

# 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 reductionss 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.

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

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

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

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

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

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

67 developing world, where future climate impacts may be the<sup>101</sup>  
68 most critical. Finally, only process-based models can capture<sup>102</sup>  
69 the growth response to novel conditions and practices that are<sup>103</sup>  
70 not represented in historical data (e.g. Pugh et al., 2016, Roberts<sup>104</sup>  
71 et al., 2017). These novel changes can include the direct fertil-<sup>105</sup>  
72 ization effect of elevated CO<sub>2</sub>, or changes in management prac-<sup>106</sup>  
73 tices that may ameliorate climate-induced damages.<sup>107</sup>

74 Interest has been rising in statistical emulation, which al-<sup>108</sup>  
75 lows combining advantageous features of both statistical and<sup>109</sup>  
76 process-based models. The approach involves constructing a<sup>110</sup>  
77 statistical representation or “surrogate model” of complicated<sup>111</sup>  
78 numerical simulations by using simulation output as the train-<sup>112</sup>  
79 ing data for a statistical model (e.g. O’Hagan, 2006, Conti et al.,<sup>113</sup>  
80 2009). Emulation is particularly useful in cases where sim-<sup>114</sup>  
81 ulations are complex and output data volumes are large, and<sup>115</sup>  
82 has been used in a variety of fields, including hydrology (e.g.<sup>116</sup>  
83 Razavi et al., 2012), engineering (e.g. Storlie et al., 2009),<sup>117</sup>  
84 environmental sciences (e.g. Ratto et al., 2012), and climate<sup>118</sup>  
85 (e.g. Castruccio et al., 2014, Holden et al., 2014). For agri-<sup>119</sup>  
86 cultural impacts studies, emulation of process-based models<sup>120</sup>  
87 allows capturing key relationships between input variables in<sup>121</sup>  
88 a lightweight, flexible form that is compatible with economic<sup>122</sup>  
89 studies.<sup>123</sup>

90 In the past decade, multiple studies have developed emula-<sup>124</sup>  
91 tors of process-based crop simulations. Early studies proposing<sup>125</sup>  
92 or describing potential crop yield emulators include Howden<sup>126</sup>  
93 & Crimp (2005), Räisänen & Ruokolainen (2006), Lobell &<sup>127</sup>  
94 Burke (2010), and Ferrise et al. (2011), who used a machine<sup>128</sup>  
95 learning approach to predict Mediterranean wheat yields. Stud-<sup>129</sup>  
96 ies developing single-model emulators include Holzkämper<sup>130</sup>  
97 et al. (2012) for the CropSyst model, Ruane et al. (2013) for<sup>131</sup>  
98 the CERES wheat model, and Oyebamiji et al. (2015) for the<sup>132</sup>  
99 LPJmL model (for multiple crops, using multiple scenarios as<sup>133</sup>  
100 a training set). More recently, emulators have begun to be used<sup>134</sup>

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  
(ISIMIP) (Warszawski et al., 2014), which simulated yields for  
maize, soy, wheat, and rice. Choices differ: Blanc & Sul-  
tan (2015) and Blanc (2017) base their emulation on histori-  
cal simulations and a single future climate/emissions scenario  
(RCP8.5), and use local weather variables and yields in their  
regression but then aggregate across broad regions; Ostberg  
et al. (2018) consider multiple future climate scenarios, using  
global mean temperature change (and CO<sub>2</sub>) as regressors but  
then pattern-scale to emulate local yields; while Mistry et al.  
(2017) attempt to compare emulated historical yearly yields to  
observed historical yields, using local weather data and a his-  
torical crop simulation. These efforts do share important com-  
mon features: all emulate annual crop yields across the entire  
scenario or scenarios, and when future scenarios are consid-  
ered, they are non-stationary, i.e. their input climate parameters  
evolve over time.

An alternative approach to climate model RCP runs is to con-  
struct a training set of multiple stationary scenarios in which pa-  
rameters are systematically varied. Such a “parameter sweep”  
offers several advantages for emulation over scenarios in which  
climate evolves over time. First, it allows separating the ef-  
fects of different variables that impact yields but that are highly  
correlated in realistic future scenarios (e.g. CO<sub>2</sub> and temper-  
ature). Second, it allows making a distinction between year-  
over-year yield variations and climatological changes, which  
may involve different responses to the particular climate regres-  
sors used (e.g. Ruane et al., 2016). For example, if year-over-  
year yield variations are driven predominantly by variations in  
the distribution of temperatures throughout the growing sea-  
son, and long-term climate changes are driven predominantly

135 by shifts in means, then regressing on the mean growing season  
136 temperature will produce different yield responses at annual vs.  
137 climatological timescales.

138 Systematic parameter sweeps have begun to be used in  
139 crop modeling, with early efforts in 2015 (Makowski et al.,  
140 2015, Pirttioja et al., 2015), and several recent studies in 2018  
141 (Fronzek et al., 2018, Snyder et al., 2018, Ruiz-Ramos et al.,  
142 2018). All three studies sample multiple perturbations to tem-  
143 perature and precipitation (with Snyder et al. (2018) and Ruiz-  
144 Ramos et al. (2018) adding CO<sub>2</sub> as well), in 132, 99 and ap-  
145 proximatlly 220 different combinations, respectively, and take  
146 advantage of the structured training set to construct emulators  
147 (aka response surfaces in this context) of climatological mean  
148 yields, omitting year-over-year variations. The main limitation  
149 in these studies is geographic extent, with each study focusing  
150 on a limited number of sites. Fronzek et al. (2018) and Ruiz-  
151 Ramos et al. (2018) simulate only wheat (over several models)  
152 and Snyder et al. (2018) analyzes analyzes four crops (maize,  
153 wheat, rice, soy) based on site-specific crop model sensitivity  
154 tests from the C3MP project (McDermid et al., 2015).

155 In this paper we describe a new comprehensive dataset de-  
156 signed to expand the parameter sweep approach still further.  
157 The Global Gridded Crop Model Intercomparison (GGCMI)  
158 Phase II experiment involves running a suite of process-based  
159 crop models across historical conditions perturbed by a set of  
160 discrete steps in different input parameters, including an ap-  
161 plied nitrogen dimension. The experimental protocol involves  
162 over 700 different parameter combinations for each model and  
163 crop, with simulations providing near-global coverage at a half<sub>169</sub>  
164 degree spatial resolution. The experiment was conducted as<sub>170</sub>  
165 part of the Agricultural Model Intercomparison and Improve-<sub>171</sub>  
166 ment Project (AgMIP) (Rosenzweig et al., 2013, 2014), an in-<sub>172</sub>  
167 ternational effort conducted under a framework similar to the<sub>173</sub>  
168 Climate Model Intercomparison Project (CMIP) (Taylor et al.,<sub>174</sub>

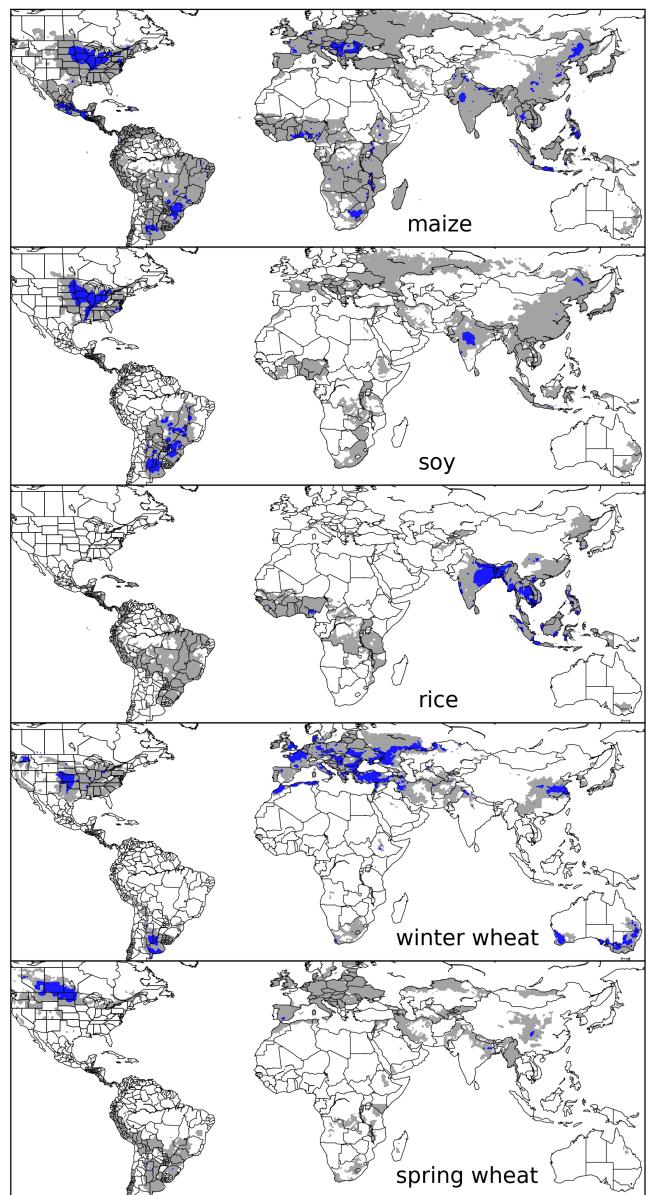


Figure 1: Presently cultivated area for rain-fed crops. Blue indicates grid cells with more than 20,000 hectares (~10% of the equatorial grid cells) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated areas for maize, rice, and soy are taken from the MIRCA2000 (“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.

2012, Eyring et al., 2016). The GGCMI protocol builds on the AgMIP Coordinated Climate-Crop Modeling Project (C3MP) (Ruane et al., 2014, McDermid et al., 2015) and will contribute to the AgMIP Coordinated Global and Regional Assessments (CGRA) (Ruane et al., 2018, Rosenzweig et al., 2018). GGCMI Phase II is designed to allow addressing goals

such as understanding where highest-yield regions may shift<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.

