ENGINEER - Vol. LIX, No. 02, pp. [75-86], 2026

DOI: https://doi.org/10.4038/engineer.v59i2.7750

ANN-Based Climate Downscaling for Future Precipitation Projections in the Mi Oya Basin Using CMIP6 Data

Sajani Dewasurendra, Nipuni Pramodya and Panduka Neluwala

Abstract: This study investigates the use of artificial neural networks (ANNs) and a unidirectional Long Short-Term Memory (LSTM) model for statistical downscaling of precipitation projections from the CanESM5 climate model in the IPCC Sixth Assessment Report (AR6), focusing on the Mi Oya River Basin, Sri Lanka. Climate downscaling addresses the mismatch between the coarse spatial resolution of General Circulation Models (GCMs) and the finer scales needed for localized climate impact assessments. Despite the global uptake of CMIP6 data sets, their application for statistical downscaling in Sri Lanka remains limited. CanESM5 was selected from a multi-model CMIP6 ensemble, based on correlation coefficient, root-mean-square error, and standard deviation metrics for the baseline period (1980–2014). Leveraging the capacity of LSTMs to capture complex, non-linear, and sequential relationships, a unidirectional many-to-one LSTM architecture was employed to align with real-time forecasting constraints using only historical inputs. The model achieved strong performance in baseline testing, with Nash–Sutcliffe Efficiency (NSE) ≈ 0.43 and low normalized RMSE, demonstrating its suitability for projection purposes. Downscaled future precipitation (2020–2100) under Shared Socioeconomic Pathway (SSP) scenarios indicates increasing rainfall trends, with SSP5-8.5 showing the largest changes, including moderate increases in seasonal rainfall (typically within 10–40% based on median values) and enhanced variability towards the end of the century. These findings demonstrate the potential of machine learning–based downscaling to enhance localized climate projections and support decision-making in water resource management, disaster preparedness, and climate-resilient agricultural planning in vulnerable tropical basins.

Keywords: ANN, CMIP6, LSTM, Machine Learning, SSPs, Statistical Downscaling

1. Introduction

The warming of the climate system is unequivocal, with widespread and rapid changes observed across the atmosphere, ocean, cryosphere, and biosphere over recent decades. Anthropogenic greenhouse gas (GHG) emissions driven by industrialization, land-use change, and energy consumption are the principal driver of these changes, with the global mean surface temperature estimated to have increased by about 1.09 °C between 1850–1900 and 2011–2020 [1]. Atmospheric concentrations of carbon dioxide, methane, and nitrous oxide now exceed those of at least the past 800,000 years [1]. These large-scale changes apparent locally through altered precipitation regimes, temperature extremes, and shifts in hydrological states that directly affect water security, food systems, and disaster risk. In South Asia and Sri Lanka, where monsoons and inter-monsoonal transitions structure hydrological seasonality, even modest perturbations in rainfall timing and intensity can cascade into significant impacts on storage reservoirs, irrigation scheduling, floodplain occupation, and drought preparedness.

To anticipate and manage such risks, climate change impact assessments rely on projections generated by GCM ensembles coordinated through the Coupled Model Intercomparison Project (CMIP) [2]. The latest CMIP6 ensemble underpins the IPCC's Sixth Assessment Report (AR6) and offers improved spatial resolution, updated process representation, and a broader range of emissions and socioeconomic pathways. Projections are contextualized using

Eng. (Ms.) ***Sajani Dewasurendra****, AMIE(SL), B.Sc.Eng. (Peradeniya), Dept. of Civil Engineering, University of Peradeniya, Sri Lanka.*
Email: dewasajani@gmail.com
Eng. (Ms.) Nipuni Pramodya*, AMIE(SL), B.Sc.Eng. (Peradeniya), Dept. of Civil Engineering, University of Peradeniya, Sri Lanka.*
Email: e18333@eng.pdn.ac.lk
Eng. (Dr.) Panduka Neluwala*, AMIE(SL), B.Sc.Eng. (Peradeniya), M.Eng., Ph.D. (Tokyo), Senior Lecturer, Dept. of Civil Engineering, University of Peradeniya, Sri Lanka.*
Email: pandukaneluwala@eng.pdn.ac.lk
https://orcid.org/0000-0002-1686-3412

Representative Concentration Pathways (RCPs) and Shared Socioeconomic Pathways (SSPs), which describe, respectively, the radiative forcing trajectories associated with GHG concentrations and the plausible global societal developments shaping emissions and adaptive capacity [3]. Although essential for understanding broad-scale climate evolution, GCM outputs are too coarse for direct application at river-basin scales [4]. The mismatch between GCM grid spacing and the spatial scales at which water-resource decisions are made tens of kilometers or less, necessitates downscaling to translate global signals into local, decision-relevant information [5].

Downscaling is commonly pursued through either dynamical or statistical approaches. Dynamical downscaling uses Regional Climate Models (RCMs) to nest higher-resolution physics within GCM boundary conditions, yielding spatially refined fields over a limited domain. Although Regional Climate Models (RCMs) offer enhanced spatial resolution, they are computationally intensive. In addition, their ability to simulate precipitation in tropical regions does not consistently exceed that of General Circulation Models (GCMs). This limitation arises from challenges in representing multiscale convective processes and complex land–sea interactions, which play a critical role in tropical rainfall dynamics [6,7]. Statistical downscaling (SD) offers a complementary and computationally efficient alternative by learning empirical relationships between large-scale predictors and local-scale predictands [8]. A variety of SD methods exist, ranging from bias-correction techniques such as quantile mapping to regression-based frameworks such as the Statistical Downscaling Model (SDSM), each with differing strengths depending on data quality, spatial resolution, and variable type [9]. When paired with bias correction, SD can substantially improve the fidelity of local climate information derived from GCMs. The NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP) archive offers bias-corrected, high-resolution (0.25°) daily outputs from multiple CMIP6 models, making it possible to apply global climate projections more effectively at regional scales, especially in data-scarce settings [10].

