

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

^rEAWAG, 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 affecting yields and considerable uncertainty remains in future yield projections. Phase II of the Global Gridded Crop Model Intercomparison (GGCMI), an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP), consists of a large simulation set with perturbations in atmospheric CO₂ concentrations, temperature, precipitation, and nitrogen inputs and 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 for twelve different process-based models. In this paper we present the simulation output database from Phase II of the GGCMI effort, a targeted experiment aimed at understanding the sensitivity to and interaction between multiple climate variables (as well as management) on yields, and illustrate some initial summary results from the model intercomparison project. We also present the construction of a simple “emulator or statistical representation of the simulated 30-year mean climatological output in each location for each crop and model. The emulator captures the response of the process-based models in a lightweight, computationally tractable form that facilitates model comparison as well as potential applications in subsequent modeling efforts such as integrated assessment.

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

1. Introduction

2 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)). For a
25 history of crop model development see the appendix of Rosen-
26 zweig et al. (2014). A half-century of improvement in both
27 models and computing resources means that researchers can
28 now run crop simulation models for many years at high spatial
29 resolution on the global scale.

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

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

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

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

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

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

74 The overall goal of this study is a better understanding of
 75 global crop model response to the major drivers in a climate
 76 change context. Most previous climate-change-focused global
 77 crop modeling studies have simulated model response to rep-
 78 resentative concentration pathways (RCPs). RCPs are likely to
 79 have strong covariance between precipitation, temperature and
 80 CO₂ that may be hard to decompose statistically. The differ-
 81 ences in year-to-year memory in the models and complexity of
 82 the changes in year-to-year distributions in weather under RCP
 83 scenarios in climate models are complications we seek to con-
 84 trol for with this study. We propose to test the response to major
 85 drivers and their interaction by isolating individual input drivers
 86 through simulations on first-moment shifts applied to the histor-
 87 ical climatology instead of RCP simulations. As emulators are
 88 a fundamentally a distillation of the process-based model down
 89 to its major drivers, the same applied to their development.

90 Statistical emulation of crop simulations has been used to
 91 combine advantageous features of both statistical and process-
 92 based models. The statistical representation of complicated nu-
 93 matical simulation (e.g. O'Hagan, 2006, Conti et al., 2009), in
 94 which simulation output acts as the training data for a statisti-
 95 cal model, has been of increasing interest with the growth of¹⁰¹
 96 simulation complexity and volume of output. Such emulators¹⁰²
 97 or "surrogate models" have been used in a variety of fields in¹⁰³
 98 cluding hydrology (e.g. Razavi et al., 2012), engineering (e.g.¹⁰⁴
 99 Storlie et al., 2009), environmental sciences (e.g. Ratto et al.,¹⁰⁵
 100 2012), and climate (e.g. Castruccio et al., 2014, Holden et al.,¹⁰⁶

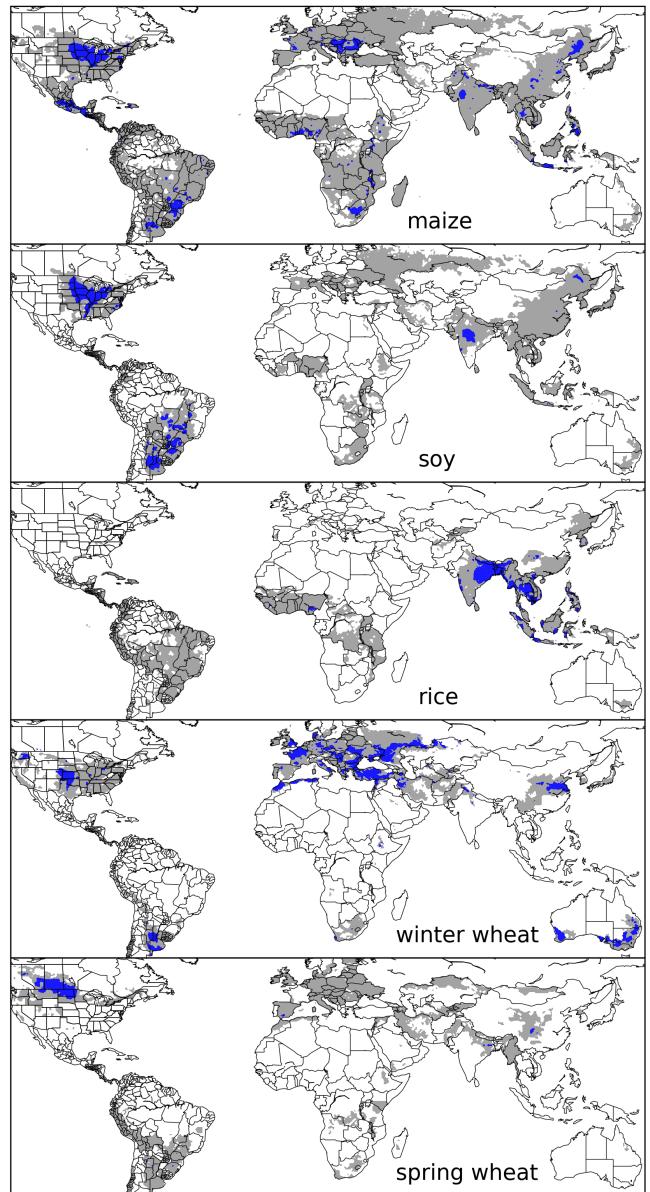


Figure 1: Presently cultivated area for rain-fed crops in the real world. Blue contour area indicates grid cells with more than 20,000 hectares (approx. 10% of the land in a grid cell near the equator for example) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated 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.

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 com-
 patible with economic studies.

In the past decade, many studies have developed emulators of

107 crop yields from process-based models. Early studies proposing or describing potential emulators include Howden & Crimp (2005), Räisänen & Ruokolainen (2006) and Lobell & Burke (2010). In an early application, Ferrise et al. (2011) used a Artificial Neural Net trained on simulation outputs to predict wheat yields in the Mediterranean. Studies developing single-model emulators include Holzkämper et al. (2012) for the CropSyst model, Ruane et al. (2013) for the CERES wheat model, Oyebamiji et al. (2015) for the LPJmL model (for multiple crops, using multiple scenarios as a training set). In recent years, emulators have begun to be used in the context of multi-model inter-comparisons, 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 & Sultan (2015) and Blanc (2017) used local weather variables (and CO₂ values) and yields but emulate across soil types using historical simulations and a future climate scenario (RCP8.5 over multiple climate models); Ostberg et al. (2018) used global mean temperature change (and CO₂) as regressors but pattern-scale to emulate local yields using multiple climate scenarios; Mistry et al. (2017) used local weather and yields and a historical simulation and compare with data.

131 Recently efforts have been made to generate datasets that allow more systematic sampling of the input variable space (the focus of this study): Makowski et al. (2015) for temperature, CO₂, and nitrogen, Pirttioja et al. (2015) and Snyder et al.

132 (2018) for temperature, water, and CO₂, and (Fronzek et al., 2018) for temperature and water, with all studies simulating selected sites for a limited number of crops.

133 The use of limited input parameter space or restricted geographic scope may impede the ability to build future projections or to understand interaction effects in global process-based crop models. 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 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).

134 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

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.

163 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.

169 2. Materials and Methods

170 2.1. GGCMI Phase II: experiment design

171 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

176 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).

184 The major goals of GGCMI Phase II are to:

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

The guiding scientific rationale of GGCMI Phase II is to provide a comprehensive, systematic evaluation of the response of process-based crop models to different values for carbon dioxide, temperature, water, and applied nitrogen (collectively known as "CTWN"). Phase II of the GGCMI project consists of a series of simulations, each with one or more of the CTWN dimensions perturbed over the 31-year historical time series (1980-2010) used in Phase I. In most cases, historical daily climate inputs are taken from the 0.5 degree NASA AgMERRA daily gridded re-analysis product specifically designed for agricultural modeling, with satellite-corrected precipitation (Ruane et al., 2015). Two models require sub-daily input data and use alternative sources. See Elliott et al. (2015) for additional details.

The experimental protocol consists of 9 levels for precipitation perturbations, 7 for temperature, 4 for CO₂, and 3 for applied nitrogen, for a total of 672 simulations for rain-fed agriculture and an additional 84 for irrigated (Table 1). For irrigated simulations, soil water is held at either field capacity or, for those models that include water-log damage, at maximum beneficial level. Temperature perturbations are applied as absolute offsets from the daily mean, minimum, and maximum temperature time series for each grid cell used as inputs. Precipitation perturbations are applied as fractional changes at the grid cell level, and carbon dioxide and nitrogen levels are specified as discrete values applied uniformly over all grid cells. Note that 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

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.

