

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 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 response of the process-based models in a lightweight, computationally tractable form that facilitates model comparison as well as potential applications in subsequent modeling efforts such as integrated assessment.

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

1. Introduction

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

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

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

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

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

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

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

67 developing world, where future climate impacts may be the¹⁰¹
68 most critical. Finally, only process-based models can capture¹⁰²
69 the growth response to novel conditions and practices that are¹⁰³
70 not represented in historical data (e.g. Pugh et al., 2016, Roberts¹⁰⁴
71 et al., 2017). These novel changes can include the direct fertil-¹⁰⁵
72 ization effect of elevated CO₂, or changes in management prac-¹⁰⁶
73 tices that may ameliorate climate-induced damages.¹⁰⁷

74 Interest has been rising in statistical emulation, which al-¹⁰⁸
75 lows combining advantageous features of both statistical and¹⁰⁹
76 process-based models. The approach involves constructing a¹¹⁰
77 statistical representation or “surrogate model” of complicated¹¹¹
78 numerical simulations by using simulation output as the train-¹¹²
79 ing data for a statistical model (e.g. O’Hagan, 2006, Conti et al.,¹¹³
80 2009). Emulation is particularly useful in cases where sim-¹¹⁴
81 ulations are complex and output data volumes are large, and¹¹⁵
82 has been used in a variety of fields, including hydrology (e.g.¹¹⁶
83 Razavi et al., 2012), engineering (e.g. Storlie et al., 2009),¹¹⁷
84 environmental sciences (e.g. Ratto et al., 2012), and climate¹¹⁸
85 (e.g. Castruccio et al., 2014, Holden et al., 2014). For agri-¹¹⁹
86 cultural impacts studies, emulation of process-based models¹²⁰
87 allows capturing key relationships between input variables in¹²¹
88 a lightweight, flexible form that is compatible with economic¹²²
89 studies.¹²³

90 In the past decade, multiple studies have developed emula-¹²⁴
91 tors of process-based crop simulations. Early studies proposing¹²⁵
92 or describing potential crop yield emulators include Howden¹²⁶
93 & Crimp (2005), Räisänen & Ruokolainen (2006), Lobell &¹²⁷
94 Burke (2010), and Ferrise et al. (2011), who used a machine¹²⁸
95 learning approach to predict Mediterranean wheat yields. Stud-¹²⁹
96 ies developing single-model emulators include Holzkämper¹³⁰
97 et al. (2012) for the CropSyst model, Ruane et al. (2013) for¹³¹
98 the CERES wheat model, and Oyebamiji et al. (2015) for the¹³²
99 LPJmL model (for multiple crops, using multiple scenarios as¹³³
100 a training set). More recently, emulators have begun to be used¹³⁴

in the context of multi-model intercomparisons, with Blanc &
Sultan (2015), Blanc (2017), Ostberg et al. (2018) and Mis-
try et al. (2017) using them to analyze the five crop models
of the Inter-Sectoral Impacts Model Intercomparison Project
(ISIMIP) (Warszawski et al., 2014), which simulated yields for
maize, soy, wheat, and rice. Choices differ: Blanc & Sul-
tan (2015) and Blanc (2017) base their emulation on histori-
cal simulations and a single future climate/emissions scenario
(RCP8.5), and use local weather variables and yields in their
regression but then aggregate across broad regions; Ostberg
et al. (2018) consider multiple future climate scenarios, using
global mean temperature change (and CO₂) as regressors but
then pattern-scale to emulate local yields; while Mistry et al.
(2017) attempt only to capture observed historical yields, using
local weather data and a historical crop simulation. These ef-
forts do share important common features: all emulate annual
crop yields across the entire scenario or scenarios, and when fu-
ture scenarios are considered, they are non-stationary, i.e. their
input climate parameters evolve over time.

An alternative approach is to construct a training set of multi-
ple stationary scenarios in which parameters are systematically
varied. Such a “parameter sweep” offers several advantages for
emulation over scenarios in which climate evolves over time.
First, it allows separating the effects of different variables that
impact yields but that are highly correlated in realistic future
scenarios (e.g. CO₂ and temperature). Second, it allows making
a distinction between year-over-year yield variations and cli-
matological changes, which may involve different responses to
the particular climate regressors used (e.g. Ruane et al., 2016).
For example, if year-over-year yield variations are driven pre-
dominantly by variations in the distribution of temperatures
throughout the growing season, and long-term climate changes
are driven predominantly by shifts in means, then regressing
on the mean growing season temperature will produce different

135 yield responses at annual vs. climatological timescales.

136 Systematic parameter sweeps have begun to be used in
137 crop modeling, with early efforts in 2015 (Makowski et al.,
138 2015, Pirttioja et al., 2015), and several recent studies in 2018
139 (Fronzek et al., 2018, Snyder et al., 2018, Ruiz-Ramos et al.,
140 2018). All three studies sample multiple perturbations to tem-
141 perature and precipitation (with Snyder et al. (2018) and Ruiz-
142 Ramos et al. (2018) adding CO₂ as well), in 132, 99 and approx-
143 imately 220 different combinations, respectively, and take ad-
144 vantage of the structured training set to construct emulators of
145 climatological mean yields, omitting year-over-year variations.
146 The main limitation in these studies is geographic extent, with
147 each study focusing on a limited number of sites, and Fronzek
148 et al. (2018) and Ruiz-Ramos et al. (2018) simulating only wheat
149 (over several models), while Snyder et al. (2018) analyzes four
150 crops but with a single model (GCAM).

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

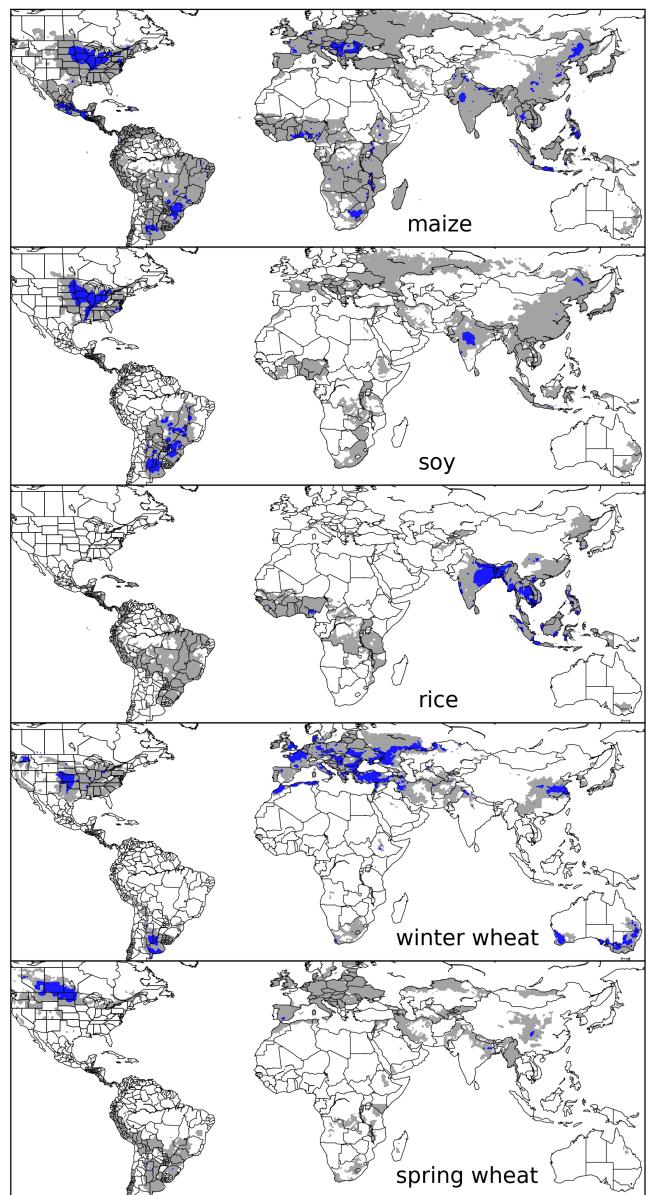


Figure 1: Presently cultivated area for rain-fed crops. Blue indicates grid cells with more than 20,000 hectares (~10% of the equatorial grid cells) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated areas for maize, rice, and soy are taken from the MIRCA2000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

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

175 drivers; and testing strategies for producing lightweight em-²⁰¹
 176 ulators of process-based models. In this paper, we describe²⁰²
 177 the GGCMI Phase II experiments, present initial results, and²⁰³
 178 demonstrate that it is tractable to emulation.²⁰⁴

179 2. Simulation – Methods

180 GGCMI Phase II is the continuation of a multi-model com-²⁰⁷
 181 parison exercise begun in 2014. The initial Phase I compared²⁰⁸
 182 harmonized yields of 21 models for 19 crops over a 31-year²⁰⁹
 183 historical (1980-2010) scenario with a primary goal of model²¹⁰
 184 evaluation (Elliott et al., 2015, Müller et al., 2017). Phase II²¹¹
 185 compares simulations of 12 models for 5 crops (maize, rice,²¹²
 186 soybean, spring wheat, and winter wheat) over the same histor-²¹³
 187 ical time series (1980-2010) used in Phase I, but with individ-²¹⁴
 188 ual climate or management inputs adjusted from their historical²¹⁵
 189 values. The reduced set of crops includes the three major global²¹⁶
 190 cereals and the major legume and accounts for over 50% of hu-²¹⁷
 191 man calories (in 2016, nearly 3.5 billion tons or 32% of total²¹⁸
 192 global crop production by weight (Food and Agriculture Orga-²¹⁹
 193 nization of the United Nations, 2018).

194 The guiding scientific rationale of GGCMI Phase II is to pro-²²⁰
 195 vide a comprehensive, systematic evaluation of the response²²¹
 196 of process-based crop models to different values for carbon²²²
 197 dioxide, temperature, water, and applied nitrogen (collectively²²³
 198 known as “CTWN”). The dataset is designed to allow re-²²⁴
 199 searchers to:

- 200 • Enhance understanding of how models work by character-²²⁷

izing their sensitivity to input climate and nitrogen drivers.