Machine learning (ML) has emerged as a particularly promising SD strategy due to its ability to capture nonlinear, multivariate relationships and efficiently extract features from high-dimensional climate datasets [11, 12]. Artificial neural networks (ANNs), in particular, can detect complex structures in atmospheric data that are often poorly represented by linear methods, making them well suited for modeling precipitation and temperature where non-linear interactions dominate [13, 14]. Other ML algorithms, including Convolutional Neural Networks (CNNs), Support Vector Machines (SVM), Relevance Vector Machines (RVM), Genetic Programming (GP), and SDSM hybrids, have also been explored for climate applications depending on data availability, spatial resolution, and the nature of predictor–predictand relationships [15]. Among ANN architectures, Long Short-Term Memory (LSTM) networks are especially effective for temporal modeling because they can retain and selectively forget information over multiple time lags, enabling the representation of memory effects such as soil–atmosphere feedbacks and moisture persistence that influence monthly rainfall [16]. Notwithstanding these advantages, applications of ML-based SD using up-to-date CMIP6 data remain limited, and few studies document a complete end-to-end work flow from GCM screening through predictor selection, model training, and scenario-based projections for basins in the dry-to-intermediate zones of the island.

This gap is particularly significant for Sri Lanka, a country highly vulnerable to shifts in precipitation patterns due to its reliance on rain-fed agriculture, limited water storage capacity, and exposure to climate-induced hazards. Addressing it requires approaches that not only leverage the strengths of modern ML architectures, such as LSTMs, but also integrate robust predictor selection and scenario-based analyses to translate coarse-scale CMIP6 outputs into basin-scale, decision-ready climate information. Such efforts can provide critical inputs for long-term water resource planning, disaster risk reduction, and climate adaptation strategies in data-scarce, climate-sensitive regions.

2. Literature Review

2.1 Neural Networks and Machine Learning in Climate Downscaling

Neural networks, a foundational class of machine learning models, consist of interconnected layers of artificial neurons that learn through training to capture complex, non-linear relationships between inputs and outputs. They have become indispensable tools across diverse domains ranging from image recognition to climate science because of their capacity for flexible pattern recognition [17–19].

In climate downscaling, statistical methods aim to bridge the gap between coarse-resolution GCM outputs and fine-scale observational needs. Traditional statistical models like regression and weather typing are limited in capturing non-linear dependencies, especially for variables like precipitation that exhibit irregular and localized variability. Neural networks, including multilayer perceptrons (MLPs), CNNs, and RNNs, provide a data-driven alternative capable of learning complex mappings without prior assumptions.

Feed-forward neural networks have shown moderate success in mapping atmospheric patterns to local rainfall, especially when augmented with relevant feature engineering. However, for tasks involving long-term memory and sequence dependence such as monthly precipitation, recurrent architectures outperform simpler models.

2.2 Long Short-Term Memory (LSTM) and Advanced Architectures

Long Short-Term Memory (LSTM) networks, a specialized form of RNN, were introduced to address the vanishing gradient problem that hampers the ability of standard RNNs to learn long-range temporal dependencies [20]. The core of the LSTM architecture lies in its gated memory cell, which comprises input, forget, and output gates that regulate the flow of information and allow the model to retain relevant historical context over extended sequences.

In climate science, LSTMs have been successfully applied to rainfall–runoff modeling, streamflow forecasting, and soil moisture prediction, often surpassing the performance of traditional process-based or purely statistical approaches [21–22]. For example, Li et al. (2020) [23] applied a sequence-to-sequence LSTM for rainfall–runoff simulation in Texas, achieving a notable improvement in Nash–Sutcliffe Efficiency (NSE) from 0.666 to 0.942 relative to physically based models. Hybrid architectures such as ConvLSTM, which integrate convolutional layers with LSTM cells, have proven effective in precipitation nowcasting by extracting spatial features from gridded data before modeling temporal evolution [24]. More recent innovations include CNN–LSTM hybrids for downscaling gridded GCM data [25], which combine spatial convolution for feature extraction with temporal sequence learning, and bidirectional LSTMs (BiLSTMs) coupled with adaptive filters such as those demonstrated by Jahangiri et al. (2025) [26] in CMIP6-based projections for Tehran, where NSE values approached 0.90 even under extreme precipitation scenarios.

2.3 ML and Deep Learning for CMIP6 Downscaling

Although ensemble and hybrid statistical approaches have been extensively explored in downscaling, the integration of CMIP6 outputs with advanced ML frameworks remains relatively limited. Recent studies, however, demonstrate the potential of such integration. Bittner et al. (2023) [27] developed a cost-efficient LSTM-based framework to downscale daily precipitation from a CMIP6 GCM across Australia under multiple SSP scenarios, incorporating quantile-based loss functions to better represent extreme rainfall events. Rampal (2024) [28], in a systematic review, documented the increasing role of ML in empirical downscaling and identified emerging architectures, including transformers and hybrid ensembles as promising avenues for improving predictive skill. In another example, Khosravi et al. (2025) [29] implemented a stacking ensemble (Stacking EML) that combined multiple ML models to downscale temperature and precipitation under CMIP6 SSPs in the Middle East, achieving an R^2 of up to 0.82 for precipitation. These examples underscore the growing recognition that combining diverse model structures within an ensemble framework can outperform single-model approaches.

2.4 Advantages and Emerging Challenges in ML-Based Downscaling

ML-based downscaling, particularly using LSTM and ensemble architectures, offers notable advantages for climate applications. These methods can capture complex, non-linear links between large-scale predictors and local hydro-meteorological responses, with LSTM's temporal memory aiding monthly and seasonal rainfall prediction. Hybrid models such as ConvLSTM and BiLSTM incorporate spatial as well as temporal features, and once trained, ML models produce downscaled outputs far more efficiently than computationally intensive Regional Climate Models (RCMs).

However, challenges remain. Overfitting is a risk in data-scarce contexts, requiring robust validation and, where possible, multi-GCM ensembles to reduce structural uncertainty. Model interpretability is limited, and reliance on a single GCM may propagate biases unless bias correction (e.g., QDM) is applied. Many models also underestimate extremes unless optimized with specialized loss functions or hybrid post-processing. Addressing these issues is essential for ML-based downscaling to reliably support climate adaptation and water resource planning.

3. Methodology

3.1 Study Area

The Mi Oya River Basin is located in Sri Lanka's North Western Province and is frequently affected by seasonal flooding associated with the monsoon cycle [30]. The basin area is approximately 1,516 km^2, and the river length is about 109 km, with most of the drainage lying within the Puttalam District (representative coordinate 8° 02′ 1.80″ N, 79° 50′ 4.19″ E). Annual rainfall volume is in the order of 2,176 million m^3, and the hydroclimate is structured by two inter-monsoon periods and the northeast and southwest monsoons. Peaks in monthly totals typically occur in April (first inter-monsoon) and October–November (second inter-monsoon and onset of the northeast monsoon), while June–August tends to be relatively dry within this basin, despite the broader southwest monsoon influence in Sri Lanka. A system of reservoirs Inginimitiya, Thabbowa, Halmilla, and Maha Wewa supports irrigation, domestic use, and partial flood moderation. The combination of flood susceptibility and water-supply dependence makes the basin representative of climate-sensitive rural catchments in Sri Lanka.

