Integration of Satellite Precipitation Products (SPPs) and Deep Learning (DL) for Extended Lead Time in Event-based Forecasts of Riverine Flooding Risk

10.36828/thesis/13929

School of Computing and Engineering

Integration of Satellite Precipitation Products (SPPs) and Deep Learning (DL) for Extended Lead Time in Event-based Forecasts of Riverine Flooding Risk

Cristiane de Fatima Donde Girotto

A thesis submitted to the University of West London in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Supervisory team

Principal: Professor Kourosh Behzadian

Secondary: Professor Massoud Zolgharni

Advisor: Dr Farzad Piadeh

Advisor: Professor Luiza Campos

16th of April 2025

Declaration

I hereby declare that the contents of this thesis are original, and the work presented is my own, except where specific reference is made to the work of others. I confirm that this thesis has not been submitted, in whole or in part, for consideration for any other degree or qualification in this, or any other university.

Cristiane de Fatima Donde Girotto

16th of April 2025

Acknowledgements

This thesis represents an extensive and complex body of research that would not have been possible without the contributions of several key individuals and wider academic community at UWL.

First and foremost, I am deeply grateful to my principal supervisor, Professor Kourosh Behzadian, for his encouragement to take on this challenge, his insightful guidance, and unwavering support. He is an excellent supervisor that is knowledgeable, accessible to all, mindful of his student's needs, generous with his time, an incredible career motivator and promotor of inclusion. Professor Behzadian provided me with numerous opportunities to learn and teach, expand my professional network, present my work internationally, engage with other students, and contribute to significant projects at national and international level. I am grateful and honoured to have been under his supervision.

I extend my sincere thanks to my secondary supervisor, Professor Zolgharni, for his support, insightful feedback and for given me access to the resources that were instrumental in raising the quality of this work. Additionally, I would like to express my deepest gratitude to Dr. Farzad Piadeh for his mentorship throughout this research and generosity on offering essential advice and sharing his profound technical knowledge. Also, for been a wonderful colleague and partner in the projects we work together in UWL. Dr Piadeh's contributions were critical to the successful completion of this PhD program.

Still within the School of Computing and Engineering at UWL, I want to thank Professor Fabio Tosti for kindly agreeing to review this work. Despite the subject being outside his main field of interest, he took the time to carefully review the extensive content, providing detailed feedback, offering pivotal suggestions, and encouraging me to make progress.

Additionally, as a recipient of the Vice Chancellor’s scholarship from the University of West London, I would like to show my gratitude to the research degrees department, particularly to the Head of the Graduate School, Professor Caroline Lafarge, and Senior Administrative Officer, Maria Pennells. Through its resources and though its members, the department offered me guidance thought the processes required for the completion of this program, provided a safe and supportive work environment, promote the engagement with the broader UWL academic community and financially supported and enabled me to showcase my research work in conferences where I made important work connections.

Outside the UWL, I am thankful to Professor Luiza Campos from UCL, which has been a long-term contributor to my research work. Since my Mres at UCL, Professor Campos has been kind in her guidance, generous with her time and a great motivator of my work often collaborating and endorsing my research projects.

Finally, I want to express my heartfelt gratitude to my family. From the older generation, that inspired and encouraged me to pursue an academic career, to the younger generation that are my me emotional backbone. My children that have recognized the depth of my efforts and appreciated my achievements, all while showing immense patience and understanding. Throughout this demanding program, the time I dedicate for them was dramatically reduced, yet they still made me feel loved unconditionally. This love is what makes me happy and sustains the resilience and determination to reach my objectives.

Above all, this research was made possible by the support of the people mentioned above, as well as many others who contributed in various ways, both to the research or to the researcher. This includes my fellow research students with whom I shared my struggles as much as learned from theirs, felling accepted and understood among an amazing group of very talented and kind people. I am sincerely thankful to all of you for making this achievement possible and just hope that everyone could have an experience like I had during my PhD program.

Abstract

Under current climate changing conditions, the urgent need for flood control and management calls for accurate event-based flooding forecasts with greater lead times to enable proactive strategies, which are more sustainable and less disruptive to the natural environment. However, increasing the lead time of predictions while keeping high accuracy is incredibly challenging due to the complexities of hydrological modelling and the limitations of rainfall data. These challenges are particularly significant in regions affected by severe weather originating from areas out of the range of land-based instruments. This thesis explores a groundbreaking alternative for increasing lead time in real-time predictions of riverine flooding caused by excessive rainfall in long-lasting weather systems from ungauged regions. The study focuses on predicting rising stream water levels - a key indicator of flood risk in natural rivers - using satellite precipitation products (SPPs), such as the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG). While IMERG data allows for the early detection of hazardous rainfall events in data-scarce areas, a Long Short-Term Memory (LSTM) deep learning (DL) algorithm is applied as a simple and fast alternative for real-time stream level predictions, relying solely on rainfall and stream level data. This minimalist data-driven approach makes the model's predictive power highly reliant on the quality and properties of the input datasets, which required the inclusion of data pre-processing steps that are crucial for the effective application of the extensive satellite-based data. The data optimization process takes into account the historical relationship between stream levels and precipitation data from the IMERG grid's units, or pixels, based on cross-correlation analysis. Additionally, the effects on modelling performance of grouping the pixel data in vertical and horizontal arrangements were also examined. The method, tailored to the characteristics of the flood prediction scenario, was fully tested and validated on predicting water level in the River Crane, near London, in the UK. The results show that the LSTM model successfully learned the historical patterns between stream levels and IMERG precipitation data collected from distances beyond 600 Km from the catchment. This ability to capture relationships over such large distances enabled real-time predictions of stream level variations with a lead time of at least 6.5 hours using IMERG data, resulting in error rates (RMSE = 6.92 mm, MAE = 3.11 mm) comparable to those of the model's using rain gauge data (RMSE = 6.30 mm, MAE =2.86 mm), which offered only a 0.5 hour of lead time. Overall, the model's performance with the NRT IMERG data was considered 'very good' (RSR < 0.5) for real-time predictions with lead times of at least 13.5 hours. Notably, the event-based forecasts using NRT data showed 'very good' model performance on lead times of up to 12 hours for the October 2022 event and up to 9 hours for the October 2023 event. The method's robustness was confirmed through similar performances in two other catchments, located at the south and north of England. Crucially, when integrated into the real-time flood forecasting decision support platform developed by Piadeh et al. (2023a), the proposed method achieved higher accuracy and longer lead times than the original approach, which relies on a NARX model using rain gauge data. Overall, the integration of LSTM modelling with IMERG data provided more accurate real-time predictions, offering at least a 2.5-hour increase in lead time compared to previous studies. However, the method has limitations for real-time application requiring lead times shorter than four hours, due to the latency in satellite-based rainfall estimates.

Keywords: Deep learning; Lead-time prediction; Real-time flood forecasting; Riverine flood risk; Satellite data; Real-time forecasts, Smart data pre-processing

Table of Contents

List of Figures

List of tables

List of Acronyms / Abbreviations

AAL	Annual Average Losses
ABI	Association of British Insures
ACC	Accuracy of true classification
AI	Artificial Intelligence
AMV	Active Microwave
ANFIS	Adaptive Neuro-fuzzy Inference System
ANN	Artificial Neural Network
CCC	Cross-Correlation Coefficient
CHIRPS	Climate Hazards center InfraRed Precipitation with Station data
CMORPH	Climate Prediction Center Morphing Method
CNN	Convolutional Neural Network
CKR	Cohen’s Kappa Rate
CSV	Comma-Separated Values
DEFRA	Department for Environment Food and Rural Affairs
DISC	Data Information Service Center
DL	Deep Learning
DPR	Dual-frequency Precipitation Radar
DT	Decision Tree
EA	Environment Agency
ELM	Extreme Learning Machine
FN	False Negative
FP	False Positive
FFNN	Feed-Forward Neural Networks
GES	Goddard Earth Sciences
GI	Green Infrastructure
GMS	Geostationary Meteorological Satellite
GOS	Global Observation System
GPCC	Global Precipitation Climatology Centre
GPU	Graphics Processing Unit
IMERG	Integrated Multi-satellitE Retrievals for Global Precipitation Measurement
ISCCP	International Satellite Cloud Climatology Project
IR	Infrared

JAXA	Japan Aerospace Exploratory Agency
KPIs	Key Performance Indicators
LEO	Low Earth Orbit
LID	Low Impact Development
LMNN	Levenberg–Marquardt Neural Network (LMNN)
LSTM	Long Short-Term Memory
MCC	Matthews Correlation Coefficient
MAE	Mean Absolute Error
MAE	Mean Absolute Percentage Error
MSE	Mean Squared Error
ML	Machine Learning
NARX	Nonlinear Auto-Regressive eXogenous
NASA	National Aeronautics and Space Administration
NBS	Nature-Based Solutions
NetCDF	Network Common Data Form
NFRA	National River Flow Archive
NRT	Near Real Time
PSONN	Particle Swarm Optimization Neural Network
PERSIANN	Precipitation Estimation from Remote Sensed Information using Artificial Neural Networks
PMV	Passive Microwave
RAF	Royal Air Force
RSR	RMSE-observations standard deviation ratio
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SDLRR	Spatiotemporal Deep Learning Rainfall-Runoff
SPPs	Satellite Precipitation Products
SuDS	Sustainable Urban Drainage Systems
SVM	Support Vector Machine
TN	True Negative
TNR	True Negative Rate (Specificity)
TP	True Positive
TPR	Total Positive Rate (Sensitivity)

TRMM	Tropical Rainfall Measuring Mission
VIS	Visible
VS	Visual Studio
WNARX	Wavelet-based Non-linear AutoregRessive with eXogenous inputs
WMO	World Meteorological Organization

List of submitted/published journal and conference papers

Girotto, Cristiane D., Piadeh, Farzad, Bkhtiari, Vahid, Behzadian, Kourosh, Chen, Albert S., Campos, Luiza C. and Zolgharni, Massoud (2023) A critical review of digital technology innovations for early warning of water-related disease outbreaks associated with climatic hazards. *International Journal of Disaster Risk Reduction,* 100. p. 104151. ISSN 22124209 http://dx.doi.org/10.1016/j.ijdrr.2023.104151

Girotto, C., Piadeh, F., Behzadian, K., Zolgharni, M., C. Campos, L., & S. Chen, A. (2024). Machine learning models for stream-level predictions using readings from satellite and ground gauging stations. *European Geosciences Union General Assembly 24*, Vienna, Austria. https://doi.org/10.5194/egusphere-egu24-13359

Girotto, C., Piadeh, F., Behzadian, K., Zolgharni, M., C. Campos, L., & S. Chen, A. (2024). Optimising oceanic rainfall estimates for increased lead time of stream level forecasting: A case study of GPM IMERG estimates application in the UK. Poster session presented at European Geosciences Union General Assembly 24, Vienna, Austria. https://doi.org/10.5194/egusphere-egu24-13509

Girotto, C., Piadeh, F., Behzadian, K., Zolgharni, M., C. Campos, L., & S. Chen, A. (2023). Role of Satellite Precipitation Products in Real-Time Predictions of Urban Rainfall-Runoff by Using Machine Learning Modelling. In EGU General Assembly Conference Abstracts (pp. EGU-10211). https://doi.org/10.5194/egusphere-egu23-10211

Girotto, C. D., Behzadian, K., Musah, A., Chen, A. S., Djordjević, S., Nichols, G., & Campos, L. C. (2024). Analysis of environmental factors influencing endemic cholera risks in sub-Saharan Africa. *Science of the Total Environment*, 926, 171896. https://doi.org/10.1016/j.scitotenv.2024.171896

Dondé, C.D. F. (1995). Desenvolvimento de método para avaliação da qualidade da apresentação gráfica dos projetos de obras de edificação. (*Development of a method for evaluating the quality of the graphic presentation of building works projects*). Salão de Iniciação Científica (07.: 1995: Porto Alegre, RS). Livro de resumos. Porto Alegre: UFRGS/PROPESQ, 1995 [Oral presentation].

1 Introduction

Climate change is progressing at increased rates with millions of people impacted by extreme weather events (WMO, 2024). The World Meteorological Organization reports that regional precipitation varied significantly in the last decade (2011-2020), with some regions abnormally wet and others abnormally dry (WMO, 2023a). Extremely wet periods were observed across the globe and record-breaking rainfall events caused severe flooding the United Kingdom (UK), Italy, Switzerland, Slovenia, Austria, Tunisia, Libya, Vietnan, India, Pakistan, China, Japan, Thailand, Cambodia, Laos, Philippines, Niger, Sudan, Somalia, Sierra Leone, Yemen, United States (US), Caribbean, Brazil, Paraguay and Australia. According to the same report, between 2011 and 2020, weather related hazards respond for 94% of disaster displacements, where flooding events were the major culprits displacing 123 million, followed by storms that displaced an estimate of 86 million.

The prospected intensification of extreme rainfall events as a result of climate change is a major source concern, as rainfall-originated floods are one of the most costly and destructive natural hazards on the planet (Westra et al., 2014). For example, to illustrate the financial burden caused by flooding, the UK experienced damages estimated at £1.6 billion from the 2015/2016 events (Environment Agency, 2018). More recently, in 2022-23, the Environment Agency (EA) reported that approximately 5.7 million properties in England were at risk of flooding (HC71, 2024). Additionally, according to the National Audit Office (2023), up to 51% of water supply infrastructure, 77% of rail infrastructure, and 21% of electricity infrastructure in the UK were assessed to be at high risk of flooding.

The excess runoff, which overcomes the drainage systems capacity, is a clear culprit in generating urban flooding that causes property damage and destroys livelihoods (Saghafian, et al., 2008), requiring urban infrastructure to react quickly. However, addressing flooding in the long term requires a global response to climate change and, under the Paris Agreement (UNFCCC, 2016), all countries are responsible for implementing mitigation and adaptation measures. Therefore, to face increasing risks from heavy rainfall and pluvial flooding due to changes in rainfall patterns, land use, and drainage capacity (Willems et al., 2012), is essential for nations to implement more sustainable solutions and enhance the resilience of large metropolitan areas while reducing CO2 emissions. This has prompted the development of concepts such as China’s 'sponge city' initiative (Chan et al., 2018) and nature-based solutions (NBS) in Europe that apply technologies like green infrastructure (GI), sustainable drainage systems (SuDS), and low-impact development (LID) to improve flood management (Wang et al., 2022). These concepts shift away from the traditional focus on flood protection, which relied heavily on hard engineering solutions such as increasing drainage capacity, runoff retention, and flood barriers. Instead, it moves towards a more proactive and sustainable approach, prioritizing flood control and risk management, and smarter urban flood drainage systems (Dong & Yang, 2019) with nature playing a greater role in prevention and management strategies (Jain et al., 2018).

Yet, the urgent global demand for changes in flood management is a long-term project that requires further learning from past events to effectively adapt government policies and urban infrastructure. Nevertheless, short-term proactive flood management can be achieved by adopting more flexible strategies and implementing new technologies. For example, areas prone to heavy rainfall can temporarily intercept and retain runoff during storm events (Mancipe et al., 2014), smart gates can be activated to divert excess water to areas with available storage capacity (Carbone et al., 2014). Sustainable Drainage Systems (SuDS), which mimic natural processes, can reduce or delay runoff (EPA, 2015). SuDS technologies like rainwater harvesting are particularly effective in controlling peak flow periods (Tavakol-Davani et al., 2016), while water conservation strategies provide a cost-effective method for reducing hydraulic loads (Nemerow et al., 2009).

However, these short-term, targeted, and more sustainable flood interventions demands high accuracy and long lead times in flood forecasting to ensure effective preparation and the timely deployment of management mechanisms. Increasing the lead time and accuracy of real-time flood forecasts is particularly challenging, as urban flood predictions rely heavily on the precision of early rainfall estimates and the complex task of modelling the catchment's hydrological response.

This thesis presents an innovative method for flood risk prediction, aimed at supporting the short-term implementation of more sustainable solutions for managing the impacts of climate change—particularly the increase in flooding. The approach combines novel data sources with advanced modelling techniques to enhance the accuracy and lead time of real-time riverine flood forecasts. In addition, by deepening our understanding of hydrological patterns, this research contributes to long-term policy development and urban planning strategies for adapting to more severe flooding events.

1.1 Research gaps and opportunities

The recent advancements of satellite precipitation products (SPPs), combined with the successful application of machine learning for capturing and predicting time series behaviour, presents a clear opportunity to address such challenges. However, many aspects of hazard modelling with machine learning still underexplored (Allen-Dumas et al., 2021), particularly on the concept of merging SPPs and deep learning modelling for real-time flood risk estimation. This research gap is confirmed by the lower number of related publications found in of some the most resourceful and widely used databases for hydrological studies, such as Scopus and Web of Science.

This study focuses on exploring the SPPs to address the shortfalls in precipitation data, given their wide coverage, high resolution, and minimal research into their application for real-time forecasting of event-based flooding risks.

To simplify the modelling tasks, data-driven models have proven to be a well-suited alternative. These models are capable of capturing non-linearity when representing physical processes response is challenging (Mosavi et al., 2018; Mounce et al, 2014; Yeditha et al., 2020; Yves et al., 2017). These models rely on the statistical interpretation of the historical relationships between the hydrological factors to predict the hydraulics response in spite of the physical mechanisms (Adamowski & Sun, 2010). They are also capable of finding regression patterns across different sets of data (Kumar & Monbey, 2019) and detecting historical relationships between flooding triggers and hydrological responses from long-term observations (Linardos et al., 2022). The application of machine learning models to extend forecast lead times was shown to be effective by Piadeh et al. (2023b), who successfully increased the lead time for water rise forecasts using an ensemble machine learning model that incorporates historical rain gauge and stream level data. Piadeh also suggested that incorporating satellite data could further enhance the model's performance.

Therefore, there is considerable evidence suggesting that the merging of SPPs and DL modelling are an alternative that must be explored to increase the lead time of real-time flooding risks forecasts.

1.2 Research Rationale

Maximizing the lead times while keeping high accuracy has been considered the ultimate goal in flood forecasting since flooding became an issue in urban areas. Consequently, a range of data sources and methodologies have been extensively explored by hydrologists and scientists in related fields. However, lead times are still small, while data sources and modelling technologies improve with the evolving technologies. Nonetheless, research on real-time flooding forecasting remains focused on refining modelling techniques, increasing the number of variables and sophisticating the data preparation while still using traditional sources of rainfall data.

Given that certain deep learning architectures can capture complex long-term dependencies between time series data (Schaefer & Zimmermann, 2008), which is the kind expected between rainfall and stream levels. Deep learning models are well-suited for the flooding forecasts based on rainfall (Thirumalaiah & Deo,1998; Campolo et al., 1999; Arnaud & Lavabre, 2002), these models are effective for modelling such relationships even when considering spatial and temporal variability (Xiang et al., 2018).

However, while reasonable accuracy has been obtained for forecasts up to three hours ahead (Piadeh et al., 2023a), the main challenge lies is improving the accuracy of event-based forecasts with lead times greater than four hours. This challenge presents an opportunity to explore the integration of SPPs and DL, which is undertaken considering the following hypothesis:

Hypothesis

The hypothesis considers that a deep learning model is capable of detecting and learning the historical patterns between precipitation estimates and stream level separated by great spatial and temporal distances.

The implications of such hypothesis for this study are explained in the following fictitious scenario:

- Assuming that an extreme rainfall event is originating over the ocean and that the first signs of the event are detected in an area out of the reach of land-based instruments, located 500 Km from the catchment.
- An event moving very fast, with an average forward speed of 50 Km/h (hurricanes average forward speed is around 30 Km/h), is estimated to reach the catchment in approximately 10 hours.
- SPP data for oceanic areas is available with a delay of four hours. This means that event's precipitation estimates observed at 500 Km are delivered when the event has already advanced 200 km toward the catchment and is now only 300 km away—expected to arrive in 6 hours.
- The subsequent SPPs data is delivered at 30 min intervals. Assuming the event maintains the same speed, the next estimate reflects its position at 475 km, though in reality, the event is already 275 km from the catchment. Similarly, when the data shows the event at 450 km, its actual position is 250 km away,

and so on. Consequently, the estimated time of arrival decreases proportionally to 5 ½ hours, 5 hours, and so forth.

- Given the previous conditions, and discounting the data delay of 4 hours, the first information of the rainfall event collected at 500 Km distance still available at least 6 hours before it can affect stream levels in the catchment.

Assuming the hypothesis is correct, the DL model has learned the patterns between remote SPPs data and stream levels at a specific catchment. Based on data collected 500 km away, the model can predict how the rainfall event will impact stream levels when it reaches the catchment 6 hours later. Given that the average forward speed is closer to 30 km/h and weather events typically do not maintain top speed throughout their entire trajectory, the actual lead time is likely to be greater than 6 hours in this scenario. Anyway, the model can be adjusted as the event progresses, with the new information provided along the data updates at every 30min.

In summary, advancements in machine learning offer a unique opportunity to explore this hypothesis. This study delves into every aspect of it to investigate its viability and effectiveness in increasing the lead time of event-based predictions using NRT data.

1.3 Aims and objectives

The aim of this research is to increase the lead time of real-time forecasts for catchments at risk of riverine flooding caused by long-lasting rainfall events originating from poorly monitored areas.

To achieve this aim, the research stablished three clear objectives:

- First objective: Investigate the suitability of satellite precipitation products (SPPs) for predicting riverine flood risks from rainfall and stream level observations using deep learning modelling.
- Second objective: Optimize the selection of the satellite precipitation dataset to minimize data volume, maximize correlation with river level rises, and ensure rapid processing for real-time predictions.
- Third objective: Assess the deep learning model's capacity to capture historical spatiotemporal relationships between SPP data and stream levels, and its performance with NRT data for real-time flood risk predictions with greater lead times.

1.4 Research questions

This study intends to address the identified gap in research while answering the following question:

"Can the integration of satellite precipitation products with deep learning data-driven models contribute to increasing the lead time in real-time predictions of riverine flooding risks?"

1.5 Research Scope

The scope of this research, depicted in Figure 1.1, is guided by the research gaps and common practices detected in the literature review and bibliographic analysis. It will explore and combine two mechanisms for increasing lead time of flooding forecasts:

- Earlier detection of rainfall events as main flood risk generators.
- Learn the historical patterns between rainfall and flooding risk increase.

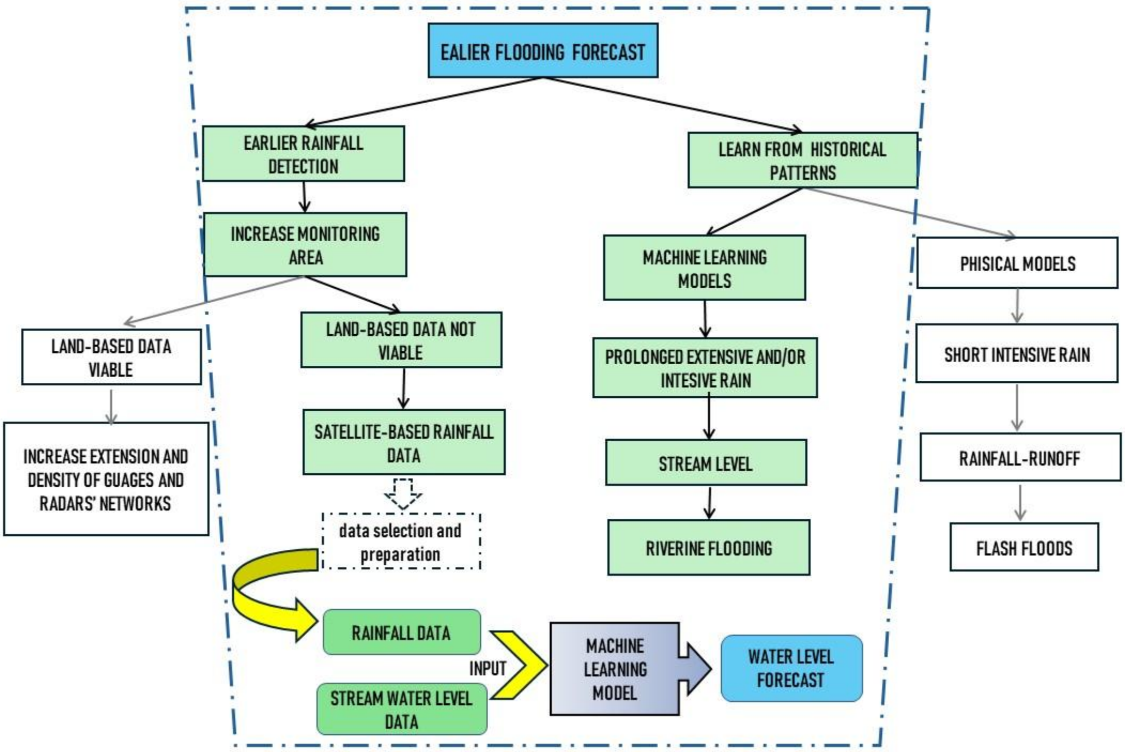


Figure 1-1 - Research scope diagram

The study focuses on rainfall as the main flooding generator and the use of satellite-based observations to extend rainfall data collection over areas where observations from land-based instruments are not available or not viable due to low accuracy. As such, the study investigates the suitability and requirements of the SPPs for use in flooding forecasts. The investigation process includes data inspection, selection, preparation and testing on forecasting models. Data inspection involves the assessment of properties and issues with the dataset; data selection is guided by ground validation and optimization criteria based on adequacy of the data for modelling; data preparation follows the requirements of the models; and performance is measured against ground-data and between models in a real case scenario.

The study is designed to address riverine flooding risk by using machine learning models to extract and learn the patterns between rain's spatial-temporal variations and stream water level changes. Flash floods are not addressed because they are usually caused by excessive runoff during intense and short events. Such events have a small time-window to be captured, and runoff is poorly monitored, data is scarce and lack continuity.

Finally, the machine learning models are trained and validated with historical data and tested for real-time predictions from NRT data. The forecasts accuracy is referenced by performance metrics like the RMSE and the MAE.

1.6 Research paradigm

Given the nature of engineering studies and the nature of this PhD, the positivist research paradigm is the most suited alternative to achieve the research objectives as "positivism is aligned with the hypothetico-deductive model of science that builds on verifying a priori hypotheses and experimentation by operationalizing variables and measures (Park et al., 2019). Data on rainfall, satellite precipitation products, and stream level data will be gathered from online sources and processed using computational and mathematical approaches in order to classify and clean up the datasets. The data provided is trusted as true replication of reality and is used to understand nature and technologies mechanisms. Findings of this research are of quantitative nature and the procedures can be replicated by other scientists that are expected to reach similar results for similar data input.

1.7 Research ethical considerations

As the research is being conducted within the engineering field, it adheres to the Statement of Ethical Principles recommended by the UK Engineering Council, which emphasizes four principles: "honesty and integrity," "respect for life, law, the environment, and public good," "accuracy and rigor," and "leadership and communication." Additionally, the research strategy and codes of practice developed by the University of West London, including the "Research Ethics Code of Practice," "Research Governance Policy," "Research Ethics Code of Practice," "Research Integrity Code of Practice," and "Procedure for Investigating Allegations of Misconduct in Research," are being followed.

The researcher has completed training on methodologies and is committed to following updating courses throughout the research period. The researcher is also dedicated to reporting their work with honesty and integrity, ensuring objectivity, carefulness, and confidentiality, and respecting the intellectual property of other researchers. Furthermore, this research also upholds research integrity as defined by the National Academy of Sciences. The researcher is committed to being honest and fair in proposing, performing, and reporting research, accurately and fairly representing contributions to research proposals and reports, showing collegiality in communications and sharing resources, and adhering to the mutual responsibilities of mentors and trainees.

Finally, the researcher understands that plagiarism is unacceptable and considers it an anti-value. They will avoid plagiarism by acknowledging others' work or ideas not only in text but also in other media, such as computer code, illustrations, graphs, etc. This applies equally to published text and data drawn from books and journals, as well as unpublished text and data, whether from lectures, theses, or other students' essays.

1.8 Thesis structure

This thesis has been organized into seven chapters:

1 - Introduction: This section describes the challenges and opportunities in the actual scenario of flooding forecast that triggered the research question and determined the aims and objectives of this study. The scope of this research is also defined in this section.

2 - Literature review: This section outlines the key studies and theories that form the foundation of the concepts and framework used in developing the methodology for this research. It also includes a bibliographic analysis of the publications consulted to identify common practices and gaps in the existing research. Furthermore, this chapter addresses the specific aspects of the research gaps aimed to be addressed in this study, including details on data sources, modelling technologies, background information, and data preparation.

3 - Methodology: This section describes the methodology structure and briefly describes the data sources, modelling technology, data processes and definitions of the criteria for performances' assessment.

4 - Case study: To assess the methodology a real case scenario was selected in a process described in this chapter. The chapter focus on the data collection, inspection, pre-processing, assessment and optimization tasks.

5 - Results and discussion: The results are presented following the order determined by the methodology. Each result is individually discussed along the presentation.

6 - Summary and Conclusions: This section presents an overall appraisal and discussion of the achievements and limitations of this research. The discussion also includes the potential application and future developments with the findings. The chapter ends with presentation of the main take aways from this thesis.

At the end of this thesis, after the list of references, an appendices sections provides extra material containing work carried out in the research and not presented in the main body. But still relevant to support future research or for clarification of the practices carried out in this study.

