as many
598 as 25 output variables for some models). Emulation studies
that are possible include a more systematic evaluation of dif-
600 ferent statistical model specifications and formal calculation of
uncertainties in derived parameters. The development of multi-
602 model ensembles such as GGCMI Phase II provides a way to
begin to better understand crop responses to a range of potential
604 climate inputs, improve process based models, and explore the
potential benefits of adaptive responses included shifting grow-
606 ing season, cultivar types and cultivar geographic extent.

7. Acknowledgments

608 We thank Michael Stein and Kevin Schwarzwald, who pro-
vided helpful suggestions that contributed to this work. This re-
610 search was performed as part of the Center for Robust Decision-
Making on Climate and Energy Policy (RDCEP) at the Univer-
612 sity of Chicago, and was supported through a variety of sources.
RDCEP is funded by NSF grant #SES-1463644 through the
614 Decision Making Under Uncertainty program. J.F. was sup-
ported by the NSF NRT program, grant #DGE-1735359. C.M.
616 was supported by the MACMIT project (01LN1317A) funded

618 through the German Federal Ministry of Education and Re-
search (BMBF). C.F. was supported by the European Research
Council Synergy grant #ERC-2013-SynG-610028 Imbalance-
620 P. P.F. and K.W. were supported by the Newton Fund through
the Met Office Climate Science for Service Partnership Brazil
(CSSP Brazil). A.S. was supported by the Office of Science
of the U.S. Department of Energy as part of the Multi-sector
622 Dynamics Research Program Area. S.O. acknowledges support
from the Swedish strong research areas BECC and MERGE to-
gether with support from LUCCI (Lund University Centre for
624 studies of Carbon Cycle and Climate Interactions). Computing
resources were provided by the University of Chicago Research
626 Computing Center (RCC).

8. Appendix A: Simulations – Assessment

632 The Müller et al. (2017) procedure evaluates response to
year-to-year temperature and precipitation variations in a con-
634 trol run driven by historical climate and compares it to de-
trended historical yields from the FAO (Food and Agriculture
Organization of the United Nations, 2018) by calculating the
Pearson product moment correlation coefficient. The procedure
636 offers no means of assessing CO₂ fertilization, since CO₂ has
been relatively constant over the historical data collection pe-
riod. Nitrogen introduces some uncertainty into the analysis,
since the GGCMI Phase II runs impose fixed, uniform nitrogen
638 application levels that are not realistic for individual countries.
We evaluate up to three control runs for each model, since some
modeling groups provide historical runs for three different ni-
640 tragenous levels.

644 Results are similar to those of GGCMI Phase I, with rea-
sonable fidelity at capturing year-over-year variation, with dif-
ferences by region and crop stronger than difference between
models. (That is, Figure 9 shows more similarity in hori-
646 zontal than vertical bars.) No single model is dominant, with each

model providing near best-in-class performance in at least one location-crop combination. For example, maize in the United States is consistently well-simulated while maize in Mexico is problematic (mean Pearson correlation coefficients of 0.X and 0.X, respectively). In some cases, especially in the developing

world, low correlation coefficients may indicate not model failure but problems in FAO yield data. For example, models have greater apparent skill for rice in India than in the neighboring Pakistan and Bangladesh; this difference may be implausible as solely a model effect. In general, Pearson correlation coeffi-

