

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

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

## 1. Introduction

2 Understanding crop yield response to a changing climate  
3 is critically important, especially as the global food produc-  
4 tion system will face pressure from increased demand over the  
5 next century. Climate-related reductions in supply could there-  
6 fore have severe socioeconomic consequences. Multiple stud-  
7 ies using different crop or climate models concur in predicting  
8 sharp yield reductions on currently cultivated cropland under  
9 business-as-usual climate scenarios, although their yield pro-  
10 jections show considerable spread (e.g. Porter et al. (IPCC),  
11 2014, Rosenzweig et al., 2014, Schauberger et al., 2017, and  
12 references therein). Modeling crop responses continues to be  
13 challenging, as crop growth is a function of complex interac-  
14 tions between climate inputs and management practices. There-  
15 fore model intercomparison projects targeting model response  
16 to important drivers are critical to improve future projections.

17 Computational models have been used to project crop yields  
18 since the 1950's, beginning with statistical models that attempt  
19 to capture the relationship between input factors and resultant  
20 yields (e.g. Heady, 1957, Heady & Dillon, 1961). These statis-  
21 tical models were typically developed on a small scale for loca-  
22 tions with extensive histories of yield data. The emergence of  
23 electronic computers allowed development of numerical mod-  
24 els that simulate the process of photosynthesis and the biology  
25 and phenology of individual crops (first proposed by de Wit  
26 (1957) and Duncan et al. (1967) and attempted by Duncan  
27 (1972); for a history of crop model development see Rosen-  
28 zweig et al. (2014)). A half-century of improvement in both  
29 models and computing resources means that researchers can  
30 now run crop simulations for many years at high spatial res-  
31 olution on the global scale.

32 Both types of models continue to be used, and compara-

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

41 Process-based models do continue to struggle with some im-  
42 portant details, including reproducing historical year-to-year  
43 variability (e.g. Müller et al., 2017), reproducing historical  
44 yields when driven by reanalysis weather (e.g. Glotter et al.,  
45 2014), and low sensitivity to extreme events (e.g. Glotter et al.,  
46 2015). These issues are driven in part by the diversity of new  
47 cultivars and genetic variants, which outstrips the ability of aca-  
48 demic modeling groups to capture them (e.g. Jones et al., 2017).  
49 Models also do not simulate many additional factors affecting  
50 production, including pests, diseases, and weeds. For these rea-  
51 sons, individual studies must generally re-calibrate models to  
52 ensure that short-term predictions reflect current cultivar mixes,  
53 and long-term projections retain considerable uncertainty (Wolf  
54 & Oijen, 2002, Jagtap & Jones, 2002, Iizumi et al., 2010, An-  
55 gulo et al., 2013, Asseng et al., 2013, 2015). Inter-model dis-  
56 crepancies can also be high in areas not yet cultivated (e.g.  
57 Challinor et al., 2014, White et al., 2011). Finally, process-  
58 based models present additional difficulties for high-resolution  
59 global studies because of their complexity and computational  
60 requirements. For economic impacts assessments, it is often  
61 impossible to integrate a set of process-based crop models di-  
62 rectly into an integrated assessment model to estimate the po-  
63 tential cost of climate change to the agricultural sector.

64 Nevertheless, process-based models are necessary for under-  
65 standing the global future yield impacts of climate change for  
66 many reasons. First, cultivation may shift to new areas, where

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

67 no yield data are currently available and therefore statistical  
 68 models cannot apply. Yield data are also often limited in the  
 69 developing world, where future climate impacts may be the  
 70 most critical. Finally, only process-based models can capture  
 71 the growth response to novel conditions and practices that are  
 72 not represented in historical data (e.g. Pugh et al., 2016, Roberts  
 73 et al., 2017). These novel changes can include the direct fertil-  
 74 ization effect of elevated CO<sub>2</sub>, or changes in management prac-  
 75 tices that may ameliorate climate-induced damages.

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

92 In the past decade, multiple studies have developed emula-  
 93 tors of process-based crop simulations. Early studies proposing  
 94 or describing potential crop yield emulators include Howden  
 95 & Crimp (2005), Räisänen & Ruokolainen (2006), Lobell &  
 96 Burke (2010), and Ferrise et al. (2011), who used a machine  
 97 learning approach to predict Mediterranean wheat yields. Stud-  
 98 ies developing single-model emulators include Holzkämper  
 99 et al. (2012) for the CropSyst model, Ruane et al. (2013) for  
 100 the CERES wheat model, and Oyebamiji et al. (2015) for the

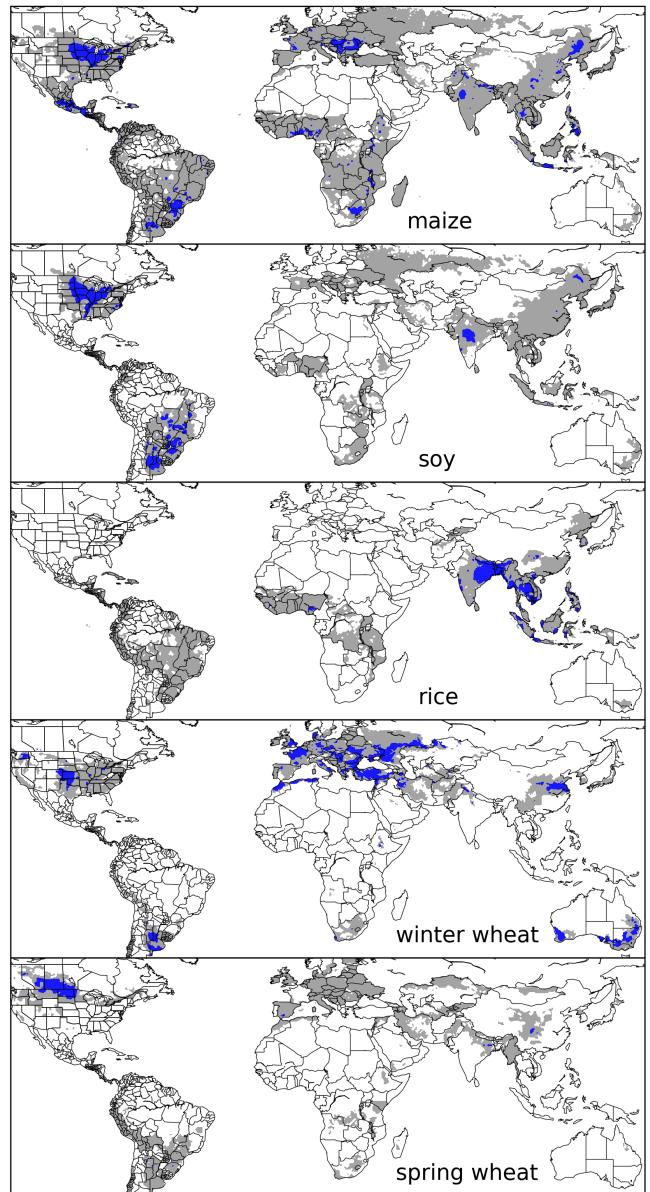


Figure 1: Presently cultivated area for rain-fed crops. Blue indicates grid cells with more than 20,000 hectares (~10% of the equatorial grid cells) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated areas for maize, rice, and soy are taken from the MIRCA2000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

LPJmL model (for multiple crops, using multiple scenarios as a training set). More recently, emulators have begun to be used in the context of multi-model intercomparisons, with Blanc & Sultan (2015), Blanc (2017), Ostberg et al. (2018) and Mis-  
 try et al. (2017) using them to analyze the five crop models of the Inter-Sectoral Impacts Model Intercomparison Project

107 (ISIMIP) (Warszawski et al., 2014), which simulated yields for<sub>141</sub> (Makowski et al., 2015, Pirttioja et al., 2015), and several re-  
108 maize, soy, wheat, and rice. Choices differ: Blanc & Sul-<sub>142</sub> cent studies in 2018 (Fronzek et al., 2018, Snyder et al., 2018,  
109 tan (2015) and Blanc (2017) base their emulation on histori-<sub>143</sub> Ruiz-Ramos et al., 2018). All three studies sample multiple per-  
110 cal simulations and a single future climate/emissions scenario<sub>144</sub> turbations to temperature and precipitation (with Snyder et al.  
111 (RCP8.5), and use local weather variables and yields in their<sub>145</sub> (2018) and Ruiz-Ramos et al. (2018) adding CO<sub>2</sub> as well), in  
112 regression but then aggregate across broad regions; Ostberg<sub>146</sub> 132, 99 and 220 different combinations, respectively, and take  
113 et al. (2018) consider multiple future climate scenarios, using<sub>147</sub> advantage of the structured training set to construct emulators  
114 global mean temperature change (and CO<sub>2</sub>) as regressors but<sub>148</sub> (“response surfaces”) of climatological mean yields, omitting  
115 then pattern-scale to emulate local yields; while Mistry et al.<sub>149</sub> year-over-year variations. All are limited in some respects and  
116 (2017) attempt to compare emulated historical yearly yields to<sub>150</sub> focus on a limited number of sites. Fronzek et al. (2018) and  
117 observed historical yields, using local weather data and a his-<sub>151</sub> Ruiz-Ramos et al. (2018) simulate only wheat (over many mod-  
118 torical crop simulation. These efforts do share important com-<sub>152</sub> els) and Snyder et al. (2018) analyzes four crops (maize, wheat,  
119 mon features: all emulate annual crop yields across the entire<sub>153</sub> rice, soy) for agricultural impacts experiments with the GCAM  
120 scenario or scenarios, and when future scenarios are consid-<sub>154</sub> (Calvin et al., 2019) model.  
121 ered, they are non-stationary, i.e. their input climate parameters  
122 evolve over time.

123 An alternative approach is to construct a training set of multi-<sub>155</sub> In this paper we describe a new comprehensive dataset de-  
124 ple stationary scenarios in which parameters are systematically<sub>156</sub> signed to expand the parameter sweep approach still further.  
125 varied. Such a “parameter sweep” offers several advantages for<sub>157</sub> The Global Gridded Crop Model Intercomparison (GGCMI)  
126 emulation over scenarios in which climate evolves over time.<sub>158</sub> Phase II experiment involves running a suite of process-based  
127 First, it allows separating the effects of different variables that<sub>159</sub> crop models across historical conditions perturbed by a set of  
128 impact yields but that are highly correlated in realistic future<sub>160</sub> discrete steps in different input parameters, including an ap-  
129 scenarios (e.g. CO<sub>2</sub> and temperature). Second, it allows making<sub>161</sub> plied nitrogen dimension. The experimental protocol involves  
130 a distinction between year-over-year yield variations and cli-<sub>162</sub> over 700 different parameter combinations for each model and  
131 matological changes, which may involve different responses to<sub>163</sub> crop, with simulations providing near-global coverage at a half  
132 the particular climate regressors used (e.g. Ruane et al., 2016).<sub>164</sub> degree spatial resolution. The experiment was conducted as  
133 For example, if year-over-year yield variations are driven pre-<sub>165</sub> part of the Agricultural Model Intercomparison and Improve-  
134 dominantly by variations in the distribution of temperatures<sub>166</sub> ment Project (AgMIP) (Rosenzweig et al., 2013, 2014), an in-  
135 throughout the growing season, and long-term climate changes<sub>167</sub> ternational effort conducted under a framework similar to the  
136 are driven predominantly by shifts in means, then regressing<sub>168</sub> Climate Model Intercomparison Project (CMIP) (Taylor et al.,  
137 on the mean growing season temperature will produce different<sub>169</sub> 2012, Eyring et al., 2016). The GGCMI protocol builds on the  
138 yield responses at annual vs. climatological timescales.<sub>170</sub> AgMIP Coordinated Climate-Crop Modeling Project (C3MP)  
139 Systematic parameter sweeps have begun to be used in crop<sub>171</sub> (Ruane et al., 2014, McDermid et al., 2015) and will con-  
140 model evaluation and emulation, with early efforts in 2015<sub>172</sub> tribute to the AgMIP Coordinated Global and Regional As-  
173 sessments (CGRA) (Ruane et al., 2018, Rosenzweig et al.,  
174 2018). GGCMI Phase II is designed to allow addressing goals

such as understanding where highest-yield regions may shift<sup>191</sup>  
 under climate change; exploring future adaptive management<sup>192</sup>  
 strategies; understanding how interacting input drivers affect<sup>193</sup>  
 crop yield; quantifying uncertainties across models and major  
 drivers; and testing strategies for producing lightweight em-<sup>194</sup>  
 ulators of process-based models. In this paper, we describe  
 the GGCMI Phase II experiments, present initial results, and  
 demonstrate that it is tractable to emulation.<sup>195</sup>

## 2. Simulation – Methods

GGCMI Phase II is the continuation of a multi-model com-<sup>211</sup>  
 parison exercise begun in 2014. The initial Phase I compared<sup>212</sup>  
 harmonized yields of 21 models for 19 crops over a 31-year<sup>213</sup>  
 historical (1980-2010) scenario with a primary goal of model<sup>214</sup>  
 evaluation (Elliott et al., 2015, Müller et al., 2017). Phase II<sup>215</sup>  
 compares simulations of 12 models for 5 crops (maize, rice,<sup>216</sup>  
 soybean, spring wheat, and winter wheat) over the same histor-<sup>217</sup>  
 ical time series (1980-2010) used in Phase I, but with individ-<sup>218</sup>  
 ual climate or management inputs adjusted from their historical<sup>219</sup>  
 values. The reduced set of crops includes the three major global<sup>220</sup>  
 cereals and the major legume and accounts for over 50% of hu-<sup>221</sup>  
 man calories (in 2016, nearly 3.5 billion tons or 32% of total<sup>222</sup>  
 global crop production by weight (Food and Agriculture Orga-<sup>223</sup>  
 nization of the United Nations, 2018).<sup>224</sup>

The guiding scientific rationale of GGCMI Phase II is to pro-<sup>225</sup>  
 vide a comprehensive, systematic evaluation of the response<sup>226</sup>  
 of process-based crop models to different values for carbon<sup>227</sup>

dioxide, temperature, water, and applied nitrogen (collectively  
 known as “CTWN”). The dataset is designed to allow re-  
 searchers to:

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

The experimental protocol consists of 9 levels for precipitation perturbations, 7 for temperature, 4 for 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

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

Table 1: GGCMI Phase II input variable test levels. Temperature and precipitation values indicate the perturbations from the historical, climatology. \* Only simulated by one model. W-percentage does not apply to the irrigated (W<sub>inf</sub>) simulations, which are all simulated at the maximum beneficial levels of water.

