EGU24-9083, updated on 08 Aug 2024
https://doi.org/10.5194/egusphere-egu24-9083
EGU General Assembly 2024

A Novel Bias-Adjustment Methodology for Streaming Global Climate Models

Ehsan Sharifi[1], Katherine Grayson[2], Sebastian Müller[1], and Stephan Thober[1]
[1]UFZ The Helmholtz Centre for Environmental Research, Department of Computational Hydrosystems, Leipzig, Germany (ehsan.sharifi@ufz.de)
[2]Barcelona Supercomputing Center (BSC)

Projections generated by global climate models (GCMs) are increasingly utilized to inform climate adaptation policies. It is known that climate models simplify the real climate system, leading to biases between simulated and observed climates. The spatial and temporal resolution of GCMs is ever increasing to provide a better representation of the Earth system and in turn, also provide higher quality information for users. To effectively handle the substantial climate data produced by these models, which can reach Terabytes to Petabytes, the Destination Earth (DestinE) initiative is exploring data streaming—a new approach that enables user applications to run Earth system models in an end-to-end workflow directly downstream of the climate simulations, eliminating the need to store entire time-series of variables to disk.

Traditional methods for quantile or percentile calculation typically involve sorting the data and directly computing the specific value corresponding to the desired quantile. These methods can be computationally intensive, especially for large datasets, as it necessitates storing and processing the entire dataset. While traditional bias-adjustment (BA) algorithms rely on data being fully available, a further challenge lies in developing bias-adjustment procedures capable of accommodating streamed data on-the-fly. In the DestinE Climate Digital Twin (CDT), we extend the quantile-mapping technique used in the ISI-MIP project (isimip.org) because it is a well-established method and preserves the trend of the original data. The technique involves aligning the CDFs of the model data with those of the observed data by adjusting the model's cumulative distribution to match that of the observed data. The enhancements of the BA method in DestinE-CDT is making use of the T-Digest algorithm, a sophisticated strategy that dynamically clusters data points into small groups, which is used to generate a summarized representation of the data distribution from streamed data and accurately calculate percentiles. This clustering technique offers an accurate estimate of percentiles while efficiently managing large and unbounded data streams where new data points are continuously added.

We apply the developed quantile-mapping BA for different variables on a global scale and compare it with the parametric distribution functions used in quantile-mapping BA from the ISI-MIP project.