

# 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 only to capture observed historical yields, using  
local weather data and a historical crop simulation. These ef-  
forts do share important common features: all emulate annual  
crop yields across the entire scenario or scenarios, and when fu-  
ture scenarios are considered, they are non-stationary, i.e. their  
input climate parameters evolve over time.

An alternative approach is to construct a training set of multi-  
ple stationary scenarios in which parameters 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 effects of different variables that  
impact yields but that are highly correlated in realistic future  
scenarios (e.g. CO<sub>2</sub> and temperature). Second, it allows making  
a distinction between year-over-year yield variations and cli-  
matological changes, which may involve different responses to  
the particular climate regressors used (e.g. Ruane et al., 2016).  
For example, if year-over-year yield variations are driven pre-  
dominantly by variations in the distribution of temperatures  
throughout the growing season, and long-term climate changes  
are driven predominantly by shifts in means, then regressing  
on the mean growing season temperature will produce different

135 yield responses at annual vs. climatological timescales.

136 Systematic parameter sweeps have begun to be used in  
137 crop modeling, with early efforts in 2015 (Makowski et al.,  
138 2015, Pirttioja et al., 2015), and several recent studies in 2018  
139 (Fronzek et al., 2018, Snyder et al., 2018, Ruiz-Ramos et al.,  
140 2018). All three studies sample multiple perturbations to tem-  
141 perature and precipitation (with Snyder et al. (2018) and Ruiz-  
142 Ramos et al. (2018) adding CO<sub>2</sub> as well), in 132, 99 and approx-  
143 imately 220 different combinations, respectively, and take ad-  
144 vantage of the structured training set to construct emulators of  
145 climatological mean yields, omitting year-over-year variations.  
146 The main limitation in these studies is of geographic extent with  
147 each study focusing on a limited number of sites, and Fronzek  
148 et al. (2018) and Ruiz-Ramos et al. (2018) simulate only wheat  
149 (over several models), while Snyder et al. (2018) analyzes four  
150 crops but with a single model (GCAM).

151 In this paper we describe a new comprehensive dataset de-  
152 signed to expand the parameter sweep approach still further.  
153 The Global Gridded Crop Model Intercomparison (GGCMI)  
154 Phase II experiment involves running a suite of process-based  
155 crop models across historical conditions perturbed by a set of  
156 discrete steps in different input parameters, including an ap-  
157 plied nitrogen dimension. The experimental protocol involves  
158 over 700 different parameter combinations for each model and  
159 crop, with simulations providing near-global coverage at a half  
160 degree spatial resolution. The experiment was conducted as  
161 part of the Agricultural Model Intercomparison and Improve-  
162 ment Project (AgMIP) (Rosenzweig et al., 2013, 2014), an in-  
163 ternational effort conducted under a framework similar to the<sub>169</sub>  
164 Climate Model Intercomparison Project (CMIP) (Taylor et al.,<sub>170</sub>  
165 2012, Eyring et al., 2016). The GGCMI protocol builds on the<sub>171</sub>  
166 AgMIP Coordinated Climate-Crop Modeling Project (C3MP)<sub>172</sub>  
167 (Ruane et al., 2014, McDermid et al., 2015) and will con-<sub>173</sub>  
168 tribute to the AgMIP Coordinated Global and Regional As-<sub>174</sub>

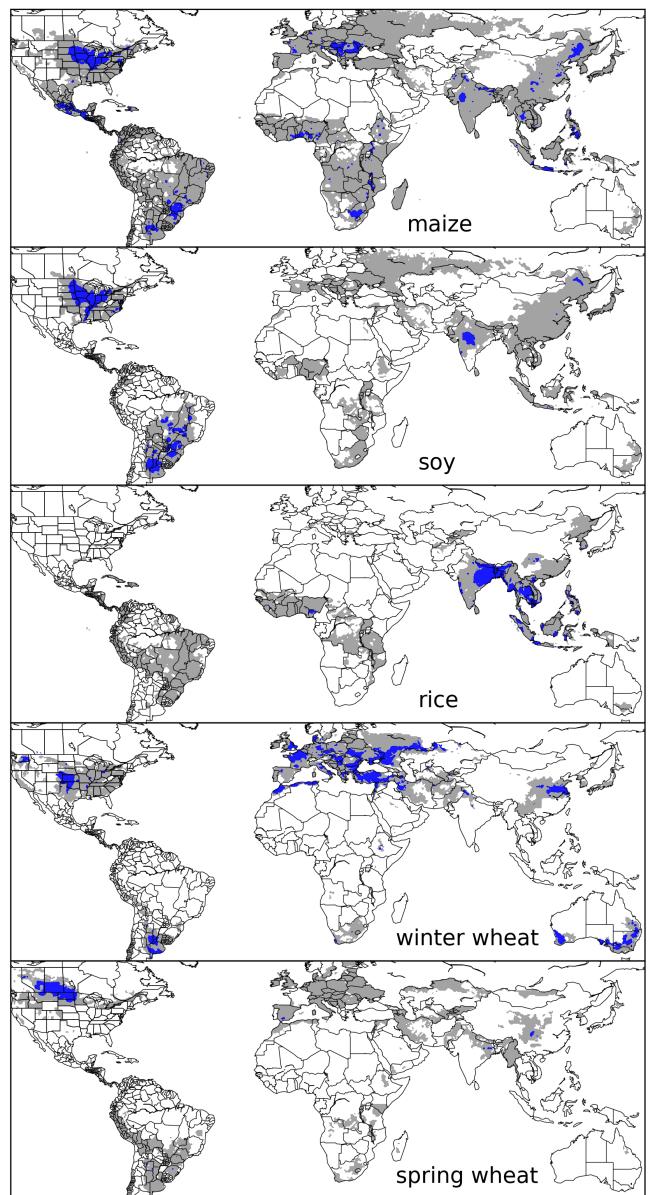


Figure 1: Presently cultivated area for rain-fed crops. Blue indicates grid cells with more than 20,000 hectares (~10% of the equatorial grid cells) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated areas for maize, rice, and soy are taken from the MIRCA2000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

sessments (CGRA) (Ruane et al., 2018, Rosenzweig et al., 2018). GGCMI Phase II is designed to allow addressing goals such as understanding where highest-yield regions may shift under climate change; exploring future adaptive management strategies; understanding how interacting input drivers affect crop yield; quantifying uncertainties across models and major

175 drivers; and testing strategies for producing lightweight em-<sup>201</sup>  
 176 ulators of process-based models. In this paper, we describe<sup>202</sup>  
 177 the GGCMI Phase II experiments, present initial results, and<sup>203</sup>  
 178 demonstrate that it is tractable to emulation.<sup>204</sup>

## 179 2. Simulation – Methods

180 GGCMI Phase II is the continuation of a multi-model com-<sup>207</sup>  
 181 parison exercise begun in 2014. The initial Phase I compared<sup>208</sup>  
 182 harmonized yields of 21 models for 19 crops over a 31-year<sup>209</sup>  
 183 historical (1980-2010) scenario with a primary goal of model<sup>210</sup>  
 184 evaluation (Elliott et al., 2015, Müller et al., 2017). Phase II<sup>211</sup>  
 185 compares simulations of 12 models for 5 crops (maize, rice,<sup>212</sup>  
 186 soybean, spring wheat, and winter wheat) over the same histor-<sup>213</sup>  
 187 ical time series (1980-2010) used in Phase I, but with individ-<sup>214</sup>  
 188 ual climate or management inputs adjusted from their historical<sup>215</sup>  
 189 values. The reduced set of crops includes the three major global<sup>216</sup>  
 190 cereals and the major legume and accounts for over 50% of hu-<sup>217</sup>  
 191 man calories (in 2016, nearly 3.5 billion tons or 32% of total<sup>218</sup>  
 192 global crop production by weight (Food and Agriculture Orga-<sup>219</sup>  
 193 nization of the United Nations, 2018).

194 The guiding scientific rationale of GGCMI Phase II is to pro-<sup>220</sup>  
 195 vide a comprehensive, systematic evaluation of the response<sup>221</sup>  
 196 of process-based crop models to different values for carbon<sup>222</sup>  
 197 dioxide, temperature, water, and applied nitrogen (collectively<sup>223</sup>  
 198 known as “CTWN”). The dataset is designed to allow re-<sup>224</sup>  
 199 searchers to:

- 200 • Enhance understanding of how models work by character-<sup>227</sup>

izing their sensitivity to input climate and nitrogen drivers.

