

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

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

^aDepartment of the Geophysical Sciences, University of Chicago, Chicago, IL, USA

^bCenter for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA

^cDepartment of Computer Science, University of Chicago, Chicago, IL, USA

^dPotsdam Institute for Climate Impact Research, Leibniz Association (Member), Potsdam, Germany

^eNASA Goddard Institute for Space Studies, New York, NY, United States

^fJoint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA

^gEcosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria

^hDepartment of Soil Science, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovak Republic

ⁱLaboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France

^jSino-French Institute of Earth System Sciences, College of Urban and Environmental Sciences, Peking University, Beijing, China

^kUnité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d'Astrophysique et de Géophysique, University of Liège, Belgium

^lMet Office Hadley Centre, Exeter, United Kingdom

^mDepartment of Geography, Ludwig-Maximilians-Universität, Munich, Germany

ⁿGeorg-August-University Göttingen, Tropical Plant Production and Agricultural Systems Modelling, Göttingen, Germany

^oDepartment of Geographical Sciences, University of Maryland, College Park, MD, USA

^pTexas AgriLife Research and Extension, Texas A&M University, Temple, TX, USA

^qDepartment of Statistics, University of Chicago, Chicago, IL, USA

^rEWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

^sDepartment of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

^tEarth Institute Center for Climate Systems Research, Columbia University, New York, NY, USA

^uKarlsruhe Institute of Technology, IMK-IFU, 82467 Garmisch-Partenkirchen, Germany.

^vSchool of Geography, Earth and Environmental Science, University of Birmingham, Birmingham, UK.

Abstract

Concerns about food security under climate change have motivated efforts to better understand the future changes in yields by using detailed process-based models in agronomic sciences. Process-based models differ on many details and considerable uncertainty remains in future yield projections. We conduct a large simulation set with perturbations in atmospheric CO₂ 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 a data-rich basis of projected yield changes across 12 models and five crops (maize, soy, rice, spring wheat, and winter wheat) using global gridded simulations. We present the simulation outputs from Phase II of the GGCMI effort, a targeted experiment aimed at understanding the interaction between multiple climate variables (as well as management), and show some initial summary results from the model intercomparison project. We then construct an “emulator or statistical representation of the simulated 30-year mean climatological output in each location for each crop and model. The emulator captures the response of the process-based models in a lightweight, computationally tractable form that facilitates model comparison as well as potential applications in subsequent modeling efforts such as integrated assessment.

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

1. Introduction

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

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

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

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

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

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

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

67 models cannot apply. Yield data are also often limited in the de-
 68 veloping world, where future climate impacts may be the most
 69 critical. Second, only process-based models can capture the
 70 growth response to elevated CO₂, novel conditions that are not
 71 represented in historical data (e.g. Pugh et al., 2016, Roberts
 72 et al., 2017). Similarly process-based models can represent
 73 novel changes in management practices (e.g. fertilizer input)
 74 that may ameliorate climate-induced damages.

75 Statistical emulation of crop simulations has been used to
 76 combine advantageous features of both statistical and process-
 77 based models. The statistical representation of complicated nu-
 78 matical simulation (e.g. O'Hagan, 2006, Conti et al., 2009), in
 79 which simulation output acts as the training data for a statisti-
 80 cal model, has been of increasing interest with the growth of
 81 simulation complexity and volume of output. Such emulators
 82 or "surrogate models" have been used in a variety of fields in-
 83 cluding hydrology (e.g. Razavi et al., 2012), engineering (e.g.
 84 Storlie et al., 2009), environmental sciences (e.g. Ratto et al.,
 85 2012), and climate (e.g. Castruccio et al., 2014, Holden et al.,
 86 2014). For agricultural impacts studies, emulation of process-
 87 based models allows exploring crop yields in regions outside
 88 ranges of current cultivation and with input variables outside
 89 historical precedents, in a lightweight, flexible form that is com-
 90 patible with economic studies.

91 In the past decade, many studies have developed emulators of
 92 crop yields from process-based models. Early studies propos-
 93 ing or describing potential emulators include Howden & Crimp
 94 (2005), Räisänen & Ruokolainen (2006) and Lobell & Burke
 95 (2010). In an early application, Ferrise et al. (2011) used a Arti-
 96 ficial Neural Net trained on simulation outputs to predict wheat₁₀₁
 97 yields in the Mediterranean. Studies developing single-model₁₀₃
 98 emulators include Holzkämper et al. (2012) for the CropSyst₁₀₄
 99 model, Ruane et al. (2013) for the CERES wheat model, Oye-₁₀₅
 100 bamiji et al. (2015) for the LPJmL model (for multiple crops,₁₀₆

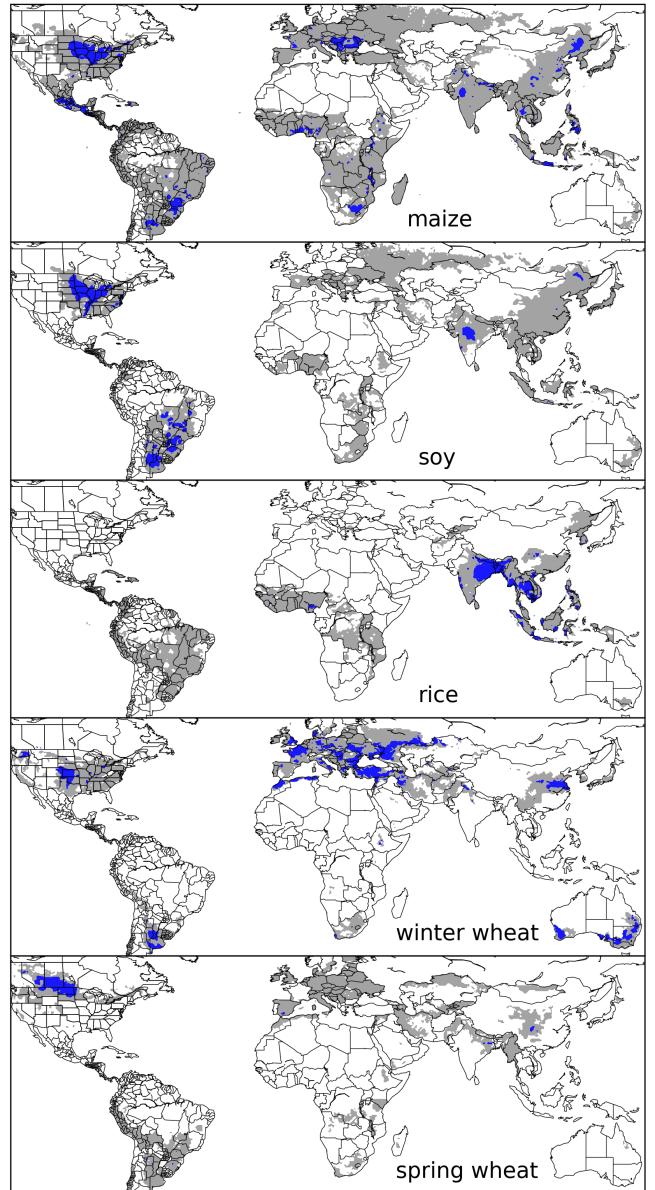


Figure 1: Presently cultivated area for rain-fed crops in the real world. Blue contour area indicates grid cells with more than 20,000 hectares (approx. 10% of the land in a grid cell near the equator for example) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated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

using multiple scenarios as a training set). In recent years, emulators have begun to be used in the context of multi-model intercomparisons, with Blanc & Sultan (2015), Blanc (2017), Ostberg et al. (2018) and Mistry et al. (2017) using them to analyze the five crop models of the Inter-Sectoral Impacts Model Intercomparison Project (ISIMIP) (Warszawski et al., 2014) (for

maize, soy, wheat, and rice). Approaches differ: Blanc & Sultan¹³⁵ (2015) and Blanc (2017) used local weather variables (and CO₂¹³⁶ values) and yields but emulate across soil types using historical¹³⁷ simulations and a future climate scenario (RCP8.5 over multiple¹³⁸ climate models); Ostberg et al. (2018) used global mean¹³⁹ temperature change (and CO₂) as regressors but pattern-scale¹⁴⁰ to emulate local yields using multiple climate scenarios; Mistry¹⁴¹ et al. (2017) used local weather and yields and a historical simulation¹⁴² and compare with data. Other studies have used the development¹⁴³ of emulators (or response surface) to analyze non-RCP¹⁴⁴ crop model simulations that sampled a suite of climate¹⁴⁵ (and management) perturbations: Makowski et al. (2015) for temperature,¹⁴⁶ CO₂, and nitrogen, Pirttioja et al. (2015) and Snyder et al.¹⁴⁷ (2018) for temperature, water, and CO₂, and (Fronzek¹⁴⁸ et al., 2018) for temperature and water, with all studies simulating¹⁴⁹ selected sites for a limited number of crops.¹⁵⁰

GGCMI protocol builds on the AgMIP Coordinated Climate-Crop Modeling Project (C3MP) (Ruane et al., 2014, McDermid et al., 2015) and will contribute to the AgMIP Coordinated Global and Regional Assessments (CGRA) (Ruane et al., 2018, Rosenzweig et al., 2018).