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

Table 2: Models included in GGCMI Phase II and the number of C, T, W, and N simulations that each performs for rain-fed crops (“Sims per Crop”), with 672 as the maximum. “N-Dim.” indicates whether the simulations include varying nitrogen levels. Two models provide only one nitrogen level, and †PROMET provides only two of the three nitrogen levels (and so is not emulated across the nitrogen dimension). All models provide the same set of simulations across all modeled crops, but some omit individual crops. (For example, APSIM does not simulate winter wheat.) Irrigated simulations are provided at the level of the other covariates for each model, i.e. an additional 84 simulations for fully-sampled models. Simulations are nearly global but their geographic extent can vary, even for different crops in an individual model, since some simulations omit regions far outside the currently cultivated area. In most cases, historical daily climate inputs are taken from the 0.5 degree NASA AgMERRA daily gridded re-analysis product specifically designed for agricultural modeling, with satellite-corrected precipitation (Ruane et al., 2015), but two models (marked with \*) require sub-daily input data and use alternative sources. See Elliott et al. (2015) for additional details.

higher temperatures. An additional, identical set of scenarios<sup>246</sup> Sacks et al. (2010) and Portmann et al. (2008, 2010)), but vary  
 (at the same C, T, W, and N levels) not shown or analyzed here<sup>247</sup> by crop and by location on the globe. For example, maize is  
 simulate adaptive agronomy under climate change by varying<sup>248</sup> sown in March in Spain, in July in Indonesia, and in December  
 the growing season for crop production. The resulting GGCMI<sup>249</sup> in Namibia. All stresses are disabled other than factors related  
 Phase II dataset captures a distribution of crop responses over<sup>250</sup> to nitrogen, temperature, and water (e.g. alkalinity and salinity).  
 the potential space of future climate conditions.<sup>251</sup> No additional nitrogen inputs, such as atmospheric deposition,  
 The 12 models included in GGCMI Phase II are all mech-<sup>252</sup> are considered, but some model treatment of soil organic matter  
 anistic process-based crop models that are widely used in im-<sup>253</sup> may allow additional nitrogen release through mineralization.  
 pacts assessments (Table 2). Although some models share a<sup>254</sup> See Rosenzweig et al. (2014), Elliott et al. (2015) and Müller  
 common base (e.g. the LPJ family or the EPIC family of mod-<sup>255</sup> et al. (2017) for further details on models and underlying as-  
 els), they have subsequently developed independently. (For<sup>256</sup> sumptions.

The participating modeling groups provide simulations at<sup>740</sup> any of four initially specified levels of participation, so the num-  
 key factors are not standardized across the experiment, includ-<sup>741</sup> ber of simulations varies by model, with some sampling only a  
 ing secondary soil nutrients, carry-over effects across growing<sup>742</sup> part of the experiment variable space. Most modeling groups  
 years including residue management and soil moisture, and the<sup>743</sup> simulate all five crops in the protocol, but some omitted one  
 extent of simulated area for different crops. Growing seasons<sup>744</sup> or more. Table 2 provides details of coverage for each model.  
 are standardized across models (with assumptions based on<sup>745</sup> Note that the three models that provide less than 50 simulations

264 are excluded from the emulator analysis.

265 Each model is run at 0.5 degree spatial resolution and cov-  
 266 ers all currently cultivated areas and much of the uncultivated  
 267 land area. (See Figure 1 for the present-day cultivated area of  
 268 rain-fed crops, and Figure S1 in the Supplemental Material for  
 269 irrigated crops.) Coverage extends considerably outside cur-  
 270 rently cultivated areas because cultivation will likely shift under  
 271 climate change. However, areas are not simulated if they are  
 272 assumed to remain non-arable even under an extreme climate  
 273 change; these regions include Greenland, far-northern Canada,  
 274 Siberia, Antarctica, the Gobi and Sahara Deserts, and central  
 275 Australia.

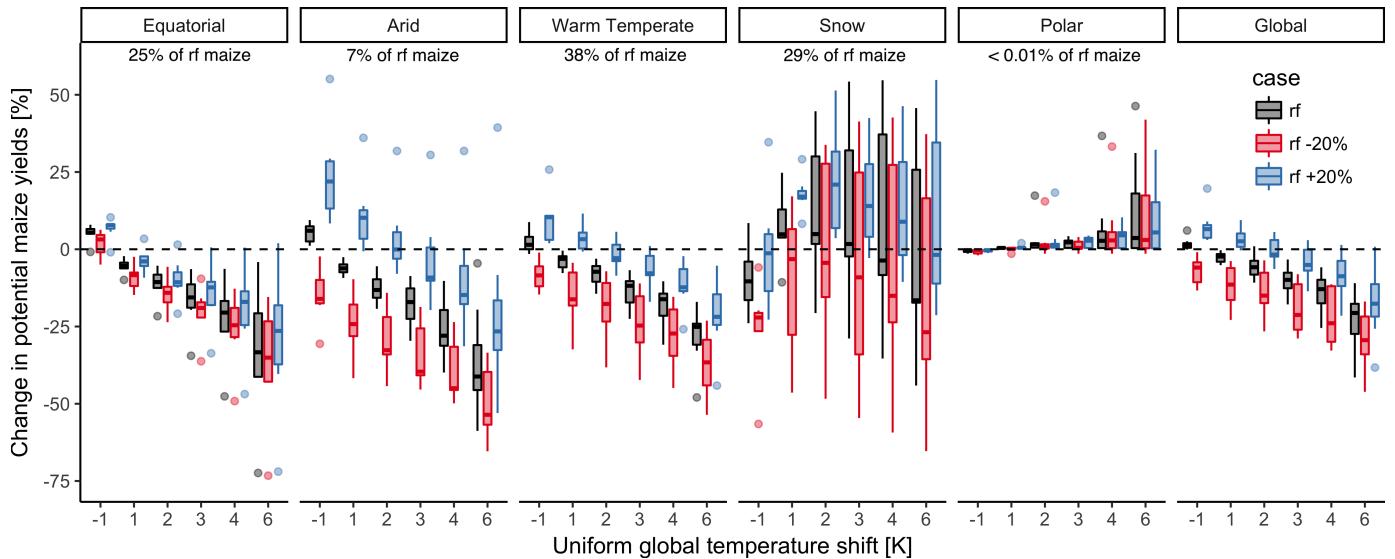
280 as the baseline the scenario with historical climatology (i.e. T  
 281 and P changes of 0), C of 360 ppm, and applied N at 200 kg  
 282 ha<sup>-1</sup>. We show absolute yields in some cases to illustrate geo-  
 283 graphic differences in yields.

284 The GGCMI Phase II simulations are designed for evaluat-  
 285 ing changes in yield but not absolute yields, since they omit  
 286 detailed calibrations. To provide some validation of the skill of  
 287 the process-based models used, we repeat the validation exer-  
 288 cises of Müller et al. (2017) for GGCMI Phase I. See Appendix  
 289 A for details on simulation model validation.

### 290 3. Simulation – Results

291 All models produce as output crop yields (tons ha<sup>-1</sup> year<sup>-1</sup>)  
 292 for each 0.5 degree grid cell. Because both yields and yield  
 293 changes vary substantially across models and across grid cells,  
 294 we primarily analyze relative change from a baseline. We take

295 Crop models in the GGCMI Phase II ensemble show broadly  
 296 consistent responses to climate and management perturbations  
 297 in most regions, with a strong negative impact of increased tem-  
 298 perature in all but the coldest regions. We illustrate this result



299 Figure 2: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek  
 300 (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation levels  
 301 (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  
 302 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  
 303 material for only currently-cultivated land. Panel text gives the percentage of rain-fed maize presently cultivated in each climate zone (data from Portmann et al.,  
 304 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  
 305 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  
 306 climate regions (other than cold continental), with projected changes larger than inter-model variance. Outliers in the tropics (strong negative impact of temperature  
 307 increases) are the pDSSAT model; outliers in the high-rainfall case (strong positive impact of precipitation increases) are the JULES model. Inter-model variance  
 308 increases in the case where precipitation is reduced, suggesting uncertainty in model response to water limitation. The right panel with global changes shows yield  
 309 responses to an globally uniform temperature shift; note that these results are not directly comparable to simulations of more realistic climate scenarios with the  
 310 same global mean change.

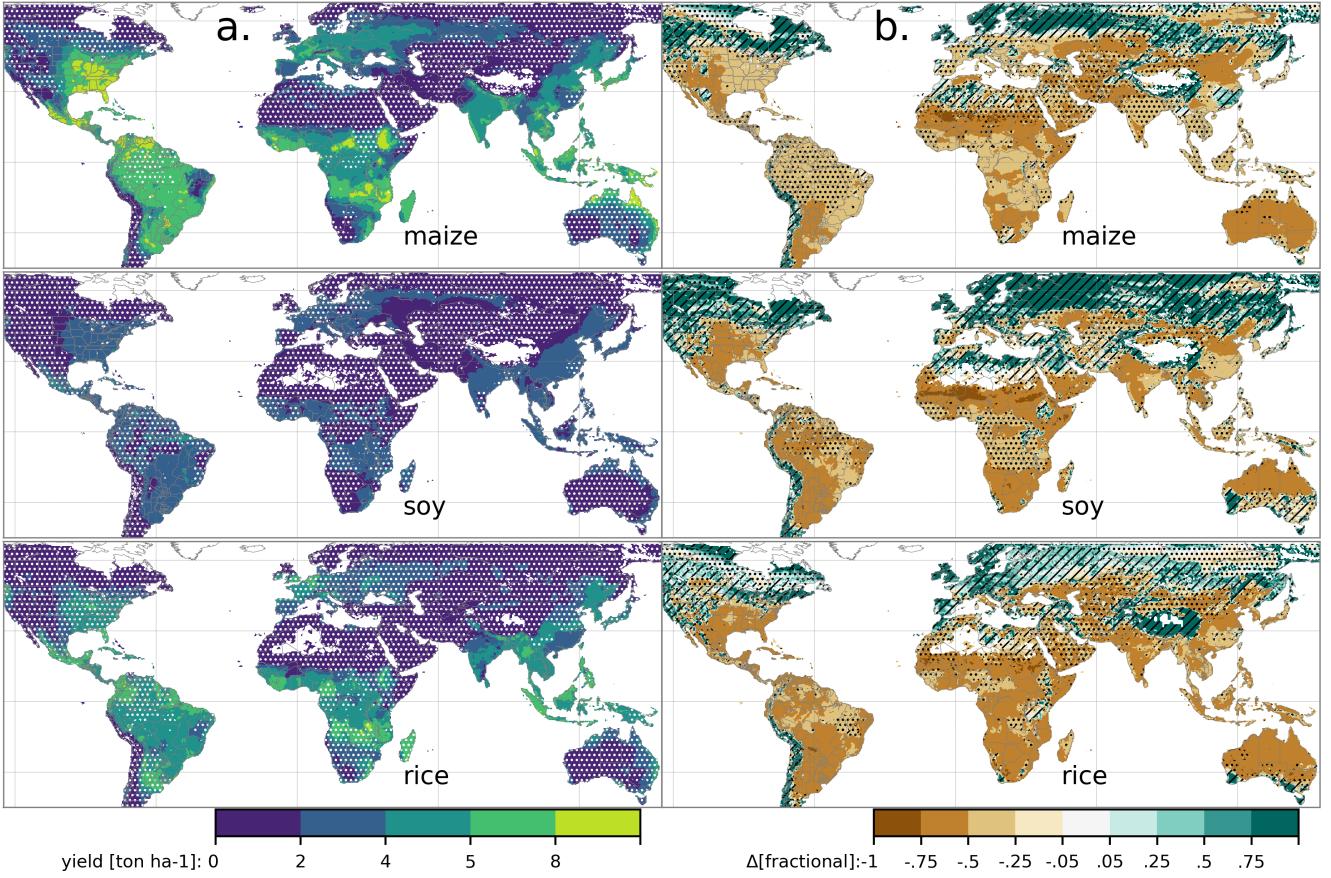


Figure 3: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure S11 in the supplemental material.) White stippling indicates areas where these crops are not currently cultivated. Absence of cultivation aligns well with the lowest yield contour ( $0\text{--}2 \text{ ton ha}^{-1}$ ). Right column (b) shows the multi-model mean fractional yield change in the extreme  $T + 4 \text{ }^{\circ}\text{C}$  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.

for rain-fed maize in Figure 2, which shows yields for the primary Köppen-Geiger climate regions (Rubel & Kottke, 2010). In warming scenarios, models show decreases in maize yield in the warm temperate, equatorial, and arid regions that account for nearly three-quarters of global maize production. These impacts are robust for even moderate climate perturbations. In the warm temperate zone, even a 1 degree temperature rise with other variables held fixed leads to a median yield reduction that outweighs the variance across models. A 6 degree temperature rise results in median loss of  $\sim 25\%$  of yields with a signal to noise ratio of nearly three to one. A notable exception is the snow region, where models disagree strongly, extending even to the sign of impacts. Other crops show similar responses

to warming, with robust yield losses in warmer locations and high inter-model variance in the cold continental regions (Figure S7).