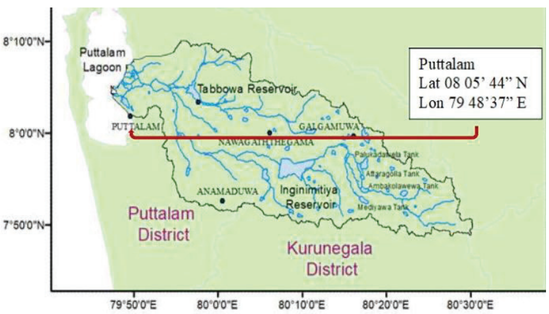

Figure 1 – Mi Oya River Basin, station network, and major reservoirs

3.2 Observational Data and Quality Control

Daily rainfall observations were obtained from 14 Department of Meteorology stations within or near the basin. To ensure statistical robustness and avoid transient biases associated with short series, we required a minimum of 30 consecutive years of data for inclusion in the baseline. Seven stations, Thabbowa, Madiyawa, Puttalam, Kootukachchiya, Atharagalla, Anamaduwa, and Palawi, met this criterion, providing a continuous record for 1980–2014. Data gaps were imputed using the normal-ratio method, which leverages neighboring station climatology, and cumulative consistency was verified with double-mass curves to detect shifts that could arise from gauge relocation, instrument changes, or procedural updates. Following quality control, daily series were aggregated to monthly totals to align with the temporal scale. The resulting multi-station monthly climatology reveals prominent peaks in April and October–November and a relative minimum during June–August.

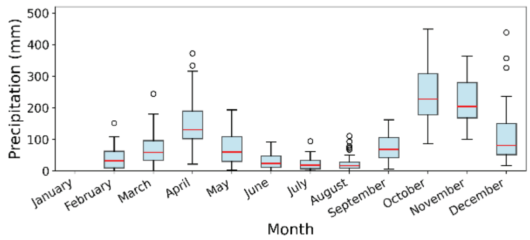

Figure 2 – Observed Precipitation

3.3 GCM Screening and Predictor Data

Ten CMIP6 GCMs commonly used in South Asian applications were shortlisted (Table 1) and evaluated against observed monthly rainfall for 1980–2014 using CC, RMSE, and SD to select the best GCM for the basin.

Predictor variables were sourced from the NEX-GDDP-CMIP6 archive at 0.25° spatial resolution. Given the basin's extent relative to this grid spacing, predictors were selected from four grid points around the study area. Nine predictors were initially extracted from the selected GCM, and a correlation analysis with observed monthly precipitation was performed to identify and retain the six variables showing the strongest physical relevance and statistical association.

Table 1 – Selected GCM models

Model	Research Center
Can-ESM5	Canadian Center for Climate Modelling and Analysis
ACCESS	Australian Community Climate and Earth System Simulator

MIROC6	Model for Interdisciplinary Research on Climate Consortium
EC-Earth3	EC-Earth Consortium
HadGEM3-GC3-LL	Met Office Hadley Centre
HadGEM3-GC3	Met Office Hadley Centre
MPI-ESM1-2-HR	Max Planck Institute for Meteorology
CNRM-ESM2	Centre National de Recherches Météorologiques (CNRM)
CNRM-CM6-1	Centre National de Recherches Météorologiques (CNRM)
NorESM2-MM	Norwegian Climate Centre (NCC)

3.4 Downscaling Model and Training Protocol

An eight-layer deep recurrent neural network (RNN), implemented as a unidirectional Long Short-Term Memory (LSTM) architecture and trained via backpropagation, was developed to statistically downscale monthly precipitation for the Mi Oya Basin. This configuration was selected for its ability to capture long-term temporal dependencies and its suitability for operational forecasting, where future information is not available.

The predictor set comprised GCM-derived variables (hurs, huss, pr, rlds, rsds, and sfcWind) selected based on their statistical relationship with observed rainfall, while the predictand was defined as the observed monthly precipitation in the Mi Oya Basin. To represent temporal dependencies, the model ingested 24-month sequences of predictors and produced precipitation for the subsequent month, a many-to-one mapping consistent with operational forecasting requirements. The dataset was chronologically divided into training (1980–2005, 75%), validation (2006–2008, 10%), and testing (2009–2014, 15%) subsets in order to preserve temporal dependence and prevent information leakage.

The network architecture comprised four sequential LSTM layers with decreasing complexity, followed by a dense hidden layer and a single-neuron output layer. The first two layers each contained 100 units with ReLU activation, L2 kernel regularization, and dropout applied to the second layer. The third and fourth layers contained 50 units with the same activation and regularization settings, with dropout rates of 0.2 and 0.3, respectively. A dense hidden layer with 25 units was placed before the final output layer, which produced the monthly precipitation estimate.

All predictors were normalized to the range [0,1] using Min–Max scaling based on the training data and applied to validation and testing sets:

$$X = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad \text{....(1)}$$

Table 2 – ANN Architecture

Component	Configuration
Optimizer	Adam
Loss Function	Mean Squared Error (MSE)
Batch Size	32
Epochs	300 (early stopping with patience = 20)
LSTM Layers	4 layers (100–100–50–50 units, with L2 regularization and dropout)
Dense Layer	25 units
Output Layer	1 unit

Model training was carried out in Google Colab using the Adam optimizer with mean squared error (MSE) as the loss function. A batch size of 32 and a maximum of 300 epochs were used, with early stopping (patience = 20) based on validation loss. Hyperparameters, including the number of LSTM layers and units, dropout rates, learning rate, and batch size, were tuned iteratively to balance predictive accuracy and generalization. The final configuration is summarized in Table 2.

3.5 Projecting Model

Monthly precipitation projections for the period 2020–2100 were generated using the finalized LSTM-based downscaling framework. The trained model was applied to outputs from the selected GCM to produce basin-scale rainfall estimates. Projections were developed under four Shared Socioeconomic Pathway (SSP) scenarios representing a spectrum of plausible futures:

- SSP1-2.6: low emissions with strong sustainability measures,
- SSP2-4.5: intermediate emissions with moderate global progress,
- SSP3-7.0: regional rivalry with moderate-to-high emissions, and
- SSP5-8.5: high emissions under fossil fuel–intensive development.

