

ACCEPTED MANUSCRIPT • OPEN ACCESS

Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios

To cite this article before publication: Christoph Müller *et al* 2021 *Environ. Res. Lett.* in press <https://doi.org/10.1088/1748-9326/abd8fc>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2020 The Author(s). Published by IOP Publishing Ltd.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios

Christoph Müller¹, James Franke^{2,3}, Jonas Jägermeyr^{4,5,1}, Alex C. Ruane⁵, Joshua Elliott⁴, Elisabeth Moyer^{2,3}, Jens Heinke¹, Pete Falloon⁶, Christian Folberth⁷, Louis Francois⁸, Tobias Hank⁹, R. César Izaurralde^{10,11}, Ingrid Jacquemin⁸, Wenfeng Liu¹², Stefan Olin¹³, Tom A. M. Pugh^{13,14,15}, Karina Williams^{6,16}, Florian Zabel⁹

¹Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, Germany

²Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA

³Center for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA

⁴NASA Goddard Institute for Space Studies, New York, NY, United States

⁵Center for Climate Systems Research, Columbia University Earth Institute, New York, USA

⁶Met Office Hadley Centre, Exeter, UK

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

⁸Unité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d'Astrophysique et de Géophysique, University of Liège, Liège, Belgium

⁹Ludwig-Maximilians-Universität München (LMU), Department of Geography, Munich, Germany

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

¹¹Texas Agrilife Research and Extension, Texas A&M University, Temple, TX, USA

¹²College of Water Resources and Civil Engineering, China Agricultural University, Beijing, China

¹³Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

¹⁴School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK

¹⁵Birmingham Institute of Forest Research, University of Birmingham, Birmingham, UK

¹⁶Global Systems Institute, University of Exeter, Exeter, UK

34 Abstract

35 Concerns over climate change are motivated in large part because of their impact on
36 human society. Assessing the effect of that uncertainty on specific potential impacts is
37 demanding, since it requires a systematic survey over both climate and impacts models.
38 We provide a comprehensive evaluation of uncertainty in projected crop yields for
39 maize, spring and winter wheat, rice, and soybean, using a suite of 9 crop models and
40 up to 45 CMIP5 and 34 CMIP6 climate projections for three different forcing scenarios.
41 To make this task computationally tractable, we use a new set of statistical crop model
42 emulators. We find that climate and crop models contribute about equally to overall
43 uncertainty. While the ranges of yield uncertainties under CMIP5 and CMIP6 projections
44 are similar, median impact in aggregate total caloric production is typically more
45 negative for the CMIP6 projections (+1 to -19%) than for CMIP5 (+5 to -13%). In the first
46 half of the 21st century and for individual crops is the spread across crop models typically
47 wider than that across climate models, but we find distinct differences between crops:
48 globally, wheat and maize uncertainties are dominated by the crop models, but soybean
49 and rice are more sensitive to the climate projections. Climate models with very similar
50 global mean warming can lead to very different aggregate impacts so that climate model
51 uncertainties remain a significant contributor to agricultural impacts uncertainty. These
52 results show the utility of large-ensemble methods that allow comprehensively
53 evaluating factors affecting crop yields or other impacts under climate change. The crop
54 model ensemble used here is unbalanced and pulls the assumption that all projections
55 are equally plausible into question. Better methods for consistent model testing, also at
56 the level of individual processes, will have to be developed and applied by the crop
57 modeling community.

58 Introduction

59 Climate change impacts on agriculture are subject to large uncertainties from a variety of
60 sources. One of the most important sources of uncertainty is associated with the severity
61 of climate change itself, even for a fixed emission scenario. For example, climate
62 projections in the CMIP5 archive (Coupled Model Intercomparison Project 5, Taylor *et al.*
63 (2012)) under the RCP8.5 scenario show a range of 3.2-4.9 K warming in mean
64 growing-season temperatures, and the more recent CMIP6 projections (Eyring *et al.*,
65 2016) show a range from 3.6-5.9 K. Climate models also differ not only in mean
66 projected changes over large regions but in the spatial patterns of those changes, with
67 precipitation an especial concern (e.g. Almazroui *et al.*, 2020; Akinsanola *et al.*, 2020).
68 Recent papers have compared CMIP6 to CMIP5 across a range of impact relevant
69 climate features such as extreme heat, precipitation, ENSO, and the monsoon (e.g. Fan
70 *et al.*, 2020; Freund *et al.*, 2020; Jiang *et al.*, 2020; Xin *et al.*, 2020; Zhu and Yang,
71 2020) to name a few. In many cases, CMIP6 has improved in skill of representing these
72 climate features, but climate models still show little improvement in some areas. In
73 general CMIP6 is noticeably more sensitive to CO₂ than CMIP5, largely due to the
74 updated representation of aerosols (e.g. Wyser *et al.*, 2020). Given these wide

75 uncertainties in climate projections, it is important to understand how they translate into
76 uncertainties in potential impacts on crop yields.

77 Process-based crop models provide a means of understanding the impact of different
78 climate changes on crop yields (Jones *et al.*, 2017). While these models were first
79 developed for application to individual sites and crop model ensembles were also used
80 at the site level to explore model-induced uncertainty (e.g., Asseng *et al.*, 2013; Palosuo
81 *et al.*, 2011), they have been extended to provide global coverage in the Global Gridded
82 Crop Model Intercomparison (GGCMI, Elliott *et al.* (2015)) of the Agricultural Model
83 Intercomparison and Improvement Project (AgMIP, Rosenzweig *et al.* (2013)). Global-
84 scale crop model applications are required for understanding future challenges to
85 agricultural production since production zones may shift under climate change, and
86 individual farms and regions are connected via agricultural markets and technological
87 development and innovation. The combination of global and regional scale analyses has
88 been shown to help in understanding the dynamics of agricultural production systems
89 (Rosenzweig *et al.*, 2018; Ruane *et al.*, 2018). Global crop simulations do suffer some
90 uncertainties since many processes cannot be fully calibrated at large scales – suitable
91 reference data and management information is not available for all regions – but global
92 crop simulations have been shown to have skill in reproducing observed historical inter-
93 annual variability and spatial patterns (Müller *et al.*, 2017).

94 Global assessments across ensembles of both crop models and climate projections
95 require some means of reducing computational demands. A comprehensive set of
96 climate projections in the CMIP5 or CMIP6 archives would consist of up to 34 and 45
97 members per radiative forcing scenario, and the GGCMI Phase II experiment alone
98 involved 12 different global crop models (Franke *et al.*, 2020a). These numbers are
99 prohibitive for computing a full set of crop yield simulations driven by different climate
100 projections. In practice, studies of future climate impacts on crop yields are often
101 performed using small and sometimes arbitrary selections of climate projections, crop
102 models, or crops. For example, McSweeney and Jones (2016) find that considering only
103 5 individual climate models in global impact assessments falls short of representing the
104 underlying uncertainty. A larger sample is required to fully characterize the uncertainty
105 range of climate models. Yet, a higher number of climate scenarios often proves
106 unpractical from the perspective of computational resources and climate change impact
107 assessments on agriculture often rely on climate projections from a small set of climate
108 models (e.g. Rosenzweig *et al.*, 2014).

109 In this work we avoid these computational bottlenecks and provide a more
110 comprehensive impacts assessment by using statistical emulators of individual crop
111 models. We present results of a global-scale assessment of potential crop yield changes
112 that explores the full range of the CMIP5 and CMIP6 climate projection archive. We use
113 a set of 9 global gridded crop model emulators (Franke *et al.*, 2020b) that were trained
114 on a very large systematic input sensitivity analysis with up to 1404 simulation data sets
115 per crop and crop model, each of 31 years in length and with near-global coverage
116 (Franke *et al.*, 2020a). The training domain represents an unprecedentedly rich data
117 base for emulator training, with perturbations in atmospheric carbon dioxide (CO₂)
118 concentrations (4 levels from 360 ppm to 810 ppm), air temperature (7 levels from -1 to
119 +6K), water supply (8 levels from -50 to +30% precipitation and full irrigation), nitrogen

1
2
3 120 (N) fertilization (3 levels from 10 to 200 kgN/ha) and adaptation (2 levels: none and
4 121 maintained growing seasons). The emulators themselves are grid-cell specific
5 122 regression models with 34 coefficients (Franke *et al.*, 2020b). Emulation allows a
6 123 computationally light-weight means of assessing crop yield impacts under arbitrary
7 124 climate and CO₂ scenarios that can be applied to the full CMIP5 and CMIP6 climate
8 125 archive. This exercise therefore allows us to evaluate the uncertainty in climate model
9 126 projections through the perspective of its implications for global food production.

12 127 In this analysis, we break down the different sources of uncertainty (greenhouse gas
13 128 concentration pathways, climate model, crop model) assess the role of the modeled
14 129 response to CO₂ fertilization and growing season adaptation and identify future
15 130 directions for crop model development and improvements.

16
17
18
19 131
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

132 Methods

133 In order to assess the current uncertainty in projections of future crop productivity on
134 current cropland, we combine the full GGCM Phase 2 crop model emulator ensemble
135 (Franke *et al.*, 2020b) with the full GCM ensemble of the CMIP5 and CMIP6 archives for
136 three different radiative forcing pathways: the representative concentration pathways
137 (RCP) 2.6, 4.6 and 8.5 (van Vuuren *et al.*, 2011).