The effects of rainfall changes on maize yields shown in Figure 2 are also as expected and are consistent across models. Increased rainfall mitigates the negative effect of higher temperatures by counteracting the increased evapo-transpiration to some degree, most strongly in arid regions. Decreased rainfall amplifies yield losses and also increases inter-model variance more strongly, suggesting that models have difficulty representing crop response to water stress or increased evapo-transpiration due to warmer temperatures. We show only rain-fed maize here; see Figure S5 for the irrigated case. As ex-

pected, irrigated crops are more resilient to temperature increases in all regions, especially so where water is limiting.  
 Mapping the distribution of baseline yields and yield changes shows the geographic dependencies that underlie these results. Figure 3 shows baseline and changes in the T+4 scenario for rain-fed maize, soy, and rice in the multi-model ensemble mean, with locations of model agreement marked. Absolute yield potentials show strong spatial variation, with much of the Earth's surface area unsuitable for any given crop. In general, models agree most on yield response in regions where yield potentials are currently high and therefore where crops are currently grown. Models show robust decreases in yields at low latitudes, and highly uncertain median increases at most high latitudes. For wheat crops see Figure S11; wheat projections are more uncertain, possible because calibration is especially important for wheat (e.g. Asseng et al., 2013).

#### 4. Emulation – Methods

As part of our demonstration of the properties of the GGCMI Phase II dataset, we construct an emulator of 30-year climatological mean yields. This approach is made possible by the structured set of simulations involving systematic perturbations. In the GGCMI Phase II dataset, the year-over-year responses are generally quantitatively distinct from (and larger than) climatological mean responses. In the example of Figure 4, responses to year-over-year temperature variations are 100% larger than those to long-term perturbations in the baseline case, and larger still under warmer conditions, rising to nearly 200% more in the T+6 case. The stronger year-over-year response under warmer conditions also manifests as a wider distribution of yields (Figure 5). As discussed previously, year-over-year and climatological responses can differ for many reasons including memory in the crop model, lurking covariants, and differing associated distributions of daily growing-season daily

weather (e.g. Ruane et al., 2016). Note that the GGCMI Phase II datasets do not capture one climatological factor, potential future distributional shifts, because all simulations are run with fixed offsets from the historical climatology. Prior work has suggested that mean changes are the dominant drivers of climatological crop yield shifts in non-arid regions (e.g. Glotter et al., 2014).

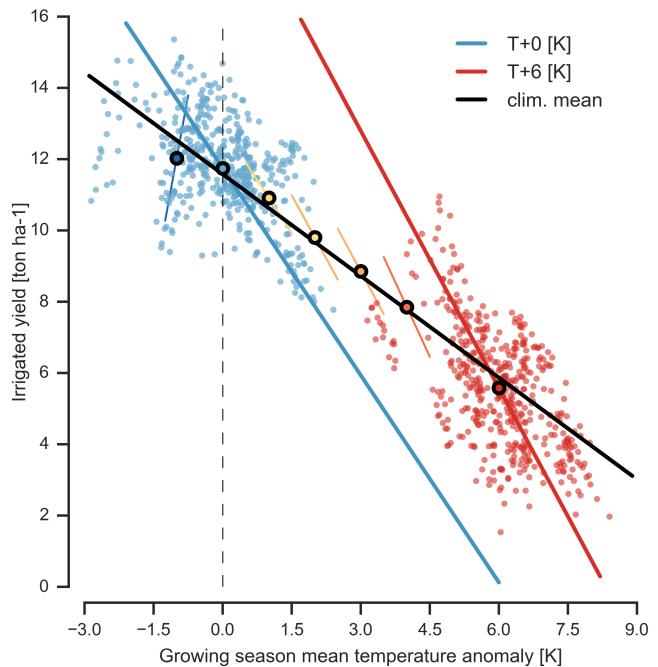


Figure 4: Example showing distinction between crop yield responses to year-to-year and climatological mean temperature shifts. Figure shows irrigated maize for a representative high-yield region (nine adjacent grid cells in northern Iowa) from the pDSSAT model, for the baseline 1981-2010 historical climate (blue) and for the scenario of maximum temperature change (+6 K, red). Other variables are held at baseline values, and the choice of irrigated yields means that precipitation is not a factor. Open black circles mark climatological mean yield values for all six temperature scenarios (T-1, +0, +1, +2, +3, +4, +6). Colored lines show total least squares linear regressions of year-over-year variations in each scenario. Black line shows the fit through the climatological mean values. Responses to year-over-year temperature variations (colored lines) are 100–200% larger than those to long-term climate perturbations, rising under warmer conditions.

Emulation involves fitting individual regression models for each crop, simulation model, and 0.5 degree geographic pixel from the GGCMI Phase II dataset; the regressors are the applied constant perturbations in CO<sub>2</sub>, temperature, water, and nitrogen (C, T, W, N). We regress 30-year climatological mean yields against a third-order polynomial in C, T, W, and N with interaction terms. (We aggregate the entire 30-year run in each case to

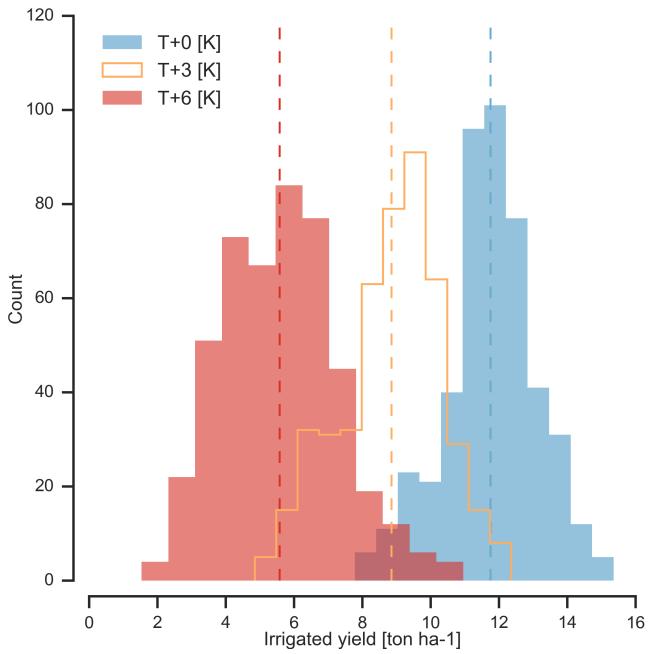


Figure 5: Example showing climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows irrigated maize for<sup>397</sup> nine adjacent high-yield grid cells of Figure 4 from the pDSSAT model, for the baseline 1981-2010 historical climate (blue) and for scenarios with temperature<sup>398</sup> shifted by T+3 (orange) and T+6 K (red), with other variables held at baseline values. The stronger year-over-year temperature response with higher temperatures seen in Figure 4 is manifested here as larger variance in annual yields even<sup>399</sup> though the variance in climate drivers is identical. In this work we emulate not the year-over-year distributions but the climatological mean response<sup>400</sup> (dashed vertical lines).

improve signal to noise ration in our model.) The higher-order<sup>402</sup> terms are necessary to capture any nonlinear responses, which<sup>403</sup> are well-documented in observations for temperature and wa-<sup>404</sup> ter perturbations (e.g. Schlenker & Roberts (2009) for T and<sup>405</sup> He et al. (2016) for W). We include interaction terms (both lin-<sup>406</sup> ear and higher-order) because past studies have shown them to<sup>407</sup> be significant effects. For example, Lobell & Field (2007) and<sup>408</sup> Tebaldi & Lobell (2008) showed that in real-world yields, the<sup>409</sup> joint distribution in T and W is needed to explain observed yield<sup>410</sup> variance. (C and N are fixed in these data.) Other observation-<sup>411</sup> based studies have shown the importance of the interaction be-<sup>412</sup> tween water and nitrogen (e.g. Aulakh & Malhi, 2005), and be-<sup>413</sup> tween nitrogen and carbon dioxide (Osaki et al., 1992, Nakamura et al., 1997). To avoid overfitting or unstable parameter estimation, we apply a feature selection procedure (described below) that reduces the potential 34-term polynomial (for the

<sup>384</sup> rain-fed case) to 23 terms.

<sup>385</sup> We do not focus on comparing different functional forms in this study, and instead choose a relatively simple parametrization that allows for some interpretation of coefficients. Some <sup>388</sup> prior studies have used more complex functional forms and <sup>389</sup> larger numbers of parameters, e.g. 39 in Blanc & Sultan (2015) and Blanc (2017), who borrow information across space by fitting <sup>391</sup> grid points simultaneously across a large region in a panel regression. The simple functional form used here allows emulation <sup>392</sup> at the grid cell level. The emulation therefore indirectly includes <sup>393</sup> any yield response to geographically distributed factors such as soil type, insolation, and the baseline climate itself. We <sup>396</sup> hold the statistical specification constant across all crops and models to facilitate parameter by parameter simulation model comparison.

#### 4.1. Feature selection procedure

Although the GGCMI Phase II sampled variable space is large, it is still sufficiently limited that use of the full polynomial expression described above can be problematic. We therefore reduce the number of terms through a feature selection cross-validation process in which terms in the polynomial are tested for importance. In this procedure higher-order and interaction terms are added successively to the regression model; we then follow the reduction of the aggregate mean squared error with increasing terms and eliminate those terms that do not contribute significant reductions. See supplemental documents for more details. We select terms by applying the feature selection process to the three models that provided the complete set of 672 rain-fed simulations (pDSSAT, EPIC-TAMU, and LPJmL); the resulting choice of terms is then applied for all emulators.

Feature importance is remarkably consistent across all three models and across all crops (see Figure S4 in the supplemental material). The feature selection process results in a final polynomial in 23 terms, with 11 terms eliminated. We omit the  $N^3$

term, which cannot be fitted because we sample only three nitrogen levels. We eliminate many of the C terms: the cubic, the CT, CTN, and CWN interaction terms, and all higher order interaction terms in C. Finally, we eliminate two 2nd-order interaction terms in T and one in W. Implication of this choice include that nitrogen interactions are complex and important, and that water interaction effects are more nonlinear than those in temperature. The resulting statistical model (Equation 1) is used for all grid cells, models, and rain-fed crops. (The regressions for irrigated crops do not contain the W terms and the

models that do not sample the nitrogen levels omit the N terms).