- Study the interactions between climate variables and nitro-<sup>120</sup>  
 121 gen inputs in driving modeled yield impacts.
- Explore differences in crop response to warming across the<sup>122</sup>  
 123 Earth’s climate regions.
- Provide a dataset that allows statistical emulation of crop<sup>124</sup>  
 125 model responses for downstream modelers.
- Illustrate differences in potential adaptation via growing<sup>126</sup>  
 127 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 higher temperatures. An additional, identical set of scenarios (at the same C, T, W, and N levels) not shown or analyzed here simulate adaptive agronomy under climate change by varying the growing season for crop production. The resulting GGCMI

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>

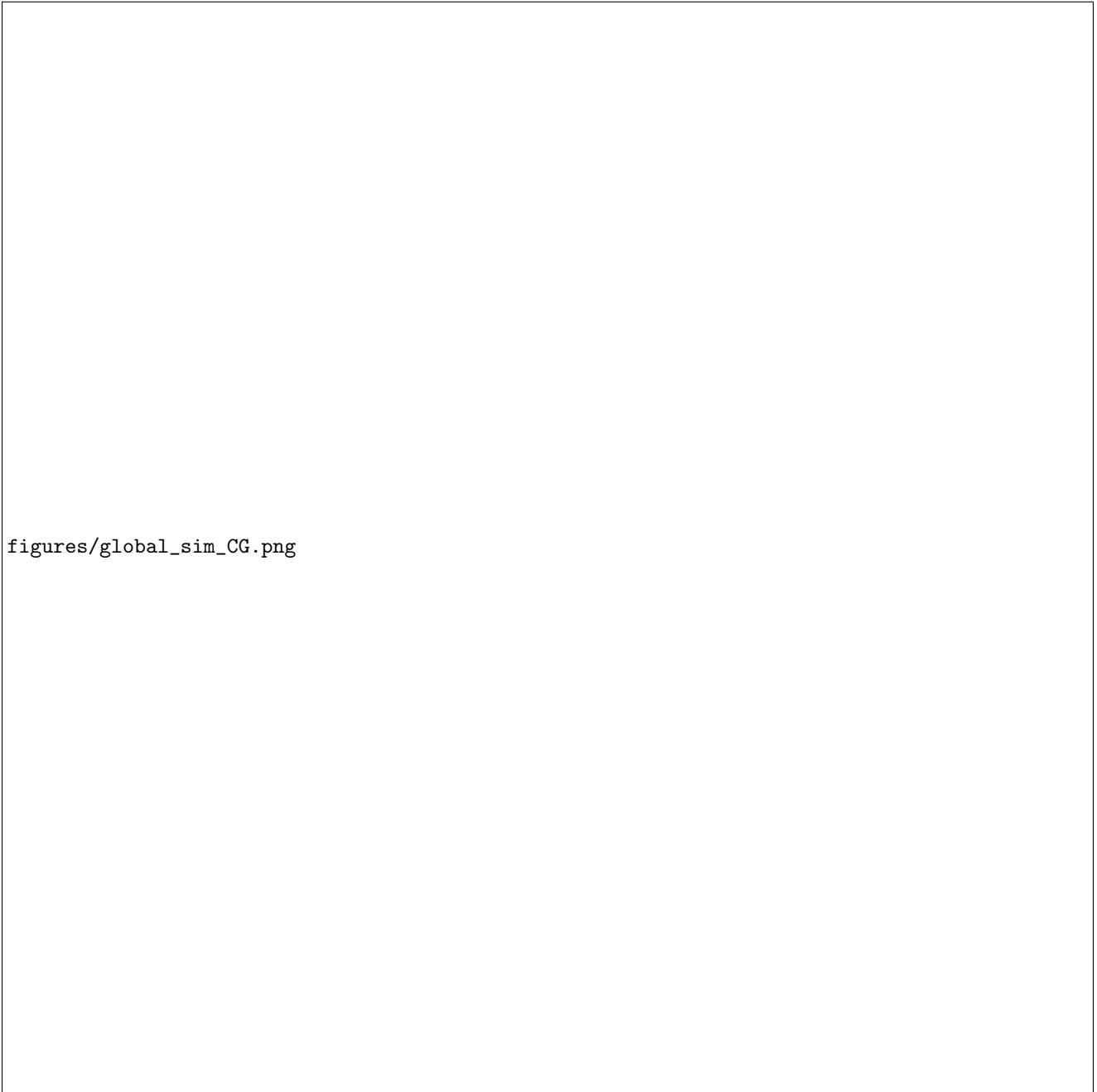
Table 1: GGCMI Phase II input variable test levels. Temperature and precipitation values indicate the perturbations from the historical, climatology. \* Only simulated by one model. W-percentage does not apply to the irrigated (W<sub>inf</sub>) simulations, which are all simulated at the maximum beneficial 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 Phase II dataset captures a distribution of crop responses over<sup>246</sup> to nitrogen, temperature, and water (e.g. alkalinity and salinity).  
 229 the potential space of future climate conditions.<sup>247</sup> No additional nitrogen inputs, such as atmospheric deposition,  
 230 The 12 models included in GGCMI Phase II are all mech-<sup>248</sup> are considered, but some model treatment of soil organic matter  
 231 anistic process-based crop models that are widely used in im-<sup>249</sup> may allow additional nitrogen release through mineralization.  
 232 pacts assessments (Table 2). Although some models share a<sup>250</sup> See Rosenzweig et al. (2014), Elliott et al. (2015) and Müller  
 233 common base (e.g. the LPJ family or the EPIC family of mod-<sup>251</sup> et al. (2017) for further details on models and underlying as-  
 234 els), they have subsequently developed independently. (For<sup>252</sup> sumptions.  
 235 more details on model genealogy, see Figure S1 in Rosenzweig<sup>253</sup> The participating modeling groups provide simulations at  
 236 et al. (2014).) Differences in model structure mean that several<sup>254</sup> any of four initially specified levels of participation, so the num-  
 237 key factors are not standardized across the experiment, includ-<sup>255</sup> ber of simulations varies by model, with some sampling only a  
 238 ing secondary soil nutrients, carry-over effects across growing<sup>256</sup> part of the experiment variable space. Most modeling groups  
 239 years including residue management and soil moisture, and the<sup>257</sup> simulate all five crops in the protocol, but some omitted one  
 240 extent of simulated area for different crops. Growing seasons<sup>258</sup> or more. Table 2 provides details of coverage for each model.  
 241 are standardized across models (with assumptions based on<sup>259</sup> Note that the three models that provide less than 50 simulations  
 242 Sacks et al. (2010) and Portmann et al. (2008, 2010)), but vary<sup>260</sup> are excluded from the emulator analysis.  
 243 by crop and by location on the globe. For example, maize is<sup>261</sup> Each model is run at 0.5 degree spatial resolution and cov-  
 244 sown in March in Spain, in July in Indonesia, and in December<sup>262</sup> ers all currently cultivated areas and much of the uncultivated  
 245 in Namibia. All stresses are disabled other than factors related<sup>263</sup> land area. (See Figure 1 for the present-day cultivated area of

<sup>264</sup> rain-fed crops, and Figure S1 in the Supplemental Material for<sup>266</sup> recently cultivated areas because cultivation will likely shift under  
<sup>265</sup> irrigated crops.) Coverage extends considerably outside cur-<sup>267</sup> climate change. However, areas are not simulated if they are



figures/global\_sim(CG.png)

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 pmm and 200 kg ha<sup>-1</sup> yr<sup>-1</sup>). Y-axis is fractional change in the regional average climatological potential yield relative to the baseline. The figure shows all modeled land area; see Figure S6 in the supplemental material for only currently-cultivated land. Panel text gives the percentage of rain-fed maize presently cultivated in each climate zone (data from Portmann et al., 2010). Box-and-whiskers plots show distribution across models, with median marked. Box edges are first and third quartiles, i.e. box height is the interquartile range (IQR). Whiskers extend to maximum and minimum of simulations but are limited at 1.5-IQR; otherwise the outlier is shown. Models generally agree in most climate regions (other than cold continental), with projected changes larger than inter-model variance. Outliers in the tropics (strong negative impact of temperature increases) are the pDSSAT model; outliers in the high-rainfall case (strong positive impact of precipitation increases) are the JULES model. Inter-model variance increases in the case where precipitation is reduced, suggesting uncertainty in model response to water limitation. The right panel with global changes shows yield responses to an globally uniform temperature shift; note that these results are not directly comparable to simulations of more realistic climate scenarios with the same global mean change.

assumed to remain non-arable even under an extreme climate<sup>301</sup>  
change; these regions include Greenland, far-northern Canada,<sup>302</sup>  
Siberia, Antarctica, the Gobi and Sahara Deserts, and central<sup>303</sup>  
Australia.<sup>304</sup>

All models produce as output crop yields (tons  $\text{ha}^{-1}$  year $^{-1}$ )<sup>305</sup>  
for each 0.5 degree grid cell. Because both yields and yield<sup>306</sup>  
changes vary substantially across models and across grid cells,<sup>307</sup>  
we primarily analyze relative change from a baseline. We take<sup>308</sup>  
as the baseline the scenario with historical climatology (i.e. T<sup>309</sup>  
and P changes of 0), C of 360 ppm, and applied N at 200 kg<sup>310</sup>  
 $\text{ha}^{-1}$ . We show absolute yields in some cases to illustrate geo-<sup>311</sup>  
graphic differences in yields for a single model.<sup>312</sup>