To capture both near-term and long-term climate signals, two 20-year projection windows were analyzed: 2020–2039 (near-future) and 2060–2079 (far-future). These time slices allow comparison of changes in monthly and seasonal rainfall patterns relative to the historical baseline (1980–2014), thereby providing insight into the potential evolution of hydrological conditions in the Mi Oya River Basin under alternative emission trajectories.

3.6 Evaluation Criteria

The performance of the downscaling model was evaluated using two widely applied hydrological metrics, Root Mean Square Error (RMSE) and Nash–Sutcliffe Efficiency (NSE). RMSE was reported in normalized precipitation units consistent with the Min–Max scaling applied during training, and it measures the average magnitude of prediction errors, with lower values indicating higher accuracy. NSE evaluates predictive skill relative to the observed mean, ranging from $-\infty$ to 1.0, where values closer to 1.0 indicate excellent model performance.

For baseline model assessment, RMSE values were reported in normalized precipitation units, consistent with the Min–Max scaling applied during training, ensuring that evaluation was not biased by scaling effects. These metrics, extensively validated in hydrological modeling, provide a robust basis for assessing both calibration accuracy and future projection reliability.

4. Results & Discussion

4.1 GCM and Predictor Selection

The GCM screening confirms that the choice of driving model is consequential for basin-scale downscaling. CanESM5 outperforms other candidates in both correlation and error magnitude (Table 3), and the Taylor diagram (Figure 2) indicates that its variance and phase of the seasonal cycle align comparatively well with observations. While some models approximate the annual cycle, their errors in amplitude or timing degrade monthly correspondence; such mismatches propagate into downscaling and motivate the up-front model selection.

Although CanESM5 was identified as the most suitable GCM for the Mi Oya Basin based on statistical performance metrics (CC, RMSE, and SD), the present study utilizes a single GCM as the driving model for the downscaling framework. Multi-model ensembles are often recommended to better capture structural uncertainties among climate models.

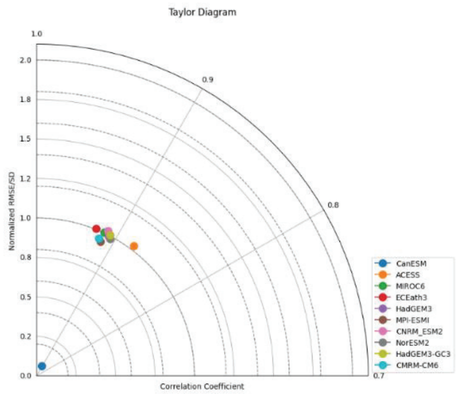

Figure 2 – Taylor Diagram

Table 3– Statistics and ranks of GCMs

GCM	RMSE /mm	SD	CC	Rank
Can-ESM5	61.47	87.37	0.61	1
ACESS	155.16	151.01	0.23	10
MIROC6	66.21	65.95	0.60	3
ECEath3	68.07	67.80	0.59	6
HadGEM3 -GC31-LL	69.82	70.20	0.55	7
MPI-ESM1-2-HR	64.90	68.91	0.59	2
CNRM_ESM2	68.04	66.54	0.59	5
NorESM2-MM	70.03	71.19	0.58	8
HadGEM3 -GC3	72.45	71.95	0.53	9
CMRM-CM6-1	67.17	70.36	0.56	4

However, incorporating multiple GCMs would require substantial additional data processing, model training, and recalibration of the machine learning framework. Therefore, this study adopts a single well-performing GCM following a performance-based screening procedure. Similar approaches have been applied in ANN-based statistical downscaling studies where a representative GCM is selected after evaluation against observed data [11], [15]. Future research could extend this framework by incorporating multi-GCM ensembles to better quantify projection uncertainty. From the nine candidate predictors, six variables, hurs, huss, pr, rlds, rsds, and sfcWindwere retained based on both statistical performance to monthly rainfall in a monsoon-dominated basin (Table 4) and their physical relevance in a monsoon-regulated basin. Among these, precipitation flux (pr) shows the strongest direct correlation, while humidity variables (hurs, huss) indicate the importance of moisture availability. Radiative fluxes (rlds, rsds) exhibit weaker but physically consistent signals through their influence on convective instability and surface energy balance.

Table 4 – Correlation Coefficients of Variables

Variable	Correlation Coefficient
hurs	0.36
huss	0.26
pr	0.60
rlds	0.11
rsds	-0.23
sfcWind	-0.48
tas	0.04
tasmax	0.04
tasmin	0.04

The negative correlation with near-surface wind speed (sfcWind) reflects the role of low-level convergence and horizontal moisture transport, which can act to suppress or enhance local rainfall depending on circulation patterns. In contrast, temperature-based predictors (tas, tasmax, tasmin) showed near-zero correlations (~0.04) and were excluded, consistent with the basin's hydroclimatic regime where rainfall variability is more sensitive to moisture and circulation than to direct temperature forcing.

4.2 Downscaling Performance for 1980–2014

The LSTM-based downscaling reproduces the seasonal cycle and the interannual modulation of monthly rainfall with reasonable reliability (Table 5). On the independent testing period (2009–2014), the model achieves NSE ≈ 0.43 with RMSE ≈ 0.18 in normalized units, similar to training and validation performance (NSE ≈ 0.48; RMSE ≈ 0.16–0.18). Although the testing NSE value of approximately 0.43 indicates moderate predictive skill, similar performance levels are commonly reported in precipitation downscaling studies due to the highly variable and nonlinear nature of rainfall processes. Precipitation is influenced by complex atmospheric interactions and localized convective dynamics that are not always fully captured by large-scale predictors derived from GCMs. Consequently, statistical and machine learning–based downscaling models often exhibit moderate efficiency values when applied to rainfall variables. Previous studies using ANN-based statistical downscaling approaches have also reported comparable performance for precipitation modelling [11], while deep learning–based rainfall downscaling frameworks similarly highlight the inherent challenges in accurately reproducing precipitation variability [21].

