# AN UNCERTAINTY-AWARE FLOOD MODELING FRAMEWORK WITH ENSEMBLE HYDROCLIMATIC DATA AND GRAPH NEURAL NETWORKS

**ANAGHA P**

**DEPARTMENT OF CIVIL & ENVIRONMENTAL ENGINEERING**

**INDIAN INSTITUTE OF TECHNOLOGY DELHI**

**MARCH 2026**

# AN UNCERTAINTY-AWARE FLOOD MODELING FRAMEWORK WITH ENSEMBLE HYDROCLIMATIC DATA AND GRAPH NEURAL NETWORKS

*by*

**ANAGHA P**

Department of Civil & Environmental Engineering

*submitted*

*In fulfilment of the requirement for the degree of*

**DOCTOR OF PHILOSOPHY**

to the

**INDIAN INSTITUTE OF TECHNOLOGY DELHI**

**MARCH 2026**

**This dissertation is dedicated to my family.**

To my parents, Ramachandran P and Baby K, for their unconditional love, sacrifices, and steadfast belief in me.

And to my spouse, Abheesh P, for his unwavering love, understanding, and encouragement.

# CERTIFICATE

This is to certify that the dissertation entitled **“AN UNCERTAINTY-AWARE FLOOD MODELING FRAMEWORK WITH ENSEMBLE HYDROCLIMATIC DATA AND GRAPH NEURAL NETWORKS”**, being submitted by **Ms. Anagha P**, to the **Indian Institute of Technology Delhi** for the award of the degree of **‘Doctor of Philosophy’** in the Department of Civil & Environmental Engineering is a record of bonafide research work carried out by her under my supervision and guidance. The dissertation work, in my opinion, has reached the requisite standards fulfilling the requirement for the degree of **Doctor of Philosophy**.

The material contained in the dissertation is original and has not been submitted in part or full to any other University or Institute for the award of any other degree or diploma.

**Prof. Manabendra Saharia**

Department of Civil & Environmental Engineering

Indian Institute of Technology Delhi

New Delhi-110016, India

# ACKNOWLEDGEMENTS

Life is a unique journey in which each of us learns through our own experiences, shaped by countless interactions and shared moments. Many individuals contribute, often in subtle yet profound ways, to my growth and understanding along the way. While it is impossible to name or thank everyone who has influenced my learning and supported me throughout this Ph.D. journey, I remain deeply grateful to all who have played a part in it.

I would like to express my heartfelt gratitude to my advisor, Prof. Manabendra Saharia, for his trust, unwavering support, kindness, and invaluable guidance throughout my doctoral journey. His insightful feedback, expertise, and encouragement have played an invaluable role in moulding and fostering my growth as a researcher. His visionary guidance has instilled in me the confidence to think boldly and aspire to achieve greater heights in my research and career.

I would also like to thank my research committee members, Prof. Arun Kumar, Prof. Dhanya C T, and Prof. Sayan Ranu, for their time, valuable suggestions, and constant support. Special thanks are due to my HydroSense team for their collaboration, shared learning, and friendship. I am also deeply grateful to my M.Tech. supervisor, Prof. Rama Rao Nidamanuri, for introducing me to the world of research, and to Prof. Gnanapazham L. and Prof. Ramiya A. M. for their encouragement in pursuing this path.

Furthermore, I would like to thank the India Meteorological Department (IMD) in Pune, where I had the opportunity to visit several times. I am sincerely thankful to Shri Mritunjay Mohapatra, the Director General of IMD and Dr. O. P. Sreejith, Head of CRS at IMD Pune, for their support and guidance. I am also grateful to the scientists at IMD Pune for their generous knowledge sharing and the valuable learning experiences I gained there.

A very special thank you to my family, who have been a constant source of love, encouragement, and support throughout my graduate studies. Finally, I would like to thank my colleagues from SIMLAB for their stimulating discussions and enduring friendships. To all those who may come across this acknowledgement, please accept my heartfelt appreciation for your mentorship, encouragement, and friendship. Your mentoring and encouragement have been invaluable in my development as a researcher and a better version of myself.

(Anagha P)

## ABSTRACT

Floods are among the worst destructive natural hazards, can result in significant fatalities, extensive property damage, large-scale displacement of people, and serious economic consequences. For effective flood monitoring and early warning systems, accurate streamflow prediction is essential. Although hydrologic models are commonly employed for streamflow simulation, their effectiveness is limited by biases and uncertainties stemming from nonlinear processes, model parameterization, and errors in meteorological inputs. This thesis aims to enhance the simulation of streamflow by addressing uncertainties related to input data, model structure, and parameters. The approach involves generating an ensemble precipitation dataset and using this as input for a hydrologic-hydrodynamic model to produce the ensemble streamflow dataset. Following this, the output of the hydrological model is post-processed by utilizing a data-driven methodology to further enhance the accuracy of streamflow predictions.