138 The crop model emulator ensemble consists of 3rd order polynomial regression models
139 for nine different global gridded crop models (GGCMs) for the major staple crops maize,
140 spring wheat, winter wheat, rice and soybean. The emulators can reproduce well the
141 response of the original crop models to changes in carbon dioxide (C), temperatures (T),
142 water supply (W), nitrogen inputs (N) and growing season adaptation (A) that the models
143 showed in a large input sensitivity study using systematic parameter sweeps along the
144 CTWN-A dimensions. All crops are simulated separately for purely rainfed and for fully
145 irrigated systems, where irrigation is one element on the water availability dimension
146 (W). The CTNW-A experiment of the GGCM Phase 2 is described in detail by Franke *et al.*
147 (2020a). Emulator design and performance is described in detail by Franke *et al.*
148 (2020b). The emulators compute crop yields per crop and geographic location
149 (geographic grid at 0.5° longitude/latitude resolution) from atmospheric carbon dioxide
150 concentrations ([CO₂]), changes in growing season temperature (ΔT) and growing
151 season precipitation (ΔP), as well as nitrogen fertilizer inputs. Separate emulators exist
152 for purely rainfed and irrigated production systems as well as for the non-adapted setting
153 (same planting dates and variety selection as in the baseline period) and the adapted
154 setting (same planting dates, but new varieties that allow for maintaining the original
155 growing season length under warming). In this analysis, we work explicitly with the crop
156 model emulators, but since these are crop model specific emulators, we refer to the
157 GGCM-specific emulators with the names of the underlying GGCMs (CARAIB (Dury *et al.*
158 *et al.*, 2011), EPIC-TAMU (Izaurrealde *et al.*, 2006), GEPIC (Folberth *et al.*, 2012), JULES
159 (Williams *et al.*, 2017), LPJ-GUESS (Olin *et al.*, 2015), LPJmL (von Bloh *et al.*, 2018),
160 PEPIC (Liu *et al.*, 2016), PROMET (Hank *et al.*, 2015), pDSSAT (Elliott *et al.*, 2014)).
161 We obtained the largest possible climate model (GCM) ensemble from the CMIP5 and
162 CMIP6 archives that provide data for the historical period and at least one of the three
163 RCPs considered here (RCP2.6, RCP4.5 and RCP85 for CMIP5; SSP126, SSP245 and
164 SSP585 from the ScenarioMIP in CMIP6 (O'Neill *et al.*, 2016)). As we only consider
165 GCMs that contribute at least one of the considered RCPs and the historical period, our
166 GCM ensemble can differ from other ensembles (e.g. Meehl *et al.* (2020)). In order to
167 compute future yield projections, we compute average growing season mean
168 temperatures and average total growing season precipitation for a baseline period
169 (1980-2010) for each grid cell that is currently used to produce any of the five crops
170 considered here, following the MIRCA2000 data set (Portmann *et al.*, 2010) and
171 distinguishing both irrigated and rainfed growing seasons. The 31-year baseline period
172 corresponds to the reference period of the AgMERRA data set (Ruane *et al.*, 2015) that
173 was used as the basis for the GGCM Phase 2 CTWN-A simulations (Franke *et al.*,
174 2020a) and previous crop model evaluation (Müller *et al.*, 2017).

175 Against this crop- and grid cell-specific baseline conditions, we compute absolute
176 differences in average growing season temperature (ΔT in K) and relative differences in
177 total growing season precipitation (ΔP , unitless) for all future 31-year moving window

178 periods in the 21st century (2011-2084). As we are only interested in changes in 31-year
 179 average T and P from the historical simulation of the same GCM, no bias correction is
 180 necessary for the computation of ΔT and ΔP .

181 Growing season T is computed as the weighted average of monthly T data from each
 182 GCM, using the days per month within the growing season as weights. Growing season
 183 P is computed in a similar way, but using growing season totals, by adding monthly
 184 precipitation sums using days per month within the growing season to compute shares
 185 of precipitation that are considered as part of the growing season. Crop- and grid-cell
 186 specific growing season start and end dates are taken from the dataset used in the
 187 GGCM simulation phases 1 and 2 (Franke *et al.*, 2020a; Elliott *et al.*, 2015) so that
 188 these are consistent with what is assumed by the emulators. We do not change growing
 189 season length with increasing warming for the computation of average growing season
 190 conditions.

191 We consider all climate model projections in the CMIP archive that provide historical and
 192 future scenarios in a consistent manner. We used monthly data rather than daily data to
 193 increase sample size, which we consider more important than daily resolution. We
 194 assume errors induced by this are small, especially since growing season conditions are
 195 computed as 31-year moving window averages, which is the time frame on which the
 196 emulators have been trained (Franke *et al.*, 2020b). We accept different
 197 parameterization schemes of the same GCM as separate models where available to
 198 further increase sample size. We always only consider one ensemble simulation set per
 199 GCM, parameterization and RCP, selecting the smallest run number in the archive if
 200 several versions are available. Detailed information on the 45 CMIP5 and 34 CMIP6
 201 models considered, including version and ensemble member numbers, are listed in the
 202 supplementary tables S1 and S2.

203
 204 As some GGCMs tend to differ in their simulated baseline crop productivity levels (see
 205 e.g. Müller *et al.*, 2017), we harmonize simulated crop yields (Y_t^*) to match observed
 206 yield patterns from Mueller *et al.* (2012) as in equation 1, where Y_t is the simulated yield
 207 in time step t, A_c is the harvested area in grid cell c, $O_{r,c}$ is the observed yield in the
 208 reference period r and cell c and $Y_{r,c}$ is the simulated yield in the reference period in cell
 209 c.

$$Y_t^* = \sum Y_{t,c} * A_c * O_{r,c} / Y_{r,c} \quad \text{Equation 1}$$

210
 211 This is a simple multiplicative bias adjustment compared to more complex approaches
 212 used for the bias adjustment of climate projections. Our analysis is based on 31-year
 213 averages so that the focus is not on inter-annual or seasonal variations. Still the
 214 adjustment of the productivity levels helps to eliminate increased variance in the crop
 215 model ensemble from differences in mean biases as we are interested in projected
 216 changes here.

217 Crop yield data are aggregated to global production (P) using crop-specific harvested
 218 area data from MIRCA2000 (Portmann *et al.*, 2010). As winter and spring wheat are not
 219 explicitly distinguished in MIRCA2000, we assume that winter wheat is grown in a
 220 specific grid cell if the average temperature of the coldest month of the year is between -
 221 10°C and +7°C or if the growing season is longer than 150 days or if the growing season
 222 includes December (Northern Hemisphere) or July (Southern Hemisphere). Otherwise
 223 we assume that spring wheat is grown (see map in supplementary figure S1). Changes

224 in production are equivalent to changes in productivity (yields) here, as the harvested
225 area data set is static in time (equation 2).

$$P_t = \sum Y_{t,c}^* * A_c \quad \text{Equation 2}$$

226

227 For the aggregation of different crops, we compute total calories, assuming net water
228 contents of 12% for maize, spring and winter wheat, 13% for rice and 9% for soybean,
229 according to Wirsenius (2000) and caloric contents of the “as purchased” biomass (i.e.
230 including the water content) of 3.56kcal/g for maize, 2.8kcal/g for rice, 3.35kcal/g for
231 soybean and of 3.34kcal/g for spring and winter wheat, following FAO (2001).

232

233 As the central metric for uncertainty in crop yield projections, we compute total variance
234 across all GCM×GGCM combinations for all crops separately and for total calorie
235 production of all 5 crops considered here. We assume that the total variance $var(total)$ is
236 the sum of the variance across all GCMs $var(GCM)$ after averaging across all GGCMs
237 and of the variance across all GGCMs $var(GGCM)$ after averaging across all GCMs,
238 plus a cross term that describes the covariance between GCM and GGCM responses.
239 This cross term cannot be directly computed but we assume it to be the difference to
240 unity (equation 3).

241

$$1 = \frac{var(GCM)}{var(total)} + \frac{var(GGCM)}{var(total)} + cross.terms \quad \text{Equation 3}$$

242

243 With this assumption, which follows a similar uncertainty decomposition in climate
244 projections by Hawkins and Sutton (2009), shares of total variance can be attributed to
245 differences in GCMs or differences in GGCMs.

246 To test the robustness of this attribution to the ensemble composition, we compute the
247 variances for all sub-sets, leaving out one GGCM each time (i.e. 11%≈1/9), testing if
248 variance attribution is sensitive to the ensemble composition.

249

250 The GGCM Phase 2 input sensitivity CTNW-A experiment tested temperature increases
251 of up to +6K and precipitation changes between -50% and +30% (Franke *et al.*, 2020a).
252 Under RCP8.5 (SSP585 for CMIP6), some GCMs exceed this temperature range for
253 some cropland areas. With the non-linear design of the GGCM crop model emulator
254 ensemble (Franke *et al.* 2020b), it is difficult to extrapolate beyond its training domain
255 range, especially in the temperature dimension, which is, together with the [CO₂]
256 dimension, typically the most powerful feature in the models. To avoid overly spurious
257 crop model projections, we capped growing season temperature changes (ΔT) at +6K
258 and changes in precipitation at -50% and +30% at the grid-cell and crop-specific growing
259 season level. As the emulators rely on the balance of the T and C terms, we
260 simultaneously kept [CO₂] constant at the grid-cell and crop-specific growing season
261 value at which +6K for ΔT was reached. The majority of GCMs has only small fractions
262 of current cropland that exceed ΔT of +6K, but for some models, this can be substantial.
263 For the CMIP5 ensemble, 7% of all cropland exceeds +6K (2% for rice to 12% for spring
264 wheat) averaged across all GCMs for RCP8.5, while this is more severe for the CMIP6
265 ensemble (22% of all cropland, ranging from 9% for rice to 29% for spring wheat; see
266 supplementary Figure S2). As we drive the emulators with 31-year moving window
267 average, the last year considered here is 2084 (2069-2099). Therefore, we did not have

1
2
3 268 to generally cap [CO₂] at 810ppm, as this concentration level is only exceeded after
4 269 2086 (Riahi *et al.*, 2011).

5 270

6 271 Of the CMIP5 archive, CESM1-CAM5-1-FV2 had to be excluded due to missing
7 272 precipitation values for Dec 2056 and in the CMIP6 archive, CIESM had to be excluded
8 273 due to implausible strong decline of temperatures at the end of the 21st century.

9
10
11 274
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Accepted Manuscript

275 Results

276 Changes in T and P projections in CMIP ensembles

277 Generally, the spread of growing season changes in temperatures and precipitation is
 278 larger in the CMIP6 ensemble with 34 members than in the CMIP5 ensemble with 45
 279 members (Figure 1). Under the high radiative forcing scenario RCP8.5, the CMIP6
 280 ensemble projects a stronger median warming of about 1K and similar changes in
 281 precipitation as the CMIP5 ensemble. Differences in projected growing season warming
 282 are less pronounced in lower radiative forcing cases (RCP2.6 and RCP4.5) and scale
 283 with the radiative forcing (Figure 1, supplementary figures S3 and S4).