229 climate conditions. 247 release additional nitrogen through mineralization. See Rosen-
 230 The 12 models included in GGCMI Phase II are all mecha- 248 zweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017)
 231 nistic process-based crop models that are widely used in im- 249 for further details on models and underlying assumptions.
 232 pacts assessments (Table 2). Although some of the models 250
 233 shares a common base (e.g. LPJmL and LPJ-GUESS and the 251
 234 EPIC models), they have developed independently from this 252
 235 shared base, for more details on the genealogy of the mod- 253
 236 els see Figure S1 in Rosenzweig et al. (2014). Differences in 254
 237 model structure does mean that several key factors are not stan- 255
 238 dardized across the experiment, including secondary soil nutri- 256
 239 ents, carry over effects across growing years including residue 257
 240 management and soil moisture, and extent of simulated area for 258
 241 different crops. Growing seasons are identical across models, 259
 242 but vary by crop and by location on the globe. All stresses 260
 243 except factors related to nitrogen, temperature, and water (e.g. 261
 244 Alkalinity, salinity) are disabled. No additional nitrogen inputs, 262
 245 such as atmospheric deposition, are considered, but some mod- 263
 246 els have individual assumptions on soil organic matter that may 264
 247 The participating modeling groups provide simulations at
 248 any of four initially specified levels of participation, so the num-
 249

ber of simulations varies by model, with some sampling only a₂₉₉ part of the experiment variable space. Most modeling groups₃₀₀ simulate all five crops in the protocol, but some omitted one₃₀₁ or more. Table 2 provides details of coverage for each model.₃₀₂

Note that the three models that provide less than 50 simulations₂₆₉
are excluded from the emulator analysis.₃₀₃

All models produce as output, crop yields (tons ha⁻¹ year⁻¹)
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⁻¹. We show absolute yields in some cases to illustrate geo-
graphic differences in yields for a single model.

2.2. Simulation model validation approach

To verify the skill of the process-based models used, we re-
peat the validation exercises presented in Müller et al. (2017)
for GGCMI Phase I. Note however that the GGCMI Phase II
simulations are designed for evaluating changes in yield but not
absolute yields, and so omit the calibrations used in predict-
ing modeling to account for cultivar, pest loss, and manage-
ment differences. The Phase II simulations also do not repro-
duce realistic nitrogen application levels for individual coun-₃₀₄
tries, since nitrogen is one of the parameters systematically var-₃₀₅
ied. The Müller et al. (2017) validation procedure evaluates re-₃₀₆
sponse to year-to-year temperature and precipitation variations₃₀₇
in a control run driven by historical climate and compares it₃₀₈
to detrended historical yields from the FAO (Food and Agri-₃₀₉
culture Organization of the United Nations, 2018) by calculat-₃₁₀
ing the Pearson correlation coefficient. The procedure offers no₃₁₁
means of assessing CO₂ fertilization, since CO₂ has been rel-₃₁₂
atively constant over the historical data collection period. Ni-₃₁₃
trogen data are limited for many countries, and as mentioned₃₁₄
the GGCMI Phase II runs impose fixed and uniform nitrogen₃₁₅

application, introducing some uncertainty into the analysis. We evaluate one or more control runs for each model, since some modeling groups provide historical runs for three different nitrogen levels.

2.3. Climatological-mean yield emulator design

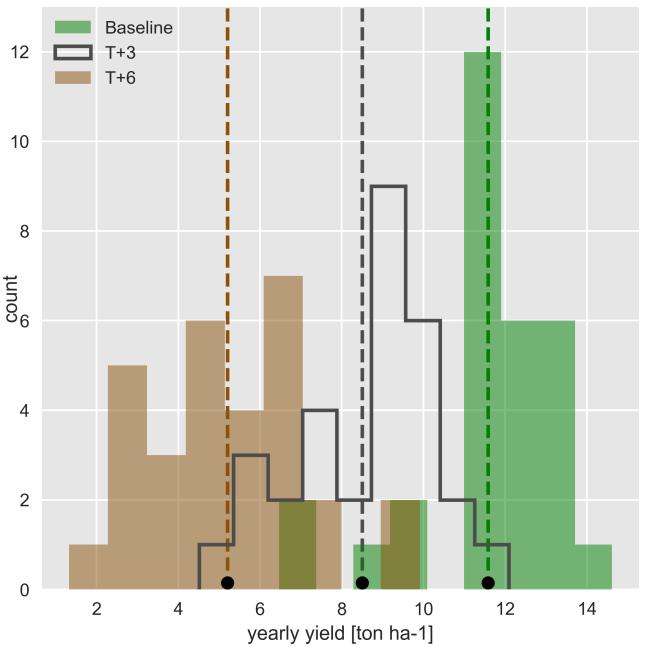


Figure 2: Example showing both climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for a grid cell in northern Iowa (a representative high-yield region) from the pDSST model, for the baseline climatology (1981–2010) and for scenarios with 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.

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

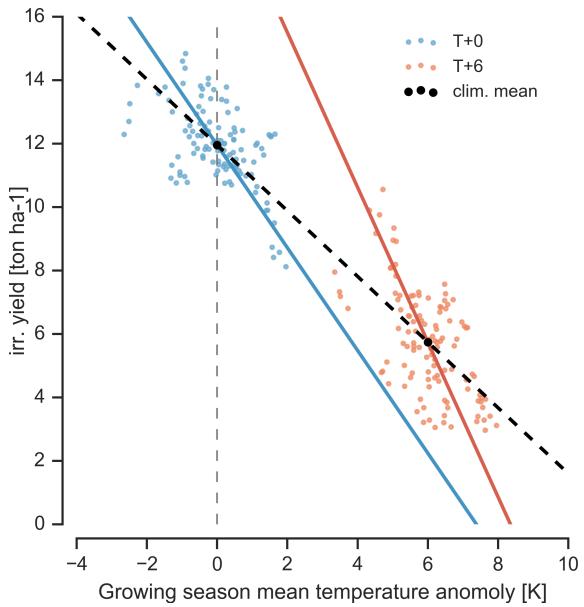


Figure 3: Example showing simple temperature relationship developed from year-to-year values vs. climatological mean values. Figure shows irrigated maize for four adjacent grid cell in northern Iowa (a representative high-yield region) from the pdSSAT model, for the baseline climatology (1981-2010) and for scenarios with temperature shifted (T) $+6^{\circ}\text{C}$, with other variables held at baseline values. Irrigated yields are shown to control for precipitation effects. Blue and red lines indicate total least squares linear regression across each temperature scenario. Black dots indicate the climatological mean yield values for each climatological temperature scenario.

316 water, nitrogen and CO_2 , we aggregate the simulation outputs
 317 in the time dimension, and regress on the 30-year mean yields.
 318 (See Figure 2 for illustration). The regression therefore omits
 319 information about yield responses to year-to-year climate per-
 320 turbations, which are more complex. Emulating inter-annual
 321 yield variations would likely require considering statistical de-
 322 tails of the historical climate time series, including changes in
 323 marginal distribution and temporal dependencies. (Future work
 324 should explore this). The climatological emulation indirectly
 325 includes any yield response to geographically distributed fac-
 326 tors such as soil type, insolation, and the baseline climate itself,
 327 because we construct separate emulators for each grid cell. The
 328 emulator parameter matrices are portable and the yield compu-
 329 tations are cheap even at the half-degree grid cell resolution, so
 330 we do not aggregate in space at this time.

333 dard polynomial for representing simulations at the yearly level
 334 across different soil types geographically (not at the grid cell
 335 level). We do not test this specification here, and instead use as
 336 a starting point a standard third-order polynomial to represent
 337 the climatological-mean response at the grid cell level as it is
 338 the simplest effective specification. We regress climatolog-
 339 ical-mean yields against a third-order polynomial in C , T , W , and N
 340 with interaction terms. The higher-order terms are necessary to
 341 capture any nonlinear responses, which are well-documented
 342 in observations for temperature and water perturbations (e.g.
 343 Schlenker & Roberts (2009) for T and He et al. (2016) for W).
 344 We include interaction terms (both linear and higher-order) be-
 cause 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 stud-
 ies have shown the importance of the interaction between water
 and nitrogen (e.g. Aulakh & Malhi, 2005), and between nitro-
 gen and carbon dioxide (Osaki et al., 1992, Nakamura et al.,
 1997). We do not focus on comparing different model spec-
 345 ifications in this study, and instead stick to a relatively simple
 346 parameterized specification that allows for some, albeit limited,
 347 coefficient interpretation.