GGCMI Phase II is designed to allow addressing goals such as understanding where highest-yield regions may shift under climate change; exploring future adaptive management strategies; understanding how interacting parameters affect crop yield; quantifying uncertainties across models and major drivers; and testing strategies for producing lightweight emulators of process-based models. In this paper, we describe the GGCMI Phase II experiments, summarize output and present initial results, demonstrate that it is tractable to emulation, and present a simple climatological emulator as a potential tool for impacts assessments.¹⁵⁰

The Global Gridded Crop Model Intercomparison (GGCMI)

Phase II experiment seeks to provide a comprehensive global dataset to allow systematically exploring how process-based crop models for the major crop respond to the main climate and management drivers and their interactions. The experiment involves running a suite of process-based crop models across historical conditions perturbed by a set of defined input parameters, and was conducted as part of the Agricultural Model Intercomparison and Improvement Project (AgMIP) (Rosenzweig et al., 2013, 2014), an international effort conducted under a framework similar to the Climate Model Intercomparison Project (CMIP) (Taylor et al., 2012, Eyring et al., 2016). The

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

| Input variable | Abbr. | Tested range | Unit |
|------------------|-------|--|---------------------|
| CO ₂ | 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 _{inf}) | % |
| Applied nitrogen | N | 10, 60, 200 | kg ha ⁻¹ |

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_{inf}) simulations.

161 their historical values. The reduced set of crops includes the¹⁹⁴
162 three major global cereals and the major legume and accounts¹⁹⁵
163 for over 50% of human calories (in 2016, nearly 3.5 billion tons¹⁹⁶
164 or 32% of total global crop production by weight (Food and¹⁹⁷
165 Agriculture Organization of the United Nations, 2018).

166 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.¹⁹⁹
- Study the interactions between climate variables and nitrogen inputs in driving modeled yield impacts.²⁰⁰
- Explore differences in crop response to warming across the Earth's climate regions.²⁰¹
- Provide a dataset that allows statistical emulation of crop model responses for downstream modelers.²⁰²
- Illustrate differences in potential adaptation via growing season changes.²⁰³

177 The guiding scientific rationale of GGCMI Phase II is to pro-

178 vide a comprehensive, systematic evaluation of the response²¹²
179 of process-based crop models to different values for carbon²¹³
180 dioxide, temperature, water, and applied nitrogen (collectively²¹⁴
181 known as "CTWN"). Phase II of the GGCMI project consists²¹⁵
182 of a series of simulations, each with one or more of the CTWN²¹⁶
183 dimensions perturbed over the 31-year historical time series²¹⁷
184 (1980-2010) used in Phase I. In most cases, historical daily cli-²¹⁸
185 mate inputs are taken from the 0.5 degree NASA AgMERRA²¹⁹
186 daily gridded re-analysis product specifically designed for agri-²²⁰
187 cultural modeling, with satellite-corrected precipitation (Ruane²²¹
188 et al., 2015). Two models require sub-daily input data and use²²²
189 alternative sources. See Elliott et al. (2015) for additional de-²²³
190 tails.

191 The experimental protocol consists of 9 levels for precipita-²²⁵
192 tion perturbations, 7 for temperature, 4 for CO₂, and 3 for ap-²²⁶
193 plied nitrogen, for a total of 672 simulations for rain-fed agri-²²⁷

culture and an additional 84 for irrigated (Table 1). For irrigated simulations, soil water is held at either field capacity or, for those models that include water-log damage, at maximum beneficial level. Temperature perturbations are applied as absolute offsets from the daily mean, minimum, and maximum temperature time series for each grid cell used as inputs. Precipitation perturbations are applied as fractional changes at the grid cell level, and carbon dioxide and nitrogen levels are specified as discrete values applied uniformly over all grid cells. Note that CO₂ changes are applied independently of changes in climate variables, so that higher CO₂ 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 shares 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 mod-

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

228 els have individual assumptions on soil organic matter that may₂₄₆ any of four initially specified levels of participation, so the num-
 229 release additional nitrogen through mineralization. See Rosen-₂₄₇ ber of simulations varies by model, with some sampling only a
 230 zweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017)₂₄₈ part of the experiment variable space. Most modeling groups
 231 for further details on models and underlying assumptions. ₂₄₉ simulate all five crops in the protocol, but some omitted one
 232 Each model is run at 0.5 degree spatial resolution and covers ₂₅₀ or more. Table 2 provides details of coverage for each model.
 233 all currently cultivated areas and much of the uncultivated land ₂₅₁ Note that the three models that provide less than 50 simulations
 234 area. Coverage extends considerably outside currently culti- ₂₅₂ are excluded from the emulator analysis.
 235 vated areas because cultivation will likely shift under climate ₂₅₃
 236 change. See Figure 1 for the present-day cultivated area of ₂₅₄ All models produce as output, crop yields ($\text{tons ha}^{-1} \text{ year}^{-1}$)
 237 rain-fed crops, and Figure S1 in the supplemental material for ₂₅₅ for each 0.5 degree grid cell. Because both yields and yield
 238 irrigated crops. Some areas such as Greenland, far-northern ₂₅₆ changes vary substantially across models and across grid cells,
 239 Canada, Siberia, Antarctica, the Gobi and Sahara deserts, and ₂₅₇ we primarily analyze relative change from a baseline. We take
 240 central Australia are not simulated as they are assumed to re- ₂₅₈ as the baseline the scenario with historical climatology (i.e. T
 241 main non-arable even under an extreme climate change. Grow- ₂₅₉ and P changes of 0). C of 360 ppm, and applied N at 200 kg
 242 ing seasons are standardized across models with data adapted ₂₆₀ ha^{-1} . We show absolute yields in some cases to illustrate geo-
 243 from several sources (Sacks et al., 2010, Portmann et al., 2008, ₂₆₁ graphic differences in yields for a single model.
 244 2010).
 245 The participating modeling groups provide simulations at₂₆₃

262 2.2. *Simulation model validation approach*

263 Simulation model validation for GGCMI Phase II builds on
 the validation efforts presented in Müller et al. (2017) for the

264 first phase. In the case presented here however, the models
 265 are not run on the best approximation of management levels
 266 (namely nitrogen application level) by region as with Phase I.
 267 As the goals of this phase of the project are focused on under-
 268 standing the sensitivity in *change* in yield to changes in input
 269 drivers –and not to simulate historical yields as accurately as
 270 possible– no direct comparison to historical yield data can be
 271 made. Additionally, even when provided with an appropriate
 272 local nitrogen level, models simulated *potential* yields that do
 273 not include reductions from pests, weeds, or diseases. Poten-
 274 tial yields represent an ideal case that is not realized in many
 275 less industrialized areas, and some models are not calibrated as
 276 they were in Phase I of the project.

277 We evaluate the models here based on the response to year-
 278 to-year temperature and precipitation variability in the histori-
 279 cal record. If the models can (somewhat) faithfully represent
 280 the the historical variability in yields (which, once detrended
 281 to account for changing management levels must be driven
 282 largely by differences in weather), then the models may provide
 283 some utility in understanding the impact on mean climatologi-
 284 cal shifts in temperature and precipitation. Specifically, we cal-
 285 culate a Pearson correlation coefficient between the detrended
 286 time series of simulations and FAO data (Food and Agricul-
 287 ture Organization of the United Nations, 2018) for the period
 288 1981-2009 at the national scale. The FAO data is detrended
 289 because much of the change in yield over this time period is
 290 due to intensification and changes in management (Ray et al.,
 291 2012). Validating the response to CO₂ and nitrogen applica-
 292 tions is more difficult because real world data is not available
 293 outside of small greenhouse and field level trials.

294 2.3. Climatological-mean yield emulator design

295 We construct our emulator at the 30-year climatological
 296 mean level. Blanc & Sultan (2015) and Blanc (2017) have
 297 shown that a emulator of a global process-based crop model can

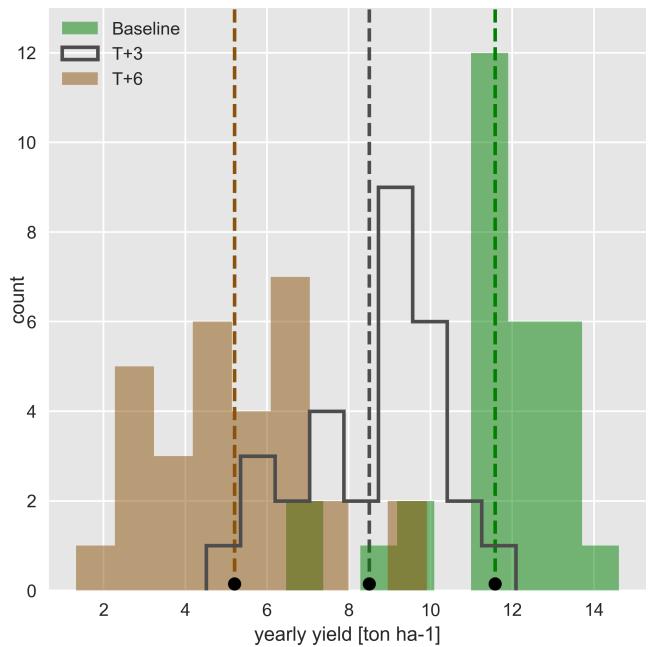


Figure 2: Example showing both climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for a grid cell in northern Iowa (a representative high-yield region) from the pDSSAT model, for the baseline climatology (1981-2010) and for scenarios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

be successfully developed at the yearly scale. Our decision to construct a climatological-mean yield emulator is driven by the target application for this analysis tool. Many impact modelers are not focused on the changes in the year-to-year variability in yields, but instead on the broad mean changes over the multi-decadal timescale. Emulation involves fitting individual regression models for each crop, simulation model, and 0.5 degree geographic pixel from the GGCMI Phase II data set. The regressors are the applied constant perturbations in temperature, water, nitrogen and CO₂, 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. The emulator parameter matrices are portable and the yield computations are cheap even at the half-degree grid cell resolution, so we do not aggregate in space at this time.

