

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

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

<sup>a</sup>Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA

<sup>b</sup>Center for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA

<sup>c</sup>Department of Computer Science, University of Chicago, Chicago, IL, USA

<sup>d</sup>Potsdam Institute for Climate Impact Research, Leibniz Association (Member), Potsdam, Germany

<sup>e</sup>NASA Goddard Institute for Space Studies, New York, NY, United States

<sup>f</sup>Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA

<sup>g</sup>Ecosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria

<sup>h</sup>Department of Soil Science, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovak Republic

<sup>i</sup>Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France

<sup>j</sup>Sino-French Institute of Earth System Sciences, College of Urban and Environmental Sciences, Peking University, Beijing, China

<sup>k</sup>Unité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d'Astrophysique et de Géophysique, University of Liège, Belgium

<sup>l</sup>Met Office Hadley Centre, Exeter, United Kingdom

<sup>m</sup>Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany

<sup>n</sup>Georg-August-University Göttingen, Tropical Plant Production and Agricultural Systems Modelling, Göttingen, Germany

<sup>o</sup>Department of Geographical Sciences, University of Maryland, College Park, MD, USA

<sup>p</sup>Texas AgriLife Research and Extension, Texas A&M University, Temple, TX, USA

<sup>q</sup>Department of Statistics, University of Chicago, Chicago, IL, USA

<sup>r</sup>EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

<sup>s</sup>Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

<sup>t</sup>Earth Institute Center for Climate Systems Research, Columbia University, New York, NY, USA

<sup>u</sup>Karlsruhe Institute of Technology, IMK-IFU, 82467 Garmisch-Partenkirchen, Germany.

<sup>v</sup>School 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. Results of these models are often used to inform subsequent model-based analyses in agricultural economics and integrated assessment, but the computational requirements of process-based models sometimes hamper the ability to utilize these models in existing impact assessment frameworks at high resolutions globally. Conducting a large simulation set with perturbations in atmospheric CO<sub>2</sub> concentrations, temperature, precipitation, and nitrogen inputs, Phase II of the Global Gridded Crop Model Intercomparison (GGCMI), an activity of the Agricultural Model Intercomparison and Improvement Project (AgMIP), constitutes an unprecedentedly data-rich basis of projected yield changes across 12 models and five crops (maize, soy, rice, spring wheat, and winter wheat) using global gridded simulations. Results from Phase II of the GGCMI effort, a targeted experiment aimed at understanding the interaction between multiple climate variables (as well as management) are presented first. Then the construction of an “emulator or statistical representation of the simulated 30-year mean climatological output in each location is outlined. The emulated response surfaces capture the details 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

Projecting crop yield response to a changing climate is of great importance, 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. Multiple studies with different crop or climate models predict sharp reduction in yields on currently cultivated cropland under business-as-usual climate scenarios, although their yield projections show considerable spread (e.g. Porter et al. (IPCC), 2014, Rosenzweig et al., 2014, Schauberger et al., 2017, and references therein). Model differences are unsurprising because crop responses in

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

models can be complex, with crop growth a function of complex interactions between climate inputs and management practices.

Computational Models have been used to project crop yields since the 1950's, beginning with statistical models (Heady, 1957, Heady & Dillon, 1961) that attempt to capture the relationship between input factors and resultant yields. These statistical models were typically developed on a small scale for locations with extensive histories of yield data. The emergence of computers allowed development of numerical models that simulate the process of photosynthesis and the biology and phenology of individual crops (first proposed by de Wit (1957), Duncan et al. (1967) and attempted by Duncan (1972)). Historical mapping of crop model development can be found in the appendix/supplementary of Rosenzweig et al. (2014). A half-century of improvement in both models and computing resources means that researchers can now run crop simulation models for many years at high spatial resolution on the global scale.

Both types of models continue to be used, and comparative studies have concluded that when done carefully, both approaches can provide similar yield estimates (e.g. Lobell & Burke, 2010, Moore et al., 2017, Roberts et al., 2017, Zhao et al., 2017). Models tend to agree broadly in major response patterns, including a reasonable representation of the spatial pattern in historical yields of major crops (e.g. Elliott et al., 2015, Müller et al., 2017) and projections of decreases in yield under future climate scenarios.

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

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

critical. Second, only process-based models can capture the growth response to elevated CO<sub>2</sub>, novel conditions that are not represented in historical data (e.g. Pugh et al., 2016, Roberts et al., 2017). Similarly, only process-based models can represent novel changes in management practices (e.g. fertilizer input) that may ameliorate climate-induced damages.

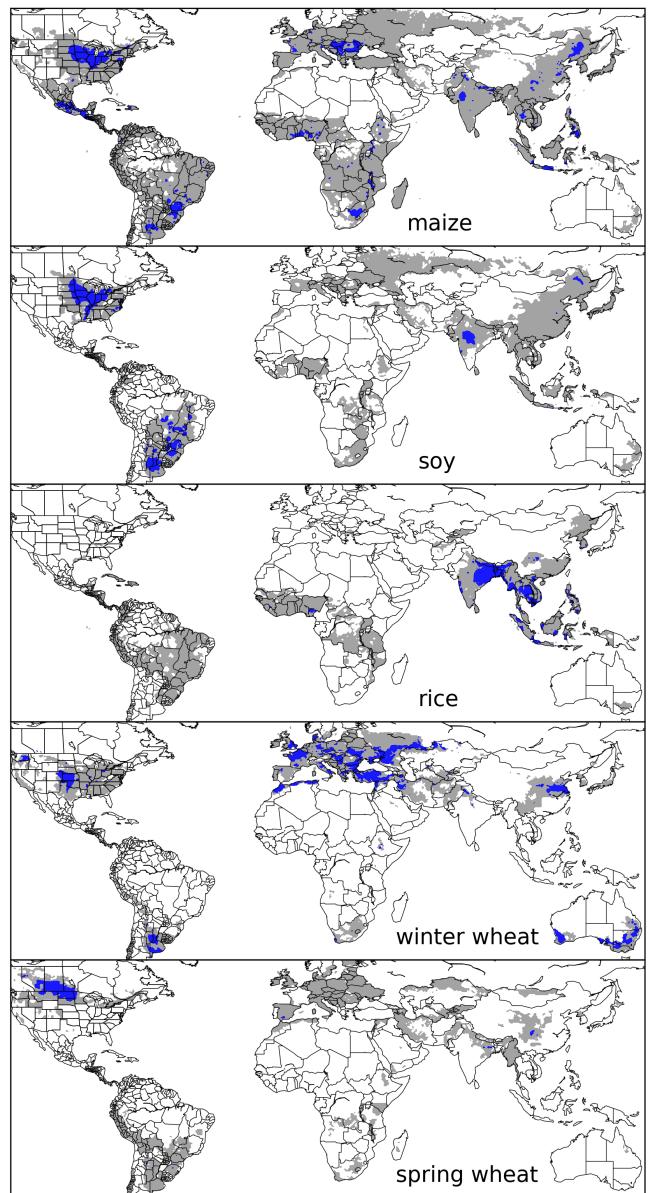


Figure 1: Presently cultivated area for rain-fed crops in the real world. Blue contour area indicates grid cells with more than 20,000 hectares (approx. 10% of the land in a grid cell near the equator for example) of crop cultivated. Gray contour shows area with more than 10 hectares cultivated. Cultivated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

Statistical emulation of crop simulations offers the possibility of combining some advantageous features of both statistical and process-based models. The statistical representation of complicated numerical simulation (e.g. O'Hagan, 2006, Conti et al.,

2009), in which simulation output acts as the training data for a<sub>126</sub> statistical model, has been of increasing interest with the growth<sub>127</sub> of simulation complexity and volume of output. Such emula<sub>128</sub> tors or “surrogate models” have been used in a variety of fields<sub>129</sub> including hydrology (Razavi et al., 2012), engineering (Storlie<sub>130</sub> et al., 2009), environmental sciences (Ratto et al., 2012), and<sub>131</sub> climate (Castruccio et al., 2014). For agricultural impacts stud<sub>132</sub> ies, emulation of process-based models allows exploring crop<sub>133</sub> yields in regions outside ranges of current cultivation and with<sub>134</sub> input variables outside historical precedents, in a lightweight,<sub>135</sub> flexible form that is compatible with economic studies.<sub>136</sub>

Crop yield emulators have been proposed and implemented<sub>137</sub> by many studies (e.g. Howden & Crimp, 2005, Räisänen &<sub>138</sub> Ruokolainen, 2006, Lobell & Burke, 2010, Iizumi et al., 2010,<sub>139</sub> Ferrise et al., 2011, Holzkämper et al., 2012, Ruane et al., 2013,<sub>140</sub> Makowski et al., 2015), and in the last several years multiple<sub>141</sub> studies have developed emulators based on a variety of crop<sub>142</sub> simulation model outputs. Several studies have developed an<sub>143</sub> emulator for a single crop model run on a RCP climate scenario<sub>144</sub> set (e.g. Oyebamiji et al., 2015). Multiple groups (e.g. Blanc<sub>145</sub> & Sultan, 2015, Blanc, 2017, Ostberg et al., 2018), successfully<sub>146</sub> constructed emulators for a 5-crop-model intercomparison ex<sub>147</sub> ercise performed as part of ISIMIP (Warszawski et al., 2014),<sub>148</sub> the Inter-Sectoral Impacts Model Intercomparison Project and<sub>149</sub> evaluated several different climate scenarios (over multiple cli<sub>150</sub> mate model runs). Several other studies (e.g. Moore et al.,<sub>151</sub> 2017, Mistry et al., 2017) utilize a hybrid simulation output and<sub>152</sub> real-world data approach to develop and emulator or damage<sub>153</sub> function. Additional recent studies have explored an impact re<sub>154</sub> sponse surface (aka. emulator when using simulated data) over<sub>155</sub> an explicit multivariate input simulation space (as opposed to<sub>156</sub> specific RCP climate model runs), with a site-based approach<sub>157</sub> (as opposed to a globally gridded model) across temperature,<sub>158</sub> water, and CO<sub>2</sub> sampling (Snyder et al., 2018), or with models<sub>159</sub> for wheat across water and temperature dimensions for different<sub>160</sub> sites in Europe (Fronzek et al., 2018).<sub>158</sub>

The Global Gridded Crop Model Intercomparison (GGCMI)<sub>159</sub> Phase II experiment is an attempt to expand upon previous<sub>160</sub> process-based crop modeling studies by running globally grid<sub>161</sub> ded crop models over a set of uniform input dimensions as op<sub>162</sub> posed to RCP climate scenarios in order to focus on testing the<sub>163</sub> sensitivity to yield drivers within and across models. GGCMI<sub>164</sub> is a multi-model exercise conducted as part of the Agricul<sub>165</sub> tural Model Intercomparison and Improvement Project (Ag<sub>166</sub> MIP, (Rosenzweig et al., 2013, 2014)), which brings together<sub>167</sub> major global crop simulation models from different research<sub>168</sub>

organizations around the world under a framework similar to the Climate Model Intercomparison Project (CMIP, Taylor et al., 2012, Eyring et al., 2016). The GGCMI analysis framework builds on the AgMIP Coordinated Climate-Crop Modeling Project (C3MP, Ruane et al., 2014, McDermid et al., 2015), and will contribute to the AgMIP Coordinated Global and Regional Assessments (CGRA, Ruane et al., 2018, Rosenzweig et al., 2018).

