

LB/TH/22/2026

TH6224

**DEVELOPING A HIGH-RESOLUTION GRIDDED
PRECIPITATION DATASET FOR SRI LANKA BASED ON
GRID SENSITIVITY TO STREAMFLOW ESTIMATIONS
UNDER CLIMATE VULNERABILITY**

Malagalage Nethmini Nadee Kaushalya Peiris

248080V

Degree of Master of Engineering in
Water Resources Engineering and Management

Department of Civil Engineering

University of Moratuwa

Sri Lanka

March 2026

**DEVELOPING A HIGH-RESOLUTION GRIDDED
PRECIPITATION DATASET FOR SRI LANKA BASED ON
GRID SENSITIVITY TO STREAMFLOW ESTIMATIONS
UNDER CLIMATE VULNERABILITY**

Malagalage Nethmini Nadee Kaushalya Peiris

248080V

Supervised by

Dr G. M. C. A. Perera

Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Engineering in Water Resources Engineering and Management

UNESCO Madanjeet Singh Centre for South
Asia Water Management (UMCSAWM)

Department of Civil Engineering
University of Moratuwa
Sri Lanka

March 2026

DECLARATION OF CANDIDATE AND SUPERVISOR

“I declare that this is my own work, and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text”.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Malagalage Nethmini Nadee
Kaushalya Peiris

11/03/2026

Date

The above candidate has carried out research for the Master’s Thesis/Dissertation under my supervision.

Dr G. M. C. A. Perera

11/03/2026

Date

ABSTRACT

Developing a High-Resolution Gridded Precipitation Dataset for Sri Lanka based on Grid Sensitivity to Streamflow Estimations under Climate Vulnerability

Ground-based rain gauges remain the benchmark for accurate precipitation measurement; however, their sparse spatial distribution limits the representation of rainfall heterogeneity. Satellite-based precipitation products (SPPs) provide consistent spatial coverage but are often affected by retrieval errors and regional biases, restricting their direct use in local-scale hydrological applications. To overcome these limitations, precipitation data merging (PDM) techniques integrating gauge and satellite observations have gained prominence. This study introduces a novel machine learning framework, GraphIDW, which combines Graph Neural Networks (GNNs) with Inverse Distance Weighting (IDW) interpolation to explicitly incorporate spatial autocorrelation into the merging process, addressing a major limitation of traditional ML-based PDM approaches. The framework was evaluated across the Wet Zone of Sri Lanka from 2001 to 2015 using two state-of-the-art SPPs (IMERG and CHIRPS) together with ground observations. IMERG data (0.1°) were first downscaled to 0.05° using CHIRPS, after which the downscaled product was merged with gauge observations through GraphIDW. A total of 60 gauges (70%) were used for training and 28 (30%) for validation. Results show that GraphIDW outperforms conventional machine learning algorithms, including Random Forest, Artificial Neural Network, Support Vector Regression, and XGBoost. It achieved the highest probability of detection (0.97) and reduced root mean square error (RMSE) and mean absolute error (MAE) by 13-41% and 9-36%, respectively, compared with the original SPPs. The results demonstrate that explicitly accounting for spatial dependence through graph-based learning significantly improves precipitation estimation, particularly in regions characterised by strong spatial heterogeneity. By embedding spatial autocorrelation directly into the merging process, GraphIDW provides a robust and computationally efficient framework for generating high-resolution rainfall datasets that are better suited for hydrological analysis in complex climatic and topographic settings.

Keywords: CHIRPS; GNN; IMERG; precipitation data merging; spatial autocorrelation

ACKNOWLEDGEMENT

I extend my sincerest gratitude to my research supervisor, Dr. G. M. C. A. Perera, whose unwavering guidance, profound knowledge, and enthusiastic support were instrumental in the successful completion of my MSc research project. I am deeply grateful for his patience, continuous motivation, and encouragement throughout this academic journey. It has been both an honour and a privilege to be mentored by such an exceptional teacher and researcher.

