

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

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

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

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

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

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

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

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

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

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

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

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

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

^lMet Office Hadley Centre, Exeter, United Kingdom

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

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

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

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

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

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

15 Computational Models have been used to project crop yields
16 since the 1950's, beginning with statistical models (Heady,
17 1957, Heady & Dillon, 1961) that attempt to capture the rela-
18 tionship between input factors and resultant yields. These sta-
19 tistical models were typically developed on a small scale for lo-
20 cations with extensive histories of yield data. The emergence of
21 computers allowed development of numerical models that sim-
22 ulate the process of photosynthesis and the biology and phe-
23 nology of individual crops (first proposed by de Wit (1957),
24 Duncan et al. (1967) and attempted by Duncan (1972)). For a
25 history of crop model development see the appendix of 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 simulation models for many years at high spatial
29 resolution 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 models do continue to struggle with some important details, including reproducing historical year-to-year variability (e.g. Müller et al., 2017), reproducing historical yields when driven by reanalysis weather (e.g. Glotter et al., 2014), and low sensitivity to extreme events (e.g. Glotter et al., 2015). These issues are driven in part by the diversity of new cultivars and genetic variants, which outstrips the ability of academic modeling groups to capture them (e.g. Jones et al., 2017). Models do not simulate many additional factors affecting production, including pests/diseases/weeds. For these reasons, individual studies must generally re-calibrate models to ensure that short-term predictions reflect current cultivar mixes, and long-term projections retain considerable uncertainty (Wolf & Oijen, 2002, Jagtap & Jones, 2002, Iizumi et al., 2010, Angulo et al., 2013, Asseng et al., 2013, 2015). Inter-model discrepancies can also be high in areas not yet cultivated (e.g. Challinor et al., 2014, White et al., 2011). Finally, process-based models present additional difficulties for high-resolution global studies because of their complexity and computational requirements. For economic impacts assessments, it is often impossible to integrate a set of process-based crop models directly into an integrated assessment model to estimate the potential cost of climate change to the agricultural sector.

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

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

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

74 Statistical emulation of crop simulations has been used to
 75 combine advantageous features of both statistical and process-
 76 based models. The statistical representation of complicated nu-
 77 matical simulation (e.g. O'Hagan, 2006, Conti et al., 2009), in
 78 which simulation output acts as the training data for a statisti-
 79 cal model, has been of increasing interest with the growth of
 80 simulation complexity and volume of output. Such emulators
 81 or "surrogate models" have been used in a variety of fields in-
 82 cluding hydrology (e.g. Razavi et al., 2012), engineering (e.g.
 83 Storlie et al., 2009), environmental sciences (e.g. Ratto et al.,
 84 2012), and climate (e.g. Castruccio et al., 2014, Holden et al.,
 85 2014). For agricultural impacts studies, emulation of process-
 86 based models allows exploring crop yields in regions outside
 87 ranges of current cultivation and with input variables outside
 88 historical precedents, in a lightweight, flexible form that is com-
 89 patible with economic studies.

90 In the past decade, many studies have developed emulators of
 91 crop yields from process-based models. Early studies propos-
 92 ing or describing potential emulators include Howden & Crimp
 93 (2005), Räisänen & Ruokolainen (2006) and Lobell & Burke
 94 (2010). In an early application, Ferrise et al. (2011) used a Arti-
 95 ficial Neural Net trained on simulation outputs to predict wheat₁₀₁
 96 yields in the Mediterranean. Studies developing single-model₁₀₂
 97 emulators include Holzkämper et al. (2012) for the CropSyst₁₀₃
 98 model, Ruane et al. (2013) for the CERES wheat model, Oye-₁₀₄
 99 bamiji et al. (2015) for the LPJmL model (for multiple crops,₁₀₅
 100 using multiple scenarios as a training set). In recent years, emu-₁₀₆

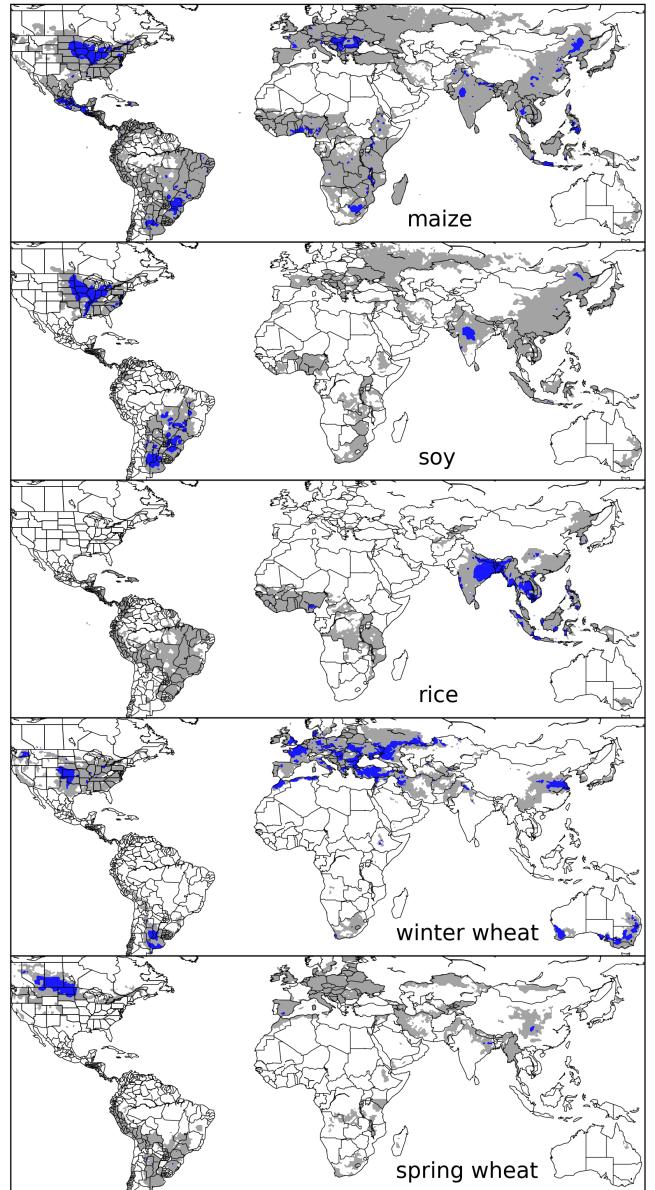


Figure 1: Presently cultivated area for rain-fed crops in the real world. Blue contour area indicates grid cells with more than 20,000 hectares (approx. 10% of the land in a grid cell near the equator for example) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated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

lators have begun to be used in the context of multi-model inter-
 comparisons, with Blanc & Sultan (2015), Blanc (2017), Ost-
 berg et al. (2018) and Mistry et al. (2017) using them to analyze
 the five crop models of the Inter-Sectoral Impacts Model Inter-
 comparison Project (ISIMIP) (Warszawski et al., 2014) (for
 maize, soy, wheat, and rice). Approaches differ: Blanc & Sultan

107 (2015) and Blanc (2017) used local weather variables (and CO₂)₁₃₅
 108 values) and yields but emulate across soil types using historical₁₃₆
 109 simulations and a future climate scenario (RCP8.5 over mul-₁₃₇
 110 tiple climate models); Ostberg et al. (2018) used global mean₁₃₈
 111 temperature change (and CO₂) as regressors but pattern-scale₁₃₉
 112 to emulate local yields using multiple climate scenarios; Mis-₁₄₀
 113 try et al. (2017) used local weather and yields and a historical₁₄₁
 114 simulation and compare with data. As an alternative approach₁₄₂
 115 to RCP climate scenarios, a systematic parameter sweep offers₁₄₃
 116 some advantages over analyses on small number of potential₁₄₄
 117 future scenarios in which climate varies over time. A parameter₁₄₅
 118 sweep across the major drivers allows highlighting the distinc-₁₄₆
 119 tion between year-over-year and climatological changes, which₁₄₇
 120 can be different (e.g. Ruane et al., 2016). It also removes the₁₄₈
 121 correlation between the key variables driving yields which may₁₄₉
 122 prove difficult to disentangle in RCP climate model runs and it₁₅₀
 123 provides fully stationary simulations.
151

124 The trend may be increasing towards more systematic pa-₁₅₂
 125 rameter sweeps in the crop modeling literature in the context₁₅₃
 126 of emulation and model improvement. Earlier efforts include₁₅₄
 127 Makowski et al. (2015) and Pirttioja et al. (2015), and more₁₅₅
 128 recently Fronzek et al. (2018) and Snyder et al. (2018). Both₁₅₆
 129 Fronzek et al. (2018) and Snyder et al. (2018) test different lev-₁₅₇
 130 els of temperature and precipitation and Snyder et al. (2018)₁₅₈
 131 adds CO₂ for 132 and 99 different combinations respectively₁₅₉
 132 and both take advantage of the output simulation data to con-₁₆₀
 133 struct climatological mean emulators (aka response surface).₁₆₁
 134 Fronzek et al. (2018) tests many different models for wheat for₁₆₂

sites in Europe and Snyder et al. (2018) analyzes four crops for
 136 the GCAM model for a variety of different sites. In this paper
 we describe a new comprehensive dataset designed to expand
 137 this approach still further. The GGCMI Phase II experiment
 provides global coverage at the half degree lat-lon grid level
 138 and adds an applied nitrogen dimension to test different lev-
 139 els of management resulting in over 700 different combinations
 140 of input parameters for each model and crop. The experiment
 involves running a suite of process-based crop models across
 141 historical conditions perturbed by a set of defined input pa-
 142 rameters, and was conducted as part of the Agricultural Model
 143 Intercomparison and Improvement Project (AgMIP) (Rosen-
 144 zweig et al., 2013, 2014), an international effort conducted un-
 145 der a framework similar to the Climate Model Intercomparison
 146 Project (CMIP) (Taylor et al., 2012, Eyring et al., 2016). The
 147 GGCMI protocol builds on the AgMIP Coordinated Climate-
 148 Crop Modeling Project (C3MP) (Ruane et al., 2014, McDer-
 149 mid et al., 2015) and will contribute to the AgMIP Coordinated
 150 Global and Regional Assessments (CGRA) (Ruane et al., 2018,
 151 Rosenzweig et al., 2018).

GGCMI Phase II is designed to allow addressing goals such
 as understanding where highest-yield regions may shift un-
 152 der climate change; exploring future adaptive management
 153 strategies; understanding how interacting input drivers affect
 154 crop yield; quantifying uncertainties across models and major
 155 drivers; and testing strategies for producing lightweight em-
 156 ulators of process-based models. In this paper, we describe
 157 the GGCMI Phase II experiments, summarize output data, and
 158

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

Table 1: Phase II input variable test levels. Temperature and precipitation values indicate the perturbations from the historical, climatology. * Only simulated by one model. W-percentage does not apply to the irrigated (W_{inf}) simulations.