The GGCMI Phase II project develops global simulations of yields of major crops under scenarios that sample a uniform parameter space. Overall goals include understanding where highest-yield regions may shift under climate change, exploring future adaptive management strategies, understanding how interacting parameters affect crop yields, quantifying uncertainties, and testing strategies for producing lightweight statistical emulations of the more detailed process-based models. In the remainder of this paper, we describe the GGCMI Phase II experiments, present the simulation database output (for public use) and initial overall results. We also present an example climatological-mean yield emulator as a distillation of the dataset and as a potential tool for impact assessments.

We do not present all the final insights to be gained from this model intercomparison project, or the best possible emulation of the year-to-year response to changes the input dimensions.

## 2. Materials and Methods

### 2.1. GGCMI Phase II: experiment design

GGCMI Phase II is the continuation of a multi-model comparison exercise begun in 2014. The initial Phase I compared harmonized yields of 21 models for 19 crops over a historical (1980-2010) scenario with a primary goal of model evaluation (Elliott et al., 2015, Müller et al., 2017). Phase II compares simulations of 12 models for 5 crops (maize, rice, soybean, spring wheat, and winter wheat) over hundreds of scenarios in which individual climate or management inputs are adjusted from their historical values. The reduced set of crops includes the three major global cereals and the major legume and accounts for over 50% of human calories (in 2016, nearly 3.5 billion tons or 32% of total global crop production by weight (Food and Agriculture Organization of the United Nations, 2018).

The major goals of GGCMI Phase II are to:

- Enhance understanding of how models work by characterizing their sensitivity to input climate and nitrogen drivers.

Input variable	Abbr.	Tested range	Unit
CO <sub>2</sub>	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 <sub>inf</sub> )	%
Applied nitrogen	N	10, 60, 200	kg ha <sup>-1</sup>

Table 1: Phase II input variable test levels. Temperature and precipitation values indicate the perturbations from the historical, climatology. \* Only simulated by one model. W-percentage does not apply to the irrigated (W<sub>inf</sub>) simulations.

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

Table 2: Models included in the GGCMI Phase II and the number of C, T, W, and N simulations performed for rain-fed crops (“Sims per Crop”), with 672 as the maximum. “N-Dim.” indicates if simulations include varying nitrogen levels; two models omit this dimension. All models provided the same set of simulations across all modeled crops, but some omitted individual crops in some cases. (For example, APSIM did not simulate winter wheat.) Irrigated simulations are provided at the level of the other covariates for each model (for an additional 84 simulations for the fully-sampled models). Geographic extent of simulation varies to some extent within a certain model for different scenarios (672 rain-fed simulations does not necessarily equal 672 climatological yields in all areas). This geographic variance only applies for areas far outside the area of currently cultivated crops. Two models (marked with \*) use non-AgMERRA climate inputs. For further details on models, see Elliott et al. (2015). †PROMET provided simulations at only two nitrogen levels so is not emulated across the nitrogen dimension.

- Study the interactions between climate variables and nitrogen inputs in driving modeled yield impacts. 197
- Explore differences in crop response to warming across the Earth’s climate regions. 199
- Provide a dataset that allows statistical emulation of crop model responses for downstream modelers. 202
- Illustrate differences in potential adaptation via growing season changes. 204

solute 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 CO<sub>2</sub> changes are applied independently of changes in climate variables, so that higher CO<sub>2</sub> is not associated with higher temperatures. An additional, identical set of scenarios (at the same C, T, W, and N levels) simulate adaptive agronomy under climate change by varying the growing season for crop production. (These adaptation simulations are not shown or analyzed here.) The resulting GGCMI data set captures a distribution of crop responses over the potential space of future climate conditions.

The 12 models included in GGCMI Phase II are all mechanistic process-based crop models that are widely used in impacts assessments (Table 2). Although some of the models share a common base (e.g. LPJmL and LPJ-GUESS and the EPIC models), they have developed independently from this shared base, for more details on the genealogy of the models see Figure S1 in Rosenzweig et al. (2014). Differences in model structure does mean that several key factors are not standardized across the experiment, including secondary soil nutrients, carry over effects across growing years including residue management and soil moisture, and extent of simulated area for different crops. Growing seasons are identical across models, but vary by crop and by location on the globe. All stresses except factors related to nitrogen, temperature, and water (e.g. Alkalinity, salinity) are disabled. No additional nitrogen inputs, such as atmospheric deposition, are considered, but some mod-

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

The experimental protocol consists of 9 levels for precipitation perturbations, 7 for temperature, 4 for CO<sub>2</sub>, 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 ab-

els have individual assumptions on soil organic matter that may<sub>283</sub> release additional nitrogen through mineralization. See Rosen<sub>284</sub> zweig et al. (2014), Elliott et al. (2015) and Müller et al. (2017)<sub>285</sub> for further details on models and underlying assumptions.

Each model is run at 0.5 degree spatial resolution and covers<sub>286</sub> all currently cultivated areas and much of the uncultivated land area. Coverage extends considerably outside currently cultivated areas because cultivation will likely shift under climate change. See Figure 1 for the present-day cultivated area of rain-fed crops, and Figure ?? in the supplemental material for irrigated crops. Some areas such as Greenland, far-northern Canada, Siberia, Antarctica, the Gobi and Sahara deserts, and central Australia are not simulated as they are assumed to remain non-arable even under an extreme climate change.

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

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

## 2.2. Simulation model validation approach

Simulation model validation for GGCMI phase II builds on the validation efforts presented in Müller et al. (2017) for the first phase. In the case presented here however, the models<sub>288</sub> are not run on the best approximation of management levels<sub>289</sub> (namely nitrogen application level) by country as with phase I.<sub>290</sub> As the goals of this phase of the project are focused on under-<sub>291</sub> standing the sensitivity in *change* in yield to changes in input<sub>292</sub> drivers –and not to simulate historical yields as accurately as<sub>293</sub> possible— no direct comparison to historical yield data can be<sub>294</sub> made. Additionally, even when provided with an appropriate<sub>295</sub> local nitrogen level, models simulated *potential* yields that do<sub>296</sub> not include reductions from pests, weeds, or diseases. Potential<sub>297</sub> yields represent an ideal case that is not realized in many<sub>298</sub> less industrialized areas. Finally, some models are not calibrated<sub>299</sub> as they were in phase I of the project.

We evaluate the models here based on the response to year-<sub>300</sub> to-year temperature and precipitation variability in the historical record. If the models can (somewhat) faithfully represent<sub>301</sub> the historical variability in yields (which, once detrended<sub>302</sub> to account for changing management levels must be driven<sub>304</sub> largely by differences in weather), then the models may provide<sub>305</sub> some utility in understanding the impact on mean climatic<sub>306</sub> shifts in temperature and precipitation. Specifically,<sub>307</sub> we calculate a Pearson correlation coefficient between the de-<sub>308</sub> trended time series of simulations and FAO data for the period<sub>309</sub>

1981-2009. Validating the response to CO<sub>2</sub> and Nitrogen applications is more difficult because real world data is not available outside of small greenhouse and field level trials.

## 2.3. Climatological-mean yield emulator design

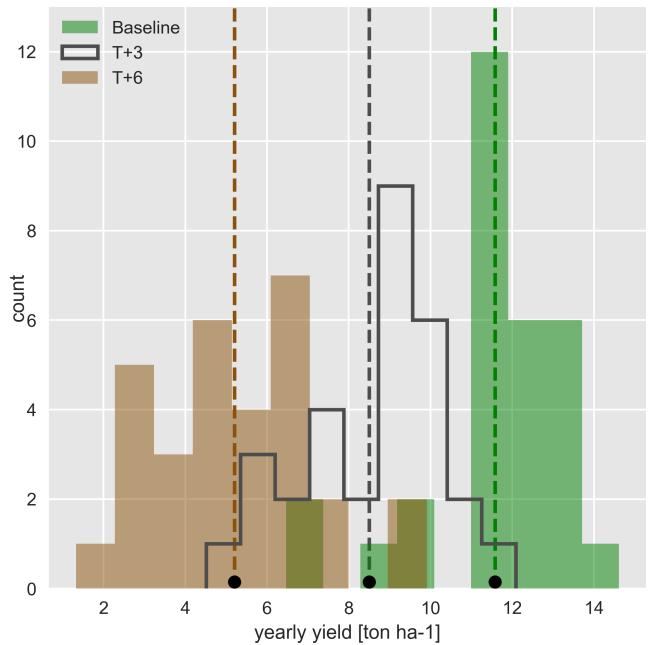


Figure 2: Example showing both climatological mean yields and distribution of yearly yields for three 30-year scenarios. Figure shows rain-fed maize for a grid cell in northern Iowa (a representative high-yield region) from the pDSSAT model, for the baseline climatology (1981-2010) and for scenarios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

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

The decision to first construct a climatological-mean yield emulator is driven by the target application for this analysis tool. Many impact modelers are not focused on the changes in the year-to-year variability in yields, but instead on the broad mean changes over the multi-decadal timescale. Emulation involves fitting individual regression models for each crop, simulation model, and 0.5 degree geographic pixel from the GGCMI Phase II data set. The regressors are the applied constant perturbations in temperature, water, nitrogen and CO<sub>2</sub>, we aggregate the simulation outputs in the time dimension, and regress on the 30-year mean yields. (See Figure 2 for illustration.) The regression therefore omits information about yield responses to year-to-year climate perturbations, which are more complex. Emulating inter-annual yield variations would likely require considering statistical details of the historical climate time series, including changes in marginal distribution and temporal dependencies. (Future work should explore this). The climatological emulation indirectly includes any yield response to geographically distributed factors such as soil type, insolation, and the

310 baseline climate itself, because we construct separate emula-<sup>339</sup>  
 311 tors for each grid cell. The emulator parameter matrices are<sup>340</sup>  
 312 portable and the yield computations are cheap even at the half-<sup>341</sup>  
 313 degree grid cell resolution, so we do not aggregate in space at<sup>342</sup>  
 314 this time.<sup>343</sup>

315 Blanc & Sultan (2015) and Blanc (2017) have shown that a<sup>344</sup>  
 316 fractional polynomial specification is more effective than a stan-<sup>345</sup>  
 317 dard polynomial for representing simulations at the yearly level<sup>346</sup>  
 318 across different soil types geographically (not at the grid cell<sup>347</sup>  
 319 level). We do not test this specification here, and instead use as<sup>348</sup>  
 320 a starting point a standard third-order polynomial to represent<sup>349</sup>  
 321 the climatological-mean response at the grid cell level as it is<sup>350</sup>  
 322 the simplest effective specification.<sup>351</sup>

323 We regress climatological-mean yields against a third-order<sup>352</sup>  
 324 polynomial in C, T, W, and N with interaction terms. The<sup>353</sup>  
 325 higher-order terms are necessary to capture any nonlinear re-<sup>354</sup>  
 326 spondes, which are well-documented in observations for tem-<sup>355</sup>  
 327 perature and water perturbations (e.g. Schlenker & Roberts<sup>356</sup>  
 328 (2009) for T and He et al. (2016) for W). We include inter-<sup>357</sup>  
 329 action terms (both linear and higher-order) because past stud-<sup>358</sup>  
 330 ies have shown them to be significant effects. For example,<sup>359</sup>  
 331 Lobell & Field (2007) and Tebaldi & Lobell (2008) showed<sup>360</sup>  
 332 that in real-world yields, the joint distribution in T and W is<sup>361</sup>  
 333 needed to explain observed yield variance (C and N are fixed<sup>362</sup>  
 334 in these data). Other observation-based studies have shown the<sup>363</sup>  
 335 importance of the interaction between water and nitrogen (e.g.<sup>364</sup>  
 336 Aulakh & Malhi, 2005), and between nitrogen and carbon diox-<sup>365</sup>  
 337 ide (Osaki et al., 1992, Nakamura et al., 1997). We do not focus<sup>366</sup>  
 338 on comparing different model specifications in this study, and<sup>367</sup>

instead stick to a relatively simple parameterized specification  
 that allows for some, albeit limited, coefficient interpretation.