The first part of this thesis focuses on creating a high-resolution ensemble precipitation dataset for the Indian mainland, addressing the need for high resolution while incorporating uncertainty estimates. To achieve this, we created the Indian Precipitation Ensemble Dataset (IPED) by utilizing the vast number of precipitation gauge stations across India and applying a locally weighted spatial regression approach. IPED is a daily ensemble precipitation dataset comprising 30 members, available at 0.25° and 0.1° resolutions for the period from 1991 to 2023. It accounts for topographical attributes such as elevation, aspect, and slope along with the input gauge data. IPED is the first observation-based ensemble precipitation product for India and is anticipated to have numerous hydrometeorological applications. After developing this uncertainty-aware dataset, the next step is to evaluate its performance in detecting extreme precipitation events and assessing its usefulness for flood modeling.

Detecting Extreme Rainfall Events (ERE) like cloudbursts is challenging as they often occur in remote locations without gauge stations or radar coverage. To assess whether finer-scale, terrain-aware interpolation can improve ERE detectability without adding new stations. We evaluate the 0.1° IPED, developed using the same station network as the India Meteorological Department (IMD) dataset but using an improved interpolation technique, against three benchmark datasets: 0.25° IMD, the 0.1° Integrated Multi-satellitE Retrievals for GPM (IMERG) satellite, and the 0.12° Indian Monsoon Data Assimilation and Analysis (IMDAA) reanalysis dataset. This study demonstrates that the new ensemble-based precipitation dataset, combined with a district-specific 99.90th percentile threshold,

significantly enhances cloudburst like ERE detection compared to the other dataset such as IMD 0.25°, IMERG 0.1°, and IMDAA 0.12°. The findings of this analysis indicate two key points: (i) ensemble-based interpolation alone can substantially enhance ERE detection, and (ii) locally tailored thresholds are crucial in regions where rainfall gradients are steep.

In the next part of this thesis, the focus was on investigating how uncertainty in precipitation affects uncertainty in streamflow. This study presented an end-to-end approach for generating ensemble streamflow for the Narmada River basin. The IPED 0.25°, dataset is used as input serves as input for a nationally-calibrated hydrologic-hydrodynamic model for producing streamflow ensembles. The analysis revealed that the uncertainties in both precipitation and streamflow exhibit distinct scale effects, influenced by spatial and temporal averaging within the basin. However, time series analysis of the streamflow showed issues with peak timing in the ensemble simulations, emphasizing the need for an effective post-processing framework to better align modeled streamflow with the observed data.

To address, the streamflow biases and peak-flow timing errors, we introduced a spatiotemporal Graph Neural Network (GNN) model for post-processing ensemble streamflow predictions generated from the hydrologic-hydrodynamic model in the Narmada River basin. The proposed GNN framework in this dissertation integrates Graph Attention Networks (GAT) to represent spatial connectivity among gauging stations and Long Short-Term Memory (LSTM) networks to model temporal relationships. This framework effectively captures both upstream-downstream interactions and temporal evolution by using previous time steps as input features, in contrast to traditional hydrological models or standalone deep learning approaches. GAT introduces an attention mechanism to assign different importance to upstream stations based on their influence on downstream flows. The proposed GNN model reached the daily highest median Kling Gupta Efficiency (KGE) of 0.74, demonstrating that incorporating graph-based spatial connectivity significantly enhances the model's ability to identify spatiotemporal hydrological dependencies and improves overall streamflow prediction accuracy.

In summary, the strategies evaluated in this thesis highlight the importance of addressing uncertainty in hydrological modeling, leading to improved flood prediction models in complex hydrological regions such as India. The public release and operational transfer of IPED dataset to IMD enhances the potential for advancing the flood research and management in the country.

