

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 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 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 Understanding crop yield response to a changing climate
3 is critically important, especially as the global food produc-
4 tion system will face pressure from increased demand over the
5 next century. Climate-related reductions in supply could there-
6 fore have severe socioeconomic consequences. Multiple stud-
7 ies using different crop or climate models concur in predicting
8 sharp yield reductionss on currently cultivated cropland under
9 business-as-usual climate scenarios, although their yield pro-
10 jections show considerable spread (e.g. Porter et al. (IPCC),
11 2014, Rosenzweig et al., 2014, Schauberger et al., 2017, and
12 references therein). Modeling crop responses continues to be
13 challenging, as crop growth is a function of complex interac-
14 tions 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 that attempt
17 to capture the relationship between input factors and resultant
18 yields (e.g. Heady, 1957, Heady & Dillon, 1961). These statis-
19 tical models were typically developed on a small scale for loca-
20 tions with extensive histories of yield data. The emergence of
21 electronic computers allowed development of numerical mod-
22 els that simulate the process of photosynthesis and the biology
23 and phenology of individual crops (first proposed by de Wit
24 (1957) and Duncan et al. (1967) and attempted by Duncan
25 (1972); for a history of crop model development see Rosen-
26 zweig et al. (2014)). A half-century of improvement in both
27 models and computing resources means that researchers can
28 now run crop simulations for many years at high spatial res-
29 olution on the global scale.

30 Both types of models continue to be used, and compara-
31 tive studies have concluded that when done carefully, both ap-
32 proaches can provide similar yield estimates (e.g. Lobell &

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

Process-based models do continue to struggle with some im-
portant 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 aca-
demic 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 rea-
sons, 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, An-
gulo et al., 2013, Asseng et al., 2013, 2015). Inter-model dis-
crepancies 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 di-
rectly into an integrated assessment model to estimate the po-
tential cost of climate change to the agricultural sector.

Nevertheless, process-based models are necessary for under-
standing the global future yield impacts of climate change for
many reasons. First, cultivation may shift to new areas, where
no yield data are currently available and therefore statistical
models cannot apply. Yield data are also often limited in the

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

67 developing world, where future climate impacts may be the¹⁰¹
68 most critical. Finally, only process-based models can capture¹⁰²
69 the growth response to novel conditions and practices that are¹⁰³
70 not represented in historical data (e.g. Pugh et al., 2016, Roberts¹⁰⁴
71 et al., 2017). These novel changes can include the direct fertil-¹⁰⁵
72 ization effect of elevated CO₂, or changes in management prac-¹⁰⁶
73 tices that may ameliorate climate-induced damages.¹⁰⁷

74 Interest has been rising in statistical emulation, which al-¹⁰⁸
75 lows combining advantageous features of both statistical and¹⁰⁹
76 process-based models. The approach involves constructing a¹¹⁰
77 statistical representation or “surrogate model” of complicated¹¹¹
78 numerical simulations by using simulation output as the train-¹¹²
79 ing data for a statistical model (e.g. O’Hagan, 2006, Conti et al.,¹¹³
80 2009). Emulation is particularly useful in cases where sim-¹¹⁴
81 ulations are complex and output data volumes are large, and¹¹⁵
82 has been used in a variety of fields, including hydrology (e.g.¹¹⁶
83 Razavi et al., 2012), engineering (e.g. Storlie et al., 2009),¹¹⁷
84 environmental sciences (e.g. Ratto et al., 2012), and climate¹¹⁸
85 (e.g. Castruccio et al., 2014, Holden et al., 2014). For agri-¹¹⁹
86 cultural impacts studies, emulation of process-based models¹²⁰
87 allows capturing key relationships between input variables in¹²¹
88 a lightweight, flexible form that is compatible with economic¹²²
89 studies.¹²³

90 In the past decade, multiple studies have developed emula-¹²⁴
91 tors of process-based crop simulations. Early studies proposing¹²⁵
92 or describing potential crop yield emulators include Howden¹²⁶
93 & Crimp (2005), Räisänen & Ruokolainen (2006), Lobell &¹²⁷
94 Burke (2010), and Ferrise et al. (2011), who used a machine¹²⁸
95 learning approach to predict Mediterranean wheat yields. Stud-¹²⁹
96 ies developing single-model emulators include Holzkämper¹³⁰
97 et al. (2012) for the CropSyst model, Ruane et al. (2013) for¹³¹
98 the CERES wheat model, and Oyebamiji et al. (2015) for the¹³²
99 LPJmL model (for multiple crops, using multiple scenarios as¹³³
100 a training set). More recently, emulators have begun to be used¹³⁴

in the context of multi-model intercomparisons, with Blanc & Sultan (2015), Blanc (2017), Ostberg et al. (2018) and Mistry 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 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) attempt only to capture 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

135 yield responses at annual vs. climatological timescales.

136 Systematic parameter sweeps have begun to be used in
137 crop modeling, with early efforts in 2015 (Makowski et al.,
138 2015, Pirttioja et al., 2015), and several recent studies in 2018
139 (Fronzek et al., 2018, Snyder et al., 2018). Both Fronzek
140 et al. (2018) and Snyder et al. (2018) sample multiple perturba-
141 tions to temperature and precipitation (with Snyder et al. (2018)
142 adding CO₂ as well), in 132 and 99 different combinations, re-
143 spectively, and both take advantage of the structured training
144 set to construct emulators of climatological mean yields, omit-
145 ting year-over-year variations. Both studies are limited in other
146 aspects, however: both simulate only a handful of individual
147 sites, and Fronzek et al. (2018) simulates only wheat (over sev-
148 eral models), while Snyder et al. (2018) analyzes four crops but
149 with a single model (GCAM).

150 In this paper we describe a new comprehensive dataset de-
151 signed to expand the parameter sweep approach still further.
152 The Global Gridded Crop Model Intercomparison (GGCMI)
153 Phase II experiment involves running a suite of process-based
154 crop models across historical conditions perturbed by a set of
155 discrete steps in different input parameters, including an ap-
156 plied nitrogen dimension. The experimental protocol involves
157 over 700 different parameter combinations for each model and
158 crop, with simulations providing near-global coverage at a half
159 degree spatial resolution. The experiment was conducted as
160 part of the Agricultural Model Intercomparison and Improve-
161 ment Project (AgMIP) (Rosenzweig et al., 2013, 2014), an in-
162 ternational effort conducted under a framework similar to the
163 Climate Model Intercomparison Project (CMIP) (Taylor et al.,
164 2012, Eyring et al., 2016). The GGCMI protocol builds on the
165 AgMIP Coordinated Climate-Crop Modeling Project (C3MP)
166 (Ruane et al., 2014, McDermid et al., 2015) and will con-
167 tribute to the AgMIP Coordinated Global and Regional As-
168 sessments (CGRA) (Ruane et al., 2018, Rosenzweig et al.,

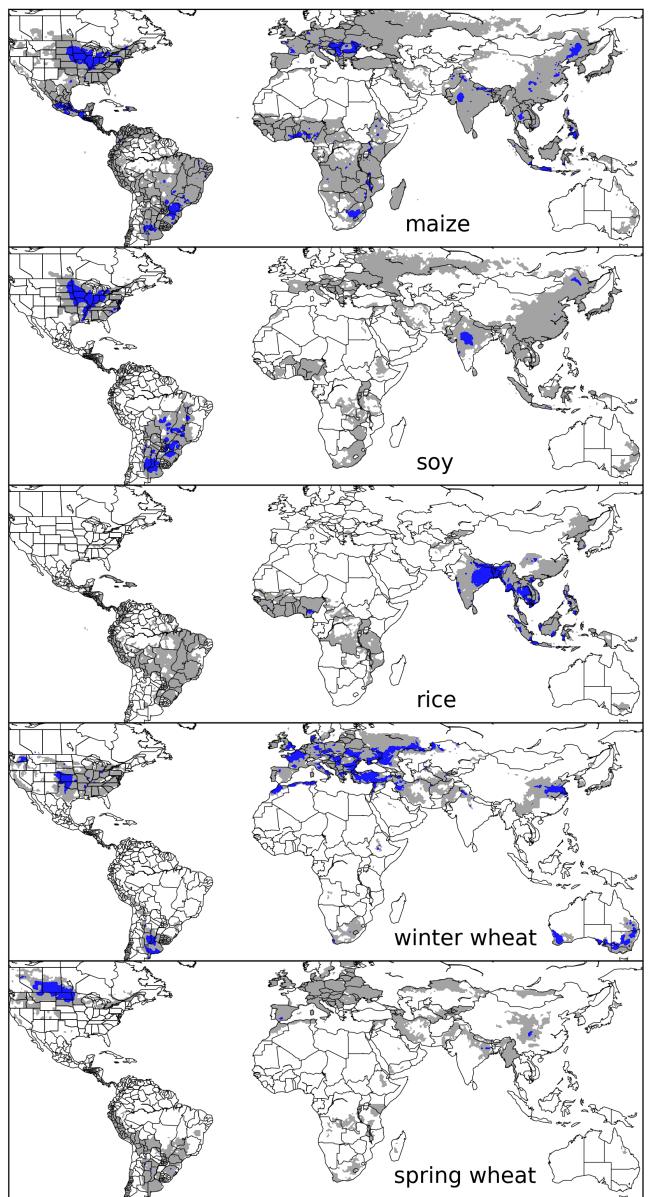


Figure 1: Presently cultivated area for rain-fed crops. Blue indicates grid cells with more than 20,000 hectares (~10% of the equatorial grid cells) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated areas for maize, rice, and soy are taken from the MIRCA2000 (“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.

