

# 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>EAWAG, 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 nitrogen inputs and constitutes a data-rich basis of projected yield changes across 12 models and five crops (maize, soy, rice, spring wheat, and winter wheat) using global gridded simulations. In this paper we present the simulation output database from Phase II of the GGCMI effort, a targeted experiment aimed at understanding the sensitivity to and interaction between multiple climate variables (as well as management) on yields, and illustrate some initial summary results from the model intercomparison project. We also present the construction of a simple “emulator or statistical representation of the simulated 30-year mean climatological output in each location for each crop and model. The emulator captures the response of the process-based models in a lightweight, computationally tractable form that facilitates model comparison as well as potential applications in subsequent modeling efforts such as integrated assessment.

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

## 1. Introduction

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

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

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

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

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

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

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

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

75 The overall goal of this study is a better understanding of  
 76 global crop model response to the major drivers in a climate  
 77 change context. Most previous climate-change-focused global  
 78 crop modeling studies have simulated model response to rep-  
 79 resentative concentration pathways (RCPs). RCPs are likely to  
 80 have strong covariance between precipitation, temperature and  
 81 CO<sub>2</sub> that may be hard to decompose statistically. The differ-  
 82 ences in year-to-year memory in the models and complexity of  
 83 the changes in year-to-year distributions in weather under RCP  
 84 scenarios in climate models are complications we seek to con-  
 85 trol for with this study. We propose to test the response to major  
 86 drivers and their interaction by isolating individual input drivers  
 87 through simulations on first-moment shifts applied to the histor-  
 88 ical climatology instead of RCP simulations. As emulators are  
 89 a fundamentally a distillation of the process-based model down  
 90 to its major drivers, the same applied to their development.

91 Statistical emulation of crop simulations has been used to  
 92 combine advantageous features of both statistical and process-  
 93 based models. The statistical representation of complicated nu-  
 94 merical simulation (e.g. O'Hagan, 2006, Conti et al., 2009), in  
 95 which simulation output acts as the training data for a statisti-<sub>101</sub>  
 96 cal model, has been of increasing interest with the growth of<sub>102</sub>  
 97 simulation complexity and volume of output. Such emulators<sub>103</sub>  
 98 or "surrogate models" have been used in a variety of fields in-<sub>104</sub>  
 99 cluding hydrology (e.g. Razavi et al., 2012), engineering (e.g.<sub>105</sub>  
 100 Storlie et al., 2009), environmental sciences (e.g. Ratto et al.,<sub>106</sub>

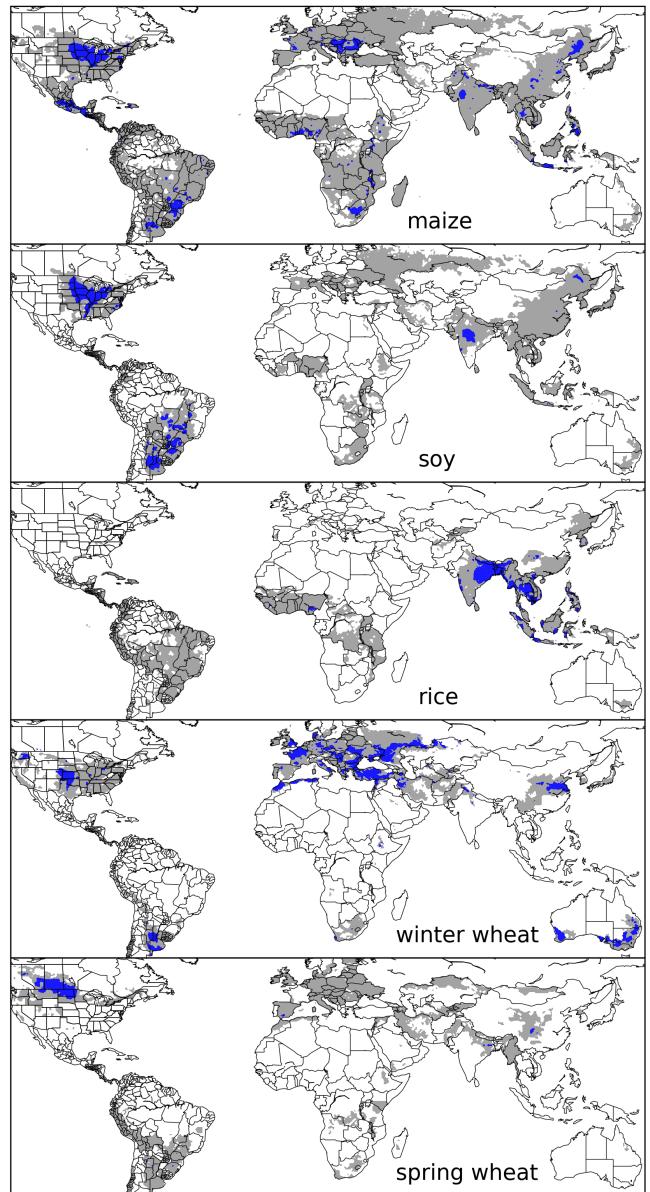


Figure 1: Presently cultivated area for rain-fed crops in the real world. Blue contour area indicates grid cells with more than 20,000 hectares (approx. 10% of the land in a grid cell near the equator for example) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated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

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

In the past decade, many studies have developed emulators of crop yields from process-based models. Early studies proposing or describing potential emulators include Howden & Crimp (2005), Räisänen & Ruokolainen (2006) and Lobell & Burke (2010). In an early application, Ferrise et al. (2011) used a Artificial Neural Net trained on simulation outputs to predict wheat yields in the Mediterranean. Studies developing single-model emulators include Holzkämper et al. (2012) for the CropSyst model, Ruane et al. (2013) for the CERES wheat model, Oyebamiji et al. (2015) for the LPJmL model (for multiple crops, using multiple scenarios as a training set). In recent years, emulators have begun to be used in the context of multi-model inter-comparisons, with Blanc & Sultan (2015), Blanc (2017), Ostberg et al. (2018) and Mistry et al. (2017) using them to analyze the five crop models of the Inter-Sectoral Impacts Model Intercomparison Project (ISIMIP) (Warszawski et al., 2014) (for maize, soy, wheat, and rice). Approaches differ: Blanc & Sultan (2015) and Blanc (2017) used local weather variables (and CO<sub>2</sub> values) and yields but emulate across soil types using historical simulations and a future climate scenario (RCP8.5 over multiple climate models); Ostberg et al. (2018) used global mean temperature change (and CO<sub>2</sub>) as regressors but pattern-scale to emulate local yields using multiple climate scenarios; Mistry et al. (2017) used local weather and yields and a historical simulation and compare with data. Other studies have used the development of an emulators (or response surface) to analyze non-RCP crop model simulations that sampled a suite of climate (and management) perturbations (the focus of this study):

Makowski et al. (2015) for temperature, CO<sub>2</sub>, and nitrogen, Pirttioja et al. (2015) and Snyder et al. (2018) for temperature, water, and CO<sub>2</sub>, and (Fronzek et al., 2018) for temperature and water, with all studies simulating selected sites for a limited number of crops.

In general, the use of limited input parameter space or restricted geographic scope impedes ability to build future projections or to understand interaction effects in global process-based crop models. The Global Gridded Crop Model Intercomparison (GGCMI) Phase II experiment seeks to provide a comprehensive global dataset to allow systematically exploring how process-based crop models for the major crop respond to the main climate and management drivers and their interactions. The experiment involves running a suite of process-based crop models across historical conditions perturbed by a set of defined input parameters, and was conducted as part of the Agricultural Model Intercomparison and Improvement Project (AgMIP) (Rosenzweig et al., 2013, 2014), an international effort conducted under a framework similar to the Climate Model Intercomparison Project (CMIP) (Taylor et al., 2012, Eyring et al., 2016). The GGCMI protocol builds on the AgMIP Coordinated Climate-Crop Modeling Project (C3MP) (Ruane et al., 2014, McDermid et al., 2015) and will contribute to the AgMIP Coordinated Global and Regional Assessments (CGRA) (Ruane et al., 2018, Rosenzweig et al., 2018).

GGCMI Phase II is designed to allow addressing goals such as understanding where highest-yield regions may shift under climate change; exploring future adaptive management

<b>Input variable</b>	<b>Abbr.</b>	<b>Tested range</b>	<b>Unit</b>
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 ( $W_{int}$ ) simulations.

strategies; understanding how interacting parameters affect<sup>195</sup>  
crop yield; quantifying uncertainties across models and major<sup>196</sup>  
drivers; and testing strategies for producing lightweight emu-<sup>197</sup>  
lators of process-based models. In this paper, we describe the<sup>198</sup>  
GGCMI Phase II experiments, summarize output and present<sup>199</sup>  
initial results, demonstrate that it is tractable to emulation, and<sup>200</sup>  
present a simple climatological emulator as a potential tool for<sup>201</sup>  
impacts assessments.<sup>202</sup>

## 2. Materials and Methods

### 2.1. GGCMI Phase II: experiment design

GGCMI Phase II is the continuation of a multi-model com-<sup>206</sup>  
parison exercise begun in 2014. The initial Phase I compared<sup>207</sup>  
harmonized yields of 21 models for 19 crops over a historical<sup>208</sup>  
(1980-2010) scenario with a primary goal of model evaluation<sup>209</sup>  
(Elliott et al., 2015, Müller et al., 2017). Phase II compares sim-<sup>210</sup>  
ulations of 12 models for 5 crops (maize, rice, soybean, spring<sup>211</sup>  
wheat, and winter wheat) over hundreds of scenarios in which<sup>212</sup>  
individual climate or management inputs are adjusted from<sup>213</sup>  
their historical values. The reduced set of crops includes the<sup>214</sup>  
three major global cereals and the major legume and accounts<sup>215</sup>  
for over 50% of human calories (in 2016, nearly 3.5 billion tons<sup>216</sup>  
or 32% of total global crop production by weight (Food and<sup>217</sup>  
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 beneficial level. Temperature perturbations are applied as absolute offsets from the daily mean, minimum, and maximum temperature time series for each grid cell used as inputs. Precipitation perturbations are applied as fractional changes at the grid cell level, and carbon dioxide and nitrogen levels are specified as discrete values applied uniformly over all grid cells. Note that CO<sub>2</sub> changes are applied independently of changes in climate variables, so that higher CO<sub>2</sub> is not associated with higher temperatures. An additional, identical set of scenarios (at the same C, T, W, and N levels) simulate adaptive agronomy under climate change by varying the growing season for crop production. (These adaptation simulations are not shown