### 3. Simulation – Results

Crop models in the GGCMI Phase II ensemble show broadly<sup>313</sup>  
consistent responses to climate and management perturbations<sup>314</sup>  
in most regions, with a strong negative impact of increased tem-<sup>315</sup>  
perature in all but the coldest regions. We illustrate this result<sup>316</sup>  
for rain-fed maize in Figure 2, which shows yields for the pri-<sup>317</sup>  
mary Köppen-Geiger climate regions (Rubel & Kottek, 2010).<sup>318</sup>  
In warming scenarios, models show decreases in maize yield in<sup>319</sup>  
the temperate, tropical, and arid regions that account for nearly<sup>320</sup>  
three-quarters of global maize production. These impacts are<sup>321</sup>  
robust for even moderate climate perturbations. In the temper-<sup>322</sup>  
ate zone, even a 1 degree temperature rise with other variables<sup>323</sup>  
held fixed leads to a median yield reduction that outweighs the<sup>324</sup>  
variance across models. A 6 degree temperature rise results in<sup>325</sup>  
median loss of ~25% of yields with a signal to noise of nearly<sup>326</sup>  
three. A notable exception is the cold continental region, where<sup>327</sup>  
models disagree strongly, extending even to the sign of impacts.<sup>328</sup>  
Other crops show similar responses to warming, with robust<sup>329</sup>  
yield losses in warmer locations and high inter-model variance<sup>330</sup>  
in the cold continental regions (Figure S7).<sup>331</sup>

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, most strongly in arid regions. Decreased rainfall  
amplifies yield losses and also increases inter-model variance  
more strongly, suggesting that models have difficulty represent-  
ing crop response to water stress XX - see reviewer comments?<sup>332</sup>  
We show only rain-fed maize here; see Figure S5 for the irri-  
gated 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<sup>333</sup>  
shows the geographic dependencies that underlie these results.<sup>334</sup>  
Figure 3 shows baseline and changes in the T+4 scenario for<sup>335</sup>  
rain-fed maize, soy, and rice in the multi-model ensemble mean,<sup>336</sup>  
with locations of model agreement marked. Absolute yield poten-<sup>337</sup>  
tials show strong spatial variation, with much of the Earth's<sup>338</sup>  
surface area unsuitable for any given crop. In general, mod-<sup>339</sup>  
els agree most on yield response in regions where yield poten-<sup>340</sup>  
tials are currently high and therefore where crops are currently<sup>341</sup>  
grown. Models show robust decreases in yields at low latitudes,<sup>342</sup>  
and highly uncertain median increases at most high latitudes.<sup>343</sup>  
For wheat crops see Figure S11; wheat projections are both<sup>344</sup>  
more uncertain and show fewer areas of increased yield in the<sup>345</sup>  
inter-model mean.

### 4. Emulation – Methods

As part of our demonstration of the properties of the GGCMI<sup>346</sup>  
Phase II dataset, we construct an emulator of 30-year clima-<sup>347</sup>  
tological mean yields. This approach is made possible by<sup>348</sup>  
the structured set of simulations involving systematic pertur-<sup>349</sup>  
bations. In the GGCMI Phase II dataset, the year-over-year re-<sup>350</sup>  
sponses are generally quantitatively distinct from (and larger<sup>351</sup>  
than) climatological mean responses. In the example of Figure<sup>352</sup>  
4, responses to year-over-year temperature variations are XX%<sup>353</sup>

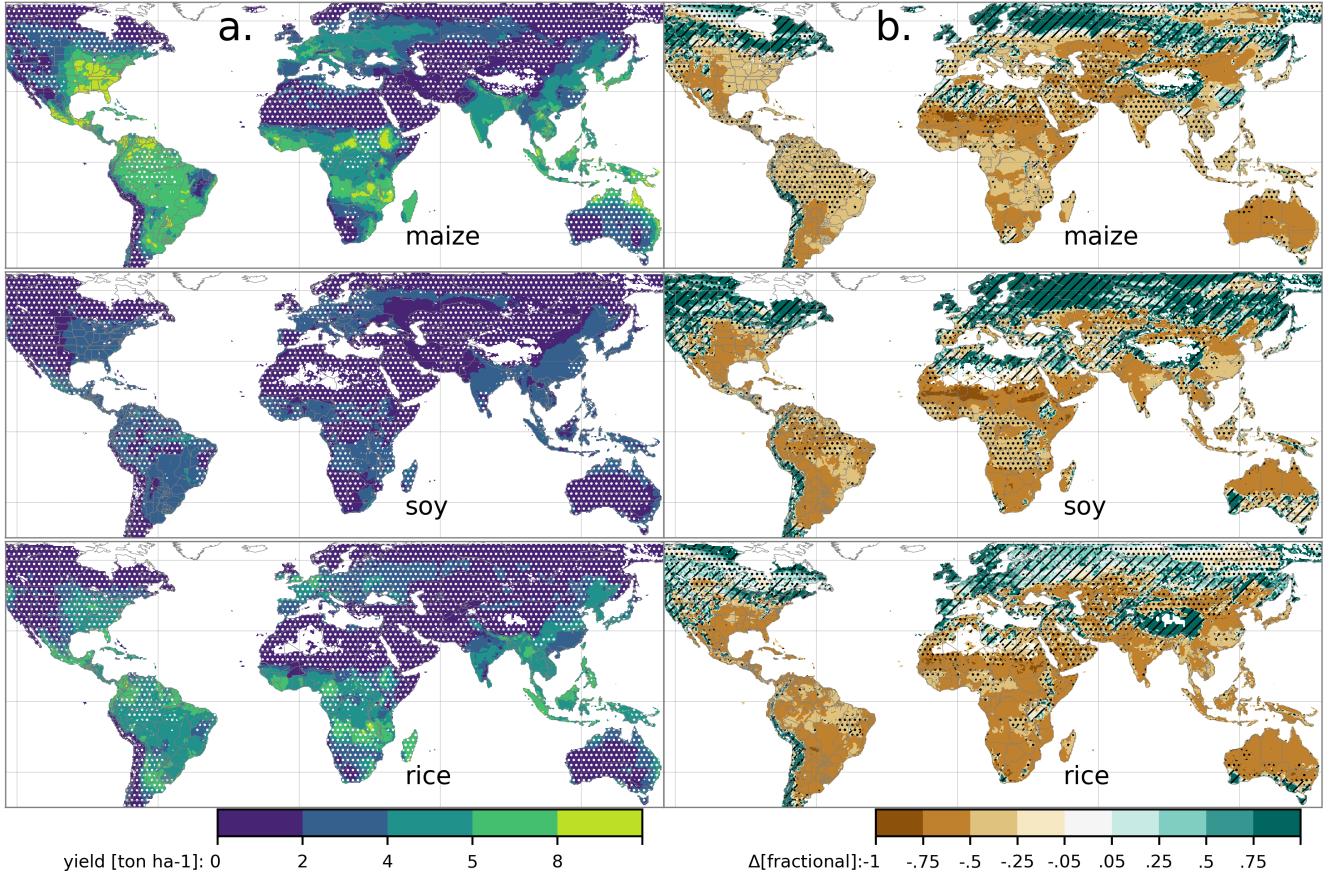


Figure 3: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure 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 ( $0.2 \text{ ton ha}^{-1}$ ). Right column (b) shows the multi-model mean fractional yield change in the extreme  $T + 4 \text{ }^{\circ}\text{C}$  scenario (with other inputs at baseline values). Areas without hatching or stippling are those where confidence in projections is high: the multi-model mean fractional change exceeds two standard deviations of the ensemble. ( $\Delta > 2\sigma$ ). Hatching indicates areas of low confidence ( $\Delta < 1\sigma$ ), and stippling areas of medium confidence ( $1\sigma < \Delta < 2\sigma$ ). Crop model results in cold areas, where yield impacts are on average positive, also have the highest uncertainty.

334 larger than those to long-term perturbations in the baseline case,<sup>347</sup>  
 335 and larger still under warmer conditions, rising to **XX%** in the<sup>348</sup>  
 336  $T+6$  case. The stronger year-over-year response under warmer<sup>349</sup>  
 337 conditions also manifests as a wider distribution of yields (Fig-<sup>350</sup>  
 338 ure 5). As discussed previously, year-over-year and climatolog-<sup>351</sup>  
 339 ical responses can differ for many reasons including memory<sup>352</sup>  
 340 in the crop model, lurking covariants, and differing associated<sup>353</sup>  
 341 distributions of daily growing-season daily weather (e.g. Ruane<sup>354</sup>  
 342 et al., 2016). Note that the GGCMI Phase II datasets do not<sup>355</sup>  
 343 capture one climatological factor, potential future distributional<sup>356</sup>  
 344 shifts, because all simulations are run with fixed offsets from<sup>357</sup>  
 345 the historical climatology. Prior work has suggested that mean<sup>358</sup>  
 346 changes are the dominant drivers of climatological crop yield<sup>359</sup>

shifts in non-arid regions (e.g. Glotter et al., 2014).