163 present initial results and demonstrate that it is tractable to em-194
164 ulation.

165 **2. Materials and Methods**

166 *2.1. GGCMI Phase II: experiment design*

167 GGCMI Phase II is the continuation of a multi-model com-200
168 parison exercise begun in 2014. The initial Phase I compared201
169 harmonized yields of 21 models for 19 crops over a historical202
170 (1980-2010) scenario with a primary goal of model evaluation203
171 (Elliott et al., 2015, Müller et al., 2017). Phase II compares sim-204
172 ulations of 12 models for 5 crops (maize, rice, soybean, spring205
173 wheat, and winter wheat) over hundreds of scenarios in which206
174 individual climate or management inputs are adjusted from207
175 their historical values. The reduced set of crops includes the208
176 three major global cereals and the major legume and accounts209
177 for over 50% of human calories (in 2016, nearly 3.5 billion tons210
178 or 32% of total global crop production by weight (Food and211
179 Agriculture Organization of the United Nations, 2018).

180 The major goals of GGCMI Phase II are to:

- 181 • Enhance understanding of how models work by character-214
182 izing their sensitivity to input climate and nitrogen drivers.215
- 183 • Study the interactions between climate variables and nitro-216
184 gen inputs in driving modeled yield impacts.217
- 185 • Explore differences in crop response to warming across the218
186 Earth's climate regions.219
- 187 • Provide a dataset that allows statistical emulation of crop220
188 model responses for downstream modelers.221
- 189 • Illustrate differences in potential adaptation via growing222
190 season changes.223

191 The guiding scientific rationale of GGCMI Phase II is to pro-225
192 vide a comprehensive, systematic evaluation of the response226
193 of process-based crop models to different values for carbon227

dioxide, temperature, water, and applied nitrogen (collectively known as “CTWN”). Phase II of the GGCMI project consists of a series of simulations, each with one or more of the CTWN dimensions perturbed over the 31-year historical time series (1980-2010) used in Phase I. In most cases, historical daily climate inputs are taken from the 0.5 degree NASA AgMERRA daily gridded re-analysis product specifically designed for agricultural modeling, with satellite-corrected precipitation (Ruane et al., 2015). Two models require sub-daily input data and use alternative sources. See Elliott et al. (2015) for additional details.

The experimental protocol consists of 9 levels for precipitation perturbations, 7 for temperature, 4 for CO₂, and 3 for applied nitrogen, for a total of 672 simulations for rain-fed agriculture and an additional 84 for irrigated (Table 1). For irrigated simulations, soil water is held at either field capacity or, for those models that include water-log damage, at maximum beneficial level. Temperature perturbations are applied as absolute offsets from the daily mean, minimum, and maximum temperature time series for each grid cell used as inputs. Precipitation perturbations are applied as fractional changes at the grid cell level, and carbon dioxide and nitrogen levels are specified as discrete values applied uniformly over all grid cells. Note that CO₂ changes are applied independently of changes in climate variables, so that higher CO₂ is not associated with higher temperatures. An additional, identical set of scenarios (at the same C, T, W, and N levels) simulate adaptive agronomy under climate change by varying the growing season for crop production. (These adaptation simulations are not shown or analyzed here.) The resulting GGCMI data set captures a distribution of crop responses over the potential space of future climate conditions.

The 12 models included in GGCMI Phase II are all mechanistic process-based crop models that are widely used in im-

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

Table 2: Models included in the GGCMI Phase II and the number of C, T, W, and N simulations performed for rain-fed crops (“Sims per Crop”), with 672 as the maximum. “N-Dim.” indicates if simulations include varying nitrogen levels; two models omit this dimension. All models provided the same set of simulations across all modeled crops, but some omitted individual crops in some cases. (For example, APSIM did not simulate winter wheat.) Irrigated simulations are provided at the level of the other covariates for each model (for an additional 84 simulations for the fully-sampled models). Geographic extent of simulation varies to some extent within a certain model for different scenarios (672 rain-fed simulations does not necessarily equal 672 climatological yields in all areas). This geographic variance only applies for areas far outside the area of currently cultivated crops. Two models (marked with *) use non-AgMERRA climate inputs. For further details on models, see Elliott et al. (2015). †PROMET provided simulations at only two nitrogen levels so is not emulated across the nitrogen dimension.

228 pacts assessments (Table 2). Although some of the models²⁴⁶
 229 shares a common base (e.g. LPJmL and LPJ-GUESS and the²⁴⁷
 230 EPIC models), they have developed independently from this²⁴⁸
 231 shared base, for more details on the genealogy of the mod-²⁴⁹
 232 els see Figure S1 in Rosenzweig et al. (2014). Differences in²⁵⁰
 233 model structure does mean that several key factors are not stan-²⁵¹
 234 dardized across the experiment, including secondary soil nutri-²⁵²
 235 ents, carry over effects across growing years including residue²⁵³
 236 management and soil moisture, and extent of simulated area for²⁵⁴
 237 different crops. Growing seasons are identical across models,²⁵⁵
 238 but vary by crop and by location on the globe. All stresses²⁵⁶
 239 except factors related to nitrogen, temperature, and water (e.g.²⁵⁷
 240 Alkalinity, salinity) are disabled. No additional nitrogen inputs,²⁵⁸
 241 such as atmospheric deposition, are considered, but some mod-
 242 els have individual assumptions on soil organic matter that may²⁵⁹
 243 release additional nitrogen through mineralization. See Rosen-²⁶⁰
 244 zweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017)²⁶¹
 245 for further details on models and underlying assumptions.

Each model is run at 0.5 degree spatial resolution and covers all currently cultivated areas and much of the uncultivated land area. Coverage extends considerably outside currently cultivated areas because cultivation will likely shift under climate change. See Figure 1 for the present-day cultivated area of rain-fed crops, and Figure S1 in the supplemental material for irrigated crops. Some areas such as Greenland, far-northern Canada, Siberia, Antarctica, the Gobi and Sahara deserts, and central Australia are not simulated as they are assumed to remain non-arable even under an extreme climate change. Growing seasons are standardized across models with data adapted from several sources (Sacks et al., 2010, Portmann et al., 2008, 2010).

The participating modeling groups provide simulations at any of four initially specified levels of participation, so the number of simulations varies by model, with some sampling only a part of the experiment variable space. Most modeling groups simulate all five crops in the protocol, but some omitted one

264 or more. Table 2 provides details of coverage for each model.²⁹⁷
 265 Note that the three models that provide less than 50 simulations²⁹⁸
 266 are excluded from the emulator analysis.

267 All models produce as output, crop yields (tons ha⁻¹ year⁻¹)²⁹⁹
 268 for each 0.5 degree grid cell. Because both yields and yield
 269 changes vary substantially across models and across grid cells,
 270 we primarily analyze relative change from a baseline. We take
 271 as the baseline the scenario with historical climatology (i.e. T
 272 and P changes of 0). C of 360 ppm, and applied N at 200 kg
 273 ha⁻¹. We show absolute yields in some cases to illustrate geo-
 274 graphic differences in yields for a single model.

275 2.2. Simulation model validation approach

276 To verify the skill of the process-based models used, we re-
 277 peat the validation exercises presented in Müller et al. (2017)
 278 for GGCMI Phase I. The Müller et al. (2017) validation pro-
 279 cedure evaluates response to year-to-year temperature and pre-
 280 cipitation variations in a control run driven by historical cli-
 281 mate and compares it to detrended historical yields from the
 282 FAO (Food and Agriculture Organization of the United Nations,
 283 2018) by calculating the Pearson correlation coefficient. The³⁰⁰
 284 procedure offers no means of assessing CO₂ fertilization, since³⁰¹
 285 CO₂ has been relatively constant over the historical data col-³⁰²
 286 lection period. Nitrogen data are limited for many countries,³⁰³
 287 and as mentioned the GGCMI Phase II runs impose fixed and³⁰⁴
 288 uniform nitrogen application, introducing some uncertainty into³⁰⁵
 289 the analysis. We evaluate one or more control runs for each³⁰⁶
 290 model, since some modeling groups provide historical runs for³⁰⁷
 291 three different nitrogen levels. Note however that the GGCMI³⁰⁸
 292 Phase II simulations are designed for evaluating changes in³⁰⁹
 293 yield but not absolute yields, and so omit the calibrations used³¹⁰
 294 in predicting modeling to account for cultivar, pest loss, and³¹¹
 295 management differences. The Phase II simulations also do³¹²
 296 not reproduce realistic nitrogen application levels for individual³¹³

countries, since nitrogen is one of the parameters systematically varied.

2.3. Climatological-mean yield emulator design

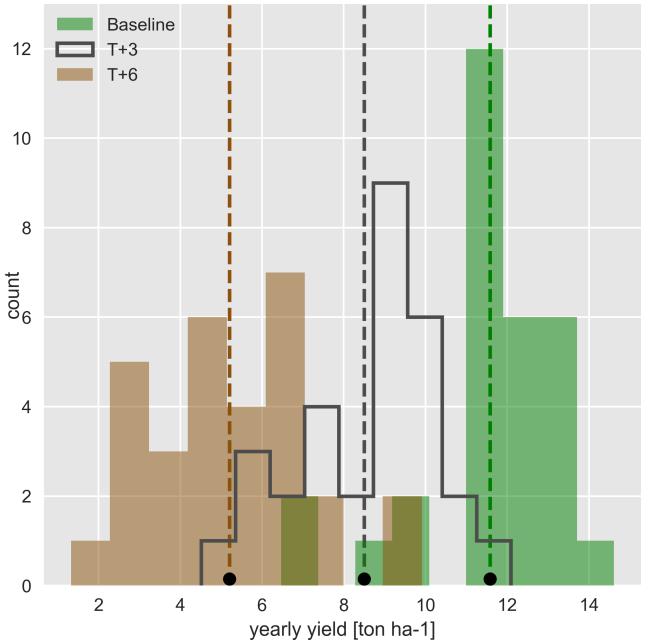


Figure 2: Example showing both climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for a grid cell in northern Iowa (a representative high-yield region) from the pDSSAT model, for the baseline climatology (1981–2010) and for scenarios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

To assist in the demonstration of the properties of the GGCMI Phase II dataset, we construct an emulator of 30-year climatological mean yields. This approach differs from previous studies of crop model emulation, which have typically emulated at the annual level. Annual emulation is required when the input training set consists of non-stationary projections of evolving yields (such as an RCP climate model run). We test the necessity for the climatological approach in this case by using the GGCMI Phase II dataset to evaluate whether year-over-year responses are quantitatively distinct from climatological mean responses. The two may differ for some reasons including: any year-to-year memory in the crop model, lurking covariants, or if the distribution of growing-season daily weather associated with interannual variability is be different

