https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

CROMES v1.0: A flexible CROp Model Emulator Suite for climate impact assessment

Christian Folberth[1], Artem Baklanov[2], Nikolay Khabarov[2], Thomas Oberleitner[1], Juraj Balkovič[1], Rastislav Skalský[1]

[1] Biodiversity and Natural Resources Program, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria

[2] Advancing Systems Analysis Program, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria

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

Abstract. Global gridded crop models (GGCMs) are simulation tools designed for global, spatially explicit estimation of crop productivity and associated externalities. Key areas for their application are climate impact and adaptation studies. As GGCMs are typically computationally costly and require comprehensive data pre- and post-processing, GGCM emulators are gaining increasing popularity. Earlier emulators have typically been published pre-trained on synthetic weather and management combinations. Here, we present a novel computational pipeline CROp Model Emulator Suite (CROMES) v1.0 that serves for flexibly training GGCM emulators on data commonly available from GGCM simulations. Essentially, CROMES consists of modules to (1) process climate data from daily resolution netCDF files to (sub-)growing season aggregates as climate features, (2) combine various feature types (climate, soil, crop management), (3) train emulators using machine-learning algorithms, and (4) produce predictions. Exemplary, we apply CROMES to train emulators on simulations for rainfed maize from the GGCM EPIC-IIASA and climate projections from a single GCM to subsequently test their skill in predicting crop yields for unseen climate projections from other GCMs. Depending on the training and target data, the regression statistics between GGCM simulations and predictions across all points in time and space are in the ranges R2=0.97 to 0.98, slope=0.99 to 1.01, and intercept=-0.06 to +0.06. The RMSE ranges between 0.49 and 0.65 t ha^{-1}. Spatially, patterns are evident with lowest performance in (semi-)arid regions where aggregation of weather data may result in higher information loss while permanent crop growth limitations may hamper evaluation statistics as well. The gain in computational speed for predictions is at more than an order of magnitude with time required to produce target features and subsequent predictions at about 30min on common hardware. We expect CROMES to be of utility in covering more comprehensively uncertainty in climate impact projections, evaluations of adaptation options, and spatio-temporal assessments of crop productivity.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

1 Introduction

Global gridded crop models (GGCMs) have become key tools in large-scale agricultural climate impact and adaptation assessments (Jägermeyr et al., 2021) and as a source of crop yield estimates for land use and integrated assessment models (Nelson et al., 2014). Yet, these combinations of large-scale spatial data frameworks and plant growth models have limitations in the volume of scenarios they can address due to computational demand or complex software and data structures. At the same time, ever larger volumes of bias-corrected climate projections become available as potential forcings for GGCMs allowing in principle for comprehensive uncertainty assessment (Gao et al., 2023; Gebrechorkos et al., 2023; Lange and Büchner, 2021; Thrasher et al., 2022). Also spatial resolutions of climate data are constantly improving with first 1k resolution global daily meteorological data available (Karger et al., 2023) but requiring vastly higher computational capacities compared to the state-of-the-art 0.5° x 0.5° (approx. 50 km x 50 km near the equator).

To allow for more comprehensive scenario analyses without exacerbating computational costs, emulators mimicking GGCMs have emerged as tools to produce reasonably accurate predictions of GGCMs' crop productivity estimates at much lower computational requirements and with sparser sets of aggregate input data. First developments in this field were common linear models trained on opportunistic samples from GGCM climate impact simulations (Blanc, 2017; Blanc and Sultan, 2015; Oyebamiji et al., 2015). Most recent emulators have been based on structured training data obtained from vast GGCM simulations for systematic perturbations of meteorologic reanalysis data combined with location-specific polynomials (Franke et al., 2020b). These have been employed extensively for comprehensive scenario analyses (Franke et al., 2022; Müller et al., 2021; Zabel et al., 2021) and analytic purposes (Müller et al., 2024).

However, emulators published thus far are subject to several limitations. E.g., inter-annual yield variability can hardly be reflected due to the use of annual or static seasonal climate features and common regression models, and predictive performance is typically still lacking robustness. Also, the frequent use of individual algorithms or parameters per pixel limits the flexibility of emulator applications across spatial scales. Structured training data furthermore require comprehensive crop model simulations and dedicated experiments (Franke et al., 2020a). This causes substantial overhead and hampers timely updates of training data with new model versions and setups that are regularly applied in climate impact studies. More complex machine-learning algorithms such as boosting, regression trees, and neural networks in turn have been shown to provide high flexibility in producing predictions similar to those of crop models if combined with covariates at moderate temporal resolutions, albeit these methods have thus far only been tested for spatial downscaling and evaluations of model training strategies (Folberth et al., 2019; Sweet et al., 2023). Yet, their high predictive performance and flexibility renders such setups promising for the development of novel emulators.

Building on these recent developments, we present herein a computational pipeline combining modules for fast climate feature engineering tailored towards the crop growing season and sub-seasons with machine-learning algorithms for the training and application of GGCM emulators. In contrast to providing pre-trained emulators, this pipeline presents a flexible tool allowing for continuous updates based on specific requirements of applications and new training data as these become available. For the demonstration experiment herein, we train

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

emulators on a set of simulation outputs for the most recent simulation round phase 3b of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) and the Global Gridded Crop Model Intercomparison (GGCMI) initiative (Jägermeyr et al., 2021). Our approach is based on the hypothesis that by using a global set of simulations spanning diverse agro-climatic and –environmental conditions, we can train emulators with high enough flexibility to mimic GGCM simulations for unseen climate projections from the same domain (here CMIP6). For practical reasons, we focus on emulators for the crop model Environmental Policy Integrated Climate (EPIC; (Williams, 1990)) that is used by the authors in the global gridded implementation EPIC-IIASA (Balkovič et al., 2013).

2 Methods

2.1 Study design and experiment setup

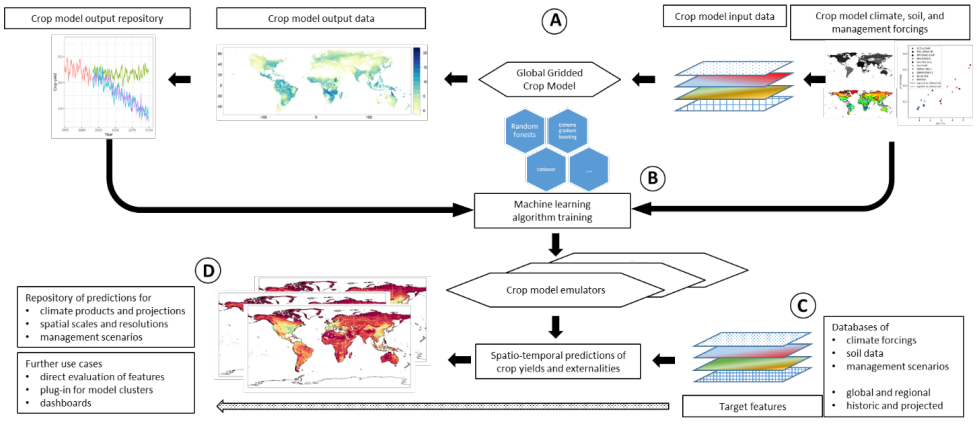

Figure 1. Study design schematic. A: Global gridded crop model simulations for a specific set of forcing data to generate a training sample for emulators, B: training of crop model emulators based on machine learning algorithms and the global GGCM training sample, C: processing of features from target forcings and predictions using emulators from (C), D: storage and evaluation of predictions and/or optional further use of climate features.

The design of CROMES and the setup for the present study is shown in Figure 1 with details provided in the subsequent sections. First, GGCM simulations - using here the EPIC-IIASA model and forcing data from ISIMIP3b - are performed to generate a training sample (Figure 1A). A climate feature processing module generates features from climate forcing datasets for various parts of the crop growing season. These are combined with the GGCM crop yield estimates as target variable and further features on soil, site characteristics, and crop management to train machine-learning algorithms as emulators (Figure 1B). The same module produces features for predictions (Figure 1C) that serve as covariates for the emulator, which eventually produces crop yield predictions (Figure 1D). The rapid generation of climate features is a core part of CROMES as it is key for the computational speed gain compared to GGCM simulations. These features may also be used directly, e.g., for analyses of growing season climate.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

The exemplary application of CROMES herein evaluates in how far emulators that are trained on GGCM simulations for a specific GCM covering the historical time period and three projections along different representative concentration pathways (RCPs; see sect. 2.9) are skilled to predict crop yields for climate scenarios from other GCMs. Essentially, we perform GGCM simulations using climate forcings from five GCMs, subsequently train emulators for each of these GCMs individually, and benchmark crop yield predictions for the other four GCMs against actual crop model simulations.

While crop nutrient supply can in principle be added to the features, we opt herein to evaluate only predictions for simulations with sufficient nutrient supply to single out the skill of the emulators to capture climate signals.