# सार

बाढ़ सबसे विनाशकारी प्राकृतिक आपदाओं में से एक है, जिससे जान माल को नुक्सान, और बड़े पैमाने पर लोगों की विस्थापन के नतीजे हो सकते है। बाढ़ की असरदार मॉनिटरिंग और शुरुआती चेतावनी सिस्टम के लिए, स्ट्रीमफ़्लो का सही अनुमान लगाना ज़रूरी है। हालाँकि स्ट्रीमफ़्लो अनुकरण करने के लिए आमतौर पर हाइड्रोलॉजिक मॉडल का इस्तेमाल किए जाते हैं, लेकिन उनका असर नॉन-लीनियर प्रोसेस, मॉडल पैरामीटराइज़ेशन और मौसम से जुड़े इनपुट में गलतियों से होने वाले बायस और अनिश्चितताओं की वजह से सीमित होता है। इस शोध प्रबंध का मकसद इनपुट डेटा, मॉडल स्ट्रक्चर और पैरामीटर से जुड़ी अनिश्चितताओं को दूर करके स्ट्रीमफ़्लो के सिमुलेशन को बेहतर बनाना है। इस तरीके में एक एन्सेम्बल प्रीसिपिटेशन डेटासेट बनाना और इसे हाइड्रोलॉजिक-हाइड्रोडायनामिक मॉडल के लिए इनपुट के तौर पर इस्तेमाल करके एन्सेम्बल स्ट्रीमफ़्लो डेटासेट बनाना शामिल है। इसके बाद, हाइड्रोलॉजिकल मॉडल के आउटपुट को स्ट्रीमफ़्लो के अंदाज़ों की सटीकता को और बेहतर बनाने के लिए डेटा-ड्रिवन तकनीक का इस्तेमाल करके पोस्ट प्रोसेस करना भी शामिल है ।

इस शोध प्रबंध का पहला भाग भारतीय मुख्य भूमि के लिए एक हाई-रिज़ॉल्यूशन एन्सेम्बल प्रीसिपिटेशन डेटासेट बनाने पर फ़ोकस करता है, जो अनिश्चितता के अनुमानों को शामिल करते हुए हाई रिज़ॉल्यूशन की ज़रूरत को पूरा करता है। इसे पाने के लिए, हमने पूरे भारत में प्रीसिपिटेशन गेज स्टेशनों के सबसे बड़े नेटवर्क का इस्तेमाल करके और लोकल वेटेड स्पेशल रिग्रेशन अप्रोच को लागू करके इंडियन प्रीसिपिटेशन एन्सेम्बल डेटासेट, IPED डेवलप किया गया। IPED एक डेली 30-मेंबर एन्सेम्बल प्रीसिपिटेशन डेटासेट है जो 1991 से 2023 तक के समय के लिए 0.1° और 0.25° रिज़ॉल्यूशन पर उपलब्ध है। यह इनपुट गेज डेटा के साथ ऊंचाई, ढलान और आस्पेक्ट जैसे टोपोग्राफ़िकल एट्रीब्यूट्स को ध्यान में रखता है। यह भारत के लिए पहला ऑब्ज़र्वेशन-बेस्ड एन्सेम्बल प्रीसिपिटेशन प्रोडक्ट है और इसके महत्वपूर्ण हाइड्रोमेटोरोलॉजिकल एप्लीकेशन होने की उम्मीद है। इस अनिश्चितता-जागरूक डेटासेट को

डेवलप करने के बाद, अगला कदम बहुत ज़्यादा प्रीसिपिटेशन इवेंट्स का पता लगाने में इसके परफॉर्मेंस को इवैल्यूएट करना और बाढ़ मॉडलिंग के लिए इसकी उपयोगिता का आकलन करना है।

बादल फटने जैसी अत्यधिक वर्षा की घटनावों (ERE) का पता लगाना मुश्किल है क्योंकि ये अक्सर दूर-दराज की जगहों पर होती हैं जहाँ गेज स्टेशन या रडार कवरेज नहीं होता। नए स्टेशन जोड़े बिना, बेहतर-स्केल वाली, टेरेन-अवेयर इंटरपोलेशन ERE का पता लगाने की क्षमता को बेहतर बना सकती है या नहीं, इसका आकलन करने के लिए, हम 0.1° IPED का मूल्यांकन करते हैं, जिसे भारतीय मौसम विज्ञान विभाग (IMD) डेटासेट के समान स्टेशन नेटवर्क का इस्तेमाल करके बनाया गया है, लेकिन एक बेहतर इंटरपोलेशन तकनीक के साथ, तीन बेंचमार्क डेटासेट के साथ: 0.25° IMD, 0.1° IMERG, और 0.12° इंडियन मॉनसून डेटा एसिमिलेशन एंड एनालिसिस (IMDAA) रीएनालिसिस डेटासेट। यह स्टडी दिखाती है कि नया एन्सेम्बल-बेस्ड बारिश का डेटासेट, ज़िले के हिसाब से 99.90th परसेंटाइल थ्रेशोल्ड के साथ मिलकर, IMD 0.25°, IMERG 0.1°, और IMDAA 0.12° जैसे दूसरे डेटासेट की तुलना में बादल फटने जैसे ERE का पता लगाने में काफ़ी सुधार करता है। इस विश्लेषण के निष्कर्ष दो खास बातें बताते हैं: (i) अकेले एन्सेम्बल-बेस्ड इंटरपोलेशन ERE का पता लगाने में काफ़ी सुधार कर सकता है, और (ii) उन इलाकों में जहाँ बारिश का ढलान ज़्यादा होता है, वहाँ लोकल तौर पर तैयार किए गए थ्रेशोल्ड बहुत ज़रूरी होते हैं।