- Study the interactions between climate variables and nitro-¹²⁰
 121 gen inputs in driving modeled yield impacts.
- Explore differences in crop response to warming across the¹²²
 123 Earth’s climate regions.
- Provide a dataset that allows statistical emulation of crop¹²⁴
 125 model responses for downstream modelers.
- Illustrate differences in potential adaptation via growing¹²⁶
 127 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, soil water is held at either field capacity or, for those models that include water-log damage, at maximum beneficial level. Temperature perturbations are applied as absolute offsets from the daily mean, minimum, and maximum temperature time series for each grid cell used as inputs. Precipitation perturbations are applied as fractional changes at the grid cell level, and carbon dioxide and nitrogen levels are specified as discrete values applied uniformly over all grid cells. Note that CO₂ changes are applied independently of changes in climate variables, so that higher CO₂ is not associated with higher temperatures. An additional, identical set of scenarios (at the same C, T, W, and N levels) not shown or analyzed here simulate adaptive agronomy under climate change by varying the growing season for crop production. The resulting GGCMI

Input variable	Abbr.	Tested range	Unit
CO 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. * Only simulated by one model. W-percentage does not apply to the irrigated (W_{inf}) simulations, which are all simulated at the maximum beneficial levels of water.

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

Table 2: Models included in GGCMI Phase II and the number of C, T, W, and N simulations that each performs for rain-fed crops (“Sims per Crop”), with 672 as the maximum. “N-Dim.” indicates whether the simulations include varying nitrogen levels. Two models provide only one nitrogen level, and †PROMET provides only two of the three nitrogen levels (and so is not emulated across the nitrogen dimension). All models provide the same set of simulations across all modeled crops, but some omit individual crops. (For example, APSIM does not simulate winter wheat.) Irrigated simulations are provided at the level of the other covariates for each model, i.e. an additional 84 simulations for fully-sampled models. Simulations are nearly global but their geographic extent can vary, even for different crops in an individual model, since some simulations omit regions far outside the currently cultivated area. In most cases, historical daily climate inputs are taken from the 0.5 degree NASA AgMERRA daily gridded re-analysis product specifically designed for agricultural modeling, with satellite-corrected precipitation (Ruane et al., 2015), but two models (marked with *) require sub-daily input data and use alternative sources. See Elliott et al. (2015) for additional details.

228 Phase II dataset captures a distribution of crop responses over²⁴⁶ to nitrogen, temperature, and water (e.g. alkalinity and salinity).
 229 the potential space of future climate conditions.²⁴⁷ No additional nitrogen inputs, such as atmospheric deposition,
 230 The 12 models included in GGCMI Phase II are all mech-²⁴⁸ are considered, but some model treatment of soil organic matter
 231 anistic process-based crop models that are widely used in im-²⁴⁹ may allow additional nitrogen release through mineralization.
 232 pacts assessments (Table 2). Although some models share a²⁵⁰ See Rosenzweig et al. (2014), Elliott et al. (2015) and Müller
 233 common base (e.g. the LPJ family or the EPIC family of mod-²⁵¹ et al. (2017) for further details on models and underlying as-
 234 els), they have subsequently developed independently. (For²⁵² sumptions.
 235 more details on model genealogy, see Figure S1 in Rosenzweig²⁵³ The participating modeling groups provide simulations at
 236 et al. (2014).) Differences in model structure mean that several²⁵⁴ any of four initially specified levels of participation, so the num-
 237 key factors are not standardized across the experiment, includ-²⁵⁵ ber of simulations varies by model, with some sampling only a
 238 ing secondary soil nutrients, carry-over effects across growing²⁵⁶ part of the experiment variable space. Most modeling groups
 239 years including residue management and soil moisture, and the²⁵⁷ simulate all five crops in the protocol, but some omitted one
 240 extent of simulated area for different crops. Growing seasons²⁵⁸ or more. Table 2 provides details of coverage for each model.
 241 are standardized across models (with assumptions based on²⁵⁹ Note that the three models that provide less than 50 simulations
 242 Sacks et al. (2010) and Portmann et al. (2008, 2010)), but vary²⁶⁰ are excluded from the emulator analysis.
 243 by crop and by location on the globe. For example, maize is²⁶¹ Each model is run at 0.5 degree spatial resolution and cov-
 244 sown in March in Spain, in July in Indonesia, and in December²⁶² ers all currently cultivated areas and much of the uncultivated
 245 in Namibia. All stresses are disabled other than factors related²⁶³ land area. (See Figure 1 for the present-day cultivated area of

rain-fed crops, and Figure S1 in the Supplemental Material for₂₈₀
irrigated crops.) Coverage extends considerably outside cur-
rently cultivated areas because cultivation will likely shift under₂₈₁
climate change. However, areas are not simulated if they are₂₈₂
assumed to remain non-arable even under an extreme climate₂₈₃
change; these regions include Greenland, far-northern Canada,₂₈₄
Siberia, Antarctica, the Gobi and Sahara Deserts, and central₂₈₅
Australia.₂₈₆

All models produce as output crop yields (tons ha^{-1} year $^{-1}$)₂₈₈
for each 0.5 degree grid cell. Because both yields and yield₂₈₉
changes vary substantially across models and across grid cells,₂₉₀
we primarily analyze relative change from a baseline. We take₂₉₁
as the baseline the scenario with historical climatology (i.e. T₂₉₂
and P changes of 0), C of 360 ppm, and applied N at 200 kg₂₉₃
 ha^{-1} . We show absolute yields in some cases to illustrate geo-₂₉₄
graphic differences in yields for a single model.₂₉₅

3. Simulation – Results

Crop models in the GGCMI Phase II ensemble show broadly consistent responses to climate and management perturbations in most regions, with a strong negative impact of increased temperature in all but the coldest regions. We illustrate this result for rain-fed maize in Figure 2, which shows yields for the primary Köppen-Geiger climate regions (Rubel & Kottek, 2010).

In warming scenarios, models show decreases in maize yield in the temperate, tropical, and arid regions that account for nearly three-quarters of global maize production. These impacts are robust for even moderate climate perturbations. In the temperate zone, even a 1 degree temperature rise with other variables held fixed leads to a median yield reduction that outweighs the variance across models. A 6 degree temperature rise results in median loss of ~25% of yields with a signal to noise of nearly three. A notable exception is the cold continental region, where

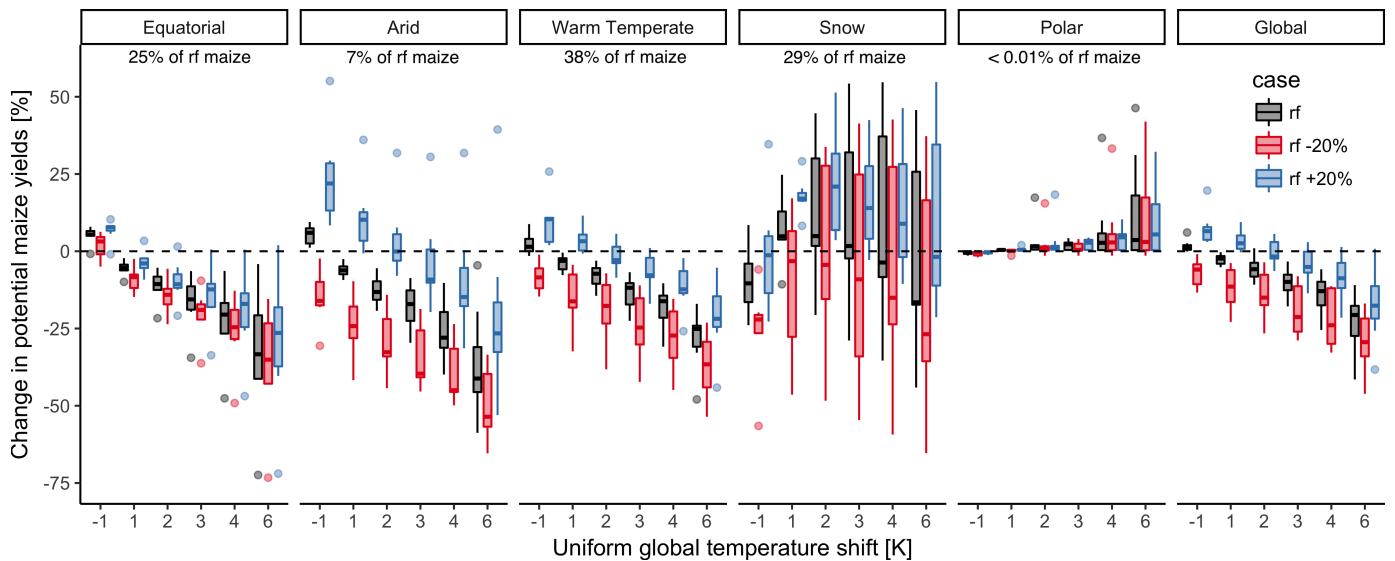


Figure 2: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation levels (rain-fed (rf) -20%, rain-fed (0), and rain-fed +20%), with CO₂ and nitrogen held constant at baseline values (360 ppm and 200 kg ha^{-1} yr^{-1}). Y-axis is fractional change in the regional average climatological potential yield relative to the baseline. The figure shows all modeled land area; see Figure S6 in the supplemental material for only currently-cultivated land. Panel text gives the percentage of rain-fed maize presently cultivated in each climate zone (data from Portmann et al., 2010). Box-and-whiskers plots show distribution across models, with median marked. Box edges are first and third quartiles, i.e. box height is the interquartile range (IQR). Whiskers extend to maximum and minimum of simulations but are limited at 1.5-IQR; otherwise the outlier is shown. Models generally agree in most climate regions (other than cold continental), with projected changes larger than inter-model variance. Outliers in the tropics (strong negative impact of temperature increases) are the pDSSAT model; outliers in the high-rainfall case (strong positive impact of precipitation increases) are the JULES model. Inter-model variance increases in the case where precipitation is reduced, suggesting uncertainty in model response to water limitation. The right panel with global changes shows yield responses to an globally uniform temperature shift; note that these results are not directly comparable to simulations of more realistic climate scenarios with the same global mean change.