2.2 Technical design of the emulator pipeline

The code implementation of CROMES is closely aligned with the study design (sect. 2.1) and detailed in the subsequent sections. CROMES handles the processing of data, feature engineering, training of emulators, and emulator evaluation in four steps:

1) conversion of netCDF climate data to binary files for rapid read access
2) processing of soil, site, crop management, and climate features
3) emulator training
4) emulator application

Implemented features are mostly generic. These include among others growing season aggregates of key climate variables, soil texture, and crop growing season information. More complex approaches are required for the estimation of potential evapotranspiration (PET), which can be based on various methods in crop models (Wartenburger et al., 2018). Herein, we use the Penman-Monteith method that is widely used within GGCMs (Jägermeyr et al., 2021) and has been implemented in the EPIC model as described in (Stockle et al., 1992). We use the CatBoost algorithm for emulator training, a computationally highly efficient algorithm that has been top-ranking in benchmarks (Prokhorenkova et al., 2018) and tested in a wide range of applications (Hancock and Khoshgoftaar, 2020).

2.3 Climate data pre-processing

Climate features are produced for an individual pixel as aggregates over specific time periods (e.g. annual growing season; see Sect. 2.4). In this calculation the whole set of values of each climatic variable needs to be made available to an aggregation function, essentially for the estimation of PET. Therefore, the original set of two-dimensional maps in the netCDF files typically used to supply spatio-temporal climate data has to be converted to a set of vectors, i.e., time series, of individual map pixels for a defined land mask. This conversion of maps to vectors is carried out in a netCDF to binary file translation routine.

The conversion carried out once per climate data set substantially speeds up the subsequent climate feature engineering process. Selecting all climatic values sequentially for each individual map pixel is infeasible due to the large size of the pixel set (here, the ISIMIP 3b cropland mask with 65797 pixels) and the large number of days (about 36500 for a 100-year dataset). Together with the number of climatic variables (here six) this leads to about

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

66000*36500*6 = 14 * 10^9 selection operations. As one selection (seek) operation on a state-of-the-art solid-state drive can take more than 0.01 to 0.2 ms, this would result in 14 * 10^9 * 0.01 / 1000 / 3600 / 24 = 2 to 40 days of processing. Carrying out the conversion through a dedicated routine that is extensively using RAM instead of intensive disk input allows to reduce that time to few minutes and less depending on specific hardware in place.

While netCDF files may vary in their configuration, the routines presently implemented in CROMES expect netCDF files compliant with data format conventions used within ISIMIP phase 3b, which are based on NetCDF Climate and Forecast (CF) Metadata Conventions CF-1.6 and a spatial resolution of 0.5° x 0.5°.

2.4 Feature engineering

2.4.1 Summary of included features

Table 1 provides an overview of implemented climate features. The first six rows (TMX to HUR) correspond to raw climate input variables for the EPIC crop growth model that are here used both directly and in the calculation of derived climate features. The latter include growing degree days (GDD; see sect. 2.4.2), the number of hot degree days (HDD), extreme degree days (EDD), numbers of wet and dry days, and the actual length of the growing season or selected key stages (see below). PET is (see sect. 2.4.3 for details) is used directly and in the calculation of the climatic moisture deficit (CMD) and days with CMD below zero (CMDlt0) as drought indicators. Further outputs of the climate feature module are the individual growing season length (GSL) and the maturity status of the crop at harvest (HUIeopv). CO2trans has a globally uniform annual value.

Table 1. Overview of climate features by climate variable and temporal reference. Actual growing season (AGS) length is dynamically estimated each season (see section 2.4.2). {agg} in the bottom part refers to average (av) or sum (sum) over the respective period. An exemplary feature descriptor would accordingly be TMXavAGS. HUIeopv as an indicator for crop maturity is only output for the whole growing season. CO2trans has an annual value and is hence not aggregated.

Abbreviation	Description
Agro-climatic features (VARs)	
TMX	Maximum temperature [°C]
TMN	Minimum temperature [°C]
PRCP	Total precipitation [mm]
RAD	Solar radiation [MJ m^{-2}]
WSD	Wind speed [m s^{-1}]
HUR	Relative humidity [-]
GDD	Growing degree days [°C]
HDD	Hot degree days (T_{av} > 30 °C) [d]
EDD	Extreme degree days (T_{av} > 1.5 crop-specific optimum temperature) [d]
PET	Potential evapotranspiration [mm]
CMD	Climatic moisture deficit (PET- PRCP) [mm]
CMDlt0	Days with CMD below zero [d]
WET	Wet days (PRCP ≥ 0.1 mm) [d]
DRY	Dry days (PRCP < 0.1 mm) [d]
GSL	Growing season length, i.e., days from planting to harvest [d]
HUIeopv	Heat unit index (HUI) at the end of the period (only produced for AGS) [-]
CO2trans	Transient atm. CO_2 concentration [ppm]
Temporal aggregates and derivatives of agro-climatic features	
VAR{agg}AGS	Aggregate for the actual growing season (AGS)
VAR{agg}AGSr	Aggregate for the reproductive phase, i.e., second half of the AGS
VAR{agg}AGSe	Aggregate for the establishment phase, i.e., first quarter of the AGS
VAR{agg}PGS	Aggregate for the pre-growing season, i.e., the 30 days prior to sowing

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

Aggregations are performed (a) for the whole actual growing season (AGS) starting with germination, (b) for the first quarter of the growing season during which the crop emerges (AGSe), (c) for the second half of the growing season – i.e., the reproductive phase during which flowers are prone to water stress (Williams et al., 1989) – (AGSr), and (d) for the 30 days prior to the growing season, during which soil water available for the crop may accumulate (PGS). This breakdown into key growth stages - while also considering growing season totals – serves for improving the information content not only with respect to growth stage-specific crop sensitivities to stresses but also with respect to synchronous or asynchronous manifestation of plant growth limitations such as drought and shading. We use the term actual growing season here to indicate that the climate feature module estimates the crop growth duration for each individual season based on growing degree day (GDD) accumulation as opposed to using a fixed calendar that would not account for earlier (later) maturing of crops in warmer (cooler) years. The estimation of the time periods is further elaborated in sect. 2.4.2.

Table 2 shows the non-climatic, temporally static features, essentially soil attributes and slope that impact soil hydrology and root space (see section 2.8). Two crop management parameters are the crop's pixel-specific length of vegetation period (LVP) based on the input planting and harvest dates and the potential heat unit (PHU) requirement.

Table 2. Static soil, site, and crop management features considered in the present setup.

Feature	Description	Category
DEPTH	Total soil depth [m]	Soil
SAND	Sand content [%]	Soil
CLAY	Clay content [%]	Soil
PH	pH [-]	Soil
SB	Sum of bases [cmol kg^{-1}]	Soil
CEC	Cation exchange capacity [cmol kg^{-1}]	Soil
EC	Electric conductivity [mmho cm^{-1}]	Soil
ROK	Coarse fragment (rock) content [%]	Soil
BD	Bulk density [g cm^{-3}]	Soil
CARB	Carbonate content [%]	Soil
OC	Organic carbon content [%]	Soil
FC	Soil water content at field capacity (at 33 kPa) [m m^{-1}]	Soil
WP	Soil water content at wilting point (at 1500 kPa) [m m^{-1}]	Soil
PAW	Total plant available water capacity [m^3 m^{-3}]	Soil
SLP	Hill slope [%]	Site
PHU	Potential heat units (syn. growing degree days) from planting to maturity [°C]	Crop management
LVP	Length of vegetation period from reported planting to harvest date [d]	Crop management

2.4.2 Estimation of growing season length and sub-seasons

The estimation of growing season length is based on GDD accumulation as implemented in the EPIC model and most other GGCMs (Jägermeyr et al., 2021; Müller et al., 2017). Any adjustments can be made in the code or input parameterization that includes parameters for crop-specific base and optimum temperatures.

Earlier crop model emulators and various analytical studies combining crop model simulations and climatic indicators for climate impact estimation have utilized monthly or annual climate features (Blanc, 2017; Folberth et al., 2019; Franke et al., 2020b; Goulart et al., 2023; Sweet et al., 2023). While annual features cannot be

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

expected to capture more than trends in climate, monthly features – typically ordered from planting - at least capture some dynamics within the growing season. Yet, neither of the two considers the effect of earlier (later) crop maturity due to warmer (cooler) than baseline average growing season temperatures. This is one of the main climate impact drivers in crop models (Minoli et al., 2019; Zabel et al., 2021). It determines for example the amount of solar radiation the crop receives for biomass accumulation and whether it is exposed to adverse weather occurring later in the reported growing season. As in the majority of crop models, the progression of crop development from planting to maturity is in CROMES estimated based on the heat unit (HU, syn. growing degree days (GDD)) accumulation approach. That is, on each day i of the growing season daily HU are calculated according to:

$$HU_i = \frac{T_{max,i}+T_{min,i}}{2} - T_b, \quad with\ HU_i \geq 0 \tag{1}$$

where T_{max} [°C] is daily maximum temperature, T_{min} [°C] is daily minimum temperature, and T_b [°C] is the crop-specific base temperature for growth, here 8 °C for maize. The sum of HU for recent historic average temperatures between reported planting and harvest dates in a location is considered a static cultivar definition termed potential heat units (PHU). Based on input planting dates and PHU, the model estimates the progression of plant phenologic development, biomass accumulation, and maturation for each individual growing season. Harvest occurs dynamically after the PHU value is reached (or at a defined cut-off, see below). To normalize plant maturation across locations, a heat unit index (HUI) is used, which is calculated as the cumulative fraction of required PHU reached on day *i* of the growing season as

$$HUI_i = \frac{\sum_{k=1}^{i} HU_k}{PHU} \tag{2}$$

The HUI at harvest serves as a feature (HUIeopv) herein to inform whether the crop has reached maturity. Prior to emergence of the crop, an additional amount of germination HU (GMHU) is required for the seed to develop to a seedling, here 100 °C for maize.