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

414 To fit the parameters  $K$ , we use a Bayesian Ridge  
 415 probabilistic estimator (MacKay, 1991), which reduces volatility in  
 416 parameter estimates when the sampling is sparse, by weight-  
 417 ing parameter estimates towards zero. The Bayesian Ridge  
 418 method is necessary to maintain a consistent functional form  
 419 across all models and locations. We use the implementation of  
 420 the Bayesian Ridge estimator from the scikit-learn package in  
 421 Python (Pedregosa et al., 2011). In the GGCMI Phase II ex-  
 422 periment, the most problematic fits are those for models that

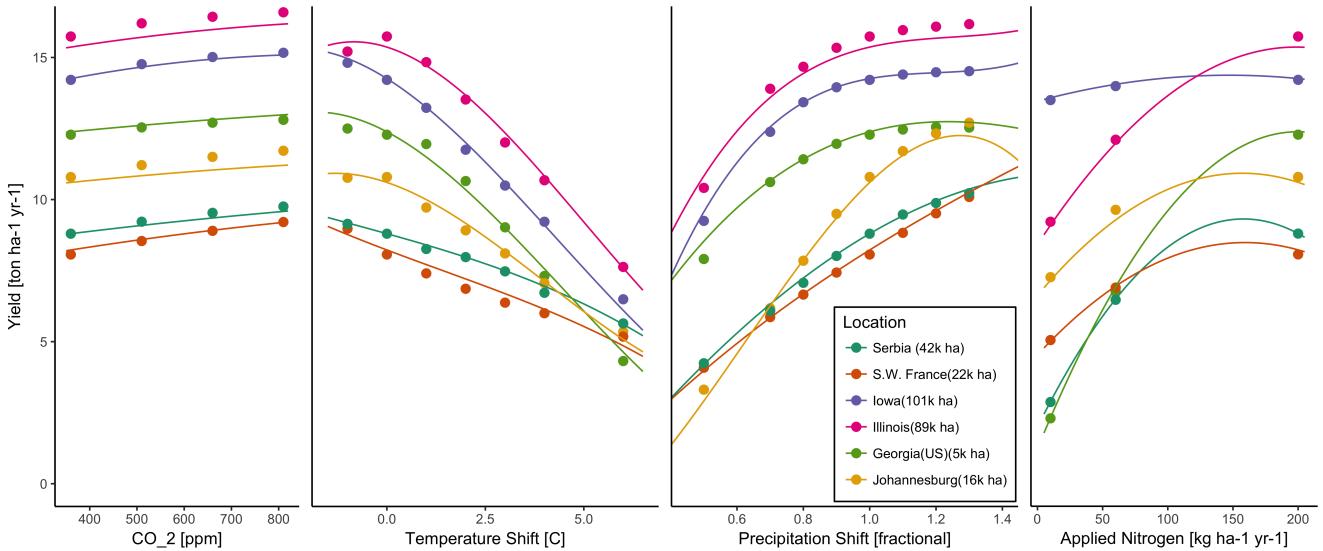


Figure 6: Illustration of spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation along a single variable, with others held at baseline values. Dots show climatological mean yields and lines the results of the full 4D emulator of Equation 1. In general the climatological response surface is sufficiently smooth that it can be represented with the sampled variable space by the simple polynomial used in this work. Extrapolation can however produce misleading results. Nitrogen fits may not be realistic at intermediate values given limited sampling. For more detailed emulator assessment, see Appendix ??.

423 provided a limited number of cases or for low-yield geographic<sup>456</sup>  
424 regions where some modeling groups did not run all scenar-<sup>457</sup>  
425 ios. We do not attempt to emulate models that provided less<sup>458</sup>  
426 than 50 simulations. The lowest number of simulations emu-<sup>459</sup>  
427 lated across the full parameter space is then 130 (for the PEPIC<sup>460</sup>  
428 model). The resulting parameter matrices for all crop model<sup>461</sup>  
429 emulators are available on request [give location?](#), as are the raw<sup>462</sup>  
430 simulation data and a Python application to emulate yields. The<sup>463</sup>  
431 yield output for a single GGCMI Phase II model that simulates<sup>464</sup>  
432 all scenarios and all five crops is  $\sim$ 12.5 GB; the emulator is<sup>465</sup>  
433  $\sim$ 100 MB, a reduction by over two orders of magnitude.

## 434 5. Emulation – Results

435 Emulation provides not only a computational tool but a<sup>469</sup>  
436 means of understanding and interpreting crop yield response<sup>470</sup>  
437 across the parameter space. Emulation is only possible when<sup>471</sup>  
438 crop yield responses are sufficiently smooth and continuous to<sup>472</sup>  
439 allow fitting with a relatively simple functional form, but this<sup>473</sup>  
440 condition largely holds in the GGCMI Phase II simulations. Re-<sup>474</sup>  
441 sponds are quite diverse across locations, crops, and models,<sup>475</sup>  
442 but in most cases local responses are regular enough to permit<sup>476</sup>  
443 emulation. We show illustrations of emulation fidelity in this<sup>477</sup>  
444 section; for more detailed discussion see Appendix ??.

445 Crop yield responses are geographically diverse, even in<sup>479</sup>  
446 high-yield and high-cultivation areas. Figure 6 illustrates ge-<sup>480</sup>  
447 ographic diversity for a single crop and model (rain-fed maize<sup>481</sup>  
448 in pDSSAT); this heterogeneity supports the choice of emulat-<sup>482</sup>  
449 ing at the grid cell level. Each panel in Figure 6 shows sim-<sup>483</sup>  
450 ulted yield output from scenarios varying only along a single<sup>484</sup>  
451 dimension ( $\text{CO}_2$ , temperature, precipitation, or nitrogen addi-<sup>485</sup>  
452 tion), with other inputs held fixed at baseline levels, compared<sup>486</sup>  
453 to the full 4D emulation across the parameter space. Yields<sup>487</sup>  
454 evolve smoothly across the space sampled, and the polynomial<sup>488</sup>  
455 fit captures the climatological response to perturbations.Crop<sup>489</sup>

yield responses generally follow similar functional forms across  
models, though with a large spread in magnitude likely due to  
the lack of calibration. Figure 7 illustrates inter-model diversity  
for a single crop and location (rain-fed maize in northern Iowa,  
also shown in Figure 6). Differences in response shape can lead  
to differences in the fidelity of emulation, though comparison  
here is complicated by the different sampling regimes across  
models. Note that models are most similar in their responses to  
temperature perturbations.

While the nitrogen dimension is important, it is also the most  
problematic to emulate in this work because of its limited sam-  
pling. The GGCMI Phase II protocol specified only three ni-  
trogen levels (10, 60 and 200  $\text{kg N y}^{-1} \text{ ha}^{-1}$ ), so a third-order  
fit would be over-determined but a second-order fit can result  
in potentially unphysical results. Steep and nonlinear declines  
in yield with lower nitrogen levels mean that some regressions  
imply a peak in yield between the 100 and 200  $\text{kg N y}^{-1} \text{ ha}^{-1}$   
levels. While it is possible that over-application of nitrogen at  
the wrong time in the growing season could lead to reduced  
yields, these features are potentially an artifact of under sam-  
pling. In addition, the polynomial fit cannot capture the well-  
documented saturation effect of nitrogen application (e.g. In-  
gestad, 1977) as accurately as would be possible with a non-  
parametric model.

The emulation fidelity demonstrated here is sufficient to al-  
low using emulated response surfaces to compare model re-  
sponses and derive insight about impacts projections. Because  
the emulator or “surrogate model” transforms the discrete sim-  
ulation sample space into a continuous response surface at any  
geographic scale, it can be used for a variety of applications,  
including construction of continuous damage functions. As an  
example, we show a damage function constructed from the 4D  
emulation, aggregated to global yield, with simulated values  
shown for comparison (Figure 8, which shows maize on cur-

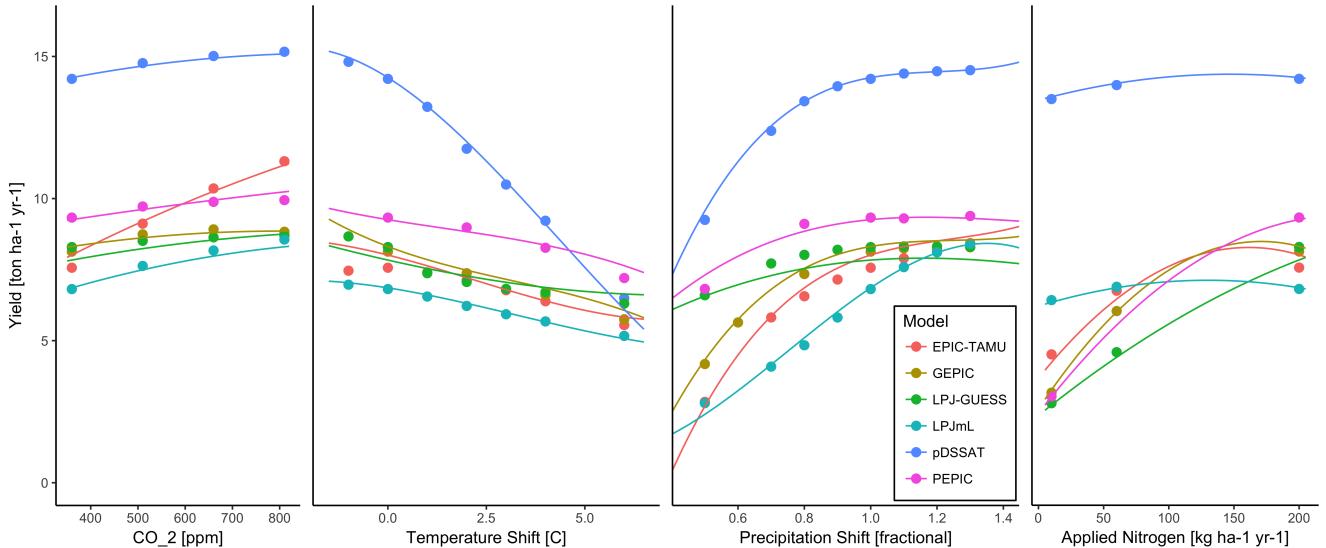


Figure 7: Illustration of across-model variations in yield response. Figures shows simulations and emulations from six models for rain-fed maize in the same Iowa grid cell shown in Figure 6, with the same plot conventions. Models that do not simulate the nitrogen dimension are omitted for clarity. Note that models are uncalibrated, increasing spread in absolute yields. While most model responses can readily emulated with a simple polynomial, some response surfaces are more complicated (e.g. LPJ-GUESS here) and lead to emulation error, though error generally remains small relative to inter-model uncertainty. For more detailed emulator assessment, see Appendix A. As in Figure 6, extrapolation out of the sample space is problematic.

rently cultivated land; see Figures S16- S19 for other crops and<sup>490</sup> factors ( $\text{CO}_2$ , temperature, precipitation, and applied nitrogen).  
 dimensions). The emulated values closely match simulations<sup>491</sup> Its global nature also allows identifying geographic shifts in  
 even at this aggregation level. Note that these functions are<sup>492</sup> high yield potential locations. We expect that the simulations  
 presented only as examples and do not represent true global<sup>493</sup> will yield multiple insights in future studies, and show here a  
 projections, because they are developed from simulation data<sup>494</sup> selection of preliminary results to illustrate their potential uses.  
 with a uniform temperature shift while increases in global mean<sup>495</sup>  
 temperature should manifest non-uniformly. The global cover-<sup>496</sup> First, the GGCMI Phase II simulations allow identifying ma-  
 age of the GGCMI Phase II simulations allows impacts mod-<sup>497</sup> jor areas of uncertainty. Across the major crops, inter-model  
 elers to apply arbitrary geographically-varying climate projec-<sup>498</sup> uncertainty is greatest for wheat and least for soy. Across fac-  
 tions, as well as arbitrary aggregation masks, to develop dam-<sup>499</sup> tors impacting yields, inter-model uncertainty is largest for  $\text{CO}_2$   
 age functions for any climate scenario and any geopolitical or<sup>500</sup> fertilization and nitrogen response effects. The  $\text{CO}_2$  response  
 geographic level.<sup>501</sup> is generally subject to large uncertainties (not evident in Fig-  
 ures 6 – 7 for maize as it is a C4 crop). All relevant  $\text{CO}_2$  pro-  
 cesses have not been studied in sufficient detail or have not been  
 implemented in models sufficiently (e.g. J. Boote et al., 2013)  
 and a broader experimental basis for model parameterization  
 is needed (Leakey et al., 2009). Efforts to improve model re-  
 sponse to  $\text{CO}_2$  are ongoing. Across geographic regions, projec-  
 tions are most uncertain in the high latitudes where yields may  
 increase, and most robust in low latitudes where yield impacts  
 are largest.

