

GGCMI Phase II: globally gridded crop simulation model response to uniform changes in CO₂, temperature, water, and nitrogen levels

James Franke^{1,2}, Joshua Elliott^{2,3}, Christoph Müller⁴, Alexander Ruane⁵, Abigail Snyder⁶, Jonas Jägermeyr^{3,2,4,5}, Juraj Balkovic^{7,8}, Philippe Ciais^{9,10}, Marie Dury¹¹, Pete Falloon¹², Christian Folberth⁷, Louis François¹¹, Tobias Hank¹³, Munir Hoffmann^{14,23}, R. Cesar Izaurralde^{15,16}, Ingrid Jacquemin¹¹, Curtis Jones¹⁵, Nikolay Khabarov⁷, Marian Koch¹⁴, Michelle Li^{2,17}, Wenfeng Liu^{9,18}, Stefan Olin¹⁹, Meridell Phillips^{5,20}, Thomas A. M. Pugh^{21,22}, Ashwan Reddy¹⁵, Xuhui Wang^{9,10}, Karina Williams¹², Florian Zabel¹³, and Elisabeth Moyer^{1,2}

¹Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA

²Center for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA

³Department of Computer Science, University of Chicago, Chicago, IL, USA

⁴Potsdam Institute for Climate Impact Research, Leibniz Association (Member), Potsdam, Germany

⁵NASA Goddard Institute for Space Studies, New York, NY, United States

⁶Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA

⁷Ecosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria

⁸Department of Soil Science, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovak Republic

⁹Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France

¹⁰Sino-French Institute of Earth System Sciences, College of Urban and Env. Sciences, Peking University, Beijing, China

¹¹Unité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d'Astrophysique et de Géophysique, University of Liège, Belgium

¹²Met Office Hadley Centre, Exeter, United Kingdom

¹³Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany

¹⁴Georg-August-University Göttingen, Tropical Plant Production and Agricultural Systems Modelling, Göttingen, Germany

¹⁵Department of Geographical Sciences, University of Maryland, College Park, MD, USA

¹⁶Texas AgriLife Research and Extension, Texas A&M University, Temple, TX, USA

¹⁷Department of Statistics, University of Chicago, Chicago, IL, USA

¹⁸EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

¹⁹Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

²⁰Earth Institute Center for Climate Systems Research, Columbia University, New York, NY, USA

²¹School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK.

²²Birmingham Institute of Forest Research, University of Birmingham, Birmingham, UK.

²³Leibniz Centre for Agricultural Landscape Research (ZALF), D-15374 Müncheberg, Germany

Correspondence: James Franke (jfranke@uchicago.edu)

Abstract. Concerns about food security under climate change motivate efforts to better understand future changes in crop yields. Process-based crop models, which represent plant physiological processes, are necessary tools for this purpose since they allow representing future conditions not sampled in the historical record and new locations where cultivation may shift.

5 However, models remain uncertain and differ in many critical details. The Global Gridded Crop Model Intercomparison (GGCMI) Phase II experiment, an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP),

is designed to allow systematic evaluation and “emulation” of model responses to multiple interacting factors, including carbon dioxide, temperature, water availability, and nitrogen application (CTWN). In this paper we describe the GGCMI Phase II experimental protocol (simulations run with systematic uniform perturbations of historical climate) and its simulations (from twelve crop models and five crops); identify responses that are robust across models and those that remain uncertain; and

5 present an emulator or statistical representation of model responses. Modeled yields show robust decreases to warmer mean climatologies in almost all regions, with a nonlinear dependence that means yield changes in warmer baseline locations are more sensitive to temperature increases. Inter-model uncertainty is qualitatively similar across all the four input dimensions, but is highest in high-latitude regions where crops may be grown in the future.

1 Introduction

10 Understanding crop yield response to a changing climate is critically important, especially as the global food production system will face pressure from increased demand over the next century (?). Climate-related reductions in supply could therefore have severe socioeconomic consequences (e.g. ??). Multiple studies using different crop or climate models concur in projecting sharp yield reductions on currently cultivated cropland under business-as-usual climate scenarios, although their yield projections show considerable spread (e.g. ???, and references therein). Modeling crop responses continues to be challenging, as crop

15 growth is a function of complex interactions between climate inputs and management practices (?). Intercomparison projects such as the Agricultural Model Intercomparison and Improvement Project (AgMIP ?) have helped to quantify uncertainties in model projections (?), developed protocols (?) for evaluating model performance (?) and analyzed selected mechanisms (e.g. ?).

Models tend to agree broadly in major response patterns, including a reasonable representation of the spatial pattern in
20 historical yields of major crops (e.g. ??) and projections of shifts in yield under future climate scenarios.

Process-based models still struggle with some important details, including reproducing historical year-to-year variability in various countries (e.g. ?), reproducing historical yields when driven by reanalysis weather (e.g. ?), and low sensitivity to extreme events (e.g. ???). These issues are driven in part by the diversity of new cultivars and genetic variants, which outstrips the ability of academic modeling groups to capture them (e.g. ?). Models also do not simulate many additional
25 factors affecting production, including but not limited to: pests, diseases, and weeds. For these reasons, individual studies must generally re-calibrate models to ensure that short-term predictions reflect current cultivars and management levels, and long-term projections retain considerable uncertainty (??????). Inter-model discrepancies can also be high in areas not yet cultivated (e.g. ??). Finally, process-based models present additional difficulties for high-resolution global studies because of their complexity and computational requirements, as well as the general inability to calibrate site-specific parameters.

30 Nevertheless, process-based models are necessary for understanding the future yield impacts of climate change as well as options for climate mitigation (?) and adaptation (?), future sustainable development (?), and food security (?). Despite the additional uncertainty and challenges to run crop models at the global scale (?), gridded, global crop model applications are much needed for consistent assessments. Climate change and international agricultural trade often require to consider the global

and local dynamics (??) and global market mechanisms may strongly modulate climate change impacts (??). Cultivation may shift to new areas, where no yield data are currently available and therefore statistical models cannot be applied. Finally, only process-based models can capture the growth response to novel conditions and practices that are not represented in historical data (e.g. ??). These novel changes can include the direct fertilization effect of elevated CO₂, and changes in management practices that may mitigate climate-induced damages (e.g. ?).

