

The GGCMI Phase II experiment: simulating and emulating 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 crop 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 applied nitrogen inputs and constitutes a data-rich basis of projected yield changes across twelve models and five crops (maize, soy, rice, spring wheat, and winter wheat) using global gridded simulations. In this paper we present the simulation output database from Phase II of the GGCMI effort, a targeted experiment aimed at understanding the sensitivity to and interaction between multiple climate variables (as well as management) on yields, and illustrate some initial summary results from the model intercomparison project. We also present the construction of a simple “emulator” or statistical representation of the simulated 30-year mean climatological output in each location for each crop and model. The emulator captures the mean-climatological response of the process-based models in a lightweight, computationally tractable form that facilitates model comparison as well as potential applications in subsequent modeling efforts such as integrated assessment.

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

1. Introduction

2 Understanding crop yield response to a changing climate
3 is critically important, especially as the global food produc-
4 tion system will face pressure from increased demand over
5 the next century. Climate-related reductions in supply could
6 therefore have severe socioeconomic consequences. Multiple
7 studies using different crop or climate models concur in pre-
8 dicting sharp yield reductions on currently cultivated cropland
9 under business-as-usual climate scenarios, although their yield
10 projections show considerable spread (e.g. Rosenzweig et al.,
11 2014, Schauberger et al., 2017, Porter et al. (IPCC), 2014, and
12 references therein). Modeling crop responses continues to be
13 challenging, as crop growth is a function of complex interac-
14 tions between climate inputs and management practices. Inter-
15 comparison projects targeting model responses to important
16 drivers are critical to improve future projections.

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

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

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

41 Process-based models do continue to struggle with some im-
42 portant details, including reproducing historical year-to-year
43 variability (e.g. Müller et al., 2017), reproducing historical
44 yields when driven by reanalysis weather (e.g. Glotter et al.,
45 2014), and low sensitivity to extreme events (e.g. Glotter et al.,
46 2015, Schewe et al., 2019). These issues are driven in part by
47 the diversity of new cultivars and genetic variants, which out-
48 strips the ability of academic modeling groups to capture them
49 (e.g. Jones et al., 2017). Models also do not simulate many ad-
50 ditional factors affecting production, including pests, diseases,
51 and weeds. For these reasons, individual studies must generally
52 re-calibrate models to ensure that short-term predictions reflect
53 current cultivar mixes, and long-term projections retain consid-
54 erable uncertainty (Wolf & Oijen, 2002, Jagtap & Jones, 2002,
55 Iizumi et al., 2010, Angulo et al., 2013, Asseng et al., 2013,
56 2015). Inter-model discrepancies can also be high in areas not
57 yet cultivated (e.g. Challinor et al., 2014, White et al., 2011).
58 Finally, process-based models present additional difficulties for
59 high-resolution global studies because of their complexity and
60 computational requirements. For global economic impacts as-
61 sessments, it is often impossible to integrate a set of process-
62 based crop models directly into an integrated assessment model
63 to estimate the potential cost of climate change to the agricul-
64 tural sector.

65 Nevertheless, process-based models are necessary for under-
66 standing the global future yield impacts of climate change for

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

many reasons. First, cultivation may shift to new areas, where
68 no yield data are currently available and therefore statistical
models cannot apply. Yield data are also often limited in the
70 developing world, where future climate impacts may be the
most critical. Finally, only process-based models can capture
72 the growth response to novel conditions and practices that are
not represented in historical data (e.g. Pugh et al., 2016, Roberts
74 et al., 2017). These novel changes can include the direct fer-
tilization effect of elevated CO₂, and changes in management
76 practices that may ameliorate climate-induced damages.

Interest has been rising in statistical emulation, which al-
78 lows combining advantageous features of both statistical and
process-based models. The approach involves constructing a
80 statistical representation or “surrogate model” of complicated
numerical simulations by using simulation output as the train-
82 ing data for a statistical model (e.g. O’Hagan, 2006, Conti et al.,
2009). Emulation is particularly useful in cases where sim-
84 ulations are complex and output data volumes are large, and
has been used in a variety of fields, including hydrology (e.g.
86 Razavi et al., 2012), engineering (e.g. Storlie et al., 2009),
environmental sciences (e.g. Ratto et al., 2012), and climate
88 (e.g. Castruccio et al., 2014, Holden et al., 2014). For agri-
cultural impacts studies, emulation of process-based models
90 allows capturing key relationships between input variables in
a lightweight, flexible form that is compatible with economic
92 studies.

In the past decade, multiple studies have developed emula-
94 tors of process-based crop simulations. Early studies proposing
or describing potential crop yield emulators include Howden
96 & Crimp (2005), Räisänen & Ruokolainen (2006), Lobell &
Burke (2010), and Ferrise et al. (2011), who used a machine
98 learning approach to predict Mediterranean wheat yields. Stud-
ies developing single-model emulators include Holzkämper
100 et al. (2012) for the CropSyst model, Ruane et al. (2013) for

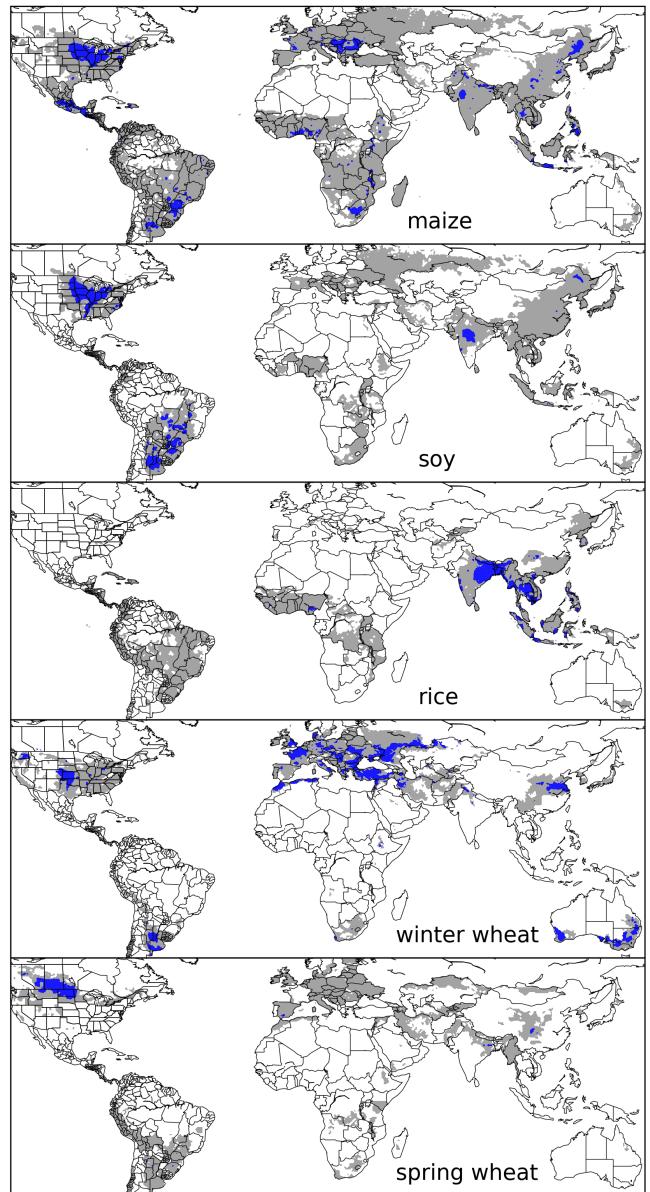


Figure 1: Presently cultivated area for rain-fed crops. Blue indicates grid cells with more than 20,000 hectares (~10% of the equatorial grid cells) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated areas for maize, rice, and soy are taken from the MIRCA2000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

the CERES wheat model, and Oyebamiji et al. (2015) for the
102 LPJmL model (for multiple crops, using multiple scenarios as
a training set). More recently, emulators have begun to be used
in the context of multi-model intercomparisons, with Blanc &
104 Sultan (2015), Blanc (2017), Ostberg et al. (2018) and Mis-
try et al. (2017) using them to analyze the five crop models

of the Inter-Sectoral Impacts Model Intercomparison Project
108 (ISIMIP) (Warszawski et al., 2014), which simulated yields
for maize, soy, wheat, and rice. Choices differ: Blanc & Sul-
110 tan (2015) and Blanc (2017) base their emulation on historical
simulations and three climate scenarios for one Representative
112 Concentration Pathway (RPC8.5), which represents a high level
of global warming; and use local weather variables and yields
114 in their regression across soil regions; Ostberg et al. (2018) use
global mean temperature change (and CO₂) as regressors then
116 pattern-scale to emulate local yields; while Mistry et al. (2017)
compare emulated and observed historical yields, using local
118 weather data and a historical crop simulation. These efforts
do share important common features: all emulate annual crop
120 yields across the entire scenario or scenarios, and when future
scenarios are considered, they are non-stationary, i.e. their input
122 climate parameters evolve over the course of the simulations.

An alternative approach is to construct a training set of multi-
124 ple stationary scenarios in which parameters are systematically
varied. Such a “parameter sweep” offers several advantages for
126 emulation over scenarios in which climate evolves over time.
First, it allows separating the effects of different variables that
128 impact yields but that are highly correlated in realistic future
scenarios (e.g. CO₂ and temperature). Second, it allows making
130 a distinction between year-over-year yield variations and cli-
matological changes, which may involve different responses to
132 the particular climate regressors used (e.g. Ruane et al., 2016).
For example, if year-over-year yield variations are driven pre-
134 dominantly by variations in the distribution of temperatures
throughout the growing season, and long-term climate changes
136 are driven predominantly by shifts in means, then regressing
on the mean growing season temperature will produce different
138 yield responses at annual vs. climatological timescales. Disad-
vantaged of this approach include neglecting changes in sea-
140 sonality and some implausible combinations of input settings

(e.g. colder temperature and high CO₂).

