

# 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. Process-based models differ on many details affecting yields and considerable uncertainty remains in future yield projections. Phase II of the Global Gridded Crop Model Intercomparison (GGCMI), an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP), consists of a large simulation set with perturbations in atmospheric CO<sub>2</sub> concentrations, temperature, precipitation, and applied nitrogen inputs and constitutes a data-rich basis of projected yield changes across twelve models and five crops (maize, soy, rice, spring wheat, and winter wheat) using global gridded simulations. In this paper we present the simulation output database from Phase II of the GGCMI effort, a targeted experiment aimed at understanding the sensitivity to and interaction between multiple climate variables (as well as management) on yields, and illustrate some initial summary results from the model intercomparison project. We also present the construction of a simple “emulator or statistical representation of the simulated 30-year mean climatological output in each location for each crop and model. The emulator captures the response of the process-based models in a lightweight, computationally tractable form that facilitates model comparison as well as potential applications in subsequent modeling efforts such as integrated assessment.

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

## 1. Introduction

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

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

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

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

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

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

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

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

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

90 In the past decade, many studies have developed emulators of  
 91 crop yields from process-based models. Early studies propos-  
 92 ing or describing potential emulators include Howden & Crimp  
 93 (2005), Räisänen & Ruokolainen (2006) and Lobell & Burke  
 94 (2010). In an early application, Ferrise et al. (2011) used a Arti-  
 95 ficial Neural Net trained on simulation outputs to predict wheat<sub>101</sub>  
 96 yields in the Mediterranean. Studies developing single-model<sub>102</sub>  
 97 emulators include Holzkämper et al. (2012) for the CropSyst<sub>103</sub>  
 98 model, Ruane et al. (2013) for the CERES wheat model, Oye-<sub>104</sub>  
 99 bamiji et al. (2015) for the LPJmL model (for multiple crops,<sub>105</sub>  
 100 using multiple scenarios as a training set). In recent years, emu-<sub>106</sub>

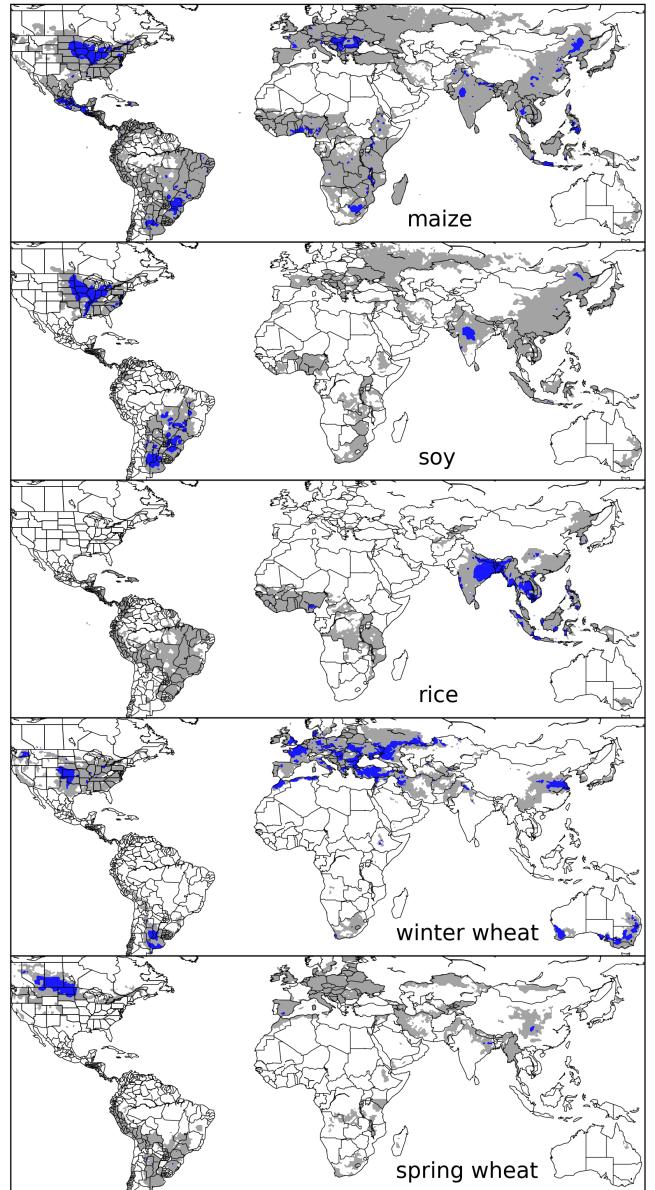


Figure 1: Presently cultivated area for rain-fed crops in the real world. Blue contour area indicates grid cells with more than 20,000 hectares (approx. 10% of the land in a grid cell near the equator for example) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated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

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

107 (2015) and Blanc (2017) used local weather variables (and CO<sub>2</sub><sup>135</sup> climatological mean emulators. however, both papers investi-  
 108 values) and yields but emulate across soil types using historical<sup>136</sup> gate just a few individual locations, Frontzek many models only  
 109 simulations and a future climate scenario (RCP8.5 over mul-<sup>137</sup> wheat, Snyder multiple crops but only one model. In this paper  
 110 tiple climate models); Ostberg et al. (2018) used global mean<sup>138</sup> we describe a new comprehensive dataset designed to expand  
 111 temperature change (and CO<sub>2</sub>) as regressors but pattern-scale<sup>139</sup> this approach still further. GGCMI Phase II experiments pro-  
 112 to emulate local yields using multiple climate scenarios; Mis-<sup>140</sup> vide global coverage, add nitrogen dimension (over 700 simu-  
 113 try et al. (2017) used local weather and yields and a historical<sup>141</sup> lations), full suite of models. [all the acronym stuff]  
 114 simulation and compare with data.<sup>142</sup> might delete or move stuff from line 136, just say what

115 limitations of all existing studies, reason for param-<sup>143</sup> GGCMI Phase II is  
 116 eter sweep - key paragraph. don't be negative as on line 131,<sup>144</sup> last para might be good as is, other than that this might be  
 117 don't end on negative. blame datasets, not authors of papers<sup>145</sup> a dangerous ending. better say that it's tractable to emulation  
 118 A systematic parameter sweep offers advantages over anal-<sup>146</sup> and resulting emulator can provide interpretable parameters and  
 119 yses on small number of realistic scenario in which climate<sup>147</sup> insight

120 varies over time.<sup>148</sup> present a simple climatological emulator as a potential tool  
 121 It allows highlighting the distinction between year-over-year<sup>149</sup> for impacts assessments.

122 and climatological changes, which can be different.<sup>150</sup> Line 276 - move caveats to later in this section- make sepa-  
 123 - removes correlation of the key variables which makes them<sup>151</sup> rate caveat para to split it up. consider shortening.

124 difficult to disentangle in realistic scenarios<sup>152</sup> Line 300 delete which are of most interest to impact modelers  
 125 - provides fully stationary simulations<sup>153</sup> what I would do is say that you make a climatological mean  
 126 - large suite of simulations allows testing ability of emula-<sup>154</sup> emulator, go immediately to saying "we test the necessity for  
 127 tor to reproduce yields by using only some in training set and<sup>155</sup> this approach by using the GGCMI Phase II dataset to evalua-  
 128 testing ability to reproduce those exclude.<sup>156</sup> tate whether year-over-year responses are quantitatively distinct  
 129 (maybe paragraph break here?)<sup>157</sup> from climatological mean responses" -*i*, these can be different  
 130 Say that trend is increasing to use these - cite Frontzek and<sup>158</sup> for many reasons -*i*, introduce Figure 3, say yes they are differ-  
 131 Snyder, respectively and Makowski and Pirttioja as "earlier ef-<sup>159</sup> ent, so we isolate the climatological signal of reponse to long-  
 132 forts incude... Both Frontzek and Snyder do tempreature and<sup>160</sup> term perturbations by emulating on the mean yield for each sce-  
 133 water and Snyder adds Co2, 50 and 100? of simulations per<sup>161</sup> nario in the parameter sweep.

134 model. Both take advantage of size of database to construct<sup>162</sup> separate paragraph - note that we don't capture distributional

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.

shift and if you wanted to do this use methods from A, B, or<sup>196</sup>  
C. Leeds and Poppick are for temporal dependence and Matz<sup>197</sup>  
is for marginal distribution in temperature. Can add Chang for<sup>198</sup>  
precipitation.<sup>199</sup>

Unclear where you cop to not having done a formal uncer-<sup>200</sup>  
tainty assessment - in Methods? In conclusion as a suggestion  
for future work?<sup>201</sup>

Conclusions - line 750-751 should be removed because you  
already did that.<sup>202</sup>

GGCMI Phase II is designed to allow addressing goals such<sup>204</sup>  
as understanding where highest-yield regions may shift un-<sup>205</sup>  
der climate change; exploring future adaptive management<sup>206</sup>  
strategies; understanding how interacting parameters affect<sup>207</sup>  
crop yield; quantifying uncertainties across models and major<sup>208</sup>  
drivers; and testing strategies for producing lightweight emu-<sup>209</sup>  
lators of process-based models. In this paper, we describe the  
GGCMI Phase II experiments, summarize output and present<sup>210</sup>  
initial results, demonstrate that it is tractable to emulation, and<sup>211</sup>  
present a simple climatological emulator as a potential tool for<sup>212</sup>  
impacts assessments.<sup>213</sup>

## 2. Materials and Methods

### 2.1. GGCMI Phase II: experiment design

GGCMI Phase II is the continuation of a multi-model com-<sup>218</sup>  
parison exercise begun in 2014. The initial Phase I compared<sup>219</sup>  
harmonized yields of 21 models for 19 crops over a historical<sup>220</sup>  
(1980-2010) scenario with a primary goal of model evaluation<sup>221</sup>  
(Elliott et al., 2015, Müller et al., 2017). Phase II compares sim-<sup>222</sup>  
ulations of 12 models for 5 crops (maize, rice, soybean, spring<sup>223</sup>  
wheat, and winter wheat) over hundreds of scenarios in which<sup>224</sup>  
individual climate or management inputs are adjusted from<sup>225</sup>  
their historical values. The reduced set of crops includes the<sup>226</sup>  
three major global cereals and the major legume and accounts<sup>227</sup>  
for over 50% of human calories (in 2016, nearly 3.5 billion tons<sup>228</sup>

or 32% of total global crop production by weight (Food and Agriculture Organization of the United Nations, 2018).

