

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

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

<sup>a</sup>Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA

<sup>b</sup>Center for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA

<sup>c</sup>Department of Computer Science, University of Chicago, Chicago, IL, USA

<sup>d</sup>Potsdam Institute for Climate Impact Research, Leibniz Association (Member), Potsdam, Germany

<sup>e</sup>NASA Goddard Institute for Space Studies, New York, NY, United States

<sup>f</sup>Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA

<sup>g</sup>Ecosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria

<sup>h</sup>Department of Soil Science, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovak Republic

<sup>i</sup>Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France

<sup>j</sup>Sino-French Institute of Earth System Sciences, College of Urban and Environmental Sciences, Peking University, Beijing, China

<sup>k</sup>Unité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d'Astrophysique et de Géophysique, University of Liège, Belgium

<sup>l</sup>Met Office Hadley Centre, Exeter, United Kingdom

<sup>m</sup>Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany

<sup>n</sup>Georg-August-University Göttingen, Tropical Plant Production and Agricultural Systems Modelling, Göttingen, Germany

<sup>o</sup>Department of Geographical Sciences, University of Maryland, College Park, MD, USA

<sup>p</sup>Texas AgriLife Research and Extension, Texas A&M University, Temple, TX, USA

<sup>q</sup>Department of Statistics, University of Chicago, Chicago, IL, USA

<sup>r</sup>EWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

<sup>s</sup>Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

<sup>t</sup>Earth Institute Center for Climate Systems Research, Columbia University, New York, NY, USA

<sup>u</sup>Karlsruhe Institute of Technology, IMK-IFU, 82467 Garmisch-Partenkirchen, Germany.

<sup>v</sup>School 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. Results of these models are often used to inform subsequent model-based analyses in agricultural economics and integrated assessment, but the computational requirements of process-based models sometimes hamper the ability to utilize these models in existing impact assessment frameworks at high resolutions globally. Conducting a large simulation set with perturbations in atmospheric CO<sub>2</sub> concentrations, temperature, precipitation, and nitrogen inputs, Phase II of the Global Gridded Crop Model Intercomparison (GGCMI), an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP), constitutes an unprecedentedly data-rich basis of projected yield changes across 12 models and five crops (maize, soy, rice, spring wheat, and winter wheat) using global gridded simulations. Results from Phase II of the GGCMI effort, a targeted experiment aimed at understanding the interaction between multiple climate variables (as well as management) are presented first. Then the construction of an “emulator or statistical representation of the simulated 30-year mean climatological output in each location is outlined. The emulated response surfaces capture the details of the process-based models in a lightweight, computationally tractable form that facilitates model comparison as well as potential applications in subsequent modeling efforts such as integrated assessment.

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

## 1. Introduction

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

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

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

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

40 Process models do continue to struggle with some important  
41 details, including reproducing historical year-to-year variabil-  
42 ity (e.g. Müller et al., 2017), reproducing historical yields when  
43 driven by reanalysis weather (e.g. Glotter et al., 2014), and low  
44 sensitivity to extreme events (e.g. Glotter et al., 2015). These  
45 issues are driven in part by the diversity of new cultivars and  
46 genetic variants, which outstrips the ability of academic mod-  
47eling groups to capture them (e.g. Jones et al., 2017). Mod-  
48 els do not simulate many additional factors affecting produc-  
49 tion, including pests/diseases/weeds. For these reasons, indi-  
50 vidual studies must generally re-calibrate models to ensure that  
51 short-term predictions reflect current cultivar mixes, and long-  
52 term projections retain considerable uncertainty (Wolf & Oijen,  
53 2002, Jagtap & Jones, 2002, Angulo et al., 2013, Asseng et al.,  
54 2013, 2015). Inter-model discrepancies can also be high in ar-  
55 eas not yet cultivated (e.g. Challinor et al., 2014, White et al.,  
56 2011). Finally, process-based models present additional diffi-  
57 culties for high-resolution global studies because of their com-  
58 plexity and computational requirements. For economic impacts  
59 assessments, it is often impossible to integrate a set of process-  
60 based crop models directly into an integrated assessment model  
61 to estimate the potential cost of climate change to the agricul-  
62 tural sector.

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

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

models cannot apply. Yield data are also often limited in the developing world, where future climate impacts may be the most critical. Second, only process-based models can capture the growth response to elevated CO<sub>2</sub>, novel conditions that are not represented in historical data (e.g. Pugh et al., 2016, Roberts et al., 2017). Similarly, only process-based models can represent novel changes in management practices (e.g. fertilizer input) that may ameliorate climate-induced damages.

Statistical emulation of crop simulations offers the possibility of combining some advantageous features of both statistical and process-based models. The statistical representation of complicated numerical simulation (e.g. O'Hagan, 2006, Conti et al., 2009), in which simulation output acts as the training data for a statistical model, has been of increasing interest with the growth of simulation complexity and volume of output. Such emulators or "surrogate models" have been used in a variety of fields including hydrology (Razavi et al., 2012), engineering (Storlie et al., 2009), environmental sciences (Ratto et al., 2012), and climate (Castruccio et al., 2014). For agricultural impacts studies, emulation of process-based models allows exploring crop yields in regions outside ranges of current cultivation and with input variables outside historical precedents, in a lightweight, flexible form that is compatible with economic studies.

Crop yield emulators have been proposed and implemented by many studies (e.g. Howden & Crimp, 2005, Räisänen & Ruokolainen, 2006, Lobell & Burke, 2010, Iizumi et al., 2010, Ferrise et al., 2011, Holzkämper et al., 2012, Ruane et al., 2013, Makowski et al., 2015), and in the last several years multiple studies have developed emulators based on a variety of crop simulation model outputs. Several studies have developed an emulator for a single crop model run on a RCP climate scenario set (e.g. Oyebamiji et al., 2015). Multiple groups (e.g. Blanc & Sultan, 2015, Blanc, 2017, Ostberg et al., 2018), successfully constructed emulators for a 5-crop-model intercomparison exercise performed as part of ISIMIP (Warszawski et al., 2014), the Inter-Sectoral Impacts Model Intercomparison Project and evaluated several different climate scenarios (over multiple climate model runs). Several other studies (e.g. Moore et al., 2017, Mistry et al., 2017) utilize a hybrid simulation output and real-world data approach to develop and emulator or damage

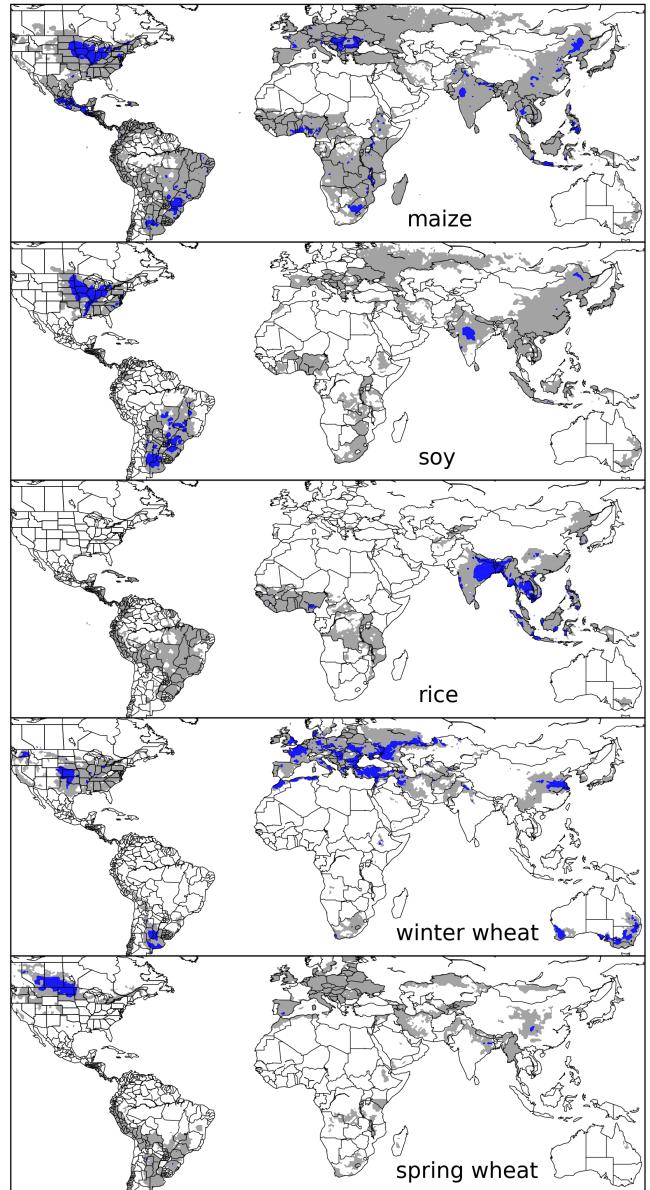


Figure 1: Presently cultivated area for rain-fed crops in the real world. Blue contour area indicates grid cells with more than 20,000 hectares (approx. 10% of the land in a grid cell near the equator for example) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated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

ercise performed as part of ISIMIP (Warszawski et al., 2014), the Inter-Sectoral Impacts Model Intercomparison Project and evaluated several different climate scenarios (over multiple climate model runs). Several other studies (e.g. Moore et al., 2017, Mistry et al., 2017) utilize a hybrid simulation output and real-world data approach to develop and emulator or damage

function. Additional recent studies have explored an impact response surface (aka. emulator when using simulated data) over an explicit multivariate input simulation space (as opposed to specific RCP climate model runs), with a site-based approach (as opposed to a globally gridded model) across temperature, water, and CO<sub>2</sub> sampling (Snyder et al., 2018), or with models for wheat across water and temperature dimensions for different sites in Europe (Fronzek et al., 2018).

The Global Gridded Crop Model Intercomparison (GGCMI) Phase II experiment is an attempt to expand upon previous process-based crop modeling studies by running globally grid-ded crop models over a set of uniform input dimensions as opposed to RCP climate scenarios in order to focus on testing the sensitivity to yield drivers within and across models. GGCMI is a multi-model exercise conducted as part of the Agricultural Model Intercomparison and Improvement Project (AgMIP, Rosenzweig et al., 2013, 2014)), which brings together major global crop simulation models from different research organizations around the world under a framework similar to the Climate Model Intercomparison Project (CMIP, Taylor et al., 2012, Eyring et al., 2016). The GGCMI analysis framework builds on the AgMIP Coordinated Climate-Crop Modeling Project (C3MP, Ruane et al., 2014, McDermid et al., 2015), and will contribute to the AgMIP Coordinated Global and Regional Assessments (CGRA, Ruane et al., 2018, Rosenzweig et al., 2018).

