

# The GGCMI Phase II experiment: global crop yield responses to changes in carbon dioxide, temperature, water, and nitrogen levels

James Franke<sup>a,b,\*</sup>, Joshua Elliott<sup>b,c</sup>, Christoph Müller<sup>d</sup>, Alexander Ruane<sup>e</sup>, Abigail Snyder<sup>f</sup>, Jonas Jägermeyr<sup>c,b,d,e</sup>, Juraj Balkovic<sup>g,h</sup>, Philippe Ciais<sup>i,j</sup>, Marie Dury<sup>k</sup>, Pete Falloon<sup>l</sup>, Christian Folberth<sup>g</sup>, Louis François<sup>k</sup>, Tobias Hank<sup>m</sup>, Munir Hoffmann<sup>n</sup>, Cesar Izaurralde<sup>o,p</sup>, Ingrid Jacquemin<sup>k</sup>, Curtis Jones<sup>o</sup>, Nikolay Khabarov<sup>g</sup>, Marian Koch<sup>n</sup>, Michelle Li<sup>b,l</sup>, Wenfeng Liu<sup>r,i</sup>, Stefan Olin<sup>s</sup>, Meridel Phillips<sup>e,t</sup>, Thomas Pugh<sup>u,v</sup>, Ashwan Reddy<sup>o</sup>, Xuhui Wang<sup>i,j</sup>, Karina Williams<sup>l</sup>, Florian Zabel<sup>m</sup>, Elisabeth Moyer<sup>a,b</sup>

<sup>a</sup>Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA

<sup>b</sup>Center for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA

<sup>c</sup>Department of Computer Science, University of Chicago, Chicago, IL, USA

<sup>d</sup>Potsdam Institute for Climate Impact Research, Leibniz Association (Member), Potsdam, Germany

<sup>e</sup>NASA Goddard Institute for Space Studies, New York, NY, United States

<sup>f</sup>Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA

<sup>g</sup>Ecosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria

<sup>h</sup>Department of Soil Science, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovak Republic

<sup>i</sup>Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France

<sup>j</sup>Sino-French Institute of Earth System Sciences, College of Urban and Environmental Sciences, Peking University, Beijing, China

<sup>k</sup>Unité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d'Astrophysique et de Géophysique, University of Liège, Belgium

<sup>l</sup>Met Office Hadley Centre, Exeter, United Kingdom

<sup>m</sup>Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany

<sup>n</sup>Georg-August-University Göttingen, Tropical Plant Production and Agricultural Systems Modelling, Göttingen, Germany

<sup>o</sup>Department of Geographical Sciences, University of Maryland, College Park, MD, USA

<sup>p</sup>Texas AgriLife Research and Extension, Texas A&M University, Temple, TX, USA

<sup>q</sup>Department of Statistics, University of Chicago, Chicago, IL, USA

<sup>r</sup>EWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

<sup>s</sup>Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

<sup>t</sup>Earth Institute Center for Climate Systems Research, Columbia University, New York, NY, USA

<sup>u</sup>Karlsruhe Institute of Technology, IMK-IFU, 82467 Garmisch-Partenkirchen, Germany.

<sup>v</sup>School of Geography, Earth and Environmental Science, University of Birmingham, Birmingham, UK.

## Abstract

Concerns about food security under climate change have motivated efforts to better understand the future changes in yields by using detailed process-based models in agronomic sciences. Results of these models are often used to inform subsequent model-based analyses in agricultural economics and integrated assessment, but the computational requirements of process-based models sometimes hamper the ability to utilize these models in existing impact assessment frameworks at high resolutions globally. Conducting a large simulation set with perturbations in atmospheric CO<sub>2</sub> concentrations, temperature, precipitation, and nitrogen inputs, Phase II of the Global Gridded Crop Model Intercomparison (GGCMI), an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP), constitutes an unprecedentedly data-rich basis of projected yield changes across 12 models and five crops (maize, soy, rice, spring wheat, and winter wheat) using global gridded simulations. Results from Phase II of the GGCMI effort, a targeted experiment aimed at understanding the interaction between multiple climate variables (as well as management) are presented first. Then the construction of an “emulator or statistical representation of the simulated 30-year mean climatological output in each location is outlined. The emulated response surfaces capture the details of the process-based models in a lightweight, computationally tractable form that facilitates model comparison as well as potential applications in subsequent modeling efforts such as integrated assessment.

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

## 1. Introduction

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

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

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

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

41 Process models do continue to struggle with some important  
42 details, including reproducing historical year-to-year variability  
43 (e.g. Müller et al., 2017), reproducing historical yields when  
44 driven by reanalysis weather (e.g. Glotter et al., 2014), and low  
45 sensitivity to extreme events (e.g. Glotter et al., 2015). These  
46 issues are driven in part by the diversity of new cultivars and  
47 genetic variants, which outstrips the ability of academic mod-  
48 eling groups to capture them (e.g. Jones et al., 2017). Mod-  
49 els do not simulate many additional factors affecting produc-  
50 tion, including pests/diseases/weeds. For these reasons, indi-  
51 vidual studies must generally re-calibrate models to ensure that  
52 short-term predictions reflect current cultivar mixes, and long-  
53 term projections retain considerable uncertainty (Wolf & Oijen,  
54 2002, Jagtap & Jones, 2002, Angulo et al., 2013, Asseng et al.,  
55 2013, 2015). Inter-model discrepancies can also be high in ar-  
56 eas not yet cultivated (e.g. Challinor et al., 2014, White et al.,  
57 2011). Finally, process-based models present additional diffi-  
58 culties for high-resolution global studies because of their com-  
59 plexity and computational requirements. For economic impacts  
60 assessments, it is often impossible to integrate a set of process-  
61 based crop models directly into an integrated assessment model  
62 to estimate the potential cost of climate change to the agricul-  
63 tural sector.

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

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

no yield data are currently available and therefore statistical models cannot apply. Yield data are also often limited in the developing world, where future climate impacts may be the most critical. Second, only process-based models can capture the growth response to elevated CO<sub>2</sub>, novel conditions that are not represented in historical data (e.g. Pugh et al., 2016, Roberts et al., 2017). Similarly, only process-based models can represent novel changes in management practices (e.g. fertilizer input) that may ameliorate climate-induced damages.

Statistical emulation of crop simulations offers the possibility of combining some advantageous features of both statistical and process-based models. The statistical representation of complicated numerical simulation (e.g. O'Hagan, 2006, Conti et al., 2009), in which simulation output acts as the training data for a statistical model, has been of increasing interest with the growth of simulation complexity and volume of output. Such emulators or "surrogate models" have been used in a variety of fields including hydrology (Razavi et al., 2012), engineering (Storlie et al., 2009), environmental sciences (Ratto et al., 2012), and climate (Castruccio et al., 2014). For agricultural impacts studies, emulation of process-based models allows exploring crop yields in regions outside ranges of current cultivation and with input variables outside historical precedents, in a lightweight, flexible form that is compatible with economic studies.

Crop yield emulators have been proposed and implemented by many studies (e.g. Howden & Crimp, 2005, Räisänen & Ruokolainen, 2006, Lobell & Burke, 2010, Iizumi et al., 2010, Ferrise et al., 2011, Holzkämper et al., 2012, Ruane et al., 2013, Howden & Crimp, 2005, Makowski et al., 2015), and in the last several years multiple studies have developed emulators based on a variety of simulation model outputs. Several studies analyzed a single crop model run on a RCP climate scenario set (e.g. Oyebamiji et al., 2015). Multiple groups (e.g. Blanc & Sultan, 2015, Blanc, 2017, Ostberg et al., 2018), constructed

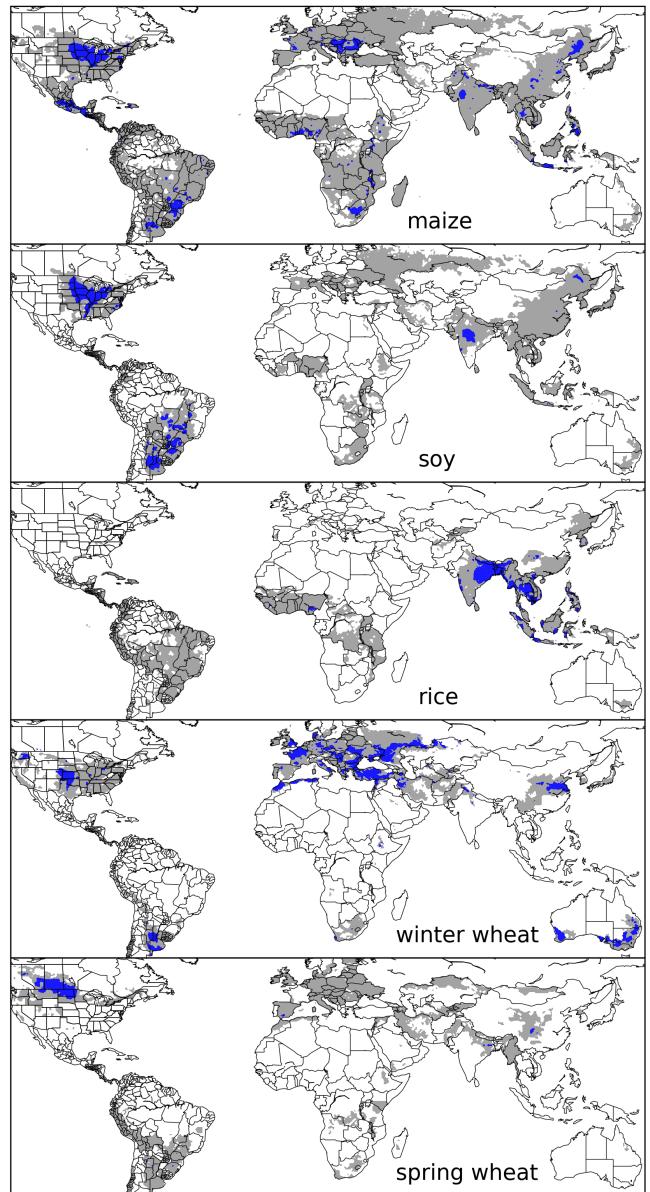


Figure 1: Presently cultivated area for rain-fed crops in the real world. Blue contour area indicates grid-cells with more than 20,000 hectares (approx. 10% of the land in a grid cell near the equator for example) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated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

emulators for a 5-model intercomparison exercise performed as part of ISIMIP (Warszawski et al., 2014), the Inter-Sectoral Impacts Model Intercomparison Project and evaluated several different climate scenarios (over multiple climate model runs). Several other studies (e.g. Moore et al., 2017, Mistry et al., 2017) utilize a hybrid simulation output and real-world data ap-

proach to develop an emulator or damage function. Additional<sub>107</sub>  
 recent studies have explored an impact response surface (aka.<sub>108</sub>  
 emulator when using simulated data) over an explicit multivari-<sub>109</sub>  
 ate input simulation space (as opposed to specific RCP climate<sub>110</sub>  
 model runs), with a site-based approach (as opposed to a glob-<sub>111</sub>  
 ally gridded model) across temperature, water, and CO<sub>2</sub> sam-<sub>112</sub>  
 pling (Snyder et al., 2018), or with models for wheat across<sub>113</sub>  
 water and temperature dimensions for different sites in Europe<sub>114</sub>  
 (Fronzek et al., 2018).<sub>115</sub>