The major goals of GGCMI Phase II are to:

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

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

The experimental protocol consists of 9 levels for precipitation perturbations, 7 for temperature, 4 for CO<sub>2</sub>, and 3 for applied nitrogen, for a total of 672 simulations for rain-fed agriculture and an additional 84 for irrigated (Table 1). For irrigated simulations, soil water is held at either field capacity or, for those models that include water-log damage, at maximum

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.

229 beneficial level. Temperature perturbations are applied as ab-247 shares a common base (e.g. LPJmL and LPJ-GUESS and the  
230 solute offsets from the daily mean, minimum, and maximum248 EPIC models), they have developed independently from this  
231 temperature time series for each grid cell used as inputs. Pre-249 shared base, for more details on the genealogy of the mod-  
232 cipitation perturbations are applied as fractional changes at the250 els see Figure S1 in Rosenzweig et al. (2014). Differences in  
233 grid cell level, and carbon dioxide and nitrogen levels are spec-251 model structure does mean that several key factors are not stan-  
234 ified as discrete values applied uniformly over all grid cells.252 dardized across the experiment, including secondary soil nutri-  
235 Note that CO<sub>2</sub> changes are applied independently of changes253 ents, carry over effects across growing years including residue  
236 in climate variables, so that higher CO<sub>2</sub> is not associated with254 management and soil moisture, and extent of simulated area for  
237 higher temperatures. An additional, identical set of scenarios255 different crops. Growing seasons are identical across models,  
238 (at the same C, T, W, and N levels) simulate adaptive agron-256 but vary by crop and by location on the globe. All stresses  
239 omy under climate change by varying the growing season for257 except factors related to nitrogen, temperature, and water (e.g.  
240 crop production. (These adaptation simulations are not shown258 Alkalinity, salinity) are disabled. No additional nitrogen inputs,  
241 or analyzed here.) The resulting GGCMI data set captures a259 such as atmospheric deposition, are considered, but some mod-  
242 distribution of crop responses over the potential space of future260 els have individual assumptions on soil organic matter that may  
243 climate conditions.261 release additional nitrogen through mineralization. See Rosen-  
244 The 12 models included in GGCMI Phase II are all mecha-263 zweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017)  
245 nistic process-based crop models that are widely used in im-  
246 pacts assessments (Table 2). Although some of the models264 for further details on models and underlying assumptions.  
265 Each model is run at 0.5 degree spatial resolution and covers

265 all currently cultivated areas and much of the uncultivated land<sup>299</sup>  
266 area. Coverage extends considerably outside currently culti-<sup>300</sup>  
267 vated areas because cultivation will likely shift under climate<sup>301</sup>  
268 change. See Figure 1 for the present-day cultivated area of<sup>302</sup>  
269 rain-fed crops, and Figure S1 in the supplemental material for<sup>303</sup>  
270 irrigated crops. Some areas such as Greenland, far-northern<sup>304</sup>  
271 Canada, Siberia, Antarctica, the Gobi and Sahara deserts, and<sup>305</sup>  
272 central Australia are not simulated as they are assumed to re-<sup>306</sup>  
273 main non-arable even under an extreme climate change. Grow-<sup>307</sup>  
274 ing seasons are standardized across models with data adapted<sup>308</sup>  
275 from several sources (Sacks et al., 2010, Portmann et al., 2008,<sup>309</sup>  
276 2010).

277 The participating modeling groups provide simulations at<sup>311</sup>  
278 any of four initially specified levels of participation, so the num-<sup>312</sup>  
279 ber of simulations varies by model, with some sampling only a<sup>313</sup>  
280 part of the experiment variable space. Most modeling groups<sup>314</sup>  
281 simulate all five crops in the protocol, but some omitted one<sup>315</sup>  
282 or more. Table 2 provides details of coverage for each model.<sup>316</sup>

283 Note that the three models that provide less than 50 simulations  
284 are excluded from the emulator analysis.

285 All models produce as output, crop yields (tons ha<sup>-1</sup> year<sup>-1</sup>)<sup>318</sup>  
286 for each 0.5 degree grid cell. Because both yields and yield<sup>319</sup>  
287 changes vary substantially across models and across grid cells,<sup>320</sup>  
288 we primarily analyze relative change from a baseline. We take<sup>321</sup>  
289 as the baseline the scenario with historical climatology (i.e. T<sup>322</sup>  
290 and P changes of 0), C of 360 ppm, and applied N at 200 kg<sup>323</sup>  
291 ha<sup>-1</sup>. We show absolute yields in some cases to illustrate geo-<sup>324</sup>  
292 graphic differences in yields for a single model.

## 293 2.2. *Simulation model validation approach*

294 To verify the skill of the process-based models used, we re-<sup>328</sup>  
295 peat the validation exercises presented in Müller et al. (2017)<sup>329</sup>  
296 for GGCMI Phase I. Note however that the GGCMI Phase II<sup>330</sup>  
297 simulations are designed for evaluating changes in yield but not<sup>331</sup>  
298 absolute yields, and so omit the calibrations used in predict-<sup>332</sup>

ing modeling to account for cultivar, pest loss, and management differences. The Phase II simulations also do not reproduce realistic nitrogen application levels for individual countries, since nitrogen is one of the parameters systematically varied. The Müller et al. (2017) validation procedure evaluates response to year-to-year temperature and precipitation variations in a control run driven by historical climate and compares it to detrended historical yields from the FAO (Food and Agriculture Organization of the United Nations, 2018) by calculating the Pearson correlation coefficient. The procedure offers no means of assessing CO<sub>2</sub> fertilization, since CO<sub>2</sub> has been relatively constant over the historical data collection period. Nitrogen data are limited for many countries, and as mentioned the GGCMI Phase II runs impose fixed and uniform nitrogen application, introducing some uncertainty into the analysis. We evaluate one or more control runs for each model, since some modeling groups provide historical runs for three different nitrogen levels.

### 317 2.3. *Climatological-mean yield emulator design*

To demonstrate the properties of the GGCMI Phase II dataset, we construct an emulator of 30-year climatological mean yields, which are of most interest to impact modelers. This approach differs from previous studies of crop model emulation, which have typically emulated at the annual level. Annual emulation is required when the input training set consists of non-stationary projections of evolving yields (such as an RCP run). Recent studies (e.g. Fronzek et al., 2018, Snyder et al., 2018) that used a training set of stationary simulations with fixed variations in parameters allow emulating the climatological mean response instead. The two can differ for multiple reasons, including any year-to-year memory in the crop model, or if the distribution of growing-season daily temperatures associated with interannual variability is different from that associated with long-term CO<sub>2</sub>-driven changes. The confounding

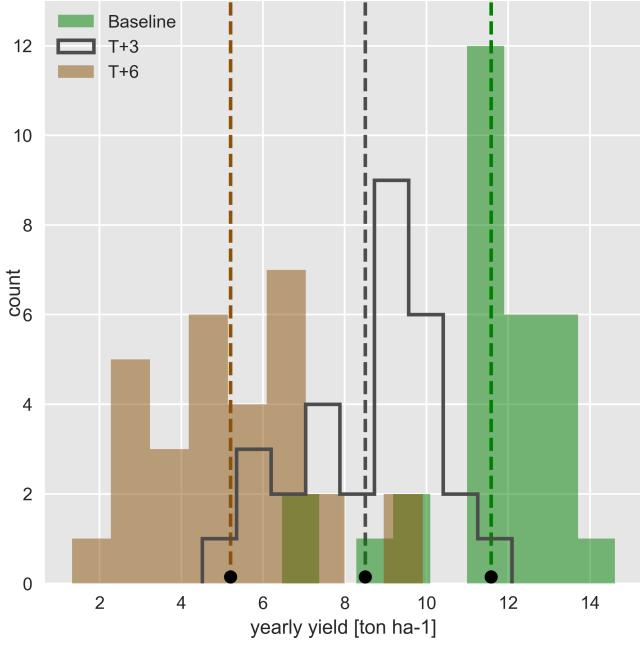


Figure 2: Example showing both climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for a grid cell in northern Iowa (a representative high-yield region) from the pDSSAT model, for the baseline climatology (1981–2010) and for scenarios 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.