228 higher temperatures. An additional, identical set of scenarios<sup>246</sup> Sacks et al. (2010) and Portmann et al. (2008, 2010)), but vary  
229 (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  
230 simulate adaptive agronomy under climate change by varying<sup>248</sup> sown in March in Spain, in July in Indonesia, and in December  
231 the growing season for crop production. The resulting GGCMI<sup>249</sup> in Namibia. All stresses are disabled other than factors related  
232 Phase II dataset captures a distribution of crop responses over<sup>250</sup> to nitrogen, temperature, and water (e.g. alkalinity and salinity).  
233 the potential space of future climate conditions.<sup>251</sup>

234 The 12 models included in GGCMI Phase II are all mech-<sup>252</sup> No additional nitrogen inputs, such as atmospheric deposition,  
235 anistic process-based crop models that are widely used in im-<sup>253</sup> are considered, but some model treatment of soil organic matter  
236 pacts assessments (Table 2). Although some models share a<sup>254</sup> may allow additional nitrogen release through mineralization.  
237 common base (e.g. the LPJ family or the EPIC family of mod-<sup>255</sup> See Rosenzweig et al. (2014), Elliott et al. (2015) and Müller  
238 els), they have subsequently developed independently. (For<sup>256</sup> et al. (2017) for further details on models and underlying as-  
239 more details on model genealogy, see Figure S1 in Rosenzweig<sup>257</sup> sumptions.

240 et al. (2014).) Differences in model structure mean that several<sup>258</sup> The participating modeling groups provide simulations at  
241 key factors are not standardized across the experiment, includ-<sup>259</sup> any of four initially specified levels of participation, so the num-  
242 ing secondary soil nutrients, carry-over effects across growing<sup>260</sup> ber of simulations varies by model, with some sampling only a  
243 years including residue management and soil moisture, and the<sup>261</sup> part of the experiment variable space. Most modeling groups  
244 extent of simulated area for different crops. Growing seasons<sup>262</sup> simulate all five crops in the protocol, but some omitted one  
245 are standardized across models (with assumptions based on<sup>263</sup> or more. Table 2 provides details of coverage for each model.  
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.

276 All models produce as output crop yields (tons  $\text{ha}^{-1}$  year $^{-1}$ )  
 277 for each 0.5 degree grid cell. Because both yields and yield  
 278 changes vary substantially across models and across grid cells,  
 279 we primarily analyze relative change from a baseline. We take

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  $\text{ha}^{-1}$ . We show absolute yields in some cases to illustrate geo-  
 283 graphic differences in yields for a single model.

### 284 3. Simulation – Results

Crop models in the GGCM Phase II ensemble show broadly  
 consistent responses to climate and management perturbations  
 in most regions, with a strong negative impact of increased tem-  
 perature in all but the coldest regions. We illustrate this result  
 for rain-fed maize in Figure 2, which shows yields for the pri-  
 mary Köppen-Geiger climate regions (Rubel & Kottek, 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 im-  
 pacts are robust for even moderate climate perturbations. In the

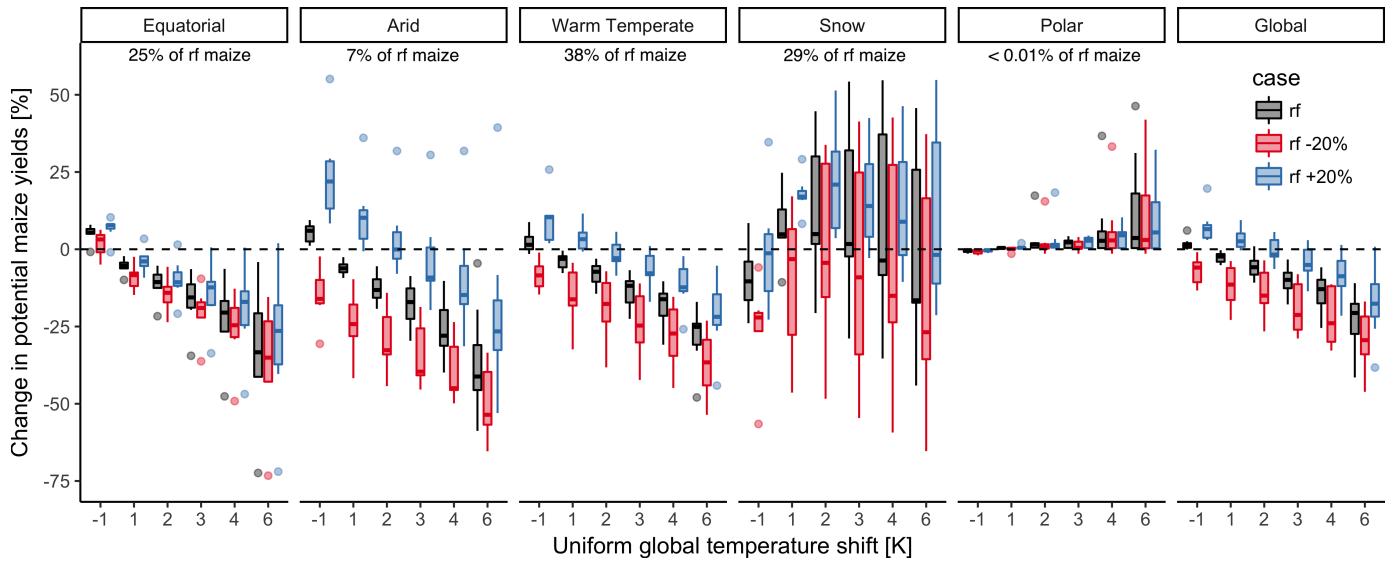


Figure 2: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation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

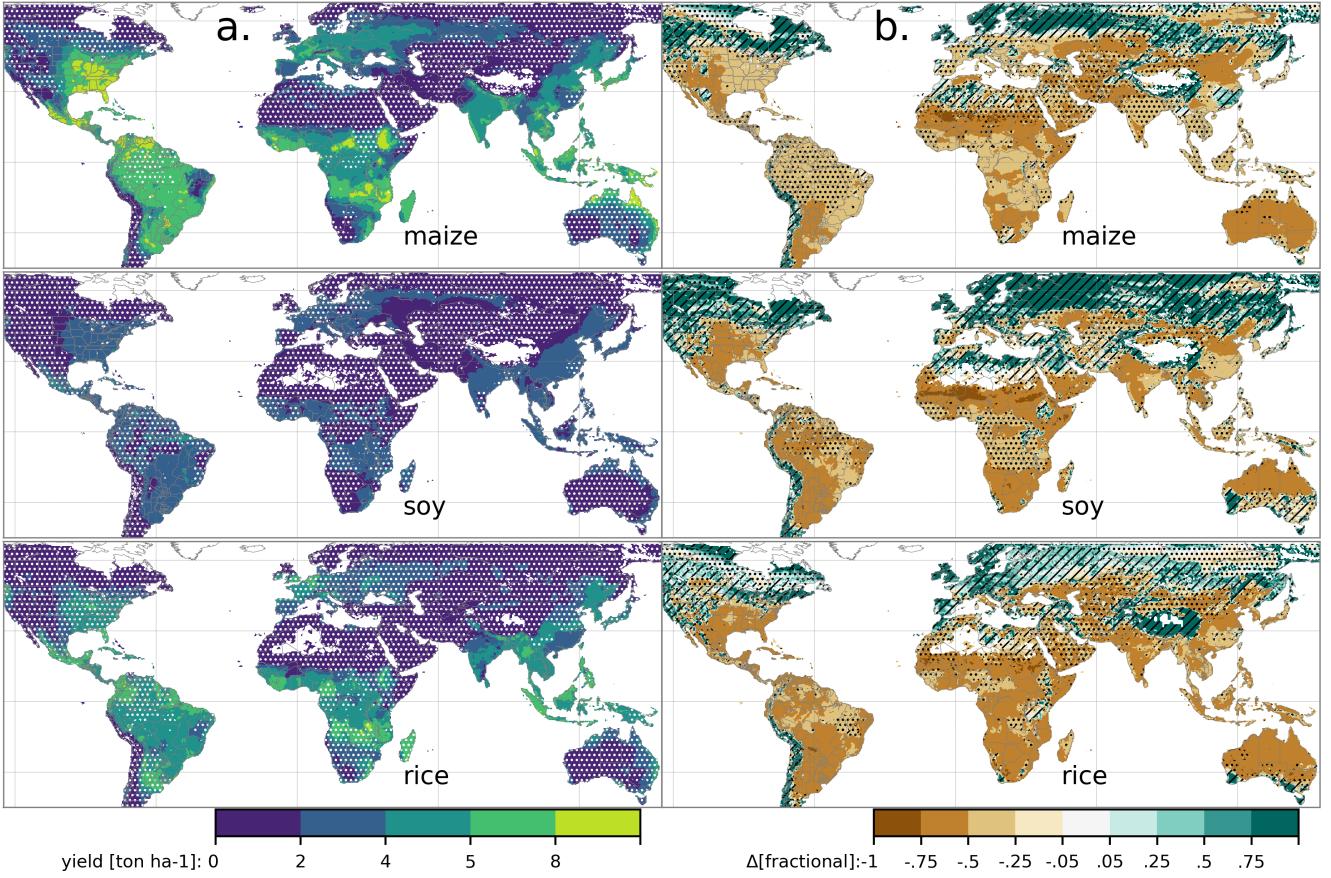


Figure 3: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure 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 (0–2 ton  $\text{ha}^{-1}$ ). Right column (b) shows the multi-model mean fractional yield change in the extreme  $T + 4$  °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.

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 (Fig-  
ure S7).

The effects of rainfall changes on maize yields shown in Fig-  
ure 2 are also as expected and are consistent across models.  
Increased rainfall mitigates the negative effect of higher tem-

peratures 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 expected, 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,

321 with locations of model agreement marked. Absolute yield po-354  
 322 tentials show strong spatial variation, with much of the Earth's355  
 323 surface area unsuitable for any given crop. In general, mod-356  
 324 els agree most on yield response in regions where yield poten-357  
 325 tials are currently high and therefore where crops are currently  
 326 grown. Models show robust decreases in yields at low latitudes,  
 327 and highly uncertain median increases at most high latitudes.  
 328 For wheat crops see Figure S11; wheat projections are more  
 329 uncertain potentially due to lack of calibration (especially im-  
 330 portant for wheat Asseng et al., 2013) and the more complicated  
 331 phenological development of winter wheat when compared to  
 332 other crops. Simulation model validation can be found in Ap-  
 333 pendix 7.

