DS_5220

Supervised Machine Learning and Learning Theory

Project Report

Name: Danish Mansoor **NU_ID: 002209538**

Let's save our planet.

Statistical Downscaling of Climate Data Using Deep Learning

1. Introduction: The past few decades have witnessed an unprecedented rise in the intensity, frequency, and duration of hydroclimatic extremes globally, such as extreme precipitation, prolonged droughts, and heatwaves as shown in *figure 1*. Observational data provides valuable insights into the trends and variability of these events, but projecting their future behavior remains a significant challenge. Earth System Models (ESMs), which simulate the interactions among the atmosphere, ocean, land, and biosphere, offer a pathway to understanding future climates (*figure 2*). However, their coarse spatial resolution, typically ranging from 100 to 250 km, limits their ability to predict localized impacts critical for urban planning, disaster preparedness, and water resource management. This poses a barrier to generating actionable insights for policymakers and stakeholders.

Figure 1: *Various hydroclimatic extremes: drought, hailstones, sea level rise, floods, cyclones, lightning, tornadoes and forest fire.*

To address this limitation, **downscaling** techniques are employed to refine the coarse-resolution outputs of ESMs into high-resolution datasets. These methods enable a better understanding of localized climate phenomena, essential for planning at regional and community levels. **Dynamical downscaling**, which embeds sub-grid processes into Regional Climate Models (RCMs), provides physically consistent projections but is computationally intensive and constrained in scalability. On the other hand, **statistical downscaling** uses statistical or machine learning models to map relationships between large-scale climate model outputs and high-resolution observations. While computationally efficient, traditional statistical methods assume stationarity and often struggle to capture non-linear and complex interactions between variables.