Systematic parameter sweeps have begun to be used in crop
model evaluation and emulation, with early efforts in 2014 and
142 2015 (Ruane et al., 2014, Makowski et al., 2015, Pirttioja et al.,
2015), and several recent studies in 2018 (Fronzek et al., 2018,
Snyder et al., 2018, Ruiz-Ramos et al., 2018). All three 2018
144 studies sample multiple perturbations to temperature and pre-
cipitation (with Snyder et al. (2018) and Ruiz-Ramos et al.
146 (2018) adding CO₂ as well), in 132, 99 and 220 different combi-
nations, respectively, and take advantage of the structured train-
ing set to construct emulators (“response surfaces”) of climati-
148 logical mean yields, omitting year-over-year variations. All
studies focus on a limited number of sites; Fronzek et al. (2018)
and Ruiz-Ramos et al. (2018) simulate only wheat (over many
models) and Snyder et al. (2018) analyzes four crops (maize,
wheat, rice, soy) for agricultural impacts experiments with the
152 GCAM (Calvin et al., 2019) model.

In this paper we describe a new comprehensive dataset de-
156 signed to expand the parameter sweep approach still further.
The Global Gridded Crop Model Intercomparison (GGCMI)
160 Phase II experiment involves running a suite of process-based
crop models across historical conditions perturbed by a set of
162 discrete steps in different input parameters, including an ap-
plied nitrogen dimension. The experimental protocol involves
over 700 different parameter combinations for each model and
crop, with simulations providing near-global coverage at a half
degree spatial resolution. The experiment was conducted as
164 part of the Agricultural Model Intercomparison and Improve-
ment 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
166 on the AgMIP Coordinated Climate-Crop Modeling Project
(C3MP) (Ruane et al., 2014, McDermid et al., 2015) and con-

tributes 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 input drivers 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, present initial results, and introduce a spatially explicit emulator for climatological time scales that allows for representing crop model responses in economic assessment models and other applications.

2. Simulation – Methods

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 31-year 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 the same historical time series (1980-2010) used in Phase I, but with individual climate or management inputs 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).

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”). The dataset is designed to allow researchers to:

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

The experimental protocol consists of 9 levels for precipitation perturbations, 7 for temperature, 4 for 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, limitations from actual water supply are not considered so that plants are optimally supplied with water. Temperature perturbations are applied as absolute offsets from the daily mean, minimum, and maximum temperature time se-

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: GGCMI Phase II input variable test levels. Temperature and precipitation values indicate the perturbations from the historical climatology and are selected to represent reasonable ranges for potential climate changes in the medium term. W-percentage does not apply to the irrigated (W_{inf}) simulations, which are all simulated at the maximum beneficial levels of water. *Only simulated by one model.

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 , Wu et al. (2016)	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 GGCMI Phase II and the number of C, T, W, and N simulations that each performs for rain-fed crops (“Simulations per Crop”), with 672 as the maximum for rain fed crops. “N-Dim.” indicates whether the simulations include varying nitrogen levels. Two models provide only one nitrogen level, and †PROMET provides only two of the three nitrogen levels (and so is not emulated across the nitrogen dimension). All models provide the same set of simulations across all modeled crops, but some omit individual crops. (For example, APSIM does not simulate winter wheat.) Irrigated simulations are provided at the level of the other covariates for each model, i.e. an additional 84 simulations for fully-sampled models. Simulations are nearly global but their geographic extent can vary, even for different crops in an individual model, since some simulations omit regions far outside the currently cultivated area. In most cases, historical daily climate inputs are taken from the 0.5 degree NASA AgMERRA daily gridded re-analysis product specifically designed for agricultural modeling, with satellite-corrected precipitation (Ruane et al., 2015), but two models (marked with *) require sub-daily input data and use alternative sources. See Elliott et al. (2015) for additional details.

ries for each grid cell. 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. The resulting GGCMI Phase II dataset 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 models share a common base (e.g. the LPJ family or the EPIC family of models), they have subsequently developed independently. Differences in model structure 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 the extent of simulated area for different crops. Growing seasons are stan-

dardized across models (with assumptions based on Sacks et al. (2010) and Portmann et al. (2008, 2010)), but vary by crop and by location on the globe. For example, maize is sown in March in Spain, in July in Indonesia, and in December in Namibia. All stresses are disabled other than factors related to nutrients, temperature, and water (e.g. alkalinity and salinity). No additional nitrogen inputs, such as atmospheric deposition, are considered, but some model treatments of soil organic matter allows additional nitrogen release through mineralization. See Rosenzweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017) for further details on models and underlying assumptions.

Christoph – This is a bit dangerous and part of why I’m hesitant to shift the focus of the paper to describing the simulation setup and results prominently. Various models used here cannot be found in these publications. Reviewers typically want to see some basic information without having to look into other papers. I’m fearing that this is a larger trap than missing to test functional forms in the emulator section.

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

Each model is run at 0.5 degree spatial resolution and covers
272 all currently cultivated areas and much of the uncultivated land
273 area. (See Figure 1 for the present-day cultivated area of rain-
274 fed crops, and Figure S1 in the Supplemental Material for irrigated
275 crops.) Coverage extends considerably outside currently
276 cultivated areas because cultivation will likely shift under cli-
277 mate change. However, areas are not simulated in some cases if
278 they are assumed to remain non-arable even under an extreme

climate change; these regions include Greenland, far-northern
280 Canada, Siberia, Antarctica, the Gobi and Sahara Deserts, and
281 central Australia.

All models produce as output crop yields ($\text{tons ha}^{-1} \text{ year}^{-1}$)
282 for each 0.5 degree grid cell. Because both yields and yield
283 changes vary substantially across models and across grid cells,
284 we primarily analyze relative change from a baseline. We take
285 as the baseline the scenario with historical climatology (i.e. T
286 and P changes of 0), C of 360 ppm, and applied N at 200 kg
287 ha^{-1} . We show absolute yields in some cases to illustrate geo-
288 graphic differences in yields.

The GGCMI Phase II simulations are designed for evaluating
290 changes in yield but not absolute yields, since they omit
291 detailed calibrations. To provide some validation of the skill of
292 the process-based models used, we repeat the validation exer-
293 cises of Müller et al. (2017) for GGCMI Phase I. See Appendix
294

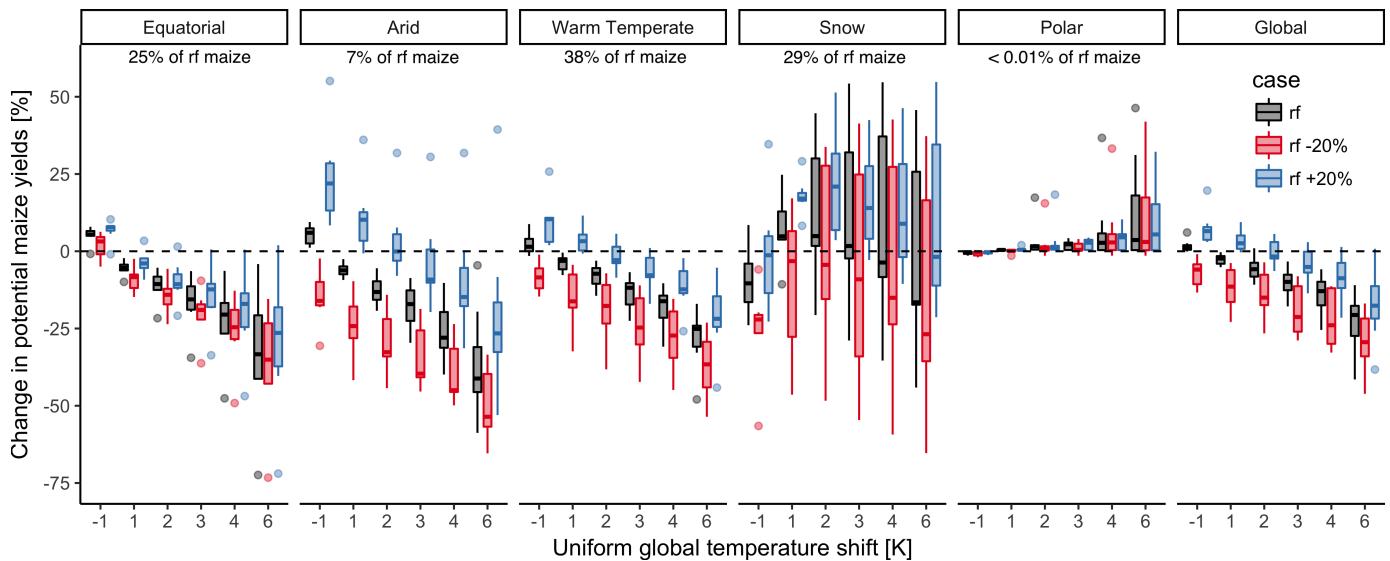


Figure 2: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek, 2010). Note that ‘Equatorial’ and ‘Snow’ regions are sometimes referred to as ‘tropical’ and ‘cold continental’ respectively elsewhere. We show 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). The y-axis is fractional change in the regional average climatological potential yield relative to the baseline. The figure shows all modeled land area; see Figure S6 in the supplemental material for only currently-cultivated land. Panel text gives the percentage of rain-fed maize presently cultivated in each climate zone (data from Portmann et al., 2010). Box-and-whiskers plots show distribution across models, with median marked. Box edges are first and third quartiles, i.e. box height is the interquartile range (IQR). Whiskers extend to maximum and minimum of simulations but are limited at 1.5-IQR; otherwise the outlier is shown. Models generally agree in most climate regions (other than ‘cold continental’ (snow)), 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. The right panel (Global) shows yield responses to a 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. Panel subtitles indicate percentage of rain-fed(rf) maize currently grow in each region.

A for details on simulation model validation.

296 3. Simulation – Results

Crop models in the GGCMI Phase II ensemble show broadly
298 consistent responses to climate and management perturbations
in most regions, with a strong negative impact of increased tem-
300 perature in all but the coldest regions. We illustrate this re-
sult for rain-fed maize in Figure 2, which shows yields across
302 all grid cells for the primary Köppen-Geiger climate regions
(Rubel & Kottek, 2010). In warming scenarios with precipi-
304 tation held constant, all models show decreases in maize yield
in the ‘warm temperate’, ‘equatorial’(tropical), and arid regions
306 that account for nearly three-quarters of global maize produc-
tion. These impacts are robust for even moderate climate per-
308 turbations. In the ‘warm temperate’ zone, even a 1 degree tem-
perature rise with other variables held fixed leads to a median
310 yield reduction that outweighs the variance across models. A
6 degree temperature rise results in median loss of ~25% of
312 yields with a signal to noise ratio of nearly three to one. A no-
table exception is the ‘snow’ (‘cold-continental’) region, where
314 models disagree strongly, extending even to the sign of impacts.
Other crops show similar responses to warming, with robust
316 yield losses in warmer locations and high inter-model variance
in the cold continental regions (Figure S7).