169 2018). GGCMI Phase II is designed to allow addressing goals
170 such as understanding where highest-yield regions may shift
171 under climate change; exploring future adaptive management
172 strategies; understanding how interacting input drivers affect
173 crop yield; quantifying uncertainties across models and major
174 drivers; and testing strategies for producing lightweight em-

ulators of process-based models. In this paper, we describe₂₀₁
 the GGCMI Phase II experiments, present initial results, and₂₀₂
 demonstrate that it is tractable to emulation.₂₀₃

2. Simulation – Methods

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

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”). The dataset is designed to allow researchers to:

- Enhance understanding of how models work by characterizing their sensitivity to input climate and nitrogen drivers.

- Study the interactions between climate variables and nitrogen inputs in driving modeled yield impacts.
- 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 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 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 data set captures a distribution of crop responses over the po-

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.

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.

228 potential space of future climate conditions.

229 The 12 models included in GGCMI Phase II are all mech-
230 anistic process-based crop models that are widely used in im-
231 pacts assessments (Table 2). Although some models share a
232 common base (e.g. LPJmL and LPJ-GUESS and the EPIC mod-
233 els), they have subsequently developed independently. (For
234 more details on model genealogy, see Figure S1 in Rosenzweig
235 et al. (2014).) Differences in model structure mean that sev-
236 eral key factors are not standardized across the experiment, in-
237 cluding secondary soil nutrients, carry-over effects across grow-
238 ing years including residue management and soil moisture, and
239 the extent of simulated area for different crops. Growing sea-
240 sons are standardized across models (with assumptions based
241 on Sacks et al. (2010), Portmann et al. (2008, 2010)), but vary
242 by crop and by location on the globe. For example, XX exam-
243 ple here. All stresses are disabled other than factors related to
244 nitrogen, temperature, and water (e.g. alkalinity and salinity).
245 No additional nitrogen inputs, such as atmospheric deposition,

246 are considered, but some model treatment of soil organic matter
247 may allow additional nitrogen release through mineralization.
248 See Rosenzweig et al. (2014), Elliott et al. (2015) and Müller
249 et al. (2017) for further details on models and underlying as-
250 sumptions.

The participating modeling groups provide simulations at
any of four initially specified levels of participation, so the num-
ber 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.

Each model is run at 0.5 degree spatial resolution and cov-
ers all currently cultivated areas and much of the uncultivated
land area. (See Figure 1 for the present-day cultivated area of
rain-fed crops, and Figure S1 in the Supplemental Material for
irrigated crops.) Coverage extends considerably outside cur-

264 recently cultivated areas because cultivation will likely shift under
 265 climate change. However, areas are not simulated if they are
 266 assumed to remain non-arable even under an extreme climate
 267 change; these regions include Greenland, far-northern Canada,
 268 Siberia, Antarctica, the Gobi and Sahara Deserts, and central
 269 Australia.

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

3. Simulation – Results

Crop models in the GGCM ensemble show broadly consistent responses to climate and management perturbations in most regions, with a strong negative impact of increased temperature in all but the coldest regions. We illustrate this result for rain-fed maize in Figure 2, which shows yields for the primary Köppen-Geiger climate regions (Rubel & Kottek, 2010). In warming scenarios, models show decreases in maize yield in the temperate, 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 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 ~25% of yields with a signal to noise of nearly three. A notable exception is the cold continental region, where models disagree strongly, extending even to the sign of impacts.

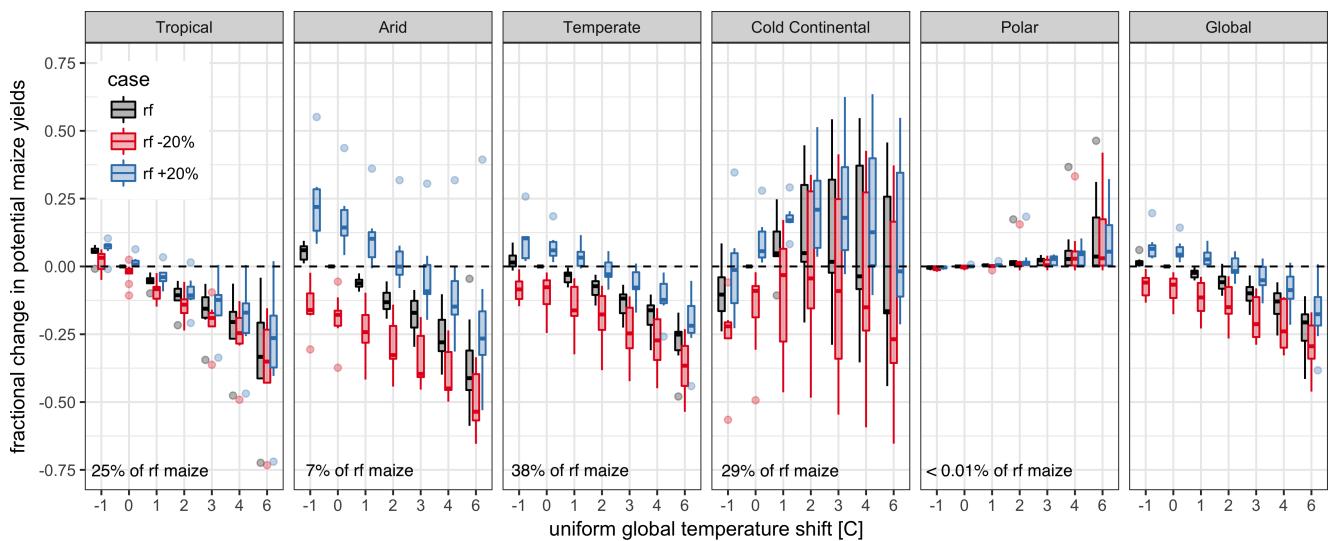


Figure 2: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation 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 (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 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.