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>GEPIG</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 or analyzed here.) The resulting GGCMI data set captures a<sub>247</sub> such as atmospheric deposition, are considered, but some mod-  
230 distribution of crop responses over the potential space of future<sub>248</sub> els have individual assumptions on soil organic matter that may  
231 climate conditions. <sub>249</sub> release additional nitrogen through mineralization. See Rosen-  
232 The 12 models included in GGCMI Phase II are all mecha-<sub>251</sub> zweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017)  
233 nistic process-based crop models that are widely used in im-  
234 pacts assessments (Table 2). Although some of the models<sub>252</sub> for further details on models and underlying assumptions.  
235 shares a common base (e.g. LPJmL and LPJ-GUESS and the<sub>253</sub>  
236 EPIC models), they have developed independently from this<sub>254</sub>  
237 shared base, for more details on the genealogy of the mod-<sub>255</sub>  
238 els see Figure S1 in Rosenzweig et al. (2014). Differences in<sub>256</sub>  
239 model structure does mean that several key factors are not stan-<sub>257</sub>  
240 dardized across the experiment, including secondary soil nutri-<sub>258</sub>  
241 ents, carry over effects across growing years including residue<sub>259</sub>  
242 management and soil moisture, and extent of simulated area for<sub>260</sub>  
243 different crops. Growing seasons are identical across models,<sub>261</sub>  
244 but vary by crop and by location on the globe. All stresses<sub>262</sub>  
245 except factors related to nitrogen, temperature, and water (e.g.,<sub>263</sub>  
246 Alkalinity, salinity) are disabled. No additional nitrogen inputs,<sub>264</sub>

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

265 The participating modeling groups provide simulations at<sup>299</sup>  
266 any of four initially specified levels of participation, so the num-<sup>300</sup>  
267 ber of simulations varies by model, with some sampling only a<sup>301</sup>  
268 part of the experiment variable space. Most modeling groups<sup>302</sup>  
269 simulate all five crops in the protocol, but some omitted one<sup>303</sup>  
270 or more. Table 2 provides details of coverage for each model.<sup>304</sup>  
271 Note that the three models that provide less than 50 simulations<sup>305</sup>  
272 are excluded from the emulator analysis.

273 All models produce as output, crop yields (tons ha<sup>-1</sup> year<sup>-1</sup>)<sup>307</sup>  
274 for each 0.5 degree grid cell. Because both yields and yield<sup>308</sup>  
275 changes vary substantially across models and across grid cells,<sup>309</sup>  
276 we primarily analyze relative change from a baseline. We take<sup>310</sup>  
277 as the baseline the scenario with historical climatology (i.e. T<sub>311</sub>  
278 and P changes of 0). C of 360 ppm, and applied N at 200 kg<sup>312</sup>  
279 ha<sup>-1</sup>. We show absolute yields in some cases to illustrate geo-<sup>313</sup>  
280 graphic differences in yields for a single model.

## 281 2.2. Simulation model validation approach

282 Simulation model validation for GGCMI Phase II continues<sup>316</sup>  
283 the validation efforts presented in Müller et al. (2017) for the<sup>317</sup>  
284 first phase of GGCMI. In the case presented here however, the<sup>318</sup>  
285 models are not run on the best approximation of management<sup>319</sup>  
286 levels (namely nitrogen application level) by region as with<sup>320</sup>  
287 Phase I. As the goals of this phase of the project are focused<sup>321</sup>  
288 on understanding the sensitivity in *change* in yield to changes<sup>322</sup>  
289 in input drivers –and not to simulate historical yields as accu-<sup>323</sup>  
290 rately as possible– no direct comparison to historical yield data<sup>324</sup>  
291 can be made. Additionally, even when provided with an appro-<sup>325</sup>  
292 priate local nitrogen level, models simulated *potential* yields<sup>326</sup>  
293 that do not include reductions from pests, weeds, or diseases.<sup>327</sup>  
294 Potential yields represent an ideal case that is not realized in<sup>328</sup>  
295 many less industrialized areas, and some models are not cali-<sup>329</sup>  
296 brated as they were in Phase I of the project.

297 We exactly replicate a portion of the validation process pre-<sup>331</sup>  
298 viously outlined in Müller et al. (2017), namely we evaluate<sup>332</sup>

the models based on the response to year-to-year temperature and precipitation variability in the historical record. If the models can (somewhat) faithfully represent the the historical variability in yields (which, once detrended to account for changing management levels must be driven largely by differences in weather), then the models may provide some utility in understanding the impact on mean climatological shifts in temperature and precipitation. Specifically, we calculate a Pearson correlation coefficient between the detrended time series of simulations and FAO data (Food and Agriculture Organization of the United Nations, 2018) for the period 1981-2009 at the national scale. The FAO data is detrended because much of the change in yield over this time period is due to intensification and changes in management (Ray et al., 2012). Validating the response to CO<sub>2</sub> and nitrogen applications is more difficult because real world data is not available outside of small greenhouse and field level trials.

## 2.3. Climatological-mean yield emulator design

We construct our emulator at the 30-year climatological mean level. Blanc & Sultan (2015) and Blanc (2017) have shown that a emulator of a global process-based crop model can be successfully developed at the yearly scale. Our decision to construct a climatological-mean yield emulator is driven by the target application for this analysis tool. Many impact modelers are not focused on the changes in the year-to-year variability in yields, but instead on the broad mean changes over the multi-decadal timescale. Emulation 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 per-

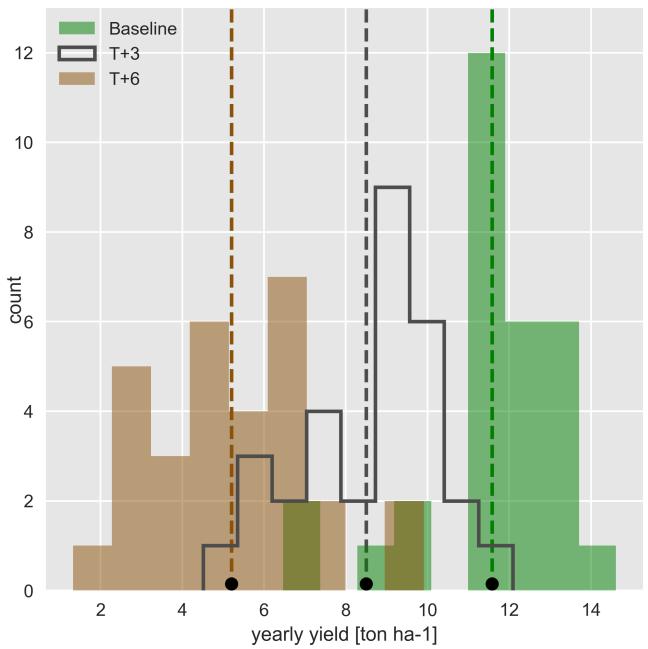


Figure 2: Example showing both climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for a<sub>363</sub> grid cell in northern Iowa (a representative high-yield region) from the pDSSAT model, for the baseline climatology (1981-2010) and 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<sub>365</sub> yield.

turbations, which are more complex. Emulating inter-annual<sub>367</sub> yield variations would likely require considering statistical de-<sub>368</sub> tails of the historical climate time series, including changes in<sub>369</sub> marginal distribution and temporal dependencies. (Future work<sub>370</sub> should explore this). The climatological emulation indirectly<sub>371</sub> includes any yield response to geographically distributed fac-<sub>372</sub> tors such as soil type, insolation, and the baseline climate itself,<sub>373</sub> because we construct separate emulators for each grid cell. The<sub>374</sub> emulator parameter matrices are portable and the yield computa-<sub>375</sub> tions are cheap even at the half-degree grid cell resolution, so<sub>376</sub> we do not aggregate in space at this time.<sub>377</sub>

Blanc & Sultan (2015) and Blanc (2017) have shown that<sub>378</sub> a fractional polynomial specification is more effective than a stan-<sub>379</sub>dard polynomial for representing simulations at the yearly level<sub>380</sub> across different soil types geographically (not at the grid cell<sub>381</sub> level). We do not test this specification here, and instead use as<sub>382</sub> a starting point a standard third-order polynomial to represent<sub>383</sub>

the climatological-mean response at the grid cell level as it is<sub>350</sub> the simplest effective specification. We regress climatological-<sub>351</sub> mean yields against a third-order polynomial in C, T, W, and N<sub>352</sub> with interaction terms. The higher-order terms are necessary to<sub>353</sub> capture any nonlinear responses, which are well-documented<sub>354</sub> in observations for temperature and water perturbations (e.g.<sub>355</sub> Schlenker & Roberts (2009) for T and He et al. (2016) for W).<sub>356</sub> We include interaction terms (both linear and higher-order) be-<sub>357</sub> cause past studies have shown them to be significant effects.<sub>358</sub> For example, Lobell & Field (2007) and Tebaldi & Lobell<sub>359</sub> (2008) showed that in real-world yields, the joint distribution<sub>360</sub> in T and W is needed to explain observed yield variance (C<sub>361</sub> and N are fixed in these data). Other observation-based stud-<sub>362</sub> ies have shown the importance of the interaction between water<sub>363</sub> and nitrogen (e.g. Aulakh & Malhi, 2005), and between nitro-<sub>364</sub> gen and carbon dioxide (Osaki et al., 1992, Nakamura et al.,<sub>365</sub> 1997). We do not focus on comparing different model speci-<sub>366</sub> fications in this study, and instead stick to a relatively simple<sub>367</sub> parameterized specification that allows for some, albeit limited,<sub>368</sub> coefficient interpretation.