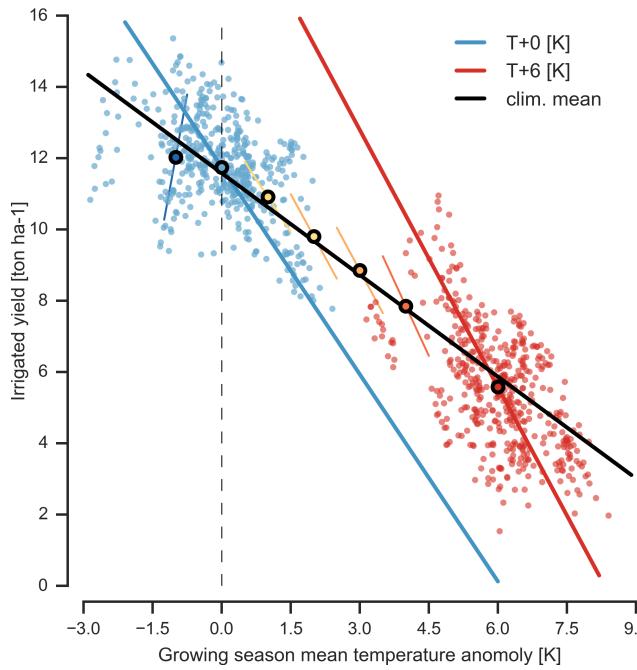


Figure 3: Example showing temperature relationship developed from year-to-year values vs. climatological mean values. Figure shows irrigated maize for nine adjacent grid cell in northern Iowa (a representative high-yield region) from the pDSSAT model, for the baseline climatology (1981–2010) and for scenario with temperature shifted (T) +6 K, with other variables held at baseline values. Irrigated yields are shown to control for precipitation effects. Blue and red lines indicate total least squares linear regression across each temperature scenario. Black ringed points indicate the climatological mean yield values for each climatological temperature scenario in the study (T -1, +0, +1, +2, +3, +4, +6 [K]). Short colored lines indicate slope of best fit (TLS) for year-to-year relationship at each different climatological mean temperature simulation. The bold black line indicates the fit through the climatological mean values.

is to facilitate these research efforts.

Emulation involves fitting individual regression models for each crop, simulation model, and 0.5 degree geographic pixel from the GGCMI Phase II data set. The regressors are the applied constant perturbations in temperature, water, nitrogen and CO_2 , we aggregate the simulation outputs in the time dimension, and regress on the 30-year mean yields. (See Figure 2 for illustration). The regression therefore omits information about yield responses to year-to-year climate perturbations, which are more complex. Emulating inter-annual yield variations would likely require considering statistical details of the historical climate time series, including changes in marginal distribution and temporal dependencies. (Future work should explore this). The climatological emulation indirectly includes any yield response to geographically distributed factors such as soil type, insolation, and the baseline climate itself, because we construct separate emulators for each grid cell. The emulator parameter matrices are portable and the yield computations are cheap even at the half-degree grid cell resolution, so we do not aggregate in space at this time.

from that associated with long-term CO_2 -driven changes (e.g. Ruane et al., 2016). The year-over-year yield response to individual factors in GGCMI Phase II do in fact often exceed the climatological-mean response (Figure 3). Note that the GGCMI Phase II datasets will not capture distributional shifts, because all simulations are run with fixed offsets from the historical climatology. Some work has shown that the mean climatological change may dominate any shifts in the weather distributions at least in non-arid regions (e.g. Glotter et al., 2014). (For methods to generate adjust historical climate data inclusive of distributional and temporal dependence changes, see Leeds et al. (2015), Poppick et al. (2016), Chang et al. (2016) and Haugen et al. (2018)). Emulation approaches are an area of active ongoing study and one of the goals of the GGCMI Phase II dataset

We regress climatological mean yields against a third-order polynomial in C , T , W , and N with interaction terms. The higher-order terms are necessary to capture any nonlinear responses, which are well-documented in observations for temperature and water perturbations (e.g. Schlenker & Roberts (2009) for T and He et al. (2016) for W). 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 diox-

ide (Osaki et al., 1992, Nakamura et al., 1997). 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. We

choose an emulation at grid-cell level in this study.

The limited GGCM1 variable sample space means that use of the full polynomial expression described above, which has 34 terms for the rain-fed case (12 for irrigated), can be problematic, and can lead to over-fitting and unstable parameter estimations. 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

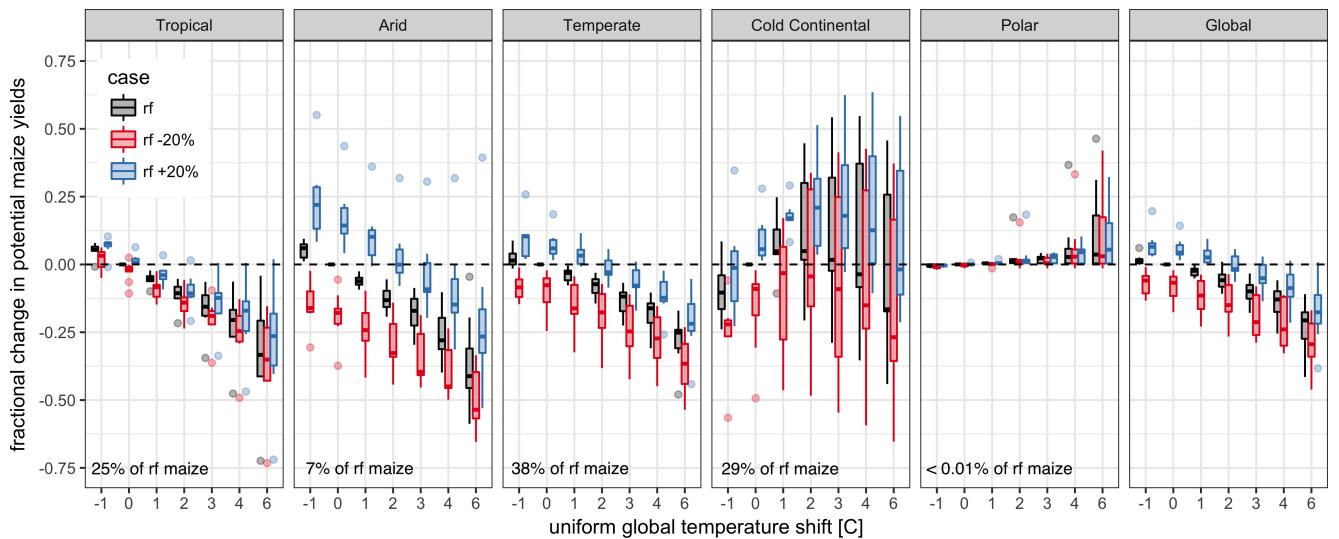


Figure 4: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation 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.

396 include that nitrogen interactions are complex and important,⁴²⁰
 397 and that water interaction effects are more nonlinear than those⁴²¹
 398 in temperature. The resulting statistical model (Equation 1) is⁴²²
 399 used for all grid cells, models, and crops:

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

400 To fit the parameters K , we use a Bayesian Ridge probabilis-
 401 tic estimator (MacKay, 1991), which reduces volatility in pa-
 402 rameter estimates when the sampling is sparse, by weighting
 403 parameter estimates towards zero. The Bayesian Ridge method
 404 is necessary to maintain a consistent functional form across all
 405 models, and locations as the linear least squares fails to pro-⁴³⁴
 406 vide a stable result in many cases. In the GGCMI Phase II⁴³⁵
 407 experiment, the most problematic fits are those for models that⁴³⁶
 408 provided a limited number of cases or for low-yield geographic⁴³⁷
 409 regions where some modeling groups did not run all scenarios.⁴³⁸
 410 Because we do not attempt to emulate models that provided less⁴³⁹
 411 than 50 simulations, the lowest number of simulations emulated⁴⁴⁰
 412 across the full parameter space is 130 (for the PEPIC model).⁴⁴¹
 413 We do not provide a formal parameter uncertainty analysis as⁴⁴²
 414 part of this study. We use the implementation of the Bayesian⁴⁴³
 415 Ridge estimator from the scikit-learn package in Python (Pe-⁴⁴⁴
 416 dregosa et al., 2011).

417 The resulting parameter matrices for all crop model emula-⁴⁴⁶
 418 tors are available on request, as are the raw simulation data and⁴⁴⁷
 419 a Python application to emulate yields. The yield output for a⁴⁴⁸

single GGCMI model that simulates all scenarios and all five crops is \sim 12.5 GB; the emulator is \sim 100 MB, a reduction by over two orders of magnitude.

423 2.4. Emulator evaluation

424 Because no general criteria exist for defining an acceptable
 425 model emulator, we develop a metric of emulator performance
 426 specific to GGCMI. For a multi-model comparison exercise like
 427 GGCMI, a reasonable criterion is what we term the “normalized
 428 error”, which compares the fidelity of an emulator for a given
 429 model and scenario to the inter-model uncertainty. We define
 430 the normalized error e for each scenario as the difference be-
 431 tween the fractional yield change from the emulator and that in
 432 the original simulation, divided by the standard deviation of the
 433 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 emulated or simulated yield from a defined baseline, in some scenario (scn.) in C, T, W, and N space; $Y_{scn.}$ and $Y_{baseline}$ are the absolute emulated or simulated mean yields. 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 9 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 intercomparison exercise. The rationale for this choice is to relate the fidelity of