318 The effects of rainfall changes on maize yields shown in Fig-
ure 2 are also as expected and are consistent across models.
320 Increased rainfall mitigates the negative effect of higher tem-
peratures by counteracting the increased evapo-transpiration to
322 some degree, most strongly in arid regions. Decreased rain-
fall amplifies yield losses and also increases inter-model vari-
324 ance more strongly, suggesting that models have difficulty rep-
resenting crop response to water stress or increased evapo-
326 transpiration due to warmer temperatures. We show only rain-
fed maize here; see Figure S5 for the irrigated case. As ex-

328 pected, irrigated crops are more resilient to temperature in-
creases in all regions, especially so where water is limiting.

330 Mapping the distribution of baseline yields and yield changes
shows the geographic dependencies that underlie these results.
332 Figure 3 shows baseline and changes in the T+4 scenario for
rain-fed maize, soy, and rice in the multi-model ensemble mean,
334 with locations of model agreement marked. Absolute yield pot-
entials show strong spatial variation, with much of the Earth’s
336 surface area unsuitable for any of these crops. In general, mod-
els agree most on yield response in regions where yield poten-
tials are currently high and therefore where crops are currently
338 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 more
340 uncertain, possibly because calibration is especially important
for wheat (e.g. Asseng et al., 2013).

344 4. Emulation – Methods

As part of our demonstration of the properties of the GGCMI
Phase II dataset, we construct an emulator of 30-year clima-
346 tological mean yields. This approach is made possible by
the structured set of simulations involving systematic pertur-
348 bations. In the GGCMI Phase II dataset, the year-over-year re-
sponses are generally quantitatively distinct from (and larger
350 than) climatological mean responses. In the example Figure
4, responses to year-over-year temperature variations are 100%
larger than those to long-term perturbations in the baseline case,
and larger still under warmer conditions, rising to nearly 200%
352 more in the T+6 case. The stronger year-over-year response
under warmer conditions also manifests as a wider distribu-
354 tion of yields (Figure 5). As discussed previously, year-over-
year and climatological responses can differ for many reasons
including memory in the crop model, lurking covariants, and
356 differing associated distributions of daily growing-season daily
358 360

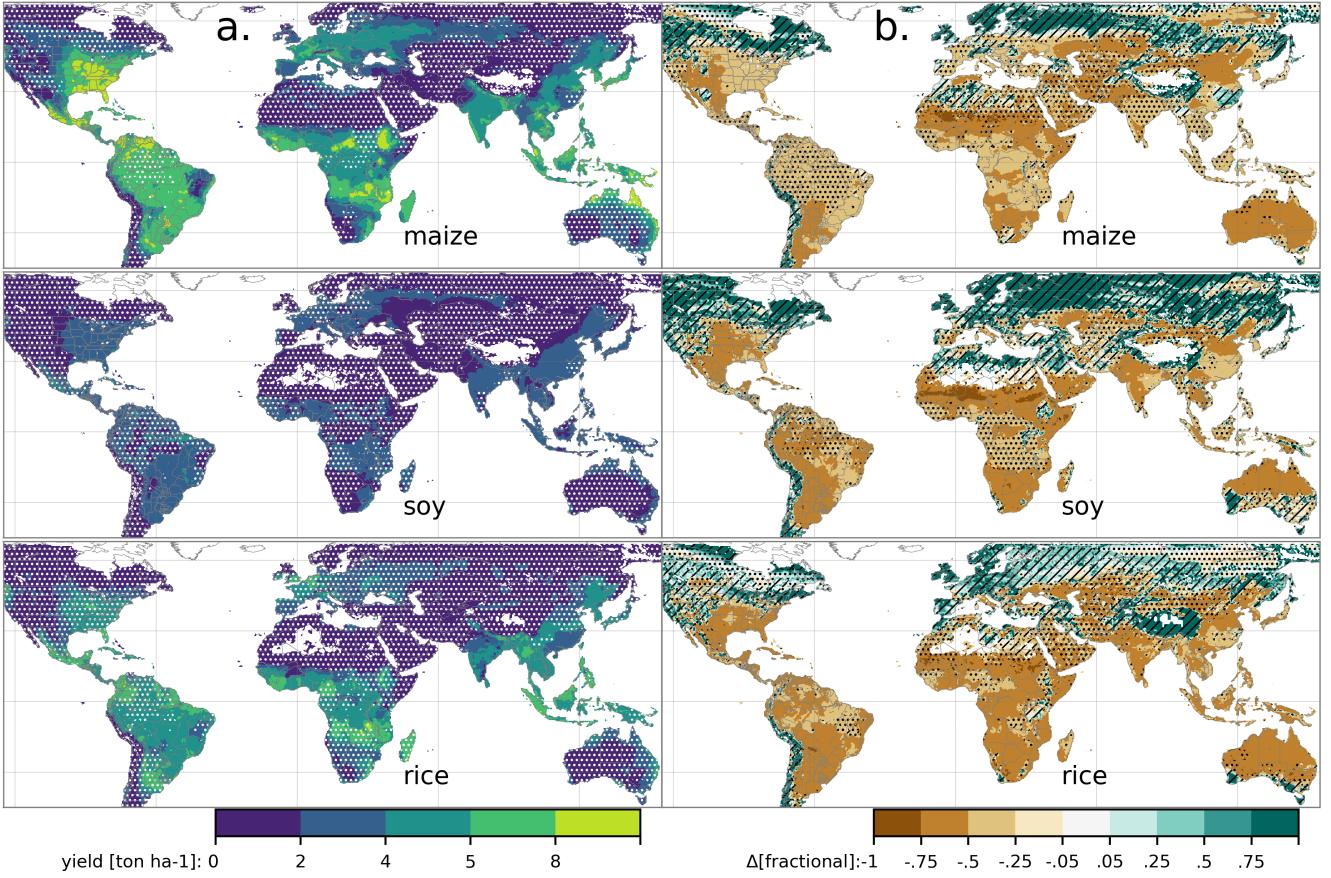


Figure 3: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (Wheat shows a qualitatively similar response, 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.

weather (e.g. Ruane et al., 2016). Note that the GGCMI Phase
362 II datasets do not capture one climatological factor, potential
future distributional shifts, because all simulations are run with
364 fixed offsets from the historical climatology. Prior work has
suggested that mean changes are the dominant drivers of clima-
366 tological crop yield shifts in non-arid regions (e.g. Glotter et al.,
2014).

Emulation involves fitting individual regression models for
368 each crop, simulation model, and 0.5 degree geographic pixel
from the GGCMI Phase II dataset; the regressors are the ap-
370 plied constant perturbations in CO_2 , temperature, water, and
372 nitrogen (C, T, W, N). We regress 30-year climatological mean
yields against a third-order polynomial in C, T, W, and N with

interaction terms. (We aggregate the entire 30-year run in
374 each case to improve signal to noise ratio.) The higher-order
terms are necessary to capture any nonlinear responses, which
are well-documented in observations for temperature and wa-
376 ter perturbations (e.g. Schlenker & Roberts (2009) for T and
He et al. (2016) for W). We include interaction terms (both lin-
378 ear 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
380 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 be-
382 tween water and nitrogen (e.g. Aulakh & Malhi, 2005), and be-

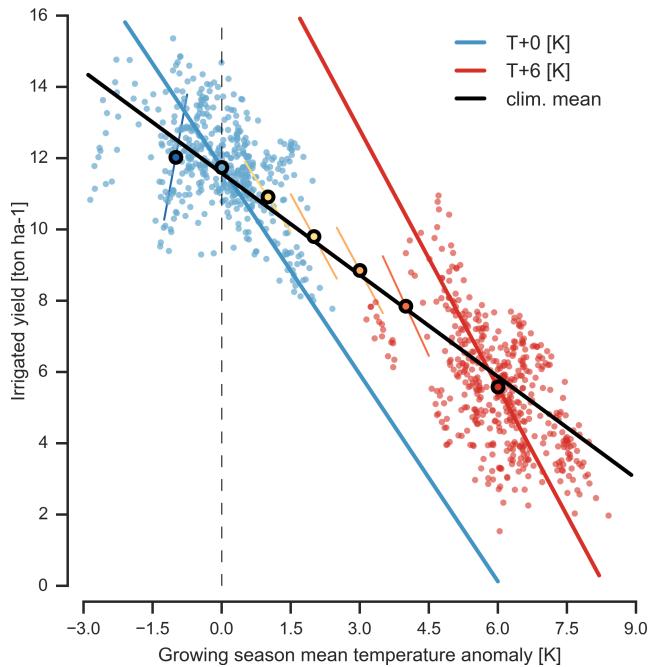


Figure 4: Example showing distinction between crop yield responses to year-to-year and climatological mean temperature shifts. Figure shows irrigated maize for a representative high-yield region (nine adjacent grid cells in northern Iowa) from the pDSSAT model, for the baseline 1981-2010 historical climate (blue) and for the scenario of maximum temperature change (+6 K, red). Other variables are held at baseline values, and the choice of irrigated yields means that precipitation is not a factor. Open black circles mark climatological mean yield values for all six temperature scenarios ($-1, +0, +1, +2, +3, +6$). Colored lines show total least squares linear regressions of year-over-year variations in each scenario. Black line shows the fit through the climatological mean values. Responses to year-over-year temperature variations (colored lines) are 100–200% larger than those to long-term climate perturbations, rising under warmer conditions. Linear fits are shown for illustration purposes and are not used in the emulation models.

tween nitrogen and carbon dioxide (Osaki et al., 1992, Nakamura et al., 1997). To avoid over-fitting or unstable parameter estimation, we apply a feature selection procedure (described below) that reduces the potential 34-term polynomial (for the rain-fed case) to 23 terms.

