Climate Data Science

Shardul Junagade
23110297
GitHub Link - ShardulJunagade/Temporal-Downscaling

1. Introduction

In this project, I addressed the challenge of downscaling coarse-resolution monthly climate data into high-resolution daily and hourly predictions. This is especially important for applications like flood forecasting and water management, where fine temporal details can significantly affect decision-making. I worked with ERA5 reanalysis data for total precipitation and explored various methods to enhance the temporal resolution of the dataset.

2. Data Preparation and Preprocessing

The I downloaded 14 years (2011–2024) of ERA5 data using the CDS API. The raw data was in GRIB format, which I converted into CSV files using xarray. The preprocessing steps included:
Data Cleaning: Removing rows with missing values and duplicates.
Aggregation: Creating summary statistics (mean, max, min, standard deviation, variance) on yearly, monthly, and daily scales.

Splitting the Data:
Training: 10 years (2011–2020)
Validation: 2 years (2021–2022)
Test: 2 years (2023–2024)

3. Model and Experimentation

Baseline Method:
For a simple baseline, I assigned the monthly mean values to all hours and days within that month. The resulting metrics were:
NSE: -0.24106728
RMSE: 0.00038496
MAE: 0.00024062

Machine Learning Approaches

I then experimented with more sophisticated methods:

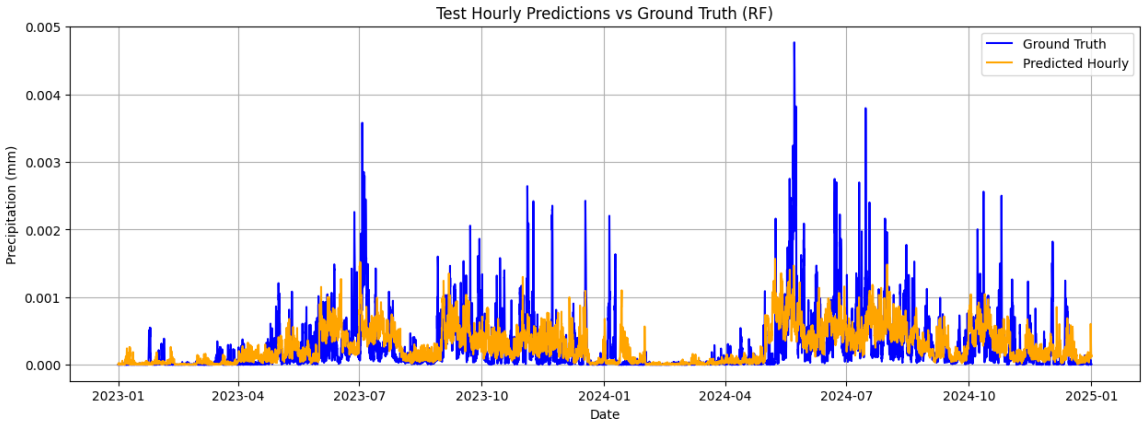
1. **Random Forest:**

   Hourly Predictions:

   RMSE = 0.00037678

   MAE = 0.00021458

   NSE = 0.2

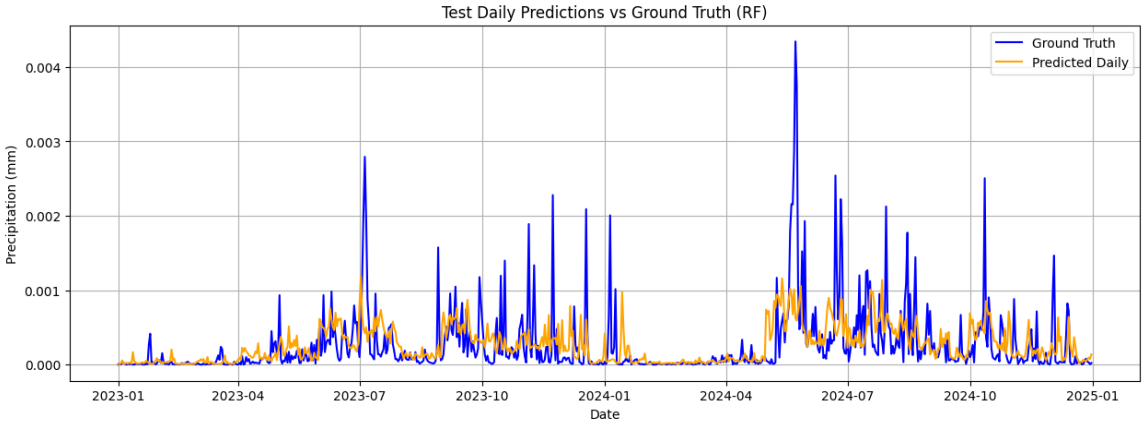


   Daily Predictions:

   RMSE = 0.00041219

   MAE = 0.00021953

   NSE = 0.19



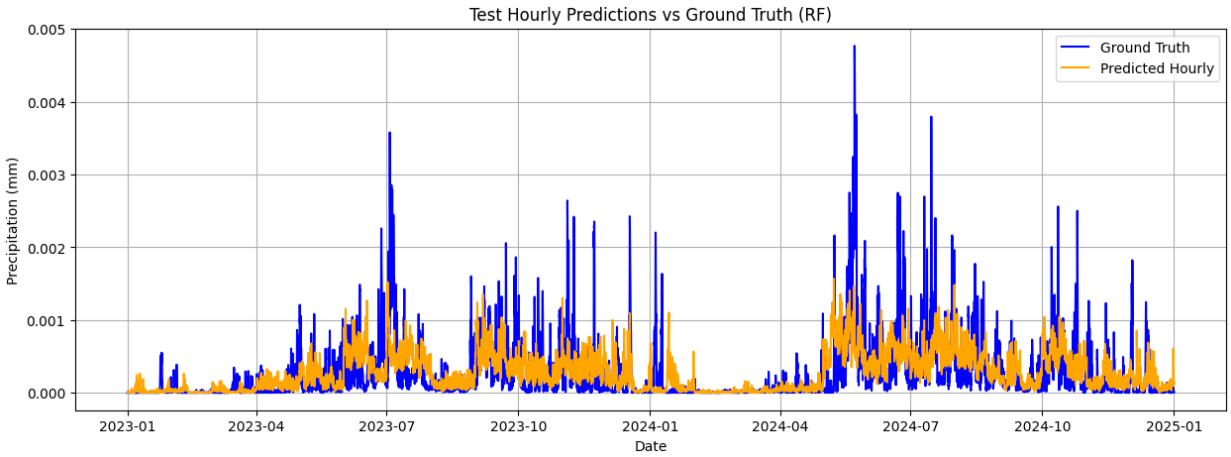
2. **XGBoost:**

Hourly Predictions:

RMSE = 0.00036710

MAE = 0.00020150

NSE = 0.24

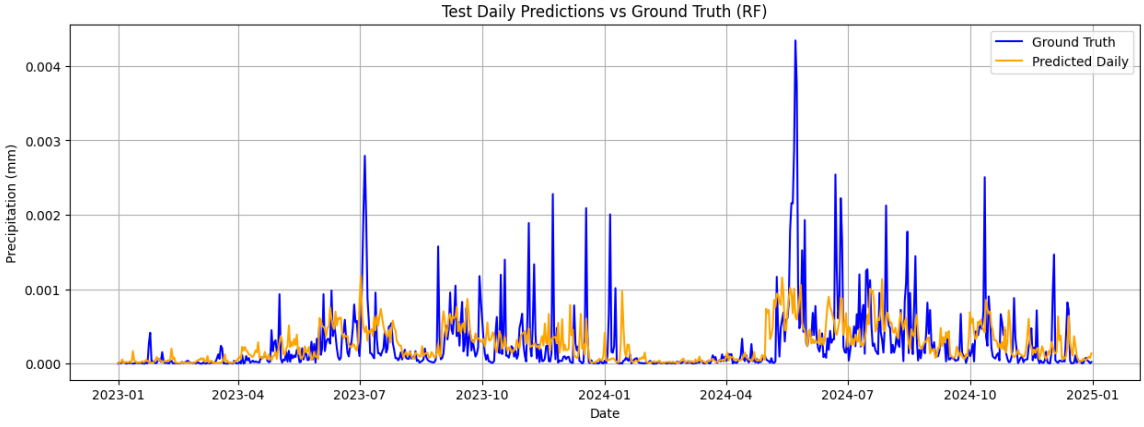


Daily Predictions:

RMSE = 0.00040691

MAE = 0.00020328

NSE = 0.21



I also implemented LSTM, but it didn't stabilize properly, due to constraint of time.

6. Results and Conclusion

The experiments clearly demonstrated that both Random Forest and XGBoost significantly improved the downscaling performance over the simple baseline, with XGBoost leading in terms of NSE for both hourly and daily predictions.