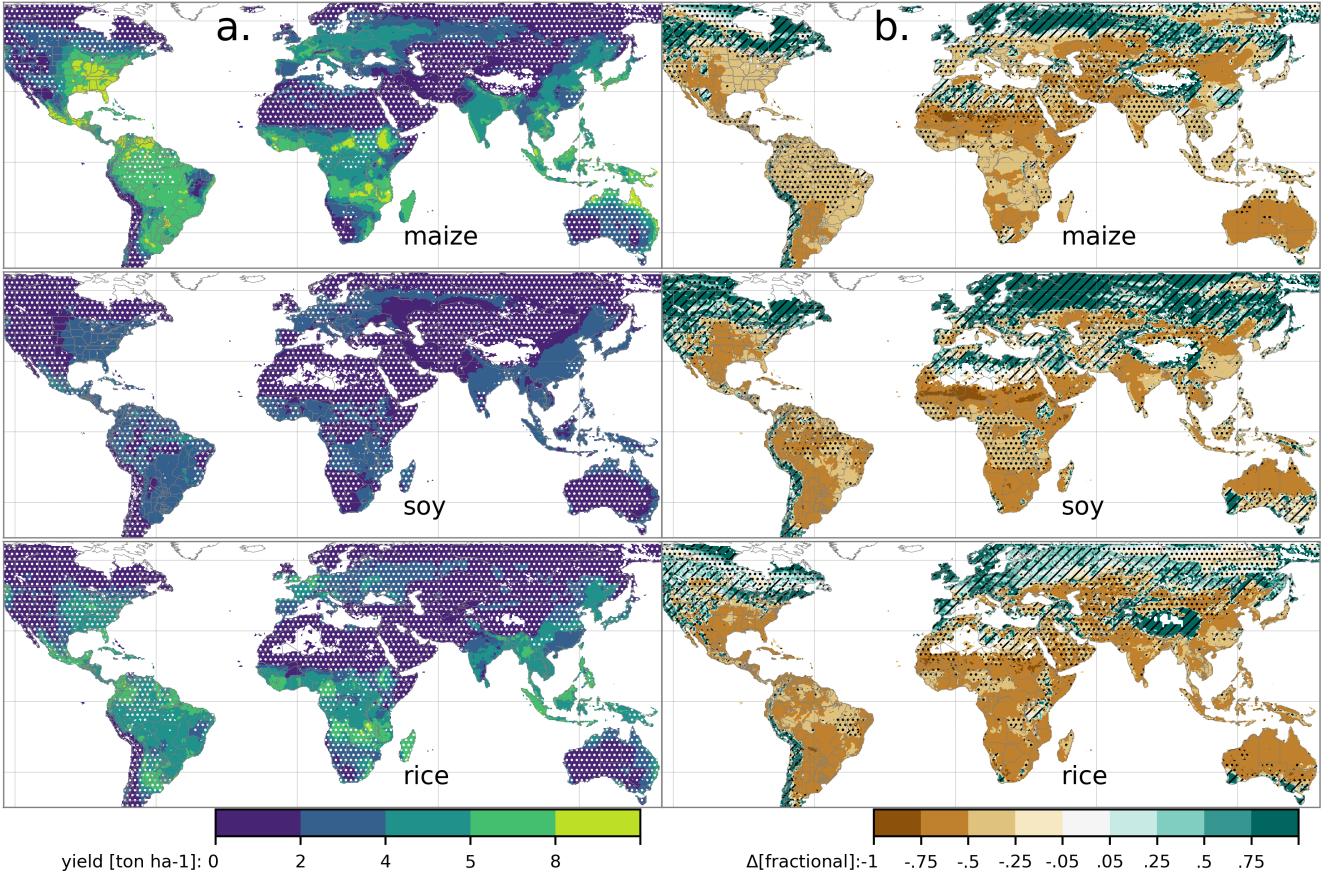


Figure 5: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure 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 (0.2 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.

449 the emulation to an estimate of true uncertainty, which we take₄₅₀
450 as the multi-model spread.

451 3. Results

452 3.1. Simulation results

453 Crop models in the GGCMI ensemble show a broadly con-
454 sistent responses to climate and management perturbations in
455 most regions, with a strong negative impact of increased tem-
456 perature in all but the coldest regions. We illustrate this result
457 for rain-fed maize in Figure 4, which shows yields for the pri-
458 mary Köppen-Geiger climate regions (Rubel & Kottek, 2010).
459 In warming scenarios, models show decreases in maize yield in
460 the temperate, tropical, and arid regions that account for nearly₄₇₁

461 three-quarters of global maize production. These impacts are
462 robust for even moderate climate perturbations. In the temper-
463 ate zone, even a 1 degree temperature rise with other variables
464 held fixed leads to a median yield reduction that outweighs the
465 variance across models. A 6 degree temperature rise results in
466 median loss of ~25% of yields with a signal to noise of nearly
467 three. A notable exception is the cold continental region, where
468 models disagree strongly, extending even to the sign of impacts.
469 Model simulations of other crops produce similar responses to
470 warming, with robust yield losses in warmer locations and high
471 inter-model variance in the cold continental regions (Figures
472 S7).

The effects of rainfall changes on maize yields are also as ex-

474 pected and are consistent across models. Increased rainfall mit-⁵⁰⁸
475 igates the negative effect of higher temperatures, most strongly⁵⁰⁹
476 in arid regions. Decreased rainfall amplifies yield losses and⁵¹⁰
477 also increases inter-model variance more strongly, suggesting⁵¹¹
478 that models have difficulty representing crop response to water⁵¹²
479 stress. We show only rain-fed maize here; see Figure S5 for the⁵¹³
480 irrigated case. As expected, irrigated crops are more resilient to⁵¹⁴
481 temperature increases in all regions, especially so where water⁵¹⁵
482 is limiting.

483 Mapping the distribution of baseline yields and yield changes⁵¹⁶
484 shows the geographic dependencies that underlie these results.⁵¹⁷
485 Figure 5 shows baseline and changes in the T+4 scenario for⁵¹⁸
486 rain-fed maize, soy, and rice in the multi-model ensemble mean,⁵¹⁹
487 with locations of model agreement marked. Absolute yield po-⁵²⁰
488 tentials are have strong spatial variation, with much of the⁵²¹
489 Earth's surface area unsuitable for any given crop. In general,⁵²²
490 models agree most on yield response in regions where yield⁵²³
491 potentials are currently high and therefore where crops are cur-⁵²⁴
492 rently grown. Models show robust decreases in yields at low⁵²⁵
493 latitudes, and highly uncertain median increases at most high⁵²⁶
494 latitudes. For wheat crops see Figure S11; wheat projections⁵²⁷
495 are both more uncertain and show fewer areas of increased yield⁵²⁸
496 in the inter-model mean.

497 3.2. *Simulation model validation results*

498 Figure 6 shows the Pearson time series correlation between⁵³²
499 the simulation model yield and FAO yield data. Figure 6 can be⁵³³
500 compared to Figures 1,2,3,4 and 6 in Müller et al. (2017). The⁵³⁴
501 results are mixed, with many regions for rice and wheat be-⁵³⁵
502 ing difficult to model. No single model is dominant, with each⁵³⁶
503 model providing near best-in-class performance in at least one⁵³⁷
504 location-crop combination. The presence of very few vertical⁵³⁸
505 dark green color bars clearly illustrates the power of a multi-⁵³⁹
506 model intercomparison project like the one presented here. The⁵⁴⁰
507 ensemble mean does not beat the best model in each case, but⁵⁴¹

shows positive correlation in over 75% of the cases presented here. The EPIC-TAMU model performs best for soy, CARIAB, EPIC-TAMU, and PEPIC perform best for maize, PROMET performs best for wheat, and the EPIC family of models perform best for rice. Reductions in skill over the performance illustrated in Müller et al. (2017) can be attributed to the nitrogen levels or lack of calibration in some models.

515 *** 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 6), 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.

531 The FAO data is at least one level of abstraction from ground truth in many cases, especially in developing countries. The 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 successful models) and India (many successful models) for rice may reside at least in part in the FAO data and not the models themselves. The same might apply to Bangladesh and India for rice. Partitioning of these contributions is impossible at this stage. Additionally, there is less year-to-year variability in

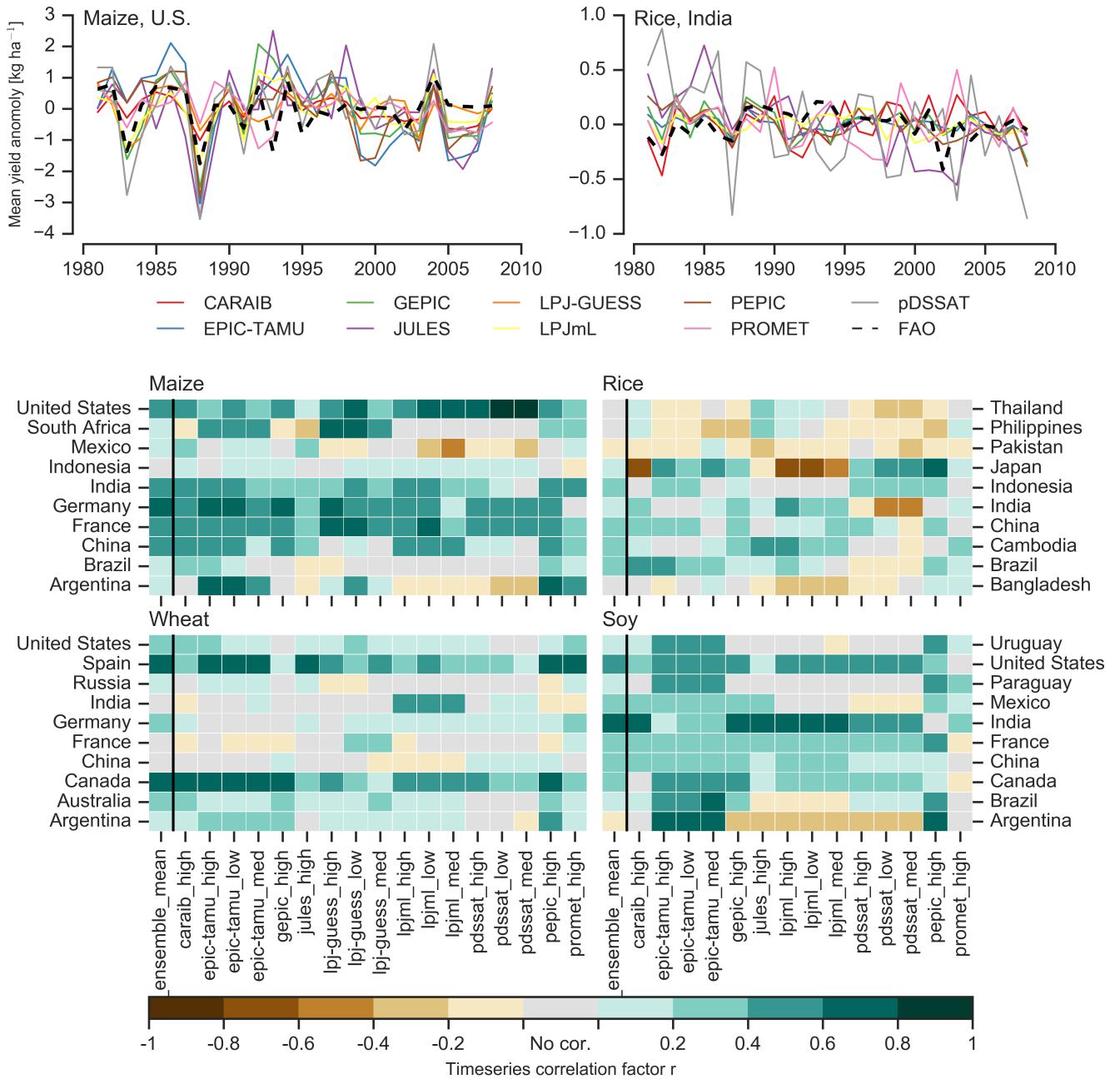


Figure 6: Time series correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen application case. The heatmaps illustrate the Pearson r correlation coefficient between the detrended simulation mean yield at the country level compared to the detrended FAO yield data for the top producing countries for each crop with continuous FAO data over the 1980–2010 period. Models that provided different nitrogen application levels are shown with low, med, and high label (models that did not simulate different nitrogen levels are analogous to a high nitrogen application level). The ensemble mean yield is also correlated with the FAO data (not the mean of the correlations). Wheat contains both spring wheat and winter wheat simulations.

rice yields (partially due to the fraction of irrigated cultivation).⁵⁴⁷ 3.3. Emulator performance

Since the Pearson r metric is scale invariant, it will tend to score

the rice models more poorly than maize and soy. An example⁵⁴⁸

of very poor performance can be seen with the pDSSAT model⁵⁴⁹

for rice in India (top right panel of Figure 6).