#### 334 4. Emulation – Methods

335 As part of our demonstration of the properties of the GGCMI  
 336 Phase II dataset, we construct an emulator of 30-year clima-  
 337 tological mean yields. This approach is made possible by  
 338 the structured set of simulations involving systematic per-  
 339 turbations. In the GGCMI Phase II dataset, the year-over-year re-  
 340 sponds are generally quantitatively distinct from (and larger  
 341 than) climatological mean responses. In the example of Figure  
 342 4, responses to year-over-year temperature variations are 100%  
 343 larger than those to long-term perturbations in the baseline case,  
 344 and larger still under warmer conditions, rising to nearly 200%359  
 345 more in the T+6 case. The stronger year-over-year response359  
 346 under warmer conditions also manifests as a wider distribu-360  
 347 tion of yields (Figure 5). As discussed previously, year-over-361  
 348 year and climatological responses can differ for many reasons362  
 349 including memory in the crop model, lurking covariants, and363  
 350 differing associated distributions of daily growing-season daily364  
 351 weather (e.g. Ruane et al., 2016). Note that the GGCMI Phase365  
 352 II datasets do not capture one climatological factor, potential366  
 353 future distributional shifts, because all simulations are run with367

354 fixed offsets from the historical climatology. Prior work has  
 355 suggested that mean changes are the dominant drivers of clima-  
 356 tological crop yield shifts in non-arid regions (e.g. Glotter et al.,  
 357 2014).

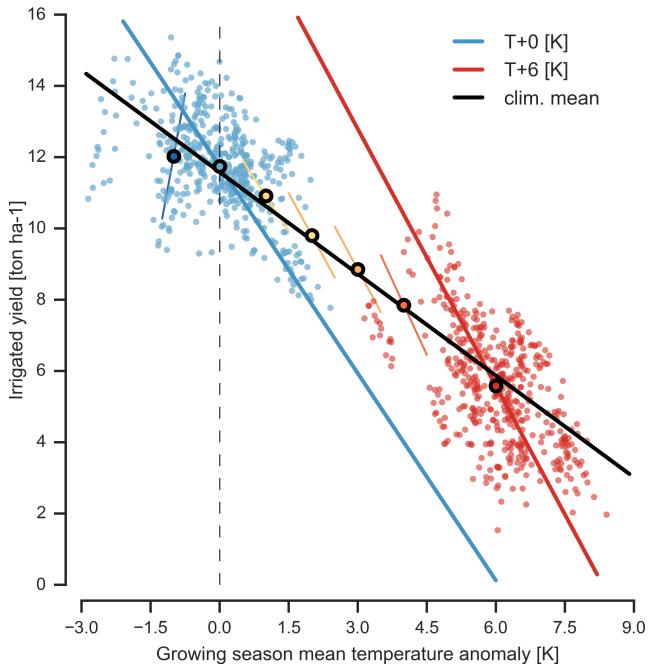


Figure 4: Example showing distinction between crop yield responses to year-to-year and climatological mean temperature shifts. Figure shows irrigated maize for a representative high-yield region (nine adjacent grid cells in northern Iowa) from the pDSSAT model, for the baseline 1981-2010 historical climate (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 improve signal to noise ratio in our model.) The higher-order terms are necessary to capture any nonlinear responses, which are well-documented in observations for temperature and wa-

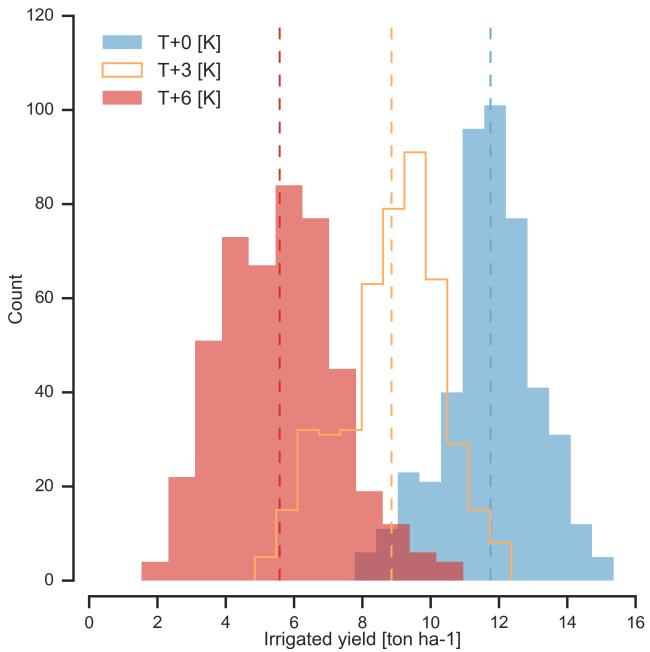


Figure 5: Example showing climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows irrigated maize for nine adjacent high-yield grid cells of Figure 4 (in northern Iowa, same as Figure 4) from the pDSSAT model, for the baseline 1981–2010 historical climate (blue) and for scenarios with temperature 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. In this work we emulate not the year-over-year distributions but the climatological mean response (dashed vertical lines).

tion that allows for some interpretation of coefficients. Some prior studies have used more complex functional forms and larger numbers of parameters, e.g. 39 in Blanc & Sultan (2015) and Blanc (2017), who borrow information across space by fitting grid points simultaneously across a large region in a panel regression. We do not aggregate in space since the emulation is not computationally demanding even at the half-degree grid cell resolution globally. The emulation therefore indirectly includes any yield response to geographically distributed factors such as soil type, insolation, and the baseline climate itself. We explicitly hold the specification constant across all crops and models to facilitate model comparison by looking at parameters directly in lieu of the much larger yield output data.

The relatively simple functional form used here allows emulation at grid-cell level with low noise? how do you quantify this?

maybe add table of parameters like stefan said?

#### 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 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

ter perturbations (e.g. Schlenker & Roberts (2009) for T and He et al. (2016) for W). We include interaction terms (both linear and higher-order) because past studies have shown them to be significant effects. For example, Lobell & Field (2007) and Tebaldi & Lobell (2008) showed that in real-world yields, the joint distribution in T and W is needed to explain observed yield variance. (C and N are fixed in these data.) Other observation-based studies have shown the importance of the interaction between water and nitrogen (e.g. Aulakh & Malhi, 2005), and between nitrogen and carbon 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 rain-fed case) to 23 terms.

We do not focus on comparing different functional forms in this study, and instead choose a relatively simple parametriza-

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 regression for irrigated crops does not contain the W terms and models that did not sample the nitrogen levels (see 2 do not contain<sup>416</sup>

any of the N terms.

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

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

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

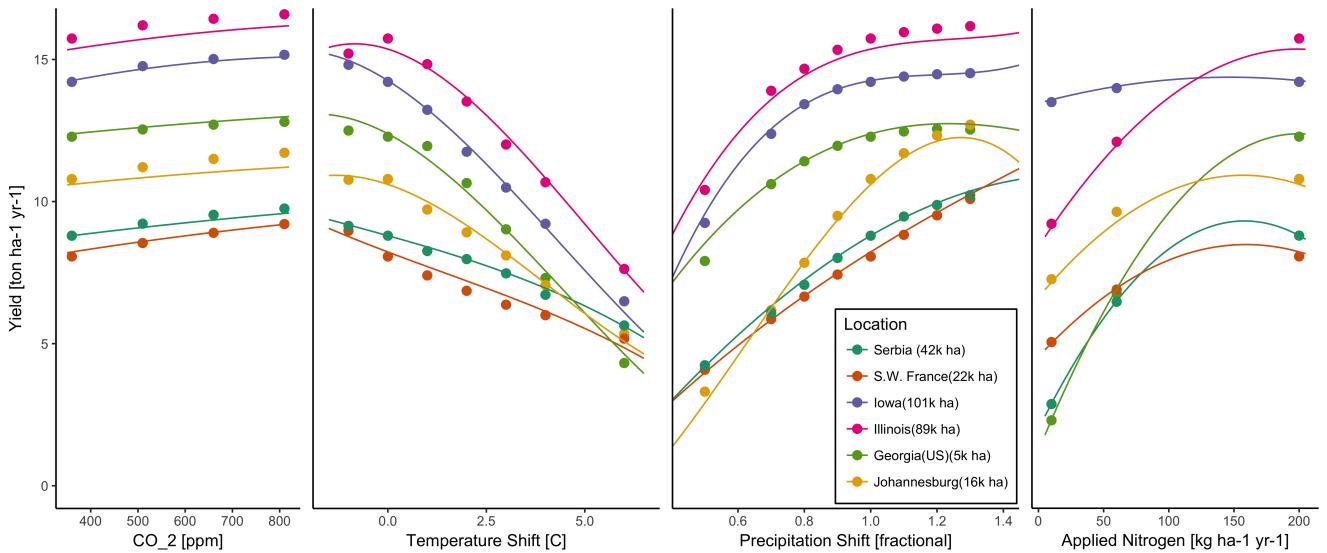


Figure 6: Illustration of spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation 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 9.

425 provided a limited number of cases or for low-yield geographic<sup>458</sup>  
426 regions where some modeling groups did not run all scenar-<sup>459</sup>  
427 ios. We do not attempt to emulate models that provided less<sup>460</sup>  
428 than 50 simulations. The lowest number of simulations emu-<sup>461</sup>  
429 lated across the full parameter space is then 130 (for the PEPIC<sup>462</sup>  
430 model). The resulting parameter matrices for all crop model<sup>463</sup>  
431 emulators are available on request [give location?](#), as are the raw<sup>464</sup>  
432 simulation data and a Python application to emulate yields. The<sup>465</sup>  
433 yield output for a single GGCMI Phase II model that simulates<sup>466</sup>  
434 all scenarios and all five crops is ~12.5 GB; the emulator is<sup>467</sup>  
435 ~100 MB, a reduction by over two orders of magnitude.