2 Literature review

This chapter is designed to present the pillars of this research's foundations, built from previous knowledge, stablished theories and concepts collected from previous studies. The source of the publications accessed extends over a comprehensive range of databases (Scopus, Web of Science, Google Scholar). These were used for access to the background information presented in the introduction section, for the identification of the main research gaps, for guidance on the construction of a methodology and for reference in the interpretation of results.

Follow a summary of the key topics addressed in this section:

- Background information: describes the key challenging elements for increasing lead time of flooding risk forecasts addressed by this research.
- Bibliography analysis and publications review: investigates the actual scenario of research on the application of SPPs and deep learning modelling techniques for flood forecasting. From that, the main focused areas and gaps in research are identified, and previous knowledge, stablished theories and concepts collected from related studies are highlighted in table A1 (Appendix section).
- Satellite-based rainfall products: looks into the key alternative source of rainfall data used in this study's methodology. It extracts information on the product's properties, features and limitations, and explores the previous applications and performance of SPPs in flooding forecasting.
- Deep learning techniques: investigates the second key element of the methodology dedicated to offer an alternative to hydrological modelling. It searches for renowned deep learning algorithms dedicated for time series

forecasting with known applications in the prediction of flooding risks, focusing particularly on risks of riverine floods induced by rainfall. This subsection includes information of the models' features, qualities, complexity, computational demand, limitations and data requirements. Finally, it looks into the applications and performances of deep learning models using satellite data in flood forecasting.

- Data processing: Identifies the data pre-processing and preparation requirements used to guide the construction of a methodology. reference in the interpretation of results.
- Performance assessment: Definition of the metrics for reference in the interpretation of results and validation of the methodology.

2.1 Background: challenges for increasing lead time of real-time forecasts

As noted by Piadeh et al. (2022), recent studies focused on real-time urban flood forecasting show that high accuracy is typically limited to short-term predictions of around 15 minutes, remains acceptable up to 60 minutes, and decreases significantly for forecasts beyond 90 minutes ahead of the event. Two of the main sources of inaccuracies in predictions with greater lead time are the difficulties related to rainfall predictions and hydrological modelling.

2.1.1 Rainfall predictions

Rain is difficult to predict and measure because rainfall has high spatial variability, is intermittent, and its onset involves atmospheric processes of chaotic nature. (Fritsch et al., 1998). Predictions of heavy rainfall, which is a common cause of urban flooding and a crucial element for flooding forecasts, are frequently associated with low accuracy (Song et al., 2019), low hit rates and false alarms (Tripathy et al., 2020). The accuracy of predictions is influenced by the characteristics of the event (such as intensity, extent, and duration) as well as by the capabilities and limitations of the instruments used for its detection.

There are several methods to detect and measure rainfall, as summarized in Figure 2.1, which shows the observation components of the Global Observing System (GOS). These include surface observations, upper-air observations, marine observations, aircraft-based observations, satellite observations, weather radar observations, and other observation platforms. The GOS, part of the WMO Integrated Global Observing System (WIGOS), is an international collaborative program for weather monitoring that has been running for over 60 years.

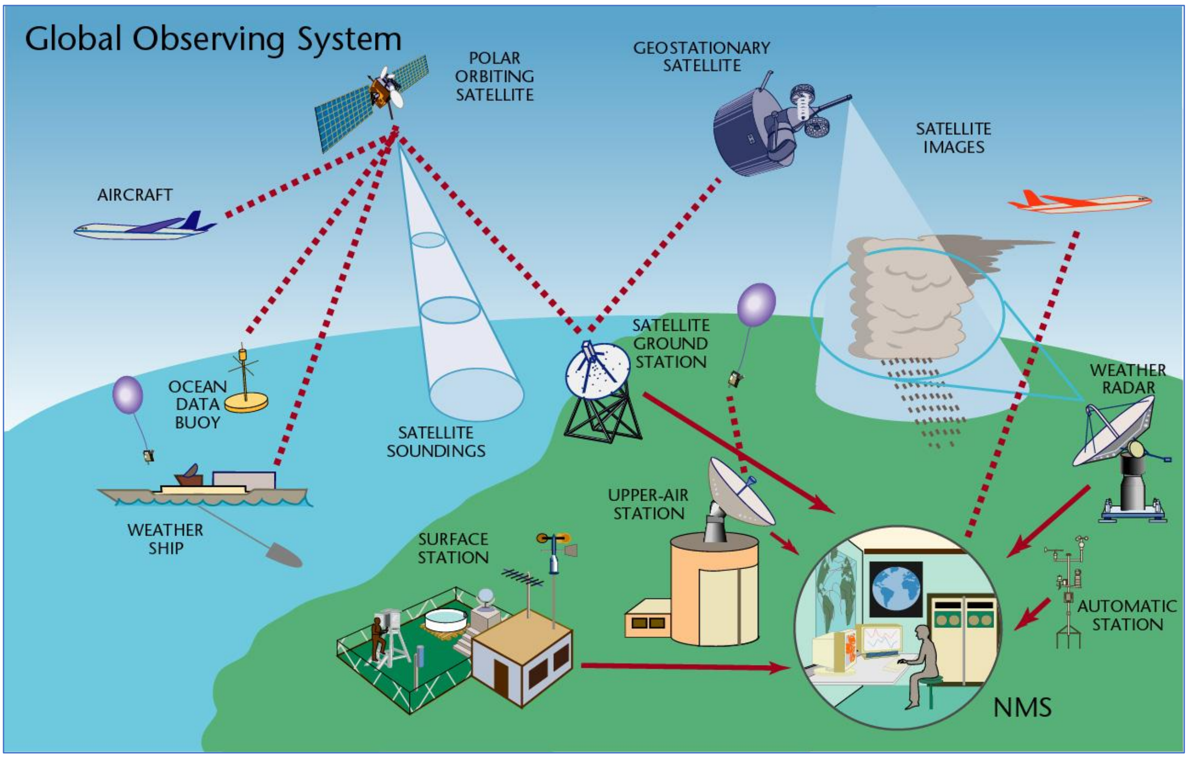


Figure 2-1 - The Global Observing System (GOS) – (WMO, 2020)

Rain gauges

Rain gauges at surface stations are the most common instruments for rainfall measurement. Although there are several different types of rain gauges, they are generally simple to install and easy to operate. They typically use a circular collector, or orifice, to capture rainfall and measure the volume of rain accumulated over a given period. In most official gauges, the orifice area ranges between 127 cm^2 and 400cm^2, with readings provided in millimetres (mm) at a minimum frequency of 15 minutes (WMO, 2023b). This instrument is considered the most accurate for measuring rainfall and is widely used to provide data for hydrological modelling. However, rain gauge measurements are subject to errors due to factors such as exposure to wind, which can cause under-capture of up to 5% of total rainfall in a typical event (Nystuen, 1999). Additionally, there are several limitations of gauge-based rainfall data that need to be considered for use in hydrological applications, as noted by Kidd et al. (2017):

- Individual gauges' readings are representative of the rain event at the point of collection and immediate surroundings, losing representativeness with increasing distance from the instrument.
- Rain gauging networks require a high density and uniform distribution of instruments to effectively capture the spatial and temporal variability or rainfall.
- The availability, consistency and frequency of rain gauges data varies considerably between years and from region to region in the Globe.
- The global coverage by rain gauges is minimal and uneven. Based on the foremost repository of gauge-based rainfall data, the Global Precipitation Climatology Centre (GPCC), is noted only 1.6% of the surface of Earth is within a 10km radius of a rain gauge, and 5.9% within a 25km radius. For inland areas

between 60°S and 60°N, where most rain gauges are located, these values increase to 6.5% and 23%, respectively. The concentration in such areas mean that rain gauges are even more sparsely distributed in other regions, particularly in oceanic areas, where only 4% is within a 100 km radius of a gauge.

- Globally, the majority of official rainfall data is provided as daily gauge accumulations, with sub-daily readings, such as 3-hourly measurements, being much less common.

Therefore, the availability of gauge-based rainfall data for hydrological applications will vary according with the purpose, requirements and targeted locations. This is particularly challenging for hydrological modelling that requires the interpolation of rain gauge data (Berne & Krajewski, 2013).

Land-based weather radars

Land-based weather radars are another commonly used instrument for the detection and quantification of rainfall. These instruments work by scanning the atmosphere in a 360-degree rotation, emitting pulses of electromagnetic waves and carrying back information of the particles in the atmosphere on their returning signal (Met Office, 2019). The reflective properties of these particles are used to classify the type of precipitation, as well as its intensity, shape and distance from the weather station. Weather radars can detect nearly all types of precipitations except for the very light events like drizzle or snow. The beams can scan up to 15 Km above ground and reach distances of up to 250 Km horizontally. However, quantitative precipitation estimates are considered reliable up to 200 Km (Sokol et al., 2021). In comparison with rain-gauges, individual weather radar instruments can capture rainfall data over much larger areas, providing more comprehensive and representative observation of rainfall variations across both space and time.

However, certain aspects of the precipitation data from weather radars must be considered for hydrological applications:

- Precipitation estimates from weather radars are often prone to underestimation or overestimation. This occurs due to the presence of common sources of errors during the instrument's scan of the atmosphere, such as beam blockages, bright band interception (Andrieu et al., 1997), and shadowing effects (Green et al., 2024). Other sources of errors include instability of electronics, poor quality of signal processing, echoes form non-meteorological targets, and even the curvature of Earth's surface (Sokol et al., 2021)
- The accuracy and detection capability of weather radars vary significantly according with the precipitation processes, the intensity of the rainfall and the distance from the instrument. For example, the primary instrument used for short-term rainfall forecasts (up to 6hs ahead) in the UK, the Nimrod radar, detects 50% of heavy rain (> 2mmh) and only 20% of light rain (< 2mmh) at a distance range of 200Km. In terms of detection lead time, the Nimrod detects 60 % of heavy rainfall between 0 to 3 hours ahead, dropping to 32% for lead times between 3 to 6 hours (Tilford, 2003).
- For streamflow modelling or forecasting using radar-based estimates, it is usually necessary to perform a radar rainfall retrieval process before applying the data in machine learning models. Is also important to quantify the uncertainty of radar-based rainfall data when used in physically-based models (Sokol et al., 2021). Furthermore, since the radar-based estimates are a product

of indirect measurements, they are subject to bias which may lead to large errors on hydrological modelling (Berne & Krajewski, 2013).

Considering these limitations, while weather radar data can offer valuable insights for short-term rainfall forecasting, it often requires to be complemented with other measurement sources to improve accuracy and reliability in hydrological applications. Recently studies have started to investigate how to integrate the key advantage of weather radar data and satellite-based cloud information higher accuracy and greater lead times on the predictions of extreme weather events like cyclones (Zhang et al., 2019).

Marine, aircraft and upper air instruments

While land-based instruments are limited by range, the uncovered areas can be reached by upper air instruments, like weather ballons and aircraft, or by marine instruments, based on drifting buoys, ships and platforms.

However, upper-air and marine instruments are often non-stationary and temporary, with readings collected sporadically, at irregular intervals, and with non-uniform spatial coverage. As a result, data from these instruments is not widely used for hydrological purposes—a situation that is unlikely to change in the foreseeable future. Precipitation data over the ocean will likely remain sparse, leading us to rely on satellites to provide such data more dependably(Serra, 2018).

Satellite-based weather instruments

An alternative rainfall estimation method comes with the advances of satellite observation capacity reaching global coverage, offering spatial resolution as fine as

0.05° (5 Km) and providing data updates at 30min intervals. Precipitation estimates produced from satellite observations have evolved in the last decades forming a comprehensive database for use in hydrological and climate studies (Kidd and Levizzani, 2011). SPPs are produced via a complex process of data collection, data transmission and data processing. The process involves a large network of satellites carrying sensors (IR, MW, etc) and other instruments to scan the atmosphere and transmit readings to a central hub that will process a large amount of data to produce the rainfall estimates. The extent and amount of the data used is the main determinant of the processing time and the quality of the results. Actually, rainfall estimates with acceptable accuracy require at least 4hs after observation to be delivered, while estimates with higher accuracy may take up to several months. Therefore, the large latency for higher accuracy in satellite-based precipitation estimates is an important limitation factor for certain hydrological applications and the use of SPPs data in flood forecasting models has been little explored.

2.1.2 Hydrological modelling for flood forecasting

The second important challenge for increasing the accuracy and lead times of flooding forecasts arises from the difficulties in modelling the complex flooding mechanisms. Intrinsically, flooding forecasts rely on the ability to understand and replicate the processes within and between the variables in a flooding event. Traditionally, forecasting models simulate flooding mechanisms and real-world scenarios using mathematical equations to represent the physical laws and processes governing the interactions between variables during an event. These models often allow for parameter adjustments to reflect the changing conditions of the numerous variables in urban environments, as well as to account for the non-linear relationships between them. Many of these variables, such as land cover, infiltration capacity, and evapotranspiration, are dynamic in nature, making it particularly challenging to accurately replicate them in physically-based models. In a real scenario this would require a deep knowledge of the internal interactions between variables as well as continuous data collection and updates to better represent changing conditions.

Furthermore, according to Lennon et al. (2014), solutions based on historical data are ineffective under the new conditions imposed by climate change and have great limitations for real time applications. Also, probabilistic methods for estimating extreme rainfall rely on a stationary assumption, which is contradicted by trend analysis of some rainfall records, incurring in poor performance of hydrological models (Garcia-Aristizabal et al., 2015). Therefore, there is a growing body of research highlighting the necessity of developing non-stationary models and frameworks (Hirsch, 2011; Cheng and Aghakouchak, 2014).

Nonetheless, research aimed at increasing the lead time of flood forecasts is starting to show results. Per example, in a recent pilot study in the UK, Piadeh et al. (2023a, 2023b) managed to improve the accuracy of event-based flooding forecasts up to 2-3 hours ahead. The author developed a system that works by continuously monitoring data and, when a potential flood event is identified, it isolates the relevant information. A specialized dataset is than generated based on past similar events and used on ML models for real time predictions. However, Piadeh faced difficulties on predicting greater lead time with higher accuracy because rainfall data is solely based on gauges readings, therefore limited to capture rain data already within the catchment area.

In search for solutions to address the challenges, the following bibliographic analysis was carried out to inform the scope and depth of this literature review.

2.2 Bibliographic analysis

This section presents a bibliographic analysis of publications available in the Scopus database carried out to scrutinise the current state of research within the scope of this study. The Scopus database was chosen due to its extensive collection of publications in the physical sciences and its compatibility with the VOSviewer software, which enables the creation and analysis of visual maps from the search results. The analysis of the visual maps allows for the identification key trends, dominant theories, innovative methodologies, emerging technologies and, more importantly, to discover new directions of research and underexplored areas. The search for publications started by looking into the broader topic of 'rainfall induced flooding risk forecast' by searching for articles in the Scopus database with words or terms linked to ((flood*) AND (rain* OR data OR precipitation) AND (predict* OR forecast* OR risk) AND (model* AND data-driven OR "neural networks" OR "machine learning" OR "deep learning" OR "artificial intelligence" OR "ai")) within articles' titles, key-words and abstracts for the terms. After filtering unrelated subject areas (i.e. social sciences, biology, veterinary, etc.), the search returned at total of 928 publications. A visual network map of these publications was created using the VOS viewer software. The map is exhibited in Figure 2.2 allows for identification of clusters and strengths of associations between the terms.

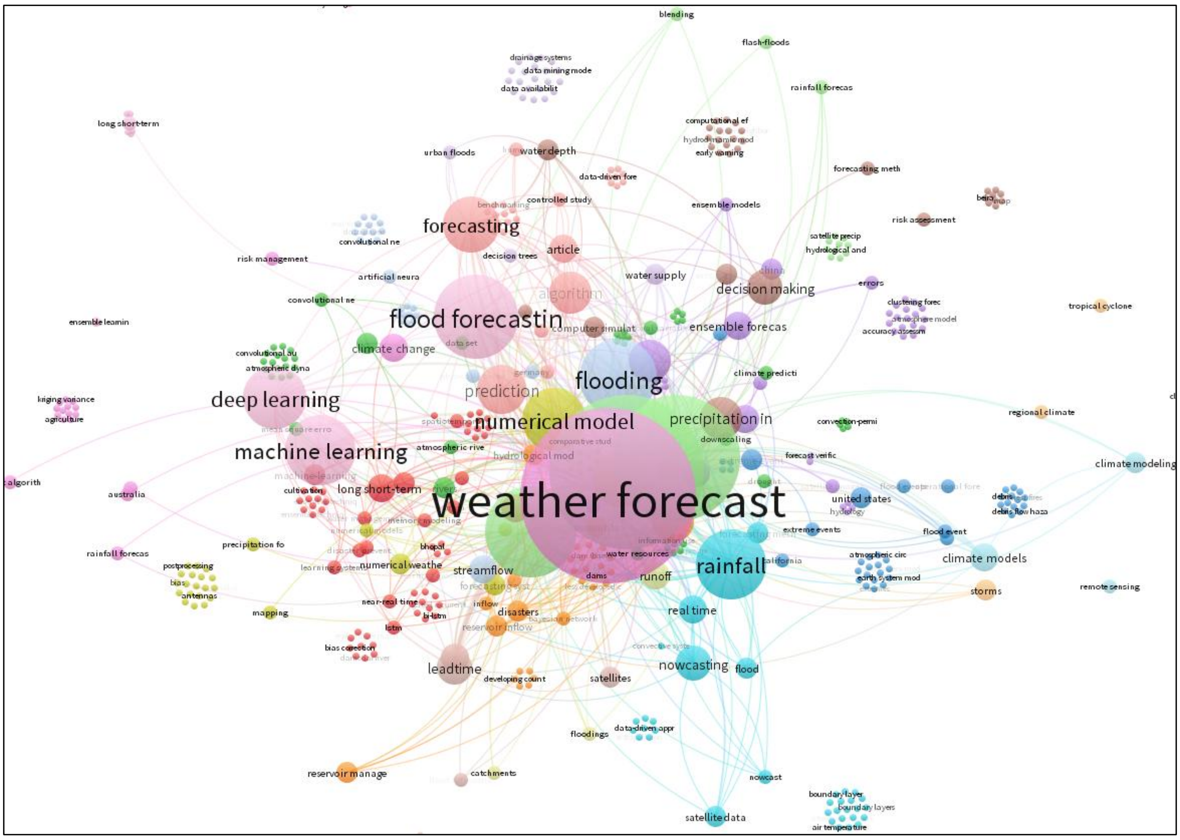
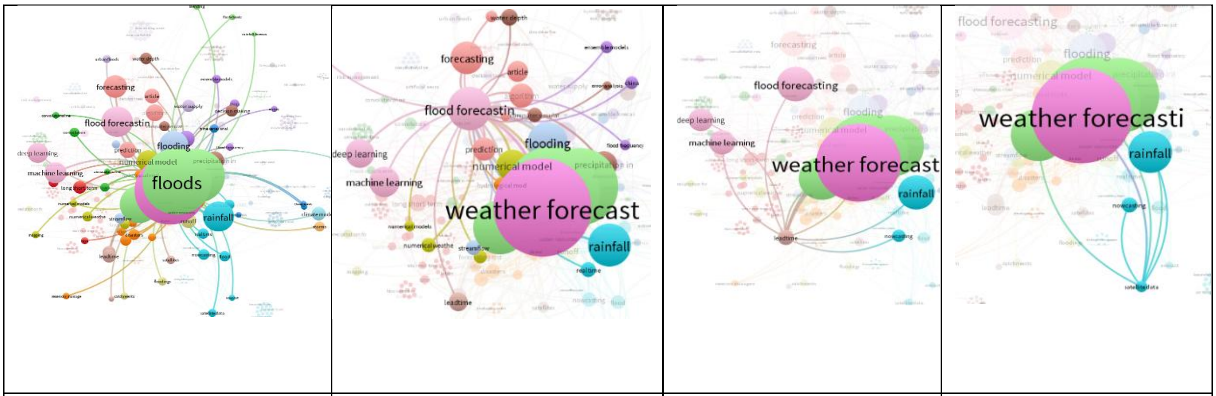


Figure 2-2 - Bibliographic analysis: all publications selected with the initial search criteria in the Scopus database.

The bibliographic network reveals a clear association between weather and flood forecasting with data-driven models, represented in Figure 2.2 by terms like ‘machine learning’, ‘ deep learning’, ‘numerical model’, etc. However, looking into the particular association with the terms that are the focus of this research like ‘flood’, ‘forecast’, ‘lead time’, ‘satellite’, etc. There are rare and weaker associations with satellite-related terms, as displayed in Figure 2.3 , which indicates that this is a field yet to be fully explored. This triggered the addition of terms related to ‘satellite’ in the search engine to gather information on how the academic community is using satellite-related products or equipment in research.



(a)	(b)	(c)	(b)

Figure 2-3 - Bibliographic analysis focusing on certain associations between the results from the initial search criteria. (a) Focused on terms associated with 'flood'. (b) Focused on terms associated with 'forecast'. (c) Focused on terms associated 'lead-time'. (d)Focused on terms associated with 'satellite data'.

As show in Figure 2.4 , the search including satellite-related 'terms' returned fewer publications but a much wider range of connections. These publications address a diverse range of applications, including for individual satellites sensors.

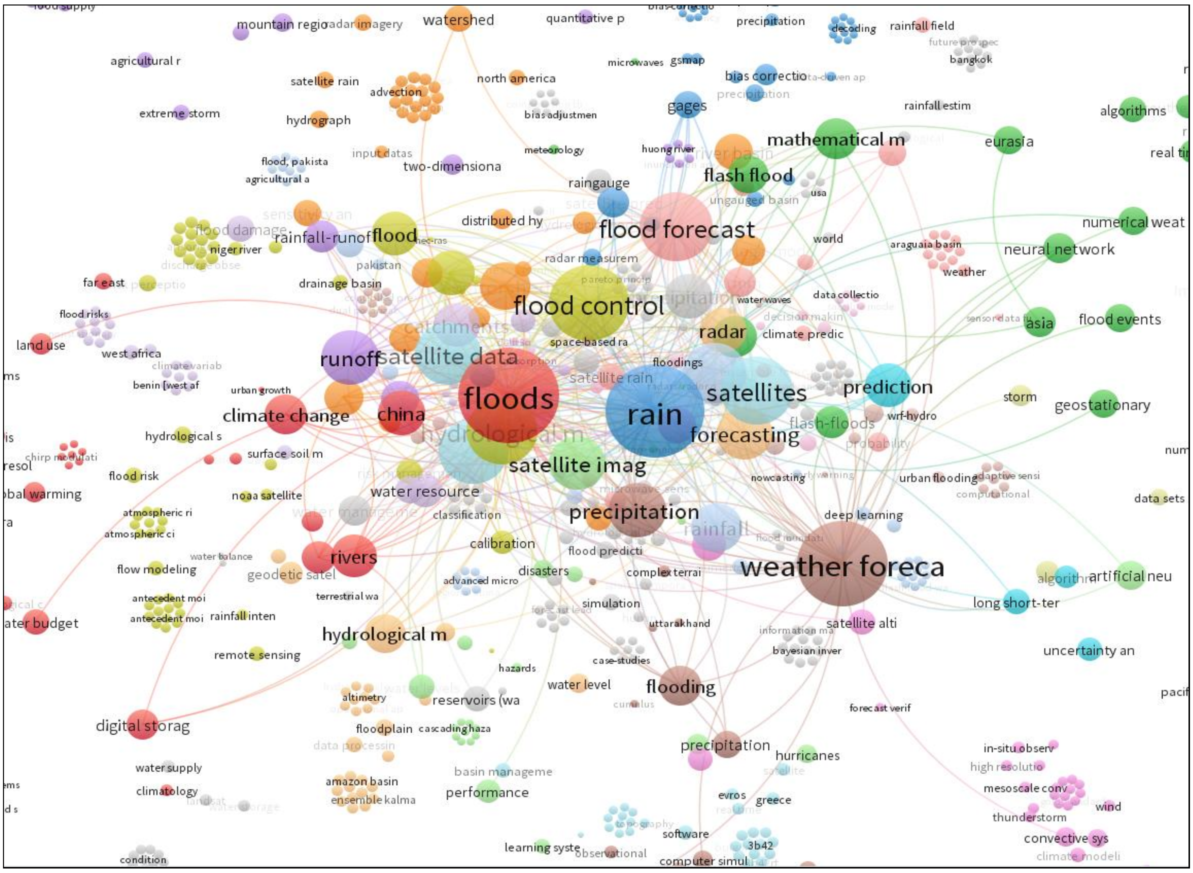


Figure 2-4- Bibliographic analysis including all publications selected with the inclusion of the terms for 'satellite' in the search criteria using the Scopus database.

To further narrow the search around the main topic, the terms associated with data-driven models were added including anything related to machine learning, neural networks, deep learning, etc. As a result, 30 publications were found addressing satellite-based rainfall products, flooding predictions and machine learning methods within their research scope. The results are displayed in Figure 2.5a which also shows the links between other terms within these publications.

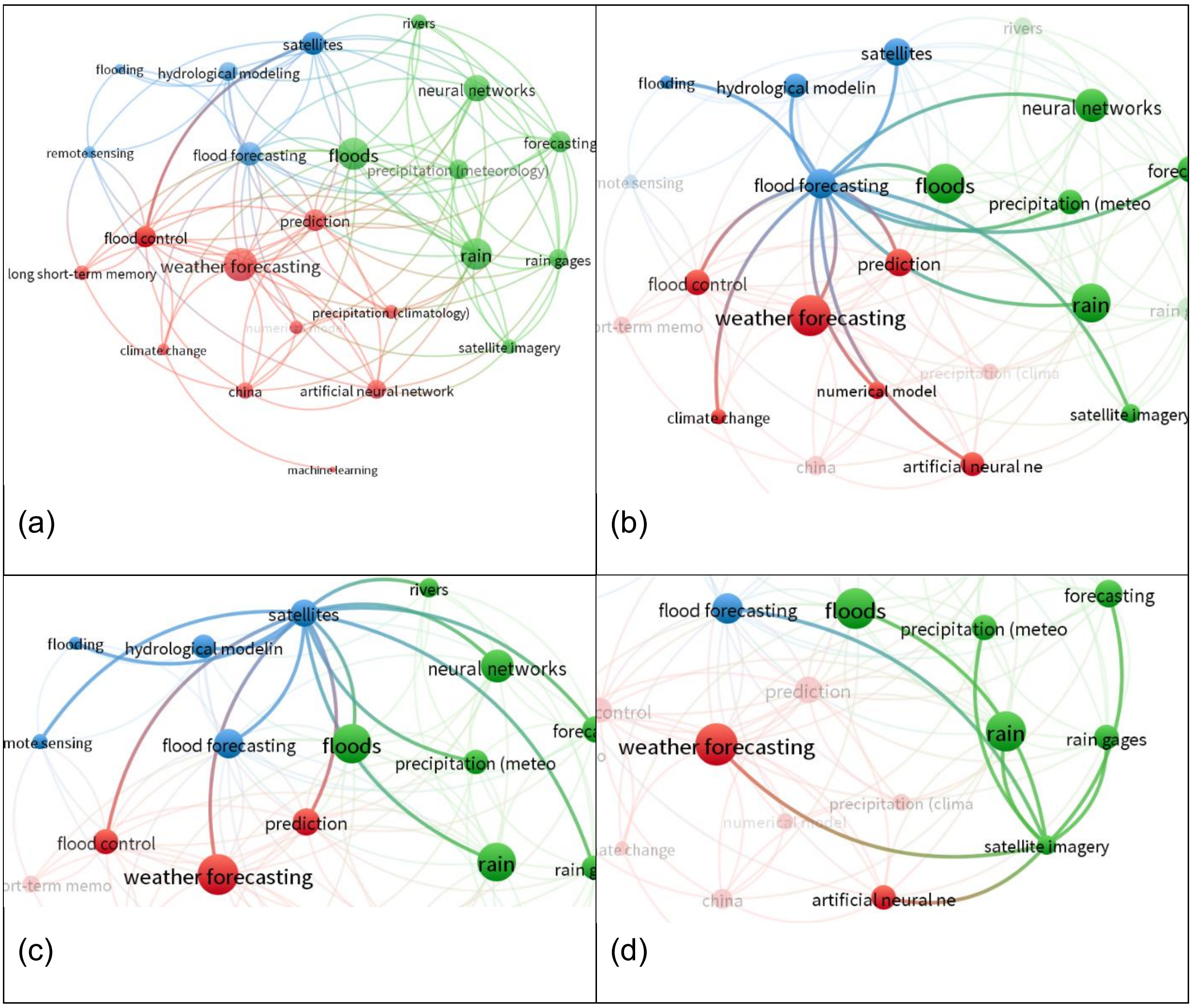


Figure 2-5 - Bibliographic analysis of publications selected with the inclusion of the terms related to 'satellite' and 'machine learning' term in the search criteria. (a) Terms from all publications. (b) Focused on terms associated with 'floods forecasting'. (c) Focused on terms associated with 'satellites.' (d) Focused on terms associated with 'satellite imagery'

The final selection returned only 26 publications that, after the exclusion of unrelated topics, revealed clear links with of satellite imagery (Figure 2.5d), often applied for the detection of flooded areas or estimation of the flood extent (Ireland et al., 2015; Tanim et al., 2022). Particularly useful for this study are the connections observed in Figure 2.5b and 2.5c, between satellites and terms related to flood or neural networks and, interesting, the link with the term ‘rivers’.