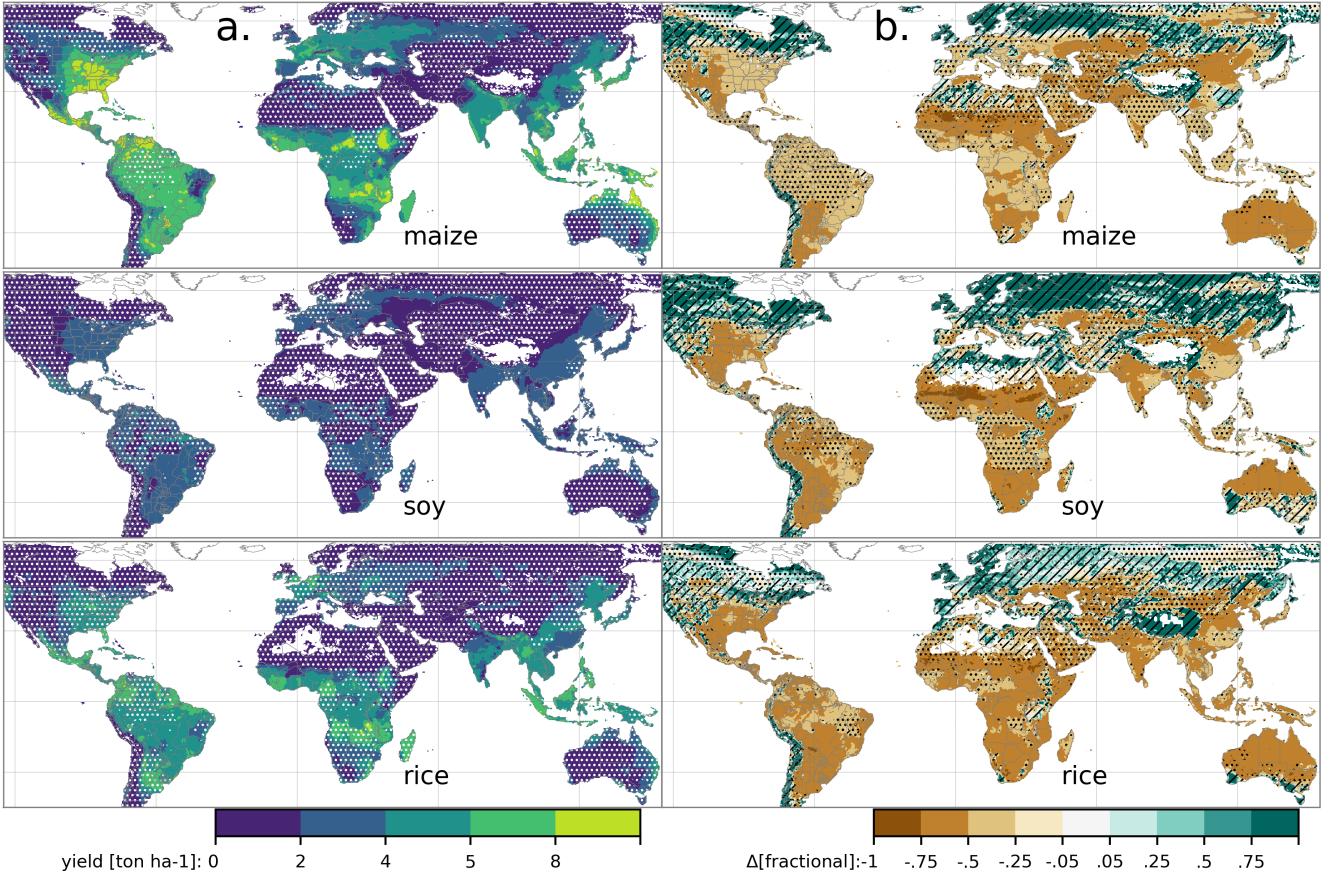


Figure 3: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure 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.

295 Other crops show similar responses to warming, with robust₃₀₈
 296 yield losses in warmer locations and high inter-model variance₃₀₉
 297 in the cold continental regions (Figure S7).

298 The effects of rainfall changes on maize yields shown in Fig-₃₁₁
 299 ure 2 are also as expected and are consistent across models.₃₁₂
 300 Increased rainfall mitigates the negative effect of higher tem-₃₁₃
 301 peratures, most strongly in arid regions. Decreased rainfall₃₁₄
 302 amplifies yield losses and also increases inter-model variance₃₁₅
 303 more strongly, suggesting that models have difficulty represent-₃₁₆
 304 ing crop response to water stress XX - see reviewer comments?₃₁₇
 305 We show only rain-fed maize here; see Figure S5 for the irri-₃₁₈
 306 gated case. As expected, irrigated crops are more resilient to₃₁₉
 307 temperature increases in all regions, especially so where water₃₂₀

is limiting.

Mapping the distribution of baseline yields and yield changes₃₁₀ 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 given crop. 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 both

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

323 4. Emulation – Methods

324 As part of our demonstration of the properties of the GGCMI
 325 Phase II dataset, we construct an emulator of 30-year climatic
 326 logical mean yields. This approach is made possible by
 327 the structured set of simulations involving systematic pertur-
 328 bations. In the GGCMI Phase II dataset, the year-over-year re-
 329 sponds are generally quantitatively distinct from (and larger
 330 than) climatological mean responses. In the example of Figure
 331 4, responses to year-over-year temperature variations are **XX%**
 332 larger than those to long-term perturbations in the baseline case,
 333 and larger still under warmer conditions, rising to **XX%** in the
 334 T+6 case. The stronger year-over-year response under warmer
 335 conditions also manifests as a wider distribution of yields (Fig-
 336 ure 5). As discussed previously, year-over-year and climatolog-
 337 ical responses can differ for many reasons including memory
 338 in the crop model, lurking covariants, and differing associated
 339 distributions of daily growing-season daily weather (e.g. Ruane
 340 et al., 2016). Note that the GGCMI Phase II datasets do not
 341 capture one climatological factor, potential future distributional
 342 shifts, because all simulations are run with fixed offsets from
 343 the historical climatology. Prior work has suggested that mean
 344 changes are the dominant drivers of climatological crop yield
 345 shifts in non-arid regions (e.g. Glotter et al., 2014).

346 Emulation involves fitting individual regression models for
 347 each crop, simulation model, and 0.5 degree geographic pixel
 348 from the GGCMI Phase II data set; the regressors are the ap-
 349 plied constant perturbations in CO₂, temperature, water, and
 350 nitrogen (C,T, W, N). We regress 30-year climatological mean
 351 yields against a third-order polynomial in C, T, W, and N with
 352 interaction terms. The higher-order terms are necessary to
 353 capture any nonlinear responses, which are well-documented

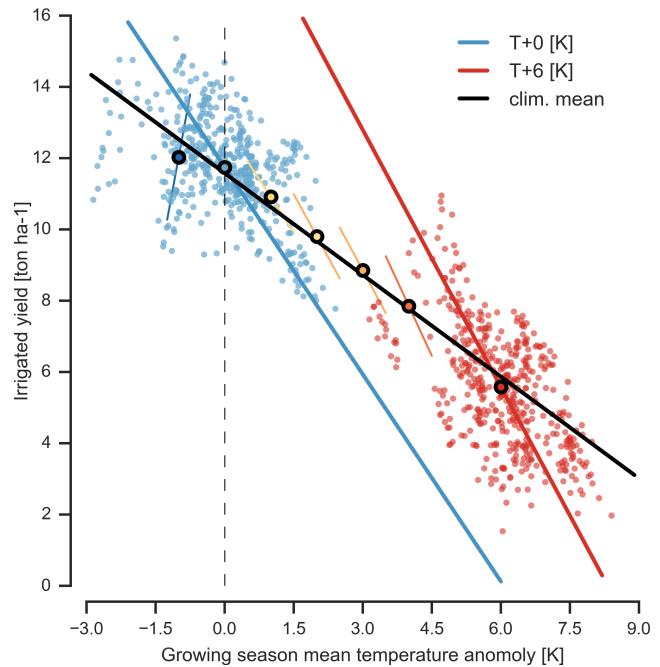


Figure 4: Example showing distinction between crop yield responses to year-to-year and climatological mean temperature shifts. Figure shows irrigated maize for a representative high-yield region (nine adjacent grid cells in northern Iowa) from the pDSSAT model, for the baseline 1981-2010 historical climate (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 **XX–XX%** larger than those to long-term climate perturbations, rising under warmer conditions.

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 overfitting 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.

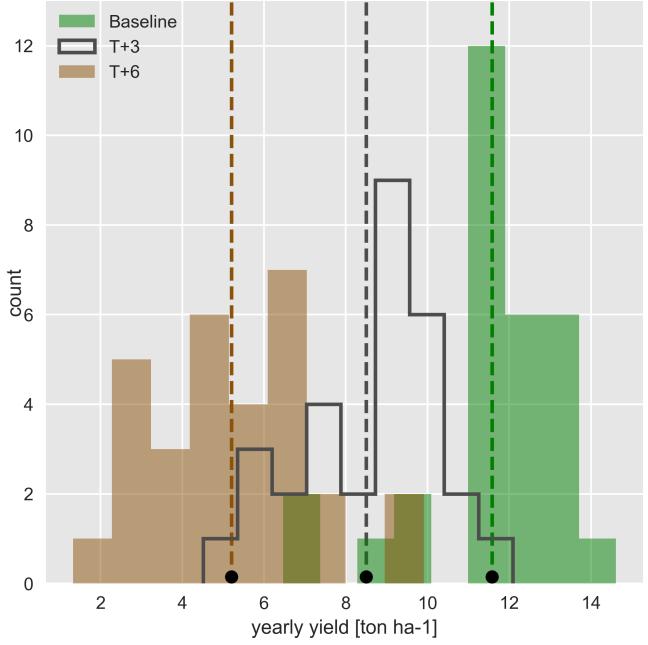


Figure 5: Example showing climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for one of the high-yield grid cells of Figure 4 (in northern Iowa) from the pDSSAT model, for the baseline 1981–2010 historical climate (green) and for scenarios with temperature shifted by T+3 (black) and T+6 K (brown), 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. In this work we emulate not the year-over-year distributions but the climatological mean response (dashed vertical lines / black dots).

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 simpler functional form used here allows emulation at grid-cell level **with low noise? how do you quantify this?** The emulation therefore indirectly includes any yield response to geographically distributed factors such as soil type, insolation, and the baseline climate itself.