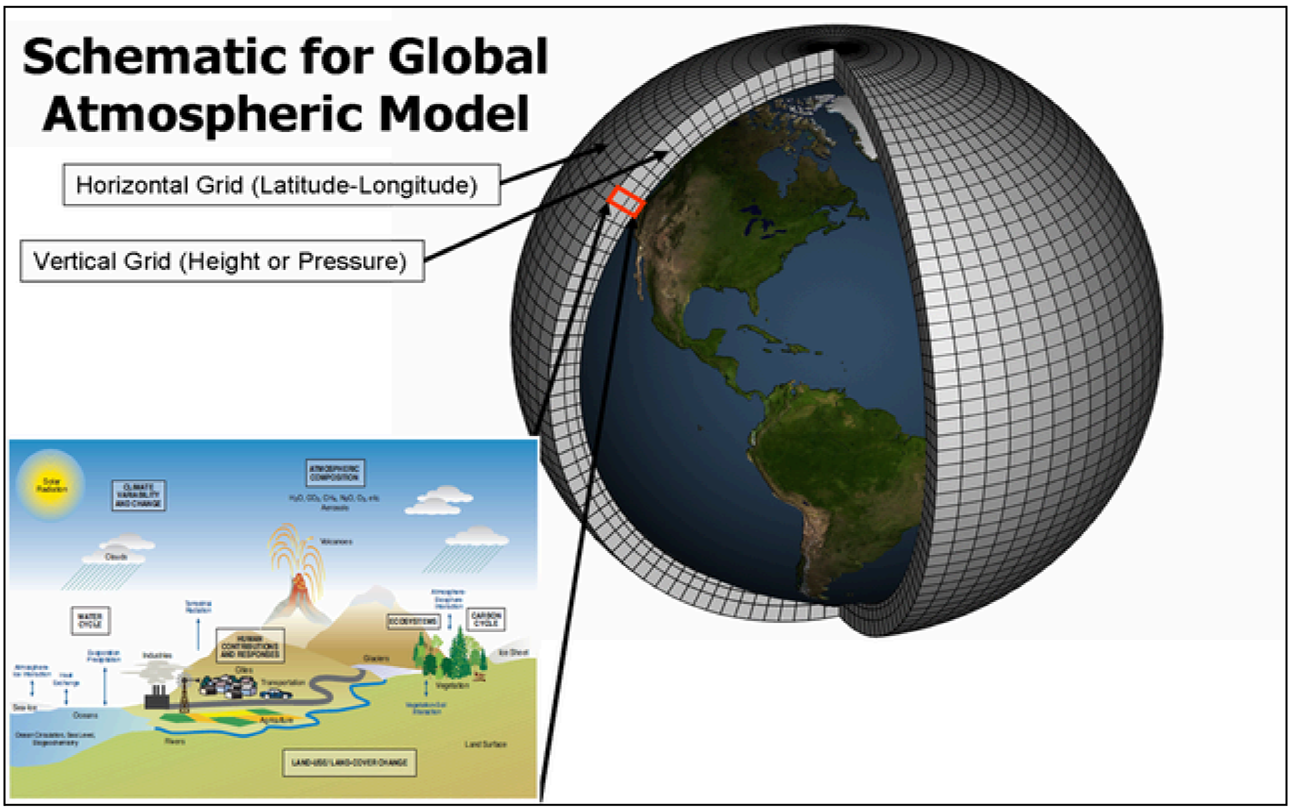

Figure 2: *Pictorial representation of Global Atmospheric Model (GCM) and its processes involved in it.*

Recent advances in deep learning offer promising tools to overcome these challenges. Techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their hybrids, such as Convolutional Long Short-Term Memory (ConvLSTM) networks, excel in capturing spatial and temporal dependencies in data. These methods have shown potential to improve the accuracy and resolution of climate projections, facilitating the generation of actionable, high-resolution climate information.

2. Research Objective: The primary objective of this research is to develop and implement an advanced statistical downscaling approach leveraging Convolutional Long Short-Term Memory (ConvLSTM) networks to enhance the spatial resolution of climate model outputs that can inform regional adaptation strategies, including urban planning, agricultural management, and disaster preparedness.

- Specifically, the study aims to downscale Temperature and Precipitation data from a coarse-resolution data (**0.5° × 0.5°**) to a finer spatial resolution (**0.25° × 0.25°**) by incorporating spatial and temporal dynamics using ConvLSTM networks to effectively model both spatial patterns and temporal dependencies in climate data, addressing the limitations of traditional statistical methods.

3. Convolutional LSTM (ConvLSTM) Architecture Overview: The ConvLSTM model is designed to downscale climate variables by capturing spatiotemporal patterns in low-resolution data. It processes a sequence of climate variables (e.g., precipitation, temperature, and other climatic variables) as a 3D tensor (10×N×M), where 10 represents variables, and N, M are spatial dimensions. The schematic representation is shown in *figure 3.*

Key Components:

1. Stacked ConvLSTM Blocks:
    - First Block: Encodes temporal and spatial dependencies from the input.
    - Second & Third Blocks: Sequentially process features, deepening spatiotemporal understanding.
2. Super-Resolution (SR) Blocks:
    - Enhance spatial resolution through convolutional layers with skip connections.
    - Two stages improve output resolution progressively.
3. Final Refinement:
    - Additional convolutional layers refine the high-resolution output for precise predictions.

This architecture effectively combines temporal encoding (ConvLSTM) with spatial enhancement (SR blocks), producing high-resolution climate projections essential for detailed analysis.