Blanc & Sultan (2015) and Blanc (2017) have shown that a fractional polynomial specification is more effective than a standard polynomial for representing simulations at the yearly level across different soil types geographically (not at the grid cell level). We do not test this specification here, and instead use as a starting point a standard third-order polynomial to represent the climatological-mean response at the grid cell level as it is the simplest effective specification. 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. Aulakh & Malhi, 2005), and between nitrogen and carbon dioxide (Osaki et al., 1992, Nakamura et al., 1997). We do not focus on comparing different model specifications in this study, and instead stick to a relatively simple parameterized specification that allows for some, albeit limited, coefficient interpretation.

The limited GGCMI variable sample space means that use

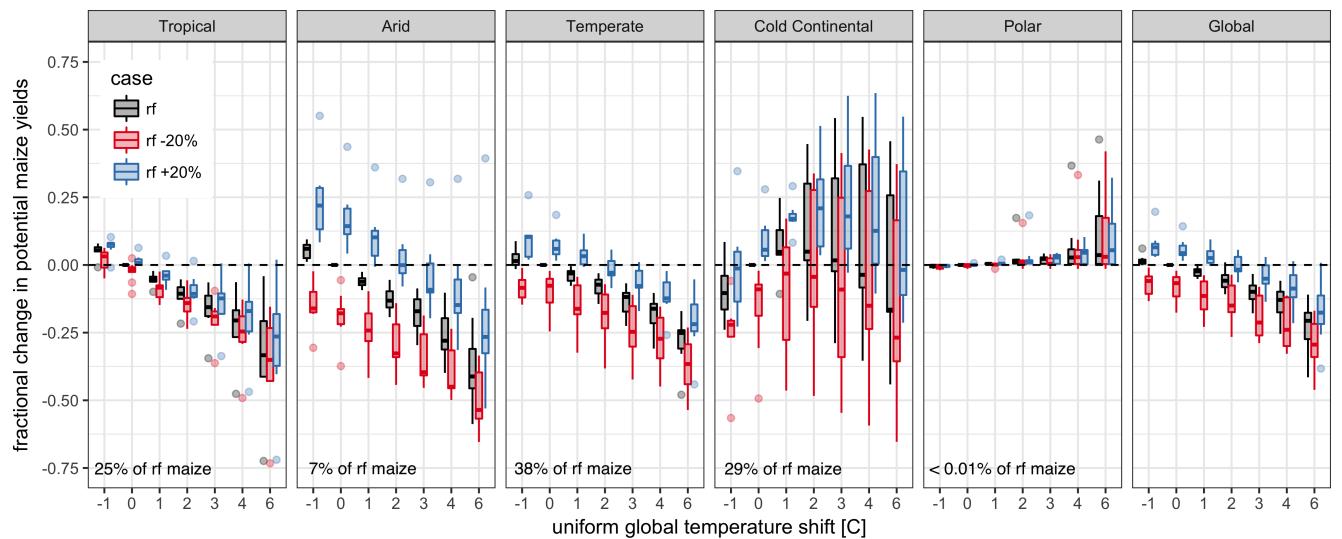


Figure 3: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation levels (rain-fed (rf) -20%, rain-fed (0), and rain-fed +20%), with CO₂ and nitrogen held constant at baseline values (360 ppm and 200 kg ha⁻¹ yr⁻¹). Y-axis is fractional change in the regional average climatological potential yield relative to the baseline. The figure shows all modeled land area; see Figure S6 in the supplemental material for only currently-cultivated land. Panel text gives the percentage of rain-fed maize presently cultivated in each climate zone (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.

349 of the full polynomial expression described above, which has
 350 34 terms for the rain-fed case (12 for irrigated), can be prob-
 351 lematic, and can lead to over-fitting and unstable parameter es-
 352 timations. We therefore reduce the number of terms through a
 353 feature selection cross-validation process in which terms in the
 354 polynomial are tested for importance. In this procedure higher-
 355 order and interaction terms are added successively to the model;
 356 we then follow the reduction of the the aggregate mean squared
 357 error with increasing terms and eliminate those terms that do
 358 not contribute significant reductions. See supplemental docu-
 359 ments for more details. We select terms by applying the feature
 360 selection process to the three models that provided the com-
 361 plete set of 672 rain-fed simulations (pDSSAT, EPIC-TAMU,
 362 and LPJmL); the resulting choice of terms is then applied for
 363 all emulators.

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

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

To fit the parameters K , we use a Bayesian Ridge probabilistic estimator (MacKay, 1991), which reduces volatility in parameter estimates when the sampling is sparse, by weighting parameter estimates towards zero. The Bayesian Ridge method is necessary to maintain a consistent functional form across all models, and locations as the linear least squares fails to provide a stable result in many cases. In the GGCMI Phase II experiment, the most problematic fits are those for models that provided a limited number of cases or for low-yield geographic regions where some modeling groups did not run all scenarios. Because we do not attempt to emulate models that provided less than 50 simulations, the lowest number of simulations emulated across the full parameter space is 130 (for the PEPIC model). We use the implementation of the Bayesian Ridge estimator from the scikit-learn package in Python (Pedregosa et al., 2011).

The resulting parameter matrices for all crop model emulators are available on request, as are the raw simulation data and a Python application to emulate yields. The yield output for a single GGCMI model that simulates all scenarios and all five crops is ~ 12.5 GB; the emulator is ~ 100 MB, a reduction by over two orders of magnitude.

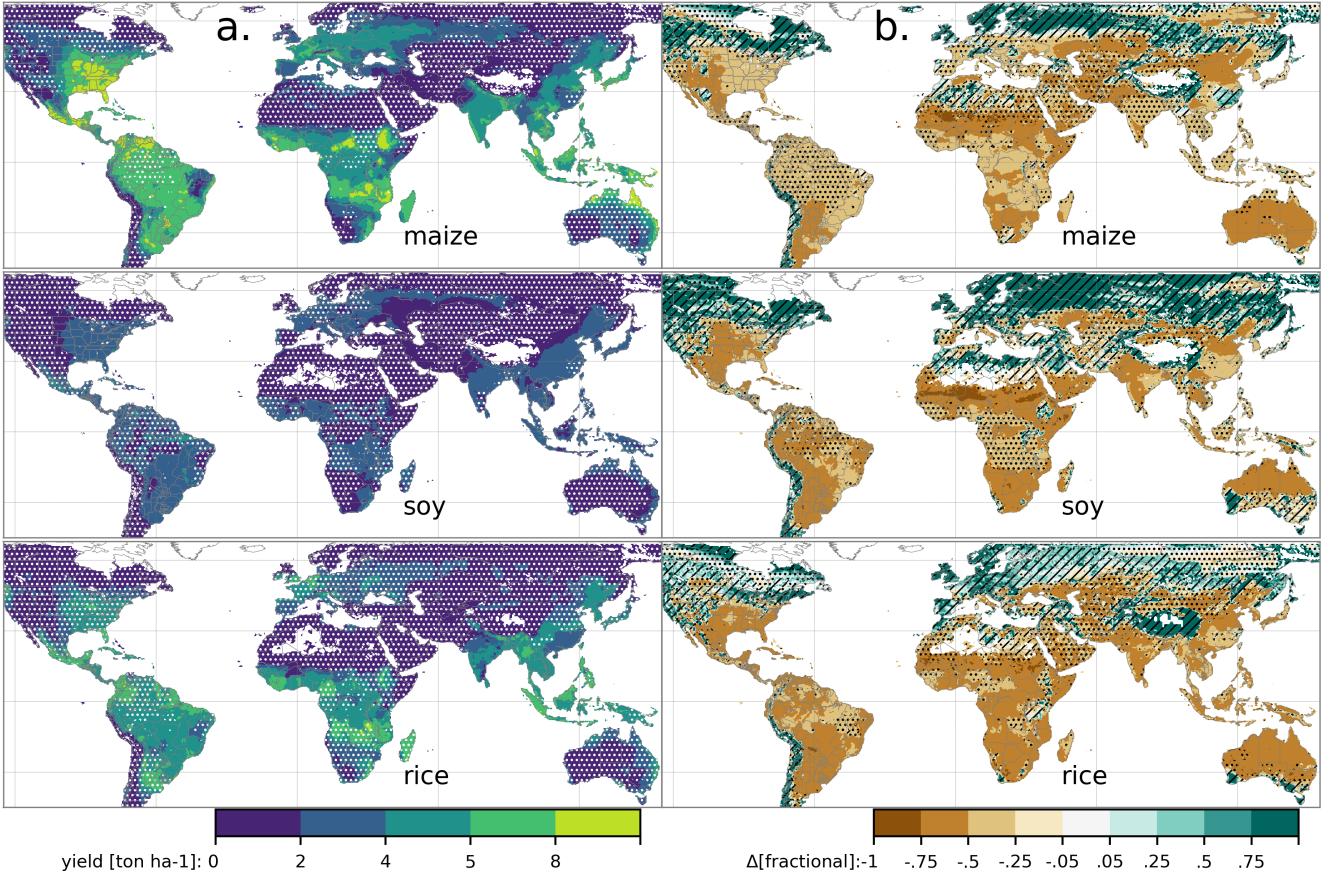


Figure 4: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure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

