

Strong regional influence of climatic forcing datasets on global crop model ensembles

Ruane, Alex C.; Phillips, Meridel; Müller, Christoph; Elliott, Joshua; Jägermeyr, Jonas; Arneth, Almut; Balkovic, Juraj; Deryng, Delphine; Folberth, Christian; Iizumi, Toshichika; Izaurralde, Roberto C.; Khabarov, Nikolay; Lawrence, Peter; Liu, Wenfeng; Olin, Stefan; Pugh, Thomas A.M.; Rosenzweig, Cynthia; Sakurai, Gen; Schmid, Erwin; Sultan, Benjamin

DOI:

[10.1016/j.agrformet.2020.108313](https://doi.org/10.1016/j.agrformet.2020.108313)

License:

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Ruane, AC, Phillips, M, Müller, C, Elliott, J, Jägermeyr, J, Arneth, A, Balkovic, J, Deryng, D, Folberth, C, Iizumi, T, Izaurralde, RC, Khabarov, N, Lawrence, P, Liu, W, Olin, S, Pugh, TAM, Rosenzweig, C, Sakurai, G, Schmid, E, Sultan, B, Wang, X, de Wit, A & Yang, H 2021, 'Strong regional influence of climatic forcing datasets on global crop model ensembles', *Agricultural and Forest Meteorology*, vol. 300, 108313. <https://doi.org/10.1016/j.agrformet.2020.108313>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

1
2 **Strong regional influence of climatic forcing datasets on global crop model ensembles**
3 Alex C. Ruane¹, Meridel Phillips^{2,1}, Christoph Müller³, Joshua Elliott⁴, Jonas Jägermeyr^{4,1},
4 Almut Arneth⁵, Juraj Balkovic^{6,7}, Delphine Deryng⁸, Christian Folberth⁷, Toshichika Iizumi⁹,
5 Roberto C. Izaurralde^{10,11}, Nikolay Khabarov⁷, Peter Lawrence¹², Wenfeng Liu^{13,14},
6 Stefan Olin¹⁵, Thomas A. M. Pugh^{16,17}, Cynthia Rosenzweig¹, Gen Sakurai⁹, Erwin Schmid¹⁸,
7 Benjamin Sultan¹⁹, Xuhui Wang^{13,20}, Allard de Wit²¹, Hong Yang^{14,22}

8
9 ¹*NASA Goddard Institute for Space Studies, New York, New York, USA*

10 ²*Columbia University, New York, New York, USA*

11 ³*Potsdam Institute for Climate Impacts Research, Member of the Leibniz Association, Potsdam,*
12 *Germany*

13 ⁴*University of Chicago, Chicago, Illinois, USA*

14 ⁵*Department Atmospheric Environmental Research, Karlsruhe Institute of Technology,*
15 *Garmisch-Partenkirchen, Germany*

16 ⁶*Department of Soil Science, Comenius University in Bratislava, Bratislava, Slovak Republic*

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

19 ⁸*IRI THESys, Humboldt-Universität zu Berlin, Berlin, Germany*

20 ⁹*National Agriculture and Food Research Organization, Tsukuba, Ibaraki, Japan*

21 ¹⁰*Department of Geographical Sciences, University of Maryland, College Park, MD, USA*

22 ¹¹*Texas AgriLife Research and Extension, Texas A&M University, Temple, TX, USA*

23 ¹²*Earth System Laboratory, National Center for Atmospheric Research, Boulder, CO, USA*

24 ¹³*Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ,*
25 *Université Paris-Saclay, Gif-sur-Yvette, France*

26 ¹⁴*Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland*

27 ¹⁵*Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden*

28 ¹⁶*School of Geography, Earth & Environmental Science, University of Birmingham,*
29 *Birmingham, United Kingdom*

30 ¹⁷*Birmingham Institute of Forest Research, University of Birmingham, Birmingham, United*
31 *Kingdom*

32 ¹⁸*University of Natural Resources and Life Sciences, Vienna, Austria*

33 ¹⁹*ESPACE-DEV, Institute of Research for Development, Montpellier, France*

34 ²⁰*Sino-French Institute of Earth System Sciences, Peking University, Beijing, China*

35 ²¹*Earth Observation and Environmental Informatics, Alterra Wageningen University and*
36 *Research Centre, Wageningen, Netherlands*

37 ²²*Department of Environmental Sciences, MGU, University of Basel, Basel, Switzerland*

38
39 December 18th, 2020

40
41
42 Corresponding Author's Address

43 Alex Ruane

44 2880 Broadway

45 New York, NY 10025, USA

46 alexander.c.ruane@nasa.gov

47 **Abstract**

48 We present results from the Agricultural Model Intercomparison and Improvement Project
49 (AgMIP) Global Gridded Crop Model Intercomparison (GGCMI) Phase I, which aligned 14 global
50 gridded crop models (GGCMs) and 11 climatic forcing datasets (CFDs) in order to understand
51 how the selection of climate data affects simulated historical crop productivity of maize, wheat,
52 rice and soybean. Results show that CFDs demonstrate mean biases and differences in the
53 probability of extreme events, with larger uncertainty around extreme precipitation and in regions
54 where observational data for climate and crop systems are scarce. Countries where simulations
55 correlate highly with reported FAO national production anomalies tend to have high correlations
56 across most CFDs, whose influence we isolate using multi-GGCM ensembles for each CFD.
57 Correlations compare favorably with the climate signal detected in other studies, although
58 production in many countries is not primarily climate-limited (particularly for rice). Bias-adjusted
59 CFDs most often were among the highest model-observation correlations, although all CFDs
60 produced the highest correlation in at least one top-producing country. Analysis of larger multi-
61 CFD-multi-GGCM ensembles (up to 91 members) shows benefits over the use of smaller subset
62 of models in some regions and farming systems, although bigger is not always better. Our analysis
63 suggests that global assessments should prioritize ensembles based on multiple crop models over
64 multiple CFDs as long as a top-performing CFD is utilized for the focus region.

65

66

67 *Keywords:* Agricultural Model Intercomparison and Improvement Project (AgMIP); Global
68 Gridded Crop Model Intercomparison (GGCMI); Climatic Forcing Datasets; Climate Impacts;
69 Agroclimate; Crop production

70

71 **1. Introduction**

72 Global agricultural systems are vulnerable to climate hazards including extreme events and long-
73 term trends that alter the growth environment. Cultivar and farm management practices are often
74 selected to produce high and stable yields within the current expected climate, but this still leads
75 to underperforming years as well as emerging pressures for adaptation as regional climates shift
76 under anthropogenic climate change (Lobell et al., 2011; Mbow et al., 2019; Porter et al., 2014;
77 Rosenzweig et al., 2014). Understanding regional agricultural systems' climate hazard profile is
78 critical to major international goals for disaster preparedness (e.g., the Sendai Framework;
79 UNISDR, 2015), greenhouse gas mitigation (e.g., the Paris Agreement, United Nations, 2015a),
80 and the Sustainable Development Goals (United Nations, 2015b). Planning for current and future
81 farming systems is therefore rooted in solid analysis of crop response to recent historical climate,
82 which then acts as a baseline for the generation of future agroclimatic scenarios to allow
83 investigation of adaptation, mitigation, and resilience-building interventions (Antle et al., 2015;
84 Lange, 2019a; Ruane et al., 2015). As many of the world's most vulnerable agricultural regions
85 are found in areas with incomplete or inconsistent meteorological observations, the Agricultural
86 Model Intercomparison and Improvement Project (AgMIP¹) has developed protocols and datasets
87 to fill in observational gaps in order to provide a consistent climatic forcing for agricultural models
88 across AgMIP and related simulation projects (Rosenzweig et al., 2013; Ruane et al., 2015; Ruane
89 et al., 2017).

90

91 In this study, we investigate the hypothesis that the selection of a climatic forcing dataset (CFD)
92 has strong influence on the fidelity of crop models simulating regional production of maize, wheat,

¹ Abbreviations: AgMIP: The Agricultural Model Intercomparison and Improvement Project; CFD: Climatic Forcing Dataset; GGCM: Global Gridded Crop Model; GGCMi: Global Gridded Crop Model Intercomparison

93 rice, and soy. To do this we utilize global agricultural model simulations conducted as part of the
94 AgMIP Global Gridded Crop Model Intercomparison (GGCMI, Elliott et al., 2015; Müller et al.,
95 2017; see Supplementary Material S1), allowing us to investigate multi-model ensembles to reduce
96 model-specific bias. Our final analysis of simulation skill is the correlation between crop model
97 ensembles and the time series of national level production (Figure 6), with the preceding figures
98 and sections providing examples and analysis approaches that help interpret differences across
99 nations, crop systems, and crop model ensembles (further bolstered by the Supplementary
100 Material). Influence of CFDs on production will depend on (i) the accuracy of climatic forcing
101 datasets (CFDs) in capturing mean climate and resolving extreme events (Section 3), (ii) the ability
102 of crop model biophysical process representations to capture important climatic responses (Section
103 4), and (iii) whether CFD differences align with critical crop model processes and structural
104 differences in a manner that would lead to noticeable differences in agricultural response (Section
105 5). In this way we may apply the agricultural impacts lens to identify important differences in
106 climate datasets that would otherwise be too subtle to distinguish. The structure of the GGCMI
107 intercomparison also allows us to investigate the role of CFD selection within the context of
108 GGCM/CFD ensembles including up to 91 members.

109
110

111 **2. Material and methods**

112 *2.1 Climatic forcing datasets*

113 Crop models typically require daily meteorological inputs including maximum and minimum
114 temperature (T_{max} and T_{min}), precipitation (P), and solar radiation (S_{rad}). Many crop models also
115 require information about humidity (relative humidity, vapor pressure deficit, or dew point

116 temperature), longwave radiation, and wind speed in order to more accurately estimate potential
117 evapotranspiration. Some models utilize hourly information to better understand processes related
118 to the diurnal cycle. High-quality *in situ* measurements remain the gold standard for model
119 simulations, with remote sensing and retrospective analyses ('reanalyses') filling in gaps in space
120 and time (Gelaro et al., 2017; Schollaert Uz et al., 2019). Agricultural applications benefit from
121 the combination of best performing products (Toreti et al., 2019), although care must be taken to
122 ensure that CFDs utilize bias adjustment techniques that maintain the statistics most relevant to
123 crop models (Famien et al., 2018; Galmarini et al., 2019; Parkes et al., 2019). CFDs created for
124 application across multiple scales, regions or sectors (e.g., Lange, 2019) may face additional
125 constraints in terms of variable and water/energy budget consistency than would be required of
126 only a single scale and sector.

127

128 Reanalyses are numerical weather prediction models reinitialized multiple times each day using
129 assimilation of observational data . These do not assimilate the specific variables needed for crop
130 models, however, so variables like maximum and minimum temperature, precipitation rate,
131 incident solar radiation, and near-surface humidity are the products of internal model processes
132 and parameterizations. Observational datasets also have uncertainties and biases, particularly in
133 regions where local observations are sparse, of poor quality, or difficult to access (Iizumi et al.,
134 2014, 2017; Ruane et al., 2015). Historical CFDs are typically generated by combining the
135 universal coverage and physical consistency of reanalysis outputs with observational data from
136 gridded *in situ* measurements and satellite remote sensing in order to create a uniform, coherent,
137 and bias-adjusted dataset to drive impact models. The resulting CFD is a globally-coherent dataset

138 with day-to-day sequences and variable relations from the reanalysis that have been adjusted to
139 ensure that monthly statistics match observational products.

140

141 **Table 1** provides an overview of the 11 climatic forcing datasets (CFDs) used in the GGCM Phase
142 1 simulations evaluated in this study, including their underlying reanalyses, key bias-adjustment
143 targets (in situ station and remote sensing products), and special notes on key aspects of the bias
144 adjustment. Many of these datasets were compared against global station data by Ruane et al.,
145 (2015a), which also includes additional distinction between bias-adjustment methods in the
146 various products. The GRASP dataset is particularly unique in that it does not adjust biases on a
147 monthly basis according to target observational datasets; rather, the 1961-1990 period was used to
148 determine time-constant adjustment factors that are then applied to reanalysis data over the entire
149 1980-2010 period (Iizumi et al., 2014).

150

151 Several CFDs share common characteristics that allow us to isolate the ramifications of particular
152 options in the CFD-generation process. AgMERRA and AgCFSR utilize the same bias-adjustment
153 methods and target observational datasets but differentiate in their selection of underlying
154 reanalysis (same monthly values but different daily sequences). AgCFSR and CFSR are driven by
155 the same reanalysis, but CFSR does not undergo any bias adjustment (same daily sequence but
156 different monthly values). Likewise, both WFDEIcru, and WFDEIgpcc are based on the ERA-
157 INTERIM reanalysis, which is also included without bias-adjustment (ERA-Interim, same daily sequence
158 but different monthly values). Additionally, WFDEIcru and WFDEIgpcc use the same bias-
159 adjustment methods and target datasets with the exception of different monthly precipitation
160 dataset targets (CRU or GPCC) (same daily sequences and monthly values except for monthly

161 precipitation). WFDEIcru and WFDEIgpcc also represent an updated application of the WATCH
 162 methodology, while PFGv2 is an update to the Princeton CFD.

163

164 **Table 1:** Overview of Climatic forcing Datasets (adapted from Elliott et al., 2015). Acronyms are explained in table
 165 footnotes. All GGCMs utilize these CFDs' daily maximum and minimum temperatures, precipitation, and solar
 166 radiation, with some models additionally using wind speed, humidity, vapor pressure, and dewpoint temperature.

Dataset	Underlying reanalysis (resolution) [reference]	Years	CFD native resolution	Bias-adjustment notes (and key reference)
AgCFSR	CFSR (≈0.3°) [Saha et al., 2010]	1980-2010	0.5°/0.25°	Monthly temperature and precipitation values match ensemble of CRUTS3.10 (Harris et al., 2014), GPCCv6 (Fuchs, 2009; Rudolf et al., 2010), and WM (Willmott and Matsuura, 1995), with adjustment to CRUTS3.00 wet day frequency and SRB solar radiation (Stackhouse, Jr et al., 2011). Diurnal temperature range matches CRU (on average). Monthly precipitation climatology from high-resolution satellite products, although that information is lost at 0.5° resolution used in this study. Includes vapor pressure, dew point temperature, and relative humidity at time of maximum temperature (Ruane et al., 2015)
AgMERRA	MERRA (0.5°x0.66°) [(Rienecker et al., 2011)]	1980-2010	0.5°/0.25°	Same bias-adjustment targets and methods as AgCFSR, but diurnal temperature range is adjusted be ¾ of the distance between MERRA and CRU (on average) and incorporates precipitation sequence from MERRA-Land (Reichle et al., 2011) dataset that utilizes GPCP observations (Ruane et al., 2015)
CFSR	CFSR (≈0.3°) [Saha et al., 2010]	1979-2011	0.3°	No bias-adjustment from original reanalysis (Saha et al., 2010)
ERA-Interim	ERA-Interim (0.75°) [ECMWF, 2009]	1979-2019	0.75°	No bias-adjustment from original reanalysis (ECMWF, 2009)
GRASP	JRA-25 (1.125°) & ERA-40 (2.5° version) [Uppala et al., 2005]	1961-2010	1.125°	Adjusts to CRU-TS3.10 for temperature and precipitation, CRUTS3.0 wet-day frequency, CRU-CL1.0 winds, and SRB solar radiation. Time-constant correction factors derived from 1961-1990 period (Iizumi et al., 2014)
GSWP3	20CR (2°) [Compo et al., 2011]	1901-2010	0.5°	Adjusts to GPCP precipitation, SRB solar radiation, and CRU temperature (Dirmeyer et al., 2006)
PGFv2	NCAR Reanalysis 1 (2.8°) [Kalnay et al., 1996]	1901-2012	0.5°	Adjusts to CRU, GPCP, SRB, and utilizes the TRMM Multi-satellite Precipitation Analysis (Sheffield et al., 2006)
Princeton	NCAR Reanalysis 1 (2.8°) [Kalnay et al., 1996]	1948-2008	0.5°	Adjusts to CRU TS2.0, GPCP, SRB, and utilizes the TRMM Multi-satellite Precipitation Analysis (Sheffield et al., 2006)

WATCH	ERA-40 (2.5°) [Uppala et al., 2005]	1958- 2001	0.5°	Adjusts to CRUTS2.1 temperature and GPCCv4 precipitation. Also known as WATCH Forcing Data (WFD) (Weedon et al., 2011); listed as element of GGCM Phase 1, but not included in further analysis given year 2001 end date.
WFDEIcru	ERA-Interim (0.75°) [Dee et al., 2011]	1979- 2012	0.5°	Monthly corrections to CRU TS3.1/CRUTS3.101/CRUTS3.21; includes longwave radiation (Weedon et al., 2018)
WFDEIgpcc	ERA-Interim (0.75°) [Dee et al., 2011]	1979- 2010	0.5°	Same as WFDEIcru but precipitation adjusted to GPCCv5/v6 (Weedon et al., 2018)

* All CFDs were applied on a common 0.5° x 0.5° grid for crop model simulations and analyses in this study; Native resolution shows highest level of distinction for CFD (AgMERRA and AgCFSR shown separately for all variables / precipitation only)

NCAR = National Center for Atmospheric Research (USA)	GPCP = Global Precipitation Climatology Project (USA)
CFSR = NCAR Climate Forecast System Reanalysis (USA)	GPCC = Global Precipitation Climatology Centre (Germany)
MERRA = Modern Era Retrospective-analysis for Research and Applications (USA)	CRU = Climatic Research Unit (University of East Anglia, UK)
JRA-25 = 25-year Japanese Reanalysis (Japan)	CMORPH = Climate Prediction Center Morphing Product (USA)
ERA-40 = European Centre for Medium-range Weather Forecasting 40 year reanalysis (UK)	PERSIANN = Precipitation Estimation using Remote-Sensing and Artificial Neural Networks (USA)
ERA-Interim = European Centre for Medium-range Weather Forecasting Interim reanalysis (UK)	TRMM = National Aeronautics and Space Administration Tropical Rainfall Measurement Mission (USA)
WM = Willmott and Matsuura, 1995	TMPA = TRMM Multi-satellite Precipitation Analysis (USA)
SRB = NASA/GEWEX Solar Radiation Budget (USA)	

167
168 This study analyzes agroclimatic aspects of CFDs using methods established in Ruane et al. (2018)
169 to target agricultural productivity. Seasonal climate factors are calculated according to the local
170 major growing seasons for maize, wheat, rice, and soybean determined by GGCM protocols for
171 planting and average harvest dates (Elliott et al., 2015). In many cases this information is
172 documented on a country-level, missing differences within a country that can be important to
173 regional production.
174
175 We evaluate CFDs for the 1980-2010 period, offering a ‘current’ climatology containing the 30
176 complete growing seasons that led to harvests from 1981-2010. This includes data from 1980 to
177 account for regions where the growing season overlaps January 1st such that planting in 1980 led
178 to a harvest in 1981. Simulations were run with CO₂ concentration data from Mauna Loa (Thoning
179 et al., 1989). This period also included substantial climatic trends in many regions owing to large-

180 scale modes of climate variability, as well as anthropogenic climate change, which required us to
181 detrend GGCM outputs when comparing against detrended FAO production anomalies (which
182 were also detrended, as described in Section 2.4 below). The WATCH forcing dataset is not
183 included in further analyses for this study given that it does not extend beyond 2001, but we do
184 include analysis of simulations driven by the Princeton dataset up to 2008.