468 yield responses generally follow similar functional forms across  
469 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. For this location and crop, CO<sub>2</sub>  
fertilization effects can range from ~5–50%, and nitrogen re-  
sponses from nearly flat to a 60% drop in the lowest-application  
simulation.

## 436 5. Emulation – Results

437 Emulation provides not only a computational tool but a<sup>471</sup>  
438 means of understanding and interpreting crop yield response<sup>472</sup>  
439 across the parameter space. Emulation is only possible when<sup>473</sup>  
440 crop yield responses are sufficiently smooth and continuous to<sup>474</sup>  
441 allow fitting with a relatively simple functional form, but this<sup>475</sup>  
442 condition largely holds in the GGCMI Phase II simulations. Re-<sup>476</sup>  
443 sponds are quite diverse across locations, crops, and models,<sup>477</sup>  
444 but in most cases local responses are regular enough to permit<sup>478</sup>  
445 emulation. We show illustrations of emulation fidelity in this<sup>479</sup>  
446 section; for more detailed discussion see Appendix 9.

447 Crop yield responses are geographically diverse, even in<sup>481</sup>  
448 high-yield and high-cultivation areas Figure 6 illustrates geo-<sup>482</sup>  
449 graphic diversity for a single crop and model (rain-fed maize<sup>483</sup>  
450 in pDSSAT); this heterogeneity supports the choice of emulat-<sup>484</sup>  
451 ing at the grid cell level. Each panel in Figure 6 shows sim-<sup>485</sup>  
452 ulted yield output from scenarios varying only along a single<sup>486</sup>  
453 dimension (CO<sub>2</sub>, temperature, precipitation, or nitrogen addi-<sup>487</sup>  
454 tion), with other inputs held fixed at baseline levels, compared<sup>488</sup>  
455 to the full 4D emulation across the parameter space. Yields<sup>489</sup>  
456 evolve smoothly across the space sampled, and the polynomial<sup>490</sup>  
457 fit captures the climatological response to perturbations. Crop<sup>491</sup>

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

484 The emulation fidelity demonstrated here is sufficient to al-  
485 low using emulated response surfaces to compare model re-  
486 sponds and derive insight about impacts projections. Because  
487 the emulator or “surrogate model” transforms the discrete sim-  
488 ulation sample space into a continuous response surface at any  
489 geographic scale, it can be used for a variety of applications,  
490 including construction of continuous damage functions. As an

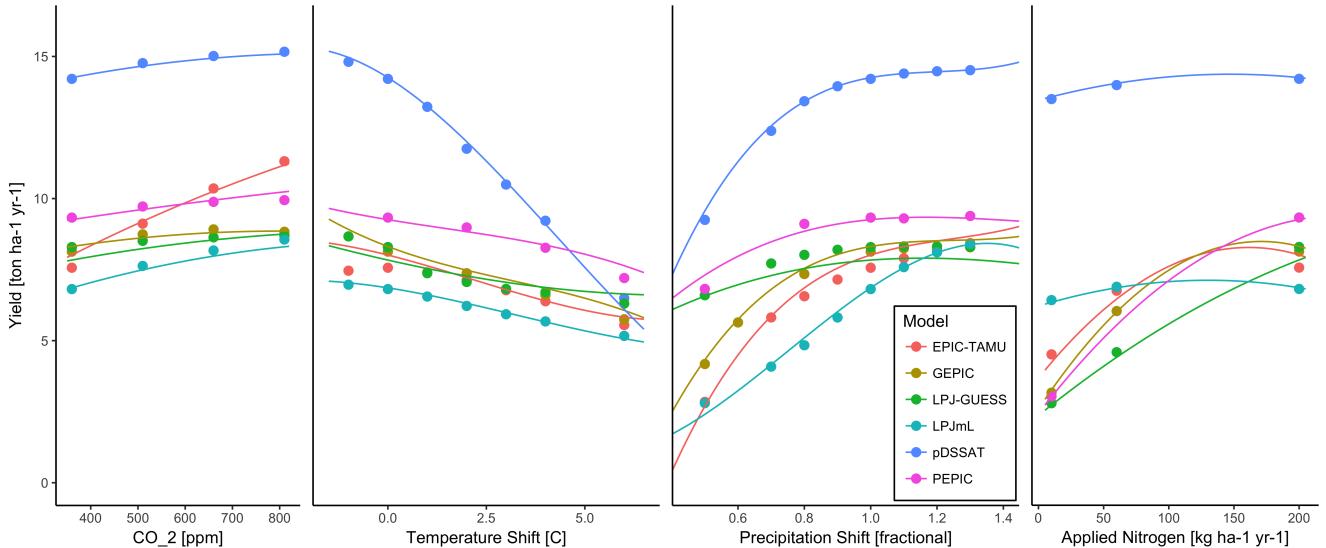


Figure 7: Illustration of across-model variations in yield response. Figures shows simulations and emulations from six models for rain-fed maize in the same Iowa grid cell shown in Figure 6, with the same plot conventions. Models that do not simulate the nitrogen dimension are omitted for clarity. Note 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 9. As in Figure 6, extrapolation out of the sample space is problematic.

example, we show a damage function constructed from the 4D<sub>511</sub> emulation, aggregated to global yield, with simulated values<sub>512</sub> shown for comparison (Figure 8, which shows maize on currently cultivated land; see Figures S16- S19 for other crops and dimensions). The emulated values closely match simulations<sub>515</sub> even at this aggregation level. Note that these functions are<sub>516</sub> presented only as examples and do not represent true global<sub>517</sub> projections, because they are developed from simulation data<sub>518</sub> with a uniform temperature shift while increases in global mean<sub>519</sub> temperature should manifest non-uniformly. The global coverage<sub>520</sub> of the GGCMI Phase II simulations allows impacts models<sub>521</sub> to apply arbitrary geographically-varying climate projections,<sub>522</sub> as well as arbitrary aggregation masks, to develop damage<sub>523</sub> functions for any climate scenario and any geopolitical or<sub>524</sub> geographic level.

## 6. Conclusions and discussion

The GGCMI Phase II experiment provides a database targeted to allow detailed study of crop yields from process-based models under climate change. The experiment is designed to

facilitate not only comparing the sensitivities of process-based crop yield models to changing climate and management inputs but also evaluating the complex interactions between driving factors (CO<sub>2</sub>, temperature, precipitation, and applied nitrogen). Its global nature also allows identifying geographic shifts in high yield potential locations. We expect that the simulations will yield multiple insights in future studies, and show here a selection of preliminary results to illustrate their potential uses.

First, the GGCMI Phase II simulations allow identifying major areas of uncertainty. Across the major crops, inter-model uncertainty is greatest for wheat and least for soy. Across factors impacting yields, inter-model uncertainty is largest for CO<sub>2</sub> fertilization and nitrogen response effects. Across geographic regions, projections are most uncertain in the high latitudes where yields may increase, and most robust in low latitudes where yield impacts are largest.

Second, the GGCMI Phase II simulations allow understanding the way that climate-driven changes and locations of cultivated land combine to produce yield impacts. One counterintuitive result immediate apparent is that irrigated maize

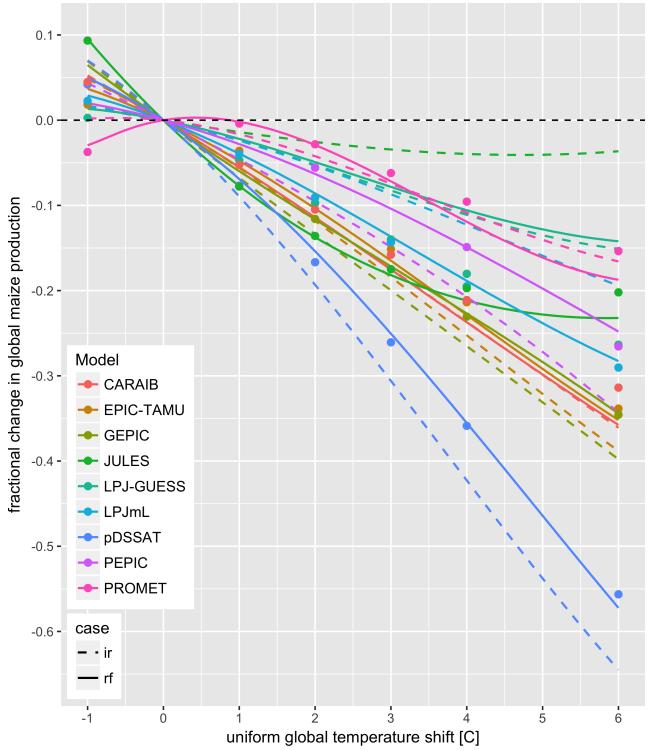


Figure 8: Global emulated damages for maize on currently cultivated lands<sup>558</sup> for the GGCMI Phase II models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating<sup>559</sup> up emulated values at the grid cell level, not from a regression of global mean yields.) Lines are emulations for rain-fed (solid) and irrigated (dashed) crops;<sup>560</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 locations of cultivated areas: irrigated crops tend to be grown in warmer areas<sup>561</sup> where impacts are more severe for a given temperature shift. (The exceptions are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures<sup>562</sup> S16- S19 in the supplemental material.

<sup>544</sup> rice is not generally grown in water-limited conditions).

<sup>545</sup> Third, we show that even the relatively limited GGCMI  
<sup>546</sup> Phase II sampling space allows emulation of the climatological  
<sup>547</sup> response of crop models with a relatively simple reduced-form  
<sup>548</sup> statistical model. The systematic parameter sampling in the  
<sup>549</sup> GGCMI Phase II procedure provides information on the influ-  
<sup>550</sup> ence of multiple interacting factors in a way that single projec-  
<sup>551</sup> tions cannot, and emulating the resulting response surface then  
<sup>552</sup> produces a tool that can aid in both physical interpretation of  
<sup>553</sup> the process-based models and in assessment of agricultural im-  
<sup>554</sup> pacts under arbitrary climate scenarios. Emulating the climato-  
<sup>555</sup> logical response isolates long-term impacts from any confound-  
<sup>556</sup> ing factors that complicate year-over-year changes, and the use  
<sup>557</sup> of simple functional forms offer the possibility of physical in-  
<sup>558</sup> terpretation of parameter values. We anticipate that systematic  
<sup>559</sup> parameter sampling will become the norm in future crop model  
<sup>560</sup> intercomparison exercises.