The GGCMI Phase II project develops global simulations of yields of major crops under scenarios that sample a uniform

parameter space. Overall goals include understanding where highest-yield regions may shift under climate change, exploring future adaptive management strategies, understanding how interacting parameters affect crop yields, quantifying uncertainties, and testing strategies for producing lightweight statistical emulations of the more detailed process-based models. In the remainder of this paper, we describe the GGCMI Phase II experiments, present the simulation database output (for public use) and initial overall results. We also present an example climatological-mean yield emulator as a distillation of the dataset and as a potential tool for impact assessments.

We do not present all the final insights to be gained from this model intercomparison project, or the best possible emulation of the year-to-year response to changes the input dimensions.

## 2. Materials and Methods

### 2.1. GGCMI Phase II: experiment design

GGCMI Phase II is the continuation of a multi-model comparison exercise begun in 2014. The initial Phase I compared harmonized yields of 21 models for 19 crops over a historical (1980-2010) scenario with a primary goal of model evaluation (Elliott et al., 2015, Müller et al., 2017). Phase II compares simulations of 12 models for 5 crops (maize, rice, soybean, spring wheat, and winter wheat) over hundreds of scenarios in which individual climate or management inputs are adjusted from their historical values. The reduced set of crops includes the three major global cereals and the major legume and accounts

Input variable	Abbr.	Tested range	Unit
CO <sub>2</sub>	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 <sub>inf</sub> )	%
Applied nitrogen	N	10, 60, 200	kg ha <sup>-1</sup>

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<sub>inf</sub>) simulations.

161 for over 50% of human calories (in 2016, nearly 3.5 billion tons<sup>194</sup>  
162 or 32% of total global crop production by weight (Food and<sup>195</sup>  
163 Agriculture Organization of the United Nations, 2018).<sup>196</sup>

164 The major goals of GGCMI Phase II are to:<sup>197</sup>

- 165 • Enhance understanding of how models work by character-<sup>198</sup>  
166 izing their sensitivity to input climate and nitrogen drivers.<sup>199</sup>
- 167 • Study the interactions between climate variables and nitro-<sup>200</sup>  
168 gen inputs in driving modeled yield impacts.<sup>201</sup>
- 169 • Explore differences in crop response to warming across the<sup>202</sup>  
170 Earth's climate regions.<sup>203</sup>
- 171 • Provide a dataset that allows statistical emulation of crop<sup>204</sup>  
172 model responses for downstream modelers.<sup>205</sup>
- 173 • Illustrate differences in potential adaptation via growing<sup>206</sup>  
174 season changes.<sup>207</sup>

175 The guiding scientific rationale of GGCMI Phase II is to pro-

176 vide a comprehensive, systematic evaluation of the response<sup>210</sup>  
177 of process-based crop models to different values for carbon<sup>211</sup>  
178 dioxide, temperature, water, and applied nitrogen (collectively<sup>212</sup>  
179 known as "CTWN"). Phase II of the GGCMI project consists<sup>213</sup>  
180 of a series of simulations, each with one or more of the CTWN<sup>214</sup>  
181 dimensions perturbed over the 31-year historical time series<sup>215</sup>  
182 (1980-2010) used in Phase I. In most cases, historical daily cli-<sup>216</sup>  
183 mate inputs are taken from the 0.5 degree NASA AgMERRA<sup>217</sup>  
184 daily gridded re-analysis product specifically designed for agri-<sup>218</sup>  
185 cultural modeling, with satellite-corrected precipitation (Ruane<sup>219</sup>  
186 et al., 2015). Two models require sub-daily input data and use<sup>220</sup>  
187 alternative sources. See Elliott et al. (2015) for additional de-<sup>221</sup>  
188 tails.<sup>222</sup>

189 The experimental protocol consists of 9 levels for precipita-<sup>223</sup>  
190 tion perturbations, 7 for temperature, 4 for CO<sub>2</sub>, and 3 for ap-<sup>224</sup>  
191 plied nitrogen, for a total of 672 simulations for rain-fed agri-<sup>225</sup>  
192 culture and an additional 84 for irrigated (Table 1). For irri-<sup>226</sup>  
193 gated simulations, soil water is held at either field capacity or,<sup>227</sup>

for those models that include water-log damage, at maximum  
196 beneficial level. Temperature perturbations are applied as ab-  
197 solute offsets from the daily mean, minimum, and maximum  
198 temperature time series for each grid cell used as inputs. Pre-  
199 cipitation perturbations are applied as fractional changes at the  
grid cell level, and carbon dioxide and nitrogen levels are spec-  
ified as discrete values applied uniformly over all grid cells.  
200 Note that CO<sub>2</sub> changes are applied independently of changes  
in climate variables, so that higher CO<sub>2</sub> is not associated with  
201 higher temperatures. An additional, identical set of scenarios  
202 (at the same C, T, W, and N levels) simulate adaptive agron-  
203 omy under climate change by varying the growing season for  
204 crop production. (These adaptation simulations are not shown  
205 or analyzed here.) The resulting GGCMI data set captures a  
206 distribution of crop responses over the potential space of future  
207 climate conditions.<sup>208</sup>

The 12 models included in GGCMI Phase II are all mechan-  
istic process-based crop models that are widely used in im-  
pacts assessments (Table 2). Although some of the models  
shares a common base (e.g. LPJmL and LPJ-GUESS and the  
EPIC models), they have developed independently from this  
shared base, for more details on the genealogy of the mod-  
els see Figure S1 in Rosenzweig et al. (2014). Differences in  
model structure does mean that several key factors are not stan-  
dardized across the experiment, including secondary soil nutri-  
ents, carry over effects across growing years including residue  
management and soil moisture, and extent of simulated area for  
different crops. Growing seasons are identical across models,  
but vary by crop and by location on the globe. All stresses  
except factors related to nitrogen, temperature, and water (e.g.  
Alkalinity, salinity) are disabled. No additional nitrogen inputs,  
such as atmospheric deposition, are considered, but some mod-  
els have individual assumptions on soil organic matter that may  
release additional nitrogen through mineralization. See Rosen-

Model (Key Citations)	Maize	Soy	Rice	Winter Wheat	Spring Wheat	N Dim.	Simulations per Crop
<b>APSIM-UGOE</b> , Keating et al. (2003), Holzworth et al. (2014)	X	X	X	–	X	Yes	37
<b>CARAIB</b> , Dury et al. (2011), Pirttioja et al. (2015)	X	X	X	X	X	No	224
<b>EPIC-IIASA</b> , Balkovi et al. (2014)	X	X	X	X	X	Yes	35
<b>EPIC-TAMU</b> , Izaurrealde et al. (2006)	X	X	X	X	X	Yes	672
<b>JULES*</b> , Osborne et al. (2015), Williams & Falloon (2015), Williams et al. (2017)	X	X	X	–	X	No	224
<b>GEPIC</b> , Liu et al. (2007), Folberth et al. (2012)	X	X	X	X	X	Yes	384
<b>LPJ-GUESS</b> , Lindeskog et al. (2013), Olin et al. (2015)	X	–	–	X	X	Yes	672
<b>LPJmL</b> , von Bloh et al. (2018)	X	X	X	X	X	Yes	672
<b>ORCHIDEE-crop</b> , Valade et al. (2014)	X	–	X	–	X	Yes	33
<b>pDSSAT</b> , Elliott et al. (2014), Jones et al. (2003)	X	X	X	X	X	Yes	672
<b>PEPIC</b> , Liu et al. (2016a,b)	X	X	X	X	X	Yes	130
<b>PROMET*†</b> , 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 zweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017)<sup>246</sup> part of the experiment variable space. Most modeling groups  
229 for further details on models and underlying assumptions. <sup>247</sup> simulate all five crops in the protocol, but some omitted one

230 Each model is run at 0.5 degree spatial resolution and covers <sup>248</sup> or more. Table 2 provides details of coverage for each model.  
231 all currently cultivated areas and much of the uncultivated land <sup>249</sup> Note that the three models that provide less than 50 simulations  
232 area. Coverage extends considerably outside currently culti- <sup>250</sup> are excluded from the emulator analysis.  
233 vated areas because cultivation will likely shift under climate <sup>251</sup> All models produce as output, crop yields ( $\text{tons ha}^{-1} \text{ year}^{-1}$ )  
234 change. See Figure 1 for the present-day cultivated area of <sup>252</sup> for each 0.5 degree grid cell. Because both yields and yield  
235 rain-fed crops, and Figure S1 in the supplemental material for <sup>253</sup> changes vary substantially across models and across grid cells,  
236 irrigated crops. Some areas such as Greenland, far-northern <sup>254</sup> we primarily analyze relative change from a baseline. We take  
237 Canada, Siberia, Antarctica, the Gobi and Sahara deserts, and <sup>255</sup> as the baseline the scenario with historical climatology (i.e. T  
238 central Australia are not simulated as they are assumed to re- <sup>256</sup> and P changes of 0). C of 360 ppm, and applied N at 200 kg  
239 main non-arable even under an extreme climate change. Grow- <sup>257</sup>  $\text{ha}^{-1}$ . We show absolute yields in some cases to illustrate geo-  
240 ing seasons are standardized across models with data adapted <sup>258</sup> graphic differences in yields for a single model.

241 from several sources (Sacks et al., 2010, Portmann et al., 2008,<sup>259</sup> 2.2. *Simulation model validation approach*

242 2010). <sup>260</sup> Simulation model validation for GGCMI phase II builds on  
243 The participating modeling groups provide simulations at<sup>261</sup> the validation efforts presented in Müller et al. (2017) for the  
244 any of four initially specified levels of participation, so the num-<sup>262</sup> first phase. In the case presented here however, the models  
245 ber of simulations varies by model, with some sampling only a<sup>263</sup> are not run on the best approximation of management levels

(namely nitrogen application level) by country as with phase I. As the goals of this phase of the project are focused on understanding the sensitivity in *change* in yield to changes in input drivers –and not to simulate historical yields as accurately as possible– no direct comparison to historical yield data can be made. Additionally, even when provided with an appropriate local nitrogen level, models simulated *potential* yields that do not include reductions from pests, weeds, or diseases. Potential yields represent an ideal case that is not realized in many less industrialized areas. Finally, some models are not calibrated as they were in phase I of the project.

We evaluate the models here based on the response to year-to-year temperature and precipitation variability in the historical record. If the models can (somewhat) faithfully represent the historical variability in yields (which, once detrended to account for changing management levels must be driven largely by differences in weather), then the models may provide some utility in understanding the impact on mean climatological shifts in temperature and precipitation. Specifically, we calculate a Pearson correlation coefficient between the detrended time series of simulations and FAO data for the period 1981-2009. Validating the response to CO<sub>2</sub> and Nitrogen applications is more difficult because real world data is not available outside of small greenhouse and field level trials.

### 2.3. Climatological-mean yield emulator design

We construct our emulator at the 30-year climatological mean level. Blanc & Sultan (2015) and Blanc (2017) have shown that a emulator of a global process-based crop simulation model can be successfully developed at the yearly scale.