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

Feature importance is remarkably consistent across all three  
 models and across all crops (see Figure ?? in the supplemental  
 material). The feature selection process results in a final poly-  
 nomial in 23 terms, with 11 terms eliminated. We omit the  $N^3$   
 term, which cannot be fitted because we sample only three ni-  
 trogen 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 in-  
 teraction 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

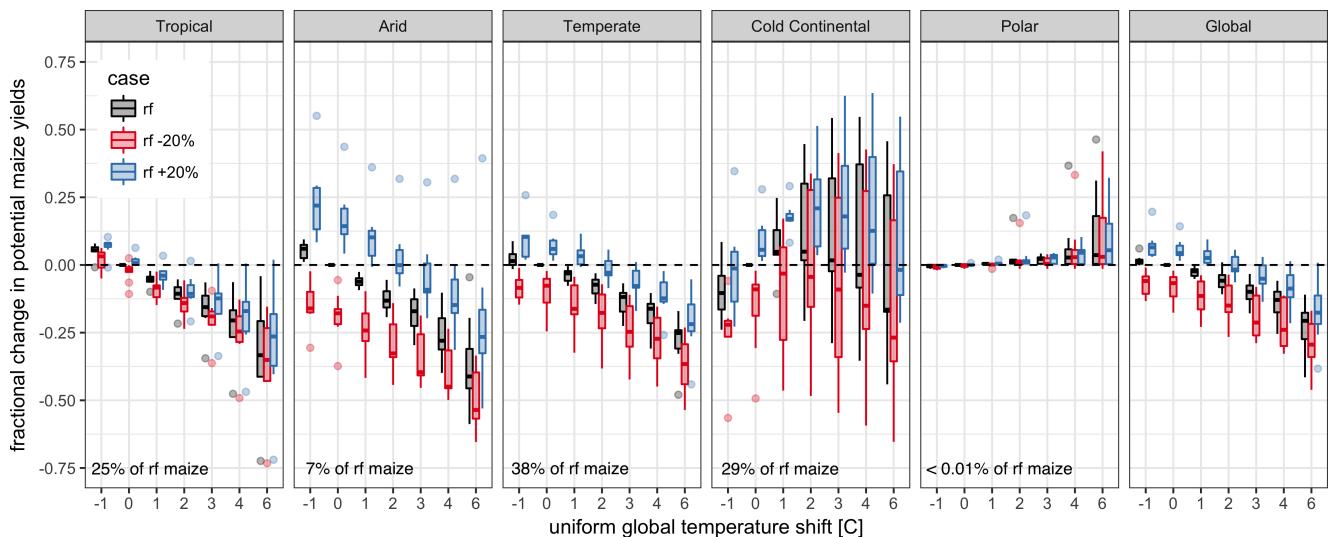


Figure 3: Illustration of the distribution of regional yield changes across the multi-model ensemble, split by Köppen-Geiger climate regions (Rubel & Kottek (2010)). We show responses of a single crop (rain-fed maize) to applied uniform temperature perturbations, for three discrete precipitation perturbation levels (rain-fed (rf) -20%, rain-fed (0), and rain-fed +20%), with CO<sub>2</sub> and nitrogen held constant at baseline values (360 ppm and 200 kg ha<sup>-1</sup> yr<sup>-1</sup>). 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 ?? in the supplemental material for only currently-cultivated land. Panel text gives the percentage of rain-fed maize presently cultivated in each climate zone (Portmann et al., 2010). Box-and-whiskers plots show distribution across models, with median marked. Box edges are first and third quartiles, i.e. box height is the interquartile range (IQR). Whiskers extend to maximum and minimum of simulations but are limited at 1.5-IQR; otherwise the outlier is shown. Models generally agree in most climate regions (other than cold continental), with projected changes larger than inter-model variance. Outliers in the tropics (strong negative impact of temperature increases) are the pDSSAT model; outliers in the high-rainfall case (strong positive impact of precipitation increases) are the JULES model. Inter-model variance increases in the case where precipitation is reduced, suggesting uncertainty in model response to water limitation. The right panel with global changes shows yield responses to an globally uniform temperature shift; note that these results are not directly comparable to simulations of more realistic climate scenarios with the same global mean change.

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

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

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

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

#### 392 2.4. Emulator evaluation

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

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

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

403 Here  $F_{scn.}$  is the fractional change in a model’s mean emulated or simulated yield from a defined baseline, in some scenario (scn.) in C, T, W, and N space;  $Y_{scn.}$  and  $Y_{baseline}$  are the

absolute emulated or simulated mean yields. The normalized error  $e$  is the difference between the emulated fractional change in yield and that actually simulated, normalized by  $\sigma_{sim.}$ , the standard deviation in simulated fractional yields  $F_{sim, scn.}$  across all models. The emulator is fit across all available simulation outputs, and then the error is calculated across the simulation scenarios provided by all nine models (Figure 8 and Figures ?? and Figures ?? in supplemental documents). Note that the normalized error  $e$  for a model depends not only on the fidelity of its emulator in reproducing a given simulation but on the particular suite of models considered in the intercomparison exercise. The rationale for this choice is to relate the fidelity of the emulation to an estimate of true uncertainty, which we take as the multi-model spread.

### 3. Results

#### 3.1. Simulation results

Crop models in the GGCMI ensemble show a broadly consistent responses to climate and management perturbations in most regions, with a strong negative impact of increased temperature in all but the coldest regions. We illustrate this result for rain-fed maize in Figure 3, which shows yields for the primary Köppen-Geiger climate regions (Rubel & Kottek, 2010). In warming scenarios, models show decreases in maize yield in the temperate, tropical, and arid regions that account for nearly three-quarters of global maize production. These impacts are robust for even moderate climate perturbations. In the temperate zone, even a 1 degree temperature rise with other variables held fixed leads to a median yield reduction that outweighs the variance across models. A 6 degree temperature rise results in median loss of  $\sim 25\%$  of yields with a signal to noise of nearly three. A notable exception is the cold continental region, where models disagree strongly, extending even to the sign of impacts. Model simulations of other crops produce similar responses to warming, with robust yield losses in warmer locations and high inter-model variance in the cold continental regions (Figures ??).

The effects of rainfall changes on maize yields are also as expected and are consistent across models. Increased rainfall mitigates the negative effect of higher temperatures, most strongly in arid regions. Decreased rainfall amplifies yield losses and also increases inter-model variance more strongly, suggesting that models have difficulty representing crop response to water stress. We show only rain-fed maize here; see Figure ?? for the irrigated case. As expected, irrigated crops are more resilient to temperature increases in all regions, especially so where water is limiting.

Mapping the distribution of baseline yields and yield changes shows the geographic dependencies that underlie these results. Figure 4 shows baseline and changes in the T+4 scenario for rain-fed maize, soy, and rice in the multi-model ensemble mean, with locations of model agreement marked. Absolute yield potentials are have strong spatial variation, with much of the Earth’s surface area unsuitable for any given crop. In general, models agree most on yield response in regions where yield

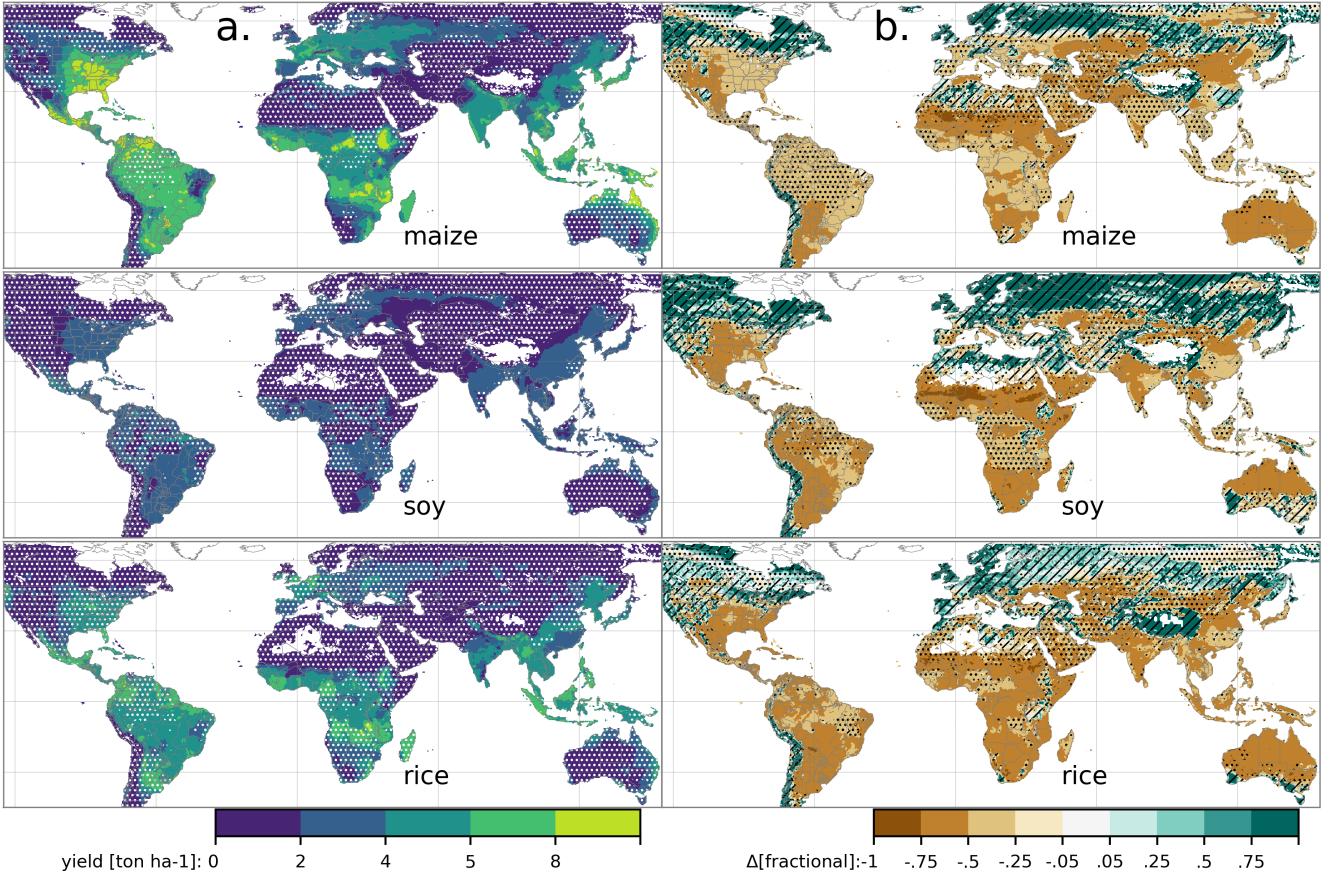


Figure 4: Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for three major crops. Left column (a) shows multi-model mean climatological yields for the baseline scenario for (top-bottom) rain-fed maize, soy, and rice. (For wheat see Figure ?? 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.

