

The GGCMI Phase II experiment: 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 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 nitrogen inputs and constitutes a data-rich basis of projected yield changes across 12 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 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

2 Projecting crop yield response to a changing climate is of
3 great importance, especially as the global food production sys-
4 tem will face pressure from increased demand over the next
5 century. Climate-related reductions in supply could therefore
6 have severe socioeconomic consequences. Multiple studies
7 with different crop or climate models predict sharp reduction in
8 yields on currently cultivated cropland under business-as-usual
9 climate scenarios, although their yield projections show consid-
10 erable spread (e.g. Porter et al. (IPCC), 2014, Rosenzweig et al.,
11 2014, Schauberger et al., 2017, and references therein). Model
12 differences are unsurprising because crop responses in models
13 can be complex, with crop growth a function of complex inter-
14 actions between climate inputs and management practices.

15 Computational Models have been used to project crop yields
16 since the 1950's, beginning with statistical models (Heady,
17 1957, Heady & Dillon, 1961) that attempt to capture the rela-
18 tionship between input factors and resultant yields. These sta-
19 tistical models were typically developed on a small scale for lo-
20 cations with extensive histories of yield data. The emergence of
21 computers allowed development of numerical models that sim-
22 ulate the process of photosynthesis and the biology and phe-
23 nology of individual crops (first proposed by de Wit (1957),
24 Duncan et al. (1967) and attempted by Duncan (1972)). His-
25 torical mapping of crop model development can be found in
26 the appendix/supplementary of Rosenzweig et al. (2014). A
27 half-century of improvement in both models and computing re-
28 sources means that researchers can now run crop simulation
29 models for many years at high spatial resolution on the global
30 scale.

31 Both types of models continue to be used, and compara-
32 tive studies have concluded that when done carefully, both ap-

33 proaches can provide similar yield estimates (e.g. Lobell &
34 Burke, 2010, Moore et al., 2017, Roberts et al., 2017, Zhao
35 et al., 2017). Models tend to agree broadly in major response
36 patterns, including a reasonable representation of the spatial
37 pattern in historical yields of major crops (e.g. Elliott et al.,
38 2015, Müller et al., 2017) and projections of decreases in yield
39 under future climate scenarios.

40 Process models do continue to struggle with some important
41 details, including reproducing historical year-to-year variabil-
42 ity (e.g. Müller et al., 2017), reproducing historical yields when
43 driven by reanalysis weather (e.g. Glotter et al., 2014), and low
44 sensitivity to extreme events (e.g. Glotter et al., 2015). These
45 issues are driven in part by the diversity of new cultivars and ge-
46 netic variants, which outstrips the ability of academic modeling
47 groups to capture them (e.g. Jones et al., 2017). Models do not
48 simulate many additional factors affecting production, includ-
49 ing pests/diseases/weeds. For these reasons, individual stud-
50 ies must generally re-calibrate models to ensure that short-term
51 predictions reflect current cultivar mixes, and long-term pro-
52 jections retain considerable uncertainty (Wolf & Oijen, 2002,
53 Jagtap & Jones, 2002, Iizumi et al., 2010, Angulo et al., 2013,
54 Asseng et al., 2013, 2015). Inter-model discrepancies can also
55 be high in areas not yet cultivated (e.g. Challinor et al., 2014,
56 White et al., 2011). Finally, process-based models present ad-
57 dditional difficulties for high-resolution global studies because
58 of their complexity and computational requirements. For eco-
59 nomic impacts assessments, it is often impossible to integrate a
60 set of process-based crop models directly into an integrated as-
61 sessment model to estimate the potential cost of climate change
62 to the agricultural sector.

63 Nevertheless, process-based models are necessary for under-
64 standing the global future yield impacts of climate change for
65 many reasons. First, cultivation may shift to new areas, where
66 no yield data are currently available and therefore statistical

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

67 models cannot apply. Yield data are also often limited in the de-
 68 veloping world, where future climate impacts may be the most
 69 critical. Second, only process-based models can capture the
 70 growth response to elevated CO₂, novel conditions that are not
 71 represented in historical data (e.g. Pugh et al., 2016, Roberts
 72 et al., 2017). Similarly process-based models can represent
 73 novel changes in management practices (e.g. fertilizer input)
 74 that may ameliorate climate-induced damages.

75 The overall goal of this study is a better understanding of
 76 global crop model response to the major drivers in a climate
 77 change context. Most previous climate-change-focused global
 78 crop modeling studies have simulated model response to rep-
 79 resentative concentration pathways (RCPs). RCPs are likely to
 80 have strong covariance between precipitation, temperature and
 81 CO₂ that may be hard to decompose statistically. The differ-
 82 ences in year-to-year memory in the models and complexity of
 83 the changes in year-to-year distributions in weather under RCP
 84 scenarios in climate models are complications we seek to con-
 85 trol for with this study. We propose to test the response to major
 86 drivers and their interaction by isolating individual input drivers
 87 through simulations on first-moment shifts applied to the histor-
 88 ical climatology instead of RCP simulations. As emulators are
 89 a fundamentally a distillation of the process-based model down
 90 to its major drivers, the same applied to their development.

91 Statistical emulation of crop simulations has been used to
 92 combine advantageous features of both statistical and process-
 93 based models. The statistical representation of complicated nu-
 94 merical simulation (e.g. O'Hagan, 2006, Conti et al., 2009), in
 95 which simulation output acts as the training data for a statisti-₁₀₁
 96 cal model, has been of increasing interest with the growth of₁₀₂
 97 simulation complexity and volume of output. Such emulators₁₀₃
 98 or "surrogate models" have been used in a variety of fields in-₁₀₄
 99 cluding hydrology (e.g. Razavi et al., 2012), engineering (e.g.₁₀₅
 100 Storlie et al., 2009), environmental sciences (e.g. Ratto et al.,₁₀₆

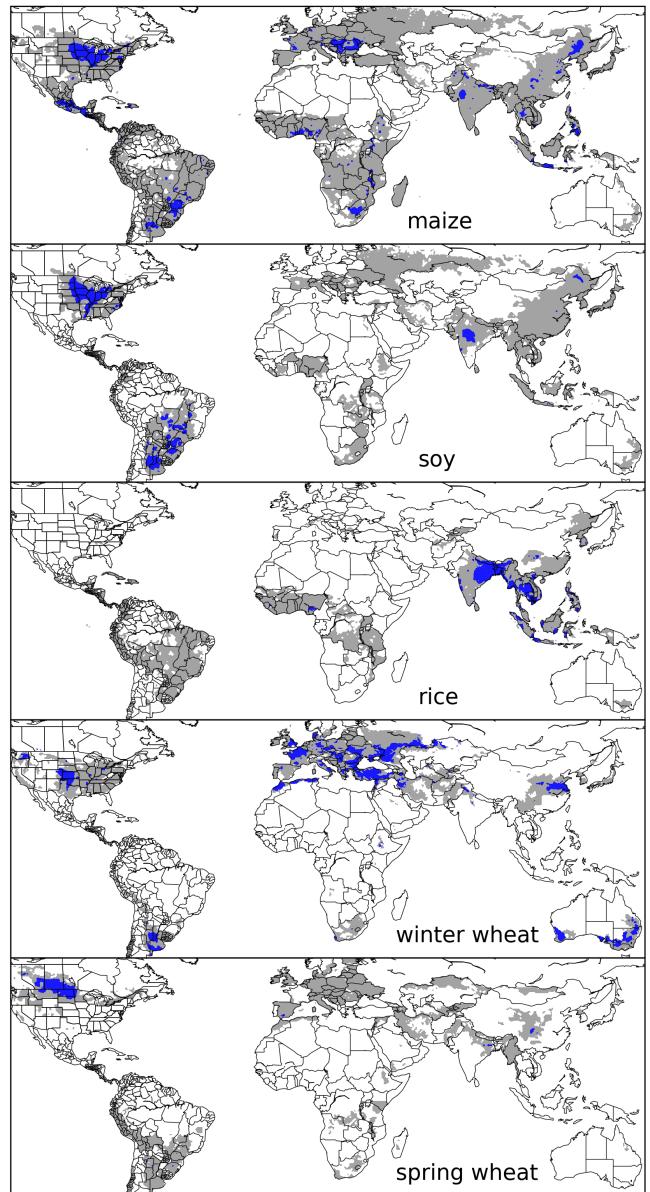


Figure 1: Presently cultivated area for rain-fed crops in the real world. Blue contour area indicates grid cells with more than 20,000 hectares (approx. 10% of the land in a grid cell near the equator for example) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated 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.

2012), and climate (e.g. Castruccio et al., 2014, Holden et al.,
 2014). For agricultural impacts studies, emulation of process-
 based models allows exploring crop yields in regions outside
 ranges of current cultivation and with input variables outside
 historical precedents, in a lightweight, flexible form that is com-
 patible with economic studies.