We do not focus on comparing different functional forms in this study, and instead choose a relatively simple parametrization that allows for some interpretation of coefficients. Some prior studies have used other statistical specifications, e.g. terms in Blanc & Sultan (2015) and Blanc (2017), who borrow information across space by fitting grid points simultaneously across soil region in a panel regression. The simple functional form used here allows emulation at the grid cell level. The emulation therefore indirectly includes any yield response to

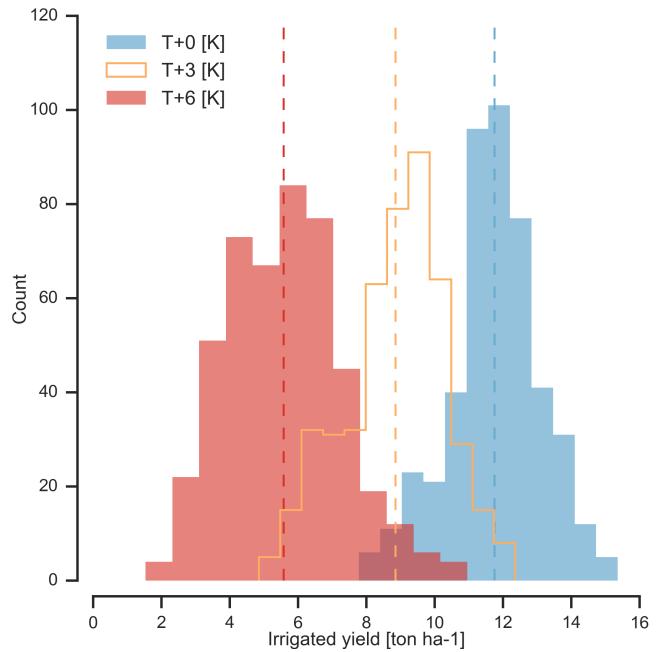


Figure 5: Example showing climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows irrigated maize for nine adjacent high-yield grid cells of Figure 4 from the pDSSAT model, for the baseline 1981-2010 historical climate (blue) and for scenarios with temperature shifted by $T+3$ (orange) and $T+6$ K (red), with other variables held at baseline values. The stronger year-over-year temperature response with higher temperatures seen in Figure 4 is manifested here as larger variance in annual yields even though the variance in climate drivers is identical. In this work we emulate not the year-over-year distributions but the climatological mean response (dashed vertical lines).

geographically distributed factors such as soil type, insolation, and the baseline climate. We hold the statistical specification constant across all crops and models to facilitate parameter by parameter simulation model comparison.

4.1. Feature selection procedure

Although the GGCMI Phase II sampled variable space is large, it is still sufficiently limited that use of the full polynomial expression described above can be problematic. We therefore reduce the number of terms through a feature selection cross-validation process in which terms in the polynomial are tested for importance. In this procedure higher-order and interaction terms are added successively to the regression model; we then follow the reduction of the aggregate mean squared error with increasing terms and eliminate those terms that do not contribute significant reductions. See section 1 in the supplemental

416 documents for more details. We select terms by applying the
 417 feature selection process to three example models that provided
 418 the complete set of 672 rain-fed simulations (pDSSAT, EPIC-
 419 TAMU, and LPJmL); the resulting choice of terms is then ap-
 420 plied for all emulators and all crops.

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 rain-fed crops. (The regressions for irrigated crops do not contain the W terms and the models that do not sample the nitrogen levels omit the N terms).

$$Y = K_1 + K_2C + K_3T + K_4W + K_5N + K_6C^2 + K_7T^2 + K_8W^2 + K_9N^2 + K_{10}CW + K_{11}CN + K_{12}TW + K_{13}TN + K_{14}WN + K_{15}T^3 + K_{16}W^3 + K_{17}TWN + K_{18}T^2W + K_{19}W^2T + K_{20}W^2N + K_{21}N^2C + K_{22}N^2T + K_{23}N^2W \quad (1)$$

To fit the parameters K , we use a Bayesian Ridge proba-
 422 bilistic estimator (MacKay, 1991), which reduces volatility in
 parameter estimates when the sampling is sparse, by weight-
 424 ing parameter estimates towards zero. The Bayesian Ridge
 method is necessary to maintain a consistent functional form

426 across all models and locations. We use the implementation of
 427 the Bayesian Ridge estimator from the scikit-learn package in
 428 Python (Pedregosa et al., 2011). In the GGCMI Phase II ex-
 429 periment, the most problematic fits are those for models that
 430 provided a limited number of cases or for low-yield geographic
 431 regions where some modeling groups did not run all scenar-
 432 ios. We do not attempt to emulate models that provided less
 433 than 50 simulations. The lowest number of simulations emu-
 434 lated across the full parameter space is then 130 (for the PEPIC
 435 model). The resulting parameter matrices for all crop model
 436 emulators are available on request [give location?](#), as are the raw
 437 simulation data and a Python application to emulate yields. The
 438 yield output for a single GGCMI Phase II model that simulates
 439 all scenarios and all five crops is \sim 12.5 GB; the emulator is
 440 \sim 100 MB, a reduction by over two orders of magnitude.

5. Emulation – Results

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

Crop yield responses are geographically diverse, even in
 451 high-yield and high-cultivation areas. Figure 6 illustrates geo-
 452 graphic diversity for a single crop and model (rain-fed maize
 453 in pDSSAT); this heterogeneity supports the choice of emulat-
 454 ing at the grid cell level. Each panel in Figure 6 shows sim-
 455 ulated yield output from scenarios varying only along a single
 456 dimension (CO_2 , temperature, precipitation, or nitrogen addi-

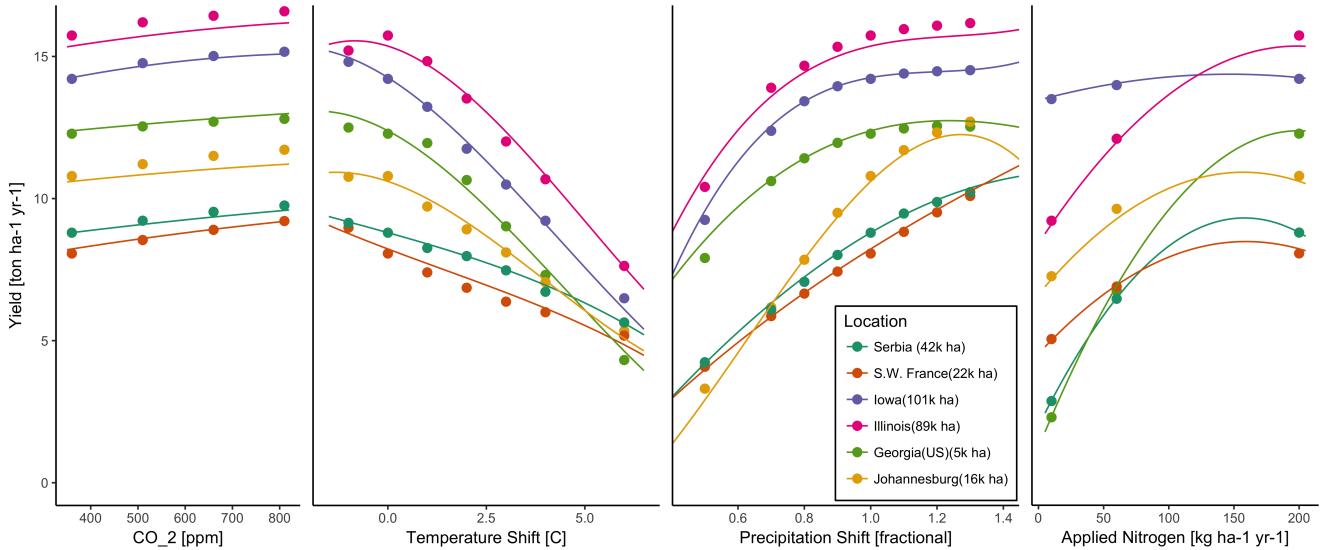


Figure 6: Illustration of spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation along a single variable, with others held at baseline values. Dots show climatological mean yields and lines the results of the full 4D emulator of Equation 1. In general the climatological response surface is sufficiently smooth that it can be represented within the sampled variable space by the simple polynomial used in this work. Extrapolation can however produce misleading results. Nitrogen fits in some cases may not be realistic at intermediate values given limited sampling. For more detailed emulator assessment, see Appendix B.

tion), with other inputs held fixed at baseline levels, compared
460 to the full 4D emulation across the parameter space. Yields
evolve smoothly across the space sampled, and the polynomial
462 fit captures the climatological response to perturbations. Crop
yield responses generally follow similar functional forms across

models, though with a large spread in magnitude partly due to
464 the lack of calibration. Figure 7 illustrates inter-model diversity
466 for a single crop and location (rain-fed maize in northern Iowa,
also shown in Figure 6). Differences in response shape can lead
468 to differences in the fidelity of emulation, though comparison