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

### 3.2. Simulation model validation results

Figure 7 shows the time series correlation between the simulation model yield and FAO yield data. The results are mixed, with many regions for rice and wheat being difficult to model. No single model is dominant, with each model providing near best-in-class performance in at least one location-crop combination. The presence of no vertical dark green color bars clearly illustrates the power of a multi-model intercomparison project like the one presented here. The ensemble mean yield is calculated across all ‘high’ nitrogen application level model simulations and correlated with the FAO data (not the mean of the correlations). The ensemble mean does not beat the best model in each case, but shows positive correlation in over 75% of the cases presented here.

Soy is qualitatively the easiest crop to represent (except in Argentina), which is likely due to the invariance of the response to nitrogen application (soy fixes atmospheric nitrogen very efficiently). Comparison to the FAO data is therefore easier than the other crops because the nitrogen application levels do not matter. US maize has the best performance across models, with nearly every model representing the historical variability to some extent. Especially good example years for US maize are 1983, 1988, and 2004 (top left panel), where every model gets the direction of the anomaly compared to surrounding years correct. 1983 and 1988 are famously bad years for US maize along with 2012 (not shown). US maize is possibly both the most uniformly industrialized (in terms of management practices) crop and the one with the best data collection in the historical period of all the cases presented here.

FAO data is at least one level of abstraction from ground truth in many cases, especially in developing countries. The failure of models to represent the year-to-year variability in rice in some countries in southeast Asia is likely partly due to model failure and partly due to lack of data. Partitioning of these contributions is impossible at this stage. Additionally, there is less

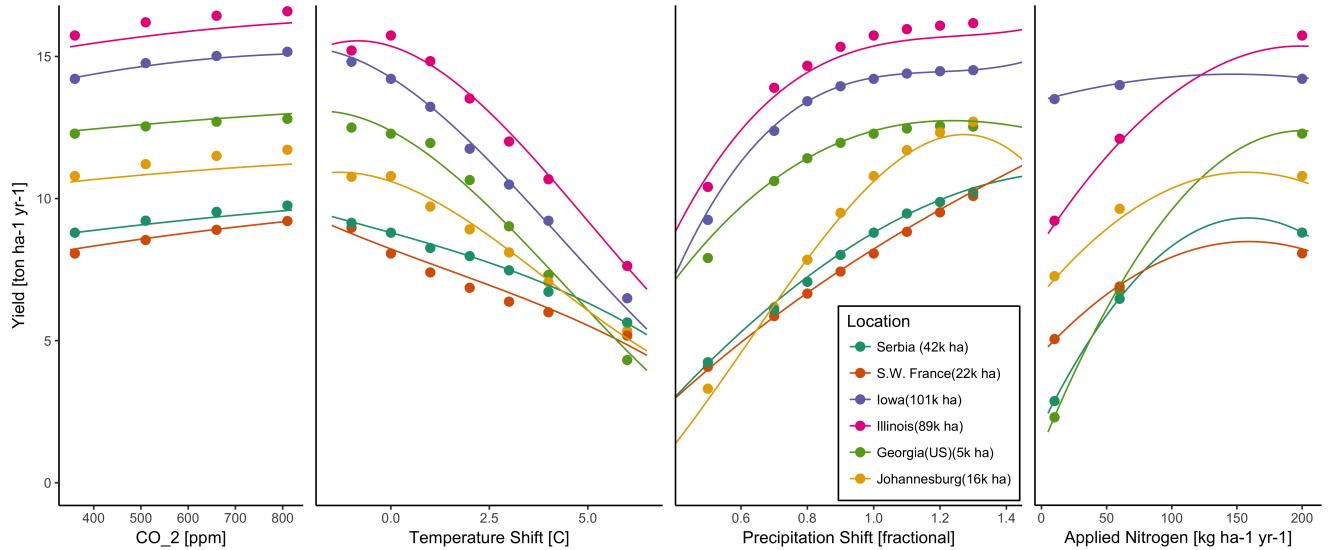


Figure 5: Example illustrating spatial variations in yield response and emulation ability. We show rain-fed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation along a single variable, with others held at baseline values. Dots show climatological mean yield, solid lines a simple polynomial fit across this 1D variation, and dashed lines the results of the full emulator of Equation 1. The climatological response is sufficiently smooth to be represented by a simple polynomial. Because precipitation changes are imposed as multiplicative factors rather than additive offsets, locations with higher baseline precipitation have a larger absolute spread in inputs sampled than do drier locations (not visible here).

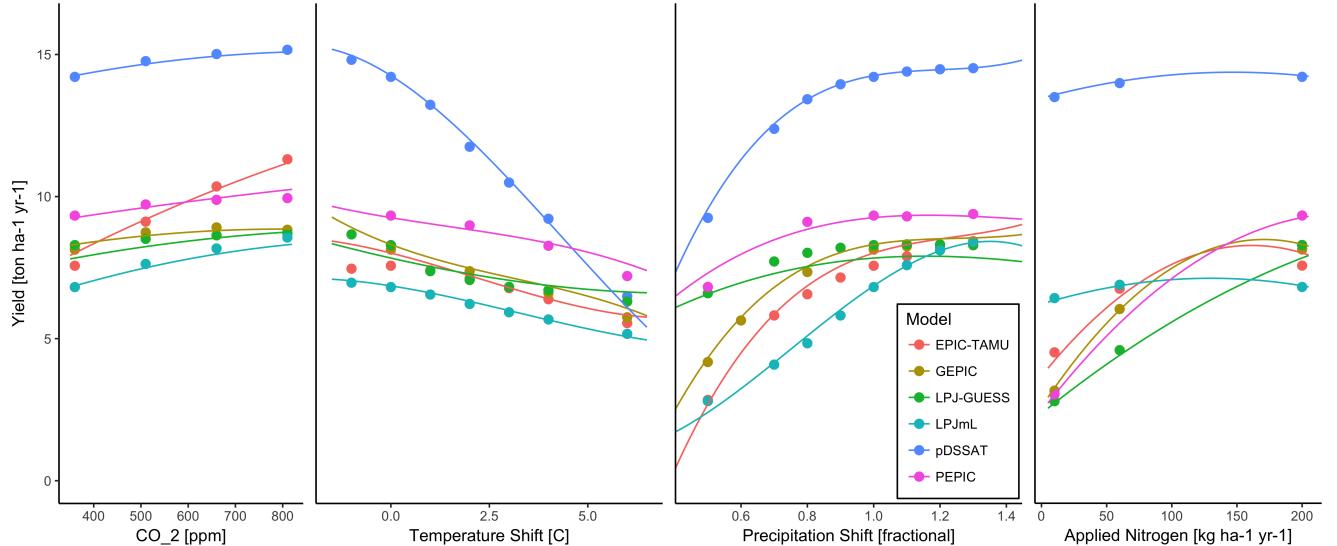


Figure 6: Example illustrating across-model variations in yield response. We show simulations and emulations from six models for rain-fed maize in the same Iowa location shown in Figure 5, with the same plot conventions. Models that did not simulate the nitrogen dimension are omitted for clarity. The inter-model standard deviation is larger for the perturbations in CO<sub>2</sub> and nitrogen than for those in temperature or precipitation illustrating that the sensitivity to weather is better constrained than to management at elevated CO<sub>2</sub>. While most model responses can readily be emulated with a simple polynomial, some response surfaces are more complicated and lead to emulation error.

501 year-to-year variability in rice yields (partially due to the fraction of irrigated cultivation). Since the Pearson r metric is scale invariant, it will tend to score the rice models more poorly than maize and soy. The pDSSAT model shows very poor performance for rice in India (top right panel).

506 One may speculate that the difference in performance between Pakistan (no successful models) and India (many successful models) for rice may lie in the FAO data and not the models themselves. What would be so different about rice pro-