## 502 6. Conclusions and Discussion

The GGCMI Phase II experiment provides a database tar-<sup>503</sup>  
 geted to allow detailed study of crop yields from process-based<sup>504</sup>  
 models under climate change. The experiment is designed to<sup>505</sup>  
 facilitate not only comparing the sensitivities of process-based<sup>506</sup>  
 crop yield models to changing climate and management inputs<sup>507</sup>  
 but also evaluating the complex interactions between driving<sup>508</sup>

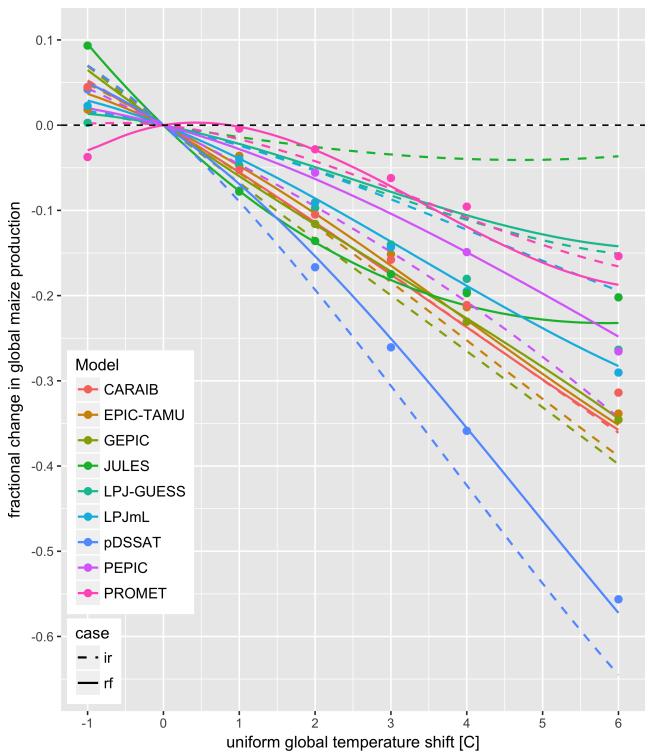


Figure 8: Global emulated damages for maize on currently cultivated lands<sup>556</sup> for the GGCMI Phase II models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating<sup>557</sup> up emulated values at the grid cell level, not from a regression of global mean yields.) Lines are emulations for rain-fed (solid) and irrigated (dashed) crops;<sup>558</sup> for comparison, dots are the simulated values for the rain-fed case. For most models, irrigated crops show a sharper reduction than do rain-fed because of the<sup>559</sup> locations of cultivated areas: irrigated crops tend to be grown in warmer areas where impacts are more severe for a given temperature shift. (The exceptions<sup>560</sup> are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures S16-S19 in the supplemental material.

<sup>542</sup> tion levels than are analogous non-irrigated crops, presumably  
<sup>543</sup> because those rain-fed crops are limited by water as well as  
<sup>544</sup> nitrogen availability (Figure S19). (Soy as an efficient atmo-  
<sup>545</sup> spheric nitrogen-fixer is relatively insensitive to nitrogen, and  
<sup>546</sup> rice is not generally grown in water-limited conditions).

<sup>547</sup> Most crops exhibit a somewhat uniform response to temper-  
<sup>548</sup> ature increase across different Köppen-Geiger when analyzed  
<sup>549</sup> over currently cultivated area (see Figure S16: i.e. equatorial  
<sup>550</sup> maize and ‘snow’ maize show similar response to a temperature  
<sup>551</sup> increase). This counterintuitive result agrees with existing liter-  
<sup>552</sup> ature including Rosenzweig et al. (2014) which shows increases  
<sup>553</sup> in yields mainly in regions where crops are not currently grown  
<sup>554</sup> and in Simona et al. (2014). A primary cause of this effect is  
<sup>555</sup> less difference in growing season temperature across Köppen-  
<sup>556</sup> Geiger regions when they are weighted by current cultivation  
<sup>557</sup> area than might be expected. Additionally, it has been proposed  
<sup>558</sup> that the growing season is shortened under warmer tempera-  
<sup>559</sup> tures in a way that is independent of baseline growing season  
<sup>560</sup> temperature (e.g. Wang et al., 2017, Rezaei et al., 2018). Cur-  
<sup>561</sup> rently most models in GGCMI include a direct linear shortening  
<sup>562</sup> of the growing season with warming, but uncertainty about the  
<sup>563</sup> exact nature of this response remains and it is an active area of  
<sup>564</sup> research.

<sup>529</sup> Second, the GGCMI Phase II simulations allow understand-<sup>563</sup>  
<sup>530</sup> ing the way that climate-driven changes and locations of cul-<sup>564</sup>  
<sup>531</sup> tivated land combine to produce yield impacts. One coun-<sup>565</sup>  
<sup>532</sup> terintuitive result immediate apparent is that irrigated maize<sup>566</sup>  
<sup>533</sup> shows steeper yield reductions under warming than does rain-<sup>567</sup>  
<sup>534</sup> fed maize when considered only over currently cultivated land.<sup>568</sup>  
<sup>535</sup> The effect results from geographic differences in cultivation. In<sup>569</sup>  
<sup>536</sup> any given location, irrigation increases crop resiliency to tem-<sup>570</sup>  
<sup>537</sup> perature increase, but irrigated maize is grown in warmer loca-<sup>571</sup>  
<sup>538</sup> tions where the impacts of warming are more severe (Figures<sup>572</sup>  
<sup>539</sup> S5-S6). The same behavior holds for rice and winter wheat,<sup>573</sup>  
<sup>540</sup> but not for soy or spring wheat (Figures S8-S10). Irrigated<sup>574</sup>  
<sup>541</sup> wheat and maize are also more sensitive to nitrogen fertiliza-<sup>575</sup>

Third, we show that even the relatively limited GGCMI Phase II sampling space allows emulation of the climatological response of crop models with a relatively simple reduced-form statistical model. The systematic parameter sampling in the GGCMI Phase II procedure provides information on the influence of multiple interacting factors in a way that single projections cannot, and emulating the resulting response surface then produces a tool that can aid in both physical interpretation of the process-based models and in assessment of agricultural impacts under arbitrary climate scenarios. Emulating the climatological response isolates long-term impacts from any confound-

576 ing factors that complicate year-over-year changes, and the use<sub>610</sub> level experimental data. The parameter space tested in GGCMI  
577 of simple functional forms offer the possibility of physical in-<sub>611</sub> Phase II will allow detailed investigations into yield variabil-  
578 interpretation of parameter values. We anticipate that systematic<sub>612</sub> ity and response to extremes under changing management and  
579 parameter sampling will become the norm in future crop model<sub>613</sub> CO<sub>2</sub> levels and allow the study of geographic shifts in opti-  
580 intercomparison exercises. <sub>614</sub> mal growing regions for different crops. The output dataset

581 While the GGCMI Phase II database should offer the foun-<sub>615</sub> also contains other runs and variables not analyzed or shown  
582 dation for multiple future studies, several cautions need to be<sub>616</sub> here. Runs include several which allowed adaptation to climate  
583 noted. Because the simulation protocol was designed to focus<sub>617</sub> changes by altering growing seasons, and additional variables  
584 on change in yield under climate perturbations and not on repli-<sub>618</sub> include above ground biomass, LAI, and root biomass (as many  
585 cating real-world yields, the models are not formally calibrated<sub>619</sub> as 25 output variables for some models). Emulation studies that  
586 so cannot be used for impacts projections unless in used in con-<sub>620</sub> are possible include a more systematic evaluation of different  
587 junction with historical data (or data products). Because the<sub>621</sub> statistical model specifications and formal calculation of uncer-  
588 GGCMI Phase II simulations apply uniform perturbations to<sub>622</sub> tainties in derived parameters.

589 historical climate inputs, they do not sample changes in higher<sub>623</sub> The development of multi-model ensembles such as GGCMI  
590 order moments, and cannot address the additional crop yield<sub>624</sub> Phase II provides a way to begin to better understand crop re-  
591 impacts of potential changes in climate variability. Although<sub>625</sub> sponses to a range of potential climate inputs, improve process  
592 distributional changes in model projections are fairly uncertain<sub>626</sub> based models, and explore the potential benefits of adaptive re-  
593 at present, follow-on experiments may wish to consider them.<sub>627</sub> sponses included shifting growing season, cultivar types and  
594 Several recent studies have described procedures for generating<sub>628</sub> cultivar geographic extent.

595 simulations that combine historical data with model projections  
596 of not only mean changes in temperature and precipitation but<sub>629</sub>  
597 changes in their marginal distributions or temporal dependence.<sub>630</sub>  
598 For methods to generate adjust historical climate data inclusive<sub>631</sub> **7. Acknowledgments**  
599 of distributional and temporal dependence changes, see Leeds<sub>632</sub>  
600 et al. (2015), Poppick et al. (2016), Chang et al. (2016) and<sub>633</sub>  
601 Haugen et al. (2018)). Emulation approaches are an area of ac-<sub>634</sub>  
602 tive ongoing study and one of the goals of the GGCMI Phase II<sub>635</sub>  
603 dataset is to facilitate these research efforts. <sub>636</sub>

604 The GGCMI Phase II output dataset invites a broad range<sub>637</sub>  
605 of potential future avenues of analysis. A major target area of<sub>638</sub>  
606 research is studying the models themselves including: a de-<sub>639</sub>  
607 tailed examination of interaction terms between the major in-<sub>640</sub>  
608 put drivers, a robust quantification of the sensitivity of differ-<sub>641</sub>  
609 ent models to the input drivers, and comparisons with field-<sub>642</sub>  
610 P. P.F. and K.W. were supported by the Newton Fund through

643 the Met Office Climate Science for Service Partnership Brazil<sup>676</sup>  
 644 (CSSP Brazil). A.S. was supported by the Office of Science<sup>677</sup>  
 645 of the U.S. Department of Energy as part of the Multi-sector<sup>678</sup>  
 646 Dynamics Research Program Area. Computing resources were<sup>679</sup>  
 647 provided by the University of Chicago Research Computing<sup>680</sup>  
 648 Center (RCC). S.O. acknowledges support from the Swedish<sup>681</sup>  
 649 strong research areas BECC and MERGE together with sup-<sup>682</sup>  
 650 port from LUCCI (Lund University Centre for studies of Car-<sup>683</sup>  
 651 bon Cycle and Climate Interactions).

## 652 8. Appendix A: Simulations – Assessment

653 The Müller et al. (2017) procedure evaluates response to  
 654 year-to-year temperature and precipitation variations in a con-<sup>684</sup>  
 655 trol run driven by historical climate and compares it to de-<sup>685</sup>  
 656 trended historical yields from the FAO (Food and Agriculture<sup>686</sup>  
 657 Organization of the United Nations, 2018) by calculating the<sup>687</sup>  
 658 Pearson correlation coefficient. The procedure offers no means  
 659 of assessing CO<sub>2</sub> fertilization, since CO<sub>2</sub> has been relatively  
 660 constant over the historical data collection period. Nitrogen in-<sup>688</sup>  
 661 troduces some uncertainty into the analysis, since the GGCMI  
 662 Phase II runs impose fixed, uniform nitrogen application levels  
 663 that are not realistic for individual countries. We evaluate up to<sup>689</sup>  
 664 three control runs for each model, since some modeling groups<sup>690</sup>  
 665 provide historical runs for three different nitrogen levels.<sup>691</sup>

666 Figure 9 shows the Pearson time series correlation between<sup>692</sup>  
 667 the simulation model yield and FOA yield data. Figure 9 can be<sup>693</sup>  
 668 compared to Figures 1,2,3,4 and 6 in Müller et al. (2017). The<sup>694</sup>  
 669 results are mixed, with many regions for rice and wheat be-<sup>695</sup>  
 670 ing difficult to model. No single model is dominant, with each<sup>696</sup>  
 671 model providing near best-in-class performance in at least one<sup>697</sup>  
 672 location-crop combination. The presence of very few vertical<sup>698</sup>  
 673 dark green color bars clearly illustrates the power of a multi-<sup>699</sup>  
 674 model intercomparison project like the one presented here. The<sup>700</sup>  
 675 ensemble mean does not beat the best model in each case, but<sup>701</sup>

643 shows positive correlation in over 75% of the cases presented  
 644 here. The EPIC-TAMU model performs best for soy, CARIAB,  
 645 EPIC-TAMU, and PEPIC perform best for maize, PROMET  
 646 performs best for wheat, and the EPIC family of models per-  
 647 form best for rice. [Reductions in skill over the performance](#)  
 648 [illustrated in Müller et al. \(2017\)](#) may be attributed to the nitro-  
 649 gen levels or lack of calibration in some models.<sup>702</sup>

650 Note that failure to reproduce year-to-year variability in the  
 651 FAO data product in some cases may not necessarily indicate  
 652 model failure as yield data in many areas in the developing  
 653 world are a level of abstraction from ground truth. The strik-  
 654 ing difference between model skill for India and Pakistan or  
 655 Bangladesh for rice must be in part attributable to this effect.  
 656 Additionally, there is less year-to-year variability in rice yields  
 657 (partially due to the fraction of irrigated cultivation). Since the  
 658 Pearson r metric is scale invariant, it will tend to score the rice  
 659 models more poorly than maize and soy.<sup>703</sup>

## 660 9. Appendix B: Emulation – Assessment

661 Because no general criteria exist for defining an acceptable  
 662 crop model emulator, we utilize a metric of emulator perfor-  
 663 mance specific to GGCMI Phase II. For a multi-model com-  
 664 parison exercise like GGCMI Phase II, one reasonable criterion  
 665 is what we term the “normalized error”, which compares the fi-  
 666 delity of an emulator for a given model and scenario to the inter-  
 667 model uncertainty. We define the normalized error  $e$  for each  
 668 scenario as the difference between the fractional yield change  
 669 from the emulator and that in the original simulation, divided  
 670 by the standard deviation of the multi-model spread (Equations  
 671 2 and 3):