इस शोध प्रबंध के अगले भाग में, इस बात पर फोकस किया गया कि बारिश में अनिश्चितता, स्ट्रीमफ्लो में अनिश्चितता को कैसे प्रभावित करती है। इस स्टडी में नर्मदा बेसिन के लिए एनसेंबल स्ट्रीमफ्लो बनाने के लिए एक संपूर्ण पद्धति प्रस्तुत की गई है। IPED 0.25°, डेटासेट का इस्तेमाल इनपुट के तौर पर किया जाता है, जो स्ट्रीमफ्लो एनसेंबल बनाने के लिए एक नेशनल लेवल पर कैलिब्रेटेड हाइड्रोलॉजिक-हाइड्रोडायनामिक मॉडल के लिए इनपुट का काम करता है। विश्लेषण से पता चला कि बारिश और स्ट्रीमफ्लो दोनों में अनिश्चितताएं अलग-अलग स्केल के प्रभाव दिखाती हैं, जो बेसिन के अंदर

स्थानिक और टेम्पोरल एवरेजिंग से प्रभावित होती हैं। हालांकि, स्ट्रीमफ्लो के टाइम सीरीज़ एनालिसिस से एनसेंबल सिमुलेशन में पीक टाइमिंग के साथ समस्याएं दिखाई दीं, जिससे मॉडल किए गए स्ट्रीमफ्लो को देखे गए डेटा के साथ बेहतर ढंग से अलाइन करने के लिए एक असरदार पोस्ट-प्रोसेसिंग फ्रेमवर्क की आवश्यकता पर ज़ोर दिया गया।

स्ट्रीमफ्लो बायस और पीक-फ्लो टाइमिंग एरर को ठीक करने के लिए, हमने नर्मदा बेसिन में हाइड्रोलॉजिक-हाइड्रोडायनामिक मॉडल से जेनरेट हुए एनसेंबल स्ट्रीमफ्लो पूर्वानुमानों की पोस्ट-प्रोसेसिंग के लिए एक स्पेशियोटेम्पोरल ग्राफ न्यूरल नेटवर्क (GNN) मॉडल पेश किया है। इस शोध प्रबंध में प्रस्तावित GNN फ्रेमवर्क, गेजिंग स्टेशनों और लॉन्ग शॉर्ट-टर्म मेमोरी (LSTM) नेटवर्क के बीच स्पेशल कनेक्टिविटी को दिखाने के लिए ग्राफ अटेंशन नेटवर्क (GAT) को इंटीग्रेट करता है ताकि टेम्पोरल डिपेंडेंसी को कैप्चर किया जा सके। यह फ्रेमवर्क पिछले टाइम स्टेप्स को इनपुट फीचर्स के तौर पर इस्तेमाल करके, ट्रेडिशनल हाइड्रोलॉजिकल मॉडल या स्टैंडअलोन डीप लर्निंग अप्रोच के उलट, अपस्ट्रीम-डाउनस्ट्रीम इंटरैक्शन और टेम्पोरल इवोल्यूशन दोनों को असरदार तरीके से कैप्चर करता है। GAT डाउनस्ट्रीम फ्लो पर उनके असर के आधार पर अपस्ट्रीम नोड्स को अलग-अलग महत्व देने के लिए एक अटेंशन मैकेनिज्म पेश करता है। प्रस्तावित GNN मॉडल 0.74 की डेली हाईएस्ट मीडियन क्लिंग गुप्ता एफिशिएंसी (KGE) तक पहुंच गया, जिससे पता चलता है कि ग्राफ-बेस्ड स्पेशल कनेक्टिविटी को शामिल करने से मॉडल की स्पेशियोटेम्पोरल हाइड्रोलॉजिकल डिपेंडेंसी को कैप्चर करने की क्षमता काफी बढ़ जाती है और ओवरऑल स्ट्रीमफ्लो प्रेडिक्शन एक्यूरेसी में सुधार होता है।