107 In the past decade, many studies have developed emulators of₁₃₅ response surface) to analyze non-RCP crop model simulations
 108 crop yields from process-based models. Early studies propos-₁₃₆ that sampled a suite of climate (and management) perturbations
 109 ing or describing potential emulators include Howden & Crimp₁₃₇ (the focus of this study): Makowski et al. (2015) for tempera-
 110 (2005), Räisänen & Ruokolainen (2006) and Lobell & Burke₁₃₈ ture, CO₂, and nitrogen, Pirttioja et al. (2015) and Snyder et al.
 111 (2010). In an early application, Ferrise et al. (2011) used a Arti-₁₃₉ (2018) for temperature, water, and CO₂, and (Fronzek et al.,
 112 ficial Neural Net trained on simulation outputs to predict wheat₁₄₀ 2018) for temperature and water, with all studies simulating se-
 113 yields in the Mediterranean. Studies developing single-model₁₄₁ lected sites for a limited number of crops.
 114 emulators include Holzkämper et al. (2012) for the CropSyst₁₄₂
 115 model, Ruane et al. (2013) for the CERES wheat model, Oye-₁₄₃
 116 bamiji et al. (2015) for the LPJmL model (for multiple crops,₁₄₄
 117 using multiple scenarios as a training set). In recent years, emu-₁₄₅
 118 lators have begun to be used in the context of multi-model inter-₁₄₆ The use of limited input parameter space or restricted geo-
 119 comparisons, with Blanc & Sultan (2015), Blanc (2017), Ost-₁₄₇ graphic scope impedes ability to build future projections or
 120 berg et al. (2018) and Mistry et al. (2017) using them to analyze₁₄₈ to understand interaction effects in global process-based crop
 121 the five crop models of the Inter-Sectoral Impacts Model In-₁₄₉ models.
 122 tercomparison Project (ISIMIP) (Warszawski et al., 2014) (for₁₅₀
 123 maize, soy, wheat, and rice). Approaches differ: Blanc & Sultan₁₅₁ The Global Gridded Crop Model Intercomparison (GGCMI)
 124 (2015) and Blanc (2017) used local weather variables (and CO₂₁₅₂
 125 values) and yields but emulate across soil types using historical₁₅₃ Phase II experiment seeks to provide a comprehensive global
 126 simulations and a future climate scenario (RCP8.5 over mul-₁₅₄ dataset to allow systematically exploring how process-based
 127 tiple climate models); Ostberg et al. (2018) used global mean₁₅₅ crop models for the major crop respond to the main climate
 128 temperature change (and CO₂) as regressors but pattern-scale₁₅₆ and management drivers and their interactions. The experiment
 129 to emulate local yields using multiple climate scenarios; Mis-₁₅₇ involves running a suite of process-based crop models across
 130 try et al. (2017) used local weather and yields and a historical₁₅₈ historical conditions perturbed by a set of defined input pa-
 131 simulation and compare with data. ₁₅₉rameters, and was conducted as part of the Agricultural Model
 132 Recently efforts have been made to generate datasets that al-₁₆₀ Intercomparison and Improvement Project (AgMIP) (Rosen-
 133 low more systematic sampling of the input var ₁₆₁zweig et al., 2013, 2014), an international effort conducted un-
 134 Other studies have used the development of an emulators (or₁₆₂ der a framework similar to the Climate Model Intercomparison
 135 ₁₆₃ project (CMIP) (Taylor et al., 2012, Eyring et al., 2016). The
 136 ₁₆₄ GGCMI protocol builds on the AgMIP Coordinated Climate-
 137 ₁₆₅ Crop Modeling Project (C3MP) (Ruane et al., 2014, McDer-
 138 ₁₆₆ mid et al., 2015) and will contribute to the AgMIP Coordinated
 139 ₁₆₇ Global and Regional Assessments (CGRA) (Ruane et al., 2018,
 140 ₁₆₈ Rosenzweig et al., 2018).

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: Phase II input variable test levels. Temperature and precipitation values indicate the perturbations from the historical, climatology. * Only simulated by one model. W-percentage does not apply to the irrigated (W_{inf}) simulations.

163 GGCMI Phase II is designed to allow addressing goals such¹⁹⁴
164 as understanding where highest-yield regions may shift un¹⁹⁵
165 der climate change; exploring future adaptive management¹⁹⁶
166 strategies; understanding how interacting parameters affect¹⁹⁷
167 crop yield; quantifying uncertainties across models and major¹⁹⁸
168 drivers; and testing strategies for producing lightweight emu-¹⁹⁹
169 lators of process-based models. In this paper, we describe the
200
170 GGCMI Phase II experiments, summarize output and present
201 initial results, demonstrate that it is tractable to emulation, and
202 present a simple climatological emulator as a potential tool for
203 impacts assessments.
204

174 2. Materials and Methods

175 2.1. GGCMI Phase II: experiment design

176 GGCMI Phase II is the continuation of a multi-model com-²⁰⁹
177 parison exercise begun in 2014. The initial Phase I compared²¹⁰
178 harmonized yields of 21 models for 19 crops over a historical²¹¹
179 (1980-2010) scenario with a primary goal of model evaluation²¹²
180 (Elliott et al., 2015, Müller et al., 2017). Phase II compares sim-²¹³
181 ulations of 12 models for 5 crops (maize, rice, soybean, spring²¹⁴
182 wheat, and winter wheat) over hundreds of scenarios in which²¹⁵
183 individual climate or management inputs are adjusted from²¹⁶
184 their historical values. The reduced set of crops includes the²¹⁷
185 three major global cereals and the major legume and accounts²¹⁸
186 for over 50% of human calories (in 2016, nearly 3.5 billion tons²¹⁹
187 or 32% of total global crop production by weight (Food and²²⁰
188 Agriculture Organization of the United Nations, 2018).
221

189 The major goals of GGCMI Phase II are to:

- Explore differences in crop response to warming across the Earth's climate regions.
- Provide a dataset that allows statistical emulation of crop model responses for downstream modelers.
- Illustrate differences in potential adaptation via growing season changes.

The guiding scientific rationale of GGCMI Phase II is to provide a comprehensive, systematic evaluation of the response of process-based crop models to different values for carbon dioxide, temperature, water, and applied nitrogen (collectively known as “CTWN”). Phase II of the GGCMI project consists of a series of simulations, each with one or more of the CTWN dimensions perturbed over the 31-year historical time series (1980-2010) used in Phase I. 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). Two models require sub-daily input data and use alternative sources. See Elliott et al. (2015) for additional details.

The experimental protocol consists of 9 levels for precipitation perturbations, 7 for temperature, 4 for CO₂, and 3 for applied nitrogen, for a total of 672 simulations for rain-fed agriculture and an additional 84 for irrigated (Table 1). For irrigated 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 absolute offsets from the daily mean, minimum, and maximum temperature time series for each grid cell used as inputs. Precipitation perturbations are applied as fractional changes at the grid cell level, and carbon dioxide and nitrogen levels are specified as discrete values applied uniformly over all grid cells. Note that CO₂ changes are applied independently of changes in climate variables, so that higher CO₂ is not associated with

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 the GGCMI Phase II and the number of C, T, W, and N simulations performed for rain-fed crops (“Sims per Crop”), with 672 as the maximum. “N-Dim.” indicates if simulations include varying nitrogen levels; two models omit this dimension. All models provided the same set of simulations across all modeled crops, but some omitted individual crops in some cases. (For example, APSIM did not simulate winter wheat.) Irrigated simulations are provided at the level of the other covariates for each model (for an additional 84 simulations for the fully-sampled models). Geographic extent of simulation varies to some extent within a certain model for different scenarios (672 rain-fed simulations does not necessarily equal 672 climatological yields in all areas). This geographic variance only applies for areas far outside the area of currently cultivated crops. Two models (marked with *) use non-AgMERRA climate inputs. For further details on models, see Elliott et al. (2015). †PROMET provided simulations at only two nitrogen levels so is not emulated across the nitrogen dimension.

higher temperatures. An additional, identical set of scenarios²⁴⁶ different crops. Growing seasons are identical across models,
 (at the same C, T, W, and N levels) simulate adaptive agron-²⁴⁷ but vary by crop and by location on the globe. All stresses
 omy under climate change by varying the growing season for²⁴⁸ except factors related to nitrogen, temperature, and water (e.g.
 crop production. (These adaptation simulations are not shown²⁴⁹ Alkalinity, salinity) are disabled. No additional nitrogen inputs,
 or analyzed here.) The resulting GGCMI data set captures a²⁵⁰ such as atmospheric deposition, are considered, but some mod-
 distribution of crop responses over the potential space of future²⁵¹ els have individual assumptions on soil organic matter that may
 climate conditions.²⁵² release additional nitrogen through mineralization. See Rosen-
²⁵³ zweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017)
²⁵⁴ for further details on models and underlying assumptions.

The 12 models included in GGCMI Phase II are all mecha-²⁵⁴ Each model is run at 0.5 degree spatial resolution and covers
 nistic process-based crop models that are widely used in im-²⁵⁵ all currently cultivated areas and much of the uncultivated land
 pacts assessments (Table 2). Although some of the models²⁵⁶ area. Coverage extends considerably outside currently culti-
 shares a common base (e.g. LPJmL and LPJ-GUESS and the²⁵⁷ vated areas because cultivation will likely shift under climate
 EPIC models), they have developed independently from this²⁵⁸ change. See Figure 1 for the present-day cultivated area of
 shared base, for more details on the genealogy of the mod-²⁵⁹ rain-fed crops, and Figure S1 in the supplemental material for
 els see Figure S1 in Rosenzweig et al. (2014). Differences in²⁶⁰ irrigated crops. Some areas such as Greenland, far-northern
 model structure does mean that several key factors are not stan-²⁶¹ Canada, Siberia, Antarctica, the Gobi and Sahara deserts, and
 dardized across the experiment, including secondary soil nutri-²⁶² central Australia are not simulated as they are assumed to re-
 ents, carry over effects across growing years including residue²⁶³
 management and soil moisture, and extent of simulated area for²⁶⁴

main non-arable even under an extreme climate change. Growing seasons are standardized across models with data adapted from several sources (Sacks et al., 2010, Portmann et al., 2008, 2010).

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.

All models produce as output, crop yields ($\text{tons ha}^{-1} \text{ year}^{-1}$) for each 0.5 degree grid cell. Because both yields and yield changes vary substantially across models and across grid cells, we primarily analyze relative change from a baseline. We take as the baseline the scenario with historical climatology (i.e. T and P changes of 0), C of 360 ppm, and applied N at 200 kg ha^{-1} . We show absolute yields in some cases to illustrate geographic differences in yields for a single model.