399 2.4. Emulator evaluation

400 Because no general criteria exist for defining an acceptable
 401 model emulator, we develop a metric of emulator performance
 402 specific to GGCMI. For a multi-model comparison exercise like
 403 GGCMI, a reasonable criterion is what we term the “normalized
 404 error”, which compares the fidelity of an emulator for a given
 405 model and scenario to the inter-model uncertainty. We define
 406 the normalized error e for each scenario as the difference be-
 407 tween the fractional yield change from the emulator and that in
 408 the original simulation, divided by the standard deviation of the
 409 multi-model spread (Equations 2 and 3):

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

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

410 Here $F_{scn.}$ is the fractional change in a model’s mean emu-
 411 lated or simulated yield from a defined baseline, in some sce-
 412 nario (scn.) in C, T, W, and N space; $Y_{scn.}$ and $Y_{baseline}$ are the
 413 absolute emulated or simulated mean yields. The normalized
 414 error e is the difference between the emulated fractional change
 415 in yield and that actually simulated, normalized by $\sigma_{sim.}$, the
 416 standard deviation in simulated fractional yields $F_{sim, scn.}$ across
 417 all models. The emulator is fit across all available simulation
 418 outputs, and then the error is calculated across the simulation
 419 scenarios provided by all nine models (Figure 8 and Figures

420 S12 and Figures S13 in supplemental documents). Note that₄₅₃
421 the normalized error e for a model depends not only on the fidelity₄₅₄
422 of its emulator in reproducing a given simulation but on₄₅₅
423 the particular suite of models considered in the intercomparison₄₅₆
424 exercise. The rationale for this choice is to relate the fidelity of₄₅₇
425 the emulation to an estimate of true uncertainty, which we take₄₅₈
426 as the multi-model spread.

427 3. Results

428 3.1. Simulation results

429 Crop models in the GGCMI ensemble show a broadly con-
430 sistent responses to climate and management perturbations in
431 most regions, with a strong negative impact of increased tem-
432 perature in all but the coldest regions. We illustrate this result
433 for rain-fed maize in Figure 3, which shows yields for the pri-
434 mary Köppen-Geiger climate regions (Rubel & Kottek, 2010).
435 In warming scenarios, models show decreases in maize yield in
436 the temperate, tropical, and arid regions that account for nearly
437 three-quarters of global maize production. These impacts are
438 robust for even moderate climate perturbations. In the temper-
439 ate zone, even a 1 degree temperature rise with other variables
440 held fixed leads to a median yield reduction that outweighs the
441 variance across models. A 6 degree temperature rise results in
442 median loss of ~25% of yields with a signal to noise of nearly
443 three. A notable exception is the cold continental region, where
444 models disagree strongly, extending even to the sign of impacts.
445 Model simulations of other crops produce similar responses to
446 warming, with robust yield losses in warmer locations and high
447 inter-model variance in the cold continental regions (Figures
448 S7).

449 The effects of rainfall changes on maize yields are also as ex-₄₈₂
450 pected and are consistent across models. Increased rainfall mit-₄₈₃
451 igates the negative effect of higher temperatures, most strongly₄₈₄
452 in arid regions. Decreased rainfall amplifies yield losses and₄₈₅

also increases inter-model variance more strongly, suggesting
that models have difficulty representing crop response to water
stress. We show only rain-fed maize here; see Figure S5 for the
irrigated case. As expected, irrigated crops are more resilient to
temperature increases in all regions, especially so where water
is limiting.

Mapping the distribution of baseline yields and yield changes
shows the geographic dependencies that underlie these results.
Figure 4 shows baseline and changes in the T+4 scenario for
rain-fed maize, soy, and rice in the multi-model ensemble mean,
with locations of model agreement marked. Absolute yield poten-
tials are have strong spatial variation, with much of the
Earth's surface area unsuitable for any given crop. In general,
models agree most on yield response in regions where yield
potentials are currently high and therefore where crops are cur-
rently grown. Models show robust decreases in yields at low
latitudes, and highly uncertain median increases at most high
latitudes. For wheat crops see Figure S11; wheat projections
are both more uncertain and show fewer areas of increased yield
in the inter-model mean.

473 3.2. Simulation model validation results

Figure 5 shows the Pearson 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 pro-
viding near best-in-class performance in at least one location-
crop combination. The presence of very few vertical dark green
color bars clearly illustrates the power of a multi-model inter-
comparison project like the one presented here. The ensemble
mean does not beat the best model in each case, but shows pos-
itive correlation in over 75% of the cases presented here. The
EPIC-TAMU model performs best for soy, CARIAB, EPIC-
TAMU, and PEPIC perform best for maize, PROMET performs

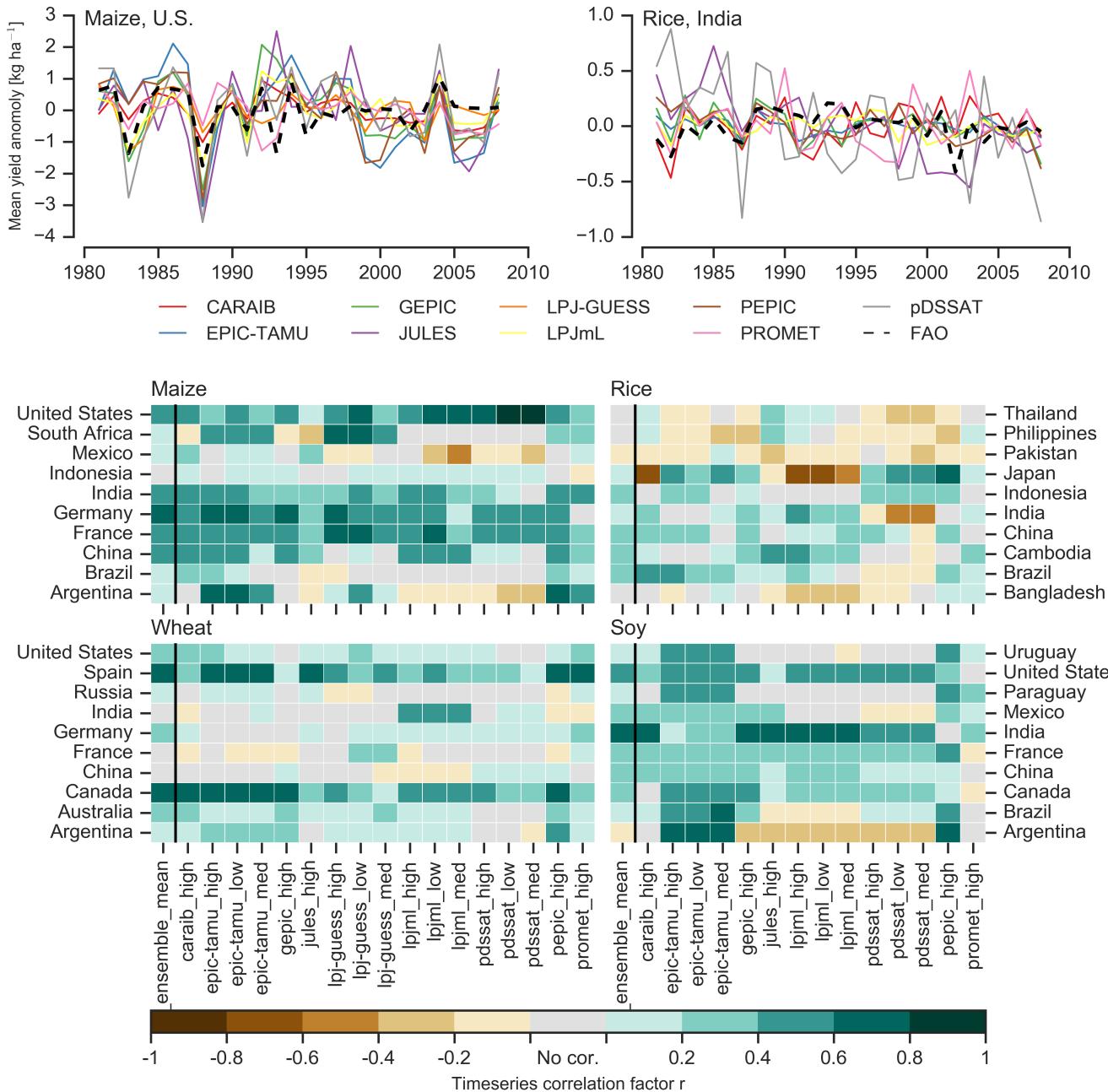


Figure 5: Time series correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen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

best for wheat, and the EPIC family of models perform best for rice.
 Soy is qualitatively the easiest crop to represent (except in Argentina), which is likely due in part to the invariance of the response to nitrogen application (soy fixes atmospheric nitrogen
 very efficiently). Comparison to the FAO data is therefore easier than the other crops because the nitrogen application levels do not matter. US maize has the best performance across models, with nearly every model representing the historical variability to a reasonable extent. Especially good example years for US

486 best for wheat, and the EPIC family of models perform best for 491
 487 rice. 492