The decision to first construct a climatological-mean yield emulator is driven by the target application for this analysis tool. Many impact modelers are not focused on the changes in the year-to-year variability in yields, but instead on the broad mean changes over the multi-decadal timescale. Emulation in-

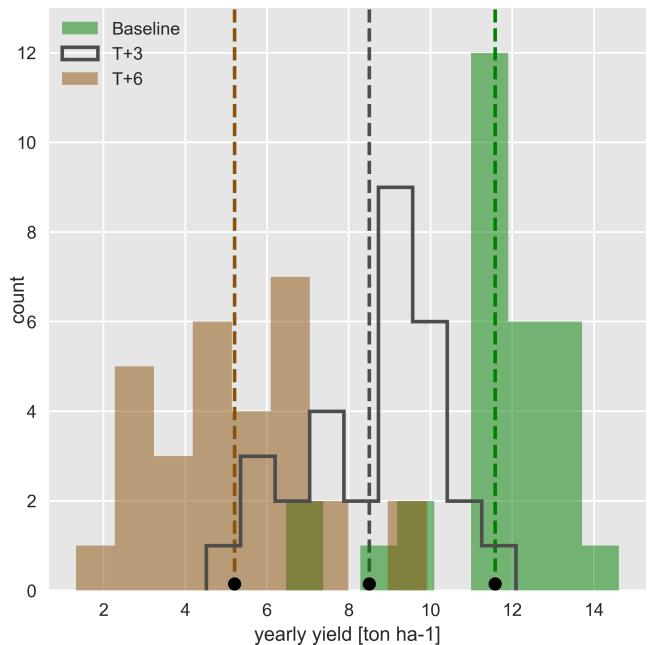


Figure 2: Example showing both climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for a grid cell in northern Iowa (a representative high-yield region) from the pDSSAT model, for the baseline climatology (1981-2010) and for scenarios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

volves 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<sub>2</sub>, 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.

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

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

ies have shown them to be significant effects. For example, Lobell & Field (2007) and Tebaldi & Lobell (2008) showed that in real-world yields, the joint distribution in T and W is needed to explain observed yield variance (C and N are fixed in these data). Other observation-based studies have shown the importance of the interaction between water and nitrogen (e.g. Aulakh & Malhi, 2005), and between nitrogen and carbon dioxide (Osaki et al., 1992, Nakamura et al., 1997). We do not focus on comparing different model specifications in this study, and instead stick to a relatively simple parameterized specification that allows for some, albeit limited, coefficient interpretation.

The limited GGCMI 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

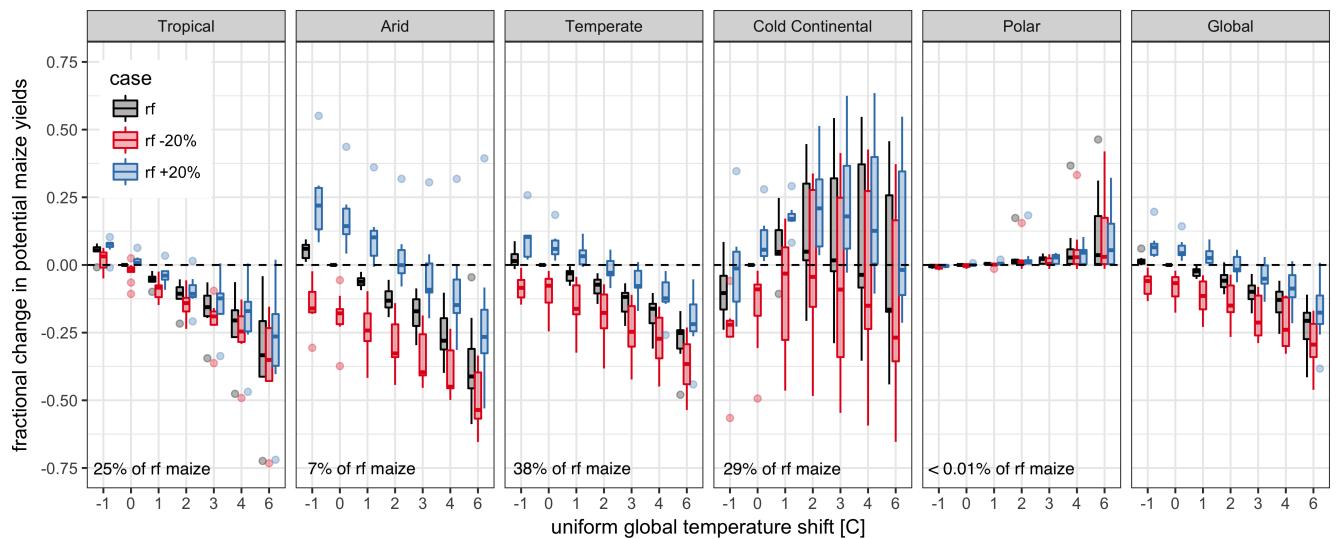


Figure 3: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation levels (rain-fed (rf) -20%, rain-fed (0), and rain-fed +20%), with CO<sub>2</sub> and nitrogen held constant at baseline values (360 ppm and 200 kg ha<sup>-1</sup> yr<sup>-1</sup>). Y-axis is fractional change in the regional average climatological potential yield relative to the baseline. The figure shows all modeled land area; see Figure S6 in the supplemental material for only currently-cultivated land. Panel text gives the percentage of rain-fed maize presently cultivated in each climate zone (Portmann et al., 2010). Box-and-whiskers plots show distribution across models, with median marked. Box edges are first and third quartiles, i.e. box height is the interquartile range (IQR). Whiskers extend to maximum and minimum of simulations but are limited at 1.5-IQR; otherwise the outlier is shown. Models generally agree in most climate regions (other than cold continental), with projected changes larger than inter-model variance. Outliers in the tropics (strong negative impact of temperature increases) are the pDSSAT model; outliers in the high-rainfall case (strong positive impact of precipitation increases) are the JULES model. Inter-model variance increases in the case where precipitation is reduced, suggesting uncertainty in model response to water limitation. The right panel with global changes shows yield responses to an globally uniform temperature shift; note that these results are not directly comparable to simulations of more realistic climate scenarios with the same global mean change.

polynomial are tested for importance. In this procedure higher-order and interaction terms are added successively to the model; we then follow the reduction of the aggregate mean squared error with increasing terms and eliminate those terms that do not contribute significant reductions. See supplemental documents for more details. We select terms by applying the feature selection process to the three models that provided the complete set of 672 rain-fed simulations (pDSSAT, EPIC-TAMU, and LPJmL); the resulting choice of terms is then applied for all emulators.

Feature importance is remarkably consistent across all three models and across all crops (see Figure S4 in the supplemental material). The feature selection process results in a final polynomial in 23 terms, with 11 terms eliminated. We omit the  $N^3$  term, which cannot be fitted because we sample only three nitrogen levels. We eliminate many of the C terms: the cubic, the CT, CTN, and CWN interaction terms, and all higher order interaction terms in C. Finally, we eliminate two 2nd-order interaction terms in T and one in W. Implication of this choice include that nitrogen interactions are complex and important, and that water interaction effects are more nonlinear than those in temperature. The resulting statistical model (Equation 1) is used for all grid cells, models, and crops:

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

To fit the parameters  $K$ , we use a Bayesian Ridge probabilistic

estimator (MacKay, 1991), which reduces volatility in parameter estimates when the sampling is sparse, by weighting parameter estimates towards zero. The Bayesian Ridge method is necessary to maintain a consistent functional form across all models, and locations as the linear least squares fails to provide a stable result in many cases. In the GGCMI Phase II experiment, the most problematic fits are those for models that provided a limited number of cases or for low-yield geographic regions where some modeling groups did not run all scenarios. Because we do not attempt to emulate models that provided less than 50 simulations, the lowest number of simulations emulated across the full parameter space is 130 (for the PEPIC model). We use the implementation of the Bayesian Ridge estimator from the scikit-learn package in Python (Pedregosa et al., 2011).

The resulting parameter matrices for all crop model emulators are available on request, as are the raw simulation data and a Python application to emulate yields. The yield output for a 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.

#### 2.4. Emulator evaluation

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

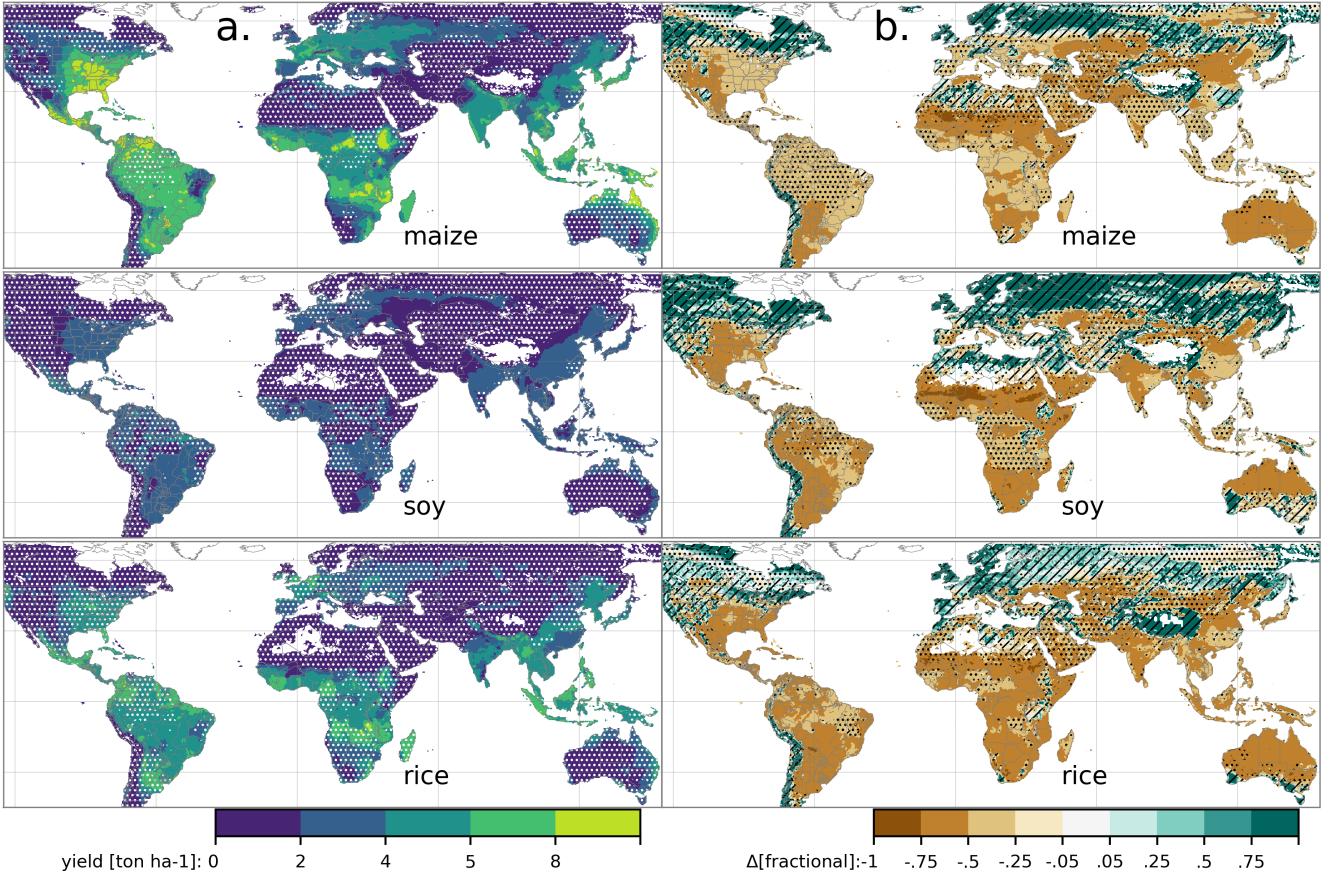


Figure 4: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure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

$$F_{scn.} = \frac{Y_{scn.} - Y_{baseline}}{Y_{baseline}}$$

$$e_{scn.} = \frac{F_{em, scn.} - F_{sim, scn.}}{\sigma_{sim, scn.}}$$

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  $\sigma_{sim}$ , 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 the emulation to an estimate of true uncertainty, which we take as the multi-model spread.