185

186 *2.2 Global gridded crop models*

187 Crop models track daily water, carbon, and nitrogen balances in the plant and field environment
188 progressing through developmental stages as determined by genotype parameters, field
189 management, and climate drivers. These models have been developed using extensive observations
190 and field and chamber trials, with many AgMIP-facilitated intercomparisons helping to elucidate
191 strengths and weaknesses associated with various modeling approaches (Martre et al., 2015;
192 McDermid et al., 2015; Ruane et al., 2017; Zhao et al., 2017).

193

194 The process-based crop models utilized in this study (Elliott et al., 2015; Müller et al., 2017) are
195 configured using information about the cultivar genotype (e.g., temperature-based phenology, heat
196 and drought resistance), soils (e.g., 1 to 2 meters of layered texture and water holding properties),
197 farm management (e.g., tillage methods, planting and harvest dates, fertilizer and irrigation
198 applications), and climate (as noted in previous section). Müller et al. (2019) and Supplementary
199 Material S2 provide a more complete description of the 14 GGCM models and 3 configuration
200 types utilized, including 2 configurations in which growing season and fertilizer levels are
201 harmonized for consistency. Irrigation is assumed to be unconstrained by water availability and
202 any soil water deficit is balanced the next day without application or conveyance losses.

203 Calibration of any model parameters was performed at the global scale, although modelers
204 configured soils, cultivars, and management practices regionally (e.g., to match GGCMI growing
205 season harmonization protocols). Observational production data were used by some models to
206 calibrate mean yields, but no models incorporated information about the observed interannual
207 anomalies in focus for this study.

208
209 The goal of this current study is to isolate the role of climatic forcing dataset and ensemble
210 selection in GGCM historical performance, and we refer readers to (Müller et al., 2017) for a more
211 detailed evaluation of GGCM-based differences in capturing historical national yield variation.
212 The group of models include several with common origins, as described by Rosenzweig et al.
213 (2014; Supplementary Information); however, large variations in included model processes,
214 configuration settings and calibration datasets mean that each of the models in the ensemble are
215 substantially distinct from one another (see Müller et al., 2019, and Supplementary Information
216 S2). Folberth et al. (2019) further evaluated differences in the 5 different modeling group
217 simulations stemming from the EPIC model, finding that yield estimates were distinguished by
218 differences in model versions, parameterization, management assumptions (beyond those
219 harmonized within GGCMI), soil attributes, and cultivar distributions.

220

221 *2.3 Simulation subsets and ensembles for analysis*

222 **Table 2:** Coverage of Global Gridded Crop Models (GGCMs), Climate Forcing Datasets (CFDs), and GGCM
223 configuration settings (see Supplementary Material S2 and S3 for configuration and model information). Underlined
224 models are used in the ‘+’ subset for each CFD, and the bolded configuration was selected for analysis when outputs
225 from multiple configurations were submitted for a given GGCM.

		Climatic Forcing Datasets										
		AgCFSR	AgMERRA	CFSR	ERA1	GRASP	GSWP3	PGFv2	Princeton	WATCH \$	WFDEIcru	WFDEIgpcc
GGCMs	CGMS-WOFOST		D									D
	CLM-Crop		H, N, D									H, N, D
	EPIC-BOKU	H, N, D	H, N, D	H, N, D	H, N, D	H, N, D	H, N, D	H, N, D	H, N, D	H, N, D	H, N, D	H, N, D
	EPIC-IIASA ^	H, N	H, N, D			H				H, D		H, N, D
	EPIC-TAMU *	H, N	H, N	H, N	H, N	H, N			H, N	H, N	H, N	H, N
	GEPIC	H	H, N, D			H		D		H		H, N, D
	LPJ-GUESS	N	N, D	N	N	N	D	N, D	N	N	N	N, D
	LPJmL	N, D	N, D	N, D	N, D	N, D	D	N, D	N, D	N, D	N, D	N, D
	pAPSIM ~	H, N, D	H, N, D	H, N, D	H, N, D	H, N, D			H, N, D	H, N, D	H, N, D	H, N, D
	pDSSAT	H, N, D	H, N, D	H, N, D	H, N, D	H, N, D			H, N, D	H, N, D	H, N, D	H, N, D
	PEGASUS ~	H	H, N, D	H	H				H	H	H	H, N, D
	ORCHIDEE-CROP		H, N, D				H, D †		H, N, D ††			H, N, D †††
	PEPIC	H, N, D	H, N, D			H, N, D	H, N, D	H, N, D			H, N, D	H, N, D
	PRYSBIZ +	D	D			D			D	D	D	D

^ EPIC-IIASA did not run Rice with AgCFSR or WFDEIgpcc

* EPIC-TAMU only ran Maize and Wheat

~ pAPSIM and PEGASUS did not run Rice

+ PRYSBIZ only ran irrigated lands

\$ WATCH only goes to 2001; not included in ensemble

Simulations Available	GGCM Configurations
1 Configuration	H = Harmonized
2 Configurations	N = No N Limitation
3 Configurations	D = Default

† No Rice with H

†† No Wheat with H,

no Rice or Wheat with N,

only Rice with D

††† Only Rice and Wheat with N

226

227

228 **Table 2** shows the complete set of GGCM Phase 1 simulations, which were run for both rainfed
 229 and irrigated conditions. Gaps in the table reflect that resource and structural constraints limited
 230 the ability of many modeling teams to run every requested combination of CFD, configuration and
 231 crop species. In order to achieve complete multi-model coverage for at least two WFDs, each
 232 GGCM was specifically requested to run the AgMERRA and WFDEIgpcc CFDs and then as many
 233 additional CFDs as resources allowed. There are relatively fewer simulation outputs submitted for
 234 the GSWP3 and PGFv2 CFDs as these were added to the GGCM protocol later in the project
 235 timeline. As our interest is in determining the response of GGCMs to the CFDs' growing season
 236 climate, we prioritize the simulations with consistent planting and harvest dates ([H and N] > D)
 237 and selected configurations that included nitrogen limitations where available (H>N), resulting in
 238 a final prioritization of H>N>D (see Supplementary S2 for further model and configuration
 239 information). Analysis here focuses on the relative seasonal anomalies for each GGCM simulation,
 240 which are a better reflection of climatic response than the raw anomalies influenced by mean bias

241 and further questions of model configuration such as soil nitrogen and cultivar characteristics
242 (Müller et al., 2017).

243

244 To isolate the implications of the CFD selection in the full ensemble, we identify two types of
245 GGCM-CFD groupings that sample across the crop model dimension:

246 ‘+’ *subset* [per CFD]: A consistent subset of GGCMs across CFDs, representing the 7 GGCMs
247 (5 for rice) that ran most CFDs (underlined in Table 2): EPIC-BOKU, EPIC-TAMU,
248 LPJ-GUESS, LPJmL, pAPSIM, pDSSAT and PEGASUS, using the bolded and
249 underlined configuration in Table 2. The ‘AgMERRA+’ subset, for example, is the
250 ensemble average of these 7 GGCMs simulating the AgMERRA CFD using the
251 specified configuration.

252 ‘All’ *subset* [per CFD]: All GGCMs that ran a given CFD, using the bolded configuration. The
253 ‘AgMERRA_all’ subset, for example, includes all GGCMs that ran the AgMERRA
254 CFD using the specified configuration.

255

256 We also form ensembles across both the climate and crop model dimensions of GGCMs in order
257 to look at overall GGCM performance:

258 ‘Ensemble+’ *subset*: All GGCMs that were included in the + ensembles across all CFDs
259 (bolded and underlined in Table 2). This represents the aggregate performance of the
260 core set of GGCMs that ran most CFDs.

261 ‘Ensemble-all’ *subset*: All GGCM/CFD combinations marked as bold in Table 2 (e.g., 91
262 model simulations in total for maize). To our knowledge this is the largest
263 GGCM/CFD ensemble to have been constructed, and we examine it here to quantify

264 the potential added benefit given that the resources required for such large community
265 efforts typically preclude their use for individual applications.

266

267 Each of these subsets is designed to build on AgMIP findings that the statistics of an ensemble of
268 models performs better than any single model when evaluated across a broad spectrum of
269 environments and systems (Bassu et al., 2014; Fleisher et al., 2017; Jägermeyr et al., 2020; Martre
270 et al., 2015; Müller et al., 2017; Nelson et al., 2014; Wallach et al., 2015; Zhao et al., 2017).
271 Consequently, no model is given more weight within any particular ensemble when calculating
272 ensemble statistics (Wallach et al., 2016). Müller et al. (2017) provide analysis of individual
273 GGCM performance, which is not our focus here,

274

275 Analysis of the ‘+’ subsets for each CFD therefore provides unprecedented insight into these
276 CFDs’ effects on agricultural simulations with a consistent crop model ensemble rather than being
277 dependent on a single crop model. Note that the ‘+’ ensemble contains 7 models for maize, wheat,
278 and soybean, but only 5 models for rice given that pAPSIM and PEGASUS did not provide data
279 for rice. The ‘+’ ensemble includes two EPIC GGCMs but these employ different core EPIC model
280 versions and a number of differences in configuration for soils and management (Folberth et al.,
281 2019). The ‘All’ subsets indicate whether the inclusion of additional GGCMs would have altered
282 the ensemble’s response to the CFD response. These contrast with the ‘*Ensemble-all*’ subset that
283 provides the overall GGCM Phase 1 ensemble performance, which benefits from both an
284 ensemble of CFDs and GGCMs although the relative weight of each depends on the outputs
285 provided (Table 2). An example of GGCM/CFD ensemble construction is provided for Romanian
286 maize production anomalies in Figure S.2.

287

288 *2.4 Production datasets and processing*

289 GGCMs simulate crop yields (t/ha) that must be converted to production (total kg) using harvested
290 area masks in order to compare against observational production datasets. We calculate national-
291 level production from the 0.5° x 0.5° grid using harvested crop areas from the Spatial Production
292 Allocation Model v2.0 (SPAM), which approximates the year 2005 and does not change from year
293 to year (You et al., 2014). We aggregate rainfed and irrigated production values separately using
294 the corresponding GGCM simulations and SPAM areas, then use the sum of rainfed and irrigated
295 production for national or global totals (following Ruane et al., 2018b; Porwollik et al., 2017).

296

297 Reference national production data are drawn from the United Nations Food and Agricultural
298 Organization (<http://www.fao.org/faostat/en/#data>). These data are reported by governments and
299 include heterogeneous cultivars, planting dates, fertilizer applications, irrigation methods, farm
300 management, and soils that cannot be fully represented by GGCM's relatively coarse resolution
301 configurations. FAOstat data also reflect agricultural trends and anomalies beyond those driven
302 solely by field-level climate such as the effects of technological improvements, mechanization,
303 agricultural sector development, labor supply, geopolitical conflict, crop pests and diseases, and
304 large-scale disasters (e.g., earthquakes, floods, hurricanes). Overall, Ray et al., (2015) estimated
305 that the climate signal explains only about a third of observed global interannual yield variability.
306 For these reasons we detrend FAOstat data and crop model outputs. GGCM has explored multiple
307 methods for detrending including first-difference, linear and polynomial fits, and there is a clear
308 tradeoff between consistency in methods and unique characteristics in production time series that
309 defy general approaches. While further efforts to isolate the climate signal in national production

310 datasets using a blend of locally-selected detrending techniques would be beneficial to GGCM
311 and the broader agricultural community, here we calculate anomalies from a 5-year moving
312 average and compare against similarly detrended GGCM outputs (as described in Müller et al.,
313 2017, and further evaluate in Supplemental Materials S8). We assign each simulated growing
314 season according to the average harvest date for the purpose of time series correlations, which can
315 cause an occasional mismatch with FAO data that assigns harvests to the growing season in which
316 the majority of the growing period occurs, leading to occasional differences for locations and
317 seasons with early or late harvests that are on the other side of New Year's day than the average
318 harvest date. Additional information on the use of production datasets is provided in
319 Supplementary Materials S4.

320
321 To understand the role of climate variability on a sub-national scale we also employ the detrended
322 United States Department of Agriculture's National Agricultural Statistics Service (USDA NASS)
323 county-level production (<https://quickstats.nass.usda.gov/>). NASS production data are collected
324 using reported and surveyed yields. We combine the average of NASS 1981-2010 county-level
325 cropped areas with simulated yields to calculate simulated county-level production for comparison
326 to NASS production anomalies.

327
328 We analyze uncertainty by determining the relative variation across ensemble members for each
329 year compared to the variation of the ensemble median across years. Anomalies of precipitation
330 and yield are first calculated as percentages to remove the effects of mean biases. We then
331 calculate a standardized anomaly, which is the ratio of (i) the standard deviation of yearly ensemble
332 member anomalies (compared to the ensemble mean) to (ii) the standard deviation of the ensemble

333 mean time series itself. Standardized anomalies >1 therefore indicate that a given annual anomaly
334 is more likely due to ensemble member differences, while standardized anomalies <1 indicate that
335 anomalies are likely representative of true interannual variation. Supplementary Material S.7
336 provides further detail on this method as well as two contrasting examples (Figure S.5).

337

338 **3. Differences between climatic forcing datasets**

339 CFD regional differences can be measured in myriad ways, including in their mean quantities,
340 statistical distributions, sequencing of events, variable relationships, modes of variability, long-
341 term trends, and spatial coherence. While a comprehensive atlas of CFD differences for each
342 growing season is beyond the scope of this paper, **Figure 1** provides the median of the *CFD-all*
343 ensemble for the rainfed maize growing season as well as biases for AgMERRA and WFDEIgpcc,
344 which are the CFDs most commonly used within GGCM Phase 1. Corresponding bias maps for
345 the other CFDs are provided in Supplementary Figures S.2-S.4. It is important to emphasize that
346 the *CFD-all* median is not necessarily the true value given common biases in observational
347 datasets and methods across CFDs. Computing the median is likely to reduce some of the more
348 outlying values, however; and therefore serves as a ‘best-guess’ basis to help us identify CFD
349 differences that are likely relevant to agricultural production. The Princeton CFD was not included
350 in these *CFD-all* climate maps because it ends in 2008, and because it displayed a checkerboard-
351 like spatial bias pattern for precipitation threshold statistics. This suggests errors in re-gridding
352 and/or interpolation of daily sequences in the GGCM processing of that dataset, although this
353 pattern was not apparent in the mean precipitation rate or other variables. The following metrics
354 are evaluated for the rainfed maize growing season and cultivation regions as an example given
355 that maize is an important staple crop with widespread cultivation.

356

357 *3.1 Mean growing season metrics*

358 Median *CFD-all* mean temperature in the rainfed maize growing season (Fig. 1a) generally follows
359 mean climatological patterns with warmer conditions in the Tropics and cooler conditions at higher
360 latitudes, as maize generally corresponds to the warm season unless part of a multi-cropped region.
361 CFD differences for mean temperature are generally low ($<1^{\circ}\text{C}$). AgMERRA (Fig. 1b) is slightly
362 cooler than *CFD-all* in most of the United States, South America, Africa, Europe, and Indonesia,
363 and is slightly warmer in South and East Asia as well as the Middle East, Mexico, and South
364 America west of the Andes. WFDEIgpcc (Fig. 1c) has generally the opposite differential pattern
365 for the United States and Asia, and is also cooler than *CFD-all* in Europe, East Africa, and southern
366 South America.