488 Soy is qualitatively the easiest crop to represent (except in 494
 489 Argentina), which is likely due in part to the invariance of the 495
 490 response to nitrogen application (soy fixes atmospheric nitrogen

maize are 1983, 1988, and 2004 (top left panel of Figure 5),⁵³⁰ where every model gets the direction of the anomaly compared⁵³¹ to surrounding years correct. 1983 and 1988 are famously bad⁵³² years for US maize along with 2012 (not shown). US maize⁵³³ is possibly both the most uniformly industrialized (in terms of⁵³⁴ management practices) crop and the one with the best data col-⁵³⁵ lection in the historical period of all the cases presented here.⁵³⁶

The FAO data is at least one level of abstraction from ground⁵³⁷ truth in many cases, especially in developing countries. The⁵³⁸ failure of models to represent the year-to-year variability in rice⁵³⁹ in some countries in southeast Asia is likely partly due to model⁵⁴⁰ failure and partly due to lack of data. It is possible to speculate⁵⁴¹ that the difference in performance between Pakistan (no suc-⁵⁴² cessful models) and India (many successful models) for rice⁵⁴³ may reside at least in part in the FAO data and not the mod-⁵⁴⁴ els themselves. The same might apply to Bangladesh and In-⁵⁴⁵ dia for rice. Partitioning of these contributions is impossible at⁵⁴⁶ this stage. Additionally, there is less year-to-year variability in⁵⁴⁷ rice yields (partially due to the fraction of irrigated cultivation).⁵⁴⁸ Since the Pearson r metric is scale invariant, it will tend to score⁵⁴⁹ the rice models more poorly than maize and soy. An example⁵⁵⁰ of very poor performance can be seen with the pDSSAT model⁵⁵¹ for rice in India (top right panel of Figure 5).

3.3. Emulator performance

Emulation provides not only a computational tool but a⁵⁵⁴ means of understanding and interpreting crop yield response⁵⁵⁵ across the parameter space. Emulation is only possible, how-⁵⁵⁶ ever, when crop yield responses are sufficiently smooth and⁵⁵⁷ continuous to allow fitting with a relatively simple functional⁵⁵⁸ form. In the GGCMI simulations, this condition largely but⁵⁵⁹ not always holds. Responses are quite diverse across locations,⁵⁶⁰ crops, and models, but in most cases local responses are reg-⁵⁶¹ular enough to permit emulation. Figure 6 illustrates the ge-⁵⁶²ographic diversity of responses even in high-yield areas for a⁵⁶³

single crop and model (rain-fed maize in pDSSAT for various high-cultivation areas). This heterogeneity validates the choice of emulating at the grid cell level.

Each panel in Figure 6 shows model yield output from sce-⁵³⁷ narios varying only along a single dimension (CO_2 , tempera-⁵³⁸ ture, precipitation, or nitrogen addition), with other inputs held⁵³⁹ fixed at baseline levels; in all cases yields evolve smoothly⁵⁴⁰ across the space sampled. For reference we show the results⁵⁴¹ of the full emulation fitted across the parameter space. The⁵⁴² polynomial fit readily captures the climatological response to⁵⁴³ perturbations.

Crop yield responses generally follow similar functional⁵⁴⁴ forms across models, though with a spread in magnitude. Fig-⁵⁴⁵ ure 7 illustrates the inter-model diversity of yield responses⁵⁴⁶ to the same perturbations, even for a single crop and location⁵⁴⁷ (rain-fed maize in northern Iowa, the same location shown in⁵⁴⁸ the Figure 6). The differences make it important to construct⁵⁴⁹ emulators separately for each individual model, and the fidelity⁵⁵⁰ of emulation can also differ across models. This figure illus-⁵⁵¹ trates a common phenomenon, that models differ more in re-⁵⁵²sponse to perturbations in CO_2 and nitrogen perturbations than⁵⁵³ to those in temperature or precipitation. (Compare also Figures⁵⁵⁴ 3 and S18.) For this location and crop, CO_2 fertilization effects⁵⁵⁵ can range from ~5–50%, and nitrogen responses from nearly⁵⁵⁶ flat to a 60% drop in the lowest-application simulation.

While the nitrogen dimension is important and uncertain, it⁵⁵⁷ is also the most problematic to emulate in this work because⁵⁵⁸ of its limited sampling. The GGCMI protocol specified only⁵⁵⁹ three nitrogen levels (10, 60 and 200 $\text{kg N y}^{-1} \text{ha}^{-1}$), so a third-⁵⁶⁰ order fit would be over-determined but a second-order fit can⁵⁶¹ result in potentially unphysical results. Steep and nonlinear de-⁵⁶²clines in yield with lower nitrogen levels means that some re-⁵⁶³gressions imply a peak in yield between the 100 and 200 kg N ⁵⁶⁴ $\text{y}^{-1} \text{ha}^{-1}$ levels. While there may be some reason to believe

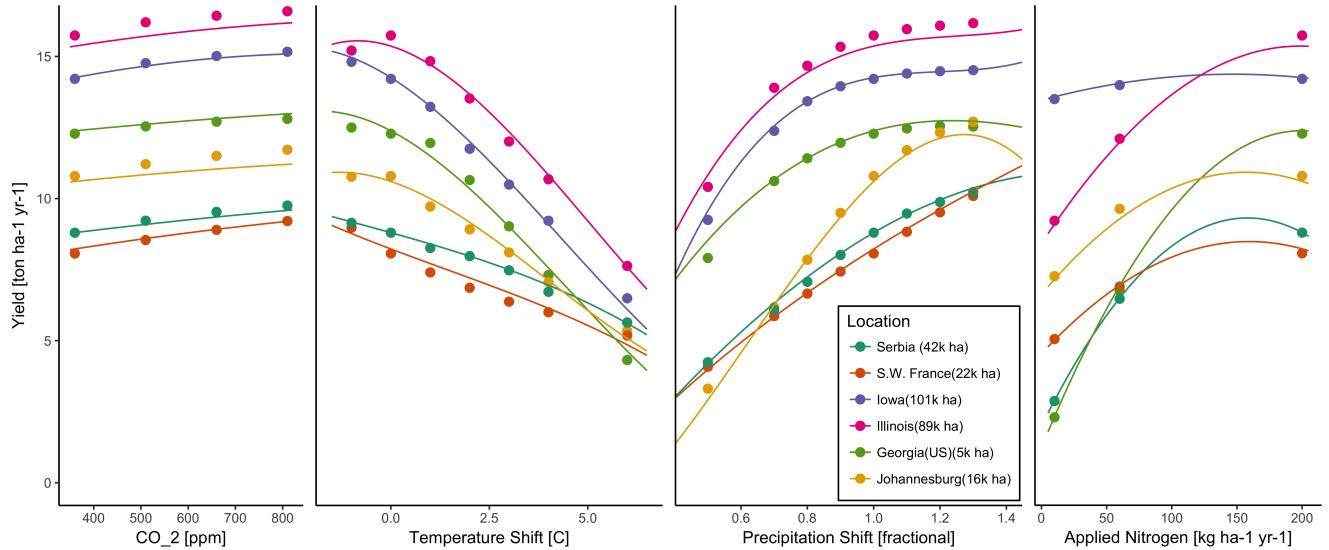


Figure 6: Example illustrating spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows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

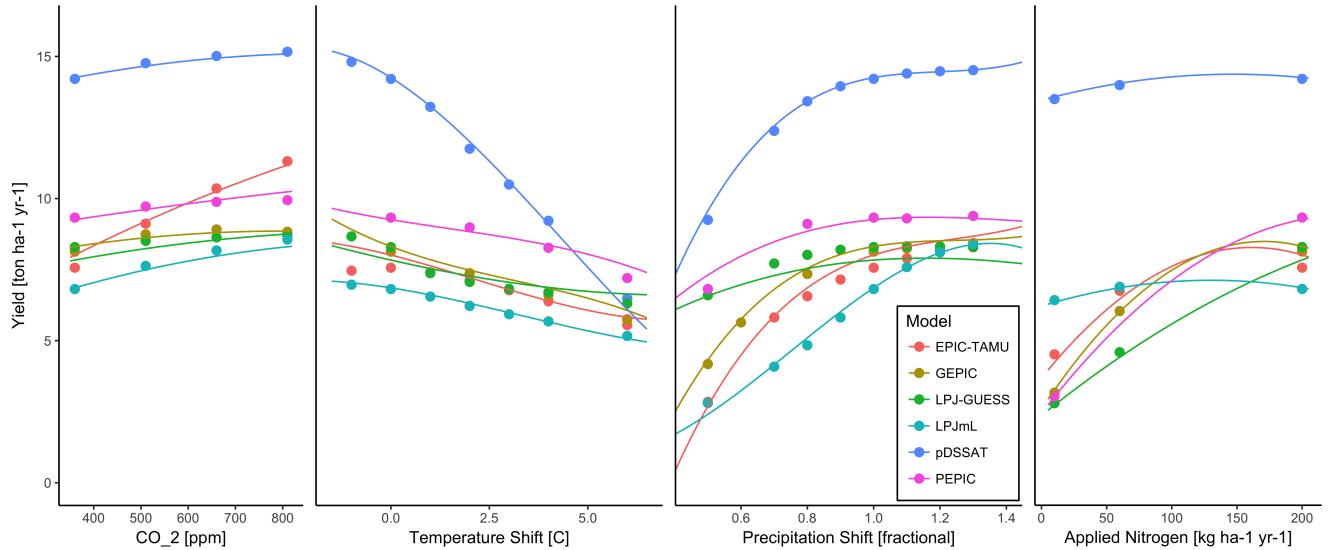


Figure 7: Example illustrating across-model variations in yield response. We show simulations and emulations from six models for rain-fed maize in the same Iowa location shown in Figure 6, with the same plot conventions. Models that did not simulate the nitrogen dimension are omitted for clarity. The inter-model standard deviation is larger for the perturbations in CO₂ and nitrogen than for those in temperature or precipitation illustrating that the sensitivity to weather is better constrained than to management at elevated CO₂. While most model responses can readily be emulated with a simple polynomial, some response surfaces are more complicated and lead to emulation error.