Understanding process-based models' responses to important drivers are thus critical to improve future projections of agricultural productivity. The Global Gridded Crop Model Intercomparison (GGCMI) of AgMIP (?) has thus designed a 2nd phase (*Phase II*) with a focus on models' response to changes in the most important model drivers: atmospheric carbon dioxide concentrations ([CO₂], C), temperature (T), water (W), and nitrogen (N). The CTWN experiment is inspired by AgMIP's Co-ordinated Climate-Crop Modeling Project (C3MP ??) and has been extended by the nitrogen dimension. To account for the importance of adaption in agricultural production to climate change, a fifth dimension, adaptation (A) has been added to the CTWN experiment. The A dimension only has two elements, namely *no adaptation* and one *adaption* setting, in which the growing season that is shortened under warming, is regained by adapted crop cultivar choice.

The Global Gridded Crop Model Intercomparison (GGCMI) Phase II experiment was conducted by a suite of process-based crop models, running across historical conditions perturbed by a set of discrete steps in the four input parameters (CTWN), for both adaptation settings (yes/no). The experimental protocol involves 756 different parameter combinations for each model and crop for the non-adapted and 648 combinations for the adapted setting (does not include T=0), with simulations providing near-global coverage at a half degree spatial resolution. The experiment was conducted as part of the Agricultural Model Intercomparison and Improvement Project (AgMIP) (??), an international effort conducted under a framework similar to the Climate Model Intercomparison Project (CMIP) (??). The GGCMI protocol builds on the AgMIP Coordinated Climate-Crop Modeling Project (C3MP) (??) and contributes to the AgMIP Coordinated Global and Regional Assessments (CGRA) (??). GGCMI Phase II is designed to allow addressing goals such as understanding where highest-yield regions may shift under climate change; exploring future adaptive management strategies; understanding how interacting input drivers affect crop yield; quantifying uncertainties across models and major drivers; and testing strategies for producing lightweight emulators of process-based models. In this paper, we describe the GGCMI Phase II model experiments and present initial summary results.

2 Methods

2.1 Modeling protocol

GGCMI Phase II is the continuation of a multi-model comparison exercise begun in 2014. The initial Phase I compared harmonized yield simulations of 21 models for 19 crops over the historical period with a primary goal of model evaluation (??) and understanding of different sources of uncertainty, such as model parameterization (?), different weather products, or aggregation (?). Phase II compares simulations of 12 models for 5 crops (maize, rice, soybean, spring wheat, and winter wheat) using only one of the weather products used in Phase I, which was then perturbed systematically for the different simulations. The reduced set of crops includes the three major global cereals and the major legume and accounts for over 50% of human

calories (in 2016, nearly 3.5 billion tons or 32% of total global crop production by weight (?), which were also in the focus of analyses in the previous analyses (???) and for which global sub-national reference data exist (??).

The guiding scientific rationale of GGCMI Phase II is to provide a comprehensive, systematic evaluation of the response of process-based crop models to [CO₂], temperature, water, and applied nitrogen (collectively referred to as “CTWN”). Given that the production of annual crops is continuously changing by making use of new varieties and technologies (?) it is plausible to assume at least some degree of adaptation to new climate conditions. We thus also include one additional simulation set across the entire CTWN space with just one simple adaptation setting (A), in which we assume that varieties are grown at each temperature level, that maintain the growing season, i.e. that reach maturity at the same time as current varieties under current temperature regimes. We acknowledge that this is an overly simplistic assumption that does not reflect the range of opportunities and limitations to adaptation (??). Still, this setting allows to demonstrate the role of adaptation and is easy to integrate into the modeling protocol as growing seasons have to be harmonized across groups anyway. The dataset is designed to allow researchers to:

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

The experimental protocol consists of 9 levels for precipitation perturbations, 7 for temperature, 4 for [CO₂], and 3 for applied nitrogen, for a total of 672 simulations for rainfed agriculture and an additional 84 for irrigated (Table 1). For irrigated simulations, limitations from actual water supply are not considered. Limits for the climate variable perturbations are selected to represent reasonable ranges for potential climate changes in the medium term. Note that [CO₂] changes are applied independently of changes in climate variables, so that higher [CO₂] is not associated with higher temperatures. The resulting GGCMI Phase II dataset captures a distribution of crop responses over the potential space of future climate conditions.

25 Each model is run at 0.5 degree spatial resolution and covers all currently cultivated areas and much of the uncultivated land area. (See Figure 1 for the present-day cultivated area of rainfed crops, and Figure S1 in the Supplemental Material for irrigated crops.) To reduce the computational burden, we provide a boolean mask, identifying regions that are unsuitable for cropland for non-climatic reasons. For this, we use a combination of unsuitable areas according to GAEZ (downloaded from www.gaez.iiasa.ac.at/w). **Tom: how did you generate the 90% suitability map?** There are a few cases, where (at the resolution of 0.5 degrees) the grid cell is classified as dominantly unsuitable but the cropland masks assign cropland to these grid cells. To ensure that all cropland currently used in the aggregation is also simulated by the groups, we exclude only pixels that are predominantly unsuitable (>=90%) and do not contain any cropland (according to MIRCA2000 (?)).

Coverage extends considerably outside currently cultivated areas because cultivation will likely shift under climate change. However, areas are not simulated in some cases if they are assumed to remain non-arable even under an extreme climate