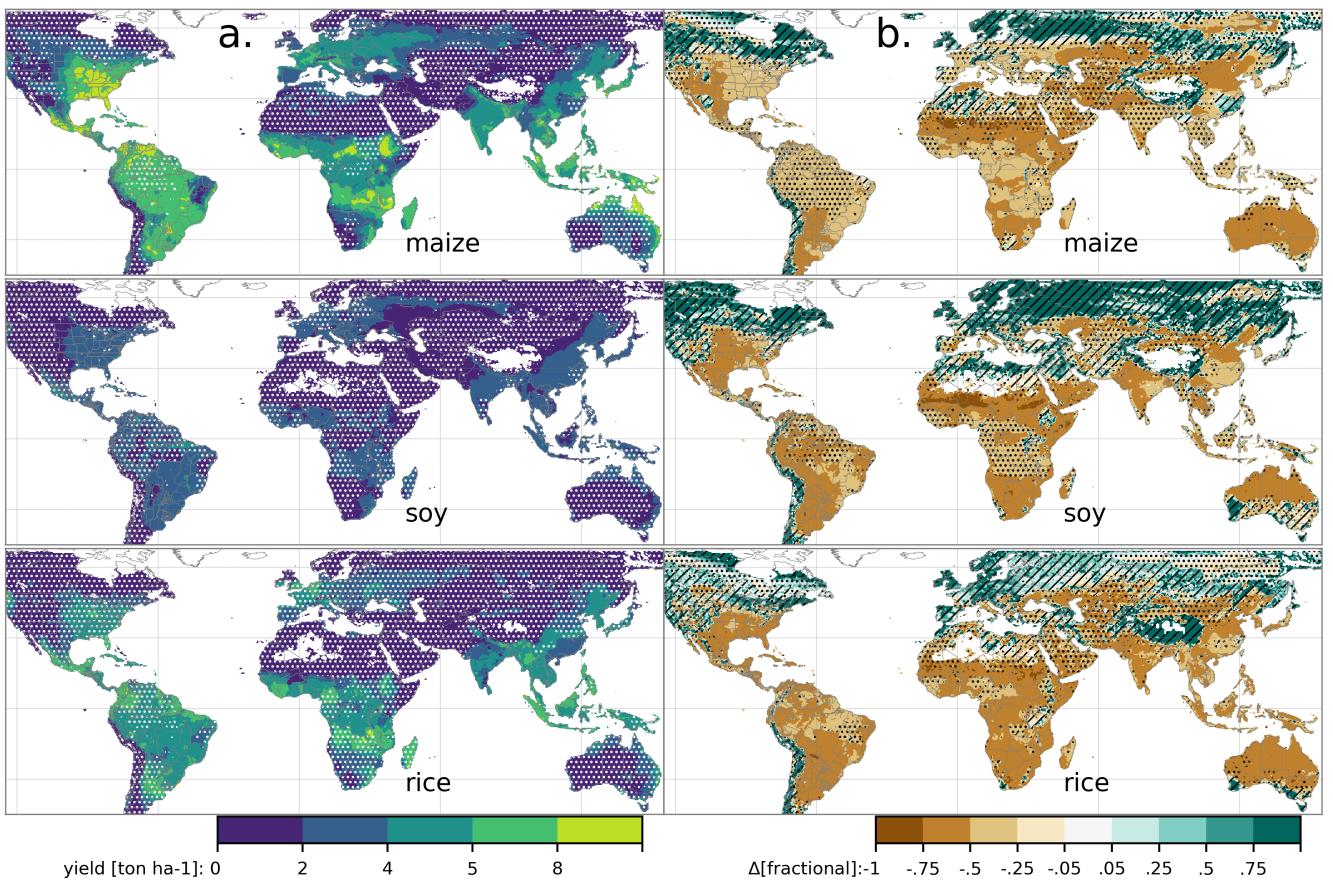


Figure 3: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top–bottom) rain-fed maize, soy, and rice. (For wheat see Figure S11 in the supplemental material.) White stippling indicates areas where these crops are not currently cultivated. Absence of cultivation aligns well with the lowest yield contour ($0\text{--}2 \text{ ton ha}^{-1}$). 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.

296 models disagree strongly, extending even to the sign of impacts.³⁰⁹
 297 Other crops show similar responses to warming, with robust³¹⁰
 298 yield losses in warmer locations and high inter-model variance³¹¹
 299 in the cold continental regions (Figure S7).

300 The effects of rainfall changes on maize yields shown in Fig-³¹³
 301 ure 2 are also as expected and are consistent across models.³¹⁴
 302 Increased rainfall mitigates the negative effect of higher tem-³¹⁵
 303 peratures, most strongly in arid regions. Decreased rainfall³¹⁶
 304 amplifies yield losses and also increases inter-model variance³¹⁷
 305 more strongly, suggesting that models have difficulty represent-³¹⁸
 306 ing crop response to water stress XX - see reviewer comments?³¹⁹
 307 We show only rain-fed maize here; see Figure S5 for the irri-³²⁰
 308 gated case. As expected, irrigated crops are more resilient to³²¹

temperature increases in all regions, especially so where water is limiting.

Mapping the distribution of baseline yields and yield changes shows the geographic dependencies that underlie these results.

Figure 3 shows baseline and changes in the $T+4$ scenario for rain-fed maize, soy, and rice in the multi-model ensemble mean, with locations of model agreement marked. Absolute yield potentials show strong spatial variation, with much of the Earth's surface area unsuitable for any given crop. In general, models agree most on yield response in regions where yield potentials are currently high and therefore where crops are currently grown. Models show robust decreases in yields at low latitudes, and highly uncertain median increases at most high latitudes.

322 For wheat crops see Figure S11; wheat projections are both
 323 more uncertain and show fewer areas of increased yield in the
 324 inter-model mean.

325 4. Emulation – Methods

326 As part of our demonstration of the properties of the GGCMI
 327 Phase II dataset, we construct an emulator of 30-year climato-
 328 logical mean yields. This approach is made possible by
 329 the structured set of simulations involving systematic per-
 330 turbations. In the GGCMI Phase II dataset, the year-over-year re-
 331 sponds are generally quantitatively distinct from (and larger
 332 than) climatological mean responses. In the example of Figure
 333 4, responses to year-over-year temperature variations are **XX%**
 334 larger than those to long-term perturbations in the baseline case,
 335 and larger still under warmer conditions, rising to **XX%** in the
 336 T+6 case. The stronger year-over-year response under warmer
 337 conditions also manifests as a wider distribution of yields (Figure
 338 5). As discussed previously, year-over-year and climatolog-
 339 ical responses can differ for many reasons including memory
 340 in the crop model, lurking covariants, and differing associated
 341 distributions of daily growing-season daily weather (e.g. Ruane
 342 et al., 2016). Note that the GGCMI Phase II datasets do not
 343 capture one climatological factor, potential future distributional
 344 shifts, because all simulations are run with fixed offsets from
 345 the historical climatology. Prior work has suggested that mean
 346 changes are the dominant drivers of climatological crop yield
 347 shifts in non-arid regions (e.g. Glotter et al., 2014).

348 Emulation involves fitting individual regression models for
 349 each crop, simulation model, and 0.5 degree geographic pixel
 350 from the GGCMI Phase II dataset; the regressors are the applied
 351 constant perturbations in CO₂, temperature, water, and nitro-
 352 gen (C,T,W,N). We regress 30-year climatological mean yields
 353 against a third-order polynomial in C, T, W, and N with inter-
 354 action terms. The higher-order terms are necessary to capture

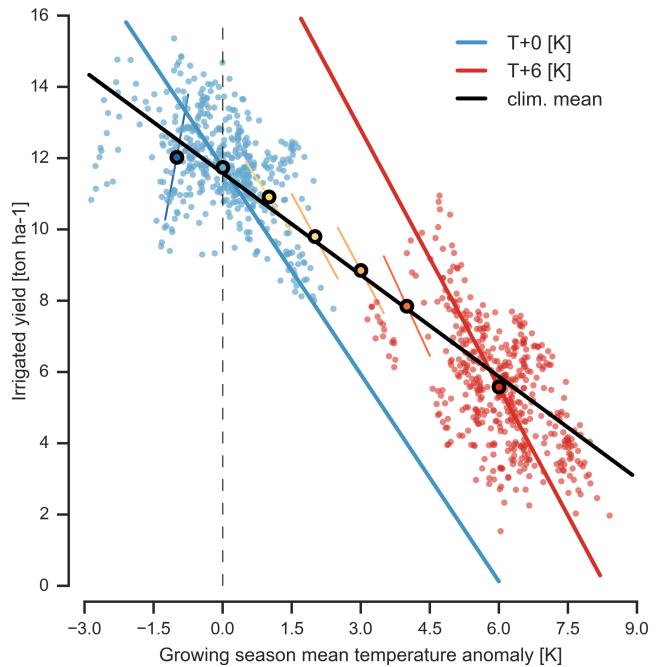


Figure 4: Example showing distinction between crop yield responses to year-to-year and climatological mean temperature shifts. Figure shows irrigated maize for a representative high-yield region (nine adjacent grid cells in northern Iowa) from the pDSSAT model, for the baseline 1981-2010 historical climate (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 (T-1, +0, +1, +2, +3, +4, +6). Colored lines show total least squares linear regressions of year-over-year variations in each scenario. Black line shows the fit through the climatological mean values. Responses to year-over-year temperature variations (colored lines) are **XX–XX%** larger than those to long-term climate perturbations, rising under warmer conditions.

any nonlinear responses, which are well-documented in obser-
 vations for temperature and water perturbations (e.g. Schlenker
 & Roberts (2009) for T and He et al. (2016) for W). We include
 interaction terms (both linear and higher-order) because past
 studies have shown them to be significant effects. For example,
 Lobell & Field (2007) and Tebaldi & Lobell (2008) showed that
 in real-world yields, the joint distribution in T and W is needed
 to explain observed yield variance. (C and N are fixed in these
 data.) Other observation-based studies have shown the impor-
 tance of the interaction between water and nitrogen (e.g. Aulakh
 & Malhi, 2005), and between nitrogen and carbon dioxide (Os-
 aki et al., 1992, Nakamura et al., 1997). To avoid overfitting
 or unstable parameter estimation, we apply a feature selection
 procedure (described below) that reduces the potential 34-term

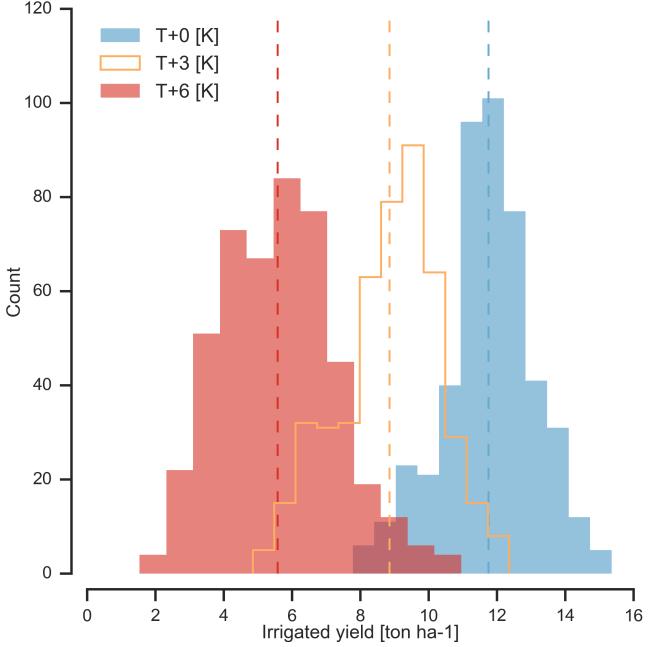


Figure 5: Example showing climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows irrigated maize for nine adjacent high-yield grid cells of Figure 4 (in northern Iowa, same as Figure 4) from the pDSSAT model, for the baseline 1981–2010 historical climate (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. In this work we emulate not the year-over-year distributions but the climatological mean response (dashed vertical lines).

polynomial (for the rain-fed case) to 23 terms.