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, however, when crop yield responses are sufficiently smooth and

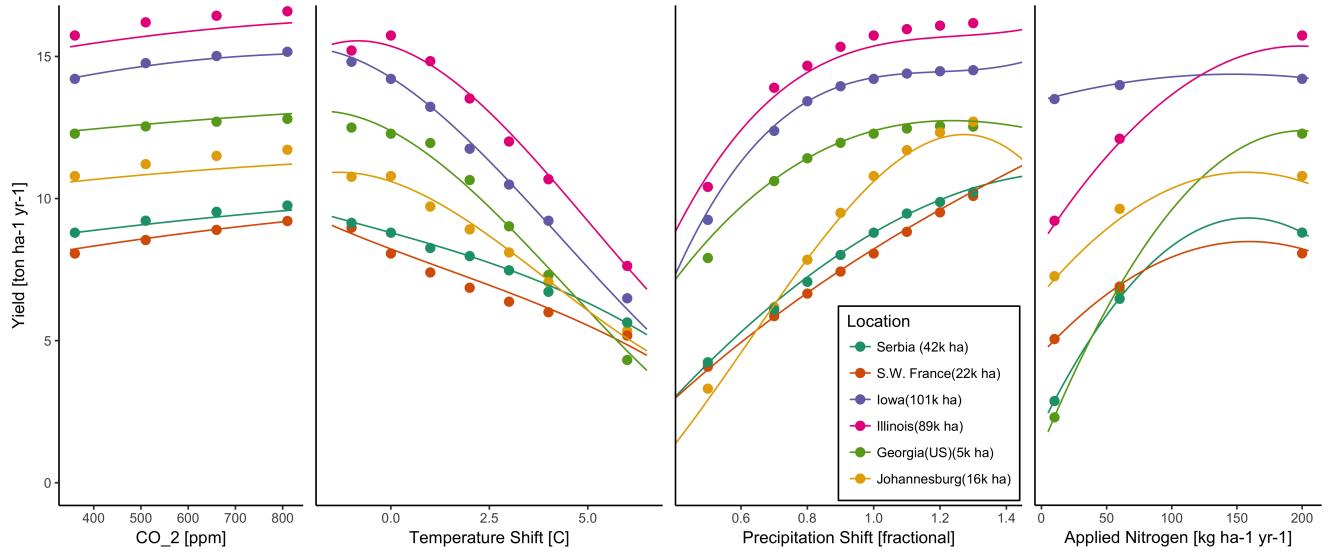


Figure 7: Example illustrating spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation along a single variable, with others held at baseline values. Dots show climatological mean yield, solid lines a simple polynomial fit across this 1D variation, and dashed lines the results of the full emulator of Equation 1. The climatological response is sufficiently smooth to be represented by a simple polynomial. Because precipitation changes are imposed as multiplicative factors rather than additive offsets, locations with higher baseline precipitation have a larger absolute spread in inputs sampled than do drier locations (not visible here).

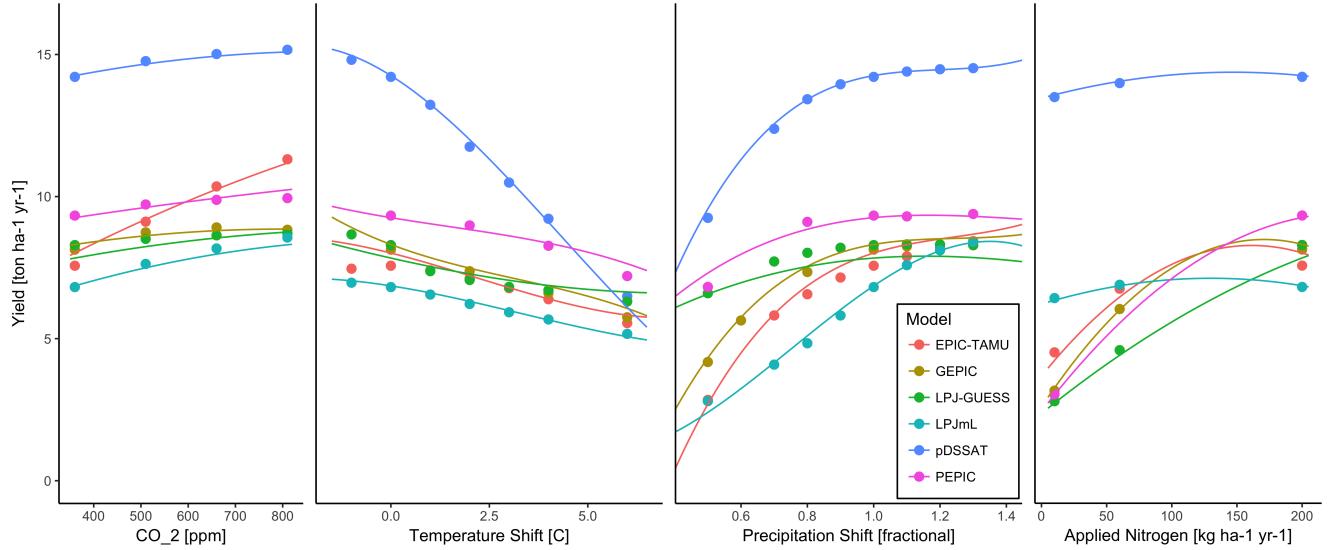


Figure 8: Example illustrating across-model variations in yield response. We show simulations and emulations from six models for rain-fed maize in the same Iowa location shown in Figure 7, with the same plot conventions. Models that did not simulate the nitrogen dimension are omitted for clarity. The inter-model standard deviation is larger for the perturbations in CO₂ and nitrogen than for those in temperature or precipitation illustrating that the sensitivity to weather is better constrained than to management at elevated CO₂. While most model responses can readily be emulated with a simple polynomial, some response surfaces are more complicated and lead to emulation error.

continuous to allow fitting with a relatively simple functional form. In the GGCMI simulations, this condition largely but not always holds. Responses are quite diverse across locations, crops, and models, but in most cases local responses are regular enough to permit emulation. Figure 7 illustrates the ge-

ographic diversity of responses even in high-yield areas for a single crop and model (rain-fed maize in pDSSAT for various high-cultivation areas). This heterogeneity validates the choice of emulating at the grid cell level.

Each panel in Figure 7 shows model yield output from sce-

562 scenarios varying only along a single dimension (CO_2 , tempera-⁵⁹⁶
563 ture, precipitation, or nitrogen addition), with other inputs held⁵⁹⁷
564 fixed at baseline levels; in all cases yields evolve smoothly⁵⁹⁸
565 across the space sampled. For reference we show the results⁵⁹⁹
566 of the full emulation fitted across the parameter space. The⁶⁰⁰
567 polynomial fit readily captures the climatological response to⁶⁰¹
568 perturbations.⁶⁰²

569 Crop yield responses generally follow similar functional⁶⁰³
570 forms across models, though with a spread in magnitude. Fig-⁶⁰⁴
571 ure 8 illustrates the inter-model diversity of yield responses⁶⁰⁵
572 to the same perturbations, even for a single crop and location⁶⁰⁶
573 (rain-fed maize in northern Iowa, the same location shown in⁶⁰⁷
574 the Figure 7). The differences make it important to construct⁶⁰⁸
575 emulators separately for each individual model, and the fidelity⁶⁰⁹
576 of emulation can also differ across models. This figure illus-⁶¹⁰
577 trates a common phenomenon, that models differ more in re-⁶¹¹
578 sponse to perturbations in CO_2 and nitrogen perturbations than⁶¹²
579 to those in temperature or precipitation. (Compare also Figures⁶¹³
580 4 and S18.) For this location and crop, CO_2 fertilization effects⁶¹⁴
581 can range from $\sim 5\text{--}50\%$, and nitrogen responses from nearly⁶¹⁵
582 flat to a 60% drop in the lowest-application simulation.⁶¹⁶

583 While the nitrogen dimension is important and uncertain, it⁶¹⁷
584 is also the most problematic to emulate in this work because⁶¹⁸
585 of its limited sampling. The GGCMI protocol specified only⁶¹⁹
586 three nitrogen levels (10, 60 and 200 $\text{kg N y}^{-1} \text{ha}^{-1}$), so a third-⁶²⁰
587 order fit would be over-determined but a second-order fit can⁶²¹
588 result in potentially unphysical results. Steep and nonlinear de-⁶²²
589 clines in yield with lower nitrogen levels means that some re-⁶²³
590 gressions imply a peak in yield between the 100 and 200 kg N_{624}
591 $\text{y}^{-1} \text{ha}^{-1}$ levels. While there may be some reason to believe⁶²⁵
592 over-application of nitrogen at the wrong time in the growing⁶²⁶
593 season could lead to reduced yields, these features are almost⁶²⁷
594 certainly an artifact of under sampling. In addition, the polyno-⁶²⁸
595 mial fit cannot capture the well-documented saturation effect⁶²⁹

of nitrogen application (e.g. Ingestad, 1977) as accurately as
would be possible with a non-parametric model.

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 9), but some individual model-crop combinations are problematic (e.g. PROMET for rice and soy, JULES for soy and winter wheat, Figures S14–S15). Normalized errors for soy are somewhat higher across all models not because emulator fidelity is worse but because models agree more closely on yield changes for soy than for other crops (see Figure S16, lowering the denominator). Emulator performance often degrades in geographic locations where crops are not currently cultivated. Figure 10 shows a CARAIB case as an example, where emulator performance is satisfactory over cultivated areas for all crops other than soy, but uncultivated regions show some problematic areas.