2.2. Simulation model validation approach

To verify the skill of the process-based models used, we repeat the validation exercises presented in Müller et al. (2017) for GGCMI Phase I. Note however that the GGCMI Phase II simulations are designed for evaluating changes in yield but not absolute yields, and so omit the calibrations used in predicting modeling to account for cultivar, pest loss, and management differences. The Phase II simulations also do not reproduce realistic nitrogen application levels for individual countries, since nitrogen is one of the parameters systematically varied. The Müller et al. (2017) validation procedure evaluates response to year-to-year temperature and precipitation variations in a control run driven by historical climate and compares it to detrended historical yields from the FAO (Food and Agri-

culture Organization of the United Nations, 2018) by calculating the Pearson correlation coefficient. The procedure offers no means of assessing CO₂ fertilization, since CO₂ has been relatively constant over the historical data collection period. Nitrogen data are limited for many countries, and as mentioned the GGCMI Phase II runs impose fixed and uniform nitrogen application, introducing some uncertainty into the analysis. We evaluate one or more control runs for each model, since some modeling groups provide historical runs for three different nitrogen levels.

2.3. Climatological-mean yield emulator design

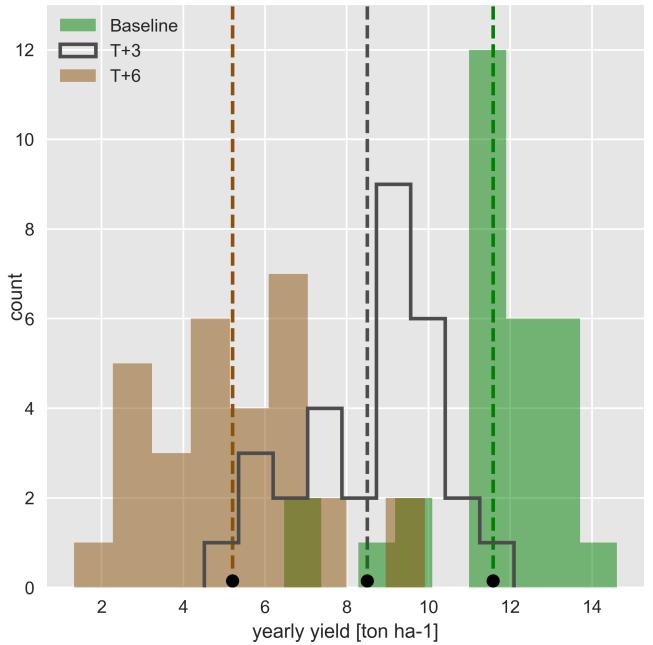


Figure 2: Example showing both climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for a grid cell in northern Iowa (a representative high-yield region) from the pDSSAT model, for the baseline climatology (1981–2010) and for scenarios with temperature shifted by (T) +3 and (T) +6 °C, with other variables held at baseline values. Dashed vertical lines and black dots indicate the climatological mean yield.

We construct our emulator at the 30-year climatological mean level. Blanc & Sultan (2015) and Blanc (2017) have shown that an emulator of a global process-based crop model can be successfully developed at the yearly scale. Our decision to construct a climatological-mean yield emulator is driven by the target application for this analysis tool. Many impact modelers

315 are not focused on the changes in the year-to-year variability in
 316 yields, but instead on the broad mean changes over the multi-
 317 decadal timescale. Emulation involves fitting individual regres-
 318 sion models for each crop, simulation model, and 0.5 degree
 319 geographic pixel from the GGCMI Phase II data set. The re-
 320 gressors are the applied constant perturbations in temperature,
 321 water, nitrogen and CO₂, we aggregate the simulation outputs
 322 in the time dimension, and regress on the 30-year mean yields.
 323 (See Figure 2 for illustration). The regression therefore omits
 324 information about yield responses to year-to-year climate per-
 325 turbations, which are more complex. Emulating inter-annual
 326 yield variations would likely require considering statistical de-
 327 tails of the historical climate time series, including changes in
 328 marginal distribution and temporal dependencies. (Future work
 329 should explore this). The climatological emulation indirectly
 330 includes any yield response to geographically distributed fac-
 331 tors such as soil type, insolation, and the baseline climate itself,
 332

because we construct separate emulators for each grid cell. The emulator parameter matrices are portable and the yield computations are cheap even at the half-degree grid cell resolution, so we do not aggregate in space at this time.

Blanc & Sultan (2015) and Blanc (2017) have shown that a fractional polynomial specification is more effective than a standard polynomial for representing simulations at the yearly level across different soil types geographically (not at the grid cell level). We do not test this specification here, and instead use as a starting point a standard third-order polynomial to represent the climatological-mean response at the grid cell level as it is the simplest effective specification. We regress climatological-mean yields against a third-order polynomial in C, T, W, and N with interaction terms. 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).

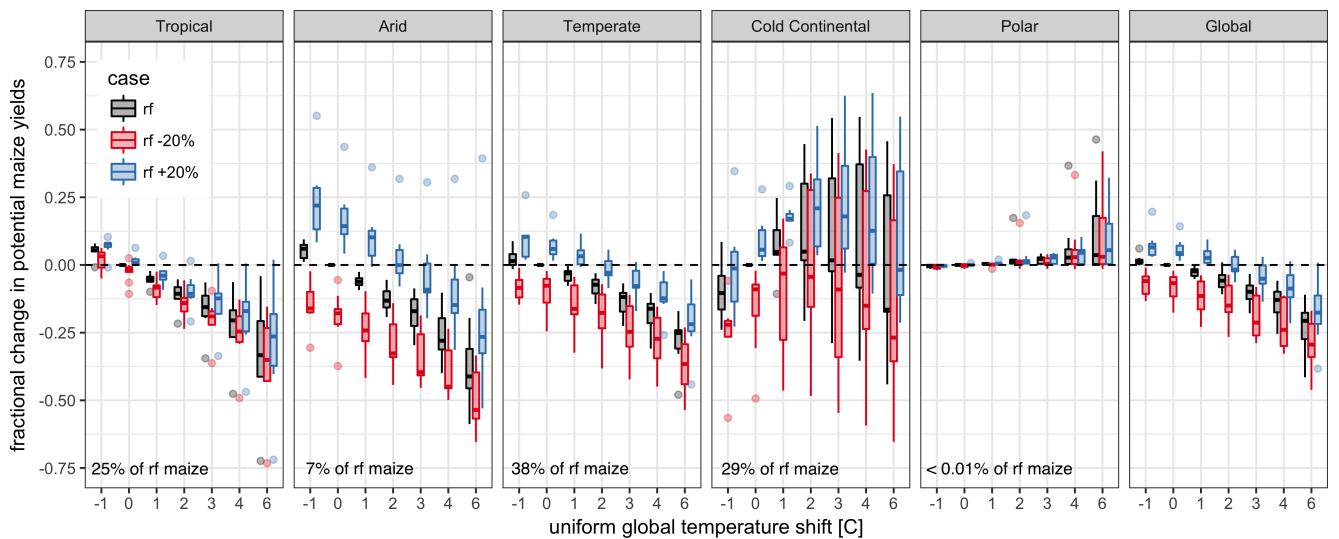


Figure 3: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation 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 (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 with global changes 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.

349 We include interaction terms (both linear and higher-order) be-₃₈₃
 350 cause past studies have shown them to be significant effects.₃₈₄
 351 For example, Lobell & Field (2007) and Tebaldi & Lobell₃₈₅
 352 (2008) showed that in real-world yields, the joint distribution₃₈₆
 353 in T and W is needed to explain observed yield variance (C₃₈₇
 354 and N are fixed in these data). Other observation-based stud-₃₈₈
 355 ies have shown the importance of the interaction between water₃₈₉
 356 and nitrogen (e.g. Aulakh & Malhi, 2005), and between nitro-₃₉₀
 357 gen and carbon dioxide (Osaki et al., 1992, Nakamura et al.,
 358 1997). We do not focus on comparing different model speci-
 359 fications in this study, and instead stick to a relatively simple
 360 parameterized specification that allows for some, albeit limited,
 361 coefficient interpretation.

362 The limited GGCMI variable sample space means that use
 363 of the full polynomial expression described above, which has
 364 34 terms for the rain-fed case (12 for irrigated), can be prob-
 365 lematic, and can lead to over-fitting and unstable parameter es-
 366 timations. We therefore reduce the number of terms through a
 367 feature selection cross-validation process in which terms in the
 368 polynomial are tested for importance. In this procedure higher-₃₉₂
 369 order and interaction terms are added successively to the model;₃₉₃
 370 we then follow the reduction of the the aggregate mean squared
 371 error with increasing terms and eliminate those terms that do
 372 not contribute significant reductions. See supplemental docu-₃₉₄
 373 ments for more details. We select terms by applying the feature
 374 selection process to the three models that provided the com-₃₉₅
 375 plete set of 672 rain-fed simulations (pDSSAT, EPIC-TAMU,₃₉₆
 376 and LPJmL); the resulting choice of terms is then applied for
 377 all emulators.₄₀₀

378 Feature importance is remarkably consistent across all three₄₀₂
 379 models and across all crops (see Figure S4 in the supplemental₄₀₃
 380 material). The feature selection process results in a final poly-₄₀₄
 381 nomial in 23 terms, with 11 terms eliminated. We omit the N³₄₀₅
 382 term, which cannot be fitted because we sample only three ni-₄₀₆

383rogen levels. We eliminate many of the C terms: the cubic,
 384 the CT, CTN, and CWN interaction terms, and all higher order
 385 interaction terms in C. Finally, we eliminate two 2nd-order in-
 386 teraction terms in T and one in W. Implication of this choice
 387 include that nitrogen interactions are complex and important,
 388 and that water interaction effects are more nonlinear than those
 389 in temperature. The resulting statistical model (Equation 1) is
 390 used for all grid cells, models, and crops:

$$\begin{aligned}
 Y = & K_1 & (1) \\
 & + K_2 C + K_3 T + K_4 W + K_5 N \\
 & + K_6 C^2 + K_7 T^2 + K_8 W^2 + K_9 N^2 \\
 & + K_{10} CW + K_{11} CN + K_{12} TW + K_{13} TN + K_{14} WN \\
 & + K_{15} T^3 + K_{16} W^3 + K_{17} TWN \\
 & + K_{18} T^2 W + K_{19} W^2 T + K_{20} W^2 N \\
 & + K_{21} N^2 C + K_{22} N^2 T + K_{23} N^2 W
 \end{aligned}$$

391 To fit the parameters K , we use a Bayesian Ridge probabilis-
 392 tic estimator (MacKay, 1991), which reduces volatility in pa-
 393 rameter estimates when the sampling is sparse, by weighting
 394 parameter estimates towards zero. The Bayesian Ridge method
 395 is necessary to maintain a consistent functional form across all
 396 models, and locations as the linear least squares fails to pro-
 397 vide a stable result in many cases. In the GGCMI Phase II
 398 experiment, the most problematic fits are those for models that
 399 provided a limited number of cases or for low-yield geographic
 400 regions where some modeling groups did not run all scenarios.
 401 Because we do not attempt to emulate models that provided
 402 less than 50 simulations, the lowest number of simulations em-
 403 ultated across the full parameter space is 130 (for the PEPIC
 404 model). We use the implementation of the Bayesian Ridge esti-
 405 mator from the scikit-learn package in Python (Pedregosa et al.,
 406 2011).