501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
7510  
7511  
7512  
7513  
7514  
7515  
7516  
7517  
7518  
7519  
7520  
7521  
7522  
7523  
7524  
7525  
7526  
7527  
7528  
7529  
7530  
7531  
7532  
7533  
7534  
7535  
7536  
7537  
7538  
7539  
7540  
7541  
7542  
7543  
7544  
7545  
7546  
7547  
7548  
7549  
7550  
7551  
7552  
7553  
7554  
7555  
7556  
7557  
7558  
7559  
7560  
7561  
7562  
7563  
7564  
7565  
7566  
7567  
7568  
7569  
7570  
7571  
7572  
7573  
7574  
7575  
7576  
7577  
7578  
7579  
7580  
7581  
7582  
7583  
7584  
7585  
7586  
7587  
7588  
7589  
7590  
7591  
7592  
7593  
7594  
7595  
7596  
7597  
7598  
7599  
75100  
75101  
75102  
75103  
75104  
75105  
75106  
75107  
75108  
75109  
75110  
75111  
75112  
75113  
75114  
75115  
75116  
75117  
75118  
75119  
75120  
75121  
75122  
75123  
75124  
75125  
75126  
75127  
75128  
75129  
75130  
75131  
75132  
75133  
75134  
75135  
75136  
75137  
75138  
75139  
75140  
75141  
75142  
75143  
75144  
75145  
75146  
75147  
75148  
75149  
75150  
75151  
75152  
75153  
75154  
75155  
75156  
75157  
75158  
75159  
75160  
75161  
75162  
75163  
75164  
75165  
75166  
75167  
75168  
75169  
75170  
75171  
75172  
75173  
75174  
75175  
75176  
75177  
75178  
75179  
75180  
75181  
75182  
75183  
75184  
75185  
75186  
75187  
75188  
75189  
75190  
75191  
75192  
75193  
75194  
75195  
75196  
75197  
75198  
75199  
75200  
75201  
75202  
75203  
75204  
75205  
75206  
75207  
75208  
75209  
75210  
75211  
75212  
75213  
75214  
75215  
75216  
75217  
75218  
75219  
75220  
75221  
75222  
75223  
75224  
75225  
75226  
75227  
75228  
75229  
75230  
75231  
75232  
75233  
75234  
75235  
75236  
75237  
75238  
75239  
75240  
75241  
75242  
75243  
75244  
75245  
75246  
75247  
75248  
75249  
75250  
75251  
75252  
75253  
75254  
75255  
75256  
75257  
75258  
75259  
75260  
75261  
75262  
75263  
75264  
75265  
75266  
75267  
75268  
75269  
75270  
75271  
75272  
75273  
75274  
75275  
75276  
75277  
75278  
75279  
75280  
75281  
75282  
75283  
75284  
75285  
75286  
75287  
75288  
75289  
75290  
75291  
75292  
75293  
75294  
75295  
75296  
75297  
75298  
75299  
75300  
75301  
75302  
75303  
75304  
75305  
75306  
75307  
75308  
75309  
75310  
75311  
75312  
75313  
75314  
75315  
75316  
75317  
75318  
75319  
75320  
75321  
75322  
75323  
75324  
75325  
75326  
75327  
75328  
75329  
75330  
75331  
75332  
75333  
75334  
75335  
75336  
75337  
75338  
75339  
75340  
75341  
75342  
75343  
75344  
75345  
75346  
75347  
75348  
75349  
75350  
75351  
75352  
75353  
75354  
75355  
75356  
75357  
75358  
75359  
75360  
75361  
75362  
75363  
75364  
75365  
75366  
75367  
75368  
75369  
75370  
75371  
75372  
75373  
75374  
75375  
75376  
75377  
75378  
75379  
75380  
75381  
75382  
75383  
75384  
75385  
75386  
75387  
75388  
75389  
75390  
75391  
75392  
75393  
75394  
75395  
75396  
75397  
75398  
75399  
75400  
75401  
75402  
75403  
75404  
75405  
75406  
75407  
75408  
75409  
75410  
75411  
75412  
75413  
75414  
75415  
75416  
75417  
75418  
75419  
75420  
75421  
75422  
75423  
75424  
75425  
75426  
75427  
75428  
75429  
75430  
75431  
75432  
75433  
75434  
75435  
75436  
75437  
75438  
75439  
75440  
75441  
75442  
75443  
75444  
75445  
75446  
75447  
75448  
75449  
75450  
75451  
75452  
75453  
75454  
75455  
75456  
75457  
75458  
75459  
75460  
75461  
75462  
75463  
75464  
75465  
75466  
75467  
75468  
75469  
75470  
75471  
75472  
75473  
75474  
75475  
75476  
75477  
75478  
75479  
75480  
75481  
75482  
75483  
75484  
75485  
75486  
75487  
75488  
75489  
75490  
75491  
75492  
75493  
75494  
75495  
75496  
75497  
75498  
75499  
75500  
75501  
75502  
75503  
75504  
75505  
75506  
75507  
75508  
75509  
75510  
75511  
75512  
75513  
75514  
75515  
75516  
75517  
75518  
75519  
75520  
75521  
75522  
75523  
75524  
75525  
75526  
75527  
75528  
75529  
75530  
75531  
75532  
75533  
75534  
75535  
75536  
75537  
75538  
75539  
75540  
75541  
75542  
75543  
75544  
75545  
75546  
75547  
75548  
75549  
75550  
75551  
75552  
75553  
75554  
75555  
75556  
75557  
75558  
75559  
75560  
75561  
75562  
75563  
75564  
75565  
75566  
75567  
75568  
75569  
75570  
75571  
75572  
75573  
75574  
75575  
75576  
75577  
75578  
75579  
75580  
75581  
75582  
75583  
75584  
75585  
75586  
75587  
75588  
75589  
75590  
75591  
75592  
75593  
75594  
75595  
75596  
75597  
75598  
75599  
75600  
75601  
75602  
75603  
75604  
75605  
75606  
75607  
75608  
75609  
75610  
75611  
75612  
75613  
75614  
75615  
75616  
75617  
75618  
75619  
75620  
75621  
75622  
75623  
75624  
75625  
75626  
75627  
75628  
75629  
75630  
75631  
75632  
75633  
75634  
75635  
75636  
75637  
75638  
75639  
75640  
75641  
75642  
75643  
75644  
75645  
75646  
75647  
75648  
75649  
75650  
75651  
75652  
75653  
75654  
75655  
75656  
75657  
75658  
75659  
75660  
75661  
75662  
75663  
75664  
75665  
75666  
75667  
75668  
75669  
75670  
75671  
75672  
75673  
75674  
75675  
75676  
75677  
75678  
75679  
75680  
75681  
75682  
75683  
75684  
75685  
75686  
75687  
75688  
75689  
75690  
75691  
75692  
75693  
75694  
75695  
75696  
75697  
75698  
75699  
75700  
75701  
75702  
75703  
75704  
75705  
75706  
75707  
75708  
75709  
75710  
75711  
75712  
75713  
75714  
75715  
75716  
75717  
75718  
75719  
75720  
75721  
75722  
75723  
75724  
75725  
75726  
75727  
75728  
75729  
75730  
75731  
75732  
75733  
75734  
75735  
75736  
75737  
75738  
75739  
75740  
75741  
75742  
75743  
75744  
75745  
75746  
75747  
75748  
75749  
75750  
75751  
75752  
75753  
75754  
75755  
75756  
75757  
75758  
75759  
75760  
75761  
75762  
75763  
75764  
75765  
75766  
75767  
75768  
75769  
75770  
75771  
75772  
75773  
75774  
75775  
75776  
75777  
75778  
75779  
75780  
75781  
75782  
75783  
75784  
75785  
75786  
75787  
75788  
75789  
75790  
75791  
75792  
75793  
75794  
75795  
75796  
75797  
75798  
75799  
75800  
75801  
75802  
75803  
75804  
75805  
75806  
75807  
75808  
75809  
75810  
75811  
75812  
75813  
75814  
75815  
75816  
75817  
75818  
75819  
75820  
75821  
75822  
75823  
75824  
75825  
75826  
75827  
75828  
75829  
75830  
75831  
75832  
75833  
75834  
75835  
75836  
75837  
75838  
75839  
75840  
75841  
75842  
75843  
75844  
75845  
75846  
75847  
75848  
75849  
75850  
75851  
75852  
75853  
75854  
75855  
75856  
75857  
75858  
75859  
75860  
75861  
75862  
75863  
75864  
75865  
75866  
75867  
75868  
75869  
75870  
75871  
75872  
75873  
75874  
75875  
75876  
75877  
75878  
75879  
75880  
75881  
75882  
75883  
75884  
75885  
75886  
75887  
75888  
75889  
75890  
75891  
75892  
75893  
75894  
75895  
75896  
75897  
75898  
75899  
75900  
75901  
75902  
75903  
75904  
75905  
75906  
75907  
75908  
75909  
75910  
75911  
75912  
75913  
75914  
75915  
75916  
75917  
75918  
75919  
75920  
75921  
75922  
75923  
75924  
75925  
75926  
75927  
75928  
75929  
75930  
75931  
75932  
75933  
75934  
75935  
75936  
75937  
75938  
75939  
75940  
75941  
75942  
75943  
75944  
75945  
75946  
75947  
75948  
75949  
75950  
75951  
75952  
75953  
75954  
75955  
75956  
75957  
75958  
75959  
75960  
75961  
75962  
75963  
75964  
75965  
75966  
75967  
75968  
75969  
75970  
75971  
75972  
75973  
75974  
75975  
75976  
75977  
75978  
75979  
75980  
75981  
75982  
75983  
75984  
75985  
75986  
75987  
75988  
75989  
75990  
75991  
75992  
75993  
75994  
75995  
75996  
75997  
75998  
75999  
75100  
75101  
75102  
75103  
75104  
75105  
75106  
75107  
75108  
75109  
75110  
75111  
75112  
75113  
75114  
75115  
75116  
75117  
75118  
75119  
75120  
75121  
75122  
75123  
75124  
75125  
75126  
75127  
75128  
75129  
75130  
75131  
75132  
75133  
75134  
75135  
75136  
75137  
75138  
75139  
75140  
75141  
75142  
75143  
75144  
75145  
75146  
75147  
75148  
75149  
75150  
75151  
75152  
75153  
75154  
75155  
75156  
75157  
75158  
75159  
75160  
75161  
75162  
75163  
75164  
75165  
75166  
75167  
75168  
75169  
75170  
75171  
75172  
75173  
75174  
75175  
75176  
75177  
75178  
75179  
75180  
75181  
75182  
75183  
75184  
75185  
75186  
75187  
75188  
75189  
75190  
75191  
75192  
75193  
75194  
75195  
75196  
75197  
75198  
75199  
75200  
75201  
75202  
75203  
75204  
75205  
75206  
75207  
75208  
75209  
75210  
75211  
75212  
75213  
75214  
75215  
75216  
75217  
75218  
75219  
75220  
75221  
75222  
75223  
75224  
75225  
75226  
75227  
75228  
75229  
75230  
75231  
75232  
75233  
75234  
75235  
75236  
75237  
75238  
75239  
75240  
75241  
75242  
75243  
75244  
75245  
75246  
75247  
75248  
75249  
75250  
75251  
75252  
75253  
75254  
75255  
75256  
75257  
75258  
75259  
75260  
75261  
75262  
75263  
75264  
75265  
75266  
75267  
75268  
75269  
75270  
75271  
75272  
75273  
75274  
75275  
75276  
75277  
75278  
75279  
75280  
75281  
75282  
75283  
75284  
75285  
75286  
75287  
75288  
75289  
75290  
75291  
75292  
75293  
75294  
75295  
75296  
75297  
75298  
75299  
75300  
75301  
75302  
75303  
75304  
75305  
75306  
75307  
75308  
75309  
75310  
75311  
75312  
75313  
75314  
75315  
7531