422 **3. Results**

423 *3.1. Simulation results*

424 Crop models in the GGCMI ensemble show a broadly con-  
425 sistent responses to climate and management perturbations in  
426 most regions, with a strong negative impact of increased tem-  
427 perature in all but the coldest regions. We illustrate this result  
428 for rain-fed maize in Figure 3, which shows yields for the pri-  
429 mary Köppen-Geiger climate regions (Rubel & Kottek, 2010).  
430 In warming scenarios, models show decreases in maize yield in  
431 the temperate, tropical, and arid regions that account for nearly  
432 three-quarters of global maize production. These impacts are  
433 robust for even moderate climate perturbations. In the temper-  
434 ate zone, even a 1 degree temperature rise with other variables  
435 held fixed leads to a median yield reduction that outweighs the  
436 variance across models. A 6 degree temperature rise results in  
437 median loss of ~25% of yields with a signal to noise of nearly  
438 three. A notable exception is the cold continental region, where  
439 models disagree strongly, extending even to the sign of impacts.  
440 Model simulations of other crops produce similar responses to  
441 warming, with robust yield losses in warmer locations and high  
442 inter-model variance in the cold continental regions (Figures  
443 S7).

444 The effects of rainfall changes on maize yields are also as ex-  
445 pected and are consistent across models. Increased rainfall mit-  
446 igates the negative effect of higher temperatures, most strongly  
447 in arid regions. Decreased rainfall amplifies yield losses and  
448 also increases inter-model variance more strongly, suggesting  
449 that models have difficulty representing crop response to water  
450 stress. We show only rain-fed maize here; see Figure S5 for the  
451 irrigated case. As expected, irrigated crops are more resilient to  
452 temperature increases in all regions, especially so where water  
453 is limiting.

454 Mapping the distribution of baseline yields and yield changes  
455 shows the geographic dependencies that underlie these results.

456 Figure 4 shows baseline and changes in the T+4 scenario for  
457 rain-fed maize, soy, and rice in the multi-model ensemble mean,  
458 with locations of model agreement marked. Absolute yield po-  
459 tentials are have strong spatial variation, with much of the  
460 Earth’s surface area unsuitable for any given crop. In general,  
461 models agree most on yield response in regions where yield  
462 potentials are currently high and therefore where crops are cur-  
463 rently grown. Models show robust decreases in yields at low  
464 latitudes, and highly uncertain median increases at most high  
465 latitudes. For wheat crops see Figure S11; wheat projections  
466 are both more uncertain and show fewer areas of increased yield  
467 in the inter-model mean.

468 *3.2. Simulation model validation results*

469 Figure 5 shows the time series correlation between the simu-  
470 lation model yield and FAO yield data. The results are mixed,  
471 with many regions for rice and wheat being difficult to model.  
472 No single model is dominant, with each model providing near  
473 best-in-class performance in at least one location-crop combi-  
474 nation. The presence of no vertical dark green color bars clearly  
475 illustrates the power of a multi-model intercomparison project  
476 like the one presented here. The ensemble mean yield is cal-  
477 culated across all ‘high’ nitrogen application level model sim-  
478 ulations and correlated with the FAO data (not the mean of the  
479 correlations). The ensemble mean does not beat the best model  
480 in each case, but shows positive correlation in over 75% of the  
481 cases presented here.

482 Soy is qualitatively the easiest crop to represent (except in  
483 Argentina), which is likely due to the invariance of the re-  
484 sponse to nitrogen application (soy fixes atmospheric nitrogen  
485 very efficiently). Comparison to the FAO data is therefore eas-  
486 ier than the other crops because the nitrogen application levels  
487 do not matter. US maize has the best performance across mod-  
488 els, with nearly every model representing the historical vari-  
489 ability to some extent. Especially good example years for US

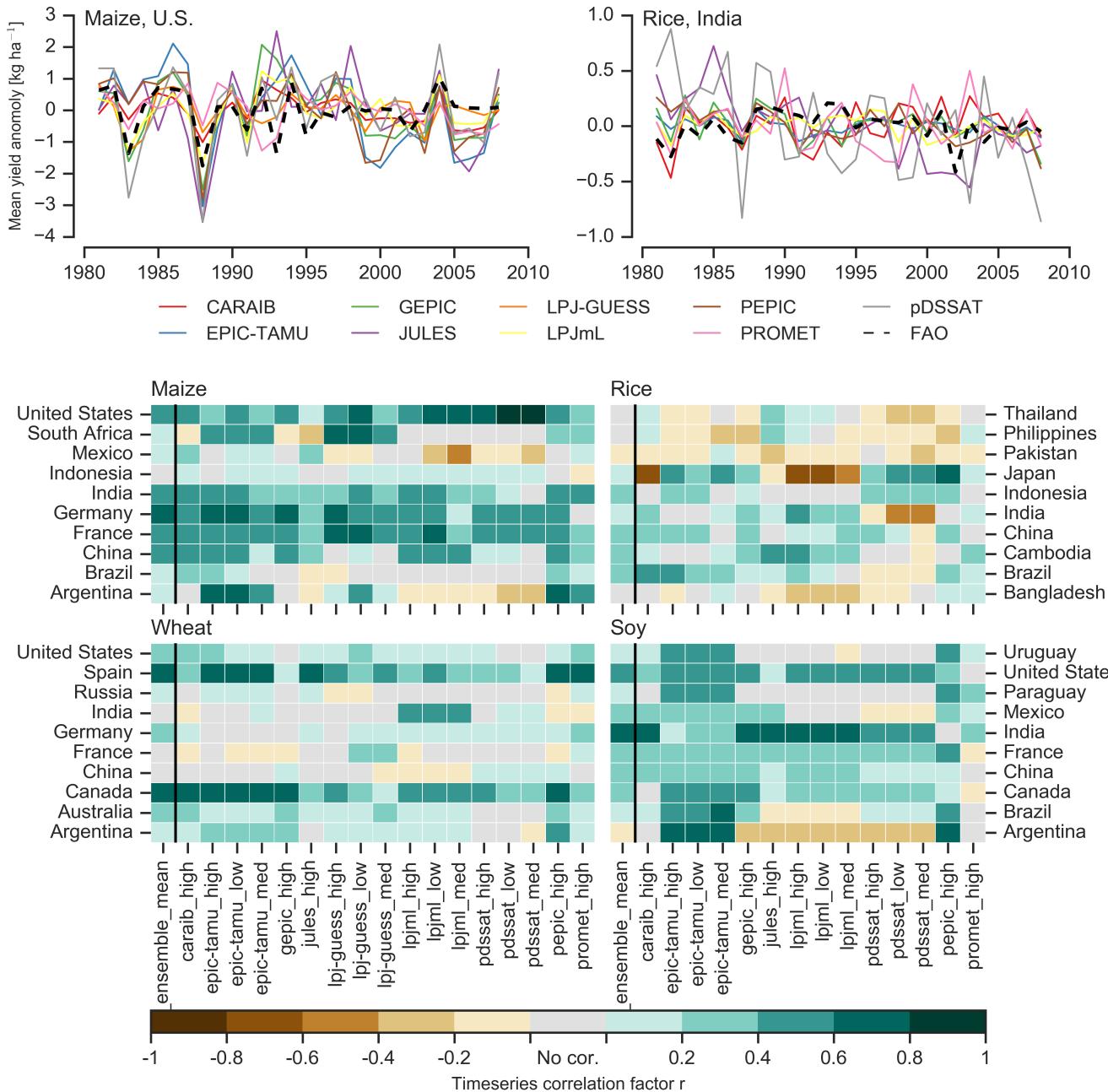


Figure 5: Time series correlation coefficients between simulated crop yield and FAO data at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen application case. The heatmaps illustrate the Pearson  $r$  correlation 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.

maize are 1983, 1988, and 2004 (top left panel), where every historical period of all the cases presented here.

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

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. Partitioning of these con-

tributions is impossible at this stage. Additionally, there is less year-to-year variability in rice yields (partially due to the fraction of irrigated cultivation). Since the Pearson r metric is scale invariant, it will tend to score the rice models more poorly than maize and soy. The pDSSAT model shows very poor performance for rice in India (top right panel).

*One may speculate that the difference in performance between Pakistan (no successful models) and India (many successful models) for rice may lie in the FAO data and not the models themselves. What would be so different about rice production across these two countries that could explain this difference??*

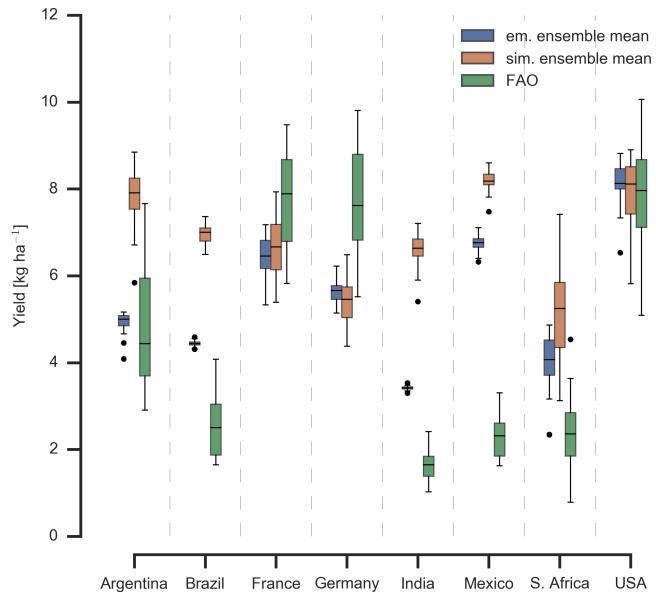


Figure 6: Distribution in historical yields (1981-2009) for maize for eight example high producing countries. FAO, simulation (high nitrogen), and emulation. Emulated values are calculated based on the additive temperature anomaly or percentage precipitation anomaly from the 1981-2009 period in each year. Note: the emulator is designed to provide the mean change in yield under climatological mean shift in temperature (or precipitation). Applying it at the year to year level should be interpreted with caution.