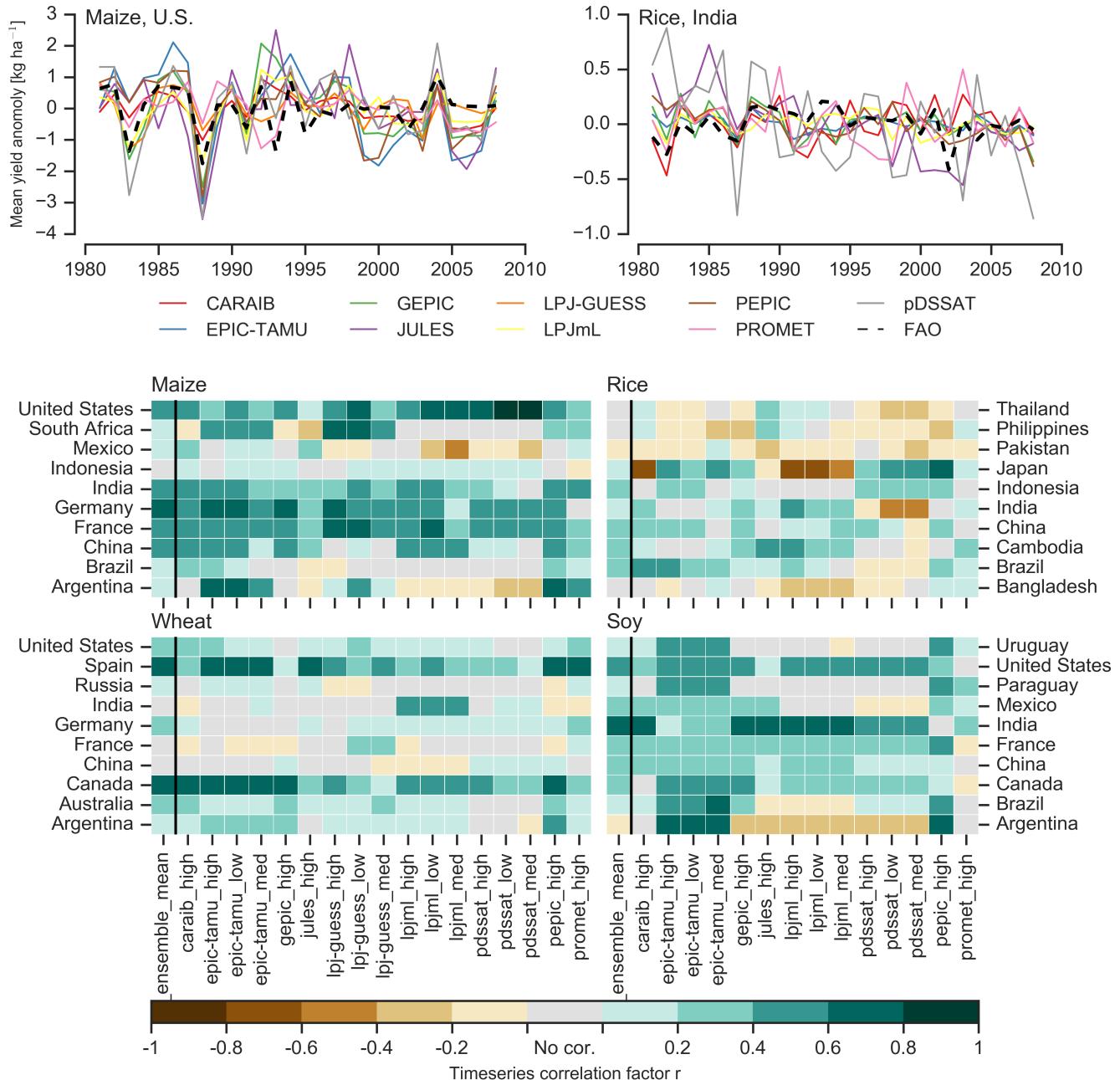


Figure 9: Time series of correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen application case. The heatmaps illustrate the Pearson r correlation coefficient between the detrended simulation mean yield at the country level compared to the detrended FAO yield data for the top producing countries for each crop with continuous FAO data over the 1980–2010 period. Models that provided different nitrogen application levels are shown with low, med, and high label (models that did not simulate different nitrogen levels are analogous to a high nitrogen application level). The ensemble mean yield is also correlated with the FAO data (not the mean of the correlations). Wheat contains both spring wheat and winter wheat simulations. The Pearson r correlation coefficients are similar to those of GGCMI Phase I, with reasonable fidelity at capturing year-over-year variation, with differences by region and crop stronger than difference between models as indicated by more horizontal bars than vertical bars of the same color.

660 clients in GGCMI Phase II are slightly below those of Phase I,
 661 likely because of unrealistic nitrogen levels and lack of country
 662 level calibration in some models. (Compare Figure 9 to Müller
 663 et al. (2017) Figures 1–4 and 6.) Note that in this methodol-
 664 ogy, simulations of crops with low year-to-year variability such
 665 as irrigated rice and wheat will tend to score more poorly than
 666 those with higher variability.