367

368 Median *CFD-all* mean precipitation rate (Fig. 1g) reflects that rainfed maize generally grows
369 during the local wet season. AgMERRA is generally very close to the median CFD, with a slight
370 dry bias ($\approx 5\%$) in Southern Russia and scattered small regions around the world. WFDEIgpcc has
371 a widespread wet bias with prominent differences $>10\%$ in the US Midwest, southern South
372 America, central Africa, Europe, and eastern India. Dry bias pockets $>10\%$ are less common, but
373 include southwest India and Myanmar.

374

375 Solar radiation in the *CFD-all* (Fig. 1p) reflects a combination of latitude, aridity, and seasonality
376 of the growing period, with cloudier conditions in the moist Tropics and reduced solar radiation in
377 the cool season maize in SE China and northern Mexico. AgMERRA has solar radiation very close
378 to the ensemble median. This is likely because many CFDs used the same NASA/GEWEX SRB

379 information (Stackhouse, Jr et al., 2011) and the others did not substantially differ on aggregate.
380 WFDEIgpcc is generally cloudier in the tropics and sunnier at mid-latitudes ($\approx \pm 1.5$ MJ/m²/day).

381

382 *3.2 Distributional statistics within the growing season*

383 Days where maximum temperature exceeds 35°C (Fig. 1d) are associated with negative impacts
384 on maize pollination and production (Hatfield and Prueger, 2015), and patterns of this extreme
385 temperature are a reasonable proxy for similar heat stress thresholds of wheat, rice, and soybean
386 (Deryng et al., 2014; Schaubberger et al., 2017). The median *CFD-all* sees more of these extreme
387 heat days along the fringes of the major growing areas, including in the Sahel, Central Asia, NE
388 Brazil, and the SW Great Plains and NE Mexico. AgMERRA is similar to *CFD-all* in major
389 breadbaskets of the Central United States, Europe, and East Asia but tends to underestimate these
390 days (by ≈ 10) in many tropical areas while overestimating them in semi-arid zones of Southern
391 Africa, Southern South America, Central and West Asia, and the western Great Plains.
392 WFDEIgpcc has an overall tendency towards more extreme heat days than *CFD-all* (by ≈ 10 -15 in
393 many regions), particularly in North America and along the fringes of the Amazon although it is
394 similar to *CFD-all* in Europe and East Asia. WFDEIgpcc has more extreme heat even in several
395 regions that showed an overall cool bias in mean temperature, suggesting a larger diurnal
396 temperature range or broader distribution of daily extremes.

397

398 The number of wet days ($P > 0$ mm) within a growing season is an important proxy for the
399 likelihood of dry spells and the overall proportion of precipitation that reaches the root zone (as
400 opposed to running off). *CFD-all* median number of precipitation days per growing season (Fig.
401 1j) has a pattern generally similar to the mean growing season precipitation rate. AgMERRA has

402 fewer wet days in most maize-growing regions (especially in Africa, Mexico and South Asia),
403 while WFDEI_{gpc} has more wet days (particularly in Africa, Southern South America, Eastern
404 Europe, and the foothills of the Hindu-Kush-Himalayas). These differences are likely due to the
405 additional bias-adjustment of the number of precipitation days within AgMERRA, AgCFSR, and
406 GRASP which corrects a common drizzle-bias in reanalyses and leads to lower numbers than the
407 *CFD-all* median.

408

409 Heavy precipitation days can be problematic for crops given that they are often associated with
410 nitrogen leaching, and a larger proportion of total precipitation that falls as heavy precipitation
411 events can reduce the overall soil infiltration and heighten the risk of low soil column spells. The
412 median *CFD-all* number of days where $P > 20$ mm (Fig. 1n) has similar spatial patterns to the
413 mean precipitation field, with the most frequent heavy events in the Amazon and monsoon regions
414 of Asia. Different crop systems and soil profiles may have distinct thresholds for pluvial flooding
415 and high runoff proportions, but we employ $P > 20$ mm as representative of the higher tail of the
416 distribution and note that these days likely consist of heavier daily totals in smaller regions within
417 the larger grid cell (see Supplementary Material S9). AgMERRA has more heavy wet days in the
418 Tropics (≈ 3 more) and Western Africa in particular, likely as a secondary consequence of the
419 reduction in drizzle days resulting in fewer (but more intense) precipitation events to match
420 monthly totals. WFDEI_{gpc} has fewer very wet days than *CFD-all* with nearly the opposite
421 geospatial pattern of bias as AgMERRA but more substantial reduction over the rainforests of
422 Central Africa.

423

424

425 4. GGCM response to interannual climate

426 In order to understand the geographical distribution of climatic uncertainty, **Figure 2a,c** shows the
427 standardized anomalies of rainfed maize growing season temperatures and precipitation from
428 *CFD-all*, revealing the places where the CFD ensemble is less clear than a typical annual anomaly.
429 High values over the Western Amazon, Central and Western Africa, and Borneo reflect the
430 difficulty of obtaining high quality observational data in these regions. Standardized temperature
431 anomalies above one, indicating CFD variance is greater than interannual variance, are also seen
432 across much of Africa, the Hindu-Kush-Himalayas and Mexico, while lower values reflect larger
433 interannual variance and consistent observational data across North America, Europe, Southeast
434 Africa, India, East Asia, and Eastern South America. Most maize-growing areas that show high
435 standardized temperature anomalies also show high standardized precipitation anomalies, with
436 additional regions of larger CFD uncertainty for precipitation over East-Central Africa, the Middle
437 East, Central Asia, and Southeast Asia.

438
439 Standardized anomalies of *Ensemble-all* rainfed maize yield simulations (**Figure 2e**) reflect many
440 of the patterns seen in the standardized anomalies of growing season temperature and precipitation,
441 underscoring the role of climate uncertainty in the overall simulation uncertainty. Standardized
442 anomalies for simulated yield (peaking above 5 in some locations) are much larger than for the
443 climate variables (which peaked closer to 2), suggesting strong interactions between uncertain
444 GGCM configurations and climate variability within the simulated yields. High uncertainties are
445 particularly prominent in developing countries, where crop simulation models are typically more
446 difficult to configure given the relative lack of observational climate, soils, and agronomic data,
447 their greater proportion of small-holder farms, and heterogeneous cultivars and management that

448 may not be consistently represented across GGCMs (Fritz et al., 2015). Regions with lower
449 fertilizer usage have additional interactions between nitrogen stress and heat or water stress driven
450 by climate, which would only be captured in GGCMs including nitrogen dynamics. Very few
451 places have standardized yield anomalies below 1.

452

453 Standardized anomalies for the wheat and soybean (Supplementary Figure S.2) have similar
454 patterns, with lower standardized anomalies for temperature than precipitation and the highest
455 standardized anomalies coming from the simulated yield. Major production regions for maize,
456 wheat, and soybean, which tend to be in the middle latitudes, typically have standardized
457 anomalies <1 for climate variables, however the major production regions for rice (**Figure 2b,d,f**)
458 in Southeast Asia have standardized precipitation anomalies >1 , corresponding with substantial
459 yield uncertainty likely dependent on CFD selection.

460

461 **Figure 3** shows the correlation between median *Ensemble-all* yields with the median *CFD-all*
462 growing season mean temperature, precipitation, and solar radiation to identify regional and crop-
463 specific agroclimatic sensitivities. These fundamental climate responses motivate agricultural
464 management decisions to reduce risk and point to areas where uncertainty in CFD variables is
465 likely to strongly affect simulated yields. Higher correlations do not necessarily mean more
466 accurate simulations, only that the GGCM simulations for a given crop have a strong and consistent
467 response to regional variation of a particular climate variable.

468

469 Rainfed maize, wheat, rice, and soybean simulations each follow a common interannual pattern
470 dominated by precipitation, with a positive correlation associating wet years with higher yields

471 and the worst-yielding years generally associated with drought. This relationship is strongest in
472 areas with marginal rainfall totals and low irrigation, including NE Brazilian maize, wheat in the
473 western Great Plains of North America, rice in the Sahel, and soybean in SE Europe. Temperature
474 correlations are broadly negative, indicating that yields are higher in cool years and depressed in
475 hotter conditions. Regional pockets show a positive correlation with temperature, indicating that
476 warmer conditions can be beneficial along the cooler poleward and high-elevation fringes. Yield
477 is often negatively correlated with seasonal solar radiation anomalies, which is likely a reflection
478 of cross-correlations in the climate system whereby higher precipitation is associated with cloudier
479 weather and droughts with clearer skies. It is also likely that high temperatures are cross-correlated
480 with drier conditions and higher potential evapotranspiration.

481
482 Exceptions to this general pattern are also illustrative, as apparent in diverse median responses and
483 a lack of consistency across GGCM/CFD combinations (represented by the hatching in Figure 3).
484 Most crops are less sensitive to seasonal climate metrics in the moist tropics, where water is less
485 often a limiting factor and interannual variations are generally small compared to the average
486 growing season total. These areas are likely more responsive to shifts in sub-seasonal
487 characteristics such as heat waves and the onset, exit, break periods, and intense precipitation
488 events within rainy seasons. Rice, which is often grown in those moist tropical regions, is the least
489 dependent on seasonal climate anomalies, a result consistent with the finding of reduced sensitivity
490 to climate variability by Ray et al. (2015).

491
492 A comparison between rainfed and irrigated maize (top and bottom rows of **Figure 3**, respectively)
493 highlights the ways in which water management affects climate response, most notably by

494 reducing the dependence on precipitation anomalies. Simulations of irrigated maize are not
495 completely absent of precipitation response, however; showing signs that modeled irrigation
496 management does not eliminate water stress in places like Texas, Spain, the Indus Basin, and
497 Northern China. Negative responses to wet seasons may reflect nutrient leaching under increased
498 runoff in Central America, Northern Europe, and India. Irrigated maize in Northern Europe and
499 the northern Great Plains has an enhanced positive response to temperature compared to the rainfed
500 maize, possibly related to a reduction in water stress that can accompany a warmer season's higher
501 evapotranspiration demand. Irrigated areas also have relatively higher correlations with solar
502 radiation as water supply diminishes the effects of the cross-relationship between sunshine and
503 drought conditions.

504

505

506 **5. Crop model performance with different climatic forcing datasets**

507 The selection of climate forcing dataset(s) for GGCM applications often depends on the
508 availability of those inputs as well as the resources allocated to exploring CFD uncertainty and/or
509 benefiting from CFD ensemble behaviors. In this section we examine how the selection of a CFD
510 compares to the use of the full CFD ensemble, examining global CFD differences, performance
511 against regional production observations, and the simulations' ability to capture national
512 production anomalies. Differences in GGCM-CFD performance also highlight the ramifications
513 of a given CFD's selection of an underlying reanalysis and specific bias-adjustment targets and
514 methods, as well as non-climatic configurations that reduce GGCM correlations regardless of the
515 CFD selected.

516

517 *5.1 Global implications of CFD selection*

518 GGCM responses to CFD differences accumulate within any given regional farming system's
519 growing season, with the aggregate effect being a CFD-dependent crop yield for each grid cell for
520 each year. The temporal correlations between GGCM simulations using different CFDs therefore
521 indicate whether the CFD selection altered the overall climate response, with low correlations
522 indicating a fundamentally different agro-climatic relationship over the 1981-2010 period.

523

524 **Figure 4** presents the correlation between individual GGCM-CFD simulations and the median of
525 the *GGCM-all* ensemble. A full intercomparison of GGCMs across all crop systems is beyond the
526 scope of this study, so here we examine pDSSAT and LPJmL to explore potential interactions
527 between CFD selection and GGCM utilized. *LPJmL-AgMERRA* correlates highly with the median
528 of the *LPJmL-all* ensemble in much of the mid-latitudes; however, lower latitudes and many
529 developing countries have lower correlation suggesting more CFD-based uncertainty (Fig. 4a;
530 $r > 0.85$; with $r > 0.9$ in many high producing areas). This is consistent with the regional patterns of
531 heightened temperature and precipitation uncertainty shown in Figure 2. Regions of high
532 correlations between *LPJmL-AgMERRA* and *CFD-all* cover the vast majority of maize-growing
533 regions including major breadbaskets in the US Midwest, Europe, China, and South America. This
534 suggests that a single *LPJmL-AgMERRA* simulation provides a broadly similar response to using
535 all CFDs and then creating an ensemble median. This is not true for all CFDs, however, as can be
536 seen for *LPJmL-CFSR* where lower regional correlations indicate a different pattern of interannual
537 response imposed by that specific CFD (Fig 4b). pDSSAT generally shows a larger difference
538 between any CFD and the *CFD-all* runs, as the highest-correlated AgMERRA and WFDEIgpcc
539 simulations still have lower correlations than were seen for LPJmL rainfed maize (Figs. 4c-d). The

540 correlations of *LPJmL-WFDEIcru* and *LPJmL-WFDEIgpcc* vs. *LPJmL-all* for rainfed rice (Figs.
541 4e-f) show increased dependence on CFD (lower correlations) over the major rice production
542 zones of SE Asia than were seen for maize breadbaskets in places like the US Midwest (Fig 4a).
543 Even as *WFDEIcru* and *WFDEIgpcc* differ only in their monthly precipitation totals, *LPJmL*
544 simulations driven by *WFDEIgpcc* follow the *LPJmL-all* median closely, while those driven by
545 *WFDEIcru* are considerably lower in much of Brazil, the Democratic Republic of the Congo, and
546 Madagascar.

547

548 *5.2 Regional implications of CFD selection*

549 Differences between CFDs are likely to be heightened on smaller scales, particularly when they
550 interact with unique vulnerabilities in regional crop systems. A focus on sub-national heterogeneity
551 is also particularly important in large countries with production regions across multiple climate
552 zones. **Figure 5** examines sub-national features of rainfed maize simulations driven by various
553 CFDs against the US NASS county-level production anomalies.

554

555 The importance of bias-adjustment is underlined by comparisons between *pDSSAT-AgCFSR* and
556 *pDSSAT-CFSR*, with the non-bias-adjusted CFSR achieving substantially lower skill over nearly
557 all US rainfed maize regions with particularly low values over the northwest Midwest (from
558 Missouri through North Dakota, Fig. 5a,b). Both CFDs use the same underlying CFSR reanalysis,
559 so differences here are related to monthly mean climate, the imposition of SRB solar radiation,
560 changes in the number of precipitation days, and adjustments to the diurnal temperature range. A
561 similar reduction in skill is seen in *LPJmL* simulations using the non-bias-adjusted the ERAI
562 reanalysis compared to the *WFDEIgpcc*, which also is based on ERAI daily sequences (Figs. 5c-

563 d). In this case the swath of low-correlation simulations extending from Nebraska to Wisconsin
564 appears in simulations run with both CFDs, indicating a bias stemming from crop model
565 configuration rather than the selection of CFDs. Jägermeyr and Frieler, (2018) identified this as a
566 problem related to erroneous planting dates and cultivars that have been updated in later LPJmL
567 configurations.

568

569 The ensemble median of *pDSSAT-all* and *LPJmL-all* are highly correlated with NASS county-
570 level production for most of the US (Fig. 5e,f). Different regions exhibit strengths and weaknesses
571 for each GGCM, indicating that national level production anomalies are the aggregate across
572 regions with heterogeneous skill. In general, *pDSSAT-AgCFSR* is not substantially different from
573 *pDSSAT-all*, and *LPJmL-WFDEIgpcc* is likewise similar to *LPJmL-all*. This indicates that rainfed
574 maize simulations over the US can utilize one of these CFDs without losing too much information
575 that would otherwise be gained from the full CFD ensemble. AgCFSR, AgMERRA, WFDEIcru,
576 and WFDEIgpcc all capture similarly high levels of correlation for LPJmL and pDSSAT rainfed
577 maize, with CFSR and ERAI (the unadjusted reanalyses) and GRASP showing lower correlations.
578 In *some* regions the best-performing CFD has higher correlations than the *CFD-all* median, but
579 *CFD-all* excels at being near the top correlations for all regions.

580

581 *5.3 National implications of CFD selection*

582 **Figure 6** displays correlations between detrended FAO national production reports and simulated
583 production (including rainfed and irrigated areas) from 1981-2010. The top 20 producing countries
584 (2013-2017) for maize, wheat, rice, and soybean are shown using the *CFD+* ensembles (featuring
585 the largest common subset of GGCMs), allowing us to identify the climate-driven signal

586 (independent of GGCM differences) and its correlation with FAO reports for each country and
587 crop type. We also include the larger *AgMERRA-all* and *WFDEIgpcc-all* ensembles to understand
588 the ramifications of including additional GGCMs, *Ensemble+* to understand how an ensemble of
589 CFDs affects performance for the common GGCM subset, and *Ensemble-all* for the complete
590 GGCM Phase 1 set of GGCM-CFD combinations (bolded configurations in Table 2). The final
591 column in Figure 6 shows correlations between the simulation ensembles and the total global
592 production of each crop. Below we highlight the main features of these results, with broader
593 interpretation provided in the discussion section that follows.