Figure 6 shows the distribution across historical maize yields for some high producing countries. The discrepancy between the simulations and FAO data is most evident in developing nations, where nitrogen application levels are far below the 200 kg ha⁻¹ applied in the simulations shown here (though the distributions are similar in those nations otherwise). The distribution

in historical yields is also calculated with the climatological-mean emulator by passing it the historical (1981-2009) anomalies in growing season precipitation and temperature, CO<sub>2</sub> concentration of 360 ppm, and spatially varying nitrogen application rates (data from: Potter et al., 2010, Mueller et al., 2012). The emulator distribution is shifted towards the FAO distribution in cases where the nitrogen levels are too high in the simulations, but this does not account for

### 3.3. Emulator performance

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 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 geographic 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 scenarios varying only along a single dimension (CO<sub>2</sub>, temperature, precipitation, or nitrogen addition), with other inputs held fixed at baseline levels; in all cases yields evolve smoothly across the space sampled. For reference we show the results of the full emulation fitted across the parameter space. The polynomial fit readily captures the climatological response to perturbations.

Crop yield responses generally follow similar functional forms across models, though with a spread in magnitude. Figure 8 illustrates the inter-model diversity of yield responses to the same perturbations, even for a single crop and location

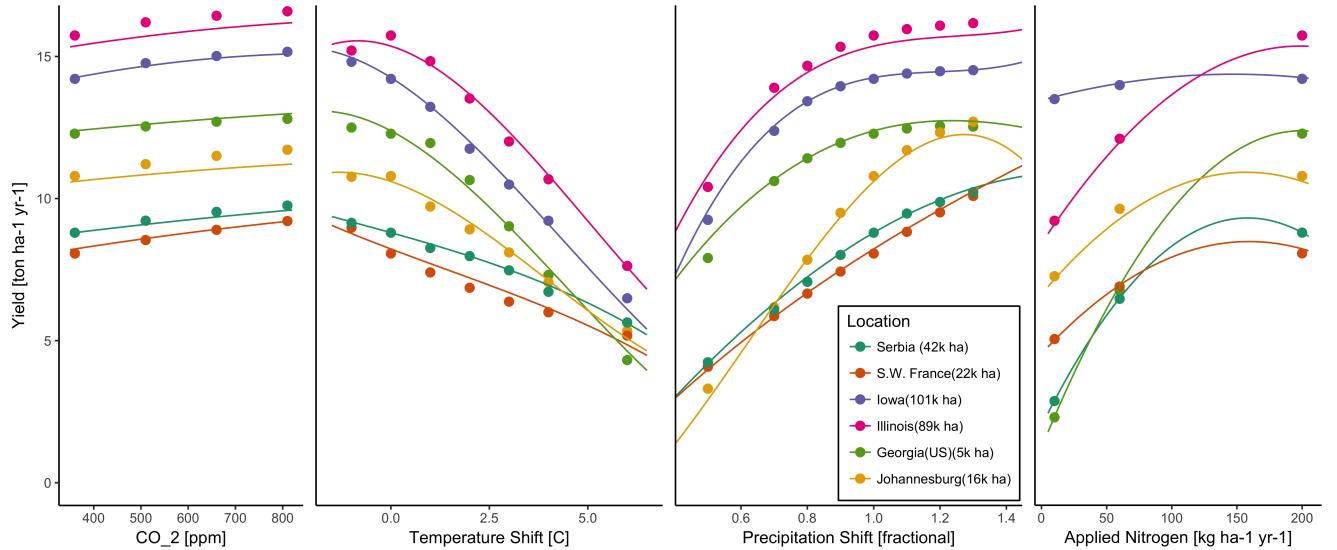


Figure 7: Example illustrating spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows 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).

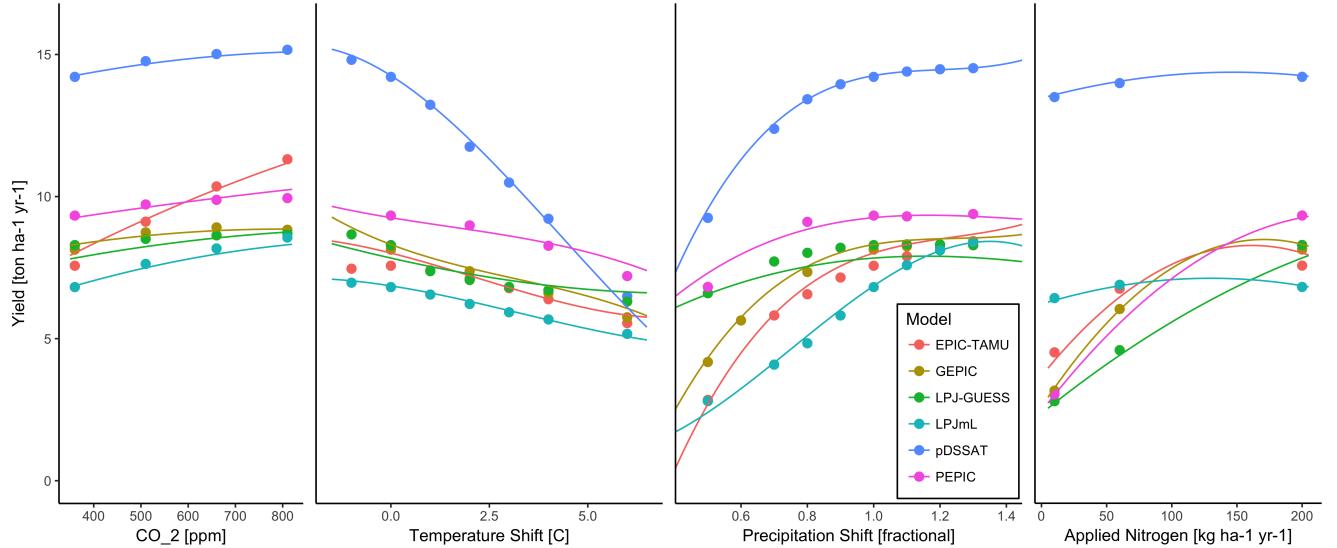


Figure 8: Example illustrating across-model variations in yield response. We show simulations and emulations from six models for rain-fed maize in the same Iowa location shown in Figure 7, with the same plot conventions. Models that did not simulate the nitrogen dimension are omitted for clarity. The inter-model standard deviation is larger for the perturbations in CO<sub>2</sub> and nitrogen than for those in temperature or precipitation illustrating that the sensitivity to weather is better constrained than to management at elevated CO<sub>2</sub>. While most model responses can readily be emulated with a simple polynomial, some response surfaces are more complicated and lead to emulation error.

554 (rain-fed maize in northern Iowa, the same location shown in 559  
 555 the Figure 7). The differences make it important to construct 560  
 561 emulators separately for each individual model, and the fidelity 561  
 562 of emulation can also differ across models. This figure illus- 562  
 563 trates a common phenomenon, that models differ more in re- 563

554 response to perturbations in CO<sub>2</sub> and nitrogen perturbations than  
 555 to those in temperature or precipitation. (Compare also Figures  
 556 3 and S18.) For this location and crop, CO<sub>2</sub> fertilization effects  
 557 can range from ~5–50%, and nitrogen responses from nearly  
 558 flat to a 60% drop in the lowest-application simulation.

564 While the nitrogen dimension is important and uncertain, it  
 565 is also the most problematic to emulate in this work because  
 566 of its limited sampling. The GGCMI protocol specified only  
 567 three nitrogen levels ( $10, 60$  and  $200 \text{ kg N y}^{-1} \text{ ha}^{-1}$ ), so a third-  
 568 order fit would be over-determined but a second-order fit can  
 569 result in potentially unphysical results. Steep and nonlinear de-  
 570 clines in yield with lower nitrogen levels means that some re-  
 571 gressions imply a peak in yield between the  $100$  and  $200 \text{ kg N}$   
 572  $\text{y}^{-1} \text{ ha}^{-1}$  levels. While there may be some reason to believe  
 573 over-application of nitrogen at the wrong time in the growing  
 574 season could lead to reduced yields, these features are almost  
 575 certainly an artifact of under sampling. In addition, the poly-  
 576 nomial fit cannot capture the well-documented saturation effect  
 577 of nitrogen application (e.g. Ingestad, 1977) as accurately as  
 578 would be possible with a non-parametric model.

579 To assess the ability of the polynomial emulation to capture  
 580 the behavior of complex process-based models, we evaluate the  
 581 normalized emulator error. That is, for each grid cell, model,  
 582 and scenario we evaluate the difference between the model yield  
 583 and its emulation, normalized by the inter-model standard de-  
 584 viation in yield projections. This metric implies that emulation  
 585 is generally satisfactory, with several distinct exceptions. Al-  
 586 most all model-crop combination emulators have normalized  
 587 errors less than one over nearly all currently cultivated hectares  
 588 (Figure 9), but some individual model-crop combinations are  
 589 problematic (e.g. PROMET for rice and soy, JULES for soy  
 590 and winter wheat, Figures S14–S15). Normalized errors for soy  
 591 are somewhat higher across all models not because emulator fi-  
 592 delity is worse but because models agree more closely on yield  
 593 changes for soy than for other crops (see Figure S16, lowering  
 594 the denominator. Emulator performance often degrades in geo-  
 595 graphic locations where crops are not currently cultivated. Fig-  
 596 ure 10 shows a CARAIB case as an example, where emulator  
 597 performance is satisfactory over cultivated areas for all crops