कुल मिलाकर, इस शोध प्रबंध में बताई गई स्ट्रेटेजी हाइड्रोलॉजिकल मॉडलिंग में अनिश्चितता को दूर करने के महत्व को दिखाती हैं, जिससे भारत जैसे मुश्किल हाइड्रोलॉजिकल इलाकों में बाढ़ की भविष्यवाणी के बेहतर मॉडल बन सकते हैं। IPED डेटासेट को पब्लिक में जारी करने और IMD को

ऑपरेशनल ट्रांसफर करने से देश में बाढ़ रिसर्च और मैनेजमेंट को आगे बढ़ाने की संभावना बढ़ जाती है।

**शोध का शीर्षक:** एनसेंबल हाइड्रोक्लाइमैटिक डेटा और ग्राफ न्यूरल नेटवर्क के साथ एक अनिश्चितता-जागरूक फ्लड मॉडलिंग फ्रेमवर्क

**छात्र का नाम:** अनाघा पी

**प्रवेश संख्या**: 2020CEZ8288

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AUC | Area Under the Curve |
| CAI | Climatologically Aided Interpolation |
| CAMELS | Catchment Attributes and Meteorology for Large-Sample Studies |
| CHIRPS | Climate Hazards Group InfraRed Precipitation with Station data |
| CP | Cumulative Probability |
| CSI | Critical Success Index |
| CWC | Central Water Commission |
| DDS | Distributed Dimension Search |
| DEM | Digital Elevation Model |
| DWE | Disastrous Weather Events |
| DWR | Doppler Weather Radar |
| ECMWF | European Centre for Medium-Range Weather Forecasts |
| EMDNA | Ensemble Meteorological Dataset for North America |
| EM-Earth | Ensemble Meteorological Dataset for Planet Earth |
| ERE | Extreme Rainfall Events |
| ERFS | Extended Range Forecast System |
| FAR | False Alarm Rate |
| FN | False Negatives |
| FP | False Positive |
| GAT | Graph Attention Networks |
| GCN | Graph Convolution Network |
| GEE | Google Earth Engine |
| GEFS | Global Ensemble Forecast System |

| | |
|---|---|
| GMET | Gridded Ensemble Meteorological Tool |
| GNN | Graph Neural Network |
| GNN | Graph Neural Network |
| GPCC | Global Precipitation Climatology Center |
| GPEP | Geospatial Probabilistic Estimation Package |
| GWN | Graph WaveNet |
| HP | Himachal Pradesh |
| HyMAP | Hydrological Modeling and Analysis Platform |
| IDW | Inverse Distance Weighing |
| ILDAS | Indian Land Data Assimilation System |
| IMD | India Meteorological Department |
| IMDAA | Indian Monsoon Data Assimilation and Analysis |
| IMERG | Integrated Multi-satellitE Retrievals for GPM |
| IPED | Indian Precipitation Ensemble Dataset |
| IQR | Inter Quartile Range |
| J&K | Jammu and Kashmir |
| JJAS | June, July, August, and September |
| KGE | Kling Gupta Efficiency |
| LDAS | Land Data Assimilation System |
| LP3 | Log-Pearson Type -III |
| LSM | Land Surface Models |
| LSM | Land Surface Model |
| LSTM | Long Short-Term Memory |
| LULC | Land Use Land Cover |
| LWLR | Locally Weighted Linear Regression |

| | |
|---|---|
| MAE | Mean Absolute Error |
| MAXD | Maximum Distance |
| MICE | Multiple Imputation by Chained Equations |
| ML | Machine Learning |
| NASA | National Aeronautics and Space Administration |
| Noah-MP | Noah land surface model with multi-parameterization options |
| NR | Not Reported |
| NRSC | National Remote Sensing Centre |
| NSE | Nash-Sutcliffe Efficiency Index |
| NWH | Northwestern Himalayas |
| NWP | Numerical Weather Prediction |
| PCP | Estimated Precipitation |
| PET | Potential EvapoTranspiration |
| POD | Probability of Detection |
| PoP | Probability of Precipitation |
| QC | Quality Control |
| RMSE | Root Mean Square Error |
| RN | Random Normal Deviate |
| RNN | Recurrent Neural Network |
| ROC | Relative Operating Characteristic |
| SCOPE Climate | Spatially Coherent Probabilistic Extended Climate dataset |
| SCRF | Spatially Correlated Random Fields |
| STGNN | Spatio-Temporal Graph Neural Network |
| TP | True positives |
| TRMM | Tropical Rainfall Measuring Mission |

| | |
|---|---|
| UCRB | Upper Colorado River Basin |
| UK | Uttarakhand |
| US | United States |
| WDF | Wet Day Fraction |