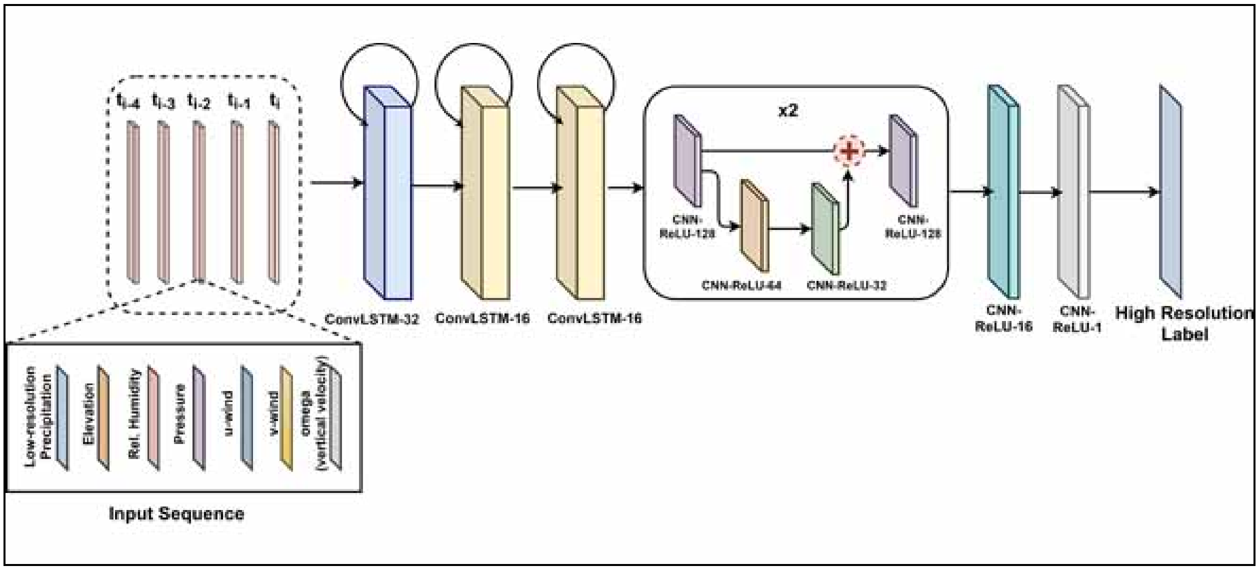

Figure 3: *Schematic representation of the ConvLSTM model: Early layers emphasize capturing temporal patterns, while later layers focus on spatial features.*

4. Study Area and Data Sources

4.1: Study Area Central India: Central India, delineated in *figure 4*, serves as the focus area for this research. Renowned as the core monsoon region, it uniquely reflects the dynamics of the Indian summer monsoon due to its advantageous plain topography. This characteristic makes it an ideal location for studying monsoonal patterns and developing a deep learning downscaling model.

4.2:Dataset: European Centre for Medium-Range Weather Forecasts (ECMWF): The ERA5 dataset, developed by the **European Centre for Medium-Range Weather Forecasts (ECMWF)** under the **Copernicus Climate Change Service (C3S)**, is a widely utilized global atmospheric reanalysis dataset. Reanalysis combines extensive historical observational data with numerical weather prediction models, offering a consistent, high-resolution depiction of the Earth's atmosphere, oceans, and land conditions over time.

In this study, we utilized the **monthly averaged ERA5 dataset** with a spatial resolution of 0.5° × 0.5° as input for training our ConvLSTM model. The input features included multiple variables, as detailed in *Table 1*, while the target variables were **temperature** and **precipitation**. For validation, we employed high-resolution data with a spatial resolution of 0.25° × 0.25° for temperature and precipitation, as illustrated in *figure 5*. This fine resolution data provided a benchmark for assessing the model's performance.