Figure 2 provides an overview of growing season-based climate feature aggregation (incl. pre-growing season (PGS)). The climate feature module first estimates for each growing season based on the input planting date, GMHU, and PHU the germination and maturity dates. If the crop does not mature due to too low growing season temperatures, a cut-off is enforced 21d after the reported planting date. Subsequently, climate features are calculated for the whole actual growing season (AGS) and the critical growing season phases for crop establishment (AGSe) and reproductive phase (AGSr). The first occurs from HUI=0 to HUI=0.25, the second from HUI=0.5 to HUI=1.0 or cut-off date. During the reproductive phase, the crop yield is most sensitive to drought. The PGS is defined as 30d prior to planting, a period that may inform on germination and early growth conditions such as soil humidity.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

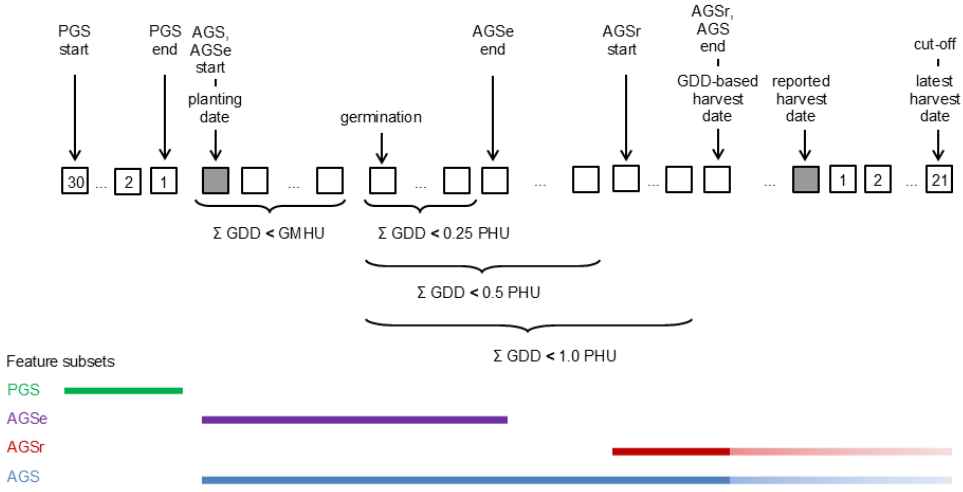

Figure 2. Conceptual definition of the crop growing season and growing season-oriented climate feature subsets. Squared boxes indicate individual days for periods that are universally pre-defined (with numbers) or flexible based on individual input growing season dates and GDD accumulation (empty). PGS=pre-growing season, AGS=actual growing season, AGSe=actual growing season emergence phase (1st quarter), AGSr=actual growing season reproductive phase (2nd half), cut-off=forced growing season cut-off if PHU are not reached 21d after reported harvest date, GDD=growing degree days (syn. heat units), PHU=potential heat units (i.e., GDD estimated for the baseline period as part of cultivar definition). Colored bars in the lower part of the figure indicate the extent of the growing season subsets. The lighter colored extensions at the end of AGS and AGSr indicate that the end of the growing season is either determined by reaching GDD ≥ 1.0 PHU or at the cut-off. The latter serves to avoid overly long growing seasons in cool years where a crop may not reach maturity in autumn and the growing season would hence extend over winter.

2.4.3 Penman-Monteith PET estimation

There are numerous methods for estimating PET employed in GGCMs (Jägermeyr et al., 2021; Liu et al., 2016; Wartenburger et al., 2018) with degrees of complexity and input data requirements. The most popular choice in a recent ensemble is Penman-Monteith (Jägermeyr et al., 2021), which is also implemented in the EPIC crop growth model based on (Stockle et al., 1992). The same approach was followed herein for PET estimation in CROMES.

Penman-Monteith requires all raw climate variables (first six rows in Table 1) as well as information on daily crop height (CHT) and leaf area index (LAI), rendering its estimation considerably complex. The underlying calculations are therefore only provided in abbreviated form and the reader is referred to the above reference and the code for further information. In short, the climate feature module estimates daily progression of CHT and LAI based on HUI and crop-specific parameters, and passes these parameters, daily climate data, and further coefficients (atm. CO_2 concentration, elevation, soil albedo, latitude) to the PET function. Whether or not a crop is growing on a day determines the use of the main equation which is

$$E_o = \frac{\delta(h_o - G) + 86.7\,AD\,(e_a - e_d)/AR}{(HV)(\delta + \gamma)} \quad (3)$$

if no crop is grown or if a crop grows

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

$$E_p = \frac{\delta(h_o - G) + 86.7\ AD(e_a - e_d) / AR}{HV(\delta + \gamma(1. + CR / AR))} \quad (4)$$

where AD is the air density [g m^{-3}], AR is the aerodynamic resistance for heat and vapor transfer [s m^{-1}], and CR is the canopy resistance for vapor transfer [s m^{-1}], HV is the latent heat of vaporization [MJ kg^{-1}], e_a is saturation vapor pressure [kPa], e_d is actual vapor pressure [kPa], δ is the slope of the saturation vapor pressure curve [kPa °C], G is soil heat flux assumed zero in the model, h_o is net solar radiation [MJ m^{-2}], and γ is the psychrometric constant [kPa °C].

2.5 Non-climatic features

Soil features (Table 2) include soil physical and chemical attributes as commonly required by crop models and provided in state-of-the-art data sources such as the one used herein (see section 2.9). Here, we used soil features stored after a spin-up run of the crop model for full consistency with crop model simulations. The first 11 rows of soil features (DEPTH to OC) in Table 2 are raw values, the remainder has been estimated based on routines implemented in the EPIC model (FC, WP, PAW). PHU have been derived as described in the prior sections.

2.6 Emulator training and feature importance

All features, including the target variable crop yield for model training, are eventually merged based on simulation unit IDs or climate grid IDs (see sect. 2.8).

For the demonstration herein, we chose CatBoost, a high-performing algorithm with GPU support that significantly speeds up the training phase (Prokhorenkova et al., 2018). Hyperparameter selection was done using cross-validation (CV) and grid search as implemented in the Python catboost package. This step should be tailored to each specific training and prediction setup. However, this would imply a high resource demand with likely similar outcomes for the datasets used herein. Therefore, we performed the procedure on only one climate dataset, UKESM1-0-LL with ssp585 (see sect. 2.9).

Provided the abundant data and high dimensionality (60 features), only two hyperparameters were selected for grid-search using 4-fold CV. These are depth of the trees (short depth) in steps of [8, 11, 14] and the maximum number of trees (short iterations) in steps of [400, 800, 1200, 1600]. The default grid-search procedure is implemented in CatBoost as follows: The dataset is split into 80% training and 20% test data. For all possible combinations of parameters (points of the grid), a model is fitted on the train dataset. Among the models, the one best performing on the test dataset is selected and sent to CV. Within the above defined grid, the first best model parameters were (14, 1600) achieving a test RMSE equal to 0.4446 (and test-RMSE-mean 0.4470 for 4-fold CV). The second-best model parameters were (14, 1200), test RMSE = 0.4682, followed by (11, 1600) with test RMSE = 0.4871. The experiments demonstrate that there is no overfitting, and results should be close to the lowest feasible generalization error for models fitted using this dataset. Even if a further small increase in accuracy is possible, it may deteriorate performance in emulator applications.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

With fixed depth = 14 and iterations = 1600, the remaining training parameters were left to default values. For further emulator training, climate scenarios (i.e., historical and three SSPs; Sect. 2.9) for each GCM were pooled and emulators trained on the whole sample as the other four GCMs not used in the training were subsequently used as novel data for benchmarking (see subsequent sections).

CatBoost provides three approaches to estimate feature importance: Prediction Values Change (PVC), Loss Function Change, and Shapley Additive Explanations (SHAP). The computational complexity of these approaches increases substantially in the same order. For example, computing SHAP values with the Python package SHAP (Lundberg et al., 2020) becomes computationally impractical for our datasets and models without further subsampling at a rate of 0.0001 and lower. PVC in turn is readily available after the training procedure. We hence select herein PVC, which quantifies the average level to which altering a feature value influences the predicted value. PVC importance values are non-negative and normalized so that their sum for all features equals 100.

2.7 Emulator evaluation metrics

In line with earlier studies on crop model emulator development (Blanc, 2017; Franke et al., 2020b; Oyebamiji et al., 2015), we use the root mean square error (RMSE) and linear regression statistics (Pearson's correlation coefficient R2, slope, and intercept) to evaluate emulator performance. The first also corresponds to the metric for the loss function in emulator training (see sect. 2.6). Objective of the performance evaluation is the emulators' skill in predicting crop yield simulation outputs for climate projections that have not been used in emulator training.

Evaluations are performed across all individual locations (simulation units) and years as well as for global and sub-continental area-weighted aggregates. For spatial aggregation, crop yields are area-weighted based on the extent of crop- and water management-specific harvested area in each 5-arcmin pixel. Harvested areas were sourced from the SPAM 2010 v2.0 dataset (International Food Policy Research Institute, 2020; Yu et al., 2020).