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

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

708 absolute emulated or simulated mean yields. The normalized  
 709 error  $e$  is the difference between the emulated fractional change  
 710 in yield and that actually simulated, normalized by  $\sigma_{sim}$ , the  
 711 standard deviation in simulated fractional yields  $F_{sim, scn}$  across  
 712 all models. The emulator is fit across all available simulation  
 713

705 Here  $F_{scn}$  is the fractional change in a model's mean emu-711  
 706 lated or simulated yield from a defined baseline, in some sce-712  
 707 nario (scn.) in C, T, W, and N space;  $Y_{scn}$  and  $Y_{baseline}$  are the

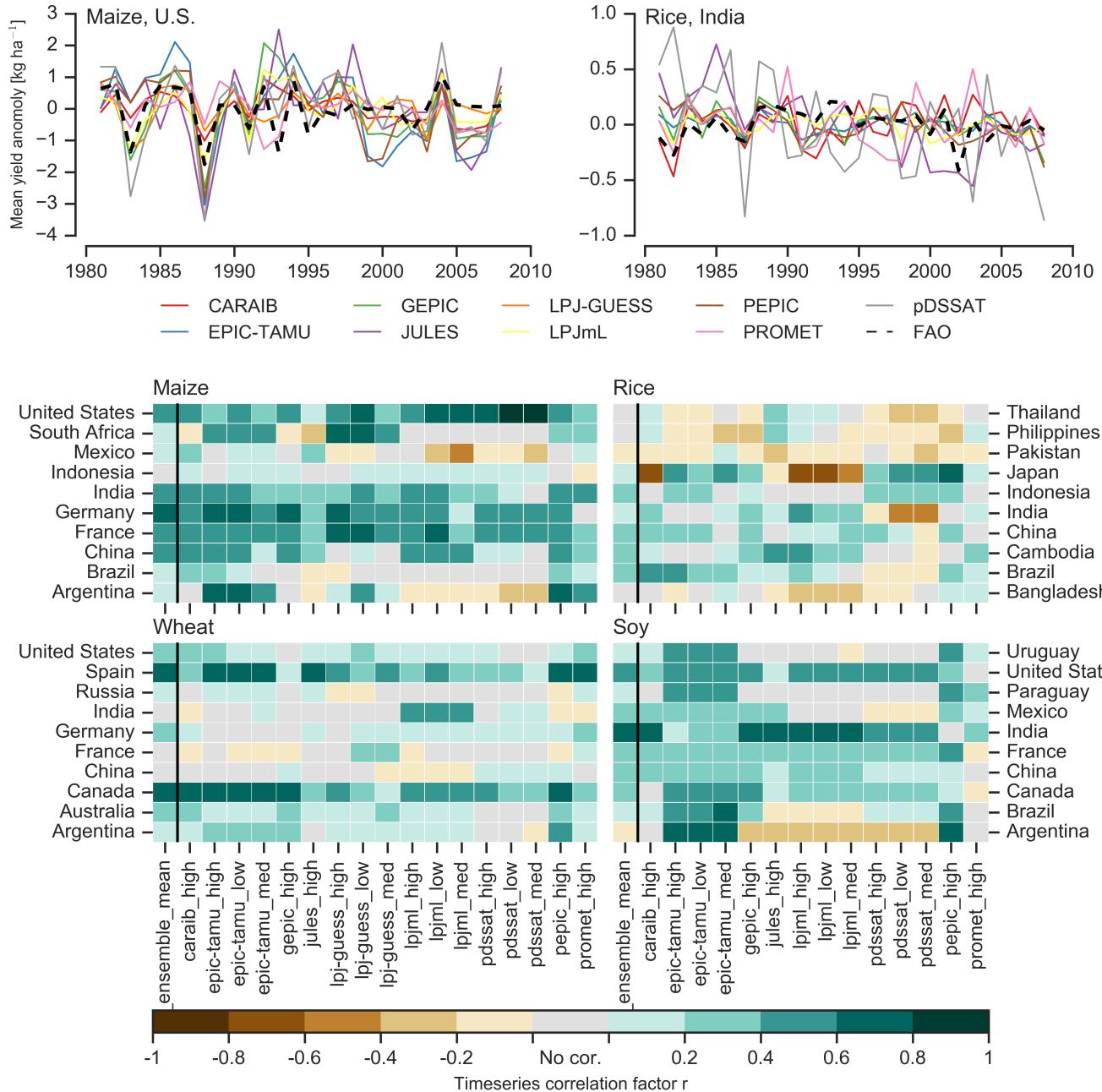


Figure 9: Time series of correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen application case. The heatmaps illustrate the Pearson  $r$  correlation coefficient between the detrended simulation mean yield at the country level compared to the detrended FAO yield data for the top producing countries for each crop with continuous FAO data over the 1980-2010 period. Models that provided different nitrogen application levels are shown with low, med, and high label (models that did not simulate different nitrogen levels are analogous to a high nitrogen application level). The ensemble mean yield is also correlated with the FAO data (not the mean of the correlations). Wheat contains both spring wheat and winter wheat simulations.

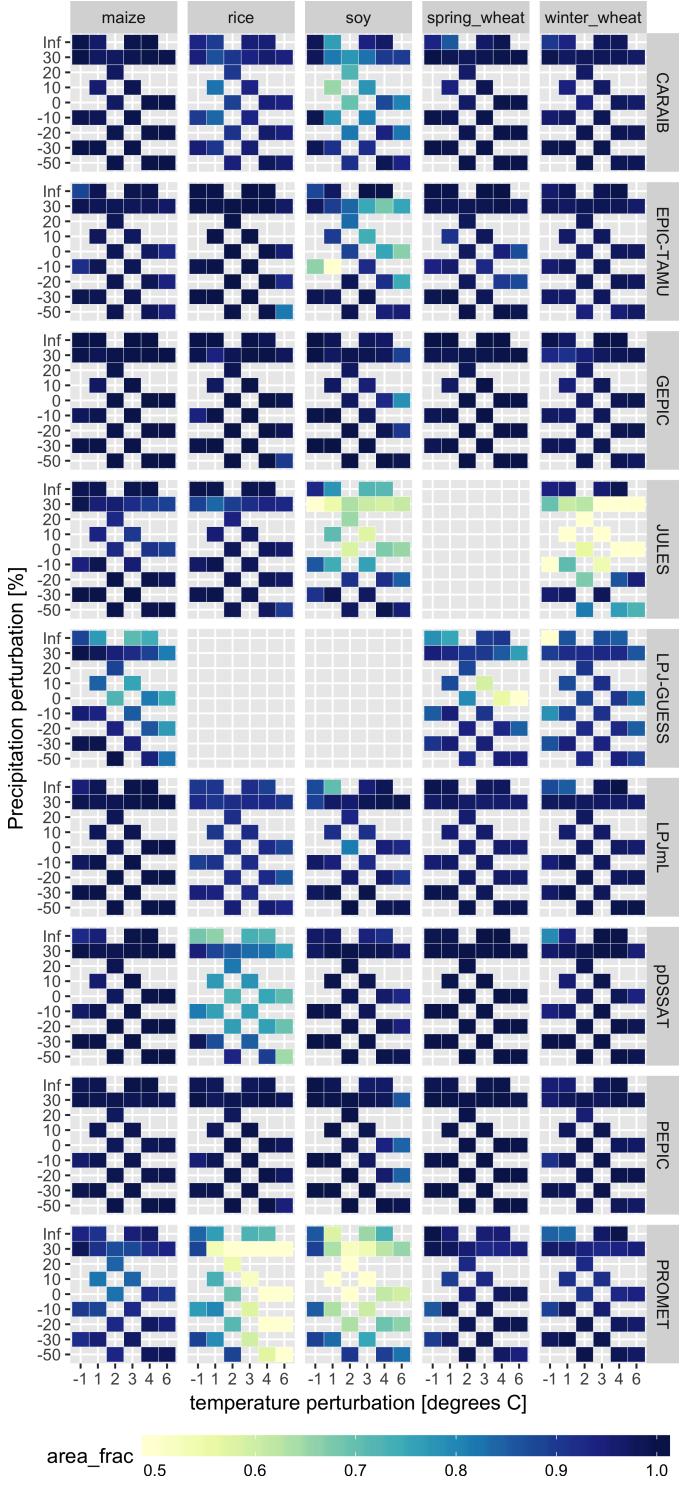


Figure 10: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO<sub>2</sub> and nitrogen held fixed at baseline values. Large columns are crops and large rows models; squares within are T,P scenario pairs. Colors denote the fraction of currently cultivated hectares ('area frac') for each crop with normalized area  $e$  less than 1 indicating the the error between the emulation and simulation less than one standard deviation of the ensemble simulation spread. Of the 756 scenarios with these CO<sub>2</sub> and N values, we consider only those for which all 9 models submitted data. JULES did not simulate rice and soy. Emulator performance is generally satisfactory, with some exceptions. Emulator failures (significant areas of poor performance) occur for individual crop-model combinations, with performance generally degrading for hotter and wetter scenarios.

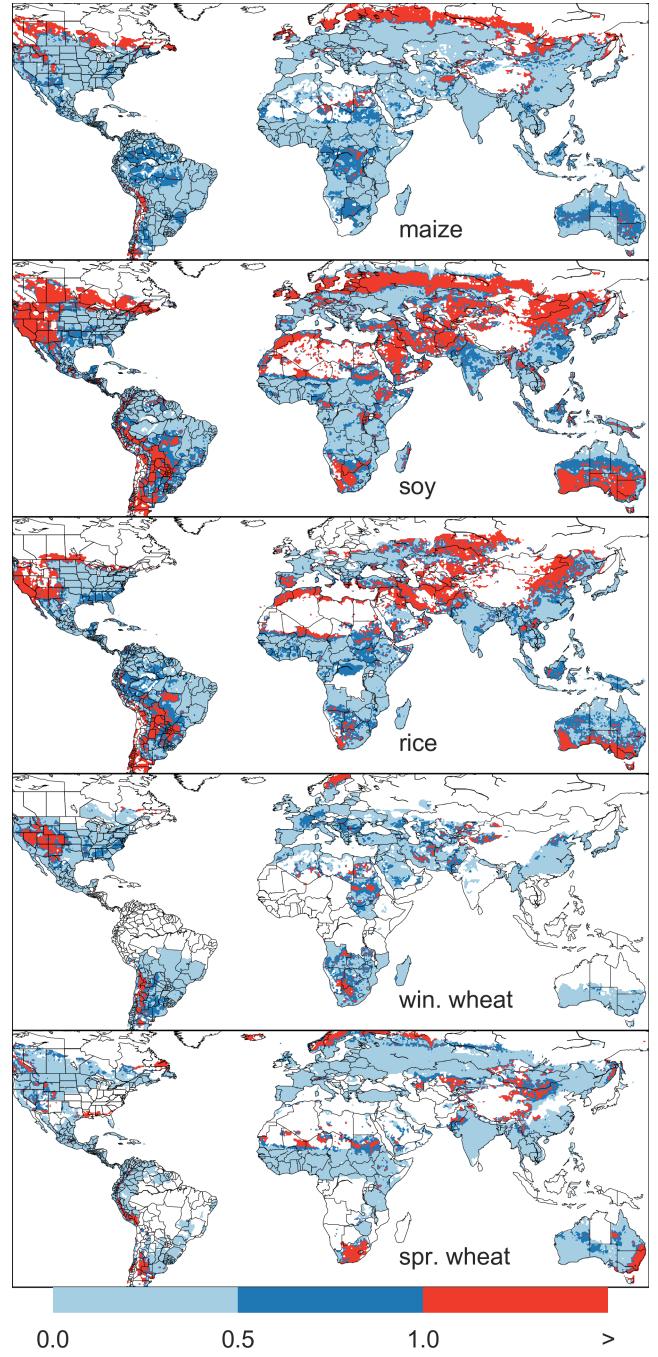


Figure 11: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error  $e$ , where  $e > 1$  means that emulator error exceeds the multi-model standard deviation. White areas are those where crops are not simulated by this model. Models differ in their areas omitted, meaning the number of samples used to calculate the multi-model standard deviation is not spatially consistent in all locations. Emulator performance is generally good relative to model spread in areas where crops are currently cultivated (compare to Figure 1) and in temperate zones in general; emulation issues occur primarily in marginal areas with low yield potentials. For CARAIB, emulation of soy is more problematic, as was also shown in Figure 10.

outputs, and then the error is calculated across the simulation<sup>747</sup> scenarios provided by all nine models (Figure 10 and Figures<sup>748</sup> S12 and Figures S13 in supplemental documents).<sup>749</sup>