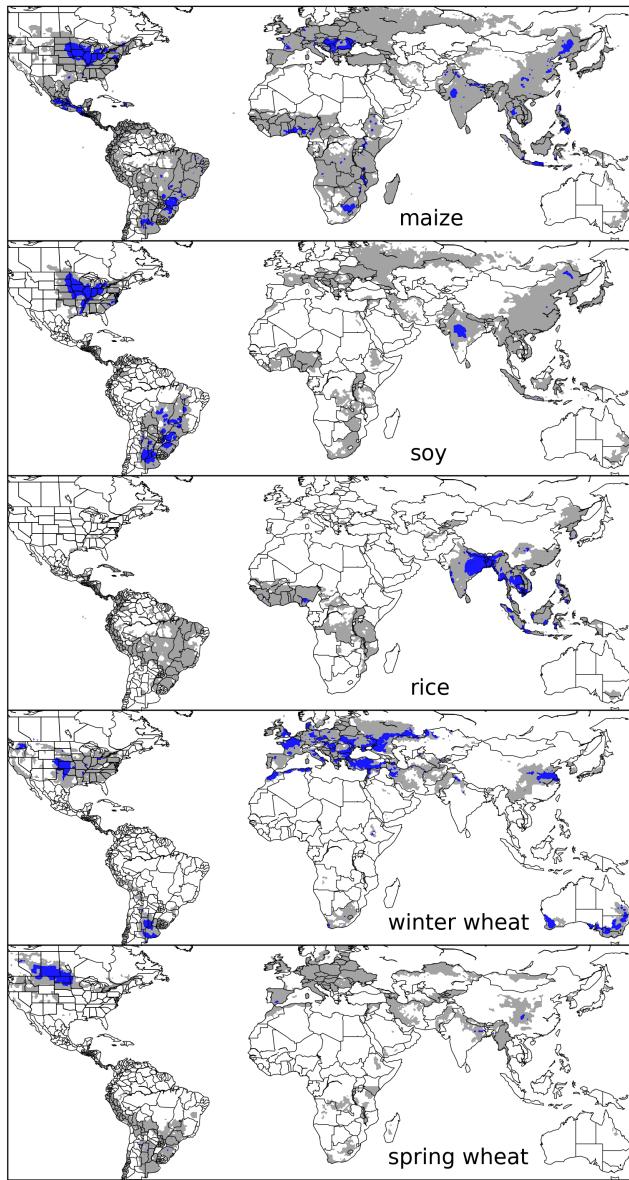


Figure 1. Presently cultivated area for rainfed crops. Blue indicates grid cells with more than 20,000 hectares ($\sim 10\%$ of an equatorial grid cell). Gray contour shows area with more than 10 hectares cultivated. Cultivated areas for maize, rice, and soy are taken from the MIRCA2000 (“monthly irrigated and rainfed crop areas around the year 2000”) dataset (?). Areas for winter and spring wheat areas are adapted from MIRCA2000 data and sorted by growing period. For analogous figure of irrigated crops, see Figure S1.

change; these regions include Greenland, far-northern Canada, Siberia, Antarctica, the Gobi and Sahara Deserts, and central Australia. All models produce as output crop yields ($\text{tons ha}^{-1} \text{ year}^{-1}$) for each 0.5 degree grid cell. Because both yields and yield changes vary substantially across models and across grid cells, we primarily analyze relative change from a baseline. We

Table 1. GGCMI Phase II input levels. Temperature and precipitation values indicate the perturbations from the historical climatology. W- percentage does not apply to the irrigated (W_{inf}) simulations, which are all simulated at the maximum beneficial levels of water. Bold font indicates the ‘baseline’ historical level. One model provided simulations at the T + 5 level. See Figure S3 in the supplement for number of simulations associated with each combination of input levels.

Input variable	Tested range	Unit
[CO ₂] (C)	360 , 510, 660, 810	ppm
Temperature (T)	-1, 0 , 1, 2, 3, 4, 6	°C
Precipitation (W)	-50, -30, -20, -10, 0 , 10, 20, 30, (and W_{inf})	%
Applied nitrogen (N)	10, 60, 200	kg ha ⁻¹
Adaptation (A)	A0: none , A1: new cultivar to maintain original growing season length	-

take as the baseline the scenario with historical climatology (i.e. T and P changes of 0), C of 360 ppm, and applied N at 200 kg ha⁻¹. The GGCMI Phase II simulations are designed for evaluating changes in yield but not absolute yields, since they omit detailed calibrations. To provide some evaluation of the skill of the process-based models used, we repeat the evaluation exercises of ? for GGCMI Phase I.

2.1.1 Input data and harmonization

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 (?), but two models (JULES and PROMET) require sub-daily input data and use alternative baseline climate inputs (WFDEI (?) and ERA-Interim (?), respectively).

Temperature perturbations are applied as absolute offsets from the daily mean, minimum, and maximum temperature time series for each grid cell. Precipitation perturbations are applied as fractional changes at the grid cell level, and [CO₂] and nitrogen levels are specified as discrete values applied uniformly over all grid cells. No other N inputs are assumed (i.e. no atmospheric deposition), but soil mineralization can supply additional plant-available N, which can vary by model.

Growing seasons are standardized across models (with assumptions based on ? and ??), but vary by crop and by location on the globe. For example, maize is sown in March in Spain, in July in Indonesia, and in December in Namibia. Only growing seasons for rainfed systems are used in GGCMI phase II, as full irrigation is treated as one perturbation level in the W dimension (“W_{inf}”). Growing seasons for maize, rice, and soybean have been used from phase 1 (?), but wheat was separated into spring wheat (swh) and winter wheat (wwh) so that new growing season data had to be provided. For this, wheat crop calendars were separated following simple rules. The SAGE crop calendar (?) provides separate spring and winter cropping seasons. This

was used as primary source of information, as for the other crops (?). In areas, where no SAGE crop calendar information is available, the MIRCA2000 crop calendar (?) was used. Wheat growing seasons were assigned to winter wheat if

- the mean monthly temperature is below freezing point (<0°C) at most for 5 months per year,
- 5 – the coldest 3 months of a year are below 10°C
- the season starts after the warmest or before the coldest month of the year on the northern hemisphere, or
- the season starts after the warmest and before the coldest month of the year on the southern hemisphere

and were assigned to spring wheat otherwise. For areas where also not MIRCA2000 crop calendar was available, simulated LPJmL growing seasons (?) were used as in phase 1 (?), applying the same rules as for MIRCA2000 to distinguish spring 10 wheat from winter wheat seasons.

As per convention, N fertilizer is to be applied in two doses to reduce losses to the environment. For simplicity, the first half of the total fertilizer input is to be applied at sowing and the other half is to be applied on day 40 after sowing for all crops except for winter wheat. For winter wheat, the application date for the second N fertilizer application was provided, because the length of dormancy during winter can be quite variable across regions. The second fertilization date for winter wheat was 15 set to the middle day of the first month with average temperatures above 5°C after sowing with a minimum distance to the maturity date of 50 days and to planting of 40 days.

All stresses are disabled other than factors related to nitrogen, temperature, and water (e.g. alkalinity, salinity, non-nitrogen nutrients). No additional nitrogen inputs, such as atmospheric deposition, are considered, but some model treatments of soil organic matter allows additional nitrogen release through mineralization.