Despite this limitation, the developed model successfully reproduces the seasonal rainfall cycle and long-term variability, which are the primary focus of the climate projection analysis. The training curves (Figure 3) show stable convergence by ~200 epochs, indicating that the selected architecture and regularization are adequate to balance fit and generalization. These scores are typical of monthly precipitation downscaling in tropical settings where convective variability and mesoscale dynamics contribute noise at the monthly aggregation scale. Importantly, the baseline skill is sufficient to serve as a foundation for scenario-based projections, with uncertainty communicated in qualitative terms and through multi-scenario comparisons.

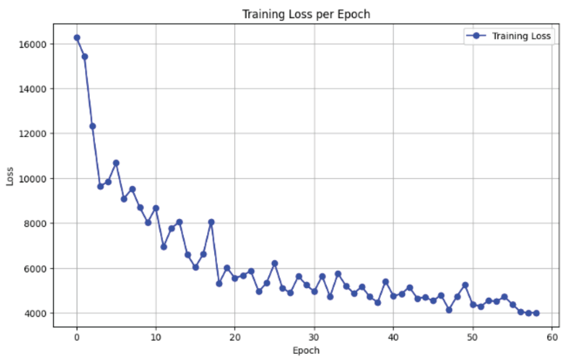

Figure 3 – Training Loss

Table 5 –Downscaling performance

Phase	NSE	RMSE
Training	0.48	0.16
Validating	0.48	0.18
Testing	0.43	0.18

4.3 Monthly Change Signals and Seasonal Structure Under SSPs

Figure 4 compares observed monthly climatology with downscaled SSP projections for 2020–2100. Across all scenarios, the seasonal structure is preserved: October–November remains the dominant wet season (second inter-monsoon plus onset of the northeast monsoon), April shows a distinct shoulder (first inter-monsoon), and December–March stays comparatively dry. Notably, the magnitude of the October–November peak increases under higher-emission scenarios, and the spread (interannual variability) within that season broadens in the far future. By contrast, July–

August remain relatively modest in this basin, with SSP2-4.5 and SSP5-8.5 indicating a tendency toward drier mid-year conditions, a signal consistent with a possible weakening or timing shift of mid-year monsoonal activity at the basin scale.

It should be noted that the present analysis focuses on long-term changes in monthly and seasonal precipitation patterns rather than extreme rainfall events. The projections were developed using monthly aggregated data to evaluate long-term hydrological trends under different climate scenarios. Extreme rainfall analysis typically requires higher temporal resolution datasets, such as daily or sub-daily rainfall, together with specialized extreme value statistical methods. Therefore, the assessment of precipitation extremes was beyond the scope of this study. However, future research could extend the developed framework to investigate extreme rainfall indices and associated flood risks using higher-resolution datasets.

4.4 Near-Future Versus Far-future Seasonal Changes

Seasonal boxplots for the near future (2020–2039) and the far future (2060–2079) quantify changes in medians and variability (Figures 5 and 6). The second inter-monsoon emerges as the wettest and most variable season in both windows, with SSP5-8.5 exhibiting the largest increases in both central tendency and spread by 2060–2079. The southwest monsoon, historically moderate in rainfall for this basin, becomes more variable under SSP3-7.0 and SSP5-8.5, suggesting sensitivity to circulation changes and moisture transport at seasonal timescales. The first inter-monsoon and the northeast monsoon also increase, but their changes are generally smaller than those of the second inter-monsoon, except under the highest-emission scenarios where northeast monsoon intensification becomes pronounced in the far future.

From a management perspective, the increased variability matters as much as the increase in means. Wider seasonal distributions imply greater year-to-year uncertainty in reservoir inflows and heightened risk of both flood-rich and flood-poor seasons. Planning horizons that rely on historical variability alone would understate future volatility, especially under SSP3-7.0 and SSP5-8.5.

4.5 Basin-Integrated Percentage Changes and Long-Term Trends

Relative to the 1980–2014 baseline, basin-integrated precipitation increases are projected to be ~19.1% (SSP1-2.6), 18.2% (SSP2-4.5), 22.3% (SSP3-7.0), and 22.8% (SSP5-8.5) over 2020–2100. These scenario-dependent differences are reinforced by the 10-year moving averages (Figure 7), which smooth interannual fluctuations and reveal a clear divergence between high- and low-emission trajectories. While even the low-emission pathway (SSP1-2.6) exhibits intensified rainfall, the amplitude and persistence of increases are larger under high emissions, especially after mid-century.

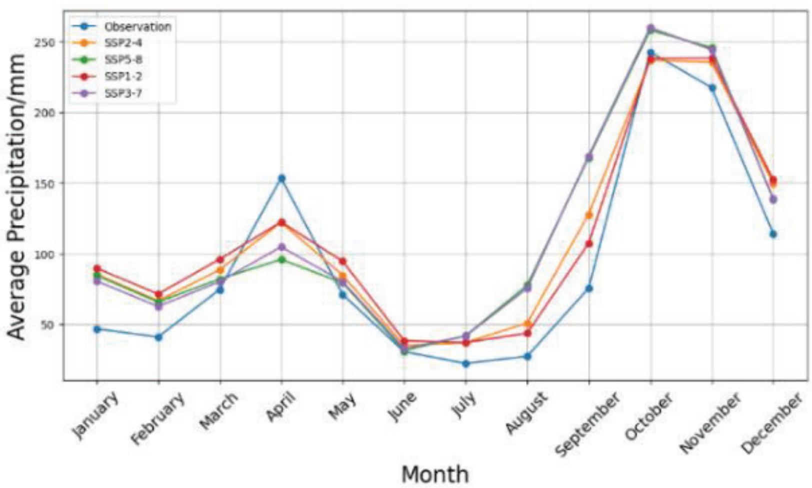

Figure 4 – Comparison of monthly precipitation of observation and downscaled data