The limited GGCMI variable sample space means that use<sub>369</sub> of the full polynomial expression described above, which has<sub>370</sub> 34 terms for the rain-fed case (12 for irrigated), can be prob-<sub>371</sub> lematic, and can lead to over-fitting and unstable parameter es-<sub>372</sub> timations. We therefore reduce the number of terms through a<sub>373</sub> feature selection cross-validation process in which terms in the<sub>374</sub> polynomial are tested for importance. In this procedure higher-<sub>375</sub> order and interaction terms are added successively to the model;<sub>376</sub> we then follow the reduction of the the aggregate mean squared<sub>377</sub> error with increasing terms and eliminate those terms that do<sub>378</sub> not contribute significant reductions. See supplemental docu-<sub>379</sub> ments for more details. We select terms by applying the feature<sub>380</sub> selection process to the three models that provided the com-<sub>381</sub> plete set of 672 rain-fed simulations (pDSSAT, EPIC-TAMU,<sub>382</sub>

and LPJmL); the resulting choice of terms is then applied for all emulators.

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

$$Y = K_1 \quad (1)$$

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

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

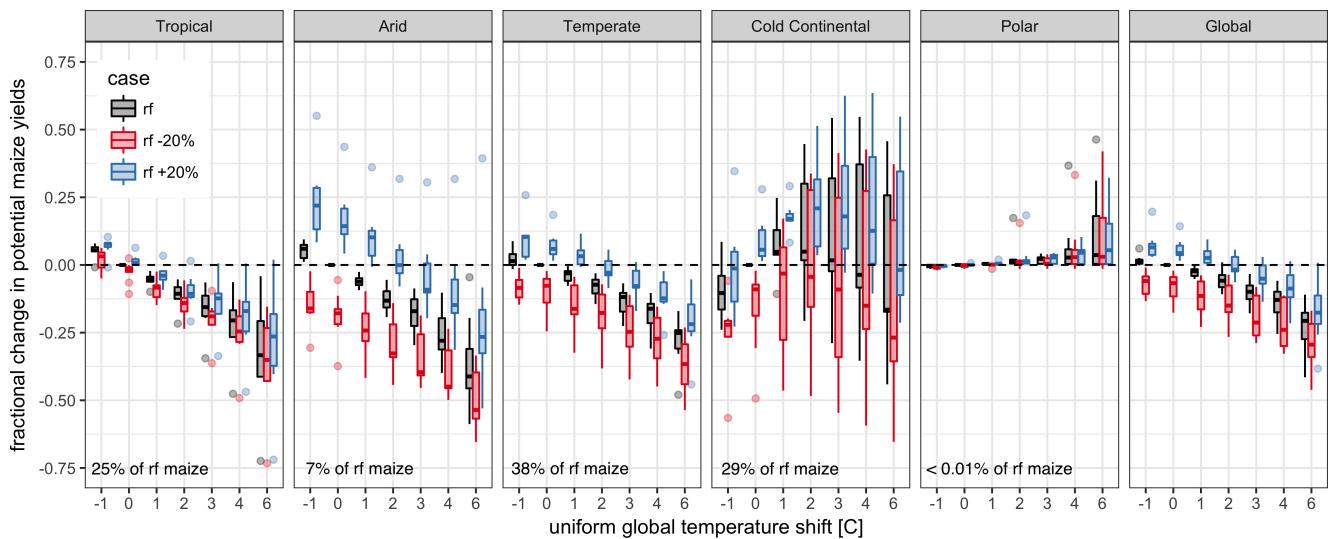


Figure 3: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation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

406 experiment, the most problematic fits are those for models that<sub>435</sub>  
 407 provided a limited number of cases or for low-yield geographic<sub>436</sub>  
 408 regions where some modeling groups did not run all scenarios.<sub>437</sub>  
 409 Because we do not attempt to emulate models that provided<sub>438</sub>  
 410 less than 50 simulations, the lowest number of simulations em-<sub>439</sub>  
 411 ulated across the full parameter space is 130 (for the PEPIC<sub>440</sub>  
 412 model). We use the implementation of the Bayesian Ridge esti-<sub>441</sub>  
 413 mator from the scikit-learn package in Python (Pedregosa et al.,<sub>442</sub>  
 414 2011).

415 The resulting parameter matrices for all crop model emula-<sub>444</sub>  
 416 tors are available on request, as are the raw simulation data and<sub>445</sub>  
 417 a Python application to emulate yields. The yield output for a<sub>446</sub>  
 418 single GGCMI model that simulates all scenarios and all five<sub>447</sub>  
 419 crops is  $\sim$ 12.5 GB; the emulator is  $\sim$ 100 MB, a reduction by<sub>448</sub>  
 420 over two orders of magnitude.

#### 421 2.4. Emulator evaluation

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

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

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

432 Here  $F_{scn.}$  is the fractional change in a model’s mean emu-<sub>445</sub>  
 433 lated or simulated yield from a defined baseline, in some sce-<sub>446</sub>  
 434 nario (scn.) in C, T, W, and N space;  $Y_{scn.}$  and  $Y_{baseline}$  are the<sub>447</sub>

absolute emulated or simulated mean yields. The normalized  
 435 error  $e$  is the difference between the emulated fractional change  
 436 in yield and that actually simulated, normalized by  $\sigma_{sim.}$ , the  
 437 standard deviation in simulated fractional yields  $F_{sim, scn.}$  across  
 438 all models. The emulator is fit across all available simulation  
 439 outputs, and then the error is calculated across the simulation  
 440 scenarios provided by all nine models (Figure 8 and Figures  
 441 S12 and Figures S13 in supplemental documents). Note that  
 442 the normalized error  $e$  for a model depends not only on the fi-  
 443 delity of its emulator in reproducing a given simulation but on  
 444 the particular suite of models considered in the intercomparison  
 445 exercise. The rationale for this choice is to relate the fidelity of  
 446 the emulation to an estimate of true uncertainty, which we take  
 447 as the multi-model spread.

## 449 3. Results

### 450 3.1. Simulation results

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

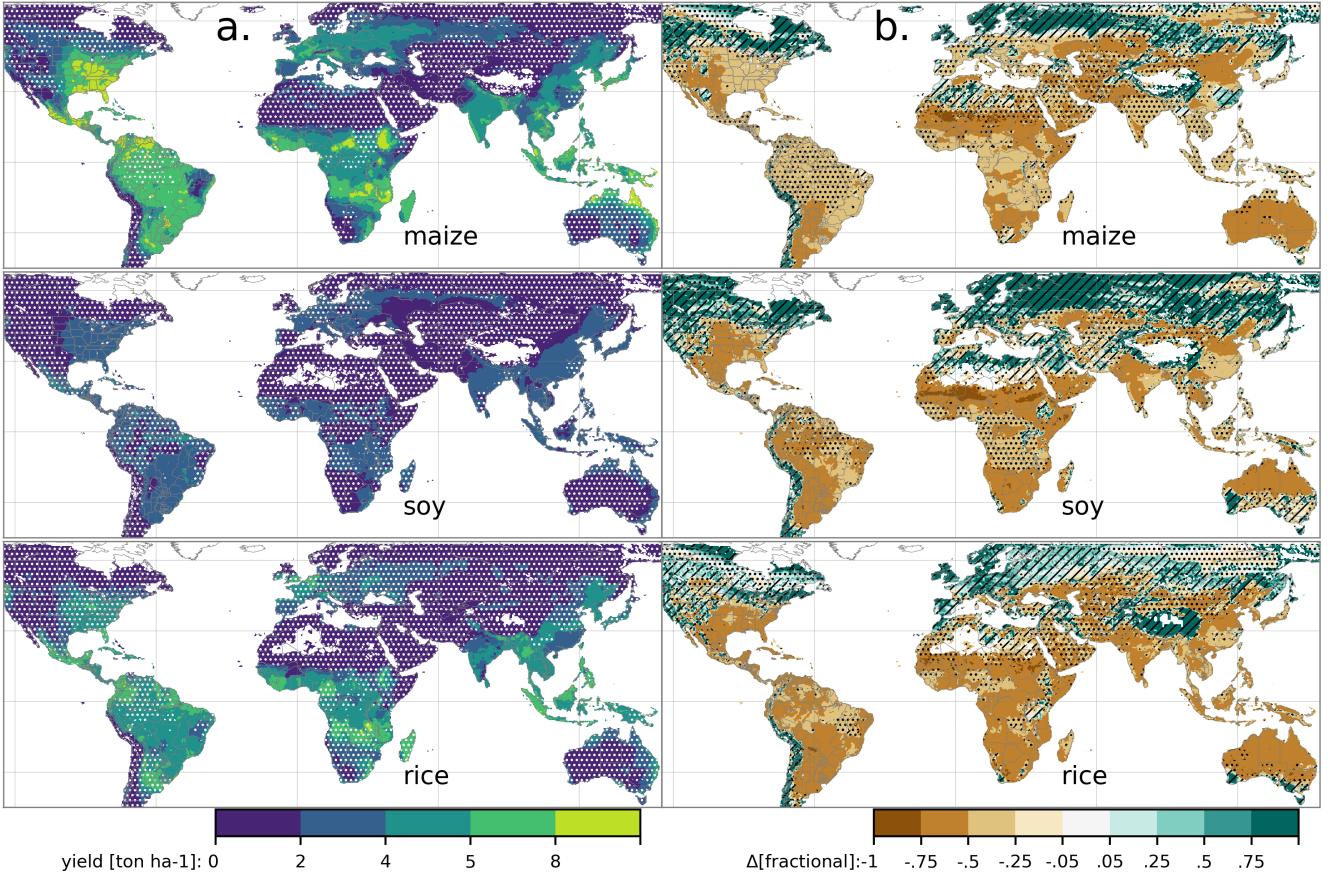


Figure 4: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure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

468 warming, with robust yield losses in warmer locations and high<sub>481</sub>  
 469 inter-model variance in the cold continental regions (Figures<sub>482</sub>  
 470 S7).