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 model; we then follow the reduction of 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 polynomial in 23 terms, with 11 terms eliminated. We omit the N^3 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 crops:

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

To fit the parameters K , we use a Bayesian Ridge proba-

396 bilistic estimator (MacKay, 1991), which reduces volatility in⁴¹⁵
 397 parameter estimates when the sampling is sparse, by weight-
 398 ing parameter estimates towards zero. The Bayesian Ridge⁴¹⁶
 399 method is necessary to maintain a consistent functional form⁴¹⁷
 400 across all models and locations. We use the implementation of⁴¹⁸
 401 the Bayesian Ridge estimator from the scikit-learn package in⁴¹⁹
 402 Python (Pedregosa et al., 2011). In the GGCMI Phase II ex-⁴²⁰
 403 periment, the most problematic fits are those for models that⁴²¹
 404 provided a limited number of cases or for low-yield geographic⁴²²
 405 regions where some modeling groups did not run all scenar-⁴²³
 406 ios. We do not attempt to emulate models that provided less⁴²⁴
 407 than 50 simulations. The lowest number of simulations emu-⁴²⁵
 408 lated across the full parameter space is then 130 (for the PEPIC⁴²⁶
 409 model). The resulting parameter matrices for all crop model⁴²⁷
 410 emulators are available on request [give location?](#), as are the raw⁴²⁸
 411 simulation data and a Python application to emulate yields. The⁴²⁹
 412 yield output for a single GGCMI model that simulates all sce-⁴³⁰
 413 narios and all five crops is ~ 12.5 GB; the emulator is ~ 100 MB,⁴³¹
 414 a reduction by over two orders of magnitude.⁴³²

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 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 [XX](#).

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 (CO_2 , temperature, precipitation, or nitrogen addition), with other inputs held fixed at baseline levels, compared

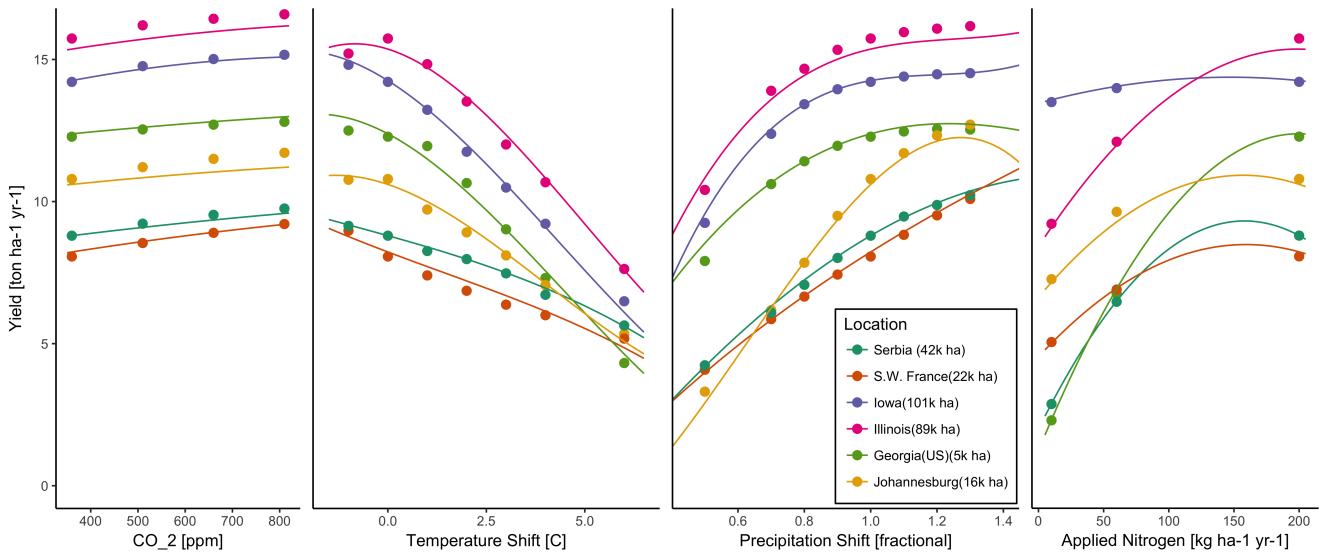


Figure 6: Illustration of spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation 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 [XX](#).

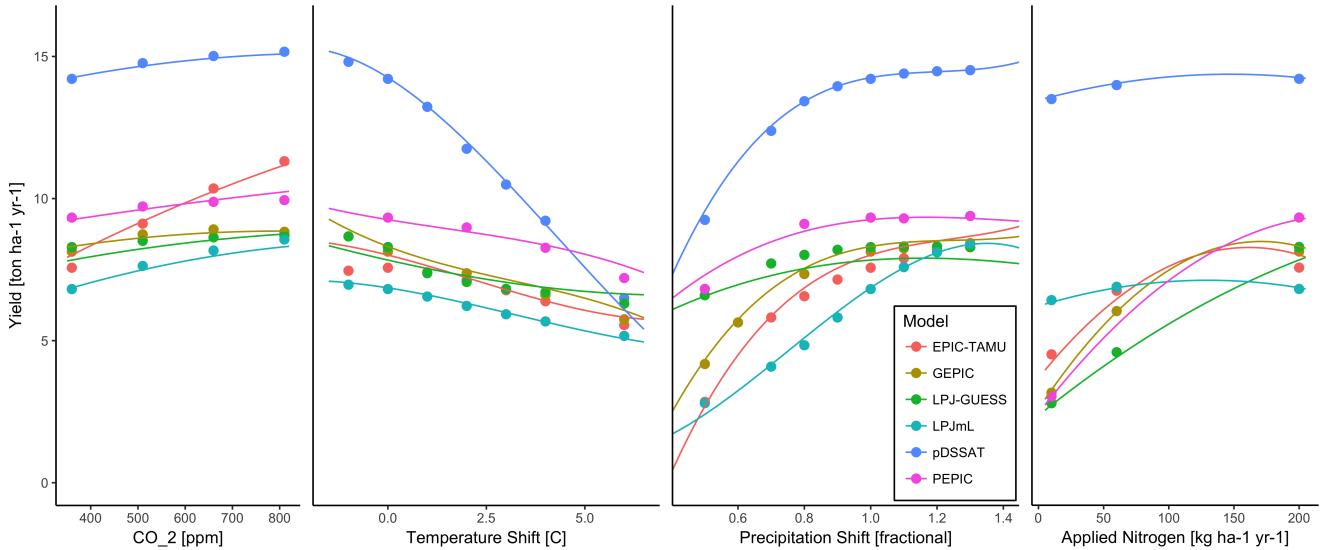


Figure 7: Illustration of across-model variations in yield response. Figures shows simulations and emulations from six models for rain-fed maize in the same Iowa grid cell shown in Figure 6, with the same plot conventions. Models that do not simulate the nitrogen dimension are omitted for clarity. Note 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 XX. As in Figure 6, extrapolation out of the sample space is problematic.

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, 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. For this location and crop, CO₂ fertilization effects 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, 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 mean that some regressions imply a peak in yield between the 100 and 200 kg N y⁻¹ 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 almost certainly an artifact of undersampling. 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 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

474 dimensions). The emulated values closely match simulations⁴⁸⁴
 475 even at this aggregation level. Note that these functions are pre-⁴⁸⁵
 476 sented only as examples and do not represent true global pro-⁴⁸⁶
 477 jections, because they are developed from simulation data with⁴⁸⁷
 478 a uniform temperature shift while increases in global mean tem-⁴⁸⁸
 479 perature should manifest non-uniformly. The global coverage⁴⁸⁹
 480 of the GGCMI simulations allows impacts modelers to apply⁴⁹⁰
 481 arbitrary geographically-varying climate projections, as well as⁴⁹¹
 482 arbitrary aggregation masks, to develop damage functions for⁴⁹²
 483 any climate scenario and any geopolitical or geographic level.⁴⁹³

6. Conclusions and discussion

The GGCMI Phase II experiment provides a database tar-
 geted to allow detailed study of crop yields from process-based
 models under climate change. The experiment is 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 major areas of uncertainty. Across the major crops, inter-model uncertainty is greatest for wheat and least for soy. Across factors impacting yields, inter-model uncertainty is largest for CO_2 fertilization and nitrogen response effects. Across geographic regions, projections are most uncertain in the high latitudes where yields may increase, and most robust in low latitudes where yield impacts are largest.

Second, the GGCMI Phase II simulations allow understanding the way that climate-driven changes and locations of cultivated land combine to produce yield impacts. One counterintuitive result immediate apparent is that irrigated maize shows steeper yield reductions under warming than does rain-fed maize when considered only over currently cultivated land.