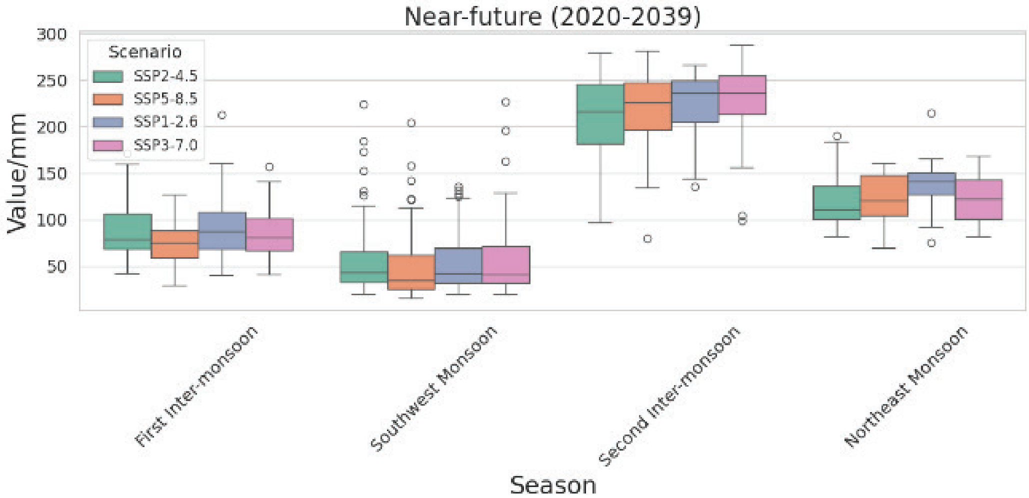

Figure 5 – Average Seasonal Precipitation boxplot for the near future

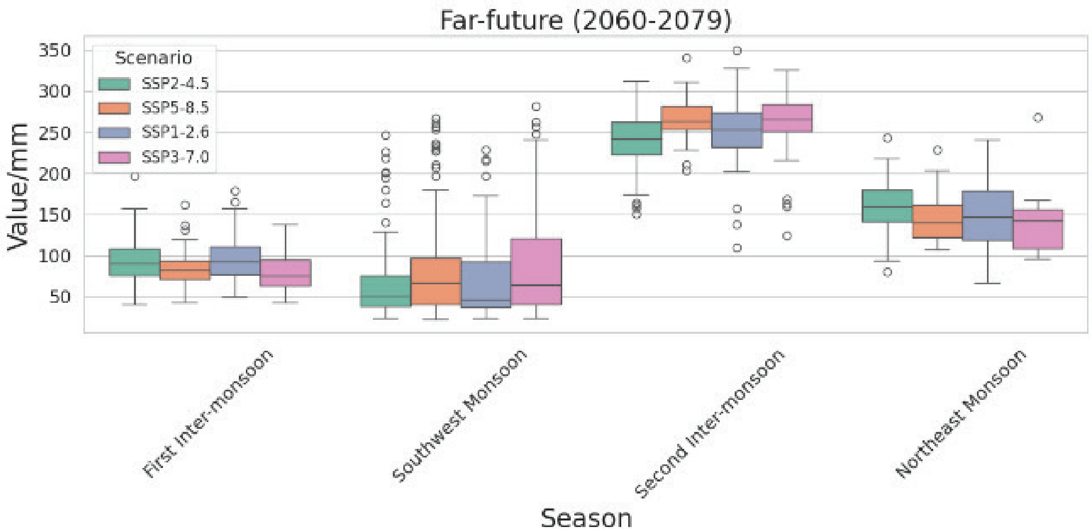

Figure 6 – Average Seasonal Precipitation boxplot for the far future

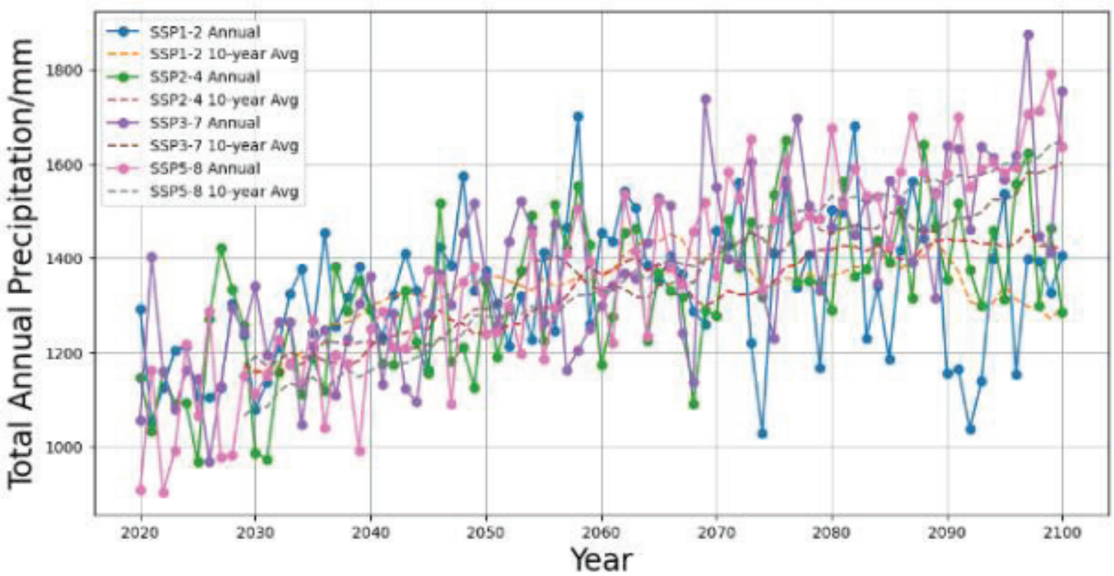

Figure 7 – 10-year Moving Average

Where seasonal magnitudes are evaluated based on median values from the projections, the southwest monsoon shows an increase of approximately 30–45% under SSP5-8.5 from the near future to the far future, while the northeast monsoon increases by about 15–30% over the same period. The first inter-monsoon exhibits a moderate increase of around 10–20%, whereas the second inter-monsoon shows a relatively smaller increase of approximately 5–15% by the far future. These results indicate that, although all seasons contribute to the overall wetting trend in the basin, the largest relative increases occur during the southwest and northeast monsoon seasons, while the second inter-monsoon continues to dominate in terms of absolute rainfall magnitude. This seasonal pattern is consistent with enhanced atmospheric moisture availability and potential changes in circulation that intensify rainfall over the northern and western regions of Sri Lanka.

4.6 Practical Implications

Physically, the projected wetting aligns with expectations of a warmer atmosphere's capacity to hold more moisture (Clausius–Clapeyron scaling) and with circulation adjustments that can alter both the timing and intensity of monsoonal inflows at regional scales. The Mi Oya Basin's response emphasizes the late-year seasons, where local orographic effects are limited but moisture fluxes and synoptic disturbances can be decisive. At the monthly scale, the LSTM's memory of the preceding 24 months helps transmit low-frequency variability related to antecedent moisture states and basin-integrated hydroclimate, providing a stable basis for scenario comparison.