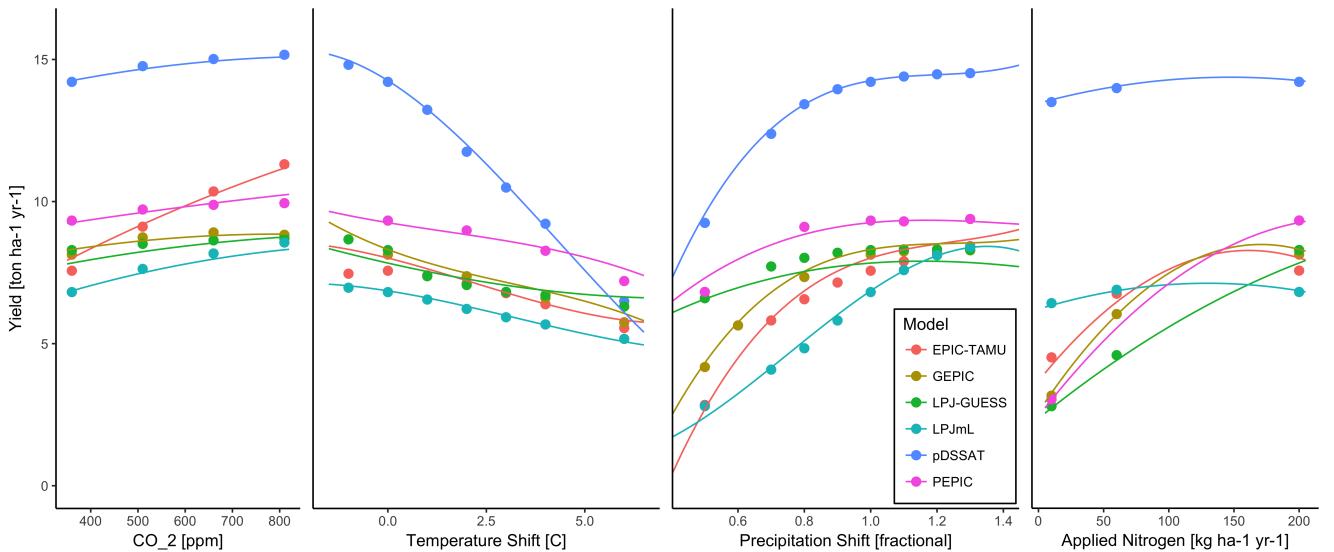


Figure 7: Illustration of across-model variations in yield response. Figures shows simulations and emulations from six models for rain-fed maize in the same Iowa grid cell shown in Figure 6, with the same plot conventions. Models that do not simulate the nitrogen dimension are omitted for clarity. Note that models are uncalibrated, increasing spread in absolute yields. While most model responses can readily emulated with a simple polynomial, some response surfaces are more complicated (e.g. LPJ-GUESS here) and lead to emulation error, though error generally remains small relative to inter-model uncertainty. For more detailed emulator assessment, see Appendix B. As in Figure 6, extrapolation out of the sample space is potentially problematic.

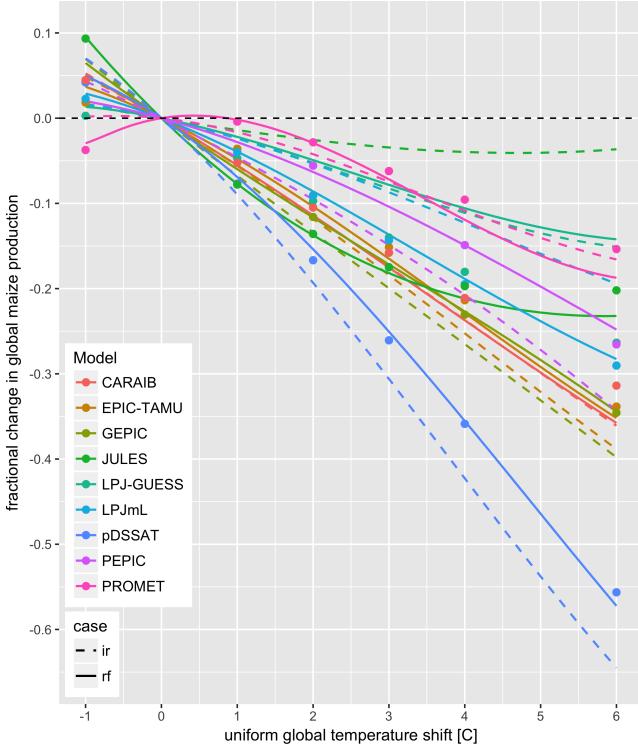


Figure 8: Global emulated damages for maize on currently cultivated lands for the GGCMI Phase II models emulated, for uniform temperature shifts with other inputs held at baseline. (The damage function is created from aggregating up emulated values at the grid cell level, not from a regression of global 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.

here is complicated by the different simulation experiment sampling regimes across models. Note that models are most similar in their responses to temperature perturbations.

While the nitrogen dimension is important, it is also the most problematic to emulate in this work because of its limited sampling. The GGCMI Phase II protocol specified only three nitrogen levels ($10, 60$ and $200 \text{ kg N y}^{-1} \text{ ha}^{-1}$), so a third-order fit would be over-determined but a second-order fit can result in potentially unphysical results. Steep and nonlinear declines in yield with lower nitrogen levels mean that some regressions imply a peak in yield between the 100 and $200 \text{ kg N y}^{-1} \text{ ha}^{-1}$ levels. While it is possible that over-application of nitrogen at the wrong time in the growing season could lead to reduced

yields, these features are potentially an artifact of under sampling. In addition, the polynomial fit cannot capture the well-documented saturation effect of nitrogen application (e.g. Ingstad, 1977) as accurately as would be possible with a non-parametric model.

The emulation fidelity demonstrated here is sufficient to allow using emulated response surfaces to compare model responses and derive insight about impacts projections. Because the emulator or “surrogate model” transforms the discrete simulation sample space into a continuous response surface at any geographic scale, it can be used for a variety of applications, including construction of continuous damage functions. As an example, we show a damage function constructed from the 4D emulation, aggregated to global yield, with simulated values shown for comparison (Figure 8, which shows maize on currently cultivated land; see Figures S16- S19 for other crops and dimensions). The emulated values closely match simulations 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 in space and distributions (Sippel et al., 2015). The global coverage of the GGCMI Phase II simulations allows impacts modelers to apply arbitrary geographically-varying climate projections, as well as arbitrary aggregation masks, to develop damage functions for any climate scenario and any geopolitical or geographic level.

6. Discussion and Conclusions

The GGCMI Phase II experiment provides a database targeted to allow detailed study of crop yields from process-based models under climate change. The systematic input parameter variations are designed to facilitate not only comparing the sensitivities of process-based crop yield models to changing cli-

mate and management inputs but also evaluating the complex interactions between driving 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.

First, the GGCMI Phase II simulations allow identifying major areas of uncertainty. Inter-model uncertainty is qualitatively similar across all four inputs tested at the globally aggregate level with some notable exceptions (see Figure S20). For example, uncertainty in the response to CO_2 and water dominates for wheat and soy is insensitive to nitrogen. 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 (Figure 8). The effect results from geographic differences in cultivation. In any given location, irrigation (or additional rainfall increases) crop resiliency to temperature increase (partly by reducing negative effects from increased evapo-transpiration), 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-fixer is rel-

atively insensitive to nitrogen, and rice is not generally grown in water-limited conditions).

Third, we show that the GGCMI Phase II input parameter sweep allows emulation of the climatological response of crop models with a relatively simple reduced-form statistical model. The systematic parameter sampling in the GGCMI Phase II procedure provides information on the influence of multiple interacting factors in a way that RCP climate model runs cannot, and emulating the resulting response surface then produces a tool that can aid in both physical interpretation of the process-based models and in assessment of agricultural impacts under arbitrary climate scenarios. Emulating the climatological response isolates long-term impacts from any confounding factors that complicate year-over-year changes, and the use of a relatively simple functional forms offer the possibility of physical interpretation of parameter values.

While the GGCMI Phase II database should offer the foundation for multiple future studies, several cautions need to be noted. Because the simulation protocol was designed to focus on change in yield under climate perturbations and not on replicating real-world yields, the models are not formally calibrated so cannot be used for impacts projections unless in used in conjunction with historical data (or data products). Because the GGCMI Phase II simulations apply uniform perturbations to historical climate inputs, they do not sample changes in seasonal variability, and cannot address the additional crop yield impacts of potential changes in climate variability. Although distributional changes in climate model projections are fairly uncertain at present citation needed, follow-on experiments may wish to consider them. Several recent studies have described procedures for generating simulations that combine historical data with model projections of changes not only in temperature and precipitation means but in their marginal distributions or temporal dependence (e.g. Leeds et al. (2015), Pop-

pick et al. (2016), Chang et al. (2016) and Haugen et al. (2018)).

584 The ranges for input perturbations for the historical climatology
585 were selected to represent the range of potential future climatic
586 changes. Using the emulator to extrapolate beyond
these ranges (Table 1) may lead to misinterpretations.

588 Finally, the GGCMI Phase II output dataset invites a broad
range of potential future avenues of analysis. A major target
589 area of research is studying the simulation models themselves
including: a detailed examination of interaction terms between
590 the major input drivers, a robust quantification of the sensitiv-
ity of different models to the input drivers, and comparisons
591 with field-level experimental data. The parameter space tested
in GGCMI Phase II may allow investigations into yield vari-
592 ability and response to extremes under changing management
and CO₂ levels and allow the study of geographic shifts in opti-
593 mal growing regions for different crops. The output dataset also
contains additional runs and variables not analyzed or shown
594 here. Simulations include several which allowed adaptation to
climate changes by altering growing seasons, and additional out-
595 variables include above ground biomass, LAI, and root biomass
(as many as 25 output variables for some models). Emulation
596 studies that are possible include a more systematic evaluation
of different statistical model specifications and formal calcula-
597 tion of uncertainties in derived parameters. The development
600 of multi-model ensembles such as GGCMI Phase II provides a
way to begin to better understand crop responses to a range of
603 potential climate inputs, improve process based models, and ex-
606 plore the potential benefits of adaptive responses included shift-
608 ing growing season, cultivar types and cultivar geographic ex-
tent.

7. Acknowledgments

614 We thank Michael Stein and Kevin Schwarzwald, who pro-
vided helpful suggestions that contributed to this work. This re-

616 search was performed as part of the Center for Robust Decision-
Making on Climate and Energy Policy (RDCEP) at the Univer-
618 sity of Chicago, and was supported through a variety of sources.
620 RDCEP is funded by NSF grant #SES-1463644 through the
Decision Making Under Uncertainty program. J.F. was sup-
622 ported by the NSF NRT program, grant #DGE-1735359. C.M.
was supported by the MACMIT project (01LN1317A) funded
624 through the German Federal Ministry of Education and Re-
search (BMBF). C.F. was supported by the European Research
Council Synergy grant #ERC-2013-SynG-610028 Imbalance-
626 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
628 of the U.S. Department of Energy as part of the Multi-sector
Dynamics Research Program Area. S.O. acknowledges support
from the Swedish strong research areas BECC and MERGE to-
630 gether with support from LUCCI (Lund University Centre for
studies of Carbon Cycle and Climate Interactions). Computing
resources were provided by the University of Chicago Research
632 Computing Center (RCC).