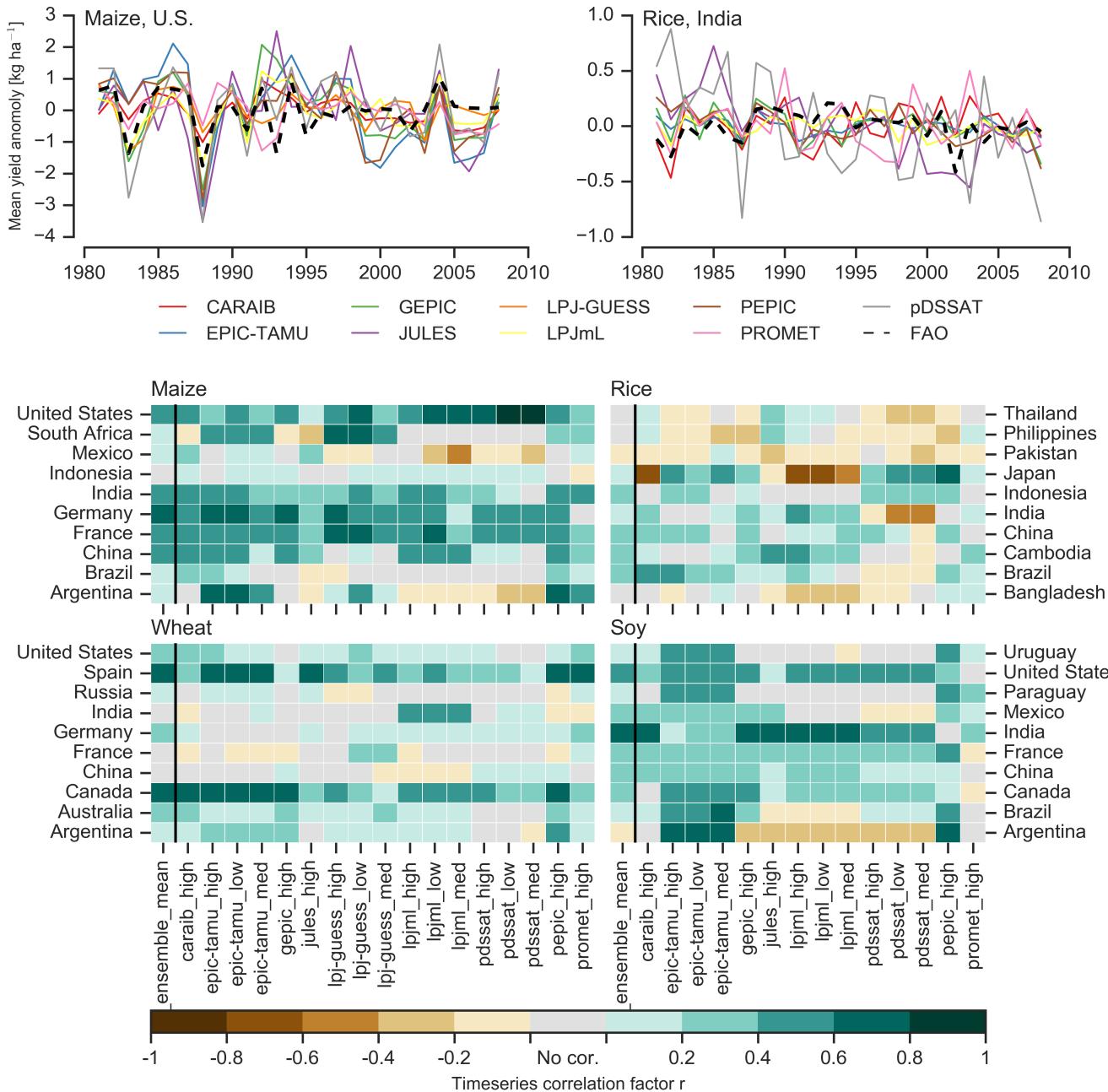


Figure 7: Time series correlation coefficients between simulated crop yield and FAO data at the country level. The top panels indicate two example cases: US maize (a good case), and rice in India (mixed case), both for the high nitrogen application case. The heatmaps illustrate the Pearson r correlation 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.

518 form. In the GGCMI simulations, this condition largely but<sub>528</sub>  
 519 not always holds. Responses are quite diverse across locations,<sub>529</sub>  
 520 crops, and models, but in most cases local responses are reg-<sub>531</sub>  
 521 ular enough to permit emulation. Figure 5 illustrates the ge-<sub>531</sub>  
 522 ographic diversity of responses even in high-yield areas for a<sub>532</sub>  
 523 single crop and model (rain-fed maize in pDSSAT for various<sub>533</sub>  
 524 high-cultivation areas). This heterogeneity validates the choice<sub>534</sub>  
 525 of emulating at the grid cell level.<sub>535</sub>

526 Each panel in Figure 5 shows model yield output from sce-<sub>536</sub>  
 527 narios varying only along a single dimension ( $\text{CO}_2$ , tempera-<sub>537</sub>

ture, precipitation, or nitrogen addition), with other inputs held fixed at baseline levels; in all cases yields evolve smoothly across the space sampled. For reference we show the results of the full emulation fitted across the parameter space. The polynomial fit readily captures the climatological response to perturbations.

Crop yield responses generally follow similar functional forms across models, though with a spread in magnitude. Figure 6 illustrates the inter-model diversity of yield responses to the same perturbations, even for a single crop and location

(rain-fed maize in northern Iowa, the same location shown in the Figure 5). The differences make it important to construct emulators separately for each individual model, and the fidelity of emulation can also differ across models. This figure illustrates a common phenomenon, that models differ more in response to perturbations in CO<sub>2</sub> and nitrogen perturbations than to those in temperature or precipitation. (Compare also Figures 3 and ??.) For this location and crop, CO<sub>2</sub> 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 and uncertain, it is also the most problematic to emulate in this work because of its limited sampling. The GGCMI protocol specified only three nitrogen levels (10, 60 and 200 kg N y<sup>-1</sup> ha<sup>-1</sup>), 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 means that some regressions imply a peak in yield between the 100 and 200 kg N y<sup>-1</sup> ha<sup>-1</sup> levels. While there may be some reason to believe over-application of nitrogen at the wrong time in the growing season could lead to reduced yields, these features are almost certainly an artifact of under sampling. In addition, the polynomial fit cannot capture the well-documented saturation effect of nitrogen application (e.g. Ingestad, 1977) as accurately as would be possible with a non-parametric model.

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 and its emulation, normalized by the inter-model standard deviation in yield projections. This metric implies that emulation is generally satisfactory, with several distinct exceptions. Almost all model-crop combination emulators have normalized errors less than one over nearly all currently cultivated hectares (Figure 8), but some individual model-crop combinations are problematic (e.g. PROMET for rice and soy, JULES for soy and winter wheat, Figures ??–??). Normalized errors for soy are somewhat higher across all models not because emulator fidelity is worse but because models agree more closely on yield changes for soy than for other crops (see Figure ??, lowering the denominator). Emulator performance often degrades in geographic locations where crops are not currently cultivated. Figure 9 shows a CARAIB case as an example, where emulator performance is satisfactory over cultivated areas for all crops other than soy, but uncultivated regions show some problematic areas.

It should be noted that this assessment metric is relatively forgiving. First, each emulation is evaluated against the simulation actually used to train the emulator. Had we used a spline interpolation the error would necessarily be zero. Second, the performance metric scales emulator fidelity not by the magnitude of yield changes but by the inter-model spread in those changes. Where models differ more widely, the standard for emulators becomes less stringent. Because models disagree on the magnitude of CO<sub>2</sub> fertilization, this effect is readily seen when comparing assessments of emulator performance in simulations at baseline CO<sub>2</sub> (Figure 8) with those at higher CO<sub>2</sub>

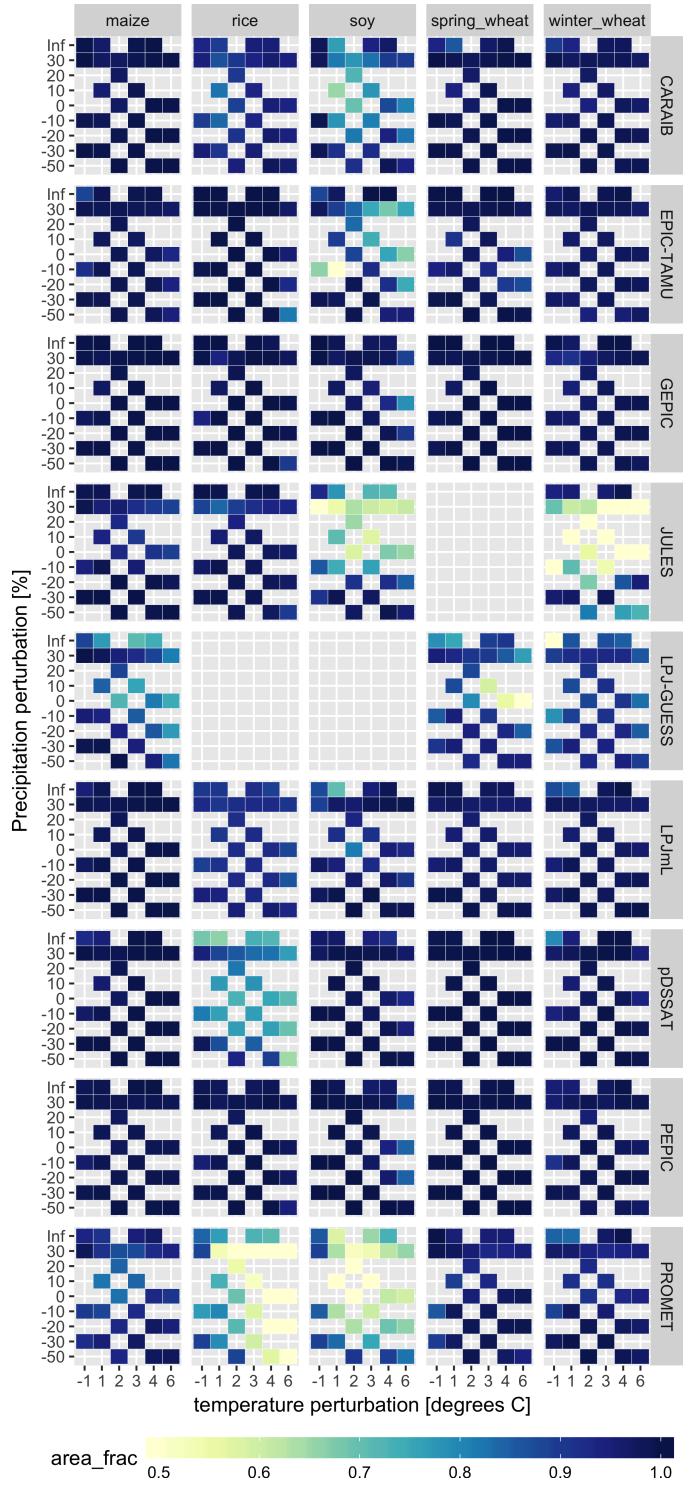


Figure 8: Assessment of emulator performance over currently cultivated areas based on normalized error (Equations 3, 2). We show performance of all 9 models emulated, over all crops and all sampled T and P inputs, but with CO<sub>2</sub> and nitrogen held fixed at baseline values. Large columns are crops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