471 The effects of rainfall changes on maize yields are also as ex-<sub>484</sub>  
 472 pected and are consistent across models. Increased rainfall mit-<sub>485</sub>  
 473 igates the negative effect of higher temperatures, most strongly<sub>486</sub>  
 474 in arid regions. Decreased rainfall amplifies yield losses and<sub>487</sub>  
 475 also increases inter-model variance more strongly, suggesting<sub>488</sub>  
 476 that models have difficulty representing crop response to water<sub>489</sub>  
 477 stress. We show only rain-fed maize here; see Figure S5 for the<sub>490</sub>  
 478 irrigated case. As expected, irrigated crops are more resilient to<sub>491</sub>  
 479 temperature increases in all regions, especially so where water<sub>492</sub>  
 480 is limiting.

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

483 Figure 4 shows baseline and changes in the T+4 scenario for  
 484 rain-fed maize, soy, and rice in the multi-model ensemble mean,  
 485 with locations of model agreement marked. Absolute yield po-  
 486 tentials are have strong spatial variation, with much of the  
 487 Earth's surface area unsuitable for any given crop. In general,  
 488 models agree most on yield response in regions where yield  
 489 potentials are currently high and therefore where crops are cur-  
 490 rently grown. Models show robust decreases in yields at low  
 491 latitudes, and highly uncertain median increases at most high  
 492 latitudes. For wheat crops see Figure S11; wheat projections  
 493 are both more uncertain and show fewer areas of increased yield

494 in the inter-model mean.

### 495 3.2. Simulation model validation results

496 Figure 5 shows the Pearson time series correlation between  
497 the simulation model yield and FAO yield data. Figure 5 can be  
498 compared to Figures 1,2,3,4 and 6 in Müller et al. (2017). The  
499 results are mixed, with many regions for rice and wheat be-  
500 ing difficult to model. No single model is dominant, with each  
501 model providing near best-in-class performance in at least one  
502 location-crop combination. The presence of very few vertical  
503 dark green color bars clearly illustrates the power of a multi-  
504 model intercomparison project like the one presented here. The  
505 ensemble mean does not beat the best model in each case, but  
506 shows positive correlation in over 75% of the cases presented  
507 here. The EPIC-TAMU model performs best for soy, CARIAB,  
508 EPIC-TAMU, and PEPIC perform best for maize, PROMET  
509 performs best for wheat, and the EPIC family of models per-  
510 form best for rice. Reductions in skill over the performance  
511 illustrated in Müller et al. (2017) can be attributed to the nitro-  
512 gen levels or lack of calibration in some models.

513 \*\*\* or harmonization \*\*\* Christoph

514 Soy is qualitatively the easiest crop to represent (except in Argentina), which is likely due in part to the invariance of the response to nitrogen application (soy fixes atmospheric nitrogen very efficiently). Comparison to the FAO data is therefore easier than the other crops because the nitrogen application levels do not matter. US maize has the best performance across models, with nearly every model representing the historical variability to a reasonable extent. Especially good example years for US maize are 1983, 1988, and 2004 (top left panel of Figure 5), where every model gets the direction of the anomaly compared to surrounding years correct. 1983 and 1988 are famously bad years for US maize along with 2012 (not shown). US maize is possibly both the most uniformly industrialized (in terms of

527 management practices) crop and the one with the best data col-  
528 lection in the historical period of all the cases presented here.

529 The FAO data is at least one level of abstraction from ground  
530 truth in many cases, especially in developing countries. The  
531 failure of models to represent the year-to-year variability in rice  
532 in some countries in southeast Asia is likely partly due to model  
533 failure and partly due to lack of data. It is possible to speculate  
534 that the difference in performance between Pakistan (no suc-  
535 cessful models) and India (many successful models) for rice  
536 may reside at least in part in the FAO data and not the mod-  
537 els themselves. The same might apply to Bangladesh and In-  
538 dia for rice. Partitioning of these contributions is impossible at  
539 this stage. Additionally, there is less year-to-year variability in  
540 rice yields (partially due to the fraction of irrigated cultivation).  
541 Since the Pearson r metric is scale invariant, it will tend to score  
542 the rice models more poorly than maize and soy. An example  
543 of very poor performance can be seen with the pDSSAT model  
544 for rice in India (top right panel of Figure 5).

### 545 3.3. Emulator performance

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

559 Each panel in Figure 6 shows model yield output from sce-  
560 narios varying only along a single dimension ( $\text{CO}_2$ , tempera-

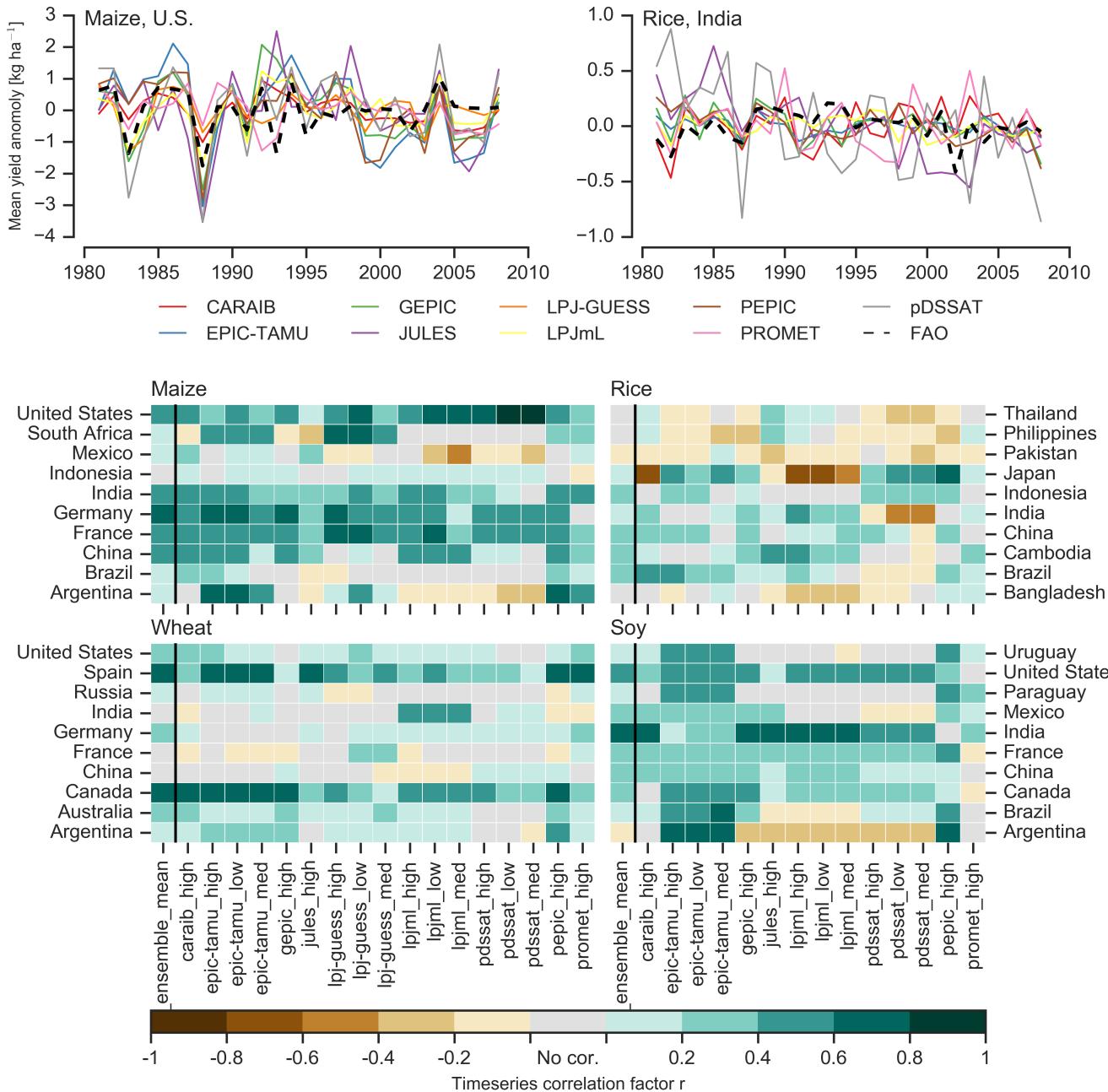


Figure 5: Time series correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen application case. The heatmaps illustrate the Pearson r correlation coefficient between the detrended simulation mean yield at the country level compared to the detrended FAO yield data for the top producing countries for each crop with continuous FAO data over the 1980-2010 period. Models that provided different nitrogen application levels are shown with low, med, and high label (models that did not simulate different nitrogen levels are analogous to a high nitrogen application level). The ensemble mean yield is also correlated with the FAO data (not the mean of the correlations). Wheat contains both spring wheat and winter wheat simulations.

ture, precipitation, or nitrogen addition), with other inputs held<sub>566</sub> perturbations.

fixed at baseline levels; in all cases yields evolve smoothly

across the space sampled. For reference we show the results<sub>567</sub>

of the full emulation fitted across the parameter space. The<sub>568</sub>

polynomial fit readily captures the climatological response to<sub>569</sub>

Crop yield responses generally follow similar functional

forms across models, though with a spread in magnitude. Fig-

ure 7 illustrates the inter-model diversity of yield responses

to the same perturbations, even for a single crop and location

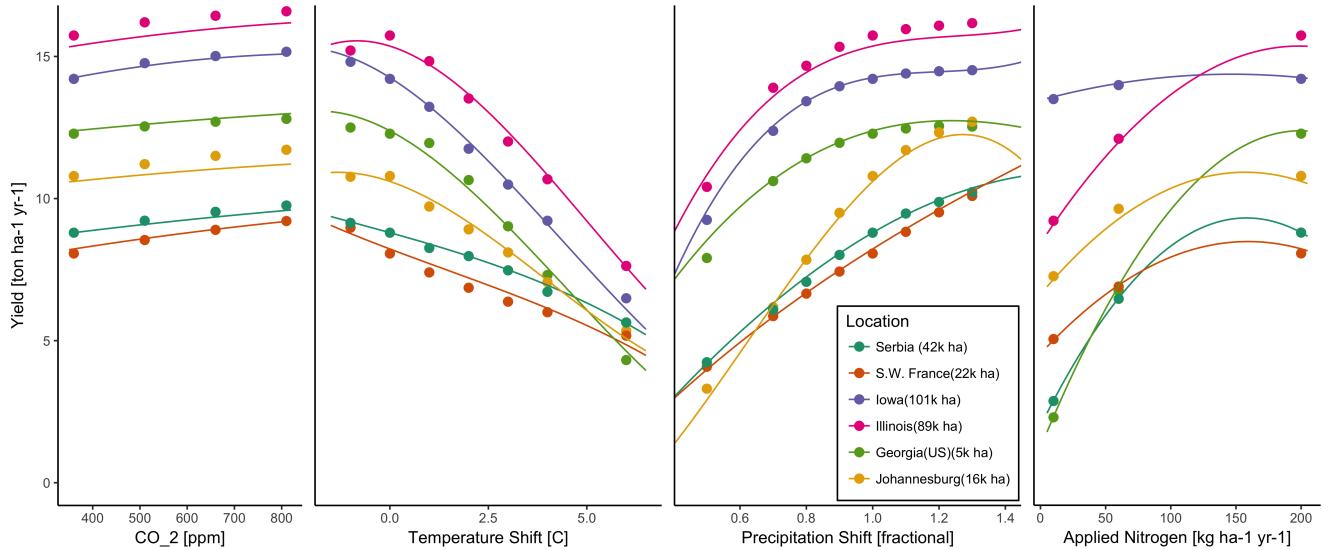


Figure 6: Example illustrating spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows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