The Global Gridded Crop Model Intercomparison (GGCMI)<sub>116</sub>  
 Phase II experiment is an attempt to expand upon previous<sub>117</sub>  
 process-based crop modeling studies by running globally grid-<sub>118</sub>  
 ded crop models over a set of uniform input dimensions as op-<sub>119</sub>  
 posed to RCP climate scenarios in order to test the sensitivity<sub>120</sub>  
 to yield drivers within and across models. GGCMI is a multi-<sub>121</sub>  
 model exercise conducted as part of the Agricultural Model In-<sub>122</sub>  
 tercomparison and Improvement Project (AgMIP, (Rosenzweig<sub>123</sub>  
 et al., 2013, 2014)), which brings together major global crop<sub>124</sub>  
 simulation models from different research organizations around<sub>125</sub>  
 the world under a framework similar to the Climate Model In-<sub>126</sub>  
 tercomparison Project (CMIP, Taylor et al., 2012, Eyring et al.,<sub>127</sub>  
 2016). The GGCMI analysis framework builds on the Ag-<sub>128</sub>  
 MIP Coordinated Climate-Crop Modeling Project (C3MP, Ru-<sub>129</sub>  
 ane et al., 2014, McDermid et al., 2015), and will contribute<sub>130</sub>  
 to the AgMIP Coordinated Global and Regional Assessments<sub>131</sub>  
 (CGRA, Ruane et al., 2018, Rosenzweig et al., 2018).<sub>132</sub>

The GGCMI Phase II project develops global simulations<sub>133</sub>  
 of yields of major crops under scenarios that sample a uni-<sub>134</sub>

form parameter space. Overall goals include understanding where highest-yield regions may shift under climate change, exploring future adaptive management strategies, understanding how interacting parameters affect crop yields, quantifying uncertainties, and testing strategies for producing lightweight statistical emulations of the more detailed process-based models. In the remainder of this paper, we describe the GGCMI Phase II experiments, present initial overall results, and release the simulation output dataset for public use. We also present a climatological-mean yield emulator as a distillation of the dataset and as a potential tool for impact assessments.<sub>143</sub>

## 2. Materials and Methods

### 2.1. GGCMI Phase II: experiment design

GGCMI Phase II is the continuation of a multi-model comparison exercise begun in 2014. The initial Phase I compared harmonized yields of 21 models for 19 crops over a historical (1980-2010) scenario with a primary goal of model evaluation (Elliott et al., 2015, Müller et al., 2017). Phase II compares simulations of 12 models for 5 crops (maize, rice, soybean, spring wheat, and winter wheat) over hundreds of scenarios in which individual climate or management inputs are adjusted from their historical values. The reduced set of crops includes the three major global cereals and the major legume and accounts for over 50% of human calories (in 2016, nearly 3.5 billion tons or 32% of total global crop production by weight (Food and Agriculture Organization of the United Nations, 2018).

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

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

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

Table 2: Models included in the GGCMI Phase II and the number of C, T, W, and N simulations performed for rain-fed crops (“Sims per Crop”), with 672 as the maximum. “N-Dim.” indicates if simulations include varying nitrogen levels; two models omit this dimension. All models provided the same set of simulations across all modeled crops, but some omitted individual crops in some cases. (For example, APSIM did not simulate winter wheat.) Irrigated simulations are provided at the level of the other covariates for each model (for an additional 84 simulations for the fully-sampled models). Geographic extent of simulation varies to some extent within a certain model for different scenarios (672 rain-fed simulations does not necessarily equal 672 climatological yields in all areas). This geographic variance only applies for areas far outside the area of currently cultivated crops. Two models (marked with \*) use non-AgMERRA climate inputs. For further details on models, see Elliott et al. (2015). †PROMET provided simulations at only two nitrogen levels so is not emulated across the nitrogen dimension.

161 The major goals of GGCMI Phase II are to:

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

172 The guiding scientific rationale of GGCMI Phase II is to pro-<sub>190</sub>  
173 vide a comprehensive, systematic evaluation of the response<sub>191</sub>  
174 of process-based crop models to different values for carbon<sub>192</sub>  
175 dioxide, temperature, water, and applied nitrogen (collectively<sub>193</sub>  
176 known as “CTWN”). Phase II of the GGCMI project consists<sub>194</sub>  
177 of a series of simulations, each with one or more of the CTWN<sub>195</sub>

178 dimensions perturbed over the 31-year historical time series  
179 (1980-2010) used in Phase I. In most cases, historical daily cli-  
180 mate inputs are taken from the 0.5 degree NASA AgMERRA  
181 daily gridded re-analysis product specifically designed for agri-  
182 cultural modeling, with satellite-corrected precipitation (Ruane  
183 et al., 2015). Two models require sub-daily input data and use  
184 alternative sources. See Elliott et al. (2015) for additional de-  
185 tails.

186 The experimental protocol consists of 9 levels for precipita-  
187 tion perturbations, 7 for temperature, 4 for CO<sub>2</sub>, and 3 for ap-  
188 plied nitrogen, for a total of 672 simulations for rain-fed agri-  
189 culture and an additional 84 for irrigated (Table 1). For irri-  
190 gated simulations, soil water is held at either field capacity or,  
191 for those models that include water-log damage, at maximum  
192 beneficial level. Temperature perturbations are applied as ab-  
193 solute offsets from the daily mean, minimum, and maximum  
194 temperature time series for each grid cell used as inputs. Pre-  
195 cipitation 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.<sup>231</sup> Note that CO<sub>2</sub> changes are applied independently of changes in climate variables, so that higher CO<sub>2</sub> is not associated with higher temperatures. An additional, identical set of scenarios (at the same C, T, W, and N levels) simulate adaptive agronomy under climate change by varying the growing season for crop production. (These adaptation simulations are not shown or analyzed here.) The resulting GGCMI data set captures a distribution of crop responses over the potential space of future climate conditions.

The 12 models included in GGCMI Phase II are all mechanistic process-based crop models that are widely used in impacts assessments (Table 2). Although some of the models share a common base (e.g. LPJmL and LPJ-GUESS and the EPIC models), they have developed independently from this shared base, for more details on the genealogy of the models see Figure S1 in Rosenzweig et al. (2014). Differences in model structure does mean that several key factors are not standardized across the experiment, including secondary soil nutrients, carry over effects across growing years including residue management and soil moisture, and extent of simulated area for different crops. Growing seasons are identical across models, but vary by crop and by location on the globe. All stresses except factors related to nitrogen, temperature, and water (e.g. Alkalinity, salinity) are disabled. No additional nitrogen inputs, such as atmospheric deposition, are considered, but some models have individual assumptions on soil organic matter that may release additional nitrogen through mineralization. See Rosenzweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017) for further details on models and underlying assumptions.

Each model is run at 0.5 degree spatial resolution and covers all currently cultivated areas and much of the uncultivated land area. Coverage extends considerably outside currently culti-

vated areas because cultivation will likely shift under climate change. See Figure 1 for the present-day cultivated area of rain-fed crops, and Figure ?? in the supplemental material for irrigated crops. Some areas such as Greenland, far-northern Canada, Siberia, Antarctica, the Gobi and Sahara deserts, and central Australia are not simulated as they are assumed to remain non-arable even under an extreme climate change.

The participating modeling groups provide simulations at any of four initially specified levels of participation, so the number of simulations varies by model, with some sampling only a part of the experiment variable space. Most modeling groups simulate all five crops in the protocol, but some omitted one or more. Table 2 provides details of coverage for each model. Note that the three models that provide less than 50 simulations are excluded from the emulator analysis.

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

## 2.2. Simulation model validation approach

Simulation model validation for GGCMI phase II builds on the validation efforts presented in Müller et al. (2017) for the first phase. In this case however, the models are not run on the best approximation of management levels (namely nitrogen application level) by country as with phase I. As the goals of this phase of the project are focused on understanding the sensitivity in *change* in yield to changes in input drivers –and not to simulate historical yields as accurately as possible– no direct comparison to historical yield data can be made. Additionally,

263 some models are not calibrated as they were in phase I of the<sup>279</sup>  
 264 project.

265 We evaluate the models here based on the response to year-<sup>281</sup>  
 266 to-year temperature and precipitation variability in the histori-<sup>282</sup>  
 267 cal record. If the models can (somewhat) faithfully represent<sup>283</sup>  
 268 the historical variability in yields (which, once detrended<sup>284</sup>  
 269 to account for changing management levels must be driven by<sup>285</sup>  
 270 differences in weather), then the models may provide some util-<sup>286</sup>  
 271 ity in understanding the impact on mean climatological shifts in<sup>287</sup>  
 272 temperature and precipitation. Specifically, we calculate a Pear-<sup>288</sup>  
 273 son correlation coefficient between the detrended time series of<sup>289</sup>  
 274 simulations and FAO data for the period 1981-2009. Validating<sup>290</sup>  
 275 the response to CO<sub>2</sub> and Nitrogen applications is more difficult<sup>291</sup>  
 276 because real world data is not available outside of small green-<sup>292</sup>  
 277 house and field level trials.

### 278 2.3. Climatological-mean yield emulator design

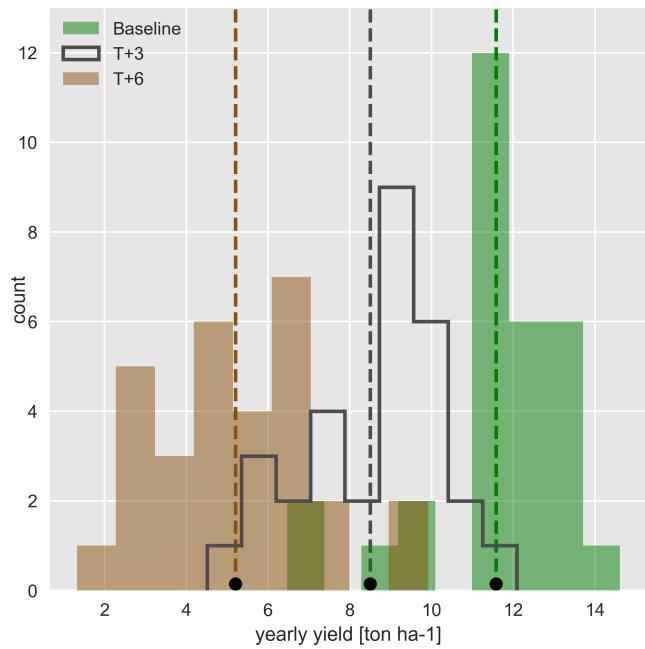


Figure 2: Example showing both climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for a<sup>309</sup>  
 grid cell in northern Iowa (a representative high-yield region) from the pDSSAT  
 model, for the baseline climatology (1981-2010) and for scenarios with tem-<sup>310</sup>  
 perature shifted by (T) +3 and (T) +6 °C, with other variables held at baseline  
 values. Dashed vertical lines and black dots indicate the climatological mean  
 yield.

