https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

Limited reducibility of CMIP6 ensemble size to bracket crop climate impacts globally

Christian Folberth[1], Artem Baklanov[2], Nikolay Khabarov[2], Thomas Oberleitner[1], Juraj Balkovic[1]

[1]Biodiversity and Natural Resources Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

[2]Advancing Systems Analysis Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

Correspondence to: Christian Folberth (folberth@iiasa.ac.at)

Abstract: Global crop–climate impact studies using gridded, process-based models are essential for understanding how climate change may affect agricultural productivity. Due to high computational demand, these models are typically driven by a limited subset of projections from up to five global climate models (GCMs) from CMIP experiments. Yet, there is limited understanding of whether these subsets adequately capture variability in climate drivers within the whole ensemble and how the captured variability in drivers translates into crop yield impacts. Here, we analyze climate and crop yield projections from 29 GCMs contributing to CMIP6 ScenarioMIP that provide all required forcing variables and have been bias-adjusted and downscaled. To address computational constraints, we apply a high-performant crop model emulator to estimate yields of four staple crops across 150k global locations under the four core emission pathways and for the whole climate ensemble from 1984–2099. We then aggregate results to major regions and use linear programming with the objective to minimize ensemble sizes while bracketing projected changes in temperature, precipitation, or crop yields. I.e., we identify the smallest GCM subsets that capture each variables' full range. We find that ensemble sizes required to bracket ranges for individual variables increase from temperature (around 5–20 GCMs) to precipitation and further to crop yields (about 10–29 GCMs each), depending on the selection of crops and scenarios, and the strictness of the boundaries. Capturing yield responses across contrasting crops and emission pathways commonly requires the full ensemble. Notably, bracketing climate variables alone fails to capture the range of yield outcomes, highlighting the limited predictive value of individual climatic drivers. Our results emphasize the need for thorough contextualization of GCM selection in impact studies and suggest that emulators can fill a gap in complementing mechanistic model simulations with comprehensive climate uncertainty quantification.

1 Introduction

Forcing process-based impact models with projections from global climate models (GCMs) is a central approach to forward-looking climate impact and adaptation planning and to inform policymaking for climate mitigation (Bezner Kerr et al., 2022). Most global process-based impact modelling - spanning sectors such as hydrology, agriculture, health, and energy systems (Frieler et al., 2024) - covers only a small fraction of climate projections

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

due to substantial computational constraints. In contrast, the number of GCMs participating in key exercises providing such projections, most prominently the climate model intercomparison project (CMIP, presently in its phase CMIP6), has grown over the past decades to presently several dozen ensemble members. These diverge substantially in their conceptualization and representation of Earth System processes, their configurations of modules, and consequently projections of future climate along anthropogenic and natural greenhouse gas (GHG) forcings. In CMIP6 ScenarioMIP these are defined by Shared Socio-economic Pathways (SSPs) as drivers of Representative Concentration Pathways (RCPs), commonly combined into SSP-RCPs (Meinshausen et al., 2020) covering four core scenarios with increasing GHG forcings, namely SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 besides several side tracks (Riahi et al., 2017). For global impact studies, it is common practice to select 3 to 5 GCMs and two concentration scenarios bracketing highest and lowest emissions (SSP5-8.5 and SSP1-2.6) or worst case and business-as-usual scenarios (SSP5-8.5 and SSP2-4.5).

The selection of GCMs is far from trivial with 39 members originally contributing to the CMIP6 scenario track (Tebaldi et al., 2021). Various methods have been proposed for GCM sampling or weighting, typically to limit computational requirements while maintaining robustness in the first case or to arrive at more robust ensemble responses in the latter (Katzenberger et al., 2026). Merrifield et al. (2023) for example provide a toolbox with various skill- and independence-based metrics to identify optimal sub-ensembles for user defined study regions. Brunner et al. (2020) applied similar metrics to weight GCMs within the ensemble as a transient approach rather than binary selection. A very common approach to select GCMs in terms of their robustness is to first exclude models that are considered to be too sensitive to increases in atm. CO_2 compared to observed past changes, which results in more conservative warming estimates both globally and regionally (Cannon, 2024; Hausfather et al., 2022; Nijsse et al., 2020; Tokarska et al., 2020). Extending this approach to precipitation, Chai et al. (2021) used observed sensitivity of precipitation change to temperature change to narrow down the likely range of changes in runoff across CMIP6 projections. In the face of these contrasting approaches to GCM selection, it needs to be noted that there is so far no commonly accepted method to climate ensemble sub-selection and any criteria chosen are eventually normative (Sobolowski et al., 2025).

In global impact modelling, the most widely used selections of GCMs in recent years are the five-member sets provided by the Intersectoral Impact Model Intercomparison Project (ISIMIP) phase 2b based on CMIP5 and phase 3b based on CMIP6. In both phases, GCMs were selected based on timeliness and completeness of available data, bracketing of a wide range of global changes in temperature and precipitation across the ensemble, as well as structural independence and skill of models (Lange, 2021). Limited research has addressed how these and other GCM sub-ensembles relate to wider CMIP projections. Two studies evaluated the ranges of changes in temperature and precipitation covered by ISIMIP phase 2b experiments compared to the CMIP5 ensemble and found that the bracketing of temperature is reasonable but that of precipitation often rather limited across major world regions (Ito et al., 2020; McSweeney and Jones, 2016). Both also found that a larger number of GCMs would be required to bracket precipitation compared to temperature. Along the same lines, (Herger et al., 2018) combined different approaches to select GCM sub-ensembles from CMIP5 for two SSP-RCPs separately based on skill metrics and coverage of the ranges in selected climate variables, finding that the optimal selection is sensitive to the selected metrics.

Few studies have evaluated how GCM selection affects impact projections in the land domain and cropping systems specifically. Teckentrup et al. (2023) conducted an extensive study modelling the land carbon cycles

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

across Australia while testing various GCM sampling, averaging, and bias-adjustment strategies out of a 21 member CMIP6 ensemble. They found that bias-adjustment methods and ensemble averaging play substantial roles in reducing projection uncertainty. Using an 18 member CMIP6 ensemble, Swaminathan et al. (2024) found that removing GCMs with a transient climate response (TCR, an indicator for GCMs' CO_2 sensitivity) considered unlikely hardly affects the occurrence of extreme climate events such as atmospheric droughts. Thereby, the ISIMIP phase 3b ensemble was largely representative for the evaluated impact metrics. Müller et al. (2021) used statistical impact emulators for an ensemble of Global Gridded Crop Models (GGCMs; Franke et al., 2020b) to compare ranges of impact projections based on the CMIP6 and CMIP5 ensembles finding a trend towards more adverse projections between these CMIP phases.

Thus far, no study has explicitly bridged the chain from global and regional changes in key climate variables over key climatic impact drivers to actual impacts, i.e., crop yields in agricultural modelling, to evaluate how ensemble sizes can be optimized on the one hand and how sampling strategies based on climate variable ranges translate into impacts on the other. While earlier research has addressed the bracketing of temperature and precipitation ranges alone it has neglected so far that complex mechanistic models account for further climate variables, interactions among changes in climate variables, and sub-seasonal patterns. Also the behavior of GCM ensembles and resulting impacts along emission scenarios has thus far received little attention with a common focus on the high emission scenario SSP5-8.5 that is as of now considered a less likely pathway (Hausfather, 2025; Wynes et al., 2024) and only one of several emission pathways of relevance for impact studies. In line with these gaps, this study addresses the three overarching research questions

1. In how far can the number of CMIP6 ensemble members be reduced while bracketing ranges in key climate attributes and yield impacts?
2. How is this affected by the choice of cropping seasons and emission pathways?
3. How does the bracketing of climate attributes translate into crop yield impacts?

To address these questions, we force a machine-learning based crop model emulator (Folberth et al., 2025a) trained on the global gridded crop model EPIC-IIASA (Balkovič et al., 2014) with bias-adjusted downscaled climate projections from 29 CMIP6 ensemble members for four SSP-RCP combinations (Thrasher et al., 2022) and the four major row crops wheat, maize, rice, and soybean. We then evaluate how GCM sampling based on climate variables ex ante or crop yields ex post affects ranges in outcomes covered globally and across major world regions. This is also done for a subset of GCMs with TCR in the IPCC likely range (Hausfather et al., 2022). Eventually, we put the set of GCMs from ISIMIP phase 3b commonly used in impact studies in the context of the whole ensemble to evaluate in how far impact ranges are covered for contrasting crops.

2 Methods

2.1 Study design

The overarching objective of the study is to evaluate to what extent the CMIP6 ensemble can be reduced with the objective of bracketing most extreme (highest and lowest) projections of changes in key climate variables and crop yields. To overcome computational challenges posed by applying a process-based crop model across comprehensive climate projections, we use a high performant crop model emulator framework. The experiment

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

is organized in three main blocks addressing (A) the training of crop model emulators for yield predictions within the CMIP6 domain (see sect. 2.2), (B) application of these emulators to a comprehensive set of CMIP6 climate projections (see sect. 2.3), and (C) the evaluation of climate features and crop yield impacts including in how far CMIP6 projections can be bracketed by smaller sub-ensembles (Figure 1; see sect. 2.4 to 2.6).