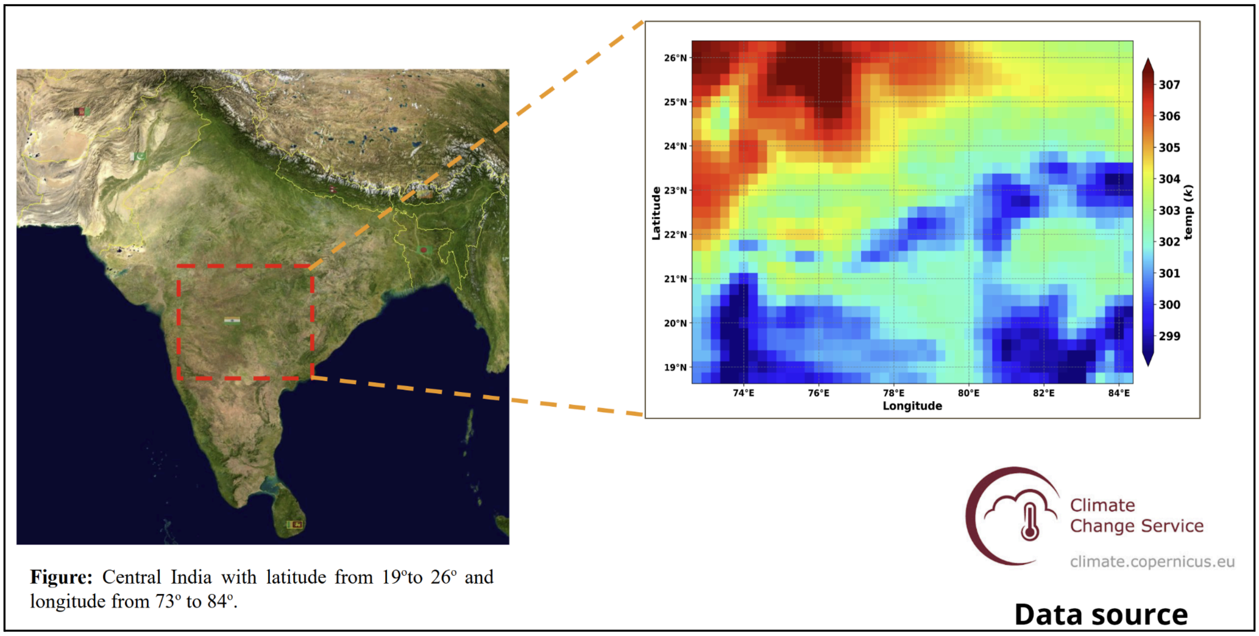

Figure 4: *Central India as the study region with a pictorial representation of the temperature data.*

5. Results and Discussion

5.1: Training and validation Loss: The ConvLSTM model was initially trained for 150 epochs with a batch size of 32 for both rainfall and temperature downscaling. However, during training, the loss curves revealed that both training and validation losses decreased exponentially within the first 30 epochs, with minimal improvement observed beyond that point. Consequently, the number of epochs was reduced to 50 to allow for efficient fine-tuning of other hyperparameters.

For precipitation, a batch size of 16 yielded the best performance due to the higher complexity of the data, while for temperature, a batch size of 32 produced optimal results. Further details regarding the hyperparameter fine-tuning process can be found within the code. *Figure 6* illustrates the training and validation loss curves for both temperature and precipitation datasets.

5.2: Downscaling of Temperature and Precipitation:

After training, the ConvLSTM model was applied to downscale temperature and precipitation for the testing dataset. The downscaled outputs for selected time stamps are illustrated in *figure 7* and *figure 8,* with a GIF file showcasing the results for a 16-month time period included in the project folder for additional visualization.

The figures clearly demonstrate that the downscaled temperature and precipitation closely follow the spatial patterns of the target data for the same months. However, some overestimation is observed along the boundaries during certain months, indicating potential limitations in the model's ability to capture edge effects accurately. Despite these localized discrepancies, the overall performance suggests that the model effectively captures the underlying patterns in temperature and precipitation fields.