While the GGCMI Phase II database should offer the foun-<sup>561</sup> dation for multiple future studies, several cautions need to be  
<sup>562</sup> noted. Because the simulation protocol was designed to focus  
<sup>563</sup> on change in yield under climate perturbations and not on repli-  
<sup>564</sup> cating real-world yields, the models are not formally calibrated  
<sup>565</sup> so cannot be used for impacts projections unless in used in con-  
<sup>566</sup> junction with historical data (or data products). Because the  
<sup>567</sup> GGCMI Phase II simulations apply uniform perturbations to  
<sup>568</sup> historical climate inputs, they do not sample changes in higher  
<sup>569</sup> order moments, and cannot address the additional crop yield  
<sup>570</sup> impacts of potential changes in climate variability. Although  
<sup>571</sup> distributional changes in model projections are fairly uncertain  
<sup>572</sup> at present, follow-on experiments may wish to consider them.  
<sup>573</sup> Several recent studies have described procedures for generating  
<sup>574</sup> simulations that combine historical data with model projections  
<sup>575</sup> of not only mean changes in temperature and precipitation but  
<sup>576</sup> changes in their marginal distributions or temporal dependence.

<sup>531</sup> shows steeper yield reductions under warming than does rain-<sup>565</sup>  
<sup>532</sup> fed maize when considered only over currently cultivated land.<sup>566</sup>  
<sup>533</sup> The effect results from geographic differences in cultivation. In<sup>567</sup>  
<sup>534</sup> any given location, irrigation increases crop resiliency to tem-<sup>568</sup>  
<sup>535</sup> perature increase, but irrigated maize is grown in warmer loca-<sup>569</sup>  
<sup>536</sup> tions where the impacts of warming are more severe (Figures<sup>570</sup>  
<sup>537</sup> S5-S6). The same behavior holds for rice and winter wheat,<sup>571</sup>  
<sup>538</sup> but not for soy or spring wheat (Figures S8-S10). Irrigated<sup>572</sup>  
<sup>539</sup> wheat and maize are also more sensitive to nitrogen fertiliza-<sup>573</sup>  
<sup>540</sup> tion levels than are analogous non-irrigated crops, presumably<sup>574</sup>  
<sup>541</sup> because those rain-fed crops are limited by water as well as<sup>575</sup>  
<sup>542</sup> nitrogen availability (Figure S19). (Soy as an efficient atmo-<sup>576</sup>  
<sup>543</sup> spheric nitrogen-fixing is relatively insensitive to nitrogen, and<sup>577</sup>

578 For methods to generate adjust historical climate data inclusive<sub>609</sub>  
579 of distributional and temporal dependence changes, see Leeds<sub>610</sub>  
580 et al. (2015), Poppick et al. (2016), Chang et al. (2016) and<sub>611</sub>  
581 Haugen et al. (2018)). Emulation approaches are an area of active<sub>612</sub>  
582 ongoing study and one of the goals of the GGCMI Phase II<sub>613</sub>  
583 dataset is to facilitate these research efforts.<sub>614</sub>

584 The GGCMI Phase II output dataset invites a broad range<sub>616</sub>  
585 of potential future avenues of analysis. A major target area of<sub>617</sub>  
586 research is studying the models themselves including: a detailed<sub>618</sub>  
587 examination of interaction terms between the major input<sub>619</sub>  
588 drivers, a robust quantification of the sensitivity of different<sub>620</sub>  
589 models to the input drivers, and comparisons with field-level<sub>621</sub>  
590 experimental data. The parameter space tested in GGCMI<sub>622</sub>  
591 Phase II will allow detailed investigations into yield variability<sub>623</sub>  
592 and response to extremes under changing management and<sub>624</sub>  
593 CO<sub>2</sub> levels and allow the study of geographic shifts in optimal<sub>625</sub>  
594 growing regions for different crops. The output dataset<sub>626</sub>  
595 also contains other runs and variables not analyzed or shown<sub>627</sub>  
596 here. Runs include several which allowed adaptation to climate<sub>628</sub>  
597 changes by altering growing seasons, and additional variables<sub>629</sub>  
598 include above ground biomass, LAI, and root biomass (as many<sub>630</sub>  
599 as 25 output variables for some models). Emulation studies that<sub>631</sub>  
600 are possible include a more systematic evaluation of different<sub>632</sub>  
601 statistical model specifications and formal calculation of uncertainties<sub>633</sub>  
602 in derived parameters.<sub>634</sub>

603 The development of multi-model ensembles such as GGCMI<sub>636</sub>  
604 Phase II provides a way to begin to better understand crop responses<sub>637</sub>  
605 to a range of potential climate inputs, improve process<sub>638</sub>  
606 based models, and explore the potential benefits of adaptive responses<sub>639</sub>  
607 included shifting growing season, cultivar types and<sub>640</sub>  
608 cultivar geographic extent.<sub>641</sub>

## 7. Acknowledgments

We thank Michael Stein and Kevin Schwarzwald, who provided helpful suggestions that contributed to this work. This research was performed as part of the Center for Robust Decision-Making on Climate and Energy Policy (RDCEP) at the University of Chicago, and was supported through a variety of sources.

RDCEP is funded by NSF grant #SES-1463644 through the Decision Making Under Uncertainty program. J.F. was supported by the NSF NRT program, grant #DGE-1735359. C.M. was supported by the MACMIT project (01LN1317A) funded through the German Federal Ministry of Education and Research (BMBF). C.F. was supported by the European Research Council Synergy grant #ERC-2013-SynG-610028 Imbalance. P. P.F. and K.W. were supported by the Newton Fund through the Met Office Climate Science for Service Partnership Brazil (CSSP Brazil). A.S. was supported by the Office of Science of the U.S. Department of Energy as part of the Multi-sector Dynamics Research Program Area. Computing resources were provided by the University of Chicago Research Computing Center (RCC). S.O. acknowledges support from the Swedish strong research areas BECC and MERGE together with support from LUCCI (Lund University Centre for studies of Carbon Cycle and Climate Interactions).

## 8. Appendix: Simulations – Assessment

The GGCMI Phase II simulations are designed for evaluating changes in yield but not absolute yields, since they omit detailed calibrations. To provide some validation of the skill of the process-based models used, we repeat the validation exercises of Müller et al. (2017) for GGCMI Phase I. The Müller et al. (2017) procedure evaluates response to year-to-year temperature and precipitation variations in a control run driven by historical climate and compares it to detrended historical yields from the FAO (Food and Agriculture Organization of the United

Nations, 2018) by calculating the Pearson correlation coefficient. The procedure offers no means of assessing CO<sub>2</sub> fertilization, since CO<sub>2</sub> has been relatively constant over the historical data collection period. Nitrogen introduces some uncertainty into the analysis, since the GGCMI Phase II runs impose fixed, uniform nitrogen application levels that are not realistic for individual countries. We evaluate up to three control runs for each model, since some modeling groups provide historical runs for three different nitrogen levels.

Figure 9 shows the Pearson time series correlation between the simulation model yield and FAO yield data. Figure 9 can be compared to Figures 1,2,3,4 and 6 in Müller et al. (2017). The results are mixed, with many regions for rice and wheat being difficult to model. No single model is dominant, with each model providing near best-in-class performance in at least one location-crop combination. The presence of very few vertical dark green color bars clearly illustrates the power of a multi-model intercomparison project like the one presented here. The ensemble mean does not beat the best model in each case, but shows positive correlation in over 75% of the cases presented here. The EPIC-TAMU model performs best for soy, CARIAB, EPIC-TAMU, and PEPIC perform best for maize, PROMET performs best for wheat, and the EPIC family of models perform best for rice. Reductions in skill over the performance illustrated in Müller et al. (2017) may be attributed to the nitrogen levels or lack of calibration in some models.

The FAO data is at least one level of abstraction from ground truth in many cases, especially in developing countries. The failure of models to represent the year-to-year variability in rice in some countries in southeast Asia is likely partly due to model failure and partly due to lack of data. It is possible to speculate that the difference in performance between Pakistan (no successful models) and India (many successful models) for rice may reside at least in part in the FAO data and not the mod-

els themselves. The same might apply to Bangladesh and India for rice. Additionally, there is less year-to-year variability in rice yields (partially due to the fraction of irrigated cultivation). Since the Pearson r metric is scale invariant, it will tend to score the rice models more poorly than maize and soy.

## 9. Appendix: Emulation – Assessment

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

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

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

Here  $F_{scn.}$  is the fractional change in a model’s mean emulated or simulated yield from a defined baseline, in some scenario (scn.) in C, T, W, and N space;  $Y_{scn.}$  and  $Y_{baseline}$  are the absolute emulated or simulated mean yields. The normalized error  $e$  is the difference between the emulated fractional change in yield and that actually simulated, normalized by  $\sigma_{sim.}$ , the standard deviation in simulated fractional yields  $F_{sim, scn.}$  across all models. The emulator is fit across all available simulation outputs, and then the error is calculated across the simulation scenarios provided by all nine models (Figure 10 and Figures S12 and Figures S13 in supplemental documents).

