CROMES - A fast and efficient machine learning emulator pipeline for gridded crop models

Christian Folberth1, Artem Baklanov2, Nikolay Khabarov2, Thomas Oberleitner1, Juraj Balkovič2, Rastislav Skalský1
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

Background
Global gridded crop models (GGCMs) have become state-of-the-art tools in large-scale climate impact and adaptation assessments. 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 structures. Emulators mimicking such models are therefore emerging as an attractive option to produce reasonable predictions of GGCMs’ crop productivity estimates at much lower computational costs. However, such emulators’ flexibility is thus far typically limited in terms of crop management flexibility and spatial resolutions among others.

Here we present a new emulator pipeline CROp model Machine learning Emulator Suite (CROMES) that serves for processing climate features from netCDF input files, combining these with site-specific features (soil, topography), and crop management specifications (planting dates, cultivars, irrigation) to train machine learning crop meta-models (MLCMs) and subsequently produce predictions (Figure 1).

Performance
Predictions require for a first used climate dataset about 45 min to convert from netCDF to a faster readable binary file format and 10 min for any subsequent scenario, including climate feature generation and predictions, compared to approx. 14 h for a GGCM simulation on the same system.

The best agreement between predictions and crop model simulations occurs in (sub-) tropical regions, the poorest in cold and arid climates (Figure 4) where both growing season length and water availability limit crop growth. The performance slightly deteriorates if fertilizer supply is considered, more so at low levels of nutrient inputs.

Outlook
Importantly, emulators produced by CROMES are virtually scale-free as all training samples, i.e., pixels, are pooled and treated as individual points without geo-referencing. This allows for applications on increasingly available high-resolution climate datasets or in regional studies for which more granular data may be available than at global scales. Using climate features based on crop growing seasons and cardinal growth stages, also adaptation studies such as growing season and cultivar shifts are facilitated. We expect CROMES to enable explorations of comprehensive climate projection ensembles, studies of dynamic climate adaptation scenarios, and cross-scale impact and adaptation assessments.

References:


Figure 1: Processing and computational pipeline to produce machine learning crop meta-model (MLCM) predictions in a fully integrated framework. The same design is used for MLCM training.

Figure 2: Framework of the project modified from (Folberth et al., 2019) starting from a multi-dimensional cube of crop model input data at the top right and resulting in climate impact and adaptation projections at the bottom. Letters A-F indicate key steps in the methodology: A: Global gridded crop model simulations for a multidimensional cube of input data to generate training data for machine learning algorithms, B: training of machine-learning crop meta-models (MLCMs) based on various machine learning algorithms and the global GGCM training sample, C: predictions for comprehensive sets of GGCMs and derived high-resolution climate data, other agro-environmental features, and management trajectories, D: storage and combination of MLCM predictions, E: extraction of training data and GGCM simulation space if required, F: processing and interpretation of outcomes from benchmarking to adaptation pathway development, G: inclusion of field observations and associated data in MLCM training data space as an add-on to A-F. Dimensions of the training data cube are: C=atm. CO2 concentration, W=precipitation (incl. sufficient water supply), T=temperature, N=mineral nitrogen input, S=soil type, V=crop variety, MLCMs=machine-learning crop meta-models.

Figure 3: (a) Comparison of global gridded crop yields for rainfed maize from EPIC crop model simulations vs predictions by an ML model that was trained on one GCM and applied to another GCM for RCP8.5 in both cases. (b) Global area-weighted crop yields for the same data over time for the original model EPIC and the emulator.

Figure 4: Same as Figure 3b but for macro-regions.