The decision to first construct a climatological-mean yield emulator is driven by the target application for this analysis tool. Many impact modelers are not focused on the changes in the year-to-year variability in yields, but instead on the broad mean changes over the multi-decadal timescale. Emulation involves fitting individual regression models for each crop, simulation model, and 0.5 degree geographic pixel from the GGCMI Phase II data set. The regressors are the applied constant perturbations in temperature, water, nitrogen and CO<sub>2</sub>, we aggregate the simulation outputs in the time dimension, and regress on the 30-year mean yields. (See Figure 2 for illustration.) The regression therefore omits information about yield responses to year-to-year climate perturbations, which are more complex. Emulating inter-annual yield variations would likely require considering statistical details of the historical climate time series, including changes in marginal distribution and temporal dependencies. (Future work should explore this.) The climatological emulation indirectly includes any yield response to geographically distributed factors such as soil type, insolation, and the baseline climate itself, because we construct separate emulators for each grid cell.

We regress 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, 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.

313 Aulakh & Malhi, 2005), and between nitrogen and carbon diox-<sub>330</sub>  
 314 ide (Osaki et al., 1992, Nakamura et al., 1997). We do not focus<sub>331</sub>  
 315 on comparing different model specifications in this study, and<sub>332</sub>  
 316 instead stick to a relatively simple parameterized specification<sub>333</sub>  
 317 that allows for some, albeit limited, coefficient interpretation.

334 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.

318 The limited GGCMI variable sample space means that use<sub>335</sub>  
 319 of the full polynomial expression described above, which has<sub>336</sub>  
 320 34 terms for the rain-fed case (12 for irrigated), can be prob-<sub>337</sub>  
 321 lematic, and can lead to over-fitting and unstable parameter es-<sub>338</sub>  
 322 estimations. We therefore reduce the number of terms through a<sub>339</sub>  
 323 feature selection cross-validation process in which terms in the<sub>340</sub>  
 324 polynomial are tested for importance. In this procedure higher-<sub>341</sub>  
 325 order and interaction terms are added successively to the model;<sub>342</sub>  
 326 we then follow the reduction of the the aggregate mean squared<sub>343</sub>  
 327 error with increasing terms and eliminate those terms that do<sub>344</sub>  
 328 not contribute significant reductions. See supplemental docu-<sub>345</sub>  
 329 ments for more details. We select terms by applying the feature<sub>346</sub>

Feature importance is remarkably consistent across all three models and across all crops (see Figure ?? 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:

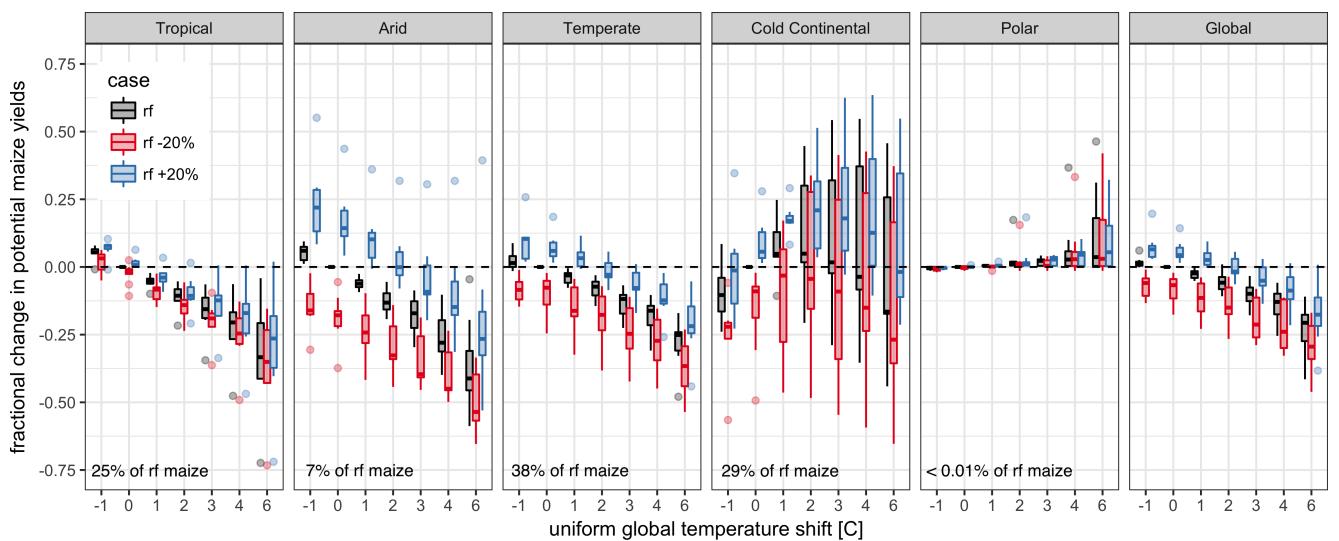


Figure 3: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation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

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

347 To fit the parameters  $K$ , we use a Bayesian Ridge probabilis-  
 348 tic estimator (MacKay, 1991), which reduces volatility in pa-  
 349 rameter estimates when the sampling is sparse, by weighting  
 350 parameter estimates towards zero. The Bayesian Ridge method  
 351 is necessary to maintain a consistent functional form across all  
 352 models, and locations as the linear least squares fails to pro-  
 353 vide a stable result in many cases. In the GGCMI Phase II<sub>380</sub>  
 354 experiment, the most problematic fits are those for models that<sub>381</sub>  
 355 provided a limited number of cases or for low-yield geographic<sub>382</sub>  
 356 regions where some modeling groups did not run all scenarios.<sub>383</sub>  
 357 Because we do not attempt to emulate models that provided<sub>384</sub>  
 358 less than 50 simulations, the lowest number of simulations em-<sub>385</sub>  
 359 ulted across the full parameter space is 130 (for the PEPIC<sub>386</sub>  
 360 model). We use the implementation of the Bayesian Ridge esti-<sub>387</sub>  
 361 mator from the scikit-learn package in Python (Pedregosa et al.,<sub>388</sub>  
 362 2011).

363 The resulting parameter matrices for all crop models are<sub>391</sub>  
 364 available on request, as are the raw simulation data and a Python<sub>392</sub>  
 365 application to emulate yields. The yield output for a single<sub>393</sub>  
 366 GGCMI model that simulates all scenarios and all five crops<sub>394</sub>  
 367 is ~12.5 GB; the emulator is ~100 MB, a reduction by over<sub>395</sub>  
 368 two orders of magnitude.

370 Because no general criteria exist for defining an acceptable  
 371 model emulator, we develop a metric of emulator performance  
 372 specific to GGCMI. For a multi-model comparison exercise like  
 373 GGCMI, a reasonable criterion is what we term the “normalized  
 374 error”, which compares the fidelity of an emulator for a given  
 375 model and scenario to the inter-model uncertainty. We define  
 376 the normalized error  $e$  for each scenario as the difference be-  
 377 tween the fractional yield change from the emulator and that in  
 378 the original simulation, divided by the standard deviation of the  
 379 multi-model spread (Equations 2 and 3):

$$F_{scn.} = \frac{Y_{scn.} - Y_{baseline}}{Y_{baseline}} \tag{2}$$

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

Here  $F_{scn.}$  is the fractional change in a model’s mean emulated or simulated yield from a defined baseline, in some scenario (scn.) in C, T, W, and N space;  $Y_{scn.}$  and  $Y_{baseline}$  are the absolute emulated or simulated mean yields. The normalized error  $e$  is the difference between the emulated fractional change in yield and that actually simulated, normalized by  $\sigma_{sim.}$  the standard deviation in simulated fractional yields  $F_{sim, scn.}$  across all models. The emulator is fit across all available simulation outputs, and then the error is calculated across the simulation scenarios provided by all nine models (Figure 9 and Figures ?? and Figures ?? in supplemental documents). Note that the normalized error  $e$  for a model depends not only on the fidelity of its emulator in reproducing a given simulation but on the particular suite of models considered in the intercomparison exercise. The rationale for this choice is to relate the fidelity of the emulation to an estimate of true uncertainty, which we take as the multi-model spread.

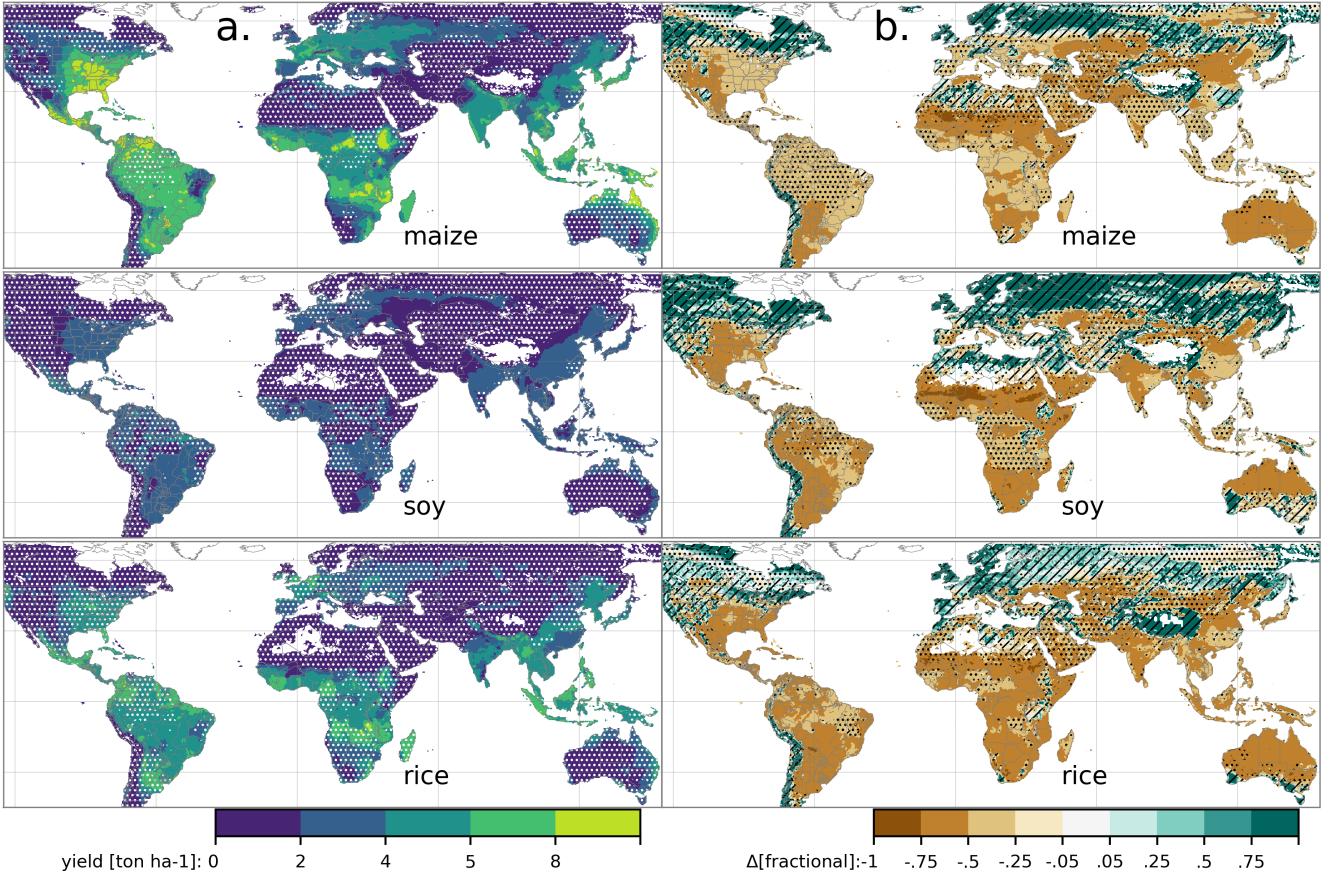


Figure 4: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top-bottom) rain-fed maize, soy, and rice. (For wheat see Figure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