The effect results from geographic differences in cultivation. In any given location, irrigation increases crop resiliency to temperature increase, but irrigated maize is grown in warmer locations where the impacts of warming are more severe (Figures S5–S6). The same behavior holds for rice and winter wheat, but not for soy or spring wheat (Figures S8–S10). Irrigated wheat and maize are also more sensitive to nitrogen fertilization levels than are analogous non-irrigated crops, presumably

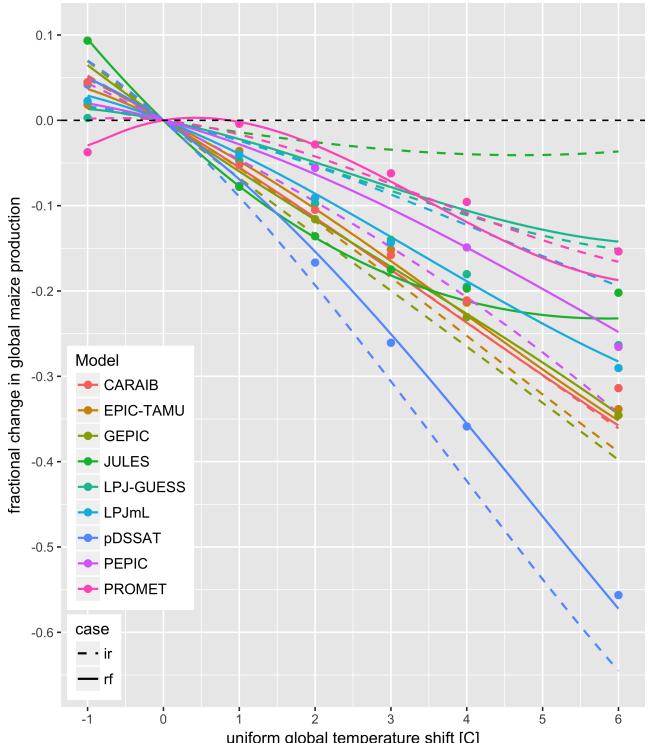


Figure 8: Global emulated damages for maize on currently cultivated lands for the GGCMI models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating up emulated values at the grid cell level, not from a regression of global mean 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.

518 because those rain-fed crops are limited by water as well as₅₅₂
519 nitrogen availability (Figure S19). (Soy as an efficient atmo-₅₅₃
520 spheric nitrogen-fixer is relatively insensitive to nitrogen, and₅₅₄
521 rice is not generally grown in water-limited conditions).

522 Third, we show that even the relatively limited GGCMI₅₅₆
523 Phase II sampling space allows emulation of the climatologi-
524 cal response of crop models with a relatively simple reduced-₅₅₇
525 form statistical model. The systematic parameter sampling in₅₅₈
526 the GGCMI Phase II procedure provides information on the in-₅₅₉
527 fluence of multiple interacting factors in a way that single pro-₅₆₀
528 jections cannot, and emulating the resulting response surface₅₆₁
529 then produces a tool that can aid in both physical interpretation₅₆₂
530 of the process-based models and in assessment of agricultural₅₆₃
531 impacts under arbitrary climate scenarios. Emulating the cli-₅₆₄
532 matological response isolates long-term impacts from any con-₅₆₅
533 founding factors that complicate year-over-year changes, and₅₆₆
534 the use of simple functional forms offer the possibility of phys-₅₆₇
535 ical interpretation of parameter values. Care should be taken in₅₆₈
536 applying relationships developed at the yearly level to shifts in₅₆₉
537 the mean climatology. We anticipate that systematic parameter₅₇₀
538 sampling will become the norm in future model intercompari-₅₇₁
539 son exercise.

540 While the GGCMI Phase II database should offer the foun-₅₇₃
541 dation for multiple future studies, several cautions need to be
542 noted. Because the simulation protocol was designed to fo-₅₇₅
543 cus on change in yield under climate perturbations and not₅₇₆
544 on replicating real-world yields, the models are not formally
545 calibrated so cannot be used for impacts projections unless in₅₇₇
546 used in conjunction with historical data (or data products). Be-₅₇₈
547 cause the GGCMI simulations apply uniform perturbations to₅₇₉
548 historical climate inputs, they do not sample changes in higher₅₈₀
549 order moments, and cannot address the additional crop yield₅₈₁
550 impacts of potential changes in climate variability. Although₅₈₂
551 distributional changes in model projections are fairly uncertain₅₈₃

555 at present, follow-on experiments may wish to consider them.
Several recent studies have described procedures for generating
simulations that combine historical data with model projections
of not only mean changes in temperature and precipitation but
changes in their marginal distributions or temporal dependence.

The GGCMI phase II output dataset invites a broad range of potential future avenues of analysis. A major target area involves studying the models themselves with a detailed examination of interaction terms between the major input drivers, a more robust quantification of the sensitivity of different models to the input drivers, and comparisons with field-level experimental data. The parameter space tested in GGCMI phase II will allow detailed investigations into yield variability and response to extremes under changing management and CO₂ levels. As mentioned previously, the database allows study of geographic shifts in optimal growing regions for different crops and studying the viability of switching crop types in some areas. The output dataset also contains other runs and variables not analyzed or shown here. Runs include several which allowed adaptation to climate changes by altering growing seasons, and additional variables include above ground biomass, LAI, and root biomass (as many as 25 output variables for some models). Emulation studies that are possible include a more systematic evaluation of different statistical model specifications and formal calculation of uncertainties in derived parameters.

The future of food security is one of the larger challenges facing humanity at present. The development of multi-model ensembles such as GGCMI Phase II provides a way to begin to better understand crop responses to a range of potential climate inputs, improve process based models, and explore the potential benefits of adaptive responses included shifting growing season, cultivar types and cultivar geographic extent.

584 **7. Acknowledgments**

585 We thank Michael Stein and Kevin Schwarzwald, who pro-
586 vided helpful suggestions that contributed to this work. This re-
587 search was performed as part of the Center for Robust Decision-
588 Making on Climate and Energy Policy (RDCEP) at the Univer-
589 sity of Chicago, and was supported through a variety of sources.
590 RDCEP is funded by NSF grant #SES-1463644 through the
591 Decision Making Under Uncertainty program. J.F. was sup-
592 ported by the NSF NRT program, grant #DGE-1735359. C.M.
593 was supported by the MACMIT project (01LN1317A) funded
594 through the German Federal Ministry of Education and Re-
595 search (BMBF). C.F. was supported by the European Research
596 Council Synergy grant #ERC-2013-SynG-610028 Imbalance-
597 P. P.F. and K.W. were supported by the Newton Fund through
598 the Met Office Climate Science for Service Partnership Brazil
599 (CSSP Brazil). A.S. was supported by the Office of Science
600 of the U.S. Department of Energy as part of the Multi-sector
601 Dynamics Research Program Area. Computing resources were
602 provided by the University of Chicago Research Computing
603 Center (RCC). S.O. acknowledges support from the Swedish
604 strong research areas BECC and MERGE together with sup-
605 port from LUCCI (Lund University Centre for studies of Car-
606 bon Cycle and Climate Interactions).

607 **8. Appendix: Simulations – Assessment**

608 The GGCMI Phase II simulations are designed for evaluat-
609 ing changes in yield but not absolute yields, since they omit
610 detailed calibrations. To provide some validation of the skill of
611 the process-based models used, we repeat the validation exer-
612 cises of Müller et al. (2017) for GGCMI Phase I. The Müller
613 et al. (2017) procedure evaluates response to year-to-year tem-
614 perature and precipitation variations in a control run driven by
615 historical climate and compares it to detrended historical yields
616 from the FAO (Food and Agriculture Organization of the United

617 Nations, 2018) by calculating the Pearson correlation coeffi-
618 cient. The procedure offers no means of assessing CO₂ fertiliza-
619 tion, since CO₂ has been relatively constant over the historical
620 data collection period. Nitrogen introduces some uncertainty
621 into the analysis, since the GGCMI Phase II runs impose fixed,
622 uniform nitrogen application levels that are not realistic for in-
623 dividual countries. We evaluate up to three control runs for each
624 model, since some modeling groups provide historical runs for
625 three different nitrogen levels.

Figure 9 shows the Pearson time series correlation between
the simulation model yield and FOA yield data. Figure 9 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 be-
ing difficult to model. No single model is dominant, with each
model providing near best-in-class performance in at least one
location-crop combination. The presence of very few vertical
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 per-
form best for rice. Reductions in skill over the performance
illustrated in Müller et al. (2017) can be attributed to the nitro-
gen levels or lack of calibration in some models.

642 *** or harmonization *** Christoph

Soy is qualitatively the easiest crop to represent (except in
Argentina), which is likely due in part to the invariance of the
response to nitrogen application (soy fixes atmospheric nitrogen
very efficiently). Comparison to the FAO data is therefore easier
than the other crops because the nitrogen application levels do
not matter. US maize has the best performance across models,
with nearly every model representing the historical variability

to a reasonable extent. Especially good example years for US maize are 1983, 1988, and 2004 (top left panel of Figure 9), where every model gets the direction of the anomaly compared to surrounding years correct. 1983 and 1988 are famously bad years for US maize along with 2012 (not shown). US maize

is possibly both the most uniformly industrialized (in terms of management practices) crop and the one with the best data collection in the historical period of all the cases presented here.