284

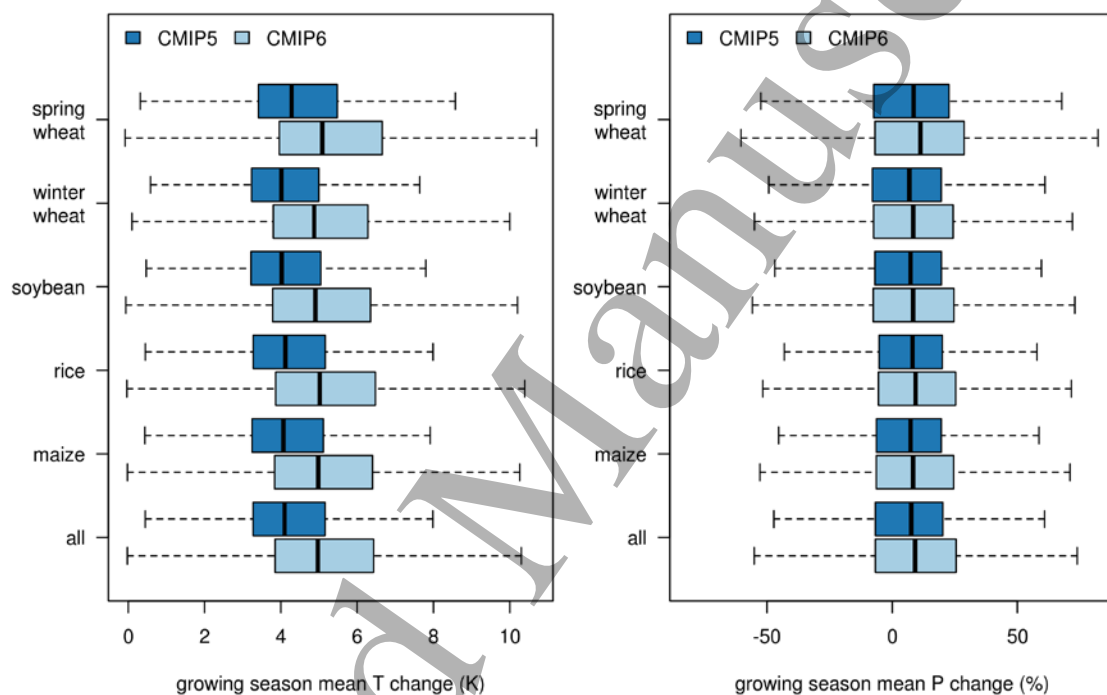


Figure 1: harvested-area weighted distribution of projected changes in crop-specific mean growing season temperatures (left) and precipitation sum (right) for the CMIP5 and the CMIP6 ensemble under RCP8.5 at the end of the 21st century (2069-2099). Colored boxes show the 25th to 75th percentile of the distribution and the thick black lines show the median. Whiskers extend to the maximum value within 1.5 times the interquartile range beyond the 25th and 75th percentile respectively. Outliers, i.e. values outside this range are not shown. Growing seasons are held constant across historical and future time periods. Figures for RCP4.5 and RCP2.6 are shown in the appendix, but show a similar pattern: warmer average conditions in CMIP6 and larger spread across the ensemble than in CMIP5.

285

286 Projected impacts

287 At the most aggregated level (across all crops, globally), the GCMxGGCM ensemble
288 projects a broad range of possible climate change impacts on crop productivity on
289 current cropland (Figure 2). The ensemble of crop model emulators projects consistently
290 more negative impacts on average (except for LPJ-GUESS where projections increase
291 by 1 percent point), so that the uncertainty range (± 1 standard deviation, colored area
292 in Figure 2) of only 3 GGCMs overlaps the zero line (CARAIB, LPJ-GUESS, PROMET)
293 for the CMIP6 ensemble, while this is the case for all but three crop models under
294 CMIP5. Still the most extreme projections for the CMIP6 scenario span farther into the
295 positive range than they do under CMIP5 (Figure 2).

296 We observe distinct differences between individual GCMs, with GEPIC and pDSSAT
297 being typically the most pessimistic models and CARIB and LPJ-GUESS the most
298 optimistic ones.

299 Projected impacts scale with the radiative forcing and with the GCMs' equilibrium climate
300 sensitivity (ECS, taken from Meehl *et al.* (2020)), which constitute an important
301 determinant of crop yield projections. Projected impacts are generally less variable at
302 lower radiative forcing (time axes in Figure 2 and different RCPs in supplementary
303 figures S5 and S6). Under RCP2.6, all but GEPIC project a positive median change for
304 the CMIP5 ensemble and all but GEPIC and pDSSAT do so for RCP2.6 and CMIP6
305 (supplementary figure S5) and for RCP4.5 and CMIP5. For RCP4.5 and CMIP6, 5 of 9
306 GGCMs project negative median impacts by the end of the 21st century (supplementary
307 figure S6).

308 The relationship between ECS and median climate change impact on crop yields is
309 stronger for the CMIP6 ensemble (Figure 3). However, the range of projected changes
310 in crop productivity can differ substantially at similar ECS values. The ECS relationship
311 with changes in crop productivity is weaker for the CMIP5-based ensemble as the GCM
312 with the lowest ECS (IPSL-CM5A-MR) shows the strongest decline in crop productivity
313 (Figure 3). The low ECS value reported by (Meehl *et al.*, 2020) is also not reflected in
314 the temperature increase of IPSL-CM5A-MR on current cropland of the 5 crops
315 considered here, where the mean temperature increase is not exceptionally high in
316 comparison to other GCMs, but certainly not at the low end (supplementary Figure S7).
317 This suggests that the IPSL-CM5A-MR model may have a different distribution of
318 warming over oceans vs. land or a much lower warming on non-cultivated land.

319
320 At the level of individual crops the GGCM ensemble shows distinct differences, even
321 though GEPIC and pDSSAT generally belong to the more pessimistic models and
322 CARAIB and LPJ-GUESS generally belong to the more optimistic models. For maize,
323 pDSSAT is the most pessimistic model, distinctly more so than the other models, with
324 end-of-the-century median projections of -32% (-41%) in comparison to -15% (-21%) for
325 GEPIC, the next most pessimistic GGCM for CMIP5 (CMIP6), see supplementary figure
326 S8), but also the ± 1 SD range of GEPIC does not overlap with that of CARIB, LPJ-
327 GUESS and PROMET. LPJ-GUESS projections broaden the projection range of the
328 GGCM ensemble substantially to the positive side for spring and winter wheat, but it also
329 covers the very pessimistic projection range for winter wheat. For these crops, LPJ-
330 GUESS is the most sensitive model to different GCMs.

331 There is no emulator for LPJ-GUESS for rice and soybean, as no simulations were
332 submitted for these crops to the GGCM Phase 2 data archive (Franke *et al.* 2020a). The

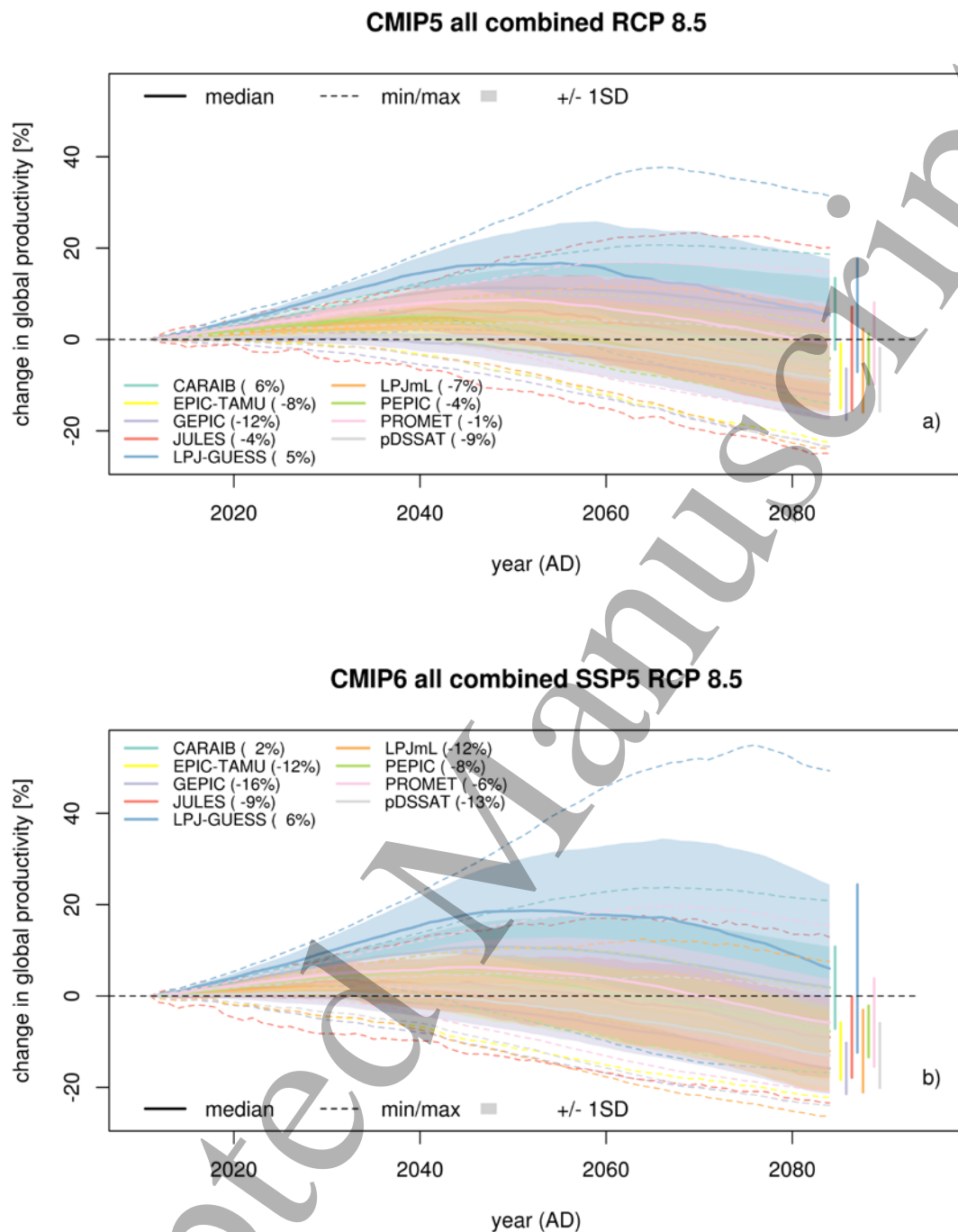


Figure 2: time series of projected impacts aggregated across the major five crops per crop model for RCP8.5. Thick lines show the median, dashed lines the minimum and maximum across all CMIP5 GCMs (panel a)) and all CMIP6 GCMs (panel b)), shaded areas represent +/- one standard deviation around the median. For better visibility, the range of +/- 1 SD per GGCM at the end of the 21st century is depicted as colored vertical lines and the median value in the last time slice (2069-2099) is given in parenthesis next to each GGCM's name.

333 +/-1SD range of all GGCMs overlap for soybean, whereas those of CARAIB and JULES
 334 for rice do not overlap with the +/-1SD ranges of EPIC-TAMU and GEPIC and that of
 335 JULES does not overlap with PEPIC in both CMIP5 and CMIP6 and with that of
 336 pDSSAT only for CMIP5 (supplementary Figures S8-S12).