I gratefully acknowledge Professor R. L. H. L. Rajapakse, Chairman of the UNESCO Madanjeet Singh Centre for South Asia Water Management (UMCSAWM), for encouraging me to pursue higher studies and for providing the opportunity to undertake the Master of Science degree under the SAF-Madanjeet Singh Scholarship Scheme. His vision, guidance, and mentorship have been invaluable to the successful completion of this research.

I would also like to express my sincere appreciation to Dr. Nimal Wijeratne, Dr. Luminda Gunawardhana, Dr. N. K. Gunasekara, and Dr. Chatura Dissanayake for their invaluable guidance, constructive feedback, and support throughout the course of this study.

I extend my gratitude to the Late Shri Madanjeet Singh, Founder of the SAF-Madanjeet Singh Scholarship Scheme, the South Asia Foundation (SAF), and the UNESCO Madanjeet Singh Centre for South Asia Water Management, University of Moratuwa, for enabling me to pursue postgraduate studies in civil engineering.

Special thanks are extended to Mr. Ajith Wijemannage, Director of the Department of Meteorology, Sri Lanka, for facilitating access to ground-based rainfall data, which formed a critical component of this research.

I also express my sincere gratitude to the Head of the Department of Civil Engineering, University of Moratuwa, and the academic and non-academic staff of the department for their support and assistance throughout my research period.

My heartfelt thanks go to Mr. Wajira Kumarasinghe, Mr. Samantha Ranaweera, Mrs. K. A. V. Kalanika, Ms. L. J. N. Silva, and Mr. W. R. Wijeweera, the staff of the

UNESCO Madanjeet Singh Centre for South Asia Water Management (UMCSAWM), for their continuous support and cooperation during the study.

I extend my deepest appreciation to my parents, my brother, and my colleagues, whose unwavering encouragement, understanding, and support have been a constant source of strength throughout this journey. Without their support, the successful completion of my studies would not have been possible.

Finally, I gratefully acknowledge the financial support provided by the Climate Action for Advanced Inclusive Education (CAFIED) Project, funded under the UNESCO Participation Programme (2024-2025), which made this research possible.

TABLE OF CONTENT

1	Introduction	1
1.1	Background of the Study	1
1.2	Problem Statement	2
1.3	Aims and Objectives	3
1.4	Scope and Limitations.....	4
1.5	Significance of the Research.....	4
1.6	Structure of the Dissertation	5
2	Literature Review.....	7
2.1	Development of High-Resolution Gridded Precipitation Dataset.....	7
2.1.1	Stage I: Spatial Interpolation Based on the Measurements of Rain Gauges	8
2.1.2	Stage II: Remote sensing retrieval	9
2.1.3	Stage III: Multi-source precipitation data merging.....	13
2.2	Machine Learning (ML) for Precipitation Data Merging	17
2.2.1	Traditional machine learning approach	17
2.2.2	Double machine learning approach	19
2.2.3	Incorporating spatial autocorrelation to machine learning approach	20
2.3	Graph Neural Networks (GNN) for Relational Data	22
2.4	Applications of Graph Neural Networks in Hydrology	23
2.5	Hybrid Interpolation for Rainfall Spatial Estimation.....	24
2.6	Training Strategies for the ML Algorithm for PDM.....	25
2.7	Satellite-based Precipitation Products (SPPs) Downscaling.....	26
2.8	Evaluation Metrics	26
2.8.1	Categorical metrics	27
2.8.2	Continuous metrics	28
2.9	Summary of the Literature Review.....	29