- 564 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.
- 565 566 567 568 569 570 571 572 573 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

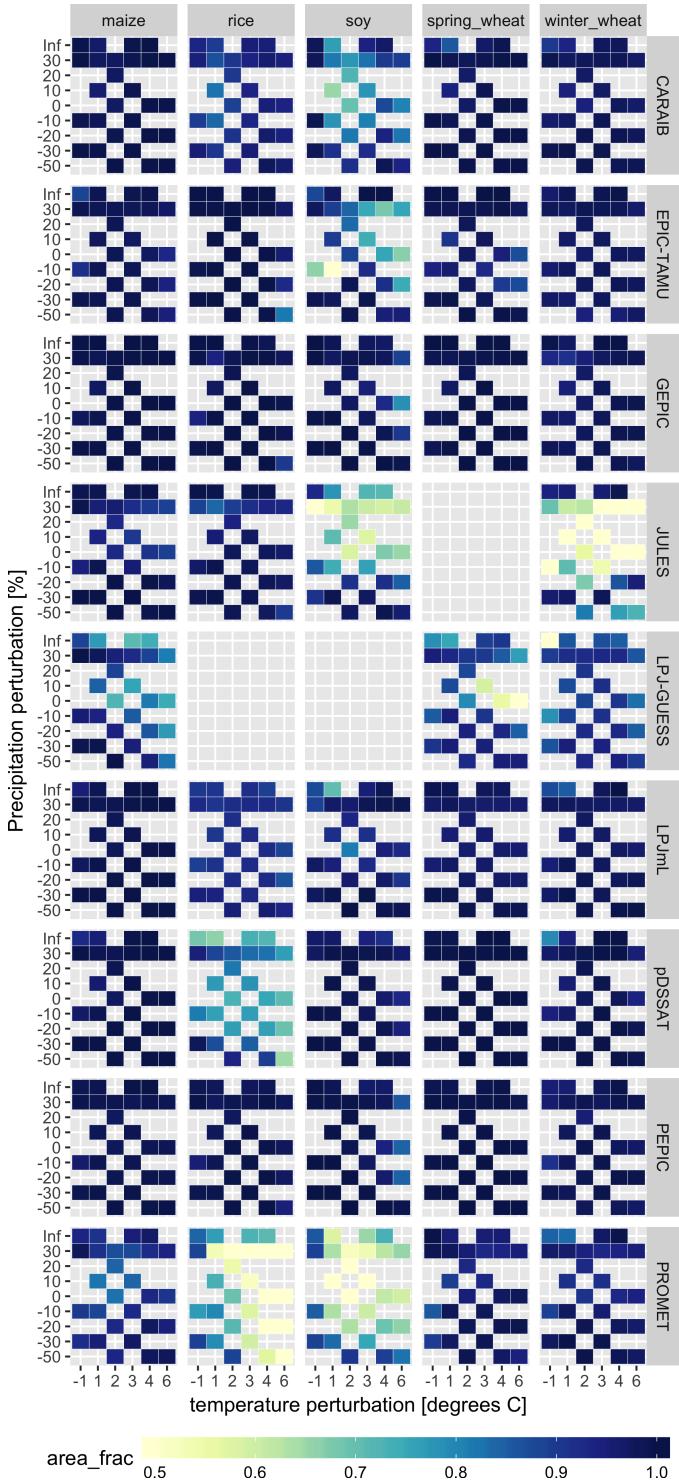


Figure 8: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO₂ and nitrogen held fixed at baseline values. Large columns are crops and large 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 CO₂ and N values, we consider only those for which all 9 models submitted data. JULES did not simulate spring wheat and LPJ-GUESS did not simulate rice and soy. Emulator performance is generally satisfactory, with some exceptions. Emulator failures (significant areas of poor performance) occur for individual crop-model combinations, with performance generally degrading for hotter and wetter scenarios.

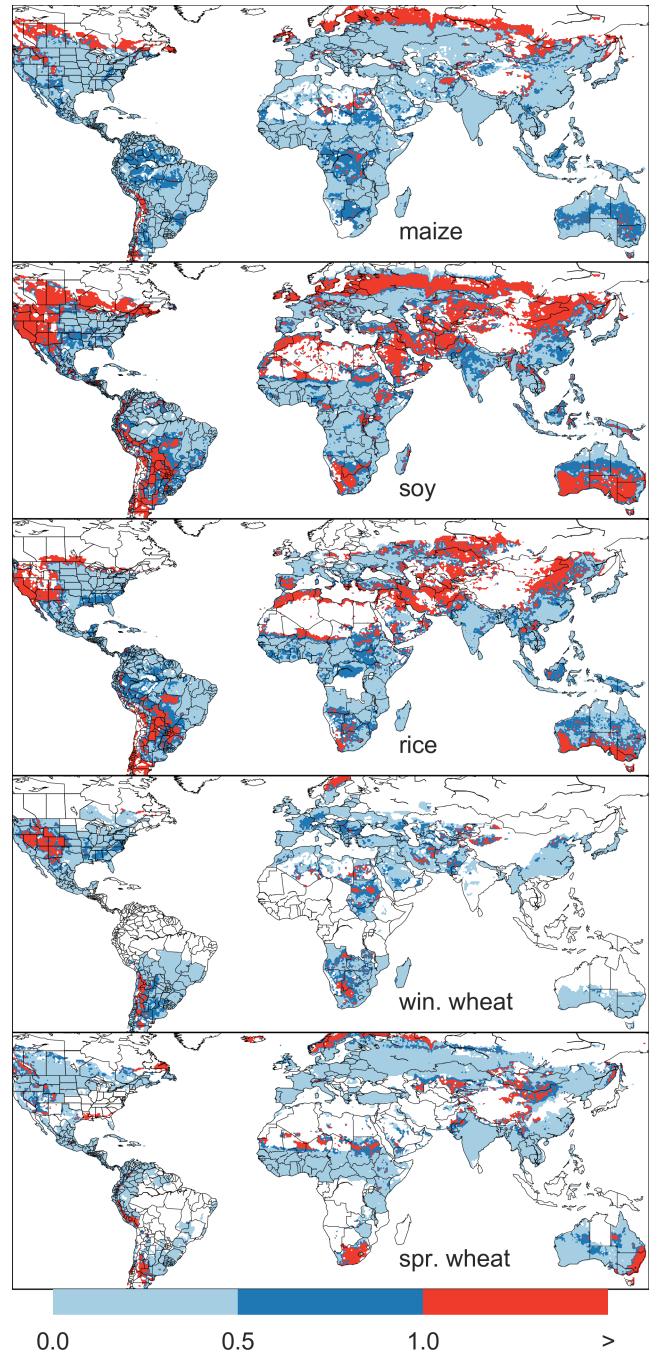


Figure 9: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error e , where $e > 1$ means that emulator error exceeds the multi-model standard deviation. White areas are those where crops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

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. Although most all model-crop combination emulators have normalized errors less than one over nearly all currently cultivated hectares (Figure 8), but some individual model-crop combinations are problematic (e.g. PROMET for rice and soy, JULES for soy and winter wheat, Figures S14–S15). 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 S16, lower-lying the denominator. Emulator performance often degrades in geographic locations where crops are not currently cultivated. Figure 9 shows a CARAIB case as an example, where emulator performance is satisfactory over cultivated areas for all crops other than soy, but uncultivated regions show some problematic areas.

It should be noted that this assessment metric is relatively forgiving. First, each emulation is evaluated against the simulation actually used to train the emulator. Had we used a spline interpolation the error would necessarily be zero. Second, the performance metric scales emulator fidelity not by the magnitude of yield changes but by the inter-model spread in those changes. Where models differ more widely, the standard for emulators becomes less stringent. Because models disagree on the magnitude of CO₂ fertilization, this effect is readily seen when comparing assessments of emulator performance in simulations at baseline CO₂ (Figure 8) with those at higher CO₂ levels (Figure S13). Widening the inter-model spread leads to an apparent increase in emulator skill.