However, only five of those concurrently include applications the application of satellite-based rainfall data and machine learning. Of the five remaining publications, four address precipitation forecast but not specifically flooding forecast (Leinonen et al., 2023; Merizalde et al, 2023; Li et al, 2021; Ehsani et al., 2022), and one uses satellite data and machine learning for runoff and streamflow predictions but there is no provision for real-time predictions (Yeditha et al., 2022).

Following the bibliographic analysis on the Scopus’ material, the number of the publications assessed was broadened by including documents identified during the review process and through a similar search process in the Web of Science and Google Scholar databases. A concise overview of the main challenges and opportunities identified these publications in provided in Table A1 in the appendices section. .

The next subsections present the most relevant findings from the literature review, on the key topics of interest for this research thesis focusing on satellites-based precipitation data and deep learning models, which are the key elements of the research hypothesis allowing for increasing lead times on predictions.

2.3 Satellite precipitation products (SPPs)

Satellites started carrying precipitation sensors from 1997 and the first dedicated precipitation mission, the Global Precipitation Measurement (GPM), was launched in 2014. All SPPs are essentially derived from infrared (IR) or visible (VIS) data from the geostationary (GEO) satellite, and passive microwave (PMW) or active microwave (AMW) data from the low earth orbit (LEO) satellite sensors (Kidd & Levizzani, 2022).

A list with information on the LEO and GEO satellites operating across the globe is available in Table 2.1.

Table 2.1 - Current operational meteorological satellites.

Low Earth Orbit – LEO					
Program / Series	**Base**	**Movement**	**Altitude**	**Resolution**	**Observation Frequency**
European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)	Europe	Orbital	850 Km	5 - 15 Km	Twice a day
National Oceanic and Atmospheric Administration (NOAA)	US				
Defence Meteorological Satellite Program (DMSP)					
Feng-Yun 3 (FY - 3)	China				
Meteor - M	Russia				
Geostationary – GEO					
Program / Satellites	**Base**	**Movement**	**Altitude**	**Resolution**	**Observation Frequency**
Geostationary Operational Environmental Satellite (3rd gen) - GOES- 15/16/17	US	Stationary - positioned along the Equator	35786 Km	From 4Km 60°N-60°S.	Every 30 minutes
Meteosat Second Generation (MSG) (Indian Ocean) - Meteosat-8/9/10/11	Europe				
Himawari - 8 (3rd gen)	Japan				
Feng-Yun 3 (FY - 3)	China				
Indian National Satellite System - INSAT – 3	India				

Note: Adapted from Kidd and Levizzani (2022)

As described by Kidd and Levizzani (2022), none of the satellite sensors can capture precipitation on its own. The VIS and IR combination collects information on cloud top (temperature, pressure, height) and cloud particles (size and phase). The microwave sensors function similarly to the ground base weather radars, the PMW use sound and image frequencies to estimate precipitation characteristics from its effects on the atmosphere, and AMW provide measurements vertical profile of the precipitation.

Therefore, SPPs combine data from multiple sensors and products are tailored according to the user’s needs. More complex products will also use gauges readings to validate and adjust the calculations of estimates. However, each satellite sensor has its own limitations and specific role in SPPs, depending on the requirements to meet the end user's needs.

While there are several SPPS products (see Table 2.2), none can be considered universally superior. Their performance vary not just according to the products’ design but also with the several characteristics of the event observed, including its location (Kidd and Levizzani, 2022).

Table 2.2 - Active programs and SPPs with global/quasi-global coverage

Programs	Precipitation products	Spatial coverage	Data period	Spatial resolution	Temporal resolution	Minimum latency	Providers / website
Climate prediction center morphing method (CMORPH)	CMORPH	60°S - 60°N	1998	8 × 8 km	30 min	18 hours	NOAA - CPC *https://www.ncei.noaa.gov/products/climate-data-records/precipitation-cmorph*
Tropical rainfall measuring mission– (TRMM)	TRMM- 3B42 V7	50°S - 50°N	1998 - 2019 (ended)	0.25° × 0.25°	3 h	7 hours	NASA & JAXA *https://disc.gsfc.nasa.gov/datasets/TRMM_3B42_Daily_7/summary*
Climate hazards group infrared precipitation with station data (CHIRPS)	CHIRPS - V2.0	60°S - 60°N	1981	0.05° × 0.05°	Daily	2 days	Climate Hazards Center - UC Santa Barbara *https://chc.ucsb.edu/data/chirps*
Precipitation estimation from remotely sensed information using artificial neural networks (PERSIANN)	PERSIANN	60°S - 60°N	2000	0.04° × 0.04°	1 h	2 days	UCI Center for Hydrometeorology & Remote Sensing (CHRS)
	PERSIANN-CDR (climate data record)	60°S - 60°N	1983	0.25° × 0.25°	Daily	3 months	*https://chrsdata.eng.uci.edu*
The global precipitation measurement mission (GPM)	IMERG V06/ V07 (early -late -final)	60°S - 60°N	2000	0.1° × 0.1°	30 min	4hs (Early Run) 12hs (Late Run) 3-4 months (Final Run)	NASA & JAXA *https://gpm.nasa.gov/data/imerg*
Global Satellite Mapping of Precipitation (GSMaP)	GSMaP - V6/V7	60°S - 60°N	2000	0.1° × 0.1°	1 h	4 hours	Japan Aerospace Explorartion Agency - JAXA *http://sharaku.eorc.jaxa.jp/*

2.3.1 SPPs performance and areas of application

Satellite products have been tested on several applications and capabilities. Per example, for capturing extreme events (Behrangi et al., 2009; Hur et al. 2016; Joyce et al., 2004; Ly et al., 2022; Xiao et al., 2020), for detecting rainfall's spatiotemporal variation (Maggioni et al., 2016; Hamza et al., 2020; Pham et al, 2024) , for runoff estimation (Admojo et al., 2018; Al-Areeq et al., 2022), for simulating streamflow (Bitew et al., 2012; Chen et al., 2020), and for other hydrological applications (Gao et al., 2018; Peng et al., 2021; Moges et al., 2022; Guo et al., 2023; Wu et al., 2023).

The application of SPP products in flooding research was recently explored by Hinge et al. (2022), who reviewed over 1,300 publications to identify studies utilizing SPPs for flood simulation or prediction. The authors found that TRMM, CMORPH, IMERG and PERSIANN were the most frequently studied products. While the first three frequently overperformed other products on various settings, PERSIANN did not show superior performance to other SPPs in any comparison. Hinge's study also noted that TRMM was by far the most tested SPPs in flooding applications, achieving high success rates, except when compared to IMERG, which outperformed all other SPPs in nine out of ten times.

Integrated Multi-satellitE Retrievals for GPM (IMERG), is one of the most successful SPPS in flood-related applications. Developed as part of NASA's Global Precipitation Measurement (GPM) mission, IMERG blends data from the TRMM mission (2000–2015) with data from the GPM mission (2000–present), covering over two decades of continuous precipitation datasets. With the conclusion of the TRMM mission in 2019, IMERG is now expected to serve as an enhanced alternative, taking over TRMM's role in various applications.

Due to its exceptional performance in flood-related application the IMERG products are the focus of this research to represent the SPPs performance in flood risk predictions. Therefore, deserving a more detailed scrutiny of its characteristics and applications.

2.3.2 The GPM-IMERG precipitation product

The GPM is a project initiate by NASA and the Japan Aerospace eXploration Agency (JAXA) that comprises a consortium of international space agencies from the US, India, Europe and Japan. The network of satellites collaborating with the project is called GPM constellation and is composed of the CORE satellite and at least eight other satellites (Figure 2.6).

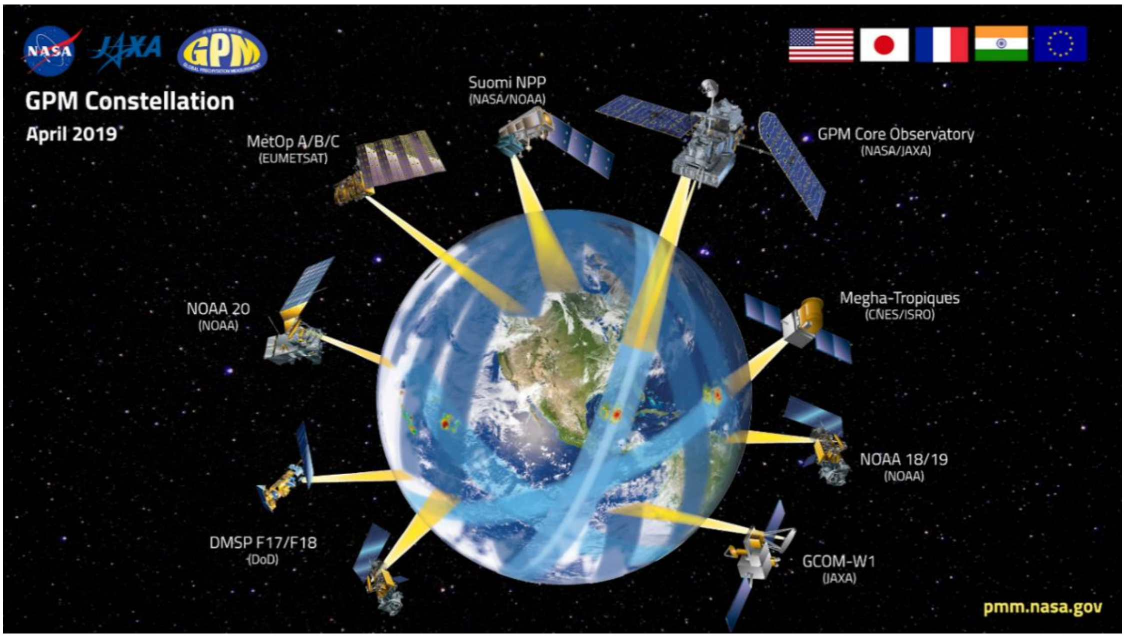


Figure 2-6 - Diagram of the GPM satellite constellation as of early 2019 (NASA GSFC, 2019).

The IMERG algorithm integrates data from multiple satellite platforms in the constellation (DPR radar, microwave imager, microwave and infrared sensors) on the estimation of precipitation for almost the entire surface of the Earth (Huffman et al., 2018). The algorithm is especially important for the regions of the globe that lacks ground-based rainfall measurement instruments.

IMERG precipitation estimates are available in three formats:

- IMERG Early Run uses data from fewer sensors in the constellation, to save in observation times, and only applies forward propagation, for quick processing. While this simplified version may compromise some accuracy, it offers NRT estimates delivered 4 hours after observation, making it the preferred IMERG product for real-time applications.
- IMERG Late Run uses all data available from the GPM constellation, meaning that all satellites have completed their observation cycle, and the information collected was transmitted to a hub. This data is processed using forward and backward propagation for better accuracy. The whole process produces a product with higher accuracy but requires 12hs to be completed.
- IMERG Final Run uses the most complete set of data and most complex process by aggregating monthly gauge analysis to adjust estimates. The Final Run product is delivered with 3 months latency but greater accuracy and is the version recommended for use in research.

In comparison with gridded rain-gauge data, IMERG Final demonstrated the best performance, followed by IMERG Late and IMERG Early (Sungmin et al., 2017). Particularly for application in flood modelling, Wang et al., (2017) suggests the use of IMERG Early Run for real-time flood forecasting and IMERG Final Run for hydrological modelling. The same suggestion is reiterated by Jiang & Bauer-Gottwein (2019) after evaluation the hydrological applications of IMERG in 300 catchments in China.

While in comparison with the other SPPs in Table 2.2, the IMERG Early is only product that combines fine temporal and spatial resolution with the lowest latency and overall better ability on detecting precipitation (Liu, 2023; Rachdane et al., 2022; Setti et al., 2023). The IMERG products also perform better than other SPPs in reproducing the variability and spatiotemporal patterns of extreme precipitation (Pradhan et al., 2022; Talchabhadel et al. 2021).

However, the IMERG algorithm incorporates data as and when it arrives, receiving it from several satellites, as seem in Figure 2.6. Each of these satellites has its particular orbital path, translation speed, communication capabilities and range of sensor instruments, which difficulties the maintenance of temporal homogeneity in the IMERG products and production of estimations over polar regions (Huffman et al., 2020). There are also some limitations related to estimating snowfall and precipitation intensity. However, the IMERG products have shown substantial improvements at each new version released since 2014 (Pradhan et al., 2022; Tang et al., 2020).

These improvements were also noticed in the results obtained in the numerous hydrological applications of the IMERG. While studies struggled with data from earlier versions (V03, V04, V05) of the IMERG (Wang et al., 2018; Yuan et al., 2018), recent studies show a superior performance of the IMERG V06 on runoff and flow simulations when compared with other SPPs (Ahmed et al., 2020; Hafizi & Sorman, 2022; Hamza et al., 2020; Jiang and Bauer-Gottwein, 2019; Le et al. 2020; Su et al., 2021; Sungmin et al., 2017; Yeditha et al., 2022).

Although the IMERG products are less commonly used for stream level (or river stage) simulations, the IMERG it is notably the sole precipitation data source in Google's operational flood forecasting system for inundation extent calculations (Nevo et al., 2022).

Overall, satellite-based precipitation products have been applied in several types of hydrological models (semi-distributed, distributed, lumped or probabilistic) carrying out flood simulations or flooding predictions (Hinge et al., 2022). Machine learning models have been using satellite images on the prediction of extreme flooding and on the identification of flood risk areas (Wardah et al., 2008; Yeditha et al., 2020).

Additionally, the application of satellite-based rainfall data, in conjunction with AI technologies for monitoring riverine water levels, has gained attention in several recent studies (Abdollahipour et al., 2021; Adane et al., 2021; Bitew et al., 2012; Nourani et al., 2021; Tam et al., 2019). These studies revealed that more research is needed to validate models, and that future research could benefit from using precipitation data from multiple sources. For more accurate flood water depth estimation, Ghorbanzadeh et al. (2018) and Elkhrachy (2022) recommend integrating multiple training datasets and machine learning algorithms. Overall, research on the integration of DL and SPPs for flooding forecasting is at the early stages and a novelty in real-time applications.

2.4 Deep learning (DL) in flooding forecast

Machine learning (ML) is a field of data-driven modelling techniques with expanding applications in urban flooding, particularly in deep learning methods. ML modelling entails programming computers to use training data or past experiences to improve performance by adjusting its parameters (Alpaydin, 2020). Prediction and classification tasks are carried out through the recognition of patterns learned from supervised or unsupervised learning techniques. Unsupervised learning models work with unlabelled data to detect data structure, clusters or hidden associations. Supervised learning models work with labelled input and output data providing a baseline reference for outputs. Supervised learning is used for regression tasks predicting continuous variables responses or for data classification (Rajoub, 2020).

The application of machine learning in flooding modelling is still a relatively new field of study. Unsupervised learning algorithms, like K-Means and Otsu, have been applied mostly on flood detection and mapping (Tanim et al., 2022; Nhangumbe et al., 2023; Foroughnia et al., 2022). Supervised machine learning applications extends to a much more diverse range of issues and there is a preference of specific algorithms for certain tasks. Per example, decision trees (DT) and random forest (RF) are mostly used in assessment of flooding risk and damage; convolutional neural network (CNN) is usually applied in the monitoring and mapping of flooding events, while support vector machine (SVM) and artificial neural networks (ANN) are the preferred ones for flooding forecasting (Kumar et al., 2023). However, although SVM is a robust and efficient method, it is not as popular as the ANNs which are used 3x more than any other ML method for flooding predictions (Mosavi et al, 2018). Additionally, the static feed-forward architecture of models like Multilayer Perceptrons (MLP) and SVM are not suitable for forecasting multiple steps ahead (Zhang et al., 2018). The ANNs, on the other hand, are popular due to its efficiency, versatility, processing speed and the higher ability for generalization between ML models (Antwi-Agyakwa et al., 2023).

Deep learning (DL) is a ramification of the ML range of ANN algorithms that can autonomously learn patterns in complex systems for prediction, detection, or classification tasks. These range of ANNs utilizes long clusters of several layers allowing for more advanced and abstract computational models that closely mimic real-world systems (Linardos et al., 2022). Additionally, DL models are also able to simplify the complex physical processes and number of variables used in hydrological modelling (Hayder et al., 2022).

The use of a DL model is based on its successful application in flooding research with rainfall inputs from several rainfall sources, including satellite-based products. Per example, DL models were able to capture the relationship between satellite images and rainfall data producing rainfall predictions with three days of lead time (Boonyue et al., 2019). The main limitation for the application of DL algorithms in forecasting is the requirement of large amounts of data, while other modelling techniques like SVM, and RF requires a strong correlation between input and outputs selection (Latif & Ahmed, 2023).

Between the DL models, the recurrent neural networks (RNN) are designed to process data in sequence and retain previous information to understand temporal patterns (Van et al., 2020). As such, the RNNs have a distinctive ability for capturing non-linear relationships and establishing time-series dependencies (Deng et al., 2022) making it highly suitable for flooding forecasting.

Two of the RNNs models are particularly suited for this study for its successful performance in forecasting multiple steps ahead: the recurrent neural nonlinear autoregressive with exogenous inputs (NARX) and the long short-term memory (LSTM) (Hayder et al., 2022). Follow a description of both models' structure and applications in flooding forecasting:

- **<u>NARX (Nonlinear AutoRegressive with eXogenous inputs):</u>** commonly used for time-series forecasting tasks. It is particularly suitable for situations where the relationship between past inputs and outputs is nonlinear and where exogenous (external) factors might influence the prediction (Lee et al, 2016). Recent flood-related applications of NARX models includes the forecast of rainfall-runoff water depth (Rjeily et al., 2017), predictions of river flow based on current rainfall (Noor et al., 2017; Hayder et al., 2022), increasing accuracy of flood risk from riverine water level predictions up to four steps ahead (Piadeh et al., 2023a) and simultaneous predictions of water level at multiple gauging sites (Lee et al., 2016).

- **Long Short-Term Memory (LSTM):** the LSTM are a version of the RNN with added memory cell and a selective mechanism using forget gates (Hayder et al., 2022). The LSTM were designed to excel at capturing long temporal dependencies in sequences and became popular due to its ability to deal with the vanishing gradient issue observed in other RNN models (Van Houdt et al., 2020). As they can capture long-range temporal patterns and are useful when temporal variances are important, LSTM are employed in forecasting of flood flow (Le et al., 2019; Hayder et al., 2022), flood timing and peak discharge (Ding et al., 2020), and estimating flood extent (Kim et al., 2022). LSTMs can be combined with CNNs in hybrid architectures for spatial-temporal data (Miau & Hang, 2020; Cho et al., 2022; Li et al., 2022;). The LSTM construction is based on Hochreiter and Schmidhuber (1997) architecture, comprising a memory cell (core unit), input, output and forget gates (gate units). A visualization of the architecture is provided in Figure 2.7 displaying the components, functions and internal routes of data flow.

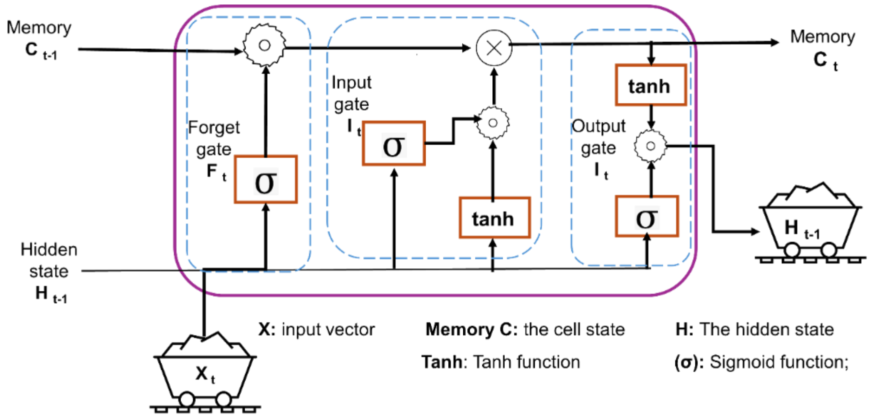


Figure 2.7 - Diagram of LSTM architecture, summary of processes and internal flow.

Finally, the applications of deep learning models in previous studies provide valuable insights on the LSTM model's performance. For example, for peak flow prediction in China, the LSTM model with only meteorological and river flow data showed a great performance in comparison to a hydrological model using a broad range of variables. Additionally, the LSTM models show a superior capability in the prediction of timeseries multiple steps ahead (Zhang et al., 2018).

2.4.1 Studies applying SPPs and machine learning in forecasting flood-related variables

The first integration of satellite-based precipitation data with machine learning algorithms was conducted by Hsu et al. (1997), who applied artificial neural networks (ANNs) to estimate rainfall from satellite infrared data. This work was done on the development of the PERSIANN satellite precipitation product, which was launched in 2000. Following this, as the number of SPPs product increased and satellite data expanded, other researchers explored similar combinations in studies with applications on flood forecasting as shown in Table 2.3.

__Table 2.3 Summary of studies with application in flooding forecasts combining satellite data and machine learning.__

Satellite Data	Machine learning	Application	Relevant outcomes	Limitations and relevant information	Reference
ISCCP-B3 dataset	ANN	Streamflow forecast (four watersheds)	The study claims hits above 60% in forecasts with 18h lead-time	-RSR values between good and satisfactory in two watersheds but unsatisfactory in the remaining two. -Lead times do not consider the period required for data collection, preparation and processing	Kim & Barros, (2001)
Infrared (IR) images from GMS-5	ANN (coupled with unit hydrograph runoff method – UH)	Flash flood forecast (ANN for rainfall estimation)	The study claims that 3–4hs forecast can be achieved from the coupled model. ANN model estimates with 16% error for 30mm of rainfall	-No metrics were presented for the rainfall-runoff. -Lead time on forecasts do not consider the period required for data collection, preparation and processing for producing the rainfall estimates or the rainfall-runoff prediction.	Wardah et al. (2008)
TRMM TRMM – RT	WNARX - dynamic neural network model	Streamflow forecast	Study claims that the TRMM-RT rainfall product could be reliably used for real-time streamflow forecasting up to 3-days lead-time. RSR = 0.46 and NSE = 0.79 indicates good performance of the model	-Three days lead time are equivalent to only 3 steps ahead predictions. - Lead times do not consider the RT data latency. - POD = 30% to 47% and FAR = 40% to 21% for 3 days forecast (Q_{90}) - TRMM products ended in 2019 -Daily-scale forecasts.	Nanda et al. (2016)
TRMM (3B42RTv7) PERSIANN	Wavelet neural network model	Streamflow forecast	Two days forecasts lead time with satisfactory range R^2 ~ 0.78 and NSE ~ 0.63 values.	- Not applied in real-time forecasting -PERSIANN product has a minimum 2-day latency	Yeditha et al. (2020)

			Predictions with both products are comparable to those using the IMD[1] data.	- Forecasts with high RSR values with both products. - Daily-scale forecasts. - TRMM products ended in 2019.	
TRMM – 3B42V7, 3B42RT PERSIANN	LMNN (basic ANN model) PSONN	Streamflow simulation	Models were tested also with temperature in the input, but performance was not significantly improved from input with only rainfall and flow data All Sat products performed well The PSONN model produced better results than the basic ANN model.	- Sat products used at daily resolution -Method requires the calibration of sat products with rain-gauge data to increase predictions' accuracy - The study suggests the investigation of sat products with higher spatial resolution (IMERG and CHIRPS) and modelling with deep-learning CNN.	Abdollahipour et al. (2021)
GPM – IMERG V06	LSTM	Google's flood forecasting system	LSTM model performance is better than linear model for river stage forecast. The model lead times varies between 8-48hs but only the overall metric NSE = 0.67 was provided. So, no info on performance for each lead time	-The requirements of large datasets is a drawback of ML models. -The study mention the wrong latency for IMERG Early Run (12hs) but does not specify if the latency is computed on lead times.	Nevo et al. (2021)

[1] *India Meteorological Department (IMD) provides gridded precipitation datasets, often constructed using ground-based observations, satellite imagery, and rainfall radar data*

TRMM – 3B42, 3B42RT CMORPH	FFNN, ANFIS, SVR (modelling) NNE (ML results ensemble)	Rainfall-runoff modelling	Combine multiple rainfall data sources is a promising option for modelling of rainfall-runoff Between sat products CMORPH performed better as individual input, but still worse than gauge ANFIS best performing ML model overall	-Model set to predict only one step ahead with daily-scale forecast -Study looked into the model's performances along the wet and dry season but didn't address effects of individual rainfall events or effect of inputs of only wet or dry periods. -The best performance was achieved when including rain-gauge data	Nourani et al. (2021)
GPM – IMERG V06	LSTM	Flash-flood simulation	The use of LSTM to merge IMERG and gauge data improved a hydrological model's performance on simulations IMERG high spatial coverage and fine temporal resolution are an advantage	SPPs data need to be assessed and corrected before use in hydrological research.	Tang et al. (2021)
GPM - IMERG V06 CHIRPS	ELM LSTM	Rainfall-runoff in two basins	Prediction's errors were smaller with the combination of CHIRPS data and ELM modelling providing one day ahead predictions with NSE ~ 0.87 Both models performed well LSTM slightly better in one basin and ELM in the other.	-Errors on predictions with SPPs data still higher than those with IMD. -Predictions were tested for only 1 step ahead (1day) -GPM-IMERG is more sensitive to light and medium rainfall -CHIRPS better on capturing heavy rainfall -Daily-scale forecasts.	Yeditha et al. (2022)

CHIRPS GPM – IMERG V06 (Final Run) TMPA	LSTM LSTM + ConvLSTM	Spatiotemporal deep learning rainfall-runoff (SDLRR) forecasting model	IMERG has better spatial representation than TMPA or CHIRPS LSTM is considered to be one of the best models for predicting streamflow The best results were achieved with the IMERG data	-TMPA overestimates rainfall of low intensity -Existing research is yet to explore the deep-learning architecture for spatiotemporal rainfall-runoff forecasting -The study did not consider lead times or real-time applications.	Zhu et al., (2023)
GPM – IMERG Early Run V06 PERSIANN-CCS	Random Forest (RF)	Extreme Runoff	The model is able to simulate extreme runoff for precipitation events of long-duration despite the event's extent (NSE = 0.72)	-The model performance for events of short duration and large extend was deemed unacceptable. -SPPs products were used to complement each other and there is no specific evaluation of the performance of the model for each product	Muñoz et al. (2024).

The summary in Table 2.3 focus on the studies' most relevant outcomes, limitations, and key factors to consider within the scope of the present research. The table highlights the focus of these studies on streamflow or runoff forecasts, the absence of sub-daily scales and the lack of real-time applications. It also reveal the frequent use of LSTM models in the most recent research works that combine satellite data and ML, suggesting a good fitness of these models to reach the objectives of this thesis .

2.5 Data handling and mining

This research deals with large time series datasets as input on DL models, which suggests the adoption of a data mining-based approach for in dept data scrutiny, optimization of datasets and efficient extraction of relevant information. Data mining is the process of extracting useful information from large datasets using a range of tools and techniques encompassing traditional and innovative methods of data analysis and processing (Jackson, 2002). While the organization and techniques of a data-mining framework vary according with the data and main objectives, most of following steps are usually included: Database creation, data inspection, data cleaning, data selection, model selection, data preparation, model construction and assessment. Some of these steps are straight forward tasks once the scope and objectives of the study are defined. Others are of more complex execution of or can be carried out in several different ways which demands better knowledge to support its execution. Follow a brief description of each:

- Database creation: data is collected and organized based on the requirements of the research objectives. Datasets sourcing and format vary greatly according to the field and aims of the research (Paradis et al., 2016). Data can be sourced online, from experiments, publication, questionnaires, observations, etc. More importantly, the sources must be reliable.

- Data inspection: investigates the datasets properties, missing data, outliers, inconsistencies, etc. It may involve statistical and mathematical analysis, or visualization of graphs, images, etc.
- Data cleaning: data cleaning deals with the issues detected during the inspection process by correcting outliers, filling data gaps, removing misleading information, eliminating data repetitions, or anything else that induces error in trend predictions.
- Data selection: feature selection of data filtering processes is applied to remove irrelevant information allowing faster processing and more objective learning in modelling tasks.
- Model definition: the ML model is selected from a wide range of algorithms to best fit the purpose of the project and the properties of the input data.
- Data preparation: the datasets are processed to attend the input requirements determined by the algorithm selected for the DL model. The processing may involve data normalization, sampling, formatting, etc. Normalization is a preprocessing technique that transforms feature values to a similar scale, ensuring all features contribute equally to the model. It is essential for datasets with features of varying ranges, units, or magnitudes. This process improves model performance, convergence, and prevents bias from features with larger values.
- Model construction and adjustment: the initial model's training options and parameters are selected, including the accuracy and precision

measurements. The model's performance is evaluated after the first run and, if not satisfactory, the initial settings can be changed. As models offer several options of parameters, there are a great number of combinations available for testing on performance improvement. The testing and adjustment process is also known as hyperparameters tuning.

- Assessment: analysis of the model's accuracy and precision in results. Comparison with other model's performances, computational demands, running time, etc. Overall evaluation of the outcome's applicability in real-world tasks.

Some data mining techniques are particularly useful for projects like this one, dealing with non-linear time series, spatial data, and temporal patterns. Spatial hotspots, heterogeneity, spatial dependency, and many discrepancies can be detected through visualization and use of statistical tools (Golmohammadi et al., 2018), while cross-correlation analysis is used for feature selection, data associations or data discrepancies. According to Chen (2015) cross-correlation analysis can detect the indirect relationships between spatial variables and "reveal the importance of the part played by geographical distances or spatial relationships". Overall, these techniques help to reduce the number of calculations, processing times, computational demands and optimize learning in ML models (Farahani, 2020).