To assess the ability of the polynomial emulation to capture the behavior of complex process-based models, we evaluate the normalized emulator error. That is, for each grid cell, model, and scenario we evaluate the difference between the model yield and its emulation, normalized by the inter-model standard deviation in yield projections. This metric implies that emulation is generally satisfactory, with several distinct exceptions. Almost all model-crop combination emulators have normalized errors less than one over nearly all currently cultivated hectares (Figure 10), but some individual model-crop combinations are

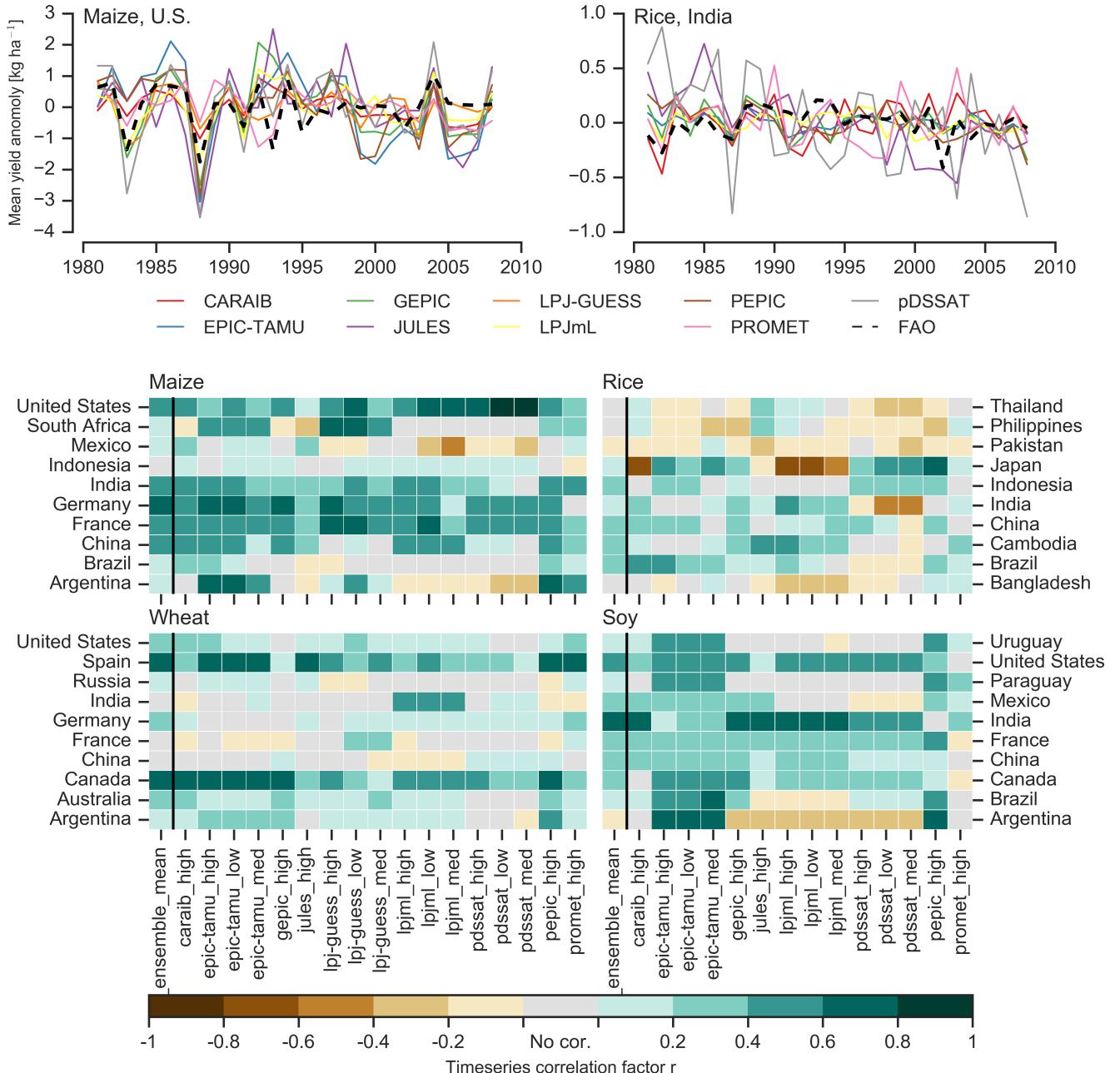


Figure 9: Time series of correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen 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.

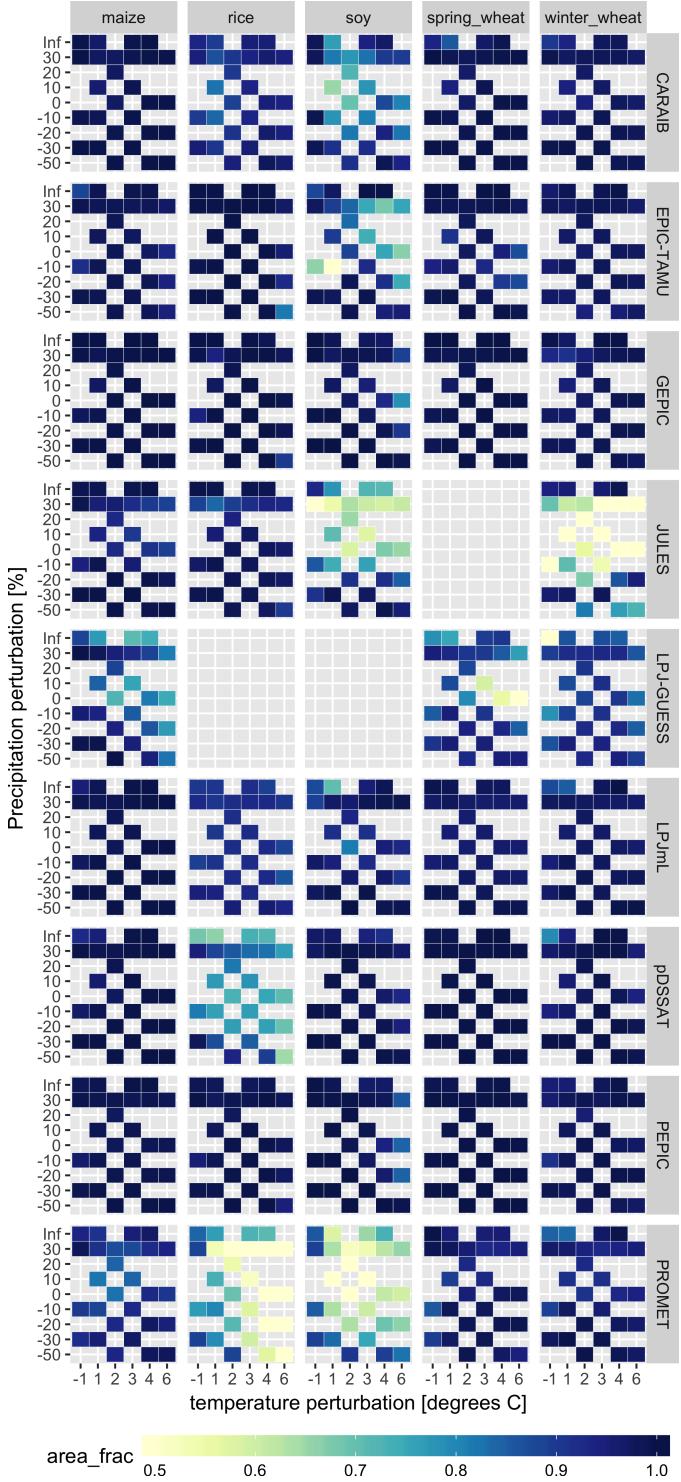


Figure 10: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO<sub>2</sub> and nitrogen held fixed at baseline values. Large columns are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

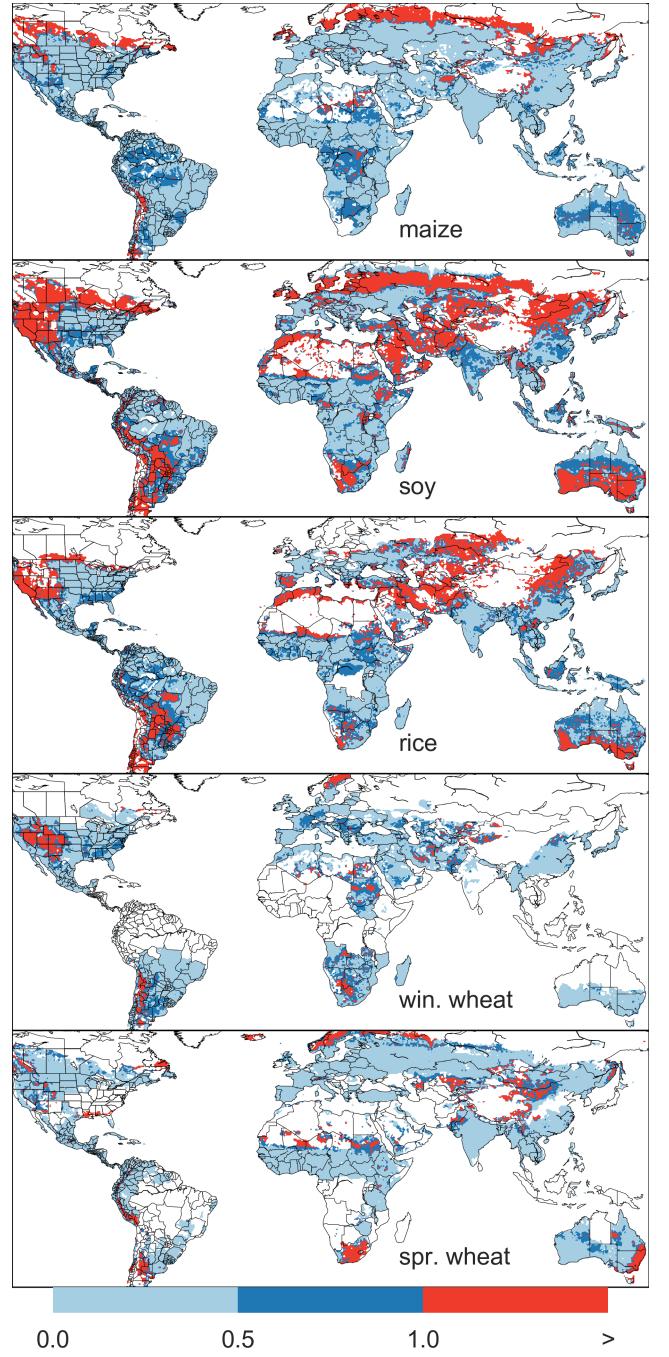


Figure 11: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error  $e$ , where  $e > 1$  means that emulator error exceeds the multi-model standard deviation. White areas are those where 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 10 and 11). In general, emulators tend to have better simulations spread in marginal areas with low yield potentials. For CARAIB, emulation of soy is more problematic, as was also shown in Figure 10.