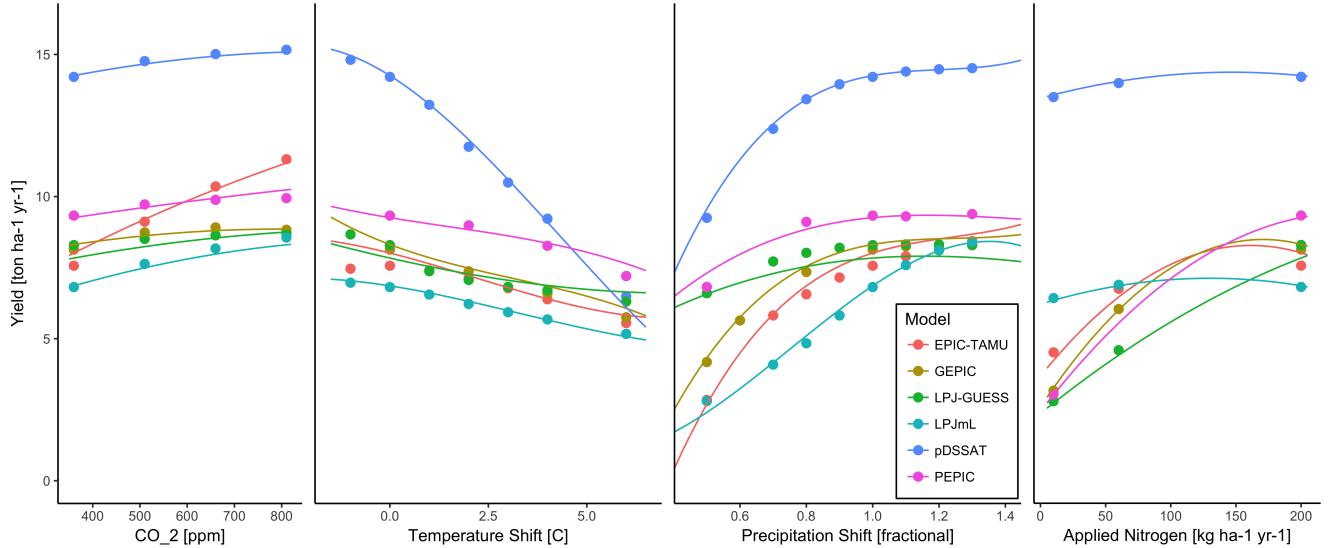


Figure 7: Example illustrating across-model variations in yield response. We show simulations and emulations from six models for rain-fed maize in the same Iowa location shown in Figure 6, with the same plot conventions. Models that did not simulate the nitrogen dimension are omitted for clarity. The inter-model standard deviation is larger for the perturbations in 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.

571 (rain-fed maize in northern Iowa, the same location shown in 576  
 572 the Figure 6). The differences make it important to construct 577  
 573 emulators separately for each individual model, and the fidelity 578  
 574 of emulation can also differ across models. This figure illus- 579  
 575 trates a common phenomenon, that models differ more in re- 580

576 response to perturbations in CO<sub>2</sub> and nitrogen perturbations than  
 577 to those in temperature or precipitation. (Compare also Figures  
 578 3 and S18.) For this location and crop, CO<sub>2</sub> fertilization effects  
 579 can range from ~5–50%, and nitrogen responses from nearly  
 580 flat to a 60% drop in the lowest-application simulation.

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

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

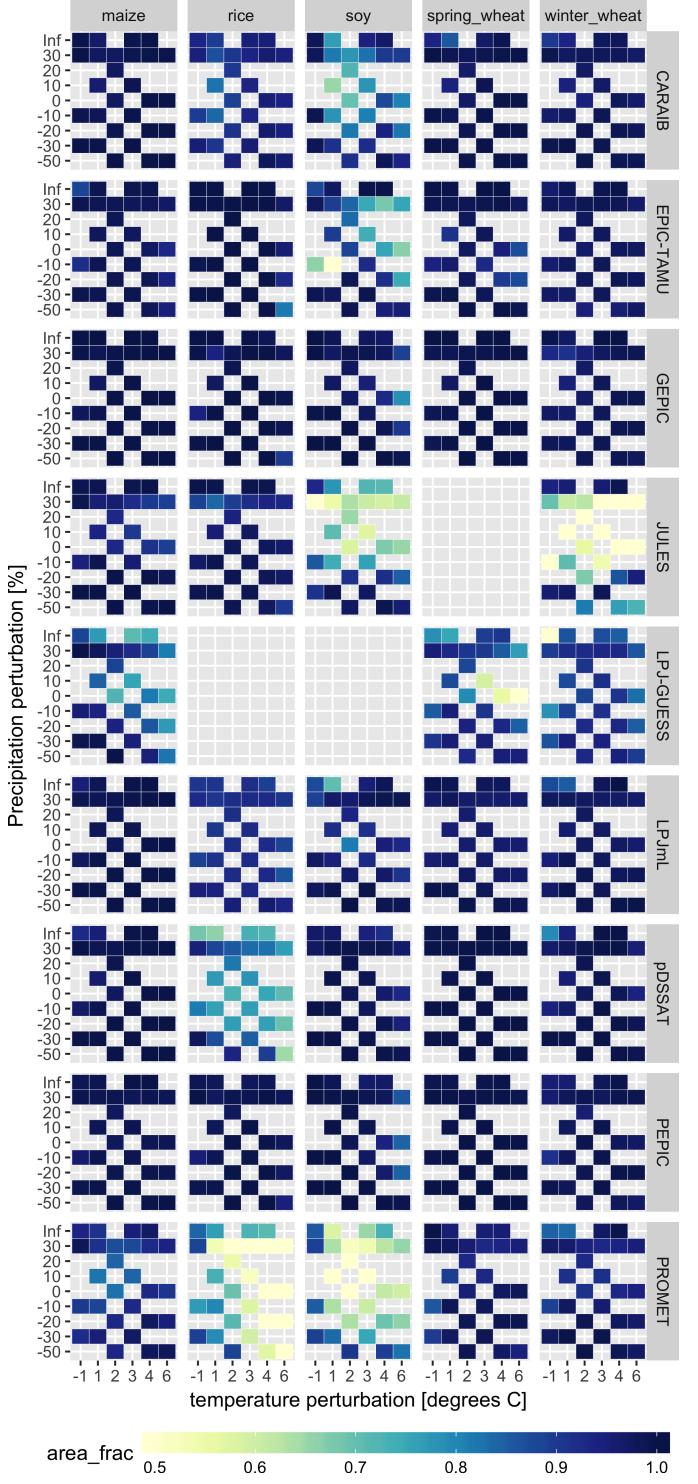


Figure 8: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with  $\text{CO}_2$  and nitrogen held fixed at baseline values. Large columns are crops and large rows models; squares within are T,P scenario pairs. Colors denote the fraction of currently cultivated hectares with  $e < 1$ . Of the 756 scenarios with these  $\text{CO}_2$  and N values, we consider only those for which all 9 models submitted data. JULES did not simulate spring wheat and LPJ-GUESS did not simulate rice and soy. Emulator performance is generally satisfactory, with some exceptions. Emulator failures (significant areas of poor performance) occur for individual crop-model combinations, with performance generally degrading for hotter and wetter scenarios.