Input Variables	**Short name**	**Vertical levels**	**Resolution**	**Units**
Mean Sea level pressure	MSLP	Mean sea level	$0.5^o \times 0.5^o$	Pa
Specific humidity	Q	850, 500 hPa	$0.5^o \times 0.5^o$	g/kg
Air Temperature	T	2m, 850 hPa	$0.5^o \times 0.5^o$	K
Zonal wind	U	850 hPa	$0.5^o \times 0.5^o$	m/s
Meridional wind	V	850 hPa	$0.5^o \times 0.5^o$	m/s
Vertical wind	W	850 hPa	$0.5^o \times 0.5^o$	m/s
Geopotential height	Z	850 hPa	$0.5^o \times 0.5^o$	m
Total precipitation	TP	surface	$0.5^o \times 0.5^o$	mm/day

Table 1 : *Input features for the training of ConvLSTM .*

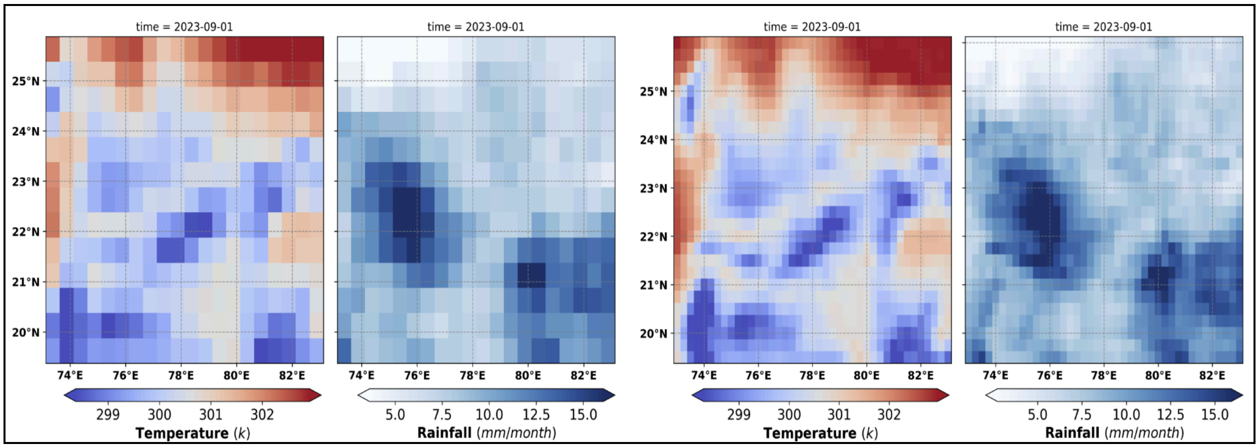

Figure 5: *Low resolution (0.5º degree) as primary input and high resolution (0.25º degree) as of target for temperature and rainfall downscaling.*