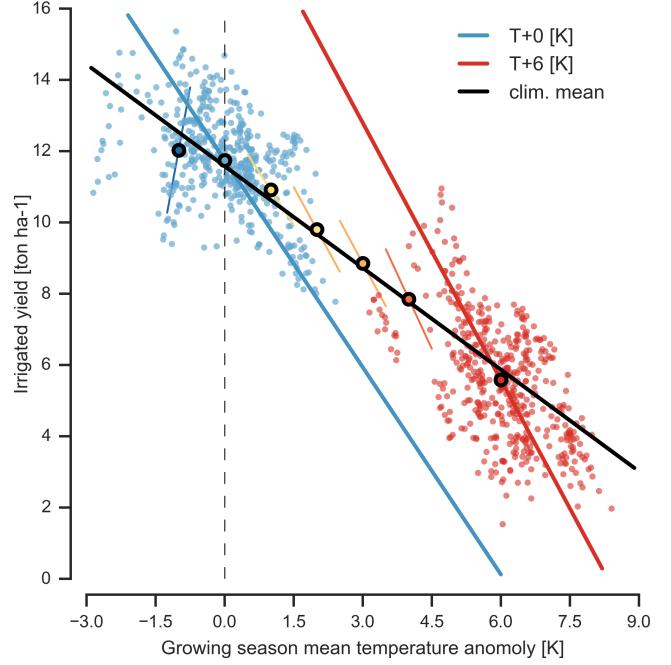


Figure 3: Example showing temperature relationship developed from year-to-year values vs. climatological mean values. Figure shows irrigated maize for nine adjacent grid cell in northern Iowa (a representative high-yield region) from the pDSSAT model, for the baseline climatology (1981–2010) and for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

of year-to-year response with climatological response might be considered a class of Simpson’s paradox. Crop yields in process based models do not respond to the mean growing season temperature, they respond to the full distribution in temperature over the growing season (or, specifically the exact temperature time series). Much of the variance is left unexplained if one tries to fit a statistical model between yields and some aggregate temperature variable (mean growing season temperature, monthly temperature etc.). Application of relationships obtained from such statistical models to mean changes in climate may provide problematic. The year-over-year yield response to individual factors in GGCMI Phase II do in fact often exceeds the climatological response (Figure 3). Note that the GGCMI Phase II datasets will not capture distributional shifts, because all simulations are run with fixed offsets from the historical climatology. (For methods to generate adjusted historical climate data inclusive of distributional changes, see Haugen et al. (2018) and Poppick et al. (2016)). Emulation approaches are an area of active ongoing study and one of the goals of the GGCMI Phase II dataset is to facilitate these efforts.

Emulation involves fitting individual regression models for each crop, simulation model, and 0.5 degree geographic pixel from the GGCMI Phase II data set. The regressors are the applied constant perturbations in temperature, water, nitrogen and CO<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

364 temporal dependencies. (Future work should explore this). The<sub>381</sub>  
 365 climatological emulation indirectly includes any yield response<sub>382</sub>  
 366 to geographically distributed factors such as soil type, insola-<sub>383</sub>  
 367 tion, and the baseline climate itself, because we construct sep-<sub>384</sub>  
 368 arate emulators for each grid cell. The emulator parameter ma-<sub>385</sub>  
 369 trices are portable and the yield computations are cheap even at<sub>386</sub>  
 370 the half-degree grid cell resolution, so we do not aggregate in<sub>387</sub>  
 371 space at this time.<sub>388</sub>

372 We regress climatological mean yields against a third-order  
 373 polynomial in C, T, W, and N with interaction terms. The  
 374 higher-order terms are necessary to capture any nonlinear re-  
 375 spondes, which are well-documented in observations for tem-  
 376 perature and water perturbations (e.g. Schlenker & Roberts  
 377 (2009) for T and He et al. (2016) for W). We include inter-  
 378 action terms (both linear and higher-order) because past stud-  
 379 ies have shown them to be significant effects. For example,  
 380 Lobell & Field (2007) and Tebaldi & Lobell (2008) showed<sub>397</sub>

that in real-world yields, the joint distribution in T and W is  
 needed to explain observed yield variance. (C and N are fixed  
 in these data.) Other observation-based studies have shown the  
 importance of the interaction between water and nitrogen (e.g.  
 Aulakh & Malhi, 2005), and between nitrogen and carbon diox-  
 ide (Osaki et al., 1992, Nakamura et al., 1997). We do not fo-  
 cuse on comparing different functional forms in this study, and  
 instead choose a relatively simple parametrization that allows  
 for some interpretation of coefficients. Some prior studies have  
 used more complex functional forms and larger numbers of pa-  
 rameters, e.g. 39 in Blanc & Sultan (2015) and Blanc (2017),  
 who borrow information across space by fitting grid points si-  
 multaneously across a large region in a panel regression. **We**  
**choose a simpler emulation at grid-cell level to avoid the re-**  
**quirement of assuming responses are uniform across space and**  
**to maximize interpretability.**

The limited GGCMI variable sample space means that use

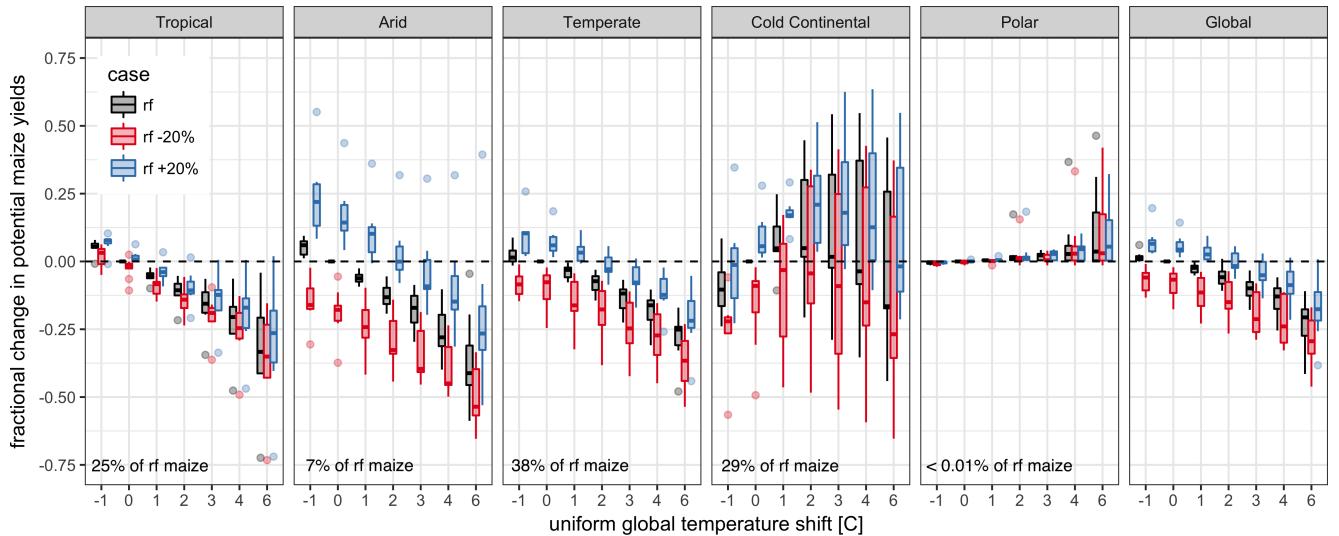


Figure 4: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation levels (rain-fed (rf) -20%, rain-fed (0), and rain-fed +20%), with 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 (data from Portmann et al., 2010). Box-and-whiskers plots show distribution across models, with median marked. Box edges are first and third quartiles, i.e. box height is the interquartile range (IQR). Whiskers extend to maximum and minimum of simulations but are limited at 1.5-IQR; otherwise the outlier is shown. Models generally agree in most climate regions (other than cold continental), with projected changes larger than inter-model variance. Outliers in the tropics (strong negative impact of temperature increases) are the pDSSAT model; outliers in the high-rainfall case (strong positive impact of precipitation increases) are the JULES model. Inter-model variance increases in the case where precipitation is reduced, suggesting uncertainty in model response to water limitation. The right panel with global changes shows yield responses to an globally uniform temperature shift; note that these results are not directly comparable to simulations of more realistic climate scenarios with the same global mean change.

398 of the full polynomial expression described above, which has  
 399 34 terms for the rain-fed case (12 for irrigated), can be prob-  
 400 lematic, and can lead to over-fitting and unstable parameter es-  
 401 timations. We therefore reduce the number of terms through a  
 402 feature selection cross-validation process in which terms in the  
 403 polynomial are tested for importance. In this procedure higher-  
 404 order and interaction terms are added successively to the model;  
 405 we then follow the reduction of the the aggregate mean squared  
 406 error with increasing terms and eliminate those terms that do  
 407 not contribute significant reductions. See supplemental docu-  
 408 ments for more details. We select terms by applying the feature  
 409 selection process to the three models that provided the com-  
 410 plete set of 672 rain-fed simulations (pDSSAT, EPIC-TAMU,  
 411 and LPJmL); the resulting choice of terms is then applied for  
 412 all emulators.

413 Feature importance is remarkably consistent across all three<sup>436</sup>  
 414 models and across all crops (see Figure S4 in the supplemental<sup>437</sup>  
 415 material). The feature selection process results in a final poly-<sup>438</sup>  
 416 nomial in 23 terms, with 11 terms eliminated. We omit the  $N^3$ <sup>439</sup>  
 417 term, which cannot be fitted because we sample only three ni-<sup>440</sup>  
 418 trogen levels. We eliminate many of the C terms: the cubic,<sup>441</sup>  
 419 the CT, CTN, and CWN interaction terms, and all higher order  
 420 interaction terms in C. Finally, we eliminate two 2nd-order in-<sup>442</sup>  
 421 teraction terms in T and one in W. Implication of this choice<sup>443</sup>  
 422 include that nitrogen interactions are complex and important,<sup>444</sup>  
 423 and that water interaction effects are more nonlinear than those<sup>445</sup>  
 424 in temperature. The resulting statistical model (Equation 1) is<sup>446</sup>  
 425 used for all grid cells, models, and crops:

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

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

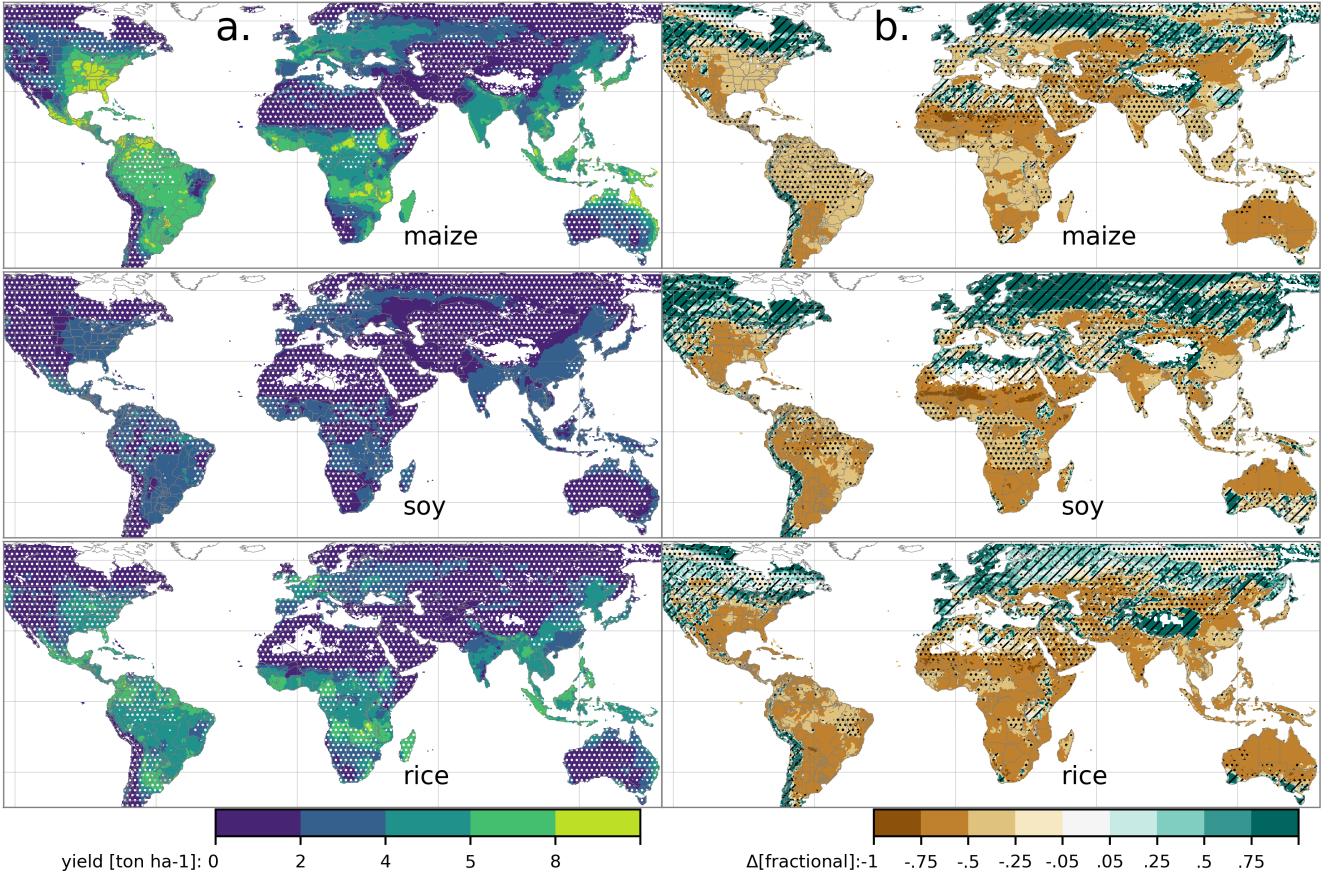


Figure 5: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure S11 in the supplemental material.) White stippling indicates areas where these crops are not currently cultivated. Absence of cultivation aligns well with the lowest yield contour ( $0.2 \text{ ton ha}^{-1}$ ). 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.

#### 448 2.4. Emulator evaluation

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

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

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

Here  $F_{scn.}$  is the fractional change in a model’s mean emulated or simulated yield from a defined baseline, in some scenario (scn.) in C, T, W, and N space;  $Y_{scn.}$  and  $Y_{baseline}$  are the absolute emulated or simulated mean yields. The normalized error  $e$  is the difference between the emulated fractional change in yield and that actually simulated, normalized by  $\sigma_{sim.}$ , 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

469 S12 and Figures S13 in supplemental documents). Note that<sub>502</sub>  
470 the normalized error  $e$  for a model depends not only on the fi-<sub>503</sub>  
471 delity of its emulator in reproducing a given simulation but on<sub>504</sub>  
472 the particular suite of models considered in the intercomparison<sub>505</sub>  
473 exercise. The rationale for this choice is to relate the fidelity of<sub>506</sub>  
474 the emulation to an estimate of true uncertainty, which we take<sub>507</sub>  
as the multi-model spread.<sub>508</sub>

### 476 3. Results

#### 477 3.1. Simulation results

478 Crop models in the GGCMI ensemble show a broadly con-<sub>509</sub>  
479 sistent responses to climate and management perturbations in  
most regions, with a strong negative impact of increased tem-<sub>510</sub>  
480 perature in all but the coldest regions. We illustrate this result  
for rain-fed maize in Figure 4, which shows yields for the pri-<sub>511</sub>  
481 mary Köppen-Geiger climate regions (Rubel & Kottek, 2010).  
In warming scenarios, models show decreases in maize yield in  
482 the temperate, tropical, and arid regions that account for nearly  
483 three-quarters of global maize production. These impacts are  
484 robust for even moderate climate perturbations. In the temper-<sub>512</sub>  
485 ate zone, even a 1 degree temperature rise with other variables<sub>513</sub>  
486 held fixed leads to a median yield reduction that outweighs the  
variance across models. A 6 degree temperature rise results in  
487 median loss of ~25% of yields with a signal to noise of nearly  
488 three. A notable exception is the cold continental region, where  
489 models disagree strongly, extending even to the sign of impacts.<sub>514</sub>  
Model simulations of other crops produce similar responses to  
490 warming, with robust yield losses in warmer locations and high  
491 inter-model variance in the cold continental regions (Figures<sub>515</sub>  
492 S7).<sub>516</sub>

493 The effects of rainfall changes on maize yields are also as ex-<sub>517</sub>  
494 pected and are consistent across models. Increased rainfall mit-<sub>518</sub>  
495 igates the negative effect of higher temperatures, most strongly<sub>519</sub>  
496 in arid regions. Decreased rainfall amplifies yield losses and<sub>520</sub>  
497

also increases inter-model variance more strongly, suggesting  
that models have difficulty representing crop response to water  
stress. We show only rain-fed maize here; see Figure S5 for the  
irrigated case. As expected, irrigated crops are more resilient to  
temperature increases in all regions, especially so where water  
is limiting.

508 Mapping the distribution of baseline yields and yield changes  
509 shows the geographic dependencies that underlie these results.  
510 Figure 5 shows baseline and changes in the T+4 scenario for  
511 rain-fed maize, soy, and rice in the multi-model ensemble mean,  
512 with locations of model agreement marked. Absolute yield po-  
513 tentials are have strong spatial variation, with much of the  
514 Earth's surface area unsuitable for any given crop. In general,  
models agree most on yield response in regions where yield  
515 potentials are currently high and therefore where crops are cur-  
516 rently grown. Models show robust decreases in yields at low  
517 latitudes, and highly uncertain median increases at most high  
518 latitudes. For wheat crops see Figure S11; wheat projections  
519 are both more uncertain and show fewer areas of increased yield  
520 in the inter-model mean.

#### 521 3.2. Simulation model validation results

522 Figure 6 shows the Pearson time series correlation between  
the simulation model yield and FOA yield data. Figure 6 can be  
523 compared to Figures 1,2,3,4 and 6 in Müller et al. (2017). The  
524 results are mixed, with many regions for rice and wheat be-  
ing difficult to model. No single model is dominant, with each  
model providing near best-in-class performance in at least one  
525 location-crop combination. The presence of very few vertical  
dark green color bars clearly illustrates the power of a multi-  
526 model intercomparison project like the one presented here. The  
ensemble mean does not beat the best model in each case, but  
527 shows positive correlation in over 75% of the cases presented  
here. The EPIC-TAMU model performs best for soy, CARIAB,  
528 EPIC-TAMU, and PEPIC perform best for maize, PROMET