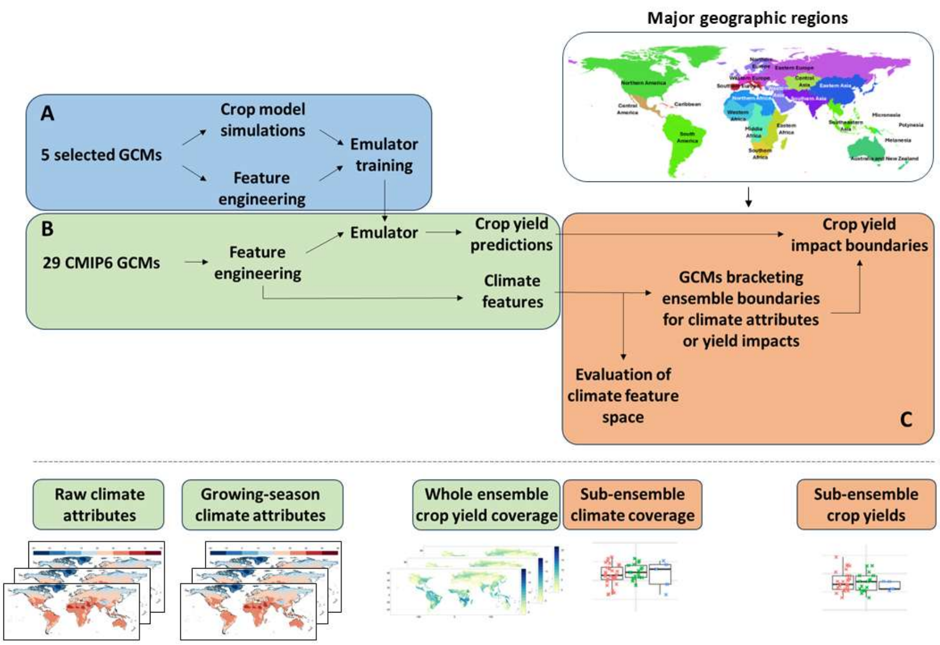


Figure 1. Study design schematic organized in three key blocks covering (A) training of crop model yield emulators for the CMIP6 application domain based on five GCMs, (B) feature engineering and crop yield predictions for the application domain of 29 GCMs, and (C) evaluation of CMIP6 ensemble climate feature characteristics, crop yield predictions, and feasibility to bracket ensemble boundaries in both climate and crop yield impacts for major geographic regions (see Figure S 1 for larger map). Items below the dashed line indicate key data produced along the chain, ranging from spatially explicit raw climate attributes (e.g., annual mean temperature over land) over growing-season climate (e.g., area-weighted maize growing season temperature) to crop yield predictions for the whole climate ensemble and eventually outcomes for ensembles bracketing climate attributes or crop yields across major geographic regions.

Results are structured in four sections: (1) As a background on the ensemble characteristics and outcomes, we provide global aggregate analyses of temperature, precipitation, and crop yield changes for selected crops. (2) We subsequently quantify the required ensemble compositions to bracket crop growing season and annual temperature and precipitation ranges of the CMIP6 ensemble across major geographic regions. (3) The same is performed for crop yield impacts directly. (4) Eventually, we evaluate how the bracketing of climate features translates into crop yield impacts and how those relate to the actual range of outcomes across the whole ensemble. Eventually, we contextualize ranges of the whole ensemble by comparing to those of the frequently used GCM selection provided by ISIMIP phase 3b.

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

To reduce the complexity of presenting outcomes across GCMs, SSP-RCPs, various variables, and crops, we focus on maize and wheat (combined spring and winter wheat or the latter only depending on the scope) as two widely cultivated crops with contrasting seasons, while providing essential information also on soybean and rice. Following a climate-impact-driver approach (Ruane et al., 2022), we focus on changes in growing season climate change particularly but also compare this to annual mean changes in climate variables as a common approach to GCM selection.

2.2 Crop model emulator, climate projections and other forcing data

The experiment conducted herein requires producing annual yield estimates for 24-29 GCMs depending on the SSP, four SSP-RCPs and the historical time period, six crops and seasons, rainfed and sufficiently irrigated water supply, on average 90 years per scenario, and about 150k location, corresponding to around 9 x 109 simulations. To cope with the large volume of crop yield simulations, we make use of a novel high accuracy and performance emulator CROp model Machine learning Emulator Suite (CROMES) (Folberth et al., 2025a). This software package serves for predicting crop yields simulated by global gridded crop models (GGCMs) by training machine-learning algorithms on forcing data and simulation outputs. Mimicking the way GGCMs simulate crop growth, daily climate data are aggregated dynamically for each growing season and cardinal sub-seasons based on growing season timing and growing degree day accumulation for each location and year. CROMES has been shown to reproduce the GGCM’s yield estimates with very high accuracy for climate projections unseen during training (Folberth et al., 2025a), qualifying it specifically for the application on climate projection uncertainty herein. In training emulators, we follow the exact same approach as in the cited study but pool training data for all five core GCMs and four SSP-RCPs for which EPIC-IIASA simulations have been produced. We assume sufficient nutrient supply in both training data simulations and predictions to focus on the climate signal.

Climate features required by the emulator and simultaneously serving as diagnostic information are produced for the whole growing season, the emergence phase (1st quarter of growing season), and the reproductive phase (2nd half of the growing season). They comprise aggregates of raw climate data as well as derivatives including potential evapotranspiration based on Penman-Monteith, climatic moisture deficit, and numbers of days crossing specific thresholds such as wet and dry days or heating degree days. Details are provided in (Folberth et al., 2025a), the code is available in a public repository (Folberth et al., 2025b). Following a climate-driver-impact approach (Ruane et al., 2022), we evaluate herein changes in (a) global mean climate variables calculated as annual arithmetic mean changes in key climate variables over the whole land mask as commonly used indicators for comparison of GCM responses and (b) changes in climate variables over crop-specific dynamic growing seasons and cropland as estimated by the above climate feature pipeline.

Crop yield estimates are based on the field-scale crop model Environmental Policy Integrated Climate v0810 (EPIC; (Izaurralde et al., 2006; Williams et al., 1989)) and the geo-spatial infrastructure of the global gridded crop model EPIC-IIASA, which is based on a global spatial base resolution of 5 arcmin x 5 arcmin (Balkovič et al., 2014). These base pixels are aggregated to simulation units reflecting similarities in soil, topography, management, and climate, which can have a maximum resolution of 30 arcmin x 30 arcmin (Skalský et al., 2008). Based on general suitability, EPIC-IIASA simulates cropping systems for approx. 150k simulation units globally, treating each as a homogenous field based on the input data. EPIC-IIASA has been positively evaluated against yield observations and is typically located around the ensemble mean response to climate change (Jägermeyr et

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

al., 2021a; Müller et al., 2017, 2024). It has also shown positive performance at the site scale when applied with higher resolution forcing data (Balkovič et al., 2020; Couëdel, 2024). To produce a training sample for the crop model emulator, we force EPIC-IIASA with the five ISIMIP priority GCMs (Frieler et al., 2025; Jägermeyr et al., 2021a; Lange, 2021) and for all four core SSP-RCPs (see sect. 2.3) We simulate six key crops and seasons commonly used in ISIMIP impact assessments, namely maize, rice (main season, second season), soybean, spring wheat, and winter wheat using crop calendars provided by GGCMI (Jägermeyr et al., 2021b). The two wheat types are considered spatially mutually exclusive and mapped to a combined wheat crop using a mask from the same source. The two rice seasons area weighted by seasonal harvested area extent according to (Laborte et al., 2017).

2.3 Climate projections

Target climate projections were obtained from the NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP-CMIP6; Thrasher et al., 2022). We selected all projections that provide the required climate quantities min. and max. temperature, precipitation, wind speed, shortwave solar radiation, and relative humidity, resulting in a total of 24-29 GCMs or GCM configurations, respectively, depending on the SSP-RCP (Table S 1). To study how ensemble behavior relates to emission scenarios, we conduct all evaluations for the four core SSP-RCPs SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5. To provide a general characterization of the ensemble, Figure S 2 shows global changes in temperature and precipitation for all GCMs and SSPs over the land mask and for the growing season and harvested area of each crop.

To account for the ‘hot model problem’, i.e., the fact that some climate models respond more sensitive to atm. CO2 increases than expected (Hausfather et al., 2022; Tebaldi et al., 2021; Tokarska et al., 2020), we complementary evaluate a sub-ensemble of GCMs with a transient climate response (TCR) considered to be in the likely range proposed earlier by the IPCC, referred to as TCR-likely hereafter (Hausfather et al., 2022). While this presents just one method among many for how to define an expectedly more robust climate ensemble, it allows for exemplary insights, how such constraints translate into ensemble properties and outcomes. The 17 GCMs remaining after this filtering are indicated in Table S 1.

To contextualize the findings, we compare impacts across the whole ensemble with those derived from an sub-ensemble frequently used in climate impact modelling, the five GCMs (GFDL-ESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR, MRI-ESM2-0, UKESM1-0-LL) provided by ISIMIP (Frieler et al., 2025). This is done using these GCMs’ bias-corrected climate projections from NEX-GDDP-CMIP6 for consistency.

2.4 Post-processing and spatial aggregation of crop yield and climate features

Prior to evaluation, gridded crop yield estimates and climate features are aggregated for macro-regions, i.e., globally and for 20 major geographic sub-regions defined by the United Nations Statistics Division (UNSD) geo-scheme (UNSD, 1996), UN regions hereafter, that provide sufficient data coverage (see Figure S 1 for map). These or similar regions are the scale of aggregation commonly used in global economic modelling and global agro-economic land use foresight, both of which frequently rely on crop model yield predictions (Stehfest et al., 2019). It is therefore considered herein an adequate functional scale of spatial aggregation to inform the relevance of outcomes for food security modelling chains. Aggregation is done using area weighting. Harvested areas for the studied crops per 5 arcmin x 5 arcmin pixel were obtained from the Spatial Production Allocation Model

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

(SPAM) 2010 v2.0 dataset (IFPRI, 2020; Yu et al., 2020). The two types of wheat and the two rice seasons simulated independently but relying on the same harvested areas are combined as described in sect. 2.2.

2.5 Identification of minimum ensemble size required to bracket climate features or yield impacts

We use linear programming to identify the minimum number of GCMs required to bracket the ranges of the NEX-GDDP-CMIP6 ensemble for key climate attributes or crop yields themselves. That is, we identify for a defined number of SSP-RCPs and crops the minimum and maximum values for an attribute globally or per UN region and subsequently minimize the ensemble size while maintaining these boundaries. Besides full coverage of such boundaries we allow some flexibility at the fringes using ε-bracketing to evaluate in how far marginal differences among GCM projections may affect results. In this case, the selected sub-ensemble must approximate both the lower and upper extremes of the full range at every location within a defined ε, a pre-defined fraction of the range that is permitted not to be bracketed. Herein, we use values of 0 (no tolerance), 0.1, and 0.2. This approach translates into a minimum cardinality set cover problem.