407 The resulting parameter matrices for all crop model emulators
 408 are available on request, as are the raw simulation data and
 409 a Python application to emulate yields. The yield output for
 410 single GGCMI model that simulates all scenarios and all five
 411 crops is ~12.5 GB; the emulator is ~100 MB, a reduction by
 412 over two orders of magnitude.

413 the normalized error e for a model depends not only on the fidelity
 414 of its emulator in reproducing a given simulation but on
 415 the particular suite of models considered in the intercomparison
 416 exercise. The rationale for this choice is to relate the fidelity of
 417 the emulation to an estimate of true uncertainty, which we take
 418 as the multi-model spread.

413 2.4. Emulator evaluation

414 Because no general criteria exist for defining an acceptable
 415 model emulator, we develop a metric of emulator performance
 416 specific to GGCMI. For a multi-model comparison exercise like
 417 GGCMI, a reasonable criterion is what we term the “normalized
 418 error”, which compares the fidelity of an emulator for a given
 419 model and scenario to the inter-model uncertainty. We define
 420 the normalized error e for each scenario as the difference be-
 421 tween the fractional yield change from the emulator and that in
 422 the original simulation, divided by the standard deviation of the
 423 multi-model spread (Equations 2 and 3):

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

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

424 Here $F_{scn.}$ is the fractional change in a model’s mean emu-
 425 lated or simulated yield from a defined baseline, in some sce-
 426 nario (scn.) in C, T, W, and N space; $Y_{scn.}$ and $Y_{baseline}$ are the
 427 absolute emulated or simulated mean yields. The normalized
 428 error e is the difference between the emulated fractional change
 429 in yield and that actually simulated, normalized by $\sigma_{sim.}$, the
 430 standard deviation in simulated fractional yields $F_{sim, scn.}$ across
 431 all models. The emulator is fit across all available simulation
 432 outputs, and then the error is calculated across the simulation
 433 scenarios provided by all nine models (Figure 8 and Figures
 434 S12 and Figures S13 in supplemental documents). Note that

441 3. Results

442 3.1. Simulation results

443 Crop models in the GGCMI ensemble show a broadly con-
 444 sistent responses to climate and management perturbations in
 445 most regions, with a strong negative impact of increased tem-
 446 perature in all but the coldest regions. We illustrate this result
 447 for rain-fed maize in Figure 3, which shows yields for the pri-
 448 mary Köppen-Geiger climate regions (Rubel & Kottek, 2010).
 449 In warming scenarios, models show decreases in maize yield in
 450 the temperate, tropical, and arid regions that account for nearly
 451 three-quarters of global maize production. These impacts are
 452 robust for even moderate climate perturbations. In the temper-
 453 ate zone, even a 1 degree temperature rise with other variables
 454 held fixed leads to a median yield reduction that outweighs the
 455 variance across models. A 6 degree temperature rise results in
 456 median loss of ~25% of yields with a signal to noise of nearly
 457 three. A notable exception is the cold continental region, where
 458 models disagree strongly, extending even to the sign of impacts.
 459 Model simulations of other crops produce similar responses to
 460 warming, with robust yield losses in warmer locations and high
 461 inter-model variance in the cold continental regions (Figures
 462 S7).

463 The effects of rainfall changes on maize yields are also as ex-
 464 pected and are consistent across models. Increased rainfall mit-
 465 igates the negative effect of higher temperatures, most strongly
 466 in arid regions. Decreased rainfall amplifies yield losses and
 467 also increases inter-model variance more strongly, suggesting

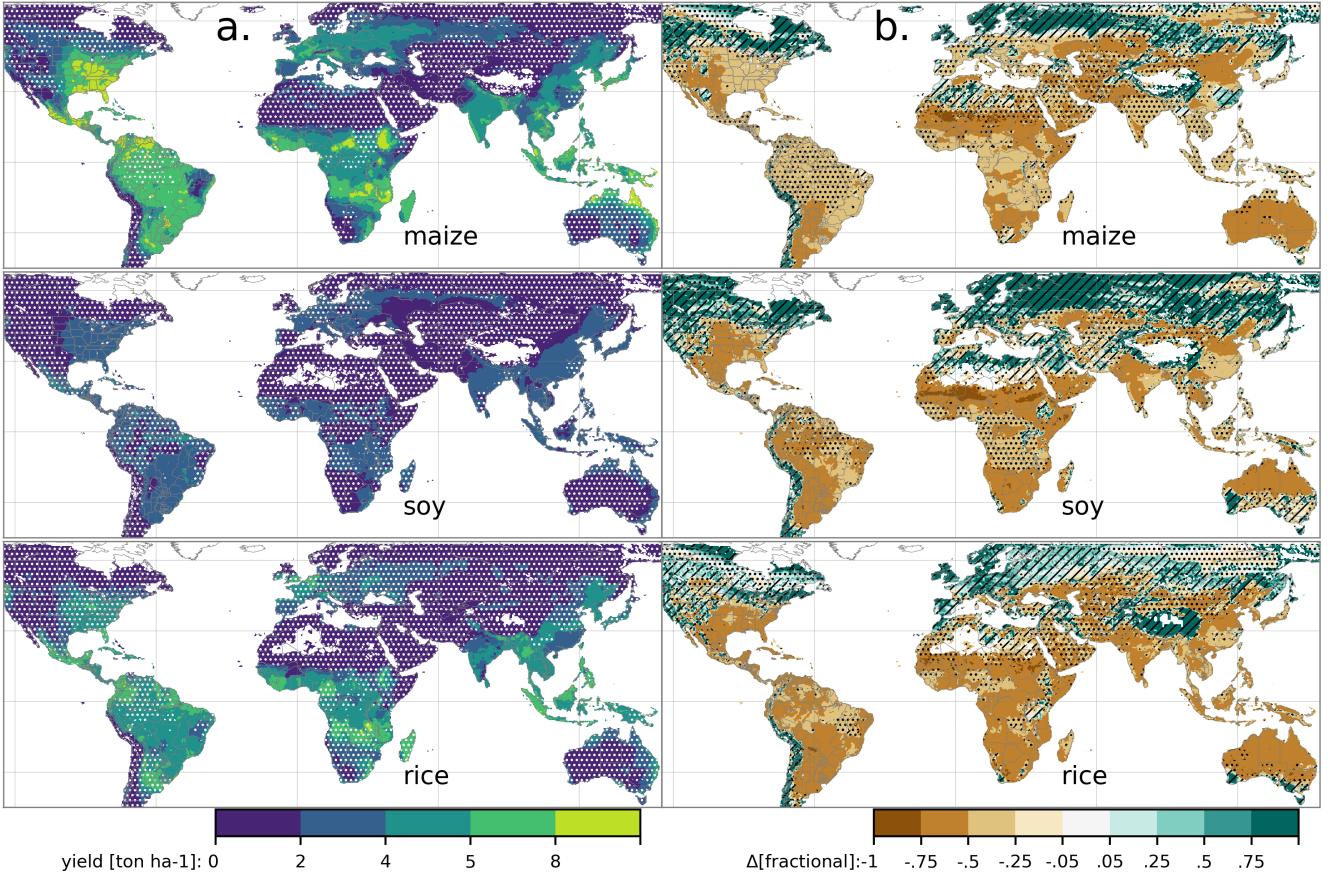


Figure 4: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure 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 $T + 4^\circ\text{C}$ 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.

468 that models have difficulty representing crop response to water₄₈₁
 469 stress. We show only rain-fed maize here; see Figure S5 for the₄₈₂
 470 irrigated case. As expected, irrigated crops are more resilient to₄₈₃
 471 temperature increases in all regions, especially so where water₄₈₄
 472 is limiting.

473 Mapping the distribution of baseline yields and yield changes₄₈₆
 474 shows the geographic dependencies that underlie these results.

475 Figure 4 shows baseline and changes in the $T+4$ scenario for
 476 rain-fed maize, soy, and rice in the multi-model ensemble mean,₄₈₈
 477 with locations of model agreement marked. Absolute yield po-₄₈₉
 478 tentials are have strong spatial variation, with much of the₄₉₀
 479 Earth's surface area unsuitable for any given crop. In general,₄₉₁
 480 models agree most on yield response in regions where yield₄₉₂

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 both more uncertain and show fewer areas of increased yield in the inter-model mean.