To assess the ability of the polynomial emulation to capture<sup>750</sup> the behavior of complex process-based models, we evaluate the<sup>751</sup> normalized emulator error. That is, for each grid cell, model,<sup>752</sup> and scenario we evaluate the difference between the model yield<sup>753</sup> and its emulation, normalized by the inter-model standard de-<sup>754</sup> viation in yield projections. This metric implies that emulation<sup>755</sup> is generally satisfactory, with several distinct exceptions. Al-<sup>756</sup> most all model-crop combination emulators have normalized<sup>757</sup> errors less than one over nearly all currently cultivated hectares<sup>758</sup> (Figure 10), but some individual model-crop combinations are<sup>759</sup> problematic (e.g. PROMET for rice and soy, JULES for soy<sup>760</sup> and winter wheat, Figures S14–S15). Normalized errors for soy<sup>761</sup>

are somewhat higher across all models not because emulator fi-

delity is worse but because models agree more closely on yield<sup>762</sup> changes for soy than for other crops (see Figure S16, lowering<sup>763</sup> the denominator. Emulator performance often degrades in geo-<sup>764</sup> graphic locations where crops are not currently cultivated. Fig-<sup>765</sup> ure 11 shows a CARAIB case as an example, where emulator<sup>766</sup> performance is satisfactory over cultivated areas for all crops<sup>767</sup> other than soy, but uncultivated regions show some problematic<sup>768</sup> areas.<sup>769</sup>

The normalized error  $e$  for a model depends not only on the<sup>770</sup> fidelity of its emulator in reproducing a given simulation but on<sup>771</sup> the particular suite of models considered in the intercomparison<sup>772</sup> exercise. The rationale for this choice is to relate the fidelity of<sup>773</sup> the emulation to an estimate of true uncertainty, which we take<sup>774</sup> as the multi-model spread. Because the inter-model spread is<sup>775</sup> large, normalized errors tend to be small. That is, any failures<sup>776</sup> of emulation are small relative to inter-model uncertainty. We<sup>777</sup> therefore do not provide a formal parameter uncertainty analy-<sup>778</sup> sis, but note that the GGCMI Phase II dataset is well-suited to<sup>779</sup>

statistical exploration of emulation approaches and quantifica-<sup>780</sup> tion of emulator fidelity.<sup>781</sup>

It should be noted that this assessment metric is relatively<sup>782</sup> forgiving. First, each emulation is evaluated against the simu-<sup>783</sup>lation actually used to train the emulator. Had we used a spline<sup>784</sup> interpolation the error would necessarily be zero. Second, the<sup>785</sup> performance metric scales emulator fidelity not by the magni-<sup>786</sup>tude of yield changes but by the inter-model spread in those<sup>787</sup> changes. Where models differ more widely, the standard for<sup>788</sup> emulators becomes less stringent. Because models disagree on<sup>789</sup> the magnitude of CO<sub>2</sub> fertilization, this effect is readily seen<sup>790</sup> when comparing assessments of emulator performance in sim-<sup>791</sup>ulations at baseline CO<sub>2</sub> (Figure 10) with those at higher CO<sub>2</sub><sup>792</sup> levels (Figure S13). Widening the inter-model spread leads to<sup>793</sup> an apparent increase in emulator skill.<sup>794</sup>

## 10. References

- Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013). Implication of crop model calibration strategies for assessing regional impacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.
- Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D., A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Alderman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Biermann, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising temperatures reduce global wheat production. *Nature Climate Change*, 5, 143–147. doi:10.1038/nclimate2470.
- Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biermann, C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, 3, 827832. doi:10.1038/nclimate1916.
- Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Efficiency, Carbon Sequestration, and Environmental Pollution. *Advances in Agronomy*, 86, 341 – 409.
- Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov, N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat

- 784 production potentials and management flexibility under the representative<sup>827</sup>  
 785 concentration pathways. *Global and Planetary Change*, 122, 107 – 121. 828  
 786 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields<sup>829</sup>  
 787 from global gridded crop models. *Agricultural and Forest Meteorology*, 236,<sup>830</sup>  
 788 145 – 161. 831  
 789 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded<sup>832</sup>  
 790 crop models using statistical estimates. *Agricultural and Forest Meteorol-*<sup>833</sup>  
 791 *ogy*, 214-215, 134 – 147. 834  
 792 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S.<sup>835</sup>  
 793 (2018). Implementing the Nitrogen cycle into the dynamic global vegeta-<sup>836</sup>  
 794 tion, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific*<sup>837</sup>  
 795 *Model Development*, 11, 2789–2812. 838  
 796 Calvin, K., Patel, P., Clarke, L., Asrar, G., Bond-Lamberty, B., Cui, R. Y.,<sup>839</sup>  
 797 Di Vittorio, A., Dorheim, K., Edmonds, J., Hartin, C., Hejazi, M., Horowitz,<sup>840</sup>  
 798 R., Iyer, G., Kyle, P., Kim, S., Link, R., McJeon, H., Smith, S. J., Snyder,<sup>841</sup>  
 799 A., Waldhoff, S., & Wise, M. (2019). Gcam v5.1: representing the linkages<sup>842</sup>  
 800 between energy, water, land, climate, and economic systems. *Geoscientific*<sup>843</sup>  
 801 *Model Development*, 12, 677–698. doi:10.5194/gmd-12-677-2019. 844  
 802 Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L.,<sup>845</sup>  
 803 & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections<sup>846</sup>  
 804 Based on Precomputed GCM Runs. *Journal of Climate*, 27, 1829–1844. 847  
 805 Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.<sup>848</sup>  
 806 (2014). A meta-analysis of crop yield under climate change and adaptation.<sup>849</sup>  
 807 *Nature Climate Change*, 4, 287 – 291. 850  
 808 Chang, W., Stein, M., Wang, J., Kotamarthi, V., & Moyer, E. (2016). Changes in<sup>851</sup>  
 809 spatio-temporal precipitation patterns in changing climate conditions. *Jour-*<sup>852</sup>  
 810 *nal of Climate*, 29. doi:10.1175/JCLI-D-15-0844.1. 853  
 811 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process<sup>854</sup>  
 812 emulation of dynamic computer codes. *Biometrika*, 96, 663–676. 855  
 813 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In<sup>856</sup>  
 814 C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth*,<sup>857</sup>  
 815 *Duke University, Durham, North Carolina* (pp. 115–118). Durham, North<sup>858</sup>  
 816 Carolina. 859  
 817 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for<sup>860</sup>  
 818 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205). 861  
 819 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdous,<sup>862</sup>  
 820 M., & François, L. (2011). Responses of European forest ecosystems to<sup>863</sup>  
 821 21st century climate: assessing changes in interannual variability and fire<sup>864</sup>  
 822 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99). 865  
 823 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,<sup>866</sup>  
 824 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating<sup>867</sup>  
 825 impact models and sectors (pSIMS). *Environmental Modelling and Soft-*<sup>868</sup>  
 826 *ware*, 62, 509–516. 869  
 827 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,  
 828 Büchner, M., Foster, I., Glotter, M., Heinke, J., Izumi, T., Izaurrealde, R. C.,  
 829 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.  
 830 (2015). The Global Gridded Crop Model Intercomparison: data and mod-  
 831 eling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8,  
 832 261–277.  
 833 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J.,  
 834 & Taylor, K. E. (2016). Overview of the coupled model intercomparison  
 835 project phase 6 (cmip6) experimental design and organization. *Geoscientific*  
 836 *Model Development*, 9, 1937–1958.  
 837 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-  
 838 mate change impacts on durum wheat in the mediterranean region. *Natural*  
 839 *Hazards and Earth System Sciences*, 11, 1293–1302.  
 840 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Re-  
 841 gionalization of a large-scale crop growth model for sub-Saharan Africa:  
 842 Model setup, evaluation, and estimation of maize yields. *Agriculture,*  
 843 *Ecosystems & Environment*, 151, 21 – 33.  
 844 Food and Agriculture Organization of the United Nations (2018). FAOSTAT  
 845 database. URL: <http://www.fao.org/faostat/en/home>.  
 846 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T.,  
 847 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P.,  
 848 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F.,  
 849 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,  
 850 Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J.,  
 851 Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nen-  
 852 del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna,  
 853 M., Semenov, M. A., Slawinski, C., Strattonovich, P., Supit, I., Waha, K.,  
 854 Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model  
 855 wheat yield impact response surfaces showing sensitivity to temperature and  
 856 precipitation change. *Agricultural Systems*, 159, 209–224.  
 857 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014).  
 858 Evaluating the utility of dynamical downscaling in agricultural impacts pro-  
 859 jections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781.  
 860 Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitiv-  
 861 ity of Agricultural Model Performance to Different Climate Inputs. *Journal*  
 862 *of Applied Meteorology and Climatology*, 55, 151113145618001.  
 863 Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported  
 864 Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous  
 865 Crop Growth and Yield: Application for Wheat in Central Europe. *Remote*  
 866 *Sensing*, 7, 3934–3965.  
 867 Haugen, M., Stein, M., Moyer, E., & Srivastava, R. (2018). Estimating changes in  
 868 temperature distributions in a large ensemble of climate simulations using  
 869 quantile regression. *Journal of Climate*, 31, 8573–8588. doi:10.1175/JCLI-

- 870 D-17-0782.1.
- 871 He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H.,<sup>914</sup>
- 872 He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water<sup>915</sup>
- 873 contents and nitrogen leaching to precipitation, management practices and<sup>916</sup>
- 874 soil hydraulic properties in semi-arid and humid regions of Canada using the<sup>917</sup>
- 875 DSSAT model. *Nutrient Cycling in Agroecosystems*, *106*, 201–215. <sup>918</sup>
- 876 Heady, E. O. (1957). An Econometric Investigation of the Technology of Agri-<sup>919</sup>
- 877 cultural Production Functions. *Econometrica*, *25*, 249–268. <sup>920</sup>
- 878 Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa<sup>921</sup>
- 879 State University Press. <sup>922</sup>
- 880 Holden, P., Edwards, N., PH, G., Fraedrich, K., Lunkeit, F., E, K., Labriet,<sup>923</sup>
- 881 M., Kanudia, A., & F, B. (2014). Plasim-entsem v1.0: A spatiotemporal<sup>924</sup>
- 882 emulator of future climate change for impacts assessment. *Geoscientific<sup>925</sup>*
- 883 Model Development
- 884 Holzkämper, A., Calanca, P., & Fuhrer, J. (2012). Statistical crop models:<sup>927</sup>
- 885 Predicting the effects of temperature and precipitation changes. *Climate<sup>928</sup>*
- 886 Research
- 887 Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I.,<sup>930</sup>
- 888 McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore,<sup>931</sup>
- 889 A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake,<sup>932</sup>
- 890 A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgliesh,<sup>933</sup>
- 891 N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp,<sup>934</sup>
- 892 J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson,<sup>935</sup>
- 893 M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T.,<sup>936</sup>
- 894 Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorf, J.,<sup>937</sup>
- 895 Wedgwood, S., & Keating, B. A. (2014). APSIM Evolution towards a new<sup>938</sup>
- 896 generation of agricultural systems simulation. *Environmental Modelling and<sup>939</sup>*
- 897 Software
- 898 Howden, S., & Crimp, S. (2005). Assessing dangerous climate change impacts<sup>941</sup>
- 899 on australia's wheat industry. *Modelling and Simulation Society of Australia<sup>942</sup>*
- 900 and New Zealand
- 901 Iizumi, T., Nishimori, M., & Yokozawa, M. (2010). Diagnostics of climate<sup>944</sup>
- 902 model biases in summer temperature and warm-season insolation for the<sup>945</sup>
- 903 simulation of regional paddy rice yield in japan. *Journal of Applied Meteo-<sup>946</sup>*
- 904 rology and Climatology
- 905 Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Ni-<sup>948</sup>
- 906 trogen Fertilizers. *Ambio*, *6*, 146–151. <sup>949</sup>
- 907 Izaurralde, R., Williams, J., McGill, W., Rosenberg, N., & Quiroga Jakas, M.<sup>950</sup>
- 908 (2006). Simulating soil C dynamics with EPIC: Model description and test-<sup>951</sup>
- 909 ing against long-term data. *Ecological Modelling*, *192*, 362–384. <sup>952</sup>
- 910 J. Boote, K., Jones, J., White, J., Asseng, S., & Lizaso, J. (2013). Putting<sup>953</sup>
- 911 mechanisms into crop production models. *Plant, cell environment*, *36*.<sup>954</sup>
- 912 doi:10.1111/pce.12119. <sup>955</sup>
- 913 Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. *Agriculture, Ecosystems & Environment*, *93*, 73 – 85.
- 914 Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijsman, A., & Ritchie, J. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, *18*, 235 – 265.
- 915 Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural Systems*, *155*, 269 – 288.
- 916 Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K., Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, *18*, 267 – 288.
- 917 Leakey, A. D. B., Bernacchi, C. J., Ainsworth, E. A., Ort, D. R., Long, S. P., & Rogers, A. (2009). Elevated CO<sub>2</sub> effects on plant carbon, nitrogen, and water relations: six important lessons from FACE. *Journal of Experimental Botany*, *60*, 2859–2876. doi:10.1093/jxb/erp096.
- 918 Leeds, W. B., Moyer, E. J., & Stein, M. L. (2015). Simulation of future climate under changing temporal covariance structures. *Advances in Statistical Climatology, Meteorology and Oceanography*, *1*, 1–14. URL: <https://www.adv-stat-clim-meteorol-oceanogr.net/1/1/2015/>. doi:10.5194/ascmo-1-1-2015.
- 919 Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., & Smith, B. (2013). Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa. *Earth System Dynamics*, *4*, 385–407.
- 920 Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling wheat yield and crop water productivity with high resolution on a global scale. *Agricultural Systems*, *94*, 478 – 493.
- 921 Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a). Global investigation of impacts of PET methods on simulating crop-water relations for maize. *Agricultural and Forest Meteorology*, *221*, 164 – 175.
- 922 Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs with yields from major crop cultivations. *Science of The Total Environment*, *572*, 526 – 537.
- 923 Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, *150*, 1443 – 1452.