2.5.1 Performance assessment of deep learning models

A machine learning model's performance is measured from the outcomes on modelling task using unseen data. Evaluation is conducted to determine if there are similarities or consistencies between the observed outcomes and the predicted results, or among the predictions generated by different models (Jackson, 2002).

In regression tasks, models are employed on the prediction of numerical continuous outcomes like stock market trends, oil prices, weather patterns, etc. While there are several metrics for the evaluation of regression models, there is no consensus on which is the ‘best’ one. The most common metrics in regression are the mean absolute error (MAE), mean squared error (MSE) or root mean square error (RMSE), and mean absolute percentage error (MAPE).

The MAE metric calculates the average difference between predictions and targets considering the absolute values of errors. MAE is a metric that is easy to interpret, as it conserves the same units as the data, and has lower sensitivity to outliers than other methods. However, the absolute values do not carry information of the error’s direction, over or underestimation.

$$MAE = (1 / n) * \Sigma|y_{true} - y_{pred}| \tag{2.5.1}$$

The MSE considers the squared values of errors on predictions, which makes it a metric highly sensitivity to the data scale and the presence of outliers. Additionally, the MSE also squares the units of the predictions which makes it difficult to interpret in real world scenarios.

$$MSE = (1 / n) * \Sigma(y_{true} - y_{pred})^2 \tag{2.5.2}$$

The MSE distortion in units is corrected in the RMSE method by taking the square root of the MSE values. Therefore, the relationship between both metrics is monotonic and keep the same order on ranking models' performance. However, RMSE is of easier interpretation.

$$RMSE = \sqrt{(MSE)} \qquad (2.5.3)$$

Guidelines on the choice of metrics to evaluate the performance of forecasts of timeseries with the same scale, recommend the use of MAE and RMSE and does not recommend the use of percentage errors, like MAPE, due to non-symmetry (Shcherbakov et al., 2013).

A model's performance is considered best when RMSE and MAE values are as close as possible to zero. However, when these metrics are greater than zero, there are no clear thresholds, and their interpretation can vary significantly. To address this issue, Singh et al. (2005) suggests that RMSE and MAE values can be considered low if they are less than half of the standard deviation of the observed data. Similarly, Moriasi et al (2007) model evaluation guidelines suggests rating thresholds based on the ratio of prediction errors to the standard deviation of observations. This ratio is known as the RMSE-observations standard deviation ratio (RSR).

$$RSR = RMSE_{Pred} / STD_{Observ} \qquad (2.5.4)$$

Table 2.4 - Model performance rating thresholds according to Moriasi et al. (2007).

RSR values	***Performance rating***
RSR < 0.5	Very Good
0.5 < RSR < 0.6	Good
0.6 < RSR < 0.7	Satisfactory
RSR > 0.7 is UNSATISFACTORY	Unsatisfactory

.

Additionally to MAE and RMSE, Shcherbakov et al., (2013) defends the use of the coefficient of determination (R^2) as an informative tool in the evaluation of the regression analysis by machine learning model. In this case, R^2 is perceived as representation of the portion of the dependent variable that can be predicted from the independent variable.

$$R^2 = 1 - \frac{\Sigma\,(y_{pred} - y_{true})^2}{\Sigma\,(\overline{y_{pred}} - y_{true})^2} \qquad (2.5.5)$$

Given the lack of consensus on a single "best" metric, the use of multiple metrics is recommended for evaluating regression models (Naser and Alavi, 2021). This approach ensures a more comprehensive assessment of model performance, providing a balanced view of accuracy, error sensitivity, and interpretability.

3 Methodology

This study considers the stream water level rises due to excessive rainfall as flooding risk in catchments. Therefore, the methodology takes advantage of the opportunity for early detection of hazardous rainfall events over ungauged areas offered by the SPPs products worldwide coverage. The early detection allows for greater lead time of predictions if there is a proven historical correlation between the rainfall events at certain region and the stream level at a specific catchment. Additionally, the methodology also explores the ability of deep learning architectures to capture complex long-term dependencies between the stream level and the remote precipitation time series datasets. As a data driven technology, it is critical to thoroughly inspect the datasets to identify its properties, features and attributes that guides the choice of the model's architecture and definition of parameters to better fit the data.

This chapter presents the main structure of the methodology, data sources, stages and processes proposed in this thesis for stream level prediction. It also describes the metrics and processes to evaluate the datasets and model's performances.

3.1 Methodology structure

This study uses the IMERG precipitation products to represent the SPPs in the methodology. The choice is based on the IMERG products properties of global coverage, long period of historical data availability, short latency in the NRT products, fine spatial and temporal resolution, and easy access to datasets. Additionally, these products have a reputation of consistent good performances in flooding related applications as described in the literature review.

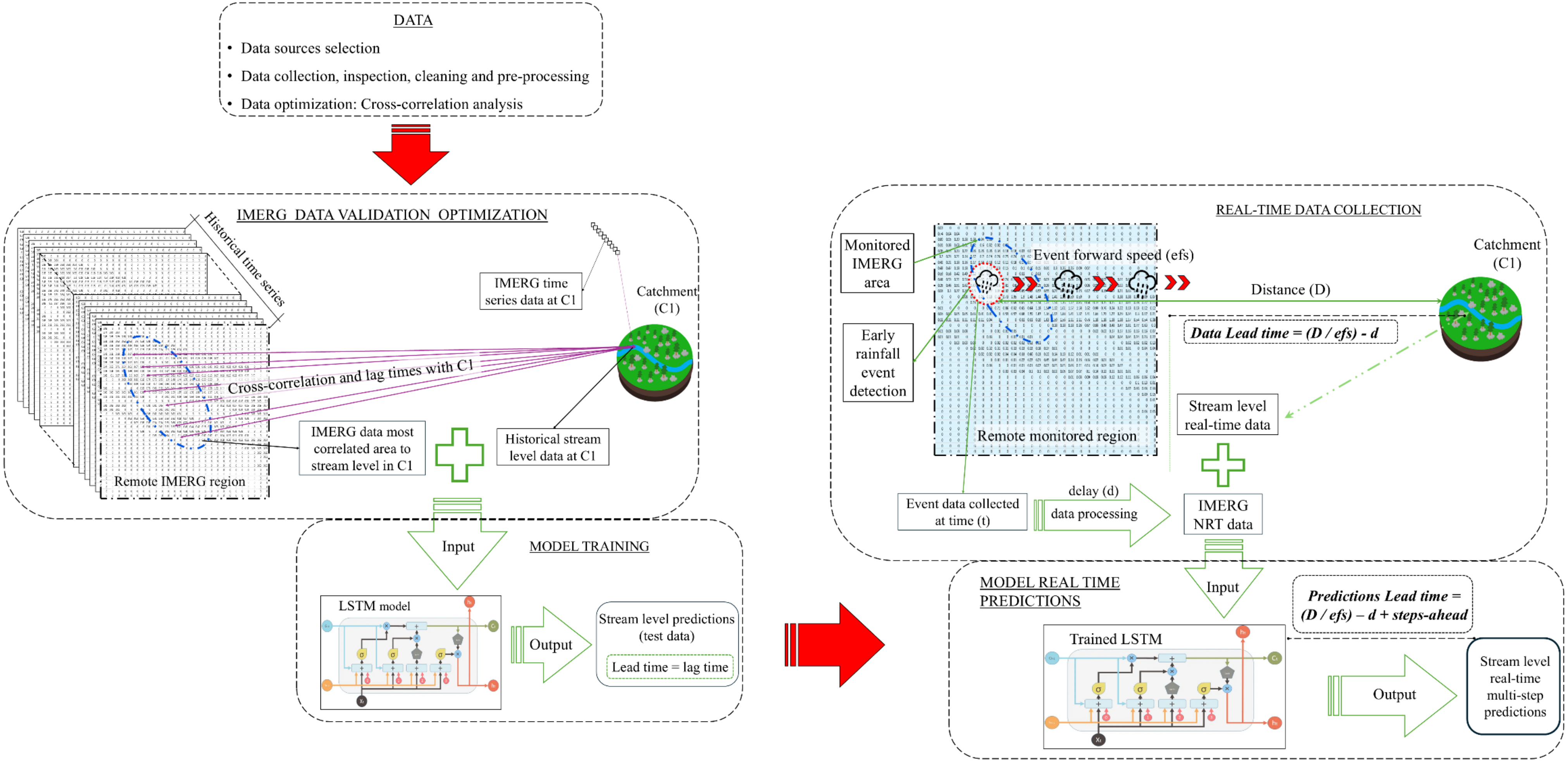


Figure 3-1 – Diagram summarizing the proposed methodology

The method starts by assessing the quality of the IMERG products along the data inspection process and progresses with the evaluation of feasibility and adequacy the local IMERG datasets to inform a deep learning model on stream level predictions using rain gauge data as reference for comparison and validation of performances. Next, the IMERG dataset is optimise and hot spot areas of its data grid selected for data extraction according with the strength of its historical correspondence with stream level at a catchment. The historical data extracted is tested for predictions with the deep learning model. Finally, the historical IMERG data is replaced by NRT IMERG data for evaluation of the real time predictions of stream level and on the performance for forecasting event-based flooding risks.

Figure 3.1 depicts the summarised methodology’s diagram, including several steps which are divided in four main stages:

1. DATA:
    - Identification of the catchment for forecasting flooding risks and a corresponding region of the IMERG global grid for data sourcing. The IMERG data must be available over the target area and the surrounding regions that are historical common routes of extreme weather events, particularly where there is no or there is restricted access to ground-based precipitation data.
    - Identify a rain-gauge at the vicinities of the catchment to provide a ground-truth reference of precipitation data that is used for the stream level predictions benchmarking the IMERG dataset suitability and performance on stream-level predictions.

- Historical datasets of catchment's stream level and respective rain-gauge data are collected to match the period of coverage of the IMERG dataset. Due to the IMERG global coverage, a region of data collection is delimited to reduce the amount of data collected while allowing the detection of rainfall events with targeted lead times.

2. IMERG data validation and optimization
   - Datasets are thoroughly inspected for detecting missing data, outliers, repetitions, and other data anomalies.
   - These anomalies are corrected using techniques of data filling, rectification of outliers, or for dealing with any other kind of misleading information in the datasets.
   - IMERG products are provided as grid data and each grid unit treated as a pixel.
   - Cross-correlation analysis: as the method deals with spatial and temporal variability between datasets, the cross-correlation function (CCF) is used to measure and compare the strength of the relationships, represented by cross-correlation coefficient (CCC) values, and lag times between each source of precipitation data and stream level at the catchment.
   - Data optimization: this task uses the CCC values and lag times of each IMERG data pixel to optimize and reduce the volume of data and processing times crucial in real-time predictions. It stablishes the regions with rainfall data most correlated with stream level variations in the catchment and with lag times superior to the target lead times for predictions.

3. <u>IMERG data modelling assessment</u>

- The complex task of predicting a catchment's response to a rainfall event is simplified by using a LSTM deep learning model and historical data from only two variables: stream level and precipitation.
- The LSTM model is constructed observing the lag times detected in the cross-correlation analysis, which allows the models to capture the temporal patterns between the datasets
- The model is trained with stream level and precipitation data as input producing stream level outputs with the designed sets of lead times.
- Model's parameters are adjusted and different data structures tested to reach best performances
- Performances of the model's predictions with SPPs and rain gauge data are compared considering the lead times and considering the spatial.
- The SPPs data is validated based on the performance of the model with data from the pixel closer to the catchment.

4. <u>IMERG data real-time application</u>

I. For early detection of rainfall events, remote regions of the IMERG data outside the catchment are selected to monitor weather events and collect data in real-time. The region is delimited considering the highest CCC values and respective lag times between the pixels and stream level historical data. The extent of the area also consider the target lead times for predictions.

II. For the modelling of real time predictions, the historical datasets of the selected SSP pixels are replaced with the NRT datasets.

III. Assessment of the method’s performance considers the error on predictions with the IMERG data and compare it with the errors using rain-gauge data for each lead-time.

The model’s algorithm is one of the most important elements of the methodology along with the IMERG data validation and optimization processes. While the model selected can be generally described in the next section, the IMERG data validation and optimization are highly dependent on the relationships between precipitation data and stream level. Therefore, the processes must be tailored considering characteristics that are particular to each catchment and will be described using a real scenario to be presented in the next chapter.

3.2 Forecasting model

The extensive coverage spanning over two decades and the large amount of data available from IMERG products offer an advantage for the application of data-hungry Deep Learning (DL) algorithms, especially when compared to other machine learning models with stricter requirements, such as a strong input-output relationship (Latif & Ahmed, 2023).

Several factors contributed to the selection of LSTM modeling techniques among deep learning algorithms, making it highly suitable for this study. Some of the most relevant factors identified through the literature review include:

- Capable of capturing complex relationships and storing long-term dependencies in sequential data (Van Houdt et al., 2020).

- Effectively addresses the vanishing gradient problem inherent in Recurrent Neural Networks (RNNs) (Van Houdt et al., 2020).
- Demonstrates strong performance in forecasting hydrological variables using Machine Learning (ML) with satellite data (Xu et al., 2020; Nevo et al., 2021; Thang et al., 2021; Yeditha et al., 2022; Zhu et al., 2023).
- Its deep learning architecture has proven successful for multi-step-ahead forecasting (Zhang et al., 2018; Hayder et al., 2022).

The Deep Learning (DL) structure of the LSTM algorithm was considered the most suitable for modeling the relationships between rainfall and stream level time series datasets. A key aspect of this methodology is the combined input of both time series into the LSTM, a strategy supported by findings that it enhances the accuracy of multi-step-ahead stream flow predictions (Badrzadeh et al., 2013; Panagoulia et al., 2017). Additionally, the spatial dependencies between the two datasets were considered by processing individual IMERG pixels, or sets of neighbouring pixels in different arrangements, and taking into account the respective lags with the stream level in the LSTM model input sequence size.

The sequence-to-sequence learning structure is employed in the LSTM model to capture the past temporal patterns of the lagged relationship between the two timeseries in the input dataset. The output sequence size reflects the number of steps ahead of the predictions. However, the LSTM model’s structure is shaped according with the properties of the datasets used for modelling and defined based the outcomes from the data inspection and selection tasks. While the data preparation processes are determined by the model's structure requirements, the architecture and parameters are adjusted according to each input dataset.

3.2.1 LSTM model performance evaluation

The quality of the model's predictions is evaluated using the R^2, RMSE, and MAE metrics. While R^2 is not directly related to error and cannot measure accuracy (Li, 2017), it serves as a useful tool for visualizing the relationship between observed data and predicted values across different forecast horizons.

Accuracy is primarily assessed through the RMSE and MAE error metrics, where the closer the values are to zero the better is the fitting of the model. However, when these metrics' values are greater than zero the interpretation of such metrics can vary significantly and in such cases the RSR value were used for rating the model performance according to Moriasi et al. (2007) thresholds shown in Table 2.4.

First, the IMERG data considered local to the catchment is validated against rain gauge data by comparing the error's metrics from the LSTM model predictions with each of these data sources. The evaluation focuses on both the overall performance with the testing subsets of the datasets and the model's ability to predict stream level responses to the individual rainfall events. A similar evaluation method is applied to the IMERG data considered non-local to the catchment, assessing both overall prediction performance and event-based accuracy using data from the pixels in different arrangements.

Next, the model is tested for real time applications. Up to this point, the models are trained and tested using data from the Final Run, which is the most accurate of the IMERG products, benefiting from the inclusion of data from a vast number of sensors and ground-based validation. As such, the performances of predictions with this product are expected to best possible with IMERG data. However, the Final Run product takes three months be delivered, which is a prohibitive length of time for real-time applications. Therefore, the dataset for testing the model for real time predictions, needs to be replaced with data from the IMERG Early Run product, which is a simpler product had the advantage of being delivered within 4hs of the observation.

Lastly, the efficacy of the LSTM model predictions is further investigated on its capacity to inform a decision support platform for real-time urban flood forecasting, developed by Piadeh et al. (2023a). This platform was proved successful on categorizing flooding and non-flooding events using stream level predictions produced by a E-NARX model which, as the LSTM model proposed in this study, used only rain-gauge and stream level data as inputs. This classification method will be referred in the study as 'DS platform' from now on.

The LSTM and E-NARX models' performance are compared based on the quality of their prediction, both using the same rain-gauge and stream-level datasets, for application in the DS platform. Additionally, to evaluate the overall methodology proposed in this study, the real-time predictions of the LSTM model using IMERG data are also integrated into the DS platform, and outcomes compared with the ones from the previous step.

The forecasts obtained with this method are compared with the true level at the catchment and categorized as follow:

- True Positive (TP): both observation and forecast indicate flooding.
- True Negative (TN): both observation and forecast indicate no flooding.
- False Negative (FN): flooding was forecasted as non-flooding (underestimation).
- False Positive (FP): non-flooding was forecasted as flooding (overestimation).

The analysis of the method's outcomes also looks into the key indicators of performance presented in Table 3.1 along with its respective formulas.

Table 3.1 - Key performance indicators for assessment of classification models.

Code	Covered concern	Range
TPR	Model sensitivity in recalling actual overflow condition, i.e. accuracy of overflow class	[0,100]
TNR	Model specificity in selecting actual non-overflow condition, i.e. accuracy of non-overflow class	[0,100]
ACC	Probability in that the model prediction is correct, i.e. interested in forecasting the right classes without caring about the type of the class or class distribution	[0,100]
MCC	Highlighting correlation and agreement between observed and predicted classes	[-1,1]
DP	Determining the likelihood of correct overflow and non-overflow conditions	-
F_1 score	Revealing the best trade-off between overflowing and not overflowing prediction by interpretation as a weighted average between PPV and TPR	[0,100]
CKR	Measuring the concordance between ACC, TPR, and TNR	[0,100]

ACC: ACcuracy of true Classification; CKR: Cohen's Kappa Rate; DP: Discriminant Power; F_1-score: Harmonic mean; MCC: Matthews Correlation Coefficient; PPV: Positive Predictive Value; TNR: True Negative Rate; TPR: Total Positive Rate

***Note**: Adapted from Piadeh et al (2023)*

These indicators are used to evaluate how well the method distinguishes two classes.

4 Case study

The performance of SPPs and ML flooding models are highly dependent on the characteristics of the rainfall and the catchment. Therefore, to seek answers for the research questions beyond the conjectures and abstract reasoning of hypothetical scenarios, data from a real-word scenario is used to test the methodology.

This section presents the selection of the real case study scenario, which serves as the key determinant for IMERG data sourcing, validation, and optimization, as well as a reference for the LSTM model's structure, parameters, and performance evaluation. The remaining sections of this chapter follows a data mining structure similar to the one presented in section 2.4.

4.1 Case study scenario selection

The UK was chosen as the regional area of the case study and three UK catchments as the focused area for flooding risk predictions. The selection is based on several attributes of the UK region:

- Vulnerability to floods: Flooding is one the most important climate change challenges in the UK (Speight and Krupska, 2021) and climate hazards are due to increase across the region with predictions of an increase in rainfall volumes and intensity in the UK of approximately 20% by 2085 (Ashley et al., 2008). Additionally, the UK Climate Projections 2009 (UKCP09) from the Met Office indicate that the intensity and severity of convective and frontal storms will likely rise (Murphy et al., 2010). Fluvial flooding is estimated to change significantly across the UK by 2070, with river flow increasing particularly at the west and south coast of England (Bates, et al. 2023).

- Flooding impact: The UK economy is significantly impacted by flooding events with some regions more affected than others. The annual financial impact estimated to be between 0.7 GBP (billions) and 2.25 GBP (Billions), depending on the source and method of calculation (Bates et al., 2023).
- Flooding strategy: The UK government approach on tackling fooding events is focused on enhancing urban resilience while meeting sustainable goals targets (Boyd, 2021), which is in line with proactive strategies for flood prevention and mitigation previously mentioned in this study.
- Regional challenges: Increasing the lead time of predictions is particularly challenging for the UK. The region receives most of its rainfall from the air masses originating in the Atlantic Ocean and travelling towards the west and southwest coast as shown in Figure 4.1.

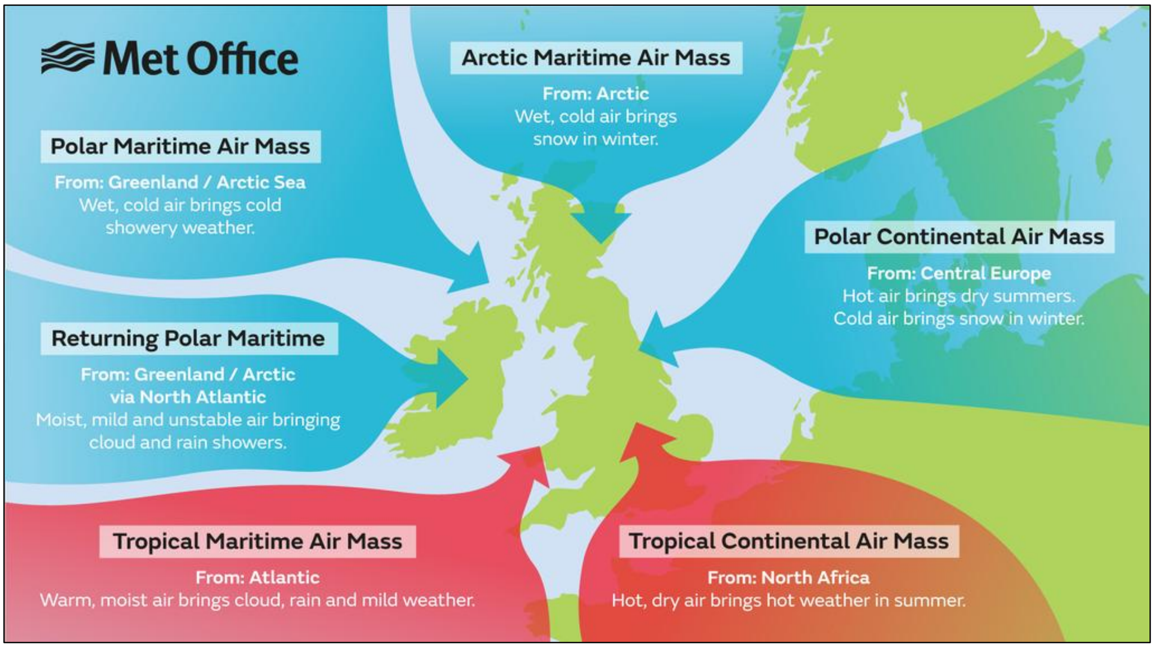


Figure 4-1 - Air masses affecting the British Isles. (Met Office, 2023a).

- Limitations: The detection of the weather systems from the west of the UK is limited by the max range of 250 Km radio of each instrument and reduced accuracy for increasing distance (see Figure 4.2). Additionally, for real time applications, analysis tools like the UK Climate Projections 2009 (UKCP09), are not capable of generating reliable future rainfall estimates for periods shorter than a day (Murphy et al., 2010).
- Data access: The UK has a large and well distributed gauging network that is able to provide consistent and continuous readings of rainfall and stream level data since at least 1961. The datasets are easily accessible online and crucial for the objectives of this research.

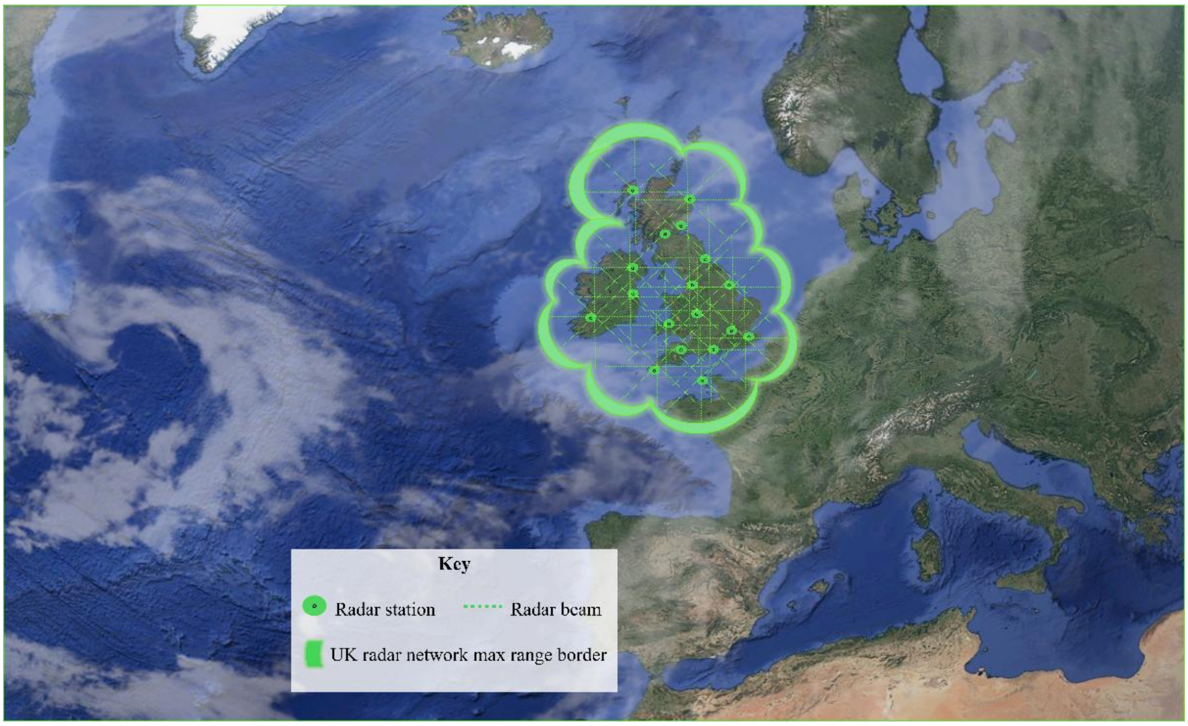


Figure 4-2 - UK weather radar network range. [Base map from ©Google maps - Map data@2024 Google]

These attributes make the UK an ideal candidate for the application of the methodology proposed in this study. The next section describes the process for selections of individual UK catchment for rainfall and stream level data collection.

4.1.1 Catchment and stream level stations selection

To ensure the applicability of this methodology is not restricted to a particular catchment or area of the UK, three catchments were selected. The catchments were chosen at different positions spread from South to North of England, to capture possible differences in weather scenarios. The choice of points of stream level data collection is extremely important to secure information that is relevant for this study and suited for modelling. Therefore, a criterion for stream level source selection was stablished considering the following aspects:

1. Flow regime: the selection should prioritize rivers with natural flow and responsive regime to avoid the interference in water level changes that are not related to rainfall, like water abstraction or recharge by reservoirs, industrial and agricultural activities, etc.
2. Data: preferable streams that can provide consistent and regular readings since 2000 to match the maximum period of data availability from the IMERG datasets.
3. Live data: is important that live data is available to secure the use of the dataset source in real-time applications.
4. Data quality: prioritize datasets with higher levels of completion and with short data gaps.
5. Corresponding rainfall data: there should be rain gauges within or at close distance of the catchment area, which are able to provide rainfall data matching the period and quality of the stream level data. These gauges are to be used as reference for analysis of the performance of SPPs.

6. Catchment size: the catchment area is preferable not too large to better capture the rainfall and stream-level relationship as smaller catchments are more sensitive to rainfall events. However, it should not be too small for better examination of relationships across a broader spatial and temporal scale, especially for rainfall datasets with limited spatial resolution (Ochoa-Rodriguez et al., 2015).
7. The north and the west of the UK face particularly higher risks of fluvial flooding (Arnell et al, 2021; Bates et al., 2023), while the London region faces the greatest average annual losses (AAL) from river flooding according with the Association of the British Insures (ABI) report (JBA Risk Management Ltd, 2021).

As a result of the criterion, three catchments were selected from the National River Flow Archives : C1 located near Heathrow airport in London, C2 located in Devon, Southwestern of England at West of Exeter, and C3 located at the Northwestern region of England, near Carlisle. Details of each catchment is provided in Table 4.1.

Table 4.1 - Main characteristics of the selected UK catchments.

Catchment	**C1**	**C2**	**C3**
River	Crane	Lew	Caldew
Station	Cranford Park (3660TH)	Gribbleford Bridge (SS50F007)	Cummersdale (NY394527)
Station location	(51.4893, -0.4164)	(50.7933, -4.0904)	(54.8656, -2.9445)
Max gauging level	1.23 m	2.07 m	3.09 m
Water level STD	113 mm	205 mm	236 mm

Operation period	Since 1974	Since 1988	Since 1997
Catchment area	61.74 Km^2	71.1 Km^2	244 Km^2
Average annual rainfall	639 mm	1192 mm	1211 mm
Live data available	Daily flow - 100% complete	Daily flow - 100% complete	Daily flow - 100% complete
Flow regime	Responsive	Responsive	Very responsive
Runoff	Natural to within 10% at the 95-percentile flow	Natural to within 10% at the 95-percentile flow	Natural to within 10% at the 95-percentile flow
Land cover: **Urban**	68.02%	1.70%	1.01%
Woodland	4.71%	15.39%	7.90%
arable/horticultural	3.27%	21.34%	8.49%
Grassland	23.92%	61.44%	77.46%
Mountain	0.00%	0.05%	5.07%

To visualize the distribution of the selected catchments in the UK, Figure 4.3 outlines their approximate locations, detailed boundaries, and the position of the stream level station in each catchment.