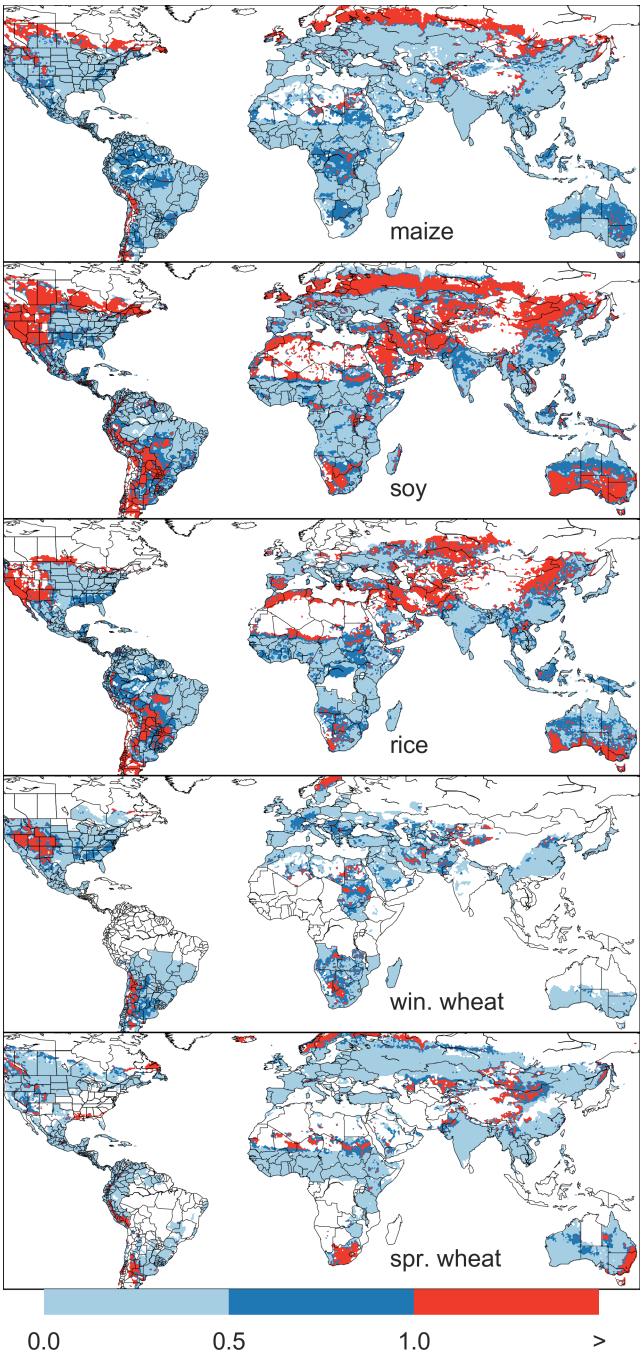


Figure 9: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error  $e$ , where  $e > 1$  means that emulator error exceeds the multi-model standard deviation. White areas are those where crops are not simulated by this model. Models differ in their areas omitted, meaning the number of samples used to calculate the multi-model standard deviation is not spatially consistent in all locations. Emulator performance is generally good relative to model spread in areas where crops are currently cultivated (compare to Figure 1) and in temperate zones in general; emulation issues occur primarily in marginal areas with low yield potentials. For CARAIB, emulation of soy is more problematic, as was also shown in Figure 8.

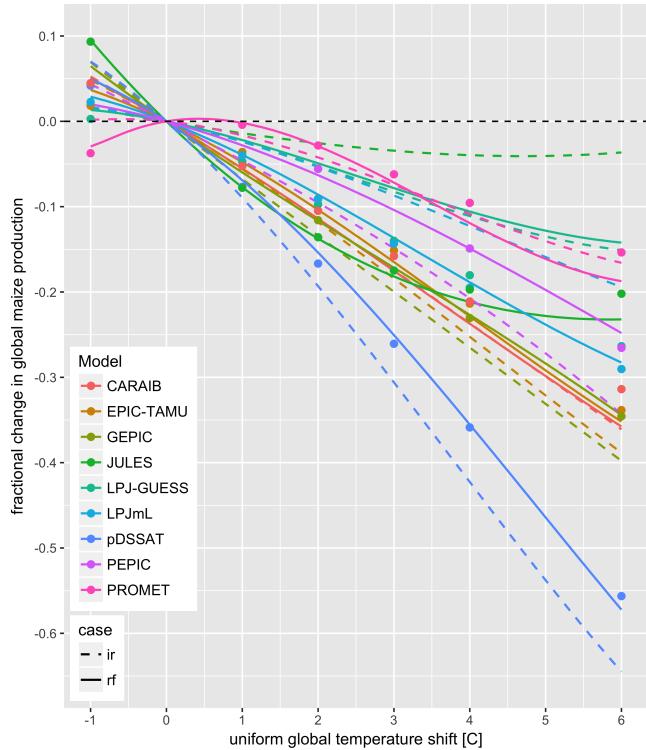
other than soy, but uncultivated regions show some problematic areas.

It should be noted that this assessment metric is relatively forgiving. First, each emulation is evaluated against the simulation actually used to train the emulator. Had we used a spline interpolation the error would necessarily be zero. Second, the performance metric scales emulator fidelity not by the magnitude of yield changes but by the inter-model spread in those changes. Where models differ more widely, the standard for emulators becomes less stringent. Because models disagree on the magnitude of CO<sub>2</sub> fertilization, this effect is readily seen when comparing assessments of emulator performance in simulations at baseline CO<sub>2</sub> (Figure 8) with those at higher CO<sub>2</sub> levels (Figure S13). Widening the inter-model spread leads to an apparent increase in emulator skill.

### 3.4. Emulator applications

Because the emulator or “surrogate model” transforms the discrete simulation sample space into a continuous response surface at any geographic scale, it can be used for a variety of applications. Emulators provide an easy way to compare a ensemble of climate or impacts projections. They also provide a means for generating continuous damage functions. As an example, we show a damage function constructed from 4D emulations for aggregated yield at the global scale, for maize on currently cultivated land, with simulated values shown for comparison. (Figure 10; see Figures S16- S19 in the supplemental material for other crops and dimensions.) The emulated values closely match simulations even at this aggregation level. Note that these functions are presented only as examples and do not represent true global projections, because they are developed from simulation data with a uniform temperature shift while increases in global mean temperature should manifest non-uniformly. The global coverage of the GGCMI simulations allows impacts modelers to apply arbitrary

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



649 Figure 10: Global emulated damages for maize on currently cultivated lands  
 650 for the GGCMI models emulated, for uniform temperature shifts with other  
 651 inputs held at baseline. (The damage function is created from aggregating up  
 652 emulated values at the grid cell level, not from a regression of global mean  
 653 yields.) Lines are emulations for rain-fed (solid) and irrigated (dashed) crops;  
 654 for comparison, dots are the simulated values for the rain-fed case. For most  
 655 models, irrigated crops show a sharper reduction than do rain-fed because of the  
 656 locations of cultivated areas: irrigated crops tend to be grown in warmer areas  
 657 where impacts are more severe for a given temperature shift. (The exceptions  
 658 are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures  
 659 S16- S19 in the supplemental material.

#### 660 4. Conclusions and discussion

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

669 from the simulations. Across the major crops, inter-model un-  
 670 certainty is greatest for wheat and least for soy. Across factors  
 671 impacting yields, inter-model-uncertainty is largest for  $\text{CO}_2$  fer-  
 672 tilization and nitrogen response effects. Across geographic re-  
 673 gions, inter-model uncertainty is largest in the high latitudes  
 674 where yields may increase, and model projections are most ro-  
 675 bust in low latitudes where yield impacts are largest.

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

685 One counterintuitive result is that irrigated maize shows  
 686 steeper yield reductions under warming than does rain-fed  
 687 maize when considered only over currently cultivated land. The  
 688 effect is the result of geographic differences in cultivated area.  
 689 In any given location, irrigation increases crop resiliency to  
 690 temperature increase, but irrigated maize is grown in warmer lo-  
 691 cations where the impacts of warming are more severe (Figures  
 692 S5-S6). The same behavior holds for rice and winter wheat,  
 693 but not for soy or spring wheat (Figures S8-S10). Irrigated  
 694 wheat and maize are also more sensitive to nitrogen fertiliza-  
 695 tion levels, presumably because growth in rain-fed crops is also  
 696 water-limited (Figure S19). (Soy as a nitrogen-fixer is relatively  
 697 insensitive to nitrogen, and rice is not generally grown in water-  
 698 limited conditions).

699 We show that emulation of the output of these complex re-  
 700 sponses is possible even with a relatively simple reduced-form  
 701 statistical model and a limited library of simulations. Emula-  
 702 tion therefore offers the opportunity of producing rapid assess-

695 ments of agricultural impacts for arbitrary climate scenarios in<sup>729</sup>  
696 a computationally non-intensive way. The resulting tool should<sup>730</sup>  
697 aid in impacts assessment, economic studies, and uncertainty<sup>731</sup>  
698 analyses. Emulator parameter values also provide a useful way<sup>732</sup>  
699 to compare sensitivities across models to different climate and<sup>733</sup>  
700 management inputs, and the terms in the polynomial fits offer<sup>734</sup>  
701 the possibility of physical interpretation of these dependencies<sup>735</sup>  
702 to some degree.

736 Because the GGCMI simulations apply uniform perturbations  
to historical climate inputs, they do not sample changes in  
higher order moments. The emulation therefore does not ad-  
dress the crop yield impacts of potential changes in climate  
variability. While some information could be extracted from  
consideration of year-over-year variability, more detailed sim-  
ulations and analysis are likely necessary to diagnose the im-  
pact of changes in variance and sub-growing-season tempo-  
ral effects. Additionally, the emulator is intended to provide  
the change in yield from a historical mean baseline value and  
should be used in conjunction with historical data (or data prod-  
ucts) or a historical mean emulator (not presented here).

703 We provide this simulation output dataset for further analysis<sup>737</sup>  
704 by the community as we have only scratched the surface with<sup>738</sup>  
705 this work. Each simulation run includes year to year variabil-<sup>739</sup>  
706 ity in yields under different climate and management regimes.<sup>740</sup>  
707 Some of the precipitation and temperature space has been lost<sup>741</sup>  
708 due to the aggregation in the time dimension for the emula-<sup>742</sup>  
709 tor presented here (i.e. the + 6 C simulation in the hottest year<sup>743</sup>  
710 of the historical period compared to the coldest historical year,<sup>744</sup>  
711 or precipitation perturbations in the driest historical year etc.<sup>745</sup>  
712 Development of a year-to-year emulator or an emulator at dif-<sup>746</sup>  
713 ferent spatial scales may provide useful for some IAM appli-<sup>747</sup>  
714 cations. More exhaustive analysis of different statistical model<sup>748</sup>  
715 specification for emulation will likely provide additional pre-<sup>749</sup>  
716 dictive skill over the specification provided here. The poten-<sup>750</sup>  
717 tially richest area for further analysis is the interactions be-  
718 tween input variable especially the Nitrogen and CO<sub>2</sub> interac-<sup>751</sup>  
719 tions with weather and with each other. More robust quantifica-<sup>752</sup>  
720 tion of the sensitivity to the input drivers (and there differences<sup>753</sup>  
721 across models), as well as quantification in differences in un-<sup>754</sup>  
722 certainty across input drivers. Adaptation via growing season<sup>755</sup>  
723 changes were also simulated and are available in the database,<sup>756</sup>  
724 though this dimension was not presented or analyzed here. The<sup>757</sup>  
725 output dataset contains many other variables other than yield to<sup>758</sup>  
726 aid in analysis including above ground biomass, LAI, and root<sup>759</sup>  
727 biomass (as many as 25 output variables for some models).<sup>760</sup>

728 The emulation approach presented here has some limitations.<sup>761</sup>

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

## 5. Acknowledgments

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

762 search (BMBF). C.F. was supported by the European Research<sup>799</sup>  
 763 Council Synergy grant #ERC-2013-SynG-610028 Imbalance-<sup>800</sup>  
 764 P. P.F. and K.W. were supported by the Newton Fund through  
 765 the Met Office Climate Science for Service Partnership Brazil<sup>801</sup>  
 766 (CSSP Brazil). A.S. was supported by the Office of Science<sup>802</sup>  
 767 of the U.S. Department of Energy as part of the Multi-sector<sup>803</sup>  
 768 Dynamics Research Program Area. Computing resources were<sup>804</sup>  
 769 provided by the University of Chicago Research Computing<sup>805</sup>  
 770 Center (RCC). S.O. acknowledges support from the Swedish<sup>806</sup>  
 771 strong research areas BECC and MERGE together with sup-<sup>807</sup>  
 772 port from LUCCI (Lund University Centre for studies of Car-<sup>808</sup>  
 773 bon Cycle and Climate Interactions).