382 4.1. Feature selection procedure

383 Although the GGCMI Phase II sampled variable space is
 384 large, it is still sufficiently limited that use of the full polynomial
 385 expression described above can be problematic. We therefore
 386 reduce the number of terms through a feature selection cross-
 387 validation process in which terms in the polynomial are tested
 388 for importance. In this procedure higher-order and interaction
 389 terms are added successively to the model; we then follow the
 390 reduction of the aggregate mean squared error with increasing
 391 terms and eliminate those terms that do not contribute signif-
 392 icant reductions. See supplemental documents for more de-
 393 tails. We select terms by applying the feature selection pro-
 394 cess to the three models that provided the complete set of 672
 395 rain-fed simulations (pDSSAT, EPIC-TAMU, and LPJmL); the
 396 resulting choice of terms is then applied for all emulators.

370 We do not focus on comparing different functional forms in
 371 this study, and instead choose a relatively simple parametriza-
 372 tion that allows for some interpretation of coefficients. Some
 373 prior studies have used more complex functional forms and
 374 larger numbers of parameters, e.g. 39 in Blanc & Sultan (2015)
 375 and Blanc (2017), who borrow information across space by fit-
 376 ting grid points simultaneously across a large region in a panel
 377 regression. The simpler functional form used here allows em-
 378 ulation at grid-cell level **with low noise? how do you quantify**
 379 **this?** The emulation therefore indirectly includes any yield re-
 380 sponse to geographically distributed factors such as soil type,
 381 insolation, and the baseline climate itself.

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

used for all grid cells, models, and crops:

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

To fit the parameters K , we use a Bayesian Ridge probabilistic estimator (MacKay, 1991), which reduces volatility in parameter estimates when the sampling is sparse, by weighting parameter estimates towards zero. The Bayesian Ridge method is necessary to maintain a consistent functional form across all models and locations. We use the implementation of the Bayesian Ridge estimator from the scikit-learn package in Python (Pedregosa et al., 2011). In the GGCMI Phase II experiment, the most problematic fits are those for models that

provided a limited number of cases or for low-yield geographic regions where some modeling groups did not run all scenarios. We do not attempt to emulate models that provided less than 50 simulations. The lowest number of simulations emulated across the full parameter space is then 130 (for the PEPIC model). The resulting parameter matrices for all crop model emulators are available on request [give location?](#), as are the raw simulation data and a Python application to emulate yields. The yield output for a single GGCMI Phase II model that simulates all scenarios and all five crops is ~ 12.5 GB; the emulator is ~ 100 MB, a reduction by over two orders of magnitude.

5. Emulation – Results

Emulation provides not only a computational tool but a means of understanding and interpreting crop yield response across the parameter space. Emulation is only possible when crop yield responses are sufficiently smooth and continuous to allow fitting with a relatively simple functional form, but this condition largely holds in the GGCMI Phase II simulations. Re-

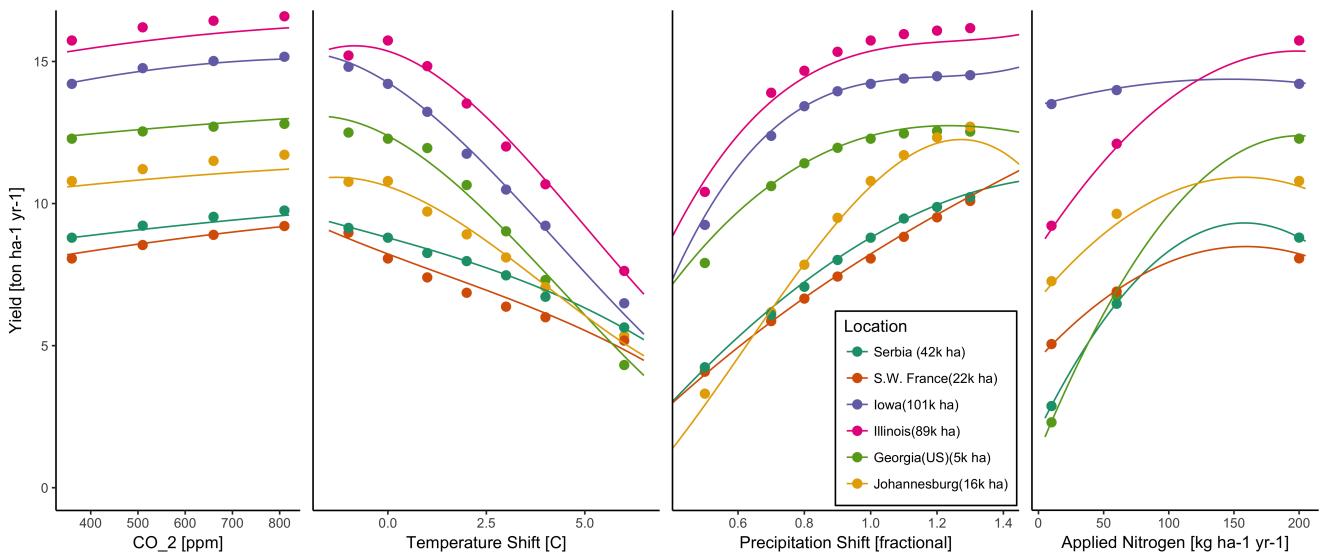


Figure 6: Illustration of spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation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

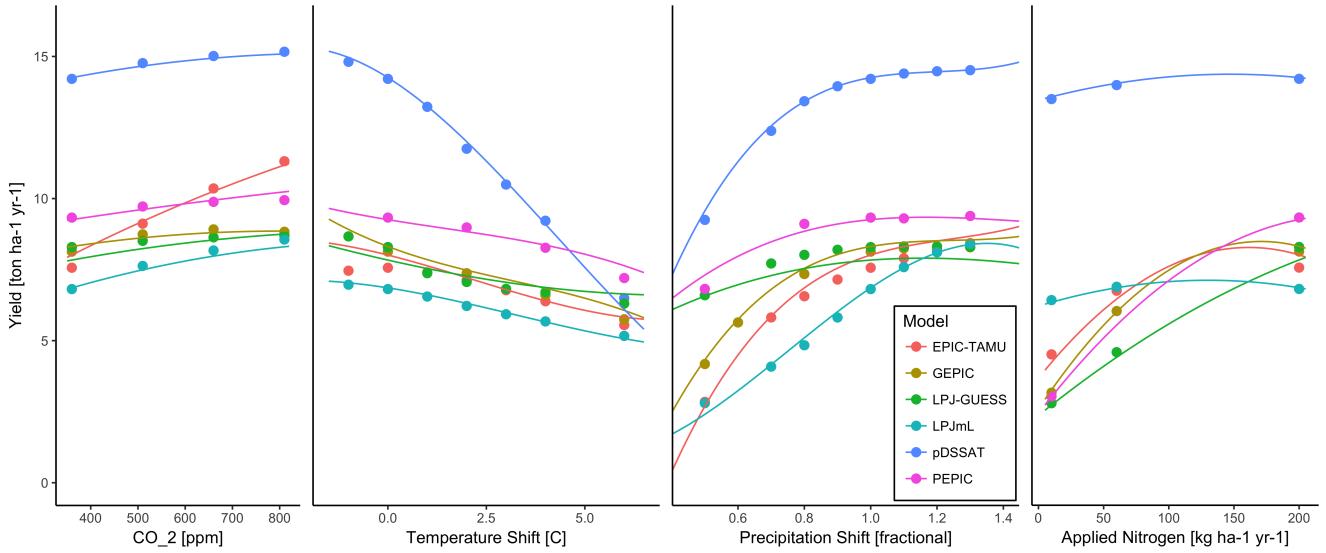


Figure 7: Illustration of across-model variations in yield response. Figures shows simulations and emulations from six models for rain-fed maize in the same Iowa grid cell shown in Figure 6, with the same plot conventions. Models that do not simulate the nitrogen dimension are omitted for clarity. Note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

424 responses are quite diverse across locations, crops, and models,⁴⁴⁴
 425 but in most cases local responses are regular enough to permit⁴⁴⁵
 426 emulation. We show illustrations of emulation fidelity in this⁴⁴⁶
 427 section; for more detailed discussion see Appendix XX.⁴⁴⁷

428 Crop yield responses are geographically diverse, even in⁴⁴⁸
 429 high-yield and high-cultivation areas Figure 6 illustrates geo-⁴⁴⁹
 430 graphic diversity for a single crop and model (rain-fed maize⁴⁵⁰
 431 in pDSSAT); this heterogeneity supports the choice of emulat-⁴⁵¹
 432 ing at the grid cell level. Each panel in Figure 6 shows sim-⁴⁵²
 433 ulated yield output from scenarios varying only along a single⁴⁵³
 434 dimension (CO₂, temperature, precipitation, or nitrogen addi-⁴⁵⁴
 435 tion), with other inputs held fixed at baseline levels, compared⁴⁵⁵
 436 to the full 4D emulation across the parameter space. Yields⁴⁵⁶
 437 evolve smoothly across the space sampled, and the polynomial⁴⁵⁷
 438 fit captures the climatological response to perturbations. Crop⁴⁵⁸
 439 yield responses generally follow similar functional forms across⁴⁵⁹
 440 models, though with a large spread in magnitude likely due to⁴⁶⁰
 441 the lack of calibration. Figure 7 illustrates inter-model diversity⁴⁶¹
 442 for a single crop and location (rain-fed maize in northern Iowa,⁴⁶²
 443 also shown in Figure 6). Differences in response shape can lead⁴⁶³

to differences in the fidelity of emulation, though comparison here is complicated by the different sampling regimes across models. Note that models are most similar in their responses to temperature perturbations. For this location and crop, CO₂ fertilization effects can range from ~5–50%, and nitrogen responses from nearly flat to a 60% drop in the lowest-application simulation.