Some models do show particular strength for particular
 668 crops. For example, the EPIC family of models, and espe-
 669 cially the EPIC-TAMU model, perform particularly well for
 670 soy across all regions. In other cases a model has particu-
 671 lar strength in only certain crop and region combinations. For ex-
 672 ample, The strongest correlation coefficient in Figure 9 is that
 673 for the pDSSAT model for maize in the U.S. (the example crop-
 674 model-location used in many example figures in this paper), but
 675 pDSSAT slightly under performs for maize in other regions.
 676 These model assessment results are similar to those for GGCMI
 677 Phase I in Müller et al. (2017).

678 9. Appendix B: Emulation – Assessment

No general criteria exist for defining an acceptable crop
 680 model emulator. For a multi-model comparison exercise like
 681 GGCMI Phase II, one reasonable criterion is what we term the
 682 “normalized error”, which compares the fidelity of an emulator
 683 for a given model and scenario to the inter-model uncertainty.
 684 We define the normalized error e for each scenario as the differ-
 685 ence between the fractional yield change from the emulator and
 686 that in the original simulation, divided by the standard deviation
 687 of the multi-model spread (Equations 2 and 3):

$$F_{scn.} = \frac{Y_{scn.} - Y_{baseline}}{Y_{baseline}} \quad (2)$$

$$e_{scn.} = \frac{F_{em, scn.} - F_{sim, scn.}}{\sigma_{sim, scn.}} \quad (3)$$

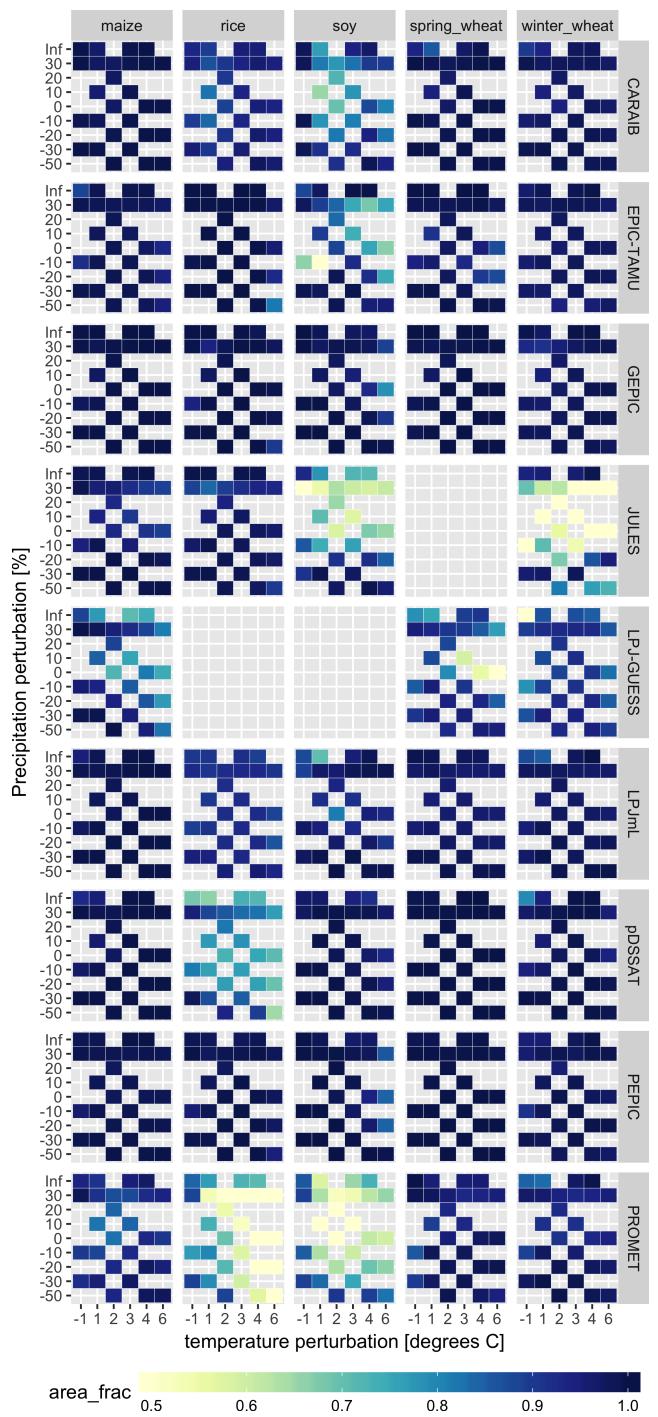