For a selected variable (a single climate feature or crop yield) A from model i in pixel j the range is bracketed by the minimum $m_j = min_i\, A_{i,j}$ and the maximum $M_j = max_i\, A_{i,j}$. The eligible GCM subsets per location are accordingly defined as

$$L_j(\varepsilon) = \left\{i\colon A_{i,j} \leq m_j + \frac{\varepsilon}{2}\right\},\quad U_j(\varepsilon) = \{i\colon\ A_{i,j} \geq M_j - \frac{\varepsilon}{2}\} \tag{1}$$

The optimization problem is to minimize the number of selected models subject to covering both ends within 0.5ε tolerance at every location according to

$$\min \textstyle\sum_i x_i \quad \text{s.t.} \quad \sum_{i \in L_j(\varepsilon)} x_i \geq 1, \sum_{i \in U_j(\varepsilon)} x_i \geq 1\ \ \forall j;\ \ x_i \in \{0,1\}. \tag{2}$$

For each location j, two covering constraints are imposed: (I) at least one selected model must fall within 0.5ε of the minimum m_j (hence serve as an approximate lower bracket), and (II) at least one must fall within 0.5ε of the maximum M_j (to be an approximate upper bracket). We chose the tolerance ε to be set uniformly, hence once matrix A^{nxm} is normalized column-wise to the range 0 to 1, 0.5ε becomes a uniform tolerance range fraction. The technical implementation is done with the PuLP linear programming toolkit for Python for high level model formulation (Mitchell et al., 2011) and the COIN-OR Branch-and-Cut solver (Forrest et al., 2026). Missing data (i.e., GCMs not present within an SSP-RCP) were filled with midpoints of the range, naturally removing missing models at location j from potential candidates to bracketing pairs at j.

2.6 Optimization of ensemble composition to bracket ensemble boundaries

While the prior section yields the minimum number of GCMs required to fully bracket ensemble boundaries as a top-down approach, we also estimate bottom-up what minimum coverage of variable ranges can be achieved across UN regions with a given number of GCMs. I.e., we apply a maximin approach that aims to maximize the minimum coverage for a variable (that is, the worst outcome) across all regions and therefore provides a best guaranteed coverage fraction.

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

This approach is implemented based on an input matrix A^{nxm} where rows correspond to n GCMs and columns to m locations. Entry $A_{i,j}$ represents the variable of interest - such as annual mean temperature - for GCM i in location j. In this context, locations are understood broadly as any distribution of interest indexed by GCMs. Depending on the scenario specifications, these also include combinations of SSP-RCPs. The full range FR of a variable at location j is

$$FR_j \coloneqq \max_i A_{i,j} - \min_i A_{i,j}. \tag{3}$$

For ensemble size k, we seek $R(k)$, the maximum worst-case fractional range covered across locations j according to

$$\mathrm{R(k)} = \max_{\mathrm{S:|S|=k}} \min_{\mathrm{j}} \frac{\max_{i \in S} A_{i,j} - \min_{i \in S} A_{i,j}}{\max_i A_{i,j} - \min_i A_{i,j}} \tag{4}$$

We convert this combinatorial problem to a 0–1 integer linear programming solvable to optimality via branch-and-bound along the following equations

$$\max \mathrm{R} \ \ s.t.$$

$$\sum_{\mathrm{i}} \mathrm{x_i} = \mathrm{k}; \quad \mathrm{x_i} \in \{0,1\} \tag{5}$$

$$\sum_{p,q} y_{j,p,q} = 1 \ \forall j; \ \ y_{j,p,q} \in \{0,1\} \tag{6}$$

$$y_{j,p,q} \leq x_p; \ y_{j,p,q} \leq x_q \quad \forall j,p,q \tag{7}$$

$$\sum_{p,q} y_{j,p,q} \left(A_{q,j} - A_{p,j} \right) \geq R \cdot FR_j \quad \forall j \tag{8}$$

where Eq. (5) selects exactly k climate models, Eq. (6) enforce one bracketing pair of climate models (p,q) per location j, and Eq. (7) ensures bracketing pairs use only selected models. To improve computational time, Eq. (7) can be equivalently streamlined to

$$\sum_q y_{j,p,q} \leq x_p \ \forall j,p \tag{7p}$$

$$\sum_p y_{j,p,q} \leq x_q \ \forall j,q \tag{7q}$$

Equation (8) requires each pair to cover at least fraction R of the full range FR_j (see Eq. (3)), with fraction R being maximized in the objective.

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

For the technical implementation, we use the same linear programming Python libraries as in sect. 2.5, i.e., PuLP (Mitchell et al., 2011) and the COIN-OR Branch-and-Cut solver (Forrest et al., 2026). Notably, for comparable targets, the relaxed coverage problem for the bracketing of ensemble ranges as in sect. 2.5 is more demanding than the maximin problem. I.e., a subset achieving maximin $R = 1 - \varepsilon$ may still fail ε-bracketing if its coverage is concentrated entirely on one side of the range. Missing data (i.e., GCMs not present within an SSP-RCP) were filled with midpoints of the range as above.

Eventually, for each GCM set optimized to bracket climate attributes at a given ensemble size, we quantify how these sub-ensembles translate into yield impact ranges by selecting the crop yield changes produced by the respective sub-ensemble. Comparing the resulting yield impact coverage to the direct bracketing of yield impacts informs in how far a climate variable is suitable to approximate changes in yield impacts.

3 Results

3.1 Global changes in temperature, precipitation and crop yield

Global aggregate changes in growing season precipitation, temperature, and crop yield estimates by the end of century vary greatly among GCMs, across SSPs, and by crop, here maize or combined (winter and spring) wheat (Figure 2). Which GCMs are located at the fringes of the distribution depends on the observed variable and dimension, and there are few cases in which individual GCMs show consistency in bracketing the ensemble. For temperature, for example, GCM 29 (see Table S 1 for GCM names) consistently shows the largest upward change in all scenarios except in SSP1-2.6 for maize, albeit being still very close to the top. GCMs 9, 16, and 17 have a tendency to lower growing season temperature increase but are not consistently bracketing the ensemble. For precipitation, the latter two tend to have the largest increases for maize but are in the mid-range for wheat where GCMs 3 and 7 show large increases and GCM 18 often a decrease under higher emission scenarios. In line with these changes in key climate variables, the ranking of GCMs in resulting yield changes has limited consistency across SSPs and can be highly divergent between crops. The most consistent GCMs are 28 and 29, which dominate low yields for maize and are located at the lower fringe for wheat. This is more so the case under higher warming scenarios where GCMs' temperature sensitivities play out more strongly. GCMs 16 and 17 result in the most positive outcomes for maize yields but are often in the mid-range for wheat.

Notably, for both growing season temperature change and changes in crop yields, results for maize show largely a correlation with GCMs' TCR values, especially in high emission scenarios. I.e., they scale with increasing TCR negatively (yield) or positively (growing season temperature). This is less the case for wheat, where some GCMs with low TCR are in the mid-range of growing season temperature change and crop yield changes show a highly divergent pattern with some of the GCMs with highest or lowest TCR at both ends of the yield change spectrum. Accordingly, there's limited with TCR for temperature change in most scenarios and merely a trend for crop yield.

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

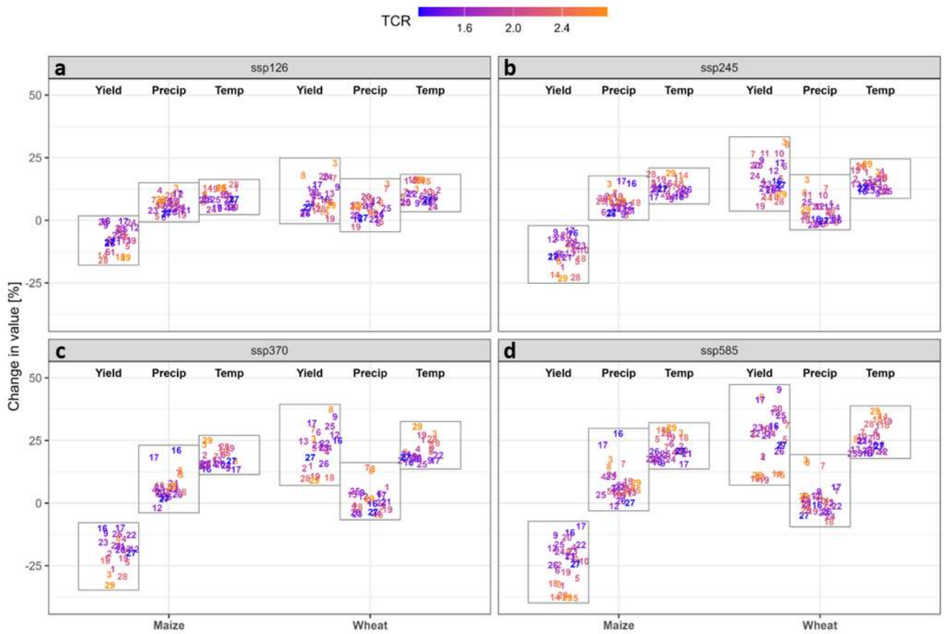


Figure 2. Global changes by end-of-century (period 2070-2099 vs 1984-2013) in area-weighted crop yield (Yield), growing season precipitation (Precip), and growing season temperature (Temp) for two contrasting crops, maize and combined spring and winter wheat, across the four SSP-RCPs and per GCM. Numbers serving here as symbols correspond to the column ID for GCMs in Table S 1.

The patterns observed in the global aggregate changes of yield, growing season precipitation and temperature are further exacerbated at the level of major geographic UN regions as shown for crop yield change only in Figure S 3 for maize and Figure S 4 for wheat. Already for the individual crops and within SSPs, hardly any GCMs provide consistently clearly high or low outcomes across regions. Several GCMs flip positions between regions such as GCMs 3 and 19 for maize in several SSPs, the first also for wheat. Furthermore, individual GCMs may also result in more positive or negative outcomes depending on the SSP such as GCM 1 in South Asia for wheat, ranking mid-range in ssp126 and constantly moving up along concentration pathways.

3.2 Requirement of GCMs to bracket climate and impact ranges

Estimating the minimum number of GCMs required to bracket ranges in growing season temperature, precipitation, or crop yield change across the whole GCM ensemble reflects varying levels of complexity in atmospheric and plant growth processes. This is further exacerbated by the coverage of different cropping seasons and SSPs on the one hand or relaxation of strictness of the extreme boundaries on the other (Figure 3). We find that, overall, ranges can be bracketed with fewest GCMs for temperature, requiring more GCMs for precipitation, and most for actual crop yields. For crop yields, the number of required GCMs is uniformly higher for SSP1-2.6 than for SSP5-8.5, indicating that it is more challenging to bracket outcomes for the lower emission scenario than the most extreme. The likely reason is that the strong global warming signal in SSP5-8.5 overrides regional noise

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

and leads to more stable GCM rankings, making the boundaries easier to identify with fewer models. Otherwise, the number of required GCMs increases with the number of SSPs accounted for (in line with theory). The impact of relaxing minimum and maximum boundaries again depends on the target variable in the same order as above. I.e., the effect is largest for temperature followed by precipitation and least pronounced for crop yields. This highlights that the complexity of crop yield responses materializes throughout the ensemble.