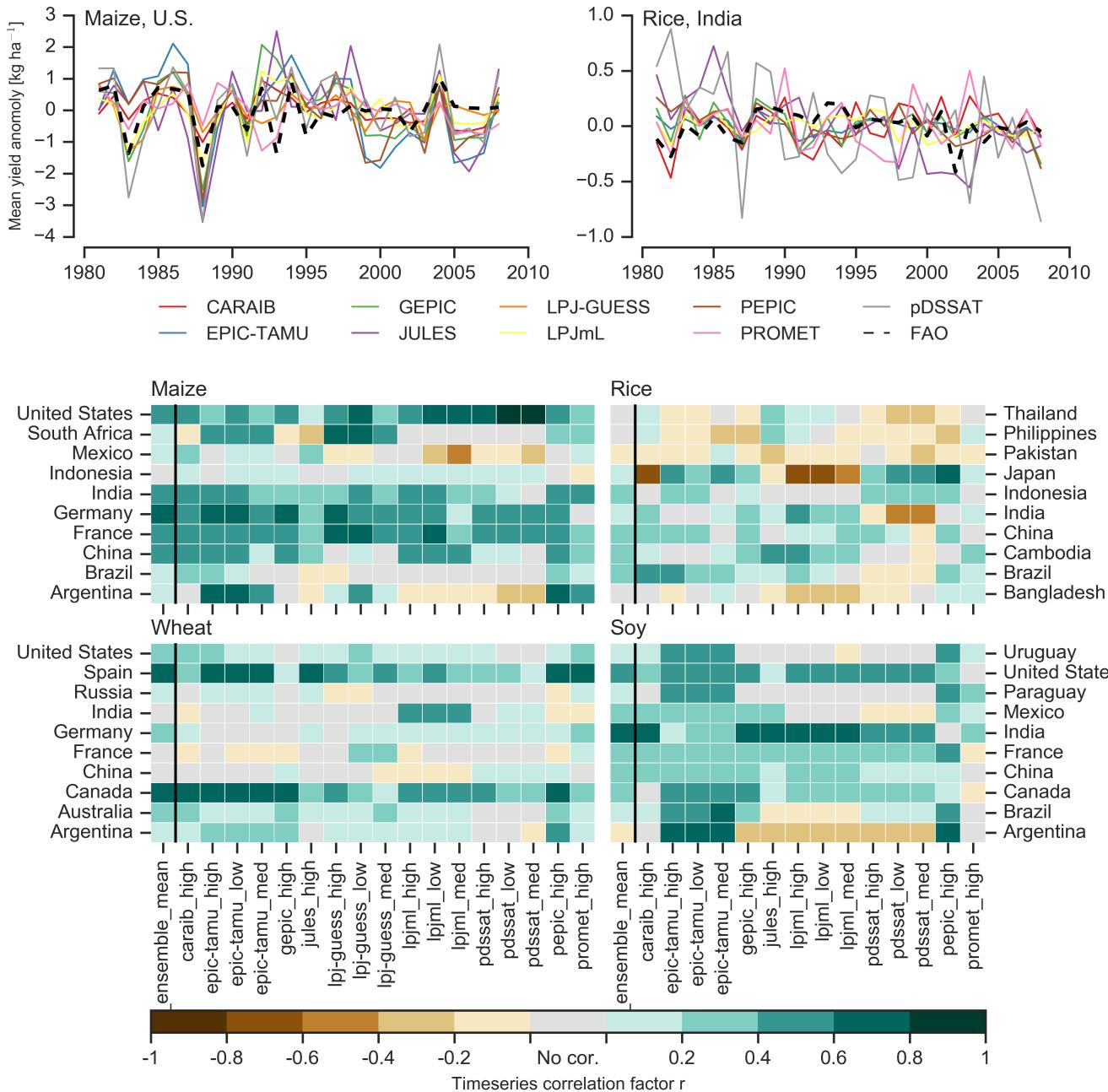


Figure 6: Time series correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen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

536 performs best for wheat, and the EPIC family of models per-  
 537 form best for rice. Reductions in skill over the performance  
 538 illustrated in Müller et al. (2017) can be attributed to the nitro-  
 539 gen levels or lack of calibration in some models.

540 \*\*\* or harmonization \*\*\* Christoph

Soy is qualitatively the easiest crop to represent (except in Argentina), which is likely due in part to the invariance of the response to nitrogen application (soy fixes atmospheric nitrogen very efficiently). Comparison to the FAO data is therefore easier than the other crops because the nitrogen application levels do

546 not matter. US maize has the best performance across models,<sup>580</sup>  
547 with nearly every model representing the historical variability<sup>581</sup>  
548 to a reasonable extent. Especially good example years for US<sup>582</sup>  
549 maize are 1983, 1988, and 2004 (top left panel of Figure 6),<sup>583</sup>  
550 where every model gets the direction of the anomaly compared<sup>584</sup>  
551 to surrounding years correct. 1983 and 1988 are famously bad<sup>585</sup>  
552 years for US maize along with 2012 (not shown). US maize<sup>586</sup>  
553 is possibly both the most uniformly industrialized (in terms of<sup>587</sup>  
554 management practices) crop and the one with the best data col-<sup>588</sup>  
555 lection in the historical period of all the cases presented here.<sup>589</sup>

556 The FAO data is at least one level of abstraction from ground<sup>590</sup>  
557 truth in many cases, especially in developing countries. The<sup>591</sup>  
558 failure of models to represent the year-to-year variability in rice<sup>592</sup>  
559 in some countries in southeast Asia is likely partly due to model<sup>593</sup>  
560 failure and partly due to lack of data. It is possible to speculate  
561 that the difference in performance between Pakistan (no suc-<sup>594</sup>  
562 cessful models) and India (many successful models) for rice<sup>595</sup>  
563 may reside at least in part in the FAO data and not the mod-<sup>596</sup>  
564 els themselves. The same might apply to Bangladesh and In-<sup>597</sup>  
565 dia for rice. Partitioning of these contributions is impossible at<sup>598</sup>  
566 this stage. Additionally, there is less year-to-year variability in<sup>599</sup>  
567 rice yields (partially due to the fraction of irrigated cultivation).<sup>600</sup>  
568 Since the Pearson r metric is scale invariant, it will tend to score  
569 the rice models more poorly than maize and soy. An example<sup>601</sup>  
570 of very poor performance can be seen with the pDSSAT model<sup>602</sup>  
571 for rice in India (top right panel of Figure 6).<sup>603</sup>

### 572 3.3. Emulator performance

573 Emulation provides not only a computational tool but a<sup>607</sup>  
574 means of understanding and interpreting crop yield response<sup>608</sup>  
575 across the parameter space. Emulation is only possible, how-<sup>609</sup>  
576 ever, when crop yield responses are sufficiently smooth and<sup>610</sup>  
577 continuous to allow fitting with a relatively simple functional<sup>611</sup>  
578 form. In the GGCMI simulations, this condition largely but<sup>612</sup>  
579 not always holds. Responses are quite diverse across locations,<sup>613</sup>

crops, and models, but in most cases local responses are reg-  
ular enough to permit emulation. Figure 7 illustrates the geo-  
graphic diversity of responses even in high-yield areas for a  
single crop and model (rain-fed maize in pDSSAT for various  
high-cultivation areas). This heterogeneity validates the choice  
of emulating at the grid cell level.

Each panel in Figure 7 shows model yield output from sce-  
narios varying only along a single dimension ( $\text{CO}_2$ , tempera-  
ture, 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. Fig-  
ure 8 illustrates the inter-model diversity of yield responses  
to the same perturbations, even for a single crop and location  
(rain-fed maize in northern Iowa, the same location shown in  
the Figure 7). The differences make it important to construct  
emulators separately for each individual model, and the fidelity  
of emulation can also differ across models. This figure illus-  
trates a common phenomenon, that models differ more in re-  
sponse to perturbations in  $\text{CO}_2$  and nitrogen perturbations than  
to those in temperature or precipitation. (Compare also Figures  
4 and S18.) For this location and crop,  $\text{CO}_2$  fertilization effects  
can range from ~5–50%, and nitrogen responses from nearly  
flat to a 60% drop in the lowest-application simulation.

While the nitrogen dimension is important and uncertain, it  
is also the most problematic to emulate in this work because  
of its limited sampling. The GGCMI protocol specified only  
three nitrogen levels (10, 60 and 200  $\text{kg N y}^{-1} \text{ha}^{-1}$ ), so a third-  
order fit would be over-determined but a second-order fit can  
result in potentially unphysical results. Steep and nonlinear de-

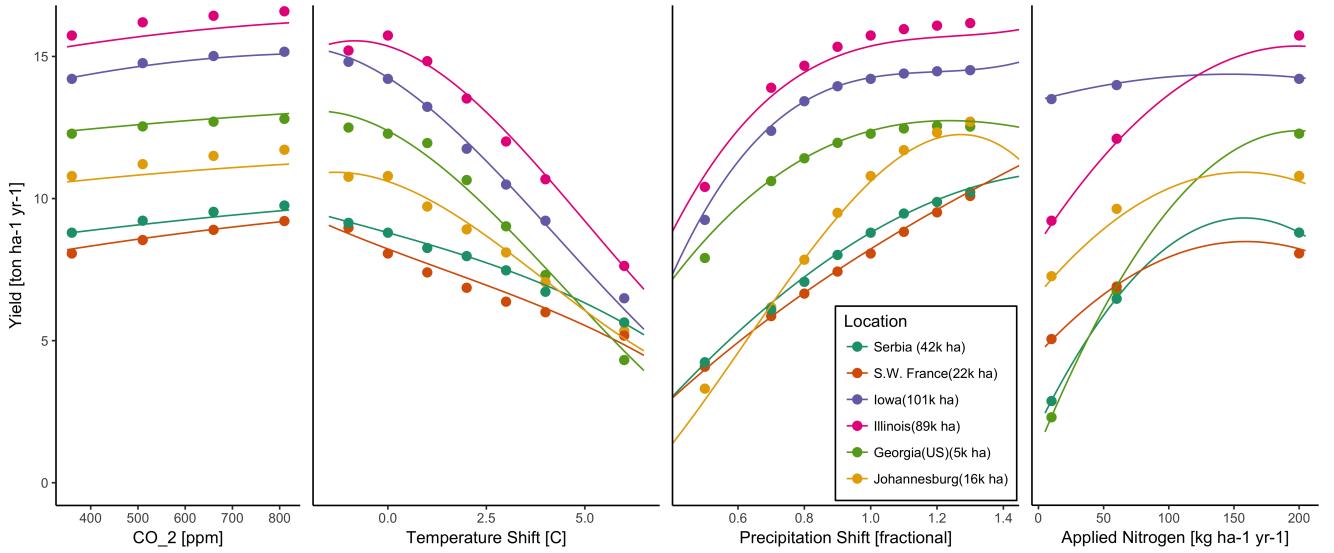


Figure 7: Example illustrating spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows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