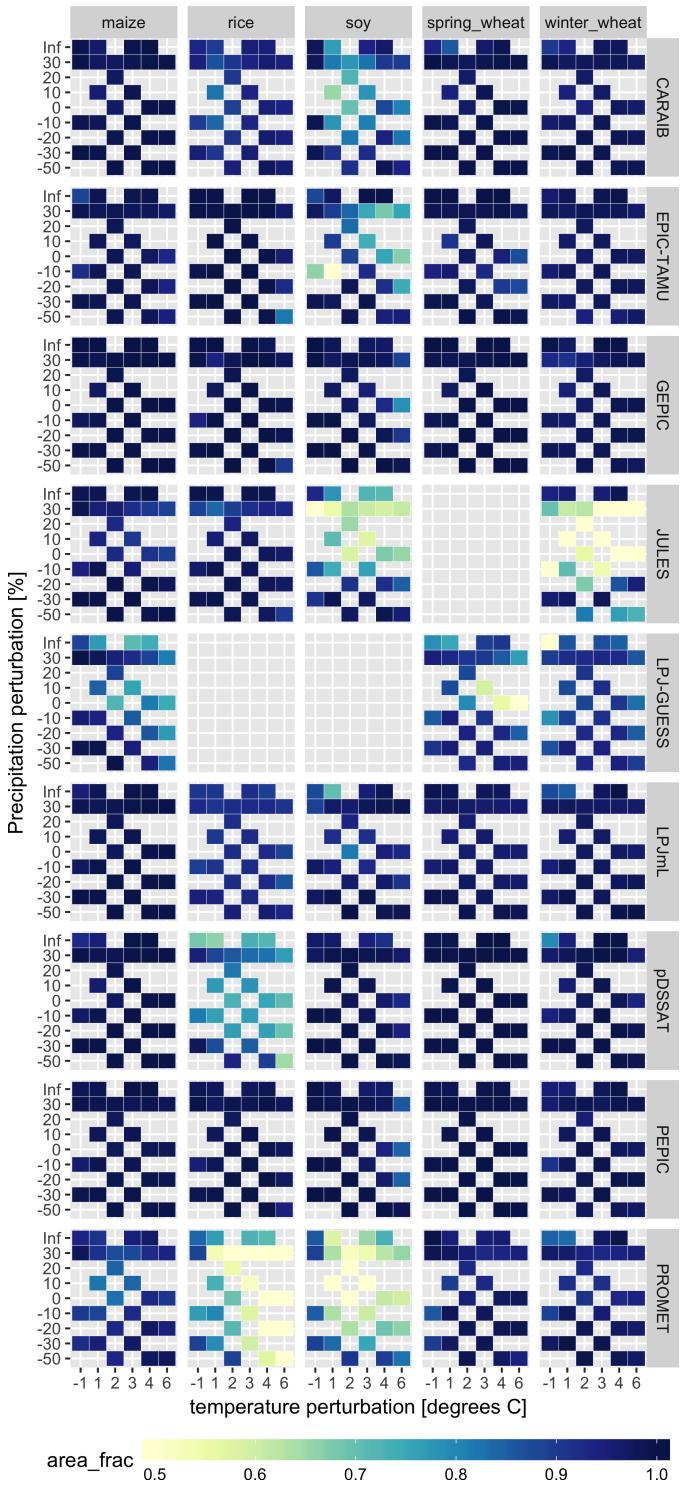


Figure 9: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with  $\text{CO}_2$  and nitrogen held fixed at baseline values. Large columns are crops and large rows models; squares within are T,P scenario pairs. Colors denote the fraction of currently cultivated hectares with  $e < 1$ . Of the 756 scenarios with these  $\text{CO}_2$  and N values, we consider only those for which all 9 models submitted data. JULES did not simulate spring wheat and LPJ-GUESS did not simulate rice and soy. Emulator performance is generally satisfactory, with some exceptions. Emulator failures (significant areas of poor performance) occur for individual crop-model combinations, with performance generally degrading for hotter and wetter scenarios.

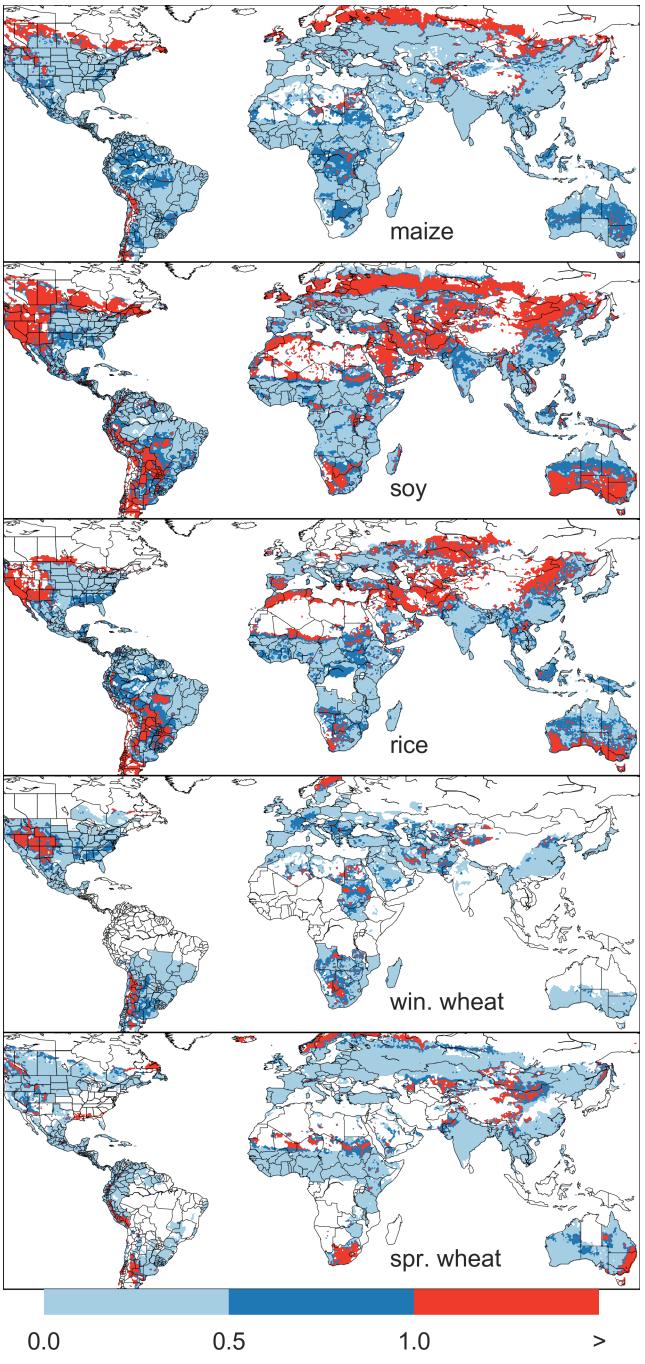


Figure 10: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error  $e$ , where  $e > 1$  means that emulator error exceeds the multi-model standard deviation. White areas are those where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

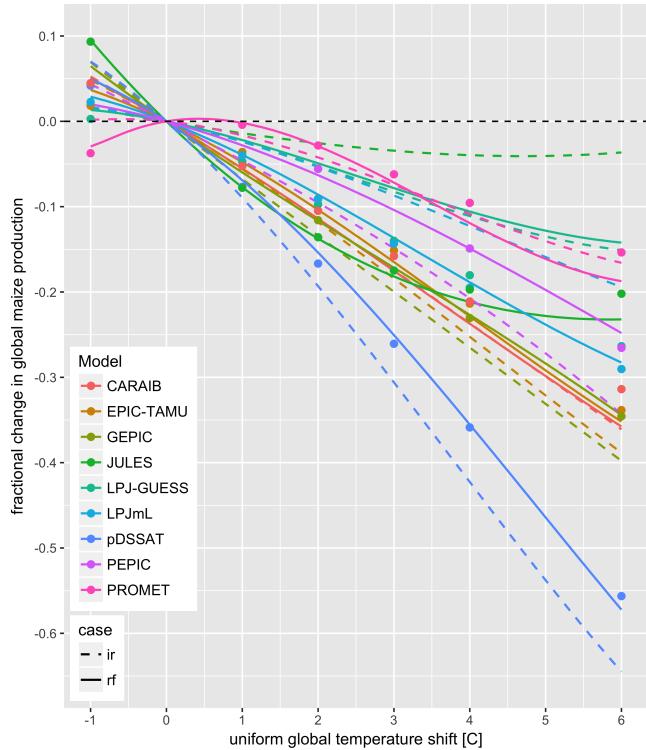
other than soy, but uncultivated regions show some problematic areas.

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<sub>2</sub> fertilization, this effect is readily seen when comparing assessments of emulator performance in simulations at baseline CO<sub>2</sub> (Figure 9) with those at higher CO<sub>2</sub> levels (Figure S13). Widening the inter-model spread leads to an apparent increase in emulator skill.

### 3.4. Emulator applications

Because the emulator or “surrogate model” transforms the discrete simulation sample space into a continuous response surface at any geographic scale, it can be used for a variety of applications. Emulators provide an easy way to compare a ensemble of climate or impacts projections. They also provide a means for generating continuous damage functions. As an example, we show a damage function constructed from 4D emulations for aggregated yield at the global scale, for maize on currently cultivated land, with simulated values shown for comparison. (Figure 11; see Figures S16- S19 in the supplemental material for other crops and dimensions.) The emulated values closely match simulations even at this aggregation level. Note that these functions are presented only as examples and do not represent true global projections, because they are developed from simulation data with a uniform temperature shift while increases in global mean temperature should manifest non-uniformly. The global coverage of the GGCMI simulations allows impacts modelers to apply arbitrary

632 geographically-varying climate projections, as well as arbitrary  
 633 aggregation mask, to develop damage functions for any climate  
 634 scenario and any geopolitical or geographic level.



601 Figure 11: Global emulated damages for maize on currently cultivated lands  
 602 for the GGCMI models emulated, for uniform temperature shifts with other  
 603 inputs held at baseline. (The damage function is created from aggregating up  
 604 emulated values at the grid cell level, not from a regression of global mean  
 605 yields.) Lines are emulations for rain-fed (solid) and irrigated (dashed) crops;  
 606 for comparison, dots are the simulated values for the rain-fed case. For most  
 607 models, irrigated crops show a sharper reduction than do rain-fed because of the  
 608 locations of cultivated areas: irrigated crops tend to be grown in warmer areas  
 609 where impacts are more severe for a given temperature shift. (The exceptions  
 610 are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures  
 611 S16- S19 in the supplemental material.

#### 635 4. Conclusions and discussion

636 The GGCMI Phase II experiment assess sensitivities of  
 637 process-based crop yield models to changing climate and man-  
 638 agement inputs, and was designed to allow not only comparison  
 639 across models but evaluation of complex interactions between  
 640 driving factors ( $\text{CO}_2$ , temperature, precipitation, and applied  
 641 nitrogen) and identification of geographic shifts in high yield  
 642 potential locations. While the richness of the dataset invites  
 643 further analysis, we show only a selection of insights derived

644 from the simulations. Across the major crops, inter-model un-  
 645 certainty is greatest for wheat and least for soy. Across factors  
 646 impacting yields, inter-model-uncertainty is largest for  $\text{CO}_2$  fer-  
 647 tilization and nitrogen response effects. Across geographic re-  
 648 gions, inter-model uncertainty is largest in the high latitudes  
 649 where yields may increase, and model projections are most ro-  
 650 bust in low latitudes where yield impacts are largest.