### 3. Results

#### 3.1. Simulation results

Crop models in the GGCMI ensemble show a broadly consistent responses to climate and management perturbations in most regions, with a strong negative impact of increased temperature in all but the coldest regions. We illustrate this result for rain-fed maize in Figure 3, which shows yields for the primary Köppen-Geiger climate regions (Rubel & Kottek, 2010).

In warming scenarios, models show decreases in maize yield in the temperate, tropical, and arid regions that account for nearly three-quarters of global maize production. These impacts are robust for even moderate climate perturbations. In the temperate zone, even a 1 degree temperature rise with other variables

held fixed leads to a median yield reduction that outweighs the variance across models. A 6 degree temperature rise results in median loss of  $\sim 25\%$  of yields with a signal to noise of nearly three. A notable exception is the cold continental region, where models disagree strongly, extending even to the sign of impacts. Model simulations of other crops produce similar responses to warming, with robust yield losses in warmer locations and high inter-model variance in the cold continental regions (Figures ??).

The effects of rainfall changes on maize yields are also as expected and are consistent across models. Increased rainfall mitigates the negative effect of higher temperatures, most strongly in arid regions. Decreased rainfall amplifies yield losses and

423 also increases inter-model variance more strongly, suggesting<sub>451</sub>  
424 that models have difficulty representing crop response to water<sub>452</sub>  
425 stress. We show only rain-fed maize here; see Figure ?? for the<sub>453</sub>  
426 irrigated case. As expected, irrigated crops are more resilient to<sub>454</sub>  
427 temperature increases in all regions, especially so where water<sub>455</sub>  
428 is limiting.<sub>456</sub>

like the one presented here. The ensemble mean yield is calculated across all ‘high’ nitrogen application level model simulations and correlated with the FAO data (not the mean of the correlations). The ensemble mean does not beat the best model in each case, but shows positive correlation in over 75% of the cases presented here.

429 Mapping the distribution of baseline yields and yield changes 457  
430 shows the geographic dependencies that underlie these results. 458  
431 Figure 4 shows baseline and changes in the T+4 scenario for 459  
432 rain-fed maize, soy, and rice in the multi-model ensemble mean, 460  
433 with locations of model agreement marked. Absolute yield po- 461  
434 tentials are have strong spatial variation, with much of the 462  
435 Earth's surface area unsuitable for any given crop. In general, 463  
436 models agree most on yield response in regions where yield 464  
437 potentials are currently high and therefore where crops are cur- 465  
438 rently grown. Models show robust decreases in yields at low 466  
439 latitudes, and highly uncertain median increases at most high 467  
440 latitudes. For wheat crops see Figure ??; wheat projections are 468  
441 both more uncertain and show fewer areas of increased yield in 469  
442 the inter-model mean. 470

### 443 3.2. Simulation model validation results

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

Figure 8: Distribution in historical yields (1981-2009) for maize for eight example high producing countries. FAO, simulation (high nitrogen), and emulation. Emulated values are calculated based on the additive temperature anomaly or percentage precipitation anomaly from the 1980-2010 period in each year. Note: the emulator is designed to provide the mean change in yield under climatological mean shift in temperature (or precipitation). Applying it at the year to year level should be interpreted with caution.

It is important to note, much FAO data is at least one level of abstraction from ground truth. This is especially applicable in developing countries. The failure of models to represent the year-to-year variability in rice in some countries in south-east Asia is likely partly due to model failure and partly due to lack of data. Partitioning of these contributions is impossible at this stage. There is less year-to-year variability in rice yields (partially due to the fraction of irrigated cultivation), since the Pearson  $r$  metric is scale invariant, it will tend to score the rice models more poorly than maize and soy. The pDSSAT model shows poor performance for rice in India (top right panel).

Figure 7 shows the time series correlation between the simulation model yield and FOA yield data. The results are mixed, with many regions for rice and wheat being difficult to model. No single model is dominant, with each model providing near best-in-class performance in at least one location-crop combination. The presence of no vertical dark green color bars clearly illustrates the power of a multi-model intercomparison project.

*One might suspect that the difference in performance between Pakistan (no successful models) and India (many suc-*

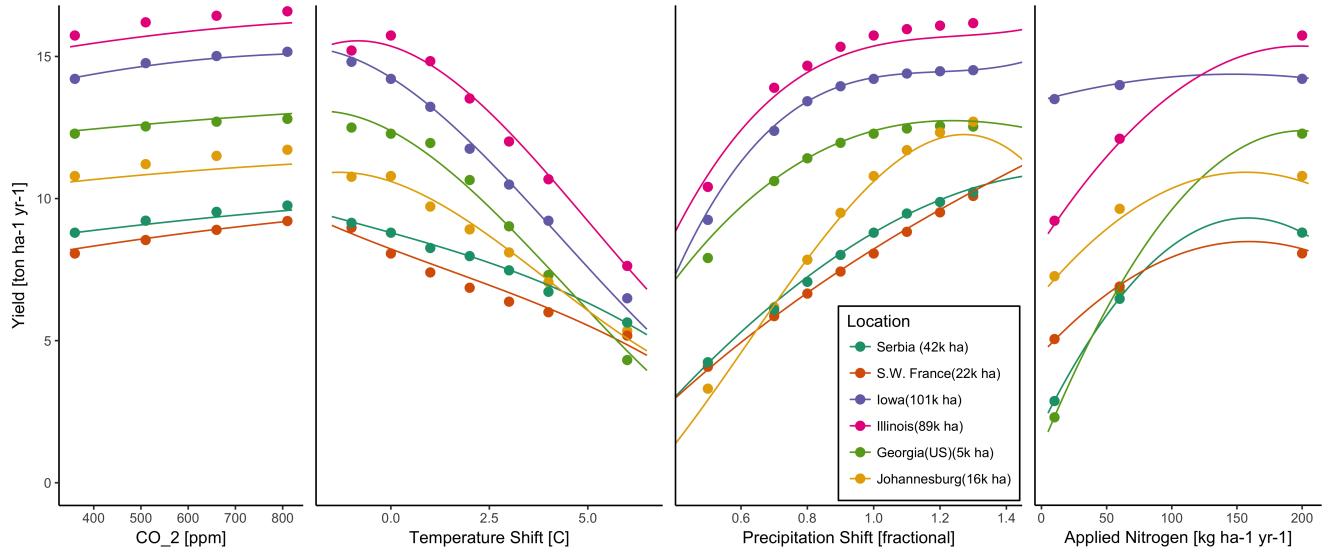


Figure 5: Example illustrating spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation along a single variable, with others held at baseline values. Dots show climatological mean yield, solid lines a simple polynomial fit across this 1D variation, and dashed lines the results of the full emulator of Equation 1. The climatological response is sufficiently smooth to be represented by a simple polynomial. Because precipitation changes are imposed as multiplicative factors rather than additive offsets, locations with higher baseline precipitation have a larger absolute spread in inputs sampled than do drier locations (not visible here).

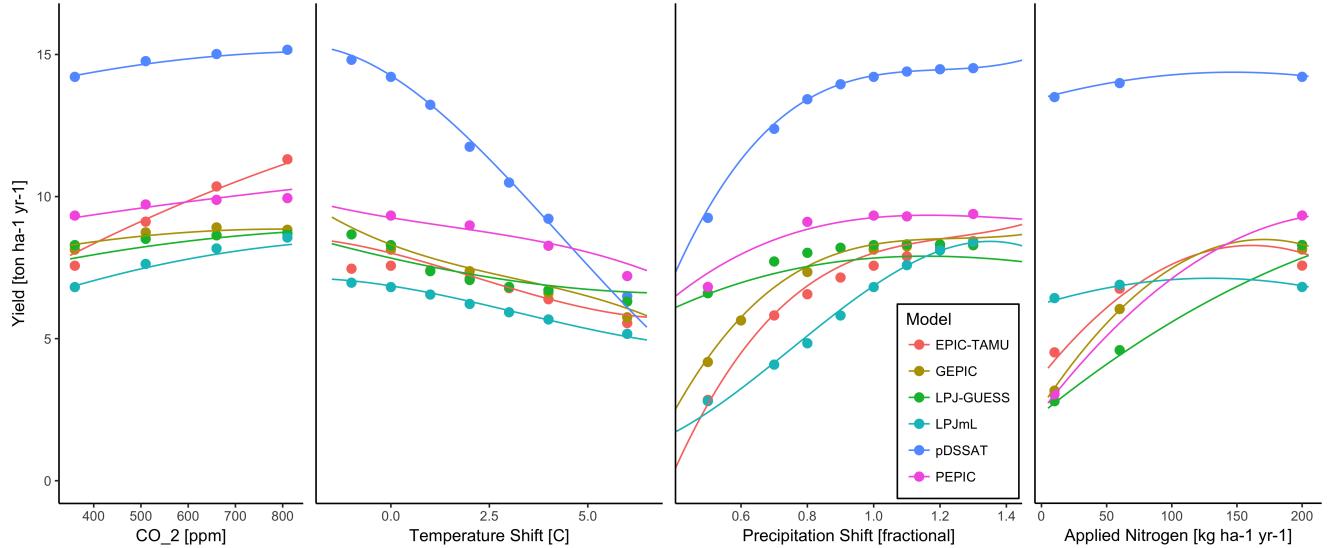


Figure 6: Example illustrating across-model variations in yield response. We show simulations and emulations from six models for rain-fed maize in the same Iowa location shown in Figure 5, with the same plot conventions. Models that did not simulate the nitrogen dimension are omitted for clarity. The inter-model standard deviation is larger for the perturbations in CO<sub>2</sub> and nitrogen than for those in temperature or precipitation illustrating that the sensitivity to weather is better constrained than to management at elevated CO<sub>2</sub>. While most model responses can readily be emulated with a simple polynomial, some response surfaces are more complicated and lead to emulation error.