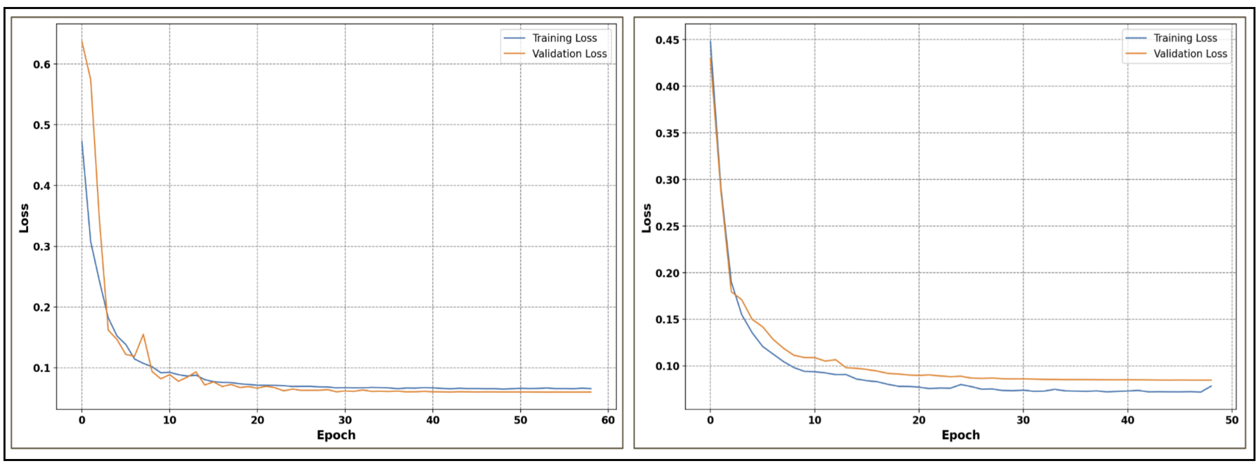

Figure 6: *Training and validation loss (RMSE) of the temperature (left) and rainfall (right)*

5.2: Performance Metrics: To evaluate the overall performance of the ConvLSTM model on the testing dataset, we utilized four performance metrics: Root Mean Square Error (RMSE), Bias, Pearson Correlation Coefficient (Corr), and Nash-Sutcliffe Efficiency (NSE). RMSE and Bias are widely used metrics in the machine learning domain, while Corr and NSE are domain-specific metrics often employed in hydrology and atmospheric sciences.

The results, visualized in the *figure 9*, reveal spatial variations in the model's performance across all grid locations. For temperature, RMSE values are predominantly low, with minimal bias, indicating accurate predictions. The high correlation and NSE values further confirm the model's ability to capture temporal patterns in temperature effectively. For rainfall, the RMSE values are slightly higher, reflecting the complexity of predicting precipitation. However, the model maintains strong correlation and NSE values, demonstrating reasonable predictive performance despite localized biases. These spatial performance maps highlight the model's strengths and areas for improvement, particularly in regions with higher error or bias in rainfall predictions.