Figure 10: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO₂ and nitrogen held fixed at baseline values. Large columns are crops and large rows models; squares within are T,P scenario pairs. Colors denote the fraction of currently cultivated hectares ('area frac') for each crop with normalized area e less than 1 indicating the the error between the emulation and simulation less than one standard deviation of the ensemble simulation spread. Of the 756 scenarios with these CO₂ and N values, we consider only those for which all 9 models submitted data (Figure S3). JULES did not simulate spring wheat and LPJ-GUESS did not simulate rice and soy. Emulator performance is generally satisfactory, with some exceptions. Emulator failures (significant areas of poor performance) occur for individual crop-model combinations, with performance generally degrading for hotter and wetter scenarios.

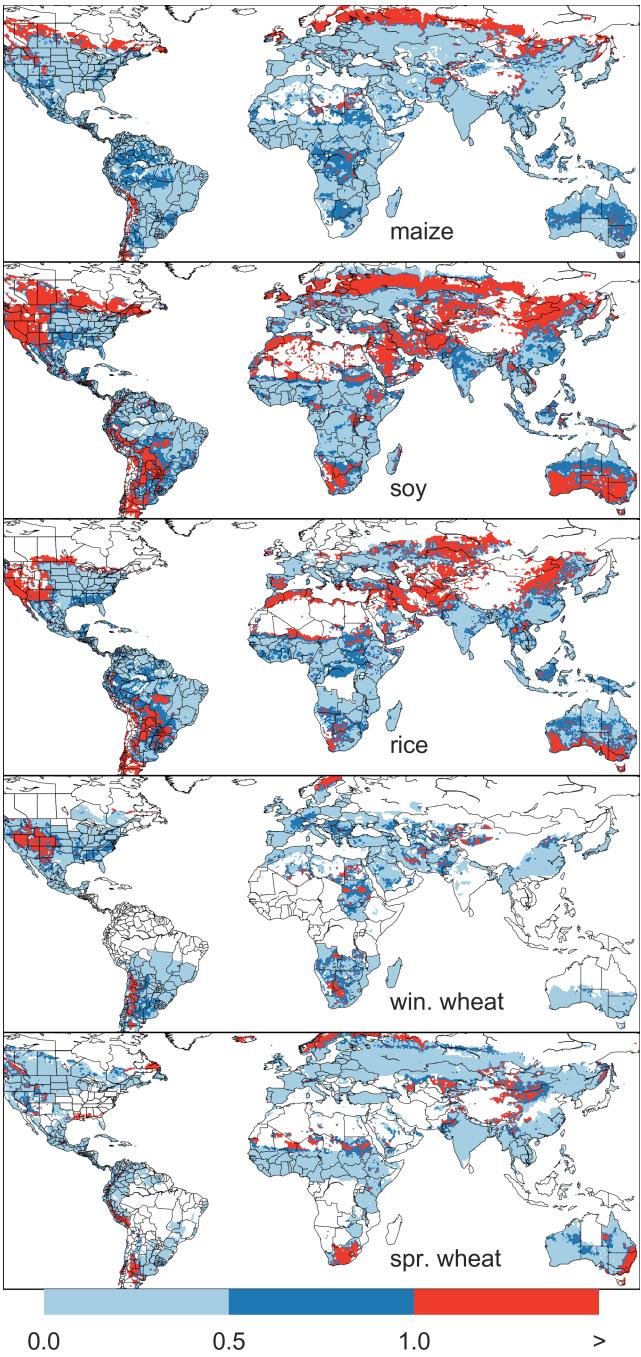


Figure 11: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error e , where $e > 1$ means that emulator error exceeds the multi-model standard deviation. White areas are those where crops are not simulated by this model. Models differ in their areas omitted, meaning the number of samples used to calculate the multi-model standard deviation is not spatially consistent in all locations. Emulator performance is generally good relative to model spread in areas where crops are currently cultivated (compare to Figure 1) and in temperate zones in general; emulation issues occur primarily in marginal areas with low yield potentials. For CARAIB, emulation of soy is more problematic, as was also shown in Figure 10.