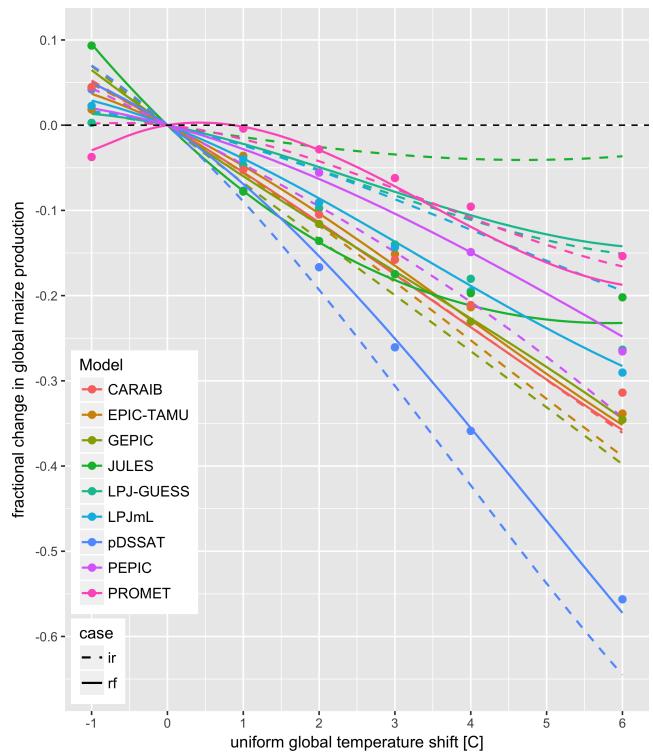
While the nitrogen dimension is important, it is also the most problematic to emulate in this work because of its limited sampling. The GGCMI Phase II protocol specified only three nitrogen levels (10, 60 and 200 kg N y⁻¹ ha⁻¹), so a third-order fit would be over-determined but a second-order fit can 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 kg N y⁻¹ 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 almost certainly an artifact of undersampling. In addition, the polynomial fit cannot capture the well-documented saturation effect of nitrogen application (e.g.

464 Ingestad, 1977) as accurately as would be possible with a non-
 465 parametric model.

466 The emulation fidelity demonstrated here is sufficient to al-
 467 low using emulated response surfaces to compare model re-
 468 sponses and derive insight about impacts projections. Because
 469 the emulator or “surrogate model” transforms the discrete sim-
 470 ulation sample space into a continuous response surface at any
 471 geographic scale, it can be used for a variety of applications,
 472 including construction of continuous damage functions. As an
 473 example, we show a damage function constructed from the 4D
 474 emulation, aggregated to global yield, with simulated values
 475 shown for comparison (Figure 8, which shows maize on cur-
 476 rently cultivated land; see Figures S16- S19 for other crops and
 477 dimensions). The emulated values closely match simulations
 478 even at this aggregation level. Note that these functions are
 479 presented only as examples and do not represent true global
 480 projections, because they are developed from simulation data
 481 with a uniform temperature shift while increases in global mean
 482 temperature should manifest non-uniformly. The global cov-
 483 erage of the GGCMI Phase II simulations allows impacts mod-
 484 elers to apply arbitrary geographically-varying climate projec-
 485 tions, as well as arbitrary aggregation masks, to develop dam-
 486 age functions for any climate scenario and any geopolitical or
 487 geographic level.

488 6. Conclusions and discussion

489 The GGCMI Phase II experiment provides a database tar-
 490 geted to allow detailed study of crop yields from process-based
 491 models under climate change. The experiment is designed to
 492 facilitate not only comparing the sensitivities of process-based
 493 crop yield models to changing climate and management inputs
 494 but also evaluating the complex interactions between driving
 495 factors (CO_2 , temperature, precipitation, and applied nitrogen).
 496 Its global nature also allows identifying geographic shifts in



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

508 high yield potential locations. We expect that the simulations
 509 will yield multiple insights in future studies, and show here a
 510 selection of preliminary results to illustrate their potential uses.

511 First, the GGCMI Phase II simulations allow identifying ma-
 512 jor areas of uncertainty. Across the major crops, inter-model
 513 uncertainty is greatest for wheat and least for soy. Across fac-
 514 tors impacting yields, inter-model uncertainty is largest for CO_2
 515 fertilization and nitrogen response effects. Across geographic
 516 regions, projections are most uncertain in the high latitudes
 517 where yields may increase, and most robust in low latitudes
 518 where yield impacts are largest.

519 Second, the GGCMI Phase II simulations allow understand-
 520 ing the way that climate-driven changes and locations of cul-

tivated land combine to produce yield impacts. One counterintuitive result immediately apparent is that irrigated maize shows steeper yield reductions under warming than does rain-fed maize when considered only over currently cultivated land. The effect results from geographic differences in cultivation. In any given location, irrigation increases crop resiliency to temperature increase, but irrigated maize is grown in warmer locations where the impacts of warming are more severe (Figures S5–S6). The same behavior holds for rice and winter wheat, but not for soy or spring wheat (Figures S8–S10). Irrigated wheat and maize are also more sensitive to nitrogen fertilization levels than are analogous non-irrigated crops, presumably because those rain-fed crops are limited by water as well as nitrogen availability (Figure S19). (Soy as an efficient atmospheric nitrogen fixer is relatively insensitive to nitrogen, and rice is not generally grown in water-limited conditions).

Third, we show that even the relatively limited GGCMI Phase II sampling space allows emulation of the climatological response of crop models with a relatively simple reduced-form 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 single projections 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 simple functional forms offer the possibility of physical interpretation of parameter values. Care should be taken in applying relationships developed at the yearly level to shifts in the mean climatology. We anticipate that systematic parameter sampling will become the norm in future model intercomparison exercise.

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 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 higher order moments, and cannot address the additional crop yield impacts of potential changes in climate variability. Although distributional changes in model projections are fairly uncertain at present, 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 not only mean changes in temperature and precipitation but changes in their marginal distributions or temporal dependence.

The GGCMI Phase II output dataset invites a broad range of potential future avenues of analysis. A major target area involves studying the models themselves with a detailed examination of interaction terms between the major input drivers, a more robust quantification of the sensitivity of different models to the input drivers, and comparisons with field-level experimental data. The parameter space tested in GGCMI Phase II will allow detailed investigations into yield variability and response to extremes under changing management and CO₂ levels. As mentioned previously, the database allows study of geographic shifts in optimal growing regions for different crops and studying the viability of switching crop types in some areas. The output dataset also contains other runs and variables not analyzed or shown here. Runs include several which allowed adaptation to climate changes by altering growing seasons, and additional variables include above ground biomass, LAI, and root biomass (as many as 25 output variables for some models).

578 Emulation studies that are possible include a more systematic⁶¹⁰
579 evaluation of different statistical model specifications and for-
580 mal calculation of uncertainties in derived parameters.⁶¹¹

581 The development of multi-model ensembles such as GGCMI⁶¹²
582 Phase II provides a way to begin to better understand crop re-⁶¹³
583 sponds to a range of potential climate inputs, improve process⁶¹⁴
584 based models, and explore the potential benefits of adaptive re-⁶¹⁵
585 sponds included shifting growing season, cultivar types and⁶¹⁶
586 cultivar geographic extent.⁶¹⁷

587 7. Acknowledgments

588 We thank Michael Stein and Kevin Schwarzwald, who pro-⁶²²
589 vided helpful suggestions that contributed to this work. This re-⁶²³
590 search was performed as part of the Center for Robust Decision-⁶²⁴
591 Making on Climate and Energy Policy (RDCEP) at the Univer-⁶²⁵
592 sity of Chicago, and was supported through a variety of sources.⁶²⁶
593 RDCEP is funded by NSF grant #SES-1463644 through the⁶²⁷
594 Decision Making Under Uncertainty program. J.F. was sup-⁶²⁸
595 ported by the NSF NRT program, grant #DGE-1735359. C.M.⁶²⁹
596 was supported by the MACMIT project (01LN1317A) funded⁶³⁰
597 through the German Federal Ministry of Education and Re-⁶³¹
598 search (BMBF). C.F. was supported by the European Research⁶³²
599 Council Synergy grant #ERC-2013-SynG-610028 Imbalance-⁶³³
600 P. P.F. and K.W. were supported by the Newton Fund through⁶³⁴
601 the Met Office Climate Science for Service Partnership Brazil⁶³⁵
602 (CSSP Brazil). A.S. was supported by the Office of Science⁶³⁶
603 of the U.S. Department of Energy as part of the Multi-sector⁶³⁷
604 Dynamics Research Program Area. Computing resources were⁶³⁸
605 provided by the University of Chicago Research Computing⁶³⁹
606 Center (RCC). S.O. acknowledges support from the Swedish⁶⁴⁰
607 strong research areas BECC and MERGE together with sup-⁶⁴¹
608 port from LUCCI (Lund University Centre for studies of Car-⁶⁴²
609 bon Cycle and Climate Interactions).

8. Appendix: Simulations – Assessment

The GGCMI Phase II simulations are designed for evaluating changes in yield but not absolute yields, since they omit detailed calibrations. To provide some validation of the skill of the process-based models used, we repeat the validation exercises of Müller et al. (2017) for GGCMI Phase I. The Müller et al. (2017) procedure evaluates response to year-to-year temperature and precipitation variations in a control run driven by historical climate and compares it to detrended historical yields from the FAO (Food and Agriculture Organization of the United Nations, 2018) by calculating the Pearson correlation coefficient. The procedure offers no means of assessing CO₂ fertilization, since CO₂ has been relatively constant over the historical data collection period. Nitrogen introduces some uncertainty into the analysis, since the GGCMI Phase II runs impose fixed, uniform nitrogen 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 nitrogen levels.

Figure 9 shows the Pearson time series correlation between the simulation model yield and FOA yield data. Figure 9 can be compared to Figures 1,2,3,4 and 6 in Müller et al. (2017). The results are mixed, with many regions for rice and wheat being difficult to model. No single model is dominant, with each model providing near best-in-class performance in at least one location-crop combination. The presence of very few vertical dark green color bars clearly illustrates the power of a multi-model intercomparison project like the one presented here. The ensemble mean does not beat the best model in each case, but shows positive correlation in over 75% of the cases presented here. The EPIC-TAMU model performs best for soy, CARIAB, EPIC-TAMU, and PEPIC perform best for maize, PROMET performs best for wheat, and the EPIC family of models perform best for rice. Reductions in skill over the performance

644 illustrated in Müller et al. (2017) can be attributed to the nitrogen levels or lack of calibration in some models.