594

595 *5.3.1 National maize production anomalies*

596 Simulations of leading national maize producers show statistically significant positive correlations
597 ($p < 0.05$) for many of the top producing countries, indicating that the simulations are capturing a
598 strong climatic signal within the FAO reports (Fig, 6a). The most apparent patterns in correlations
599 come from differences between countries, whereby simulations tend to have similarly high (or
600 low) correlations in all ensembles for a given country. This leads to stark differences between, e.g.,
601 Romania (relatively high correlations for nearly all ensembles) and Nigeria (relatively low and
602 insignificant correlations for nearly all ensembles). Due to Serbia's independence and separation
603 from Montenegro in 2006, only 5 years of FAO-reported production overlap with the 1981-2010
604 climatology, despite being a top-producer for maize and soybean in the 2013-2017 period;
605 therefore, correlations for Serbia have been excluded from Figures 6a,d.

606

607 Bias-adjusted CFDs tend to produce higher correlations in Figure 6 than the raw reanalyses (CFSR
608 and ERAI) and the GRASP dataset that adjusted according to fixed parameters determined from a

609 previous climatological period. *AgMERRA+* and *WFDEIgpcc+* are typically among the highest
610 *CFD+* correlations. The addition of GGCMs for *AgMERRA-all* and *WFDEIgpcc-all* did not show
611 clear benefits over the corresponding *AgMERRA+* and *WFDEIgpcc-all* (correlations improved in
612 10 and 8 of the 19 countries, respectively) This is similar to expectations given that there is a
613 reduced benefit when adding to an ensemble that already has 6 GGCMs unless a unique simulation
614 feature is added, which seems to be the case in Brazil given higher correlations for both although
615 the additional models lower correlations in Nigeria. The ensemble of the GGCM subset and CFDs
616 in *Ensemble+* is nearly identical to the full *Ensemble-all*, with the latter showing higher
617 correlations in 13 of 19 maize countries.

618

619 Several ensembles produce significant correlations with FAO global production reports. These
620 include *AgCFSR+*, *AgMERRA+*, *ERA+*, *WFDEIcru+*, *WFDEIgpcc+*, *AgMERRA-all*,
621 *WFDEIgpcc-all*, *Ensemble+*, and *Ensemble-all*. *WFDEIgpcc-all* has the highest global correlation
622 ($r=0.682$) as well the highest correlation out of all ensembles in 5 of the top 8 maize production
623 countries. *AgMERRA-all* correlations are significant for 16 of the 19 countries, with significantly
624 higher skill than any other ensemble in the Philippines and Ethiopia. These results highlight the
625 potential for broader GGCM application for national and global maize production decision
626 making. *Ensemble-all* had an increase in global correlation (+0.094) compared to *Ensemble+*.

627

628 5.3.2 National wheat production anomalies

629 Wheat simulations generally have lower correlations than were seen for maize, indicating a
630 comparatively smaller agroclimatic signal or common biases in the structure or configuration of
631 wheat models (Fig, 6b). Correlation levels are once again highly related to the various nations,

632 with simulation ensembles of the top two producing countries, China and India, not significantly
633 correlated to their FAO production statistics (with the exception of *WFDEIgpcc-all* in China) even
634 as positive correlations dominate most of the other countries. This may be due, in part, to the large
635 area devoted to irrigated wheat in these countries, which lowers the response to drought hazards
636 and therefore overall climate sensitivity. Diseases are also not included in GGCM simulations but
637 can play a major role in wheat breadbaskets (Savary et al., 2019). Intensified systems in the United
638 States, France, Germany, the United Kingdom, and the Ukraine also have mostly insignificant
639 correlations even as weather data are likely of good quality, indicating a large role of irrigation
640 and perhaps a muddled signal in grid cells where both spring wheat and winter wheat is present.
641 GGCM Phase 1 simulations only ran one wheat season per grid cell, which can miss second
642 season production anomalies and underrepresent vernalization requirement effects. Subsequent
643 GGCM phases have conducted separate simulations for winter and spring wheat in order to better
644 capture production in regions where both systems are prominent (Franke et al., 2020, 2019;
645 Jägermeyr et al., 2020). Simulations capture high correlations indicating a strong climate response
646 for Australian wheat, which is dominated by rainfed winter wheat demonstrating a strong
647 precipitation response (Fig 3e). Simulated wheat in European countries showed little response to
648 growing season temperature, precipitation, and solar radiation in Fig. 3, however; which is
649 consistent with relatively low national-level correlations to FAOstat.

650

651 The bias-adjusted CFDs largely outperform the raw reanalyses and GRASP for most wheat
652 countries. *WFDEIgpcc-all* increases correlations for China and Germany in comparison to
653 *WFDEIgpcc+* likely due to high correlations in at least one of the added GGCMs, although a
654 decrease in correlation is seen for Poland and the United States. *AgMERRA-all* similarly improves

655 upon *AgMERRA+* correlations in Canada and the Ukraine. Overall, *AgMERRA-all* and
656 *WFDEIgpcc-all* both improved correlations in half of the countries. Although the *Ensemble+* and
657 *Ensemble-all* have higher wheat correlations in Pakistan, there is otherwise little difference
658 between *AgMERRA+*, *WFDEIgpcc+*, *Ensemble+*, and *Ensemble-all* which have significant
659 correlations in 13, 13, 12, and 12 of the top 20 wheat producing countries, respectively. Global
660 wheat anomalies are fairly consistently and significantly simulated by all ensembles, with
661 *WFDEIgpcc-all* producing the highest global correlation ($r=0.603$) aided by relatively strong
662 performance in China, Germany, and the United Kingdom.

663

664 *5.3.3 National rice production anomalies*

665 Rice simulations have the lowest FAO correlations of the four simulated crops (Fig. 6c).
666 Significant correlations are highest for Japan, which Ray et al. (2015) also noted as being strongly
667 driven by temperature variation, as is also evident in Figure 3. Significant correlations are also
668 broadly seen for Bangladesh, Vietnam, Philippines, United States, North Korea, Egypt and
669 Madagascar, but there are no clear patterns identifying geographic regions with cohesively high or
670 low correlations.

671

672 Rice is largely irrigated across top producing countries, with a smaller weather signal in
673 interannual yield fluctuations. Yet, insignificant rice correlations in many countries could be an
674 indication of incomplete FAO data, inaccurate CFDs, poor GGCM simulation, or a realistically
675 small agroclimatic response that may reflect regional farming systems or limiting factors beyond
676 direct climate conditions. Ray et al. (2015) and identified that interannual rice variability was
677 driven less by climate than were maize, wheat and soybean, which may also reflect the substantial

678 influence of geopolitical events and socioeconomic limitations in major rice producing countries
679 over the 1981-2010 period that would influence FAO production data. Iizumi et al., (2018)
680 similarly found weak attribution of climate change impacts in long-term rice trends. The
681 simulation ensemble demonstrated only weak response to growing-season mean temperature and
682 precipitation over the major rice baskets of East, South, and Southeast Asia (Fig. 3g-i). These are
683 among the only major breadbaskets in the Tropics, which tend to have lower interannual variability
684 of mean temperature and total precipitation than mid-latitude breadbaskets. These rice areas also
685 have more uncertain climate information (Fig. 2) and have a higher proportion of total production
686 coming from heterogeneous farming systems that are difficult to configure within GCMs.
687 GCM configurations may also simulate upland (non-flooded) rice systems in areas where rice is
688 grown in paddies (flooded), and only contain a maximum of one rainfed and one irrigated season
689 even as it is common for some rice-growing areas to have two or three seasons in a given year
690 (e.g., the *aus*, *aman*, and *boro* seasons in Bangladesh). Major flood events that can destroy large
691 rice harvests in the Mekong, Indus, Ganges, and other river basins, as well as the influence of large
692 hurricanes and typhoons, are also not resolved by crop models despite being substantial climate
693 disasters (Lesk et al., 2016).

694

695 There is no substantial benefit in bias adjustment for national rice applications, with no clear
696 differences in correlation levels between the raw reanalyses (CFSR, ERAI), GRASP, and the other
697 CFDs adjusted to match monthly observations. The bias-adjustments within *AgCFSR+*,
698 *AgMERRA+*, *WFDEIcru+*, and *WFDEIgpcc+* (but not *Princeton+*) lower correlations in Japan,
699 although high correlations are seen when all GCMs are included in *AgMERRA-all* and
700 *WFDEIgpcc-all*. The top two rice production countries, China and India, are only significantly

701 simulated in the *AgMERRA+* and *AgMERRA-all* ensembles. Compared to *AgMERRA+* and
702 *WFDEIgpcc+*, respectively, the additional GCMs increase correlations for many countries in
703 *AgMERRA-all* (notably Japan, Vietnam, the Philippines, the United States, and China but not India
704 or Madagascar) and *WFDEIgpcc-all* (notably Japan and the United States but not Egypt or the
705 Philippines). While the signal was mixed for *WFDEIgpcc-all*, 14 out of 20 *AgMERRA-all* country
706 correlations were higher than *AgMERRA+*, including 10 that increased by ≥ 0.1 compared to only
707 2 where correlations dropped by ≥ 0.1 . *Ensemble+* and *Ensemble-all* capture many of the stronger
708 correlations from rice simulations, but both also see reductions in some country correlations (e.g.,
709 *Ensemble+* in Vietnam and *Ensemble-all* in North Korea). The highest global correlation is found
710 in *AgMERRA-all* ($r=0.347$), aided by higher correlations in China, Vietnam and Thailand, with
711 other ensembles unable to capture significant correlations with global rice production.

712

713 *5.3.4 National soybean production anomalies*

714 Soybean simulations have higher correlations overall than rice, with higher producing countries
715 tending to have higher correlations and the lower producing countries tending to not be
716 significantly correlated (Fig. 6d). The highest correlations are associated with the United States,
717 Brazil, Argentina, Paraguay, South Africa and Indonesia, while Ukraine, Bolivia Russia are top-
718 10 high-producing countries where relatively few ensembles capture a significant interannual
719 signal.

720

721 The bias-adjusted CFDs have a larger number of significant correlations than the raw reanalysis
722 (*CFSR+* and *ERA+*) and *GRASP+* ensembles, which signifies a benefit to bias adjustment
723 particularly in the highest producing countries. *AgMERRA-all* and *WFDEIgpcc-all* have slightly

724 reduced correlations compared to *AgMERRA+* (lower in 13 out of 19 countries) and *WFDEIgpcc+*
725 (lower in 11 out of 19 countries) as the inclusion of additional GGCMs reduces the captured
726 climate signal particularly for China, India, Paraguay, and Uruguay. *Ensemble+* and *AgMERRA+*
727 produce a significant correlation in each of the top 7 countries, and *Ensemble-all* loses significant
728 signals in China, India, and Uruguay.

729

730 Global correlations are generally positive but weaker than those seen for maize and wheat.
731 Significant correlations are captured by *AgCFSR+*, *AgMERRA+*, *ERA1+* (top correlation at
732 $r=0.416$), *GRASP+* and *Ensemble+*. The low global correlation compared to the top countries'
733 high correlation is surprising, possibly indicating inter-breadbasket anti-correlations that act to
734 offset a larger global signal. *Ensemble-all* global correlation is 0.313 lower than for *Ensemble+*,
735 indicating a substantial loss of signal within the additional CFD/GGCM combinations.

736

737 **6. Discussion**

738 The analyses above demonstrate many ways that the selection of CFD strongly influences regional
739 crop production simulations. Although it is not practical to analyze every combination of specific
740 nations, cropping systems and crop model ensemble sets in this manuscript, the examples,
741 approaches, supplementary material, and open data access of the GGCM Phase 1 dataset provide
742 a roadmap for further analysis. The extent of CFD influence depends on differences between CFD
743 characteristics, crop models' biophysical responses to these differences, attributes of national and
744 global production for each crop species, and the use of multi-GGCM and multi-CFD ensembles.
745 Key findings are discussed below, with additional uncertainties in climate and crop model
746 information described in Supplemental Material S8.

747

748 *Regional differences in climate information and responses.* CFDs differ most strongly in regions
749 where in situ observations are sparse, inconsistent or incomplete (Figure 2), and can have nearly
750 global differences in distributional or extreme characteristics (Figures 1 and S.3-5). Regional
751 cropping system models have different fundamental responses to climate variability in ways that
752 can make them more sensitive to CFD differences (Figure 3). The selection of CFDs is therefore
753 most influential in regions where agricultural systems respond strongly to a climatic variable with
754 strong observational uncertainties. Further analysis, and indeed GGCM development, is required
755 to investigate cropping system response to variables beyond the growing season mean climate
756 indices, as considerable variance is likely from sub-seasonal patterns, acute heat, drought and flood
757 extremes, severe storms, and connected impacts from sequential or compound hazards (Ben-Ari
758 et al., 2018; Grotjahn, 2020; Li et al., 2019; Raymond et al., 2020; Schewe et al., 2019).
759 Fundamental climate responses also help prioritize observational network and agricultural
760 resilience investments even as interannual response is not always a clear predictor of long-term
761 climate change risks (Ruane et al., 2016).

762

763 *GGCM/CFD abilities to capture observed interannual variance:* The selection of CFDs is only
764 able to influence a fraction of interannual production variations. GGCM results (e.g., Figure 6)
765 are broadly consistent with the findings of Ray et al. (2015), who found that climate variation
766 explains only about one third of global observed yield variability, with upwards of 60% of variation
767 explained in some highly intensified breadbaskets and lower fundamental climate responses for
768 rice than maize, wheat or soybean. Lower correlations may also be related to non-representative
769 model configurations, including incorrect planted area fractions which can change from year to

770 year, growing season dates and cultivars (Jägermeyr and Frieler, 2018), the presence of multiple
771 growing seasons (e.g., short and long rains), multi-cropping, sub-grid scale heterogeneity in
772 climate and crop systems, soil types and textures, and alternative irrigation management strategies
773 (Hoffmann et al., 2016; Lopez et al., 2017). High correlations between FAO data and simulation
774 outputs are therefore indicative of strong climate forcing in national production anomalies and an
775 ability of the GGCMs (driven by CFDs) to capture those anomalies. In some cases the GGCM
776 climate-driven ensemble captures a higher proportion of the FAO production variability that was
777 evident in Ray et al., (2015), including for maize in Mexico, wheat in Iran, rice in Madagascar,
778 and soybean in Paraguay.

779

780 *Some crop species and countries are not as clearly limited by climate.* GGCM simulations
781 generally produced the highest FAO correlations for maize, followed by wheat, soy, and rice. For
782 each species there were countries with high and low correlations. High correlation countries tend
783 to feature some combination of large-scale intensified farming, mid-latitude climates, less
784 uncertainty in climate and farm configuration information, and consolidated production regions.
785 Lower correlation countries tend to have a relatively large proportion of heterogeneous and small-
786 holder farming systems, are situated in tropical regions with lower interannual variability, and lie
787 in areas with more uncertain climate anomalies and field data (Figure 2). We would expect these
788 process-based crop models to be more climate-limited than observations, as factors not included
789 in the models reduce the coherence with the seasonal climate signal (e.g., sociopolitical events,
790 labor or machine shortages, river floods, pests and diseases) (Ray et al., 2015; van Ittersum et al.,
791 2016). Many of these non-climatic impact factors are more widespread in developing countries
792 than in intensified agricultural regions of developed countries (van Bussel et al., 2015).

793

794 *Overall performance of CFDs.* This study further confirms the utility of climatic forcing datasets
795 for agricultural applications (Toreti et al., 2019) and elucidates ways that CFD differences can
796 affect crop model simulations (Figs. 4, 5, 6). Normalized anomalies between CFDs are larger for
797 precipitation than for temperature, and differences between CFDs are larger for distributional
798 characteristics and extreme events than for mean response (Figs. 1,2). The use of bias adjustment
799 (AgCFSR vs. CFSR and WFDEI vs. ERAI) improved crop model simulation in many regions and
800 countries, while the sequence of sub-monthly weather patterns (AgCFSR vs. AgMERRA) had a
801 smaller impact (Figs. 5,6). The selection of large-scale precipitation datasets (WFDEIgpcc vs.
802 GPCCcru) did not have a substantial overall effect on performance. These conclusions for complex
803 biophysical models are consistent with those found by (Parkes et al., 2019) for empirical models.
804 We advise those planning crop model applications for a given country and crop species to examine
805 Figure 6 to ensure that their CFD is associated with high correlations against FAO production
806 variability.