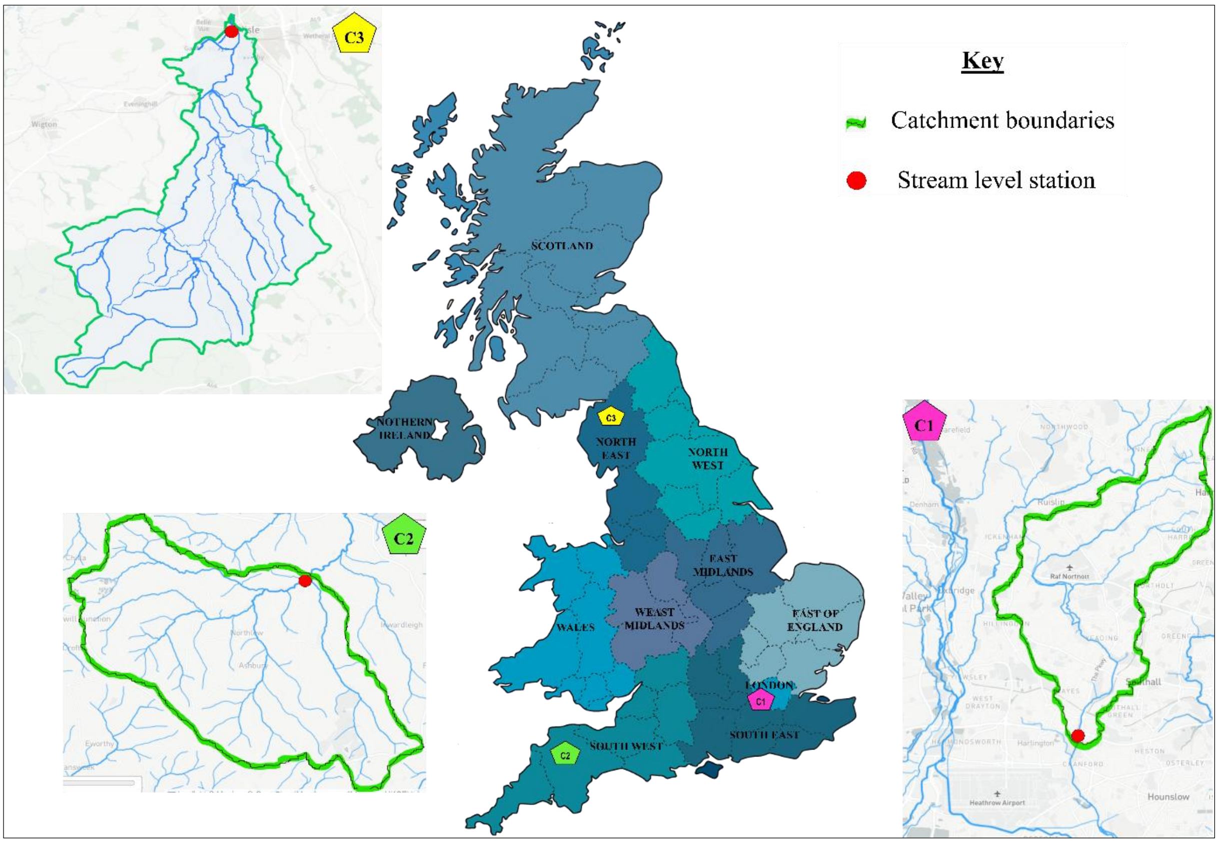


Figure 4-3 – Approximate location and detailed boundaries of the three selected catchments (C1, C2, C3). Adapted from EA (2023a)

Following the definition of the catchments, the stream level historical time series datasets from each station was retrieved from the online data service platform of the Department of Environment Food & Rural Affairs (DEFRA, 2023) accessed online at: https://environment.data.gov.uk/hydrology/explore

4.1.2 Rain-gauges selection

The selection of ground gauge stations providing the historical rainfall datasets for the study also followed a careful process. First, the rain-gauges closest to the station were considered due to their ability to best capture the rain event's characteristics within the catchment. Next, the rain-gauges were chosen to provide data matching the period of coverage, consistency and reliability required from the stream level data.

Following this process, the Northolt RAF - 247344TP (51.5527, -0.41690) rain-gauge was selected. The gauge is located close to the centre of catchment C1, approximately 7Km to the North of Cranford Park station, as shown in Figure 4.4. Along with Northolt, two other nearby gauges were included for complementary sourcing of rain data: Chertsey (51.3979, 0.5428) and Camberley (51.3285, 0.7667), both situated south-west of Cranford station and approximately 13.5Km and 30 Km respectively. The data from the extra gauges is used as other sources of references for the comparison and assessment of the IMERG data.

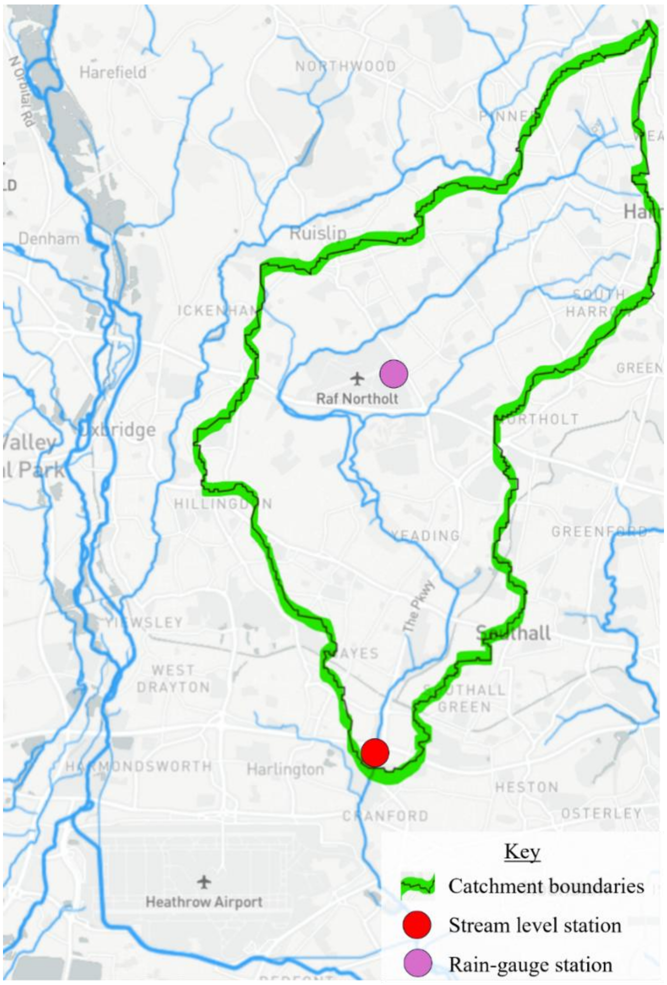
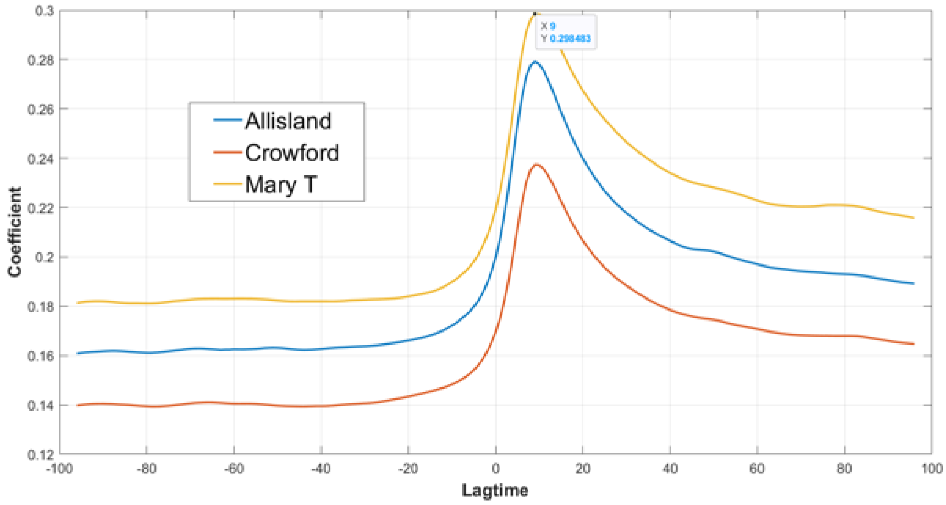


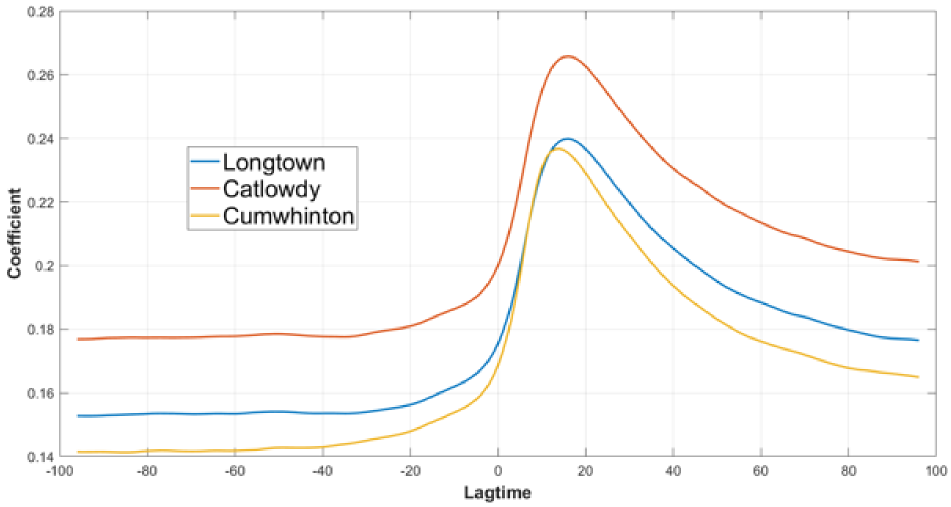
Figure 4-4 - C1 rain gauge location - Northolt RAF station. Adapted from EA (2023a).

The process for rain-gauge selection in catchment C1 was repeated in the other two catchments. However, it did not return any adequate gauging source within the C2 and C3 boundaries, which incurred on the selection of other suitable rain-gauges at the vicinities of each the catchments. For catchment C2 the rain-gauge stations at Allisland (50.86838, -4.15306), Crowford Bridge (50.75470, -4.42887) and Mary Tavy (50.56928, -4.10349) were pre-selected; and for catchment C2 the rain-gauges stations at Cumwhinton (54.8678, -2.83764), Longtown (55.0059, -3.0002) and Catlowdy (55.0829, -2.8484). Similar to stream level datasets, the historical rain-gauge time series were retrieved from the DEFRA (DEFRA, 2023) website available through the following link: https://environment.data.gov.uk/hydrology/explore

To minimize the amount of data processing, for catchments C2 and C3 the most relevant of the rain-gauge source was determine using an addition extra selective step that considered the historical relationship between the stream level and rain-gauge datasets. These relationships were evaluated based on the cross-correlation coefficient (CCC), which measures the strength and direction of the relationship between the time series, and the lag values representing the time period between when rainfall is detected by the gauge and when it affects the stream level.



(a)



(b)

Figure 4-5 - Cross-correlation between rain-gauges data and stream level at the catchments C2 (a) and C3 (b)

As a result of the cross-correlation analysis, the Mary Tavy rain-gauge station was chosen as the ground source of rainfall data for C2, due to better historical data correlation with the level at Gribbleford Bridge (see Figure 4.5a); and the Catlowdy gauge for C3, due to better historical data correlation with the level at Cummersdale (see Figure 4.5b).

4.2 Satellite-based precipitation source

The selection of the IMERG products as the source of SPPs data in this study was based on properties of the data (global coverage, long period of availability, fine spatial and temporal resolution, and easy access), and on the good and consistent performance of the products in flooding related applications.

As suggested by previous studies (Jiang and Bauer-Gottwein, 2019; Wang et al., 2017), and considering its higher accuracy, the IMERG Final Run was the version of the product used in this study to assess the data suitability, and for the training and testing of the deep learning models. While, for its lower latency, the IMERG Early Run was selected for testing the model's accuracy in real-time predictions. Both products, Early and Final Run, provide estimates since June 2000 at intervals of 30 minutes and spatial resolution of 0.1^{o} x 0.1^{o}, or approximately 10Km x 10Km. The time series data is free and available, in ASCII or netCDF format, for online download from the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

Additionally, is important to mention that this research started at the end of 2021, when V06 was the latest IMERG version available, approaching the end of the research, in April 2024, the IMERG V07 was launched. Therefore, the whole methodology was initially developed with the IMERG V06 and, at the later stages of the research, replaced with the IMERG V07. The data analysis and calculations repeated, and the entire methodology adapted for the new IMERG V07 data. Nonetheless, the IMERG V06 dataset still played a part on the evaluation process considered for the assessment of the IMERG data suitability that will be presented in the results section.

4.2.1 IMERG subset

The global coverage of the IMERG dataset offers a vast amount of gridded data, most not relevant or useful for use in the prediction event-based flooding risks for a specific catchment. Therefore, only part of global grid area was selected for monitoring and collection of SPPs data.

The delimitation of the area for data collection considered the desired lead time on predictions which requires the events to be detected early while still at great spatial distances from the catchment. Based on this requirement, the area's boundaries were determined to allow for predictions with lead-times of at least 4hs after accounting for the latency of the dataset and the characteristics of extreme weather events in the region.

According to the Met Office (2023b) extreme events typically move at forward speeds of 30 Km/h., but can reach up to 65 Km/h. However, is known that slow-moving storms have more time to accumulate and redistribute water (Camus et al., 2024). Therefore, this study focuses on a data collection area where the distance between the catchment and the point of detection allows predictions to be made with a lead time of at least 4 hours, and up to 12 hours, in the case of a weather event moving at a top forward speed of 60 km/h along its entire trajectory from the Atlantic to the catchment.

Consider the following equation for the calculation of the minimum distance between rainfall event and catchment to allow for different target lead times:

$$\mathbf{D = (d + T) * efs} \qquad (4.2.1)$$

D = distance between rainfall event and catchment

d = data delay or latency (4hs for IMERG Early Run)

T = target lead time

efs = rainfall event forward speed towards catchment

Table 4.2 - Calculation of minimum distance (D) for detection of rainfall event to allow for predictions with IMERG data with certain lead time

T	D	Efs	D
4hs	4hs	60 Km/h	480 Km
6hs	4hs	60 Km/h	600 Km
12hs	**4hs**	**60 Km/h**	**960 Km**
4hs	**4hs**	**30 Km/h**	**240 Km**
6hs	4hs	30 Km/h	300 Km
12hs	4hs	30 Km/h	480 Km

Based on the calculations in Table 4.2, the monitoring region for IMERG data collection over the Atlantic extends from 5°W to 21°W and 45°N to 61°N, as displayed in Figure 4.6. This region extends to at least 1200 Km to the west, 600 Km to the North, and 600 Km to the South of each catchment location. Data is also collected from the unit of the IMERG grid corresponding to the location of Cranford Park station, identified as centred at coordinates 51.45°N - 0.45°W and from now on called IMERG-Cranford pixel. This pixel's data is used in the assessment of the IMERG products against precipitation data from local rain-gauges.

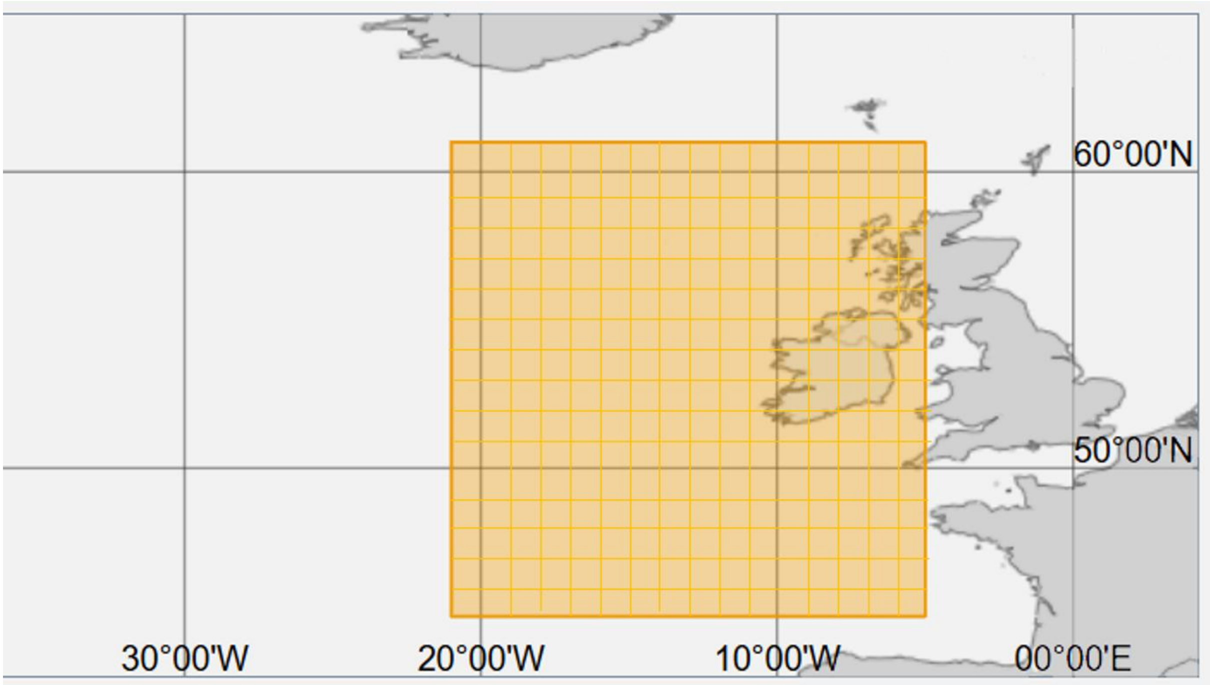


Figure 4-6 – Delimitation of monitoring region selected for the IMERG data collection - within the orange rectangle.

The historical IMERG precipitation time series corresponding to the monitoring region of the data grid and to the IMERG-Cranford pixel unit, were accessed online from the Nasa GES DISC archives (NASA GES DISC, 2023) and downloaded in netCDF format following the instructions provided in the data portal at: https://disc.gsfc.nasa.gov

The IMERG data is provided in NetCDF format with a specific URL link for each time step. The downloaded files contain the precipitation estimates at each time step for each unit of the IMERG data grid which is identified by the respective longitude (Lon), latitude (Lat) coordinates.

However, download from the GES DISC archives was a time-consuming task taking between 1 to 5 seconds for each URL (each time step) downloaded. Additionally, the download process, which took about 96 hours, randomly skipped about 50% of the total of 411,985 URLs in the first attempts. To address this issue a code was developed to identify which of the URLs were successfully downloaded and exclude it from the original list updating it with only non-downloaded URLs before the next download, reducing the downloading time by half at each attempt. In total, accounting for eventual breaks in between downloads, time to run the code to exclude URLs and to organize the new list, the entire process took over around a full week to complete.

Further information on the processing of IMERG and gauge-based datasets will be disclosed along the data inspection detailed in the next section.

4.2.2 Datasets inspection and pre-processing

Prior to any numerical analysis the datasets were inspected focusing on the aspects of quality, consistency, and properties of the data to inform the preparation process and guide the choice of modelling techniques. For reference, a summary of the datasets collected and used in this study is provided in Table 4.3.

Table 4.3 - Summary of properties of each the datasets downloaded.

Data (Source)	Period of data	Unit	Data interval	Timeseries steps	Data format
Precipitation (Rain-gauges)	01.06.2000 to 30.11.2023	mm/15min	15min	823,968	CSV (punctual)
Water level (River gauges)	01.06.2000 to 30.11.2023	Meter	15min	823,968	CSV (punctual)
Precipitation (IMERG V06 Final Run)	01.06.2000 to 30.09.2021	mm/hour	30min	364,016	netCDF (Grid)
Precipitation (IMERG V07 Final Run)	01.06.2000 to 30.11.2023	mm/hour	30min	411,984	netCDF (Grid)
Precipitation (IMERG V07 Early Run)	01.06.2022 to 30.11.2023	mm/hour	30min	26,304	netCDF (Grid)

Starting by the ground-based data sources, the stream level time series were examined for missing values, null values, and outliers. Missing and null values were found in rare occasions and short periods, usually less than one missing day of observations. The missing values in the dataset were addressed using linear interpolation, a common approach for filling short gaps in continuous datasets (Zhang & Thorburn, 2022). A singular and significant outlier was detected in the dataset from Cranford Park station. On 06/07/2021, the station readings showed water levels varying between 1.263m and 2.643m, whereas the station's historical record is 1.232m.

To correct the outlier, data was sourced from the River Crane's downstream station, at Marsh Farm, located just 5 km from Cranford. From the water level records at each of these stations, shown in Figure 4.7, is noted that the offset between both timeseries kept constant along the historical data. This stablishes a reliable relationship between the timeseries which allows for readings in Cranford to be roughly estimated based on the readings in Marsh Farm. Therefore, the outliers in the Cranford dataset on 06/07/2021 were adjusted based on the respective readings from Marsh Farm considering the average offset between both stations observed in the previous day.

Similarly, for periods ‘of missing data or significant outliers in gauges’ datasets, the gaps was filled with the imputation of values based in the readings registered by closest available rain-gauge data.

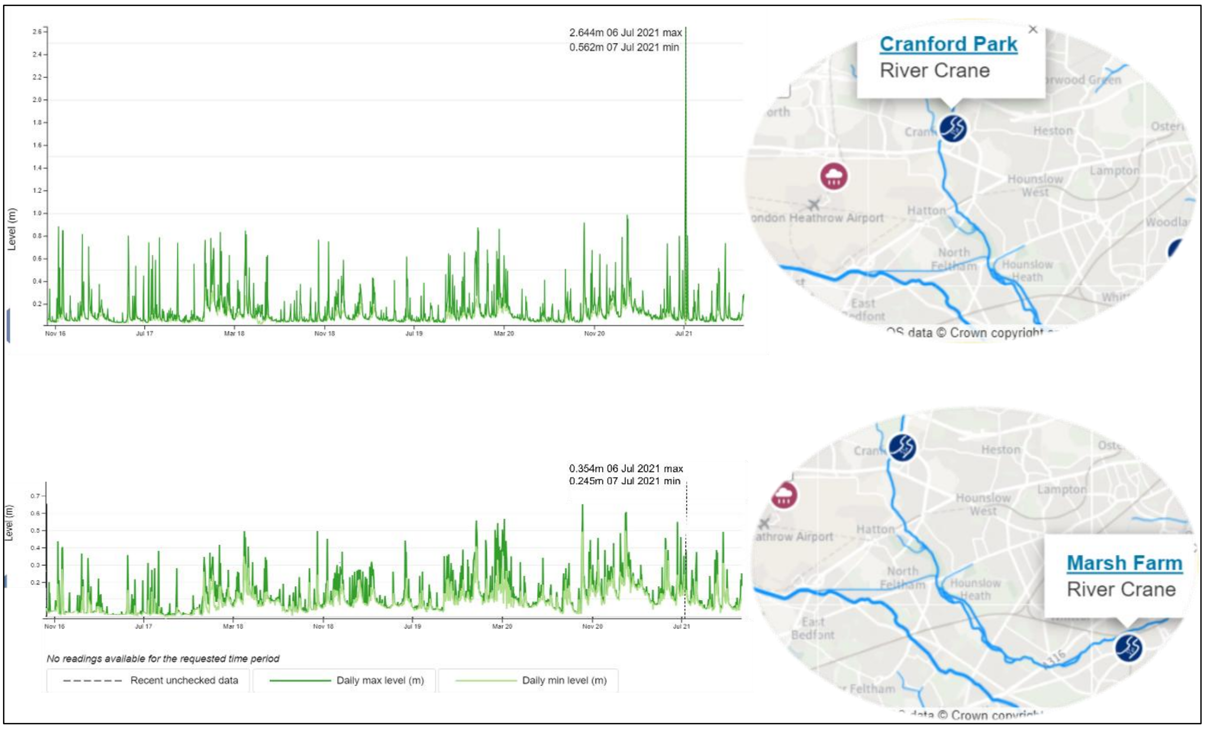


Figure 4-7 – Plot of water level time series from two stations of the River Crane, Cranford (upstream) and Marsh Farm (downstream). Adapted from EA (2023a).

Next, the inspection of the satellite-based datasets commenced with the visualization of the gridded data from the netCDF files, plotted using Python coding in the Visual Studio (VS) Code editor. Visualization is the most useful method for initial assessment of the quality of the IMERG dataset which is provided in a gridded format with one grid for each time step and precipitation values in each grid cell.

At close inspection, several plots of individual time-steps revealed images with inconsistencies, areas of distortion or random broken patterns. Figure 4.8 shows a sample of the patterns observed at the same time step in plots with the data from the IMERG V06 and from the IMERG V07. From the comparison of both, the latest version the IMERG products seems to have considerable improvements, particularly on the areas of missing data identified in a couple of the IMERG V06 images.

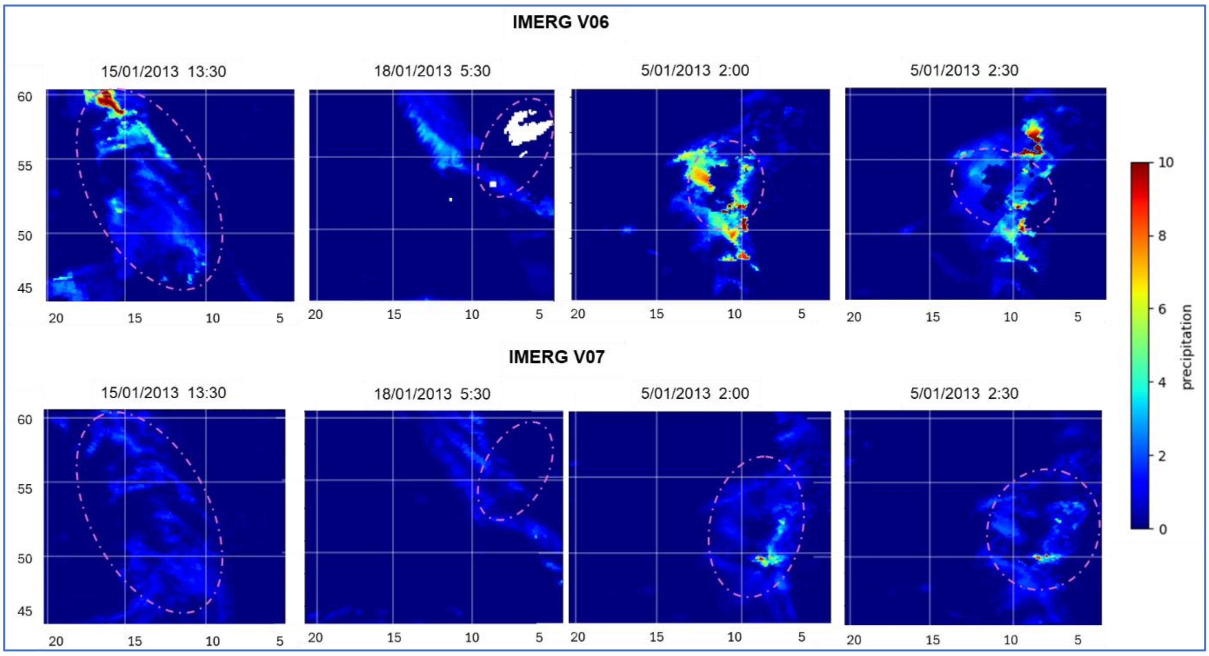


Figure 4-8 - Sample of the plots of netCDF files from the IMERG V06 and IMERG V07 with the areas of anomalies or discrepancy marked within the dotted lines.

Nonetheless, anomalous areas still occur in the IMERG V07 product as noticeable in the sample of images in Figure 4.9.

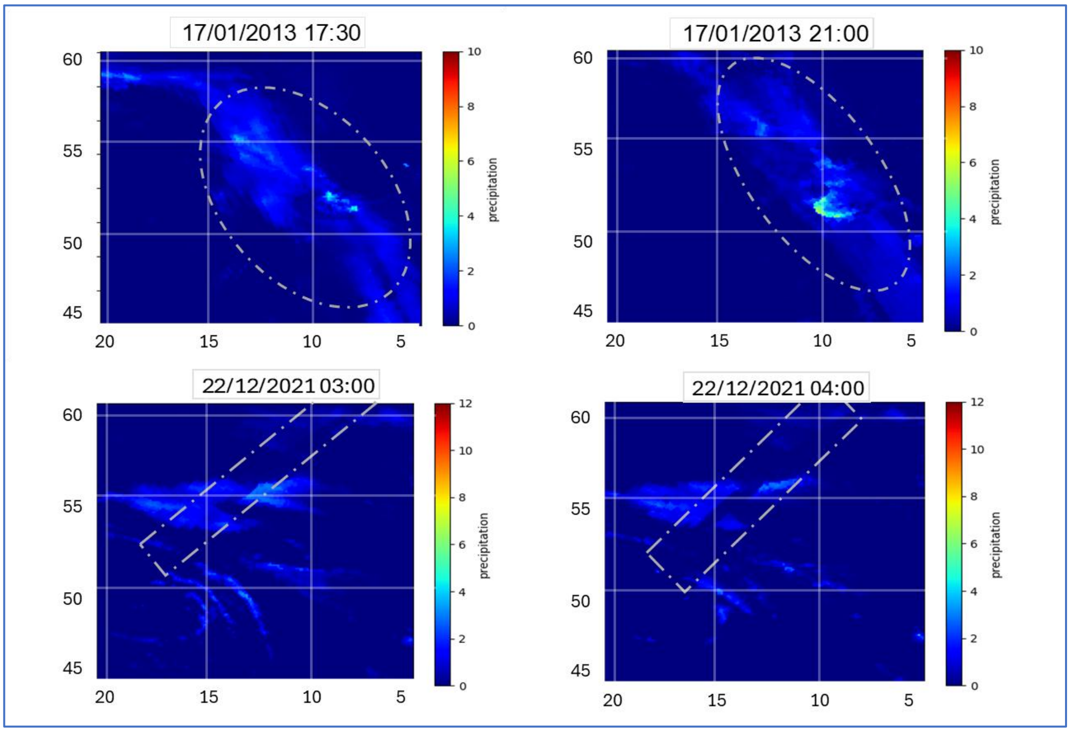


Figure 4-9 - Sample of plots of IMERG V07 Final Run precipitation data with areas of discrepancy marked within the dashed lines.