3	Methodology.....	31
3.1	Study Area.....	32
3.2	Data Collection.....	34
3.2.1	Observed rainfall data.....	34
3.2.2	Satellite rainfall data.....	34
3.2.3	Satellite-derived environmental variables	35
4	Data Checking and Analysis.....	37
4.1	Evaluation of Native Satellite-based Precipitation Products.....	37
4.2	Satellite-based Precipitation Products Downscaling	39
5	Model Development and Applications.....	41
5.1	Precipitation Data Merging (PDM) Problem Formulation.....	41
5.2	Graph Neural Network (GNN).....	42
5.3	GNN-IDW Model (GraphIDW) Structure	43
5.4	Merging Procedure.....	44
5.5	Traditional ML Algorithms.....	45
5.5.1	Support vector machines (SVM)	45
5.5.2	Random forest (RF).....	46
5.5.3	Extreme gradient boosting (XGBoost)	46
5.5.4	Artificial neural networks (ANN).....	46
6	Results.....	49
6.1	Overall Performance	49
6.2	Seasonal Evaluation	54
6.3	Intensity Evaluation.....	58
6.4	Impact of Localised Events	60
6.5	Performance during Flood Events	62
6.6	Spatial Evaluation	65

7	Discussion.....	69
7.1	Comparison of Performance of Native Satellite Precipitation Products.....	69
7.2	Performance of the GraphIDW Framework.....	71
7.3	Spatial Autocorrelation as a Missing Ingredient in Standard ML Merging.....	72
7.4	Comparison of Computational Efficiency of ML Algorithms.....	74
7.5	Impact of Rain Gauge Density.....	75
7.6	Limitations and Uncertainties	76
8	Conclusions, Future Works and Recommendations	77
8.1	Conclusions.....	77
8.2	Future Works and Recommendations	79
	Bibliography	81
	Annexure 1.....	97

LIST OF FIGURES

Figure 3-1. Methodology flowchart for the study -----	32
Figure 3-2. The topography of the Wet Zone in Sri Lanka and the distribution of rain gauges -----	33
Figure 4-1. Box plots of spatial variability of (a) Pearson correlation coefficient and (b) root mean square error at daily scale -----	38
Figure 4-2. Box plots of spatial variability of (a) Pearson correlation coefficient and (b) root mean square error at the monthly scale-----	39
Figure 4-3. Box plots of the variation of correlation between NDVI, $LST_{(day)}$, $LST_{(night)}$ and the observed rainfall in the study area -----	39
Figure 4-4. Box plots of spatial variability of (a) Pearson correlation coefficient and (b) root mean square error between IM-CH, IMERG and CHIRPS and observed rainfall -----	40
Figure 5-1. Graph construction in the GNN model -----	44
Figure 6-1. Boxplot of spatial variation of Pearson correlation coefficient (PCC) for six products, including original $IMERG_{0.05^{\circ}}$ (IM-CH) and ML-based merged products -----	51
Figure 6-2. Boxplot of spatial variation of RMSE for six products, including original $IMERG_{0.05^{\circ}}$ (IM-CH) and ML-based merged products-----	51
Figure 6-3. Boxplot of spatial variation of MAE for six products, including original $IMERG_{0.05^{\circ}}$ (IM-CH) and ML-based merged products-----	52
Figure 6-4. Roebber's performance diagram for the original $IMERG_{0.05^{\circ}}$ (IM-CH) and the merged precipitation products. The green dashed lines represent the bias score (BS), while the black contour lines indicate the critical success index (CSI). The four diagrams present the detection performance for the precipitation thresholds of 1, 5, 10, and 15 mm/day, respectively. -----	54
Figure 6-5. Box plots of the variation of PCC ((a)-(d)), RMSE ((e)-(h)) and MAE ((i)-(l)) over Wet Zone during four seasons; Northeast monsoon, First inter monsoon, Southwest monsoon and Second inter monsoon -----	57
Figure 6-6. Histogram comparison of (a) PCC, (b) RMSE, and (c) RB% of six precipitation products against ground observations under different rain intensities. (Heavy rain: >50 mm/day; Moderate rain: 15–50 mm/day; Light rain: 1–15 mm/day). -----	59
Figure 6-7. Box Plots of RMSE variation in localised, non-localised, and all heavy rainfall events (>50 mm) across all the test gauges -----	62
Figure 6-8. Comparison of ground-observed and model-predicted rainfall amounts from the original $IMERG_{0.05^{\circ}}$ (IM-CH) and GraphIDW-merged products, along with the Thiessen-weighted average rainfall for the Kelani River Basin during the flood events of (a) May 2011, (b) July 2012, (c) October 2012, and (d) June 2014. -----	64
Figure 6-9. Comparison of ground-observed and model-predicted rainfall amounts from the original $IMERG_{0.05^{\circ}}$ (IM-CH) and GraphIDW-merged products, along with the Thiessen-weighted average rainfall for the Kalu River Basin during the flood events of (a) April 2011, (b) July 2012, (c) October 2012, and (d) May 2014 -----	65