485 successful models) for rice may lie in the FAO data and not the  
486 models themselves. What would be so different about rice pro-  
487 duction across these two countries that could explain this dif-  
488 ference??

489 Figure 3.2 shows the distribution across

### 3.3. Emulator performance

490 Emulation provides not only a computational tool but a  
491 means of understanding and interpreting crop yield response  
492 across the parameter space. Emulation is only possible, how-  
493 ever, when crop yield responses are sufficiently smooth and

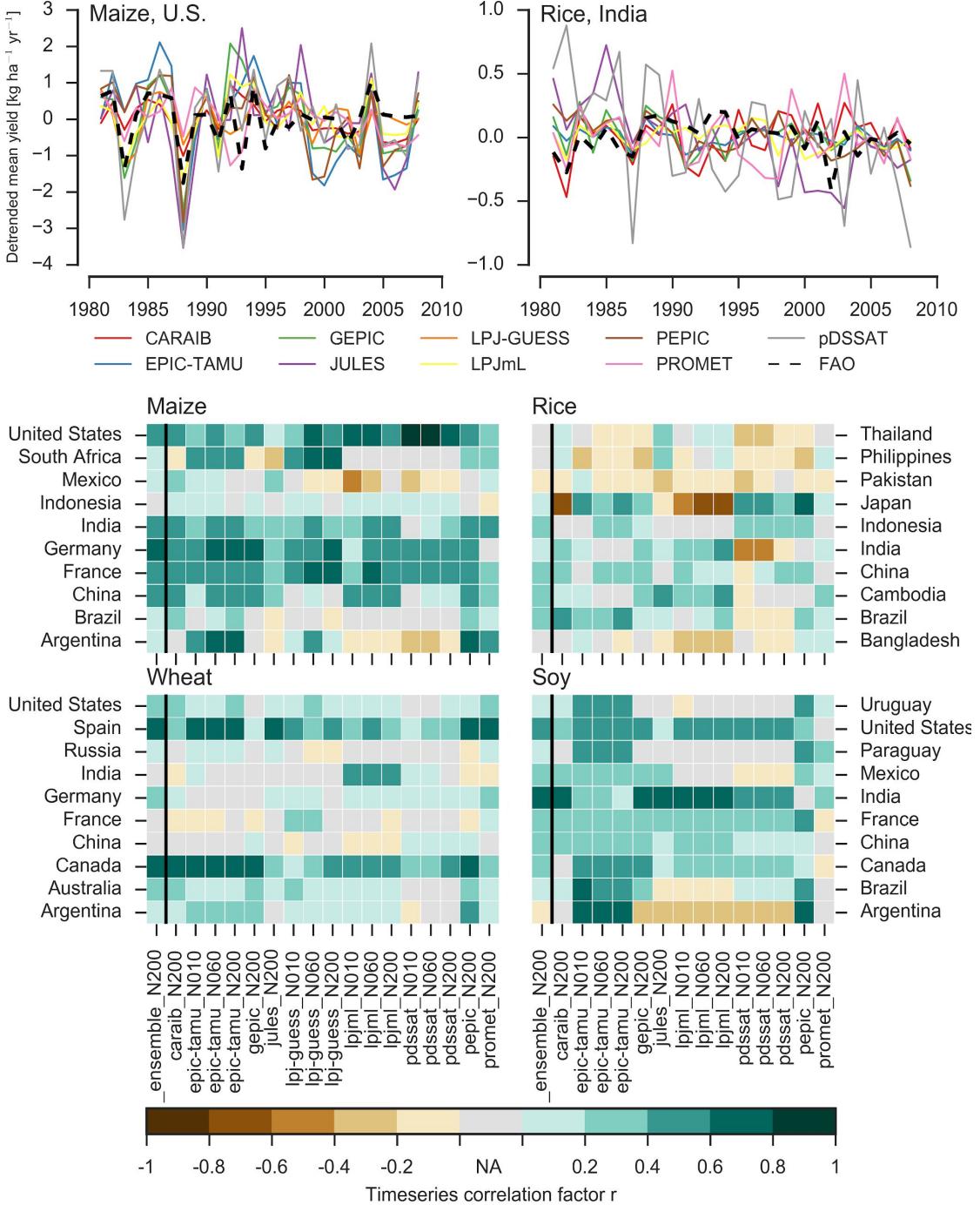
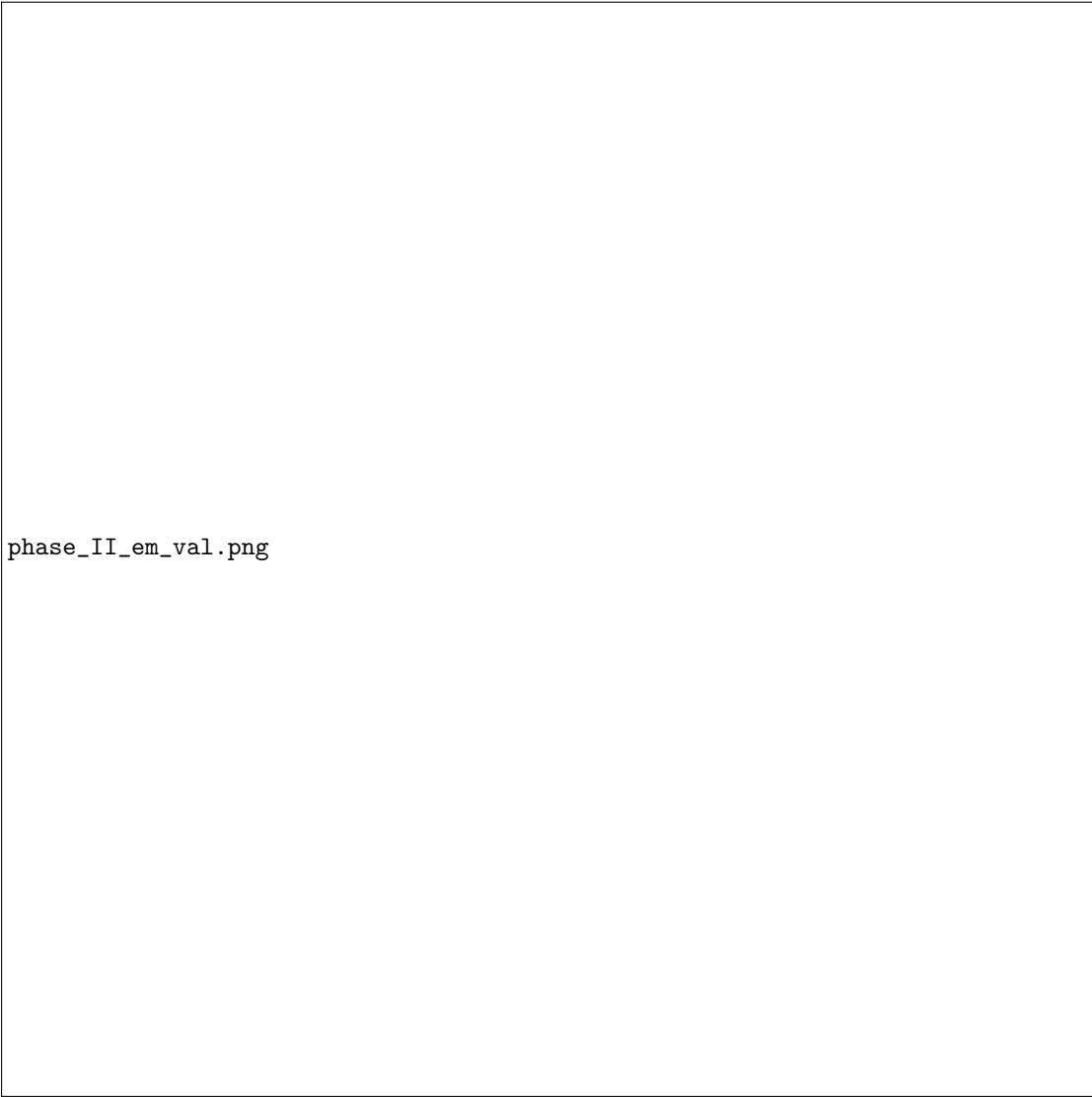


Figure 7: Time series correlation coefficients between simulated crop yield and FAO data at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen application case. The heatmaps illustrate the Pearson  $r$  correlation 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.

continuous to allow fitting with a relatively simple functional form. In the GGCMI simulations, this condition largely but not always holds. Responses are quite diverse across locations, crops, and models, but in most cases local responses are regular enough to permit emulation. Figure 5 illustrates the geographic diversity of responses even in high-yield areas for a single crop and model (rain-fed maize in pDSSAT for various high-cultivation areas). This heterogeneity validates the choice



phase\_II\_em\_val.png

Note: the emulator is designed to provide the mean change in yield under climatological mean shift in temperature (or precipitation). Applying it at the year to year level should be interpreted with caution.

503 of emulating at the grid cell level.

504 Each panel in Figure 5 shows model yield output from sce-<sup>515</sup>  
505 narios varying only along a single dimension ( $\text{CO}_2$ , tempera-<sup>516</sup>  
506 ture, precipitation, or nitrogen addition), with other inputs held<sup>517</sup>  
507 fixed at baseline levels; in all cases yields evolve smoothly<sup>518</sup>  
508 across the space sampled. For reference we show the results<sup>519</sup>  
509 of the full emulation fitted across the parameter space. The<sup>520</sup>  
510 polynomial fit readily captures the climatological response to<sup>521</sup>  
511 perturbations.

512 Crop yield responses generally follow similar functional  
513 forms across models, though with a spread in magnitude. Fig-

514 ure 6 illustrates the inter-model diversity of yield responses  
515 to the same perturbations, even for a single crop and location  
516 (rain-fed maize in northern Iowa, the same location shown in  
517 the Figure 5). The differences make it important to construct  
518 emulators separately for each individual model, and the fidelity  
519 of emulation can also differ across models. This figure illus-  
520 trates a common phenomenon, that models differ more in re-  
521 sponse to perturbations in  $\text{CO}_2$  and nitrogen perturbations than  
522 to those in temperature or precipitation. (Compare also Figures  
523 3 and ??.) For this location and crop,  $\text{CO}_2$  fertilization effects  
524 can range from ~5–50%, and nitrogen responses from nearly

flat to a 60% drop in the lowest-application simulation.

While the nitrogen dimension is important and uncertain, it is also the most problematic to emulate in this work because of its limited sampling. The GGCMI protocol specified only three nitrogen levels ( $10, 100$  and  $200 \text{ kg N y}^{-1} \text{ ha}^{-1}$ ), so a third-order fit would be over-determined but a second-order fit can result in potentially unphysical results. Steep and nonlinear declines in yield with lower nitrogen levels means that some regressions imply a peak in yield between the  $100$  and  $200 \text{ kg N y}^{-1} \text{ ha}^{-1}$  levels. While there may be some reason to believe 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.