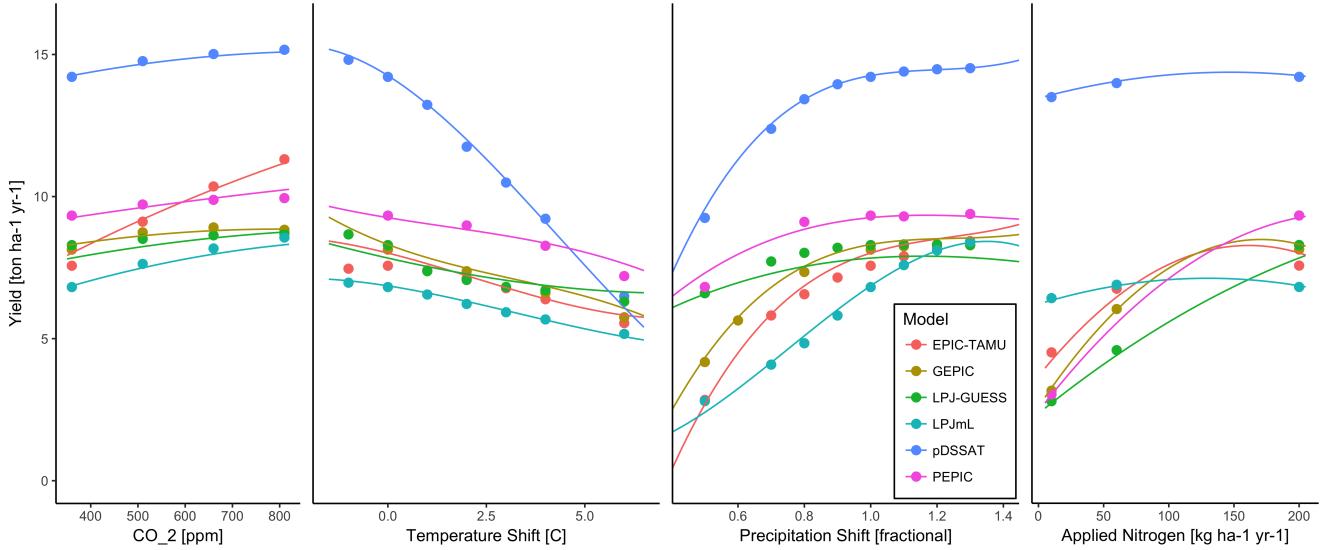


Figure 8: Example illustrating across-model variations in yield response. We show simulations and emulations from six models for rain-fed maize in the same Iowa location shown in Figure 7, with the same plot conventions. Models that did not simulate the nitrogen dimension are omitted for clarity. The 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.

614 clines in yield with lower nitrogen levels means that some re-<sub>619</sub>  
 615 gressions imply a peak in yield between the 100 and 200 kg N<sub>620</sub>  
 616 y<sup>-1</sup> ha<sup>-1</sup> levels. While there may be some reason to believe<sub>621</sub>  
 617 over-application of nitrogen at the wrong time in the growing<sub>622</sub>  
 618 season could lead to reduced yields, these features are almost

certainly an artifact of under sampling. In addition, the polynomial fit cannot capture the well-documented saturation effect of nitrogen application (e.g. Ingestad, 1977) as accurately as would be possible with a non-parametric model.

623 To assess the ability of the polynomial emulation to capture

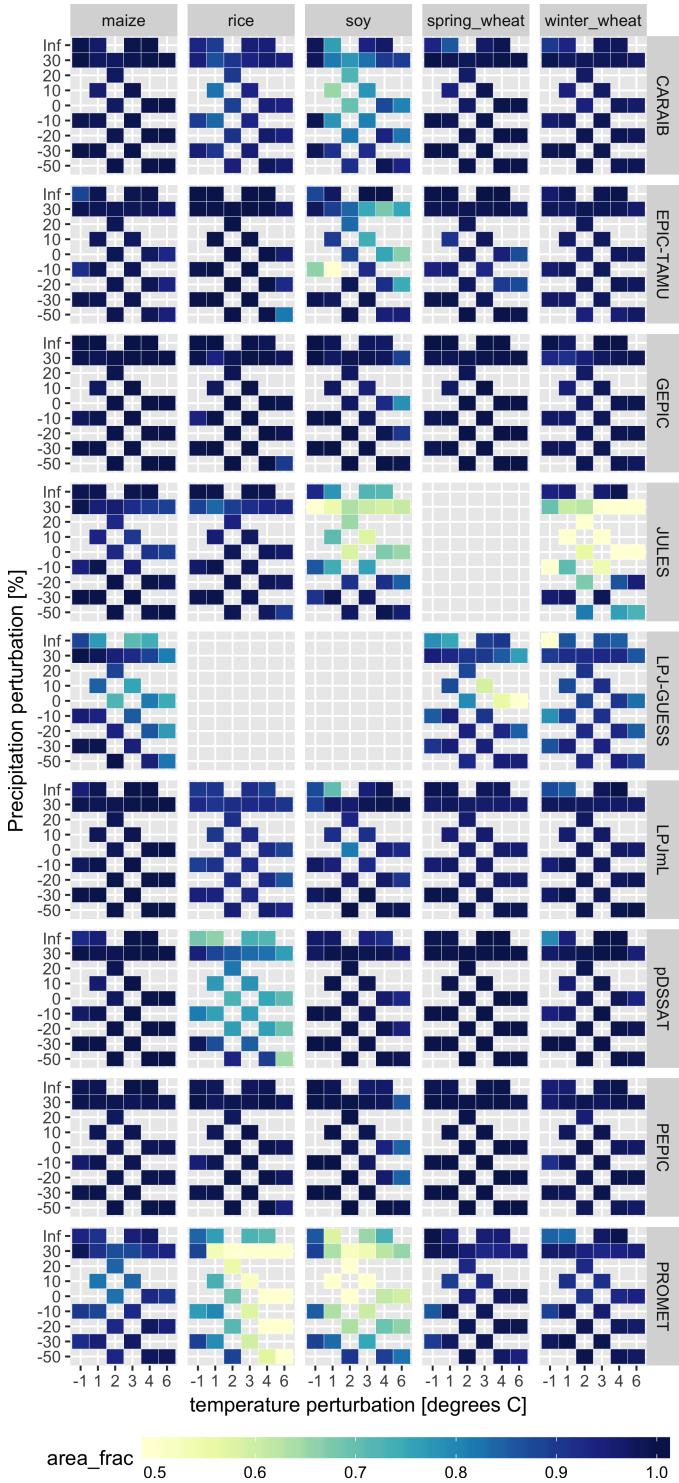


Figure 9: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO<sub>2</sub> and nitrogen held fixed at baseline values. Large columns are crops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

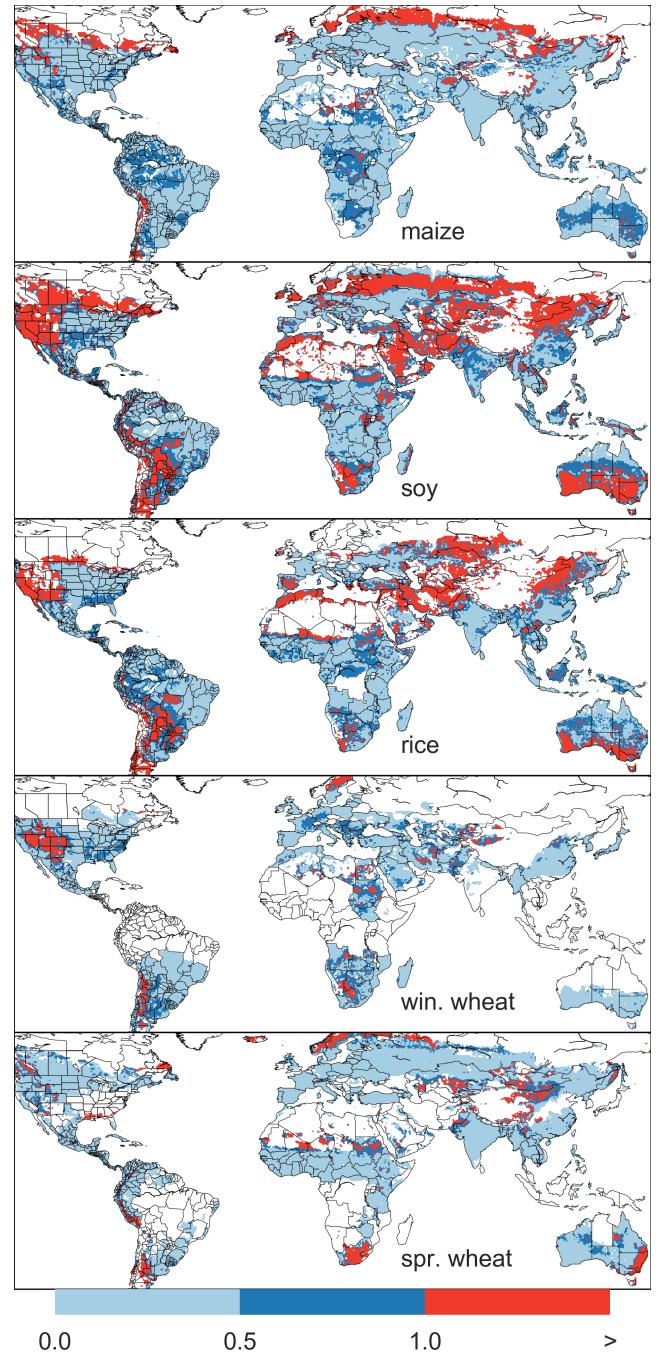


Figure 10: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error  $e$ , where  $e > 1$  means that emulator error exceeds the multi-model standard deviation. White areas are those where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

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

It should be noted that this assessment metric is relatively  
forgiving. First, each emulation is evaluated against the simu-  
lation 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 magni-  
tude 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 sim-  
ulations 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 a easy way to compare a  
ensembles of climate or impacts projections. They also pro-  
vide 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 supple-  
mental material for other crops and dimensions.) The emu-  
lated values closely match simulations even at this aggrega-  
tion level. Note that these functions are presented only as  
examples and do not represent true global projections, be-  
cause they are developed from simulation data with a uniform  
temperature shift while increases in global mean temperature  
should manifest non-uniformly. The global coverage of the  
GGCMI simulations allows impacts modelers to apply arbitrary  
geographically-varying climate projections, as well as arbitrary  
aggregation mask, to develop damage functions for any climate  
scenario and any geopolitical or geographic level.