Besides prediction performance, we also approximate the computational time requirements for data pre-processing, crop model simulations, feature processing, and emulator predictions to provide an estimate of speed gain when using emulators. This is done by doing all processing and simulations on a computational cluster with an Oracle ZS5 network storage system and computational nodes equipped with Intel Xeon Gold 2.1 GHz CPUs. All processes are performed on single cores to ensure comparability. An exception is the emulator training, which is done on a GPU (Nvidia RTX A6000) as it would require unreasonably more time on a common CPU.

2.8 Global gridded crop model and simulation setup

EPIC-IIASA (Balkovič et al., 2014) is a GGCM based on the field-scale process-based crop model Environmental Policy Integrated Climate (EPIC) v0810 (Izaurralde et al., 2012; Williams et al., 1989). EPIC-IIASA has been applied extensively in global climate impact studies and has shown good skill in reproducing both historic absolute yields under business-as-usual management and inter-annual yield variability (Balkovič et al., 2018, 2013; Müller et al., 2017). Key processes of the core model EPIC are available from the prior references and summarized in (Folberth et al., 2016).

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

EPIC-IIASA is based on a 5 x 5' spatial grid (equivalent to about 8.3 km x 8.3 km near the equator) for soil characteristics and topography that are aggregated to homogenous response units based on classification of key land surface characteristics (soil, slope, elevation). These are intersected with a 30 x 30' climate grid (about 50 km x 50 km near the equator) and national administrative boundaries to define simulation units for each of which the crop model is eventually run (Skalský et al., 2008). Accordingly, simulation units vary in size from 5' x 5' to 30' x 30' depending on local heterogeneity. Globally, this results in nearly 162k simulation units within 66k climate pixels. Out of these, around 151k simulation units are included here based on general suitability for crop cultivation (i.e., soil present and sufficient temperature).

The setup and parameterization of the EPIC-IIASA GGCM was kept the same as in ISIMIP3b (Jägermeyr et al., 2021) except that we used here sufficient nitrogen fertilizer inputs to focus on climate signals. We selected maize as a model crop due to its nearly ubiquitous cultivation globally. All simulations assumed rainfed water supply only. The time period for simulations and evaluation is 1980-2099, spanning the historical climate baseline 1980-2014 and projections from 2015-2099. We skip the last year 2100 as outputs are reported by the year of planting (Müller et al., 2017) and no harvest takes place in the last simulation year if the crop is planted in autumn and harvested the following spring.

2.9 Input data

The same raw data were used for both GGCM simulations and emulator training and predictions. Several key input data (soil attributes, growing season dates, climate data), have been provided by the most recent phase 3b of ISIMIP and GGCMI initiative as documented in (Jägermeyr et al., 2021). Soil data are originally derived from the Harmonized World Soil Database (FAO et al., 2012) and have been processed for crop land by ISIMIP and GGCMI (Volkholz and Müller, 2020). For the experiment herein, we used soil attributes stored after a spin-up run of EPIC-IIASA, which had been used in the crop model climate impact simulations as well. Slope and elevation had earlier been derived from GTOPO30 (US Geological Survey, 2002).

Climate data were sourced from five global climate models GFDL-ESM4 (Dunne et al., 2020), IPSL-CM6A-LR (Boucher et al., 2020), MPI-ESM1-2-HR (Gutjahr et al., 2019), MRI-ESM2-0 (Yukimoto et al., 2019), and UKESM1-0-LL (Sellar et al., 2019) that span a representative range of equilibrium climate sensitivities (ECS) and transient climate response (TCS). Thereby, MPI-ESM1-2-HR and GFDL-ESM4 are at the low end, MRI-ESM2-0 is in the lower mid-range, and IPSL-CM6A-LR and UKESM1-0-LL present the high end of warming levels at the end of century. For each GCM, we use outputs for the historical time period, as well as the three RCPs 2.6, 7.0, and 8.5. In line with the source climate data combining identifiers for shared socio-economic pathways (SSPs) and RCPs without separators, we refer to the climate scenarios as ssp126 (SSP1 with RCP2.6), ssp370 (SSP3 with RCP7.0), and ssp585 (SSP5 with RCP8.5). Simulations were performed with transient annual atm. CO2 concentrations corresponding to those of the respective RCPs.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

3 Results

3.1 Training metrics

Individual emulators are trained on the pooled climate scenarios of each GCM and subsequently applied to each climate scenario of the same GCM individually. Regression statistics for the training show a near perfect fit with slope and intercept uniformly at 1.00 and -0.01 (except for UKESM1-0-LL with ssp585 at intercept=0.00) and R2 ranging between 0.982 and 0.986 (Table 3). The RMSE varies between 0.41 and 0.49 t ha^{-1}, apparently scaling with absolute yields. These are highest on average during the historical period and lowest under ssp585 (see also Figure 4). This is more so the case for the two GCMs with high ECS and consequently higher levels of global warming, namely IPSL-CM6A-LR and UKESM1-0-LL (see sect. 2.9).

Table 3. Regression statistics and RMSE for each emulator trained on all climate scenarios of a specific GCM and applied to each of the source GCM's climate scenarios.

GCM	Climate scenario	R2	Slope	Intercept	RMSE
GFDL-ESM4	historical	0.985	1.00	-0.01	0.48
IPSL-CM6A-LR	historical	0.986	1.00	-0.01	0.47
MPI-ESM1-2-HR	historical	0.985	1.00	-0.01	0.49
MRI-ESM2-0	historical	0.985	1.00	-0.01	0.48
UKESM1-0-LL	historical	0.986	1.00	-0.01	0.48
GFDL-ESM4	ssp126	0.984	1.00	-0.01	0.47
IPSL-CM6A-LR	ssp126	0.985	1.00	-0.01	0.45
MPI-ESM1-2-HR	ssp126	0.984	1.00	-0.01	0.48
MRI-ESM2-0	ssp126	0.985	1.00	-0.01	0.46
UKESM1-0-LL	ssp126	0.984	1.00	-0.01	0.45
GFDL-ESM4	ssp370	0.983	1.00	-0.01	0.44
IPSL-CM6A-LR	ssp370	0.984	1.00	-0.01	0.43
MPI-ESM1-2-HR	ssp370	0.983	1.00	-0.01	0.46
MRI-ESM2-0	ssp370	0.984	1.00	-0.01	0.44
UKESM1-0-LL	ssp370	0.983	1.00	-0.01	0.42
GFDL-ESM4	ssp585	0.983	1.00	-0.01	0.44
IPSL-CM6A-LR	ssp585	0.984	1.00	-0.01	0.42
MPI-ESM1-2-HR	ssp585	0.983	1.00	-0.01	0.45
MRI-ESM2-0	ssp585	0.984	1.00	-0.01	0.44
UKESM1-0-LL	ssp585	0.982	1.00	0.00	0.41

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

3.2 Prediction performance

3.2.1 Global prediction performance

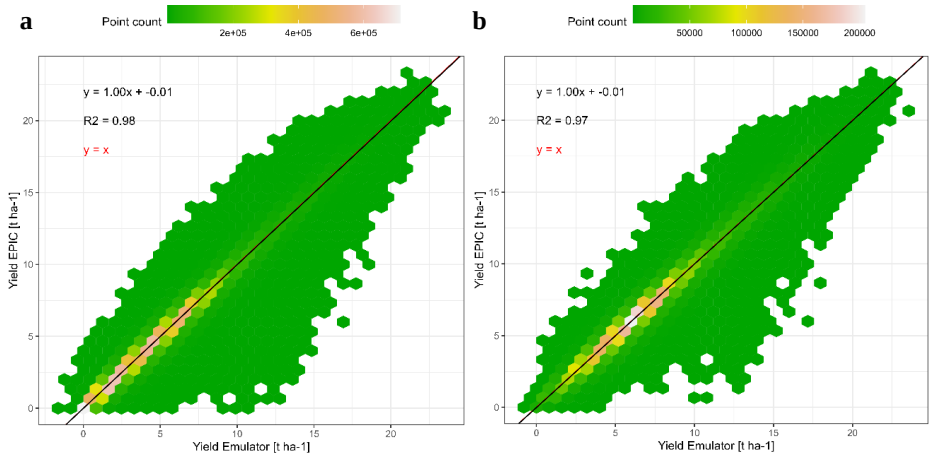

Figure 3. Comparison of exemplary global gridded crop yields for rainfed maize from EPIC-IIASA crop model simulations vs predictions by an emulator that was trained on the GCM IPSL-CM6A-LR and applied to the GCM GFDL-ESM4 for RCP8.5 in both cases with (a) all simulation units and (b) simulation units with >100 ha maize harvested area.

Applying the emulators to climate scenarios from GCMs not seen during training results in only slightly worse regression and RMSE statistics (see Table 4 for overview and Figure 3 for exemplary visualization). The R^2 now ranges between 0.974 and 0.980, the slope between 0.99 and 1.01, and the intercept between -0.05 and -0.01. The RMSE is between 0.49 and 0.62 t ha^{-1}. For both latter metrics, larger deviations from the training results occur for the historical time period and in GCMs and scenarios with lower levels of global warming. While the absolute difference is small, the change in RMSE presents an increase by 20 to 27% and indicates a slight overfitting of the emulators.