## 774 6. References

- 775 Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013).  
 776 Implication of crop model calibration strategies for assessing regional im-<sup>818</sup>  
 777 pacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.  
 778 Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D.,  
 779 A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Al-<sup>819</sup>  
 780 derman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Bier-<sup>820</sup>  
 781 nath, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising tempera-<sup>821</sup>  
 782 tures reduce global wheat production. *Nature Climate Change*, 5, 143–147.  
 783 doi:10.1038/nclimate2470.
- 784 Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A.,  
 785 J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N.,  
 786 Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath,  
 787 C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simula-<sup>822</sup>  
 788 ting wheat yields under climate change. *Nature Climate Change*, 3, 827832.  
 789 doi:10.1038/nclimate1916.
- 790 Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other  
 791 Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Ef-<sup>833</sup>  
 792 ficiency, Carbon Sequestration, and Environmental Pollution. *Advances in*  
 793 *Agronomy*, 86, 341 – 409.
- 794 Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov,  
 795 N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat  
 796 production potentials and management flexibility under the representative  
 797 concentration pathways. *Global and Planetary Change*, 122, 107 – 121.
- 798 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields  
 841 from global gridded crop models. *Agricultural and Forest Meteorology*, 236,  
 145 – 161.
- 801 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded  
 802 crop models using statistical estimates. *Agricultural and Forest Meteorol-<sup>803</sup>  
 803 ogy*, 214-215, 134 – 147.
- 804 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S.  
 805 (2018). Implementing the Nitrogen cycle into the dynamic global vegeta-<sup>806</sup>  
 806 tion, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific*  
 807 *Model Development*, 11, 2789–2812.
- 808 Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L.,  
 809 & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections  
 810 Based on Precomputed GCM Runs. *Journal of Climate*, 27, 1829–1844.
- 811 Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.  
 812 (2014). A meta-analysis of crop yield under climate change and adaptation.  
 813 *Nature Climate Change*, 4, 287 – 291.
- 814 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process  
 815 emulation of dynamic computer codes. *Biometrika*, 96, 663–676.
- 816 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In  
 817 C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth,*  
 818 *Duke University, Durham, North Carolina* (pp. 115–118). Durham, North  
 819 Carolina.
- 820 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for  
 821 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).
- 822 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdoos,  
 823 M., & François, L. (2011). Responses of European forest ecosystems to  
 824 21st century climate: assessing changes in interannual variability and fire  
 825 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).
- 826 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,  
 827 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating  
 828 impact models and sectors (pSIMS). *Environmental Modelling and Soft-<sup>829</sup>  
 829 ware*, 62, 509–516.
- 830 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,  
 831 Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurrealde, R. C.,  
 832 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.  
 833 (2015). The Global Gridded Crop Model Intercomparison: data and mod-<sup>834</sup>  
 834 eling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8,  
 835 261–277.
- 836 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J.,  
 837 & Taylor, K. E. (2016). Overview of the coupled model intercomparison  
 838 project phase 6 (cmip6) experimental design and organization. *Geoscientific*  
 839 *Model Development*, 9, 1937–1958.
- 840 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-<sup>841</sup>  
 841 mate change impacts on durum wheat in the mediterranean region. *Natural*

- 842 *Hazards and Earth System Sciences*, 11, 1293–1302. 885 *Research*, 51, 11–21.

843 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Re-886  
844 gionalization of a large-scale crop growth model for sub-Saharan Africa:887  
845 Model setup, evaluation, and estimation of maize yields. *Agriculture,888  
846 Ecosystems & Environment*, 151, 21 – 33. 889

847 Food and Agriculture Organization of the United Nations (2018). FAOSTAT890  
848 database. URL: <http://www.fao.org/faostat/en/home>. 891

849 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T.,892  
850 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P.,893  
851 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F.,894  
852 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,895  
853 Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J.,896  
854 Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nen-897  
855 del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna,898  
856 M., Semenov, M. A., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K.,899  
857 Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model900  
858 wheat yield impact response surfaces showing sensitivity to temperature and901  
859 precipitation change. *Agricultural Systems*, 159, 209–224. 902

860 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014).903  
861 Evaluating the utility of dynamical downscaling in agricultural impacts pro-904  
862 jections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781.905

863 Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitiv-906  
864 ity of Agricultural Model Performance to Different Climate Inputs. *Journal907  
865 of Applied Meteorology and Climatology*, 55, 151113145618001. 908

866 Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported909  
867 Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous910  
868 Crop Growth and Yield: Application for Wheat in Central Europe. *Remote911  
869 Sensing*, 7, 3934–3965. 912

870 He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H.,913  
871 He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water914  
872 contents and nitrogen leaching to precipitation, management practices and915  
873 soil hydraulic properties in semi-arid and humid regions of Canada using the916  
874 DSSAT model. *Nutrient Cycling in Agroecosystems*, 106, 201–215. 917

875 Heady, E. O. (1957). An Econometric Investigation of the Technology of Agri-918  
876 cultural Production Functions. *Econometrica*, 25, 249–268. 919

877 Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa920  
878 State University Press. 921

879 Holden, P., Edwards, N., PH, G., Fraedrich, K., Lunkeit, F., E, K., Labriet,922  
880 M., Kanudia, A., & F, B. (2014). Plasim-entsem v1.0: A spatiotemporal923  
881 emulator of future climate change for impacts assessment. *Geoscientific924  
882 Model Development*, 7, 433–451. doi:10.5194/gmd-7-433-2014. 925

883 Holzkämper, A., Calanca, P., & Fuhrer, J. (2012). Statistical crop models:926  
884 Predicting the effects of temperature and precipitation changes. *Climate927  
885 Research*, 51, 11–21.

886 Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I.,  
887 McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore,  
888 A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake,  
889 A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgliesh,  
890 N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp,  
891 J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson,  
892 M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T.,  
893 Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorf, J.,  
894 Wedgwood, S., & Keating, B. A. (2014). APSIM Evolution towards a new  
895 generation of agricultural systems simulation. *Environmental Modelling and  
896 Software*, 62, 327 – 350.

897 Howden, S., & Crimp, S. (2005). Assessing dangerous climate change impacts  
898 on australia's wheat industry. *Modelling and Simulation Society of Australia  
899 and New Zealand*, (pp. 505–511).

900 Izumi, T., Nishimori, M., & Yokozawa, M. (2010). Diagnostics of climate  
901 model biases in summer temperature and warm-season insolation for the  
902 simulation of regional paddy rice yield in japan. *Journal of Applied Meteorology  
903 and Climatology*, 49, 574–591.

904 Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Ni-905  
905 tragen Fertilizers. *Ambio*, 6, 146–151.

906 Izaurralde, R., Williams, J., McGill, W., Rosenberg, N., & Quiroga Jakas, M.  
907 (2006). Simulating soil C dynamics with EPIC: Model description and test-  
908 ing against long-term data. *Ecological Modelling*, 192, 362–384.

909 Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the  
910 CROPGRO-soybean model to predict regional yield and production. *Agri-911  
911 culture, Ecosystems & Environment*, 93, 73 – 85.

912 Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L.,  
913 Wilkens, P., Singh, U., Gijsman, A., & Ritchie, J. (2003). The DSSAT  
914 cropping system model. *European Journal of Agronomy*, 18, 235 – 265.

915 Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I.,  
916 Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A.,  
917 Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017).  
918 Toward a new generation of agricultural system data, models, and knowl-  
919 edge products: State of agricultural systems science. *Agricultural Systems*,  
920 155, 269 – 288.

921 Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth,  
922 D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Ver-  
923 burg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K.,  
924 Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003).  
925 An overview of APSIM, a model designed for farming systems simulation.  
926 *European Journal of Agronomy*, 18, 267 – 288.