807

808 *Effects of model ensemble statistics.* GGCM uses the 1980-2010 period to benchmark the
809 performance of global gridded crop models (Müller et al., 2017), and this study has further
810 demonstrated the utility of this period to elucidate the strengths and weaknesses of various
811 GGCM/CFD ensembles through comparison against FAO anomalies. Comparing across minimal
812 multi-GGCM ensembles for each *CFD+*, a major finding is that the difference between countries
813 > difference between CFDs > difference between *CFD+* and *CFD-all* ensembles (the effect of
814 more GGCMs on top of the multi-GGCM ensemble) > difference between *Ensemble-all* and
815 *Ensemble+* ensembles (the effect of adding further GGCM/CFD combinations on top of the multi-

816 GGCM/multi-CFD *Ensemble+*). Differences between countries emphasizes the importance of
817 improving data collection for climate, soils, cultivars, and field management which can vary
818 widely by nation. Differences between CFDs can be substantial in some parts of the world (Figure
819 2), but our overall finding is that the bias-adjusted datasets (e.g., AgMERRA and WFDEIgpcc)
820 capture the bulk of the signal captured in the GGCMI ensemble. In light of previous AgMIP studies
821 on the benefits of small multi-crop model ensembles (Wallach et al., 2016), we recommend that
822 resources are likely better focused on additional configuration information and the inclusion of a
823 multi-GGCM ensemble (3-7 models) before conducting a multi-CFD ensemble. Here the maize,
824 wheat, and soybean *CFD+* ensembles had 7 GGCMs (5 for rice), and the further addition of
825 GGCMs was not consistently helpful to the extent that would justify investment for larger GGCM
826 ensembles (Figures 2, 3, S.2). Given that *Ensemble+* has 56 GGCM/CFD combinations for maize,
827 the lack of clear benefit from the full 91 GGCM/CFD combination *Ensemble-all* underscores that
828 the full GGCMI ensemble is not typically needed for practical application.

829

830 A number of agricultural system applications stand to benefit from more accurate climate
831 observation, modeling, bias-adjustment, and methods to merge these into CFDs, including
832 seasonal forecasting (Schauberger et al., 2017), disaster preparedness (Cottrell et al., 2019;
833 Jägermeyr et al., 2020; Lunt et al., 2016), climate change resilience (Franke et al., 2019; Hasegawa
834 et al., 2018; Rosenzweig et al., 2014; Ruane et al., 2018; Zhao et al., 2017), and the development
835 of more robust and sustainable markets and farming systems (Snyder et al., 2019; Valdivia et al.,
836 2015). A new generation of CFDs are now possible given updated reanalyses (Gelaro et al., 2017;
837 Hersbach et al., 2019) and observational products (Funk et al., 2015; Lange, 2019), which will
838 enable further crop modeling applications (e.g., Iizumi et al., 2017; Lange, 2019b, 2019c). CFD

839 characteristics also propagate into climate scenarios that use the CFD as a bias-adjustment target,
840 so CFD deviations presented in Figure 1 and Figures S.3-5 may help explain differences in regional
841 projections among studies. We include a similar comparison of the W5E5 dataset to the GGCM
842 CFD ensemble in Supplemental Figure S.5 given its application in forthcoming ISIMIP Phase 3
843 simulations. Improvements in CFDs, and the selection of a CFD particularly suited for a given
844 regional farming system, are therefore important elements of a crop model application even as they
845 are a limited element of broader application improvement efforts. Further opportunities for model
846 development and application motivated by this study are described in Supplementary Material S9.

847

848

849 **Acknowledgements**

850 Funding: Support for this study was provided by NASA NNX16AK38G (INCA), and the NASA
851 Earth Sciences Directorate/GISS Climate Impacts Group funding. DD acknowledges the HPC
852 Cluster supported by the Research and Specialist Computing Support service at the University of
853 East Anglia for running PEGASUS. TAMP and AA acknowledge support from European Union
854 FP7 Grant LUC4C (Grant 603542), and the Helmholtz Association in its ATMO programme.
855 Helpful initial analyses and insights were provided by Monica Morales, Nicholas Hudson, and
856 Kevin Schwarzwald at the Columbia University Center for Climate Systems Research/NASA
857 GISS.

858

859 References

- 860 Antle, J.M., Valdivia, R.O., Boote, K.J., Janssen, S., Jones, J.W., Porter, C.H., Rosenzweig, C., Ruane, A.C.,
861 Thorburn, P.J., 2015. AgMIP's Transdisciplinary Agricultural Systems Approach to Regional Integrated
862 Assessment of Climate Impacts, Vulnerability, and Adaptation. *Handb. Clim. Chang. Agroecosystems* 27–44.
863 https://doi.org/10.1142/9781783265640_0002
- 864 Bassu, S., Brisson, N., Durand, J.L., Boote, K., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M.,
865 Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler,
866 S., Grassini, P., Hatfield, J., Hoek, S., Izaurrealde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim,
867 S.H., Kumar, N.S., Makowski, D., Müller, C., Nendel, C., Priesack, E., Pravia, M.V., Sau, F., Shcherbak, I.,
868 Tao, F., Teixeira, E., Timlin, D., Waha, K., 2014. How do various maize crop models vary in their responses
869 to climate change factors? *Glob. Chang. Biol.* 20, 2301–2320. <https://doi.org/10.1111/gcb.12520>
- 870 Ben-Ari, T., Boé, J., Ciais, P., Lecerf, R., Van Der Velde, M., Makowski, D., 2018. Causes and implications of the
871 unforeseen 2016 extreme yield loss in the breadbasket of France. *Nat. Commun.* 9, 1–10.
872 <https://doi.org/10.1038/s41467-018-04087-x>
- 873 Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S.,
874 Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y.,
875 Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo,
876 R.M., Wang, X.L., Woodruff, S.D., Worley, S.J., 2011. The Twentieth Century Reanalysis Project. *Q. J. R.*
877 *Meteorol. Soc.* 137, 1–28. <https://doi.org/10.1002/qj.776>
- 878 Cottrell, R.S., Nash, K.L., Halpern, B.S., Remenyi, T.A., Corney, S.P., Fleming, A., Fulton, E.A., Hornborg, S.,
879 Johne, A., Watson, R.A., Blanchard, J.L., 2019. Food production shocks across land and sea. *Nat. Sustain.* 2,
880 130–137. <https://doi.org/10.1038/s41893-018-0210-1>
- 881 Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A.,
882 Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C.,
883 Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. V., Isaksen, L.,
884 Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K.,
885 Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis:
886 Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 553–597.
887 <https://doi.org/10.1002/qj.828>
- 888 Deryng, D., Conway, D., Ramankutty, N., Price, J., Warren, R., 2014. Global crop yield response to extreme heat
889 stress under multiple climate change futures. *Environ. Res. Lett.* 9, 034011. [https://doi.org/10.1088/1748-](https://doi.org/10.1088/1748-9326/9/3/034011)
890 [9326/9/3/034011](https://doi.org/10.1088/1748-9326/9/3/034011)
- 891 Dirmeyer, P.A., Gao, X., Zhao, M., Guo, Z., Oki, T., Hanasaki, N., 2006. GSWP-2: Multimodel analysis and
892 implications for our perception of the land surface. *Bull. Am. Meteorol. Soc.* 87, 1381–1397.
893 <https://doi.org/10.1175/BAMS-87-10-1381>
- 894 Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K.J., Büchner, M., Foster, I., Glotter, M., Heinke,
895 J., Iizumi, T., Izaurrealde, R.C., Mueller, N.D., Ray, D.K., Rosenzweig, C., Ruane, A.C., Sheffield, J., 2015.
896 The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0),
897 Geoscientific Model Development. Copernicus GmbH. <https://doi.org/10.5194/gmd-8-261-2015>
- 898 European Centre for Medium-Range Weather Forecasts, 2009. ERA-Interim Project. Research Data Archive at the
899 National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder,
900 CO. <https://doi.org/10.5065/D6CR5RD9>
- 901 Famien, A.M., Janicot, S., Ochou, A.D., Vrac, M., Defrance, D., Sultan, B., Noël, T., 2018. A bias-corrected CMIP5
902 dataset for Africa using the CDF-t method – a contribution to agricultural impact studies. *Earth Syst. Dyn.* 9,
903 313–338. <https://doi.org/10.5194/esd-9-313-2018>
- 904 Fleisher, D.H., Condori, B., Quiroz, R., Alva, A., Asseng, S., Barreda, C., Bindi, M., Boote, K.J., Ferrise, R.,
905 Franke, A.C., Govindakrishnan, P.M., Harahagazwe, D., Hoogenboom, G., Naresh Kumar, S., Merante, P.,
906 Nendel, C., Olesen, J.E., Parker, P.S., Raes, D., Raymundo, R., Ruane, A.C., Stockle, C., Supit, I.,
907 Vanuytrecht, E., Wolf, J., Woli, P., 2017. A potato model intercomparison across varying climates and
908 productivity levels. *Glob. Chang. Biol.* 23, 1258–1281. <https://doi.org/10.1111/gcb.13411>
- 909 Folberth, C., Elliott, J., Müller, C., Balkovič, J., Chryssanthacopoulos, J., Izaurrealde, R.C., Jones, C.D., Khabarov,
910 N., Liu, W., Reddy, A., Schmid, E., Skalský, R., Yang, H., Arneth, A., Ciais, P., Deryng, D., Lawrence, P.J.,
911 Olin, S., Pugh, T.A.M., Ruane, A.C., Wang, X., 2019. Parameterization-induced uncertainties and impacts of
912 crop management harmonization in a global gridded crop model ensemble, *PLoS ONE*.
913 <https://doi.org/10.1371/journal.pone.0221862>
- 914 Franke, J., Müller, C., Elliott, J., Ruane, A., Jagermeyr, J., Balkovic, J., Ciais, P., Dury, M., Falloon, P., Folberth, C.,

915 Francois, L., Hank, T., Hoffmann, M., Jacquemin, I., Jones, C., Khabarov, N., Koch, M., Li, M., Liu, W.,
916 Olin, S., Phillips, M., Pugh, T.A., Reddy, A., Wang, X., Williams, K., Zabel, F., Moyer, E., 2019. The
917 GGCM Phase II experiment: global gridded crop model simulations under uniform changes in CO₂,
918 temperature, water, and nitrogen levels (protocol version 1.0). *Geosci. Model Dev. Discuss.* 1–30.
919 <https://doi.org/10.5194/gmd-2019-237>

920 Franke, J., Müller, C., Elliott, J., Ruane, A.C., Jagermeyr, J., Snyder, A., Dury, M., Falloon, P., Folberth, C.,
921 Francois, L., Hank, T., Izaurralde, R.C., Jacquemin, I., Jones, C., Li, M., Liu, W., Olin, S., Phillips, M.M.,
922 Pugh, T.A.M., Reddy, A.D., Williams, K., Wang, Z., Zabel, F., Moyer, E.J., 2020. The GGCM phase II
923 emulators : global gridded crop model responses to changes in CO₂, temperature, water, and nitrogen. *Geosci.*
924 *Model Dev. Discuss.* <https://doi.org/10.5194/gmd-2019-365>

925 Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger,
926 C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C.,
927 Hansen, M., Gong, P., Abdel Aziz, S., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S.,
928 Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun,
929 A., Ortner, S., Rajak, D.R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A.,
930 Vancutsem, C., Vintrou, E., Wenbin, W., van der Velde, M., Dunwoody, A., Kraxner, F., Obersteiner, M.,
931 2015. Mapping global cropland and field size. *Glob. Chang. Biol.* 21, 1980–1992.
932 <https://doi.org/10.1111/gcb.12838>

933 Fuchs, T., 2009. GPCC annual report for year 2008: Development of the GPCC data base and analysis products.
934 DWD Rep.

935 Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L.,
936 Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations - A new
937 environmental record for monitoring extremes. *Sci. Data* 2, 1–21. <https://doi.org/10.1038/sdata.2015.66>

938 Galmarini, S., Cannon, A.J., Ceglar, A., Christensen, O.B., de Noblet-Ducoudré, N., Dentener, F., Doblas-Reyes,
939 F.J., Dosio, A., Gutierrez, J.M., Iturbide, M., Jury, M., Lange, S., Loukos, H., Maiorano, A., Maraun, D.,
940 McGinnis, S., Nikulin, G., Riccio, A., Sanchez, E., Solazzo, E., Toreti, A., Vrac, M., Zampieri, M., 2019.
941 Adjusting climate model bias for agricultural impact assessment: How to cut the mustard. *Clim. Serv.* 13, 65–
942 69. <https://doi.org/10.1016/J.CLISER.2019.01.004>

943 Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A.,
944 Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty,
945 A., da Silva, A.M., Gu, W., Kim, G.K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G.,
946 Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The modern-era
947 retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* 30, 5419–5454.
948 <https://doi.org/10.1175/JCLI-D-16-0758.1>

949 Grotjahn, R., 2020. Weather extremes that impact various agricultural commodities, in: Castillo, F., Wehner, M.,
950 Stone, D. (Eds.), *Extreme Events and Climate Change: A Multidisciplinary 28 Approach*. John Wiley & Sons,
951 Inc. for American Geophysical Union, p. (in press).

952 Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic
953 observations - the CRU TS3.10 Dataset. *Int. J. Climatol.* 34, 623–642. <https://doi.org/10.1002/joc.3711>

954 Hasegawa, T., Fujimori, S., Havlík, P., Valin, H., Bodirsky, B.L., Doelman, J.C., Fellmann, T., Kyle, P., Koopman,
955 J.F.L., Lotze-Campen, H., Mason-D’Croz, D., Ochi, Y., Pérez Domínguez, I., Stehfest, E., Sulser, T.B.,
956 Tabeau, A., Takahashi, K., Takakura, J., van Meijl, H., van Zeist, W.J., Wiebe, K., Witzke, P., 2018. Risk of
957 increased food insecurity under stringent global climate change mitigation policy. *Nat. Clim. Chang.*
958 <https://doi.org/10.1038/s41558-018-0230-x>

959 Hatfield, J.L., Prueger, J.H., 2015. Temperature extremes: Effect on plant growth and development. *Weather Clim.*
960 *Extrem.* 10, 4–10. <https://doi.org/10.1016/j.wace.2015.08.001>

961 Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Sabater, J.M., Nicolas, J., Radu, R., Schepers, D., Simmons, A.,
962 Soci, C., Dee, D., 2019. Global reanalysis: goodbye ERA-Interim, hello ERA5. *ECMWF Newsl.* 17–24.
963 <https://doi.org/10.21957/vf291hehd7>

964 Hoffmann, H., Zhao, G., Asseng, S., Bindi, M., Biernath, C., Constantin, J., Coucheney, E., Dechow, R., Doro, L.,
965 Eckersten, H., Gaiser, T., Grosz, B., Heinlein, F., Kassie, B.T., Kersebaum, K.-C., Klein, C., Kuhnert, M.,
966 Lewan, E., Moriondo, M., Nendel, C., Priesack, E., Raynal, H., Roggero, P.P., Rötter, R.P., Siebert, S.,
967 Specka, X., Tao, F., Teixeira, E., Trombi, G., Wallach, D., Weihermüller, L., Yeluripati, J., Ewert, F., 2016.
968 Impact of Spatial Soil and Climate Input Data Aggregation on Regional Yield Simulations. *PLoS One* 11.
969 <https://doi.org/10.1371/journal.pone.0151782>

970 Iizumi, T., Okada, M., Yokozawza, M., 2014. A meteorological forcing data set for global crop modeling:

971 Development, evaluation, and intercomparison. *J. Geophys. Res. Atmos.* 119, 363–384.
972 <https://doi.org/10.1002/2013JD020130>

973 Iizumi, T., Shiogama, H., Imada, Y., Hanasaki, N., Takikawa, H., Nishimori, M., 2018. Crop production losses
974 associated with anthropogenic climate change for 1981–2010 compared with preindustrial levels. *Int. J.*
975 *Climatol.* 38, 5405–5417. <https://doi.org/10.1002/joc.5818>

976 Iizumi, T., Takikawa, H., Hirabayashi, Y., Hanasaki, N., Nishimori, M., 2017. Contributions of different bias-
977 correction methods and reference meteorological forcing data sets to uncertainty in projected temperature and
978 precipitation extremes. *J. Geophys. Res.* 122, 7800–7819. <https://doi.org/10.1002/2017JD026613>

979 Jägermeyr, J., Frieler, K., 2018. Spatial variations in crop growing seasons pivotal to reproduce global fluctuations
980 in maize and wheat yields. *Sci. Adv.* 4. <https://doi.org/10.1126/sciadv.aat4517>

981 Jägermeyr, J., Robock, A., Elliott, J., Müller, C., Xia, L., Khabarov, N., Folberth, C., Schmid, E., Liu, W., Zabel, F.,
982 Rabin, S., Puma, M.J., Heslin, A.C., Franke, J., Foster, I., Asseng, S., Bardeen, C.G., Toon, O.B., Rosenzweig,
983 C., 2020. A regional nuclear conflict would compromise global food security. *PNAS* (in press).