3.2. Simulation model validation results

Figure 5 shows the Pearson time series correlation between the simulation model yield and FOA yield data. Figure 5 can be compared to Figures 1,2,3,4 and 6 in Müller et al. (2017). The results are mixed, with many regions for rice and wheat being difficult to model. No single model is dominant, with each

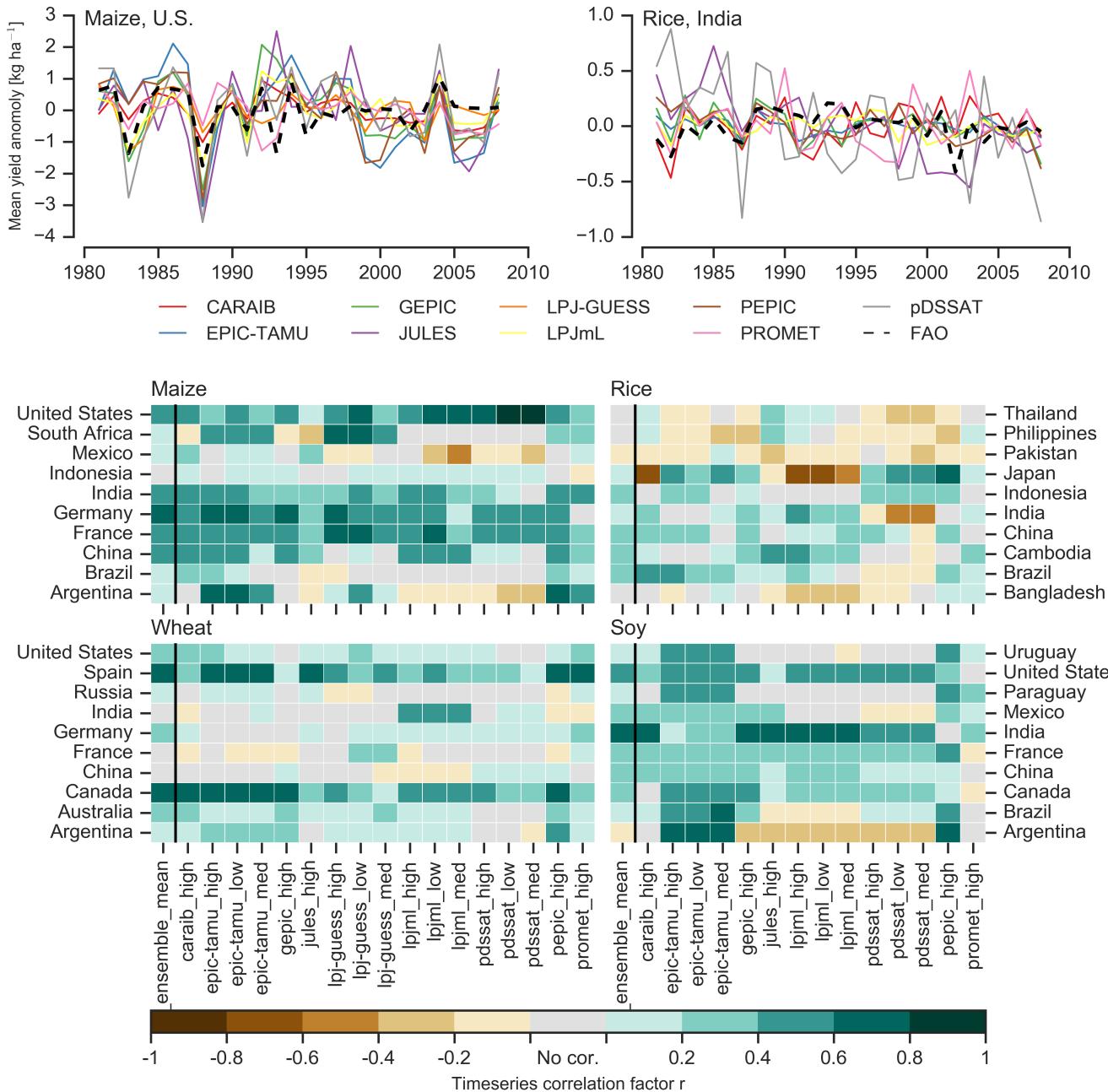


Figure 5: Time series correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen 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.

model providing near best-in-class performance in at least one location-crop combination. The presence of very few dark green color bars clearly illustrates the power of a multi-model intercomparison project like the one presented here. The ensemble mean does not beat the best model in each case, but

shows positive correlation in over 75% of the cases presented here. The EPIC-TAMU model performs best for soy, CARIAB, EPIC-TAMU, and PEPIC perform best for maize, PROMET performs best for wheat, and the EPIC family of models perform best for rice. Reductions in skill over the performance

503 illustrated in Müller et al. (2017) can be attributed to the nitro-⁵³⁷
504 gen levels or lack of calibration in some models.

505 *** or harmonization *** Christoph

506 Soy is qualitatively the easiest crop to represent (except in ⁵⁴⁰
507 Argentina), which is likely due in part to the invariance of the ⁵⁴¹
508 response to nitrogen application (soy fixes atmospheric nitrogen ⁵⁴²
509 very efficiently). Comparison to the FAO data is therefore easier ⁵⁴³
510 than the other crops because the nitrogen application levels do ⁵⁴⁴
511 not matter. US maize has the best performance across models, ⁵⁴⁵
512 with nearly every model representing the historical variability ⁵⁴⁶
513 to a reasonable extent. Especially good example years for US ⁵⁴⁷
514 maize are 1983, 1988, and 2004 (top left panel of Figure 5), ⁵⁴⁸
515 where every model gets the direction of the anomaly compared ⁵⁴⁹
516 to surrounding years correct. 1983 and 1988 are famously bad ⁵⁵⁰
517 years for US maize along with 2012 (not shown). US maize ⁵⁵¹
518 is possibly both the most uniformly industrialized (in terms of ⁵⁵²
519 management practices) crop and the one with the best data col- ⁵⁵³
520 lection in the historical period of all the cases presented here. ⁵⁵⁴

521 The FAO data is at least one level of abstraction from ground ⁵⁵⁵
522 truth in many cases, especially in developing countries. The ⁵⁵⁶
523 failure of models to represent the year-to-year variability in rice ⁵⁵⁷
524 in some countries in southeast Asia is likely partly due to model ⁵⁵⁸
525 failure and partly due to lack of data. It is possible to speculate ⁵⁵⁹
526 that the difference in performance between Pakistan (no suc- ⁵⁶⁰
527 cessful models) and India (many successful models) for rice ⁵⁶¹
528 may reside at least in part in the FAO data and not the mod- ⁵⁶²
529 els themselves. The same might apply to Bangladesh and In- ⁵⁶³
530 dia for rice. Partitioning of these contributions is impossible at ⁵⁶⁴
531 this stage. Additionally, there is less year-to-year variability in ⁵⁶⁵
532 rice yields (partially due to the fraction of irrigated cultivation). ⁵⁶⁶
533 Since the Pearson r metric is scale invariant, it will tend to score ⁵⁶⁷
534 the rice models more poorly than maize and soy. An example ⁵⁶⁸
535 of very poor performance can be seen with the pDSSAT model ⁵⁶⁹
536 for rice in India (top right panel of Figure 5).

3.3. Emulator performance

538 Emulation provides not only a computational tool but a
539 means of understanding and interpreting crop yield response
across the parameter space. Emulation is only possible, how-
ever, when crop yield responses are sufficiently smooth and
continuous to allow fitting with a relatively simple functional
form. In the GGCMI simulations, this condition largely but
not always holds. Responses are quite diverse across locations,
crops, and models, but in most cases local responses are reg-
ular enough to permit emulation. Figure 6 illustrates the geo-
graphic diversity of responses even in high-yield areas for a
single crop and model (rain-fed maize in pDSSAT for various
high-cultivation areas). This heterogeneity validates the choice
of emulating at the grid cell level.

540 Each panel in Figure 6 shows model yield output from sce-
541 narios varying only along a single dimension (CO_2 , tempera-
542 ture, precipitation, or nitrogen addition), with other inputs held
543 fixed at baseline levels; in all cases yields evolve smoothly
544 across the space sampled. For reference we show the results
545 of the full emulation fitted across the parameter space. The
546 polynomial fit readily captures the climatological response to
547 perturbations.

548 Crop yield responses generally follow similar functional
549 forms across models, though with a spread in magnitude. Fig-
550 ure 7 illustrates the inter-model diversity of yield responses
551 to the same perturbations, even for a single crop and location
552 (rain-fed maize in northern Iowa, the same location shown in
553 the Figure 6). The differences make it important to construct
554 emulators separately for each individual model, and the fidelity
555 of emulation can also differ across models. This figure illus-
556 trates a common phenomenon, that models differ more in re-
557 sponse to perturbations in CO_2 and nitrogen perturbations than
558 to those in temperature or precipitation. (Compare also Figures
559 3 and S18.) For this location and crop, CO_2 fertilization effects

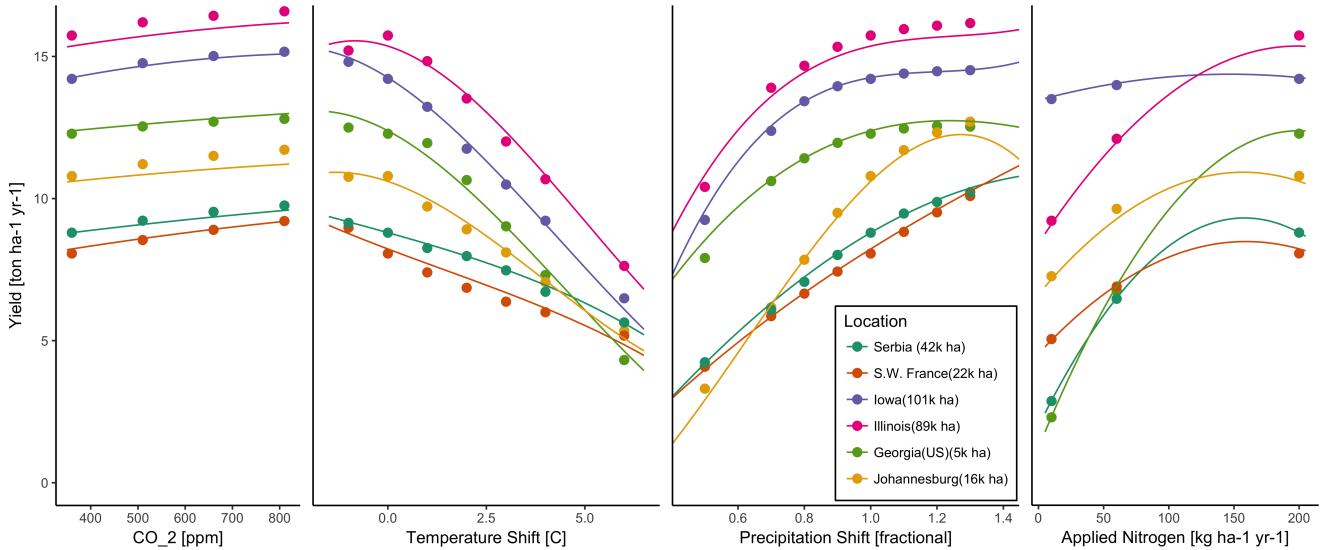


Figure 6: Example illustrating spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation along a single variable, with others held at baseline values. Dots show climatological mean yield, solid lines a simple polynomial fit across this 1D variation, and dashed lines the results of the full emulator of Equation 1. The climatological response is sufficiently smooth to be represented by a simple polynomial. Because precipitation changes are imposed as multiplicative factors rather than additive offsets, locations with higher baseline precipitation have a larger absolute spread in inputs sampled than do drier locations (not visible here).