Considering only simulation units with rainfed maize harvested area > 100 ha slightly deteriorates the regression statistics (Figure 3b). Yet, this is at a lower number of samples (n=36 x 10^6 compared to n=127 x 10^6 in Figure 3a) and the point density indicates a more pronounced concentration of samples in the yield range 3 to 10 t ha^{-1} which may affect the regression compared to the wider distribution towards the origin if all pixels are included (Figure 3a).

Finally, both panels show that predicted yields may include negative values, which occurs in this example for 0.8% of samples in the whole dataset (minimum -1.4 t ha^{-1}; mean -0.04 t ha^{-1}) and 0.007% when masking by harvested area (minimum -0.7 t ha^{-1}; mean -0.07 t ha^{-1}). Emulator applications hence need to ensure that predictions are zeroed if valid prediction ranges cannot be defined *a priori* as is the case for the algorithm employed here.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

Table 4. Ranges of regression statistics and RMSEs for each emulator trained on a specific GCM and applied to all GCMs and climate scenario combinations in the demonstration example. Emulators based on the target GCM are excluded. E.g., the first row shows results of predictions for GFDL-ESM4 x historical from the emulators trained on the GCMs IPSL-CM6A-LR, MPI-ESM1-2-HR, MRI-ESM2-0, and UKESM1-0-LL and all climate scenarios (see Methods sect. 2.9).

GCM	Climate scenario	R2	Slope	Intercept	RMSE
GFDL-ESM4	historical	0.977 - 0.978	0.99 - 1.01	-0.05	0.59 - 0.60
IPSL-CM6A-LR	historical	0.977 - 0.979	1.00 - 1.01	-0.04	0.58 - 0.61
MPI-ESM1-2-HR	historical	0.976 - 0.978	1.00 - 1.01	-0.03	0.60 - 0.62
MRI-ESM2-0	historical	0.976 - 0.977	0.99 - 0.99	-0.03	0.59 - 0.61
UKESM1-0-LL	historical	0.977 - 0.978	0.99 - 1.01	-0.05	0.58 - 0.61
GFDL-ESM4	ssp126	0.978 - 0.979	1.00 - 1.01	-0.05	0.54 - 0.55
IPSL-CM6A-LR	ssp126	0.979 - 0.980	1.00 - 1.01	-0.03	0.52 - 0.54
MPI-ESM1-2-HR	ssp126	0.977 - 0.979	1.00 - 1.01	-0.03	0.55 - 0.57
MRI-ESM2-0	ssp126	0.977 - 0.978	0.99 - 1.00	-0.04	0.55 - 0.56
UKESM1-0-LL	ssp126	0.977 - 0.978	0.99 - 1.00	-0.03	0.52 - 0.53
GFDL-ESM4	ssp370	0.976 - 0.977	0.99 - 1.00	-0.03	0.52 - 0.53
IPSL-CM6A-LR	ssp370	0.975 - 0.977	1.00 - 1.01	-0.03	0.51 - 0.53
MPI-ESM1-2-HR	ssp370	0.977 - 0.978	1.00 - 1.01	-0.03	0.53 - 0.54
MRI-ESM2-0	ssp370	0.977 - 0.977	0.99 - 1.00	-0.03	0.52 - 0.53
UKESM1-0-LL	ssp370	0.974 - 0.976	0.99 - 1.00	-0.04	0.49 - 0.50
GFDL-ESM4	ssp585	0.976 - 0.977	1.00 - 1.00	-0.02	0.52 - 0.52
IPSL-CM6A-LR	ssp585	0.974 - 0.977	1.00 - 1.00	-0.03	0.50 - 0.53
MPI-ESM1-2-HR	ssp585	0.976 - 0.977	1.00 - 1.01	-0.02	0.53 - 0.54
MRI-ESM2-0	ssp585	0.976 - 0.977	0.99 - 1.00	-0.01	0.52 - 0.53
UKESM1-0-LL	ssp585	0.972 - 0.975	0.99 - 1.00	-0.05	0.49 - 0.52

Global area-weighted mean crop yields show equally a high agreement both between emulator predictions and outputs from the crop model and among the different emulators (Figure 4). Mean correlation coefficients range between 0.879 and 0.994 with higher values in scenarios with higher levels of warming, i.e. for 'hotter' GCMs such as UKESM1-0-LL or IPSL-CM6A-LR and the high concentration pathway ssp585. The lowest values occur at the opposite end of the spectrum (MPI-ESM1-2-HR and MRI-ESM2-0 with ssp126). Notably, the yield trends may also have an impact here as larger variance facilitates higher R^2. The ranges of R^2 values among the emulators applied to the same scenarios is marginal, indicating that the choice of the emulator has little impact on this global metric. Values for RMSE do not show this pattern, while there appears to be a trend towards similar values for the same target GCM (c.f. IPSL-CM6A-LR vs MRI-ESM2-0).

Noticeable deviations occur for specific periods and climate projections, such as the 2050s in ssp370 for IPSL-CM6A-LR and MPI-ESM1-2-HR. In these two instances, there is a high agreement among emulators but not with EPIC simulations. From the 2080s towards the end of century, there is a deviation in yield predictions from the emulator based on MPI-ESM1-2-HR compared to the EPIC simulations for UKESM1-0-LL x ssp585. In the first case, this may indicate particular climate patterns in the target dataset. In the latter, the high-end warming occurring for this scenario may not be reflected in any of the other scenarios used for emulator training.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

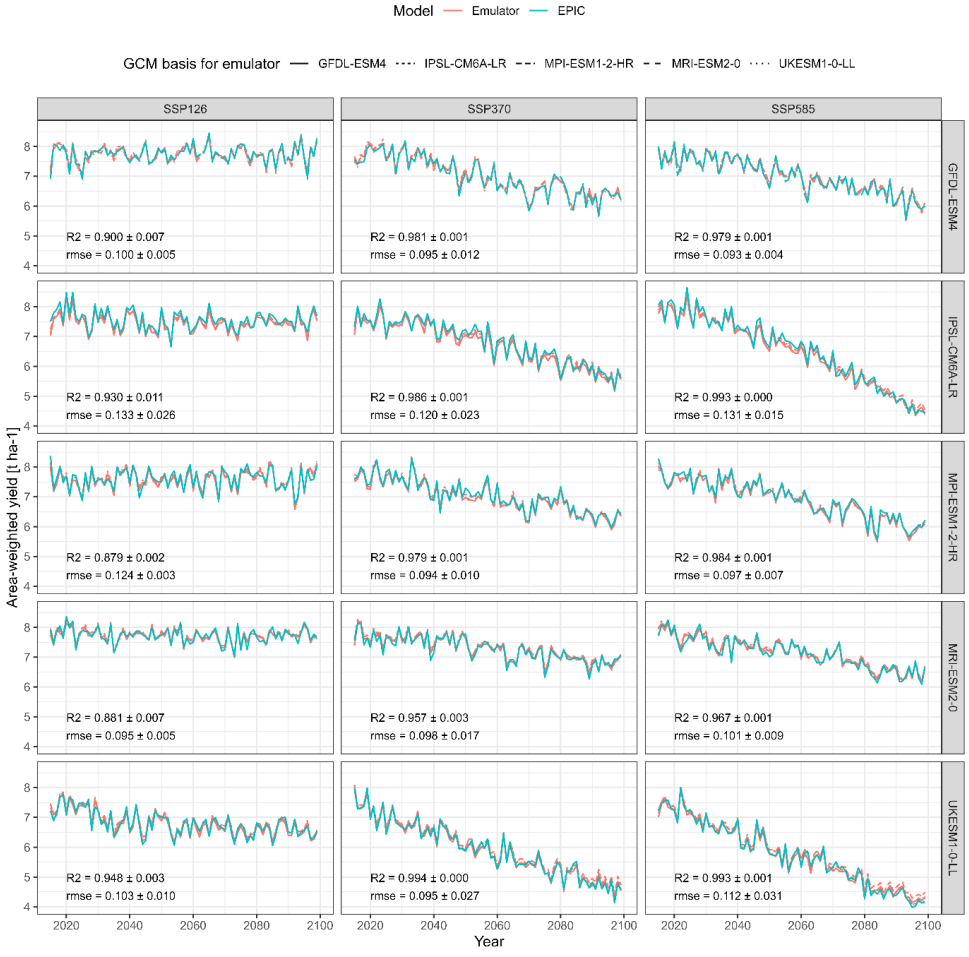

Figure 4. Global annual area-weighted yields of rainfed maize from the GGCM EPIC or predicted by the emulators between the years 2015-2099 for the five priority GCMs used in ISIMIP3b and three SSP-RCP combinations. Each panel shows predictions from four emulators trained on each of the five GCMs except the one providing the target features.