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  $\text{CO}_2$ , 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. The higher-order terms are necessary to capture any nonlinear responses, which are well-documented in observations for temperature and water perturbations (e.g. Schlenker & Roberts (2009) for T and He et al. (2016) for W). We include interaction terms (both linear and higher-order) because past studies have shown them to be significant effects. For example,

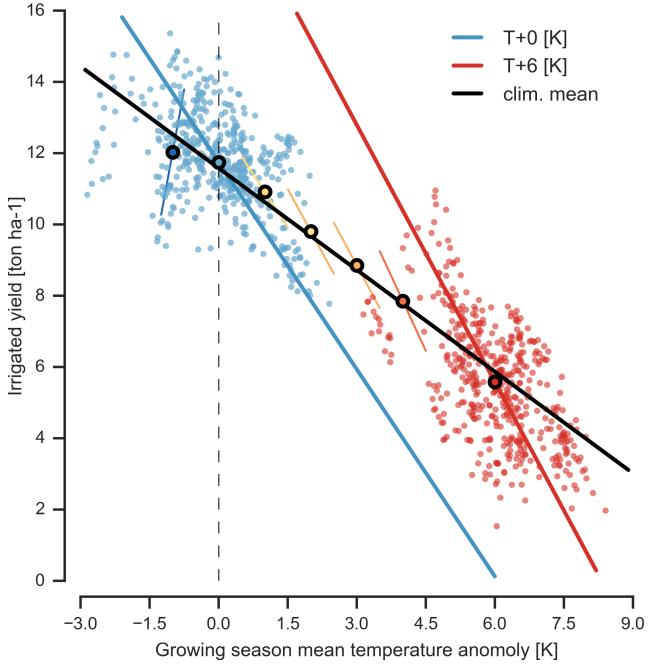


Figure 4: Example showing distinction between crop yield responses to year-to-year and climatological mean temperature shifts. Figure shows irrigated maize for a representative high-yield region (nine adjacent grid cells in northern Iowa) from the pDSSAT model, for the baseline 1981-2010 historical climate (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 ( $-1, +0, +1, +2, +3, +6$ ). 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 XX-XX% larger than those to long-term climate perturbations, rising under warmer conditions.

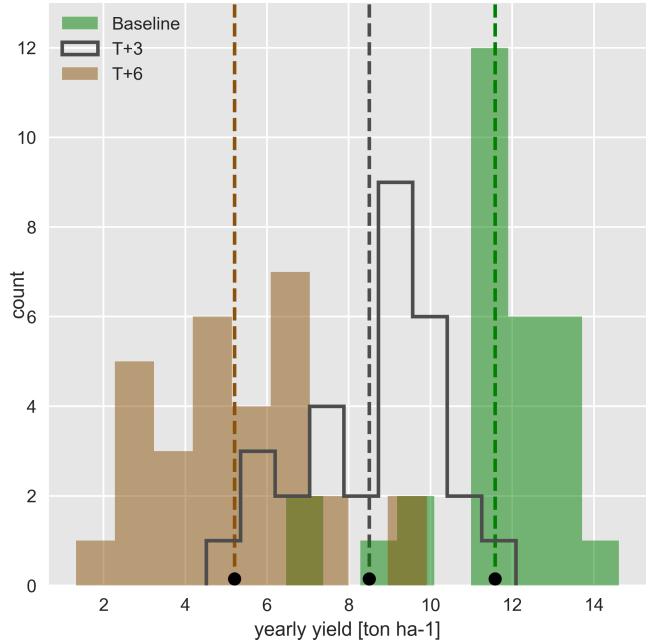


Figure 5: Example showing climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for one of the high-yield grid cells of Figure 4 (in northern Iowa) from the pDSSAT model, for the baseline 1981-2010 historical climate (green) and for scenarios with temperature shifted by T+3 (black) and T+6 K (brown), 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 / black dots).

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 (Ozaki 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 parametrization 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. The simpler functional form used here allows emulation at grid-cell level **with low noise? how do you quantify this?**. The emulation therefore indirectly includes any yield response to geographically distributed factors such as soil type, insolation, and the baseline climate itself.

#### 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 the aggregate mean squared error with increasing terms and eliminate those terms that do not contribute significant reductions. See supplemental documents for more details. We select terms by applying the feature selection process to the three models that provided the complete set of 672 rain-fed simulations (pDSSAT, EPIC-TAMU, and LPJmL); the resulting choice of terms is then applied for all emulators.

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

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

To fit the parameters  $K$ , we use a Bayesian Ridge probabilistic estimator (MacKay, 1991), which reduces volatility in parameter estimates when the sampling is sparse, by weight-ing 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 provided a limited number of cases or for low-yield geographic regions where some modeling groups did not run all scenarios. We do not attempt to emulate models that provided less than 50 simulations. The lowest number of simulations emulated across the full parameter space is then 130 (for the PEPIC model). The resulting parameter matrices for all crop model emulators are available on request [give location?](#), as are the raw simulation data and a Python application to emulate yields. The yield output for a single GGCMI Phase II model that simulates all scenarios and all five crops is  $\sim 12.5$  GB; the emulator is  $\sim 100$  MB, a reduction by over two orders of magnitude.

## 5. Emulation – Results

Emulation provides not only a computational tool but a means of understanding and interpreting crop yield response across the parameter space. Emulation is only possible when crop yield responses are sufficiently smooth and continuous to allow fitting with a relatively simple functional form, but this condition largely holds in the GGCMI Phase II simulations. Responses are quite diverse across locations, crops, and models, but in most cases local responses are regular enough to permit emulation. We show illustrations of emulation fidelity in this section; for more detailed discussion see Appendix XX.

Crop yield responses are geographically diverse, even in high-yield and high-cultivation areas Figure 6 illustrates geographic diversity for a single crop and model (rain-fed maize in pDSSAT); this heterogeneity supports the choice of emulating at the grid cell level. Each panel in Figure 6 shows simulated yield output from scenarios varying only along a single dimension ( $CO_2$ , temperature, precipitation, or nitrogen addi-

tion), with other inputs held fixed at baseline levels, compared<sup>454</sup> to the full 4D emulation across the parameter space. Yields<sup>455</sup> evolve smoothly across the space sampled, and the polynomial<sup>456</sup> fit captures the climatological response to perturbations. Crop<sup>457</sup> yield responses generally follow similar functional forms across<sup>458</sup> models, though with a large spread in magnitude likely due to<sup>459</sup> the lack of calibration. Figure 7 illustrates inter-model diversity<sup>460</sup> for a single crop and location (rain-fed maize in northern Iowa,<sup>461</sup> also shown in Figure 6). Differences in response shape can lead<sup>462</sup> to differences in the fidelity of emulation, though comparison<sup>463</sup> here is complicated by the different sampling regimes across<sup>464</sup> models. Note that models are most similar in their responses<sup>465</sup> to temperature perturbations. For this location and crop, CO<sub>2</sub><sup>466</sup> fertilization effects can range from ~5–50%, and nitrogen re-<sup>467</sup>sponses from nearly flat to a 60% drop in the lowest-application<sup>468</sup> simulation.<sup>469</sup>

While the nitrogen dimension is important, it is also the most<sup>470</sup> problematic to emulate in this work because of its limited sam-<sup>471</sup>pling. The GGCMI Phase II protocol specified only three ni-<sup>472</sup>

rogen levels (10, 60 and 200 kg N y<sup>-1</sup> ha<sup>-1</sup>), so a third-order fit would be over-determined but a second-order fit can result in potentially unphysical results. Steep and nonlinear declines in yield with lower nitrogen levels mean that some regressions imply a peak in yield between the 100 and 200 kg N y<sup>-1</sup> ha<sup>-1</sup> levels. While it is possible that over-application of nitrogen at the wrong time in the growing season could lead to reduced yields, these features are almost certainly an artifact of under-sampling. In addition, the polynomial fit cannot capture the well-documented saturation effect of nitrogen application (e.g. Ingestad, 1977) as accurately as would be possible with a non-parametric model.

The emulation fidelity demonstrated here is sufficient to allow using emulated response surfaces to compare model responses and derive insight about impacts projections. Because the emulator or “surrogate model” transforms the discrete simulation sample space into a continuous response surface at any geographic scale, it can be used for a variety of applications, including construction of continuous damage functions. As an

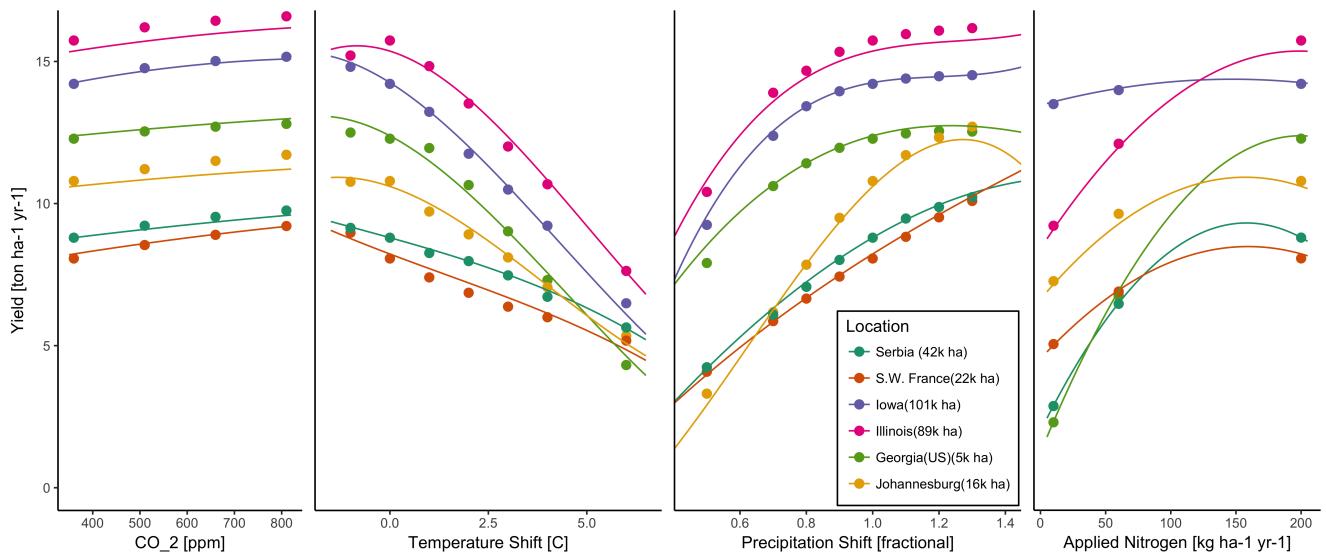


Figure 6: Illustration of spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation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