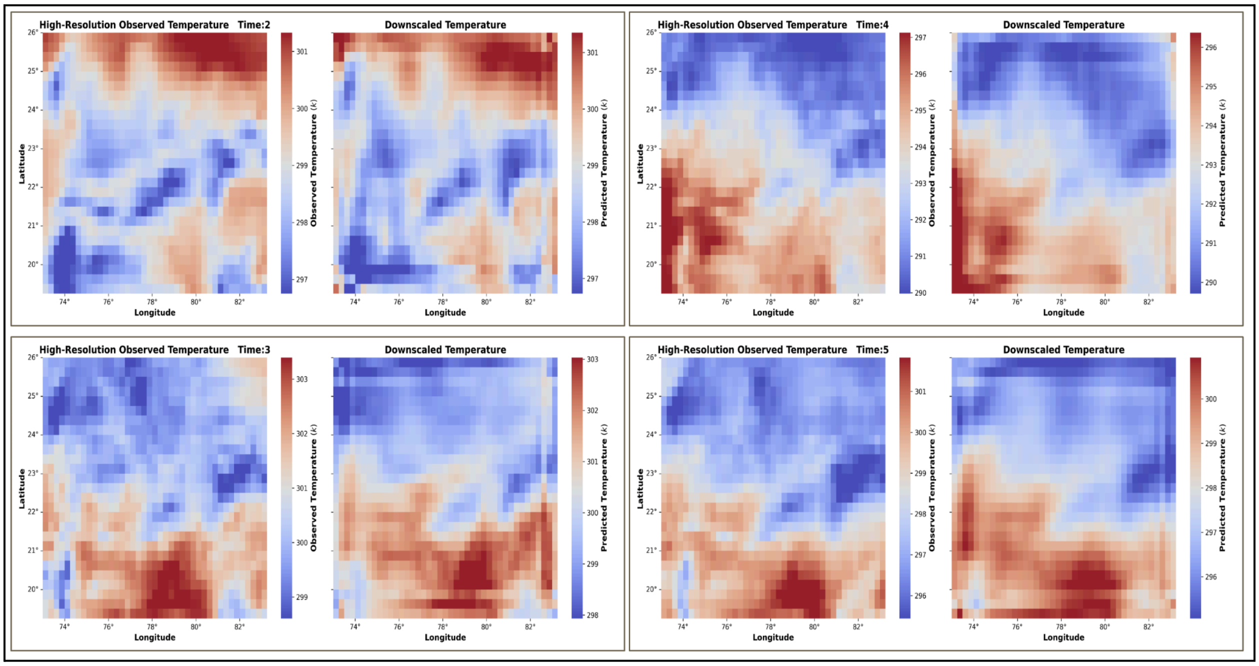

Figure 7: *Target high resolution and the downscaled temperature.*

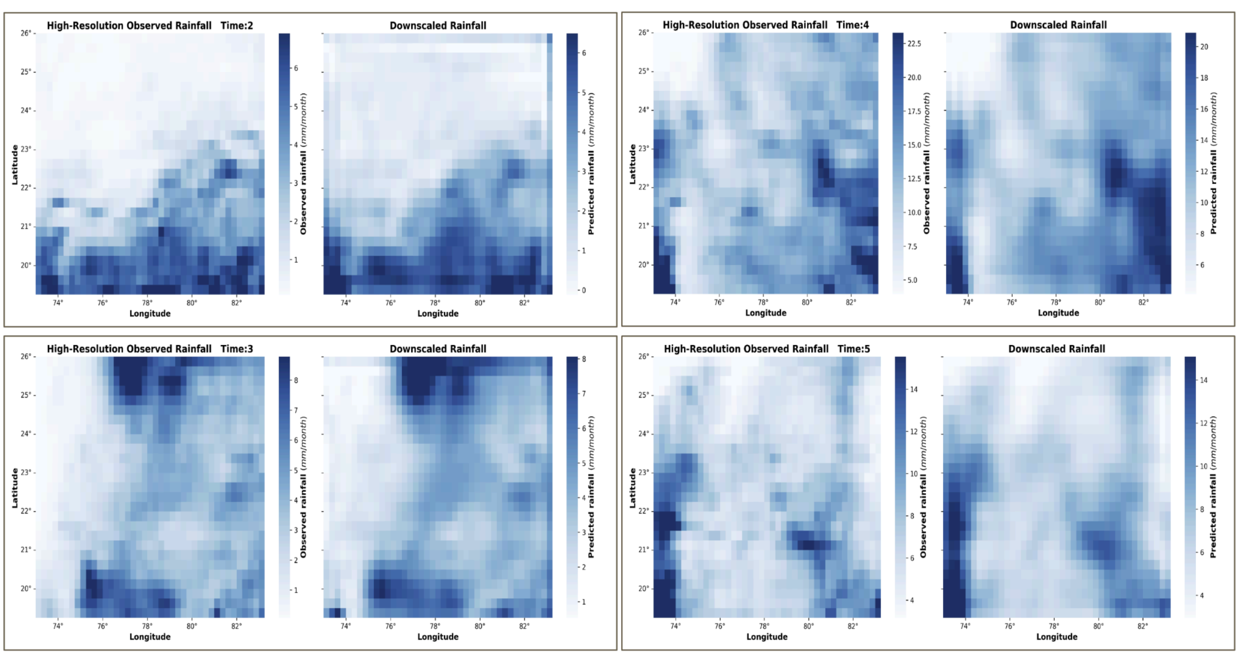

Figure 8: *Target high resolution and downscaled rainfall*

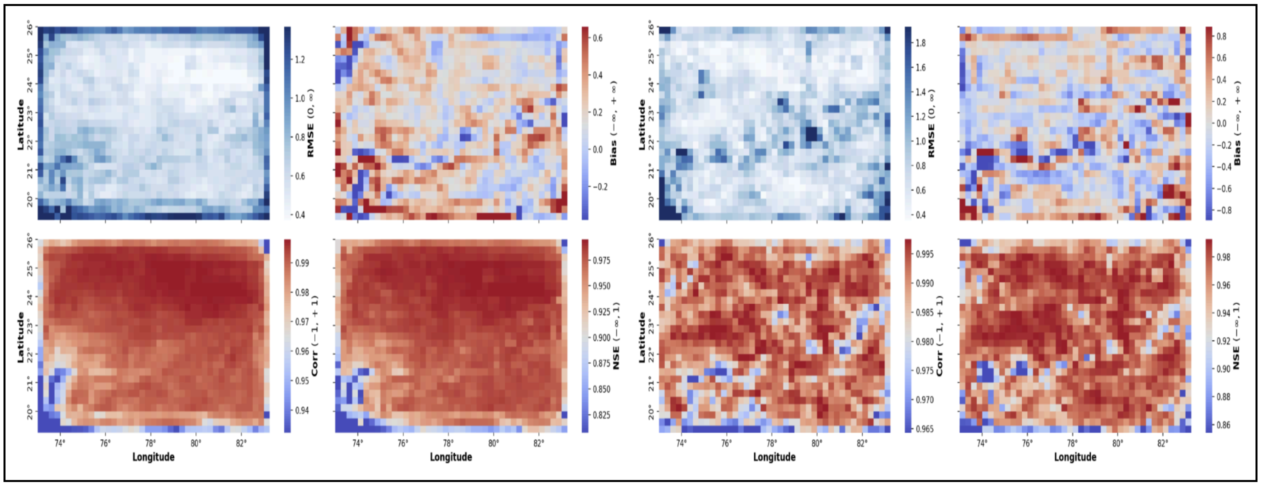

Figure 9: *Performance metrics of RMSE, bias, correlation and NSE*

6. Conclusion and Future Work

In this study, a ConvLSTM-based framework was implemented for downscaling rainfall and temperature at a monthly scale using multiple climate features. The model demonstrated its effectiveness in capturing seasonal patterns and spatiotemporal variations, offering valuable insights into regional climate dynamics. Despite its overall success, boundary-related inaccuracies were observed, highlighting areas for further improvement. Some of the areas where special attention needs to be given in future are;

1. **Enhancing Boundary Performance:**
   Incorporate advanced padding strategies during pooling operations to mitigate boundary-related inaccuracies, particularly for the temperature variable.
2. **Model Comparison:**
   Conduct a comprehensive performance assessment by comparing ConvLSTM with other advanced models to identify the most effective approach for downscaling.
3. **Refinement of Input Features:**
   Optimize the selection of input features based on regional climate characteristics to further improve model accuracy and efficiency.