20 2.1.2 Output data conventions and file names

All output data is supplied as netCDF version 4 files as in GGCMI phase 1 (?) following a similar naming convention. As the ISIMIP file naming convention now includes the identifier *global* to distinguish regional from global model output (?), we follow that convention here as well:

[model]_[climate]_hist_fullharm_[irrig.scenario]_[variable]_[crop]_global_annual_[start-year]_[end-year].nc4
25 where [model] is the GGCM name, [climate] is the original climate input data set (typically AgMERRA), [irrig.scenario] is the irrigation setting ("firr" for fully irrigated and "noirr" for fully rainfed), [variable] is the output variable (see Table 2), [crop] is the crop abbreviation ("mai" for maize, "ric" for rice, "soy" for soybean, "swh" for spring wheat, "wwh" for winter wheat), [start - year]_[end - year] specify the first and last year recorded on file. The entire file name needs to be in small caps. The first entry in the file is the result of the first simulated cropping cycle that is entirely within the given climate input.
30 For AgMERRA, where the first year provided is 1980, the first harvest record is thus of 1980, when the prescribed sowing and harvest dates are in 1980 (e.g. sowing in March and harvest in September 1980) but is of 1981 if sowing is later in a calendar year than harvest (e.g. sowing in September 1980 and harvest in March 1981). To avoid distortions in harvest events, output files report the sequence of growing periods rather than calendar years. In most cases, this is equivalent, as there is always only

one sowing event per calendar year. As harvest events are internally determined as a function of mostly temperature, these can vary between individual years. If harvest events are around the end of the calendar year (Dec. 31), reported values could contain 2 (one in early January and one in late December), 1 (normal) or none (last was in December of previous year and next is in 5 January of the following year) harvest event if reported per calendar year. As in phase 1, all output data need to be provided on a regular geographic grid, spanning latitudes from -89.75 to 89.75° and longitudes from -179.75 to 179.75° (referring to the center of each grid cell). Missing values, i.e. where no crop growth has been simulated, such as ocean grid cells, need to be reported as 1.e+20f whereas crop failure should be reported as zero yield. Following NetCDF standards, latitude, longitude and time must be included as separate variables in ascending order. Units of these axes are "degrees north", "degrees east", and 10 "growing seasons since YYYY-01-01 00:00:00", where YYYY indicates the first year (1980 for AgMERRA).

Next to simulated yield data, crop modelers are requested to provide additional output variables (Table 2).

A script for initial data quality screening has been provided to the modeling groups so that these could test plausible data range and spatial coverage prior to data submission.

2.2 Models contributing

15 Given the substantial computational requirements for participation in GGCMI Phase II with 756 (A0) + 648 (A1) global 31-year simulations for 5 different crops, we specified different participation tiers to allow submission of smaller sub-sets. These subsets were designed as orthogonal samples across the 4 dimensions of the CTWN space. These are designed by combinations of selected pairs in the CxN space (all 12 vs. 4 pairs: C360_N10, C360_N200, C660_N60, C810_N200) and the TxW space:

- 20 – a *minimum set* with an orthogonal, stratified 9-step sampling (T-1W-10,T0W10,T1W-30,T2W-50,T2W20,T3W30,T4W0,T4Winf,T6W20)
- a *reduced set* of alternating combinations, which uses different W for even Ts than for odd Ts. For even Ts (i.e. T0,T2,T4,T6), we use W = -50,-20,0,+30 = 4*4, for odd Ts (i.e. T-1,T1,T3) , we use W = -30,-10,+10,+30,inf = 3*5,
- and the *full set* (all 7Tx9W)

The different participation levels are thus defined by combining the CxN sets with the TxW sets:

- 25 – **full**: all 756 (648) A0 (A1) simulations (all 12 CxN * all 63 TxW)
- **high tier**: 362 simulations (all 12 CxN combinations * *reduced* TxW set of 31 combinations)
- **mid tier**: 124 simulations (low 4 CxN combinations * *reduced* TxW set of 31 combinations)
- **low tier**: 36 simulations (low 4 CxN combinations * *minimum* TxW set of 9 combinations)

30 The 12 models included in GGCMI Phase II are all process-based crop models that are widely used in impacts assessments (Table 3). Although some models share a common base (e.g. the LPJ family or the EPIC family of models), they have subsequently developed independently. Differences in model structure mean that several key factors are not standardized across the

Table 2. Main output variables requested per crop. Mandatory outputs (if available in that GGCM) in **bold**. Some models provide additional variables, such as more detailed separation of total nitrogen losses.

Variable	variable name	units
Yield	yield_<crop>	t ha⁻¹ yr⁻¹ (dry r
Total above ground biomass yield	biom_<crop>	t ha⁻¹ yr⁻¹ (dry r
Actual planting date	plant-day_<crop>	day of year
Anthesis date	anth-day_<crop>	days from plantin
Maturity date	maty-day_<crop>	days from plantin
Applied irrigation water	pirww_<crop>	mm yr ⁻¹
Evapotranspiration (growing season sum)	etransp_<crop>	mm yr ⁻¹ (firr sce
Transpiration (growing season sum)	transp_<crop>	mm yr ⁻¹
Evaporation (growing season sum)	evap_<crop>	mm yr ⁻¹
Runoff (total growing season sum, subsurface + surface)	runoff_<crop>	mm yr ⁻¹
Total available soil moisture in root zone (growing season sum, basis for computing average soil moisture)	trzpah2o_<crop>	mm yr ⁻¹
Total root biomass	rootm_<crop>	t ha ⁻¹ yr ⁻¹ (dry m
Total Nr uptake (total growing season sum)	tnrup_<crop>	kg ha ⁻¹ yr ⁻¹
Total Nr inputs (total growing season sum)	t nrin_<crop>	kg ha ⁻¹ yr ⁻¹
Total Nr losses (total growing season sum)	t nrloss_<crop>	kg ha ⁻¹ yr ⁻¹
Gross primary production (GPP)	gpp_<crop>	gC m ⁻² yr ⁻¹
Net primary production (NPP)	npp_<crop>	gC m ⁻² yr ⁻¹
CO ₂ response scaler on NPP	co2npp_<crop>	- {0..inf}
water response scaler on NPP	h2onpp_<crop>	- {0..1}
temperature response scaler on NPP	tnpp_<crop>	- {0..1}
Nr response scaler on NPP	nrnpp_<crop>	- {0..1}
Other nutrient response scaler on NPP	ornpp_<crop>	- {0..1}
CO ₂ response scaler on transpiration	co2trans_<crop>	- {0..1}
maximum stress response scaler	maxstress_<crop>	- {0..1}
Maximum LAI	laimax_<crop>	m ² m ⁻²

Table 3. Models included in GGCMI Phase II and the number of C, T, W, and N simulations that each performs, with 756 as the maximum for A0 (no adaptation) and 648 as the maximum for A1 (maintaining growing season adaptation). “N-Dim.” indicates whether the simulations include varying nitrogen levels. Two models provide only one nitrogen level. 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.)