- 956 Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relation-999  
 957 ships and the impacts of recent warming. *Environmental Research Letters*, 1000  
 958 2, 014002. 1001
- 959 MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447<sub>1002</sub>
- 960 Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P.<sub>1003</sub>  
 961 Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi<sub>1004</sub>  
 962 Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano<sub>1005</sub>  
 963 D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of<sub>1006</sub>  
 964 large simulated yield datasets for studying climate effects. (p. 1100)<sub>1007</sub>  
 965 doi:10.13140/RG.2.1.5173.8328. 1008
- 966 Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrolog<sub>1009</sub>  
 967 ical modelling to study the impact of climate change on the water flows of<sub>1010</sub>  
 968 mountain watersheds. *Journal of Hydrology*, 376, 362 – 377. 1011
- 969 Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B.<sub>1012</sub>  
 970 & Calzadilla, A. (2009). Global biomass production potentials exceed ex<sub>1013</sub>  
 971 pected future demand without the need for cropland expansion. *Nature Com<sub>1014</sub>  
 972 munications*, 6. 1015
- 973 McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srini<sub>1016</sub>  
 974 vasa, C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S.<sub>1017</sub>  
 975 (2015). Integrated assessments of the impacts of climate change on agricult<sub>1018</sub>  
 976 ture: An overview of AgMIP regional research in South Asia. *Chapter in<sub>1019</sub>  
 977 Handbook of Climate Change and Agroecosystems*, (pp. 201–218). 1020
- 978 Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical<sub>1021</sub>  
 979 weather responsiveness of crop yields: US evidence and implications for<sub>1022</sub>  
 980 the agricultural impacts of climate change. *Environmental Research Letters*, 1023  
 981 12. 1024
- 982 Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate<sub>1025</sub>  
 983 change impacts on agriculture implies higher social cost of carbon. *Nature<sub>1026</sub>  
 984 Communications*, 8. 1027
- 985 Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais<sub>1028</sub>  
 986 P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde<sub>1029</sub>  
 987 R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T.<sub>1030</sub>  
 988 A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G.<sub>1031</sub>  
 989 Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H.<sub>1032</sub>  
 990 (2017). Global gridded crop model evaluation: benchmarking, skills, de<sub>1033</sub>  
 991 ficiencies and implications. *Geoscientific Model Development*, 10, 1403<sub>1034</sub>  
 992 1422. 1035
- 993 Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T.<sub>1036</sub>  
 994 (1997). Effect of CO<sub>2</sub> enrichment on carbon and nitrogen interaction in<sub>1037</sub>  
 995 wheat and soybean. *Soil Science and Plant Nutrition*, 43, 789–798. 1038
- 996 O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial<sub>1039</sub>  
 997 *Reliability Engineering & System Safety*, 91, 1290 – 1300. 1040
- 998 Olin, S., Schurges, G., Lindeskog, M., Wårlind, D., Smith, B., Bodin, P.<sub>1041</sub>
- Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue  
 C:N to changes in atmospheric CO<sub>2</sub> and N management in the main wheat  
 regions of western europe. *Biogeosciences*, 12, 2489–2515. doi:10.5194/bg-  
 12-2489-2015.
- Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in  
 field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.
- Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., &  
 Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK  
 Land Environment Simulator. *Geoscientific Model Development*, 8, 1139–  
 1155.
- Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop  
 yields and their variability at different levels of global warming. *Earth Sys-  
 tem Dynamics*, 9, 479–496.
- Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff,  
 S., & Gerten, D. (2015). Emulating global climate change impacts on crop  
 yields. *Statistical Modelling*, 15, 499–525.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,  
 Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas-  
 sos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011).  
 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning  
 Research*, 12, 2825–2830.
- Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T.,  
 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski,  
 P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.,  
 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,  
 Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet,  
 J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C.,  
 Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M.,  
 Semenov, M., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang,  
 E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation  
 effects on wheat yield across a European transect: a crop model ensemble  
 analysis using impact response surfaces. *Climate Research*, 65, 87–105.
- Poppick, A., McInerney, D. J., Moyer, E. J., & Stein, M. L. (2016). Temper-  
 atures in transient climates: Improved methods for simulations with  
 evolving temporal covariances. *Ann. Appl. Stat.*, 10, 477–505. URL:  
<https://doi.org/10.1214/16-AOAS903>. doi:10.1214/16-AOAS903.
- Porter et al. (IPCC) (2014). Food security and food production systems. Cli-  
 mate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global  
 and Sectoral Aspects. Contribution of Working Group II to the Fifth Assess-  
 ment Report of the Intergovernmental Panel on Climate Change. In C. F.  
 et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK:  
 Cambridge University Press.
- Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of

- 1042 monthly growing areas of 26 irrigated crops. 1085
- 1043 Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly 1086  
1044 Irrigated and Rainfed crop Areas around the Year 2000: A New High+087  
1045 Resolution Data Set for Agricultural and Hydrological Modeling. *Global 1088  
1046 Biogeochemical Cycles*, 24, GB1011. 1089
- 1047 Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E.+090  
1048 & Arneth, A. (2016). Climate analogues suggest limited potential for inten+091  
1049 sification of production on current croplands under climate change. *Nature 1092  
1050 Communications*, 7, 12608. 1093
- 1051 Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term cli+094  
1052 mate change based on a resampling ensemble technique. *Tellus A: Dynamic 1095  
1053 Meteorology and Oceanography*, 58, 461–472. 1096
- 1054 Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the+097  
1055 reduction and sensitivity analysis of complex environmental models. *Envi+098  
1056 ronmental Modelling & Software*, 34, 1 – 4. 1099
- 1057 Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling+100  
1058 in water resources. *Water Resources Research*, 48. 1101
- 1059 Rezaei, E., Siebert, S., Hging, H., & Ewert, F. (2018). Climate change effect on+102  
1060 wheat phenology depends on cultivar change. *Scientific Reports*, 8, 4891103  
1061 doi:10.1038/s41598-018-23101-2. 1104
- 1062 Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017)+105  
1063 Comparing and combining process-based crop models and statistical models+106  
1064 with some implications for climate change. *Environmental Research Letters*+107  
1065 12. 1108
- 1066 Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A.+109  
1067 Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek+110  
1068 F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014)+111  
1069 Assessing agricultural risks of climate change in the 21st century in a global+112  
1070 gridded crop model intercomparison. *Proceedings of the National Academy+113  
1071 of Sciences*, 111, 3268–3273. 1114
- 1072 Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P.+115  
1073 Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ew+116  
1074 ert, F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural+117  
1075 Model Intercomparison and Improvement Project (AgMIP): Protocols and+118  
1076 pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182. 1119
- 1077 Rosenzweig, C., Ruane, A. C., Antle, J., Elliott, J., Ashfaq, M., Chatta+120  
1078 A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G.+121  
1079 Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M.+122  
1080 Müller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M.+123  
1081 Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., & Wiebe, K.+124  
1082 (2018). Coordinating AgMIP data and models across global and regional+125  
1083 scales for 1.5°C and 2.0°C assessments. *Philosophical Transactions of the+126  
1084 Royal Society of London A: Mathematical, Physical and Engineering Sci+127  
1085 ences*, 376.
- 1086 Ruane, A., I. Hudson, N., Asseng, S., Camarrano, D., Ewert, F., Martre, P.,  
1087 J. Boote, K., Thorburn, P., Aggarwal, P., Angulo, C., Basso, B., Bertuzzi,  
1088 P., Biernath, C., Brisson, N., Challinor, A., Doltra, J., Gayler, S., Goldberg,  
1089 R., Grant, R., & Wolf, J. (2016). Multi-wheat-model ensemble responses to  
1090 interannual climate variability. *Environmental Modelling and Software*, 81,  
1091 86–101.
- 1092 Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason-  
1093 D'Croz, D., Müller, C., Porter, C., Phillips, M. M., Raymundo, R. M.,  
1094 Sands, R., Valdivia, R. O., White, J. W., Wiebe, K., & Rosenzweig, C.  
1095 (2018). Biophysical and economic implications for agriculture of +1.5° and  
1096 +2.0°C global warming using AgMIP Coordinated Global and Regional As-  
1097 sessments. *Climate Research*, 76, 17–39.
- 1098 Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown,  
1099 D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013).  
1100 Climate change impact uncertainties for maize in panama: Farm informa-  
1101 tion, climate projections, and yield sensitivities. *Agricultural and Forest  
1102 Meteorology*, 170, 132 – 145.
- 1103 Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forc-  
1104 ing datasets for agricultural modeling: Merged products for gap-filling and  
1105 historical climate series estimation. *Agric. Forest Meteorol.*, 200, 233–248.
- 1106 Ruane, A. C., McDermid, S., Rosenzweig, C., Baigorria, G. A., Jones, J. W.,  
1107 Romero, C. C., & Cecil, L. D. (2014). Carbon-temperature-water change  
1108 analysis for peanut production under climate change: A prototype for the  
1109 agmip coordinated climate-crop modeling project (c3mp). *Glob. Change  
1110 Biol.*, 20, 394–407. doi:10.1111/gcb.12412.
- 1111 Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901-  
1112 2100 depicted by world maps of the Köppen-Geiger climate classification.  
1113 *Meteorologische Zeitschrift*, 19, 135–141.
- 1114 Ruiz-Ramos, M., Ferrise, R., Rodrguez, A., Lorite, I., Bindi, M., Carter, T.,  
1115 Fronzek, S., Palosuo, T., Pirttioja, N., Baranowski, P., Buis, S., Cam-  
1116 marano, D., Chen, Y., Dumont, B., Ewert, F., Gaiser, T., Hlavinka, P.,  
1117 Hoffmann, H., Hhn, J., Jurecka, F., Kersebaum, K., Krzyszczak, J., Lana,  
1118 M., Mechiche-Alami, A., Minet, J., Montesino, M., Nendel, C., Porter,  
1119 J., Ruget, F., Semenov, M., Steinmetz, Z., Strattonovich, P., Supit, I.,  
1120 Tao, F., Trnka, M., de Wit, A., & Ritter, R. (2018). Adaptation re-  
1121 sponse surfaces for managing wheat under perturbed climate and co2 in  
1122 a mediterranean environment. *Agricultural Systems*, 159, 260 – 274.  
1123 doi:doi.org/10.1016/j.agysy.2017.01.009.
- 1124 Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting  
1125 dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19,  
1126 607–620.
- 1127 Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng,

- 1128 D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frieler, K. (2017). Consistent negative response of US crops to high temperatures in observations and crop models. *Nature Communications*, 8, 13931.
- 1129 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.
- 1130 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.
- 1131 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.
- 1132 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Landbouwkundige Onderzoeken* : 64.6, .
- 1133 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.
- 1134 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., Mller, 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.
- 1135 Simona, B., Nadine, B., Jean-Louis, D., Kenneth, B., Jon, L., W., J. J., Cynthia, R., C., R. A., Myriam, A., Christian, B., Bruno, B., Christian, B., Hendrik, B., Sjaak, C., Marc, C., Delphine, D., Giacomo, S., Sebastian, G., Patricio, G., Jerry, H., Steven, H., Cesar, I., Raymond, J., R., K. A., Christian, K. K., Soo-Hyung, K., S., K. N., David, M., Christoph, M., Claas, N., Eckart, P., Virginia, P. M., Federico, S., Iurii, S., Fulu, T., Edmar, T., Dennis, T., & Katharina, W. (2014). How do various maize crop models vary in their responses to climate change factor? *Global Change Biology*, 20, 2301–2320.
- 1136 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.
- 1137 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.
- 1138 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.
- 1139 Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters*, 35.
- 1140 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.
- 1141 Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Ritter, R. P., A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Alderman, P., Aggarwal, P., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A., De Sanctis, G., & Asseng, S. (2017). The uncertainty of crop yield projections is reduced by improved temperature response functions. *Nature Plants*, 3, 17102. doi:10.1038/nplants.2017.102.
- 1142 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.
- 1143 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.
- 1144 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.
- 1145 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.
- 1146 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Landbouwkundige Onderzoeken* : 64.6, .
- 1147 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.
- 1148 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., Mller, 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.