4. **Temporal Resolution Improvement:**
   Extend the downscaling framework to finer temporal resolutions, such as the daily scale, to capture short-term variability and enhance its applicability for high-frequency climate studies.
5. **Broader Application:**
   Expand the framework to encompass additional regions with diverse climatic conditions to test

its generalizability and robustness in varying contexts.

Code: All the analysis was performed in Python and the code is available in github.

References:

1. Kourtis, Ioannis M., et al. "Sustainable water resources management in small Greek islands under changing climate." *Water* 11.8 (2019): 1694.
2. Prein, Andreas F., et al. "The future intensification of hourly precipitation extremes." *Nature climate change* 7.1 (2017): 48-52.
3. Kourtis, I. M., Tsihrintzis, V. A., & Baltas, E. (2020). A robust approach for comparing conventional and sustainable flood mitigation measures in urban basins. *Journal of environmental management*, *269*, 110822.
4. Kumar, P., Sharma, M. C., Saini, R., & Singh, G. K. (2020). Climatic variability at Gangtok and Tadong weather observatories in Sikkim, India, during 1961–2017. *Scientific reports*, *10*(1), 15177.
5. Harilal, Nidhin, Mayank Singh, and Udit Bhatia. "Augmented convolutional LSTMs for generation of high-resolution climate change projections." *IEEE Access* 9 (2021): 25208-25218.