## 4. Conclusions and discussion

The GGCMI Phase II experiment provides a database tar-  
geted to allow detailed study of crop yields from process-based  
models under climate change. The experiment is designed to  
facilitate not only comparing the sensitivities of process-based  
crop yield models to changing climate and management inputs  
but also evaluating the complex interactions between driving  
factors (CO<sub>2</sub>, temperature, precipitation, and applied nitrogen).  
Its global nature also allows identifying geographic shifts in  
high yield potential locations. We expect that the simulations

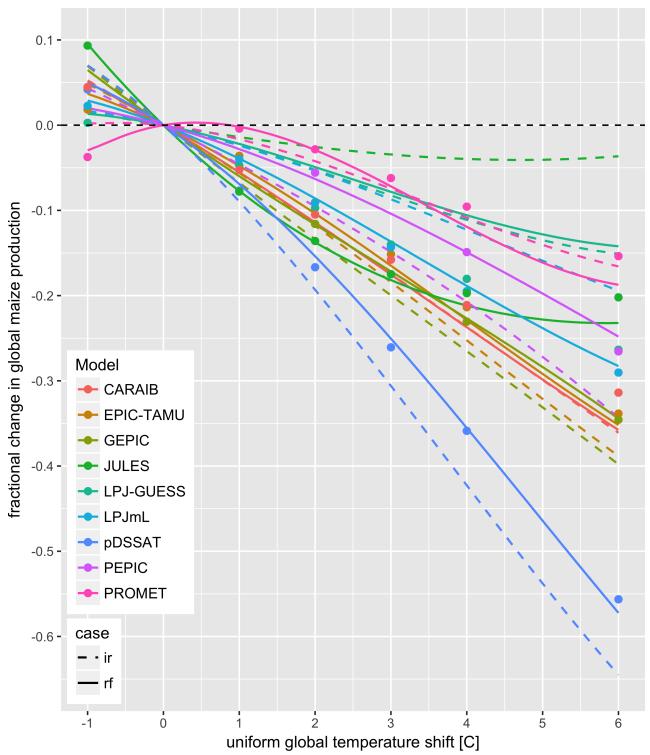


Figure 11: Global emulated damages for maize on currently cultivated lands for the GGCMI models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating up<sup>717</sup> emulated values at the grid cell level, not from a regression of global mean yields.) Lines are emulations for rain-fed (solid) and irrigated (dashed) crops;<sup>718</sup> for comparison, dots are the simulated values for the rain-fed case. For most locations of cultivated areas: irrigated crops tend to be grown in warmer areas where impacts are more severe for a given temperature shift. (The exceptions<sup>720</sup> are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures S16-S19 in the supplemental material.

<sup>689</sup> will yield multiple insights in future studies, and show here a<sub>723</sub>  
<sup>690</sup> selection of preliminary results to illustrate their potential uses.<sub>724</sub>

<sup>691</sup> First, the GGCMI Phase II simulations allow identifying ma-<sub>725</sub>  
<sup>692</sup> jor areas of uncertainty. Across the major crops, inter-model<sub>726</sub>  
<sup>693</sup> uncertainty is greatest for wheat and least for soy. Across fac-<sub>727</sub>  
<sup>694</sup> tors impacting yields, inter-model uncertainty is largest for CO<sub>2</sub><sub>728</sub>  
<sup>695</sup> fertilization and nitrogen response effects. Across geographic<sub>729</sub>  
<sup>696</sup> regions, projections are most uncertain in the high latitudes<sub>730</sub>  
<sup>697</sup> where yields may increase, and most robust in low latitudes<sub>731</sub>  
<sup>698</sup> where yield impacts are largest.

<sup>699</sup> Second, the GGCMI Phase II simulations allow understand-<sub>733</sub>  
<sup>700</sup> ing the way that climate-driven changes and locations of cul-<sub>734</sub>  
<sup>701</sup> tivated land combine to produce yield impacts. One coun-<sub>735</sub>

terintuitive result immediate apparent is that irrigated maize shows steeper yield reductions under warming than does rain-fed maize when considered only over currently cultivated land. The effect results from geographic differences in cultivation. In any given location, irrigation increases crop resiliency to temperature increase, but irrigated maize is grown in warmer locations where the impacts of warming are more severe (Figures S5-S6). The same behavior holds for rice and winter wheat, but not for soy or spring wheat (Figures S8-S10). Irrigated wheat and maize are also more sensitive to nitrogen fertilization levels than are analogous non-irrigated crops, presumably because those rain-fed crops are limited by water as well as nitrogen availability (Figure S19). (Soy as an efficient atmospheric nitrogen-fixer is relatively insensitive to nitrogen, and rice is not generally grown in water-limited conditions).

Third, we show that even the relatively limited GGCMI Phase II sampling space allows emulation of the climatological response of crop models with a relatively simple reduced-form statistical model. The systematic parameter sampling in the GGCMI Phase II procedure provides information on the influence of multiple interacting factors in a way that single projections cannot, and emulating the resulting response surface then produces a tool that can aid in both physical interpretation of the process-based models and in assessment of agricultural impacts under arbitrary climate scenarios. Emulating the climatological response isolates long-term impacts from any confounding factors that complicate year-over-year changes, and the use of simple functional forms offer the possibility of physical interpretation of parameter values. Care should be taken in applying relationships developed at the yearly level to shifts in the mean climatology. We anticipate that systematic parameter sampling will become the norm in future model intercomparison exercise.

While the GGCMI Phase II database should offer the foun-

736 dation for multiple future studies, several cautions need to be<sup>770</sup>  
737 noted. Because the simulation protocol was designed to fo-<sup>771</sup>  
738 cus on change in yield under climate perturbations and not<sup>772</sup>  
739 on replicating real-world yields, the models are not formally<sup>773</sup>  
740 calibrated so cannot be used for impacts projections unless in<sup>774</sup>  
741 used in conjunction with historical data (or data products). Be-<sup>775</sup>  
742 cause the GGCMI simulations apply uniform perturbations to<sup>776</sup>  
743 historical climate inputs, they do not sample changes in higher<sup>777</sup>  
744 order moments, and cannot address the additional crop yield<sup>778</sup>  
745 impacts of potential changes in climate variability. Although<sup>779</sup>  
746 distributional changes in model projections are fairly uncertain<sup>780</sup>  
747 at present, follow-on experiments may wish to consider them.

748 Several recent studies have described procedures for generating<sup>781</sup>  
749 simulations that combine historical data with model projections  
750 of not only mean changes in temperature and precipitation but  
751 changes in their marginal distributions (e.g. Chang et al., 2016)  
752 or temporal dependence (e.g. Leeds et al., 2015).

753 The GGCMI phase II output dataset invites a broad range of<sup>786</sup>  
754 potential future avenues of analysis. A major target area in-<sup>787</sup>  
755 volves studying the models themselves with a detailed exami-<sup>788</sup>  
756 nation of interaction terms between the major input drivers, a<sup>789</sup>  
757 more robust quantification of the sensitivity of different models<sup>790</sup>  
758 to the input drivers, and comparisons with field-level experi-<sup>791</sup>  
759 mental data. The parameter space tested in GGCMI phase II<sup>792</sup>  
760 will allow detailed investigations into yield variability and re-<sup>793</sup>  
761 sponse to extremes under changing management and CO<sub>2</sub> lev-<sup>794</sup>  
762 els. As mentioned previously, the database allows study of ge-<sup>795</sup>  
763 ographic shifts in optimal growing regions for different crops<sup>796</sup>  
764 and studying the viability of switching crop types in some ar-<sup>797</sup>  
765 eas. The output dataset also contains other runs and variables<sup>798</sup>  
766 not analyzed or shown here. Runs include several which al-<sup>799</sup>  
767 lowed adaptation to climate changes by altering growing sea-<sup>800</sup>  
768 sons, and additional variables include above ground biomass,<sup>801</sup>  
769 LAI, and root biomass (as many as 25 output variables for some<sup>802</sup>

models). Emulation studies that are possible include study of year-over-year vs climatological emulation, and more systematic evaluation of different statistical model specifications and formal calculation of uncertainties in derived parameters.

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

## 5. Acknowledgments

We thank Michael Stein and Kevin Schwarzwald, who provided helpful suggestions that contributed to this work. This research was performed as part of the Center for Robust Decision-Making on Climate and Energy Policy (RDCEP) at the University of Chicago, and was supported through a variety of sources. RDCEP is funded by NSF grant #SES-1463644 through the Decision Making Under Uncertainty program. J.F. was supported by the NSF NRT program, grant #DGE-1735359. C.M. was supported by the MACMIT project (01LN1317A) funded through the German Federal Ministry of Education and Research (BMBF). C.F. was supported by the European Research Council Synergy grant #ERC-2013-SynG-610028 Imbalance-P. P.F. and K.W. were supported by the Newton Fund through the Met Office Climate Science for Service Partnership Brazil (CSSP Brazil). A.S. was supported by the Office of Science of the U.S. Department of Energy as part of the Multi-sector Dynamics Research Program Area. Computing resources were provided by the University of Chicago Research Computing Center (RCC). S.O. acknowledges support from the Swedish strong research areas BECC and MERGE together with support from LUCCI (Lund University Centre for studies of Car-