8. Appendix A: Simulations – Assessment

636 The Müller et al. (2017) procedure evaluates response to
year-to-year temperature and precipitation variations in a con-
638 trol run driven by historical climate and compares it to de-
trended historical yields from the FAO (Food and Agriculture
Organization of the United Nations, 2018) by calculating the
640 Pearson product moment correlation coefficient. The procedure
offers no means of assessing CO₂ fertilization, since CO₂ has
been relatively constant over the historical data collection pe-
642 riod. Nitrogen introduces some uncertainty into the analysis,
since the GGCMI Phase II runs impose fixed, uniform nitrogen
644 application levels that are not realistic for individual countries.
We evaluate up to three control runs for each model, since some

modeling groups provide historical runs for three different ni-
650 trogen levels.

Results are similar to those of GGCMI Phase I, with rea-
652 sonable fidelity at capturing year-over-year variation, with dif-
ferences by region and crop stronger than difference between

models. (That is, Figure 9 shows more similarity in horizontal
654 than vertical bars.) No single model is dominant, with each
model providing near best-in-class performance in at least one
location-crop combination. For example, maize in the United
States is consistently well-simulated while maize in Mexico is
658

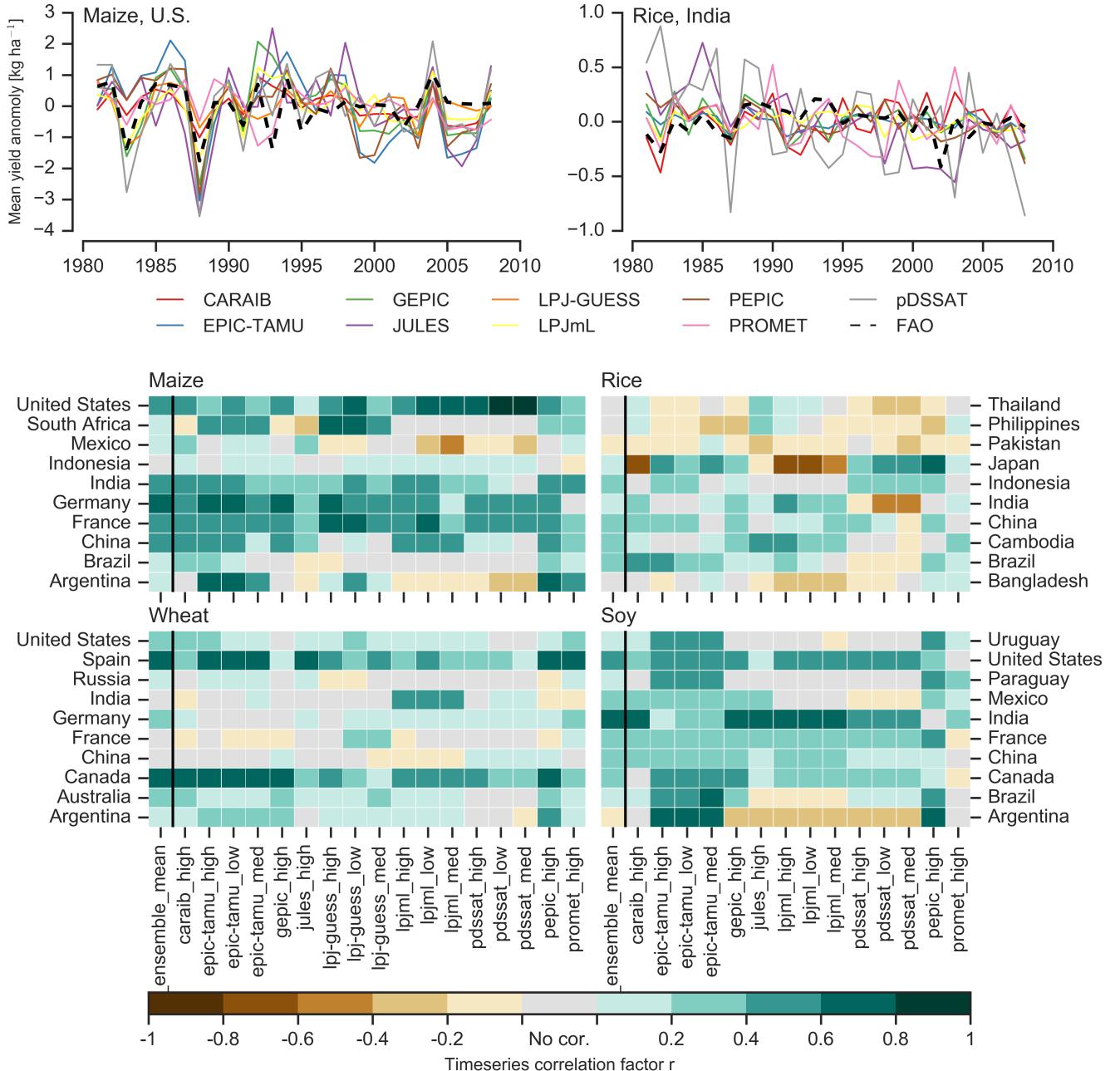


Figure 9: Time series of correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen application case. The heatmaps illustrate the Pearson r correlation coefficient between the detrended simulation mean yield at the country level compared to the detrended FAO yield data for the top producing countries for each crop with continuous FAO data over the 1980–2010 period. Models that provided different nitrogen application levels are shown with low, med, and high label (models that did not simulate different nitrogen levels are analogous to a high nitrogen application level). The ensemble mean yield is also correlated with the FAO data (not the mean of the correlations). Wheat contains both spring wheat and winter wheat simulations. The Pearson r correlation coefficients are similar to those of GGCMI Phase I, with reasonable fidelity at capturing year-over-year variation, with differences by region and crop stronger than difference between models as indicated by more horizontal bars than vertical bars of the same color.

problematic (mean Pearson correlation coefficients of 0.55 and -0.04, respectively). In some cases, especially in the developing world, low correlation coefficients may indicate not model failure but problems in FAO yield data. For example, models have greater apparent skill for rice in India than in the neighboring Pakistan and Bangladesh; this difference may be implausible as solely a model effect. In general, correlation coefficients in GGCMI Phase II are slightly below those of Phase I, likely because of unrealistic nitrogen levels and lack of country level calibration in some models. (Compare Figure 9 to Müller et al. (2017) Figures 1–4 and 6.) Note that in this methodology, simulations of crops with low year-to-year variability such as irrigated rice and wheat will tend to score more poorly than those with higher variability.

Some models do show particular strength for particular crops. For example, the EPIC family of models, and especially the EPIC-TAMU model, perform particularly well for soy across all regions. In other cases a model has particular strength in only certain crop and region combinations. For example, the strongest correlation coefficient in Figure 9 is that for the pDSSAT model for maize in the U.S. (the example crop-model-location used in many example figures in this paper), but pDSSAT slightly under performs for maize in other regions. These model assessment results are similar to those for GGCMI Phase I in Müller et al. (2017).

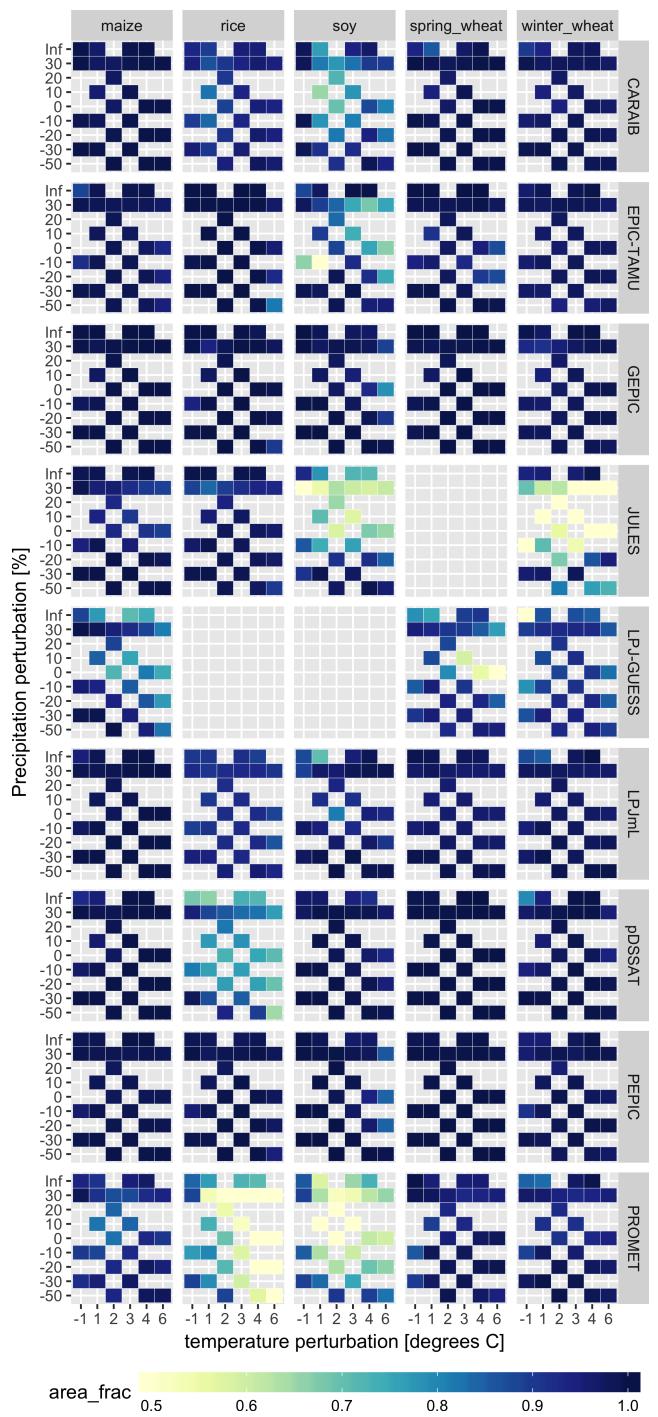


Figure 10: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO₂ and nitrogen held fixed at baseline values. Large columns are crops and large rows models; squares within are T,P scenario pairs. Colors denote the fraction of currently cultivated hectares ('area frac') for each crop with normalized area e less than 1 indicating the the error between the emulation and simulation less than one standard deviation of the ensemble simulation spread. Of the 756 scenarios with these CO₂ and N values, we consider only those for which all 9 models submitted data (Figure S3). 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.