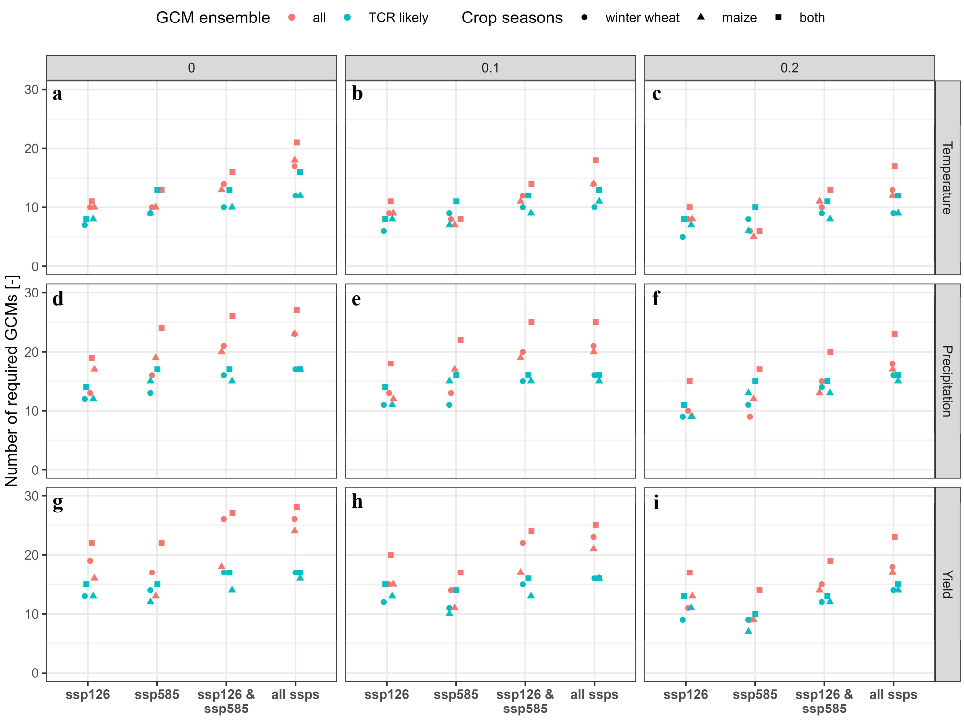


Figure 3. Minimum requirement for number of GCMs to bracket projected changes in area-weighted growing season temperature, precipitation, and crop yield (rows) for three levels of boundary relaxation (columns). I.e., values greater than zero allow for some flexibility if a GCM is close to the boundary to acknowledge that not necessarily the most extreme case needs to be selected.

Essentially, around 10 GCMs out of the whole ensemble are required to bracket ranges of temperature change in individual SSPs and crops but more than 15 for all SSPs (panel a). This increases to 16 and 24 GCMs for precipitation (panel d) and further to 22 to 29 GCMs (i.e., the whole ensemble) for crop yields (panel g) if both seasons, maize and winter wheat are considered. This reflects increasing complexity in the underlying physics where temperature increases across large regions are a primary effect in GCMs, precipitation is the outcome of chains of far more complex atmospheric processes and spatial patterns of land-atmosphere feedbacks, including also aerosol and cloud dynamics. Crop yields eventually integrate six climate variables at daily time steps with various interactions and spatio-temporal changes of individual variables' degree of influence.

For the sub-ensemble constrained by likely TCR, the numbers of GCMs are lower but scale with the ensemble size. I.e., this sub-ensemble has 17 members as compared to the 29 in the whole ensemble and all 17 are required

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

in the most comprehensive combination of crops and SSPs to bracket crop yields (panel g). Notably, further relaxation of the boundaries results in only a minor reduction of required GCMs by one to two models (panels h, i). Also, the selection of SSPs plays a minor role for this sub-ensemble compared to the whole set of GCMs.

3.3 Feasibility to bracket climate and impacts with incremental inclusion of GCMs

To evaluate how well a defined number of GCMs can bracket ranges of outcomes across the UN regions, we perform an optimization of GCM sub-ensemble composition bottom up with increasing numbers of GCMs with the objective of maximizing the minimum range fraction covered across regions by the respective GCM ensemble (Figure 4). In contrast to other statistics such as the median, which is shown complementary, this informs whether the selection misses out substantially on any of the regions.

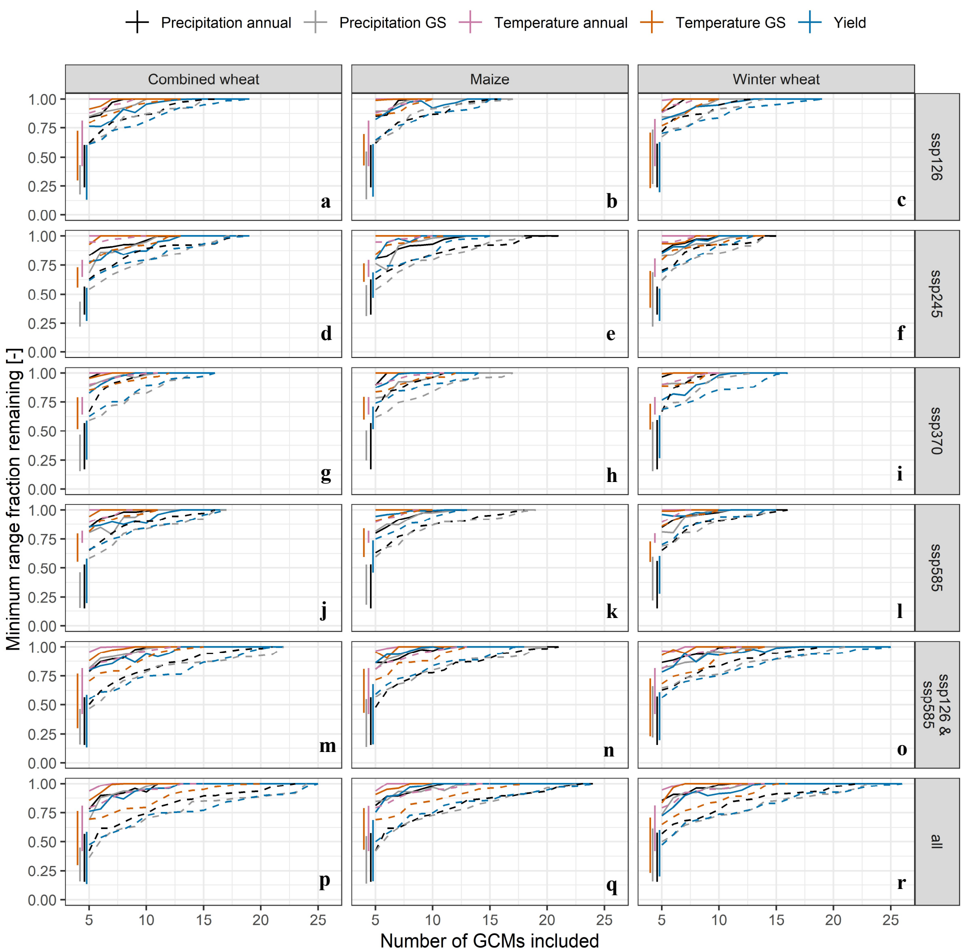


Figure 4. The median (solid line) and minimum (dashed line) range fraction (y-axis) of selected climate attributes and crop yield outcomes along an increasing number of GCMs (x-axis) for contrasting crops (columns) and emission scenarios (rows). Remaining minimum range fractions were quantified by linear

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

optimization with the objective of maximizing the lowest achievable range fraction covered across all UN regions (see sect. 2.6). E.g., a value of 0.4 indicates that across all UN regions at least this fraction is covered by the selected GCMs. Vertical lines at the left end of each panel show the minimum and median range covered by the five ISIMIP priority GCMs for each variable.

Across all scenarios and crops, the climate variable feasible to bracket with fewest GCMs is annual mean temperature over land with closure of minimum range fractions achieved (i.e., value of 1) already with less than 10 GCMs in all individual SSPs except SSP3-7.0, and less than 15 for the combined SSPs 1-2.6 and 5-8.5. This is closely followed by crop-specific growing season temperature, which requires up to 15 GCMs and shows some distinction between crops, requiring more GCMs for wheat in both cases where either winter wheat is evaluated alone or in combination with spring wheat. Notably, winter wheat, which has a very distinct growing season from summer crops with initial growth in autumn and a dormancy period over winter.
Precipitation is substantially more challenging to bracket, again for the growing season more so than annual changes. This requires up to 20 or more GCMs in SSP2-4.5 or the combined SSPs for maize (Figure 4e, n). For winter wheat typically less than 15 GCMs are required to fully bracket precipitation ranges, except in SSP5-8.5, but >15 for the combined wheat indicating that summer seasons are more complex to capture in a sub-ensemble. Notably, if only few GCMs are considered, precipitation indicators are the most challenging to bracket for maize and partly for the combined wheat, even more so than crop yields themselves (see below), and very dependent on SSPs with SSP2-4.5, SSP5-8.5, and the combined SSPs showing most pronounced effects. In these cases, precipitation in fact remains below crop yields along the trajectories of included GCMs, indicating that precipitation coverage is consistently poorer.
Maximizing the minimum coverage of crop yields themselves *ex post* shows accordingly divergent patterns between the two crops and seasons. For maize, crop yields are often less challenging to bracket than precipitation and more challenging than temperatures regardless of the emission scenarios if only few GCMs are included (Figure 4b, e, h, k, n). A minimum range fraction of 1 is achieved with less than 15 GCMs, except for the combined SSPs where nearly 20 are required. For wheat, crop yield is in turn more challenging to bracket or ranking equally with precipitation indicators. At least 15 GCMs are required throughout, except for winter wheat alone in the lower emission SSPs, and more than 20 for the combined SSPs (Figure 4m, o).

