Project: **1515**

Project title: **Digital solutions for attributing climate change impacts on child nutrition and health (CLIMAKID)**

Principal investigator: **Sabine Undorf**

Report period: **2025-05-01 to 2026-04-30**

Maximum of 2 pages including figures. 9 pt minimum font size.

Overview

During the first allocation period, the e-CLIMAKID project established the core technical and computational foundations required to support large-scale climate–agriculture–health attribution analyses. The work focused on developing, testing, and integrating workflows for climate data processing and impact modelling within the DKRZ HPC environment. This phase represents a transition from initial system design to operational capability, including the generation of first datasets and the preparation of scalable workflows for subsequent production runs. Method papers based on the BASD performed at Levante are expected to submit within the next three moths. We have succesfully used the first set of processed climate data and crop modelling setups for a science workshop in Delhi, India, with 18 participants, and four regional climate-agriculture-health attribution papers are currently in preparation based on this work, and two PhD students at Kassel University are using the climate data for climate-agriculture atribution studies regarding the focus regions.

Climate data processing

A central component of the project is the bias correction and statistical downscaling of large-ensemble CMIP6 climate data to make them suitable for impact modelling and attribution studies. During the reporting period, the original ISIMIP3BASD method was for validation and comparison purposes applied to one climate model large ensemble within the DKRZ environment, and the ISIMIP3BASD-LE workflow was implemented, tested, and applied to initial sets of simulations to validate the methodology, assess computational performance, and ensure compatibility with downstream workflows. This work included refinement of preprocessing steps such as regridding and reference period selection, as well as validation of outputs against observational datasets. The result is a set of reproducible workflows that are now technically robust and ready to be scaled to the full ensemble in the next allocation period. A first set of BASD-ed 50-member MIROC6 ensembles (hist, hist-nat, ssp245) has been handed to the impact modellers.

Crop model integration

Progress was also made in integrating agricultural impact models into the computational pipeline. APSIM workflows were partially parallelized and linked to the climate processing outputs, while DSSAT was successfully deployed in a containerized environment and connected to SLURM for batch execution. Integration of InfoCrop into a highly parallelized workflow that can then be scaled to multiple nodes has also been achieved using a Singularity container running a barebones Windows OS that runs the Windows based InfoCrop executable. The Docker image used was WINE. The Wine application makes it possible to run Windows programs alongside any Unix-like operating system, particularly Linux. Initial calibration runs using observational climate data were conducted, and workflows were tested for compatibility with bias-corrected climate inputs. These steps are essential for enabling ensemble-based impact simulations and attribution analyses in the next allocation period. The crop modellers (especially using APSIM) have performed some of the project simulations on Levante, constrained by limited DKRZ compute time.

Workflow orchestration and HPC integration

To manage the complexity of the multi-step workflows, an orchestration framework based on Apache Airflow is currently still in development. At the moment, we are developing the components that will be used before AirFlow is implemented and integrated with SLURM. This will enable automated execution of climate processing and impact modelling pipelines through structured task dependencies. The system supports reproducible data pipelines and facilitates the generation and reuse of intermediate datasets across analyses.

This integration significantly improved workflow efficiency and reliability, allowing for coordinated execution of computational tasks and monitoring of job status across the DKRZ infrastructure. It also provides the basis for scaling workflows to larger datasets and more complex analyses in the next phase.

Scientific integration and user engagement

In parallel with technical development, the project initiated interdisciplinary integration across climate, agricultural, and health domains. The scripts underlying the first climate-agriculture-health study in preparation have been developed and repositoried in GitLab, and tested, applied, and further developed by the science workshop users with the help of DKRZ workshop accounts. So far, the scripts do climate data subsetting, computing a range of climatic indices on the gridbox level, aggregation to spatial domains, deriving climate attribution statements, mapping to district and matching with household level-health data, and similar functionalities at the agriculture/health and health modelling and attribution nexus. Early-stage workshops and collaborative activities were conducted to define user requirements, test workflow usability, and identify priority use cases for the MILK platform. These interactions informed both the design of computational workflows and the development of user-facing capabilities..

Data generation and storage usage

The allocation period resulted in the generation of initial datasets required for testing and validating the integrated workflows. Storage usage reached approximately 116 TiB out of 162 TiB allocated. This reflects substantial progress while remaining below the full projected volume, as processing was intentionally limited to selected ensemble members and regions during workflow validation.

Compute resources were also used efficiently, with approximately 7,729 node-hours consumed out of 8,320 node-hours allocated, indicating near-full utilization. Compute usage was primarily associated with development and testing of bias correction and downscaling pipelines, initial ensemble processing, and crop model calibration and pilot simulations. These activities also enabled performance testing, demonstrating strong scalability of both climate data processing and selected crop model workflows.

Challenges and outlook

The main challenges encountered during this allocation period related to the coordination and integration of workflows across multiple domains and systems. In particular, harmonizing data formats and variables across climate and impact models, managing large data volumes and I/O constraints, and integrating heterogeneous crop models required iterative refinement of workflows and processing strategies. Additionally, we found that random variable/member/scenario/model CMIP6 data are missing from the DKRZ file system, and reaching out to modelling groups and fellow ata users to fill these gaps takes time and effort.

Despite these challenges, the project has successfully established a functional and scalable computational framework. Derived data, models, and scripts are being used for scientific studies in the climate-agriculture-health impact attribution area. If again granted resources from DKRZ, we will in the next allocation period build on this foundation by scaling to full ensemble processing, expanding crop model simulations, and enabling user-driven analyses through the MILK platform within the project and beyond.

Selected papers submitted/in prep.

- Mora, C. et al., Attributing global food security impacts to anthropogenic climate change and trade policy: The 2010 Russian wheat failure, ERL, submitted
- Blum, T., et al., Tropentag 2026, Göttingen, Germany
- Abigaba, D., et al., Micronutrient Forum's 7th Global Conference, Accra, Ghana
- Nübler, L., et al., EGU
- Brouillet, A., et al., EGU
- Undorf, S., et al., EGU