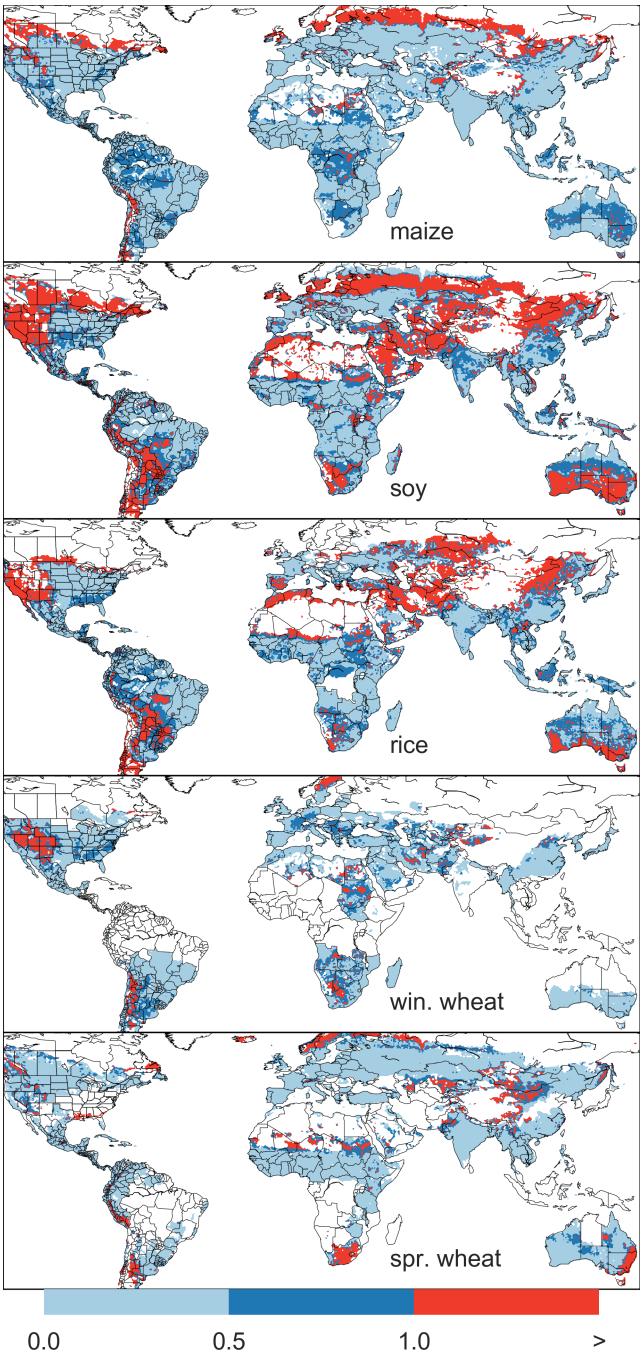


Figure 9: Illustration of our test of emulator performance, applied to the CARAIB model for the T+4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error  $e$ , where  $e > 1$  means that emulator error exceeds the multi-model standard deviation. White areas are those where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

levels (Figure ??). Widening the inter-model spread leads to an apparent increase in emulator skill.

### 3.4. Emulator applications

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

## 4. Conclusions and discussion

The GGCMI Phase II experiment assess sensitivities of process-based crop yield models to changing climate and management inputs, and was designed to allow not only comparison across models but evaluation of complex interactions between driving factors ( $\text{CO}_2$ , temperature, precipitation, and applied nitrogen) and identification of geographic shifts in high yield potential locations. While the richness of the dataset invites further analysis, we show only a selection of insights derived from the simulations. Across the major crops, inter-model uncertainty is greatest for wheat and least for soy. Across factors impacting yields, inter-model-uncertainty is largest for  $\text{CO}_2$  fertilization and nitrogen response effects. Across geographic regions, inter-model uncertainty is largest in the high latitudes where yields may increase, and model projections are most robust in low latitudes where yield impacts are largest.

Model performance when compared to historical data is mixed, with models performing better for maize and soy than for rice and wheat. The value of utilizing multiple models is illustrated by the distribution in performance skill across different countries and crops. An end-user of the simulation outputs or emulator tool may pick and choose models based on historical skill to provide the most faithful temperature and precipitation response depending on their application. The nitrogen and  $\text{CO}_2$  responses were not validated in this work.

One counterintuitive result is that irrigated maize shows steeper yield reductions under warming than does rain-fed maize when considered only over currently cultivated land. The effect is the result of geographic differences in cultivated area.

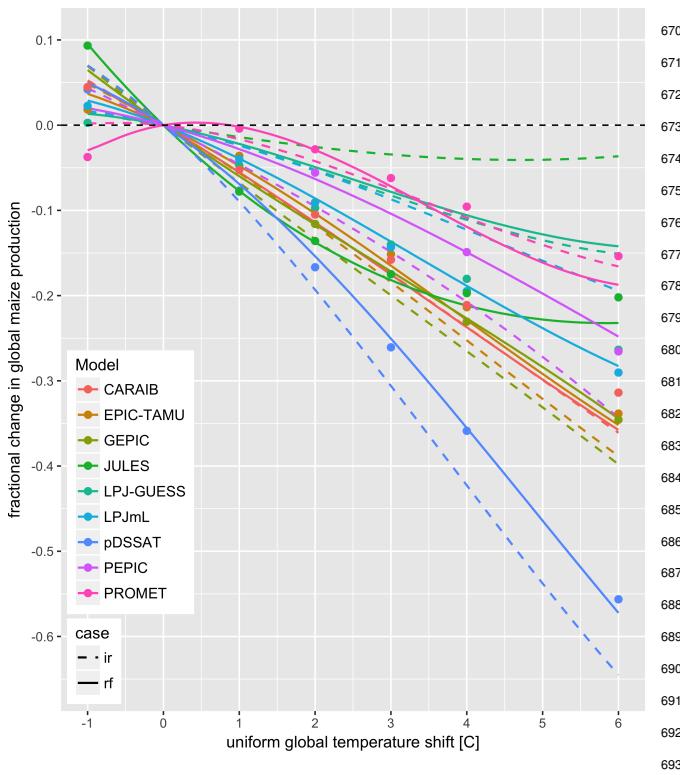


Figure 10: Global emulated damages for maize on currently cultivated lands<sup>694</sup> for the GGCMI models emulated, for uniform temperature shifts with other<sup>695</sup> inputs held at baseline. (The damage function is created from aggregating up<sup>696</sup> emulated values at the grid cell level, not from a regression of global mean<sup>697</sup> yields.) Lines are emulations for rain-fed (solid) and irrigated (dashed) crops;<sup>698</sup> for comparison, dots are the simulated values for the rain-fed case. For most<sup>699</sup> models, irrigated crops show a sharper reduction than do rain-fed because of the<sup>700</sup> locations of cultivated areas: irrigated crops tend to be grown in warmer areas<sup>701</sup> where impacts are more severe for a given temperature shift. (The exceptions<sup>702</sup> are PROMET, JULES, and LPJmL.) For other crops and scenarios see Figures<sup>703</sup> ??-?? in the supplemental material.

In any given location, irrigation increases crop resiliency to<sup>705</sup> temperature increase, but irrigated maize is grown in warmer lo-<sup>706</sup>cations where the impacts of warming are more severe (Figures<sup>707</sup> ??-??). The same behavior holds for rice and winter wheat,<sup>708</sup> but not for soy or spring wheat (Figures ??-??). Irrigated wheat<sup>709</sup> and maize are also more sensitive to nitrogen fertilization lev-<sup>710</sup>els, presumably because growth in rain-fed crops is also water-<sup>711</sup>limited (Figure ??). (Soy as a nitrogen-fixer is relatively in-<sup>712</sup>sensitive to nitrogen, and rice is not generally grown in water-<sup>713</sup>limited conditions.)<sup>714</sup>

We show that emulation of the output of these complex re-<sup>715</sup>sponses is possible even with a relatively simple reduced-form<sup>716</sup> statistical model and a limited library of simulations. Emula-<sup>717</sup>tion therefore offers the opportunity of producing rapid assess-<sup>718</sup>ments of agricultural impacts for arbitrary climate scenarios in<sup>719</sup>a computationally non-intensive way. The resulting tool should<sup>720</sup>aid in impacts assessment, economic studies, and uncertainty<sup>721</sup>analyses. Emulator parameter values also provide a useful way<sup>722</sup>to compare sensitivities across models to different climate and<sup>723</sup>management inputs, and the terms in the polynomial fits offer<sup>724</sup>the possibility of physical interpretation of these dependencies<sup>725</sup>to some degree.

We provide this simulation output dataset for further analysis by the community as we have only scratched the surface with this work. Each simulation run includes year to year variability in yields under different climate and management regimes. Some of the precipitation and temperature space has been lost due to the aggregation in the time dimension for the emulator presented here (i.e. the + 6 C simulation in the hottest year of the historical period compared to the coldest historical year, or precipitation perturbations in the driest historical year etc). Development of a year-to-year emulator or an emulator at different spatial scales may provide useful for some IAM applications. More exhaustive analysis of different statistical model specification for emulation will likely provide additional predictive skill over the specification provided here. The potentially richest area for further analysis is the interactions between input variable especially the Nitrogen and CO<sub>2</sub> interactions with weather and with each other. More robust quantification of the sensitivity to the input drivers (and there differences across models), as well as quantification in differences in uncertainty across input drivers. Adaptation via growing season changes were also simulated and are available in the database, though this dimension was not presented or analyzed here.

The emulation approach presented here has some limitations. Because the GGCMI simulations apply uniform perturbations to historical climate inputs, they do not sample changes in higher order moments. The emulation therefore does not address the crop yield impacts of potential changes in climate variability. While some information could be extracted from consideration of year-over-year variability, more detailed simulations and analysis are likely necessary to diagnose the impact of changes in variance and sub-growing-season temporal effects. Additionally, the emulator is intended to provide the change in yield from a historical mean baseline value and should be used in conjunction with historical data (or data products) or a historical mean emulator (not presented here).

The future of food security is one of the larger challenges facing humanity at present. The development (and emulation) of multi-model ensembles such as GGCMI Phase II provides a way to begin to quantify uncertainties in crop responses to a range of potential climate inputs and explore the potential benefits of adaptive responses. Emulation also allow making state-of-the-art simulation results available to a wide research community as simple, computationally tractable tools that can be used by downstream modelers to understand the socioeconomic impacts of crop response to climate change.

## 5. Acknowledgments

We thank Michael Stein and Kevin Schwarzwald, who provided helpful suggestions that contributed to this work. This research was performed as part of the Center for Robust Decision-Making on Climate and Energy Policy (RDCEP) at the University of Chicago, and was supported through a variety of sources. RDCEP is funded by NSF grant #SES-1463644 through the Decision Making Under Uncertainty program. J.F. was supported by the NSF NRT program, grant #DGE-1735359. C.M. was supported by the MACMIT project (01LN1317A) funded

through the German Federal Ministry of Education and Research (BMBF). C.F. was supported by the European Research Council Synergy grant #ERC-2013-SynG-610028 Imbalance-P. P.F. and K.W. were supported by the Newton Fund through the Met Office Climate Science for Service Partnership Brazil (CSSP Brazil). A.S. was supported by the Office of Science of the U.S. Department of Energy as part of the Multi-sector Dynamics Research Program Area. Computing resources were provided by the University of Chicago Research Computing Center (RCC). S.O. acknowledges support from the Swedish strong research areas BECC and MERGE together with support from LUCCI (Lund University Centre for studies of Carbon Cycle and Climate Interactions).