Model (Key Citations)	Maize	Soy	Rice	Winter wheat	Spring wheat	N dim.	Simulations per crop (A0/A1)
APSIM-UGOE, ??	X	X	X	X	X	X	44/36
CARAIB, ??	X	X	X	X	X	-	252/216
EPIC-IIASA, ?	X	X	X	X	X	X	39/0
EPIC-TAMU, ?	X	X	X	X	X	X	765/648
JULES, ???	X	X	X	-	X	-	252/0
GEPIC, ??	X	X	X	X	X	X	430/181
LPJ-GUESS, ??	X	-	-	X	X	X	756/648
LPJmL, ?	X	X	X	X	X	X	756/648
ORCHIDEE-crop, ?	X	-	X	X	-	X	33/0
pDSSAT, ??	X	X	X	X	X	X	756/648
PEPIC, ??	X	X	X	X	X	X	149/121
PROMET, ??	X	X	X	X	X	X [†]	261/232
Totals	12	10	11	11	11	10	5240

experiment, such as carry-over effects across growing years including residue management and soil moisture, and the extent of simulated area for different crops. Not all modeling teams provide the full simulation protocol, for instance, CARAIB and JULES do not simulate the nitrogen dimension and some crops are not parameterized in each model (see Table 3 for details).

5 2.3 Simulation Assessment

The ? 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 (?) by calculating the Pearson product moment correlation coefficient. The procedure is sensitive to the detrending method and the area mask used to aggregate yields. Here we use a 5-year running mean removal and the MICRA area mask for aggregation. In some cases the time series are shifted by one year to account for errors in FAO or model year reporting. The procedure offers no means of assessing CO₂ fertilization, since

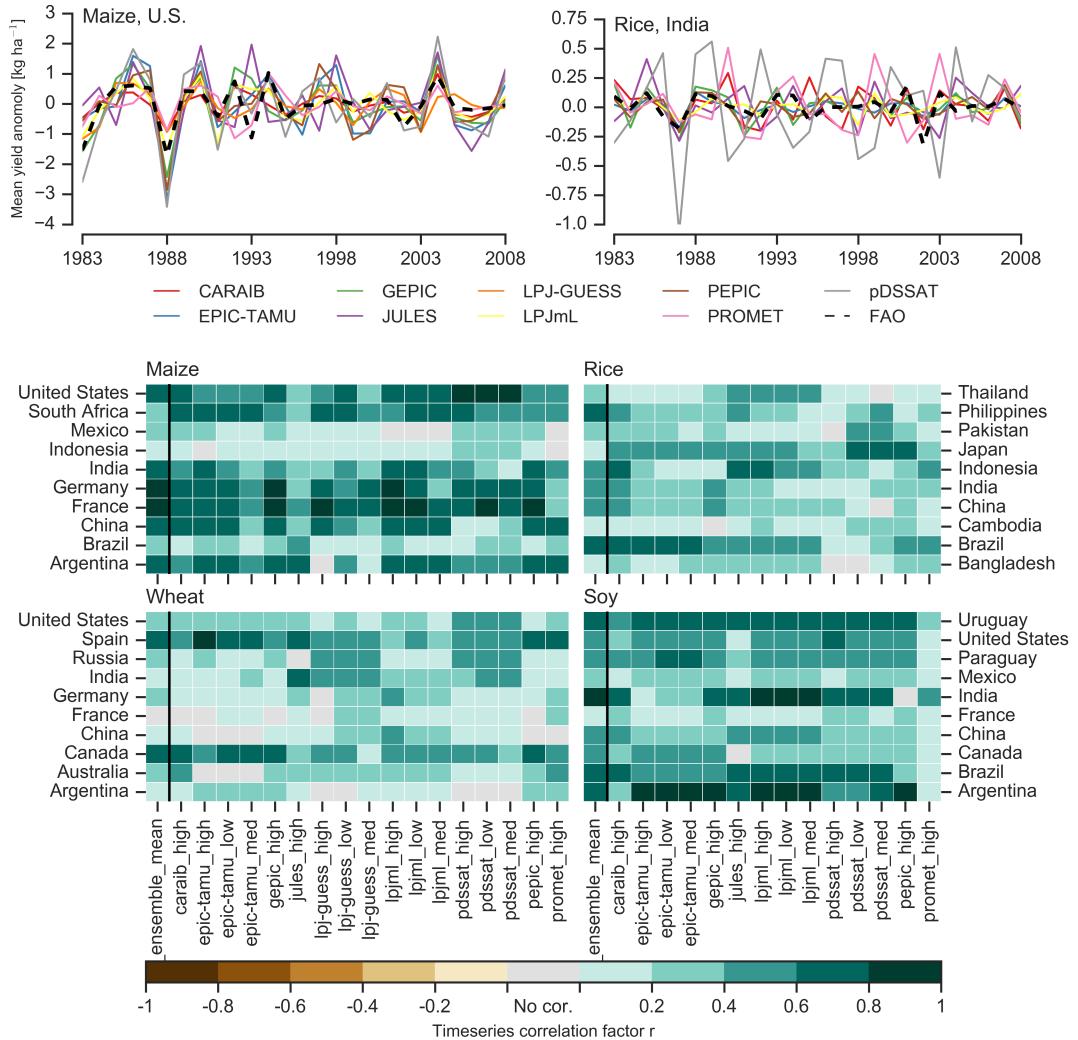


Figure 2. Time series of correlation coefficients between simulated crop yield and FAO data (?) at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen application case. The heatmaps illustrate the Pearson r correlation 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 1981-2010 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 where supplied, else one or the other (see Table 2). The Pearson r correlation coefficients are similar to those of GGCMI Phase I, with reasonable fidelity at capturing year-over-year variation, with differences by region and crop stronger than difference between models as indicated by more horizontal bars than vertical bars of the same color.

[CO₂] has been relatively constant over the historical data collection period. Nitrogen introduces another source of 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.

Results are similar to those of GGCMI Phase I, with reasonable fidelity at capturing year-over-year variation, with differences by region and crop stronger than difference between models. (That is, Figure 9 shows more similarity in horizontal than vertical bars.) No single model is dominant, with each model providing near best-in-class performance in at least one location-crop combination. For example, maize in the United States is consistently well-simulated while maize in Indonesia is problematic (mean Pearson correlation coefficients of 0.68 and 0.18, respectively). In some cases, especially in the developing world, low correlation coefficients may indicate not only model failure but also problems in FAO yield data.