645

646 *** or harmonization *** Christoph

647 Soy is qualitatively the easiest crop to represent (except in Argentina), which is likely due in part to the invariance of the

648

649 response to nitrogen application (soy fixes atmospheric nitrogen very efficiently). Comparison to the FAO data is therefore easier than the other crops because the nitrogen application levels do not matter. US maize has the best performance across models, with nearly every model representing the historical variability

650

651

652

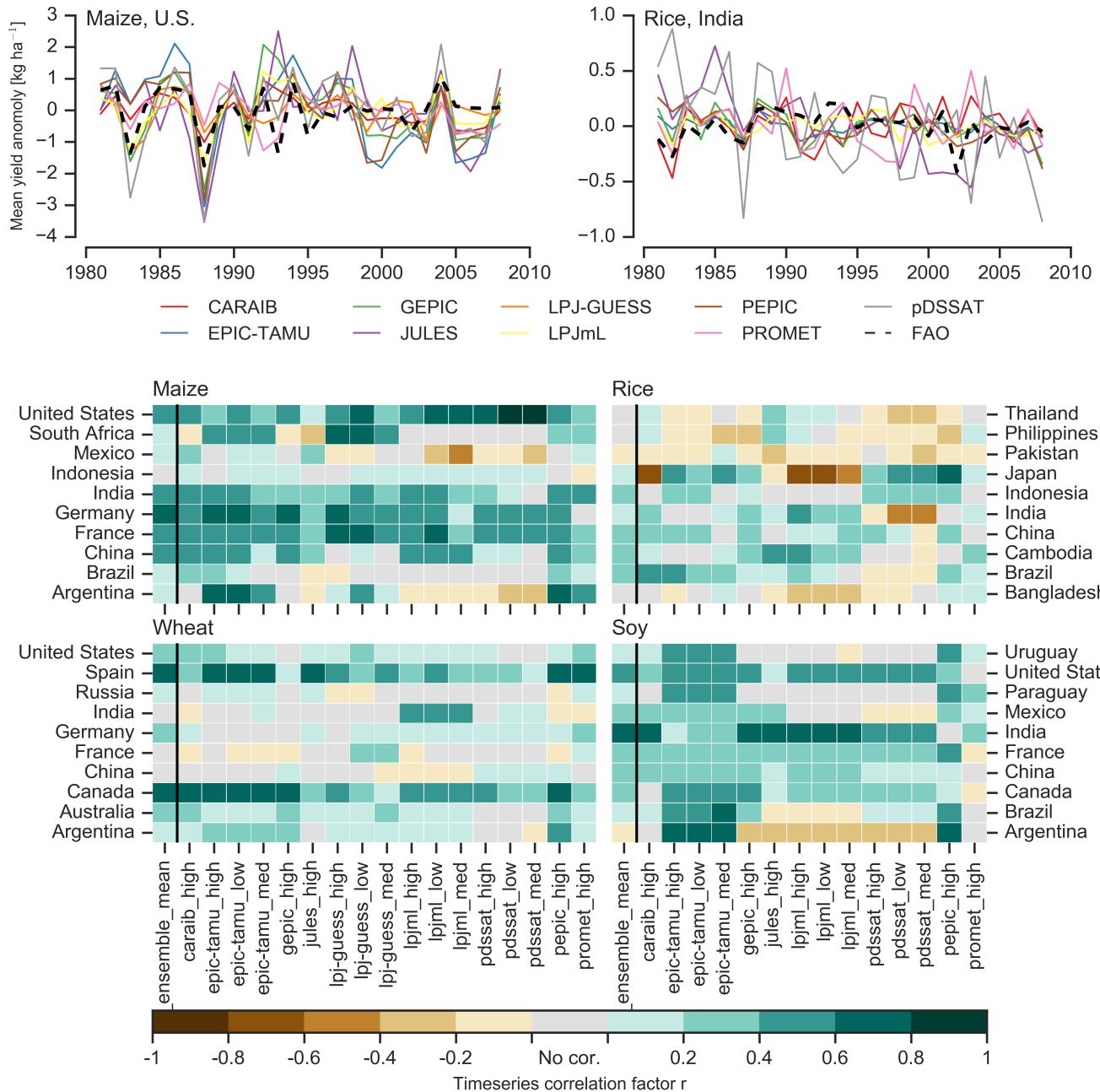


Figure 9: Time series of correlation coefficients between simulated crop yield and FAO data (Food and Agriculture Organization of the United Nations, 2018) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen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

654 to a reasonable extent. Especially good example years for US
 655 maize are 1983, 1988, and 2004 (top left panel of Figure 9),
 656 where every model gets the direction of the anomaly compared
 657 to surrounding years correct. 1983 and 1988 are famously bad
 658 years for US maize along with 2012 (not shown). US maize
 659 is possibly both the most uniformly industrialized (in terms of
 660 management practices) crop and the one with the best data col-
 661 lection in the historical period of all the cases presented here.

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

678 9. Appendix: Emulation – Assessment

679 Because no general criteria exist for defining an acceptable
 680 model emulator, we develop a metric of emulator performance
 681 specific to GGCMI Phase II. For a multi-model comparison ex-
 682 ercise like GGCMI Phase II, one reasonable criterion is what
 683 we term the “normalized error”, which compares the fidelity of
 684 an emulator for a given model and scenario to the inter-model
 685 uncertainty. We define the normalized error e for each scenario
 686 as the difference between the fractional yield change from the

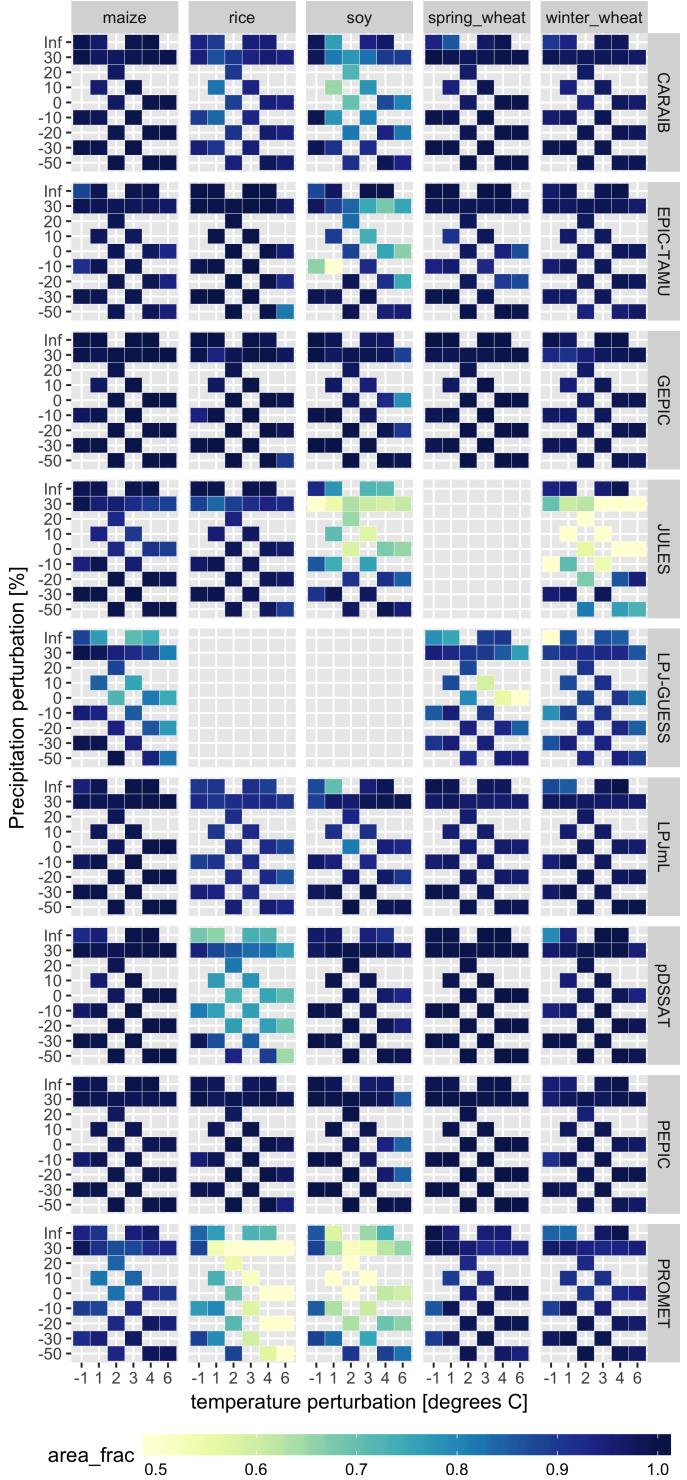


Figure 10: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO₂ and nitrogen held fixed at baseline values. Large columns are crops and large rows models; squares within are T,P scenario pairs. Colors denote the fraction of currently cultivated hectares with $e < 1$. Of the 756 scenarios with these CO₂ and N values, we consider only those for which all 9 models submitted data. JULES did not simulate spring wheat and LPJ-GUESS did not simulate rice and soy. Emulator performance is generally satisfactory, with some exceptions. Emulator failures (significant areas of poor performance) occur for individual crop-model combinations, with performance generally degrading for hotter and wetter scenarios.

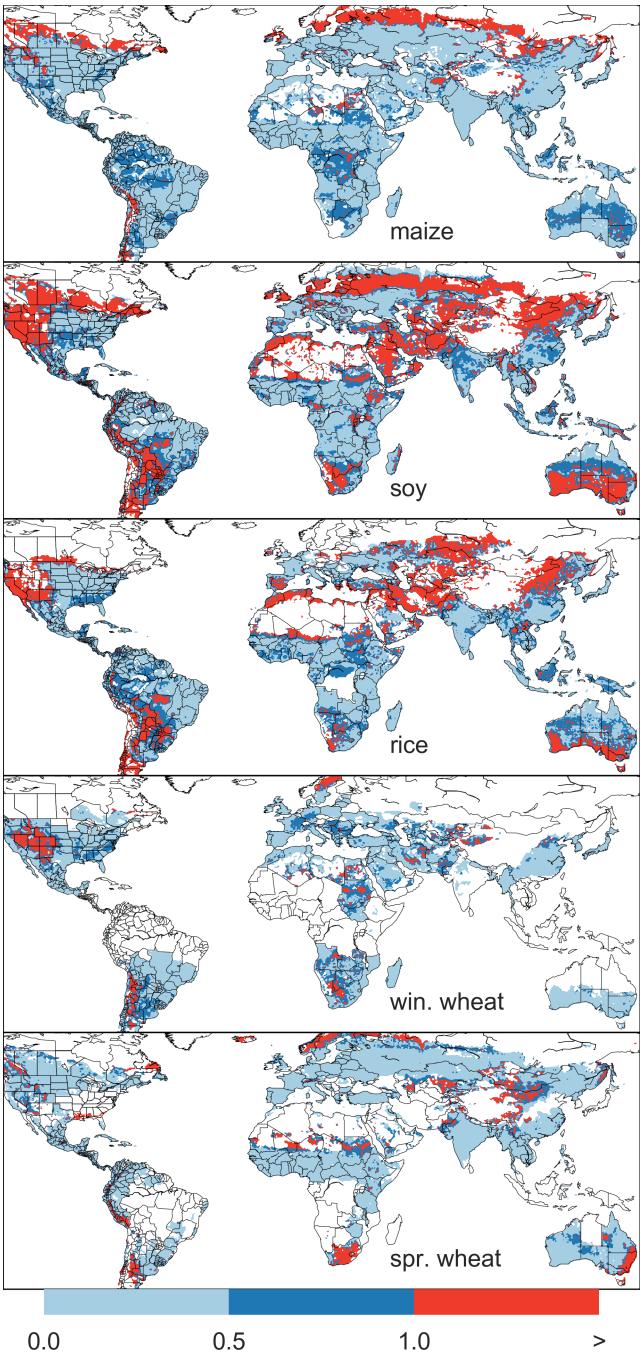


Figure 11: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error e , where $e > 1$ means that emulator error exceeds the multi-model standard deviation. White areas are those where crops 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.