337
 338

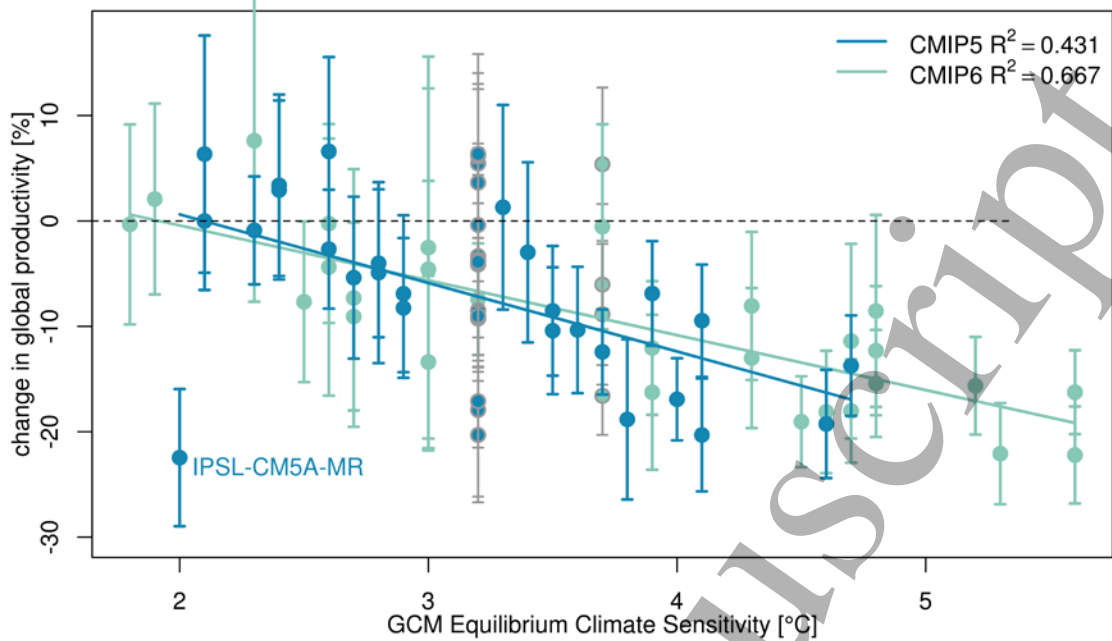


Figure 3: GGCM-ensemble projected changes in global crop productivity (%) for the CMIP5 (blue) and CMIP6 (green) ensembles for RCP8.5 at the end of the 21st century (2069-2099). Dots indicate the median projections, whiskers extend to ± 1 standard deviation from that median. Not all GCMs included in this analysis have reported ECS values in Meehl *et al.* (2020) do not report ECS values for all GCMs included here and we substituted these missing values with the CMIP ensemble mean (3.2 for CMIP5, 3.7 for CMIP6) in the figure, but exclude these in the fitting of the regression model (solid lines) here. These values are indicated by a grey border around the dot and grey whiskers.

339

340

341

342

343 Sources of uncertainty

344 We find substantial differences in overall variance in projected changes in crop
 345 productivity between the CMIP5 and CMIP6 ensembles. Total variance of the full crop
 346 model emulator and climate projections ensemble, as a measure for uncertainty, is
 347 larger for CMIP6 than for CMIP5 (Figure 4) for RCP2.6 and 8.5, but similar for RCP4.5.
 348 In the CMIP6 ensemble, the variance of both wheats, but especially winter wheat
 349 increases compared to the CMIP5 ensemble under the high radiative forcing pathway
 350 RCP8.5, while that of soybean decreases. The overall variance of crop yield projections
 351 of the ensemble increases with the radiative forcing (RCP, time) in both the CMIP5 and
 352 CMIP6 ensembles (Figure 4). This increase is strongest in the middle of the 21st century
 353 and levels off towards the end of the 21st century. This leveling-off effect can be
 354 observed at all RCPs (Figure 5), but is less strong for simulations where the effect of
 355 CO₂ fertilization is ignored or where growing season adaptation is considered (Figure 6).

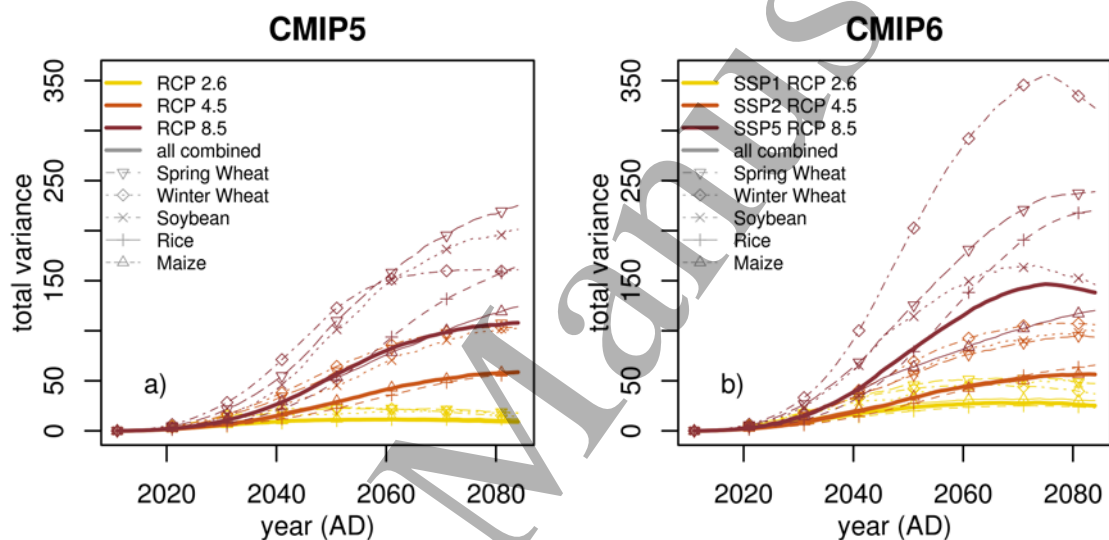


Figure 4: total variance in global productivity on current cropland across all GGCM and GCM combinations per RCP (colors) and crop (line type and symbol) for CMIP5 (panel a) and CMIP6 (panel b)). Thick colored, solid lines represent the calorie-weighted aggregation of all crops.

356

357 Breaking down overall variance in projections into a GGCM and a GCM component, we
 358 find that the GGCM component dominates in the first half of the 21st century and the
 359 GCM component gradually increases after a peak in GGCM component, typically
 360 between 2020 and 2030 (Figure 5). The shares of GGCM and GCM-induced variance
 361 are largely independent and cross-terms typically account for only a small fraction of the
 362 overall variance. The peak in GGCM-induced variance is less pronounced in the CMIP6
 363 ensemble than in CMIP5 ensemble, because the GCM-induced variance increases
 364 strongly only in the second half of the 21st century in the CMIP5 ensemble, but increases
 365 more steeply (relative to the GGCM-induced variance) from 2020 onwards in the CMIP6
 366 ensemble.

367

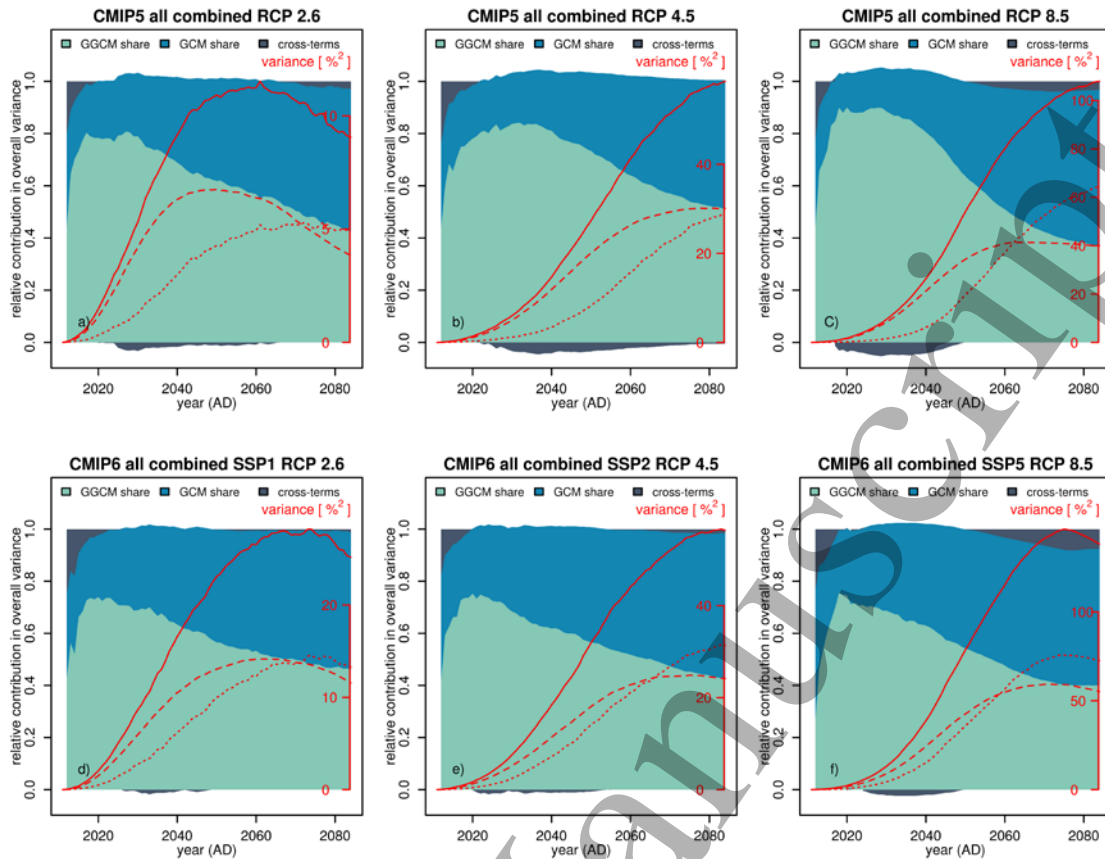


Figure 5: relative contributions of GGCMs and GCMs to the overall ensemble crop yield projections under CMIP5 (top) and CMIP6 (bottom). Red lines indicate absolute variance of the total ensemble (solid), the GGCM share (dashed) and the GCM share (dotted). Relative contributions are fairly similar across RPCs, but absolute variance increases significantly with the radiative forcing (see right-hand red axis). Scales for absolute variance are adjusted per panel and are thus not directly comparable. The variance shares of GGCMs and GCMs do not always add up to the total variance as these two sources of uncertainty are not fully independent. The difference to total variance is shown in dark blue and referred to as “cross-terms” (see equation 3).

368

369

370

371

372 While overall variance can be substantially decreased if the CO₂ fertilization effect is

373 ignored, the share of GGCM-induced variance tends to increase under this setting,

374 especially in the CMIP6 ensemble (Figure 6). Ignoring the CO₂ fertilization effect does

375 not provide plausible future crop yield projections, but it helps to analyze where the

376 GGCM-induced variance originates from. We find that crop models agree more strongly,

377 if the process of CO₂ fertilization is ignored. In other words, the simulated effects of CO₂

378 fertilization on crop yields are an important source of crop model disagreement.

379 Adaptation of cultivars to regain the growing season length that would otherwise be lost

380 due to accelerated phenological development (Franke *et al.*, 2020a; Minoli *et al.*, 2019;381 Zabel *et al.*, under review) on the other hand increases the GGCM-induced variance

382 share and overall variance substantially. This is because crop models show very

383 different responses to this adaptation measure so that overall uncertainty is increased if

384 cultivar adaptation (as implemented in the GGCM Phase 2 simulations) is considered

385 (Minoli *et al.*, 2019).