Here F_{scn} is the fractional change in a model's mean emulated or simulated yield from a defined baseline, in some scenario (scn.) in C, T, W, and N space; Y_{scn} and $Y_{baseline}$ are the absolute emulated or simulated mean yields. To assess the ability of the polynomial emulation to capture the behavior of complex process-based models, we evaluate the normalized emulator error. That is, for each grid cell, model, and scenario we evaluate the difference between the model yield and its emulation, normalized by the inter-model standard deviation in yield projections. The normalized error e is the difference between the emulated fractional change in yield and that actually simulated, normalized by σ_{sim} , the standard deviation in simulated fractional yields change $F_{sim, scn}$ across all models. The emulator is fitted across all available simulation outputs, and then the error is calculated across the each of the simulation scenarios provided by all nine models (Figure S3).

This metric implies that emulation is generally satisfactory, with several distinct exceptions. Almost all model-crop combination emulators have normalized errors less than one over nearly all currently cultivated hectares (Figure 10), but some individual model-crop combinations are problematic (e.g. PROMET for rice and soy, JULES for soy and winter wheat, Figures S14–S15). Normalized errors for soy are somewhat higher across all models not because emulator fidelity is worse but because models agree more closely on yield changes for soy than for other crops (see Figure S16), lowering the denominator. Emulator performance often degrades in geographic locations where crops are not currently cultivated. Figure 11 shows a CARAIB case as an example, where emulator performance is satisfactory over cultivated areas for all crops other than soy, but uncultivated regions show some problematic areas (see also Figure S12).

This assessment procedure is relatively forgiving for several reason. First, each emulation is evaluated against the simulation

722 actually used to train the emulator. Had we used a spline interpolation the error would necessarily be zero. Second, the performance metric scales emulator fidelity not by the magnitude
 724 of yield changes but by the inter-model spread in those changes.
 726 The normalized error e for a model depends not only on the fidelity of its emulator in reproducing a given simulation but on
 728 the particular suite of models considered in the intercomparison exercise. Where models differ more widely, the standard
 730 for emulators becomes less stringent. This effect is readily seen when comparing assessments of emulator performance in simulations at baseline CO₂ (Figure 10) with those at higher CO₂
 732 levels (Figure S13) because models disagree on the magnitude of CO₂ fertilization. The rationale for this choice of assessment metric is to relate the fidelity of the emulation to an estimate of
 734 true uncertainty, which we take as the multi-model spread. We therefore do not provide a formal parameter uncertainty analysis,
 736 but note that the GGCMI Phase II dataset is well-suited to statistical exploration of emulation approaches and quantification
 738 of emulator fidelity. More rigorous emulator assessments that may be preformed in future work include: testing other statistical specifications including non-parametric models, cross-validation procedures where the emulator is trained on some
 740 portion of data and tested on a held-out portion, and calculating standard error on emulator parameters.