## 6. References

- Angulo, C., Ritter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013). Implication of crop model calibration strategies for assessing regional impacts of climate change in europe. *Agric. For. Meteorol.*, 170, 32 – 46.
- Asseng, S., Ewert, F., Martre, P., Ritter, R. P., B. Lobell, D., Cammarano, D., A. Kimball, B., Ottman, M., W. Wall, G., White, J., Reynolds, M., D. Alderman, P., Prasad, P. V. V., Aggarwal, P., Anothai, J., Basso, B., 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.
- Asseng, S., Ewert, F., Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., 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.
- 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.
- 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.
- Blanc, E. (2017). Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models. *Agricultural and Forest Meteorology*, 236, 145 – 161.
- Blanc, E., & Sultan, B. (2015). Emulating maize yields from global gridded crop models using statistical estimates. *Agricultural and Forest Meteorology*, 214-215, 134 – 147.
- 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.
- 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.
- 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.
- Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process emulation of dynamic computer codes. *Biometrika*, 96, 663–676.
- Duncan, W. (1972). SIMCOT: a simulation of cotton growth and yield. In C. Murphy (Ed.), *Proceedings of a Workshop for Modeling Tree Growth*, Duke University, Durham, North Carolina (pp. 115–118). Durham, North Carolina.
- Duncan, W., Loomis, R., Williams, W., & Hanau, R. (1967). A model for simulating photosynthesis in plant communities. *Hilgardia*, (pp. 181–205).
- Dury, M., Hambuckers, A., Warnant, P., Henrot, A., Favre, E., Ouberdoorn, M., & François, L. (2011). Responses of European forest ecosystems to 21st century climate: assessing changes in interannual variability and fire intensity. *iForest - Biogeosciences and Forestry*, (pp. 82–99).
- Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K., Best, N., Wilde, M., & Foster, I. (2014). The parallel system for integrating impact models and sectors (pSIMS). *Environmental Modelling and Software*, 62, 509–516.
- Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Izumi, T., Izaurralde, R. C., Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., & Sheffield, J. (2015). The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8, 261–277.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958.
- Ferrise, R., Moriondo, M., & Bindi, M. (2011). Probabilistic assessments of climate change impacts on durum wheat in the mediterranean region. *Natural Hazards and Earth System Sciences*, 11, 1293–1302.
- Folberth, C., Gaiser, T., Abbaspour, K. C., Schulin, R., & Yang, H. (2012). Regionalization of a large-scale crop growth model for sub-Saharan Africa: Model setup, evaluation, and estimation of maize yields. *Agriculture, Ecosystems & Environment*, 151, 21 – 33.
- Food and Agriculture Organization of the United Nations (2018). FAOSTAT database. URL: <http://www.fao.org/faostat/en/home>.
- Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.-F., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kerssebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J., Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna, M., Semenov, M. A., Slawinski, C., Strattonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., & Rötter, R. P. (2018). Classifying multi-model wheat yield impact response surfaces showing sensitivity to temperature and precipitation change. *Agricultural Systems*, 159, 209–224.
- Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., & Moyer, E. J. (2014). Evaluating the utility of dynamical downscaling in agricultural impacts projections. *Proceedings of the National Academy of Sciences*, 111, 8776–8781.
- Glotter, M., Moyer, E., Ruane, A., & Elliott, J. (2015). Evaluating the Sensitivity of Agricultural Model Performance to Different Climate Inputs. *Journal of Applied Meteorology and Climatology*, 55, 151113145618001.
- Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous Crop Growth and Yield: Application for Wheat in Central Europe. *Remote Sensing*, 7, 3934–3965.
- 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.
- 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., Dalglish, N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson, M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T., Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., & Keating, B. A. (2014). APSIM Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling and Software*, 62, 327 – 350.
- 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).
- Izumi, T., Nishimori, M., & Yokozawa, M. (2010). Diagnostics of climate

- model biases in summer temperature and warm-season insolation for the simulation of regional paddy rice yield in japan. *Journal of Applied Meteorology and Climatology*, 49, 574–591.
- Ingestad, T. (1977). Nitrogen and Plant Growth; Maximum Efficiency of Nitrogen Fertilizers. *Ambio*, 6, 146–151.
- Izaurralde, R., Williams, J., McGill, W., Rosenberg, N., & Quiroga Jakas, M. (2006). Simulating soil C dynamics with EPIC: Model description and testing against long-term data. *Ecological Modelling*, 192, 362–384.
- Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. *Agriculture, Ecosystems & Environment*, 93, 73 – 85.
- Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijsman, A., & Ritchie, J. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18, 235 – 265.
- Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural Systems*, 155, 269 – 288.
- Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J., Silburn, M., Wang, E., Brown, S., Bristow, K., Asseng, S., Chapman, S., McCown, R., Freebairn, D., & Smith, C. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, 18, 267 – 288.
- Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., & Smith, B. (2013). Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa. *Earth System Dynamics*, 4, 385–407.
- Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIG - modelling wheat yield and crop water productivity with high resolution on a global scale. *Agricultural Systems*, 94, 478 – 493.
- Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., & Schulin, R. (2016a). Global investigation of impacts of PET methods on simulating crop-water relations for maize. *Agricultural and Forest Meteorology*, 221, 164 – 175.
- Liu, W., Yang, H., Liu, J., Azevedo, L. B., Wang, X., Xu, Z., Abbaspour, K. C., & Schulin, R. (2016b). Global assessment of nitrogen losses and trade-offs with yields from major crop cultivations. *Science of The Total Environment*, 572, 526 – 537.
- Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, 150, 1443 – 1452.
- Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2, 014002.
- MacKay, D. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447.
- Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J., Martre, P., Adam, M., Aggarwal, P., Angulo, C., Baron, C., Basso, B., Bertuzzi, P., Biernath, C., Boogaard, H., Boote, K., Brisson, N., Cammarano, D., Challinor, A., Conijn, J., & Wolf, J. (2015). Statistical analysis of large simulated yield datasets for studying climate effects. (p. 1100). doi:10.13140/RG.2.1.5173.8328.
- Mauser, W., & Bach, H. (2015). PROMET - Large scale distributed hydrological modelling to study the impact of climate change on the water flows of mountain watersheds. *Journal of Hydrology*, 376, 362 – 377.
- Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., & Calzadilla, A. (2009). Global biomass production potentials exceed expected future demand without the need for cropland expansion. *Nature Communications*, 6.
- McDermid, S., Dileepkumar, G., Murthy, K., Nedumaran, S., Singh, P., Srivastava, C., Gangwar, B., Subash, N., Ahmad, A., Zubair, L., & Nissanka, S. (2015). Integrated assessments of the impacts of climate change on agriculture: An overview of AgMIP regional research in South Asia. *Chapter in: Handbook of Climate Change and Agroecosystems*, (pp. 201–218).
- Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical weather responsiveness of crop yields: US evidence and implications for the agricultural impacts of climate change. *Environmental Research Letters*, 12.
- Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, 8.
- Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., & Yang, H. (2017). Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, 10, 1403–1422.
- Nakamura, T., Osaki, M., Koike, T., Hanba, Y. T., Wada, E., & Tadano, T. (1997). Effect of CO<sub>2</sub> enrichment on carbon and nitrogen interaction in wheat and soybean. *Soil Science and Plant Nutrition*, 43, 789–798.
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91, 1290 – 1300.
- Olin, S., Schurgers, G., Lindeskog, M., Wårild, D., Smith, B., Bodin, P., Holmér, J., & Arneth, A. (2015). Modelling the response of yields and tissue C:N to changes in atmospheric CO<sub>2</sub> and N management in the main wheat regions of western europe. *Biogeosciences*, 12, 2489–2515. doi:10.5194/bg-12-2489-2015.
- Osaki, M., Shinano, T., & Tadano, T. (1992). Carbon-nitrogen interaction in field crop production. *Soil Science and Plant Nutrition*, 38, 553–564.
- Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., & Wheeler, T. (2015). JULES-crop: a parametrisation of crops in the Joint UK Land Environment Simulator. *Geoscientific Model Development*, 8, 1139–1155.
- Ostberg, S., Schewe, J., Childers, K., & Frieler, K. (2018). Changes in crop yields and their variability at different levels of global warming. *Earth System Dynamics*, 9, 479–496.
- Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff, S., & Gerten, D. (2015). Emulating global climate change impacts on crop yields. *Statistical Modelling*, 15, 499–525.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pirttioja, N., Carter, T., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K., Kollas, C., Krzyszczak, J., Lorite, I., Minet, J., Minguez, M., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodriguez, A., Ruane, A., Ruget, F., Sanna, M., Semenov, M., Slawinski, C., Strattonovich, P., Sutip, I., Waha, K., Wang, E., Wu, L., Zhao, Z., & Rötter, R. (2015). Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. *Climate Research*, 65, 87–105.
- Porter et al. (IPCC) (2014). Food security and food production systems. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In C. F. et al. (Ed.), *IPCC Fifth Assessment Report* (pp. 485–533). Cambridge, UK: Cambridge University Press.
- Pottmann, F., Siebert, S., & Doell, P. (2010). MIRCA2000 - Global Monthly Irrigated and Rainfed crop Areas around the Year 2000: A New High-Resolution Data Set for Agricultural and Hydrological Modeling. *Global Biogeochemical Cycles*, 24, GB1011.
- Pugh, T., Müller, C., Elliott, J., Deryng, D., Folberth, C., Olin, S., Schmid, E., & Arneth, A. (2016). Climate analogues suggest limited potential for intensification of production on current croplands under climate change. *Nature Communications*, 7, 12608.
- Räisänen, J., & Ruokolainen, L. (2006). Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus A: Dynamic Meteorology and Oceanography*, 58, 461–472.
- 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*,

- 1003 12. Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A. 1074  
 1004 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) 1076  
 1005 Assessing agricultural risks of climate change in the 21st century in a global 1078  
 1006 gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, 111, 3268–3273. 1079  
 1007 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. 1081  
 1008 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. 1090  
 1009 (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. 1091  
 1010 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. 1092  
 1011 Ruane, A. C., Cecil, L. D., Horton, R. M., Gordon, R., McCollum, R., Brown, D., Killough, B., Goldberg, R., Greeley, A. P., & Rosenzweig, C. (2013). Climate change impact uncertainties for maize in panama: Farm information, climate projections, and yield sensitivities. *Agricultural and Forest Meteorology*, 170, 132 – 145. 1093  
 1012 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. 1094  
 1013 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. 1095  
 1014 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. 1096  
 1015 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. 1097  
 1016 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. 1098  
 1017 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. 1099  
 1018 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. 1100  
 1019 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. 1101  
 1020 Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters*, 35. 1102  
 1021 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. 1103  
 1022 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. 1104  
 1023 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. 1105  
 1024 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. 1106  
 1025 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. 1107  
 1026 de Wit, C. (1957). Transpiration and crop yields. *Verslagen van Landbouwkundige Onderzoeken* : 64.6. 1108  
 1027 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. 1109  
 1028 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., Müller, 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. 1110