

PROJENE: High-Resolution Bias-Corrected Climate Projections Dataset (CMIP6) for Northeast Brazil

Dataset Overview

This dataset contains high-resolution ($0.25^\circ \times 0.25^\circ$) daily climate projections for Brazil, with a particular focus on capturing the pronounced climate variability of the Northeast region (NEB). The data originates from Global Climate Models (GCMs) of the CMIP6, which were statistically bias-corrected using the Quantile Delta Mapping (QDM) method against the Brazilian Daily Weather Gridded Data (BR-DWGD).

Directory Structure

The files provided in this repository are structured chronologically and categorised by variable, processing level, and model type.

```
dataset/
├── {variable_id}/                                # E.g., 'pr', 'tasmax', 'tasmin'
│   ├── raw/                                     # Original GCM data (Regridded but NOT corrected)
│   └── bc_qdm/                                  # Bias-Corrected Data via QDM (Main dataset)
│       ├── ensemble/                           # Multi-Model Ensemble averages
│       │   ├── {variable}_day_ENSEMBLE_historical_bc_v1.1.0.nc
│       │   ├── {variable}_day_ENSEMBLE_ssp245_bc_v1.1.0.nc
│       │   └── ...
│       └── models/                             # Individual CMIP6 models
│           ├── hist/                           # Historical baseline (1961-2014)
│           ├── ssp245/                         # Intermediate mitigation scenario (2015-2100)
│           └── ssp585/                         # High emission scenario (2015-2100)
```

Variables Included

- pr: Daily Precipitation (mm/day)
- tasmax: Maximum Daily Near-Surface Air Temperature ($^\circ\text{C}$)
- tasmin: Minimum Daily Near-Surface Air Temperature ($^\circ\text{C}$)
- rsds: Surface Downwelling Shortwave Radiation ($\text{MJ}/\text{m}^2/\text{day}$)
- hurs: Near-Surface Relative Humidity (%)
- sfcwind: Near-Surface Wind Speed (m/s)

File Naming Convention

The NetCDF4 files follow a strict semantic naming convention to ease programmatic sorting and filtering:

`<variable>_day_<model_name>_<scenario>_bc_v1.1.0.nc`

Example: `pr_day_ACCESS-CM2_historical_bc_v1.1.0.nc`

- pr: Variable (Precipitation)
- day: Temporal resolution (Daily)
- ACCESS-CM2: CMIP6 Model identifier
- historical: Scenario (Historical Run)
- bc: Indicates the file underwent Bias Correction
- v1.1.0: Internal dataset (release) version

Processing Tiers & Validation

1. **bc_qdm/ (Strongly Recommended):** These files have undergone rigorous Quantile Delta Mapping (QDM). Systemic biases are removed, and physical delta trends are preserved. These are intended for hydrological modelling, impact assessments, and derivation of ETCCDI extremes.
2. **raw/:** These files were regridded to the 0.25° resolution but did NOT undergo bias correction. They are provided solely for validation purposes (e.g., assessing the QDM performance against the raw GCM anomalies).
3. **validation/ (Available for 'pr', 'tasmin' and 'tasmax'):** Within the validation directory, users will find out-of-sample skill evaluations for the indicated variables. These resources explicitly compare the post-processed data (Historical scenario) against independent observational records spanning the 1996–2014 period, providing quantitative validation metrics and, for precipitation, how the QDM algorithm improved the representation of daily extremes (ETCCDI) and mean regimes compared to the raw outputs.

Usage & Python Examples (xarray)

Given the size of the temporal arrays (daily steps up to 2100), the dataset is provided in NetCDF4 format. We strongly recommend using the xarray library in Python for multi-dimensional operations.

```
import xarray as xr

# Load a Bias-Corrected precipitation file for the SSP5-8.5 scenario
ds = xr.open_dataset('bc_qdm/models/ssp585/pr_day_BCC-CSM2-MR_ssp585_bc_v1.1.0.nc')

# Inspect dimensions (time, lat, lon) and the 'pr' variable
print(ds['pr'])

# Subset data for a specific year and bounding box (e.g., Northeast Brazil)
neb_slice = ds.sel(
    time=slice('2050-01-01', '2050-12-31'),
    lat=slice(-2.0, -18.0), # Example coordinates
    lon=slice(-48.0, -34.0)
)
```

Contact & Platform

This dataset is deeply integrated into the official PROJENE platform. To explore time-series, plot interactive regional maps, or dynamically calculate WMO Extreme Indices (ETCCDI) without downloading the entire volume natively, access the project portal: <https://projene.xhidro.com/>.