984 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G.,
985 Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C.,
986 Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis
987 project. *Bull. Am. Meteorol. Soc.* 77, 437–471. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)

988 Lange, Stefan, 2019a. Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0).
989 *Geosci. Model Dev.* 12, 3055–3070. <https://doi.org/10.5194/gmd-12-3055-2019>

990 Lange, S., 2019. Earth2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI)
991 [WWW Document]. GFZ Data Serv. <https://doi.org/10.5880/pik.2016.004>

992 Lange, Stefan, 2019b. WFDE5 over land merged with ERA5 over the ocean (W5E5). V. 1.0. [WWW Document].
993 GFZ Data Serv. <https://doi.org/10.5880/pik.2019.023>

994 Lesk, C., Rowhani, P., Ramankutty, N., 2016. Influence of extreme weather disasters on global crop production.
995 *Nature* 529, 84–87. <https://doi.org/10.1038/nature16467>

996 Li, Y., Guan, K., Schnitkey, G.D., DeLucia, E., Peng, B., 2019. Excessive rainfall leads to maize yield loss of a
997 comparable magnitude to extreme drought in the United States. *Glob. Chang. Biol.* 25, gcb.14628.
998 <https://doi.org/10.1111/gcb.14628>

999 Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate trends and global crop production since 1980. *Science*
1000 (80-). <https://doi.org/10.1126/science.1204531>

1001 Lopez, J.R., Winter, J.M., Elliott, J., Ruane, A.C., Porter, C., Hoogenboom, G., 2017. Integrating growth stage
1002 deficit irrigation into a process based crop model. *Agric. For. Meteorol.* 243, 84–92.
1003 <https://doi.org/10.1016/J.AGRFORMET.2017.05.001>

1004 Lunt, T., Jones, A.W., Mulhern, W.S., Lezaks, D.P.M., Jahn, M.M., 2016. Vulnerabilities to agricultural production
1005 shocks: An extreme, plausible scenario for assessment of risk for the insurance sector. *Clim. Risk Manag.* 13,
1006 1–9. <https://doi.org/10.1016/j.crm.2016.05.001>

1007 Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W.J.W., Rötter, R.P.R.P., Boote, K.J.K.J., Ruane, A.C.A.C.,
1008 Thorburn, P.J.P.J.P.J., Cammarano, D., Hatfield, J.L.J.L.J.L., Rosenzweig, C., Aggarwal, P.K.P.K.P.K.,
1009 Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J.A.J.A.J., Doltra, J., Gayler, S.,
1010 Goldberg, R., Grant, R.F.R.F.R.F., Heng, L., Hooker, J., Hunt, L.A.L.A.L.A., Ingwersen, J., Izaurralde,
1011 R.C.R.C., Kersebaum, K.C.K.C., Müller, C., Kumar, S.N.S.N., Nendel, C., O’leary, G., Olesen, J.E.J.E.,
1012 Osborne, T.M.T.M., Palosuo, T., Priesack, E., Ripoché, D., Semenov, M.A.M.A.M.A., Shcherbak, I., Steduto,
1013 P., Stöckle, C.O.C.O.C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White,
1014 J.W.J.W., Wolf, J., 2015. Multimodel ensembles of wheat growth: many models are better than one. *Glob.*
1015 *Chang. Biol.* 21. <https://doi.org/10.1111/gcb.12768>

1016 Mbow, C., Rosenzweig, C., Barioni, L.G., Benton, T.G., Herrero, M., Krishnapillai, M., Liwenga, E., Pradhan, P.,
1017 Rivera-Ferre, M.G., Sapkota, T., Tubiello, F.N., Xu, Y., Contreras, E.M., Pereira, J. P., Blanchard, J., Fanzo,
1018 J., Frank, S., Kriewald, S., Lanigan, G., López, D., Mason-D’Croz, D., Neofotis, P., Pant, L., Rodrigues, R.,
1019 Ruane, A.C., Waha, K., 2019. Food Security, in: Shukla, P.R., Skea, J., Buendia, E.C., Masson-Delmotte, V.,
1020 Pörtner, H.-O., Roberts, D.C., Zhai, P., Slade, R., Connors, S., Diemen, R. van, Ferrat, M., Haughey, E., Luz,
1021 S., Neogi, S., Pathak, M., Petzold, J., Pereira, J. Portugal, Vyas, P., Huntley, E., Kissick, K., Belkacemi, M.,
1022 Malley, J. (Eds.), *Climate Change and Land: An IPCC Special Report on Climate Change, Desertification,*
1023 *Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial*
1024 *Ecosystems.*

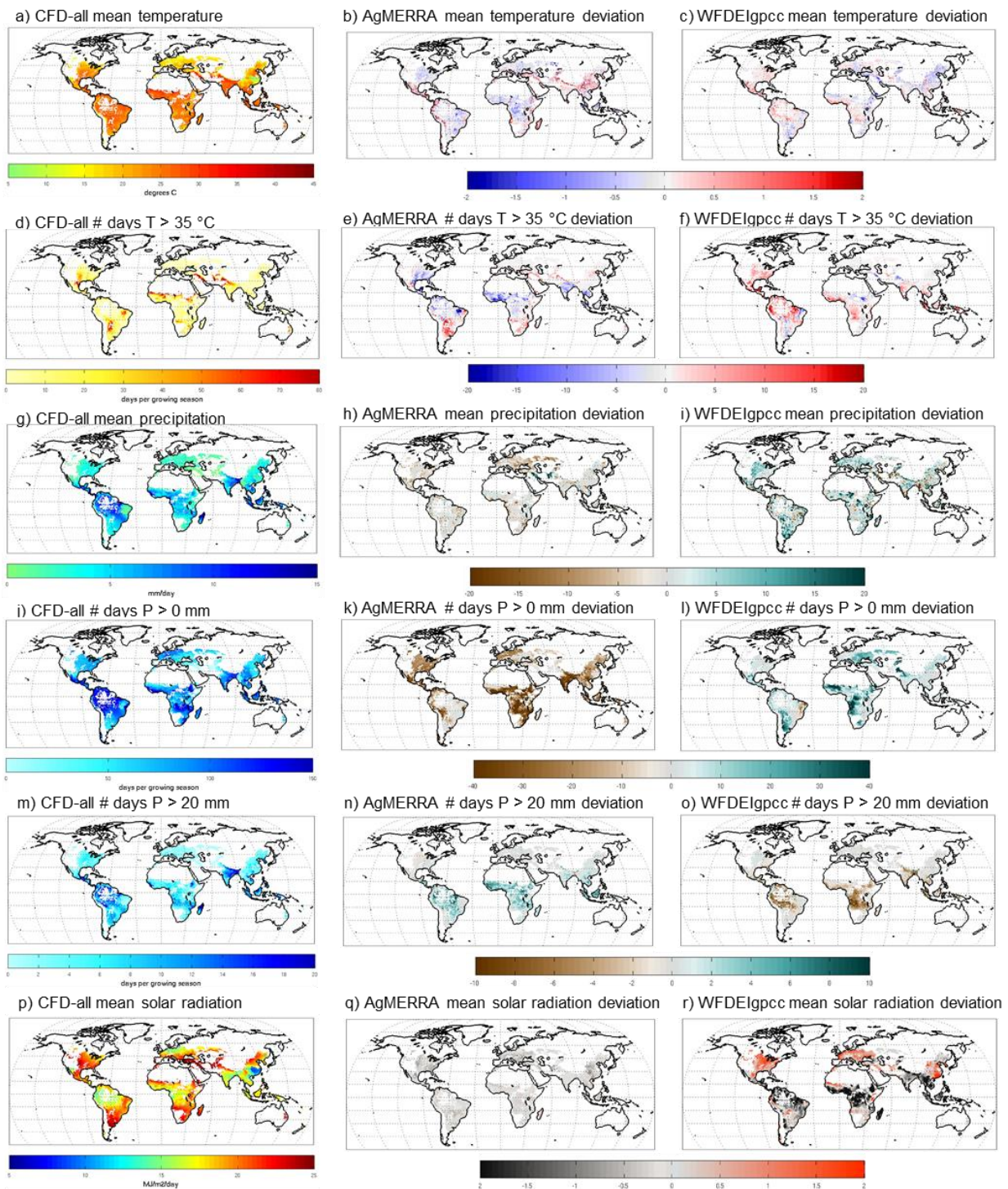
1025
1026 McDermid, S.P., Ruane, A.C., Rosenzweig, C., Hudson, N.I., Morales, M.D., Agalawatte, P., Ahmad, S., Ahuja,

1027 L.R., Amien, I., Anapalli, S.S., Anothai, J., Asseng, S., Biggs, J., Bert, F., Bertuzzi, P., Bhatia, V.S., Bindi,
1028 M., Broad, I., Cammarano, D., Carretero, R., Chattha, A.A., Chung, U., Debats, S., Deligios, P., De Sanctis,
1029 G., Dhliwayo, T., Dumont, B., Estes, L., Ewert, F., Ferrise, R., Gaiser, T., Garcia, G., Gbegbelegbe, S.,
1030 Geethalakshmi, V., Gerardeaux, E., Goldberg, R., Grant, B., Guevara, E., Hickman, J., Hoffmann, H., Huang,
1031 H., Hussain, J., Justino, F.B., Karunaratne, A.S., Koehler, A.-K., Kouakou, P.K., Kumar, S.N., Lakshmanan,
1032 A., Lieffering, M., Lin, X., Luo, Q., Magrin, G., Mancini, M., Marin, F.R., Marta, A.D., Masutomi, Y.,
1033 Mavromatis, T., McLean, G., Meira, S., Mohanty, M., Moriondo, M., Nasim, W., Negm, L., Orlando, F.,
1034 Orlandini, S., Ozturk, I., Soares Pinto, H.M., Podesta, G., Qi, Z., Ramarohetra, J., ur Rahman, M.H., Raynal,
1035 H., Rodriguez, G., Rötter, R., Sharda, V., Shuo, L., Smith, W., Snow, V., Soltani, A., Srinivas, K., Sultan, B.,
1036 Swain, D.K., Tao, F., Tesfaye, K., Travasso, M.I., Trombi, G., Topaj, A., Vanuytrecht, E., Viscarra, F.E.,
1037 Aftab Wajid, S., Wang, E., Wang, H., Wang, J., Wijekoon, E., Byun-Woo, L., Xiaoguang, Y., Young, B.H.,
1038 Yun, J.I., Zhao, Z., Zubair, L., 2015. The AgMIP Coordinated Climate-Crop Modeling Project (C3MP):
1039 Methods and Protocols, in: Rosenzweig, C., Hillel, D. (Eds.), Handbook of Climate Change and
1040 Agroecosystems: The Agricultural Model Intercomparison and Improvement Project (AgMIP) Integrated Crop
1041 and Economic Assessments, Part 1. pp. 191–220. https://doi.org/10.1142/9781783265640_0008
1042 Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter,
1043 M., Hoek, S., Iizumi, T., Izaurralde, R.C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh,
1044 T.A.M., Ray, D.K., Reddy, A., Rosenzweig, C., Ruane, A.C., Sakurai, G., Schmid, E., Skalsky, R., Song,
1045 C.X., Wang, X., De Wit, A., Yang, H., 2017. Global gridded crop model evaluation: Benchmarking, skills,
1046 deficiencies and implications. *Geosci. Model Dev.* 10, 1403–1422. <https://doi.org/10.5194/gmd-10-1403-2017>
1047 Müller, C., Elliott, J., Kelly, D., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Hoek, S., Izaurralde,
1048 R.C., Jones, C.D., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T.A.M., Reddy, A., Rosenzweig, C.,
1049 Ruane, A.C., Sakurai, G., Schmid, E., Skalsky, R., Wang, X., de Wit, A., Yang, H., 2019. The Global Gridded
1050 Crop Model Intercomparison phase 1 simulation dataset. *Sci. Data* 6, 50. [https://doi.org/10.1038/s41597-019-](https://doi.org/10.1038/s41597-019-0023-8)
1051 0023-8
1052 Nelson, G.C., van der Mensbrugge, D., Ahammad, H., Blanc, E., Calvin, K., Hasegawa, T., Havlik, P., Heyhoe, E.,
1053 Kyle, P., Lotze-Campen, H., von Lampe, M., Mason d’Croz, D., van Meijl, H., Müller, C., Reilly, J.,
1054 Robertson, R., Sands, R.D., Schmitz, C., Tabeau, A., Takahashi, K., Valin, H., Willenbockel, D., 2014.
1055 Agriculture and climate change in global scenarios: why don’t the models agree. *Agric. Econ.* 45, 85–101.
1056 <https://doi.org/10.1111/agec.12091>
1057 Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N.,
1058 Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., Taira, R.,
1059 2007. The JRA-25 reanalysis. *J. Meteorol. Soc. Japan* 85, 369–432. <https://doi.org/10.2151/jmsj.85.369>
1060 Parkes, B., Higginbottom, T.P., Hufkens, K., Ceballos, F., Kramer, B., Foster, T., 2019. Weather dataset choice
1061 introduces uncertainty to estimates of crop yield responses to climate variability and change. *Environ. Res.*
1062 *Lett* 14, 124089. <https://doi.org/10.1088/1748-9326/ab5ebb>
1063 Porter, J.R., Xie, L., Challinor, A.J., Cochrane, K., Howden, S.M., Iqbal, M.M., Lobell, D.B., Travasso, M.I., 2014.
1064 Food security and food production systems, in: Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J.,
1065 Mastrandrea, M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel,
1066 E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., White, L.L. (Eds.), *Climate Change 2014: Impacts,*
1067 *Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the*
1068 *Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* [https://doi.org/10.1111/j.1728-](https://doi.org/10.1111/j.1728-4457.2009.00312.x)
1069 4457.2009.00312.x
1070 Porwollik, V., Müller, C., Elliott, J., Chryssanthacopoulos, J., Iizumi, T., Ray, D.K., Ruane, A.C., Arneth, A.,
1071 Balkovič, J., Ciais, P., Deryng, D., Folberth, C., Izaurralde, R.C., Jones, C.D., Khabarov, N., Lawrence, P.J.,
1072 Liu, W., Pugh, T.A.M., Reddy, A., Sakurai, G., Schmid, E., Wang, X., de Wit, A., Wu, X., 2017. Spatial and
1073 temporal uncertainty of crop yield aggregations. *Eur. J. Agron.* 88, 10–21.
1074 <https://doi.org/10.1016/j.eja.2016.08.006>
1075 Ray, D.K., Gerber, J.S., Macdonald, G.K., West, P.C., 2015. Climate variation explains a third of global crop yield
1076 variability. *Nat. Commun.* 6, 1–9. <https://doi.org/10.1038/ncomms6989>
1077 Raymond, C., Horton, R.M., Zscheischler, J., Martius, O., AghaKouchak, A., Balch, J., Bowen, S.G., Camargo, S.J.,
1078 Hess, J., Kornhuber, K., Oppenheimer, M., Ruane, A.C., Wahl, T., White, K., 2020. Understanding and
1079 Managing Connected Extreme Events. *Nat. Clim. Chang.*
1080 Reichle, R.H., Koster, R.D., De Lannoy, G.J.M., Forman, B.A., Liu, Q., Mahanama, S.P.P., Toure, A., 2011.
1081 Assessment and enhancement of MERRA land surface hydrology estimates. *J. Clim.* 24, 6322–6338.
1082 <https://doi.org/10.1175/JCLI-D-10-05033.1>