386 We also find that the ensemble of crop models is very sensitive to the selection of

ensemble members. If one of the nine crop models is excluded from the ensemble, the

1
2
3 387 relative contribution to overall variance from crop models can vary strongly (figure 6).
4 388 Which GGCM has strong effects on the overall variance attribution is crop specific. If
5 389 random sets of climate models that constitute a similar share of the ensemble size ($n=4$
6 390 of 34 for CMIP6, roughly equivalent of 1 in 9 crop models), we find that results on the
7 391 GCM- and GGCM-induced variance shares change less than if individual GGCMs are
8 392 excluded in the first half of the 21st century, but can be affected similarly strongly at the
9 393 end of the century (supplementary Figure S13), suggesting that the distribution of
10 394 changes in the GCM ensemble is more balanced in short-term projections than that
11 395 within the GGCM ensemble.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

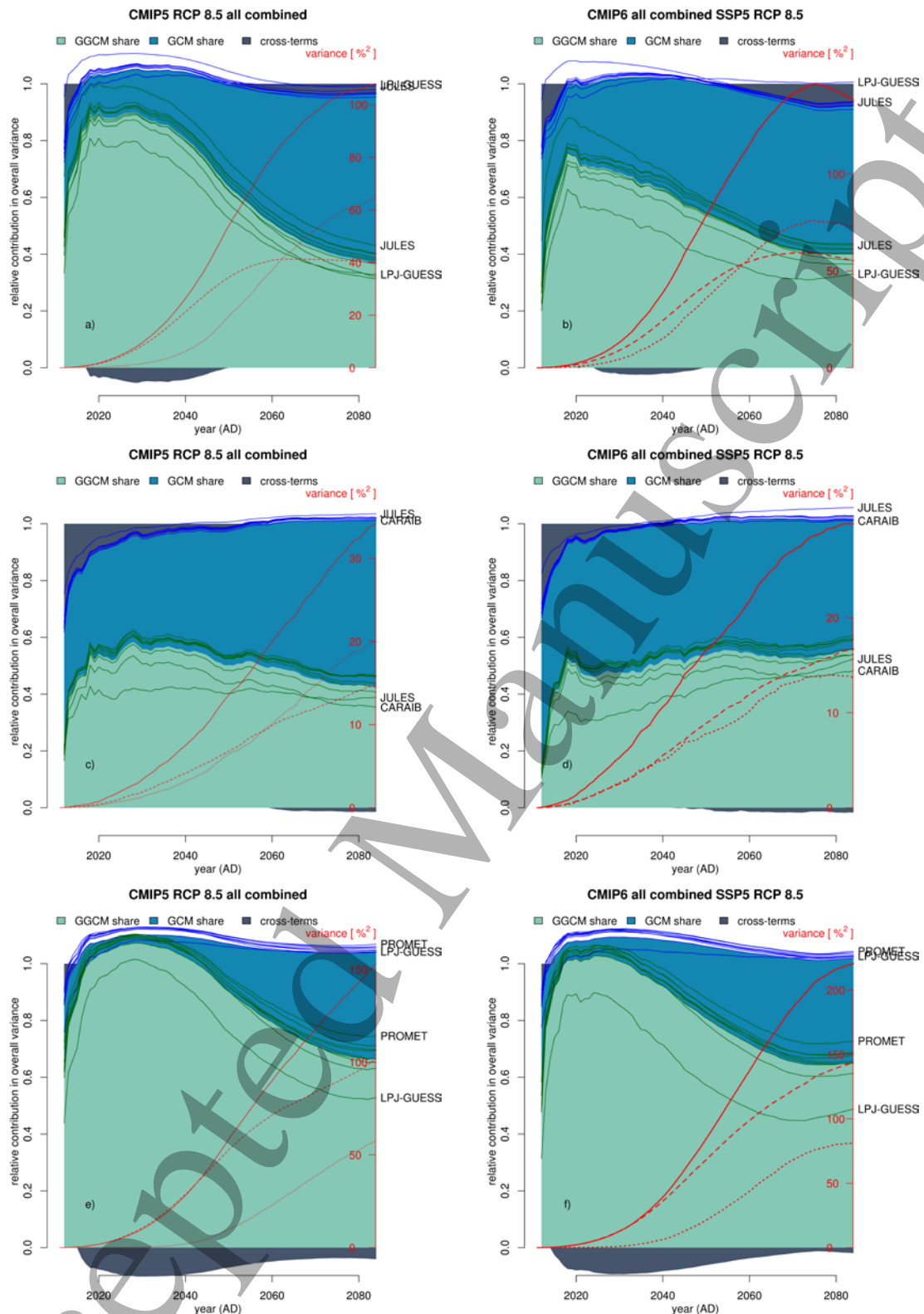


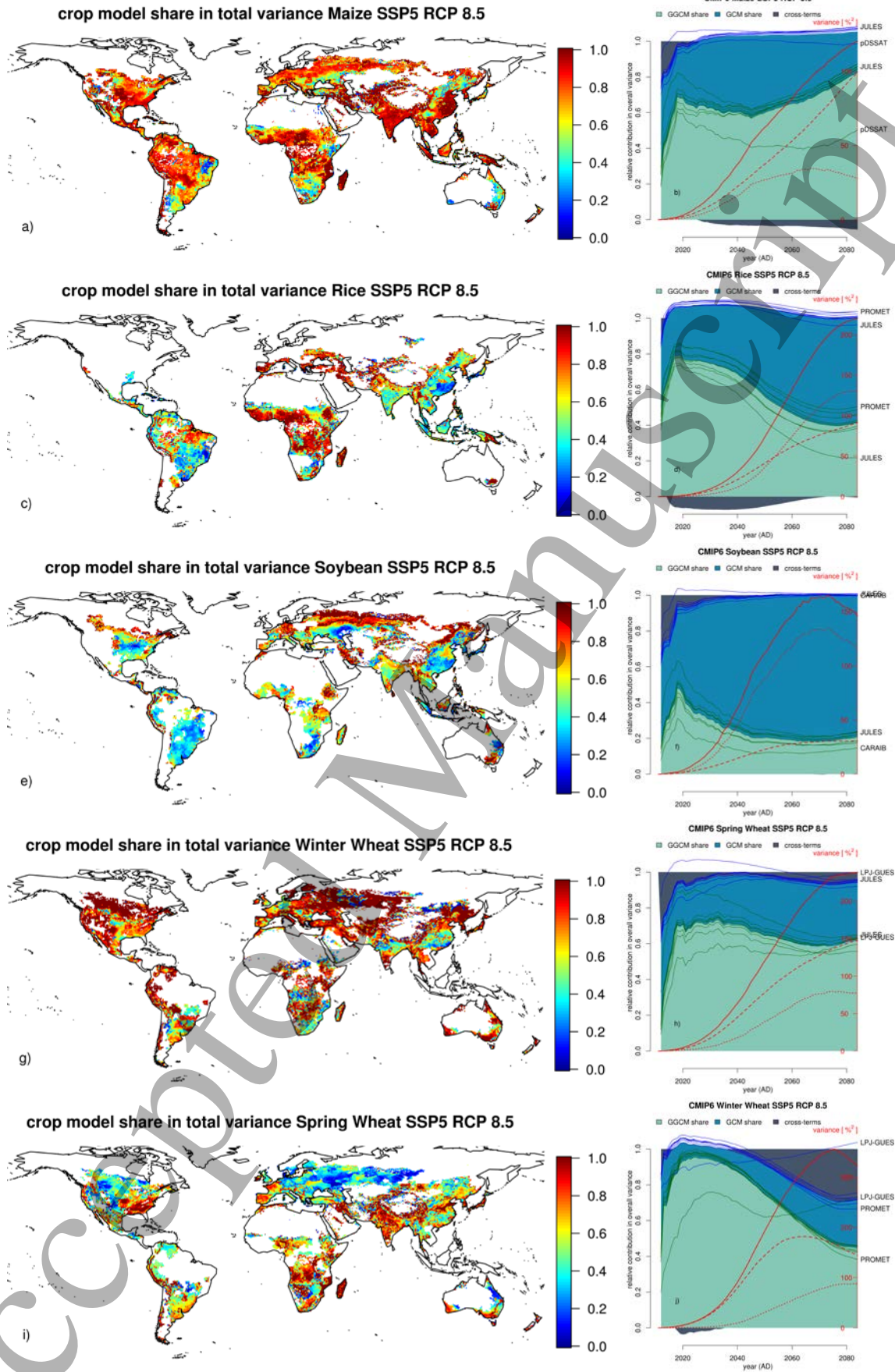
Figure 6: as figure 5, but for the standard setting (top row; panels are equivalent to right hand panels in figure 5), the projections ignoring the CO₂ fertilization effect (middle row) and the projections including the variety adaptation to regain the growing season (bottom row) for CMIP5 (left) and CMIP6 (right). Thin lines show how GCM- (blue) and GGCM-induced shares (green) in overall variance would change if one GGCM were excluded from the ensemble. The exclusion of individual GCMs can also affect the contribution of cross-terms, i.e. higher or lower co-variance between the GCM- and GGCM-shares (e.g. thin blue lines above 1.0). Red lines indicate absolute variance of the total ensemble (solid), the GGCM share (dashed) and the GCM share (dotted). Relative contributions are fairly similar across RPCs, but absolute variance increases significantly with the radiative forcing.

397 Crop specific differences

398 For individual crops, we observe substantial differences in the share of variance that can
399 be attributed to crop models. For maize and spring wheat, the GGCM-induced variance
400 shares clearly dominate the overall variance. GCM-induced variance is clearly the most
401 important contribution to overall variance in soybean yield projections and to lesser
402 extent in rice projections. Winter wheat shows a strong contribution of cross terms to the
403 overall variance, which is also true to some extent for spring wheat. This cross-term
404 contribution can be substantially reduced by excluding LPJ-GUESS from the winter
405 wheat GGCM ensemble. Excluding JULES from the spring wheat GGCM ensemble
406 would increase the GGCM-induced variance share in the first half of the 21st century and
407 would introduce negative cross-terms. Excluding LPJ-GUESS from the spring-wheat
408 ensemble on the other hand would do the opposite and reduce the GGCM-induced
409 variance share throughout most of the 21st century and would introduce larger positive
410 cross-term shares (Figure 7).