It should be noted that this assessment metric is relatively forgiving. First, each emulation is evaluated against the simulation actually used to train the emulator. Had we used a spline interpolation the error would necessarily be zero. Second, the performance metric scales emulator fidelity not by the magnitude of yield changes but by the inter-model spread in those changes. Where models differ more widely, the standard for emulators becomes less stringent. Because models disagree on the magnitude of CO_2 fertilization, this effect is readily seen when comparing assessments of emulator performance in simulations at baseline CO_2 (Figure 9) with those at higher CO_2

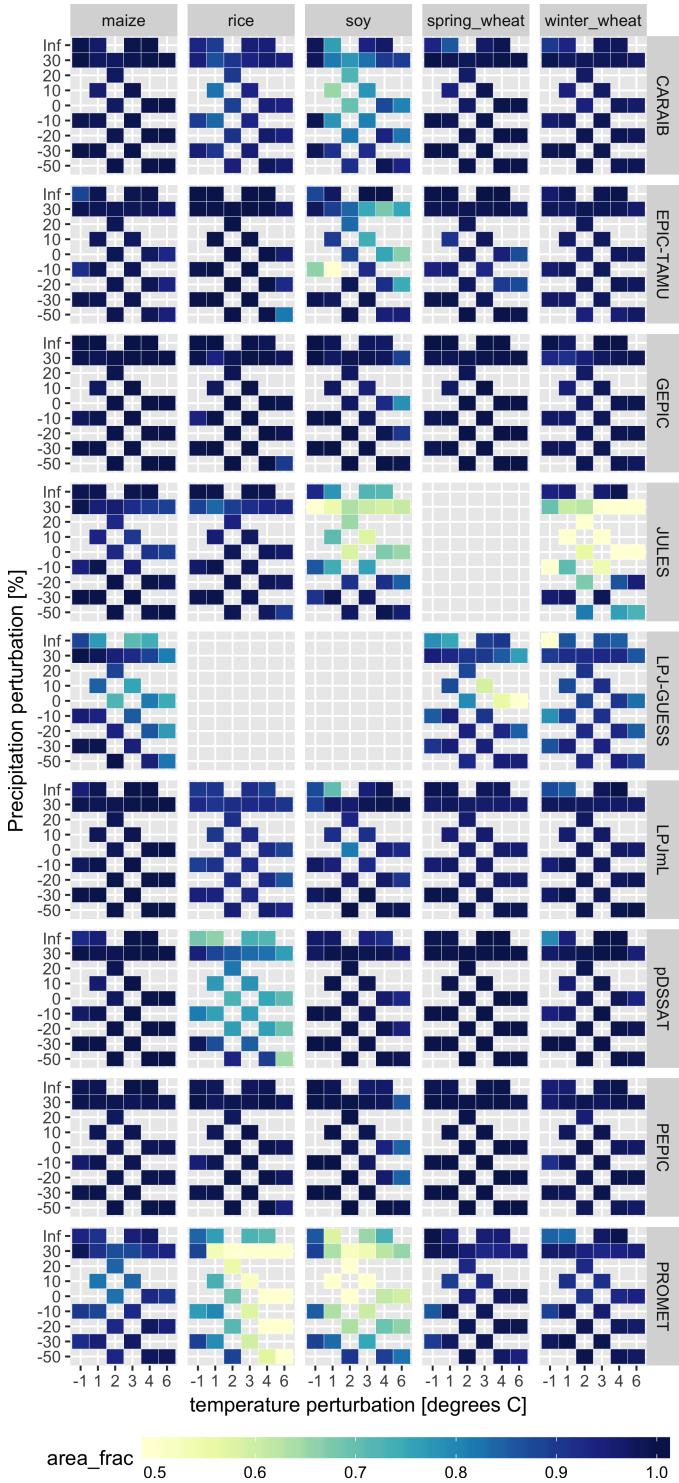


Figure 9: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO₂ and nitrogen held fixed at baseline values. Large columns are crops and large 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.

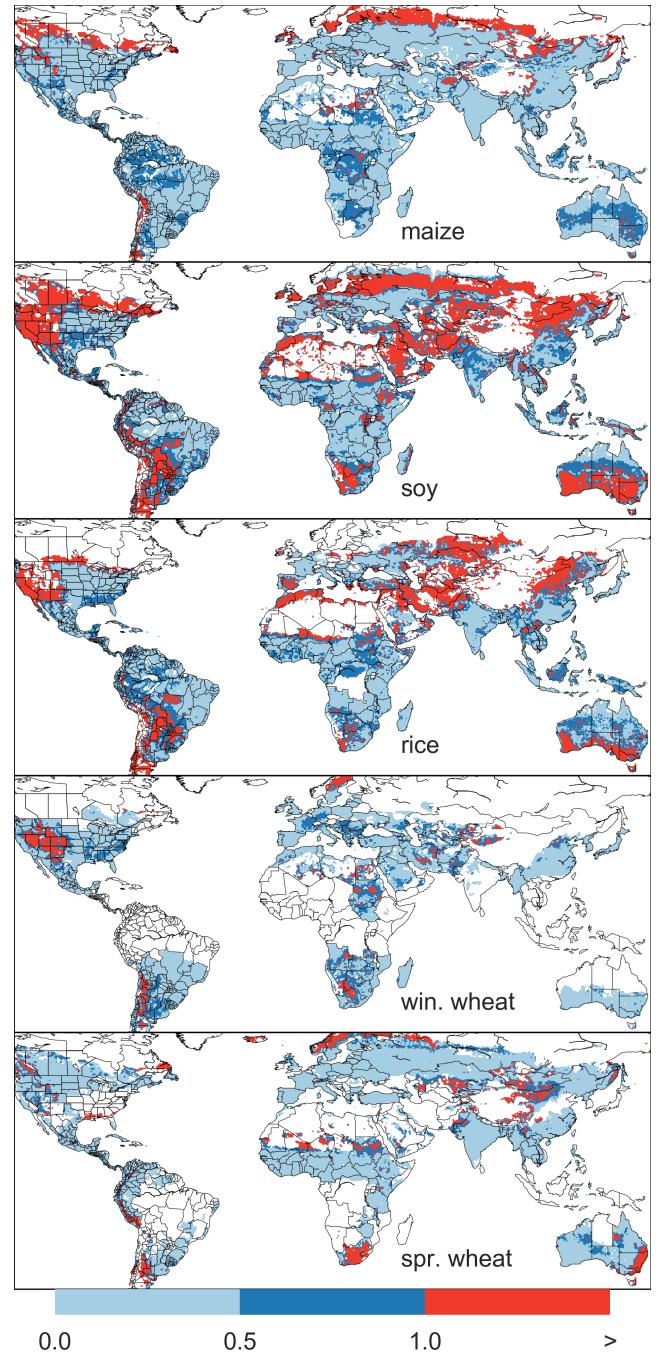


Figure 10: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error e , where $e > 1$ means that emulator error exceeds the multi-model standard deviation. White areas are those where crops 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 9.

630 levels (Figure S13). Widening the inter-model spread leads to
 631 an apparent increase in emulator skill.

632 3.4. Emulator applications

633 Because the emulator or “surrogate model” transforms the
 634 discrete simulation sample space into a continuous response
 635 surface at any geographic scale, it can be used for a variety
 636 of applications. Emulators provide an easy way to compare a
 637 ensembles of climate or impacts projections. They also pro-
 638 vide a means for generating continuous damage functions. As
 639 an example, we show a damage function constructed from 4D
 640 emulations for aggregated yield at the global scale, for maize
 641 on currently cultivated land, with simulated values shown for
 642 comparison. (Figure 11; see Figures S16- S19 in the sup-
 643 plemental material for other crops and dimensions.) The emu-
 644 lated values closely match simulations even at this aggrega-
 645 tion level. Note that these functions are presented only as
 646 examples and do not represent true global projections, be-
 647 cause they are developed from simulation data with a uniform
 648 temperature shift while increases in global mean temperature
 649 should manifest non-uniformly. The global coverage of the
 650 GGCMI simulations allows impacts modelers to apply arbitrary
 651 geographically-varying climate projections, as well as arbitrary
 652 aggregation mask, to develop damage functions for any climate
 653 scenario and any geopolitical or geographic level.

654 4. Conclusions and discussion

655 The GGCMI Phase II experiment provides a database tar-⁶⁶⁸
 656 geted to allow detailed study of crop yields from process-based⁶⁶⁹
 657 models under climate change. The experiment is designed to⁶⁷⁰
 658 facilitate not only comparing the sensitivities of process-based⁶⁷¹
 659 crop yield models to changing climate and management inputs⁶⁷²
 660 but also evaluating the complex interactions between driving⁶⁷³
 661 factors (CO_2 , temperature, precipitation, and applied nitrogen).⁶⁷⁴

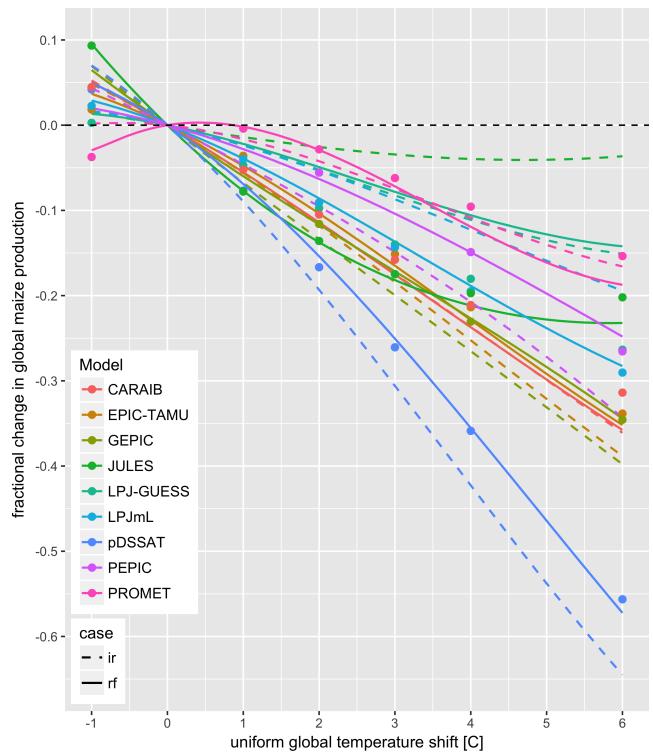


Figure 11: Global emulated damages for maize on currently cultivated lands for the GGCMI models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating up emulated values at the grid cell level, not from a regression of global mean 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.