687 emulator and that in the original simulation, divided by the standard deviation of the multi-model spread (Equations 2 and 3):
688

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

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

689 Here $F_{scn.}$ is the fractional change in a model's mean emulated or simulated yield from a defined baseline, in some scenario (scn.) in C, T, W, and N space; $Y_{scn.}$ and $Y_{baseline}$ are the absolute emulated or simulated mean yields. The normalized
690 error e is the difference between the emulated fractional change
691 in yield and that actually simulated, normalized by $\sigma_{sim.}$ the
692 standard deviation in simulated fractional yields $F_{sim, scn.}$ across
693 all models. The emulator is fit across all available simulation
694 outputs, and then the error is calculated across the simulation
695 scenarios provided by all nine models (Figure 10 and Figures
696 S12 and Figures S13 in supplemental documents).

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

To assess the ability of the polynomial emulation to capture the behavior of complex process-based models, we evaluate the normalized emulator error. That is, for each grid cell, model, and scenario we evaluate the difference between the model yield

716 and its emulation, normalized by the inter-model standard deviation in yield projections. This metric implies that emulation
 717 is generally satisfactory, with several distinct exceptions. Almost all model-crop combination emulators have normalized
 718 errors less than one over nearly all currently cultivated hectares
 719 (Figure 10), but some individual model-crop combinations are
 720 problematic (e.g. PROMET for rice and soy, JULES for soy
 721 and winter wheat, Figures S14–S15). Normalized errors for soy
 722 are somewhat higher across all models not because emulator fidelity
 723 is worse but because models agree more closely on yield
 724 changes for soy than for other crops (see Figure S16, lowering
 725 the denominator. Emulator performance often degrades in geo-
 726 graphic locations where crops are not currently cultivated. Fig-
 727 ure 11 shows a CARAIB case as an example, where emulator
 728 performance is satisfactory over cultivated areas for all crops
 729 other than soy, but uncultivated regions show some problematic
 730 areas.

731 It should be noted that this assessment metric is relatively
 732 forgiving. First, each emulation is evaluated against the simu-
 733 lation actually used to train the emulator. Had we used a spline
 734 interpolation the error would necessarily be zero. Second, the
 735 performance metric scales emulator fidelity not by the mag-
 736 nitude of yield changes but by the inter-model spread in those
 737 changes. Where models differ more widely, the standard for
 738 emulators becomes less stringent. Because models disagree on
 739 the magnitude of CO₂ fertilization, this effect is readily seen
 740 when comparing assessments of emulator performance in sim-
 741 ulations at baseline CO₂ (Figure 10) with those at higher CO₂
 742 levels (Figure S13). Widening the inter-model spread leads to
 743 an apparent increase in emulator skill.

744 10. References

745 Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013). Implication of crop model calibration strategies for assessing regional impacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.

746 Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D., A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Alderman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., Biernath, C., Challinor, A., De Sanctis, G., & Zhu, Y. (2015). Rising temperatures reduce global wheat production. *Nature Climate Change*, 5, 143–147. doi:10.1038/nclimate2470.

747 Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., J. Boote, K., Thorburn, P., Ritter, R. P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A., Doltra, J., & Wolf, J. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, 3, 827832. doi:10.1038/nclimate1916.

748 Aulakh, M. S., & Malhi, S. S. (2005). Interactions of Nitrogen with Other Nutrients and Water: Effect on Crop Yield and Quality, Nutrient Use Efficiency, Carbon Sequestration, and Environmental Pollution. *Advances in Agronomy*, 86, 341 – 409.

749 Balkovi, J., van der Velde, M., Skalsk, R., Xiong, W., Folberth, C., Khabarov, N., Smirnov, A., Mueller, N. D., & Obersteiner, M. (2014). Global wheat production potentials and management flexibility under the representative concentration pathways. *Global and Planetary Change*, 122, 107 – 121.

750 Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models. *Agricultural and Forest Meteorology*, 236, 145 – 161.

751 Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded crop models using statistical estimates. *Agricultural and Forest Meteorology*, 214-215, 134 – 147.

752 von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., & Zaehle, S. (2018). Implementing the Nitrogen cycle into the dynamic global vegetation, hydrology and crop growth model LPJmL (version 5.0). *Geoscientific Model Development*, 11, 2789–2812.

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

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

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

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

757 Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for

- 793 simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205). 836
- 794 Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdous,⁸³⁷
- 795 M., & François, L. (2011). Responses of European forest ecosystems to⁸³⁸
- 796 21st century climate: assessing changes in interannual variability and fire⁸³⁹
- 797 intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99). 840
- 798 Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K.,⁸⁴¹
- 799 Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating⁸⁴²
- 800 impact models and sectors (pSIMS). *Environmental Modelling and Soft-⁸⁴³*
- 801 ware, 62, 509–516. 844
- 802 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J.,⁸⁴⁵
- 803 Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C.,⁸⁴⁶
- 804 Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J.⁸⁴⁷
- 805 (2015). The Global Gridded Crop Model Intercomparison: data and mod-⁸⁴⁸
- 806 eling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8,⁸⁴⁹
- 807 261–277. 850
- 808 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J.,⁸⁵¹
- 809 & Taylor, K. E. (2016). Overview of the coupled model intercomparison⁸⁵²
- 810 project phase 6 (cmip6) experimental design and organization. *Geoscientific⁸⁵³*
- 811 *Model Development*, 9, 1937–1958. 854
- 812 Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of cli-⁸⁵⁵
- 813 mate change impacts on durum wheat in the mediterranean region. *Natural-⁸⁵⁶*
- 814 *Hazards and Earth System Sciences*, 11, 1293–1302. 857
- 815 Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Re-⁸⁵⁸
- 816 gionalization of a large-scale crop growth model for sub-Saharan Africa:⁸⁵⁹
- 817 Model setup, evaluation, and estimation of maize yields. *Agriculture,⁸⁶⁰*
- 818 *Ecosystems & Environment*, 151, 21 – 33. 861
- 819 Food and Agriculture Organization of the United Nations (2018). FAOSTAT⁸⁶²
- 820 database. URL: <http://www.fao.org/faostat/en/home>. 863
- 821 Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T.,⁸⁶⁴
- 822 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P.,⁸⁶⁵
- 823 Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F.,⁸⁶⁶
- 824 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,⁸⁶⁷
- 825 Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J.,⁸⁶⁸
- 826 Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nen-⁸⁶⁹
- 827 del, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna,⁸⁷⁰
- 828 M., Semenov, M. A., Slawinski, C., Stratonovitch, P., Supit, I., Waha, K.,⁸⁷¹
- 829 Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model⁸⁷²
- 830 wheat yield impact response surfaces showing sensitivity to temperature and⁸⁷³
- 831 precipitation change. *Agricultural Systems*, 159, 209–224. 874
- 832 Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014).⁸⁷⁵
- 833 Evaluating the utility of dynamical downscaling in agricultural impacts pro-⁸⁷⁶
- 834 jections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781.⁸⁷⁷
- 835 Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitiv-⁸⁷⁸
- ity 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.
- He, W., Yang, J., Zhou, W., Drury, C., Yang, X., D. Reynolds, W., Wang, H., He, P., & Li, Z.-T. (2016). Sensitivity analysis of crop yields, soil water contents and nitrogen leaching to precipitation, management practices and soil hydraulic properties in semi-arid and humid regions of Canada using the DSSAT model. *Nutrient Cycling in Agroecosystems*, 106, 201–215.
- Heady, E. O. (1957). An Econometric Investigation of the Technology of Agricultural Production Functions. *Econometrica*, 25, 249–268.
- Heady, E. O., & Dillon, J. L. (1961). *Agricultural production functions*. Iowa State University Press.
- Holden, P., Edwards, N., PH, G., Fraedrich, K., Lunkeit, F., E, K., Labriet, M., Kanudia, A., & F, B. (2014). Plasim-entsem v1.0: A spatiotemporal emulator of future climate change for impacts assessment. *Geoscientific Model Development*, 7, 433–451. doi:10.5194/gmd-7-433-2014.
- Holzkämper, A., Calanca, P., & Fuhrer, J. (2012). Statistical crop models: Predicting the effects of temperature and precipitation changes. *Climate Research*, 51, 11–21.
- Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgliesh, N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson, M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T., Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., & Keating, B. A. (2014). APSIM Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling and Software*, 62, 327 – 350.
- Howden, S., & Crimp, S. (2005). Assessing dangerous climate change impacts on australia's wheat industry. *Modelling and Simulation Society of Australia and New Zealand*, (pp. 505–511).
- Iizumi, T., Nishimori, M., & Yokozawa, M. (2010). Diagnostics of climate model biases in summer temperature and warm-season insolation for the simulation of regional paddy rice yield in japan. *Journal of Applied Meteorology and Climatology*, 49, 574–591.
- Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Nitrogen Fertilizers. *Ambio*, 6, 146–151.
- Izaurralde, R., Williams, J., McGill, W., Rosenberg, N., & Quiroga Jakas, M.