Figure 6-10. Spatial variability of single mass curve (SMC) gradients and corresponding RMSE distribution of the original IMERG _{0.05°} across the Wet Zone of Sri Lanka. -----	66
Figure 6-11. Spatial variation of RMSE in the original IMERG _{0.05°} (a) and GraphIDW-predicted rainfall (b) along with rainfall contours across the Wet Zone of Sri Lanka. -----	67
Figure 7-1. Computation time of five ML models for precipitation data merging-----	74
Figure 7-2. Performance comparison of the original IMERG _{0.05°} and ML-based merged products across varying training sample sizes. The percentages represent the proportion of training samples relative to the total available dataset. -----	75

LIST OF TABLES

Table 2-1. Applications of spatial interpolation methods widely used in past studies	10
Table 2-2. Details about widely utilized satellite-based precipitation products in the literature	12
Table 2-3. Comparison of satellite-based precipitation products	15
Table 6-1. Percentage of localised events in test gauges in the Kelani and Kalu River Basins.....	61
Table 6-2. Summary of the floods in the Kelani and Kalu River Basins	62
Table A-1. List of rain gauges used for the model training along with their station ID and location.....	97
Table A-2. List of rain gauges used for the model testing along with their station ID and location.....	99

LIST OF ABBREVIATIONS

AMW	Active microwave
ANN	Artificial neural networks
BS	Bias score
CHIRPS	Climate hazards group infrared precipitation with stations
CMORPH	Climate prediction center morphing technique
CPC	Climate prediction center
CSI	Critical success index
DEM	Digital elevation model
DL	Deep learning
DML	Double machine learning
ELM	Extreme learning machines
GBDT	Gradient boosting decision tree
GCN	Graph convolutional network
GNN	Graph neural networks
GSMaP	Global satellite mapping of precipitation
GWL	Groundwater level
GWR	Geographically weighted regression
IDW	Inverse distance weighting
IM	Inter monsoon
IMERG	Integrated multi-satellite retrievals for GPM
KED	Kriging external drift
KNN	K-nearest neighbour
LSTM	Long short-term memory
LST	Land surface temperature
LSTday	Daytime land surface temperature
LSTnight	Nighttime land surface temperature
MAE	Mean absolute error
MGWR	Multiscale geographically weighted regression
ML	Machine learning

MLP	Multilayer perceptron
MPE	Multi-sensor precipitation estimation
MSWEP	Multi-source weighted-ensemble precipitation
NDVI	Normalised difference vegetation index
NEM	Northeast monsoon
OCK	Ordinary cokriging
OK	Ordinary kriging
PCC	Pearson's correlation coefficient
PERSIANN	Precipitation estimation from remotely sensed information using artificial neural networks
PERSIANN-CCS	PERSIANN cloud classification system
PMW	Passive microwave
POD	Probability of detection
PRISM	Parameter-elevation regression of independent slopes model
RB%	Relative bias
RBF	Radial basis function
RF	Random forest
RFSI	Random forest spatial interpolation
RMSE	Root mean squared error
SPPs	Satellite-based precipitation products
SR	Success ratio
SVM	Support vector machines
SVR	Support vector regression
SWM	Southwest monsoon
TH	Thiessen polygon
TMPA	TRMM multi-satellite precipitation analysis
TPS	Thin plate splines
UK	Universal kriging
VIS/IR	Visible and infrared
XGBoost	Extreme gradient boosting