411 For most crops, there is a clear outlier model that, if excluded, strongly changes the
412 contribution of GCMs, GGCMs or cross-terms to overall variance. For maize, this is
413 pDSSAT, which projects the most pessimistic yield declines in the GGCM Phase 2
414 emulator ensemble (see Appendix Figure A1). An exclusion of pDSSAT from the
415 ensemble would reduce overall variance by more than half and substantially reduce the
416 GGCM-induced contribution. If PROMET, LPJmL or pDSSAT were excluded from the
417 rice model ensemble, the GGCM induced variance would increase, whereas it would
418 substantially decrease if JULES were excluded. The exclusion of JULES would also
419 substantially reduce overall variance of the full GCM×GGCM ensemble. Even though
420 there is generally much less GGCM-induced variance in soybean yield projections, the
421 exclusion of CARAIB would lead to a further reduction of overall variance and of the
422 GGCM-induced share.

423



1
2
3 Figure 7: crop-specific variance attribution for CMIP6 and RCP8.5 only. Right column (panels b, d, f, h, j) shows
4 changes in GCM and GGCM induced variance shares (colored areas) as well as the sensitivity of these shares to
5 exclusion of individual GGCMs from the ensemble (thin lines), the most sensitive ensemble members are labeled. Red
6 lines indicate absolute variance of the total ensemble (solid), the GGCM share (dashed) and the GCM share (dotted).
7 Scales for variance are adjusted per crop and are thus not directly comparable. Maps in the left column (panels a, c, e,
8 g, i) show the GGCM-induced variance share at the grid cell level in the last time step (2084).

424

425

Accepted Manuscript

426 Discussion

427 This unprecedentedly large ensemble of climate projections, crop model (emulators) and
428 crops allows to explore the importance of ensemble composition for climate change
429 impact analyses on crop yields and examine the uncertainty in climate model ensembles
430 through the lense of climate impacts. We find that climate projections can have a
431 substantial influence on crop yield projections, especially in marginal and dry regions,
432 but spatial patterns differ by crop (Figure 7).

433 The use of computationally efficient crop model emulators in place of the process-based
434 crop models is the only option to conduct this large ensemble analysis. While the
435 emulators have very good skill in reproducing the underlying crop models (Franke *et al.*,
436 2020b), they are no perfect reproduction of the crop models' dynamics. Our results are
437 thus only indicative of the actual contributions of crop models to overall uncertainty in
438 crop yield projections.

439 Across the full CMIP5 and CMIP6 archives, there is substantial spread in crop yield
440 projections, independent of the radiative forcing (RCP2.6, 4.5, or 8.5). At the end of the
441 21st century, climate model-induced variance is often dominant over crop model-induced
442 variance, i.e. the uncertainties in climate projections are more important for projections
443 of changes in crop yields than the uncertainties in crop models – at least at the most
444 aggregate level (combined global productivity of all crops considered here). For
445 individual crops, crop-model induced variance is larger than the climate model-induced
446 variance for maize, spring wheat and winter wheat, which jointly contribute the majority
447 of calories from the 5 crops considered here. As such, it is surprising to see that climate
448 model-induced variance is dominant over crop model-induced variance when the five
449 crops are aggregated to overall production. This suggests that there is some cancelation
450 of signals when different crops are aggregated. One example of such mutual
451 compensation of variance is the combination of predominantly negative projections for
452 maize productivity (supplementary Figure S8) and the predominantly positive projections
453 for spring and especially winter wheat (supplementary Figure S12). This may illustrate
454 compensatory responses between crops within the crop models and/or changing
455 patterns of warming within the climate models. Similarly, variances in space can cancel
456 out in the aggregation to global productivity if some regions are projected to see positive
457 effects and others to see negative impacts of climate change (e.g. winter wheat in Figure
458 7). Looking at the distribution of projected changes in global crop productivity as done
459 here via the variability metric does thus not represent the full scope of disagreement
460 among simulations. The aggregation of data across space or crops can lead to
461 cancelation of variance at the underlying level of detail that is not visible at the level of
462 analysis here. Still, the analysis provides a unique overview of the breadth of projections
463 of global crop productivity under climate change.

464 Differences across crops do not necessarily only represent differences in the simulated
465 dynamics and processes of these crops, but can also reflect the differences in the crop
466 model ensemble. LPJ-GUESS for example, which is at the most positive side of the
467 projected yield changes for spring and winter wheat did not supply data for soybean and
468 rice in the CTWN-A experiment (Franke *et al.*, 2020a) and is thus not included in the
469 emulator ensemble for these crops (Franke *et al.*, 2020b). However, the exclusion of
470 LPJ-GUESS from the wheat ensembles does not make the uncertainty attribution for
471 spring and winter wheat more similar to that of the other crops.

1
2
3 472 More crops need to be explicitly considered in climate change impact assessments, as
4 473 individual crops show distinctly different spatial patterns, uncertainties in crop yield
5 474 projections and the relative contribution of GCM- vs. GGCM-induced variance. As we
6 475 find that impact projections (supplementary Figures S8-S12) as well as drivers of
7 476 uncertainty (Figure 7) differ between different cereal crops, other crops like legumes,
8 477 tree or other perennial crops must be explicitly analyzed. Therefore, the behavior of
9 478 crops other than the major five considered here can likely not be well represented by
10 479 these. Considering the comparative high amount of research attention these 5 crops have
11 480 received, uncertainty must be very high for other crops. It is thus of fundamental
12 481 importance to broaden the range of simulated crops, also because there is the need to
13 482 represent a much broader set of crops in economic analyses of agricultural markets and
14 483 land-use dynamics under climate change. The current practice to derive climate change
15 484 impacts of crops that are not modeled by crop models from a small set of crops that is
16 485 modeled (Nelson *et al.*, 2014; Müller and Robertson, 2014) thus needs to be challenged,
17 486 even though there may be little alternative under current constraints on data availability.
18 487 The next round of AgMIP/ISIMIP future projections (Jägermeyr *et al.* in prep.) also aims
19 488 at broadening the scope of simulated crops, but many models are not available for less
20 489 ubiquitously grown crops.

21 490 For short- and mid-term projections, GGCM-induced variance dominates the overall
22 491 variance across all scenarios and crops, except for soybean, where crop models
23 492 generally contribute only a small share to overall variance and where also overall
24 493 variance is relatively low (2nd after maize). This dominance of the GGCM signal in the
25 494 first half of the 21st century is likely because of the relatively small differences in radiative
26 495 forcing in this period, which is also largely independent of the RCP trajectory (van
27 496 Vuuren *et al.*, 2011).

28 497 Future crop yields are determined by counteracting drivers. Climate change impacts
29 498 (warming, changes in precipitation) lead to overall negative impacts on crop yields that
30 499 amplify unequivocally with the radiative forcing at the global aggregation level. However,
31 500 the main cause of climate change, increasing atmospheric CO₂ concentrations from
32 501 anthropogenic emissions, also lead to increased crop productivity. There is substantial
33 502 uncertainty connected to the effects of CO₂ fertilization in models, especially at high
34 503 concentrations as projected for the end of the 21st century under RCP8.5, where also
35 504 little experimental evidence can guide model parameterization and development (Toreti
36 505 *et al.*, 2020). Nonetheless, the modeled response to elevated atmospheric CO₂
37 506 concentrations requires more attention from the modeling community.

38 507 In this analysis, we focused on changes in the CTW dimensions of the emulated CTNW-
39 508 A experiment (Franke *et al.*, 2020a; Franke *et al.*, 2020b), ignoring the N dimension,
40 509 which can also contribute to overall uncertainty. We kept N inputs at historical patterns
41 510 across regions and crops (Elliott *et al.*, 2015) throughout the simulations. It is plausible
42 511 to assume that N fertilization would change under changing crop yield potentials, market
43 512 access and dynamics, or environmental regulation. To our knowledge, there are no such
44 513 projections available, especially not any that would account for the changes in potential
45 514 yields under the multitude of climate projections used here. Long-term crop projections
46 515 here do not account for other technical and management changes in addition to
47 516 nitrogen, which additionally artificially suppresses the crop model-induced component of
48 517 uncertainty. This is somewhat analogous to the 'pathway' uncertainty in the SSP-RCP
49 518 framework.

1
2
3 519 Still, we find that the GGCM ensemble contributes relatively little to overall variance in
4 520 regions with intensive agriculture (supplementary figure S14) as well as for soybean (a
5 521 nitrogen fixing plant) more generally, suggesting that the response to N inputs is also an
6 522 important driver of uncertainty in crop yield projections. The relationship between
7 523 nutrient limitations and susceptibility to climate change impacts as well as how nutrient
8 524 limitation is modeled at different levels of nutrient supply need further scrutiny.
9
10 525 Our results for the end of the 21st century need to be interpreted with some caveats, as
11 526 we had to cap CTW drivers to the training domain of the CTNW-A experiment (Franke *et al.*,
12 527 2020a), because of the non-linear functional form of the emulators (Franke *et al.*,
13 528 2020b), which makes extrapolation beyond the training domain volatile and error-prone.
14
15 529 For the majority of GCMs and harvested areas, this is not a major caveat as most areas
16 530 do not exceed +6K. However, for some GCMs, especially under CMIP6, large fractions
17 531 of the crops' harvested areas exceed the +6K warming level (supplementary figure S2).
18 532 This leads to an artificial reduction of the GCM-induced variance in results. The
19 533 plausibility of the very high ECS in climate projections has been challenged (Tokarska *et al.*,
20 534 2020) and the ensemble could be pruned on this basis to avoid very warm climate
21 535 projections. However, the selection of climate scenarios provided to climate impact
22 536 modeling community in e.g. ISIMIP does not necessarily follow such pre-selection
23 537 approaches and we thus kept the full CMIP6 archive here. The saturating overall
24 538 variance that can be observed towards the end of the 21st century could suggest that the
25 539 capping of the warming at +6K leads to an artificial reduction of the end-of-the-century
26 540 variance, however we observe the same general feature (steepest increase in variance
27 541 in mid-century) also in the other RCPs that are not subject to the capping of temperature
28 542 signals as warming levels are generally lower (Figure 5). The observed saturation of
29 543 variance towards the end of the 21st century cannot be attributed to a saturation in
30 544 drivers of climate change as global mean cropland temperatures under RCP8.5 show no
31 545 sign of levelling off (supplementary Figure S7) and also [CO₂] and radiative forcing do no
32 546 level off under RCP8.5 (van Vuuren *et al.*, 2011).