3.4. Emulator applications

Because the emulator or “surrogate model” transforms the discrete simulation sample space into a continuous response surface at any geographic scale, it can be used for a variety

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

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 (CO₂, temperature, precipitation, and applied nitrogen) and identification of geographic shifts in high yield potential locations. While the richness of the dataset invites 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 CO₂ 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 ro-

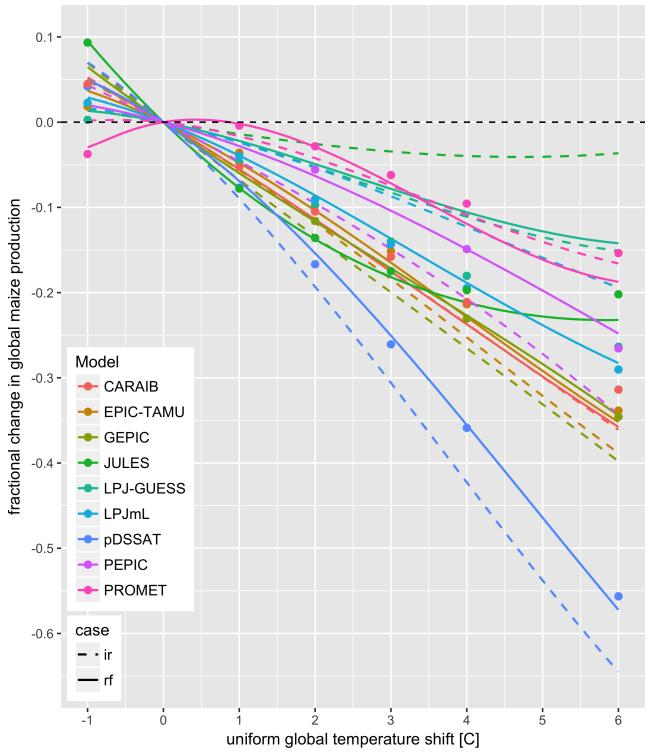


Figure 10: Global emulated damages for maize on currently cultivated lands⁶⁶⁸ for the GGCMI models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating up⁶⁶⁹ emulated values at the grid cell level, not from a regression of global mean yields.) Lines are emulations for rain-fed (solid) and irrigated (dashed) crops;⁶⁷⁰ 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 exceptions⁶⁷² are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures S16- S19 in the supplemental material.

bust in low latitudes where yield impacts are largest.

Model performance when compared to historical data is⁶⁷⁶ mixed, with models performing better for maize and soy than⁶⁷⁷ for rice and wheat. The value of utilizing multiple models is⁶⁷⁸ illustrated by the distribution in performance skill across differ-⁶⁷⁹ent countries and crops. An end-user of the simulation outputs⁶⁸⁰ or emulator tool may pick and choose models based on histori-⁶⁸¹cal skill to provide the most faithful temperature and precipita-⁶⁸²tion response depending on their application. The nitrogen and⁶⁸³ CO₂ responses were not validated in this work.

One counterintuitive result is that irrigated maize shows⁶⁸⁵ steeper yield reductions under warming than does rain-fed⁶⁸⁶ maize when considered only over currently cultivated land. The⁶⁸⁷

effect is the result of geographic differences in cultivated area.⁶⁵⁴ In any given location, irrigation increases crop resiliency to⁶⁵⁵ temperature increase, but irrigated maize is grown in warmer lo-⁶⁵⁶cations where the impacts of warming are more severe (Figures⁶⁵⁷ S5-S6). The same behavior holds for rice and winter wheat,⁶⁵⁸ but not for soy or spring wheat (Figures S8-S10). Irrigated⁶⁵⁹ wheat and maize are also more sensitive to nitrogen fertiliza-⁶⁶⁰tion levels, presumably because growth in rain-fed crops is also⁶⁶¹ water-limited (Figure S19). (Soy as a nitrogen-fixer is relatively⁶⁶² insensitive to nitrogen, and rice is not generally grown in water-⁶⁶³ limited conditions).

We show that emulation of the output of these complex re-⁶⁶⁵sponses is possible even with a relatively simple reduced-form⁶⁶⁶ statistical model and a limited library of simulations. Emula-⁶⁶⁷tion therefore offers the opportunity of producing rapid assess-⁶⁶⁸ments of agricultural impacts for arbitrary climate scenarios in⁶⁶⁹ a computationally non-intensive way. The resulting tool should⁶⁷⁰ aid in impacts assessment, economic studies, and uncertainty⁶⁷¹ analyses. Emulator parameter values also provide a useful way⁶⁷² to compare sensitivities across models to different climate and⁶⁷³ management inputs, and the terms in the polynomial fits offer⁶⁷⁴ the possibility of physical interpretation of these dependencies⁶⁷⁵ to some degree.

We provide this simulation output dataset for further analysis⁶⁷⁶ by the community as we have only scratched the surface with⁶⁷⁷ this work. Each simulation run includes year to year variabil-⁶⁷⁸ity in yields under different climate and management regimes.⁶⁷⁹ Some of the precipitation and temperature space has been lost⁶⁸⁰ due to the aggregation in the time dimension for the emula-⁶⁸¹tor presented here (i.e. the + 6 C simulation in the hottest year⁶⁸² of the historical period compared to the coldest historical year,⁶⁸³ or precipitation perturbations in the driest historical year etc).⁶⁸⁴ Development of a year-to-year emulator or an emulator at dif-⁶⁸⁵ferent spatial scales may provide useful for some IAM appli-⁶⁸⁶cations.

688 cations. More exhaustive analysis of different statistical model⁷²²
689 specification for emulation will likely provide additional pre-⁷²³
690 dictive skill over the specification provided here. The poten-⁷²⁴
691 tially richest area for further analysis is the interactions be-⁷²⁵
692 tween input variable especially the Nitrogen and CO₂ interac-⁷²⁶
693 tions with weather and with each other. More robust quantifica-⁷²⁷
694 tion of the sensitivity to the input drivers (and there differences⁷²⁸
695 across models), as well as quantification in differences in un-⁷²⁹
696 certainty across input drivers. Adaptation via growing season⁷³⁰
697 changes were also simulated and are available in the database,⁷³¹
698 though this dimension was not presented or analyzed here.⁷³²

699 The emulation approach presented here has some limitations.⁷³³
700 Because the GGCMI simulations apply uniform perturbations⁷³⁴
701 to historical climate inputs, they do not sample changes in⁷³⁵
702 higher order moments. The emulation therefore does not ad-⁷³⁶
703 dress the crop yield impacts of potential changes in climate⁷³⁷
704 variability. While some information could be extracted from⁷³⁸
705 consideration of year-over-year variability, more detailed sim-⁷³⁹
706 ulations and analysis are likely necessary to diagnose the im-⁷⁴⁰
707 pact of changes in variance and sub-growing-season tempo-⁷⁴¹
708 ral effects. Additionally, the emulator is intended to provide⁷⁴²
709 the change in yield from a historical mean baseline value and⁷⁴³
710 should be used in conjunction with historical data (or data prod-⁷⁴⁴
711 ucts) or a historical mean emulator (not presented here).⁷⁴⁵

712 The future of food security is one of the larger challenges⁷⁴⁵
713 facing humanity at present. The development (and emulation)⁷⁴⁶
714 of multi-model ensembles such as GGCMI Phase II provides⁷⁴⁷
715 a way to begin to quantify uncertainties in crop responses to⁷⁴⁸
716 a range of potential climate inputs and explore the potential⁷⁴⁹
717 benefits of adaptive responses. Emulation also allow making⁷⁵⁰
718 state-of-the-art simulation results available to a wide research⁷⁵²
719 community as simple, computationally tractable tools that can⁷⁵³
720 be used by downstream modelers to understand the socioeco-⁷⁵⁵
721 nomic impacts of crop response to climate change.⁷⁵⁶

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 (CSSP Brazil). A.S. was supported by the Office of Science of the U.S. Department of Energy as part of the Multi-sector Dynamics Research Program Area. Computing resources were provided by the University of Chicago Research Computing Center (RCC). S.O. acknowledges support from the Swedish strong research areas BECC and MERGE together with support from LUCCI (Lund University Centre for studies of Carbon Cycle and Climate Interactions).

6. References

- Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013). Implication of crop model calibration strategies for assessing regional impacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.
- Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D., A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Alderman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Biermann, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising temperatures reduce global wheat production. *Nature Climate Change*, 5, 143–147. doi:10.1038/nclimate2470.
- Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N.,

- 757 Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath,⁸⁰⁰
 758 C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, *3*, 827–832.⁸⁰¹
 759 doi:10.1038/nclimate1916.⁸⁰²
- 760 Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Efficiency, Carbon Sequestration, and Environmental Pollution. *Advances in Agronomy*, *86*, 341 – 409.⁸⁰³
- 761 Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov,⁸⁰⁴
 762 N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat production potentials and management flexibility under the representative concentration pathways. *Global and Planetary Change*, *122*, 107 – 121.⁸⁰⁵
- 763 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models. *Agricultural and Forest Meteorology*, *236*, 145 – 161.⁸⁰⁶
- 764 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded crop models using statistical estimates. *Agricultural and Forest Meteorology*, *214-215*, 134 – 147.⁸⁰⁷
- 765 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S. (2018). Implementing the Nitrogen cycle into the dynamic global vegetation, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific Model Development*, *11*, 2789–2812.⁸⁰⁸
- 766 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*, 1829–1844.⁸⁰⁹
- 767 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 – 291.⁸¹⁰
- 768 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process emulation of dynamic computer codes. *Biometrika*, *96*, 663–676.⁸¹¹
- 769 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.⁸¹²
- 770 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).⁸¹³
- 771 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).⁸¹⁴
- 772 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*, 509–516.⁸¹⁵
- 773 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Izumi, T., Izaurrealde, 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.
- 774 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.
- 775 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.
- 776 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.
- 777 Food and Agriculture Organization of the United Nations (2018). FAOSTAT database. URL: <http://www.fao.org/faostat/en/home>.
- 778 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J., Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna, M., Semenov, M. A., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model wheat yield impact response surfaces showing sensitivity to temperature and precipitation change. *Agricultural Systems*, *159*, 209–224.
- 779 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014). Evaluating the utility of dynamical downscaling in agricultural impacts projections. *Proceedings of the National Academy of Sciences*, *111*, 8776–8781.
- 780 Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitivity of Agricultural Model Performance to Different Climate Inputs. *Journal of Applied Meteorology and Climatology*, *55*, 151113145618001.
- 781 Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous Crop Growth and Yield: Application for Wheat in Central Europe. *Remote Sensing*, *7*, 3934–3965.
- 782 He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H., He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water

- contents and nitrogen leaching to precipitation, management practices and soil hydraulic properties in semi-arid and humid regions of Canada using the DSSAT model. *Nutrient Cycling in Agroecosystems*, 106, 201–215.