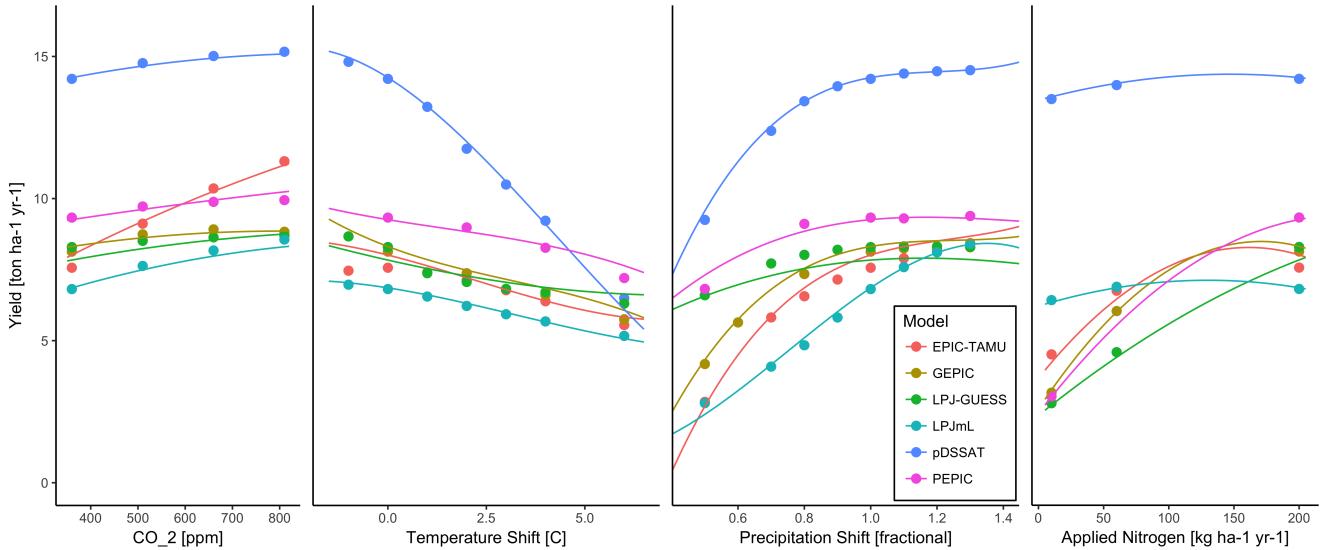


Figure 7: Illustration of across-model variations in yield response. Figures shows simulations and emulations from six models for rain-fed maize in the same Iowa grid cell shown in Figure 6, with the same plot conventions. Models that do not simulate the nitrogen dimension are omitted for clarity. Note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

example, we show a damage function constructed from the 4D<sub>492</sub> emulation, aggregated to global yield, with simulated values<sub>493</sub> shown for comparison (Figure 8, which shows maize on cur-<sub>494</sub> rently cultivated land; see Figures S16- S19 for other crops and<sub>495</sub> dimensions). The emulated values closely match simulations<sub>496</sub> even at this aggregation level. Note that these functions are<sub>497</sub> presented only as examples and do not represent true global<sub>498</sub> projections, because they are developed from simulation data<sub>499</sub> with a uniform temperature shift while increases in global mean<sub>500</sub> temperature should manifest non-uniformly. The global cover-<sub>501</sub> age of the GGCMI Phase II simulations allows impacts mod-<sub>502</sub> elers to apply arbitrary geographically-varying climate projec-<sub>503</sub> tions, as well as arbitrary aggregation masks, to develop dam-<sub>504</sub> age functions for any climate scenario and any geopolitical or<sub>505</sub> geographic level.

## 6. Conclusions and discussion

The GGCMI Phase II experiment provides a database tar-<sub>509</sub> geted to allow detailed study of crop yields from process-based<sub>510</sub> models under climate change. The experiment is designed to<sub>511</sub>

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

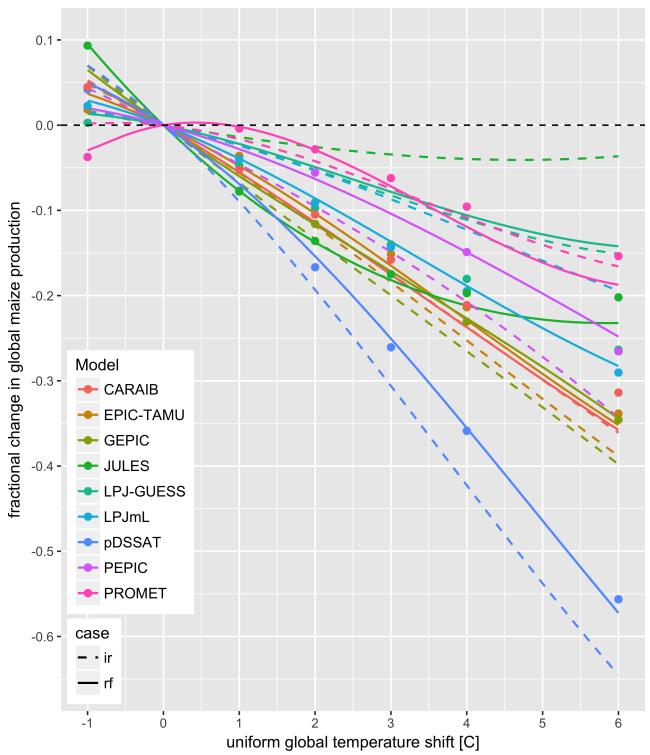


Figure 8: Global emulated damages for maize on currently cultivated lands<sup>539</sup> for the GGCMI Phase II models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating<sup>540</sup> up emulated values at the grid cell level, not from a regression of global mean yields.) Lines are emulations for rain-fed (solid) and irrigated (dashed) crops;<sup>541</sup> for comparison, dots are the simulated values for the rain-fed case. For most models, irrigated crops show a sharper reduction than do rain-fed because of the<sup>542</sup> locations of cultivated areas: irrigated crops tend to be grown in warmer areas where impacts are more severe for a given temperature shift. (The exceptions<sup>543</sup> are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures S16- S19 in the supplemental material.

<sup>525</sup> rice is not generally grown in water-limited conditions).

<sup>526</sup> Third, we show that even the relatively limited GGCMI  
<sup>527</sup> Phase II sampling space allows emulation of the climatological  
<sup>528</sup> response of crop models with a relatively simple reduced-  
<sup>529</sup> form statistical model. The systematic parameter sampling in  
<sup>530</sup> the GGCMI Phase II procedure provides information on the in-  
<sup>531</sup> fluence of multiple interacting factors in a way that single pro-  
<sup>532</sup> jections cannot, and emulating the resulting response surface  
<sup>533</sup> then produces a tool that can aid in both physical interpretation  
<sup>534</sup> of the process-based models and in assessment of agricultural  
<sup>535</sup> impacts under arbitrary climate scenarios. Emulating the cli-  
<sup>536</sup> matological response isolates long-term impacts from any con-  
<sup>537</sup> founding factors that complicate year-over-year changes, and  
<sup>538</sup> the use of simple functional forms offer the possibility of phys-  
<sup>539</sup> ical interpretation of parameter values. Care should be taken in  
<sup>540</sup> applying relationships developed at the yearly level to shifts in  
<sup>541</sup> the mean climatology. We anticipate that systematic parameter  
<sup>542</sup> sampling will become the norm in future model intercompari-  
<sup>543</sup> son exercise.

<sup>544</sup> While the GGCMI Phase II database should offer the foun-  
<sup>545</sup> dation for multiple future studies, several cautions need to be  
<sup>546</sup> noted. Because the simulation protocol was designed to focus  
<sup>547</sup> on change in yield under climate perturbations and not on repli-  
<sup>548</sup> cating real-world yields, the models are not formally calibrated  
<sup>549</sup> so cannot be used for impacts projections unless in used in con-  
<sup>550</sup> junction with historical data (or data products). Because the  
<sup>551</sup> GGCMI Phase II simulations apply uniform perturbations to  
<sup>552</sup> historical climate inputs, they do not sample changes in higher  
<sup>553</sup> order moments, and cannot address the additional crop yield  
<sup>554</sup> impacts of potential changes in climate variability. Although  
<sup>555</sup> distributional changes in model projections are fairly uncertain  
<sup>556</sup> at present, follow-on experiments may wish to consider them.  
<sup>557</sup> Several recent studies have described procedures for generating  
<sup>558</sup> simulations that combine historical data with model projections