651 Model performance when compared to historical data is  
 652 mixed, with models performing better for maize and soy than  
 653 for rice and wheat. The value of utilizing multiple models is  
 654 illustrated by the distribution in performance skill across differ-  
 655 ent countries and crops. An end-user of the simulation outputs  
 656 or emulator tool may pick and choose models based on histori-  
 657 cal skill to provide the most faithful temperature and precipita-  
 658 tion response depending on their application. The nitrogen and  
 659  $\text{CO}_2$  responses were not validated in this work.

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

674 We show that emulation of the output of these complex re-  
 675 sponds is possible even with a relatively simple reduced-form  
 676 statistical model and a limited library of simulations. Emula-  
 677 tion therefore offers the opportunity of producing rapid assess-

678 ments of agricultural impacts for arbitrary climate scenarios in<sup>712</sup>  
679 a computationally non-intensive way. The resulting tool should<sup>713</sup>  
680 aid in impacts assessment, economic studies, and uncertainty<sup>714</sup>  
681 analyses. Emulator parameter values also provide a useful way<sup>715</sup>  
682 to compare sensitivities across models to different climate and<sup>716</sup>  
683 management inputs, and the terms in the polynomial fits offer<sup>717</sup>  
684 the possibility of physical interpretation of these dependencies<sup>718</sup>  
685 to some degree.

719 dress the crop yield impacts of potential changes in climate  
variability. While some information could be extracted from  
consideration of year-over-year variability, more detailed sim-  
ulations and analysis are likely necessary to diagnose the im-  
pact of changes in variance and sub-growing-season tempo-  
ral effects. Additionally, the emulator is intended to provide  
the change in yield from a historical mean baseline value and  
should be used in conjunction with historical data (or data prod-  
ucts) or a historical mean emulator (not presented here).

720 We provide this simulation output dataset for further analysis  
686 by the community as we have only scratched the surface with<sup>721</sup>  
687 this work. Each simulation run includes year to year variabil-<sup>722</sup>  
688 ity in yields under different climate and management regimes.<sup>723</sup>  
689 Some of the precipitation and temperature space has been lost<sup>724</sup>  
690 due to the aggregation in the time dimension for the emula-<sup>725</sup>  
691 tor presented here (i.e. the + 6 C simulation in the hottest year<sup>726</sup>  
692 of the historical period compared to the coldest historical year,<sup>727</sup>  
693 or precipitation perturbations in the driest historical year etc.<sup>728</sup>).  
694 Development of a year-to-year emulator or an emulator at dif-<sup>729</sup>  
695 ferent spatial scales may provide useful for some IAM appli-<sup>730</sup>  
696 cations. More exhaustive analysis of different statistical model  
697 specification for emulation will likely provide additional pre-<sup>731</sup>  
698 dictive skill over the specification provided here. The poten-<sup>732</sup>  
699 tially richest area for further analysis is the interactions be-<sup>733</sup>  
700 tween input variable especially the Nitrogen and CO<sub>2</sub> interac-<sup>734</sup>  
701 tions with weather and with each other. More robust quantifica-<sup>735</sup>  
702 tion of the sensitivity to the input drivers (and there differences<sup>736</sup>  
703 across models), as well as quantification in differences in un-<sup>737</sup>  
704 certainty across input drivers. Adaptation via growing season<sup>738</sup>  
705 changes were also simulated and are available in the database,<sup>739</sup>  
706 though this dimension was not presented or analyzed here.<sup>740</sup>

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

## 5. Acknowledgments

We thank Michael Stein and Kevin Schwarzwald, who pro-  
vided helpful suggestions that contributed to this work. This re-  
search was performed as part of the Center for Robust Decision-  
Making on Climate and Energy Policy (RDCEP) at the Univer-  
sity of Chicago, and was supported through a variety of sources.  
RDCEP is funded by NSF grant #SES-1463644 through the  
Decision Making Under Uncertainty program. J.F. was sup-  
ported by the NSF NRT program, grant #DGE-1735359. C.M.  
was supported by the MACMIT project (01LN1317A) funded  
through the German Federal Ministry of Education and Re-  
search (BMBF). C.F. was supported by the European Research  
Council Synergy grant #ERC-2013-SynG-610028 Imbalance-  
P. P.F. and K.W. were supported by the Newton Fund through

745 the Met Office Climate Science for Service Partnership Brazil<sup>784</sup>  
 746 (CSSP Brazil). A.S. was supported by the Office of Science<sup>785</sup>  
 747 of the U.S. Department of Energy as part of the Multi-sector  
 748 Dynamics Research Program Area. Computing resources were<sup>788</sup>  
 749 provided by the University of Chicago Research Computing<sup>789</sup>  
 750 Center (RCC). S.O. acknowledges support from the Swedish<sup>790</sup>  
 751 strong research areas BECC and MERGE together with sup-<sup>791</sup>  
 752 port from LUCCI (Lund University Centre for studies of Car-<sup>793</sup>  
 753 bon Cycle and Climate Interactions).<sup>794</sup>

## 754 6. References

- 755 Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013).  
 756 Implication of crop model calibration strategies for assessing regional im-  
 757 pacts of climate change in europe. *Agric. For. Meteorol.*, *170*, 32 – 46.  
 758 Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D.,  
 759 A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Al-  
 760 derman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Bier-  
 761 nath, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising tempera-  
 762 tures reduce global wheat production. *Nature Climate Change*, *5*, 143–147.  
 763 doi:10.1038/nclimate2470.  
 764 Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A.,  
 765 J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N.,  
 766 Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath,  
 767 C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simulat-  
 768 ing wheat yields under climate change. *Nature Climate Change*, *3*, 827832.  
 769 doi:10.1038/nclimate1916.  
 770 Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other  
 771 Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Ef-  
 772 ficiency, Carbon Sequestration, and Environmental Pollution. *Advances in*  
 773 *Agronomy*, *86*, 341 – 409.  
 774 Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov,  
 775 N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat  
 776 production potentials and management flexibility under the representative  
 777 concentration pathways. *Global and Planetary Change*, *122*, 107 – 121.  
 778 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields  
 779 from global gridded crop models. *Agricultural and Forest Meteorology*, *236*,  
 780 145 – 161.  
 781 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded  
 782 crop models using statistical estimates. *Agricultural and Forest Meteorol-*  
 783 *ogy*, *214-215*, 134 – 147.  
 784 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaeble, S.  
 785 (2018). Implementing the Nitrogen cycle into the dynamic global vegeta-  
 786 tion, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific*  
 787 *Model Development*, *11*, 2789–2812.  
 788 Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L.,  
 789 & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections  
 790 Based on Precomputed GCM Runs. *Journal of Climate*, *27*, 1829–1844.  
 791 Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.  
 792 (2014). A meta-analysis of crop yield under climate change and adaptation.  
 793 *Nature Climate Change*, *4*, 287 – 291.  
 794 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process  
 795 emulation of dynamic computer codes. *Biometrika*, *96*, 663–676.  
 796 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In  
 797 C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth,*  
 798 *Duke University, Durham, North Carolina* (pp. 115–118). Durham, North  
 799 Carolina.  
 800 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for  
 801 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).  
 802 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberrous,  
 803 M., & François, L. (2011). Responses of European forest ecosystems to  
 804 21st century climate: assessing changes in interannual variability and fire  
 805 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).  
 806 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,  
 807 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating  
 808 impact models and sectors (pSIMS). *Environmental Modelling and Soft-*  
 809 *ware*, *62*, 509–516.  
 810 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,  
 811 Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurrealde, R. C.,  
 812 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.  
 813 (2015). The Global Gridded Crop Model Intercomparison: data and mod-  
 814 eling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, *8*,  
 815 261–277.  
 816 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J.,  
 817 & Taylor, K. E. (2016). Overview of the coupled model intercomparison  
 818 project phase 6 (cmip6) experimental design and organization. *Geoscientific*  
 819 *Model Development*, *9*, 1937–1958.  
 820 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-  
 821 mate change impacts on durum wheat in the mediterranean region. *Natural*  
 822 *Hazards and Earth System Sciences*, *11*, 1293–1302.  
 823 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulz, R., & Yang, H. (2012). Re-  
 824 gionalization of a large-scale crop growth model for sub-Saharan Africa:  
 825 Model setup, evaluation, and estimation of maize yields. *Agriculture,*  
 826 *Ecosystems & Environment*, *151*, 21 – 33.