684 9. Appendix B: Emulation – Assessment

No general criteria exist for defining an acceptable crop model emulator. For a multi-model comparison exercise like GGCMI Phase II, one reasonable criterion is what we term the “normalized error”, which compares the fidelity of an emulator for a given model and scenario to the inter-model uncertainty. We define the normalized error e for each scenario as the difference between the fractional yield change from the emulator and

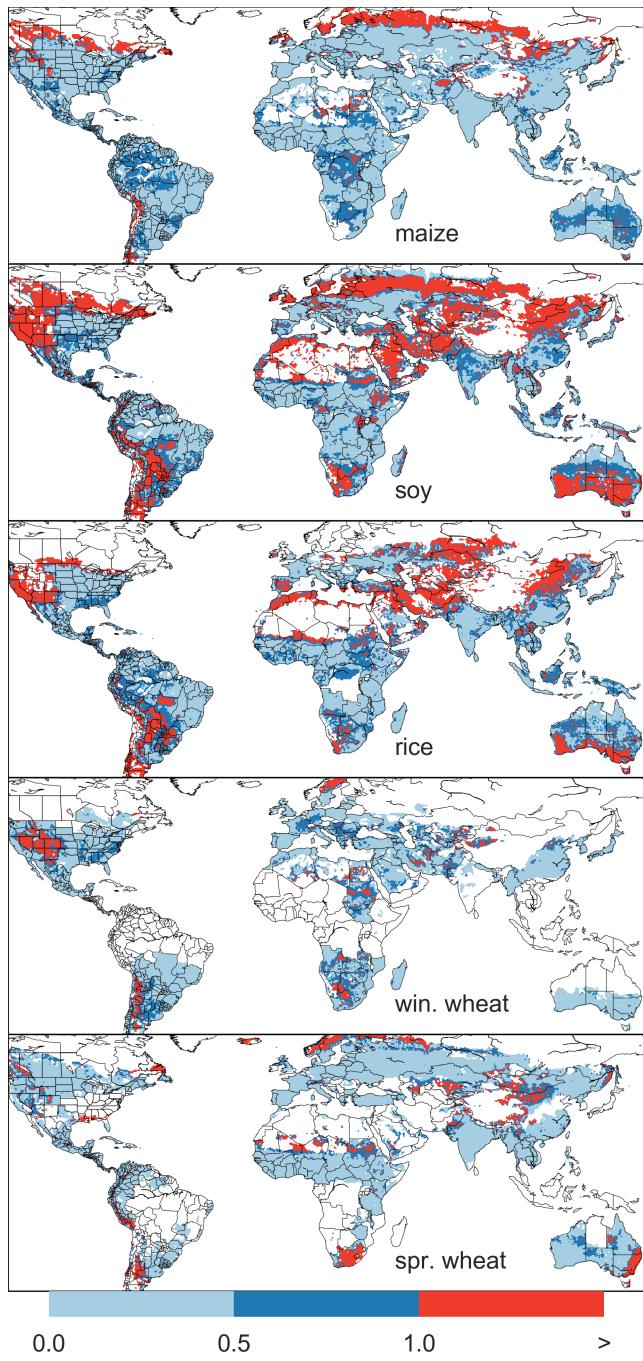


Figure 11: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error e , where $e > 1$ means that emulator error exceeds the multi-model standard deviation. White areas are those where crops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

that in the original simulation, divided by the standard deviation
692 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 a certain setting or 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 change $F_{sim, scn.}$ across all models. The emulator is fitted across all available simulation outputs for each grid cell, model, and crop, and then the error is calculated across the each of the simulation scenarios provided by all nine models (Figure S3).

This metric implies that emulation is generally satisfactory, with several distinct exceptions. Almost all model-crop combination emulators have normalized errors less than one over nearly all currently cultivated hectares (Figure 10), but some individual model-crop combinations are 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), lowering the denominator. Emulator performance often degrades in geographic locations where crops are not currently cultivated. Figure 11 shows a CARAIB model 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 (see also Figure S12).

This assessment procedure is relatively forgiving for several
722 reason. First, each emulation is evaluated against the simulation
actually used to train the emulator. Had we used a spline inter-
724 polation the error would necessarily be zero. Second, the per-
formance metric scales emulator fidelity not by the magnitude
726 of yield changes but by the inter-model spread in those changes.
The normalized error e for a model depends not only on the fi-
728 delity of its emulator in reproducing a given simulation but on
the particular suite of models considered in the intercompari-
730 son exercise. Where models differ more widely, the standard
for emulators becomes less stringent. This effect is readily seen
732 when comparing assessments of emulator performance in sim-
ulations at baseline CO₂ (Figure 10) with those at higher CO₂
734 levels (Figure S13) because models disagree on the magnitude
of CO₂ fertilization. The rationale for this choice of assessment
736 metric is to relate the fidelity of the emulation to an estimate of
true uncertainty, which we take as the multi-model spread. We
738 therefore do not provide a formal parameter uncertainty anal-
ysis, but note that the GGCMI Phase II dataset is well-suited
740 to statistical exploration of emulation approaches and quan-
tification of emulator fidelity. More rigorous emulator assess-
742 ments that could be preformed in future work include: testing
other statistical specifications including non-parametric mod-
744 els, cross-validation procedures where the emulator is trained
on some portion of data and tested on a held-out portion, and
746 calculating standard error on emulator parameters.

10. References

- 748 Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013).
750 Implication of crop model calibration strategies for assessing regional im-
pacts of climate change in europe. *Agric. For. Meteorol.*, *170*, 32–46.
- 752 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. Alder-
754 man, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Biernath, C.,
Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising temperatures reduce
global wheat production. *Nature Climate Change*, *5*, 143–147.
- 756 Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A.,
J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N., Basso,
758 B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath, C., Challinor,
A., Doltra, J., & Wolf, J. (2013). Uncertainty in simulating wheat yields
under climate change. *Nature Climate Change*, *3*, 827832.
- Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other
762 Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Ef-
ficiency, Carbon Sequestration, and Environmental Pollution. *Advances in
Agronomy*, *86*, 341–409.
- Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov,
764 N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat
production potentials and management flexibility under the representative
766 concentration pathways. *Global and Planetary Change*, *122*, 107–121.
- Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields
768 from global gridded crop models. *Agricultural and Forest Meteorology*, *236*,
145–161.
- Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded
770 crop models using statistical estimates. *Agricultural and Forest Meteorol-*
ogy, *214–215*, 134–147.
- von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S.
772 (2018). Implementing the Nitrogen cycle into the dynamic global vegeta-
tion, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific
Model Development*, *11*, 2789–2812.
- Calvin, K., Patel, P., Clarke, L., Asrar, G., Bond-Lamberty, B., Cui, R. Y.,
774 Di Vittorio, A., Dorheim, K., Edmonds, J., Hartin, C., Hejazi, M., Horowitz,
R., Iyer, G., Kyle, P., Kim, S., Link, R., McJeon, H., Smith, S. J., Snyder,
A., Waldhoff, S., & Wise, M. (2019). Gcam v5.1: representing the linkages
776 between energy, water, land, climate, and economic systems. *Geoscientific
Model Development*, *12*, 677–698.
- Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L.,
778 & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections
Based on Precomputed GCM Runs. *Journal of Climate*, *27*, 1829–1844.
- Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N.
780 (2014). A meta-analysis of crop yield under climate change and adaptation.
Nature Climate Change, *4*, 287–291.
- Chang, W., Stein, M., Wang, J., Kotamarthi, V., & Moyer, E. (2016). Changes in
782 spatio-temporal precipitation patterns in changing climate conditions. *Jour-
nal of Climate*, *29*.
- Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process
784 emulation of dynamic computer codes. *Biometrika*, *96*, 663–676.
- Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In
C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth,
Duke University, Durham, North Carolina* (pp. 115–118). Durham, North
786 Carolina.
- Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for
788 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).
- Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdoous,
790 M., & François, L. (2011). Responses of European forest ecosystems to
21st century climate: assessing changes in interannual variability and fire
intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).
- Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,
792 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating
impact models and sectors (pSIMS). *Environmental Modelling and Soft-
ware*, *62*, 509–516.
- Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,
794 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 mod-
eling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, *8*,
796 261–277.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J.,
798 & Taylor, K. E. (2016). Overview of the coupled model intercomparison
project phase 6 (cmip6) experimental design and organization. *Geoscientific
Model Development*, *9*, 1937–1958.
- Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-
799 mate change impacts on durum wheat in the mediterranean region. *Natural
Hazards and Earth System Sciences*, *11*, 1293–1302.
- Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Re-
800 gionalization of a large-scale crop growth model for sub-Saharan Africa:
Model setup, evaluation, and estimation of maize yields. *Agriculture,
Ecosystems & Environment*, *151*, 21–33.
- Food and Agriculture Organization of the United Nations (2018). FAOSTAT
802 database. URL: <http://www.fao.org/faostat/en/home>.
- Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T.,
804 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.,
806

- 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., Strattonovich, 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.
- 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.
- 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.
- 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.
- Haugen, M., Stein, M., Moyer, E., & Srivastava, R. (2018). Estimating changes in temperature distributions in a large ensemble of climate simulations using quantile regression. *Journal of Climate*, 31, 8573–8588.
- 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.
- 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., Harsdorff, 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.
- Leeds, W. B., Moyer, E. J., & Stein, M. L. (2015). Simulation of future climate under changing temporal covariance structures. *Advances in Statistical Climatology, Meteorology and Oceanography*, 1, 1–14.
- 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). GEPIC - modelling wheat yield and crop water productivity with high resolution on a global scale. *Agricultural Systems*, 94, 478 – 493.
- Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a). Global investigation of impacts of PET methods on simulating crop-water relations for maize. *Agricultural and Forest Meteorology*, 221, 164 – 175.
- Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs with yields from major crop cultivations. *Science of The Total Environment*, 572, 526 – 537.
- Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest 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).
- 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 expected future demand without the need for cropland expansion. *Nature Communications*, 6.
- McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srinivas, C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S. (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).
- Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical weather responsiveness of crop yields: US evidence and implications for the agricultural impacts of climate change. *Environmental Research Letters*, 12.
- Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, 8.
- Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Cais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. (2017). Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, 10, 1403–1422.
- Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T. (1997). Effect of 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., Wärnlund, 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-