The FAO data is at least one level of abstraction from ground truth in many cases, especially in developing countries. The

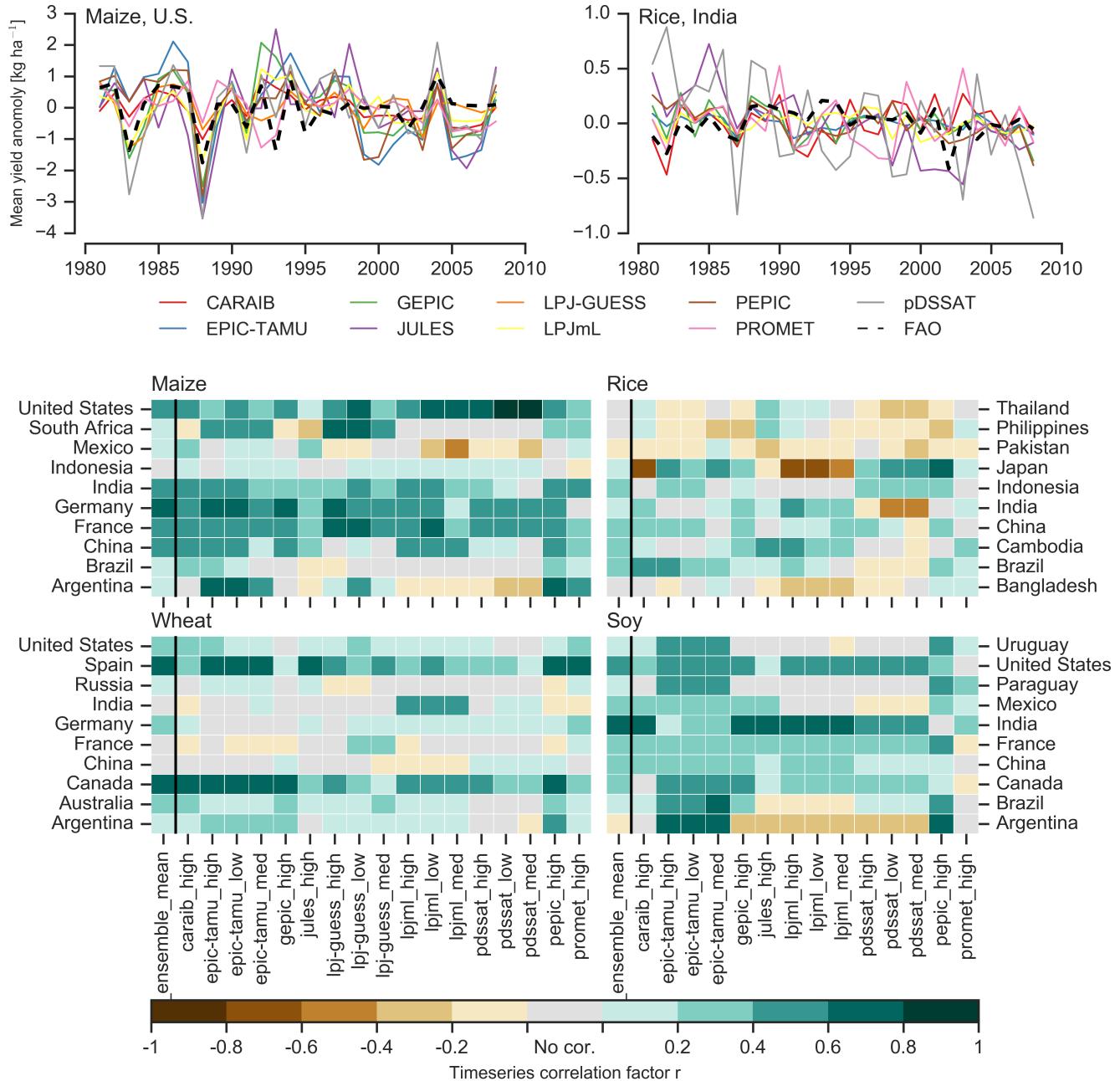


Figure 9: Time series of correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen 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.

failure of models to represent the year-to-year variability in rice
 in some countries in southeast Asia is likely partly due to model
 failure and partly due to lack of data. It is possible to speculate
 that the difference in performance between Pakistan (no suc-
 cessful models) and India (many successful models) for rice
 may reside at least in part in the FAO data and not the mod-
 els themselves. The same might apply to Bangladesh and In-
 dia for rice. Partitioning of these contributions is impossible at
 this stage. Additionally, there is less year-to-year variability in
 rice yields (partially due to the fraction of irrigated cultivation).
 Since the Pearson r metric is scale invariant, it will tend to score
 the rice models more poorly than maize and soy. An example
 of very poor performance can be seen with the pDSSAT model
 for rice in India (top right panel of Figure 9).

675 9. Appendix: Emulation – Assessment

Because no general criteria exist for defining an acceptable
 model emulator, we develop a metric of emulator performance
 specific to GGCMI. For a multi-model comparison exercise like
 GGCMI, one reasonable criterion is what we term the “nor-
 malized error”, which compares the fidelity of an emulator for a
 given model and scenario to the inter-model uncertainty. We
 define the normalized error e for each scenario as the difference
 between the fractional yield change from the emulator and that
 in the original simulation, divided by the standard deviation of
 the multi-model spread (Equations 2 and 3):

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

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

Here $F_{scn.}$ is the fractional change in a model’s mean emu-
 lated or simulated yield from a defined baseline, in some sce-
 nario (scn.) in C, T, W, and N space; $Y_{scn.}$ and $Y_{baseline}$ are the

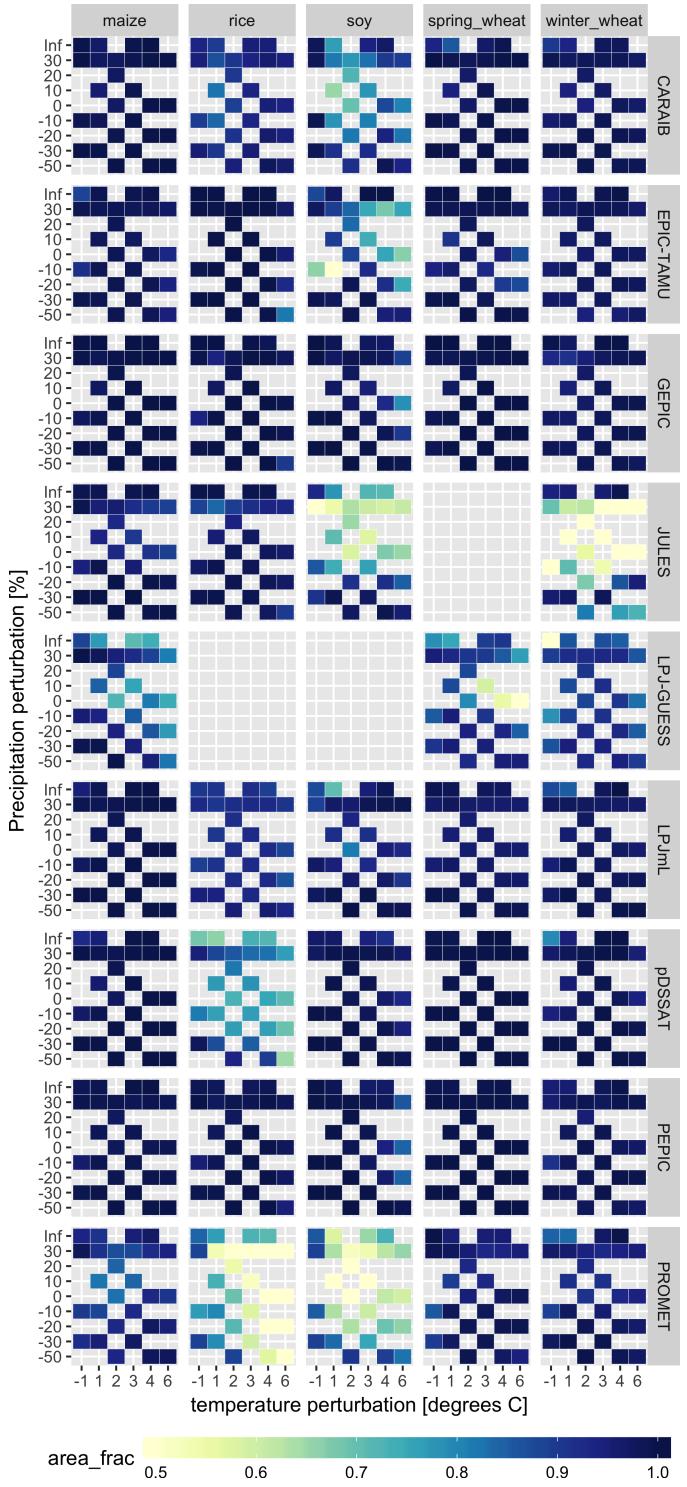


Figure 10: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO₂ and nitrogen held fixed at baseline values. Large columns are crops and large rows models; squares within are T,P scenario pairs. Colors denote the fraction of currently cultivated hectares with $e < 1$. Of the 756 scenarios with these CO₂ 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.