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 understand-

675 ing the way that climate-driven changes and locations of cul-⁷⁰⁹
676 tivated land combine to produce yield impacts. One coun-⁷¹⁰
677 terintuitive result immediate apparent is that irrigated maize⁷¹¹
678 shows steeper yield reductions under warming than does rain-⁷¹²
679 fed maize when considered only over currently cultivated land.⁷¹³
680 The effect results from geographic differences in cultivation. In⁷¹⁴
681 any given location, irrigation increases crop resiliency to tem-⁷¹⁵
682 perature increase, but irrigated maize is grown in warmer loca-⁷¹⁶
683 tions where the impacts of warming are more severe (Figures⁷¹⁷
684 S5–S6). The same behavior holds for rice and winter wheat,⁷¹⁸
685 but not for soy or spring wheat (Figures S8–S10). Irrigated⁷¹⁹
686 wheat and maize are also more sensitive to nitrogen fertiliza-⁷²⁰
687 tion levels than are analogous non-irrigated crops, presumably⁷²¹
688 because those rain-fed crops are limited by water as well as⁷²²
689 nitrogen availability (Figure S19). (Soy as an efficient atmo-⁷²³
690 spheric nitrogen-fixer is relatively insensitive to nitrogen, and⁷²⁴
691 rice is not generally grown in water-limited conditions).⁷²⁵

692 Third, we show that even the relatively limited GGCMI⁷²⁶
693 Phase II sampling space allows emulation of the climatologi-⁷²⁷
694 cal response of crop models with a relatively simple reduced-⁷²⁸
695 form statistical model. The systematic parameter sampling in⁷²⁹
696 the GGCMI Phase II procedure provides information on the in-⁷³⁰
697 fluence of multiple interacting factors in a way that single pro-⁷³¹
698 jections cannot, and emulating the resulting response surface⁷³²
699 then produces a tool that can aid in both physical interpretation⁷³³
700 of the process-based models and in assessment of agricultural⁷³⁴
701 impacts under arbitrary climate scenarios. Emulating the cli-⁷³⁵
702 matological response isolates long-term impacts from any con-⁷³⁶
703 founding factors that complicate year-over-year changes, and⁷³⁷
704 the use of simple functional forms offer the possibility of phys-⁷³⁸
705 ical interpretation of parameter values. Care should be taken in⁷³⁹
706 applying relationships developed at the yearly level to shifts in⁷⁴⁰
707 the mean climatology. We anticipate that systematic parameter⁷⁴¹
708 sampling will become the norm in future model intercompari-⁷⁴²

son exercise.

While the GGCMI Phase II database should offer the foundation for multiple future studies, several cautions need to be noted. Because the simulation protocol was designed to focus on change in yield under climate perturbations and not on replicating real-world yields, the models are not formally calibrated so cannot be used for impacts projections unless in used in conjunction with historical data (or data products). Because the GGCMI simulations apply uniform perturbations to historical climate inputs, they do not sample changes in higher order moments, and cannot address the additional crop yield impacts of potential changes in climate variability. Although distributional changes in model projections are fairly uncertain 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

743 root biomass (as many as 25 output variables for some models).⁷⁷⁶ bon Cycle and Climate Interactions).

744 Emulation studies that are possible include a more systematic
745 evaluation of different statistical model specifications and for-⁷⁷⁷

746 mal calculation of uncertainties in derived parameters.

747 The future of food security is one of the larger challenges
748 facing humanity at present. The development of multi-model⁷⁸⁰
749 ensembles such as GGCMI Phase II provides a way to begin⁷⁸¹
750 to better understand crop responses to a range of potential cli-⁷⁸²
751 mate inputs, improve process based models, and explore the po-⁷⁸³
752 tential benefits of adaptive responses included shifting growing⁷⁸⁴
753 season, cultivar types and cultivar geographic extent.⁷⁸⁶

754 5. Acknowledgments

755 We thank Michael Stein and Kevin Schwarzwald, who pro-⁷⁹¹
756 vided helpful suggestions that contributed to this work. This re-⁷⁹²
757 search was performed as part of the Center for Robust Decision-⁷⁹³
758 Making on Climate and Energy Policy (RDCEP) at the Univer-⁷⁹⁴
759 sity of Chicago, and was supported through a variety of sources.⁷⁹⁵

760 RDCEP is funded by NSF grant #SES-1463644 through the⁷⁹⁷
761 Decision Making Under Uncertainty program. J.F. was sup-⁷⁹⁸
762 ported by the NSF NRT program, grant #DGE-1735359. C.M.⁷⁹⁹
763 was supported by the MACMIT project (01LN1317A) funded⁸⁰⁰
764 through the German Federal Ministry of Education and Re-⁸⁰²
765 search (BMBF). C.F. was supported by the European Research⁸⁰³
766 Council Synergy grant #ERC-2013-SynG-610028 Imbalance-⁸⁰⁴
767 P. P.F. and K.W. were supported by the Newton Fund through⁸⁰⁶
768 the Met Office Climate Science for Service Partnership Brazil⁸⁰⁷
769 (CSSP Brazil). A.S. was supported by the Office of Science⁸⁰⁸
770 of the U.S. Department of Energy as part of the Multi-sector⁸⁰⁹
771 Dynamics Research Program Area. Computing resources were⁸¹¹
772 provided by the University of Chicago Research Computing⁸¹²
773 Center (RCC). S.O. acknowledges support from the Swedish⁸¹³
774 strong research areas BECC and MERGE together with sup-⁸¹⁴
775 port from LUCCI (Lund University Centre for studies of Car-⁸¹⁶

776 6. References

- 778 Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013).
779 Implication of crop model calibration strategies for assessing regional im-
780 pacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.
- 781 Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D.,
782 A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Al-
783 derman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Bier-
784 nath, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising tempera-
785 tures reduce global wheat production. *Nature Climate Change*, 5, 143–147.
786 doi:10.1038/nclimate2470.
- 787 Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A.,
788 J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N.,
789 Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath,
790 C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simula-
791 tion wheat yields under climate change. *Nature Climate Change*, 3, 827832.
792 doi:10.1038/nclimate1916.
- 793 Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other
794 Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Ef-
795 ficiency, Carbon Sequestration, and Environmental Pollution. *Advances in*
796 *Agronomy*, 86, 341 – 409.
- 797 Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov,
798 N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat
799 production potentials and management flexibility under the representative
800 concentration pathways. *Global and Planetary Change*, 122, 107 – 121.
- 801 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields
802 from global gridded crop models. *Agricultural and Forest Meteorology*, 236,
803 145 – 161.
- 804 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded
805 crop models using statistical estimates. *Agricultural and Forest Meteorol-*
806 *ogy*, 214-215, 134 – 147.
- 807 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S.
808 (2018). Implementing the Nitrogen cycle into the dynamic global vegeta-
809 tion, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific*
810 *Model Development*, 11, 2789–2812.
- 811 Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L.,
812 & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections
813 Based on Precomputed GCM Runs. *Journal of Climate*, 27, 1829–1844.
- 814 Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.
815 (2014). A meta-analysis of crop yield under climate change and adaptation.
816 *Nature Climate Change*, 4, 287 – 291.

- 817 Chang, W., Stein, M., Wang, J., Kotamarthi, V., & Moyer, E. (2016). Changes in⁸⁶⁰
 818 spatio-temporal precipitation patterns in changing climate conditions. *Journal of Climate*, 29. doi:10.1175/JCLI-D-15-0844.1.⁸⁶¹
 819
 820 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process⁸⁶³
 821 emulation of dynamic computer codes. *Biometrika*, 96, 663–676.⁸⁶⁴
 822 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In⁸⁶⁵
 823 C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth*,⁸⁶⁶
 824 Duke University, Durham, North Carolina (pp. 115–118). Durham, North⁸⁶⁷
 825 Carolina.⁸⁶⁸
 826 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for⁸⁶⁹
 827 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).⁸⁷⁰
 828 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdous,⁸⁷¹
 829 M., & François, L. (2011). Responses of European forest ecosystems to⁸⁷²
 830 21st century climate: assessing changes in interannual variability and fire⁸⁷³
 831 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).⁸⁷⁴
 832 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,⁸⁷⁵
 833 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating⁸⁷⁶
 834 impact models and sectors (pSIMS). *Environmental Modelling and Software*,⁸⁷⁷
 835 62, 509–516.⁸⁷⁸
 836 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,⁸⁷⁹
 837 Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C.,⁸⁸⁰
 838 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.,⁸⁸¹
 839 (2015). The Global Gridded Crop Model Intercomparison: data and mod-⁸⁸²
 840 eling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8,⁸⁸³
 841 261–277.⁸⁸⁴
 842 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J.,⁸⁸⁵
 843 & Taylor, K. E. (2016). Overview of the coupled model intercomparison⁸⁸⁶
 844 project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958.⁸⁸⁷
 845
 846 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-⁸⁸⁹
 847 mate change impacts on durum wheat in the mediterranean region. *Natural Hazards and Earth System Sciences*, 11, 1293–1302.⁸⁹⁰
 848
 849 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Re-⁸⁹²
 850 gionalization of a large-scale crop growth model for sub-Saharan Africa:⁸⁹³
 851 Model setup, evaluation, and estimation of maize yields. *Agriculture, Ecosystems & Environment*, 151, 21 – 33.⁸⁹⁴
 852
 853 Food and Agriculture Organization of the United Nations (2018). FAOSTAT⁸⁹⁶
 854 database. URL: <http://www.fao.org/faostat/en/home>.⁸⁹⁷
 855 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T.,⁸⁹⁸
 856 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P.,⁸⁹⁹
 857 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F.,⁹⁰⁰
 858 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,⁹⁰¹
 859 Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J.,⁹⁰²
 860 Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nen-
 861 del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna,
 862 Semenov, M. A., Slawinski, C., Strattonovich, P., Supit, I., Waha, K.,
 863 Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model
 864 wheat yield impact response surfaces showing sensitivity to temperature and
 865 precipitation change. *Agricultural Systems*, 159, 209–224.
 866 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014).
 867 Evaluating the utility of dynamical downscaling in agricultural impacts pro-
 868 jections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781.
 869 Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitiv-
 870 ity of Agricultural Model Performance to Different Climate Inputs. *Journal of Applied Meteorology and Climatology*, 55, 151113145618001.
 871 Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported
 872 Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous
 873 Crop Growth and Yield: Application for Wheat in Central Europe. *Remote
 874 Sensing*, 7, 3934–3965.
 875 Haugen, M., Stein, M., Moyer, E., & Sriver, R. (2018). Estimating changes in
 876 temperature distributions in a large ensemble of climate simulations using
 877 quantile regression. *Journal of Climate*, 31, 8573–8588. doi:10.1175/JCLI-
 878 D-17-0782.1.
 879 He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H.,
 880 He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water
 881 contents and nitrogen leaching to precipitation, management practices and
 882 soil hydraulic properties in semi-arid and humid regions of Canada using the
 883 DSSAT model. *Nutrient Cycling in Agroecosystems*, 106, 201–215.
 884 Heady, E. O. (1957). An Econometric Investigation of the Technology of Agri-
 885 cultural Production Functions. *Econometrica*, 25, 249–268.
 886 Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa
 887 State University Press.
 888 Holden, P., Edwards, N., PH, G., Fraedrich, K., Lunkeit, F., E, K., Labriet,
 889 M., Kanudia, A., & F, B. (2014). Plasim-entsem v1.0: A spatiotemporal
 890 emulator of future climate change for impacts assessment. *Geoscientific Model Development*, 7, 433–451. doi:10.5194/gmd-7-433-2014.
 891 Holzkämper, A., Calanca, P., & Fuhrer, J. (2012). Statistical crop models:
 892 Predicting the effects of temperature and precipitation changes. *Climate Research*, 51, 11–21.
 893 Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I.,
 894 McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore,
 895 A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake,
 896 A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalglish,
 897 N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp,
 898 J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson,
 899 M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T.,
 900