3.4 Ranges of crop yield outcomes resulting from bracketing of climate variables

Translating the bracketing of climate attributes into crop yield outcomes presents a contrasting picture and reveals no clear pattern which climate attributes best reflect crop yields (Figure 5). As annual and growing-season temperature range fractions are fully covered with lower numbers of GCMs (Figure 4), the climate-variable-derived coverage of yields plateaus at 10 to 15 GCMs with a range fraction of around 0.7 remaining for maize whereas growing season temperature is in most cases more informative than annual temperature. For wheat, more so winter wheat than for the combined types, the remaining minimum range fraction remains often around 0.5 and larger yield ranges can be covered where more GCMs are required to bracket the climate variables themselves. Annual mean temperature tends to in part outperform the growing season equivalent for combined wheat and low emission SSPs if less than 10 GCMs are included (Figure 5a, d). Precipitation allows overall for a larger fraction yield range coverage, in part due to the larger number of GCMs required to bracket precipitation uncertainties. At a given number of GCMs it depends on the SSPs whether temperature or precipitation provide better coverage. Typically, it's the first for higher emission pathways and the latter for lower, where growing season precipitation emerges as the best indicator of crop yield ranges for wheat (Figure 5a, c). Minimum range fractions larger than 0.8 can hardly be achieved through bracketing by any of the climate variables.

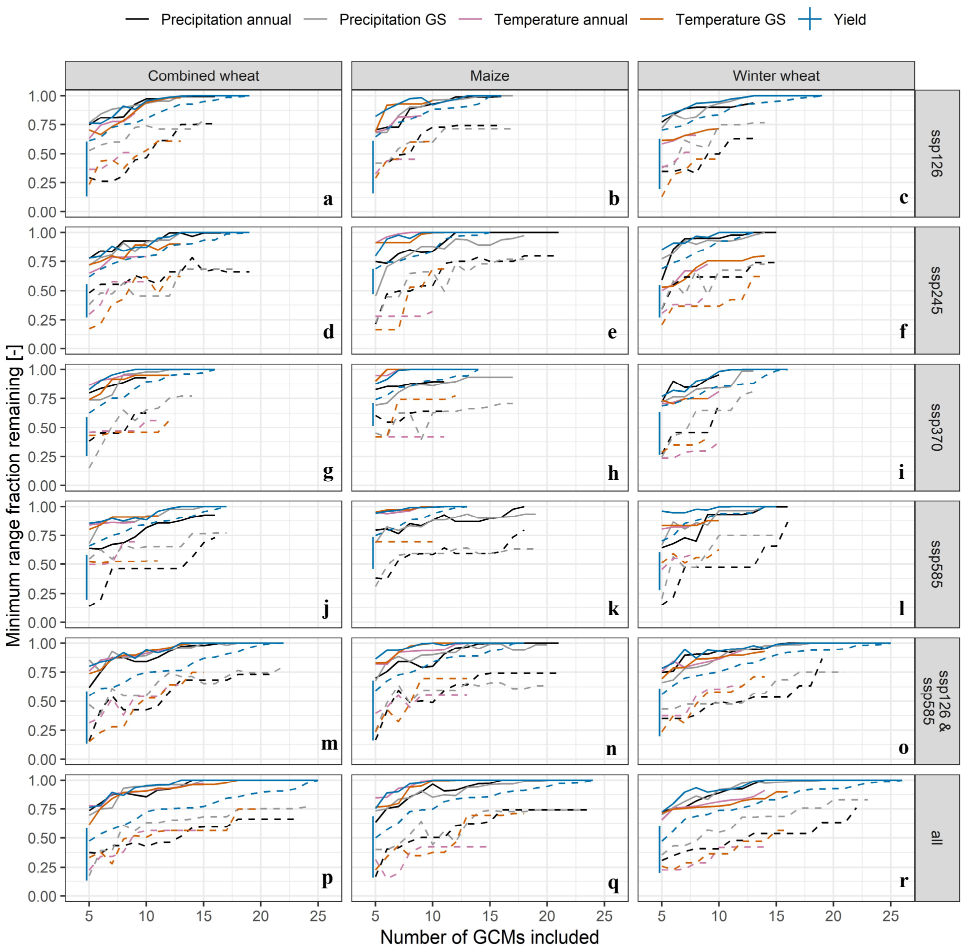


Figure 5. Same as Figure 4 but showing the mapped yield outcomes for each optimally selected ensemble of the target climate and impact variables in the earlier figure. Accordingly, ranges for crop yield coverage itself are identical to those in Figure 4 and only the yield range fraction is shown for the five ISIMIP GCMs at the left end of each panel.

Using the GCM sub-ensemble informed by TCR in the likely range slightly reduces the requirement of GCMs to bracket temperatures or leaves the number of required GCMs constant (Figure S 5). However, for SSP5-8.5 alone and the combined SSPs, the remaining range fraction is larger if only few GCMs are included. For the lower emission scenarios, also precipitation ranges can be covered more readily with less or equal 15 GCMs required for all crops. In SSP5-8.5 or the combined SSPs in turn, (nearly) the whole sub-ensemble of 17 GCMs is required

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

to bracket annual precipitation and slightly less for growing season precipitation. Crop yields again require typically fewer GCMs to bracket, except for the lower emission scenarios and wheat (compare Figure S 5 and Figure 5a, c, d, f). How the bracketing of climate indicators translates into crop yields changes most substantially for SSP5-8.5 (Figure S 6j, k, l). Here, the remaining minimum range fractions between crop yields directly and precipitation indicators start to converge from 10 GCMs on. Apart from that, no clear differences to the results derived from the complete ensemble emerge, considering that the constraint ensemble is substantially smaller. This is apparently the main driver in achieving an earlier closure of minimum range fractions.

3.5 Regional patterns in yield range fractions covered by climate variable bracketing

The prior results highlight the complexities and challenges in bracketing CMIP6 ensemble climate characteristics and crop yield outcomes. To illustrate spatial (in-)consistencies between the bracketing of specific climate features and crop yields, we map the fraction of instances in which the covered yield range is below 80% in major UN regions (Figure 6). This covers all SSPs and numbers of GCMs considered in Figure 5, using maize as an exemplary crop.

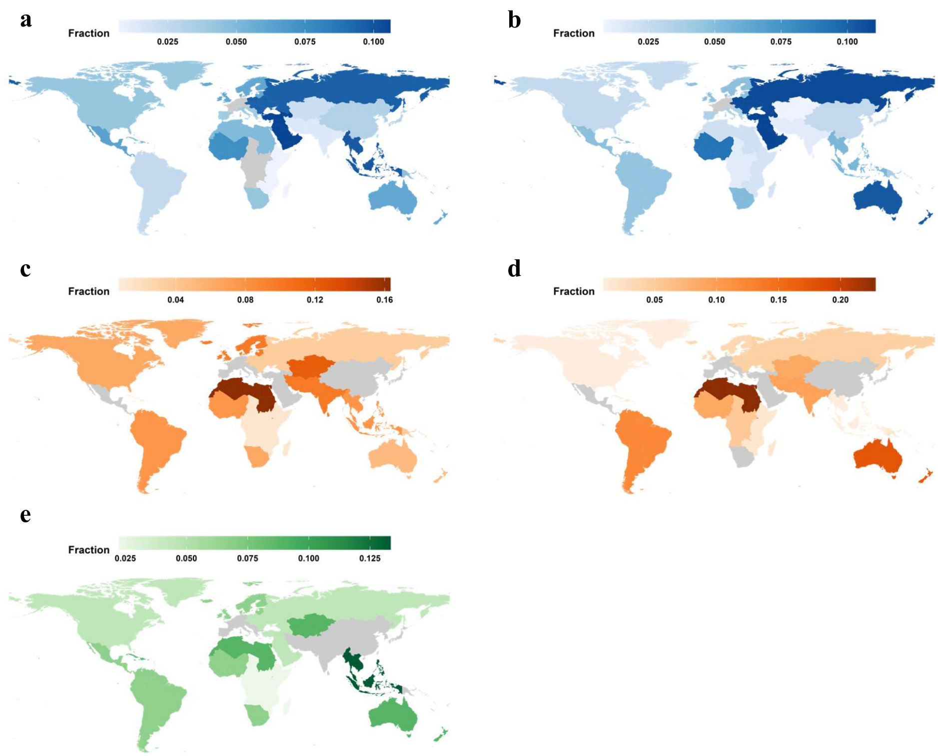


Figure 6. Fractions of instances in which <80% of the maize yield range is covered within each major geo-region when bracketing for specific climate variables with an increasing number of GCMs across SSPs. Essentially, this is a summary of Figure 5 panels b, e, h, and k using an arbitrary threshold of 80% for illustrative purposes. Panels herein show bracketing by (a) annual precipitation, (b) growing season

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

precipitation, (c) annual mean temperature, (d) growing season temperature, and (e) crop yield itself. Grey regions indicate that no values are under the threshold. See Figure S 7 for absolute numbers. See Figure S 1 for map of major geo-regions.

Bracketing crop yields directly following the maximin approach (i.e., maximizing the minimum covered yield range across regions) leaves Southeastern Asia the region most challenging to cover (Figure 6e). Yet, overall the number of instances with yield range fractions below 80% is low as already evident from Figure 5 where this threshold is solely undercut for maize yields across all SSPs if only 5-6 GCMs are considered (see also Figure S 7 for mapped absolute numbers). This consequently indicates a similar pattern across SSPs. Bracketing by annual or growing season precipitation sums and translating this into crop yield coverages (a, b) results in similar patterns among the two scenarios but shows substantial divergence from the crop yield bracketing (Figure 6e). While in the case of annual precipitation Southeastern Asia and to a lesser extent Oceania remain hotspots of lacking coverage, other regions become equally or more important such as Eastern Europe (incl. Russia), Western Asia, West Africa, and Central America. Among SSPs as well, results diverge here more substantially (not shown). Eastern Europe, for example, has several instances falling under the threshold within all SSPs, while Southeastern Asia falls into this category only for the lower emission scenarios and Western Africa for the two central ones SSP245 and SSP370. For both annual and growing season temperatures (Figure 6c,d), the pattern changes substantially, rendering Northern Africa a hotspot in total and across SSPs (not shown), followed to a lesser extent by Central Asia, South America, and Oceania. The latter especially for growing season temperatures and in lower emission scenarios.

These divergent patterns underpin that there are no particular regions that render it challenging to bracket climate impacts across the ensemble and SSPs, and different climate characteristics play out differently in individual regions. Noteworthy, in two predominantly temperate regions, Western Europe and Eastern Asia, crop yields are in most scenarios well bracketed suggesting that these are regions that can be reasonably robust represented even with smaller sets of GCMs selected based on climate features. However, it is also important to note that these results strongly depend on the methodology applied herein to cover crop yield impact ranges across regions through integer linear programming. Results of such an optimization approach are inherently also subject to "interactions" among regions and the overall objective to maximize the minimum yield range fraction at a given ensemble size, which prioritizes models helping the worst-performing region. The analysis is therefore not equivalent to a general assessment of the robustness of climate impacts across regions.