3.2.2 Spatial patterns

Aggregating area-weighted crop yields and predictions to geographic macro-regions - exemplary for UKESM1-0-LL x ssp585 - shows a similar pattern as the global performance but with a poorer turnout for both R^2 and RMSE in regions that have predominantly dry climate, i.e., Northern Africa and to lesser extents Australia and Central Asia with R^2=0.713, 0.889, and 0.900, respectively (Figure 5). Further deviations may at least in part be due to the selection of this high warming scenario.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

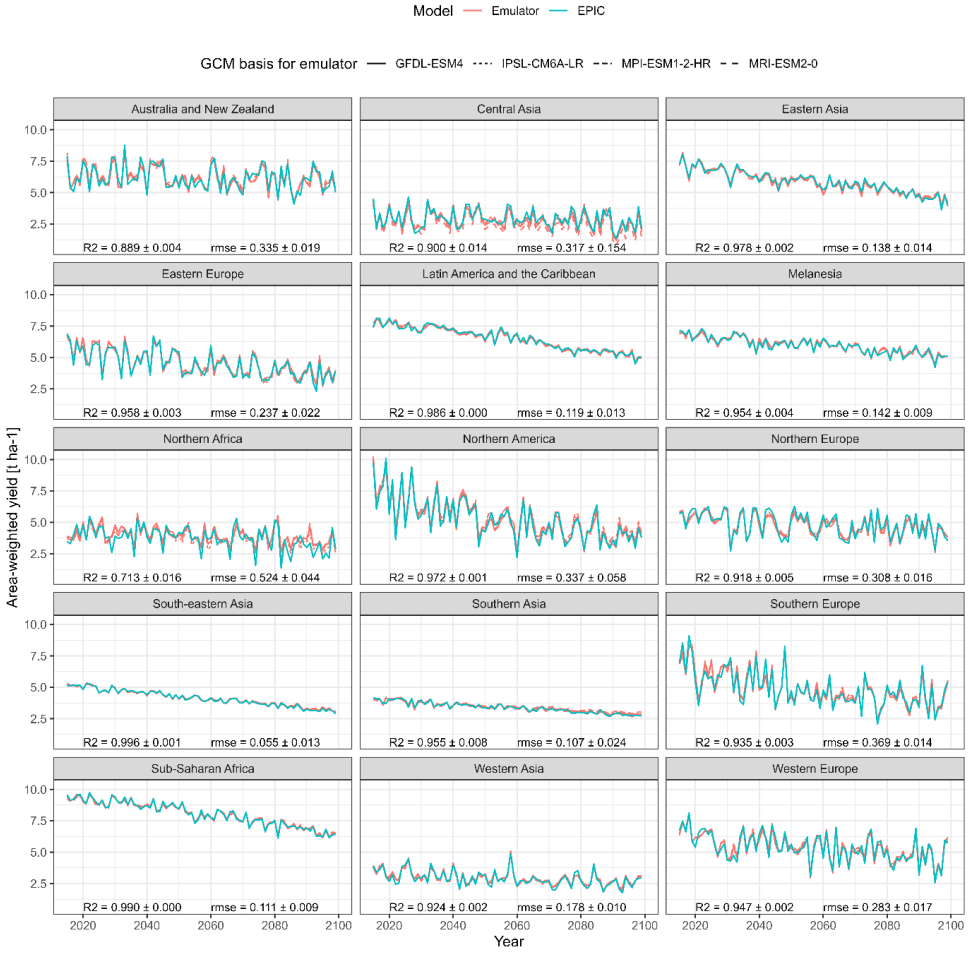

Figure 5. Same as Figure 4 but for 15 macro regions and target climate dataset UKESM1-0-LL x SSP585 only.

455 Within individual simulation units mapped to 5' x 5' pixels, high R^2 values dominate as well (Figure 6). These are mixed with very poor outcomes if the whole land mask is considered (Figure 6a) compared to masking by relevant cultivation regions (Figure 6b). In the first case, the median R^2 is 0.794, in the latter case 0.847. Hotspots for poor outcomes are arid regions - especially of the Sahel zone and West Asia - where permanently dry conditions cause constantly low yields with little variability. This also affects the outcomes of regression metrics.

460

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

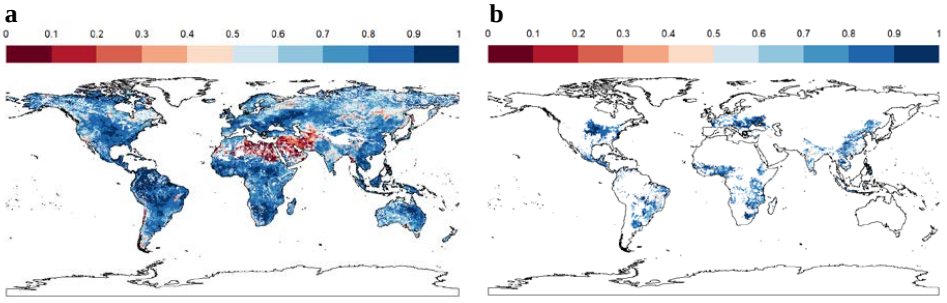

Figure 6. R2 of regressions between simulated and predicted rainfed maize yields over 85 years per pixel (i.e., simulation unit) exemplary shown for an emulator based on IPSL-CM6A-LR and applied to GFDL-ESM4 x SSP3-70 for (a) all land mask pixels and (b) pixels with >100 ha harvested area.

3.3 Feature importance

The importance of individual features shows overall good agreement among the emulators trained on different GCMs with slight variations (Table 5). While the top 10 features ranked by median importance are quite consistent, the agreement tends to decrease with decreasing importance of the features. The uniformly most important feature is the sum of shortwave solar radiation over the growing season (RADsumAGS), a direct aggregate of the photosynthetically active energy received by the crop. This is followed by the growing season precipitation sum (PRCPsumAGS). CMDlt0sumAGS, the number of days with a climatic moisture deficit, presents a drought indicator with similar ranking. Already beyond these three top ranking features, the numeric difference among prediction value change (PVC) outcomes is less discernible and shows a transient decline.

Notably, most of the climate features present in the top 20 refer to growing season aggregates, followed by drought-related features for the reproductive phase (PETsumAGSr, CMDsumAGSr, PRCPsumAGSr, CMDlt0sumAGSr), during which flowering and consequently yield formation is most sensitive to water deficit. Only one feature refers to the pre-growing season period (PGS), the average minimum daily temperature (TMNavPGS), which is not straightforward to interpret.

Non-climatic features include most importantly the crop's heat unit requirement (PHU), a spatially explicit cultivar constant, the closely related length of vegetation period (LVP), and the soil features PAW, PH, and DEPTH. While the first and the last of these relate to soil water storage and therefore modulate water deficit in interaction with weather, pH has typically little impact in the crop model and may hence be correlated with other features.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

Table 5. Feature importance for the 20 overall top-ranking features out of 60 features measured as prediction value change (PVC; see sect. 2.6). Median importance is the median of feature importance estimated for each of five individual emulators based on each of the GCMs.

Feature	Median importance	Range of importance	Rank of median importance	Range of rank
RADsumAGS	14.14	11.82 - 15.40	1	1 - 1
PRCPsumAGS	10.44	8.20 - 12.58	2	2 - 3
CMDlt0sumAGS	8.39	6.55 - 10.26	3	2 - 4
PHU	5.00	3.20 - 6.27	4	4 - 11
CMDsumAGS	4.91	3.26 - 5.24	5	4 - 10
TMXavAGS	4.38	3.82 - 7.05	6	3 - 8
PETsumAGSr	4.11	2.62 - 4.39	7	5 - 14
PAW	3.94	2.97 - 4.21	8	7 - 11
LVP	3.59	2.83 - 3.67	9	8 - 12
CMDsumAGSr	3.27	2.78 - 3.71	10	9 - 11
HURavAGS	3.10	2.48 - 3.91	11	8 - 13
GDDsumAGS	2.75	1.87 - 5.15	12	5 - 16
TMNavAGS	2.74	2.65 - 4.10	13	7 - 13
PRCPsumAGSr	2.31	1.32 - 3.33	14	10 - 19
CMDlt0sumAGSr	1.77	0.64 - 2.33	15	15 - 34
HDDsumAGS	1.64	1.49 - 2.84	16	13 - 18
PH	1.61	0.64 - 2.24	17	14 - 33
DEPTH	1.60	0.24 - 1.93	18	14 - 49
GSLsumAGSr	1.52	0.66 - 4.69	19	5 - 31
TMNavPGS	1.51	1.27 - 2.17	20	15 - 22

3.4 Computational performance

The total time required for producing EPIC simulations or an equivalent of crop yield predictions is not straightforward to compare as both approaches require different computational infrastructures (e.g. graphical processing unit (GPU) needed for reasonable performance in model training) and have their specific bottlenecks relating to computational and storage (read/write) demands. As time gain is a key advantage of emulators, we still provide a rough estimate of time required for key tasks within the modelling and data processing chains of both approaches (Figure 7) to allow for basic contextualization.

In the setup used herein, both approaches require first a conversion of netCDF files to binary files that provide substantially faster read access. This takes about 0.5h. Further production of daily weather files for the EPIC model – individual text files for each pixel - takes approx. 2h. The largest time requirement occurs for the EPIC simulation itself, which here takes 12h but can vary on the shared cluster between 6h and 18h on a single core. The crop model produces single output files for each simulation unit from which the extraction of outputs to a compilation file requires 1h. Once a climate dataset has been processed, only the last two steps crop model run and post-processing are required for each simulation.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

Within the CROMES pipeline, generation of climate features for one climate scenario for emulator training or predictions requires about 0.25h. Model training on a GPU using the CatBoost algorithm requires 0.5h and predictions, i.e., the combination of climate and other features with subsequent evaluation of the trained algorithm on the feature set, about 0.15h. Once an emulator has been trained, again only the last two steps are required, i.e., processing of climate features for a target dataset and evaluation of the emulator over the combined feature set.