The limited GGCMI variable sample space means that use
 of the full polynomial expression described above, which has
 34 terms for the rain-fed case (12 for irrigated), can be prob-
 35 lematic, and can lead to over-fitting and unstable parameter es-
 36 timations. We therefore reduce the number of terms through a
 feature selection cross-validation process in which terms in the
 polynomial are tested for importance. In this procedure higher-
 order and interaction terms are added successively to the model;
 we then follow the reduction of the the aggregate mean squared
 error with increasing terms and eliminate those terms that do

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

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

in temperature. The resulting statistical model (Equation 1) is used for all grid cells, models, and crops:

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

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

To fit the parameters K , we use a Bayesian Ridge probabilistic estimator (MacKay, 1991), which reduces volatility in parameter estimates when the sampling is sparse, by weighting parameter estimates towards zero. The Bayesian Ridge method is necessary to maintain a consistent functional form across all

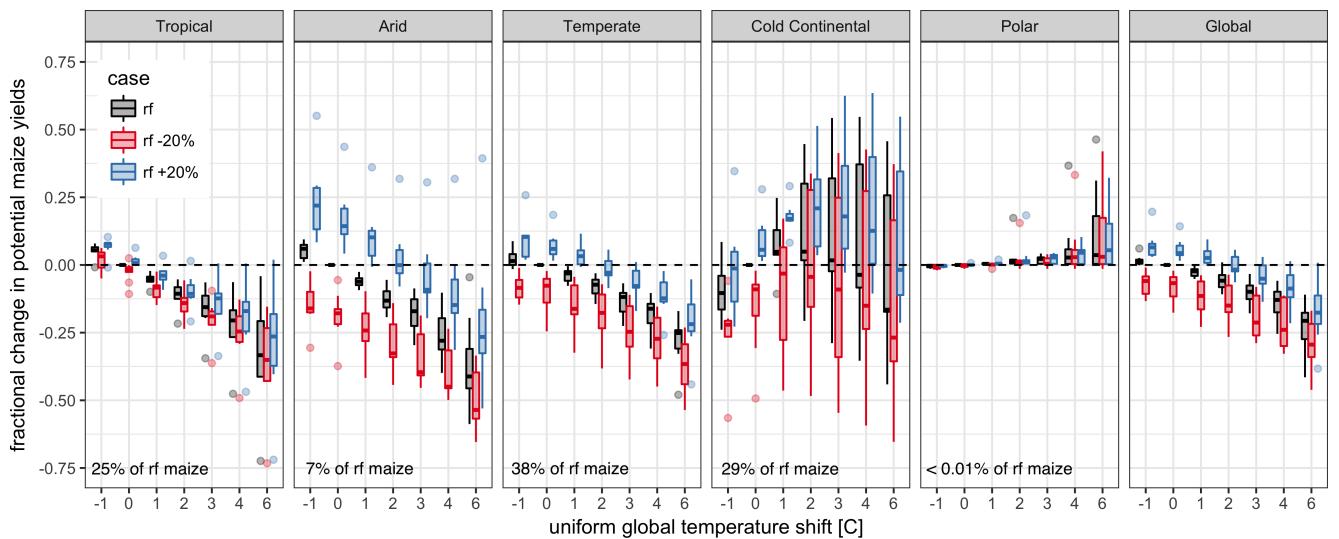


Figure 4: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation 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.

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 over two orders of magnitude.

2.4. Emulator evaluation

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

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

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

Here $F_{scn.}$ is the fractional change in a model’s mean emulated or simulated yield from a defined baseline, in some scenario (scn.) in C, T, W, and N space; $Y_{scn.}$ and $Y_{baseline}$ are the absolute emulated or simulated mean yields. The normalized error e is the difference between the emulated fractional change in yield and that actually simulated, normalized by $\sigma_{sim.}$, the standard deviation in simulated fractional yields $F_{sim, scn.}$ across all models. The emulator is fit across all available simulation outputs, and then the error is calculated across the simulation scenarios provided by all nine models (Figure 9 and Figures S12 and Figures S13 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.

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 4, 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

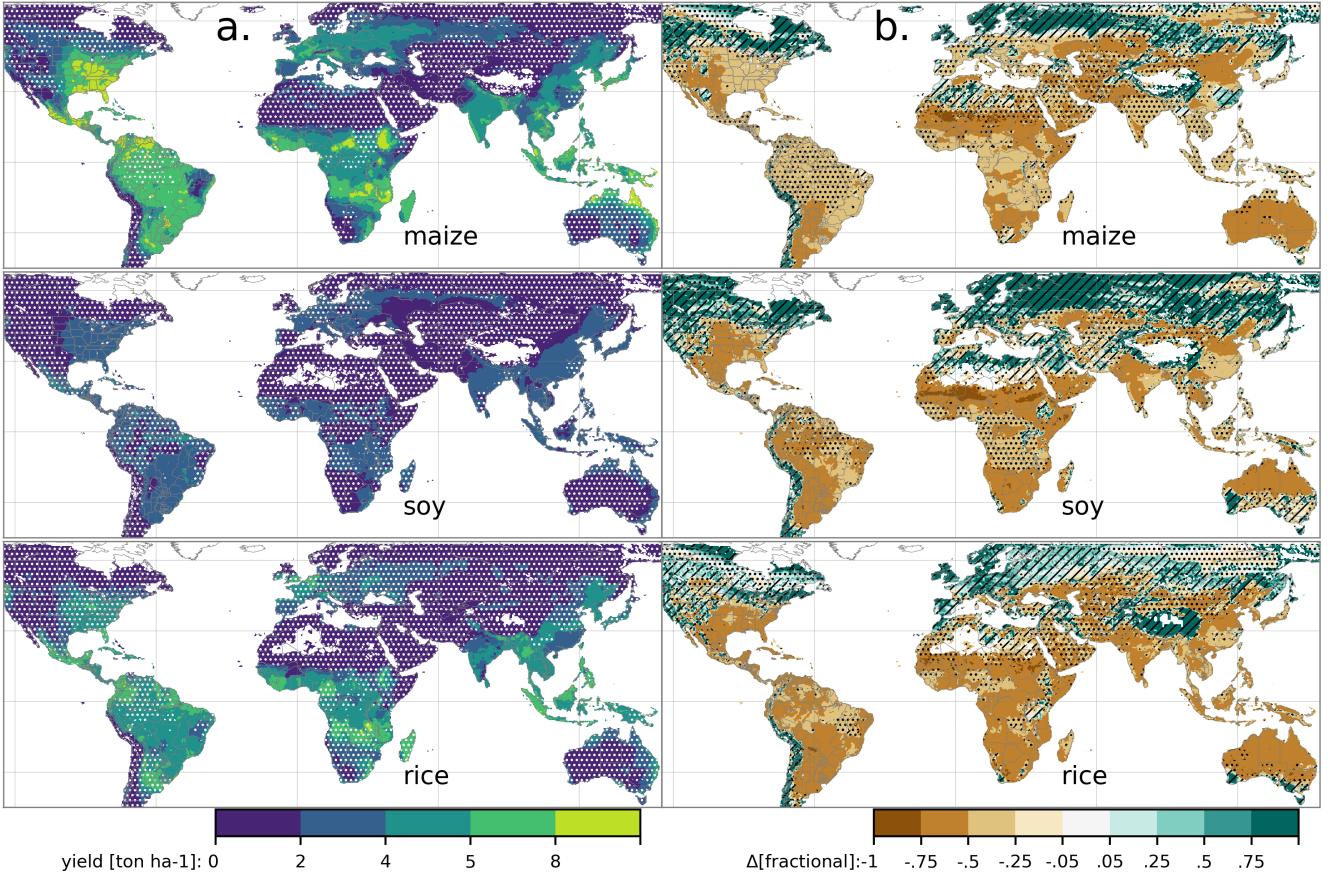


Figure 5: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure S11 in the supplemental material.) White stippling indicates areas where these crops are not currently cultivated. Absence of cultivation aligns well with the lowest yield contour (0.2 ton 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.

452 three. A notable exception is the cold continental region, where₄₆₅
 453 models disagree strongly, extending even to the sign of impacts.₄₆₆
 454 Model simulations of other crops produce similar responses to₄₆₇
 455 warming, with robust yield losses in warmer locations and high₄₆₈
 456 inter-model variance in the cold continental regions (Figures₄₆₉
 457 S7).

458 The effects of rainfall changes on maize yields are also as ex-₄₇₁
 459 pected and are consistent across models. Increased rainfall mit-₄₇₂
 460 igates the negative effect of higher temperatures, most strongly₄₇₃
 461 in arid regions. Decreased rainfall amplifies yield losses and₄₇₄
 462 also increases inter-model variance more strongly, suggesting₄₇₅
 463 that models have difficulty representing crop response to water₄₇₆
 464 stress. We show only rain-fed maize here; see Figure S5 for the₄₇₇

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 5 shows baseline and changes in the $T+4$ scenario for rain-fed maize, soy, and rice in the multi-model ensemble mean, with locations of model agreement marked. Absolute yield potentials 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 currently grown. Models show robust decreases in yields at low

478 latitudes, and highly uncertain median increases at most high₅₁₂
479 latitudes. For wheat crops see Figure S11; wheat projections₅₁₃
480 are both more uncertain and show fewer areas of increased yield₅₁₄
481 in the inter-model mean.