3.6 Crop yield outcomes covered by the ISIMIP priority GCMs

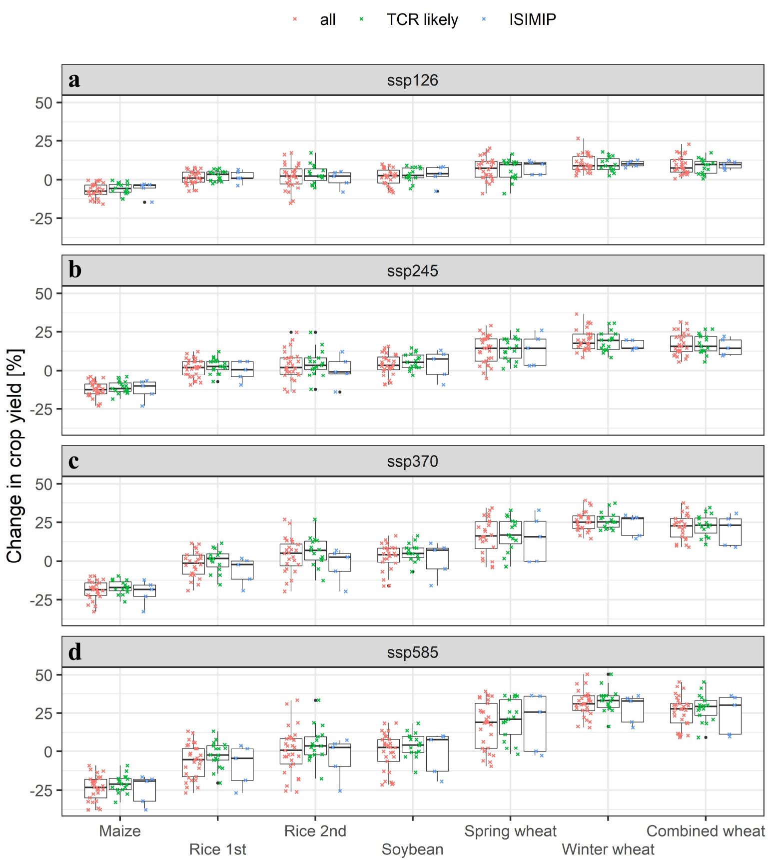
Practical implications of the limited feasibility to bracket the ensemble arise also for commonly used sub-ensembles such as the CMIP6 priority GCMs selected for ISIMIP phase3b. Outcomes in crop yield changes between this ensemble and the whole CMIP6 ensemble show that ensemble averages (here median) are less affected and consequently surprisingly robust but the range of potential outcomes can be substantially affected depending on crop and SSP-RCP (Figure 7).

Mean and median coverages of yield change ranges across all crops and SSPs by the ISIMIP ensemble compared to all GCMs are 65% and 70%, respectively, indicating overall reasonable representation of yield change ranges. We find the largest coverages on average for soybean and maize across SSPs with 84% and 81%, the lowest for second season rice and winter wheat with 56% and 41%. Among emission scenarios, SSP1-2.6 has on average

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

the lowest coverage (50%) and SSP3-7.0 the highest (75%). For all crops and scenarios, the lowest coverage occurs for winter wheat in SSP1-2.6 and the highest for soybean in SSP2-4.5. Notably, for spring wheat, the smaller ISIMIP ensemble solely indicates positive yield changes in SSP1-2.6 and SSP2-4.5, while the whole ensemble also shows potential negative outcomes. This is to a lesser extent also the case for the higher emission scenarios. Vice versa, the sub-ensemble tends towards substantially lower values for the upper end of the range for the two rice seasons and partly winter wheat under high emission scenarios suggesting a likely more adverse outcome compared to the whole ensemble.

We primarily include the TCR likely ensemble here for context as the selection of priority GCMs for ISIMIP do not account for such constraints and includes two GCMs (IPSL-CM6A-LR, UKESM1-0-LL) that are outside the likely TCR range (Table S 1). Consequently, in a nutshell the sub-ensemble with likely TCR results in further shift towards higher crop yield outcomes for most regions and crops.



505 **Figure 7. Relative changes in global area-weighted rainfed crop yields from 2070-2099 compared to 1984-2013 for seven crop types and seasons across four SSPs for three sets of CMIP6 GCM selection covering the whole ensemble (all), a sub-ensemble based on likely TCR (TCR likely), and the five priority GCMs used in ISIMIP.**

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

4 Discussion and conclusions

Our analyses indicate limited scope to optimize GCM ensemble composition towards a small subset while maintaining comprehensive coverage of outcomes. While this is achievable for basic climate variables to some extent it is hardly feasible for crop yields. Even where it applies to individual climate variables or emission pathways, as has been subject to earlier research (Ito et al., 2020; McSweeney and Jones, 2016), it becomes more and more challenging with increasing dimensionality of covering different growing seasons and emission scenarios (Figure 3; Figure 4). This also indicates that selecting GCM subsets based on narrow assumptions may be misleading regarding their representativeness for wider scenario variations; even more so when using high-end emission scenarios only for GCM selection, such as SSP5-8.5. The latter also obfuscates that it can be more challenging to bracket climate projections and impacts in lower emission pathways where individual GCMs may not decisively dominate globally in terms of temperature response or precipitation amounts.

Most importantly, bracketing of key climate variables does not go along with bracketing of yield impacts. Neither annual nor growing season aggregates for temperature or precipitation provide robust indications across crops and scenarios on how well crop yield change will be covered compared to the complete GCM ensemble. This fails for two reasons. First, complex non-linear impact functions as commonly implemented in mechanistic impact models do not in general preserve rank order. That is, bracketing an ensemble shows whether the members ranked lowest and highest in the climate space are also ranked lowest and highest in the impact space, which fails whenever the impact function re-orders the ensemble — a situation that arises generically for threshold-based or unimodal relationships between climate and impact. This is also evident in the fact that precipitation is in some cases more challenging to bracket than crops yields, which is likely due to the limited sensitivity of the applied crop model – and most other GGCMs – to excess precipitation (Oberleitner et al., 2025). Second, even if rank orders were preserved under the single-variable relationship, actual impacts are typically driven by multiple interacting climate variables simultaneously, so that correct ranking along any one climate dimension carries no guarantee of correct ranking in the full impact space.

This will be further exacerbated if other impact models are included or model intercomparison studies are performed. Process-based GGCMs alone diverge substantially in their sensitivity to climatic drivers (Franke et al., 2020a; Müller et al., 2024; Oberleitner et al., 2025). This can be expected based on the complexity of crop growth processes that not only depend on mean temperature and precipitation sums but start with solar radiation as the key driver of photosynthesis, while atmospheric water demand as a driver of crop-water relations is also subject to wind speed and air humidity. Their importances and relative contributions to crop growth and yield vary in space and time globally (Oberleitner et al., 2025) rendering a reduction to few variables overly simplistic, even at aggregate scales. The diverse conceptualizations and implementations of the underlying Earth system physics within GCMs in turn can result in a corresponding wide variety of outcomes for atmospheric processes and variables. This is shown herein for a high degree of uncoupling between temperature and precipitation change (Figure 2; Figure S 2), which extends to and interact with differences in atmospheric aerosol dynamics, cloud formation and resulting differences in incoming solar radiation among others (Wild, 2020). In line with this, earlier studies addressing the bracketing of climate attributes within CMIP ensembles found similarly to our evaluations that temperatures are easier to bracket than precipitation in SSP5-8.5 (Ito et al., 2020; McSweeney and Jones, 2016).

Comparing outcomes for the GCM selection used in ISIMIP with those for the whole ensemble indicates that this frequently used sub-ensemble largely reflects average impacts but can substantially underestimate the ranges of potential outcomes (Figure 7). This goes along with a bias towards more negative outcomes, which has implications for how climate impact uncertainties are conveyed where sub-ensembles are applied in impact studies. For impact studies making use of smaller sets of GCMs, our analyses stress that it is necessary to provide sufficient context on how the selected GCM projections relate to the wider ensemble. The best option would apparently be to quantify full ranges in outcomes across the ensemble. Where impact models themselves are computationally too costly, emulators as applied herein increasingly provide a viable option. These have thus far been developed in various forms for crop models, mostly based on machine-learning (Folberth et al., 2025a) or parametric response functions (Franke et al., 2020a; Liu et al., 2025) that can be applied at different scales and for different purposes. For other sectors such as hydrology and integrated land use, emulation of impact models is more challenging due to spatial interactions. Still, first emulators are being developed that reflect key functional relationships at aggregate spatial and temporal scales for land use modelling (Awais et al., 2024) and hydrologic modelling (Liu et al., 2018) or use deep learning to represent spatial interconnectedness and hierarchy of hydrologic models (Bennett et al., 2024).

Code availability

A frozen version of the code and data required to reproduce the bracketing of climate impacts by major UN regions is provided in a public repository (Baklanov and Folberth, 2026). Crop yield predictions have been produced using the CROMES crop model emulator suite (Folberth et al., 2025b).

Data availability

The study is based on various publicly available data sources. Climate data have been sourced from the ISIMIP data repository (Lange and Büchner, 2021) for crop model simulations and emulator training. Climate data for yield predictions have been sources from the NASA NEX GDDP CMIP6 repository (Thrasher et al., 2022).

Author contribution

CF and AB designed the experiments and carried them out. AB, NK, TO, and CF developed computational tools and performed analyses. All authors contributed to the interpretation of results. CF prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare no conflict of interest.

Acknowledgements

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

This research was funded in whole by the Austrian Science Fund (FWF) (10.55776/P36220). For the purpose of open access, the author has applied a CC BY public copyright licence to any author accepted manuscript version arising from this submission.

References

Awais, M., Vinca, A., Byers, E., Frank, S., Fricko, O., Boere, E., Burek, P., Poblete Cazenave, M., Kishimoto, P. N., Mastrucci, A., Satoh, Y., Palazzo, A., McPherson, M., Riahi, K., and Krey, V.: MESSAGEix-GLOBIOM nexus module: integrating water sector and climate impacts, Geoscientific Model Development, 17, 2447–2469, https://doi.org/10.5194/gmd-17-2447-2024, 2024.

Baklanov, A. and Folberth, C.: Bracketing crop climate impacts, https://doi.org/10.5281/zenodo.21103210, 2026.

Balkovič, J., van der Velde, M., Skalský, R., Xiong, W., Folberth, C., Khabarov, N., Smirnov, A., Mueller, N. D., and Obersteiner, M.: Global wheat production potentials and management flexibility under the representative concentration pathways, Global and Planetary Change, 122, 107–121, https://doi.org/10.1016/j.gloplacha.2014.08.010, 2014.