<sup>512</sup> shows steeper yield reductions under warming than does rain-<sup>546</sup>  
<sup>513</sup> fed maize when considered only over currently cultivated land.<sup>547</sup>  
<sup>514</sup> The effect results from geographic differences in cultivation. In<sup>548</sup>  
<sup>515</sup> any given location, irrigation increases crop resiliency to tem-<sup>549</sup>  
<sup>516</sup> perature increase, but irrigated maize is grown in warmer loca-<sup>550</sup>  
<sup>517</sup> tions where the impacts of warming are more severe (Figures<sup>551</sup>  
<sup>518</sup> S5-S6). The same behavior holds for rice and winter wheat,<sup>552</sup>  
<sup>519</sup> but not for soy or spring wheat (Figures S8-S10). Irrigated<sup>553</sup>  
<sup>520</sup> wheat and maize are also more sensitive to nitrogen fertiliza-<sup>554</sup>  
<sup>521</sup> tion levels than are analogous non-irrigated crops, presumably<sup>555</sup>  
<sup>522</sup> because those rain-fed crops are limited by water as well as<sup>556</sup>  
<sup>523</sup> nitrogen availability (Figure S19). (Soy as an efficient atmo-<sup>557</sup>  
<sup>524</sup> spheric nitrogen-fixing is relatively insensitive to nitrogen, and<sup>558</sup>

559 of not only mean changes in temperature and precipitation but<sub>592</sub> changes in their marginal distributions or temporal dependence.<sub>593</sub>

561 The GGCMI Phase II output dataset invites a broad range of<sub>594</sub> potential future avenues of analysis. A major target area in<sub>595</sub> involves studying the models themselves with a detailed exami-<sub>596</sub> nation of interaction terms between the major input drivers, a<sub>597</sub> more robust quantification of the sensitivity of different models<sub>598</sub> to the input drivers, and comparisons with field-level experi-<sub>599</sub> mental data. The parameter space tested in GGCMI Phase II<sub>600</sub> will allow detailed investigations into yield variability and re-<sub>601</sub> sponse to extremes under changing management and CO<sub>2</sub> lev-<sub>602</sub> els. As mentioned previously, the database allows study of geo-<sub>603</sub> graphic shifts in optimal growing regions for different crops and<sub>604</sub> studying the viability of switching crop types in some areas.<sub>605</sub> The output dataset also contains other runs and variables not<sub>606</sub> analyzed or shown here. Runs include several which allowed<sub>607</sub> adaptation to climate changes by altering growing seasons, and<sub>608</sub> additional variables include above ground biomass, LAI, and<sub>609</sub> root biomass (as many as 25 output variables for some models).

577 Emulation studies that are possible include a more systematic<sub>610</sub> evaluation of different statistical model specifications and for-<sub>611</sub> mal calculation of uncertainties in derived parameters.

581 The development of multi-model ensembles such as GGCMI<sub>612</sub> Phase II provides a way to begin to better understand crop re-<sub>613</sub> sponses to a range of potential climate inputs, improve process<sub>614</sub> based models, and explore the potential benefits of adaptive re-<sub>615</sub> sponses included shifting growing season, cultivar types and<sub>616</sub> cultivar geographic extent.

## 587 7. Acknowledgments

588 We thank Michael Stein and Kevin Schwarzwald, who pro-<sub>621</sub> vided helpful suggestions that contributed to this work. This re-<sub>622</sub> search was performed as part of the Center for Robust Decision-<sub>623</sub> Making on Climate and Energy Policy (RDCEP) at the Univer-<sub>624</sub>

sity of Chicago, and was supported through a variety of sources. RDCEP is funded by NSF grant #SES-1463644 through the Decision Making Under Uncertainty program. J.F. was 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).

## 578 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,

625 uniform nitrogen application levels that are not realistic for in-<sup>659</sup>  
626 individual countries. We evaluate up to three control runs for each<sup>660</sup>  
627 model, since some modeling groups provide historical runs for<sup>661</sup>  
628 three different nitrogen levels.

629 Figure 9 shows the Pearson time series correlation between<sup>663</sup>  
630 the simulation model yield and FAO yield data. Figure 9 can be<sup>664</sup>  
631 compared to Figures 1,2,3,4 and 6 in Müller et al. (2017). The<sup>665</sup>  
632 results are mixed, with many regions for rice and wheat be-<sup>666</sup>  
633 ing difficult to model. No single model is dominant, with each<sup>667</sup>  
634 model providing near best-in-class performance in at least one<sup>668</sup>  
635 location-crop combination. The presence of very few vertical<sup>669</sup>  
636 dark green color bars clearly illustrates the power of a multi-<sup>670</sup>  
637 model intercomparison project like the one presented here. The<sup>671</sup>  
638 ensemble mean does not beat the best model in each case, but<sup>672</sup>  
639 shows positive correlation in over 75% of the cases presented<sup>673</sup>  
640 here. The EPIC-TAMU model performs best for soy, CARIAB,<sup>674</sup>  
641 EPIC-TAMU, and PEPIC perform best for maize, PROMET<sup>675</sup>  
642 performs best for wheat, and the EPIC family of models per-<sup>676</sup>  
643 form best for rice. Reductions in skill over the performance<sup>677</sup>  
644 illustrated in Müller et al. (2017) can be attributed to the nitro-  
645 gen levels or lack of calibration in some models.

646 \*\*\* or harmonization \*\*\* Christoph

647 Soy is qualitatively the easiest crop to represent (except in<sup>679</sup>  
648 Argentina), which is likely due in part to the invariance of the<sup>680</sup>  
649 response to nitrogen application (soy fixes atmospheric nitrogen<sup>681</sup>  
650 very efficiently). Comparison to the FAO data is therefore easier<sup>682</sup>  
651 than the other crops because the nitrogen application levels do<sup>683</sup>  
652 not matter. US maize has the best performance across models,<sup>684</sup>  
653 with nearly every model representing the historical variability<sup>685</sup>  
654 to a reasonable extent. Especially good example years for US<sup>686</sup>  
655 maize are 1983, 1988, and 2004 (top left panel of Figure 9),<sup>687</sup>  
656 where every model gets the direction of the anomaly compared<sup>688</sup>  
657 to surrounding years correct. 1983 and 1988 are famously bad  
658 years for US maize along with 2012 (not shown). US maize

662 is possibly both the most uniformly industrialized (in terms of  
663 management practices) crop and the one with the best data col-  
664 lection in the historical period of all the cases presented here.

665 The FAO data is at least one level of abstraction from ground  
666 truth in many cases, especially in developing countries. The  
667 failure of models to represent the year-to-year variability in rice  
668 in some countries in southeast Asia is likely partly due to model  
669 failure and partly due to lack of data. It is possible to speculate  
670 that the difference in performance between Pakistan (no suc-  
671 cessful models) and India (many successful models) for rice  
672 may reside at least in part in the FAO data and not the mod-  
673 els themselves. The same might apply to Bangladesh and In-  
674 dia for rice. Partitioning of these contributions is impossible at  
675 this stage. Additionally, there is less year-to-year variability in  
676 rice yields (partially due to the fraction of irrigated cultivation).  
677 Since the Pearson r metric is scale invariant, it will tend to score  
678 the rice models more poorly than maize and soy. An example  
679 of very poor performance can be seen with the pDSSAT model  
680 for rice in India (top right panel of Figure 9).

## 678 9. Appendix: Emulation – Assessment

681 Because no general criteria exist for defining an acceptable  
682 model emulator, we develop a metric of emulator performance  
683 specific to GGCMI Phase II. For a multi-model comparison ex-  
684 ercise like GGCMI Phase II, one reasonable criterion is what  
685 we term the “normalized error”, which compares the fidelity of  
686 an emulator for a given model and scenario to the inter-model  
687 uncertainty. We define the normalized error  $e$  for each scenario  
688 as the difference between the fractional yield change from the  
689 emulator and that in the original simulation, divided by the stan-  
690 dard deviation of the multi-model spread (Equations 2 and 3):

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

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

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

692 absolute emulated or simulated mean yields. The normalized  
 693 error  $e$  is the difference between the emulated fractional change  
 694 in yield and that actually simulated, normalized by  $\sigma_{sim}$ , the  
 695 standard deviation in simulated fractional yields  $F_{sim, scn}$  across  
 696 all models. The emulator is fit across all available simulation