927 Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., &  
928

- 928 Smith, B. (2013). Implications of accounting for land use in simulations of<sup>971</sup>  
 929 ecosystem carbon cycling in Africa. *Earth System Dynamics*, *4*, 385–407. 972  
 930 Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling<sup>973</sup>  
 931 wheat yield and crop water productivity with high resolution on a global<sup>974</sup>  
 932 scale. *Agricultural Systems*, *94*, 478 – 493. 975  
 933 Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a).<sup>976</sup>  
 934 Global investigation of impacts of PET methods on simulating crop-water<sup>977</sup>  
 935 relations for maize. *Agricultural and Forest Meteorology*, *221*, 164 – 175. 978  
 936 Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C.,<sup>979</sup>  
 937 & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs<sup>980</sup>  
 938 with yields from major crop cultivations. *Science of The Total Environment*,<sup>981</sup>  
 939 *572*, 526 – 537. 982  
 940 Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict<sup>983</sup>  
 941 crop yield responses to climate change. *Agricultural and Forest Meteorol-*<sup>984</sup>  
 942 *ogy*, *150*, 1443 – 1452. 985  
 943 Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relation-<sup>986</sup>  
 944 ships and the impacts of recent warming. *Environmental Research Letters*,<sup>987</sup>  
 945 *2*, 014002. 988  
 946 MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, *4*, 415–447.<sup>989</sup>  
 947 Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P.,<sup>990</sup>  
 948 Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi,<sup>991</sup>  
 949 P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano,<sup>992</sup>  
 950 D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of<sup>993</sup>  
 951 large simulated yield datasets for studying climate effects. (p. 1100).<sup>994</sup>  
 952 doi:10.13140/RG.2.1.5173.8328. 995  
 953 Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrolog-<sup>996</sup>  
 954 ical modelling to study the impact of climate change on the water flows of<sup>997</sup>  
 955 mountain watersheds. *Journal of Hydrology*, *376*, 362 – 377. 998  
 956 Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B.,<sup>999</sup>  
 957 & Calzadilla, A. (2009). Global biomass production potentials exceed ex-<sup>1000</sup>  
 958 pected future demand without the need for cropland expansion. *Nature Com-*<sup>1001</sup>  
 959 *munications*, *6*. 1002  
 960 McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srini-<sup>1003</sup>  
 961 vasa, C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S.,<sup>1004</sup>  
 962 (2015). Integrated assessments of the impacts of climate change on agricul-<sup>1005</sup>  
 963 ture: An overview of AgMIP regional research in South Asia. *Chapter in*<sup>1006</sup>  
 964 *Handbook of Climate Change and Agroecosystems*, (pp. 201–218). 1007  
 965 Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical<sup>1008</sup>  
 966 weather responsiveness of crop yields: US evidence and implications for<sup>1009</sup>  
 967 the agricultural impacts of climate change. *Environmental Research Letters*,<sup>1010</sup>  
 968 *12*. 1011  
 969 Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate<sup>1012</sup>  
 970 change impacts on agriculture implies higher social cost of carbon. *Nature*,<sup>1013</sup>  
 971 *Communications*, *8*.  
 972 Müller, C., Elliott, J., Chrysanthacopoulos, J., Arneth, A., Balkovic, J., Ciais,  
 973 P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde,  
 974 R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T.  
 975 A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G.,  
 976 Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H.  
 977 (2017). Global gridded crop model evaluation: benchmarking, skills, de-  
 978 ficiencies and implications. *Geoscientific Model Development*, *10*, 1403–  
 979 1422.  
 980 Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T.  
 981 (1997). Effect of CO<sub>2</sub> enrichment on carbon and nitrogen interaction in  
 982 wheat and soybean. *Soil Science and Plant Nutrition*, *43*, 789–798.  
 983 O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial.  
 984 *Reliability Engineering & System Safety*, *91*, 1290 – 1300.  
 985 Olin, S., Schurgers, G., Lindeskog, M., Wårldin, D., Smith, B., Bodin, P.,  
 986 Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue  
 987 C:N to changes in atmospheric CO<sub>2</sub> and N management in the main wheat  
 988 regions of western europe. *Biogeosciences*, *12*, 2489–2515. doi:10.5194/bg-  
 989 12-2489-2015.  
 990 Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in  
 991 field crop production. *Soil Science and Plant Nutrition*, *38*, 553–564.  
 992 Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., &  
 993 Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK  
 994 Land Environment Simulator. *Geoscientific Model Development*, *8*, 1139–  
 995 1155.  
 996 Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop  
 997 yields and their variability at different levels of global warming. *Earth Sys-  
 998 tem Dynamics*, *9*, 479–496.  
 999 Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff,  
 1000 S., & Gerten, D. (2015). Emulating global climate change impacts on crop  
 1001 yields. *Statistical Modelling*, *15*, 499–525.  
 1002 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,  
 1003 Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas-  
 1004 soss, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011).  
 1005 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning  
 1006 Research*, *12*, 2825–2830.  
 1007 Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T.,  
 1008 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski,  
 1009 P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.,  
 1010 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,  
 1011 Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet,  
 1012 J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C.,  
 1013 Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M.,

- 1014 Semenov, M., Slawinski, C., Stratnovitch, P., Supit, I., Waha, K., Wang<sub>1057</sub>  
 1015 E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation<sub>1058</sub>  
 1016 effects on wheat yield across a European transect: a crop model ensemble<sub>1059</sub>  
 1017 analysis using impact response surfaces. *Climate Research*, 65, 87–105. <sub>1060</sub>  
 1018 Porter et al. (IPCC) (2014). Food security and food production systems. Cli<sub>1061</sub>  
 1019 mate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global<sub>1062</sub>  
 1020 and Sectoral Aspects. Contribution of Working Group II to the Fifth Assess<sub>1063</sub>  
 1021 ment Report of the Intergovernmental Panel on Climate Change. In C. F<sub>1064</sub>  
 1022 et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK<sub>1065</sub>  
 1023 Cambridge University Press. <sub>1066</sub>
- 1024 Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of<sub>1067</sub>  
 1025 monthly growing areas of 26 irrigated crops. <sub>1068</sub>
- 1026 Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly<sub>1069</sub>  
 1027 Irrigated and Rainfed crop Areas around the Year 2000: A New High<sub>1070</sub>  
 1028 Resolution Data Set for Agricultural and Hydrological Modeling. *Global*<sub>1071</sub>  
 1029 *Biogeochemical Cycles*, 24, GB1011. <sub>1072</sub>
- 1030 Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E.<sub>1073</sub>  
 1031 & Arneth, A. (2016). Climate analogues suggest limited potential for inten<sub>1074</sub>  
 1032 sification of production on current croplands under climate change. *Nature*<sub>1075</sub>  
 1033 *Communications*, 7, 12608. <sub>1076</sub>
- 1034 Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term cli<sub>1077</sub>  
 1035 mate change based on a resampling ensemble technique. *Tellus A: Dynamical*<sub>1078</sub>  
 1036 *Meteorology and Oceanography*, 58, 461–472. <sub>1079</sub>
- 1037 Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the<sub>1080</sub>  
 1038 reduction and sensitivity analysis of complex environmental models. *Environ*<sub>1081</sub>  
 1039 *mental Modelling & Software*, 34, 1 – 4. <sub>1082</sub>
- 1040 Ray, D., Ramankutty, N., Mueller, N., West, P., & Foley, J. (2012). Recent<sub>1083</sub>  
 1041 patterns of crop yield growth and stagnation. *Nature communications*, 3<sub>1084</sub>  
 1042 1293. doi:10.1038/ncomms2296. <sub>1085</sub>
- 1043 Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling<sub>1086</sub>  
 1044 in water resources. *Water Resources Research*, 48. <sub>1087</sub>
- 1045 Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017)<sub>1088</sub>  
 1046 Comparing and combining process-based crop models and statistical models<sub>1089</sub>  
 1047 with some implications for climate change. *Environmental Research Letters*<sub>1090</sub>  
 1048 12. <sub>1091</sub>
- 1049 Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A.<sub>1092</sub>  
 1050 Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek<sub>1093</sub>  
 1051 F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014)<sub>1094</sub>  
 1052 Assessing agricultural risks of climate change in the 21st century in a global<sub>1095</sub>  
 1053 gridded crop model intercomparison. *Proceedings of the National Academy*<sub>1096</sub>  
 1054 *of Sciences*, 111, 3268–3273. <sub>1097</sub>
- 1055 Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P.<sub>1098</sub>  
 1056 Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ew<sub>1099</sub>  
 1067 ert, F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural  
 1068 Model Intercomparison and Improvement Project (AgMIP): Protocols and  
 1069 pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182.
- Rosenzweig, C., Ruane, A. C., Antle, J., Elliott, J., Ashfaq, M., Chatta,  
 1070 A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G.,  
 1071 Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M.,  
 1072 Müller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M.,  
 1073 Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., & Wiebe, K.  
 1074 (2018). Coordinating AgMIP data and models across global and regional  
 1075 scales for 1.5°C and 2.0°C assessments. *Philosophical Transactions of the*  
 1076 *Royal Society of London A: Mathematical, Physical and Engineering*  
 1077 *Sciences*, 376.
- Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason-  
 1078 D'Croz, D., Müller, C., Porter, C., Phillips, M. M., Raymundo, R. M.,  
 1079 Sands, R., Valdivia, R. O., White, J. W., Wiebe, K., & Rosenzweig, C.  
 1080 (2018). Biophysical and economic implications for agriculture of +1.5° and  
 1081 +2.0°C global warming using AgMIP Coordinated Global and Regional As-  
 1082 sessments. *Climate Research*, 76, 17–39.
- Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown,  
 1083 D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013).  
 1084 Climate change impact uncertainties for maize in panama: Farm informa-  
 1085 tion, climate projections, and yield sensitivities. *Agricultural and Forest*  
 1086 *Meteorology*, 170, 132 – 145.
- Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forc-  
 1087 ing datasets for agricultural modeling: Merged products for gap-filling and  
 1088 historical climate series estimation. *Agric. Forest Meteorol.*, 200, 233–248.
- Ruane, A. C., McDermid, S., Rosenzweig, C., Baigorria, G. A., Jones, J. W.,  
 1089 Romero, C. C., & Cecil, L. D. (2014). Carbon-temperature-water change  
 1090 analysis for peanut production under climate change: A prototype for the  
 1091 agmip coordinated climate-crop modeling project (c3mp). *Glob. Change*  
 1092 *Biol.*, 20, 394–407. doi:10.1111/gcb.12412.
- Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901–  
 1093 2100 depicted by world maps of the Köppen-Geiger climate classification.  
 1094 *Meteorologische Zeitschrift*, 19, 135–141.
- Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting  
 1095 dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19,  
 1096 607–620.
- Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng,  
 1097 D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolin-  
 1098 sksi, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frier, K.  
 1099 (2017). Consistent negative response of US crops to high temperatures in  
 1100 observations and crop models. *Nature Communications*, 8, 13931.
- Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate

- severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106, 15594–15598.
- Snyder, A., Calvin, K. V., Phillips, M., & Ruane, A. C. (2018). A crop yield change emulator for use in gcam and similar models: Persephone v1.0. *Geoscientific Model Development Discussions*, 2018, 1–42.
- Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety*, 94, 1735 – 1763.
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93, 485–498.
- Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters*, 35.
- 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.
- 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.
- 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.
- 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.
- 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.
- de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Landbouwkundige Onderzoeken* : 64.6, .
- Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato yields on changes in climate and co2. *Agricultural and Forest Meteorology*, 112, 217 – 231.
- 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., Miller, 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.