33
34
35
36
37
38 547 Generally, climate and crop models should be selected on a fit-for-purpose basis. While
39 548 the climate community has established the standard that the same model versions that
40 549 provide future projections also provide historical simulations for evaluation purposes, this
41 550 procedure has not generally been adopted by the crop modeling community. The ISIMIP
42 551 project is promoting a similar structure in the individual simulation rounds (Frieler *et al.*,
43 552 2017), but crop models need to more rigorously provide meta information on the model
44 553 version and parameterization, which can greatly affect simulated dynamics (Folberth *et al.*,
45 554 2019). The common practice to reduce the uncertainty space by selecting a small
46 555 number of climate scenarios by e.g. first availability has already been challenged by
47 556 McSweeney and Jones (2016). We show that, at the global scale, the selection of
48 557 individual crop models can greatly affect the outcomes and even the exclusion of one
49 558 out of an ensemble of nine can have substantial effects on results. This pulls the general
50 559 assumption into question if we can consider all GGCM projections as equally plausible,
51 560 or if the skewed distribution suggests that some models should indeed be excluded prior
52 561 to the interpretation of ensemble results. More and also different global gridded crop
53 562 models are expected to contribute to the new round of global crop model simulations of
54 563 AgMIP and ISIMIP (Jägermeyr *et al.*, in prep.). However, it may not necessarily be
55 564 desirable to increase the ensemble size to a point where the exclusion of sub-samples

1
2
3 565 no longer affects the overall ensemble response if the unbalanced ensemble may be
4 566 caused by inclusion of non-plausible projections.

5
6 567 Thus, we call for intensified efforts to understand why crop models differ and to build
7 568 strategies on how models can be improved – or that lead to a better understanding why
8 569 it is plausible to have an as broad distribution as our current full ensemble suggests.
9 570 While better model agreement is not an appropriate aim in itself, model disagreement
10 571 can be used to identify aspects for coordinated model improvement, as e.g. described
11 572 by Maiorano *et al.* (2017). Also, the assessment of crop models based on their ability to
12 573 reproduce spatial and temporal patterns of historical crop yields (Müller *et al.*, 2017)
13 574 needs to be expanded by plausibility tests in individual model components and
14 575 processes. Given that crop yields are determined by many interacting processes
15 576 (Schauberger *et al.*, 2016), which have not been all implemented or sufficiently tested in
16 577 crop models (Boote *et al.*, 2013), we need to do everything possible to minimize the
17 578 chance of getting the right answer for the wrong reason as shown e.g. by Zhu *et al.*
18 579 (2019) for maize yields in the USA. As such, model performance needs to be also
19 580 assessed at the level of individual processes before errors in these can mutually cancel
20 581 out and are not traceable in the yield projections.

21
22 582 Toreti *et al.* (2020) call for a set of standard tests on crop models' response to elevated
23 583 [CO₂] that should be made accessible as meta-data for each model. Building on this
24 584 idea, we call for a set of standard tests for crop models across all major drivers of crop
25 585 yield simulations ([CO₂], temperatures, precipitation, nutrients, management aspects)
26 586 with respect to single driver effects as well as with respect to their interaction. The
27 587 CTWN-A experiment (Franke *et al.*, 2020a) that also covers more crop growth metrics
28 588 than just yields, provides a suitable basis for such tests, even though the computational
29 589 requirements are too high to qualify for a standard test.

30
31 590 Protocols for such standard model tests need to be developed in close collaboration with
32 591 experimentalists as they need to reflect the evolving understanding of physiological
33 592 processes, and need to include more aspects than just end-of-season yields. Even
34 593 though global crop model results are difficult to compare to data from experimental sites
35 594 (Deryng *et al.*, 2016), global (and field-scale) crop models need to be tested at the site
36 595 level for plausible response types (e.g. direction of change) and ranges (e.g. size of
37 596 effects). The comparison of global crop model results with site data has been shown to
38 597 allow for ex-post corrections of the range of simulated crop yield projections (Wang *et*
39 598 *al.*, 2020).

599 Conclusions

600 We find that future crop yield projections are subject to substantial uncertainties. These
601 increase with the radiative forcing, i.e. over time and also with the emission pathway
602 considered. Crop model-induced uncertainty dominates the overall uncertainty in the first
603 half of the projections for the 21st century and more efforts are needed to improve crop
604 model skill and testing procedures. In the second half of the 21st century, the overall
605 uncertainty surges, mainly driven by a steeper increase of uncertainty from climate
606 models. Long-term projections are thus of mainly academic value that can help to derive
607 insights from comparing scenarios and assumptions but should not be confused with
608 predictions of future developments. This is especially true as modifications in
609 management that can be expected to be implemented by farmers are often ignored due
610 to a lack of data on management systems and missing tools to project these into the
611 future. The unbalanced nature of the crop model ensemble, where often individual
612 models strongly affect the overall ensemble behavior call for intensified research on
613 climate change impact modeling for agriculture. This has been pleaded for by Rötter *et*
614 *al.* (2011) before and the various activities in AgMIP, MACSUR, ISIMIP and elsewhere
615 have helped to move in that direction. Still, more efforts are needed, especially with
616 respect to model evaluation standards and testing of other aspects than crop yields, as
617 e.g. by Kimball *et al.* (2019).

618

Acknowledgements

619
620 CMIP5 model output data provided by the WHOI CMIP5 Community Storage Server,
621 Woods Hole Oceanographic Institution, Woods Hole, MA, USA from their website
622 at <http://cmip5.who.edu/>.
623 J.A.F. was supported by the NSF NRT program (grant no. DGE-1735359) and the NSF
624 Graduate Research Fellowship Program (grant no. DGE-1746045). RDCEP is funded by
625 NSF through the Decision Making Under Uncertainty program (grant #SES-1463644).
626 KW was supported by the Newton Fund through the Met Office program Climate
627 Science for Service Partnership Brazil (CSSP Brazil).
628 The publication of this article was partially funded by the Open Access Fund of the
629 Leibniz Association

630

631

632

References

- 633 Akinsanola A A, Kooperman G J, Reed K A, Pendergrass A G and Hannah W M 2020
 634 Projected changes in seasonal precipitation extremes over the United States in
 635 CMIP6 simulations *Environmental Research Letters* **15** 104078
- 636 Almazroui M, Saeed F, Saeed S, Nazrul Islam M, Ismail M, Klutse N A B and Siddiqui M
 637 H 2020 Projected Change in Temperature and Precipitation Over Africa from
 638 CMIP6 *Earth Systems and Environment* **4** 455-75
- 639 Asseng S, Ewert F, Rosenzweig C, Jones J W, Hatfield J L, Ruane A C, Boote K J,
 640 Thorburn P J, Rotter R P, Cammarano D, et al. 2013 Uncertainty in simulating
 641 wheat yields under climate change *Nature Climate Change* **3** 827-32
- 642 Boote K J, Jones J W, White J W, Asseng S and Lizaso J I 2013 Putting mechanisms
 643 into crop production models *Plant Cell Environ* **36** 1658-72
- 644 Deryng D, Elliott J, Folberth C, Müller C, Pugh T A M, Boote K J, Conway D, Ruane A C,
 645 Gerten D, Jones J W, et al. 2016 Regional disparities in the beneficial effects of
 646 rising CO2 concentrations on crop water productivity *Nature Climate Change* **6**
 647 786-90
- 648 Dury M, Hambuckers A, Warnant P, Henrot A, Favre E, Ouberdous M and Francois L
 649 2011 Responses of European forest ecosystems to 21st century climate:
 650 assessing changes in interannual variability and fire intensity *iForest -*
 651 *Biogeosciences and Forestry* **4** 82-99
- 652 Elliott J, Kelly D, Chryssanthacopoulos J, Glotter M, Jhunjhnuwala K, Best N, Wilde M
 653 and Foster I 2014 The parallel system for integrating impact models and sectors
 654 (pSIMS) *Environmental Modelling & Software* **62** 509-16
- 655 Elliott J, Müller C, Deryng D, Chryssanthacopoulos J, Boote K J, Büchner M, Foster I,
 656 Glotter M, Heinke J, Iizumi T, et al. 2015 The Global Gridded Crop Model
 657 Intercomparison: data and modeling protocols for Phase 1 (v1.0) *Geoscientific*
 658 *Model Development* **8** 261-77
- 659 Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016
 660 Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6)
 661 experimental design and organization *Geosci. Model Dev.* **9** 1937-58
- 662 Fan X, Miao C, Duan Q, Shen C and Wu Y 2020 The Performance of CMIP6 Versus
 663 CMIP5 in Simulating Temperature Extremes Over the Global Land Surface
 664 *Journal of Geophysical Research: Atmospheres* **125** e2020JD033031
- 665 FAO 2001 *Food Balance Sheets: a handbook* (Rome: FAO)
- 666 Folberth C, Elliott J, Müller C, Balkovič J, Chryssanthacopoulos J, Izaurrealde R C, Jones
 667 C D, Khabarov N, Liu W, Reddy A, et al. 2019 Parameterization-induced
 668 uncertainties and impacts of crop management harmonization in a global gridded
 669 crop model ensemble *PLOS ONE* **14** e0221862
- 670 Folberth C, Gaiser T, Abbaspour K C, Schulin R and Yang H 2012 Regionalization of a
 671 large-scale crop growth model for sub-Saharan Africa: Model setup, evaluation,
 672 and estimation of maize yields *Agric Ecosyst Environ* **151** 21-33
- 673 Franke J A, Müller C, Elliott J, Ruane A C, Jägermeyr J, Balkovic J, Ciais P, Dury M,
 674 Falloon P D, Folberth C, et al. 2020a The GGCM Phase 2 experiment: global
 675 gridded crop model simulations under uniform changes in CO2, temperature,
 676 water, and nitrogen levels (protocol version 1.0) *Geosci. Model Dev.* **13** 2315-36
- 677 Franke J A, Müller C, Elliott J, Ruane A C, Jägermeyr J, Snyder A, Dury M, Falloon P D,
 678 Folberth C, François L, et al. 2020b The GGCM Phase 2 emulators: global
 679 gridded crop model responses to changes in CO2, temperature, water, and
 680 nitrogen (version 1.0) *Geosci. Model Dev.* **13** 3995-4018
- 681 Freund M B, Brown J R, Henley B J, Karoly D J and Brown J N 2020 Warming Patterns
 682 Affect El Niño Diversity in CMIP5 and CMIP6 Models *Journal of Climate* **33** 8237-
 683 60