- 974 12-2489-2015.
- 976 Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in
field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.
- 978 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.
- 980 Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop
yields and their variability at different levels of global warming. *Earth Sys-
tem Dynamics*, 9, 479–496.
- 982 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.
- 984 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).
990 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning
Research*, 12, 2825–2830.
- 994 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., Fran ois, L., Gaiser, T., Hlavinka, P.,
Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet,
J., Minguez, M., Montesino, M., Moriondo, M., M ller, C., Nendel, C.,
998  zt rk, I., Perego, A., Rodr guez, A., Ruane, A., Ruget, F., Sanna, M.,
Semenov, M., Slawinski, C., Strattonovich, P., Supit, I., Waha, K., Wang,
E., Wu, L., Zhao, Z., & R tter, R. (2015). Temperature and precipitation
1000 effects on wheat yield across a European transect: a crop model ensemble
analysis using impact response surfaces. *Climate Research*, 65, 87–105.
- 1002 Poppick, A., McInerney, D. J., Moyer, E. J., & Stein, M. L. (2016). Tempera-
1004 tures in transient climates: Improved methods for simulations with evolving
temporal covariances. *Ann. Appl. Stat.*, 10, 477–505.
- 1006 Porter et al. (IPCC) (2014). Food security and food production systems. Cli-
mate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global
1008 and Sectoral Aspects. Contribution of Working Group II to the Fifth Assess-
ment Report of the Intergovernmental Panel on Climate Change. In C. F.
1010 et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK:
Cambridge University Press.
- 1012 Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of
monthly growing areas of 26 irrigated crops.
- 1014 Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly
Irrigated and Rainfed crop Areas around the Year 2000: A New High-
1016 Resolution Data Set for Agricultural and Hydrological Modeling. *Global
Biogeochemical Cycles*, 24, GB1011.
- 1018 Pugh, T., M ller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E.,
& Arneth, A. (2016). Climate analogues suggest limited potential for inten-
1020 sification of production on current croplands under climate change. *Nature
Communications*, 7, 12608.
- 1022 R ais nen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term cli-
1024 mate change based on a resampling ensemble technique. *Tellus A: Dynamic
Meteorology and Oceanography*, 58, 461–472.
- 1026 Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the
reduction and sensitivity analysis of complex environmental models. *Environ-
1028 mental Modelling & Software*, 34, 1 – 4.
- 1030 Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling
in water resources. *Water Resources Research*, 48.
- 1032 Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017).
Comparing and combining process-based crop models and statistical models
with some implications for climate change. *Environmental Research Letters*,
12.
- 1034 Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., M ller, C., Arneth, A.,
Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek,
F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014).
Assessing agricultural risks of climate change in the 21st century in a global
1038 gridded crop model intercomparison. *Proceedings of the National Academy
of Sciences*, 111, 3268–3273.
- 1040 Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P.,
Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert,
F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural
1042 Model Intercomparison and Improvement Project (AgMIP): Protocols and
pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182.
- Rosenzweig, C., Ruane, A. C., Antle, J., Elliott, J., Ashfaq, M., Chatta,
A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G.,
Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M.,
M ller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M.,
Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., & Wiebe, K.
(2018). Coordinating AgMIP data and models across global and regional
scales for 1.5 C and 2.0 C assessments. *Philosophical Transactions of the
Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 376.
- Ruane, A., I. Hudson, N., Asseng, S., Camarrano, D., Ewert, F., Martre, P.,
J. Boote, K., Thorburn, P., Aggarwal, P., Angulo, C., Basso, B., Bertuzzi,
P., Biernath, C., Brisson, N., Challinor, A., Doltra, J., Gayler, S., Goldberg,
R., Grant, R., & Wolf, J. (2016). Multi-wheat-model ensemble responses to
interannual climate variability. *Environmental Modelling and Software*, 81,
86–101.
- Ruane, A. C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Mason-
D'Croz, D., M ller, 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 C
and +2.0 C global warming using AgMIP Coordinated Global and Regional
Assessments. *Climate Research*, 76, 17–39.
- Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown,
D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013).
Climate change impact uncertainties for maize in panama: Farm informa-
tion, climate projections, and yield sensitivities. *Agricultural and Forest
1066 Meteorology*, 170, 132 – 145.
- Ruane, A. C., Goldberg, R., & Chrissanthacopoulos, J. (2015). Climate forcing
1068 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
1070 analysis for peanut production under climate change: A prototype for the
agmip coordinated climate-crop modeling project (c3mp). *Glob. Change
Biology*, 20, 394–407.
- Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901–
1072 2100 depicted by world maps of the K ppen-Geiger climate classification.
Meteorologische Zeitschrift, 19, 135–141.
- Ruiz-Ramos, M., Ferrise, R., Rodriguez, A., Lorite, I., Bindi, M., Carter, T.,
Fronzek, S., Palosuo, T., Pirttioja, N., Baranowski, P., Buis, S., Cammarano,
D., Chen, Y., Dumont, B., Ewert, F., Gaiser, T., Hlavinka, P., Hoffmann, H.,
Hhn, J., Jurecka, F., Kersebaum, K., Krzyszczak, J., Lana, M., Mechiche-
Alami, A., Minet, J., Montesino, M., Nendel, C., Porter, J., Ruget, F., Se-
menov, M., Steinmetz, Z., Strattonovich, P., Supit, I., Tao, F., Trnka, M.,
de Wit, A., & R tter, R. (2018). Adaptation response surfaces for manag-
ing wheat under perturbed climate and co2 in a mediterranean environment.
Agricultural Systems, 159, 260 – 274.
- Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting
1074 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., M ller, 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
1076 observations and crop models. *Nature Communications*, 8, 13931.
- Schewe, J., Gosling, S. N., Reyer, C., Zhao, F., Ciais, P., Elliott, J., Francois,
L., Huber, V., Lotze, H. K., Seneviratne, S. I., van Vliet, M. T. H., Vautard,
R., Wada, Y., Breuer, L., Bchner, M., Carozza, D. A., Chang, J., Coll, M.,
Deryng, D., de Wit, A., Eddy, T. D., Folberth, C., Frieler, K., Friend, A. D.,
Gerten, D., Gudmundsson, L., Hanasaki, N., Ito, A., Khabarov, N., Kim,
H., Lawrence, P., Morfopoulos, C., Mller, C., Mller Schmied, H., Orth, R.,
Ostberg, S., Pokhrel, Y., Pugh, T. A. M., Sakurai, G., Satoh, Y., Schmid, E.,
Stacke, T., Steenbeek, J., Steinkamp, J., Tang, Q., Tian, H., Tittensor, D. P.,
Volkholz, J., Wang, X., & Warszawski, L. (2019). State-of-the-art global
models underestimate impacts from climate extremes. *Nature Communica-
tions*, 10, 1005–.
- 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.
- Sippel, S., Zscheischler, J., Heimann, M., Otto, F. E. L., Peters, J., & Mahecha,
M. D. (2015). Quantifying changes in climate variability and extremes:
Pitfalls and their overcoming. *Geophysical Research Letters*, 42, 9990–

- 1116 9998.
- 1118 Snyder, A., Calvin, K. V., Phillips, M., & Ruane, A. C. (2018). A crop yield
change emulator for use in gcam and similar models: Persephone v1.0. *Accepted for publication in Geoscientific Model Development*, (pp. 1–42). In
1120 open review.
- 1122 Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implemen-
1124 tation and evaluation of nonparametric regression procedures for sensitivity
analysis of computationally demanding models. *Reliability Engineering &
System Safety*, *94*, 1735 – 1763.
- 1126 Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5
and the experiment design. *Bulletin of the American Meteorological Society*,
93, 485–498.
- 1128 Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of cli-
mate change impacts on global crop yields. *Geophysical Research Letters*,
1130 35.
- 1132 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe,
J. (2014). The Inter-Sectoral Impact Model Intercomparison Project
(ISI-MIP): Project framework. *Proceedings of the National Academy of
Sciences*, *111*, 3228–3232.
- 1134 White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Method-
1136 ologies for simulating impacts of climate change on crop production. *Field
Crops Research*, *124*, 357 – 368.
- 1138 Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T.,
Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance
1140 against site observations of irrigated maize from Mead, Nebraska. *Geosci-
entific Model Development*, *10*, 1291–1320.
- 1142 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield vari-
ability in JULES-crop and implications for forcing with seasonal weather
forecasts. *Geoscientific Model Development*, *8*, 3987–3997.
- 1144 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Land-
1146 bouwkundige Onderzoeken* : 64.6. .
- 1148 Wolf, J., & Oijen, M. (2002). Modelling the dependence of european potato
yields on changes in climate and co2. *Agricultural and Forest Meteorology*,
112, 217 – 231.
- 1150 Wu, X., Vuichard, N., Ciais, P., Viovy, N., de Noblet-Ducoudré, N., Wang,
X., Magliulo, V., Wattenbach, M., Vitale, L., Di Tommasi, P., Moors, E. J.,
1152 Jans, W., Elbers, J., Ceschia, E., Tallec, T., Bernhofer, C., Grünwald, T.,
Moureaux, C., Manise, T., Ligne, A., Cellier, P., Loubet, B., Larmanou, E.,
1154 & Ripoche, D. (2016). ORCHIDEE-CROP (v0), a new process-based agro-
land surface model: model description and evaluation over europe. *Geosci-
entific Model Development*, *9*, 857–873.
- 1156 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao,
Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A.,
Li, T., Lin, E., Liu, Q., Martre, P., Miller, C., Peng, S., Peuelas, J., Ruane,
A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng,
1160 S. (2017). Temperature increase reduces global yields of major crops in four
independent estimates. *Proc. Natl. Acad. Sci.*, *114*, 9326–9331.