- 879 (2006). Simulating soil C dynamics with EPIC: Model description and test-⁹²²
 880 ing against long-term data. *Ecological Modelling*, *192*, 362–384. ⁹²³
- 881 Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the⁹²⁴
 882 CROPGRO-soybean model to predict regional yield and production. *Agri-*⁹²⁵
883 culture, Ecosystems & Environment, *93*, 73 – 85. ⁹²⁶
- 884 Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L.,⁹²⁷
 885 Wilkens, P., Singh, U., Gijsman, A., & Ritchie, J. (2003). The DSSAT⁹²⁸
 886 cropping system model. *European Journal of Agronomy*, *18*, 235 – 265. ⁹²⁹
- 887 Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I.,⁹³⁰
 888 Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A.,⁹³¹
 889 Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017).⁹³²
 890 Toward a new generation of agricultural system data, models, and knowl-⁹³³
 891 edge products: State of agricultural systems science. *Agricultural Systems*,⁹³⁴
 892 *155*, 269 – 288. ⁹³⁵
- 893 Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth,⁹³⁶
 894 D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Ver-⁹³⁷
 895 burg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K.,⁹³⁸
 896 Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003).⁹³⁹
 897 An overview of APSIM, a model designed for farming systems simulation.⁹⁴⁰
 898 *European Journal of Agronomy*, *18*, 267 – 288. ⁹⁴¹
- 899 Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., &⁹⁴²
 900 Smith, B. (2013). Implications of accounting for land use in simulations of⁹⁴³
 901 ecosystem carbon cycling in Africa. *Earth System Dynamics*, *4*, 385–407. ⁹⁴⁴
- 902 Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling⁹⁴⁵
 903 wheat yield and crop water productivity with high resolution on a global⁹⁴⁶
 904 scale. *Agricultural Systems*, *94*, 478 – 493. ⁹⁴⁷
- 905 Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., & Schulin, R. (2016a).⁹⁴⁸
 906 Global investigation of impacts of PET methods on simulating crop-water⁹⁴⁹
 907 relations for maize. *Agricultural and Forest Meteorology*, *221*, 164 – 175. ⁹⁵⁰
- 908 Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C.,⁹⁵¹
 909 & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs⁹⁵²
 910 with yields from major crop cultivations. *Science of The Total Environment*,⁹⁵³
 911 *572*, 526 – 537. ⁹⁵⁴
- 912 Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict⁹⁵⁵
 913 crop yield responses to climate change. *Agricultural and Forest Meteorol-*⁹⁵⁶
914 ogy, *150*, 1443 – 1452. ⁹⁵⁷
- 915 Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relation-⁹⁵⁸
 916 ships and the impacts of recent warming. *Environmental Research Letters*,⁹⁵⁹
 917 *2*, 014002. ⁹⁶⁰
- 918 MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, *4*, 415–447.⁹⁶¹
- 919 Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P.,⁹⁶²
 920 Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi,⁹⁶³
 921 P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano,⁹⁶⁴
 D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of
 923 large simulated yield datasets for studying climate effects. (p. 1100).
 doi:10.13140/RG.2.1.5173.8328.
- Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrological
 924 modelling to study the impact of climate change on the water flows of
 925 mountain watersheds. *Journal of Hydrology*, *376*, 362 – 377.
- Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B.,
 926 & Calzadilla, A. (2009). Global biomass production potentials exceed ex-
 927 pected future demand without the need for cropland expansion. *Nature Com-
 928 munications*, *6*.
- McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srinivasan,
 929 C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S.
 930 (2015). Integrated assessments of the impacts of climate change on agricul-
 931 ture: An overview of AgMIP regional research in South Asia. *Chapter in:*
932 Handbook of Climate Change and Agroecosystems, (pp. 201–218).
- Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical
 933 weather responsiveness of crop yields: US evidence and implications for
 934 the agricultural impacts of climate change. *Environmental Research Letters*,
 935 *12*.
- Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate
 936 change impacts on agriculture implies higher social cost of carbon. *Nature
 937 Communications*, *8*.
- Müller, C., Elliott, J., Chrysanthacopoulos, J., Arneth, A., Balkovic, J., Ciais,
 938 P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurrealde,
 939 R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T.
 940 A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G.,
 941 Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H.
 942 (2017). Global gridded crop model evaluation: benchmarking, skills, defi-
 943 ciencies and implications. *Geoscientific Model Development*, *10*, 1403–
 944 1422.
- Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T.
 945 (1997). Effect of CO₂ enrichment on carbon and nitrogen interaction in
 946 wheat and soybean. *Soil Science and Plant Nutrition*, *43*, 789–798.
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial.
Reliability Engineering & System Safety, *91*, 1290 – 1300.
- Olin, S., Schurgers, G., Lindeskog, M., Wårliind, D., Smith, B., Bodin, P.,
 947 Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue
 948 C:N to changes in atmospheric CO₂ and N management in the main wheat
 949 regions of western europe. *Biogeosciences*, *12*, 2489–2515. doi:10.5194/bg-
 950 12-2489-2015.
- Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in
 951 field crop production. *Soil Science and Plant Nutrition*, *38*, 553–564.
- Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., &

- 965 Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK008
 966 Land Environment Simulator. *Geoscientific Model Development*, 8, 1139–1155.
 967 1155.
- 968 Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop011
 969 yields and their variability at different levels of global warming. *Earth Sys012
 970 tem Dynamics*, 9, 479–496.
- 971 Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff,014
 972 S., & Gerten, D. (2015). Emulating global climate change impacts on crop015
 973 yields. *Statistical Modelling*, 15, 499–525.
- 974 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,017
 975 Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas018
 976 sos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011),019
 977 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning020
 978 Research*, 12, 2825–2830.
- 979 Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T.,022
 980 Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski,023
 981 P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.,024
 982 Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P.,025
 983 Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Mineti,026
 984 J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C.,027
 985 Öztürk, I., Perego, A., Rodríguez, A., Ruane, A., Ruget, F., Sanna, M.,028
 986 Semenov, M., Slawinski, C., Strattonovich, P., Supit, I., Waha, K., Wang,029
 987 E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation030
 988 effects on wheat yield across a European transect: a crop model ensemble031
 989 analysis using impact response surfaces. *Climate Research*, 65, 87–105.
- 990 Porter et al. (IPCC) (2014). Food security and food production systems. Cli033
 991 mate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global034
 992 and Sectoral Aspects. Contribution of Working Group II to the Fifth Assess035
 993 ment Report of the Intergovernmental Panel on Climate Change. In C. Fi036
 994 et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK037
 995 Cambridge University Press.
- 996 Portmann, F., Siebert, S., Bauer, C., & Doell, P. (2008). Global dataset of039
 997 monthly growing areas of 26 irrigated crops.
- 998 Portmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly041
 999 Irrigated and Rainfed crop Areas around the Year 2000: A New High042
 1000 Resolution Data Set for Agricultural and Hydrological Modeling. *Global043
 1001 Biogeochemical Cycles*, 24, GB1011.
- 1002 Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E.,045
 1003 & Arneth, A. (2016). Climate analogues suggest limited potential for inten046
 1004 sification of production on current croplands under climate change. *Nature047
 1005 Communications*, 7, 12608.
- 1006 Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term cli049
 1007 mate change based on a resampling ensemble technique. *Tellus A: Dynamical050
 1008 Meteorology and Oceanography*, 58, 461–472.
- Ratto, M., Castelletti, A., & Pagano, A. (2012). Emulation techniques for the reduction and sensitivity analysis of complex environmental models. *Environmental Modelling & Software*, 34, 1 – 4.
- Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, 48.
- Roberts, M., Braun, N., R Sinclair, T., B Lobell, D., & Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, 12.
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, 111, 3268–3273.
- Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P., Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., & Winter, J. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology*, 170, 166 – 182.
- Rosenzweig, C., Ruane, A. C., Antle, J., Elliott, J., Ashfaq, M., Chatta, A. A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G., Lotze-Campen, H., MacCarthy, D. S., Mason-D'Croz, D., Contreras, E. M., Müller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R. M., Sands, R. D., Schleussner, C.-F., Valdivia, R. O., Valin, H., & Wiebe, K. (2018). Coordinating AgMIP data and models across global and regional scales for 1.5°C and 2.0°C assessments. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 376.
- Ruane, A., I. Hudson, N., Asseng, S., Camarrano, D., Ewert, F., Martre, P., 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., Müller, C., Porter, C., Phillips, M. M., Raymundo, R. M., Sands, R., Valdivia, R. O., White, J. W., Wiebe, K., & Rosenzweig, C. (2018). Biophysical and economic implications for agriculture of +1.5° and +2.0°C global warming using AgMIP Coordinated Global and Regional Assessments. *Climate Research*, 76, 17–39.
- Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown,

- 1051 D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013). Climate change impact uncertainties for maize in panama: Farm information, climate projections, and yield sensitivities. *Agricultural and Forest Meteorology*, 170, 132 – 145.
- 1052 Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agric. Forest Meteorol.*, 200, 233–248.
- 1053 Ruane, A. C., McDermid, S., Rosenzweig, C., Baigorria, G. A., Jones, J. W., Romero, C. C., & Cecil, L. D. (2014). Carbon-temperature-water change analysis for peanut production under climate change: A prototype for the agmip coordinated climate-crop modeling project (c3mp). *Glob. Change Biol.*, 20, 394–407. doi:10.1111/gcb.12412.
- 1054 Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift*, 19, 135–141.
- 1055 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-Mechiche-Alami, A., Minet, J., Montesino, M., Nendel, C., Porten, J., Ruget, F., Semenov, M., Steinmetz, Z., Strattonovitch, P., Supit, I., Tao, F., Trnka, M., de Wit, A., & Ritter, R. (2018). Adaptation response surfaces for managing wheat under perturbed climate and co₂ in a mediterranean environment. *Agricultural Systems*, 159, 260 – 274. doi:doi.org/10.1016/j.agsy.2017.01.009.
- 1056 Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19, 607–620.
- 1057 Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., A. M. Pugh, T., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., & Frieler, K. (2017). Consistent negative response of US crops to high temperatures in observations and crop models. *Nature Communications*, 8, 13931.
- 1058 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.
- 1059 Snyder, A., Calvin, K. V., Phillips, M., & Ruane, A. C. (2018). A crop yield change emulator for use in gcam and similar models: Persephone v1.0. *Geoscientific Model Development Discussions*, 2018, 1–42.
- 1060 Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety*, 94, 1735 – 1763.
- 1061 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.
- 1062 Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters*, 35.
- 1063 Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., & Martin, J. F. (2014). Modeling sugarcane yield with a process-based model from site to continental scale: Uncertainties arising from model structure and parameter values. *Geoscientific Model Development*, 7, 1225–1245.
- 1064 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.
- 1065 White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Methodologies for simulating impacts of climate change on crop production. *Field Crops Research*, 124, 357 – 368.
- 1066 Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T., Arkebauer, T., & Scoby, D. (2017). Evaluation of JULES-crop performance against site observations of irrigated maize from Mead, Nebraska. *Geoscientific Model Development*, 10, 1291–1320.
- 1067 Williams, K. E., & Falloon, P. D. (2015). Sources of interannual yield variability in JULES-crop and implications for forcing with seasonal weather forecasts. *Geoscientific Model Development*, 8, 3987–3997.
- 1068 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Landbouwkundige Onderzoeken* : 64.6, .
- 1069 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.
- 1070 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, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci.*, 114, 9326–9331.