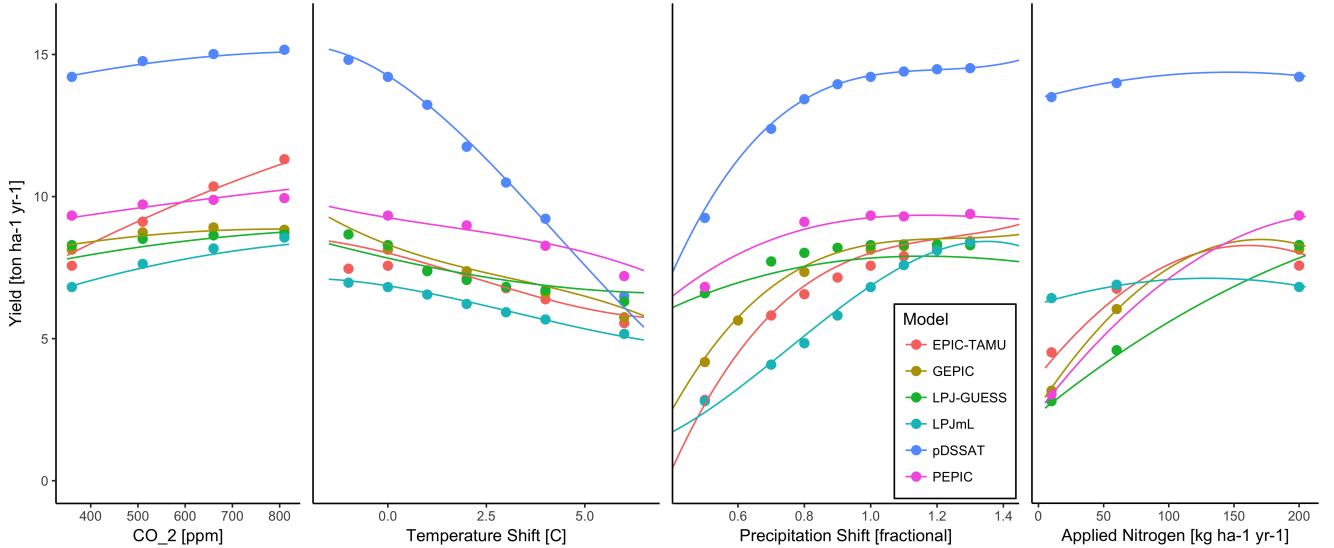


Figure 7: Example illustrating across-model variations in yield response. We show simulations and emulations from six models for rain-fed maize in the same Iowa location shown in Figure 6, with the same plot conventions. Models that did not simulate the nitrogen dimension are omitted for clarity. The inter-model standard deviation is larger for the perturbations in CO₂ and nitrogen than for those in temperature or precipitation illustrating that the sensitivity to weather is better constrained than to management at elevated CO₂. While most model responses can readily be emulated with a simple polynomial, some response surfaces are more complicated and lead to emulation error.

can range from ~5–50%, and nitrogen responses from nearly flat to a 60% drop in the lowest-application simulation.

While the nitrogen dimension is important and uncertain, it is also the most problematic to emulate in this work because of its limited sampling. The GGCMI 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 in potentially unphysical results. Steep and nonlinear declines in yield with lower nitrogen levels means that some regressions imply a peak in yield between the 100 and 200 kg N

581 $\text{y}^{-1} \text{ ha}^{-1}$ levels. While there may be some reason to believe
 582 over-application of nitrogen at the wrong time in the growing
 583 season could lead to reduced yields, these features are almost
 584 certainly an artifact of under sampling. In addition, the polyno-
 585 mial fit cannot capture the well-documented saturation effect
 586 of nitrogen application (e.g. Ingestad, 1977) as accurately as
 587 would be possible with a non-parametric model.

588 To assess the ability of the polynomial emulation to capture
 589 the behavior of complex process-based models, we evaluate the
 590 normalized emulator error. That is, for each grid cell, model,
 591 and scenario we evaluate the difference between the model yield
 592 and its emulation, normalized by the inter-model standard de-
 593 viation in yield projections. This metric implies that emulation
 594 is generally satisfactory, with several distinct exceptions. Al-
 595 most all model-crop combination emulators have normalized
 596 errors less than one over nearly all currently cultivated hectares
 597 (Figure 8), but some individual model-crop combinations are
 598 problematic (e.g. PROMET for rice and soy, JULES for soy
 599 and winter wheat, Figures S14–S15). Normalized errors for soy
 600 are somewhat higher across all models not because emulator fi-
 601 delity is worse but because models agree more closely on yield
 602 changes for soy than for other crops (see Figure S16, lower-
 603 ing the denominator). Emulator performance often degrades in
 604 geographic locations where crops are not currently cultivated.
 605 Figure 9 shows a CARAIB case as an example, where emulator
 606 performance is satisfactory over cultivated areas for all crops
 607 other than soy, but uncultivated regions show some problematic
 608 areas.

609 It should be noted that this assessment metric is relatively
 610 forgiving. First, each emulation is evaluated against the sim-
 611 ulation actually used to train the emulator. Had we used a spline
 612 interpolation the error would necessarily be zero. Second, the
 613 performance metric scales emulator fidelity not by the magni-
 614 tude of yield changes but by the inter-model spread in those

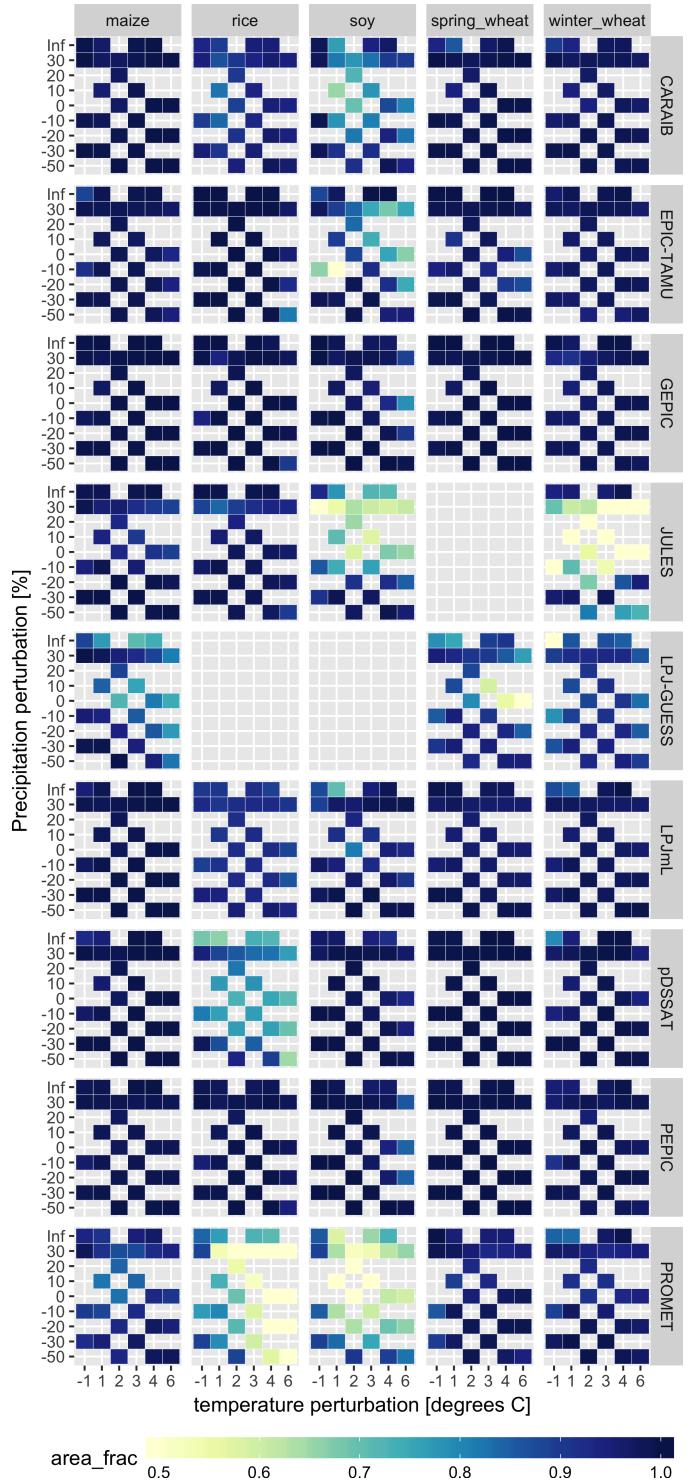


Figure 8: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO_2 and nitrogen held fixed at baseline values. Large columns are crops and large rows models; squares within are T,P scenario pairs. Colors denote the fraction of currently cultivated hectares with $e < 1$. Of the 756 scenarios with these CO_2 and N values, we consider only those for which all 9 models submitted data. 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.