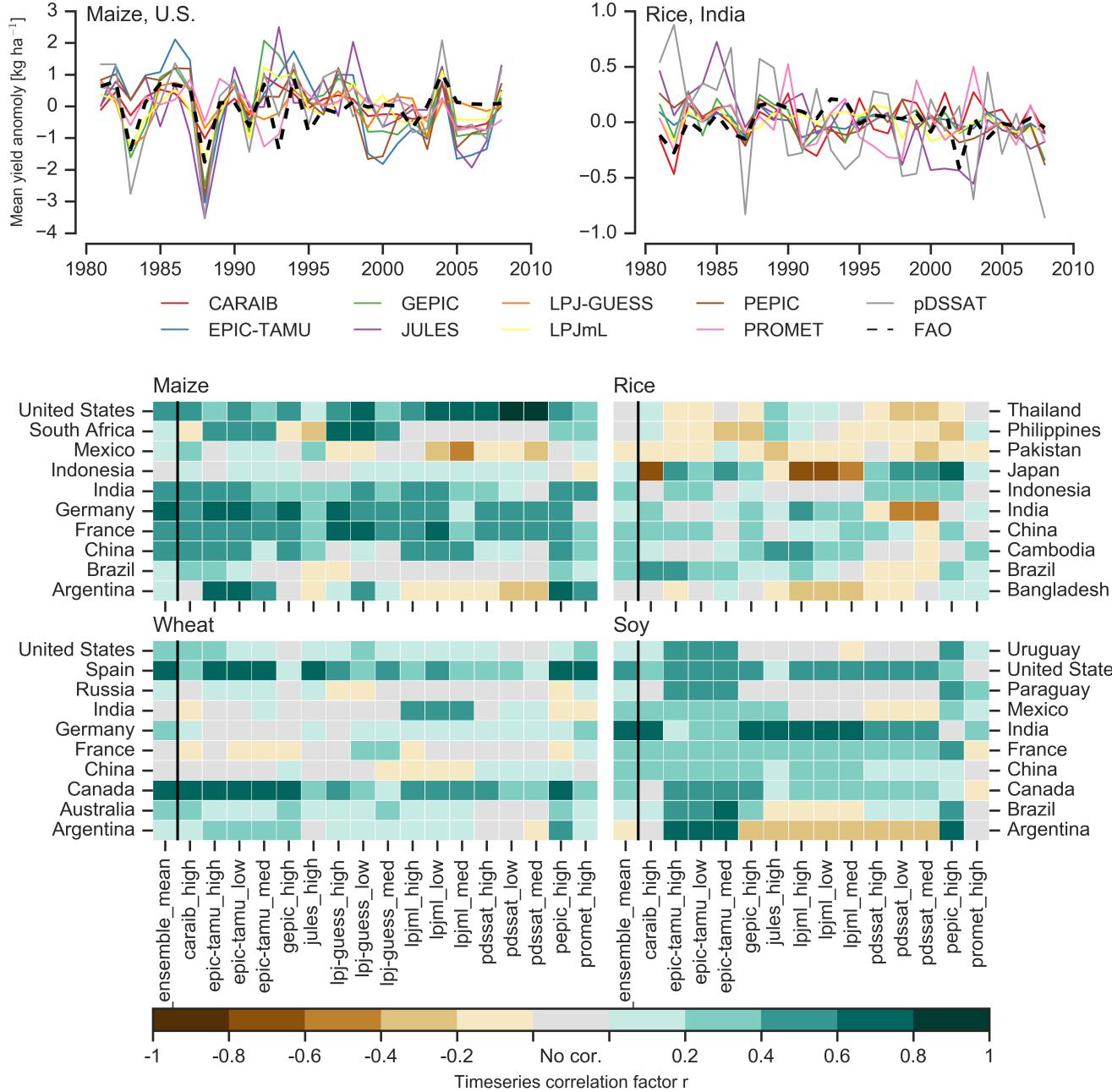


Figure 9: Time series of correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen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

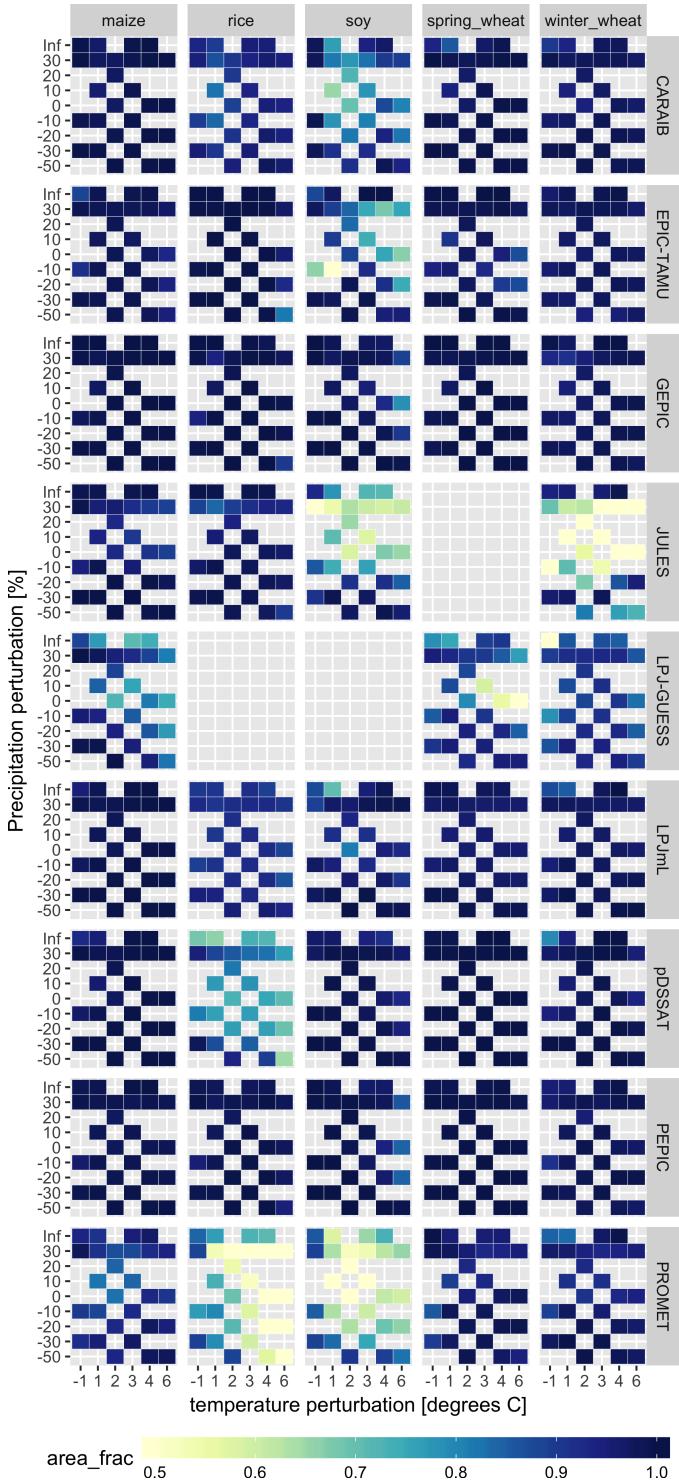


Figure 10: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO<sub>2</sub> and nitrogen held fixed at baseline values. Large columns are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

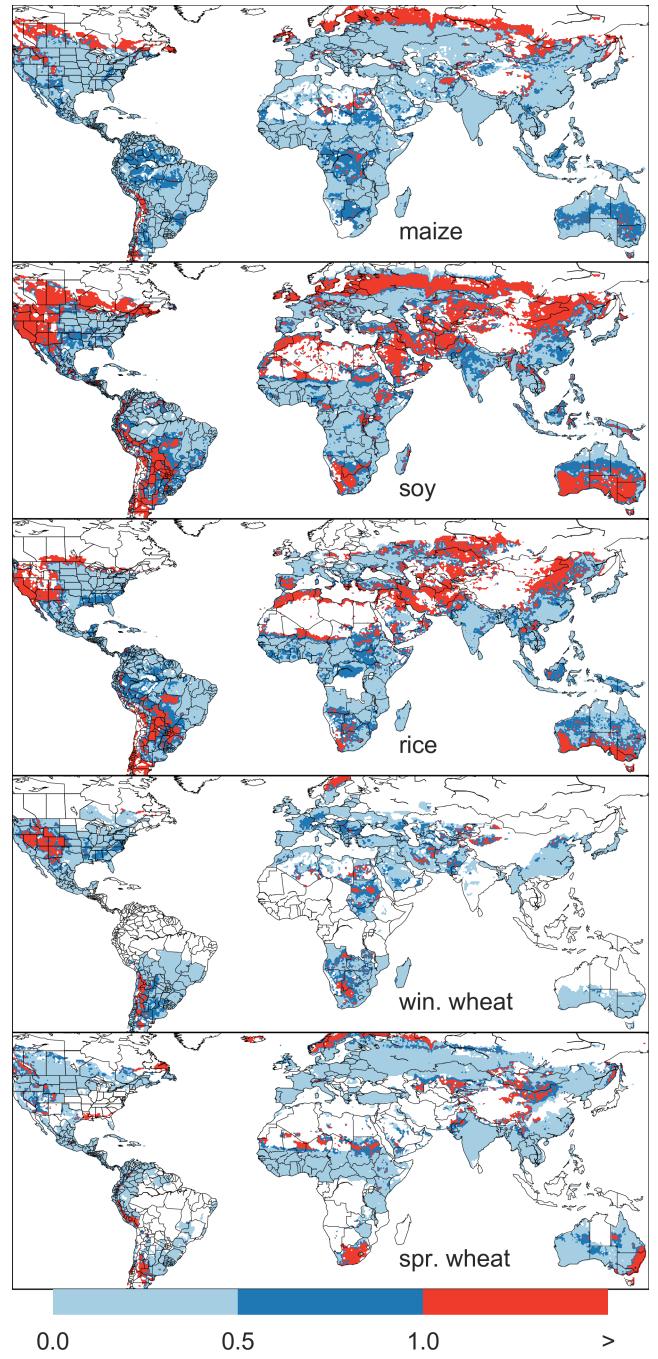


Figure 11: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error  $e$ , where  $e > 1$  means that emulator error exceeds the multi-model standard deviation. White areas are those where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

697 outputs, and then the error is calculated across the simulation<sup>731</sup>  
698 scenarios provided by all nine models (Figure 10 and Figures<sup>732</sup>  
699 S12 and Figures S13 in supplemental documents).