515 years for US maize along with 2012 (not shown). US maize
is possibly both the most uniformly industrialized (in terms of
management practices) crop and the one with the best data col-
lection in the historical period of all the cases presented here.

482 3.2. *Simulation model validation results*

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

500 *** or harmonization *** Christoph

501 Soy is qualitatively the easiest crop to represent (except in₅₃₅
502 Argentina), which is likely due in part to the invariance of the₅₃₆
503 response to nitrogen application (soy fixes atmospheric nitrogen₅₃₇
504 very efficiently). Comparison to the FAO data is therefore easier₅₃₈
505 than the other crops because the nitrogen application levels do₅₃₉
506 not matter. US maize has the best performance across models,₅₄₀
507 with nearly every model representing the historical variability₅₄₁
508 to a reasonable extent. Especially good example years for US₅₄₂
509 maize are 1983, 1988, and 2004 (top left panel of Figure 6),₅₄₃
510 where every model gets the direction of the anomaly compared₅₄₄
511 to surrounding years correct. 1983 and 1988 are famously bad₅₄₅

516 The FAO data is at least one level of abstraction from ground
517 truth in many cases, especially in developing countries. The
518 failure of models to represent the year-to-year variability in rice
519 in some countries in southeast Asia is likely partly due to model
520 failure and partly due to lack of data. It is possible to speculate
521 that the difference in performance between Pakistan (no suc-
522 cessful models) and India (many successful models) for rice
523 may reside at least in part in the FAO data and not the mod-
524 els themselves. The same might apply to Bangladesh and In-
525 dia for rice. Partitioning of these contributions is impossible at
526 this stage. Additionally, there is less year-to-year variability in
527 rice yields (partially due to the fraction of irrigated cultivation).
528 Since the Pearson r metric is scale invariant, it will tend to score
529 the rice models more poorly than maize and soy. An example
530 of very poor performance can be seen with the pDSSAT model
531 for rice in India (top right panel of Figure 6).

533 3.3. *Emulator performance*

534 Emulation provides not only a computational tool but a
means of understanding and interpreting crop yield response
across the parameter space. Emulation is only possible, how-
ever, when crop yield responses are sufficiently smooth and
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 reg-
ular enough to permit emulation. Figure 7 illustrates the geo-
graphic 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
of emulating at the grid cell level.

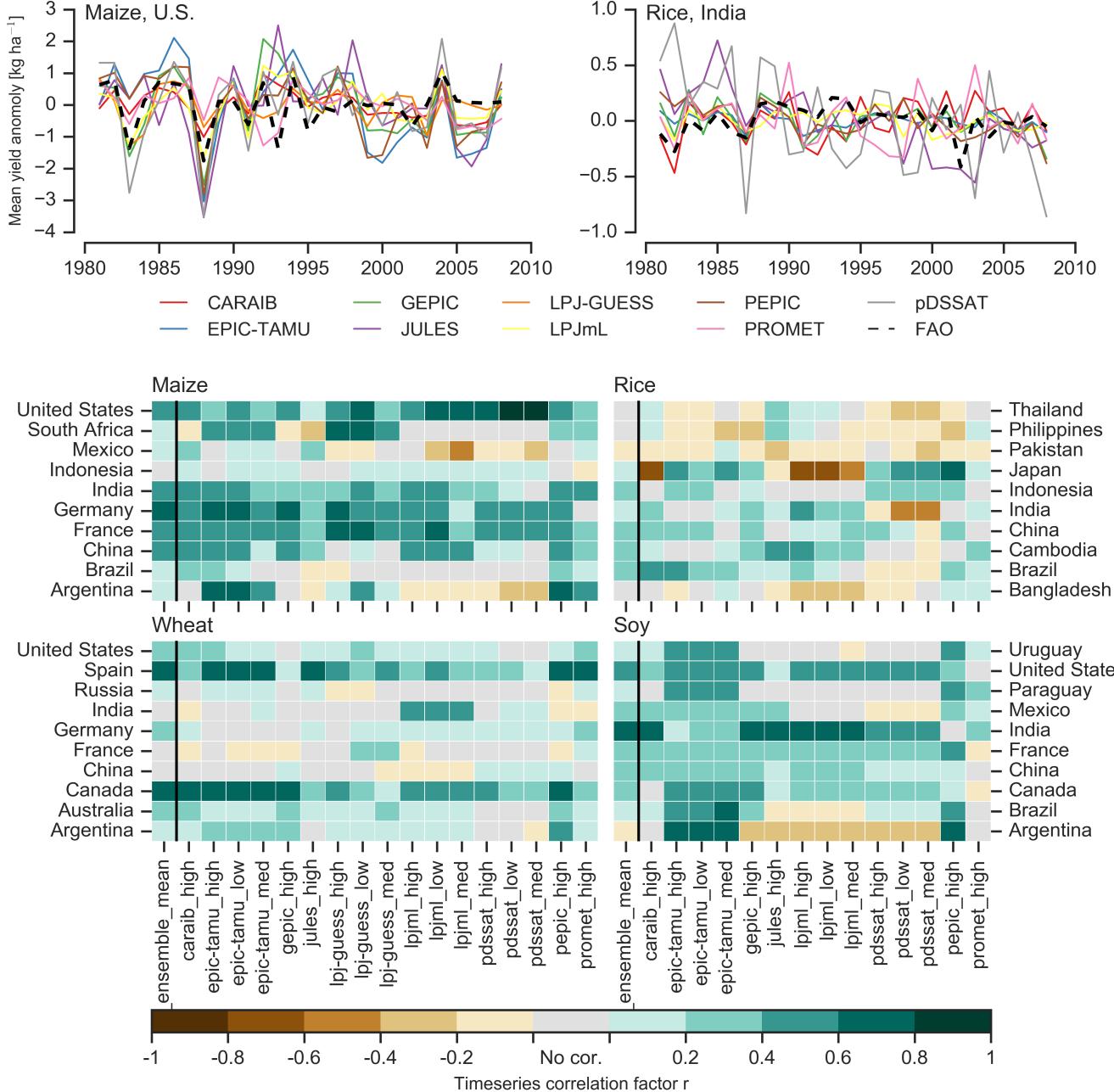


Figure 6: Time series correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen 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.

546 Each panel in Figure 7 shows model yield output from sce-551 of the full emulation fitted across the parameter space. The
 547 scenarios varying only along a single dimension (CO_2 , tempera-552 polynomial fit readily captures the climatological response to
 548 ture, precipitation, or nitrogen addition), with other inputs held553 perturbations.
 549 fixed at baseline levels; in all cases yields evolve smoothly
 550 across the space sampled. For reference we show the results554 Crop yield responses generally follow similar functional
 555 forms across models, though with a spread in magnitude. Fig-

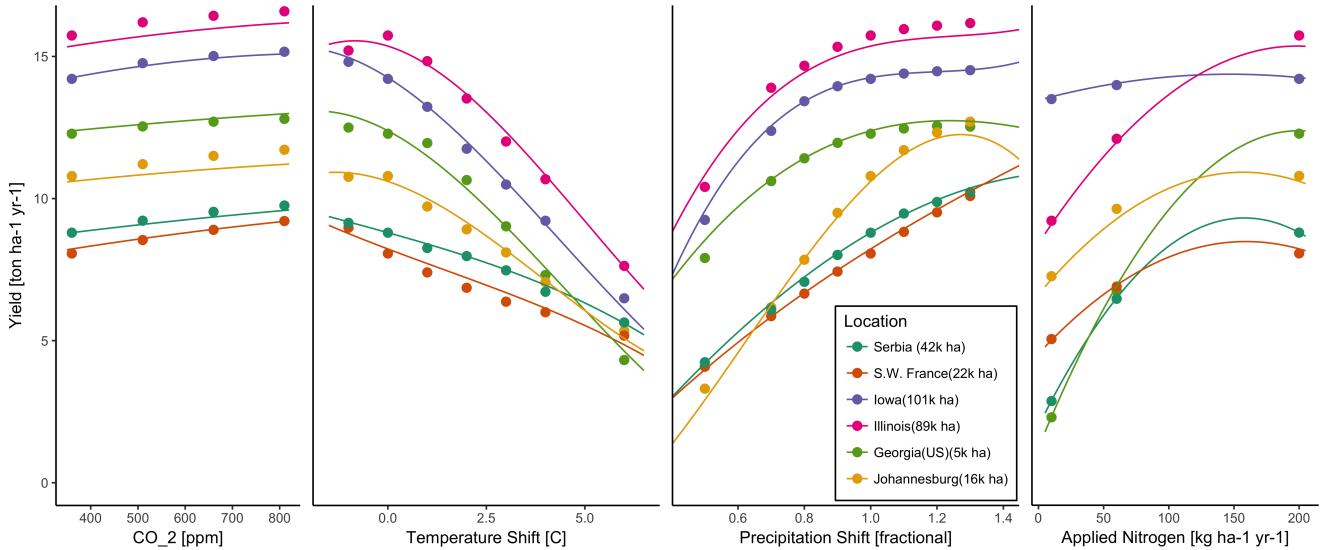


Figure 7: Example illustrating spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows 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).