Heady, E. O. (1957). An Econometric Investigation of the Technology of Agricultural Production Functions. *Econometrica*, 25, 249–268.

Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa State University Press.

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

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

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

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).

Iizumi, 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.

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

Izaurralde, 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.

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.

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

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.

Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzwirth, 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.

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.

Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIG - modelling wheat yield and crop water productivity with high resolution on a global scale. *Agricultural Systems*, 94, 478 – 493.

Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a). Global investigation of impacts of PET methods on simulating crop-water relations for maize. *Agricultural and Forest Meteorology*, 221, 164 – 175.

Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs with yields from major crop cultivations. *Science of The Total Environment*, 572, 526 – 537.

Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, 150, 1443 – 1452.

Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2, 014002.

MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447.

Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P., Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi, P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano, D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of large simulated yield datasets for studying climate effects. (p. 1100). doi:10.13140/RG.2.1.5173.8328.

Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrological modelling to study the impact of climate change on the water flows of mountain watersheds. *Journal of Hydrology*, 376, 362 – 377.

Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., & Calzadilla, A. (2009). Global biomass production potentials exceed ex-

- 929 pected future demand without the need for cropland expansion. *Nature Communications*, 6.
- 930 972
- 931 McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srinivas, C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S. 975
- 932 (2015). Integrated assessments of the impacts of climate change on agriculture: An overview of AgMIP regional research in South Asia. *Chapter in: Handbook of Climate Change and Agroecosystems*, (pp. 201–218).
- 933 978
- 934 Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical weather responsiveness of crop yields: US evidence and implications for the agricultural impacts of climate change. *Environmental Research Letters*, 12.
- 935 981
- 936 Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, 8.
- 937 983
- 938 Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. 989
- 939 (2017). Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, 10, 1403–1422.
- 940 991
- 941 Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T. 994
- 942 (1997). Effect of CO₂ enrichment on carbon and nitrogen interaction in wheat and soybean. *Soil Science and Plant Nutrition*, 43, 789–798.
- 943 995
- 944 O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91, 1290 – 1300.
- 945 998
- 946 Olin, S., Schurges, 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 CO₂ and N management in the main wheat regions of western europe. *Biogeosciences*, 12, 2489–2515. doi:10.5194/bg-12-2489-2015.
- 947 1001
- 948 Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.
- 949 1005
- 950 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.
- 951 1008
- 952 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.
- 953 1011
- 954 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.
- 955 1014
- 956 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.
- 957 1018
- 958 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.
- 959 1022
- 960 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.
- 961 1025
- 962 Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of monthly growing areas of 26 irrigated crops.
- 963 1028
- 964 Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly Irrigated and Rainfed crop Areas around the Year 2000: A New High-Resolution Data Set for Agricultural and Hydrological Modeling. *Global Biogeochemical Cycles*, 24, GB1011.
- 965 1031
- 966 Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E., & Arneth, A. (2016). Climate analogues suggest limited potential for intensification of production on current croplands under climate change. *Nature Communications*, 7, 12608.
- 967 1034
- 968 Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus A: Dynamic Meteorology and Oceanography*, 58, 461–472.
- 969 1037
- 970 Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the reduction and sensitivity analysis of complex environmental models. *Environmental Modelling & Software*, 34, 1 – 4.
- 971 1040
- 972 Ray, D., Ramankutty, N., Mueller, N., West, P., & Foley, J. (2012). Recent patterns of crop yield growth and stagnation. *Nature communications*, 3, 1293. doi:10.1038/ncomms2296.
- 973 1043
- 974 Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling

- in water resources. *Water Resources Research*, 48. 1058
- Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017) 1059
Comparing and combining process-based crop models and statistical models 1060
with some implications for climate change. *Environmental Research Letters* 1061
12. 1062
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A. 1063
Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek 1064
F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014) 1065
Assessing agricultural risks of climate change in the 21st century in a global 1066
gridded crop model intercomparison. *Proceedings of the National Academy 1067
of Sciences*, 111, 3268–3273. 1068
- Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P. 1069
Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ew 1070
ert, F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural 1071
Model Intercomparison and Improvement Project (AgMIP): Protocols and 1072
pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182. 1073
- Rosenzweig, C., Ruane, A. C., Antle, J., Elliott, J., Ashfaq, M., Chatta 1074
A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G. 1075
Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M. 1076
Müller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M. 1077
Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., & Wiebe, K. 1078
(2018). Coordinating AgMIP data and models across global and regional 1079
scales for 1.5°C and 2.0°C assessments. *Philosophical Transactions of the 1080
Royal Society of London A: Mathematical, Physical and Engineering Sci 1081
ences*, 376. 1082
- Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason 1083
D'Croz, D., Müller, C., Porter, C., Phillips, M. M., Raymundo, R. M. 1084
Sands, R., Valdivia, R. O., White, J. W., Wiebe, K., & Rosenzweig, C. 1085
(2018). Biophysical and economic implications for agriculture of +1.5° and 1086
+2.0°C global warming using AgMIP Coordinated Global and Regional As 1087
sessments. *Climate Research*, 76, 17–39. 1088
- Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown 1089
D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013) 1090
Climate change impact uncertainties for maize in panama: Farm informa 1091
tion, climate projections, and yield sensitivities. *Agricultural and Forest 1092
Meteorology*, 170, 132 – 145. 1093
- Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forc 1094
ing datasets for agricultural modeling: Merged products for gap-filling and 1095
historical climate series estimation. *Agric. Forest Meteorol.*, 200, 233–248. 1096
- Ruane, A. C., McDermid, S., Rosenzweig, C., Baigorria, G. A., Jones, J. W. 1097
Romero, C. C., & Cecil, L. D. (2014). Carbon-temperature-water change 1098
analysis for peanut production under climate change: A prototype for the 1099
agmip coordinated climate-crop modeling project (c3mp). *Glob. Change* 1100
Biol., 20, 394–407. doi:10.1111/gcb.12412.
- Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901– 1058
2100 depicted by world maps of the Köppen-Geiger climate classification. 1059
Meteorologische Zeitschrift, 19, 135–141.
- Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting 1060
dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19, 1061
607–620.
- Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, 1062
D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolin- 1063
ski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frieler, K. 1064
(2017). Consistent negative response of US crops to high temperatures in 1065
observations and crop models. *Nature Communications*, 8, 13931.
- Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate 1066
severe damages to U.S. crop yields under climate change. *Proceedings of 1067
the National Academy of Sciences*, 106, 15594–15598.
- Snyder, A., Calvin, K. V., Phillips, M., & Ruane, A. C. (2018). A crop yield 1068
change emulator for use in gcam and similar models: Persephone v1.0. *Geo- 1069
scientific Model Development Discussions*, 2018, 1–42.
- Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implemen- 1070
tation and evaluation of nonparametric regression procedures for sensitivity 1071
analysis of computationally demanding models. *Reliability Engineering & 1072
System Safety*, 94, 1735 – 1763.
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 1073
and the experiment design. *Bulletin of the American Meteorological Society*, 1074
93, 485–498.
- Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of cli- 1075
mate change impacts on global crop yields. *Geophysical Research Letters*, 1076
35.
- Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., & 1077
Martin, J. F. (2014). Modeling sugarcane yield with a process-based model 1078
from site to continental scale: Uncertainties arising from model structure 1079
and parameter values. *Geoscientific Model Development*, 7, 1225–1245.
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, 1080
J. (2014). The Inter-Sectoral Impact Model Intercomparison Project 1081
(ISI-MIP): Project framework. *Proceedings of the National Academy of 1082
Sciences*, 111, 3228–3232.
- White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Methodo- 1083
logies for simulating impacts of climate change on crop production. *Field 1084
Crops Research*, 124, 357 – 368.
- Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T., 1085
Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance 1086
against site observations of irrigated maize from Mead, Nebraska. *Geosci- 1087
entific Model Development*, 10, 1291–1320.

- 1101 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield vari-
1102 ability in JULES-crop and implications for forcing with seasonal weather
1103 forecasts. *Geoscientific Model Development*, 8, 3987–3997.
- 1104 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Land-*
1105 *bouwkundige Onderzoeken* : 64.6, .
- 1106 Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato
1107 yields on changes in climate and co2. *Agricultural and Forest Meteorology*,
1108 112, 217 – 231.
- 1109 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao,
1110 Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A.,
1111 Li, T., Lin, E., Liu, Q., Martre, P., Miller, C., Peng, S., Peuelas, J., Ruane,
1112 A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng,
1113 S. (2017). Temperature increase reduces global yields of major crops in four
1114 independent estimates. *Proc. Natl. Acad. Sci.*, 114, 9326–9331.