Balkovič, J., Madaras, M., Skalský, R., Folberth, C., Smatanová, M., Schmid, E., van der Velde, M., Kraxner, F., and Obersteiner, M.: Verifiable soil organic carbon modelling to facilitate regional reporting of cropland carbon change: A test case in the Czech Republic, Journal of Environmental Management, 274, 111206, https://doi.org/10.1016/j.jenvman.2020.111206, 2020.

Bennett, A., Tran, H., De la Fuente, L., Triplett, A., Ma, Y., Melchior, P., Maxwell, R. M., and Condon, L. E.: Spatio-Temporal Machine Learning for Regional to Continental Scale Terrestrial Hydrology, Journal of Advances in Modeling Earth Systems, 16, e2023MS004095, https://doi.org/10.1029/2023MS004095, 2024.

Bezner Kerr, R., Hasegawa, T., Lasco, R., Bhatt, I., Deryng, D., Farrell, A., Gurney-Smith, H., Ju, H., Lluch-Cota, S., Meza, F., Nelson, G., Neufeldt, H., and Thornton, P.: Food, Fibre, and Other Ecosystem Products, edited by: Pörtner, H. O., Roberts, D. C., Tignor, M., Poloczanska, E. S., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., and Rama, B., Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 713–906, https://doi.org/10.1017/9781009325844.007.714, 2022.

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., and Knutti, R.: Reduced global warming from CMIP6 projections when weighting models by performance and independence, Earth System Dynamics, 11, 995–1012, https://doi.org/10.5194/esd-11-995-2020, 2020.

Cannon, A. J.: The Impact of "Hot Models" on a CMIP6 Ensemble Used by Climate Service Providers in Canada: Do Global Constraints Lead to Appreciable Differences in Regional Projections?, Journal of Climate, 37, 2141–2154, https://doi.org/10.1175/JCLI-D-23-0459.1, 2024.

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

Chai, Y., Berghuijs, W. R., Naudts, K., Janssen, T. A. J., Yao, Y., and Dolman, H.: Using precipitation sensitivity to temperature to adjust projected global runoff, Environ. Res. Lett., 16, 124032, https://doi.org/10.1088/1748-9326/ac3795, 2021.

Couëdel, A.: Long-term soil organic carbon and crop yield feedbacks differ between 16 soil-crop models in sub-Saharan Africa, European Journal of Agronomy, 2024.

Folberth, C., Baklanov, A., Khabarov, N., Oberleitner, T., Balkovič, J., and Skalský, R.: CROMES v1.0: a flexible CROp Model Emulator Suite for climate impact assessment, Geoscientific Model Development, 18, 5759–5779, https://doi.org/10.5194/gmd-18-5759-2025, 2025a.

Folberth, C., Baklanov, A., Khabarov, N., Oberleitner, T., Balkovic, J., and Skalsky, R.: CROMES v1.0: A flexible CROp Model Emulator Suite for climate impact assessment - Frozen code repository and example for training EPIC-IIASA global gridded crop model emulators, , https://doi.org/10.5281/zenodo.14901127, 2025b.

Forrest, J., Ralphs, T., Santos, H. G., Vigerske, S., Forrest, J., Hafer, L., Kristjansson, B., jpfasano, Jan-Willem, EdwinStraver, Lubin, M., rlougee, dfylstra, a-andre, jpgoncal1, Brito, S., h-i-gassmann, Cullmann, C., Cristina, Saltzman, M., raller09, tosttost, Pitrus, B., Panov, D., MATSUSHIMA, F., Vossler, P., SWGY, R. @, and to-st: coin-or/Cbc: Release releases/2.10.13, , https://doi.org/10.5281/zenodo.18971618, 2026.

Franke, J. A., Müller, C., Elliott, J., Ruane, A. C., Jägermeyr, J., Snyder, A., Dury, M., Falloon, P. D., Folberth, C., François, L., Hank, T., Izaurralde, R. C., Jacquemin, I., Jones, C., Li, M., Liu, W., Olin, S., Phillips, M., Pugh, T. A. M., Reddy, A., Williams, K., Wang, Z., Zabel, F., and Moyer, E. J.: The GGCMI Phase 2 emulators: global gridded crop model responses to changes in CO_2, temperature, water, and nitrogen (version 1.0), Geoscientific Model Development, 13, 3995–4018, https://doi.org/https://doi.org/10.5194/gmd-13-3995-2020, 2020a.

Franke, J. A., Müller, C., Elliott, J., Ruane, A. C., Jägermeyr, J., Balkovic, J., Ciais, P., Dury, M., Falloon, P. D., Folberth, C., François, L., Hank, T., Hoffmann, M., Izaurralde, R. C., Jacquemin, I., Jones, C., Khabarov, N., Koch, M., Li, M., Liu, W., Olin, S., Phillips, M., Pugh, T. A. M., Reddy, A., Wang, X., Williams, K., Zabel, F., and Moyer, E. J.: The GGCMI Phase 2 experiment: global gridded crop model simulations under uniform changes in CO_2, temperature, water, and nitrogen levels (protocol version 1.0), Geoscientific Model Development, 13, 2315–2336, https://doi.org/https://doi.org/10.5194/gmd-13-2315-2020, 2020b.

Frieler, K., Volkholz, J., Lange, S., Schewe, J., Mengel, M., del Rocío Rivas López, M., Otto, C., Reyer, C. P. O., Karger, D. N., Malle, J. T., Treu, S., Menz, C., Blanchard, J. L., Harrison, C. S., Petrik, C. M., Eddy, T. D., Ortega-Cisneros, K., Novaglio, C., Rousseau, Y., Watson, R. A., Stock, C., Liu, X., Heneghan, R., Tittensor, D., Maury, O., Büchner, M., Vogt, T., Wang, T., Sun, F., Sauer, I. J., Koch, J., Vanderkelen, I., Jägermeyr, J., Müller, C., Rabin, S., Klar, J., Vega del Valle, I. D., Lasslop, G., Chadburn, S., Burke, E., Gallego-Sala, A., Smith, N., Chang, J., Hantson, S., Burton, C., Gädeke, A., Li, F., Gosling, S. N., Müller Schmied, H., Hattermann, F., Wang, J., Yao, F., Hickler, T., Marcé, R., Pierson, D., Thiery, W., Mercado-Bettín, D., Ladwig, R., Ayala-Zamora, A. I., Forrest, M., and Bechtold, M.: Scenario setup and forcing data for impact model evaluation and impact attribution within the third round of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP3a), Geoscientific Model Development, 17, 1–51, https://doi.org/10.5194/gmd-17-1-2024, 2024.

Frieler, K., Lange, S., Schewe, J., Mengel, M., Treu, S., Otto, C., Volkholz, J., Reyer, C. P. O., Heinicke, S., Jones, C., Blanchard, J. L., Harrison, C. S., Petrik, C. M., Eddy, T. D., Ortega-Cisneros, K., Novaglio, C., Heneghan, R., Tittensor, D. P., Maury, O., Büchner, M., Vogt, T., Quesada Chacón, D., Emanuel, K., Lee, C.-Y., Camargo, S. J., Jägermeyr, J., Rabin, S., Klar, J., Vega Del Valle, I. D., Novak, L., Sauer, I. J., Lasslop, G., Chadburn, S., Burke, E., Gallego-Sala, A., Smith, N., Chang, J., Hantson, S., Burton, C., Gädeke, A., Li, F., Gosling, S. N., Müller Schmied, H., Hattermann, F., Hickler, T., Marcé, R., Pierson, D., Thiery, W., Mercado-Bettín, D., Ladwig, R., Ayala-Zamora, A. I., Forrest, M., Bechtold, M., Reinecke, R., De Graaf, I., Kaplan, J. O., Koch, A., and Lengaigne, M.: Scenario set-up and the new CMIP6-based climate-related forcings provided within the third round of the Inter-Sectoral Model Intercomparison Project (ISIMIP3b, group I and II), https://doi.org/10.5194/egusphere-2025-2103, 2 June 2025.

Hausfather, Z.: An assessment of current policy scenarios over the 21st century and the reduced plausibility of high-emissions pathways, Dialogues on Climate Change, 2, 26–32, https://doi.org/10.1177/29768659241304854, 2025.

Hausfather, Z., Marvel, K., Schmidt, G. A., Nielsen-Gammon, J. W., and Zelinka, M.: Climate simulations: recognize the 'hot model' problem, Nature, 605, 26–29, https://doi.org/10.1038/d41586-022-01192-2, 2022.

Herger, N., Abramowitz, G., Knutti, R., Angélil, O., Lehmann, K., and Sanderson, B. M.: Selecting a climate model subset to optimise key ensemble properties, Earth System Dynamics, 9, 135–151, https://doi.org/10.5194/esd-9-135-2018, 2018.

IFPRI: Global Spatially-Disaggregated Crop Production Statistics Data for 2010 Version 2.0, https://doi.org/10.7910/DVN/PRFF8V, 2020.

Ito, R., Shiogama, H., Nakaegawa, T., and Takayabu, I.: Uncertainties in climate change projections covered by the ISIMIP and CORDEX model subsets from CMIP5, Geoscientific Model Development, 13, 859–872, https://doi.org/10.5194/gmd-13-859-2020, 2020.

Izaurralde, R. C., Williams, J. R., McGill, W. B., Rosenberg, N. J., and Jakas, M. C. Q.: Simulating soil C dynamics with EPIC: Model description and testing against long-term data, Ecological Modelling, 192, 362–384, https://doi.org/10.1016/j.ecolmodel.2005.07.010, 2006.

Jägermeyr, J., Müller, C., Ruane, A. C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J. A., Fuchs, K., Guarin, J. R., Heinke, J., Hoogenboom, G., Iizumi, T., Jain, A. K., Kelly, D., Khabarov, N., Lange, S., Lin, T.-S., Liu, W., Mialyk, O., Minoli, S., Moyer, E. J., Okada, M., Phillips, M., Porter, C., Rabin, S. S., Scheer, C., Schneider, J. M., Schyns, J. F., Skalsky, R., Smerald, A., Stella, T., Stephens, H., Webber, H., Zabel, F., and Rosenzweig, C.: Climate impacts on global agriculture emerge earlier in new generation of climate and crop models, Nat Food, 2, 873–885, https://doi.org/10.1038/s43016-021-00400-y, 2021a.