From an applications standpoint, these results argue for adaptive reservoir rule curves that anticipate higher inflow magnitudes and greater intra-seasonal variability; expanded flood early-warning capacity, particularly for October–November; and refined agricultural calendars that incorporate a probabilistic view of seasonal rainfall under alternative emissions pathways. Because increases are not uniform across months, operational strategies should be stress-tested under multiple seasonal configurations rather than relying on annual totals alone.

Regarding robustness, we note that the results represent a single-model downscaling driven by one GCM (CanESM5). While the up-front screening was rigorous, structural uncertainties across GCMs remain and would ideally be sampled through a multi-model ensemble downscaling to bracket possible futures. Predictor selection used correlation analysis; alternative feature selection techniques such as mutual information, tree-based importance, or SHAP values could refine input parsimony and interpretability. Finally, while the LSTM is well suited to monthly temporal dependence, benchmarking against additional learners (e.g., gradient-boosted trees, SVMs, ANFIS) or hybrid/ensemble strategies could further improve skill and uncertainty characterization.

5. Conclusions

This study presents a complete, statistical downscaling framework for translating CMIP6-scale climate information into basin-scale precipitation projections in a Sri Lankan catchment. A multi-criteria screening identified CanESM5 as the most skillful GCM for the Mi Oya Basin during 1980–2014. Predictors from NEX-GDDP-CMIP6 were pared to a six-variable set (hurs, huss, pr, rlds, rsds, sfcWind) that captures moisture, radiation, and low-level dynamics relevant to rainfall. A unidirectional, many-to-one LSTM with a 24-month look-back mapped these predictors to next-month precipitation, trained with Min–Max scaling and a chronological split into training, validation, and testing periods. Baseline performance was reasonable for monthly downscaling in a tropical basin (testing NSE ≈ 0.43; RMSE ≈ 0.18 in normalized units), providing sufficient fidelity for scenario analysis.

Projections for 2020–2100 under SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 indicate basin-wide precipitation increases of ~19.1%, 18.2%, 22.3%, and 22.8%, respectively, relative to 1980–2014. The seasonal cycle persists, but the second inter-monsoon and northeast monsoon exhibit the strongest increases, and interannual variability widens particularly under SSP3-7.0 and SSP5-8.5 in the far future. These changes imply greater flood risk during late-year seasons, increased inflow uncertainty for reservoirs, and the need to update water allocation and cropping calendars to accommodate wetter yet more volatile conditions.

The analysis has three principled limitations. First, reliance on a single GCM constrains structural uncertainty; future work should employ a multi-model downscaling ensemble to bound plausible outcomes and facilitate probabilistic risk assessments. Second, predictor selection based on correlation can miss nonlinear dependencies; integrating mutual information, tree-based feature importance, or SHAP analysis would improve both parsimony and interpretability. Third, while the LSTM architecture is suitable for monthly time series, comparisons with additional learners (e.g., gradient boosting, SVM, ANFIS) and multi-model ensembling could further enhance robustness, particularly for extremes.

Despite these limitations, the framework demonstrates how modern ML can be embedded within a transparent, decision-oriented downscaling workflow for a data-scarce basin. The approach is readily transferable to other Sri Lankan catchments and can be extended to multi-basin early-warning design, reservoir rule-curve optimization under climate uncertainty, and integrated water–energy–food planning that reflects the shifting hydroclimatic baseline projected for the 21st century.

Acknowledgement

The authors would like to acknowledge the Meteorological Department of Sri Lanka for providing the required rainfall data for the research study.

References

1. Climate Change 2021: The Physical Science Basis. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, 2021. https://doi.org/10.1017/9781009157896

2. Nashwan, M.S. & Shahid, S., "A Novel Framework for Selecting General Circulation Models Based on the Spatial Patterns of Climate", *International Journal of Climatology*, Vol. 40, No. 10, 2019, pp. 4422–4443. https://doi.org/10.1002/joc.6465

3. Frame, B., Lawrence, J., Ausseil, A.-G., Reisinger, A. & Daigneault, A., "Adapting Global Shared Socio-Economic Pathways for National and Local Scenarios", *Climate Risk Management*, Vol. 22, 2018, pp. 103–115. https://doi.org/10.1016/j.crm.2018.05.001

4. Perera, P.L.L.N., Neluwala, N.G.P.B. and Wijetunge, J.J., "Influence of Weather Research and Forecasting Model Microphysics and Cumulus Schemes for Forecasting Monsoon Rainfall Over the Kelani River Basin, Sri Lanka", *Journal of Water and Climate Change*, Vol. 15, No. 10, 2024, pp. 5189–5206. https://doi.org/10.2166/wcc.2024.291

5. Mora, D.E., et al., "Climate Changes of Hydrometeorological and Hydrological Extremes in the Paute Basin, Ecuadorean Andes", *Hydrology and Earth System Sciences*, Vol. 18, No. 2, 2014, pp. 631–648. https://doi.org/10.5194/hess-18-631-2014

6. Buytaert, W., Celleri, R., Willems, P., De Bièvre, B. & Wyseure, G., "Spatial And Temporal Rainfall Variability in Mountainous Areas: A Case Study from the South Ecuadorian Andes", *Journal of Hydrology*, Vol. 329, 2010, pp. 16–29. https://doi.org/10.1016/j.jhydrol.2006.02.031

7. Cloke, H. L., Wetterhall, F., He, Y., Freer, J. & Pappenberger, F., "Modelling Climate Impact on Floods with Ensemble Climate Projections", *Environmental Modelling & Software*, Vol. 41, 2013, pp. 133–145. https://doi.org/10.1002/qj.1998

8. Peng, S., Wang, C., Li, Z., Mihara, K., Kuramochi, K., Toma, Y. & Hatano, R., "Climate Change Multi-Model Projections in CMIP6 Scenarios in Central Hokkaido, Japan", *Scientific Reports*, Vol. 13, 2023, p. 230. https://doi.org/10.1038/s41598-022-27357-7

9. Balasooriya, M. and Neluwala, P., "Statistical Approach to Selecting Global Circulation Models (GCMs)–A Study Based in the Kalu Ganga Catchment of Sri Lanka", *Engineer: Journal of the Institution of Engineers, Sri Lanka*, Vol. 58, No. 3, 2025. https://doi.org/10.4038/engineer.v58i3.7704

10. Thrasher, B., Wang, W., Michaelis, A., Melton, F., Lee, T. & Nemani, R., "NASA Global Daily Downscaled Projections, CMIP6", *Scientific Data*, Vol. 9, 2022, p. 262. https://doi.org/10.1038/s41597-022-01393-4