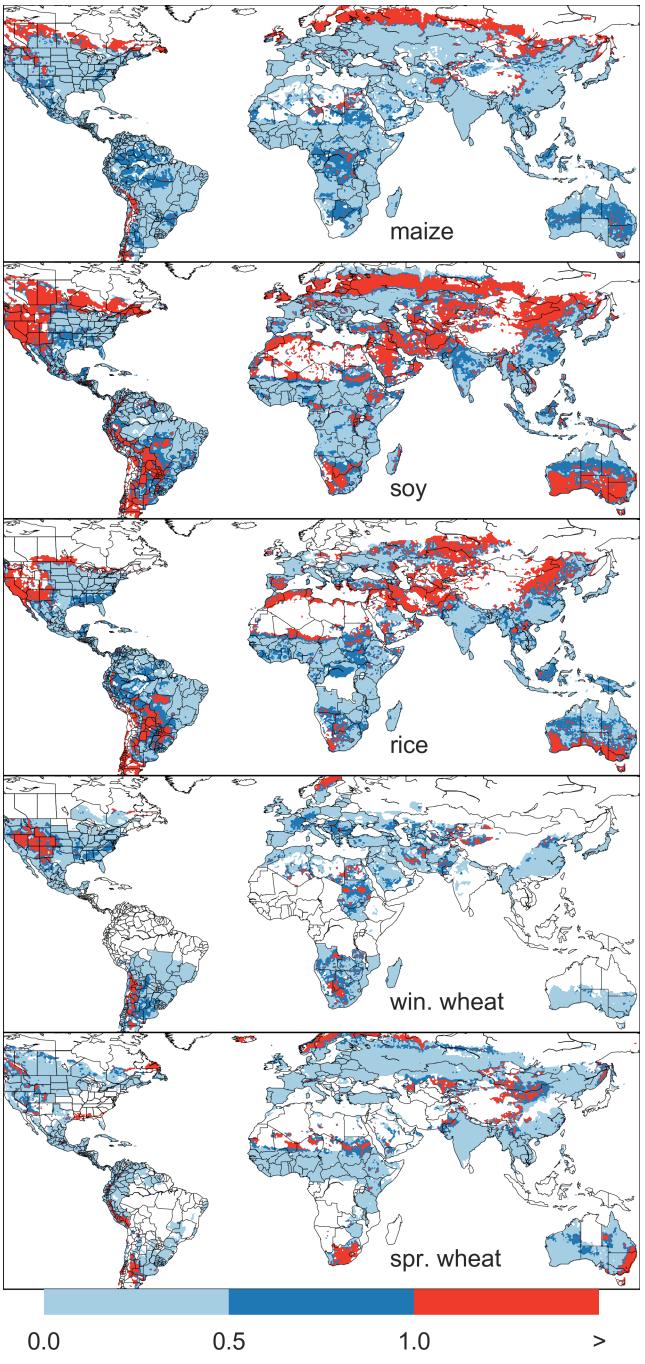


Figure 9: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error e , where $e > 1$ means that emulator error exceeds the multi-model standard deviation. White areas are those where crops 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 8.

changes. Where models differ more widely, the standard for emulators becomes less stringent. Because models disagree on the magnitude of CO₂ fertilization, this effect is readily seen when comparing assessments of emulator performance in simulations at baseline CO₂ (Figure 8) with those at higher CO₂ levels (Figure S13). Widening the inter-model spread leads to an apparent increase in emulator skill.

3.4. Emulator applications

Because 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. Emulators provide an easy way to compare a ensembles of climate or impacts projections. They also provide a means for generating continuous damage functions. As an example, we show a damage function constructed from 4D emulations for aggregated yield at the global scale, for maize on currently cultivated land, with simulated values shown for comparison. (Figure 10; see Figures S16- S19 in the supplemental material 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 simulations allows impacts modelers to apply arbitrary geographically-varying climate projections, as well as arbitrary aggregation mask, to develop damage functions for any climate scenario and any geopolitical or geographic level.

4. Conclusions and discussion

The GGCMI Phase II experiment assess sensitivities of process-based crop yield models to changing climate and management inputs, and was designed to allow not only comparison

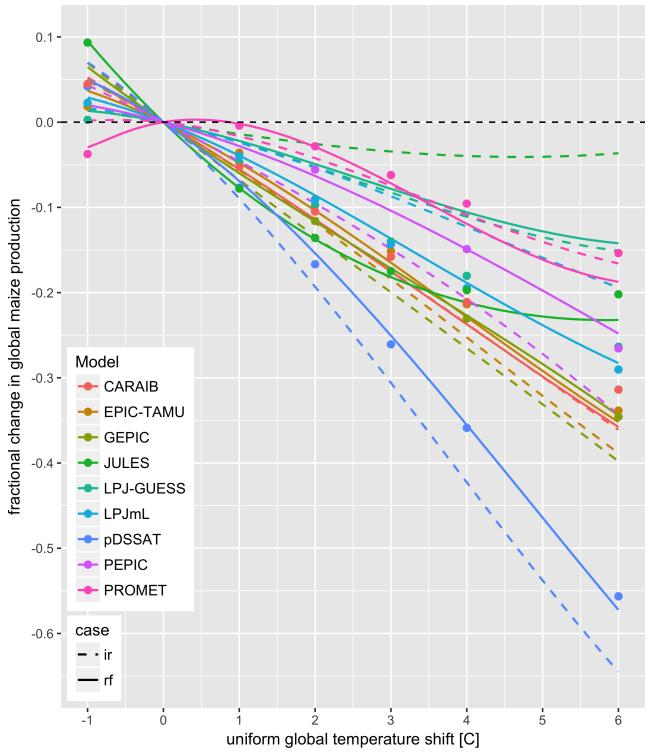


Figure 10: Global emulated damages for maize on currently cultivated lands⁶⁷⁵ for the GGCMI models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating up⁶⁷⁶ emulated values at the grid cell level, not from a regression of global mean yields.) Lines are emulations for rain-fed (solid) and irrigated (dashed) crops;⁶⁷⁷ 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 the⁶⁷⁸ locations of cultivated areas: irrigated crops tend to be grown in warmer areas where impacts are more severe for a given temperature shift. (The exceptions⁶⁷⁹ are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures S16-S19 in the supplemental material.

⁶⁶¹ mixed, with models performing better for maize and soy than
⁶⁶² for rice and wheat. The value of utilizing multiple models is
⁶⁶³ illustrated by the distribution in performance skill across differ-
⁶⁶⁴ ent countries and crops. An end-user of the simulation outputs
⁶⁶⁵ or emulator tool may pick and choose models based on histori-
⁶⁶⁶ cal skill to provide the most faithful temperature and precipita-
⁶⁶⁷ tion response depending on their application. The nitrogen and
⁶⁶⁸ CO₂ responses were not validated in this work.

⁶⁶⁹ One counterintuitive result is that irrigated maize shows
⁶⁷⁰ steeper yield reductions under warming than does rain-fed
⁶⁷¹ maize when considered only over currently cultivated land. The
⁶⁷² effect is the result of geographic differences in cultivated area.
⁶⁷³ In any given location, irrigation increases crop resiliency to
⁶⁷⁴ temperature increase, but irrigated maize is grown in warmer lo-
⁶⁷⁵cations where the impacts of warming are more severe (Figures S5-S6). The same behavior holds for rice and winter wheat,
⁶⁷⁶ but not for soy or spring wheat (Figures S8-S10). Irrigated wheat and maize are also more sensitive to nitrogen fertiliza-
⁶⁷⁷tion levels, presumably because growth in rain-fed crops is also
⁶⁷⁸ water-limited (Figure S19). (Soy as a nitrogen-fixing crop is relatively
⁶⁷⁹ insensitive to nitrogen, and rice is not generally grown in water-
⁶⁸⁰ limited conditions).

⁶⁴⁸ across models but evaluation of complex interactions between⁶⁸²
⁶⁴⁹ driving factors (CO₂, temperature, precipitation, and applied
⁶⁵⁰ nitrogen) and identification of geographic shifts in high yield
⁶⁵¹ potential locations. While the richness of the dataset invites
⁶⁵² further analysis, we show only a selection of insights derived
⁶⁵³ from the simulations. Across the major crops, inter-model un-
⁶⁵⁴certainty is greatest for wheat and least for soy. Across factors
⁶⁵⁵ impacting yields, inter-model-uncertainty is largest for CO₂ fer-
⁶⁵⁶tilization and nitrogen response effects. Across geographic re-
⁶⁵⁷gions, inter-model uncertainty is largest in the high latitudes
⁶⁵⁸where yields may increase, and model projections are most ro-
⁶⁵⁹bust in low latitudes where yield impacts are largest.

⁶⁶⁰ Model performance when compared to historical data is⁶⁹⁴

We show that emulation of the output of these complex re-
⁶⁸³sponses is possible even with a relatively simple reduced-form
⁶⁸⁴statistical model and a limited library of simulations. Emula-
⁶⁸⁵tion therefore offers the opportunity of producing rapid assess-
⁶⁸⁶ments of agricultural impacts for arbitrary climate scenarios in
⁶⁸⁷a computationally non-intensive way. The resulting tool should
⁶⁸⁸aid in impacts assessment, economic studies, and uncertainty
⁶⁸⁹analyses. Emulator parameter values also provide a useful way
⁶⁹⁰to compare sensitivities across models to different climate and
⁶⁹¹management inputs, and the terms in the polynomial fits offer
⁶⁹²the possibility of physical interpretation of these dependencies
⁶⁹³to some degree.