To assess the ability of the polynomial emulation to capture the behavior of complex process-based models, we evaluate the normalized emulator error. That is, for each grid cell, model, and scenario we evaluate the difference between the model yield and its emulation, normalized by the inter-model standard deviation in yield projections. This metric implies that emulation is generally satisfactory, with several distinct exceptions. Almost all model-crop combination emulators have normalized errors less than one over nearly all currently cultivated hectares (Figure 9), but some individual model-crop combinations are problematic (e.g. PROMET for rice and soy, JULES for soy and winter wheat, Figures ??–??). Normalized errors for soy are somewhat higher across all models not because emulator fidelity is worse but because models agree more closely on yield changes for soy than for other crops (see Figure ??, lowering the denominator). Emulator performance often degrades in geographic locations where crops are not currently cultivated. Figure 10 shows a CARAIB case as an example, where emulator

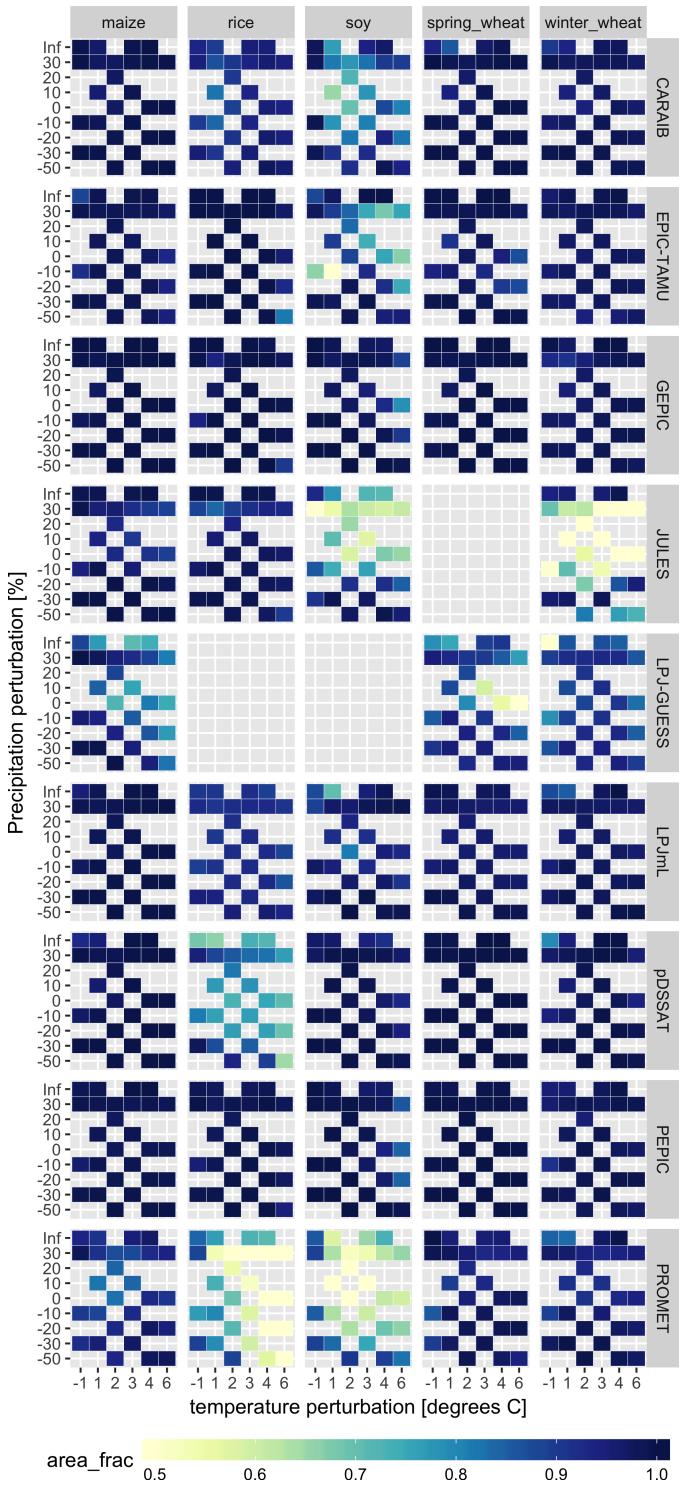


Figure 9: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with  $\text{CO}_2$  and nitrogen held fixed at baseline values. Large columns are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

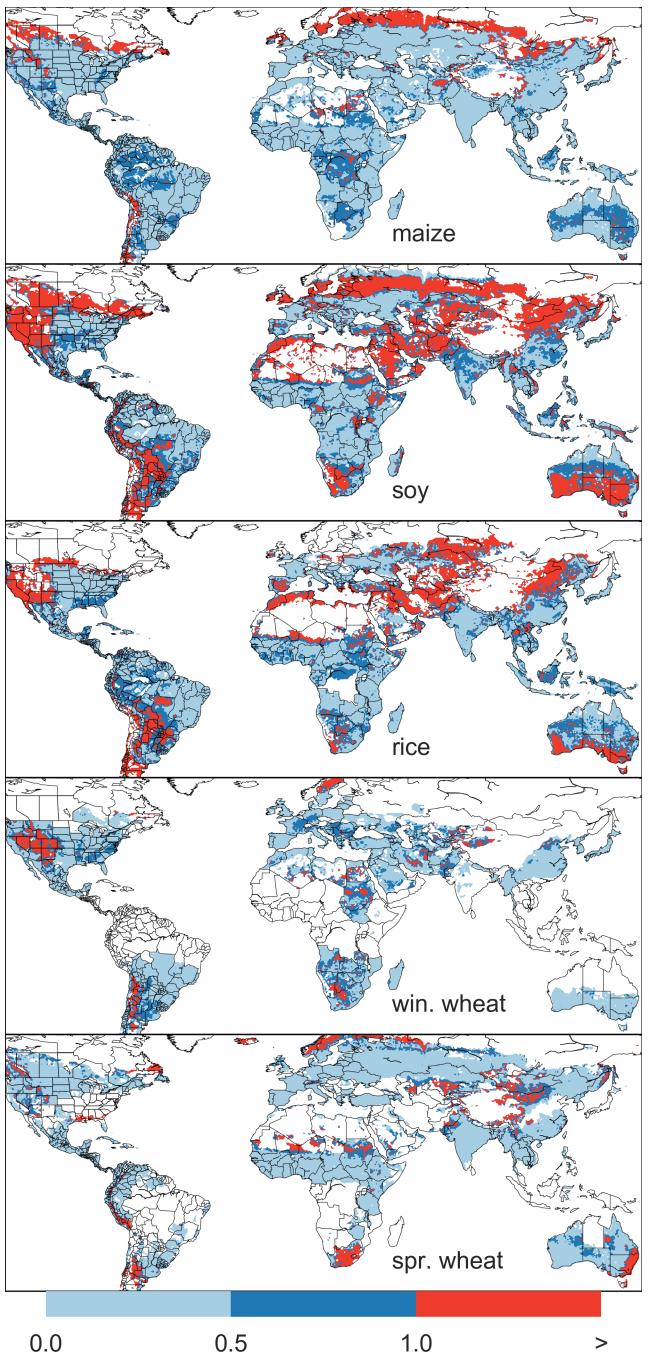


Figure 10: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error  $e$ , where  $e > 1$  means that emulator error exceeds the multi-model standard deviation. White areas are those where crops are not simulated by this model. Models differ in their areas omitted, meaning the number of samples used to calculate the multi-model standard deviation is not spatially consistent in all locations. Emulator performance is generally good relative to model spread in areas where crops are currently cultivated (compare to Figure 1) and in temperate zones in general; emulation issues occur primarily in marginal areas with low yield potentials. For CARAIB, emulation of soy is more problematic, as was also shown in Figure 9.

559 performance is satisfactory over cultivated areas for all crops  
560 other than soy, but uncultivated regions show some problematic  
561 areas.

562 It should be noted that this assessment metric is relatively  
563 forgiving. First, each emulation is evaluated against the simu-  
564 lation actually used to train the emulator. Had we used a spline  
565 interpolation the error would necessarily be zero. Second, the  
566 performance metric scales emulator fidelity not by the magni-  
567 tude of yield changes but by the inter-model spread in those  
568 changes. Where models differ more widely, the standard for  
569 emulators becomes less stringent. Because models disagree on  
570 the magnitude of CO<sub>2</sub> fertilization, this effect is readily seen  
571 when comparing assessments of emulator performance in sim-  
572 ulations at baseline CO<sub>2</sub> (Figure 9) with those at higher CO<sub>2</sub>  
573 levels (Figure ??). Widening the inter-model spread leads to an  
574 apparent increase in emulator skill.

### 575 3.4. Emulator applications

576 Because the emulator or “surrogate model” transforms the  
577 discrete simulation sample space into a continuous response  
578 surface at any geographic scale, it can be used for a variety  
579 of applications. Emulators provide an easy way to compare a  
580 ensembles of climate or impacts projections. They also pro-  
581 vide a means for generating continuous damage functions. As  
582 an example, we show a damage function constructed from 4D  
583 emulations for aggregated yield at the global scale, for maize  
584 on currently cultivated land, with simulated values shown for  
585 comparison. (Figure 11; see Figures ??- ?? in the supple-  
586 mental material for other crops and dimensions.) The emu-  
587 lated values closely match simulations even at this aggrega-  
588 tion level. Note that these functions are presented only as  
589 examples and do not represent true global projections, be-  
590 cause they are developed from simulation data with a uniform  
591 temperature shift while increases in global mean temperature  
592 should manifest non-uniformly. The global coverage of the

593 GGCMI simulations allows impacts modelers to apply arbitrary<sup>605</sup>  
 594 geographically-varying climate projections, as well as arbitrary<sup>606</sup>  
 595 aggregation mask, to develop damage functions for any climate<sup>607</sup>  
 596 scenario and any geopolitical or geographic level.