To examine the cumulative effect of these anomalies, another code was created for plotting monthly and yearly precipitation totals from the IMERG data . A sample of these plots is provided in Figure 4.10, illustrating how inconsistencies observed at individual time steps can compound into more significant issues when data is accumulated over time. These problems are especially evident in the top images of Figure 4.10, retrieved from IMERG V06, and partially resolved in the IMERG V07 (see images at the bottom) where issues like white patches with missing data seem to have been eliminated.

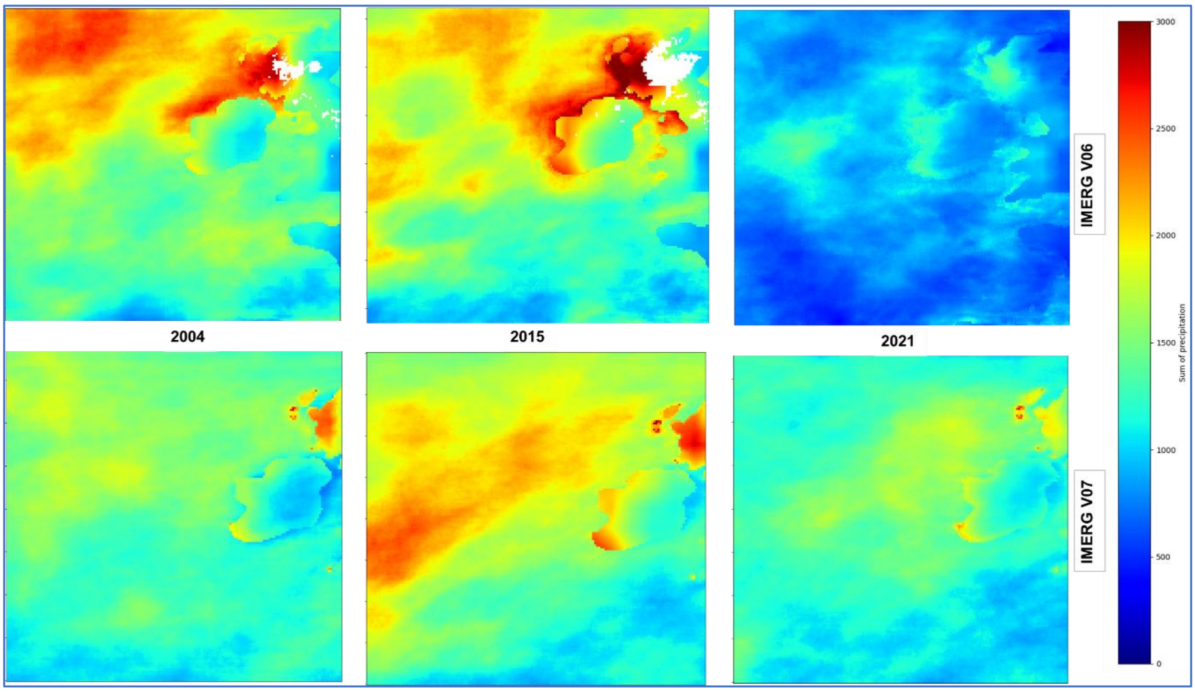


Figure 4-10 - Annual cumulative precipitation in the years 2004, 2015 and 2021. Top half of the figure, plots from IMERG V06 data and at the bottom half from IMERG V07 data.

However, the patterns in the cumulative precipitation data, observed at the top right area of the images, remains as a significant issue in the IMERG V07. The images show sudden changes in the volumes of annually accumulated rainfall outlined by sharp, well-defined borders, where more smooth transitions would be expected. This suggests that, despite noticeable improvements, the issues observed in the IMERG V06 dataset persist to some extent in the IMERG V07 version.

To rule out that the issues are due to potential errors in the code or data processing, plots were also generated using NASA's Giovanni platform—a tool designed for searching and visualizing Earth science data. Figure 4.11 shows both versions of the plot displaying cumulative precipitation for the year 2013, based on the IMERG V07 data. Even though each version uses a different colour scheme, the similarity in patterns is evident, indicating that the anomalies are not the result of coding errors or data processing issues.

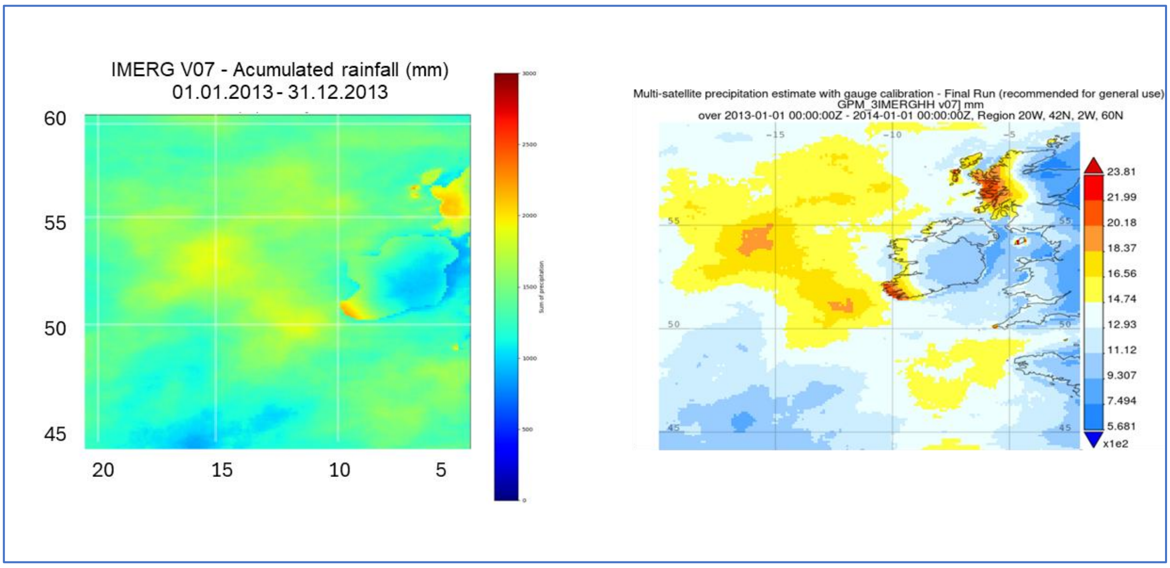


Figure 4.11 - Plots of cumulative precipitation in 2013 from IMERG V07. At the left plots generated using python and VS code, at the right plots produce in the Giovanni platform.

This allows for further investigation into the quality of the data by compiling plots of each time step in 2013 into a mp4 video portraying how rainfall events are formed, evolve and move across the monitoring region. Two samples of the videos produced with the IMERG data are available online, in the You Tube platform, and can be accessed through the links provided in the Appendix B at the end of this thesis. Similarly to the image analysis in Figure 4.11, one video sample was produced with Python in VS Code and the other with the Giovanni online tool. The videos were used to verify data consistency and the correspondence of the SPPs estimations weather with the patterns expected in real weather systems. Both videos corroborate anomalous behaviour in the data, occasionally accompanied by 'flashing' effects in the images. The flashing effect appears when the rainfall system moving across the study area suddenly intensifies, transitioning from blue or green to yellow or red within a few time steps, before abruptly returning to normal levels (light or dark blue).

Based on the plots and video analysis, the most significant and consistent anomalies observed are confined to specific areas within the monitoring region — namely, along the coast of the island of Ireland and the northwestern coast of Great Britain, as shown in Figure 4.12.

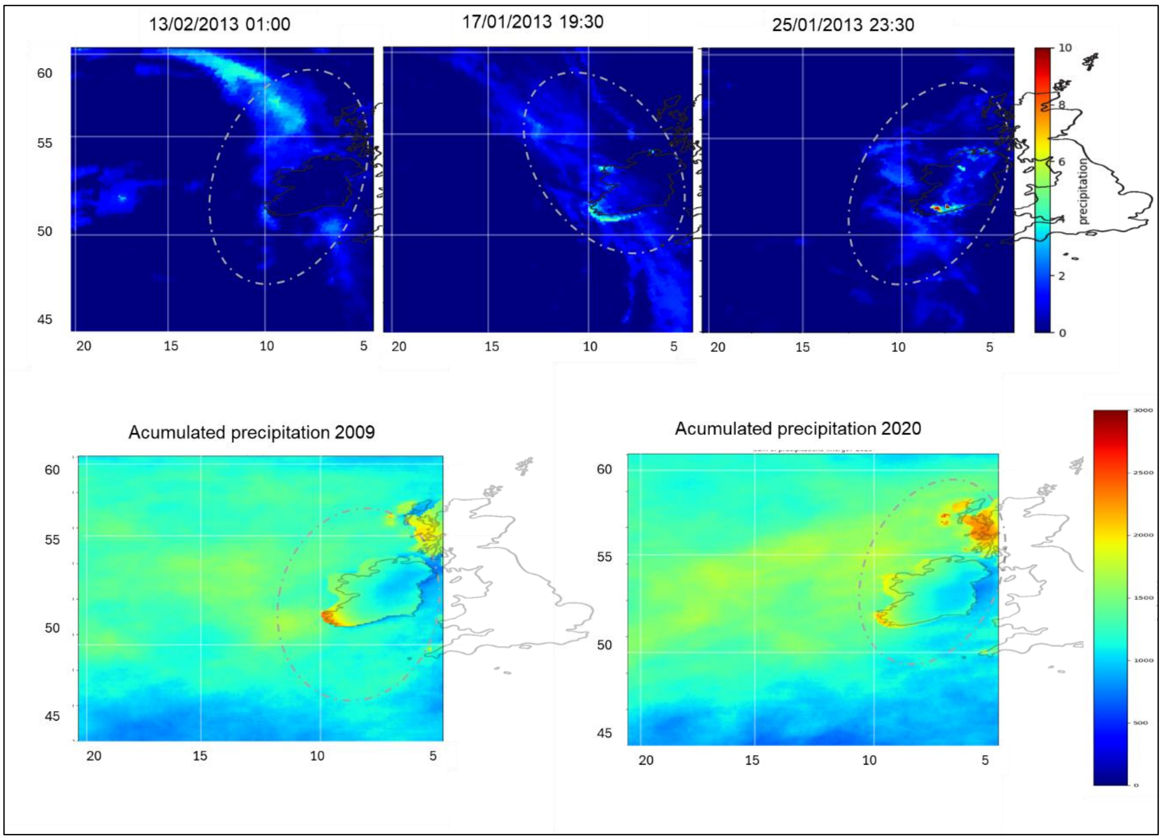


Figure 4-12 - Plots of IMERG V07 precipitation data for individual time-steps and cumulative amounts in 2009 and 2020.

Although these anomalies were first identified in the images from IMERG V06, available since 2019, no official publications directly address this particular issue. It is known that the IMERG GPM Microwave Imager (GMI), or IMERG-GMI, exhibits a high bias of +50% and a low bias of −41% over high-latitude regions and tropical oceans (Watters et al., 2025). It is also known that the GPCC gauge analyses do not extend over oceanic regions, which may lead to transition issues between areas validated and non-validated by gauge data within the IMERG dataset. While both factors could potentially contribute to the anomalies observed, their exact impact cannot be estimated from the information available. Nevertheless, although this issue limits the use of the IMERG data near the coasts of Ireland and the UK, it does not necessarily undermine the use of the data over the oceanic area, which is the primary focus of this study.

On the other hand, more random issues observed in the IMERG data, such as the broken and distorted patterns in the images (Figure 4.9) and the flashing effects in the videos (Appendix B), may be attributed to the algorithm's use of multiple data sources. As explained by Huffman et al. (2023), due to the successive inclusion of infrared (IR) data from various geo-IR satellites in the IMERG algorithm, effects like these occurs when data collected by a satellite with a direct overhead view is replaced by high-zenith-angle data from an adjacent geo-IR satellite. As such, these effects and random patterns can interfere or mislead the learning process of the deep learning models, making the use of the IMERG datasets in image format less suitable for machine learning modelling.

Therefore, the netCDF data was processed using another method of spatial time series analysis, often used remote sense, where the same statistical tool is applied to each of the IMERG pixels separately (Ozdogan, 2017). Pixels, which are the smaller spatial unit in the IMERG gridded data with (0.1°x 0.1°) dimension, are individually identified by its spatial coordinates (Lon, Lat). The structure of the IMERG data in the monitoring region is displayed in Figure 4.13 with detailed information of the pixel's distribution and data volumes.

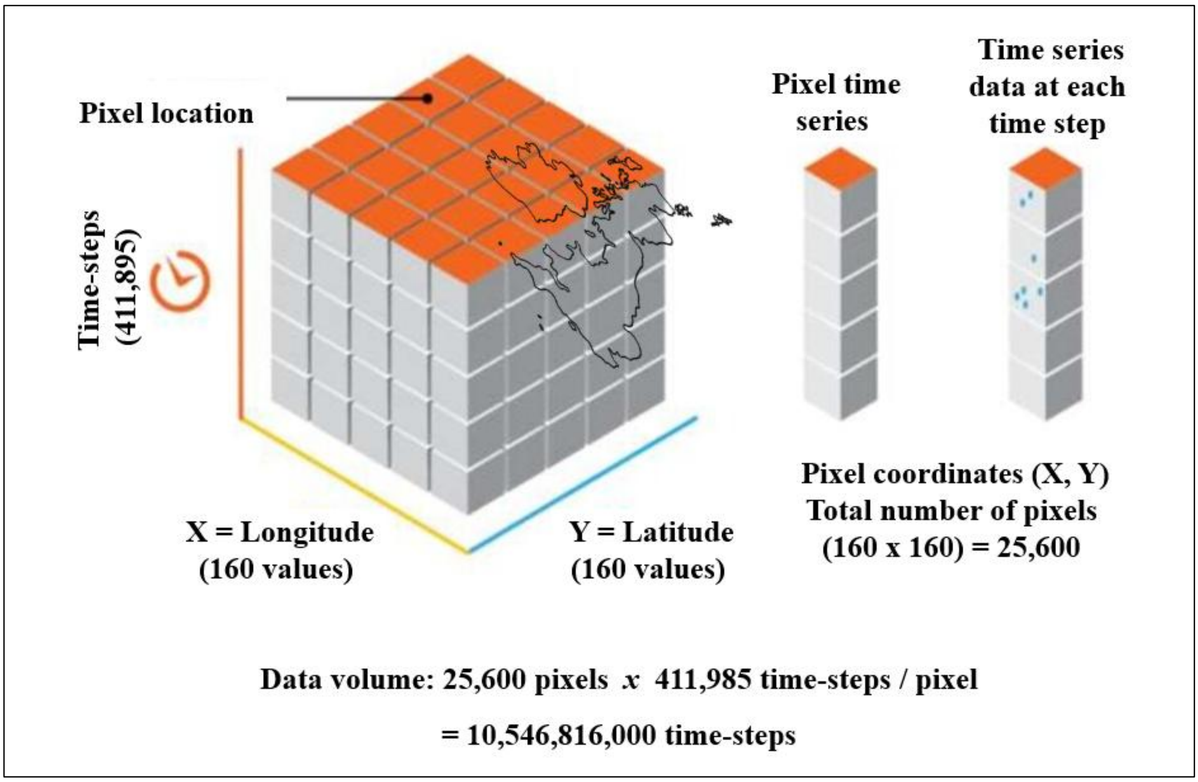


Figure 4-13 - Space and time representation of the IMERG data volume and pixels distribution in the monitoring region. Adapted from Esri 2016.

Given the extensive monitoring area and the long period of observations, the total amount of IMERG data for processing exceeds 10.5 billion data points. Which suggests for the IMERG data to be assess locally before advancing towards deeper shores and dedicating resources to process such large amount of data.

4.2.3 Comparison between the properties of the IMERG and rain-gauge datasets for stream level forecasts

The IMERG data from the pixel closer to the stream level station is expected to best represents the rainfall over the catchment. As such, the data is used for comparison with gauges data to assess its suitability for stream level predictions. From this point in the study, all analyses and calculations are based solely on the IMERG V07 dataset.

The following steps were applied to uniformize the datasets for effective comparison:

1. A python code was created to extract the time series of precipitation values in each pixel of the IMERG grid converting it from NetCDF to CSV format. The precipitation values were organized in an Excel table where each column represents a pixel unit, identified by its Lon and Lat coordinates, and each row represents one time step.
2. The water level and rain gauges time series datasets, initially provided within 15 min intervals, were upscaled to 30min intervals to match the IMERG data temporal resolution. For rainfall, the readings at the previous 15 min are added to the next value for upscaling, while for water level, consecutive values of 15 min readings were averaged for upscaling.
3. The IMERG estimates were converted from intensive units (mm/h) to extensive units (mm/30min) to match the units of the upscaled rain-gauge data.

The processes of upscaling and handling missing or null values prepared the datasets for the cross-correlation analysis between the rainfall time series from each data source and stream levels time series for each catchment. The cross-correlation coefficient (CCC) and corresponding lags were calculated using the Matlab (2022) software. The outcomes are presented in Table 4.4 as reference of the values used along the decision process defining the next steps of the methodology.

Table 4.4 - Cross-correlation coefficients (CCC) and lag between each source of rainfall data and stream level in the catchments

Catchment	Rain data source	CCC	Lag (time-steps)
C1 (Cranford Park)	Northolt RAF – gauge	0.3645	19
	Chertsey – gauge	0.3463	20
	Camberley – gauge	0.3412	20
	Cranford pixel – IMERG	0.3597	21
C2 (Gribbleford Bridge)	Allisland – gauge	0.2792	9
	Crowford- gauge	0.2374	9
	Mary T – gauge	0.2985	9
	Gribbleford pixel – IMERG	0.2542	9
C3 (Cummersdale)	Longtown – gauge	0.2245	12
	Catlowdy- gauge	0.2490	10
	Cumwhinton- gauge	0.2247	10
	Cummersdale pixel – IMERG	0.2545	12

At glance, the CCC and lag times values for the IMERG pixels over the station area closely aligned with the values for the rain-gauge data in all three catchments.

4.3 Optimization of the IMERG dataset for stream level predictions

The consistent alignment between the IMERG data and gauge-based rainfall across different scenarios is promising for the continued use of IMERG in this research. However, the analysis so far only covers a single pixel out of the 25,600 pixels in the dataset. The vast volume of IMERG data and the anomalies observed along the coastlines, suggests for the discarding of non-relevant or anomalous areas of data for improving the speed and efficiency of the data processing.

Yet, given that the origins of the anomalies have not been identified, there is insufficient information to stablish an efficient criterion for the disposal of data. As a result, the focus shifts towards developing a method to detect pixel areas in the IMERG grid that provide the most relevant data for predicting flood risk in the catchments. This is done following a similar approach to that of the previous section, and evaluating data suitability based on the strength of historical relationships between datasets, as indicated by the CCC and lag values from cross-correlation analysis.

Therefore, the CCC and corresponding lags were calculated between the time series of precipitation estimates from each IMERG pixel and the stream level readings in catchments C1 and C3. These two catchments were selected to reduce the volume of data processing while maintaining a diverse set of catchments that vary in size, land cover, distance from the coast, and geographical location within the UK.

Figure 4.14 replicates the 25,600 cells of the selected IMERG grid area in an Excel table, where each cell represents a grid unit and displays the CCC value between the stream level at the River Crane in Cranford (C1) and the time series of the corresponding IMERG pixel. For reference, across the region the max CCC value is 0.3438 (lag 18.5hs), the min CCC value is 0.1758 (lag 30hs) and the average CCC value is 0.2187. The strength of the correspondence between the IMERG and the stream level timeseries is represented by the colour code attached to the values.

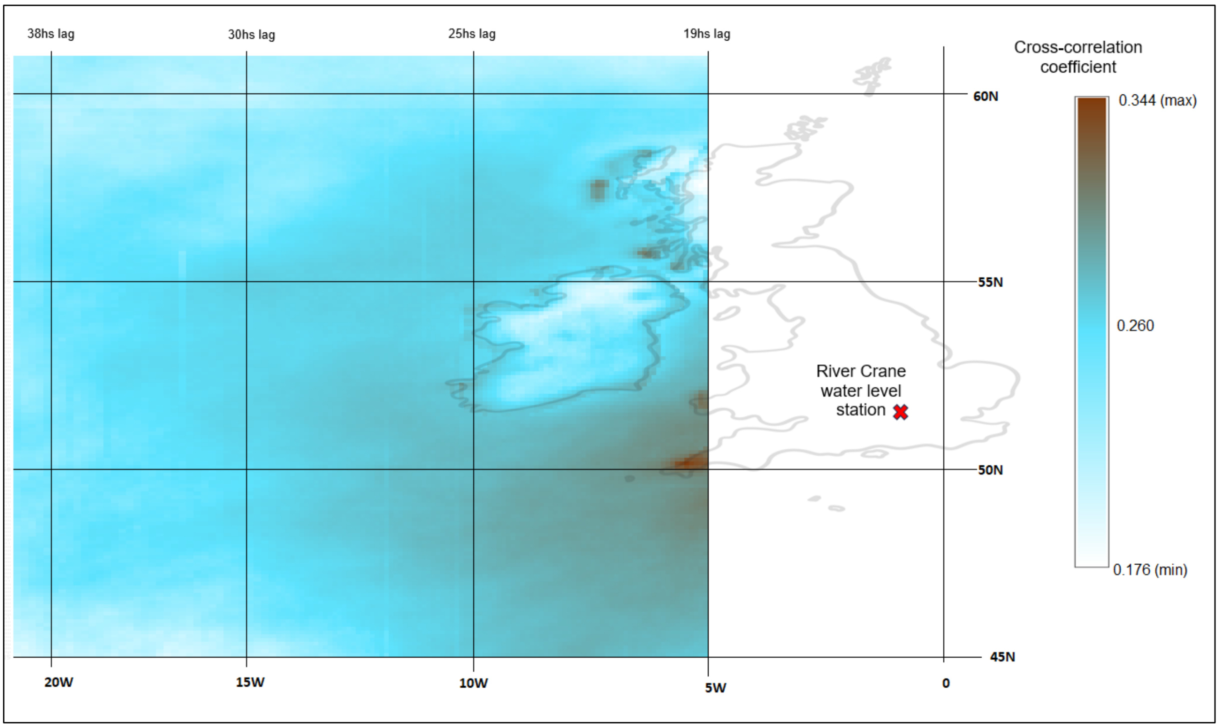


Figure 4-14 – Color-coded representation of the CCC values between the timeseries rain data at each IMEG pixel location and historical water level readings at Cranford (C1)

The process is repeated with the C3 data from the river Caldew at Cummersdale station, these outcomes are displayed in Figure 4.15. For C3 the max CCC value is 0.3849 (lag 14.5hs), the min CCC value is 0.1923 (lag 18hs) and the average CCC value is 0.2508.

The figure shows a delineation of patterns similar to those observed in the images of cumulative yearly precipitation extracted from the netCDF files (see Figures 4.10 and 4.11), suggesting that areas along the North coast of England and over Ireland should also be avoided when using the IMERG data in CSV format.

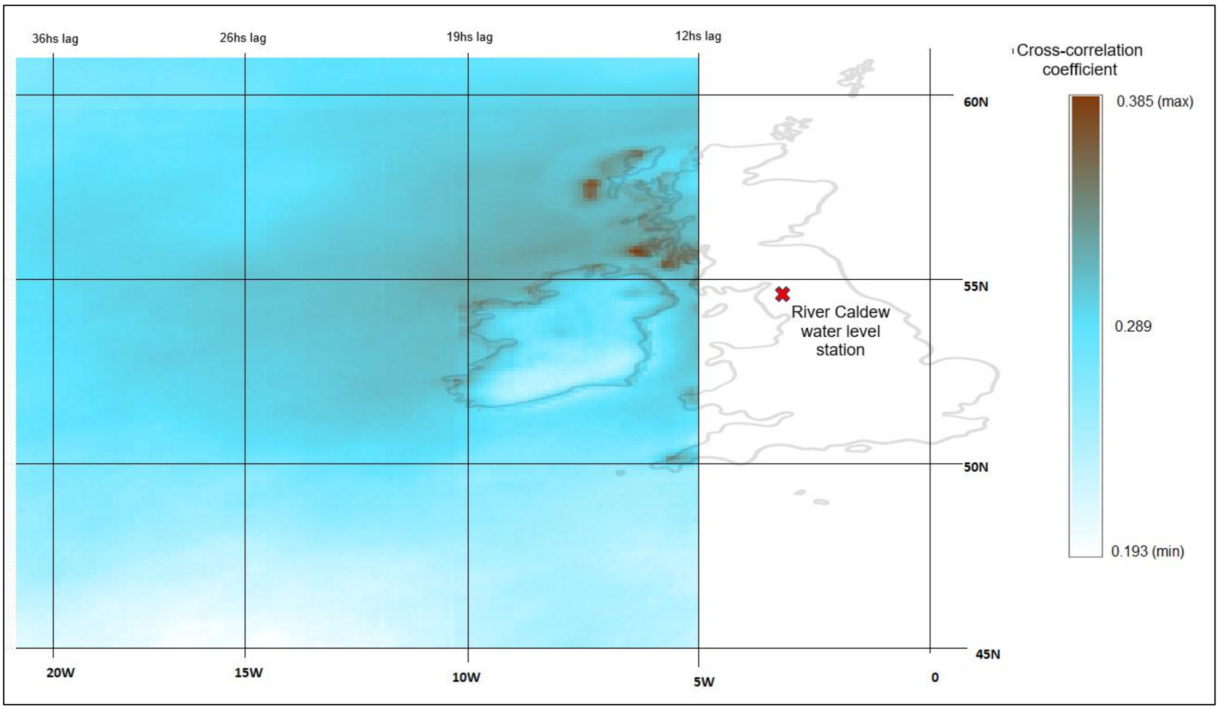


Figure 4-15- Color-coded representation of the CCC values between the timeseries rain data at each IMERG pixel location and historical water level readings at Cummersdale (C3).

Further analysis of the cross-correlation results focused on the distribution of the top CCC values along longitudinal lines to identify key areas for data collection as the distance from the catchments increased. This was done by selecting the top 10% of CCC values among pixels sharing the same longitude coordinate—returning the 16 most highly correlated pixels at each longitudinal line.

As shown in Figure 4.16, for C1 the highest CCC values are in dispersed clusters near the western border of the monitoring region, converging to a more cohesive arrangement as it approaches the catchment. These values range from 0.34379 (pixel 50.15°N, 5.45°W) at the south coast of England to 0.206866 (pixel 56.25°N, 20.95°W) at the western boundary of the monitored region), at more than 1,400 km from the catchment.

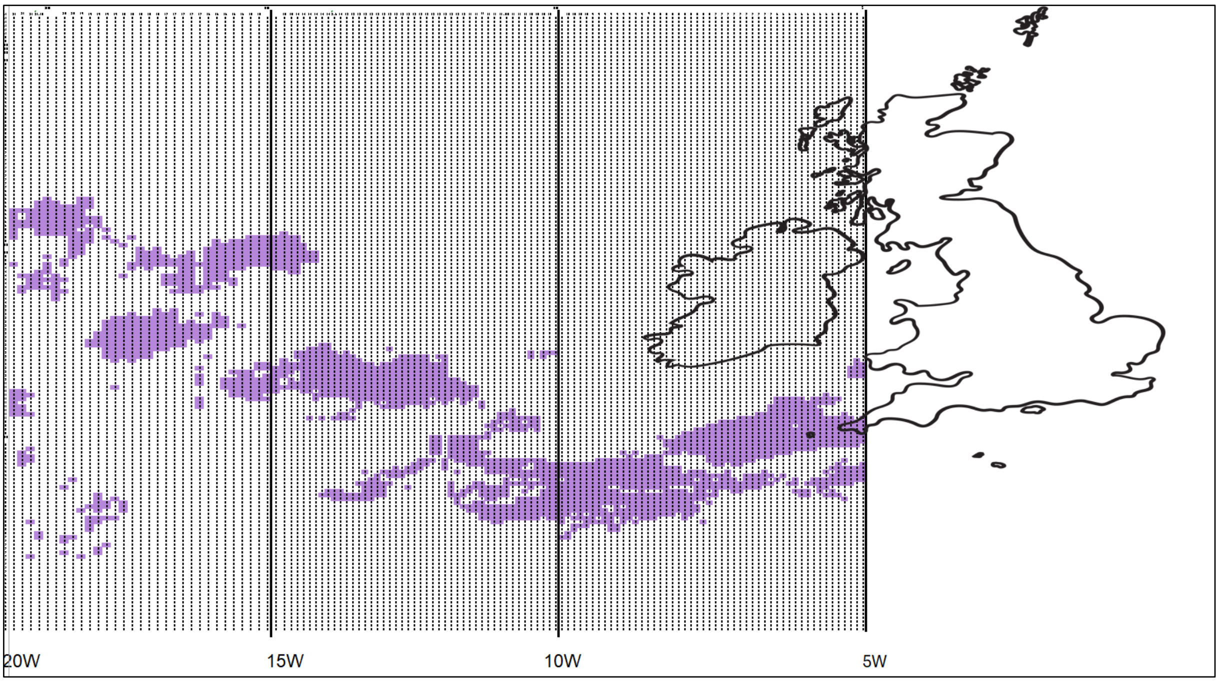


Figure 4-16 - Top 10% CCC values at increasing distances from the catchment C1.