- 827 Food and Agriculture Organization of the United Nations (2018). FAOSTAT<sup>870</sup>  
 828 database. URL: <http://www.fao.org/faostat/en/home>. <sup>871</sup>
- 829 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T.,<sup>872</sup>  
 830 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P.,<sup>873</sup>  
 831 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F.,<sup>874</sup>  
 832 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,<sup>875</sup>  
 833 Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J.,<sup>876</sup>  
 834 Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nen,<sup>877</sup>  
 835 del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna,<sup>878</sup>  
 836 M., Semenov, M. A., Slawinski, C., Stratonovitch, P., Supit, I., Waha, K.,<sup>879</sup>  
 837 Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-models<sup>880</sup>  
 838 wheat yield impact response surfaces showing sensitivity to temperature and<sup>881</sup>  
 839 precipitation change. *Agricultural Systems*, 159, 209–224. <sup>882</sup>
- 840 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014).<sup>883</sup>  
 841 Evaluating the utility of dynamical downscaling in agricultural impacts pro-<sup>884</sup>  
 842 jections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781.<sup>885</sup>
- 843 Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitiv-<sup>886</sup>  
 844 ity of Agricultural Model Performance to Different Climate Inputs. *Journal*<sup>887</sup>  
 845 of Applied Meteorology and Climatology
- 846 Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported<sup>889</sup>  
 847 Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous<sup>890</sup>  
 848 Crop Growth and Yield: Application for Wheat in Central Europe. *Remote*<sup>891</sup>  
 849 *Sensing*, 7, 3934–3965. <sup>892</sup>
- 850 He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H.,<sup>893</sup>  
 851 He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water<sup>894</sup>  
 852 contents and nitrogen leaching to precipitation, management practices and<sup>895</sup>  
 853 soil hydraulic properties in semi-arid and humid regions of Canada using the<sup>896</sup>  
 854 DSSAT model. *Nutrient Cycling in Agroecosystems*, 106, 201–215. <sup>897</sup>
- 855 Heady, E. O. (1957). An Econometric Investigation of the Technology of Agri-<sup>898</sup>  
 856 cultural Production Functions. *Econometrica*, 25, 249–268. <sup>899</sup>
- 857 Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa<sup>900</sup>  
 858 State University Press. <sup>901</sup>
- 859 Holzkämper, A., Calanca, P., & Fuhrer, J. (2012). Statistical crop models:<sup>902</sup>  
 860 Predicting the effects of temperature and precipitation changes. *Climate*<sup>903</sup>  
 861 *Research*, 51, 11–21. <sup>904</sup>
- 862 Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I.,<sup>905</sup>  
 863 McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore,<sup>906</sup>  
 864 A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake,<sup>907</sup>  
 865 A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgliesh,<sup>908</sup>  
 866 N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp,<sup>909</sup>  
 867 J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson,<sup>910</sup>  
 868 M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T.,<sup>911</sup>  
 869 Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorf, J.,<sup>912</sup>  
 870 Wedgwood, S., & Keating, B. A. (2014). APSIM Evolution towards a new  
 871 generation of agricultural systems simulation. *Environmental Modelling and  
 872 Software*, 62, 327 – 350.
- 873 Howden, S., & Crimp, S. (2005). Assessing dangerous climate change impacts  
 874 on australia's wheat industry. *Modelling and Simulation Society of Australia  
 875 and New Zealand*, (pp. 505–511).
- 876 Izumi, T., Nishimori, M., & Yokozawa, M. (2010). Diagnostics of climate  
 877 model biases in summer temperature and warm-season insolation for the  
 878 simulation of regional paddy rice yield in japan. *Journal of Applied Meteor-  
 879 ology and Climatology*, 49, 574–591.
- 880 Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Ni-  
 881 tragenous Fertilizers. *Ambio*, 6, 146–151.
- 882 Izaurralde, R., Williams, J., McGill, W., Rosenberg, N., & Quiroga Jakas, M.  
 883 (2006). Simulating soil C dynamics with EPIC: Model description and test-  
 884 ing against long-term data. *Ecological Modelling*, 192, 362–384.
- 885 Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the  
 886 CROPGRO-soybean model to predict regional yield and production. *Agri-  
 887 culture, Ecosystems & Environment*, 93, 73 – 85.
- 888 Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L.,  
 889 Wilkens, P., Singh, U., Gijsman, A., & Ritchie, J. (2003). The DSSAT  
 890 cropping system model. *European Journal of Agronomy*, 18, 235 – 265.
- 891 Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I.,  
 892 Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A.,  
 893 Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017).  
 894 Toward a new generation of agricultural system data, models, and knowl-  
 895 edge products: State of agricultural systems science. *Agricultural Systems*,  
 896 155, 269 – 288.
- 897 Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth,  
 898 D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Ver-  
 899 burg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K.,  
 900 Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003).  
 901 An overview of APSIM, a model designed for farming systems simulation.  
 902 *European Journal of Agronomy*, 18, 267 – 288.
- 903 Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., &  
 904 Smith, B. (2013). Implications of accounting for land use in simulations of  
 905 ecosystem carbon cycling in Africa. *Earth System Dynamics*, 4, 385–407.
- 906 Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling  
 907 wheat yield and crop water productivity with high resolution on a global  
 908 scale. *Agricultural Systems*, 94, 478 – 493.
- 909 Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a).  
 910 Global investigation of impacts of PET methods on simulating crop-water  
 911 relations for maize. *Agricultural and Forest Meteorology*, 221, 164 – 175.
- 912 Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C.,

- 913 & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs<sub>956</sub>  
 914 with yields from major crop cultivations. *Science of The Total Environment*,<sub>957</sub>  
 915 572, 526 – 537. <sub>958</sub>
- 916 Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict<sub>959</sub>  
 917 crop yield responses to climate change. *Agricultural and Forest Meteorology*,<sub>960</sub>  
 918 150, 1443 – 1452. <sub>961</sub>
- 919 Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relation-<sub>962</sub>  
 920 ships and the impacts of recent warming. *Environmental Research Letters*,<sub>963</sub>  
 921 2, 014002. <sub>964</sub>
- 922 MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447.<sub>965</sub>
- 923 Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P.,<sub>966</sub>  
 924 Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi,<sub>967</sub>  
 925 P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano,<sub>968</sub>  
 926 D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of<sub>969</sub>  
 927 large simulated yield datasets for studying climate effects. (p. 1100).<sub>970</sub>  
 928 doi:10.13140/RG.2.1.5173.8328. <sub>971</sub>
- 929 Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrolog-<sub>972</sub>  
 930 ical modelling to study the impact of climate change on the water flows of<sub>973</sub>  
 931 mountain watersheds. *Journal of Hydrology*, 376, 362 – 377. <sub>974</sub>
- 932 Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B.,<sub>975</sub>  
 933 & Calzadilla, A. (2009). Global biomass production potentials exceed ex-<sub>976</sub>  
 934 pected future demand without the need for cropland expansion. *Nature Com-<sub>977</sub>  
 935 munications*, 6. <sub>978</sub>
- 936 McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srinivas, C.,<sub>979</sub>  
 937 Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S. <sub>980</sub>
- 938 (2015). Integrated assessments of the impacts of climate change on agricul-<sub>981</sub>  
 939 ture: An overview of AgMIP regional research in South Asia. *Chapter in:*<sub>982</sub>  
 940 *Handbook of Climate Change and Agroecosystems*, (pp. 201–218). <sub>983</sub>
- 941 Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical<sub>984</sub>  
 942 weather responsiveness of crop yields: US evidence and implications for<sub>985</sub>  
 943 the agricultural impacts of climate change. *Environmental Research Letters*,<sub>986</sub>  
 944 12. <sub>987</sub>
- 945 Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate<sub>988</sub>  
 946 change impacts on agriculture implies higher social cost of carbon. *Nature*,<sub>989</sub>  
 947 *Communications*, 8. <sub>990</sub>
- 948 Mueller, N., Gerber, J., Johnston, M., Ray, D., Ramankutty, N., & Foley, J. <sub>991</sub>
- 949 (2012). Closing yield gaps through nutrient and water management. *Nature*,<sub>992</sub>  
 950 490, 254–7. <sub>993</sub>
- 951 Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais,<sub>994</sub>  
 952 P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde,<sub>995</sub>  
 953 R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T.,<sub>996</sub>  
 954 A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G.,<sub>997</sub>  
 955 Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. <sub>998</sub>
- 956 (2017). Global gridded crop model evaluation: benchmarking, skills, de-  
 957 ficiencies and implications. *Geoscientific Model Development*, 10, 1403–  
 958 1422.
- 959 Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T.  
 960 (1997). Effect of CO<sub>2</sub> enrichment on carbon and nitrogen interaction in  
 961 wheat and soybean. *Soil Science and Plant Nutrition*, 43, 789–798.
- 962 O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial.  
 963 *Reliability Engineering & System Safety*, 91, 1290 – 1300.
- 964 Olin, S., Schurgers, G., Lindeskog, M., Wårldin, D., Smith, B., Bodin, P.,  
 965 Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue  
 966 C:N to changes in atmospheric CO<sub>2</sub> and N management in the main wheat  
 967 regions of western europe. *Biogeosciences*, 12, 2489–2515. doi:10.5194/bg-  
 968 12-2489-2015.
- 969 Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in  
 970 field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.
- 971 Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., &  
 972 Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK  
 973 Land Environment Simulator. *Geoscientific Model Development*, 8, 1139–  
 974 1155.
- 975 Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop  
 976 yields and their variability at different levels of global warming. *Earth Sys-  
 977 tem Dynamics*, 9, 479–496.
- 978 Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff,  
 979 S., & Gerten, D. (2015). Emulating global climate change impacts on crop  
 980 yields. *Statistical Modelling*, 15, 499–525.
- 981 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,  
 982 Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas-  
 983 soss, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011).  
 984 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning  
 985 Research*, 12, 2825–2830.
- 986 Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T.,  
 987 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P.,  
 988 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.,  
 989 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,  
 990 Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet, J.,  
 991 Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C.,  
 992 Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M.,  
 993 Semenov, M., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang, E.,  
 994 Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation  
 995 effects on wheat yield across a European transect: a crop model ensemble  
 996 analysis using impact response surfaces. *Climate Research*, 65, 87–105.
- 997 Porter et al. (IPCC) (2014). Food security and food production systems. Cli-  
 998 mate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global

- 999 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.
- 1000 Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of monthly growing areas of 26 irrigated crops.
- 1001 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.
- 1002 Potter, P., Ramankutty, N., Bennett, E., & Donner, S. (2010). Characterizing the spatial patterns of global fertilizer application and manure production. *Earth Interactions - EARTH INTERACT*, 14. doi:10.1175/2010EI288.1.
- 1003 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.
- 1004 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.
- 1005 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.
- 1006 Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, 48.
- 1007 Roberts, M., Braun, N., Sinclair, T., 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.
- 1008 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.
- 1009 Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P., Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182.
- 1010 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.
- 1011 Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason-D'Croz, D., Müller, 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.
- 1012 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.
- 1013 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.
- 1014 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.
- 1015 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.
- 1016 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.
- 1017 Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Cais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frier, K. (2017). Consistent negative response of US crops to high temperatures in observations and crop models. *Nature Communications*, 8, 13931.
- 1018 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.
- 1019 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.
- 1020 Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implement-

1085 tation and evaluation of nonparametric regression procedures for sensitivity  
1086 analysis of computationally demanding models. *Reliability Engineering &*  
1087 *System Safety*, 94, 1735 – 1763.

1088 Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5  
1089 and the experiment design. *Bulletin of the American Meteorological Society*,  
1090 93, 485–498.

1091 Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of cli-  
1092 mate change impacts on global crop yields. *Geophysical Research Letters*,  
1093 35.

1094 Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., &  
1095 Martin, J. F. (2014). Modeling sugarcane yield with a process-based model  
1096 from site to continental scale: Uncertainties arising from model structure  
1097 and parameter values. *Geoscientific Model Development*, 7, 1225–1245.

1098 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe,  
1099 J. (2014). The Inter-Sectoral Impact Model Intercomparison Project  
1100 (ISI-MIP): Project framework. *Proceedings of the National Academy of  
1101 Sciences*, 111, 3228–3232.

1102 White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Method-  
1103 ologies for simulating impacts of climate change on crop production. *Field  
1104 Crops Research*, 124, 357 – 368.

1105 Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T.,  
1106 Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance  
1107 against site observations of irrigated maize from Mead, Nebraska. *Geosci-  
1108 entific Model Development*, 10, 1291–1320.

1109 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield vari-  
1110 ability in JULES-crop and implications for forcing with seasonal weather  
1111 forecasts. *Geoscientific Model Development*, 8, 3987–3997.

1112 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Land-  
1113 bouwkundige Onderzoeken : 64.6.*.

1114 Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato  
1115 yields on changes in climate and co2. *Agricultural and Forest Meteorology*,  
1116 112, 217 – 231.

1117 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao,  
1118 Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A.,  
1119 Li, T., Lin, E., Liu, Q., Martre, P., Mller, C., Peng, S., Peuelas, J., Ruane,  
1120 A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng,  
1121 S. (2017). Temperature increase reduces global yields of major crops in four  
1122 independent estimates. *Proc. Natl. Acad. Sci.*, 114, 9326–9331.