In general, correlation coefficients in GGCMI Phase II are slightly below those of Phase I, likely because of unrealistic nitrogen levels, lack of country level calibration in some models, and restriction to only the MICRA aggregation mask in this study. (Compare Figure 9 to ? Figures 1–4 and 6.) Additionally, the time period used in this case is slightly different from the time period used in ?. Note that in this methodology, simulations of crops with low year-to-year variability such as irrigated rice and wheat will tend to score more poorly than those with higher variability.

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

3 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 rainfed maize in Figure 2, which shows yields across all grid cells for the primary Köppen-Geiger climate regions (?). In warming scenarios with precipitation held constant, all models show decreases in maize yield in the ‘warm temperate’, ‘equatorial’, and ‘arid’ regions that account for nearly three-quarters of global maize production. These impacts are robust for even moderate climate perturbations. In the ‘warm temperate’ zone, even a 1 degree temperature rise with other variables held fixed leads to a median yield reduction that exceeds the variance across models. A 6 degree temperature rise results in median loss of ~25% of yields with a signal to noise ratio of nearly three to one. A notable exception is the ‘cold continental’ region, where models disagree strongly, extending even to the sign of impacts. Other crops show similar responses to warming, with robust yield losses in warmer locations and high inter-model variance in the ‘cold continental’ regions (Figure S5).

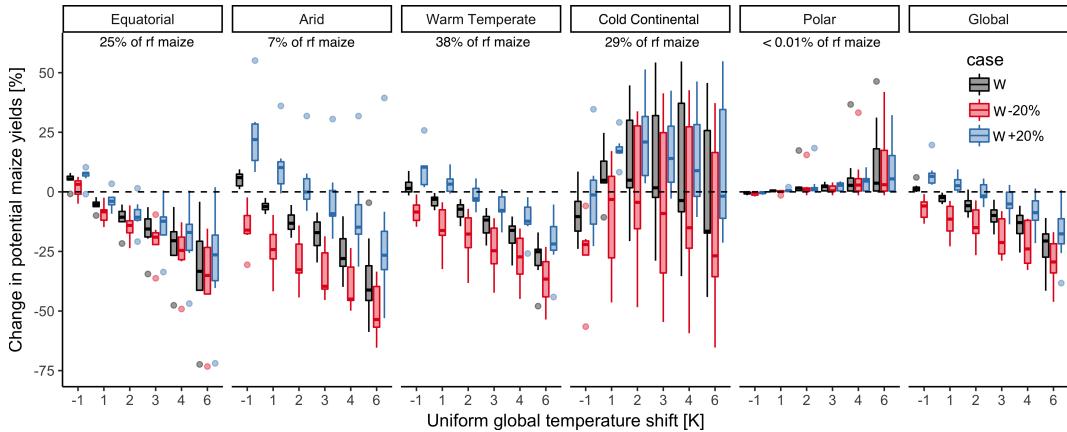


Figure 3. Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (?). We show responses of a single crop (rainfed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation levels (-20%, 0%, and +20%), with $[CO_2]$ and nitrogen held constant at baseline values (360 pmm and $200 \text{ kg ha}^{-1} \text{ yr}^{-1}$). Y-axis is fractional change in the regional average climatological potential yield relative to the baseline. Box-and-whiskers plots show distribution across models, with median marked; edges are first and third quartiles, i.e. box height is the interquartile range (IQR). If all models lie within 1.5-IQR then whiskers extend to maximum and minimum of simulations, else the outlier is shown separately. 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. Figure shows all modeled land area; see Figure S4 in the supplemental material for currently-cultivated land and Figure S5 for other crops. Panel text gives the percentage of rainfed maize presently cultivated in each climate zone (data from ?). Note that ? use the name ‘Snow’ rather than ‘Cold continental’. Outside high-latitude regions (‘Cold continental’ and ‘Polar’), models generally agree, with projected declines under increasing temperatures larger than inter-model variance. The right panel (Global) shows yield responses to a globally uniform temperature shift; note that these results are not directly comparable to simulations of more realistic climate scenarios.

The effects of rainfall changes on maize yields shown in Figure 2 are also as expected and are consistent across models. Increased rainfall mitigates the negative effect of higher temperatures by counteracting the increased evapo-transpiration to some degree, most strongly in arid regions. Decreased rainfall amplifies yield losses and also increases inter-model variance;

10 i.e. models agree that the response to decreased water availability is negative in sign but disagree on its magnitude. We show only rainfed maize here; see Figure S6 for comparison between rainfed and irrigated case. As expected, irrigated crops are more resilient to temperature increases in all regions, especially so where water is limiting. See Figures S7-15 in the supplement for other crops.

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

15 Crop cultivation areas and yield changes with respect to the T+4 scenario show distinct geographic pattern (Figure 3). Absolute yield potentials show strong spatial variation, with much of the Earth’s surface area unsuitable for any of these crops. 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,