746 10. References

- 748 Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013). Implication of crop model calibration strategies for assessing regional impacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.
- 750 Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D., A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Alderman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Biernath, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising temperatures reduce global wheat production. *Nature Climate Change*, 5, 143–147.
- 754 Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, 3, 827832.
- 760 Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Efficiency, Carbon Sequestration, and Environmental Pollution. *Advances in Agronomy*, 86, 341 – 409.
- 764 Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov, N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat production potentials and management flexibility under the representative concentration pathways. *Global and Planetary Change*, 122, 107 – 121.
- 766 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models. *Agricultural and Forest Meteorology*, 236, 145 – 161.
- 768 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded crop models using statistical estimates. *Agricultural and Forest Meteorology*, 214–215, 134 – 147.
- 770 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S. (2018). Implementing the Nitrogen cycle into the dynamic global vegetation, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific Model Development*, 11, 2789–2812.
- 772 Calvin, K., Patel, P., Clarke, L., Asrar, G., Bond-Lamberty, B., Cui, R. Y., Di Vittorio, A., Dorheim, K., Edmonds, J., Hartin, C., Hejazi, M., Horowitz, R., Iyer, G., Kyle, P., Kim, S., Link, R., McJeon, H., Smith, S. J., Snyder, A., Waldhoff, S., & Wise, M. (2019). Gcam v5.1: representing the linkages between energy, water, land, climate, and economic systems. *Geoscientific Model Development*, 12, 677–698.
- 774 Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L., & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs. *Journal of Climate*, 27, 1829–1844.
- 776 Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, 4, 287 – 291.
- 778 Chang, W., Stein, M., Wang, J., Kotamarthi, V., & Moyer, E. (2016). Changes in spatio-temporal precipitation patterns in changing climate conditions. *Journal of Climate*, 29.
- 780 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process emulation of dynamic computer codes. *Biometrika*, 96, 663–676.
- 782 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth, Duke University, Durham, North Carolina* (pp. 115–118). Durham, North Carolina.
- 784 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).
- 786 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdoorn, M., & François, L. (2011). Responses of European forest ecosystems to 21st century climate: assessing changes in interannual variability and fire intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).
- 788 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K., Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating impact models and sectors (pSIMS). *Environmental Modelling and Software*, 62, 509–516.
- 790 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C., Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J. (2015). The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8, 261–277.
- 792 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958.
- 794 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of climate change impacts on durum wheat in the mediterranean region. *Natural Hazards and Earth System Sciences*, 11, 1293–1302.
- 796 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Regionalization of a large-scale crop growth model for sub-Saharan Africa: Model setup, evaluation, and estimation of maize yields. *Agriculture, Ecosystems & Environment*, 151, 21 – 33.
- 798 Food and Agriculture Organization of the United Nations (2018). FAOSTAT database. URL: <http://www.fao.org/faostat/en/home>.
- 800 Fronzek, S., Pirttioja, N., Carter, T. R., Bind, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J., Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna, 802