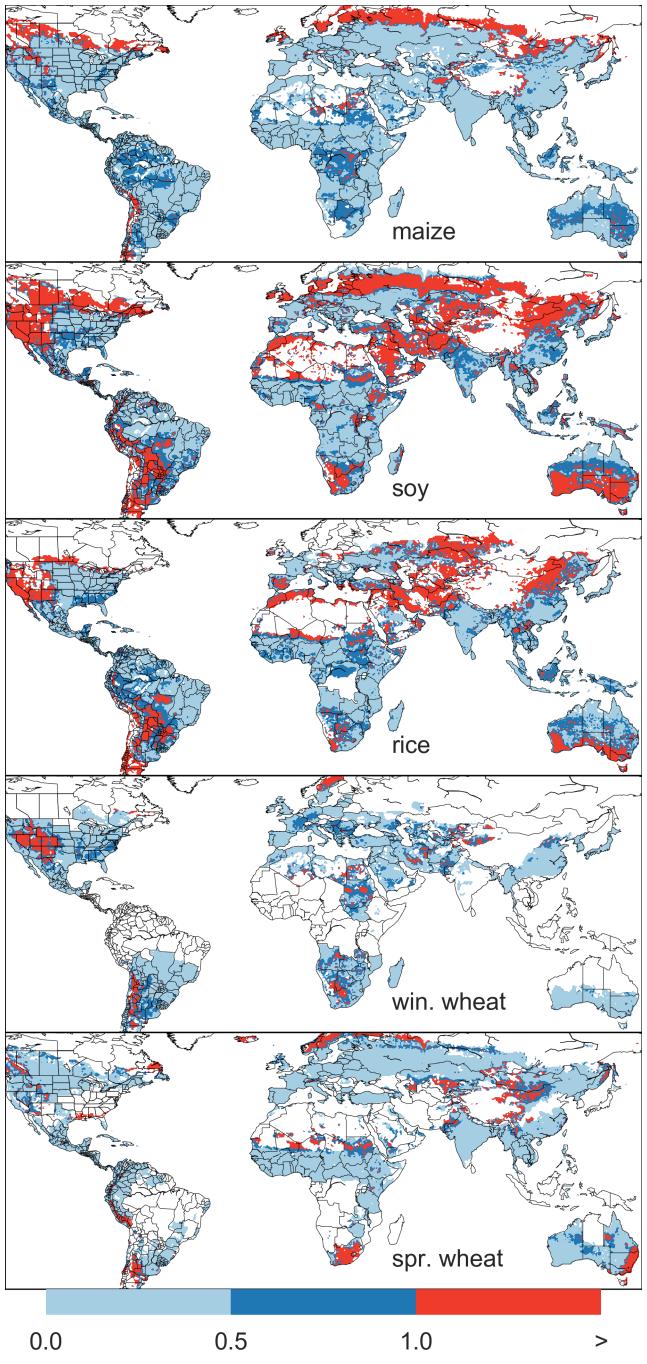


Figure 11: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error e , where $e > 1$ means that emulator error exceeds the multi-model standard deviation. White areas are those where crops 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.

absolute emulated or simulated mean yields. 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 $F_{sim, scn}$ across all models. The emulator is fit across all available simulation outputs, and then the error is calculated across the simulation scenarios provided by all nine models (Figure 10 and Figures S12 and Figures S13 in supplemental documents).

Note that the normalized error e for a model depends not only on the fidelity of its emulator in reproducing a given simulation but on the particular suite of models considered in the inter-comparison exercise. The rationale for this choice is to relate the fidelity of the emulation to an estimate of true uncertainty, which we take as the multi-model spread. **Because the inter-model spread is large, normalized errors tend to be small.** That is, any failures of emulation are small relative to inter-model uncertainty. We therefore do not provide a formal parameter uncertainty analysis, but note that the GGCMI Phase II dataset is well-suited to statistical exploration of emulation approaches and quantification of emulator fidelity.

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. 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

723 changes for soy than for other crops (see Figure S16, lowering
724 the denominator. Emulator performance often degrades in geo-
725 graphic locations where crops are not currently cultivated. Fig-
726 ure 11 shows a CARAIB case as an example, where emulator
727 performance is satisfactory over cultivated areas for all crops
728 other than soy, but uncultivated regions show some problematic
729 areas.

730 It should be noted that this assessment metric is relatively
731 forgiving. First, each emulation is evaluated against the simu-
732 lation actually used to train the emulator. Had we used a spline
733 interpolation the error would necessarily be zero. Second, the
734 performance metric scales emulator fidelity not by the magni-
735 tude of yield changes but by the inter-model spread in those
736 changes. Where models differ more widely, the standard for
737 emulators becomes less stringent. Because models disagree on
738 the magnitude of CO₂ fertilization, this effect is readily seen
739 when comparing assessments of emulator performance in sim-
740 ulations at baseline CO₂ (Figure 10) with those at higher CO₂
741 levels (Figure S13). Widening the inter-model spread leads to
742 an apparent increase in emulator skill.

743 10. References

- 744 Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013).
745 Implication of crop model calibration strategies for assessing regional im-
746 pacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.
747 Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D.,
748 A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Al-
749 derman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Bier-
750 nath, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising tempera-
751 tures reduce global wheat production. *Nature Climate Change*, 5, 143–147.
752 doi:10.1038/nclimate2470.
- 753 Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A.,
754 J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N.,
755 Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath,
756 C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simulat-
757 ing wheat yields under climate change. *Nature Climate Change*, 3, 827832.
758 doi:10.1038/nclimate1916.
- 759 Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other
760 Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Ef-
761 ficiency, Carbon Sequestration, and Environmental Pollution. *Advances in
762 Agronomy*, 86, 341 – 409.
- 763 Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov,
764 N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat
765 production potentials and management flexibility under the representative
766 concentration pathways. *Global and Planetary Change*, 122, 107 – 121.
- 767 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields
768 from global gridded crop models. *Agricultural and Forest Meteorology*, 236,
769 145 – 161.
- 770 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded
771 crop models using statistical estimates. *Agricultural and Forest Meteorol-
772 ogy*, 214-215, 134 – 147.
- 773 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S.
774 (2018). Implementing the Nitrogen cycle into the dynamic global vegeta-
775 tion, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific
776 Model Development*, 11, 2789–2812.
- 777 Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L.,
778 & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections
779 Based on Precomputed GCM Runs. *Journal of Climate*, 27, 1829–1844.
- 780 Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.
781 (2014). A meta-analysis of crop yield under climate change and adaptation.
782 *Nature Climate Change*, 4, 287 – 291.
- 783 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process
784 emulation of dynamic computer codes. *Biometrika*, 96, 663–676.
- 785 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In
786 C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth,
787 Duke University, Durham, North Carolina* (pp. 115–118). Durham, North
788 Carolina.
- 789 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for
790 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).
- 791 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberrous,
792 M., & François, L. (2011). Responses of European forest ecosystems to
793 21st century climate: assessing changes in interannual variability and fire
794 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).
- 795 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,
796 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating
797 impact models and sectors (pSIMS). *Environmental Modelling and Soft-
798 ware*, 62, 509–516.
- 799 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,
800 Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurrealde, R. C.,
801 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.

- 802 (2015). The Global Gridded Crop Model Intercomparison: data and mod-845
 803 elicing protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8, 846
 804 261–277. 847
- 805 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., 848
 806 & Taylor, K. E. (2016). Overview of the coupled model intercomparison 849
 807 project phase 6 (cmip6) experimental design and organization. *Geoscientific 850*
 808 *Model Development*, 9, 1937–1958. 851
- 809 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-852
 810 mate change impacts on durum wheat in the mediterranean region. *Natural 853*
 811 *Hazards and Earth System Sciences*, 11, 1293–1302. 854
- 812 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Re-855
 813 gionalization of a large-scale crop growth model for sub-Saharan Africa:856
 814 Model setup, evaluation, and estimation of maize yields. *Agriculture, 857*
 815 *Ecosystems & Environment*, 151, 21 – 33. 858
- 816 Food and Agriculture Organization of the United Nations (2018). FAOSTAT 859
 817 database. URL: <http://www.fao.org/faostat/en/home>. 860
- 818 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T., 861
 819 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., 862
 820 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F., 863
 821 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., 864
 822 Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J., 865
 823 Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nen-866
 824 del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna, 867
 825 Semenov, M. A., Slawinski, C., Stratonovitch, P., Supit, I., Waha, K., 868
 826 Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model 869
 827 wheat yield impact response surfaces showing sensitivity to temperature and 870
 828 precipitation change. *Agricultural Systems*, 159, 209–224. 871
- 829 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014). 872
 830 Evaluating the utility of dynamical downscaling in agricultural impacts pro-873
 831 jections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781. 874
- 832 Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitiv-875
 833 ity of Agricultural Model Performance to Different Climate Inputs. *Journal 876*
 834 *of Applied Meteorology and Climatology*, 55, 151113145618001. 877
- 835 Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported 878
 836 Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous 879
 837 Crop Growth and Yield: Application for Wheat in Central Europe. *Remote 880*
 838 *Sensing*, 7, 3934–3965. 881
- 839 He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H., 882
 840 He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water 883
 841 contents and nitrogen leaching to precipitation, management practices and 884
 842 soil hydraulic properties in semi-arid and humid regions of Canada using the 885
 843 DSSAT model. *Nutrient Cycling in Agroecosystems*, 106, 201–215. 886
- 844 Heady, E. O. (1957). An Econometric Investigation of the Technology of Agri-887
 845 cultural 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.
- Izaurrealde, 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 knowl-