1083 Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D.,
1084 Takacs, L., Kim, G.K., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster,
1085 R.D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C.R., Reichle, R., Robertson, F.R.,
1086 Ruddick, A.G., Sienkiewicz, M., Woollen, J., 2011. MERRA: NASA's modern-era retrospective analysis for
1087 research and applications. *J. Clim.* 24, 3624–3648. <https://doi.org/10.1175/JCLI-D-11-00015.1>
1088 Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C.A.C., Müller, C., Arnoeth, A., Boote, K.J.K.J., Folberth, C.,
1089 Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A.M.T.A.M., Schmid, E., Stehfest, E., Yang,
1090 H., Jones, J.W.J.W., 2014. Assessing agricultural risks of climate change in the 21st century in a global
1091 gridded crop model intercomparison. *Proc. Natl. Acad. Sci. U. S. A.* 111, 3268–3273.
1092 <https://doi.org/10.1073/pnas.1222463110>
1093 Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C.,
1094 Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The
1095 Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agric.*
1096 *For. Meteorol.* 170, 166–182. <https://doi.org/10.1016/j.agrformet.2012.09.011>
1097 Ruane, Alex C., Antle, J., Elliott, J., Folberth, C., Hoogenboom, G., Croz, D.M., Müller, C., Porter, C., Phillips,
1098 M.M., Raymundo, R.M., Sands, R., Valdivia, R.O., White, J.W., Wiebe, K., Rosenzweig, C., 2018.
1099 Biophysical and economic implications for agriculture of +1.5 ° and +2.0 ° C global warming using
1100 AgMIP Coordinated Global and Regional Assessments. *Clim. Res.* 76, 17–39.
1101 Ruane, Alex C., Goldberg, R., Chryssanthacopoulos, J., 2015. Climate forcing datasets for agricultural modeling:
1102 Merged products for gap-filling and historical climate series estimation. *Agric. For. Meteorol.* 200, 233–248.
1103 <https://doi.org/10.1016/j.agrformet.2014.09.016>
1104 Ruane, A.C., Hudson, N.I., Asseng, S., Camarrano, D., Ewert, F., Martre, P., Boote, K.J., Thorburn, P.J., Aggarwal,
1105 P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S.,
1106 Goldberg, R., Grant, R.F., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C.,
1107 Kumar, S.N., Müller, C., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E.,
1108 Ripoche, D., Rötter, R.P., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck,
1109 T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Wolf, J., 2016. Multi-wheat-model
1110 ensemble responses to interannual climate variability. *Environ. Model. Softw.* 81, 86–101.
1111 <https://doi.org/10.1016/j.envsoft.2016.03.008>
1112 Ruane, Alex C., Phillips, M.M., Rosenzweig, C., 2018. Climate shifts within major agricultural seasons for +1.5 and
1113 +2.0 °C worlds: HAPPI projections and AgMIP modeling scenarios. *Agric. For. Meteorol.* 259, 329–344.
1114 <https://doi.org/10.1016/j.agrformet.2018.05.013>
1115 Ruane, A.C., Rosenzweig, C., Asseng, S., Boote, K.J., Elliott, J., Ewert, F., Jones, J.W., Martre, P., McDermid, S.P.,
1116 Müller, C., Snyder, A., Thorburn, P.J., 2017. An AgMIP framework for improved agricultural representation
1117 in integrated assessment models. *Environ. Res. Lett.* 12. <https://doi.org/10.1088/1748-9326/aa8da6>
1118 Ruane, Alexander C., Winter, J.M., McDermid, S.P., Hudson, N.I., 2015. AgMIP Climate Data and Scenarios for
1119 Integrated Assessment, in: Rosenzweig, C., Hillel, D. (Eds.), *Handbook of Climate Change and*
1120 *Agroecosystems: The Agricultural Model Intercomparison and Improvement Project (AgMIP) Integrated Crop*
1121 *and Economic Assessments, Part 1.* pp. 45–78. https://doi.org/10.1142/9781783265640_0003
1122 Rudolf, B., Becker, A., Schneider, U., Meyer-Christoffer, A., Ziese, M., 2010. GPCP Status Report December 2010
1123 (On the most recent gridded global data set issued in fall 2010 by the Global Precipitation Climatology Centre
1124 (GPCP)). DWD/GPCP.
1125 Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, Jie, Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D.,
1126 Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, Jun, Hou, Y.-T., Chuang, H.-Y., Juang, H.-M.H., Sela,
1127 J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang,
1128 R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm,
1129 J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y.,
1130 Cucurull, L., Reynolds, R.W., Rutledge, G., Goldberg, M., 2010. NCEP Climate Forecast System Reanalysis
1131 (CFRS) 6-hourly Products, January 1979 to December 2010. <https://doi.org/10.5065/D69K487J>
1132 Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N., Nelson, A., 2019. The global burden of
1133 pathogens and pests on major food crops. *Nat. Ecol. Evol.* 3, 430–439. [https://doi.org/10.1038/s41559-018-](https://doi.org/10.1038/s41559-018-0793-y)
1134 [0793-y](https://doi.org/10.1038/s41559-018-0793-y)
1135 Schauburger, B., Gornott, C., Wechsung, F., 2017. Global evaluation of a semiempirical model for yield anomalies
1136 and application to within-season yield forecasting. *Glob. Chang. Biol.* 23, 4750–4764.
1137 <https://doi.org/10.1111/gcb.13738>
1138 Schewe, J., Gosling, S.N., Reyser, C., Zhao, F., Ciais, P., Elliott, J., Francois, L., Huber, V., Lotze, H.K.,

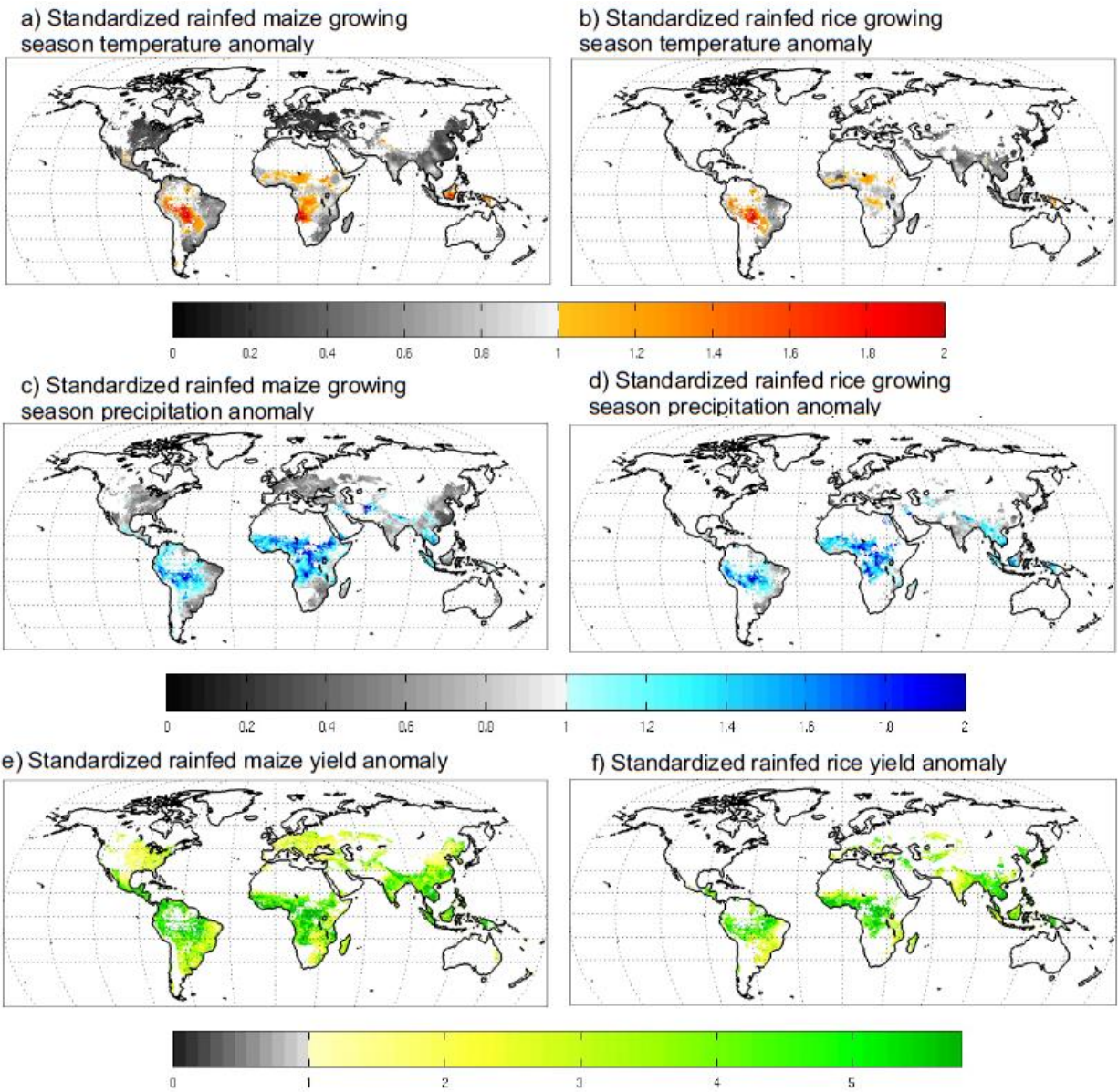
1139 Seneviratne, S.I., van Vliet, M.T.H., Vautard, R., Wada, Y., Breuer, L., Büchner, M., Carozza, D.A., Chang,
 1140 J., Coll, M., Deryng, D., de Wit, A., Eddy, T.D., Folberth, C., Frieler, K., Friend, A.D., Gerten, D.,
 1141 Gudmundsson, L., Hanasaki, N., Ito, A., Khabarov, N., Kim, H., Lawrence, P., Morfopoulos, C., Müller, C.,
 1142 Müller Schmied, H., Orth, R., Ostberg, S., Pokhrel, Y., Pugh, T.A.M., Sakurai, G., Satoh, Y., Schmid, E.,
 1143 Stacke, T., Steenbeek, J., Steinkamp, J., Tang, Q., Tian, H., Tittensor, D.P., Volkholz, J., Wang, X.,
 1144 Warszawski, L., 2019. State-of-the-art global models underestimate impacts from climate extremes. *Nat.*
 1145 *Commun.* 10, 1–14. <https://doi.org/10.1038/s41467-019-08745-6>
 1146 Schollaert Uz, S., Ruane, A.C., Duncan, B.N., Compton, &, Tucker, J., Huffman, G.J., Mladenova, I.E., Osmanoglu,
 1147 B., Holmes, T.R.H., McNally, A., Peters-Lidard, C., Bolten, J.D., Das, N., Rodell, M., McCartney, S.,
 1148 Anderson, M.C., Doorn, B., 2019. Earth observations and integrative models in support of food and water
 1149 security. *Remote Sens. Earth Syst. Sci.* <https://doi.org/10.1007/s41976-019-0008-6>
 1150 Sheffield, J., Goteti, G., Wood, E.F., 2006. Development of a 50-Year High-Resolution Global Dataset of
 1151 Meteorological Forcings for Land Surface Modeling. *J. Clim.* 19, 3088–3111.
 1152 <https://doi.org/10.1175/JCLI3790.1>
 1153 Snyder, A., Calvin, K. V., Phillips, M., Ruane, A.C., 2019. A crop yield change emulator for use in GCAM and
 1154 similar models: Persephone v1.0. *Geosci. Model Dev.* 12, 1319–1350. [https://doi.org/10.5194/gmd-12-1319-](https://doi.org/10.5194/gmd-12-1319-2019)
 1155 2019
 1156 Stackhouse, Jr, P.W., Gupta, S.K., Cox, S.J., Zhang, T., Mikovitz, J.C., Hinkelman, L.M., 2011. The
 1157 NASA/GEWEX surface radiation budget release 3.0: 24.5-year dataset. *Gewex news* 21, 10–12.
 1158 Thoning, K.W., Tans, P.P., Komhyr, W.D., 1989. Atmospheric carbon dioxide at Mauna Loa Observatory: 2.
 1159 Analysis of the NOAA GMCC data, 1974-1985. *J. Geophys. Res. Atmos.* 94, 8549–8565.
 1160 <https://doi.org/10.1029/JD094iD06p08549>
 1161 Toreti, A., Maiorano, A., De Sanctis, G., Webber, H., Ruane, A.C., Fumagalli, D., Ceglar, A., Niemeyer, S.,
 1162 Zampieri, M., 2019. Using reanalysis in crop monitoring and forecasting systems. *Agric. Syst.* 168, 144–153.
 1163 <https://doi.org/10.1016/j.agsy.2018.07.001>
 1164 UNISDR, 2015. Sendai Framework for Disaster Risk Reduction 2015-2030.
 1165 United Nations, 2015a. Adoption of the Paris Agreement, in: Conference of the Parties on Its Twenty-First Session.
 1166 United Nations, 2015b. Transforming our world: The 2030 Agenda for Sustainable Development [WWW
 1167 Document]. United Nations. URL http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E
 1168 Uppala, S.M., Kållberg, P.W., Simmons, A.J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J.K.,
 1169 Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E.,
 1170 Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier,
 1171 F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L.,
 1172 Janssen, P.A.E.M., Jenne, R., McNally, A.P., Mahfouf, J.F., Morcrette, J.J., Rayner, N.A., Saunders, R.W.,
 1173 Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J., 2005. The ERA-40
 1174 re-analysis. *Q. J. R. Meteorol. Soc.* <https://doi.org/10.1256/qj.04.176>
 1175 Valdivia, R.O., Antle, J.M., Rosenzweig, C., Ruane, A.C., Vervoort, J., Ashfaq, M., Hathie, I., Tui, S.H.-K., Mulwa,
 1176 R., Nhemachena, C., Ponnusamy, P., Rasnayaka, H., Singh, H., 2015. Representative Agricultural Pathways
 1177 and Scenarios for Regional Integrated Assessment of Climate Change Impacts, Vulnerability, and Adaptation,
 1178 in: Rosenzweig, C., Hillel, D. (Eds.), *Handbook of Climate Change and Agroecosystems: The Agricultural
 1179 Model Intercomparison and Improvement Project (AgMIP) Integrated Crop and Economic Assessments, Part
 1180 1.* pp. 101–145. https://doi.org/10.1142/9781783265640_0005
 1181 van Bussel, L.G.J., Grassini, P., van Wart, J., Wolf, J., Claessens, L., Yang, H., Boogaard, H., de Groot, H., Saito,
 1182 K., Cassman, K.G., van Ittersum, M.K., 2015. From field to atlas: Upscaling of location-specific yield gap
 1183 estimates. *F. Crop. Res.* 177, 98–108. <https://doi.org/10.1016/j.fcr.2015.03.005>
 1184 van Ittersum, M.K., van Bussel, L.G.J., Wolf, J., Grassini, P., van Wart, J., Guilpart, N., Claessens, L., De Groot, H.,
 1185 Wiebe, K., Mason-D’Croz, D., Yang, H., Boogaard, H., Van Oort, P.A.J., Van Loon, M.P., Saito, K., Adimo,
 1186 O., Adjei-Nsiah, S., Agali, A., Bala, A., Chikowo, R., Kaizzi, K., Kouressy, M., Makoi, J.H.J.R., Ouattara, K.,
 1187 Tesfaye, K., Cassman, K.G., 2016. Can sub-Saharan Africa feed itself? *Proc. Natl. Acad. Sci. U. S. A.* 113,
 1188 14964–14969. <https://doi.org/10.1073/pnas.1610359113>
 1189 Wallach, D., Mearns, L.O., Rivington, M., Antle, J.M., Ruane, A.C., 2015. Uncertainty in Agricultural Impact
 1190 Assessment, in: Rosenzweig, C., Hillel, D. (Eds.), *Handbook of Climate Change and Agroecosystems: The
 1191 Agricultural Model Intercomparison and Improvement Project (AgMIP) Integrated Crop and Economic
 1192 Assessments, Part 1.* pp. 223–259. https://doi.org/10.1142/9781783265640_0009
 1193 Wallach, D., Mearns, L.O., Ruane, A.C., Rötter, R.P., Asseng, S., 2016. Lessons from climate modeling on the
 1194 design and use of ensembles for crop modeling. *Clim. Change* 139, 551–564. <https://doi.org/10.1007/s10584->

1195 016-1803-1
1196 Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J., Viterbo, P., 2018. The WFDEI Meteorological
1197 Forcing Data. <https://doi.org/10.5065/486N-8109>
1198 Weedon, G.P., Gomes, S., Viterbo, P., Shuttleworth, W.J., Blyth, E., Österle, H., Adam, J.C., Bellouin, N., Boucher,
1199 O., Best, M., 2011. Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional
1200 Reference Crop Evaporation over Land during the Twentieth Century. *J. Hydrometeorol.* 12, 823–848.
1201 <https://doi.org/10.1175/2011JHM1369.1>
1202 Willmott, C.J., Matsuura, K., 1995. Smart Interpolation of Annually Averaged Air Temperature in the United States.
1203 *J. Appl. Meteorol.* 34, 2577–2586. [https://doi.org/10.1175/1520-0450\(1995\)034<2577:SIOAAA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1995)034<2577:SIOAAA>2.0.CO;2)
1204 You, L., Wood-Sichra, U., Fritz, S., Guo, Z., See, L., Koo, J., 2014. Spatial Production Allocation Model (SPAM)
1205 2005 v2.0. MapSPAM [WWW Document]. URL <http://mapspam.info>
1206 Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D.B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.-
1207 L., Elliott, J., Ewert, F., Janssens, I.A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., Peng, S., Peñuelas, J.,
1208 Ruane, A.C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., Asseng, S., 2017. Temperature
1209 increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci.* 201701762.
1210 <https://doi.org/10.1073/pnas.1701762114>
1211

Figures:

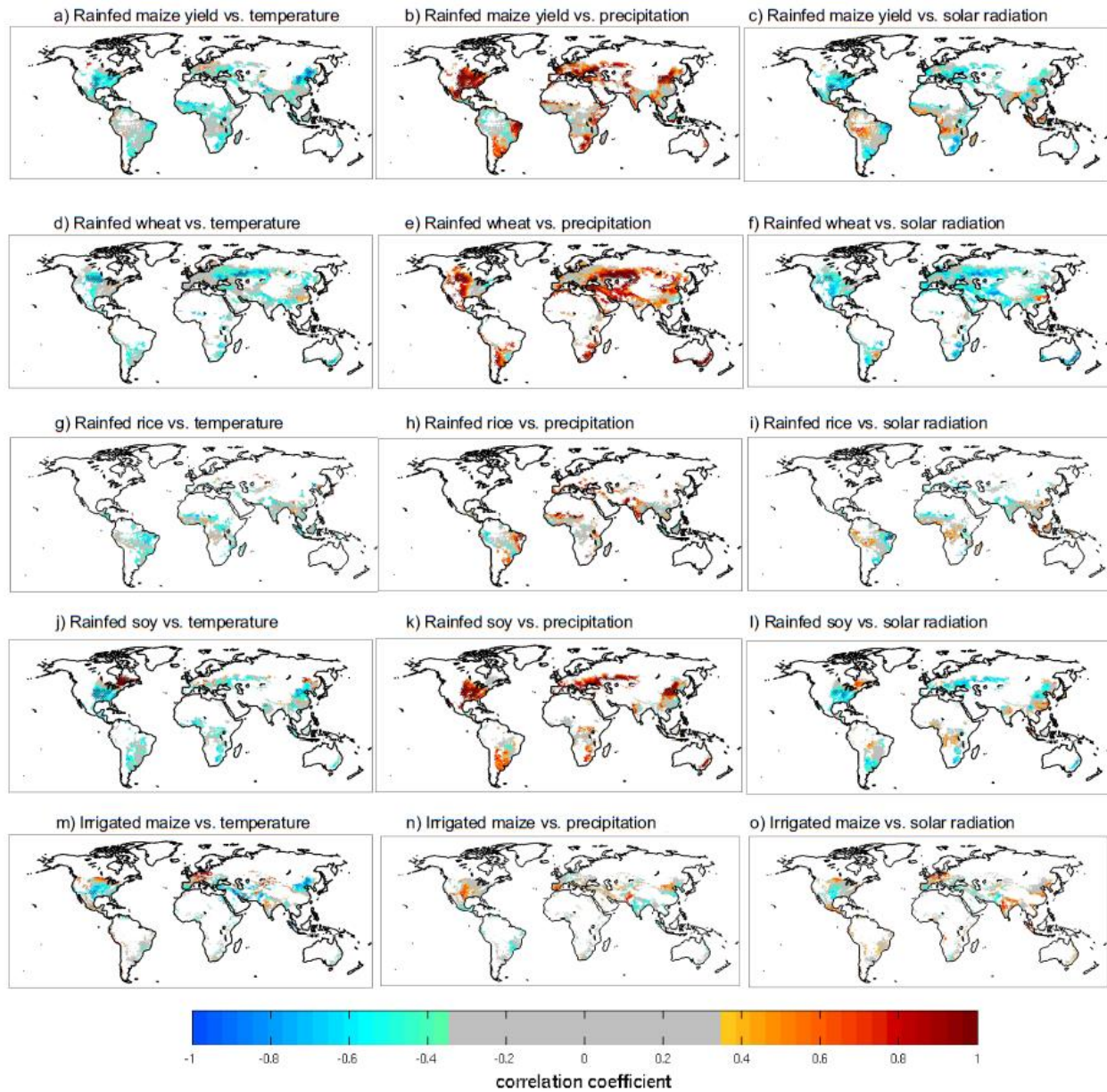
1213
1214
1215
1216
1217
1218
1219
1220

Figure 1: Rainfed maize growing season (1981-2010) mean and extreme climatologies over maize-growing areas (>10 ha) for (left) mean of all climatic forcing dataset (*CFD-all*) ensemble, (center) deviation of AgMERRA compared to *CFD-all*, and (right) deviation of WFDEIgpcc compared to *CFD-all*. From top to bottom, rows are deviations in growing season mean temperature (°C), mean precipitation (%), mean solar radiation (MJ m⁻² day⁻¹), mean number of days where $T_{max} > 35$ °C, mean number of days where $P > 0$ mm/day, mean number of days where $P > 20$ mm/day. AgMERRA and WFDEIgpcc are the most commonly simulated CFDs from GGCM Phase 1; corresponding deviation maps for other CFDs are shown in Figures S.3-S.5.



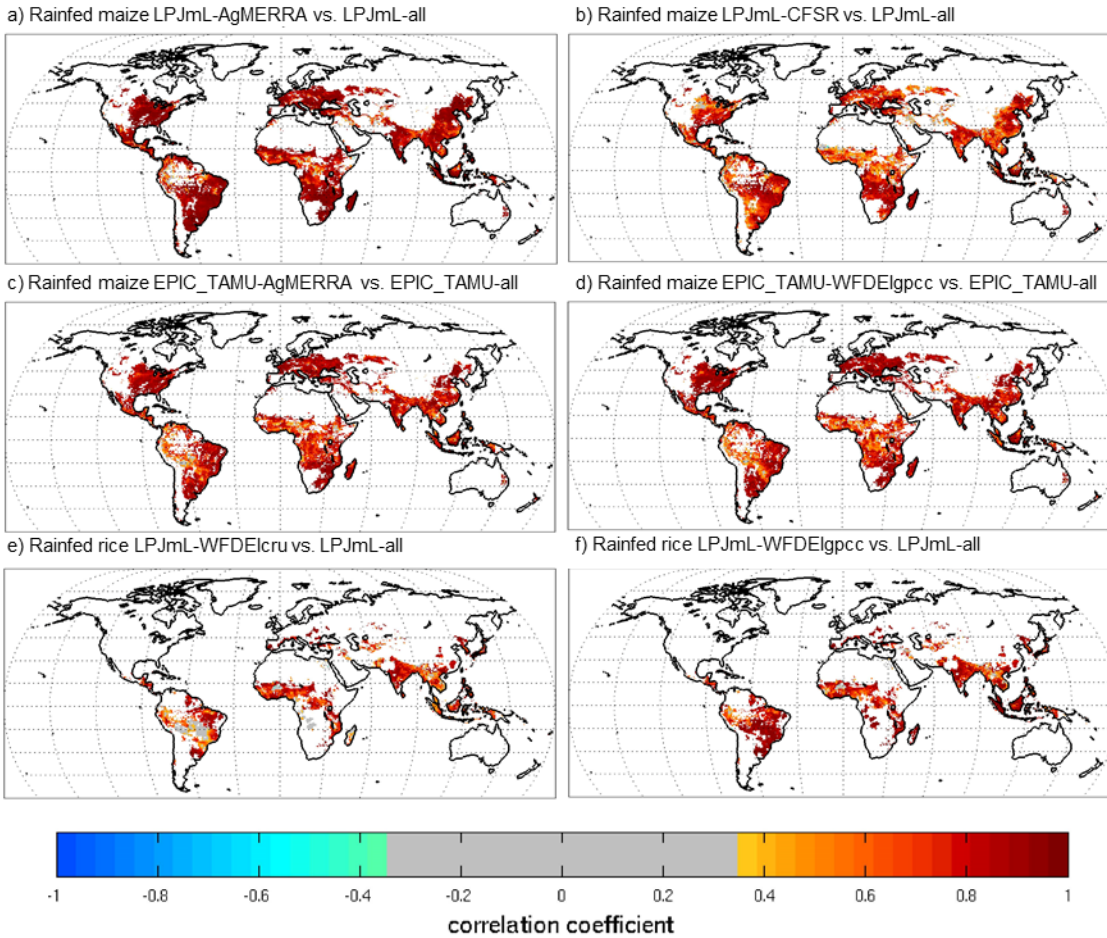
1222
 1223 **Figure 2:** Standardized anomalies (unitless) for 1981-2010 rainfed maize growing season (left) and rainfed rice
 1224 growing season (right) mean (a,b) temperature and (c,d) precipitation (across all climatic forcing datasets) as well as
 1225 for (d,e) yield (across all GGCMIxCFD combinations). Standardized anomalies are the ratio of (i) the standard
 1226 deviation of yearly ensemble member anomalies (compared to the ensemble mean) to (ii) the standard deviation of
 1227 the ensemble mean time series itself. Only regions with >10 ha of harvested area (You et al., 2014) are presented;
 1228 note that many areas with high standardized anomalies have low planted areas (Figure S1).

1229
 1230
 1231

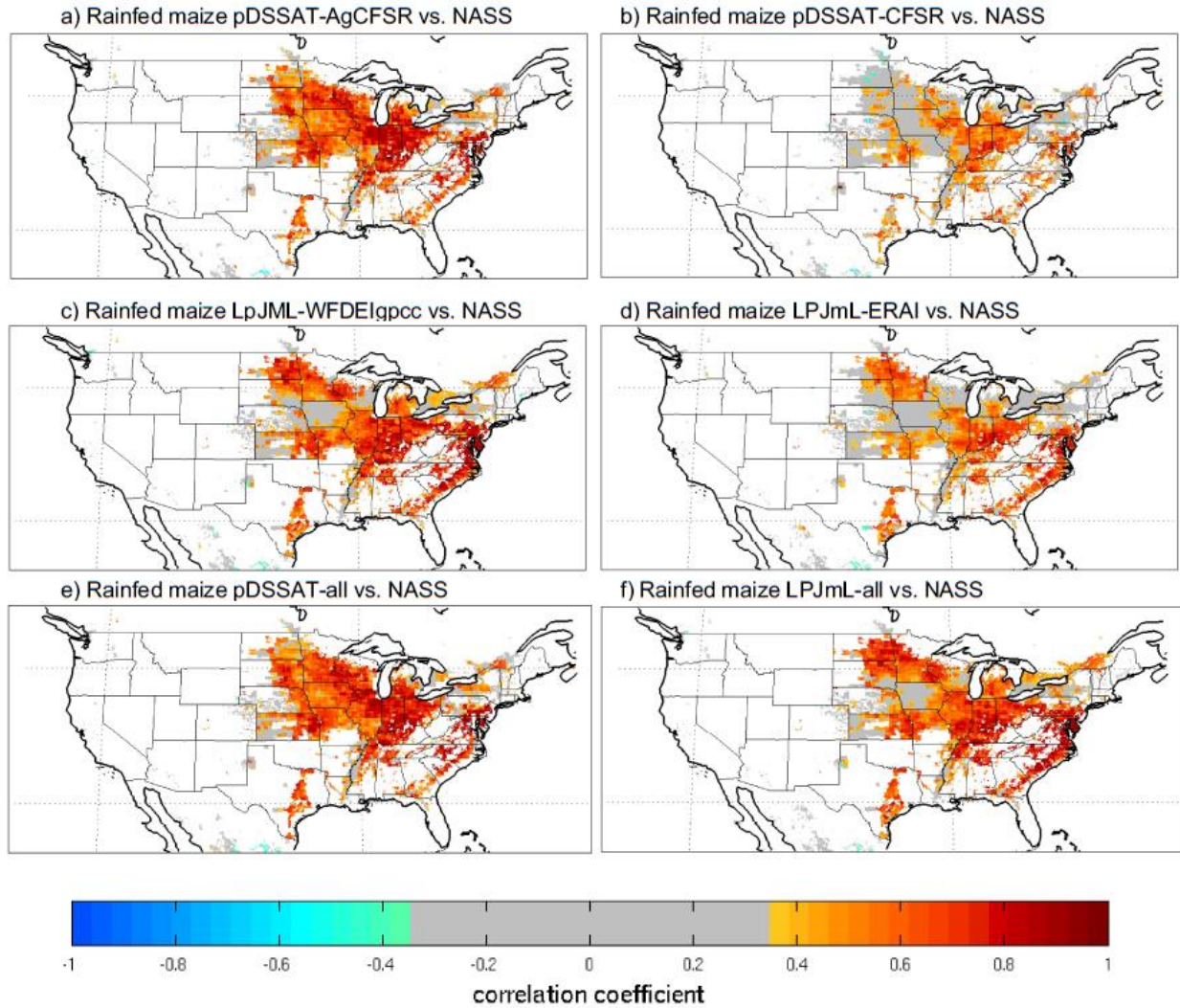


1232
 1233
 1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241

Figure 3: Regional and crop system-dependent GGCM responses to climatic forcing dataset (CFD) growing season anomalies (1981-2010), expressed as Pearson's correlations between the medians of all GGCMxCFD ensemble members (*Ensemble-all*) compared to the ensemble of all CFDs (*CFD-all*). Rows are rainfed maize, wheat, rice, and soybean, as well as irrigated maize; columns are growing season mean correlations for temperature (left), precipitation (center), and solar radiation (right). Only correlations that are significant at $p < 0.05$ level are colored and hatched areas indicate that 2/3 of GGCMxCFD combinations agree on a significant correlation in the same direction. Only regions with >10 ha of harvested area (You et al., 2014) are presented.

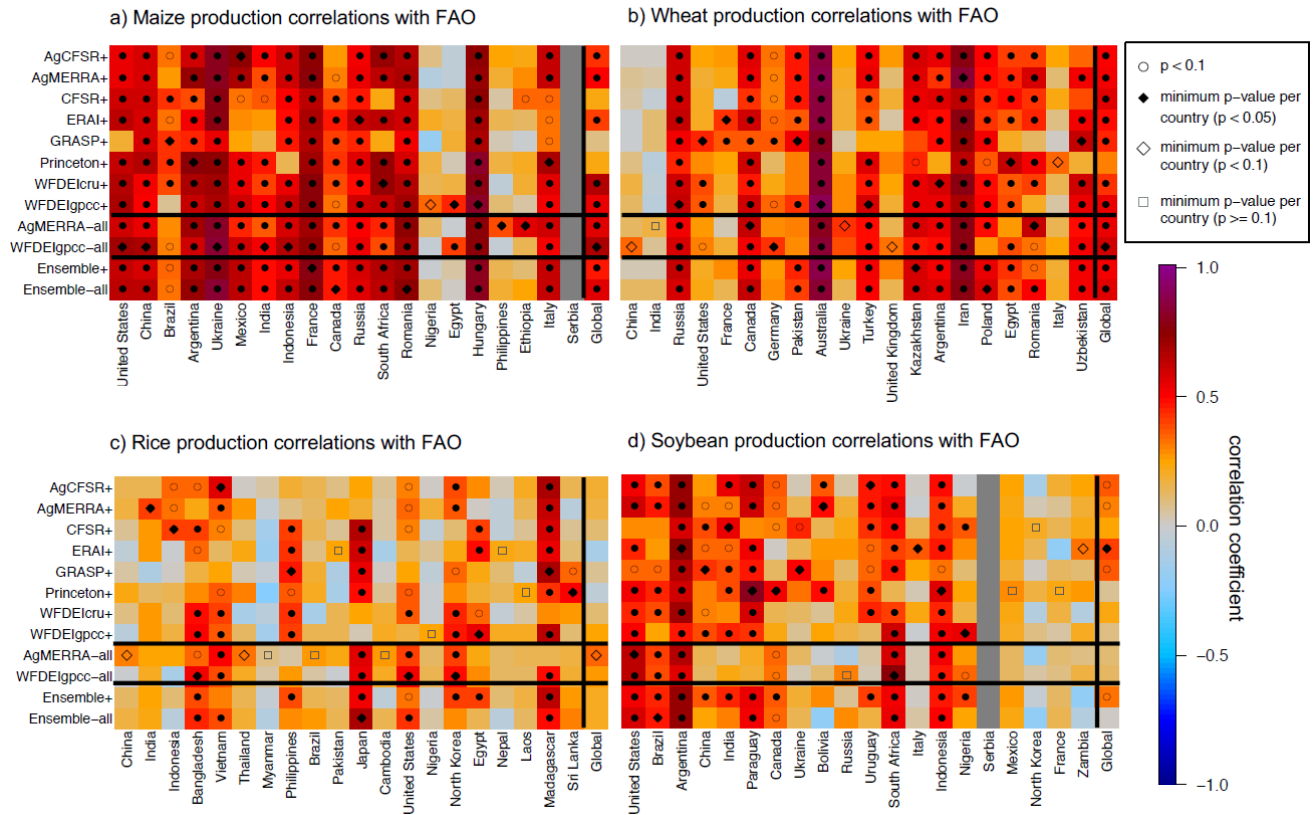


1242
 1243 **Figure 4:** 1981-2010 correlations (r) between the LPJmL GGCM simulation driven by an individual climatic
 1244 forcing dataset (CFD) and the ensemble of the simulations using all CFDs (*LPJmL-all*). a) *LPJmL-AgMERRA*
 1245 simulations vs. *LPJmL-all* for rainfed maize; b) *LPJmL-CFSR* simulations vs. *LPJmL-all* for rainfed maize; c)
 1246 *EPIC_TAMU-AgMERRA* simulations vs. *EPIC_TAMU-all* for rainfed maize; d) *EPIC_TAMU-WFDEIgpcc*
 1247 simulations vs. *EPIC_TAMU-all* for rainfed maize; e) *LPJmL-WFDEIcru* simulations vs. *LPJmL-all* for rainfed rice;
 1248 f) *LPJmL-WFDEIgpcc* simulations vs. *LPJmL-all* for rainfed rice. Only correlations that are significant at $p < 0.05$
 1249 level are colored.



1250
 1251 **Figure 5:** 1981-2010 correlations (r) between NASS county-level yield observations and GGCM yield simulations
 1252 driven by various CFDs. a) *pDSSAT-AgCFSR*, b) *pDSSAT-CFSR*, c) *LPJmL-WFDEIqpc*, d) *LPJmL-ERA1*, and
 1253 median across all GGCM simulations using each CFD e) *pDSSAT-all*, and f) *LPJmL-all*. Only regions with >10 ha
 1254 of planted area (You et al., 2014) are presented, and only correlations that are significant at $p < 0.05$ level are colored
 1255 rather than gray.

1256
 1257



1258
 1259
 1260
 1261
 1262
 1263
 1264
 1265
 1266
 1267
 1268
 1269

Figure 6: Comparison of simulated GGCM-CFD subset production anomalies with 1981-2010 FAO national production anomalies for the top 20 producer countries (production-ranked from left to right) of a) maize; b) wheat, c) rice, and d) soybean. Thick black lines separate the *CFD+* ensembles, *CFD-all* ensembles, *Ensemble+*, and *Ensemble-all*, and the columns showing the top 20 producing countries and the global production response. Symbols indicate levels of significance (filled symbols are significant at 95th percentile level, open at 90th percentile level) as well as the highest correlation for each country (square indicates highest national correlation was not significant at 90th percentile level). Serbia maize and soybean are not shown (colored gray) as Serbia's recent independence makes for insufficient national production reports from 1981-2010; Ukraine (maize, wheat), Kazakhstan (wheat), and Uzbekistan (wheat) have only 18 years with FAO statistics available. GSWP3 and PGFv2 are not shown as not enough GGCMs simulated these CFDs.