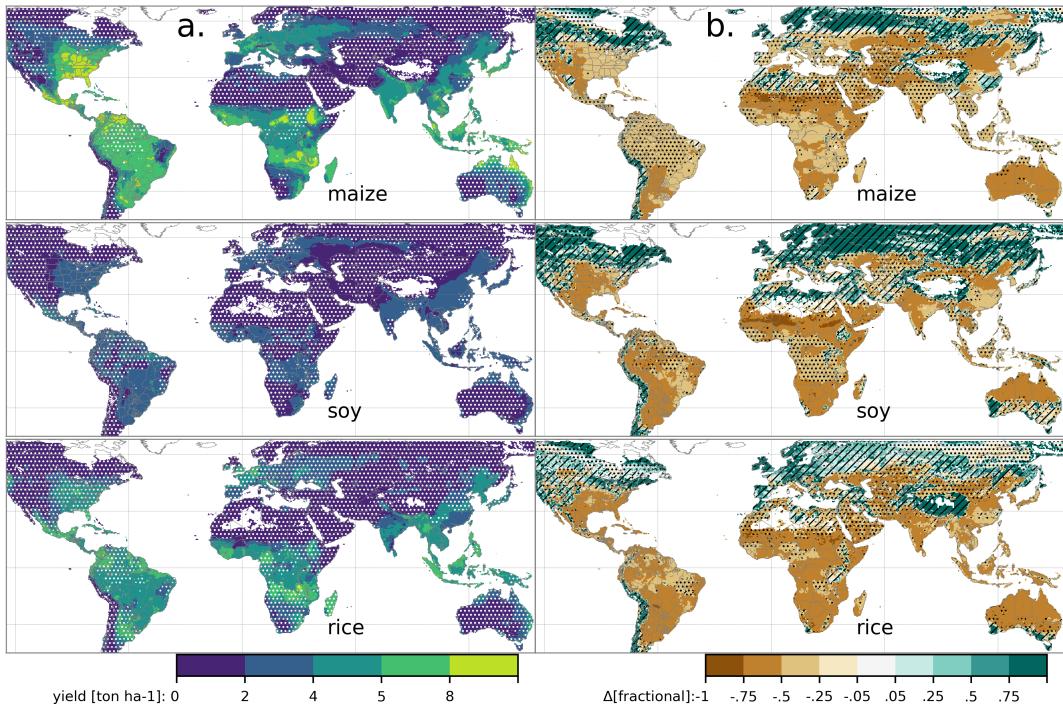


Figure 4. Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top-bottom) rainfed maize, soy, and rice. Wheat shows a qualitatively similar response, see Figure S16 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.

possibly due to how crop failures are considered across different models. For wheat crops see Figure S16 wheat projections are more uncertain, possibly because simulation calibration is especially important (e.g. ?).

4 Discussion and Conclusions

The GGCMI Phase II experiment provides a database designed to allow detailed study of crop yields from process-based models under climate change. The use of systematic input parameter variations facilitates not only comparing the sensitivities 5 of process-based crop yield models to changing climate and management inputs but also evaluating the complex interactions between driving factors (CO_2 , temperature, precipitation, and applied nitrogen). Its global extent also allows identifying geo-

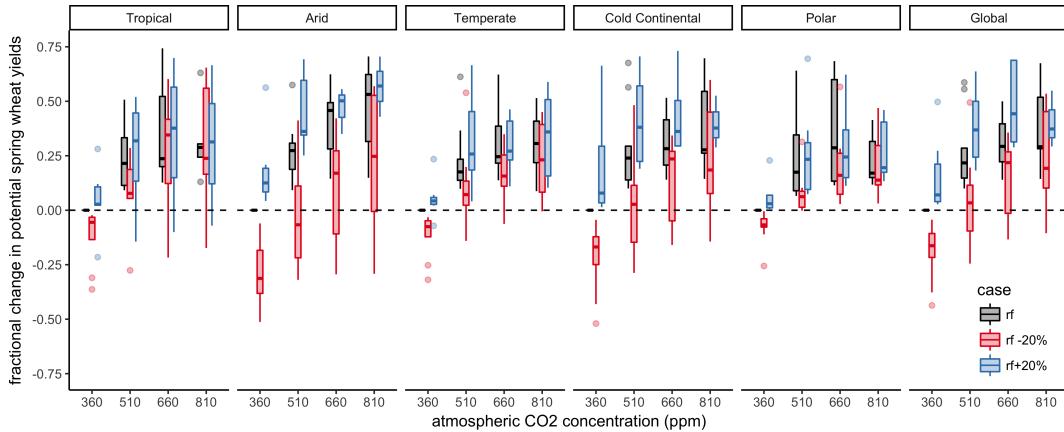


Figure 5. Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions for Carbon.

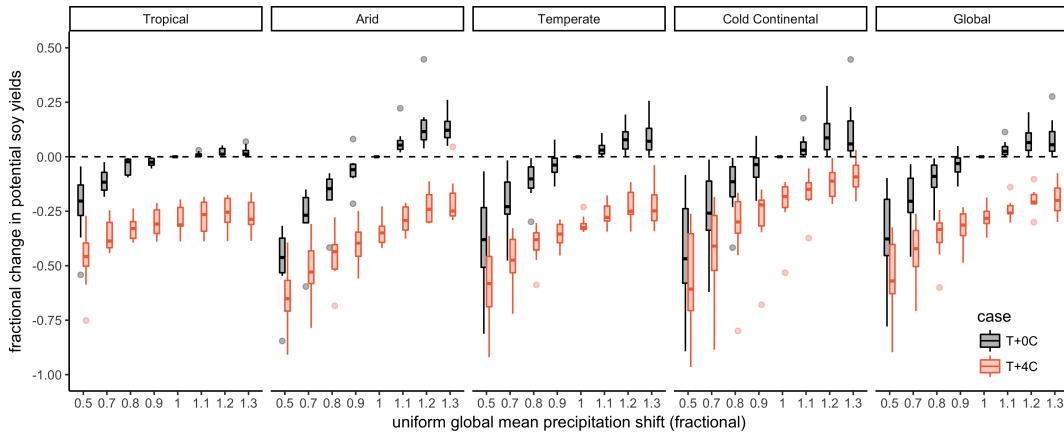


Figure 6. Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions for precip.

graphic shifts in high yield potential locations. We expect that the simulations will yield multiple insights in future studies, and show a selection of preliminary results. We discuss below implications from XXX

First, the GGCMI Phase II simulations allow identifying major areas of uncertainty. Inter-model uncertainty is qualitatively similar across all four inputs tested at the globally aggregate level with some notable exceptions. For example, soy, a nitrogen-fixing legume, is insensitive to nitrogen addition, while wheat is particularly uncertain in its response to CO_2 , levels and water availability (Figure S22). Across geographic regions, projections are most robust in the low latitudes where yield impacts are largest, and most uncertain in the high latitudes where yields may increase. Model differences in projected high-latitude yield changes appear driven more by differences in baseline than in future yields. PROMET, for example, involves a stronger response to cold than does LPJmL, with frost below -8°C irreversibly killing non-winter crops and prolonged periods of

Figure 7. Illustration of the distribution of regional irrigation water demand across the multi-model ensemble, split by Köppen-Geiger climate regions for....