In total, the emulator provides a speed improvement of at least an order of magnitude, regardless of whether the whole computational chain is considered or only the last two steps producing the actual outputs.

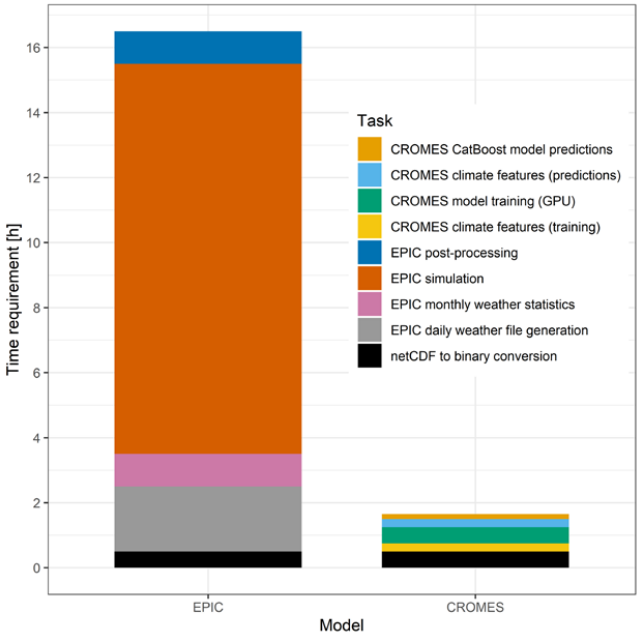

Figure 7. Time requirement for key tasks required to produce global crop model simulations with EPIC or crop yield predictions with CROMES. Some tasks only have to be performed once, essentially the bottom three of the legend or those relating to CROMES emulator training, depending on the specific purpose. The numbers shown here are therefore primarily for illustrative purposes.

4 Discussion

In principle, model emulators or meta-models present a trade in higher speed for less accuracy. Our evaluation of the CROMES pipeline for an exemplary application highlights that a substantial speed gain is in fact feasible at a comparably low cost in accuracy with most benchmark indicators pointing to a near perfect fit. The lowest agreement between predictions and crop yield simulations occurs in regions with predominantly arid climate where the aggregation of daily weather to climate features potentially fails to capture the effects of timing and volume of precipitation events. These can markedly affect crop yields as do interactions between temperature, atmospheric moisture deficit, and water availability (Schauberger et al., 2017). Yet, rainfed agriculture is typically

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

of limited importance in such regions and the constantly low crop yields pose a challenge to achieving a good regression fit while the absolute error can be considered minor.

To the authors' best knowledge, complex machine-learning algorithms have not been applied prior to train emulators for a GGCM using opportunistic training samples, i.e. data that are readily available from earlier experiments. The performance achieved herein is hence not straightforward to compare to that found in earlier studies. Most recently, (Sweet et al., 2023) evaluated cross-validation strategies for training machine-learning algorithms to predict crop yields from GGCMs. They reported a maximum R2 of 0.82 on the training set and far lower values around 0.4 on holdout data. However, their application case covered only the historic period and focused on holdout years and regions, which may be more challenging to capture than multi-year and –location climate change projections as herein. (Oyebamiji et al., 2015) developed a similar emulator approach as the one herein but using various regression methods and with the objective of predicting changes in decadal mean crop yields. Applied to an older version of the GGCM LPJmL (Bondeau et al., 2007), they found an agreement with R2=0.72 to 0.86 for unseen climate projections combining RCPs 4.5 and 8.5. Similarly, (Blanc, 2017) trained statistical emulators for crop yield changes under climate change based on various regression models for several GGCMs and samples from climate impact projections. This resulted in an R2 of 0.43 to 0.78 for multi-year average yield changes depending on the GGCM with R2 0.48 to 0.56 for an EPIC-based GGCM GEPIC. Finally, (Franke et al., 2020b) trained GGCM emulators using pixel-specific polynomials for a range GGCMs that had simulated a structured training sample with systematic changes in temperature, precipitation, CO2, and fertilizer application. Applied to an exemplary climate change projection (HadGEM2-ES with RCP8.5) this resulted in RMSE of 0.9 to 2.7 t ha-1 and 1.8 to 2.4 t ha-1 for two EPIC-based GGCMs compared to herein R2=0.97 to 0.98 and RMSE=0.50 to 0.66 on holdout data.

Identifying the reasons for deviations among different emulator development is beyond the scope of this study. Yet, we expect that feature engineering has a key influence on the outcomes. Essentially, earlier studies employed seasonal, monthly, or annual aggregates of climate variables. In turn, CROMES dynamically estimates the actual length of each growing season and its sub-phases based on GDD accumulation. This has earlier been found to be a key determinant of crop yields in GGCMs, especially under high levels of global warming. Essentially, crops mature earlier and have less time for biomass accumulation but may simultaneously not be affected by adverse weather events later in the year (Zabel et al., 2021).

Computational speed is challenging to compare between emulators and GGCMs (see sect. 3.4) and even more so among different studies. These may cover varying GGCMs with highly diverse computational demands or use publicly available training data that do not provide this information. Herein, we estimate a speed gain of conservatively an order of magnitude. (Oyebamiji et al., 2015) estimate a speed gain by a factor of 60 for their LPJmL emulator, yet without further specifications of considered steps in the modelling chain. Essentially, in both cases the time requirement decreases from hours to minutes. Based on our results, the largest gain in computational speed is achieved if an emulator is applied for comprehensive scenario analyses, e.g., across large sets of climate projections, which requires a large number of repeated runs of the same emulator.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

5 Conclusions and outlook

We expect the crop model emulator pipeline presented herein to bear great potential in various applications including complex climate impact modelling clusters or comprehensive scenario analyses. The loss of information compared to the gain in speed indicates that outcomes can be considered robust as long as predictors are part of the training domain. Quantifying this validity domain remains a prevailing issue in machine learning and will have to be characterized on a case-by-case basis until robust methods are developed. This will be an important subject for future research. Meanwhile, compared to static emulators CROMES allows for continuous updating of training data such as for the next generation of CMIP7 climate projections, with new GGCM versions, or for applications with very specific feature domains such as global cooling scenarios from geoengineering or nuclear winter.

Beyond the crop model emulation, we expect CROMES to be useful in two ways: (a) as the input data are quite generic, CROMES can also be used to efficiently train machine learning models on observations to develop observation-based machine-learning crop models; and (b) the climate features as an intermediary product of the pipeline allow for comprehensive analyses of growing season climate itself.

Code availability

A frozen version of the code required to reproduce the study is available at https://doi.org/10.5281/zenodo.14901127.

Data availability

Data derived from crop model simulations, pre-processed features, and other data required to reproduce the results presented herein are available at https://doi.org/10.5281/zenodo.14894075. Raw data sources are provided in the repository and in the text.

Author contribution

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

Competing interests

CF is a topic editor of the journal Geoscientific Model Development.

Acknowledgements

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.

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

References

Balkovič, J., Skalský, R., Folberth, C., Khabarov, N., Schmid, E., Madaras, M., Obersteiner, M., Velde, M. van der, 2018. Impacts and Uncertainties of +2°C of Climate Change and Soil Degradation on European Crop Calorie Supply. Earth's Future 6, 373–395. https://doi.org/10.1002/2017EF000629

Balkovič, J., van der Velde, M., Schmid, E., Skalský, R., Khabarov, N., Obersteiner, M., Stürmer, B., Xiong, W., 2013. Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation. Agricultural Systems 120, 61–75. https://doi.org/10.1016/j.agsy.2013.05.008

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

Blanc, É., 2017. Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models. Agricultural and Forest Meteorology 236, 145–161. https://doi.org/10.1016/j.agrformet.2016.12.022

Blanc, E., Sultan, B., 2015. Emulating maize yields from global gridded crop models using statistical estimates. Agricultural and Forest Meteorology 214–215, 134–147. https://doi.org/10.1016/j.agrformet.2015.08.256

Bondeau, A., Smith, P.C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Müller, C., Reichstein, M., Smith, B., 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. Global Change Biology 13, 679–706. https://doi.org/10.1111/j.1365-2486.2006.01305.x

Boucher, O., Servonnat, J., Albright, A.L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, E., Lionel, Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A.K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., Vuichard, N., 2020. Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. Journal of Advances in Modeling Earth Systems 12, e2019MS002010. https://doi.org/10.1029/2019MS002010

Dunne, J.P., Horowitz, L.W., Adcroft, A.J., Ginoux, P., Held, I.M., John, J.G., Krasting, J.P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C.A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K.A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P.P.G., Griffies, S.M., Guo, H., Hallberg, R.W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly, P.C.D., Nikonov, S., Paynter, D.J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B.G., Robinson, T., Schwarzkopf, D.M., Sentman, L.T., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A.T., Wyman, B., Zeng, Y., Zhao, M., 2020. The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): Overall Coupled Model Description and

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

Simulation Characteristics. Journal of Advances in Modeling Earth Systems 12, e2019MS002015. https://doi.org/10.1029/2019MS002015

FAO, IIASA, ISRIC, ISSCAS, JRC, 2012. Harmonized World Soil Database (version 1.2).