Jägermeyr, J., Müller, C., Minoli, S., Ray, D., and Siebert, S.: GGCMI Phase 3 crop calendar, https://doi.org/10.5281/zenodo.5062513, 2021b.

Katzenberger, A., Pérez-Carrasquilla, J. S., Gemmell, K., Galytska, E., Leclerc, C., P, P., Roy, I., Varuolo-Clarke, A., Tošić, M., and Črnivec, N.: Developing Guidelines for working with Multi-Model Ensembles in CMIP, Earth System Dynamics, 17, 495–532, https://doi.org/10.5194/esd-17-495-2026, 2026.

Laborte, A. G., Gutierrez, M. A., Balanza, J. G., Saito, K., Zwart, S. J., Boschetti, M., Murty, M. V. R., Villano, L., Aunario, J. K., Reinke, R., Koo, J., Hijmans, R. J., and Nelson, A.: RiceAtlas, a spatial database of global rice calendars and production, Sci Data, 4, 170074, https://doi.org/10.1038/sdata.2017.74, 2017.

Lange, S.: ISIMIP3b bias adjustment fact sheet, 2021.

Lange, S. and Büchner, M.: ISIMIP3b bias-adjusted atmospheric climate input data (v1.1), , https://doi.org/10.48364/ISIMIP.842396.1, 2021.

Liu, X., Gasser, T., Ma, J., Liu, J., Jägermeyr, J., Müller, C., Folberth, C., Zabel, F., Jain, A., Liu, W., and Webber, H.: A food crop yield emulator for integration in the compact Earth system model OSCAR (OSCAR-crop v1.0), EGUsphere, 1–38, https://doi.org/10.5194/egusphere-2025-4805, 2025.

Liu, Y., Hejazi, M., Li, H., Zhang, X., and Leng, G.: A hydrological emulator for global applications – HE v1.0.0, Geoscientific Model Development, 11, 1077–1092, https://doi.org/10.5194/gmd-11-1077-2018, 2018.

McSweeney, C. F. and Jones, R. G.: How representative is the spread of climate projections from the 5 CMIP5 GCMs used in ISI-MIP?, Climate Services, 1, 24–29, https://doi.org/10.1016/j.cliser.2016.02.001, 2016.

Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P. B., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P. J., Reimann, S., Smith, S. J., Van Den Berg, M., Velders, G. J. M., Vollmer, M. K., and Wang, R. H. J.: The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500, Geosci. Model Dev., 13, 3571–3605, https://doi.org/10.5194/gmd-13-3571-2020, 2020.

Merrifield, A. L., Brunner, L., Lorenz, R., Humphrey, V., and Knutti, R.: Climate model Selection by Independence, Performance, and Spread (ClimSIPS v1.0.1) for regional applications, Geoscientific Model Development, 16, 4715–4747, https://doi.org/10.5194/gmd-16-4715-2023, 2023.

Mitchell, S., OSullivan, M., and Dunning, I.: Pulp: a linear programming toolkit for python, The University of Auckland, Auckland, New Zealand, 65, 2011, 2011.

Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., and Yang, H.: Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications, Geosci. Model Dev., 10, 1403–1422, https://doi.org/10.5194/gmd-10-1403-2017, 2017.

Müller, C., Franke, J., Jägermeyr, J., Ruane, A. C., Elliott, J., Moyer, E., Heinke, J., Falloon, P. D., Folberth, C., Francois, L., Hank, T., Izaurralde, R. C., Jacquemin, I., Liu, W., Olin, S., Pugh, T. A. M., Williams, K., and Zabel, F.: Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios, Environ. Res. Lett., 16, 034040, https://doi.org/10.1088/1748-9326/abd8fc, 2021.

Müller, C., Jägermeyr, J., Franke, J. A., Ruane, A. C., Balkovic, J., Ciais, P., Dury, M., Falloon, P., Folberth, C., Hank, T., Hoffmann, M., Izaurralde, R. C., Jacquemin, I., Khabarov, N., Liu, W., Olin, S., Pugh, T. A. M., Wang, X., Williams, K., Zabel, F., and Elliott, J. W.: Substantial Differences in Crop Yield Sensitivities Between Models Call for Functionality-Based Model Evaluation, Earth's Future, 12, e2023EF003773, https://doi.org/10.1029/2023EF003773, 2024.

Nijsse, F. J. M. M., Cox, P. M., and Williamson, M. S.: Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models, Earth Syst. Dynam., 11, 737–750, https://doi.org/10.5194/esd-11-737-2020, 2020.

Oberleitner, T., Baklanov, A., Ferreira, T. B., Hoogenboom, G., Jägermeyr, J., Jain, A., Lin, T.-S., Mialyk, O., Müller, C., Ruane, A. C., Zabel, F., Balkovič, J., Wang, C., Faye, B., Guarin, J. R., Iizumi, T., Khabarov, N., Liu, W., Okada, M., Rabin, S. S., Scheer, C., Skalský, R., and Folberth, C.: Explaining climatic drivers of yield anomalies in global crop models through metamodel-based attribution, ESS Open Archive, https://doi.org/10.22541/essoar.176071967.73925210/v1, 2025.

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L. A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J. C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., and Tavoni, M.: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview, Global Environmental Change, 42, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009, 2017.

Ruane, A. C., Vautard, R., Ranasinghe, R., Sillmann, J., Coppola, E., Arnell, N., Cruz, F. A., Dessai, S., Iles, C. E., Islam, A. K. M. S., Jones, R. G., Rahimi, M., Carrascal, D. R., Seneviratne, S. I., Servonnat, J., Sörensson, A. A., Sylla, M. B., Tebaldi, C., Wang, W., and Zaaboul, R.: The Climatic Impact-Driver Framework for Assessment of Risk-Relevant Climate Information, Earth's Future, 10, e2022EF002803, https://doi.org/10.1029/2022EF002803, 2022.

Skalský, R., Tarasovičová, Z., Balkovič, J., Schmid, E., Fuchs, M., Moltchanova, E., and Scholtz, P.: GEO-BENE global database for bio-physical modeling, GEOBENE project, 2008.

Sobolowski, S., Somot, S., Fernandez, J., Evin, G., Brands, S., Maraun, D., Kotlarski, S., Jury, M., Benestad, R. E., Teichmann, C., Christensen, O. B., Bülow, K., Buonomo, E., Katragkou, E., Steger, C., Sørland, S., Nikulin, G., McSweeney, C., Dobler, A., Palmer, T., Wilcke, R., Boé, J., Brunner, L., Ribes, A., Qasmi, S., Nabat, P., Sevault, F., and Oudar, T.: GCM Selection and Ensemble Design: Best Practices and Recommendations from the EURO-CORDEX Community, Bulletin of the American Meteorological Society, 106, E1834–E1850, https://doi.org/10.1175/BAMS-D-23-0189.1, 2025.

Stehfest, E., Zeist, W.-J. van, Valin, H., Havlik, P., Popp, A., Kyle, P., Tabeau, A., Mason-D'Croz, D., Hasegawa, T., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fujimori, S., Humpenöder, F., Lotze-Campen, H., Meijl, H. van, and Wiebe, K.: Key determinants of global land-use projections, Nature Communications, 10, 2166, https://doi.org/10.1038/s41467-019-09945-w, 2019.

Swaminathan, R., Schewe, J., Walton, J., Zimmermann, K., Jones, C., Betts, R. A., Burton, C., Jones, C. D., Mengel, M., Reyer, C. P. O., Turner, A. G., and Weigel, K.: Regional Impacts Poorly Constrained by Climate Sensitivity, Earth's Future, 12, e2024EF004901, https://doi.org/10.1029/2024EF004901, 2024.

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K. B., Hurtt, G., Kriegler, E., Lamarque, J.-F., Meehl, G., Moss, R., Bauer, S. E., Boucher, O., Brovkin, V., Byun, Y.-H., Dix, M., Gualdi, S., Guo, H., John, J. G., Kharin, S., Kim, Y., Koshiro, T., Ma, L., Olivié, D., Panickal, S., Qiao, F., Rong, X., Rosenbloom, N., Schupfner, M., Séférian, R., Sellar, A., Semmler, T., Shi, X., Song, Z., Steger, C., Stouffer, R., Swart, N., Tachiiri, K., Tang, Q., Tatebe, H., Voldoire, A., Volodin, E., Wyser, K., Xin, X., Yang, S., Yu, Y., and Ziehn, T.: Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6, Earth System Dynamics, 12, 253–293, https://doi.org/10.5194/esd-12-253-2021, 2021.

Teckentrup, L., De Kauwe, M. G., Abramowitz, G., Pitman, A. J., Ukkola, A. M., Hobeichi, S., François, B., and Smith, B.: Opening Pandora's box: reducing global circulation model uncertainty in Australian simulations of the carbon cycle, Earth System Dynamics, 14, 549–576, https://doi.org/10.5194/esd-14-549-2023, 2023.

Thrasher, B., Wang, W., Michaelis, A., Melton, F., Lee, T., and Nemani, R.: NASA Global Daily Downscaled Projections, CMIP6, Sci Data, 9, 262, https://doi.org/10.1038/s41597-022-01393-4, 2022.

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., and Knutti, R.: Past warming trend constrains future warming in CMIP6 models, Science Advances, 6, eaaz9549, https://doi.org/10.1126/sciadv.aaz9549, 2020.

UNSD: Standard country or area codes for statistical use (Rev. 3), UN, New York: United Nations, 1996.

Wild, M.: The global energy balance as represented in CMIP6 climate models, Clim Dyn, 55, 553–577, https://doi.org/10.1007/s00382-020-05282-7, 2020.

Williams, J. R., Jones, C. A., Kiniry, J. R., and Spanel, D. A.: The EPIC crop growth model, Transactions of the ASAE, 32, 497–0511, 1989.

https://doi.org/10.5194/egusphere-2026-4018
Preprint. Discussion started: 4 August 2026

Wynes, S., Davis, S. J., Dickau, M., Ly, S., Maibach, E., Rogelj, J., Zickfeld, K., and Matthews, H. D.: Perceptions of carbon dioxide emission reductions and future warming among climate experts, Commun Earth Environ, 5, 498, https://doi.org/10.1038/s43247-024-01661-8, 2024.

Yu, Q., You, L., Wood-Sichra, U., Ru, Y., Joglekar, A. K. B., Fritz, S., Xiong, W., Lu, M., Wu, W., and Yang, P.: A cultivated planet in 2010 – Part 2: The global gridded agricultural-production maps, 28, 2020.