- 903 Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorf, J.,⁹⁴⁶
 904 Wedgwood, S., & Keating, B. A. (2014). APSIM Evolution towards a new⁹⁴⁷
 905 generation of agricultural systems simulation. *Environmental Modelling and⁹⁴⁸*
 906 *Software*, 62, 327 – 350.⁹⁴⁹
- 907 Howden, S., & Crimp, S. (2005). Assessing dangerous climate change impacts⁹⁵⁰
 908 on australia's wheat industry. *Modelling and Simulation Society of Australia⁹⁵¹*
 909 and New Zealand, (pp. 505–511).⁹⁵²
- 910 Izumi, T., Nishimori, M., & Yokozawa, M. (2010). Diagnostics of climate⁹⁵³
 911 model biases in summer temperature and warm-season insolation for the⁹⁵⁴
 912 simulation of regional paddy rice yield in japan. *Journal of Applied Meteo⁹⁵⁵*
 913 *rology and Climatology*, 49, 574–591.⁹⁵⁶
- 914 Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Ni⁹⁵⁷
 915 trogen Fertilizers. *Ambio*, 6, 146–151.⁹⁵⁸
- 916 Izaurralde, R., Williams, J., McGill, W., Rosenberg, N., & Quiroga Jakas, M.⁹⁵⁹
 917 (2006). Simulating soil C dynamics with EPIC: Model description and test⁹⁶⁰
 918 ing against long-term data. *Ecological Modelling*, 192, 362–384.⁹⁶¹
- 919 Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the⁹⁶²
 920 CROPGRO-soybean model to predict regional yield and production. *Agri⁹⁶³*
 921 *culture, Ecosystems & Environment*, 93, 73 – 85.⁹⁶⁴
- 922 Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L.,⁹⁶⁵
 923 Wilkens, P., Singh, U., Gijsman, A., & Ritchie, J. (2003). The DSSAT⁹⁶⁶
 924 cropping system model. *European Journal of Agronomy*, 18, 235 – 265.⁹⁶⁷
- 925 Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I.,⁹⁶⁸
 926 Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A.,⁹⁶⁹
 927 Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017).⁹⁷⁰
 928 Toward a new generation of agricultural system data, models, and knowl⁹⁷¹
 929 edge products: State of agricultural systems science. *Agricultural Systems*,⁹⁷²
 930 155, 269 – 288.⁹⁷³
- 931 Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth,⁹⁷⁴
 932 D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Ver⁹⁷⁵
 933 burg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K.,⁹⁷⁶
 934 Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003).⁹⁷⁷
 935 An overview of APSIM, a model designed for farming systems simulation.⁹⁷⁸
 936 *European Journal of Agronomy*, 18, 267 – 288.⁹⁷⁹
- 937 Leeds, W. B., Moyer, E. J., & Stein, M. L. (2015). Simulation of future⁹⁸⁰
 938 climate under changing temporal covariance structures. *Advances in⁹⁸¹*
 939 *Statistical Climatology, Meteorology and Oceanography*, 1, 1–14. URL:⁹⁸²
 940 <https://www.adv-stat-clim-meteorol-oceanogr.net/1/1/2015/>⁹⁸³
 941 doi:10.5194/ascmo-1-1-2015.⁹⁸⁴
- 942 Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., &⁹⁸⁵
 943 Smith, B. (2013). Implications of accounting for land use in simulations of⁹⁸⁶
 944 ecosystem carbon cycling in Africa. *Earth System Dynamics*, 4, 385–407.⁹⁸⁷
- 945 Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling⁹⁸⁸
 wheat yield and crop water productivity with high resolution on a global
 scale. *Agricultural Systems*, 94, 478 – 493.
- Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a). Global investigation of impacts of PET methods on simulating crop-water relations for maize. *Agricultural and Forest Meteorology*, 221, 164 – 175.
- Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs with yields from major crop cultivations. *Science of The Total Environment*, 572, 526 – 537.
- Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, 150, 1443 – 1452.
- Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2, 014002.
- MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447.
- Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P., Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi, P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano, D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of large simulated yield datasets for studying climate effects. (p. 1100). doi:10.13140/RG.2.1.5173.8328.
- Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrological modelling to study the impact of climate change on the water flows of mountain watersheds. *Journal of Hydrology*, 376, 362 – 377.
- Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., & Calzadilla, A. (2009). Global biomass production potentials exceed expected future demand without the need for cropland expansion. *Nature Communications*, 6.
- McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srinivas, C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S. (2015). Integrated assessments of the impacts of climate change on agriculture: An overview of AgMIP regional research in South Asia. *Chapter in: Handbook of Climate Change and Agroecosystems*, (pp. 201–218).
- Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical weather responsiveness of crop yields: US evidence and implications for the agricultural impacts of climate change. *Environmental Research Letters*, 12.
- Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, 8.
- Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Izumi, T., Izaurralde,

- R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T.,
 A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G.,
 Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H.,
 (2017). Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, *10*, 1403–1422.
- Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T.,
 (1997). Effect of CO₂ enrichment on carbon and nitrogen interaction in wheat and soybean. *Soil Science and Plant Nutrition*, *43*, 789–798.
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, *91*, 1290 – 1300.
- Olin, S., Schurges, G., Lindeskog, M., Wårlind, D., Smith, B., Bodin, P.,
 Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue C:N to changes in atmospheric CO₂ and N management in the main wheat regions of western europe. *Biogeosciences*, *12*, 2489–2515. doi:10.5194/bg-12-2489-2015.
- Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in field crop production. *Soil Science and Plant Nutrition*, *38*, 553–564.
- Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., & Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator. *Geoscientific Model Development*, *8*, 1139–1155.
- Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop yields and their variability at different levels of global warming. *Earth System Dynamics*, *9*, 479–496.
- Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff, S., & Gerten, D. (2015). Emulating global climate change impacts on crop yields. *Statistical Modelling*, *15*, 499–525.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pasos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
- Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minetti, J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M., Semenov, M., Slawinski, C., Strattonovich, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. *Climate Research*, *65*, 87–105.
- Poppick, A., McInerney, D. J., Moyer, E. J., & Stein, M. L. (2016). Temperatures in transient climates: Improved methods for simulations with evolving temporal covariances. *Ann. Appl. Stat.*, *10*, 477–505. URL: <https://doi.org/10.1214/16-AOAS903>. doi:10.1214/16-AOAS903.
- Porter et al. (IPCC) (2014). Food security and food production systems. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In C. F. et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK: Cambridge University Press.
- Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of monthly growing areas of 26 irrigated crops.
- Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly Irrigated and Rainfed crop Areas around the Year 2000: A New High-Resolution Data Set for Agricultural and Hydrological Modeling. *Global Biogeochemical Cycles*, *24*, GB1011.
- Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E., & Arneth, A. (2016). Climate analogues suggest limited potential for intensification of production on current croplands under climate change. *Nature Communications*, *7*, 12608.
- Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus A: Dynamic Meteorology and Oceanography*, *58*, 461–472.
- Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the reduction and sensitivity analysis of complex environmental models. *Environmental Modelling & Software*, *34*, 1 – 4.
- Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, *48*.
- Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, *12*.
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, *111*, 3268–3273.
- Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P., Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and

- 1075 pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182. 1118
- 1076 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. 1127
- 1085 Ruane, A., I. Hudson, N., Asseng, S., Camarrano, D., Ewert, F., Martre, P., 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. 1132
- 1091 Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason-D'Croz, D., Müller, C., Porter, C., Phillips, M. M., Raymundo, R. M., Sands, R., Valdivia, R. O., White, J. W., Wiebe, K., & Rosenzweig, C. (2018). Biophysical and economic implications for agriculture of +1.5° and +2.0°C global warming using AgMIP Coordinated Global and Regional Assessments. *Climate Research*, 76, 17–39. 1137
- 1097 Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown, D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013). Climate change impact uncertainties for maize in panama: Farm information, climate projections, and yield sensitivities. *Agricultural and Forest Meteorology*, 170, 132 – 145. 1143
- 1102 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. 1147
- 1108 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. 1152
- 1113 Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift*, 19, 135–141. 1155
- 1118 Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19, 607–620. 1158
- 1123 Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Cais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolin, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frieler, K. (2017). Consistent negative response of US crops to high temperatures in observations and crop models. *Nature Communications*, 8, 13931.
- 1128 Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106, 15594–15598.
- 1133 Snyder, A., Calvin, K. V., Phillips, M., & Ruane, A. C. (2018). A crop yield change emulator for use in gcam and similar models: Persephone v1.0. *Geoscientific Model Development Discussions*, 2018, 1–42.
- 1138 Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety*, 94, 1735 – 1763.
- 1143 Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93, 485–498.
- 1148 Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters*, 35.
- 1153 Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., & Martin, J. F. (2014). Modeling sugarcane yield with a process-based model from site to continental scale: Uncertainties arising from model structure and parameter values. *Geoscientific Model Development*, 7, 1225–1245.
- 1158 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences*, 111, 3228–3232.
- 1163 White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Methodologies for simulating impacts of climate change on crop production. *Field Crops Research*, 124, 357 – 368.
- 1168 Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T., Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance against site observations of irrigated maize from Mead, Nebraska. *Geoscientific Model Development*, 10, 1291–1320.
- 1173 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield variability in JULES-crop and implications for forcing with seasonal weather forecasts. *Geoscientific Model Development*, 8, 3987–3997.
- 1178 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Landbouwkundige Onderzoeken* : 64.6, .
- 1183 Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato yields on changes in climate and co2. *Agricultural and Forest Meteorology*, 112, 217 – 231.
- 1188 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao,

1161 Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A.,
1162 Li, T., Lin, E., Liu, Q., Martre, P., Miller, C., Peng, S., Peuelas, J., Ruane,
1163 A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng,
1164 S. (2017). Temperature increase reduces global yields of major crops in four
1165 independent estimates. *Proc. Natl. Acad. Sci.*, 114, 9326–9331.