For catchment C3, the path of influential rainfall is well defined across the monitoring region (see Figure 4.17), and the correlation with the catchment remains relatively strong with values up to 0.2564 (pixel 56.25°N, 20.95°W). The findings, reproduced in Figures 4.16 and 4.17, helped trace the pathways across the data grid where rainfall events strongly or frequently influenced changes in stream water levels at the C1 and C3, respectively.

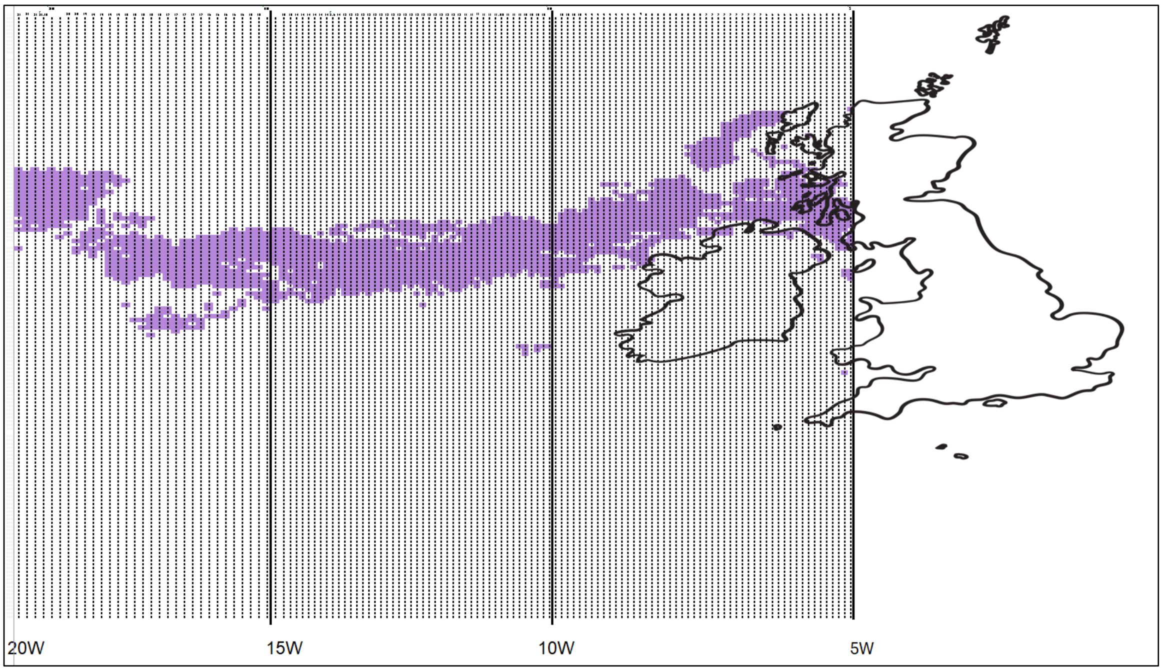


Figure 4-17 - Top 10% CCC values at increasing distances from the catchment C3.

However, a clearer understanding of the areas of precipitation data most relevant to flood risk predictions in the catchments was achieved by analysing the mean and standard deviation of the cross-correlation coefficients (CCC) across the IMERG data grid, as illustrated in Figures 4.18 and 4.19 for catchments C1 and C3, respectively.

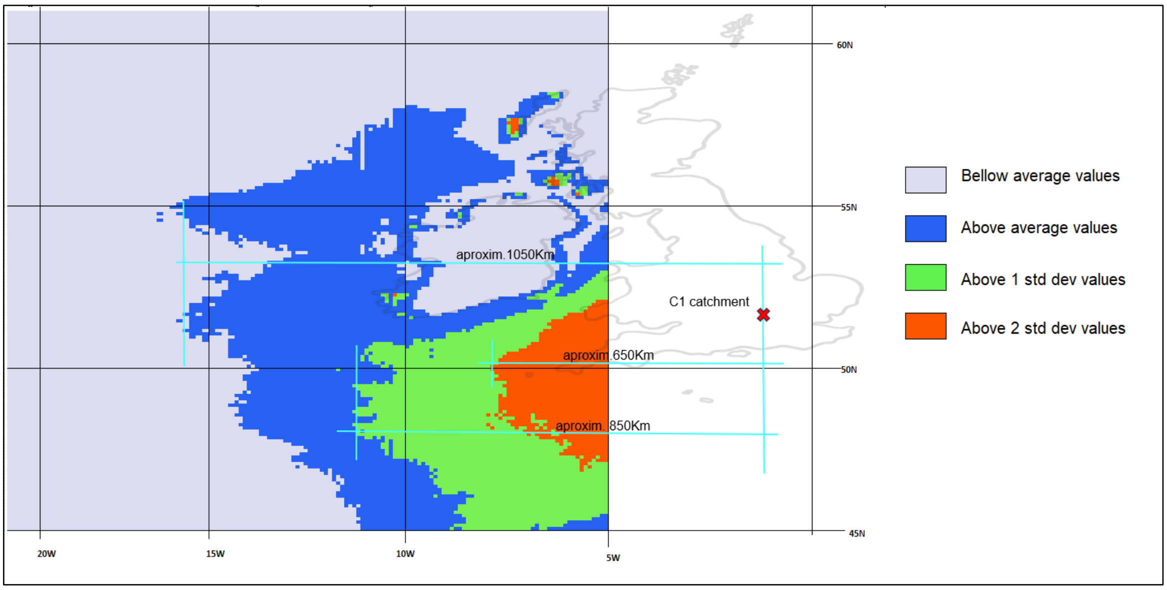


Figure 4-18 – Average and standard deviation of CCC values for C1 across the IMERG grid.

The standard deviation of CCC values is 0.0212 for C1 and 0.024 for C3, indicating low variability across the grid for both catchments, and suggesting a relatively consistent relationship between rainfall and stream level across the monitoring region.

The areas of CCC with highest standard deviation values extend up to around 8°W and 10°W for C1 and C3, respectively. Considering CCC values above 1 standard deviation, is extended up to 11°W or around 850Km distant from C1 and 15°W or 750Km distance from C3.

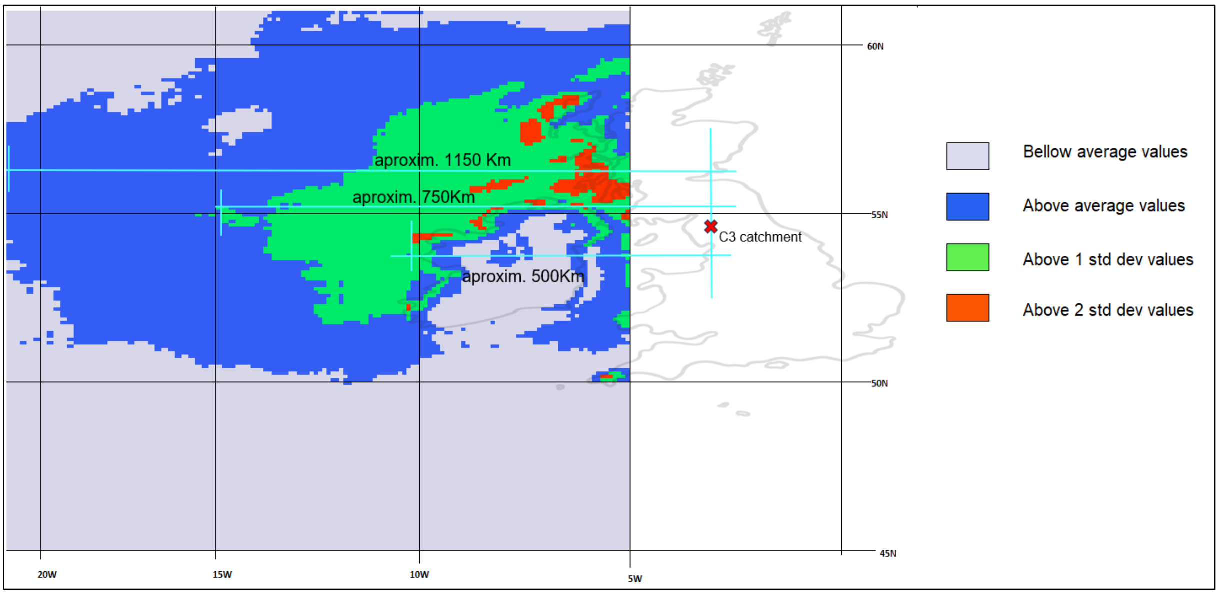


Figure 4-19 - Average and standard of deviation of CCC values for C3 across the IMERG grid.

Yet, the most important aspect of the analysis is the identification of the long-range areas of the grid where rainfall historically has the greatest impact on stream levels at the catchments (see blue area in Figures 4.18 and 4.19). While large regions of the grid show below-average correlation, this study focuses on the areas with the highest correlations, represented by green and red in both figures. These regions are particularly important where they intersect with the purple areas identified in the previous analysis of CCC values.

The intersection zone for C1, represented by the dotted square in Figure 4.20, is considered an optimal region for the case study's data collection. This zone maximizes the distance from the catchment while retaining relevant rainfall data, with CCC values above 1 standard deviation. Within this intersection zone, the yellow pixels represent the most relevant rainfall data. The dotted boundary surrounding these pixels defines the area for data collection, which will be used for testing in the deep learning models. This area will hereafter be referred to as the 'focus area' for IMERG data collection used in the models predicting stream levels at C1.

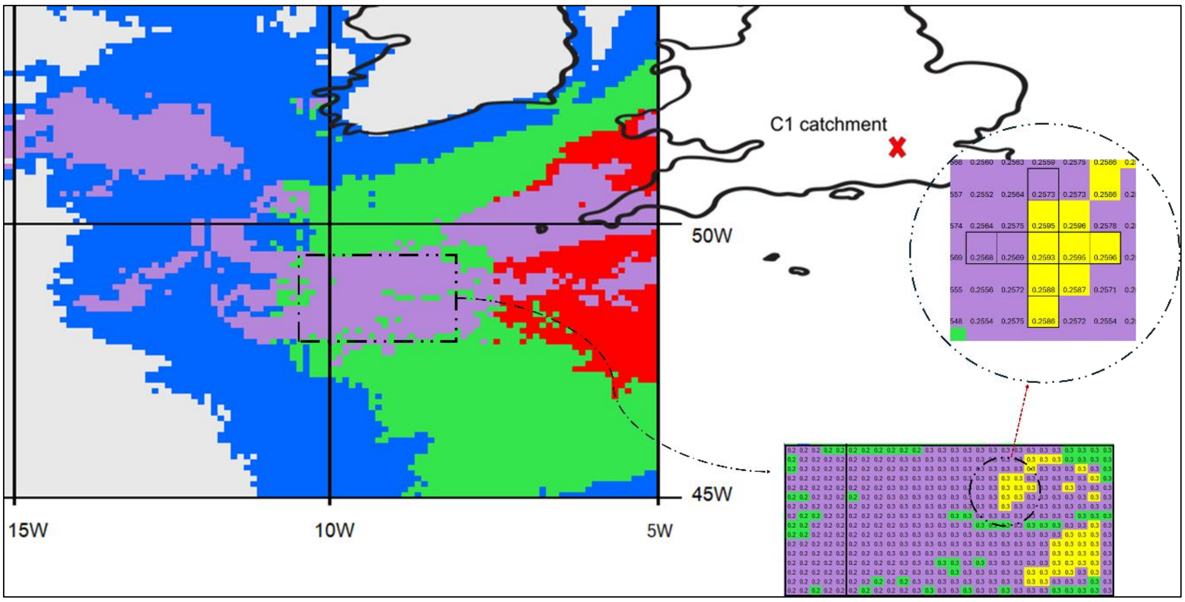


Figure 4-20 – Boundaries of the intersection zone withing the dotted square in the large map. Region of IMERG rainfall data to focus data collection for deep learning modelling, extracted from the detailed image of the intersection zone.

A total of 25 pixels were selected within the focus area of the IMERG grid shown in Figure 4.20. This selection allows for the exploration of the model's ability to capture the patterns between rainfall and stream level in various configurations, while minimizing computational demands and processing times.

The following two main arrangements were tested in this study:

1. Pixels spread horizontally: precipitation data collected from five neighbouring pixels that are highly correlated to stream level at the catchment. These pixels are at the same latitude and differ in longitude coordinates.
2. Pixels spread vertically: precipitation data collected from five neighbouring pixels that are highly correlated to stream level at the catchment. These pixels are the same longitude and differ in latitude coordinates..

This test aims to determine whether a vertical or horizontal sequence of pixels provides a better data structure for predictions. Both arrangements include the pixel centred at the coordinates (49.05°N, 8.75°W), as shown in Table 4.5, which is positioned at the centre of the focus area, located over 650 km from catchment C1. Given an estimated maximum forward speed of 60 km/h, this pixel offers rainfall information with a minimum of 10 hours lead before the event reaches the catchment. Accounting for the IMERG data delay, the lead time for real-time applications is of at least 6 hours.

Table 4.5 - First two arrangements of pixels for input on deep learning models

Group	Pixel coordinates	CCC	Lag (time-steps)	Group	Pixel coordinates	CCC	Lag (time-steps)
1	49.05°N - 8.95°W	0.25682	46	2	49.25°N - 8.75°W	0.25735	46
	49.05°N - 8.85°W	0.25693	46		49.15°N - 8.75°W	0.25952	46
	49.05°N - 8.75°W	0.25934	46		49.05°N - 8.75°W	0.25934	46
	49.05°N - 8.65°W	0.25958	46		48.95°N - 8.75°W	0.25884	46
	49.05°N - 8.65°W	0.25957	45		48.85°N - 8.75°W	0.259	46

In summary, from selecting the case study to identifying the most relevant areas for IMERG data collection, this method was designed to ensure a thorough understanding of the key characteristics of the variables and datasets. This approach establishes a solid foundation for the selection of LSTM deep learning algorithms and structure of the model used for predictions in the following section.

4.4 Datasets preparation and model setup

The initial LSTM code setup comprises four layers: two are LSTM layer with 64 neurons each, one dropout layer with 0.25 dropout rate, and a fully connected (dense) layer to produce the output. Additional settings include Adam optimizer and default MSE as loss function.

The input sequences were designed to account for the lag between the variable’s time series, while producing output sequences according with the number of steps ahead. The initial LSTM code setup comprises four layers: two are LSTM layer with 64 neurons each, one dropout layer with 0.25 dropout rate, and a fully connected (dense) layer to produce the output. Additional settings include Adam optimizer and default MSE as loss function.

Although data processing and modelling tasks were carried out for all three catchments, C1 was selected as the primary source of stream level data for modelling tasks and discussion of results. This selection is due to C1's complexity: with the largest portion of urban area (68%) among the selected catchments, it presents a greater challenge for predictions, as urban streams are known to rise quicker during storms and produce a higher total volume of water discharged during floods than rural streams (Konrad, 2003). Therefore, results for catchments C2 and C3 are provided in Appendices E and D, respectively, while the main text describes the processes, results and discussions based on C1.

As such, the stream level time series at Cranford in the C1 catchment is defined as the model's target variable and is combined with the precipitation variables in the input sequences. Before constructing the sequences, the datasets were split into subsets and normalized, as is commonly required for LSTM models. The normalization process used the min-max feature scaling technique, to transform the values in each time series to fall within the range [0, 1].

The historical datasets were divided into training, validation, and testing subsets as follows:

- Training subset: Data from June 2000 to November 2021.
- Validation subset: The last 10% of the training subset, reserved for validation.
- Testing subset: Data from December 2021 to November 2023.

The datasets were split while preserving the original sequence to maintain the temporal dependency of the data and prevent information leakage from the future during the training process. The IMERG Early Run dataset was normalized but not split, as it will be used solely for evaluating the model's performance after training.

The final step of the data preparation involved creating the input sequences for the LSTM model. The datasets were grouped into several sets of input data, as sets with different data combinations allows for explore the SPPs performance with different data arrangements.

Input sequences were created for each of the following sets of time-series data:

- Set 1: Cranford stream level and IMERG-Cranford pixel precipitation data time series. To assess how the IMERG precipitation data performs on the

LSTM model for future comparison with the local rain-gauge data.

- Set 2: Cranford stream level and precipitation data from the Northolt-RAF rain-gauge time series. To provide a reference for assessment and validation of the IMERG data performance for the LSTM predictions.
- Set 3: Cranford stream level and IMERG precipitation data from pixels in group 1 (see Table 4.5). To test the performance of the model when the data is collected from pixels spread horizontally.
- Set 4: Cranford stream level and IMERG precipitation data from pixels in group 2 (see Table 4.5). To test the performance of the model when the data is collected from pixels spread horizontally. .
- Sets 5, 6 and 7: Cranford stream level and IMERG precipitation data from 5 pixels selected based on the performance of the horizontal and vertical arrangements tested with sets 3 and 4 and. More details about the sets are provided in the following results section.
- Set 8, 9 and 10: Cranford stream level and IMERG Early Run precipitation data from 5 pixels. These sets follow the same structure and criteria used for sets 5, 6 and 7, and differ IMERG product moving from the IMERG Final Run used in the training, to the IMERG Early Run which is the NRT product used to test how the trained model performs in real-time predictions data.

The model's input sequences were created using the function in the following snippet code:

```
def create_sequences(train_data, n_steps_in, n_steps_out)
```

Where:

n_steps_in = number of input time-steps. This number is defined from the lag estimated between the stream level and each of the rainfall time series from the cross-correlation analysis. In set 1 the lag is 21 time-steps, in set 2 lag is 19 time-steps and for the other sets is the lag is 46 except one where the lag is 45. As the correlation analysis reflects the historical lag the n_steps_in was designed to extend it a up to 2 five time-steps to fit in eventual variances.

n_steps_out = number of output time-steps. This number was set to produce output sequences with predictions up to 24 time-steps head in most of the model runs. Exceptionally, the model was set for 12 time-steps ahead to lower computing demands for the assessment of the model's performance with pixels in different arrangements.

Model setup: The model was initially set for 100 epochs with batch size of 128 and tuned along several runs of the model with each set of data sequences. The final number of epochs was reduced to 50 and the batch size increased to 256 for faster and more efficient processing.

4.5 Events selection

This section identifies the events selected for the event-based evaluation of the LSTM model performance.

Records of severe weather in the UK for the years 2022 and 2023 were retrieved from the Met Office list of past weather events (Met Office, 2024). From these, two events were selected prioritizing the ones describing affected areas near or within the catchments C1, C2 and C3 and specifying the incidence of extreme or intense rainfall. Additionally, an event observed in London in October 2022 and reported by the European Severe Storms Laboratory (ESWD, 2023), was added to the selection. This event is missing from the Met Office's list but made local news with impressive headlines at the time.

Follows a list of the events selected with a brief description as provided by the Met Office and the European Severe Storms Laboratory:

<u>Storms Dudley, Eunice, and Franklin (16 – 21 February 2022)</u> - Three named storms affected the UK within the space of a week, the first time this has occurred since storm naming was introduced in 2015/2016. Two rare red warnings were issued for storm Eunice, the most severe and damaging storm to affect England and Wales since February 2014. The persistent heavy rain from the successive storms, together with snow at higher levels, brought significant flooding problems in parts of England, Wales, and Northern Ireland. Much of Wales and northern England received the whole-month February 1991-2020 average rainfall, with some locations more than 150% of average. By 23 February, West Yorkshire had already recorded 224% of the February 1991-2020 long term average rainfall.

<u>Storm Beatrice (On 23 October 2022)</u> - spawned multiple severe thunderstorms across parts of Europe, including tornadic supercells in parts of France and embedded circulations in the mesoscale convective system that impacted parts of England. Eleven tornadoes were confirmed as a result of the outbreak, some of which struck Hampshire and Greater London. The local news described the event with headlines such as "Downpours batter UK with more heavy rain and thunderstorms on the way" (Metro, 23.10.2022), "Watch London being battered by 'biblical rain' as the UK is hit with epic thunderstorm" (Euro Weekly News, 24.10.2022), "Firefighters called to flooding across London after storm batters southern England" (Evening Standard, 24.10. 2022). This event was named as storm Beatrice by the Spanish State Meteorological Agency (AEMET) (Kendon et al., 2023).

Storm Babet (18 to 21 October 2023) - Storm Babet brought exceptional rainfall to parts of eastern Scotland with 150 to 200mm falling in the wettest areas and the Met Office issuing two red warnings for rain. For the county of Angus 19 October 2023 was, by a wide margin, the wettest day on record in a series from 1891. Heavy, persistent, and widespread rain also affected much of England, Wales and Northern Ireland from 18th to 20th, with 100mm falling fairly widely. This was the third-wettest independent 3-day period for England and Wales in a series from 1891, while the Midlands provisionally recorded its wettest 3-day period on record. This rain came on top of very wet weather earlier in October with some central and eastern parts of England and Scotland recording more than twice the October whole-month average rainfall in the first three weeks of the month.

The river level variations at each of the three selected catchments during these rainfall events is displayed in Table 4.6.

Table 4.6 – Catchment's water level variation during the selected rainfall events

Catchment	**C1**			**C2**			**C3**		
Event	**Feb-22**	**Oct-22**	**Oct-23**	**Feb-22**	**Oct-22**	**Oct-23**	**Feb-22**	**Oct-22**	**Oct-23**
Initial water level (mm)	133	157	168	421	193	399	1154	1531	961
Peak water level (mm)	425	752	780	888	508	1093	2812	448	395
Percentual rise	320%	479%	464%	211%	263%	274%	237%	342%	243%
Rising period	9hs	13hs	8hs	6hs	9hs	5hs	5hs	14hs	10hs
STD	75.3mm	206mm	237mm	315mm	62mm	33mm	444mm	62mm	144mm

The selected events will be referred to along the next sections of this thesis as event February 2022, October 2022 and October 2023.

4.6 Assessment of the LSTM model and the proposed methodology

The LSTM model is first assessed following the criteria presented and discussed in the sub-section 3.2.1 of the methodology. It considers the R^2, RMSE, and MAE metrics of the model's predictions and rates performance according to the RSR thresholds presented in Table 2.4. With the case study stablished, these thresholds are defined for each of the catchments and a summary of it provided in Table 4.7 to be used as guide in the interpretation of results.

Table 4.7 - Overall and event-based performance rating thresholds for each catchment

Model performance rating			Very Good RSR<0.5	Good RSR<0.6	Satisfactory RSR<0.7
Catchment	**Event**	**STD**	**RMSE up to**	**RMSE up to**	**RMSE up to**
C1	**Overall**	113mm	56.5mm	67.8mm	79.1mm
	Feb-22	75.3mm	37.65mm	45.18mm	52.71mm
	Oct-22	206mm	103mm	123.6mm	144.2mm
	Oct-23	237mm	118.5mm	142.2mm	165.9mm
C2	**Overall**	205mm	102.5mm	123mm	143.5mm
	Feb-22	315mm	157.5mm	189mm	220.5mm
	Oct-22	62mm	31mm	37.2mm	43.4mm
	Oct-23	33mm	16.5mm	19.8mm	23.1mm
C3	**Overall**	236mm	118mm	141.6mm	165.2mm
	Feb-22	444mm	222mm	266.4mm	310.8mm
	Oct-22	62mm	31mm	37.2mm	43.4mm
	Oct-23	144mm	72mm	86.4mm	100.8mm

The LSTM model is also assessed against an E-NARX model by comparing both model's prediction performance with the same datasets. Additionally, the predictions from each model are used to inform a decision support platform for real-time flooding forecast. This platform was developed for a catchment in the vicinities of C1 and the method relies on the E-NARX model for target dataset generation from stream level and rain-gauges data (Piadeh et al, 2023a). As such, the method can easily incorporate this study's dataset as input in the E-NARX model or replaced it entirely with the LSTM model. Performances are assessed first using the original E-NARX model with the Cranford stream level and Northolt RAF rain-gauge datasets. From the historical level at Cranford, the water level threshold for classifying flooding events is estimate at 0.2m. After the assessment with local data, the predictions of the LSTM model with NRT IMERG data from the remote pixels is also assessed for application in the DS platform.

As a classification method, the outcomes from the DS platform are assessed using key performance indicators and a confusion matrix analysis. These indicators will be further described along with the method's outcomes in the results section .

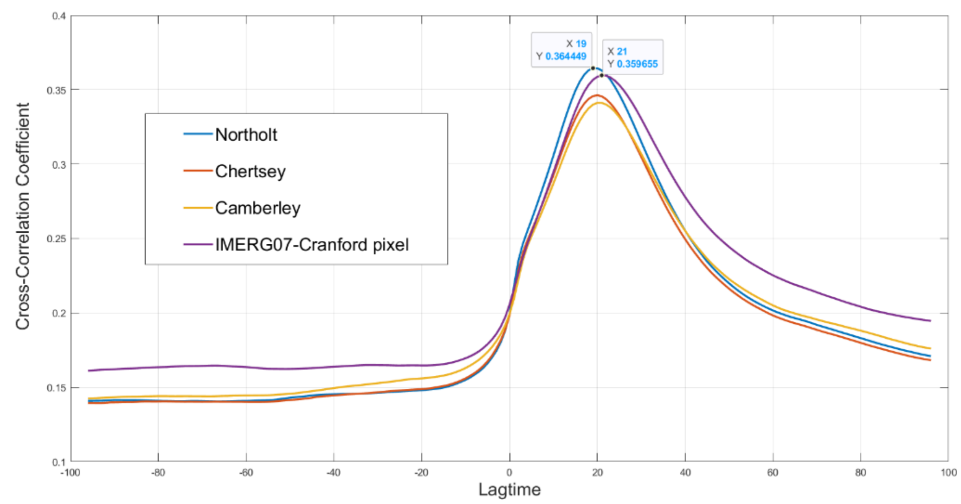
5 Results and discussion

The presentation of the results in this section follows the sequency of the processes stablished in the methodology section.

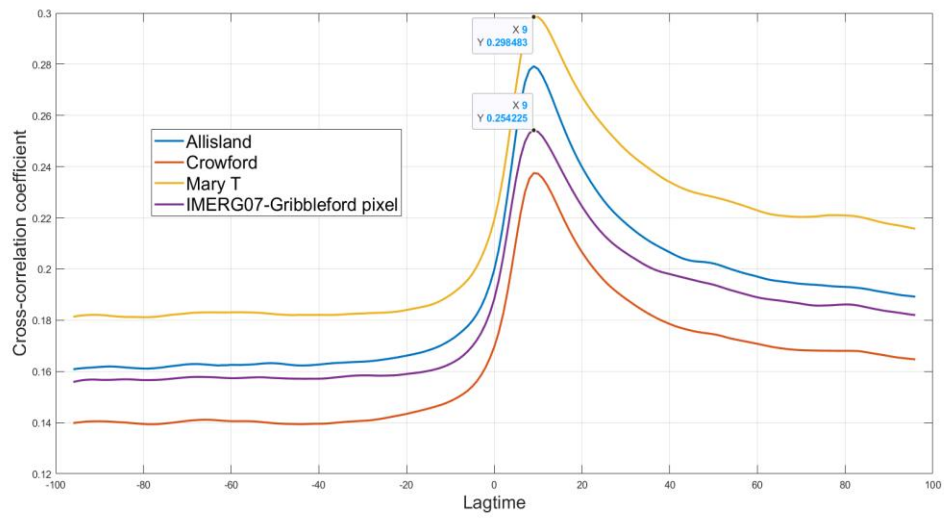
5.1 IMERG data: feasibility, limitations, and requirements

The first assessment of the IMERG data focus on the comparison with rain-gauge data in the three catchment scenarios. The relationship between the rainfall datasets and stream level is analysed through cross-correlation analysis.