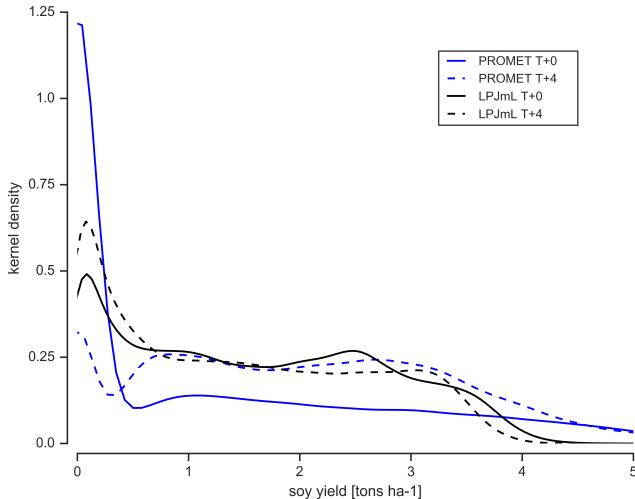


Figure 8. Kernel density estimate of soy yields north of 45° latitude for the PROMET and LPJmL models. Solid lines show the historical climatology and dashed lines show the T+4 (K) case. Note strong reduction in the lowest bin for the T+4 case for PROMET that is spread almost equally over the rest of the distribution.

below-optimum temperatures also leading to complete crop failure. Over the high-latitudes regions simulated by both models, 52% of grid cells in PROMET report 0 yield in the present climate vs. 11% of cells in the T+4 scenario, leading to a strong yield gain in warmer future climates. In LPJmL, the same high-latitude area is suitable for cultivation even in baseline climate, with crop failure rates of 4% and 5% in present and T+4 cases, so that projected yield changes are modest (Figure S23.)

Second, the GGCMI Phase II simulations demonstrate the sensitivity of climate-driven yield impacts to the locations of cultivated land. One counterintuitive result apparent in the simulations is that warmer temperatures drive steeper yield reductions in irrigated than rainfed maize when considered only over currently cultivated land, even though water availability increases 5 crop resiliency to temperature increases at any given location (compare Figure ?? to Figure 2 and Figures S6 to S7). The effect results from geographic differences in cultivation: irrigated maize is grown in warmer locations where the impacts of warming are more severe. (See Figures S8-S15 for other crops.) Geographic effects also mean that nitrogen fertilization produces stronger responses in irrigated than non-irrigated wheat and maize, presumably because those rainfed crops are limited by water availability (Figure S21).

In general, the development of multi-model ensembles involving systematic parameters sweeps has large promise for increasing understanding of potential future crop responses and for improving process-based crop models.

Code and data availability. The simulation outputs of the mandatory 7 output variables (Table 2) are available on zenodo.org. See Appendix A1 for data DOIs. All other simulation output variables are available upon request to the corresponding author. The scripts for generating the spring wheat and winter wheat growing seasons and second fertilizer dates and the quality screening script is available
15 at <https://github.com/RDCEP/ggcmi/blob/phase2/>. All input data are available via globus.org (registration required, free of charge): Minimum cropland mask is available at https://www.globus.org/app/transfer?origin_id=e4c16e81-6d04-11e5-ba46-22000b92c6ec&origin_path=%2FAGMIP.input%2FCTWN%2F choose the file boolean_cropmask_ggcmi_phase2.nc4 Growing period data for wheat is now divided up into winter and spring wheat, available at https://www.globus.org/app/transfer?origin_id=e4c16e81-6d04-11e5-ba46-22000b92c6ec&origin_path=%2FAGMIP.input%2Fother.inputs%2FAGMIP_GROWING_SEASON.HARM.version2.0%2F whereas all other growing season data (maize, rice, soybean) are the same as in phase 1 (version 1.25), available at https://www.globus.org/app/transfer?origin_id=e4c16e81-6d04-11e5-ba46-22000b92c6ec&origin_path=%2FAGMIP.input%2Fother.inputs%2FAGMIP_GROWING_SEASON.HARM.version1.25%2F

Appendix A

A1 Data Access

25 Simulation yield output dataset can be found at the DOIs located in table A1.

Author contributions. J.E., C.M, A.R., J.F., and E.M. designed the research. C.M., J.J., J.B., P.C., M.D., P.F., C.F., L.F., M.H., C.I., I.J., C.J., N.K., M.K., W.L., S.O., M.P., T.P., A.R., X.W., K.W., and F.Z. performed the simulations. J.F., J.J., A.S., M.L., and E.M. performed the analysis and J.F. and E.M. prepared the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. We thank Michael Stein and Kevin Schwarzwald, who provided helpful suggestions that contributed to this work. This research was performed as part of the Center for Robust Decision-Making on Climate and Energy Policy (RDCEP) at the University of Chicago, and was supported through a variety of sources. RDCEP is funded by NSF grant #SES-1463644 through the Decision Making
5 Under Uncertainty program. J.F. was supported by the NSF NRT program, grant #DGE-1735359. C.M. was supported by the MACMIT project (01LN1317A) funded through the German Federal Ministry of Education and Research (BMBF). C.F. was supported by the European Research Council Synergy grant #ERC-2013-Syng-610028 Imbalance-P. P.F. and K.W. were supported by the Newton Fund through the Met Office Climate Science for Service Partnership Brazil (CSSP Brazil). K.W. was supported by the IMPREX research project supported by the European Commission under the Horizon 2020 Framework programme, grant #641811. A.S. was supported by the Office of Science of
10 the U.S. Department of Energy as part of the Multi-sector Dynamics Research Program Area. S.O. acknowledges support from the Swedish strong research areas BECC and MERGE together with support from LUCCI (Lund University Centre for studies of Carbon Cycle and Climate Interactions). R.C.I. acknowledges support from the Texas Agrilife Research and 634 Extension, Texas A &M University. This is

Table A1. DOI's for model yield data outputs. All yield output data can be found at <https://doi.org/10.5281/zenodo/XX>. Where XX is the value found in the table.

Model	Maize	Soy	Rice	Winter wheat	Spring wheat
APSIM-UGOE	2582531	258235	2582533	2582537	2582539
CARAIB	2582522	2582508	2582504	2582516	2582499
EPIC-IIASA	2582453	2582461	2582457	2582463	2582465
EPIC-TAMU	2582349	2582367	2582352	2582392	2582418
JULES	2582543	2582547	2582545	–	2582551
GEPIC	2582247	258225	2582251	2582260	2582263
LPJ-GUESS	2581625	–	–	2581638	2581640
LPJmL	2581356	2581498	2581436	2581565	2581606
ORCHIDEE-crop	2582441	–	2582445	2582449	–
pDSSAT	2582111	2582147	2582127	2582163	2582178
PEPIC	2582341	2582433	2582343	2582439	2582455
PROMET	2582467	2582488	2582479	2582490	2582492

paper number 35 of the Birmingham Institute of Forest Research. Computing resources were provided by the University of Chicago Research Computing Center (RCC).