- 803 bon Cycle and Climate Interactions).
- 804
- ## 6. References
- 805 Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013).  
 806 Implication of crop model calibration strategies for assessing regional im-  
 807 pacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.  
 808 Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D.,  
 809 A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Al-  
 810 derman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Bier-  
 811 nath, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising tempera-  
 812 tures reduce global wheat production. *Nature Climate Change*, 5, 143–147.  
 813 doi:10.1038/nclimate2470.
- 814 Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A.,  
 815 J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N.,  
 816 Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath,  
 817 C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simula-  
 818 ting wheat yields under climate change. *Nature Climate Change*, 3, 827832.  
 819 doi:10.1038/nclimate1916.
- 820 Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other  
 821 Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Ef-  
 822 ficiency, Carbon Sequestration, and Environmental Pollution. *Advances in*  
 823 *Agronomy*, 86, 341 – 409.
- 824 Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov,  
 825 N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat  
 826 production potentials and management flexibility under the representative  
 827 concentration pathways. *Global and Planetary Change*, 122, 107 – 121.
- 828 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields  
 829 from global gridded crop models. *Agricultural and Forest Meteorology*, 236,  
 830 145 – 161.
- 831 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded  
 832 crop models using statistical estimates. *Agricultural and Forest Meteorol-*  
 833 *ogy*, 214-215, 134 – 147.
- 834 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S.  
 835 (2018). Implementing the Nitrogen cycle into the dynamic global vegeta-  
 836 tion, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific*  
 837 *Model Development*, 11, 2789–2812.
- 838 Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L.,  
 839 & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections  
 840 Based on Precomputed GCM Runs. *Journal of Climate*, 27, 1829–1844.
- 841 Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.  
 842 (2014). A meta-analysis of crop yield under climate change and adaptation.  
 843 *Nature Climate Change*, 4, 287 – 291.
- 844 Chang, W., Stein, M., Wang, J., Kotamarthi, V., & Moyer, E. (2016). Changes in  
 845 spatio-temporal precipitation patterns in changing climate conditions. *Jour-*  
 846 *nal of Climate*, 29. doi:10.1175/JCLI-D-15-0844.1.
- 847 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process  
 848 emulation of dynamic computer codes. *Biometrika*, 96, 663–676.
- 849 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In  
 850 C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth,*  
 851 *Duke University, Durham, North Carolina* (pp. 115–118). Durham, North  
 852 Carolina.
- 853 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for  
 854 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).
- 855 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdoos,  
 856 M., & François, L. (2011). Responses of European forest ecosystems to  
 857 21st century climate: assessing changes in interannual variability and fire  
 858 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).
- 859 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,  
 860 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating  
 861 impact models and sectors (pSIMS). *Environmental Modelling and Soft-*  
 862 *ware*, 62, 509–516.
- 863 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,  
 864 Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C.,  
 865 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.  
 866 (2015). The Global Gridded Crop Model Intercomparison: data and mod-  
 867 eling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8,  
 868 261–277.
- 869 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-  
 870 mate change impacts on durum wheat in the mediterranean region. *Natural*  
 871 *Hazards and Earth System Sciences*, 11, 1293–1302.
- 872 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Re-  
 873 gionalization of a large-scale crop growth model for sub-Saharan Africa:  
 874 Model setup, evaluation, and estimation of maize yields. *Agriculture,*  
 875 *Ecosystems & Environment*, 151, 21 – 33.
- 876 Food and Agriculture Organization of the United Nations (2018). FAOSTAT  
 877 database. URL: <http://www.fao.org/faostat/en/home>.
- 878 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T.,  
 879 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P.,  
 880 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F.,  
 881 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,  
 882 Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J.,  
 883 Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nen-  
 884 del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna,  
 885 M., Semenov, M. A., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K.,  
 886 Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model

- wheat yield impact response surfaces showing sensitivity to temperature and precipitation change. *Agricultural Systems*, 159, 209–224.
- Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014). Evaluating the utility of dynamical downscaling in agricultural impacts projections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781.
- Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitivity of Agricultural Model Performance to Different Climate Inputs. *Journal of Applied Meteorology and Climatology*, 55, 151113145618001.
- Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous Crop Growth and Yield: Application for Wheat in Central Europe. *Remote Sensing*, 7, 3934–3965.
- Haugen, M., Stein, M., Moyer, E., & Sriver, R. (2018). Estimating changes in temperature distributions in a large ensemble of climate simulations using quantile regression. *Journal of Climate*, 31, 8573–8588. doi:10.1175/JCLI-D-17-0782.1.
- He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H., He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water contents and nitrogen leaching to precipitation, management practices and soil hydraulic properties in semi-arid and humid regions of Canada using the DSSAT model. *Nutrient Cycling in Agroecosystems*, 106, 201–215.
- Heady, E. O. (1957). An Econometric Investigation of the Technology of Agricultural Production Functions. *Econometrica*, 25, 249–268.
- Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa State University Press.
- Holden, P., Edwards, N., PH, G., Fraedrich, K., Lunkeit, F., E, K., Labriet, M., Kanudia, A., & F, B. (2014). Plasim-entsem v1.0: A spatiotemporal emulator of future climate change for impacts assessment. *Geoscientific Model Development*, 7, 433–451. doi:10.5194/gmd-7-433-2014.
- Holzkämper, A., Calanca, P., & Fuhrer, J. (2012). Statistical crop models: Predicting the effects of temperature and precipitation changes. *Climate Research*, 51, 11–21.
- Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgliesh, N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson, M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T., Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., & Keating, B. A. (2014). APSIM Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling and Software*, 62, 327 – 350.
- Howden, S., & Crimp, S. (2005). Assessing dangerous climate change impacts on australia's wheat industry. *Modelling and Simulation Society of Australia and New Zealand*, (pp. 505–511).
- Iizumi, T., Nishimori, M., & Yokozawa, M. (2010). Diagnostics of climate model biases in summer temperature and warm-season insolation for the simulation of regional paddy rice yield in japan. *Journal of Applied Meteorology and Climatology*, 49, 574–591.
- Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Nitrogen Fertilizers. *Ambio*, 6, 146–151.
- Izaurrealde, R., Williams, J., McGill, W., Rosenberg, N., & Quiroga Jakas, M. (2006). Simulating soil C dynamics with EPIC: Model description and testing against long-term data. *Ecological Modelling*, 192, 362–384.
- Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. *Agriculture, Ecosystems & Environment*, 93, 73 – 85.
- Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijsman, A., & Ritchie, J. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18, 235 – 265.
- Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural Systems*, 155, 269 – 288.
- Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K., Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, 18, 267 – 288.
- Leeds, W. B., Moyer, E. J., & Stein, M. L. (2015). Simulation of future climate under changing temporal covariance structures. *Advances in Statistical Climatology, Meteorology and Oceanography*, 1, 1–14. URL: <https://www.adv-stat-clim-meteorol-oceanogr.net/1/1/2015/>. doi:10.5194/ascmo-1-1-2015.
- Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., & Smith, B. (2013). Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa. *Earth System Dynamics*, 4, 385–407.
- Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling wheat yield and crop water productivity with high resolution on a global scale. *Agricultural Systems*, 94, 478 – 493.
- Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a). Global investigation of impacts of PET methods on simulating crop-water

- relations for maize. *Agricultural and Forest Meteorology*, 221, 164 – 175. doi:10.1016/j.agrformet.2015.07.016
- Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs with yields from major crop cultivations. *Science of The Total Environment*, 572, 526 – 537.
- Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, 150, 1443 – 1452.
- Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2, 014002.
- MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447.
- Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrological modelling to study the impact of climate change on the water flows of mountain watersheds. *Journal of Hydrology*, 376, 362 – 377.
- Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., & Calzadilla, A. (2009). Global biomass production potentials exceed expected future demand without the need for cropland expansion. *Nature Communications*, 6.
- Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical weather responsiveness of crop yields: US evidence and implications for the agricultural impacts of climate change. *Environmental Research Letters*, 12.
- Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, 8.
- Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. (2017). Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, 10, 1403–1422.
- Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T. (1997). Effect of CO<sub>2</sub> enrichment on carbon and nitrogen interaction in wheat and soybean. *Soil Science and Plant Nutrition*, 43, 789–798.
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91, 1290 – 1300.
- Olin, S., Schurges, G., Lindeskog, M., Wårlind, D., Smith, B., Bodin, P., Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue C:N to changes in atmospheric CO<sub>2</sub> and N management in the main wheat regions of western europe. *Biogeosciences*, 12, 2489–2515. doi:10.5194/bg-12-2489-2015.
- Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.
- Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., & Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator. *Geoscientific Model Development*, 8, 1139–1155.
- Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop yields and their variability at different levels of global warming. *Earth System Dynamics*, 9, 479–496.
- Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff, S., & Gerten, D. (2015). Emulating global climate change impacts on crop yields. *Statistical Modelling*, 15, 499–525.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas-sos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet, J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M., Semenov, M., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. *Climate Research*, 65, 87–105.
- Poppick, A., McInerney, D. J., Moyer, E. J., & Stein, M. L. (2016). Temperatures in transient climates: Improved methods for simulations with evolving temporal covariances. *Ann. Appl. Stat.*, 10, 477–505. URL: <https://doi.org/10.1214/16-AOAS903>. doi:10.1214/16-AOAS903.
- Porter et al. (IPCC) (2014). Food security and food production systems. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In C. F. et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK: Cambridge University Press.
- Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of monthly growing areas of 26 irrigated crops.
- Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly Irrigated and Rainfed crop Areas around the Year 2000: A New High-

- 1059 Resolution Data Set for Agricultural and Hydrological Modeling. *Global Biogeochemical Cycles*, 24, GB1011.
- 1060 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.
- 1061 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.
- 1062 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.
- 1063 Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus A: Dynamical Meteorology and Oceanography*, 58, 461–472.
- 1064 Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety*, 94, 1735 – 1763.
- 1065 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.
- 1066 Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, 48.
- 1067 Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, 12.
- 1068 Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., & Martin, J. F. (2014). Modeling sugarcane yield with a process-based model from site to continental scale: Uncertainties arising from model structure and parameter values. *Geoscientific Model Development*, 7, 1225–1245.
- 1069 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences*, 111, 3228–3232.
- 1070 White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Methodologies for simulating impacts of climate change on crop production. *Field Crops Research*, 124, 357 – 368.
- 1071 Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T., Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance against site observations of irrigated maize from Mead, Nebraska. *Geoscientific Model Development*, 10, 1291–1320.
- 1072 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield variability in JULES-crop and implications for forcing with seasonal weather forecasts. *Geoscientific Model Development*, 8, 3987–3997.
- 1073 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Landbouwkundige Onderzoeken* : 64.6, .
- 1074 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.
- 1075 Williams, K., & Oijen, M. (2002). Modelling the dependence of european potato yields on changes in climate and co2. *Agricultural and Forest Meteorology*, 112, 217 – 231.
- 1076 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.
- 1077 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., Peng, S., Peuelas, J., Ruane, A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci.*, 114, 9326–9331.
- 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101