P03

# Bias Correction of the d4PDF Dataset Using Scarce Observational Data for Future Extreme Precipitation Analysis: The Case of El Salvador

○Mariana Beatriz AVALOS CABRERA, Masafumi YAMADA, Tomoharu HORI

## Introduction

Studies over the past two decades have demonstrated that the impacts of climate change on hydrological extremes exhibit high spatiotemporal variability and are associated with significant uncertainty and complexity[1]. A crucial step in these studies to reduce uncertainty is applying bias correction to climate model simulations using high-quality observations.

However, in regions lacking high-quality observations, researchers must resort to reanalysis data, which inevitably degrades the accuracy of the results due to the bias present in these datasets.

This challenge is evident in Central America, where previous studies have been (1) using mostly reanalysis data to bias-correct future precipitation, (2) have been limited to a coarse spatial and temporal scale analysis and/or (3) have lacked a specific emphasis on El Salvador. In addition, the impact of climate change on regional flood risk remains severely understudied[2].

To address this research gap, a study was conceived to assess the impact of climate change on flooding risk and evacuation time during extreme rainfall events in the slums of San Salvador, El Salvador, under different warming scenarios. The bias correction results presented here are the first phase of said study.

## Study Basin

The study area is a section of the upper Acelhuate River basin, with an approximate area of 217 km$^2$, where the Metropolitan Area of San Salvador (MASS) is located. The MASS is home to 1.7 million people, 32% of whom live in slums[3], many of which are in the Acelhuate River floodplains or inside its river channels.

## Reanalysis Datasets

To evaluate the performance of the available reanalysis datasets over the MASS, three datasets with hourly time steps were chosen: ERA5[4] (30 km grid), MERRA2[5] (56 x 70 km grid), and CMORPH[6] (8 km grid).

Observed data were obtained from 10 gauges within the study basin, provided by the Ministry of Environment and Natural Resources of El Salvador (MARN). These records span 6 to 20 years, with data completeness ranging from 47% to 91%. The datasets were cleaned to remove errors. The clean data was used to generate three hourly Area Mean Precipitation (AMP) grids matching the reanalysis resolutions.

The following evaluation metrics were calculated to evaluate the bias in the reanalysis datasets in the rainy season: Peak Error (PE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Percentage of Bias (PBIAS), and Pearson's Correlation Coefficient (R). The results are shown in **Table 1.**

**Table 1.** Datasets Evaluation Metrics

| Metric | CMORPH | MERRA2 | ERA5 |
|---|---|---|---|
| PE | 0.49 | 0.68 | 0.28 |
| RMSE | 2.0 | 2.18 | 1.52 |
| MAE | 0.43 | 0.68 | 0.43 |
| PBIAS | -25.38 | 98.93 | -2.19 |
| R | 0.04 | 0.03 | 0.11 |

Although ERA5 exhibits low percent bias, none of the reanalysis datasets proved to be viable substitutes for observations, primarily due to weak correlations (R) and a systematic underestimation of precipitation intensities (PE). Consequently, we selected the observed dataset. Despite its limitations, it preserves genuine extreme precipitation events, which are indispensable for effective bias correction.

## Bias Correction Methodology

The d4PDF dataset[7] was selected due to its large ensemble size and hourly temporal resolution, features that are essential for the analysis of extreme events. It consists of 2,000 years of past climate simulations (HPB) and future projections (HFB) under the 4K, 2K, and 1.5K warming scenarios, comprising 5,490, 3,294, and 1,782 years of data, respectively.

A hybrid Empirical Quantile Mapping (EQM) approach was chosen for its computational efficiency, as the study focuses on applying data to flood impacts rather than refining downscaling methods. EQM is a proven tool for bias correction both globally and within Central America[8]. The methodology can be summarized in the following steps:

(1) Clean the 10-gauge data.
(2) Interpolate the d4PDF ensemble datasets from a 60 km grid to a 1 km grid using the Second Inverse Distance Method.
(3) Extract 1 km grid cell coordinates covering the study basin from the d4PDF interpolated dataset.
(4) Resample the 10-gauge data into an AMP dataset spanning from 2005 to 2024 (20 years) with the same 1 km grid coordinates as d4PDF.
(5) Cut the last 20 years (1992-2011) in the HPB ensembles.
(6) Get the empirical distributions of the HPB / HFB ensembles and observed datasets.
(7) Using the HPB and observed empirical distributions, calculate the correction rules between them using a hybrid approach: if d4PDF = 0, the correction = observed – d4PDF; if d4PDF > 0, the correction = observed / d4PDF.
(8) Transfer the correction using the bias correction rules from Observed – Historical d4PDF to each value in the HPB / HFB empirical distributions.
(9) Reassemble the corrected time series to its original time steps to get the final BC datasets.

## Results

**Figure 1** illustrates the efficacy of the bias correction by comparing the raw simulated data (60 km resolution) to the corrected ensemble (1 km resolution). The results of calculating PBIAS at the observed grid cell in the basin with the highest historical precipitation intensity improved from -50.86 (raw HPB) to 0.15 (BC HPB). This demonstrates that the methodology successfully eliminates the original model bias while effectively downscaling the data to the observed spatial scale.

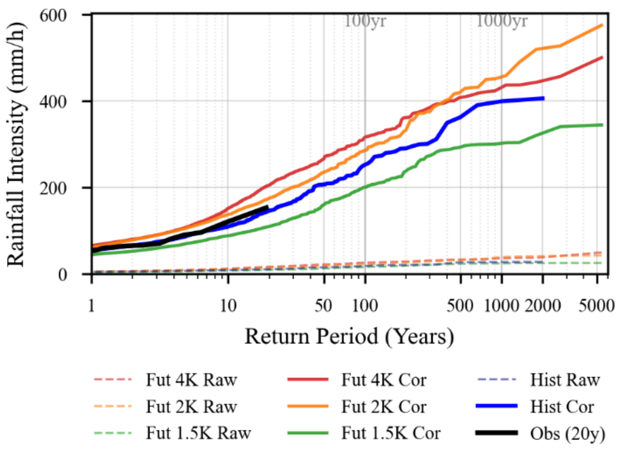


**Figure 1.** Bias Correction Results.

## Limitations

This method assumes bias stationarity and acknowledges that independent grid corrections do not preserve spatial coherence.

## References


**[1]** Meresa et al. (2023). https://doi.org/10.1007/s10712-022-09737-w **[2]** Eccles et al. (2019). https://doi.org/10.2166/wcc.2019.175 **[3]** MARN (2019). *National Climate Change Adaptation Plan*. El Salvador: Ministry of Environment and Natural Resources. **[4]** Hersbach et al. (2020). https://doi.org/10.1002/qj.3803 **[5]** Gelaro et al. (2017). https://doi.org/10.1175/JCLI-D-16-0758.1 **[6]** Xie et al. (2017) https://doi.org/10.25921/w9va-q159 **[7]** Ishii et al. (2020) https://doi.org/10.1186/s40645-020-00367-7 **[8]** Mendez et al. (2020). http://dx.doi.org/10.3390/w12020482