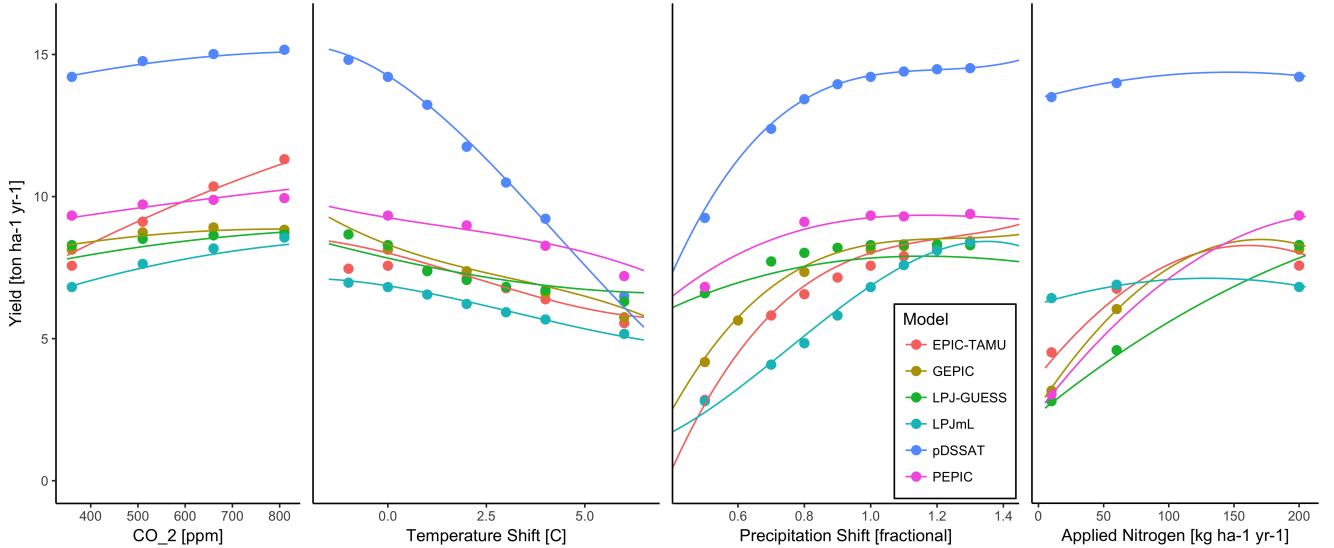


Figure 8: Example illustrating across-model variations in yield response. We show simulations and emulations from six models for rain-fed maize in the same Iowa location shown in Figure 7, with the same plot conventions. Models that did not simulate the nitrogen dimension are omitted for clarity. The 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.

ure 8 illustrates the inter-model diversity of yield responses₅₆₁ to the same perturbations, even for a single crop and location₅₆₂ (rain-fed maize in northern Iowa, the same location shown in₅₆₃ the Figure 7). The differences make it important to construct₅₆₄ emulators separately for each individual model, and the fidelity₅₆₅

of emulation can also differ across models. This figure illustrates a common phenomenon, that models differ more in response to perturbations in CO₂ and nitrogen perturbations than to those in temperature or precipitation. (Compare also Figures 4 and S18.) For this location and crop, CO₂ fertilization effects

566 can range from ~5–50%, and nitrogen responses from nearly
 567 flat to a 60% drop in the lowest-application simulation.

568 While the nitrogen dimension is important and uncertain, it
 569 is also the most problematic to emulate in this work because
 570 of its limited sampling. The GGCMI protocol specified only
 571 three nitrogen levels (10, 60 and 200 kg N y^{-1} ha $^{-1}$), so a third-
 572 order fit would be over-determined but a second-order fit can
 573 result in potentially unphysical results. Steep and nonlinear de-
 574 clines in yield with lower nitrogen levels means that some re-
 575 gressions imply a peak in yield between the 100 and 200 kg N
 576 y^{-1} ha $^{-1}$ levels. While there may be some reason to believe
 577 over-application of nitrogen at the wrong time in the growing
 578 season could lead to reduced yields, these features are almost
 579 certainly an artifact of under sampling. In addition, the poly-
 580 nomial fit cannot capture the well-documented saturation effect
 581 of nitrogen application (e.g. Ingestad, 1977) as accurately as
 582 would be possible with a non-parametric model.

583 To assess the ability of the polynomial emulation to capture
 584 the behavior of complex process-based models, we evaluate the
 585 normalized emulator error. That is, for each grid cell, model,
 586 and scenario we evaluate the difference between the model yield
 587 and its emulation, normalized by the inter-model standard de-
 588 viation in yield projections. This metric implies that emulation
 589 is generally satisfactory, with several distinct exceptions. Al-
 590 most all model-crop combination emulators have normalized
 591 errors less than one over nearly all currently cultivated hectares
 592 (Figure 9), but some individual model-crop combinations are
 593 problematic (e.g. PROMET for rice and soy, JULES for soy
 594 and winter wheat, Figures S14–S15). Normalized errors for soy
 595 are somewhat higher across all models not because emulator fi-
 596 delity is worse but because models agree more closely on yield
 597 changes for soy than for other crops (see Figure S16, lowering
 598 the denominator. Emulator performance often degrades in geo-
 599 graphic locations where crops are not currently cultivated. Fig-

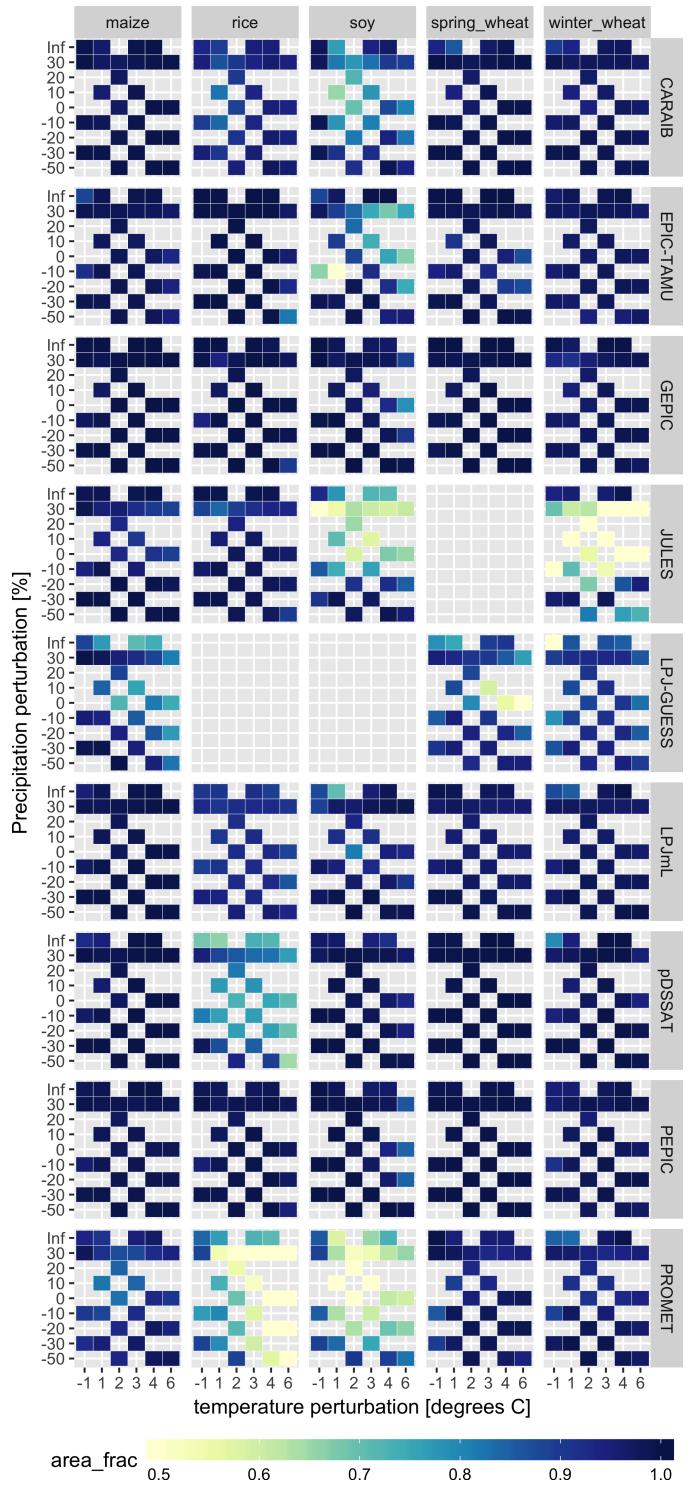


Figure 9: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO₂ and nitrogen held fixed at baseline values. Large columns are crops and large 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.