- 1
2
3 684 Frieler K, Lange S, Piontek F, Reyer C P O, Schewe J, Warszawski L, Zhao F, Chini L,
4 685 Denvil S, Emanuel K, et al. 2017 Assessing the impacts of 1.5 °C global warming
5 686 – simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project
6 687 (ISIMIP2b) *Geosci. Model Dev.* **10** 4321-45
7 688 Hank T B, Bach H and Mauser W 2015 Using a remote sensing-supported hydro-
8 689 agroecological model for field-scale simulation of heterogeneous crop growth
9 690 and yield: Application for wheat in central Europe *Remote Sensing* **7** 3934-65
10 691 Hawkins E and Sutton R 2009 The potential to narrow uncertainty in regional climate
11 692 predictions *Bulletin of the American Meteorological Society* **90** 1095-107
12 693 Izaurrealde R C, Williams J R, McGill W B, Rosenberg N J and Jakas M C Q 2006
13 694 Simulating soil C dynamics with EPIC: Model description and testing against
14 695 long-term data *Ecological Modelling* **192** 362-84
15 696 Jägermeyr J, Müller C, Ruane A C, Elliott J, Balkovic J, Faye B, Franke J A, Folberth C,
16 697 Iizumi T, Khabarov N, et al. in prep. Climate change signal in agriculture
17 698 emerges earlier in new generation of projections
18 699 Jiang D, Hu D, Tian Z and Lang X 2020 Differences between CMIP6 and CMIP5 Models
19 700 in Simulating Climate over China and the East Asian Monsoon *Adv. Atmos. Sci.*
20 701 **37** 1102-18
21 702 Jones J W, Antle J M, Basso B, Boote K J, Conant R T, Foster I, Godfray H C J, Herrero
22 703 M, Howitt R E, Janssen S, et al. 2017 Brief history of agricultural systems
23 704 modeling *Agricultural Systems* **155** 240-54
24 705 Kimball B A, Boote K J, Hatfield J L, Ahuja L R, Stockle C, Archontoulis S, Baron C,
25 706 Basso B, Bertuzzi P, Constantin J, et al. 2019 Simulation of maize
26 707 evapotranspiration: An inter-comparison among 29 maize models *Agricultural*
27 708 *and Forest Meteorology* **271** 264-84
28 709 Liu W, Yang H, Folberth C, Wang X, Luo Q and Schulin R 2016 Global investigation of
29 710 impacts of PET methods on simulating crop-water relations for maize *Agricultural*
30 711 *and Forest Meteorology* **221** 164-75
31 712 Maiorano A, Martre P, Asseng S, Ewert F, Müller C, Rötter R P, Ruane A C, Semenov M
32 713 A, Wallach D, Wang E, et al. 2017 Crop model improvement reduces the
33 714 uncertainty of the response to temperature of multi-model ensembles *Field Crops*
34 715 *Research* **202** 5-20
35 716 McSweeney C F and Jones R G 2016 How representative is the spread of climate
36 717 projections from the 5 CMIP5 GCMs used in ISI-MIP? *Climate Services* **1** 24-9
37 718 Meehl G A, Senior C A, Eyring V, Flato G, Lamarque J-F, Stouffer R J, Taylor K E and
38 719 Schlund M 2020 Context for interpreting equilibrium climate sensitivity and
39 720 transient climate response from the CMIP6 Earth system models *Science*
40 721 *Advances* **6** eaba1981
41 722 Minoli S, Müller C, Elliott J, Ruane A C, Jägermeyr J, Zabel F, Dury M, Folberth C,
42 723 François L, Hank T, et al. 2019 Global Response Patterns of Major Rainfed
43 724 Crops to Adaptation by Maintaining Current Growing Periods and Irrigation
44 725 *Earth's Future* **7** 1464-80
45 726 Mueller N D, Gerber J S, Johnston M, Ray D K, Ramankutty N and Foley J A 2012
46 727 Closing yield gaps through nutrient and water management *Nature* **490** 254-7
47 728 Müller C, Elliott J, Chryssanthacopoulos J, Arneth A, Balkovic J, Ciais P, Deryng D,
48 729 Folberth C, Glotter M, Hoek S, et al. 2017 Global gridded crop model evaluation:
49 730 benchmarking, skills, deficiencies and implications *Geoscientific Model*
50 731 *Development* **10** 1403-22
51 732 Müller C and Robertson R 2014 Projecting future crop productivity for global economic
52 733 modeling *Agric. Econ.* **45** 37-50
53 734 Nelson G C, Valin H, Sands R D, Havlík P, Ahammad H, Deryng D, Elliott J, Fujimori S,
54 735 Hasegawa T, Heyhoe E, et al. 2014 Climate change effects on agriculture:
55 736 Economic responses to biophysical shocks *Proceedings of the National*
56 737 *Academy of Sciences* **111** 3274-9
57
58
59
60

- 1
2
3 738 O'Neill B C, Tebaldi C, van Vuuren D P, Eyring V, Friedlingstein P, Hurtt G, Knutti R,
4 739 Kriegler E, Lamarque J F, Lowe J, et al. 2016 The Scenario Model
5 740 Intercomparison Project (ScenarioMIP) for CMIP6 *Geosci. Model Dev.* **9** 3461-82
6 741 Olin S, Schurgers G, Lindeskog M, Wårlind D, Smith B, Bodin P, Holmér J and Arneith A
7 742 2015 Modelling the response of yields and tissue C : N to changes in
8 743 atmospheric CO₂ and N management in the main wheat regions of western
9 744 Europe *Biogeosciences* **12** 2489-515
10 745 Palosuo T, Kersebaum K C, Angulo C, Hlavinka P, Moriondo M, Olesen J E, Patil R H,
11 746 Ruget F, Rumbaur C, Takáč J, et al. 2011 Simulation of winter wheat yield and
12 747 its variability in different climates of Europe: A comparison of eight crop growth
13 748 models *European Journal of Agronomy* **35** 103-14
14 749 Portmann F T, Siebert S and Döll P 2010 MIRCA2000-Global monthly irrigated and
15 750 rainfed crop areas around the year 2000: A new high-resolution data set for
16 751 agricultural and hydrological modeling *Global Biogeochemical Cycles* **24** Gb1011
17 752 Riahi K, Rao S, Krey V, Cho C, Chirkov V, Fischer G, Kindermann G, Nakicenovic N and
18 753 Rafaj P 2011 RCP 8.5—A scenario of comparatively high greenhouse gas
19 754 emissions *Climatic Change* **109** 33-57
20 755 Rosenzweig C, Elliott J, Deryng D, Ruane A C, Müller C, Arneith A, Boote K J, Folberth
21 756 C, Glotter M, Khabarov N, et al. 2014 Assessing agricultural risks of climate
22 757 change in the 21st century in a global gridded crop model intercomparison
23 758 *Proceedings of the National Academy of Sciences* **111** 3268-73
24 759 Rosenzweig C, Jones J W, Hatfield J L, Ruane A C, Boote K J, Thorburne P, Antle J M,
25 760 Nelson G C, Porter C, Janssen S, et al. 2013 The Agricultural Model
26 761 Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies
27 762 *Agricultural and Forest Meteorology* **170** 166-82
28 763 Rosenzweig C, Ruane A C, Antle J, Elliott J, Ashfaq M, Chatta A A, Ewert F, Folberth C,
29 764 Hathie I, Havlik P, et al. 2018 Coordinating AgMIP data and models across
30 765 global and regional scales for 1.5°C and 2.0°C assessments *Philosophical
31 766 Transactions of the Royal Society A: Mathematical, Physical and Engineering
32 767 Sciences* **376**
33 768 Rötter R P, Carter T R, Olesen J E and Porter J R 2011 Crop-climate models need an
34 769 overhaul *Nature Clim. Change* **1** 175-7
35 770 Ruane A C, Antle J, Elliott J, Folberth C, Hoogenboom G, Mason-D'Croz D, Müller C,
36 771 Porter C, Phillips M M, Raymundo R M, et al. 2018 Biophysical and economic
37 772 implications for agriculture of +1.5° and +2.0°C global warming using AgMIP
38 773 Coordinated Global and Regional Assessments *Climate Research* **76** 17-39
39 774 Ruane A C, Goldberg R and Chryssanthacopoulos J 2015 Climate forcing datasets for
40 775 agricultural modeling: Merged products for gap-filling and historical climate series
41 776 estimation *Agricultural and Forest Meteorology* **200** 233-48
42 777 Schauburger B, Rolinski S and Müller C 2016 A network-based approach for semi-
43 778 quantitative knowledge mining and its application to yield variability
44 779 *Environmental Research Letters* **11** 123001
45 780 Taylor K E, Stouffer R J and Meehl G A 2012 An Overview of CMIP5 and the
46 781 Experiment Design *Bulletin of the American Meteorological Society* **93** 485-98
47 782 Tokarska K B, Stolpe M B, Sippel S, Fischer E M, Smith C J, Lehner F and Knutti R
48 783 2020 Past warming trend constrains future warming in CMIP6 models *Science
49 784 Advances* **6** eaaz9549
50 785 Toretí A, Deryng D, Tubiello F N, Müller C, Kimball B A, Moser G, Boote K, Asseng S,
51 786 Pugh T A M, Vanuytrecht E, et al. 2020 Narrowing uncertainties in the effects of
52 787 elevated CO₂ on crops *Nature Food* **1** 775-82
53 788 van Vuuren D, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt G, Kram
54 789 T, Krey V, Lamarque J-F, et al. 2011 The representative concentration pathways:
55 790 an overview *Climatic Change* **109** 5-31
56 791 von Bloh W, Schaphoff S, Müller C, Rolinski S, Waha K and Zaehle S 2018
57 792 Implementing the nitrogen cycle into the dynamic global vegetation, hydrology,

- 1
2
3 793 and crop growth model LPJmL (version 5.0) *Geoscientific Model Development*
4 794 **11** 2789-812
- 5 795 Wang X, Zhao C, Müller C, Wang C, Ciais P, Janssens I, Peñuelas J, Asseng S, Li T,
6 796 Elliott J, et al. 2020 Emergent constraint on crop yield response to warmer
7 797 temperature from field experiments *Nature Sustainability*
- 8 798 Williams K, Gornall J, Harper A, Wiltshire A, Hemming D, Quaife T, Arkebauer T and
9 799 Scoby D 2017 Evaluation of JULES-crop performance against site observations
10 800 of irrigated maize from Mead, Nebraska *Geosci. Model Dev.* **10** 1291-320
- 11 801 Wirsenius S 2000 Human Use of Land and Organic materials. In: *Department of*
12 802 *Physical Resource Theory*, (Göteborg, Sweden: Chalmers University of
13 803 Technology and Göteborg University) p 255
- 14 804 Wyser K, van Noije T, Yang S, von Hardenberg J, O'Donnell D and Döschner R 2020 On
15 805 the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6
16 806 *Geosci. Model Dev.* **13** 3465-74
- 17 807 Xin X, Wu T, Zhang J, Yao J and Fang Y 2020 Comparison of CMIP6 and CMIP5
18 808 simulations of precipitation in China and the East Asian summer monsoon
19 809 *International Journal of Climatology* **early online**
- 20 810 Zabel F, Müller C, Elliott J, Minoli S, Jägermeyr J, Schneider J, Franke J, Moyer E, Dury
21 811 M, Francois L, et al. under review Large potential for crop production adaptation
22 812 depends on available future varieties *Science Advances*
- 23 813 Zhu P, Zhuang Q, Archontoulis S V, Bernacchi C and Müller C 2019 Dissecting the
24 814 nonlinear response of maize yield to high temperature stress with model-data
25 815 integration *Global Change Biology* **25** 2470-84
- 26 816 Zhu Y-Y and Yang S 2020 Evaluation of CMIP6 for historical temperature and
27 817 precipitation over the Tibetan Plateau and its comparison with CMIP5 *Advances*
28 818 *in Climate Change Research*
29 819
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60