700 Note that the normalized error  $e$  for a model depends not only<sup>734</sup>  
701 on the fidelity of its emulator in reproducing a given simulation<sup>735</sup>  
702 but on the particular suite of models considered in the inter-<sup>736</sup>  
703 comparison exercise. The rationale for this choice is to relate<sup>737</sup>  
704 the fidelity of the emulation to an estimate of true uncertainty,<sup>738</sup>  
705 which we take as the multi-model spread. **Because the inter-**<sup>739</sup>  
706 **model spread is large, normalized errors tend to be small. That**<sup>740</sup>  
707 **is, any failures of emulation are small relative to inter-model**<sup>741</sup>  
708 **uncertainty. We therefore do not provide a formal parameter**<sup>742</sup>  
709 **uncertainty analysis, but note that the GGCMI Phase II dataset**<sup>743</sup>  
710 **is well-suited to statistical exploration of emulation approaches**<sup>744</sup>  
711 **and quantification of emulator fidelity.**

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

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

712 To assess the ability of the polynomial emulation to capture  
713 the behavior of complex process-based models, we evaluate the  
714 normalized emulator error. That is, for each grid cell, model,<sup>746</sup>  
715 and scenario we evaluate the difference between the model yield<sup>748</sup>  
716 and its emulation, normalized by the inter-model standard de-<sup>749</sup>  
717 viation in yield projections. This metric implies that emulation<sup>750</sup>  
718 is generally satisfactory, with several distinct exceptions. Al-<sup>752</sup>  
719 most all model-crop combination emulators have normalized<sup>753</sup>  
720 errors less than one over nearly all currently cultivated hectares<sup>754</sup>  
721 (Figure 10), but some individual model-crop combinations are<sup>756</sup>  
722 problematic (e.g. PROMET for rice and soy, JULES for soy<sup>757</sup>  
723 and winter wheat, Figures S14–S15). Normalized errors for soy<sup>758</sup>  
724 are somewhat higher across all models not because emulator fi-<sup>759</sup>  
725 delity is worse but because models agree more closely on yield<sup>760</sup>  
726 changes for soy than for other crops (see Figure S16, lowering<sup>762</sup>  
727 the denominator. Emulator performance often degrades in geo-<sup>763</sup>  
728 graphic locations where crops are not currently cultivated. Fig-<sup>764</sup>  
729 ure 11 shows a CARAIB case as an example, where emulator<sup>766</sup>  
730 performance is satisfactory over cultivated areas for all crops<sup>767</sup>

## 10. References

- Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013). Implication of crop model calibration strategies for assessing regional impacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.
- Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D., A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Alderman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Biernath, 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

- 768 production potentials and management flexibility under the representative<sup>811</sup>  
 769 concentration pathways. *Global and Planetary Change*, 122, 107 – 121. 812  
 770 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields<sup>813</sup>  
 771 from global gridded crop models. *Agricultural and Forest Meteorology*, 236,<sup>814</sup>  
 772 145 – 161. 815  
 773 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded<sup>816</sup>  
 774 crop models using statistical estimates. *Agricultural and Forest Meteorol-*<sup>817</sup>  
 775 *ogy*, 214-215, 134 – 147. 818  
 776 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S.<sup>819</sup>  
 777 (2018). Implementing the Nitrogen cycle into the dynamic global vegeta-<sup>820</sup>  
 778 tion, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific*<sup>821</sup>  
 779 *Model Development*, 11, 2789–2812. 822  
 780 Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L.,<sup>823</sup>  
 781 & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections<sup>824</sup>  
 782 Based on Precomputed GCM Runs. *Journal of Climate*, 27, 1829–1844. 825  
 783 Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.<sup>826</sup>  
 784 (2014). A meta-analysis of crop yield under climate change and adaptation.<sup>827</sup>  
 785 *Nature Climate Change*, 4, 287 – 291. 828  
 786 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process<sup>829</sup>  
 787 emulation of dynamic computer codes. *Biometrika*, 96, 663–676. 830  
 788 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In<sup>831</sup>  
 789 C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth*,<sup>832</sup>  
 790 Duke University, Durham, North Carolina (pp. 115–118). Durham, North<sup>833</sup>  
 791 Carolina. 834  
 792 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for<sup>835</sup>  
 793 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205). 836  
 794 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdoos,<sup>837</sup>  
 795 M., & François, L. (2011). Responses of European forest ecosystems to<sup>838</sup>  
 796 21st century climate: assessing changes in interannual variability and fire<sup>839</sup>  
 797 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99). 840  
 798 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,<sup>841</sup>  
 799 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating<sup>842</sup>  
 800 impact models and sectors (pSIMS). *Environmental Modelling and Soft-*<sup>843</sup>  
 801 *ware*, 62, 509–516. 844  
 802 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,<sup>845</sup>  
 803 Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C.,<sup>846</sup>  
 804 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.<sup>847</sup>  
 805 (2015). The Global Gridded Crop Model Intercomparison: data and mod-<sup>848</sup>  
 806 eling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8,<sup>849</sup>  
 807 261–277. 850  
 808 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J.,<sup>851</sup>  
 809 & Taylor, K. E. (2016). Overview of the coupled model intercomparison<sup>852</sup>  
 810 project phase 6 (cmip6) experimental design and organization. *Geoscientific*<sup>853</sup>  
 811 *Model Development*, 9, 1937–1958.  
 812 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-  
 813 mate change impacts on durum wheat in the mediterranean region. *Natural*  
 814 *Hazards and Earth System Sciences*, 11, 1293–1302.  
 815 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Re-  
 816 gionalization of a large-scale crop growth model for sub-Saharan Africa:  
 817 Model setup, evaluation, and estimation of maize yields. *Agriculture,*  
 818 *Ecosystems & Environment*, 151, 21 – 33.  
 819 Food and Agriculture Organization of the United Nations (2018). FAOSTAT  
 820 database. URL: <http://www.fao.org/faostat/en/home>.  
 821 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T.,  
 822 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P.,  
 823 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F.,  
 824 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,  
 825 Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J.,  
 826 Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nen-  
 827 del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna,  
 828 M., Semenov, M. A., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K.,  
 829 Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model  
 830 wheat yield impact response surfaces showing sensitivity to temperature and  
 831 precipitation change. *Agricultural Systems*, 159, 209–224.  
 832 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014).  
 833 Evaluating the utility of dynamical downscaling in agricultural impacts pro-  
 834 jections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781.  
 835 Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitiv-  
 836 ity of Agricultural Model Performance to Different Climate Inputs. *Journal*  
 837 *of Applied Meteorology and Climatology*, 55, 151113145618001.  
 838 Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported  
 839 Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous  
 840 Crop Growth and Yield: Application for Wheat in Central Europe. *Remote*  
 841 *Sensing*, 7, 3934–3965.  
 842 He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H.,  
 843 He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water  
 844 contents and nitrogen leaching to precipitation, management practices and  
 845 soil hydraulic properties in semi-arid and humid regions of Canada using the  
 846 DSSAT model. *Nutrient Cycling in Agroecosystems*, 106, 201–215.  
 847 Heady, E. O. (1957). An Econometric Investigation of the Technology of Agri-  
 848 cultural Production Functions. *Econometrica*, 25, 249–268.  
 849 Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa  
 850 State University Press.  
 851 Holden, P., Edwards, N., PH, G., Fraedrich, K., Lunkeit, F., E, K., Labriet,  
 852 M., Kanudia, A., & F, B. (2014). Plasim-entsem v1.0: A spatiotemporal  
 853 emulator of future climate change for impacts assessment. *Geoscientific*



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., Chrysanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. (2017). Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, 10, 1403–1422.

Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T. (1997). Effect of CO<sub>2</sub> enrichment on carbon and nitrogen interaction in wheat and soybean. *Soil Science and Plant Nutrition*, 43, 789–798.

O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91, 1290 – 1300.

Olin, S., Schurges, G., Lindeskog, M., Wårliind, 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+003-12-2489-2015.

Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.

Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., & Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator. *Geoscientific Model Development*, 8, 1139–1155.

Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop yields and their variability at different levels of global warming. *Earth System Dynamics*, 9, 479–496.

Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff, S., & Gerten, D. (2015). Emulating global climate change impacts on crop yields. *Statistical Modelling*, 15, 499–525.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M., Dumont, B., Ewert, F., Ferrise, R., Francois, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet, J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M., Semenov, M., Slawinski, C., 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.

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., Chatta, A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G., Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M., Müller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M., Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., & Wiebe, K. (2018). Coordinating AgMIP data and models across global and regional scales for 1.5°C and 2.0°C assessments. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 376.
- 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.,

- 1112 Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance  
1113 against site observations of irrigated maize from Mead, Nebraska. *Geosci-  
1114 entific Model Development*, *10*, 1291–1320.
- 1115 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield vari-  
1116 ability in JULES-crop and implications for forcing with seasonal weather  
1117 forecasts. *Geoscientific Model Development*, *8*, 3987–3997.
- 1118 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Land-  
1119 bouwkundige Onderzoeken* : 64.6, .
- 1120 Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato  
1121 yields on changes in climate and co2. *Agricultural and Forest Meteorology*,  
1122 *112*, 217 – 231.
- 1123 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao,  
1124 Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A.,  
1125 Li, T., Lin, E., Liu, Q., Martre, P., Mller, C., Peng, S., Peuelas, J., Ruane,  
1126 A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng,  
1127 S. (2017). Temperature increase reduces global yields of major crops in four  
1128 independent estimates. *Proc. Natl. Acad. Sci.*, *114*, 9326–9331.