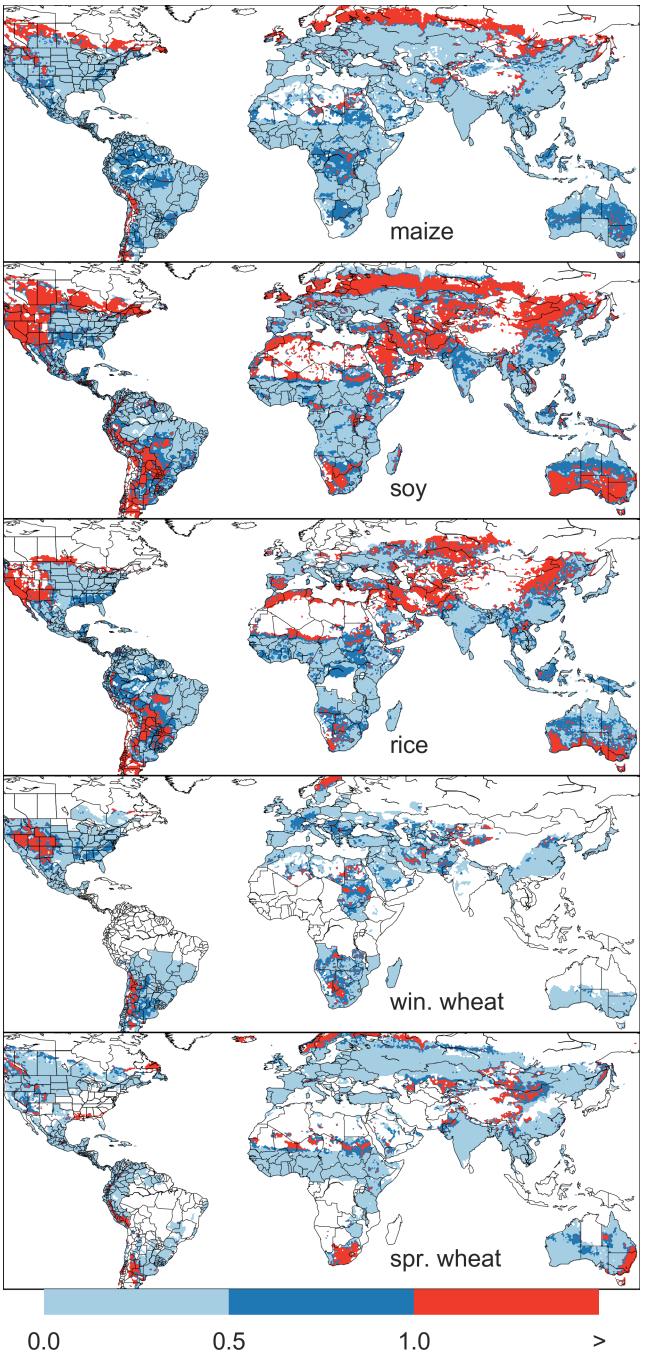


Figure 10: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error e , where $e > 1$ means that emulator error exceeds the multi-model standard deviation. White areas are those where crops 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.

ure 10 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 9) 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 an easy way to compare a ensemble 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 11; 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.

factors (CO_2 , temperature, precipitation, and applied nitrogen). Its global nature also allows identifying geographic shifts in high yield potential locations. We expect that the simulations will yield multiple insights in future studies, and show here a selection of preliminary results to illustrate their potential uses.

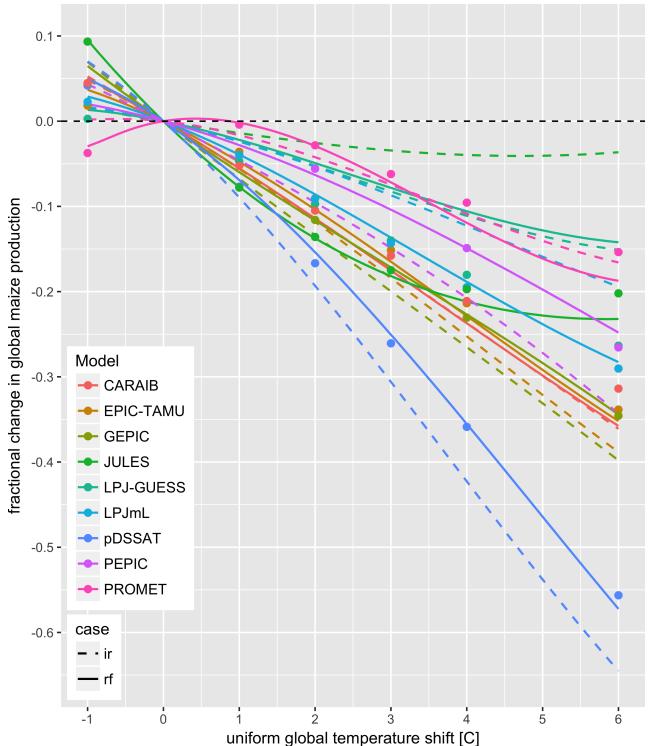


Figure 11: Global emulated damages for maize on currently cultivated lands for the GGCMI models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating up emulated values at the grid cell level, not from a regression of global mean 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 models, irrigated crops show a sharper reduction than do rain-fed because of the locations of cultivated areas: irrigated crops tend to be grown in warmer areas where impacts are more severe for a given temperature shift. (The exceptions are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures S16–S19 in the supplemental material.

First, the GGCMI Phase II simulations allow identifying major areas of uncertainty. Across the major crops, inter-model uncertainty is greatest for wheat and least for soy. Across factors impacting yields, inter-model uncertainty is largest for CO_2 fertilization and nitrogen response effects. Across geographic regions, projections are most uncertain in the high latitudes where yields may increase, and most robust in low latitudes where yield impacts are largest.

Second, the GGCMI Phase II simulations allow understanding the way that climate-driven changes and locations of cultivated land combine to produce yield impacts. One counterintuitive result immediate apparent is that irrigated maize shows steeper yield reductions under warming than does rain-fed maize when considered only over currently cultivated land.

The effect results from geographic differences in cultivation. 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 S5–S6). The same behavior holds for rice and winter wheat, but not for soy or spring wheat (Figures S8–S10). Irrigated wheat and maize are also more sensitive to nitrogen fertilization levels than are analogous non-irrigated crops, presumably because those rain-fed crops are limited by water as well as nitrogen availability (Figure S19). (Soy as an efficient atmospheric nitrogen-fixing is relatively insensitive to nitrogen, and rice is not generally grown in water-limited conditions).

Third, we show that even the relatively limited GGCMI Phase II sampling space allows emulation of the climatological response of crop models with a relatively simple reduced-form

4. Conclusions and discussion

The GGCMI Phase II experiment provides a database targeted to allow detailed study of crop yields from process-based models under climate change. The experiment is designed to facilitate not only comparing the sensitivities of process-based crop yield models to changing climate and management inputs but also evaluating the complex interactions between driving

680 statistical model. The systematic parameter sampling in the⁷¹⁴
681 GGCMI Phase II procedure provides information on the influ-⁷¹⁵
682 ence of multiple interacting factors in a way that single projec-⁷¹⁶
683 tions cannot, and emulating the resulting response surface then⁷¹⁷
684 produces a tool that can aid in both physical interpretation of⁷¹⁸
685 the process-based models and in assessment of agricultural im-⁷¹⁹
686 pacts under arbitrary climate scenarios. Emulating the climato-⁷²⁰
687 logical response isolates long-term impacts from any confound-⁷²¹
688 ing factors that complicate year-over-year changes, and the use⁷²²
689 of simple functional forms offer the possibility of physical in-⁷²³
690 terpretation of parameter values. We anticipate that systematic⁷²⁴
691 parameter sampling will become the norm in future model in-⁷²⁵
692 tercomparison exercise.