714 problematic (e.g. PROMET for rice and soy, JULES for soy<sup>748</sup>  
715 and winter wheat, Figures S14–S15). Normalized errors for soy<sup>749</sup>  
716 are somewhat higher across all models not because emulator fi-  
717 delity is worse but because models agree more closely on yield  
718 changes for soy than for other crops (see Figure S16, lowering  
719 the denominator. Emulator performance often degrades in geo-<sup>751</sup>  
720 graphic locations where crops are not currently cultivated. Fig-<sup>752</sup>  
721 ure 11 shows a CARAIB case as an example, where emulator  
722 performance is satisfactory over cultivated areas for all crops  
723 other than soy, but uncultivated regions show some problematic  
724 areas.<sup>755</sup>

725 Note that the normalized error  $e$  for a model depends not only  
726 on the fidelity of its emulator in reproducing a given simulation<sup>760</sup>  
727 but on the particular suite of models considered in the inter-<sup>761</sup>  
728 comparison exercise. The rationale for this choice is to relate  
729 the fidelity of the emulation to an estimate of true uncertainty,<sup>764</sup>  
730 which we take as the multi-model spread. Because the inter-<sup>765</sup>  
731 model spread is large, normalized errors tend to be small. That  
732 is, any failures of emulation are small relative to inter-model  
733 uncertainty. We therefore do not provide a formal parameter<sup>769</sup>  
734 uncertainty analysis, but note that the GGCMI Phase II dataset<sup>770</sup>  
735 is well-suited to statistical exploration of emulation approaches  
736 and quantification of emulator fidelity.<sup>773</sup>

737 It should be noted that this assessment metric is relatively  
738 forgiving. First, each emulation is evaluated against the simu-<sup>774</sup>  
739 lation actually used to train the emulator. Had we used a spline<sup>777</sup>  
740 interpolation the error would necessarily be zero. Second, the<sup>778</sup>  
741 performance metric scales emulator fidelity not by the magni-<sup>779</sup>  
742 tude of yield changes but by the inter-model spread in those  
743 changes. Where models differ more widely, the standard for<sup>782</sup>  
744 emulators becomes less stringent. Because models disagree on<sup>783</sup>  
745 the magnitude of CO<sub>2</sub> fertilization, this effect is readily seen  
746 when comparing assessments of emulator performance in sim-<sup>785</sup>  
747 ulations at baseline CO<sub>2</sub> (Figure 10) with those at higher CO<sub>2</sub><sup>787</sup>

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

## 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., Biernath, 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 production potentials and management flexibility under the representative concentration pathways. *Global and Planetary Change*, 122, 107 – 121.
- Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models. *Agricultural and Forest Meteorology*, 236, 145 – 161.
- Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded crop models using statistical estimates. *Agricultural and Forest Meteorology*, 214-215, 134 – 147.
- von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S. (2018). Implementing the Nitrogen cycle into the dynamic global vegetation, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific Model Development*, 11, 2789–2812.
- Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L., & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs. *Journal of Climate*, 27, 1829–1844.
- Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.

- 788 (2014). A meta-analysis of crop yield under climate change and adaptation.<sup>831</sup>  
 789 *Nature Climate Change*, *4*, 287 – 291. <sup>832</sup>
- 790 Chang, W., Stein, M., Wang, J., Kotamarthi, V., & Moyer, E. (2016). Changes in<sup>833</sup>  
 791 spatio-temporal precipitation patterns in changing climate conditions. *Journal of Climate*, *29*. doi:10.1175/JCLI-D-15-0844.1. <sup>834</sup>  
 792 <sup>835</sup>
- 793 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process<sup>836</sup>  
 794 emulation of dynamic computer codes. *Biometrika*, *96*, 663–676. <sup>837</sup>
- 795 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In<sup>838</sup>  
 796 C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth*,<sup>839</sup>  
 797 *Duke University, Durham, North Carolina* (pp. 115–118). Durham, North<sup>840</sup>  
 798 Carolina. <sup>841</sup>
- 799 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for<sup>842</sup>  
 800 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205). <sup>843</sup>
- 801 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdous,<sup>844</sup>  
 802 M., & François, L. (2011). Responses of European forest ecosystems to<sup>845</sup>  
 803 21st century climate: assessing changes in interannual variability and fire<sup>846</sup>  
 804 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99). <sup>847</sup>
- 805 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,<sup>848</sup>  
 806 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating<sup>849</sup>  
 807 impact models and sectors (pSIMS). *Environmental Modelling and Software*, *62*, 509–516. <sup>850</sup>
- 808 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,<sup>851</sup>  
 809 Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C.,<sup>852</sup>  
 810 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.,<sup>853</sup>  
 811 (2015). The Global Gridded Crop Model Intercomparison: data and modeling<sup>854</sup>  
 812 protocols for Phase 1 (v1.0). *Geoscientific Model Development*, *8*, 261–277. <sup>855</sup>
- 813 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J.,<sup>856</sup>  
 814 & Taylor, K. E. (2016). Overview of the coupled model intercomparison project<sup>857</sup>  
 815 phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*, 1937–1958. <sup>858</sup>
- 816 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of climate<sup>859</sup>  
 817 change impacts on durum wheat in the mediterranean region. *Natural Hazards and Earth System Sciences*, *11*, 1293–1302. <sup>860</sup>
- 818 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Regionalization<sup>861</sup>  
 819 of a large-scale crop growth model for sub-Saharan Africa: Model setup, evaluation, and estimation of maize yields. *Agriculture, Ecosystems & Environment*, *151*, 21 – 33. <sup>862</sup>
- 820 Food and Agriculture Organization of the United Nations (2018). FAOSTAT database. URL: <http://www.fao.org/faostat/en/home>. <sup>863</sup>
- 821 Fronzek, S., Pirttioja, N., Carter, T. R., 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.-F., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J., Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna, M., Semenov, M. A., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model wheat yield impact response surfaces showing sensitivity to temperature and precipitation change. *Agricultural Systems*, *159*, 209–224.
- 822 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014). Evaluating the utility of dynamical downscaling in agricultural impacts projections. *Proceedings of the National Academy of Sciences*, *111*, 8776–8781.
- 823 Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitivity of Agricultural Model Performance to Different Climate Inputs. *Journal of Applied Meteorology and Climatology*, *55*, 151113145618001.
- 824 Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous Crop Growth and Yield: Application for Wheat in Central Europe. *Remote Sensing*, *7*, 3934–3965.
- 825 Haugen, M., Stein, M., Moyer, E., & Srivastava, R. (2018). Estimating changes in temperature distributions in a large ensemble of climate simulations using quantile regression. *Journal of Climate*, *31*, 8573–8588. doi:10.1175/JCLI-D-17-0782.1.
- 826 He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H., He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water contents and nitrogen leaching to precipitation, management practices and soil hydraulic properties in semi-arid and humid regions of Canada using the DSSAT model. *Nutrient Cycling in Agroecosystems*, *106*, 201–215.
- 827 Heady, E. O. (1957). An Econometric Investigation of the Technology of Agricultural Production Functions. *Econometrica*, *25*, 249–268.
- 828 Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa State University Press.
- 829 Holden, P., Edwards, N., PH, G., Fraedrich, K., Lunkeit, F., E, K., Labriet, M., Kanudia, A., & F, B. (2014). Plasim-entsem v1.0: A spatiotemporal emulator of future climate change for impacts assessment. *Geoscientific Model Development*, *7*, 433–451. doi:10.5194/gmd-7-433-2014.
- 830 Holzkämper, A., Calanca, P., & Fuhrer, J. (2012). Statistical crop models: Predicting the effects of temperature and precipitation changes. *Climate Research*, *51*, 11–21.
- 831 Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalglish, N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp,