Folberth, C., Baklanov, A., Balkovič, J., Skalský, R., Khabarov, N., Obersteiner, M., 2019. Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning. Agricultural and Forest Meteorology 264, 1–15. https://doi.org/10.1016/j.agrformet.2018.09.021

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L.B., Obersteiner, M., Velde, M. van der, 2016. Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. Nature Communications 7, 11872. https://doi.org/10.1038/ncomms11872

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., Moyer, E.J., 2020a. 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/10.5194/gmd-13-2315-2020

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., Moyer, E.J., 2020b. 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/10.5194/gmd-13-3995-2020

Franke, J.A., Müller, C., Minoli, S., Elliott, J., Folberth, C., Gardner, C., Hank, T., Izaurralde, R.C., Jägermeyr, J., Jones, C.D., Liu, W., Olin, S., Pugh, T.A.M., Ruane, A.C., Stephens, H., Zabel, F., Moyer, E.J., 2022. Agricultural breadbaskets shift poleward given adaptive farmer behavior under climate change. Global Change Biology 28, 167–181. https://doi.org/10.1111/gcb.15868

Gao, X., Sokolov, A., Schlosser, C.A., 2023. A Large Ensemble Global Dataset for Climate Impact Assessments. Sci Data 10, 801. https://doi.org/10.1038/s41597-023-02708-9

Gebrechorkos, S., Leyland, J., Slater, L., Wortmann, M., Ashworth, P.J., Bennett, G.L., Boothroyd, R., Cloke, H., Delorme, P., Griffith, H., Hardy, R., Hawker, L., McLelland, S., Neal, J., Nicholas, A., Tatem, A.J., Vahidi, E., Parsons, D.R., Darby, S.E., 2023. A high-resolution daily global dataset of statistically downscaled CMIP6 models for climate impact analyses. Sci Data 10, 611. https://doi.org/10.1038/s41597-023-02528-x

Goulart, H.M.D., van der Wiel, K., Folberth, C., Boere, E., van den Hurk, B., 2023. Increase of Simultaneous Soybean Failures Due To Climate Change. Earth's Future 11, e2022EF003106. https://doi.org/10.1029/2022EF003106

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J.H., von Storch, J.-S., Brüggemann, N., Haak, H., Stössel, A., 2019. Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP). Geoscientific Model Development 12, 3241–3281. https://doi.org/10.5194/gmd-12-3241-2019

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

690 Hancock, J.T., Khoshgoftaar, T.M., 2020. CatBoost for big data: an interdisciplinary review. J Big Data 7, 94. https://doi.org/10.1186/s40537-020-00369-8

International Food Policy Research Institute, 2020. Global Spatially-Disaggregated Crop Production Statistics Data for 2010 Version 2.0. https://doi.org/10.7910/DVN/PRFF8V

Izaurralde, R.C., McGill, W.B., Williams, J.R., 2012. Development and application of the EPIC model for carbon
695 cycle, greenhouse gas mitigation, and biofuel studies, in: Managing Agricultural Greenhouse Gases. Elsevier, pp. 293–308.

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,
700 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., Rosenzweig, C., 2021. 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

Karger, D.N., Lange, S., Hari, C., Reyer, C.P.O., Conrad, O., Zimmermann, N.E., Frieler, K., 2023. CHELSA-
705 W5E5: daily 1 km meteorological forcing data for climate impact studies. Earth Syst. Sci. Data 15, 2445–2464. https://doi.org/10.5194/essd-15-2445-2023

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

Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., Schulin, R., 2016. Global investigation of impacts of PET
710 methods on simulating crop-water relations for maize. Agricultural and Forest Meteorology 221, 164–175. https://doi.org/10.1016/j.agrformet.2016.02.017

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9

715 Minoli, S., Müller, C., Elliott, J., Ruane, A.C., Jägermeyr, J., Zabel, F., Dury, M., Folberth, C., François, L., Hank, T., Jacquemin, I., Liu, W., Olin, S., Pugh, T.A.M., 2019. Global Response Patterns of Major Rainfed Crops to Adaptation by Maintaining Current Growing Periods and Irrigation. Earth's Future 7, 1464–1480. https://doi.org/10.1029/2018EF001130

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

725 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., Zabel, F., 2021. 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

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

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., Elliott, J.W., 2024. 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

Nelson, G.C., Valin, H., Sands, R.D., Havlík, P., Ahammad, H., Deryng, D., Elliott, J., Fujimori, S., Hasegawa, T., Heyhoe, E., Kyle, P., Lampe, M.V., Lotze-Campen, H., d'Croz, D.M., Meijl, H. van, Mensbrugghe, D. van der, Müller, C., Popp, A., Robertson, R., Robinson, S., Schmid, E., Schmitz, C., Tabeau, A., Willenbockel, D., 2014. Climate change effects on agriculture: Economic responses to biophysical shocks. PNAS 111, 3274–3279. https://doi.org/10.1073/pnas.1222465110

Oyebamiji, O.K., Edwards, N.R., Holden, P.B., Garthwaite, P.H., Schaphoff, S., Gerten, D., 2015. Emulating global climate change impacts on crop yields. Statistical Modelling 15, 499–525. https://doi.org/10.1177/1471082X14568248

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features, in: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.

Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., Pugh, T.A.M., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., Frieler, K., 2017. Consistent negative response of US crops to high temperatures in observations and crop models. Nature Communications 8, 13931. https://doi.org/10.1038/ncomms13931

Sellar, A.A., Jones, C.G., Mulcahy, J.P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F.M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S.T., Kelley, D.I., Ellis, R., Johnson, C.E., Walton, J., Abraham, N.L., Andrews, M.B., Andrews, T., Archibald, A.T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G.A., Gedney, N., Griffiths, P.T., Harper, A.B., Hendry, M.A., Hewitt, A.J., Johnson, B., Jones, A., Jones, C.D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R.J., Predoi, V., Robertson, E., Siahaan, A., Smith, R.S., Swaminathan, R., Woodhouse, M.T., Zeng, G., Zerroukat, M., 2019. UKESM1: Description and Evaluation of the U.K. Earth System Model. Journal of Advances in Modeling Earth Systems 11, 4513–4558. https://doi.org/10.1029/2019MS001739

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

Stockle, C.O., Williams, J.R., Rosenberg, N.J., Jones, C.A., 1992. A method for estimating the direct and climatic effects of rising atmospheric carbon dioxide on growth and yield of crops: Part I—Modification of the EPIC model for climate change analysis. Agricultural Systems 38, 225–238. https://doi.org/10.1016/0308-521X(92)90067-X

Sweet, L., Müller, C., Anand, M., Zscheischler, J., 2023. Cross-validation strategy impacts the performance and interpretation of machine learning models. Artificial Intelligence for the Earth Systems 1–35. https://doi.org/10.1175/AIES-D-23-0026.1

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

https://doi.org/10.5194/egusphere-2025-862
Preprint. Discussion started: 24 March 2025

US Geological Survey, 2002. GTOPO30 - 30 arc seconds digital elevation model from US Geological Survey. https://doi.org/10.5066/F7DF6PQS

Volkholz, J., Müller, C., 2020. ISIMIP3 soil input data. https://doi.org/10.48364/ISIMIP.942125

Wartenburger, R., Seneviratne, S.I., Hirschi, M., Chang, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Gosling, S.N., Gudmundsson, L., Henrot, A.-J., Hickler, T., Ito, A., Khabarov, N., Kim, H., Leng, G., Liu, J., Liu, X., Masaki, Y., Morfopoulos, C., Müller, C., Schmied, H.M., Nishina, K., Orth, R., Pokhrel, Y., Pugh, T.A.M., Satoh, Y., Schaphoff, S., Schmid, E., Sheffield, J., Stacke, T., Steinkamp, J., Tang, Q., Thiery, W., Wada, Y., Wang, X., Weedon, G.P., Yang, H., Zhou, T., 2018. Evapotranspiration simulations in ISIMIP2a-Evaluation of spatio-temporal characteristics with a comprehensive ensemble of independent datasets. Environmental Research Letters 13. https://doi.org/10.1088/1748-9326/aac4bb

Williams, J.R., 1990. The erosion-productivity impact calculator (EPIC) model: a case history. Phil. Trans. R. Soc. Lond. B 329, 421–428. https://doi.org/10.1098/rstb.1990.0184

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

Yu, Q., You, L., Wood-Sichra, U., Ru, Y., Joglekar, A.K.B., Fritz, S., Xiong, W., Lu, M., Wu, W., Yang, P., 2020. A cultivated planet in 2010 – Part 2: The global gridded agricultural-production maps. Earth System Science Data 12, 3545–3572. https://doi.org/10.5194/essd-12-3545-2020

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y., Ishii, M., 2019. The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: Description and Basic Evaluation of the Physical Component. Journal of the Meteorological Society of Japan. Ser. II 97, 931–965. https://doi.org/10.2151/jmsj.2019-051

Zabel, F., Müller, C., Elliott, J., Minoli, S., Jägermeyr, J., Schneider, J.M., Franke, J.A., Moyer, E., Dury, M., Francois, L., Folberth, C., Liu, W., Pugh, T.A.M., Olin, S., Rabin, S.S., Mauser, W., Hank, T., Ruane, A.C., Asseng, S., 2021. Large potential for crop production adaptation depends on available future varieties. Glob Change Biol gcb.15649. https://doi.org/10.1111/gcb.15649