- edge products: State of agricultural systems science. *Agricultural Systems*, 931
155, 269 – 288. 932
- Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, 933
D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Ver-934
burg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K., 935
Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003). 936
An overview of APSIM, a model designed for farming systems simulation. 937
European Journal of Agronomy, 18, 267 – 288. 938
- Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., & 939
Smith, B. (2013). Implications of accounting for land use in simulations of 940
ecosystem carbon cycling in Africa. *Earth System Dynamics*, 4, 385–407. 941
- Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling 942
wheat yield and crop water productivity with high resolution on a global 943
scale. *Agricultural Systems*, 94, 478 – 493. 944
- Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a). 945
Global investigation of impacts of PET methods on simulating crop-water 946
relations for maize. *Agricultural and Forest Meteorology*, 221, 164 – 175. 947
- Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., 948
& Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs 949
with yields from major crop cultivations. *Science of The Total Environment*, 950
572, 526 – 537. 951
- Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict 952
crop yield responses to climate change. *Agricultural and Forest Meteorol-* 953
ogy, 150, 1443 – 1452. 954
- Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relation-955
ships and the impacts of recent warming. *Environmental Research Letters*, 956
2, 014002. 957
- MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447. 958
- Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P., 959
Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi, 960
P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano, 961
D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of 962
large simulated yield datasets for studying climate effects. (p. 1100). 963
doi:10.13140/RG.2.1.5173.8328. 964
- Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrolog-965
ical modelling to study the impact of climate change on the water flows of 966
mountain watersheds. *Journal of Hydrology*, 376, 362 – 377. 967
- Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., 968
& Calzadilla, A. (2009). Global biomass production potentials exceed ex-969
pected future demand without the need for cropland expansion. *Nature Com-* 970
munications, 6. 971
- McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srini-972
vasa, C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S. 973
- (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 974
weather responsiveness of crop yields: US evidence and implications for 975
the agricultural impacts of climate change. *Environmental Research Letters*, 976
12.
- Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate 977
change impacts on agriculture implies higher social cost of carbon. *Nature 978
Communications*, 8.
- Müller, C., Elliott, J., Chrysanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, 979
P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurrealde, 980
R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. 981
A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., 982
Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. 983
(2017). Global gridded crop model evaluation: benchmarking, skills, 984
deficiencies and implications. *Geoscientific Model Development*, 10, 1403– 985
1422.
- Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T. 986
(1997). Effect of CO₂ enrichment on carbon and nitrogen interaction in 987
wheat and soybean. *Soil Science and Plant Nutrition*, 43, 789–798.
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. 988
Reliability Engineering & System Safety, 91, 1290 – 1300.
- Olin, S., Schurgers, G., Lindeskog, M., Wårliind, D., Smith, B., Bodin, P., 989
Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue 990
C:N to changes in atmospheric CO₂ and N management in the main wheat 991
regions of western europe. *Biogeosciences*, 12, 2489–2515. doi:10.5194/bg- 992
12-2489-2015.
- Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in 993
field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.
- Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., & 994
Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK 995
Land Environment Simulator. *Geoscientific Model Development*, 8, 1139– 996
1155.
- Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop 997
yields and their variability at different levels of global warming. *Earth Sys- 998
tem Dynamics*, 9, 479–496.
- Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff, 999
S., & Gerten, D. (2015). Emulating global climate change impacts on crop 1000
yields. *Statistical Modelling*, 15, 499–525.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., 1001
Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas- 1002
sos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011).

- 974 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. 1017
 975 1018
- 976 Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T. 1019
 977 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski 1020
 978 P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M. 1021
 979 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P. 1022
 980 Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minetti 1023
 981 J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C. 1024
 982 Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M. 1025
 983 Semenov, M., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang 1026
 984 E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation 1027
 985 effects on wheat yield across a European transect: a crop model ensemble 1028
 986 analysis using impact response surfaces. *Climate Research*, 65, 87–105. 1029
 987 Porter et al. (IPCC) (2014). Food security and food production systems. Cli 1030
 988 mate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global 1031
 989 and Sectoral Aspects. Contribution of Working Group II to the Fifth Assess 1032
 990 ment Report of the Intergovernmental Panel on Climate Change. In C. Fi 1033
 991 et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK 1034
 992 Cambridge University Press. 1035
 993 Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of 1036
 994 monthly growing areas of 26 irrigated crops. 1037
 995 Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly 1038
 996 Irrigated and Rainfed crop Areas around the Year 2000: A New High 1039
 997 Resolution Data Set for Agricultural and Hydrological Modeling. *Global 1040
 998 Biogeochemical Cycles*, 24, GB1011. 1041
 999 Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E. 1042
 1000 & Arneth, A. (2016). Climate analogues suggest limited potential for inten 1043
 1001 sification of production on current croplands under climate change. *Nature 1044
 1002 Communications*, 7, 12608. 1045
 1003 Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term cli 1046
 1004 mate change based on a resampling ensemble technique. *Tellus A: Dynamical 1047
 1005 Meteorology and Oceanography*, 58, 461–472. 1048
 1006 Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the 1049
 1007 reduction and sensitivity analysis of complex environmental models. *Envi 1050
 1008 ronmental Modelling & Software*, 34, 1 – 4. 1051
 1009 Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling 1052
 1010 in water resources. *Water Resources Research*, 48. 1053
 1011 Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017) 1054
 1012 Comparing and combining process-based crop models and statistical models 1055
 1013 with some implications for climate change. *Environmental Research Letters* 1056
 1014 12. 1057
 1015 Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A. 1058
 1016 Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek 1059
 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 As-
 sessments. *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 informa-
 tion, 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.

- 1060 Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901⁺₁₀₃ 2100 depicted by world maps of the Köppen-Geiger climate classification¹⁰⁴
- 1061 *Meteorologische Zeitschrift*, 19, 135–141.
- 1062 ability in JULES-crop and implications for forcing with seasonal weather forecasts. *Geoscientific Model Development*, 8, 3987–3997.
- 1063 Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting¹⁰⁵ dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19¹⁰⁷
- 1064 607–620.
- 1065 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Land-*
- 1066 *bouwkundige Onderzoeken* : 64.6, .
- 1067 Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng¹⁰⁹
- 1068 D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolin¹¹⁰
- 1069 ski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frieler, K¹¹¹
- 1070 (2017). Consistent negative response of US crops to high temperatures in¹¹²
- 1071 observations and crop models. *Nature Communications*, 8, 13931.
- 1072 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao,
- 1073 Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A.,
- 1074 Li, T., Lin, E., Liu, Q., Martre, P., Mller, C., Peng, S., Peuelas, J., Ruane,
- 1075 A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng,
- 1076 S. (2017). Temperature increase reduces global yields of major crops in four
- 1077 independent estimates. *Proc. Natl. Acad. Sci.*, 114, 9326–9331.
- 1078 Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate¹¹⁴
- 1079 severe damages to U.S. crop yields under climate change. *Proceedings of¹¹⁵*
- 1080 the National Academy of Sciences
- 1081 Snyder, A., Calvin, K. V., Phillips, M., & Ruane, A. C. (2018). A crop yield¹¹⁶
- 1082 change emulator for use in gcam and similar models: Persephone v1.0. *Geo-*
- 1083 *scientific Model Development Discussions*, 2018, 1–42.
- 1084 Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implemen-¹¹⁷
- 1085 tation and evaluation of nonparametric regression procedures for sensitivity¹¹⁸
- 1086 analysis of computationally demanding models. *Reliability Engineering &*
- 1087 *System Safety*, 94, 1735 – 1763.
- 1088 Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5¹¹⁹
- 1089 and the experiment design. *Bulletin of the American Meteorological Society*,
- 1090 93, 485–498.
- 1091 Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of cli-¹²⁰
- 1092 mate change impacts on global crop yields. *Geophysical Research Letters*,
- 1093 35.
- 1094 Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., &¹²¹
- 1095 Martin, J. F. (2014). Modeling sugarcane yield with a process-based model¹²²
- 1096 from site to continental scale: Uncertainties arising from model structure¹²³
- 1097 and parameter values. *Geoscientific Model Development*, 7, 1225–1245.
- 1098 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe,¹²⁴
- 1099 J. (2014). The Inter-Sectoral Impact Model Intercomparison Project¹²⁵
- 1100 (ISI-MIP): Project framework. *Proceedings of the National Academy of¹²⁶*
- 1101 Sciences
- 1102 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield vari-¹²⁷