693 While the GGCMI Phase II database should offer the foun-⁷²⁷
694 dation for multiple future studies, several cautions need to be⁷²⁸
695 noted. Because the simulation protocol was designed to fo-⁷²⁹
696 cus on change in yield under climate perturbations and not⁷³⁰
697 on replicating real-world yields, the models are not formally⁷³¹
698 calibrated so cannot be used for impacts projections unless in⁷³²
699 used in conjunction with historical data (or data products). Be-⁷³³
700 cause the GGCMI simulations apply uniform perturbations to⁷³⁴
701 historical climate inputs, they do not sample changes in higher⁷³⁵
702 order moments, and cannot address the additional crop yield⁷³⁶
703 impacts of potential changes in climate variability. Although⁷³⁷
704 distributional changes in model projections are fairly uncertain
705 at present, follow-on experiments may wish to consider them.⁷³⁸

706 Several recent studies have described procedures for generating⁷³⁹
707 simulations that combine historical data with model projections⁷⁴⁰
708 of not only mean changes in temperature and precipitation but⁷⁴¹
709 changes in their marginal distributions (cite) or temporal depen-⁷⁴²
710 dence ().

711 — NEEDS EDIT – still theres lots of future work can be⁷⁴⁴
712 done even with this database—

713 Detailed examination of interaction terms More robust quan-⁷⁴⁶

tification of the sensitivity of different models to the input drivers. Comparison with field-level experimental data can then aid in model evaluation. As mentioned previously, database allows study of geographic shifts in optimal growing regions for different crops. The output dataset also contains other runs and variables not analyzed or shown here. Runs include several which allowed adaptation to climate changes by altering growing seasons, and additional variables include above ground biomass, LAI, and root biomass (as many as 25 output variables for some models). Emulation studies that are possible include study of year-over-year vs climatological emulation, and more systematic evaluation of different statistical model specifications and formal calculation of uncertainties in derived parameters.

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 allow 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.

747 was supported by the MACMIT project (01LN1317A) funded⁷⁸⁵
 748 through the German Federal Ministry of Education and Re-⁷⁸⁶
 749 search (BMBF). C.F. was supported by the European Research⁷⁸⁷
 750 Council Synergy grant #ERC-2013-SynG-610028 Imbalance-⁷⁸⁸
 751 P. P.F. and K.W. were supported by the Newton Fund through⁷⁸⁹
 752 the Met Office Climate Science for Service Partnership Brazil⁷⁹¹
 753 (CSSP Brazil). A.S. was supported by the Office of Science⁷⁹²
 754 of the U.S. Department of Energy as part of the Multi-sector⁷⁹³
 755 Dynamics Research Program Area. Computing resources were⁷⁹⁵
 756 provided by the University of Chicago Research Computing⁷⁹⁶
 757 Center (RCC). S.O. acknowledges support from the Swedish⁷⁹⁷
 758 strong research areas BECC and MERGE together with sup-⁷⁹⁸
 759 port from LUCCI (Lund University Centre for studies of Car-⁸⁰⁰
 760 bon Cycle and Climate Interactions).

761 6. References

- 762 Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013).
 763 Implication of crop model calibration strategies for assessing regional im-⁸⁰⁵
 764 pacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.
 765 Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D.,
 766 A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Al-⁸⁰⁶
 767 derman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Bier-⁸⁰⁷
 768 nath, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising tempera-⁸⁰⁸
 769 tures reduce global wheat production. *Nature Climate Change*, 5, 143–147.
 770 doi:10.1038/nclimate2470.
 771 Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A.,
 772 J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N.,
 773 Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath,
 774 C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simula-⁸¹⁴
 775 ting wheat yields under climate change. *Nature Climate Change*, 3, 827832.
 776 doi:10.1038/nclimate1916.
 777 Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other
 778 Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Ef-⁸²⁰
 779 ficiency, Carbon Sequestration, and Environmental Pollution. *Advances in*
 780 *Agronomy*, 86, 341 – 409.
 781 Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov,
 782 N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat
 783 production potentials and management flexibility under the representative
 784 concentration pathways. *Global and Planetary Change*, 122, 107 – 121.
 785 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields
 786 from global gridded crop models. *Agricultural and Forest Meteorology*, 236,
 787 145 – 161.
 788 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded
 789 crop models using statistical estimates. *Agricultural and Forest Meteorol-
 790 ogy*, 214–215, 134 – 147.
 791 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S.
 792 (2018). Implementing the Nitrogen cycle into the dynamic global vegeta-
 793 tion, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific
 794 Model Development*, 11, 2789–2812.
 795 Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L.,
 796 & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections
 797 Based on Precomputed GCM Runs. *Journal of Climate*, 27, 1829–1844.
 798 Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.
 799 (2014). A meta-analysis of crop yield under climate change and adaptation.
 800 *Nature Climate Change*, 4, 287 – 291.
 801 Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process
 802 emulation of dynamic computer codes. *Biometrika*, 96, 663–676.
 803 Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In
 804 C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth,
 805 Duke University, Durham, North Carolina* (pp. 115–118). Durham, North
 806 Carolina.
 807 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for
 808 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).
 809 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberrous,
 810 M., & François, L. (2011). Responses of European forest ecosystems to
 811 21st century climate: assessing changes in interannual variability and fire
 812 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).
 813 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,
 814 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating
 815 impact models and sectors (pSIMS). *Environmental Modelling and Soft-
 816 ware*, 62, 509–516.
 817 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,
 818 Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurrealde, R. C.,
 819 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.
 820 (2015). The Global Gridded Crop Model Intercomparison: data and mod-
 821 eling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8,
 822 261–277.
 823 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J.,
 824 & Taylor, K. E. (2016). Overview of the coupled model intercomparison
 825 project phase 6 (cmip6) experimental design and organization. *Geoscientific
 826 Model Development*, 9, 1937–1958.
 827 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-
 828

- 828 mate change impacts on durum wheat in the mediterranean region. *Natural Hazards and Earth System Sciences*, 11, 1293–1302. 871
- 829 Holberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Regionalization of a large-scale crop growth model for sub-Saharan Africa: 873
830 Model setup, evaluation, and estimation of maize yields. *Agriculture, Ecosystems & Environment*, 151, 21 – 33. 876
- 831 Food and Agriculture Organization of the United Nations (2018). FAOSTAT database. URL: <http://www.fao.org/faostat/en/home>. 877
- 832 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., Nen-del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna, M., Semenov, M. A., Slawinski, C., Stratonovitch, 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. 888
- 840 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. 891
- 842 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. 894
- 844 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. 896
- 846 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. 901
- 848 Heady, E. O. (1957). An Econometric Investigation of the Technology of Agricultural Production Functions. *Econometrica*, 25, 249–268. 905
- 850 Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa State University Press. 907
- 852 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. 910
- 854 Holzkämper, A., Calanca, P., & Fuhrer, J. (2012). Statistical crop models: Predicting the effects of temperature and precipitation changes. *Climate Research*, 51, 11–21. 913
- 856 Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalglish, N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, 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.
- 858 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).
- 860 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.
- 862 Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Nitrogen Fertilizers. *Ambio*, 6, 146–151.
- 864 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.
- 866 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.
- 868 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.
- 870 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.
- 872 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.

- 914 Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., & 957
 915 Smith, B. (2013). Implications of accounting for land use in simulations of 958
 916 ecosystem carbon cycling in Africa. *Earth System Dynamics*, 4, 385–407. 959
 917 Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling 960
 918 wheat yield and crop water productivity with high resolution on a global 961
 919 scale. *Agricultural Systems*, 94, 478 – 493. 962
 920 Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a). 963
 921 Global investigation of impacts of PET methods on simulating crop-water 964
 922 relations for maize. *Agricultural and Forest Meteorology*, 221, 164 – 175. 965
 923 Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., 966
 924 & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs 967
 925 with yields from major crop cultivations. *Science of The Total Environment*, 968
 926 572, 526 – 537. 969
 927 Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict 970
 928 crop yield responses to climate change. *Agricultural and Forest Meteorol-* 971
 929 *ogy*, 150, 1443 – 1452. 972
 930 Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relation-973
 931 ships and the impacts of recent warming. *Environmental Research Letters*, 974
 932 2, 014002. 975
 933 MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447. 976
 934 Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P., 977
 935 Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi, 978
 936 P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano, 979
 937 D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of 980
 938 large simulated yield datasets for studying climate effects. (p. 1100). 981
 939 doi:10.13140/RG.2.1.5173.8328. 982
 940 Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrolog-983
 941 ical modelling to study the impact of climate change on the water flows of 984
 942 mountain watersheds. *Journal of Hydrology*, 376, 362 – 377. 985
 943 Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., 986
 944 & Calzadilla, A. (2009). Global biomass production potentials exceed ex-987
 945 pected future demand without the need for cropland expansion. *Nature Com-* 988
 946 *munications*, 6. 989
 947 McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srinivasan, C., 990
 948 Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S. 991
 949 (2015). Integrated assessments of the impacts of climate change on agricul-992
 950 ture: An overview of AgMIP regional research in South Asia. *Chapter in:* 993
 951 *Handbook of Climate Change and Agroecosystems*, (pp. 201–218). 994
 952 Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical 995
 953 weather responsiveness of crop yields: US evidence and implications for 996
 954 the agricultural impacts of climate change. *Environmental Research Letters*, 997
 955 12. 998
 956 Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate 999
 change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, 8.
 Müller, C., Elliott, J., Chrysanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurrealde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. (2017). Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, 10, 1403–1422.
 Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T. (1997). Effect of CO₂ 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., Wårliind, 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.
 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-sos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
 Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet, J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C.,