695 We provide this simulation output dataset for further analysis⁷²⁹
696 by the community as we have only scratched the surface with⁷³⁰
697 this work. Each simulation run includes year to year variabil-⁷³¹
698 ity in yields under different climate and management regimes.⁷³²
699 Some of the precipitation and temperature space has been lost⁷³³
700 due to the aggregation in the time dimension for the emula-⁷³⁴
701 tor presented here (i.e. the + 6 C simulation in the hottest year⁷³⁵
702 of the historical period compared to the coldest historical year,⁷³⁶
703 or precipitation perturbations in the driest historical year etc).⁷³⁷
704 Development of a year-to-year emulator or an emulator at dif-⁷³⁸
705 ferent spatial scales may provide useful for some IAM appli-⁷³⁹
706 cations. More exhaustive analysis of different statistical model⁷⁴⁰
707 specification for emulation will likely provide additional pre-⁷⁴¹
708 dictive skill over the specification provided here. The poten-⁷⁴²
709 tially richest area for further analysis is the interactions be-
710 tween input variable especially the Nitrogen and CO₂ interac-⁷⁴³

711 tions with weather and with each other. More robust quantifica-
712 tion of the sensitivity to the input drivers (and there differences
713 across models), as well as quantification in differences in un-
714 certainty across input drivers. Adaptation via growing season
715 changes were also simulated and are available in the database,
716 though this dimension was not presented or analyzed here. The
717 output dataset contains many other variables other than yield to
718 aid in analysis including above ground biomass, LAI, and root
719 biomass (as many as 25 output variables for some models).

720 The emulation approach presented here has some limitations.⁷⁵³
721 Because the GGCMI simulations apply uniform perturbations⁷⁵⁴
722 to historical climate inputs, they do not sample changes in⁷⁵⁵
723 higher order moments. The emulation therefore does not ad-⁷⁵⁶
724 dress the crop yield impacts of potential changes in climate⁷⁵⁷
725 variability. While some information could be extracted from⁷⁵⁸
726 consideration of year-over-year variability, more detailed sim-⁷⁵⁹
727 ulations and analysis are likely necessary to diagnose the im-⁷⁶⁰
728 pact of changes in variance and sub-growing-season tempo-⁷⁶¹

ral effects. Additionally, the emulator is intended to provide the change in yield from a historical mean baseline value and should be used in conjunction with historical data (or data products) or a historical mean emulator (not presented here).

The future of food security is one of the larger challenges facing humanity at present. The development (and emulation) of multi-model ensembles such as GGCMI Phase II provides a way to begin to quantify uncertainties in crop responses to a range of potential climate inputs and explore the potential benefits of adaptive responses. Emulation also allow making state-of-the-art simulation results available to a wide research community as simple, computationally tractable tools that can be used by downstream modelers to understand the socioeconomic impacts of crop response to climate change.

5. Acknowledgments

We thank Michael Stein and Kevin Schwarzwald, who provided helpful suggestions that contributed to this work. This research was performed as part of the Center for Robust Decision-Making on Climate and Energy Policy (RDCEP) at the University of Chicago, and was supported through a variety of sources. RDCEP is funded by NSF grant #SES-1463644 through the Decision Making Under Uncertainty program. J.F. was supported by the NSF NRT program, grant #DGE-1735359. C.M. was supported by the MACMIT project (01LN1317A) funded through the German Federal Ministry of Education and Research (BMBF). C.F. was supported by the European Research Council Synergy grant #ERC-2013-SynG-610028 Imbalance-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 Dynamics Research Program Area. Computing resources were provided by the University of Chicago Research Computing

- 762 Center (RCC). S.O. acknowledges support from the Swedish⁸⁰³
 763 strong research areas BECC and MERGE together with sup-⁸⁰⁴
 764 port from LUCCI (Lund University Centre for studies of Car-⁸⁰⁵
 765 bon Cycle and Climate Interactions).
- 806 Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.
 807 (2014). A meta-analysis of crop yield under climate change and adaptation.
Nature Climate Change, 4, 287 – 291.
- 808 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process
 809 emulation of dynamic computer codes. *Biometrika*, 96, 663–676.
- 810 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In
 811 C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth,*
Duke University, Durham, North Carolina (pp. 115–118). Durham, North
 812 Carolina.
- 813 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for
 814 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).
- 815 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdoos,
 816 M., & François, L. (2011). Responses of European forest ecosystems to
 817 21st century climate: assessing changes in interannual variability and fire
 818 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).
- 819 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,
 820 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating
 821 impact models and sectors (pSIMS). *Environmental Modelling and Soft-
 822 ware*, 62, 509–516.
- 823 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,
 824 Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C.,
 825 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.
 826 (2015). The Global Gridded Crop Model Intercomparison: data and mod-
 827 eling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8,
 828 261–277.
- 829 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J.,
 830 & Taylor, K. E. (2016). Overview of the coupled model intercomparison
 831 project phase 6 (cmip6) experimental design and organization. *Geoscientific
 832 Model Development*, 9, 1937–1958.
- 833 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-
 834 mate change impacts on durum wheat in the mediterranean region. *Natural
 835 Hazards and Earth System Sciences*, 11, 1293–1302.
- 836 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Re-
 837 gionalization of a large-scale crop growth model for sub-Saharan Africa:
 838 Model setup, evaluation, and estimation of maize yields. *Agriculture,
 839 Ecosystems & Environment*, 151, 21 – 33.
- 840 Food and Agriculture Organization of the United Nations (2018). FAOSTAT
 841 database. URL: <http://www.fao.org/faostat/en/home>.
- 842 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T.,
 843 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P.,
 844 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F.,
 845 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,
 846 Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J.,
 847 Based on Precomputed GCM Runs. *Journal of Climate*, 27, 1829–1844.

- Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nen-del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna, M., Semenov, M. A., Slawinski, C., Strattonovich, 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.

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. doi:10.5194/gmd-7-433-2014.

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., Dalgliesh, 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., Harsdorf, 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.

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). GEPIG - 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.

- 932 Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict⁹⁷⁵
 933 crop yield responses to climate change. *Agricultural and Forest Meteorology*, 150, 1443 – 1452.
 934
 935 Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relation⁹⁷⁸
 936 ships and the impacts of recent warming. *Environmental Research Letters*, 2, 014002.
 937
 938 MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447.
 939 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 of⁹⁸⁵
 940 large simulated yield datasets for studying climate effects. (p. 1100).⁹⁸⁶
 941 doi:10.13140/RG.2.1.5173.8328.
 942
 943 Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrolog⁹⁸⁸
 944 ical modelling to study the impact of climate change on the water flows of⁹⁸⁹
 945 mountain watersheds. *Journal of Hydrology*, 376, 362 – 377.
 946
 947 Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., & Calzadilla, A. (2009). Global biomass production potentials exceed ex⁹⁹¹
 948 pected future demand without the need for cropland expansion. *Nature Communications*, 6.⁹⁹³
 949
 950 McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srini⁹⁹⁵
 951 vasa, C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S.⁹⁹⁶
 952 (2015). Integrated assessments of the impacts of climate change on agricul⁹⁹⁷
 953 ture: An overview of AgMIP regional research in South Asia. *Chapter in: Handbook of Climate Change and Agroecosystems*, (pp. 201–218).⁹⁹⁸
 954
 955 Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical¹⁰⁰⁰
 956 weather responsiveness of crop yields: US evidence and implications for¹⁰⁰¹
 957 the agricultural impacts of climate change. *Environmental Research Letters*, 12.¹⁰⁰²
 958
 959 Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate¹⁰⁰⁴
 960 change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, 8.¹⁰⁰⁶
 961 Müller, C., Elliott, J., Chryssanthacopoulos, 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, 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, 1403–1422.¹⁰¹³
 962
 963 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, 789–798.¹⁰¹⁵
 964
 965 O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91, 1290 – 1300.
 966
 967 Olin, S., Schurges, G., Lindeskog, M., Wårliind, 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.
 968
 969 Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.
 970 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.
 971
 972 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.
 973
 974 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.
 975
 976 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas-
 977 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.
 978
 979 Pirttioja, N., Carter, T., Fronzek, S., Bindi, 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., Ferrié, 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., Strattonovitch, 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.
 980
 981 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.
 982
 983 Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of monthly growing areas of 26 irrigated crops.
 984

- 1018 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.
- 1019 1064
- 1020 1065
- 1021 1066
- 1022 1067
- 1023 1068
- 1024 1069
- 1025 1070
- 1026 1071
- 1027 1072
- 1028 1073
- 1029 1074
- 1030 1075
- 1031 1076
- 1032 1077
- 1033 1078
- 1034 1079
- 1035 1080
- 1036 1081
- 1037 1082
- 1038 1083
- 1039 1084
- 1040 1085
- 1041 1086
- 1042 1087
- 1043 1088
- 1044 1089
- 1045 1090
- 1046 1091
- 1047 1092
- 1048 1093
- 1049 1094
- 1050 1095
- 1051 1096
- 1052 1097
- 1053 1098
- 1054 1099
- 1055 1100
- 1056 1101
- 1057 1102
- 1058 1103
- 1059 1104
- 1060 1105
- (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. doi:10.1111/gcb.12412.
- 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.
- 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*, 2018, 1–42.
- 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.

- 1104 Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., &
1105 Martin, J. F. (2014). Modeling sugarcane yield with a process-based model
1106 from site to continental scale: Uncertainties arising from model structure
1107 and parameter values. *Geoscientific Model Development*, 7, 1225–1245.
- 1108 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe,
1109 J. (2014). The Inter-Sectoral Impact Model Intercomparison Project
1110 (ISI-MIP): Project framework. *Proceedings of the National Academy of
1111 Sciences*, 111, 3228–3232.
- 1112 White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Method-
1113 ologies for simulating impacts of climate change on crop production. *Field
1114 Crops Research*, 124, 357 – 368.
- 1115 Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T.,
1116 Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance
1117 against site observations of irrigated maize from Mead, Nebraska. *Geosci-
1118 entific Model Development*, 10, 1291–1320.
- 1119 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield vari-
1120 ability in JULES-crop and implications for forcing with seasonal weather
1121 forecasts. *Geoscientific Model Development*, 8, 3987–3997.
- 1122 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Land-
1123 bouwkundige Onderzoeken* : 64.6, .
- 1124 Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato
1125 yields on changes in climate and co2. *Agricultural and Forest Meteorology*,
1126 112, 217 – 231.
- 1127 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao,
1128 Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A.,
1129 Li, T., Lin, E., Liu, Q., Martre, P., Miller, C., Peng, S., Peuelas, J., Ruane,
1130 A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng,
1131 S. (2017). Temperature increase reduces global yields of major crops in four
1132 independent estimates. *Proc. Natl. Acad. Sci.*, 114, 9326–9331.