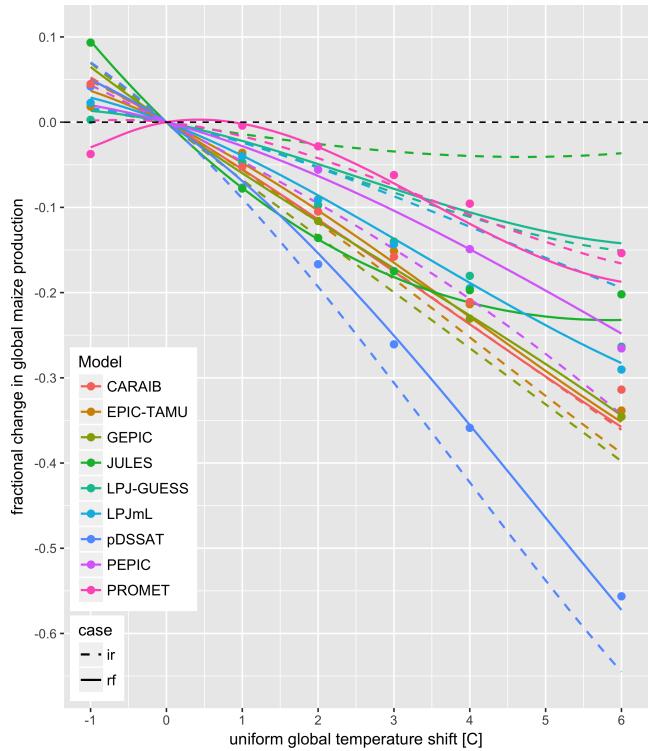


Figure 11: Global emulated damages for maize on currently cultivated lands for the GGCMI models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating up emulated values at the grid-cell level, not from a regression of global mean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

further analysis, we show only a selection of insights derived from the simulations. 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, inter-model uncertainty is largest in the high latitudes where yields may increase, and model projections are most robust in low latitudes where yield impacts are largest.

Model performance when compared to historical data is mixed, with models performing better for maize and soy than for rice and wheat. The value of utilizing multiple models is illustrated by the distribution in performance skill across different countries and crops. An end-user of the simulation outputs or emulator tool may pick and choose models based on historical skill to provide the most faithful temperature and precipitation response depending on their application. The nitrogen and CO<sub>2</sub> responses were not validated in this work.

One counterintuitive result is that irrigated maize shows steeper yield reductions under warming than does rain-fed maize when considered only over currently cultivated land. The effect is the result of geographic differences in cultivated area. In any given location, irrigation increases crop resiliency to temperature increase, but irrigated maize is grown in warmer locations where the impacts of warming are more severe (Figures ??–??). The same behavior holds for rice and winter wheat, but not for soy or spring wheat (Figures ??–??). Irrigated wheat and maize are also more sensitive to nitrogen fertilization levels, presumably because growth in rain-fed crops is also water-limited (Figure ??). (Soy as a nitrogen-fixer is relatively insensitive to nitrogen, and rice is not generally grown in water-limited conditions.)

We show that emulation of the output of these complex responses is possible even with a relatively simple reduced-form statistical model and a limited library of simulations. Emula-

#### 597 4. Conclusions and discussion

The GGCMI Phase II experiment assess sensitivities of process-based crop yield models to changing climate and management inputs, and was designed to allow not only comparison across models but evaluation of complex interactions between driving factors (CO<sub>2</sub>, temperature, precipitation, and applied nitrogen) and identification of geographic shifts in high yield potential locations. While the richness of the dataset invites

639 tion therefore offers the opportunity of producing rapid assessments  
640 of agricultural impacts for arbitrary climate scenarios in a computationally non-intensive way. The resulting tool should aid in impacts assessment, economic studies, and uncertainty analyses. Emulator parameter values also provide a useful way to compare sensitivities across models to different climate and management inputs, and the terms in the polynomial fits offer the possibility of physical interpretation of these dependencies to some degree.

648 We open up this simulation output dataset for further analysis by the community as we have only scratched the surface with this work, and all simulation output data are readily available. Each simulation run includes year to year variability in yields under different climate and management regimes. Some of the precipitation and temperature space has been lost due to the aggregation in the time dimension (i.e. the + 6 C simulation in the hottest year of the historical period compared to the coldest historical year, or precipitation perturbations in the driest historical year etc.) Development of a year-to-year emulator, or an emulator at different spatial scales may provide useful for some IAM applications. More exhaustive analysis of different statistical model specification for emulation may likely provide additional predictive skill over the specification provided here. The potentially richest area for analysis is the interactions space between input variable especially the Nitrogen and CO<sub>2</sub> interactions with weather and with each other. Adaptation via growing season changes were also simulated and are available in the database, though this dimension was not presented or analyzed here.

668 The emulation approach presented here has some limitations. Because the GGCMI simulations apply uniform perturbations to historical climate inputs, they do not sample changes in higher order moments. The emulation therefore does not address the crop yield impacts of potential changes in climate

673 variability. While some information could be extracted from consideration of year-over-year variability, more detailed simulations and analysis are likely necessary to diagnose the impact of changes in variance and sub-growing-season temporal effects. Additionally, the emulator is intended to provide the change in yield from a historical mean baseline value and should be used in conjunction with historical data (or data products) or a historical mean emulator (not presented here).

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

## 5. Acknowledgments

We thank Michael Stein and Kevin Schwarzwald, who provided helpful suggestions that contributed to this work. This research was performed as part of the Center for Robust Decision-Making on Climate and Energy Policy (RDCEP) at the University of Chicago, and was supported through a variety of sources. RDCEP is funded by NSF grant #SES-1463644 through the Decision Making Under Uncertainty program. J.F. was supported by the NSF NRT program, grant #DGE-1735359. C.M. was supported by the MACMIT project (01LN1317A) funded through the German Federal Ministry of Education and Research (BMBF). C.F. was supported by the European Research Council Synergy grant #ERC-2013-SynG-610028 Imbalance-P. P.F. and K.W. were supported by the Newton Fund through the Met Office Climate Science for Service Partnership Brazil

706 (CSSP Brazil). A.S. was supported by the Office of Science<sup>745</sup>  
 707 of the U.S. Department of Energy as part of the Multi-sector<sup>746</sup>  
 708 Dynamics Research Program Area. Computing resources were<sup>747</sup>  
 709 provided by the University of Chicago Research Computing<sup>749</sup>  
 710 Center (RCC). S.O. acknowledges support from the Swedish<sup>750</sup>  
 711 strong research areas BECC and MERGE together with sup-<sup>751</sup>  
 712 port from LUCCI (Lund University Centre for studies of Car-<sup>752</sup>  
 713 bon Cycle and Climate Interactions).<sup>753</sup>

(2018). Implementing the Nitrogen cycle into the dynamic global vegetation, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific Model Development*, *11*, 2789–2812.

Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L., & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs. *Journal of Climate*, *27*, 1829–1844.

Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, *4*, 287 – 291.

Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process emulation of dynamic computer codes. *Biometrika*, *96*, 663–676.

Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth, Duke University, Durham, North Carolina* (pp. 115–118). Durham, North Carolina.

Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).

Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdoos, M., & François, L. (2011). Responses of European forest ecosystems to 21st century climate: assessing changes in interannual variability and fire intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).

Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K., Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating impact models and sectors (pSIMS). *Environmental Modelling and Software*, *62*, 509–516.

Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C., Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J. (2015). The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, *8*, 261–277.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*, 1937–1958.

Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of climate change impacts on durum wheat in the mediterranean region. *Natural Hazards and Earth System Sciences*, *11*, 1293–1302.

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

Food and Agriculture Organization of the United Nations (2018). FAOSTAT

## 714 6. References

- 715 Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013).  
 716 Implication of crop model calibration strategies for assessing regional im-<sup>759</sup>  
 717 pacts of climate change in europe. *Agric. For. Meteorol.*, *170*, 32 – 46.
- 718 Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D.,  
 719 A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Al-<sup>761</sup>  
 720 derman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Bier-<sup>763</sup>  
 721 nath, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising tempera-<sup>764</sup>  
 722 tures reduce global wheat production. *Nature Climate Change*, *5*, 143–147.<sup>765</sup>  
 723 doi:10.1038/nclimate2470.
- 724 Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A.,  
 725 J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N.,<sup>767</sup>  
 726 Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath,<sup>769</sup>  
 727 C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simulat-<sup>770</sup>  
 728 ing wheat yields under climate change. *Nature Climate Change*, *3*, 827832.<sup>771</sup>  
 729 doi:10.1038/nclimate1916.
- 730 Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other  
 731 Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Ef-<sup>774</sup>  
 732 ficiency, Carbon Sequestration, and Environmental Pollution. *Advances in*  
 733 *Agronomy*, *86*, 341 – 409.
- 734 Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov,<sup>777</sup>  
 735 N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat  
 736 production potentials and management flexibility under the representative  
 737 concentration pathways. *Global and Planetary Change*, *122*, 107 – 121.
- 738 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields  
 739 from global gridded crop models. *Agricultural and Forest Meteorology*, *236*,  
 740 145 – 161.
- 741 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded  
 742 crop models using statistical estimates. *Agricultural and Forest Meteorol-*  
 743 *ogy*, *214-215*, 134 – 147.
- 744 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S.<sup>787</sup>

- 788 database. URL: <http://www.fao.org/faostat/en/home>. 831
- 789 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T.,<sup>832</sup>
- 790 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P.,<sup>833</sup>
- 791 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F.,<sup>834</sup>
- 792 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,<sup>835</sup>
- 793 Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J.,<sup>836</sup>
- 794 Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nen-<sup>837</sup>
- 795 del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna,<sup>838</sup>
- 796 M., Semenov, M. A., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K.,<sup>839</sup>
- 797 Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model<sup>840</sup>
- 798 wheat yield impact response surfaces showing sensitivity to temperature and<sup>841</sup>
- 799 precipitation change. *Agricultural Systems*, 159, 209–224. 842
- 800 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014).<sup>843</sup>
- 801 Evaluating the utility of dynamical downscaling in agricultural impacts pro-<sup>844</sup>
- 802 jections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781.<sup>845</sup>
- 803 Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitiv-<sup>846</sup>
- 804 ity of Agricultural Model Performance to Different Climate Inputs. *Journal*<sup>847</sup>
- 805 of Applied Meteorology and Climatology
- 806 Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported<sup>849</sup>
- 807 Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous<sup>850</sup>
- 808 Crop Growth and Yield: Application for Wheat in Central Europe. *Remote*<sup>851</sup>
- 809 *Sensing*, 7, 3934–3965. 852
- 810 He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H.,<sup>853</sup>
- 811 He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water<sup>854</sup>
- 812 contents and nitrogen leaching to precipitation, management practices and<sup>855</sup>
- 813 soil hydraulic properties in semi-arid and humid regions of Canada using the<sup>856</sup>
- 814 DSSAT model. *Nutrient Cycling in Agroecosystems*, 106, 201–215. 857
- 815 Heady, E. O. (1957). An Econometric Investigation of the Technology of Agri-<sup>858</sup>
- 816 cultural Production Functions. *Econometrica*, 25, 249–268. 859
- 817 Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa<sup>860</sup>
- 818 State University Press. 861
- 819 Holzkämper, A., Calanca, P., & Fuhrer, J. (2012). Statistical crop models:<sup>862</sup>
- 820 Predicting the effects of temperature and precipitation changes. *Climate*<sup>863</sup>
- 821 *Research*, 51, 11–21. 864
- 822 Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I.,<sup>865</sup>
- 823 McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore,<sup>866</sup>
- 824 A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake,<sup>867</sup>
- 825 A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgliesh,<sup>868</sup>
- 826 N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp,<sup>869</sup>
- 827 J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson,<sup>870</sup>
- 828 M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T.,<sup>871</sup>
- 829 Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorf, J.,<sup>872</sup>
- 830 Wedgwood, S., & Keating, B. A. (2014). APSIM Evolution towards a new<sup>873</sup>
- 831 generation of agricultural systems simulation. *Environmental Modelling and Software*, 62, 327 – 350.
- 832 Howden, S., & Crimp, S. (2005). Assessing dangerous climate change impacts on australia's wheat industry. *Modelling and Simulation Society of Australia and New Zealand*, (pp. 505–511).
- 833 Izumi, T., Nishimori, M., & Yokozawa, M. (2010). Diagnostics of climate model biases in summer temperature and warm-season insolation for the simulation of regional paddy rice yield in japan. *Journal of Applied Meteorology and Climatology*, 49, 574–591.
- 834 Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Nitrogen Fertilizers. *Ambio*, 6, 146–151.
- 835 Izaurrealde, R., Williams, J., McGill, W., Rosenberg, N., & Quiroga Jakas, M. (2006). Simulating soil C dynamics with EPIC: Model description and testing against long-term data. *Ecological Modelling*, 192, 362–384.
- 836 Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. *Agriculture, Ecosystems & Environment*, 93, 73 – 85.
- 837 Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijssman, A., & Ritchie, J. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18, 235 – 265.
- 838 Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural Systems*, 155, 269 – 288.
- 839 Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K., Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, 18, 267 – 288.
- 840 Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., & Smith, B. (2013). Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa. *Earth System Dynamics*, 4, 385–407.
- 841 Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling wheat yield and crop water productivity with high resolution on a global scale. *Agricultural Systems*, 94, 478 – 493.
- 842 Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a). Global investigation of impacts of PET methods on simulating crop-water relations for maize. *Agricultural and Forest Meteorology*, 221, 164 – 175.
- 843 Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs

- 874 with yields from major crop cultivations. *Science of The Total Environment*, 917  
 875 572, 526 – 537. 918  
 876 Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict 919  
 877 crop yield responses to climate change. *Agricultural and Forest Meteorol-* 920  
 878 *ogy*, 150, 1443 – 1452. 921  
 879 Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relation- 922  
 880 ships and the impacts of recent warming. *Environmental Research Letters*, 923  
 881 2, 014002. 924  
 882 MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447. 925  
 883 Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P., 926  
 884 Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi, 927  
 885 P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano, 928  
 886 D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of 929  
 887 large simulated yield datasets for studying climate effects. (p. 1100). 930  
 888 doi:10.13140/RG.2.1.5173.8328. 931  
 889 Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrolog- 932  
 890 ical modelling to study the impact of climate change on the water flows of 933  
 891 mountain watersheds. *Journal of Hydrology*, 376, 362 – 377. 934  
 892 Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., 935  
 893 & Calzadilla, A. (2009). Global biomass production potentials exceed ex- 936  
 894 pected future demand without the need for cropland expansion. *Nature Com-* 937  
 895 *munications*, 6. 938  
 896 McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srinivas, 939  
 897 C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S. 940  
 898 (2015). Integrated assessments of the impacts of climate change on agricul- 941  
 899 ture: An overview of AgMIP regional research in South Asia. *Chapter in:* 942  
*900 Handbook of Climate Change and Agroecosystems*, (pp. 201–218). 943  
 901 Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical 944  
 902 weather responsiveness of crop yields: US evidence and implications for 945  
 903 the agricultural impacts of climate change. *Environmental Research Letters*, 946  
 904 12. 947  
 905 Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate 948  
 906 change impacts on agriculture implies higher social cost of carbon. *Nature* 949  
 907 *Communications*, 8. 950  
 908 Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, 951  
 909 P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, 952  
 910 R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. 953  
 911 A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., 954  
 912 Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. 955  
 913 (2017). Global gridded crop model evaluation: benchmarking, skills, de- 956  
 914 ficiencies and implications. *Geoscientific Model Development*, 10, 1403–957  
 915 1422. 958  
 916 Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T. 959  
 (1997). Effect of CO<sub>2</sub> enrichment on carbon and nitrogen interaction in wheat and soybean. *Soil Science and Plant Nutrition*, 43, 789–798.  
 O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91, 1290 – 1300.  
 Olin, S., Schurgers, G., Lindeskog, M., Wårldin, D., Smith, B., Bodin, P., Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue C:N to changes in atmospheric CO<sub>2</sub> and N management in the main wheat regions of western europe. *Biogeosciences*, 12, 2489–2515. doi:10.5194/bg-12-2489-2015.  
 Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.  
 Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., & Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator. *Geoscientific Model Development*, 8, 1139–1155.  
 Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop yields and their variability at different levels of global warming. *Earth System Dynamics*, 9, 479–496.  
 Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff, S., & Gerten, D. (2015). Emulating global climate change impacts on crop yields. *Statistical Modelling*, 15, 499–525.  
 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas-  
 slos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.  
 Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet, J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M., Semenov, M., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. *Climate Research*, 65, 87–105.  
 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.

- 960 Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly  
961 Irrigated and Rainfed crop Areas around the Year 2000: A New High  
962 Resolution Data Set for Agricultural and Hydrological Modeling. *Global  
963 Biogeochemical Cycles*, 24, GB1011.
- 964 Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E.  
965 & Arneth, A. (2016). Climate analogues suggest limited potential for inten  
966 sification of production on current croplands under climate change. *Nature  
967 Communications*, 7, 12608.
- 968 Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term cli  
969 mate change based on a resampling ensemble technique. *Tellus A: Dynamical  
970 Meteorology and Oceanography*, 58, 461–472.
- 971 Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the  
972 reduction and sensitivity analysis of complex environmental models. *Environ  
973 mental Modelling & Software*, 34, 1 – 4.
- 974 Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling  
975 in water resources. *Water Resources Research*, 48.
- 976 Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017)  
977 Comparing and combining process-based crop models and statistical models  
978 with some implications for climate change. *Environmental Research Letters*  
979 12.
- 980 Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A.  
981 Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek,  
982 F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014)  
983 Assessing agricultural risks of climate change in the 21st century in a global  
984 gridded crop model intercomparison. *Proceedings of the National Academy  
985 of Sciences*, 111, 3268–3273.
- 986 Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P.  
987 Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ew  
988 ert, F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural  
989 Model Intercomparison and Improvement Project (AgMIP): Protocols and  
990 pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182.
- 991 Rosenzweig, C., Ruane, A. C., Antle, J., Elliott, J., Ashfaq, M., Chatta  
992 A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G.,  
993 Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M.,  
994 Müller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M.,  
995 Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., & Wiebe, K.  
996 (2018). Coordinating AgMIP data and models across global and regional  
997 scales for 1.5°C and 2.0°C assessments. *Philosophical Transactions of the  
998 Royal Society of London A: Mathematical, Physical and Engineering Sci  
999 ences*, 376.
- 1000 Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason  
1001 D'Croz, D., Müller, C., Porter, C., Phillips, M. M., Raymundo, R. M.,  
1002 Sands, R., Valdivia, R. O., White, J. W., Wiebe, K., & Rosenzweig, C.  
1003 (2018). Biophysical and economic implications for agriculture of +1.5° and  
1004 +2.0°C global warming using AgMIP Coordinated Global and Regional As  
1005 sessments. *Climate Research*, 76, 17–39.
- 1006 Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown,  
1007 D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013).  
1008 Climate change impact uncertainties for maize in panama: Farm informa  
1009 tion, climate projections, and yield sensitivities. *Agricultural and Forest  
1010 Meteorology*, 170, 132 – 145.
- 1011 Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forcing  
1012 datasets for agricultural modeling: Merged products for gap-filling and  
1013 historical climate series estimation. *Agric. Forest Meteorol.*, 200, 233–248.
- 1014 Ruane, A. C., McDermid, S., Rosenzweig, C., Baigorria, G. A., Jones, J. W.,  
1015 Romero, C. C., & Cecil, L. D. (2014). Carbon-temperature-water change  
1016 analysis for peanut production under climate change: A prototype for the  
1017 agmip coordinated climate-crop modeling project (c3mp). *Glob. Change  
1018 Biol.*, 20, 394–407. doi:10.1111/gcb.12412.
- 1019 Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901–  
1020 2100 depicted by world maps of the Köppen-Geiger climate classification.  
1021 *Meteorologische Zeitschrift*, 19, 135–141.
- 1022 Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng,  
1023 D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolin  
1024ski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frieler, K.  
1025 (2017). Consistent negative response of US crops to high temperatures in  
1026 observations and crop models. *Nature Communications*, 8, 13931.
- 1027 Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate  
1028 severe damages to U.S. crop yields under climate change. *Proceedings of  
1029 the National Academy of Sciences*, 106, 15594–15598.
- 1030 Snyder, A., Calvin, K. V., Phillips, M., & Ruane, A. C. (2018). A crop yield  
1031 change emulator for use in gcam and similar models: Persephone v1.0. *Geo  
1032 cientific Model Development Discussions*, 2018, 1–42.
- 1033 Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implemen  
1034 tation and evaluation of nonparametric regression procedures for sensitivity  
1035 analysis of computationally demanding models. *Reliability Engineering &  
1036 System Safety*, 94, 1735 – 1763.
- 1037 Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5  
1038 and the experiment design. *Bulletin of the American Meteorological Society*,  
1039 93, 485–498.
- 1040 Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of cl  
1041 imate change impacts on global crop yields. *Geophysical Research Letters*,  
1042 35.
- 1043 Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., &  
1044 Martin, J. F. (2014). Modeling sugarcane yield with a process-based model  
1045 from site to continental scale: Uncertainties arising from model structure

- 1046 and parameter values. *Geoscientific Model Development*, 7, 1225–1245.
- 1047 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe,  
1048 J. (2014). The Inter-Sectoral Impact Model Intercomparison Project  
1049 (ISI-MIP): Project framework. *Proceedings of the National Academy of  
1050 Sciences*, 111, 3228–3232.
- 1051 White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Method-  
1052 ologies for simulating impacts of climate change on crop production. *Field  
1053 Crops Research*, 124, 357 – 368.
- 1054 Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T.,  
1055 Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance  
1056 against site observations of irrigated maize from Mead, Nebraska. *Geosci-  
1057 entific Model Development*, 10, 1291–1320.
- 1058 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield vari-  
1059 ability in JULES-crop and implications for forcing with seasonal weather  
1060 forecasts. *Geoscientific Model Development*, 8, 3987–3997.
- 1061 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Land-  
1062 bouwkundige Onderzoeken* : 64.6, .
- 1063 Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato  
1064 yields on changes in climate and co2. *Agricultural and Forest Meteorology*,  
1065 112, 217 – 231.
- 1066 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao,  
1067 Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A.,  
1068 Li, T., Lin, E., Liu, Q., Martre, P., Mller, C., Peng, S., Peuelas, J., Ruane,  
1069 A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng,  
1070 S. (2017). Temperature increase reduces global yields of major crops in four  
1071 independent estimates. *Proc. Natl. Acad. Sci.*, 114, 9326–9331.