- 1000 Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M.,¹⁰⁴³
 1001 Semenov, M., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang,¹⁰⁴⁴
 1002 E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation¹⁰⁴⁵
 1003 effects on wheat yield across a European transect: a crop model ensemble¹⁰⁴⁶
 1004 analysis using impact response surfaces. *Climate Research*, 65, 87–105.¹⁰⁴⁷
 1005 Porter et al. (IPCC) (2014). Food security and food production systems. Cli¹⁰⁴⁸
 1006 mate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global¹⁰⁴⁹
 1007 and Sectoral Aspects. Contribution of Working Group II to the Fifth Assess¹⁰⁵⁰
 1008 ment Report of the Intergovernmental Panel on Climate Change. In C. Fio¹⁰⁵¹
 1009 et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK¹⁰⁵²
 1010 Cambridge University Press.¹⁰⁵³
- 1011 Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of¹⁰⁵⁴
 1012 monthly growing areas of 26 irrigated crops.¹⁰⁵⁵
- 1013 Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly¹⁰⁵⁶
 1014 Irrigated and Rainfed crop Areas around the Year 2000: A New High¹⁰⁵⁷
 1015 Resolution Data Set for Agricultural and Hydrological Modeling. *Global*¹⁰⁵⁸
1016 Biogeochemical Cycles, 24, GB1011.¹⁰⁵⁹
- 1017 Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E.,¹⁰⁶⁰
 1018 & Arneth, A. (2016). Climate analogues suggest limited potential for inten¹⁰⁶¹
 1019 sification of production on current croplands under climate change. *Nature*¹⁰⁶²
1020 Communications, 7, 12608.¹⁰⁶³
- 1021 Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term cli¹⁰⁶⁴
 1022 mate change based on a resampling ensemble technique. *Tellus A: Dynamical*¹⁰⁶⁵
1023 Meteorology and Oceanography, 58, 461–472.¹⁰⁶⁶
- 1024 Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the¹⁰⁶⁷
 1025 reduction and sensitivity analysis of complex environmental models. *Envi¹⁰⁶⁸
 1026 ronmental Modelling & Software*, 34, 1 – 4.¹⁰⁶⁹
- 1027 Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling¹⁰⁷⁰
 1028 in water resources. *Water Resources Research*, 48.¹⁰⁷¹
- 1029 Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017)¹⁰⁷²
 1030 Comparing and combining process-based crop models and statistical models¹⁰⁷³
 1031 with some implications for climate change. *Environmental Research Letters*¹⁰⁷⁴
 1032 12.¹⁰⁷⁵
- 1033 Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A.,¹⁰⁷⁶
 1034 Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek,¹⁰⁷⁷
 1035 F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014)¹⁰⁷⁸
 1036 Assessing agricultural risks of climate change in the 21st century in a global¹⁰⁷⁹
 1037 gridded crop model intercomparison. *Proceedings of the National Academy*¹⁰⁸⁰
1038 of Sciences, 111, 3268–3273.¹⁰⁸¹
- 1039 Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P.,¹⁰⁸²
 1040 Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Bass, B., Ew¹⁰⁸³
 1041 ert, F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural¹⁰⁸⁴
 1042 Model Intercomparison and Improvement Project (AgMIP): Protocols and¹⁰⁸⁵
 pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182.
- Rosenzweig, C., Ruane, A. C., Antle, J., Elliott, J., Ashfaq, M., Chatta,
 A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G.,
 Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M.,
 Müller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M.,
 Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., & Wiebe, K.
 (2018). Coordinating AgMIP data and models across global and regional
 scales for 1.5°C and 2.0°C assessments. *Philosophical Transactions of the
 Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 376.
- Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason-
 D'Croz, D., Müller, C., Porter, C., Phillips, M. M., Raymundo, R. M.,
 Sands, R., Valdivia, R. O., White, J. W., Wiebe, K., & Rosenzweig, C.
 (2018). Biophysical and economic implications for agriculture of +1.5° and
 +2.0°C global warming using AgMIP Coordinated Global and Regional As-
 sessments. *Climate Research*, 76, 17–39.
- Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown,
 D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013).
 Climate change impact uncertainties for maize in panama: Farm informa-
 tion, climate projections, and yield sensitivities. *Agricultural and Forest
 Meteorology*, 170, 132 – 145.
- Ruane, A. C., Goldberg, R., & Chrysanthacopoulos, J. (2015). Climate forcing
 datasets for agricultural modeling: Merged products for gap-filling and
 historical climate series estimation. *Agric. Forest Meteorol.*, 200, 233–248.
- Ruane, A. C., McDermid, S., Rosenzweig, C., Baigorria, G. A., Jones, J. W.,
 Romero, C. C., & Cecil, L. D. (2014). Carbon-temperature-water change
 analysis for peanut production under climate change: A prototype for the
 agmip coordinated climate-crop modeling project (c3mp). *Glob. Change
 Biol.*, 20, 394–407. doi:10.1111/gcb.12412.
- Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901–
 2100 depicted by world maps of the Köppen-Geiger climate classification.
Meteorologische Zeitschrift, 19, 135–141.
- Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting
 dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19,
 607–620.
- Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng,
 D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolin-
 ski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frieler, K.
 (2017). Consistent negative response of US crops to high temperatures in
 observations and crop models. *Nature Communications*, 8, 13931.
- Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate
 severe damages to U.S. crop yields under climate change. *Proceedings of
 the National Academy of Sciences*, 106, 15594–15598.

- 1086 Snyder, A., Calvin, K. V., Phillips, M., & Ruane, A. C. (2018). A crop yield
1087 change emulator for use in gcam and similar models: Persephone v1.0. *Geo-*
1088 *scientific Model Development Discussions*, 2018, 1–42.
- 1089 Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implemen-
1090 tation and evaluation of nonparametric regression procedures for sensitivity
1091 analysis of computationally demanding models. *Reliability Engineering &*
1092 *System Safety*, 94, 1735 – 1763.
- 1093 Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5
1094 and the experiment design. *Bulletin of the American Meteorological Society*,
1095 93, 485–498.
- 1096 Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of cli-
1097 mate change impacts on global crop yields. *Geophysical Research Letters*,
1098 35.
- 1099 Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., &
1100 Martin, J. F. (2014). Modeling sugarcane yield with a process-based model
1101 from site to continental scale: Uncertainties arising from model structure
1102 and parameter values. *Geoscientific Model Development*, 7, 1225–1245.
- 1103 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe,
1104 J. (2014). The Inter-Sectoral Impact Model Intercomparison Project
1105 (ISI-MIP): Project framework. *Proceedings of the National Academy of*
1106 *Sciences*, 111, 3228–3232.
- 1107 White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Method-
1108 ologies for simulating impacts of climate change on crop production. *Field*
1109 *Crops Research*, 124, 357 – 368.
- 1110 Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T.,
1111 Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance
1112 against site observations of irrigated maize from Mead, Nebraska. *Geosci-
1113 entific Model Development*, 10, 1291–1320.
- 1114 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield vari-
1115 ability in JULES-crop and implications for forcing with seasonal weather
1116 forecasts. *Geoscientific Model Development*, 8, 3987–3997.
- 1117 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Land-
1118 bouwkundige Onderzoeken* : 64.6, .
- 1119 Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato
1120 yields on changes in climate and co2. *Agricultural and Forest Meteorology*,
1121 112, 217 – 231.
- 1122 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao,
1123 Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A.,
1124 Li, T., Lin, E., Liu, Q., Martre, P., Miller, C., Peng, S., Peuelas, J., Ruane,
1125 A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng,
1126 S. (2017). Temperature increase reduces global yields of major crops in four
1127 independent estimates. *Proc. Natl. Acad. Sci.*, 114, 9326–9331.