- 874 J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson,<sup>917</sup>  
 875 M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T.,<sup>918</sup>  
 876 Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorff, J.,<sup>919</sup>  
 877 Wedgwood, S., & Keating, B. A. (2014). APSIM Evolution towards a new<sup>920</sup>  
 878 generation of agricultural systems simulation. *Environmental Modelling and*<sup>921</sup>  
 879 *Software*, *62*, 327 – 350.<sup>922</sup>
- 880 Howden, S., & Crimp, S. (2005). Assessing dangerous climate change impacts<sup>923</sup>  
 881 on australia's wheat industry. *Modelling and Simulation Society of Australia*<sup>924</sup>  
 882 *and New Zealand*, (pp. 505–511).<sup>925</sup>
- 883 Izumi, T., Nishimori, M., & Yokozawa, M. (2010). Diagnostics of climate<sup>926</sup>  
 884 model biases in summer temperature and warm-season insolation for the<sup>927</sup>  
 885 simulation of regional paddy rice yield in japan. *Journal of Applied Meteo-*<sup>928</sup>  
 886 *rology and Climatology*, *49*, 574–591.<sup>929</sup>
- 887 Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Ni<sup>930</sup>  
 888 trogen Fertilizers. *Ambio*, *6*, 146–151.<sup>931</sup>
- 889 Izaurrealde, R., Williams, J., McGill, W., Rosenberg, N., & Quiroga Jakas, M.<sup>932</sup>  
 890 (2006). Simulating soil C dynamics with EPIC: Model description and test<sup>933</sup>  
 891 against long-term data. *Ecological Modelling*, *192*, 362–384.<sup>934</sup>
- 892 Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the<sup>935</sup>  
 893 CROPGRO-soybean model to predict regional yield and production. *Agri-*<sup>936</sup>  
 894 *culture, Ecosystems & Environment*, *93*, 73 – 85.<sup>937</sup>
- 895 Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L.,<sup>938</sup>  
 896 Wilkens, P., Singh, U., Gijsman, A., & Ritchie, J. (2003). The DSSAT<sup>939</sup>  
 897 cropping system model. *European Journal of Agronomy*, *18*, 235 – 265.<sup>940</sup>
- 898 Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I.,<sup>941</sup>  
 899 Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A.,<sup>942</sup>  
 900 Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017).<sup>943</sup>  
 901 Toward a new generation of agricultural system data, models, and knowl-<sup>944</sup>  
 902 edge products: State of agricultural systems science. *Agricultural Systems*,<sup>945</sup>  
 903 *155*, 269 – 288.<sup>946</sup>
- 904 Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth,<sup>947</sup>  
 905 D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Ver-<sup>948</sup>  
 906 burg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K.,<sup>949</sup>  
 907 Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003).<sup>950</sup>  
 908 An overview of APSIM, a model designed for farming systems simulation.<sup>951</sup>  
 909 *European Journal of Agronomy*, *18*, 267 – 288.<sup>952</sup>
- 910 Leeds, W. B., Moyer, E. J., & Stein, M. L. (2015). Simulation of future<sup>953</sup>  
 911 climate under changing temporal covariance structures. *Advances in*<sup>954</sup>  
 912 *Statistical Climatology, Meteorology and Oceanography*, *1*, 1–14. URL:<sup>955</sup>  
 913 <https://www.adv-stat-clim-meteorol-oceanogr.net/1/1/2015/><sup>956</sup>  
 914 doi:10.5194/ascmo-1-1-2015.<sup>957</sup>
- 915 Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., &<sup>958</sup>  
 916 Smith, B. (2013). Implications of accounting for land use in simulations of<sup>959</sup>  
 ecosystem carbon cycling in Africa. *Earth System Dynamics*, *4*, 385–407.
- Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling  
 wheat yield and crop water productivity with high resolution on a global  
 scale. *Agricultural Systems*, *94*, 478 – 493.
- Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a).  
 Global investigation of impacts of PET methods on simulating crop-water  
 relations for maize. *Agricultural and Forest Meteorology*, *221*, 164 – 175.
- Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C.,  
 & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs  
 with yields from major crop cultivations. *Science of The Total Environment*,  
*572*, 526 – 537.
- Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict  
 crop yield responses to climate change. *Agricultural and Forest Meteorol-*  
*ogy*, *150*, 1443 – 1452.
- Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relation-  
 ships and the impacts of recent warming. *Environmental Research Letters*,  
*2*, 014002.
- MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, *4*, 415–447.
- Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P.,  
 Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi,  
 P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano,  
 D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of  
 large simulated yield datasets for studying climate effects. (p. 1100).  
 doi:10.13140/RG.2.1.5173.8328.
- Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrolog-  
 ical modelling to study the impact of climate change on the water flows of  
 mountain watersheds. *Journal of Hydrology*, *376*, 362 – 377.
- Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B.,  
 & Calzadilla, A. (2009). Global biomass production potentials exceed ex-  
 pected future demand without the need for cropland expansion. *Nature Com-  
 munications*, *6*.
- McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Sriniv-  
 asa, C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S.  
 (2015). Integrated assessments of the impacts of climate change on agricul-  
 ture: An overview of AgMIP regional research in South Asia. *Chapter in:*  
*Handbook of Climate Change and Agroecosystems*, (pp. 201–218).
- Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical  
 weather responsiveness of crop yields: US evidence and implications for  
 the agricultural impacts of climate change. *Environmental Research Letters*,  
*12*.
- Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate  
 change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, *8*.

- Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. (2017). Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, *10*, 1403–1422.
- Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T. (1997). Effect of CO<sub>2</sub> enrichment on carbon and nitrogen interaction in wheat and soybean. *Soil Science and Plant Nutrition*, *43*, 789–798.
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, *91*, 1290 – 1300.
- Olin, S., Schurgers, G., Lindeskog, M., Wårldin, D., Smith, B., Bodin, P., Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue C:N to changes in atmospheric CO<sub>2</sub> and N management in the main wheat regions of western europe. *Biogeosciences*, *12*, 2489–2515. doi:10.5194/bg-12-2489-2015.
- Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in field crop production. *Soil Science and Plant Nutrition*, *38*, 553–564.
- Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., & Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator. *Geoscientific Model Development*, *8*, 1139–1155.
- Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop yields and their variability at different levels of global warming. *Earth System Dynamics*, *9*, 479–496.
- Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff, S., & Gerten, D. (2015). Emulating global climate change impacts on crop yields. *Statistical Modelling*, *15*, 499–525.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pasos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
- Pirttioja, N., Carter, T., Fronzek, S., Bind, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minetti, J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M., Semenov, M., Slawinski, C., Strattonovich, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. *Climate Research*, *65*, 87–105.
- Poppick, A., McInerney, D. J., Moyer, E. J., & Stein, M. L. (2016). Temperatures in transient climates: Improved methods for simulations with evolving temporal covariances. *Ann. Appl. Stat.*, *10*, 477–505. URL: <https://doi.org/10.1214/16-AOAS903>. doi:10.1214/16-AOAS903.
- Porter et al. (IPCC) (2014). Food security and food production systems. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In C. F. et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK: Cambridge University Press.
- Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of monthly growing areas of 26 irrigated crops.
- Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly Irrigated and Rainfed crop Areas around the Year 2000: A New High-Resolution Data Set for Agricultural and Hydrological Modeling. *Global Biogeochemical Cycles*, *24*, GB1011.
- Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E., & Arneth, A. (2016). Climate analogues suggest limited potential for intensification of production on current croplands under climate change. *Nature Communications*, *7*, 12608.
- Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus A: Dynamic Meteorology and Oceanography*, *58*, 461–472.
- Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the reduction and sensitivity analysis of complex environmental models. *Environmental Modelling & Software*, *34*, 1 – 4.
- Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, *48*.
- Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, *12*.
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, *111*, 3268–3273.
- Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P., Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ew-

- ert, F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182.
- Rosenzweig, C., Ruane, A. C., Antle, J., Elliott, J., Ashfaq, M., Chattha, A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G., Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M., Müller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M., Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., & Wiebe, K. (2018). Coordinating AgMIP data and models across global and regional scales for 1.5°C and 2.0°C assessments. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 376.
- Ruane, A., I. Hudson, N., Asseng, S., Camarrano, D., Ewert, F., Martre, P., Boote, K., Thorburn, P., Aggarwal, P., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A., Doltra, J., Gayler, S., Goldberg, R., Grant, R., & Wolf, J. (2016). Multi-wheat-model ensemble responses to interannual climate variability. *Environmental Modelling and Software*, 81, 86–101.
- Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason-D'Croz, D., Müller, C., Porter, C., Phillips, M. M., Raymundo, R. M., Sands, R., Valdivia, R. O., White, J. W., Wiebe, K., & Rosenzweig, C. (2018). Biophysical and economic implications for agriculture of +1.5° and +2.0°C global warming using AgMIP Coordinated Global and Regional Assessments. *Climate Research*, 76, 17–39.
- Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown, D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013). Climate change impact uncertainties for maize in panama: Farm information, climate projections, and yield sensitivities. *Agricultural and Forest Meteorology*, 170, 132 – 145.
- Ruane, A. C., Goldberg, R., & Chrysanthacopoulos, J. (2015). Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agric. Forest Meteorol.*, 200, 233–248.
- Ruane, A. C., McDermid, S., Rosenzweig, C., Baigorria, G. A., Jones, J. W., Romero, C. C., & Cecil, L. D. (2014). Carbon-temperature-water change analysis for peanut production under climate change: A prototype for the agmip coordinated climate-crop modeling project (c3mp). *Glob. Change Biol.*, 20, 394–407. doi:10.1111/gcb.12412.
- Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift*, 19, 135–141.
- Ruiz-Ramos, M., Ferrise, R., Rodriguez, A., Lorite, I., Bindi, M., Carter, T., Fronzek, S., Palosuo, T., Pirttioja, N., Baranowski, P., Buis, S., Cammarano, D., Chen, Y., Dumont, B., Ewert, F., Gaiser, T., Hlavinka, P., Hoffmann, H., Hhn, J., Jurecka, F., Kersebaum, K., Krzyszczak, J., Lana, M., Mechiche-Alami, A., Minet, J., Montesino, M., Nendel, C., Porter, J., Ruget, F., Semenov, M., Steinmetz, Z., Strattonovich, P., Supit, I., Tao, F., Trnka, M., de Wit, A., & Ritter, R. (2018). Adaptation response surfaces for managing wheat under perturbed climate and co2 in a mediterranean environment. *Agricultural Systems*, 159, 260 – 274. doi:doi.org/10.1016/j.agbsy.2017.01.009.
- Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19, 607–620.
- Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frieler, K. (2017). Consistent negative response of US crops to high temperatures in observations and crop models. *Nature Communications*, 8, 13931.
- Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106, 15594–15598.
- Snyder, A., Calvin, K. V., Phillips, M., & Ruane, A. C. (2018). A crop yield change emulator for use in gcam and similar models: Persephone v1.0. *Geoscientific Model Development Discussions*, 2018, 1–42.
- Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety*, 94, 1735 – 1763.
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93, 485–498.
- Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters*, 35.
- Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., & Martin, J. F. (2014). Modeling sugarcane yield with a process-based model from site to continental scale: Uncertainties arising from model structure and parameter values. *Geoscientific Model Development*, 7, 1225–1245.
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences*, 111, 3228–3232.
- White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Methodologies for simulating impacts of climate change on crop production. *Field Crops Research*, 124, 357 – 368.
- Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T.,

- 1132 Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance  
1133 against site observations of irrigated maize from Mead, Nebraska. *Geosci-  
1134 entific Model Development*, *10*, 1291–1320.
- 1135 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield vari-  
1136 ability in JULES-crop and implications for forcing with seasonal weather  
1137 forecasts. *Geoscientific Model Development*, *8*, 3987–3997.
- 1138 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Land-  
1139 bouwkundige Onderzoeken* : 64.6, .
- 1140 Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato  
1141 yields on changes in climate and co2. *Agricultural and Forest Meteorology*,  
1142 *112*, 217 – 231.
- 1143 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao,  
1144 Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A.,  
1145 Li, T., Lin, E., Liu, Q., Martre, P., Mller, C., Peng, S., Peuelas, J., Ruane,  
1146 A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng,  
1147 S. (2017). Temperature increase reduces global yields of major crops in four  
1148 independent estimates. *Proc. Natl. Acad. Sci.*, *114*, 9326–9331.