11. Nourani, V., Razzaghzadeh, Z., Hosseini Baghanam, A. & Molajou, A., "ANN-Based Statistical Downscaling of Climatic Parameters Using Decision Tree Predictor Screening Method", *Theoretical and Applied Climatology*, Vol. 137, 2019, pp. 1729–1746. https://doi.org/10.1007/s00704-018-2686-z

12. Seo, G.-Y. and Ahn, J.-B., "Comparison of Bias Correction Methods for Summertime Daily Rainfall in South Korea Using Quantile Mapping and Machine Learning Model", *Atmosphere*, Vol. 14, No. 7, 2023, p. 1057. https://doi.org/10.3390/atmos14071057

13. Sahastrabuddhe, R., Ghausi, S. A., Joseph, J. and Ghosh, S., "Indian Summer Monsoon Rainfall in a Changing Climate: A Review", *Journal of Water and Climate Change*, Vol. 14, No. 4, 2023, pp. 1061–1088. https://doi.org/10.2166/wcc.2023.127

14. Rhymee, H., Ratnayake, U., Abdul Rahman, E. K. and Shams, S., "Application of Normalized Difference Vegetation Index in Agriculture to Estimate Rice Yield", *AIP Conference Proceedings*, Vol. 2643, 2023. https://doi.org/10.1063/5.0115666

15. Nourani, V., et al., "ANN-Based Statistical Downscaling of Climatic Parameters Using Decision Tree Predictor Screening Method", *Theoretical and Applied Climatology*, Vol. 137, No. 3–4, 2019, pp. 1729–1746. https://doi.org/10.1007/s00704-018-2686-z

16. Singh, D., Vardhan, M., Sahu, R., Chatterjee, D., Chauhan, P. & Liu, S., "Machine-Learning- And Deep-Learning-Based Streamflow Prediction in a Hilly Catchment for Future Scenarios Using CMIP6 GCM Data", *Hydrology and Earth System Sciences*, Vol. 27, 2023, pp. 1047–1075. https://doi.org/10.5194/hess-27-1047-2023

17. S. P. Maurya, A. K. Yadav, & R. Singh (Eds.), "Methodologies of Scenario Development for Water Resource Management: A Review", *Modeling and Simulation of Environmental Systems*, 2022, pp. 303–316.

18. Koswaththa, H.M.S.A., Ranasinghe, S.K., Ekanayake, I., Herath, D. and Neluwala, N.G.P.B., "Downscaling Future Precipitation Over Mi Oya River Basin Using Artificial Neural Networks", *Engineer: Journal of the Institution of Engineers, Sri Lanka*, Vol. 57, No. 2, 2024. https://doi.org/10.4038/engineer.v57i2.7649

19. Tran Anh, D., Van, S. P., Dang, T. D. and Hoang, L. P., "Downscaling Rainfall Using Deep Learning Long Short-Term Memory and Feedforward Neural Network", *International Journal of Climatology*, Vol. 39, No. 10, 2019, pp. 4170–4188. https://doi.org/10.1002/joc.6066

20. Shibuya, E. and Hotta, K., "Cell Image Segmentation by Using Feedback and Convolutional LSTM", *The Visual Computer*, Vol. 38, No. 11, 2022, pp. 3791–3801. https://doi.org/10.1007/s00371-021-02221-3

21. Amrutha, K., Patnaik, R., Sandeep, A. S. & Pattanaik, J. K., "Climate Change Impact on Major River Basins in the Indian Himalayan Region: Risk Assessment and Sustainable Management", *Climate Change Adaptation, Risk Management and Sustainable Practices in the Himalaya*, Springer, 2023, pp. 45–63. https://doi.org/10.1007/978-3-031-24659-3_3

22. Li, W., Kiaghadi, A. and Dawson, C., "High Temporal Resolution Rainfall–Runoff Modeling Using Long-Short-Term-Memory (LSTM) Networks", *Neural Computing and Applications*, Vol. 33, No. 4, 2021, pp. 1261–1278. https://doi.org/10.1007/s00521-020-05010-6

23. Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.Y., Wong, W.K. and Woo, W.C., "Deep Learning for Precipitation Nowcasting: A Benchmark and a New Model", *Advances in Neural Information Processing Systems*, Vol. 30, 2017. https://doi.org/10.48550/arXiv.1706.03458

24. Nourani, V., Bahghanam, A.H., Pourali, H., Bejani, M. and Gebremichael, M., "Bias Correcting the Precipitation Dynamics of Regional Climate Models Via Kernel-Aware 2D Convolutional-Long Short-Term Memory", *Journal of Hydrology*, Vol. 657, 2025, p. 133068. https://doi.org/10.1016/j.jhydrol.2025.133068

25. Jahangiri, M., Asghari, M., Niksokhan, M.H. et al., "Bilstm-Kalman Framework for Precipitation Downscaling Under Multiple Climate Change Scenarios", Scientific Reports, Vol. 15, 2025, p. 24354. https://doi.org/10.1038/s41598-025-08264-z

26. Bittner, M., Hobeichi, S., Zawish, M., Diatta, S., Ozioko, R., Xu, S. and Jantsch, A., "An LSTM-Based Downscaling Framework for Australian Precipitation Projections", *in NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*, 2023.

27. Rampal, N., Hobeichi, S., Gibson, P.B., Baño-Medina, J., Abramowitz, G., Beucler, T., González-Abad, J., Chapman, W., Harder, P. and Gutiérrez, J.M., "Enhancing Regional Climate Downscaling Through Advances in Machine Learning", *Artificial Intelligence for the Earth Systems*, Vol. 3, No. 2, 2024. https://doi.org/10.1175/AIES-D-23-0066.1

28. Khosravi, Y., Ouarda, T.B. and Homayouni, S., "Developing an ensemble machine learning framework for enhanced climate projections using CMIP6 data in the Middle East", npj Climate and Atmospheric Science, Vol. 8, 2025, p. 174. https://doi.org/10.1038/s41612-025-01033-9

29. Riflan, M., Gunathilake, M.B., Samarasinghe, J.T., Bandara, I.M., Chathuranika, I.M. and Rathnayake, U., "Analysis of Statistical Downscaling Model (SDSM) projected future rainfall in Northwestern, Western and Southern provinces of Sri Lanka", Ruhuna Journal of Science, Vol. 13, No. 1, 2022, p. 70. https://doi.org/10.4038/rjs.v13i1.117