- M., Semenov, M. A., Slawinski, C., Stratnovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model wheat yield impact response surfaces showing sensitivity to temperature and precipitation change. *Agricultural Systems*, 159, 209–224.
- Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014). Evaluating the utility of dynamical downscaling in agricultural impacts projections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781.
- Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitivity of Agricultural Model Performance to Different Climate Inputs. *Journal of Applied Meteorology and Climatology*, 55, 151113145618001.
- Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous Crop Growth and Yield: Application for Wheat in Central Europe. *Remote Sensing*, 7, 3934–3965.
- Haugen, M., Stein, M., Moyer, E., & Sriver, R. (2018). Estimating changes in temperature distributions in a large ensemble of climate simulations using quantile regression. *Journal of Climate*, 31, 8573–8588.
- He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H., He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water contents and nitrogen leaching to precipitation, management practices and soil hydraulic properties in semi-arid and humid regions of Canada using the DSSAT model. *Nutrient Cycling in Agroecosystems*, 106, 201–215.
- Heady, E. O. (1957). An Econometric Investigation of the Technology of Agricultural Production Functions. *Econometrica*, 25, 249–268.
- Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa State University Press.
- Holden, P., Edwards, N., PH, G., Fraedrich, K., Lunkeit, F., E, K., Labriet, M., Kanudia, A., & F, B. (2014). Plasim-entsem v1.0: A spatiotemporal emulator of future climate change for impacts assessment. *Geoscientific Model Development*, 7, 433–451.
- Holzkämper, A., Calanca, P., & Fuhrer, J. (2012). Statistical crop models: Predicting the effects of temperature and precipitation changes. *Climate Research*, 51, 11–21.
- Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgiesh, N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson, M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T., Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorff, J., Wedgwood, S., & Keating, B. A. (2014). APSIM Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling and Software*, 62, 327 – 350.
- Howden, S., & Crimp, S. (2005). Assessing dangerous climate change impacts on australia's wheat industry. *Modelling and Simulation Society of Australia and New Zealand*, (pp. 505–511).
- Iizumi, T., Nishimori, M., & Yokozawa, M. (2010). Diagnostics of climate model biases in summer temperature and warm-season insolation for the simulation of regional paddy rice yield in japan. *Journal of Applied Meteorology and Climatology*, 49, 574–591.
- Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Nitrogen Fertilizers. *Ambio*, 6, 146–151.
- Izaurralde, R., Williams, J., McGill, W., Rosenberg, N., & Quiroga Jakas, M. (2006). Simulating soil C dynamics with EPIC: Model description and testing against long-term data. *Ecological Modelling*, 192, 362–384.
- Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. *Agriculture, Ecosystems & Environment*, 93, 73 – 85.
- Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijsman, A., & Ritchie, J. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18, 235 – 265.
- Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural Systems*, 155, 269 – 288.
- Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K., Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, 18, 267 – 288.
- Leeds, W. B., Moyer, E. J., & Stein, M. L. (2015). Simulation of future climate under changing temporal covariance structures. *Advances in Statistical Climatology, Meteorology and Oceanography*, 1, 1–14.
- Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., & Smith, B. (2013). Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa. *Earth System Dynamics*, 4, 385–407.
- Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling wheat yield and crop water productivity with high resolution on a global scale. *Agricultural Systems*, 94, 478 – 493.
- Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a). Global investigation of impacts of PET methods on simulating crop-water relations for maize. *Agricultural and Forest Meteorology*, 221, 164 – 175.
- Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs with yields from major crop cultivations. *Science of The Total Environment*, 572, 526 – 537.
- Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, 150, 1443 – 1452.
- Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2, 014002.
- MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447.
- Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P., Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi, P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano, D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of large simulated yield datasets for studying climate effects. (p. 1100).
- Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrological modelling to study the impact of climate change on the water flows of mountain watersheds. *Journal of Hydrology*, 376, 362 – 377.
- Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., & Calzadilla, A. (2009). Global biomass production potentials exceed expected future demand without the need for cropland expansion. *Nature Communications*, 6.
- McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srinivas, C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S. (2015). Integrated assessments of the impacts of climate change on agriculture: An overview of AgMIP regional research in South Asia. *Chapter in: Handbook of Climate Change and Agroecosystems*, (pp. 201–218).
- Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical weather responsiveness of crop yields: US evidence and implications for the agricultural impacts of climate change. *Environmental Research Letters*, 12.
- Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, 8.
- Müller, C., Elliott, J., Chrysanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. (2017). Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, 10, 1403–1422.
- Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T. (1997). Effect of CO₂ enrichment on carbon and nitrogen interaction in wheat and soybean. *Soil Science and Plant Nutrition*, 43, 789–798.
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91, 1290 – 1300.
- Olin, S., Schurgers, G., Lindeskog, M., Wårild, D., Smith, B., Bodin, P., Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue C:N to changes in atmospheric CO₂ and N management in the main wheat regions of western europe. *Biogeosciences*, 12, 2489–2515. doi:10.5194/bg-12-2489-2015.
- Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.
- Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., &

- Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator. *Geoscientific Model Development*, 8, 1139–1155.
- Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop yields and their variability at different levels of global warming. *Earth System Dynamics*, 9, 479–496.
- Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff, S., & Gerten, D. (2015). Emulating global climate change impacts on crop yields. *Statistical Modelling*, 15, 499–525.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas-
sos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pirttioja, N., Carter, T., Fronzek, S., Bindl, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet, J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M., Semenov, M., Slawinski, C., Strattonovich, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. *Climate Research*, 65, 87–105.
- Poppick, A., McInerney, D. J., Moyer, E. J., & Stein, M. L. (2016). Temperatures in transient climates: Improved methods for simulations with evolving temporal covariances. *Ann. Appl. Stat.*, 10, 477–505.
- Porter et al. (IPCC) (2014). Food security and food production systems. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In C. F. et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK: Cambridge University Press.
- Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of monthly growing areas of 26 irrigated crops.
- Portmann, F., Siebert, S., & Doell, 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. *Global Biogeochemical Cycles*, 24, GB1011.
- Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E., & Arneth, A. (2016). Climate analogues suggest limited potential for intensification of production on current croplands under climate change. *Nature Communications*, 7, 12608.
- Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus A: Dynamic Meteorology and Oceanography*, 58, 461–472.
- Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the reduction and sensitivity analysis of complex environmental models. *Environmental Modelling & Software*, 34, 1 – 4.
- Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, 48.
- Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, 12.
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, 111, 3268–3273.
- Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P., Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182.
- Rosenzweig, C., Ruane, A. C., Antle, J., Elliott, J., Ashfaq, M., Chatta, A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G., Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M., Müller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M., Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., & Wiebe, K. (2018). Coordinating AgMIP data and models across global and regional scales for 1.5°C and 2.0°C assessments. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 376.
- Ruane, A., I. Hudson, N., Asseng, S., Camarrano, D., Ewert, F., Martre, P., J. Boote, K., Thorburn, P., Aggarwal, P., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A., Doltra, J., Gayler, S., Goldberg, R., Grant, R., & Wolf, J. (2016). Multi-wheat-model ensemble responses to interannual climate variability. *Environmental Modelling and Software*, 81, 86–101.
- Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason-D'Croz, D., Müller, C., Porter, C., Phillips, M. M., Raymundo, R. M., Sands, R., Valdivia, R. O., White, J. W., Wiebe, K., & Rosenzweig, C. (2018). Biophysical and economic implications for agriculture of +1.5° and +2.0°C global warming using AgMIP Coordinated Global and Regional Assessments. *Climate Research*, 76, 17–39.
- Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown, D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013). Climate change impact uncertainties for maize in panama: Farm information, climate projections, and yield sensitivities. *Agricultural and Forest Meteorology*, 170, 132 – 145.
- Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agric. Forest Meteorol.*, 200, 233–248.
- Ruane, A. C., McDermid, S., Rosenzweig, C., Baigorria, G. A., Jones, J. W., Romero, C. C., & Cecil, L. D. (2014). Carbon-temperature-water change analysis for peanut production under climate change: A prototype for the agmip coordinated climate-crop modeling project (c3mp). *Glob. Change Biol.*, 20, 394–407.
- Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift*, 19, 135–141.
- Ruiz-Ramos, M., Ferrise, R., Rodriguez, A., Lorite, I., Bindl, M., Carter, T., Fronzek, S., Palosuo, T., Pirttioja, N., Baranowski, P., Buis, S., Cammarano, D., Chen, Y., Dumont, B., Ewert, F., Gaiser, T., Hlavinka, P., Hoffmann, H., Hhn, J., Jurecka, F., Kersebaum, K., Krzyszczak, J., Lana, M., Mechiche-Alami, A., Minet, J., Montesino, M., Nendel, C., Porter, J., Ruget, F., Semenov, M., Steinmetz, Z., Strattonovich, P., Supit, I., Tao, F., Trnka, M., de Wit, A., & Rötter, R. (2018). Adaptation response surfaces for managing wheat under perturbed climate and co2 in a mediterranean environment. *Agricultural Systems*, 159, 260 – 274.
- Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19, 607–620.
- Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frieler, K. (2017). Consistent negative response of US crops to high temperatures in observations and crop models. *Nature Communications*, 8, 13931.
- Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106, 15594–15598.
- Snyder, A., Calvin, K. V., Phillips, M., & Ruane, A. C. (2018). A crop yield change emulator for use in gcam and similar models: Persephone v1.0. *Geoscientific Model Development Discussions*, (pp. 1–42). In open review.
- Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety*, 94, 1735 – 1763.
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93, 485–498.
- Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters*, 35.
- Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., & Martin, J. F. (2014). Modeling sugarcane yield with a process-based model from site to continental scale: Uncertainties arising from model structure and parameter values. *Geoscientific Model Development*, 7, 1225–1245.
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe,

- J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences*, 111, 3228–3232.
- White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Methodologies for simulating impacts of climate change on crop production. *Field Crops Research*, 124, 357 – 368.
- Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T., Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance against site observations of irrigated maize from Mead, Nebraska. *Geoscientific Model Development*, 10, 1291–1320.
- Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield variability in JULES-crop and implications for forcing with seasonal weather forecasts. *Geoscientific Model Development*, 8, 3987–3997.
- de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Landbouwkundige Onderzoeken* : 64.6.,
- Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato yields on changes in climate and co2. *Agricultural and Forest Meteorology*, 112, 217 – 231.
- Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A., Li, T., Lin, E., Liu, Q., Martre, P., Miller, C., Peng, S., Peuelas, J., Ruane, A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci.*, 114, 9326–9331.