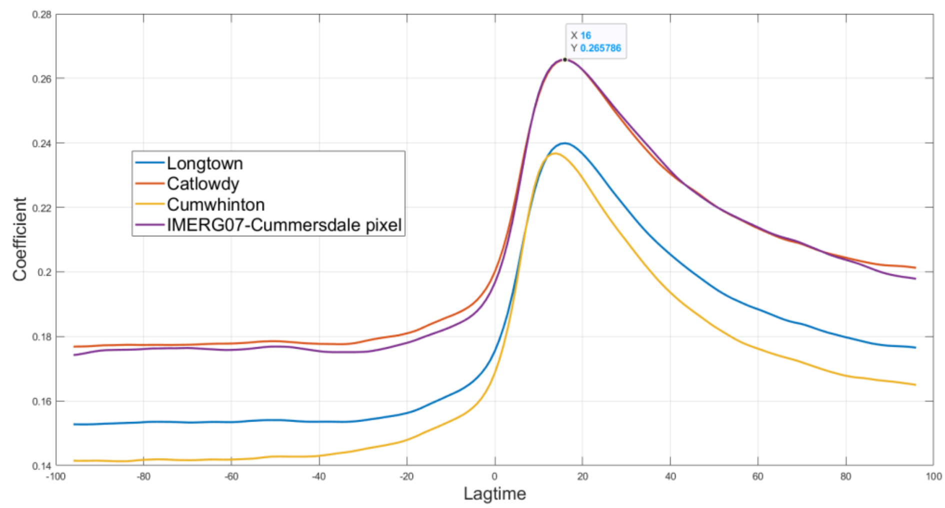
For a visual interpretation of the results, the numerical results first presented in Table 4.4, are reproduced here in Figure 5.1 where the cross-correlation graphs reveal that the curves representing the CCC values are very similar between the datasets in each catchment. Also, that the lag times for all datasets fall within a typical range for catchments of this size in the UK (Black et al., 2021).



(C1)



(C2)



(C3)

Figure 5-1 - Graphs of cross-correlation between the rain data sources (gauges and IMERG) and stream-level at Cranford Park (C1), Gribbleford Bridge (C2) and Cummersdale (C3)

These characteristics are favourable for the study as it confirms a consistent relationship between rainfall and stream level in all catchments, and that these relationships have temporal stability, both of which are important factor in the forecasts. More importantly, that the IMERG data shows a relationship between with the stream level the closely replicate those or the rain-gauges data.

On the interpretation of the CCC, lies a more difficult task as thresholds for categorizing the strength of relationships are variable, and set on arbitrary and inconsistent cutoff points. While most scholars would agree that a coefficient less than 0.1 suggests a trivial or absent relationship and one of 0.9 suggests a very strong relationship, values in the between are debatable. For instance, the correlation value of 0.65 can be interpreted as "good" or "moderate" under different guidelines. It is also unduly simplistic to say that a coefficient of 0.39 indicates a "weak" relationship, whereas 0.40 is considered "moderate.” Hence, coefficient values should be interpreted within the context of the posed research question and properties of the variables (Schober et al., 2018)

Therefore, given the proximity of the cross-correlation coefficient (CCC) values derived from IMERG data to those obtained from rain-gauge data, and considering the presence of underlying factors in the catchment not captured in the cross-correlation analysis, the values of 0.25 and 0.37 can be interpreted as the representation of a moderate to good relationship between the IMERG local data and stream level.

More importantly, the correlation curve of the IMERG data blends in well with the curves of rain-gauges data, as shown in Figure 5.1, and consistently shows CCC values within the top range for each catchment. Therefore, based on the extended and successful application of rain-gauges data in forecasting tasks, this study considers that there is sufficient evidence to supports the suitability of IMERG local data as a reliable source of precipitation data for similar forecasting applications.

Nonetheless, before progressing to the modelling tasks, additional important aspects of the historical relationship between precipitation events and water levels were revealed through the cross-correlation analysis. These aspects warrant further discussion, particularly regarding the observed common trajectories of precipitation events moving from the Atlantic toward the UK catchments, as previously illustrated in Figures 4.16 and 4.17.

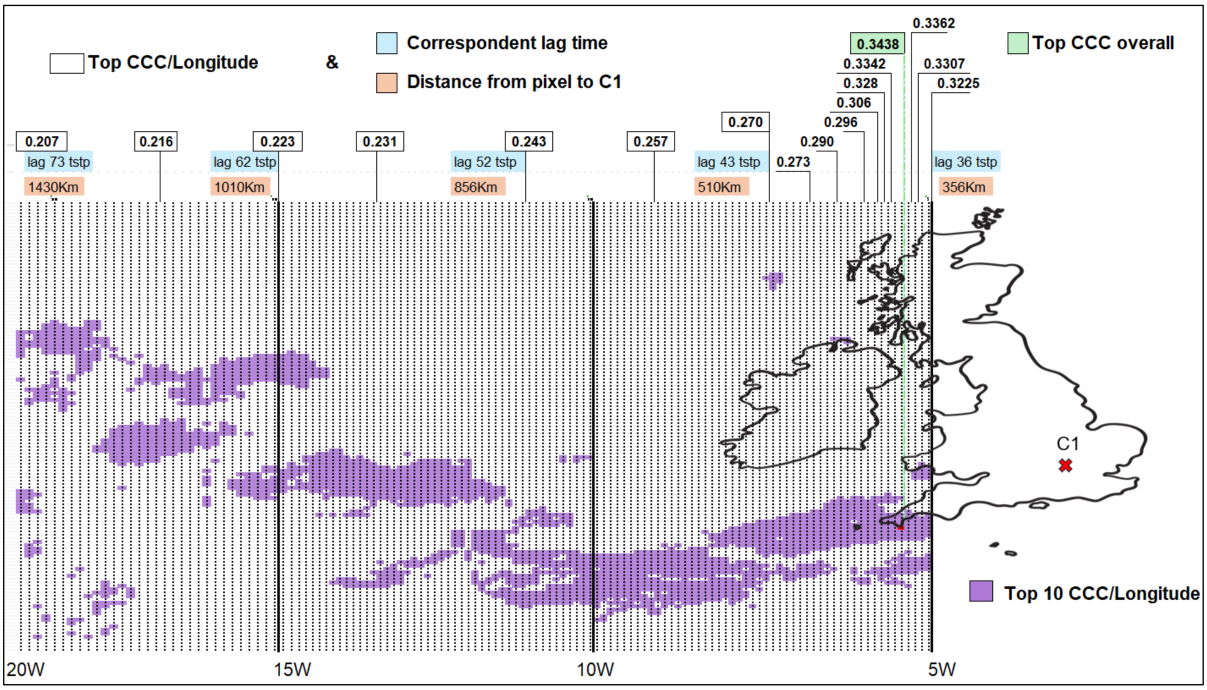


Figure 5-2 – Sample of pixels with the top CCC values with stream level at C1 for several longitudes, accompanied by the respective lag time and distance between the pixel and the C1. In purple, the position of the top 10% CCC values for each longitude.

As shown in Figure 5.2, the cross-correlation analysis for catchment C1 reveals that the precipitation events influencing water levels at Cranford predominantly travelled through a specific range of latitudes, forming a path across the longitudinal lines in the monitored area. This path is well-defined from at the coast of the UK along the oceanic region below the island of Ireland, up to approximately 12°W. Beyond this point, the path begins to spread into blocks of correlated areas and becomes more dispersed at greater longitudinal distances.

Also, as the longitudinal distance increases, the correlation coefficient decreases, and longer lag times are observed. Interestingly, the region of IMERG data most correlated with C1 is not the one closest to the catchment, along the 5°W line, as might be expected. Instead, the IMERG data most strongly correlated with stream level at C1 is located around the pixel with coordinates (50.15, -5.45), which exhibits the highest CCC value. This pixel, referred to as the ‘optimum’ pixel, has a CCC value of 0.3438 with the stream level at Cranford and a historical lag of 37 time-steps (18.5 hours). For real-time applications, taking into account the historical lag and the 4-hour latency of the IMERG Early Run, the data from the optimum pixel for C1 corresponds to an event that typically reaches the catchment 14 hours after the observation.

However, the historical lag is derived from over 20 years of observations, representing the optimal or most likely lag between the two datasets. t's important to note that the lag for individual events may vary, depending on the specific characteristics of each event and the conditions within the catchment area (Wu et al., 2016).

Given that this study aims to use data from rainfall events detected outside the catchment, the total lag includes the time it takes for the rainfall to reach the catchment area. In the case of the data from the optimum pixel for C1, the rainfall event must travel 385 km to reach the catchment's border. This journey could take 6.5 hours if driven by an exceptionally fast weather system or 13 hours for a system moving at the average forward speed. This means that, when accounting for the 4 hours delay in the IMERG data, effectively 2.5hs and 9hs can respectively be added to the lag within the catchment before the rainfall affects water level at Cranford.

Table 5.1 - Time ahead of data availability for predictions using of IMERG rainfall estimates from pixels at different locations for events with fast and average forward speeds towards the C1 catchment. (4hs IMERG data delay considered)

Distance between IMERG pixel and C1	Time for rainfall to reach C1 considering forward speed		Time ahead of the event at the moment that data is available	
	60Km/h	30Km/h	60Km/h	30Km/h
1430 Km	24 hs	48 hs	20 hs	44 hs
1010 Km	17 hs	34 hs	13 hs	30 hs
856 Km	14 hs	28 hs	10 hs	24 hs
510 Km	8.5 hs	17 hs	4.5 hs	13 hs
356 Km	6 hs	12 hs	2 hs	8 hs

Table 5.1 provides the lead times in data collected from some of the best correlated pixels with C1 extracted from Figure 5.2, at variable distance from the catchment.

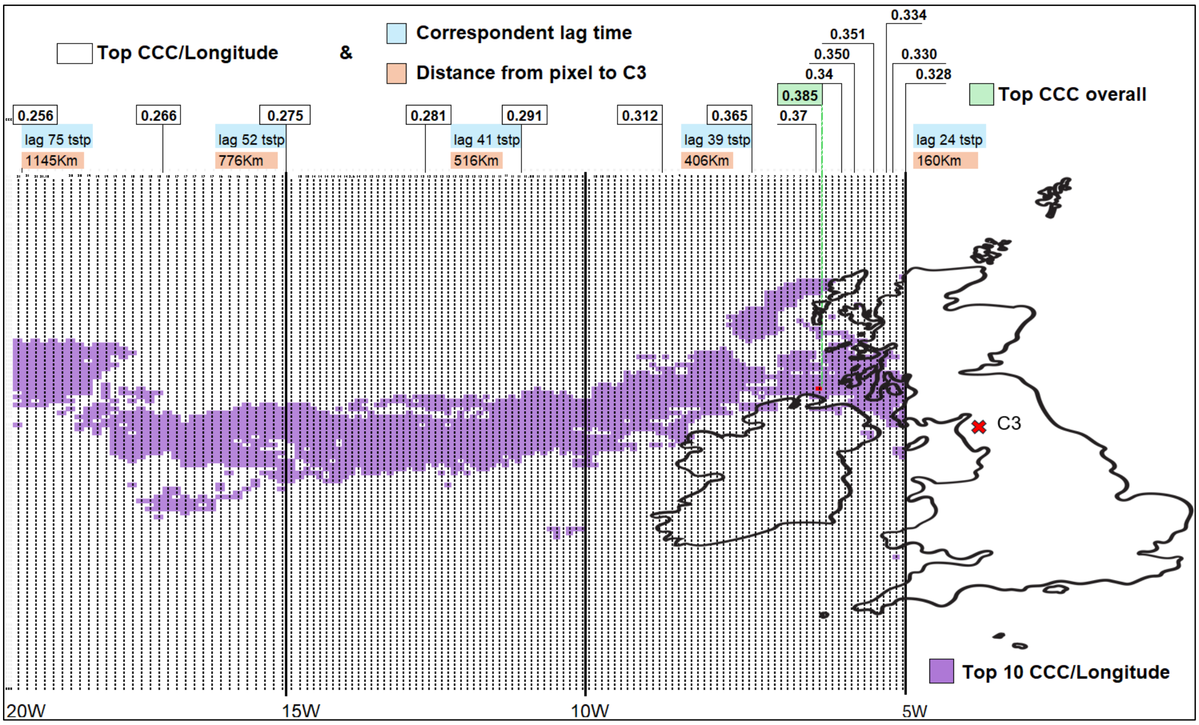


Figure 5-3 - Sample of pixels with the top CCC values with stream level at C3 for several longitudes, accompanied by the respective lag time and distance between the pixel and the C3. In purple, the position of the top 10% CCC values for each longitude.

Similarly to what was observed in C1, the rainfall events follow a common historical path along the Atlantic Ocean towards C3, which can be visualized in the purplecoloured pixels in Figure 5.3. Compared to C1, the path for C3 is better define and more consistent across the entire monitoring region, the CCC values are slightly higher and lag times shorter. However, these is somehow expected as C1 is a catchment with more complex land use, large urban area and located twice as far from the monitoring region compared to C3.

Regarding the lead time for data availability, the optimum pixel for C3, located at coordinates (55.75, -6.35), has a lag of 29 time-steps (14.5 hours). Following the same rationale applied in the C1 cross-correlation analysis, the optimum pixel for C3 provides data 10.5 hours before the event reaches the catchment, based on the historical lag.

Given the variable conditions influencing the lag of individual events, the optimum IMERG pixel for C3, located 238 km from the catchment, provides information after the event has reached the catchment if the event moves at 60 km/h. However, if the event travels at the average speed of 30 km/h, the information is provided with 4 hours lead. Similarly, several pixels near the C3 catchment will have their potential for increasing lead time limited by the event's forward speed. Table 5.2 presents the time ahead of data provision for a selection of the best-correlated pixels with C3, extracted from Figure 5.3, and located at variable distance from the catchment..

Table 5.2 – Time ahead of data availability for predictions using of IMERG rainfall estimates from pixels at different locations for events with fast and average forward speeds towards the C3 catchment. (4hs IMERG data delay considered)

Distance between IMERG pixel and C3	Time for rainfall to reach C3 considering forward speed		Time ahead of the event at the moment that data is available	
	60Km/h	30Km/h	60Km/h	30Km/h
1145 Km	19 hs	38 hs	15 hs	34 hs
776 Km	13 hs	26 hs	9 hs	22 hs
516 Km	8.6 hs	17 hs	4.6 hs	13 hs
406 Km	6.8 hs	13.5 hs	2.8 hs	9.5 hs
160 Km	2.6 hs	5.3 hs	-	1.3 hs

Additional aspects of the relationship between the IMERG precipitation estimates and stream level are revealed along further analysis of the cross-correlation results in both catchments.

Figure 5.4 displays the distribution of pixels across the monitoring regions based on their correlation values. With a standard deviation of 0.0212 and an average CCC value of 0.2187 in C1, and a standard deviation of 0.024 with an average CCC value of 0.2508 in C3, the variation in CCC values is limited to approximately 9.7% in both correlation grids.

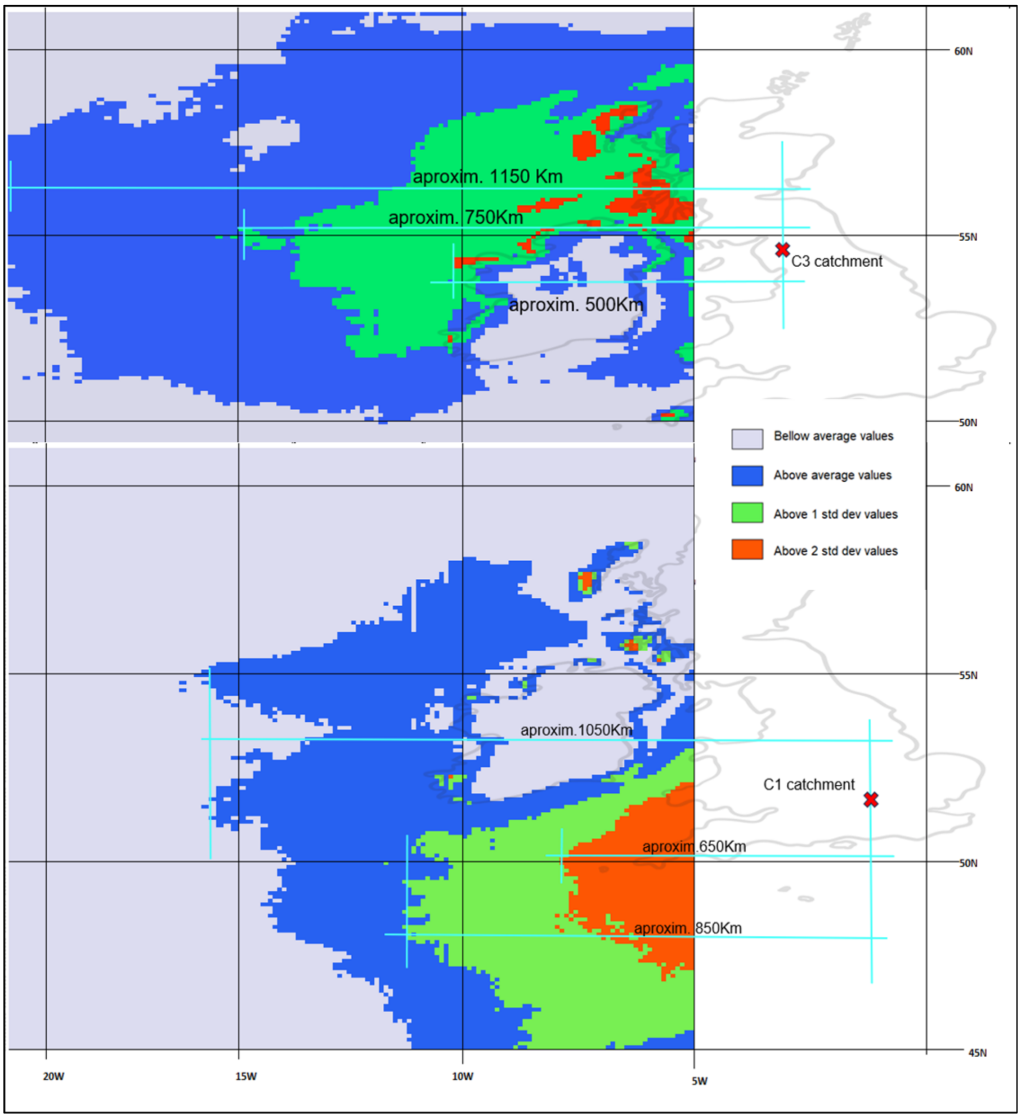


Figure 5-4 – Distribution of IMERG pixels across the monitoring region based on the average and standard deviation of CCC values with C3 and C1.

While observing Figure 5.4, the large red region of pixels for C1 stands out, along with considerably smaller and sparser red areas for C3. These red areas represent the most correlated pixels and potentially the most valuable data for stream level predictions. However, due to the relatively low standard deviation across the grid, the green zones may be just as relevant as the red ones, particularly for the goal of achieving greater lead time in detecting flood-risk-related rainfall events. Anyway, the extensive green and red areas over the oceanic regions, both below and above the island of Ireland, represent hot spot areas for monitoring precipitation to predict flood risk at C1 and C3, respectively.

It is also interesting that the region above Ireland shows the weakest correlation with both catchments, despite having the highest density of rain gauges for the IMERG validation and being relatively close to C1 and C3. Still, this aligns with the anomalies identified during the data inspection process in Section 4.2.2, demonstrated here in Figure 5.5. Yet, the area falls within the range of land-based weather instruments for higher-quality rainfall data, therefore, this area is of lower priority for this study.

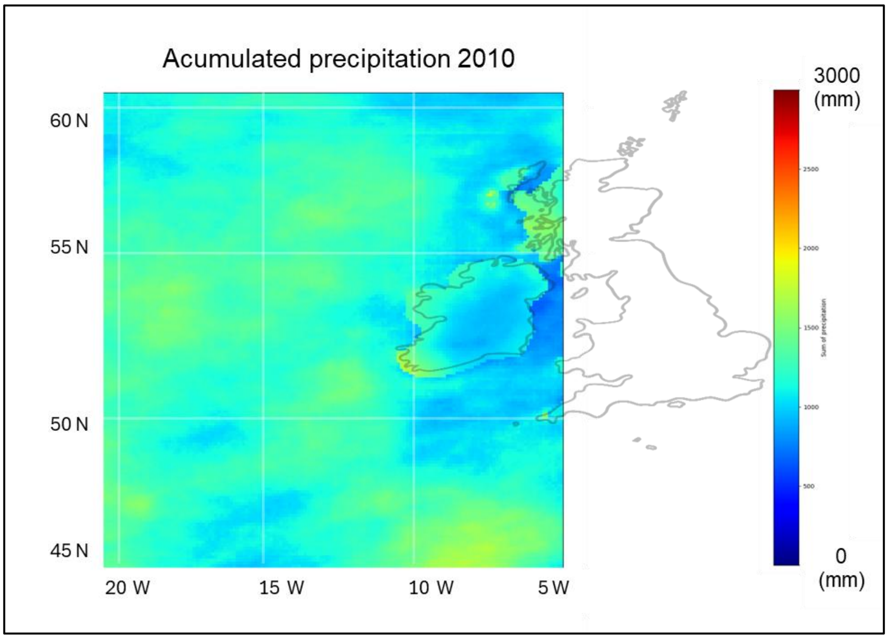


Figure 5-5 - Total annual precipitation in 2010 according to IMERG V07 estimates.

In comparison, beyond the Irish borders further into the Atlantic, there is a common region in the IMERG data seems to be relevant for both catchments. This region, detailed in the left side of Figure 5.6, is marked by the dark area within the dashed lines. It represents the intersection of the areas of the IMERG data with above average correlation with the stream levels in both catchments.

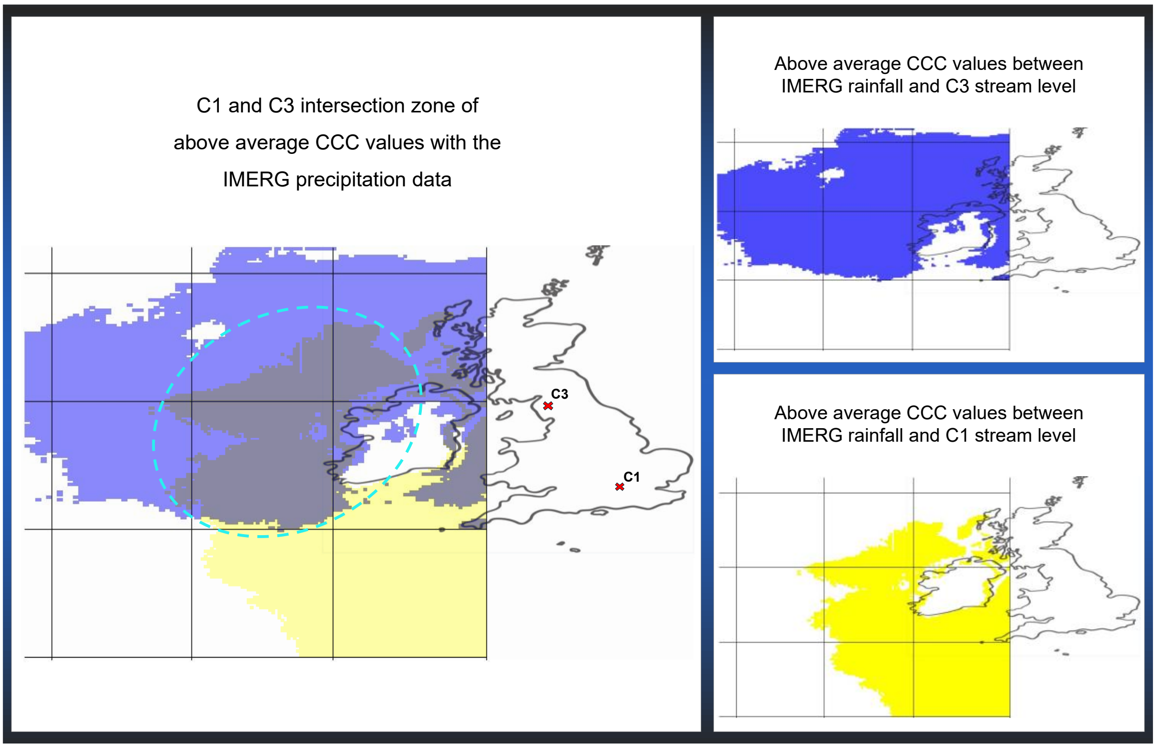


Figure 5-6 – Area of the IMERG precipitation data in the monitoring region that has above average correlation with both catchments' stream level

The size of the intersection zone is large advancing up to around 15°W and relates to catchments of different characteristics and separated by more than 410Km. As such, the is a possibility of the region being also relevant to other catchments in the vicinities or between C1 and C3. If so, this would mean an opportunity to extend the method for predictions to several other catchments. Yet, it depends on the learning and prediction capability of the DL models with the IMERG data that is yet to be tested.

5.2 LSTM model performance with precipitation data from rain-gauge and the local IMERG pixel

This subsection presents the stream level predictions using the LSTM model with IMERG data from the Cranford pixel, which is considered local to the stream gauge station. These predictions are compared and validated against the model's results using rain-gauge data.

The model is designed to predict up to 24 time-steps (12 hours) ahead, using input sequences from two sets: Set 1 (IMERG-Cranford pixel and stream level at Cranford) and Set 2 (rain gauge at Northolt RAF and stream level at Cranford). The LSTM model structure is identical for both sets, differing only in the size of the input sequences, which is adjusted to account for the lags between variables - 21 for Set 1 and 19 for Set 2 (see Table 4.4). Model's training time on a GPU NVIDIA GeForce RTX 4090 is 10 seconds per epoch, and predictions running time of 1 second.

Since both sets use the same stream level data and differ only in the rainfall data, results will be discussed by referring to Set 1 as IMERG-Cranford and Set 2 as Northolt gauge. The performance metrics of the LSTM model are shown in Table 5.3 and interpreted based on the RSR thresholds stablished in Table 4.7.

Table 5.3 – Performance metrics for LSTM model predictions of stream levels using the Northolt rain-gauge data and IMERG-Cranford precipitation data. Rating thresholds: Very good (green), Good (blue), Satisfactory (orange).

RMSE (mm)		*MAE (mm)*		*Lead-time*	
Northolt gauge	**IMERG-Cranford**	**Northolt gauge**	**IMERG-Cranford**	***Hours***	***Steps ahead***
6.30	**6.23**	2.86	**2.52**	*0.5*	*1*
8.44	9.26	3.26	3.27	*1.0*	*2*
11.38	11.75	4.14	4.50	*1.5*	*3*
14.38	**14.21**	5.18	5.50	*2.0*	*4*
17.34	**16.82**	6.36	6.57	*2.5*	*5*

20.32	**19.49**	7.60	7.82	*3.0*	*6*
23.07	**22.28**	8.73	9.13	*3.5*	*7*
25.72	**25.05**	9.82	10.26	*4.0*	*8*
28.36	**27.76**	10.90	11.42	*4.5*	*9*
30.94	**30.45**	12.00	12.64	*5.0*	*10*
33.48	**33.06**	13.04	13.91	*5.5*	*11*
35.93	**35.55**	14.04	15.13	*6.0*	*12*
38.22	**37.94**	14.97	16.25	*6.5*	*13*
40.42	**40.27**	15.88	17.32	*7.0*	*14*
42.53	**42.48**	16.73	18.41	*7.5*	*15*
44.61	**44.55**	17.59	19.50	*8.0*	*16*
46.62	**46.53**	18.44	20.60	*8.5*	*17*
48.53	**48.40**	19.24	21.56	*9.0*	*18*
50.38	**50.22**	19.98	22.55	*9.5*	*19*
52.20	**51.95**	20.71	23.43	*10.0*	*20*
53.99	**53.65**	21.42	24.35	*10.5*	*21*
55.72	**55.33**	22.14	25.21	*11.0*	*22*
57.44	**56.95**	22.87	25.99	*11.5*	*23*
59.13	**58.52**	23.60	26.82	*12.0*	*24*
35.22	***34.94***	***13.81***	***15.19***	***Overall***	

Considering that water levels at Cranford typically range between 60 mm and 1,100 mm with a standard deviation (STD) of 113 mm, the RMSE and MAE values are considered low. The model's performance is considered 'very good' (RSR < 0.5) on predictions with lead times up to 11 hours with both rainfall data sources, remaining 'good' (RSR < 0.6) up to the max lead time of 12 hours. Overall, RMSE values are better in the predictions with the IMERG-Cranford data (RMSE = 34.94mm), while MAE values are better for predictions with the Northolt gauge data (MAE = 13.81mm).

Given that the MAE metric treats errors equally according to its magnitude, and that the RMSE is more error sensitive, emphasizes large errors, the results in Table 5.3 indicate that the IMERG-Cranford dataset had fewer large errors or overall low range in the error's magnitude. The dataset from the Northolt gauge, on the other hand, probably had a few errors of greater magnitude, but errors were in average lower.

Interpreting the distribution of errors is crucial for evaluating the model, particularly in terms of how its predictions will be applied in real-world scenarios. In this research, avoiding large errors is critical, as the magnitude of the error could determine whether a flood risk alert is triggered. Consequently, the lower RMSE obtained with the IMERG-Cranford data, show in ‘**bold**’ values in Table 5.3, makes it preferable for stream level predictions.

It is also observed that difference between the MAE values from results of both datasets consistently grow along increasing lead times, while the difference between the RMSE values fluctuates within a small range of values. This suggests that, although the model using IMERG-Cranford data shows overall increase in average errors as the prediction lead time extends, the magnitude of these errors does not grow significantly. Additionally, is expected that the accuracy decreases as lead times increase in regression models, which uses past predictions as inputs for predicting the next step. Yet, the LSTM model deals better with it than other recurring networks due to the inclusion of the forget gate filter on its structure (Kratzert et al, 2018).

Similar, deduction can be drawn from Figure 5.7 which displays the variance plots between the results from the LSTM model using the IMERG-Cranford data and target values for different lead times. The plots show that the errors do not increase at random but within a certain range of values.

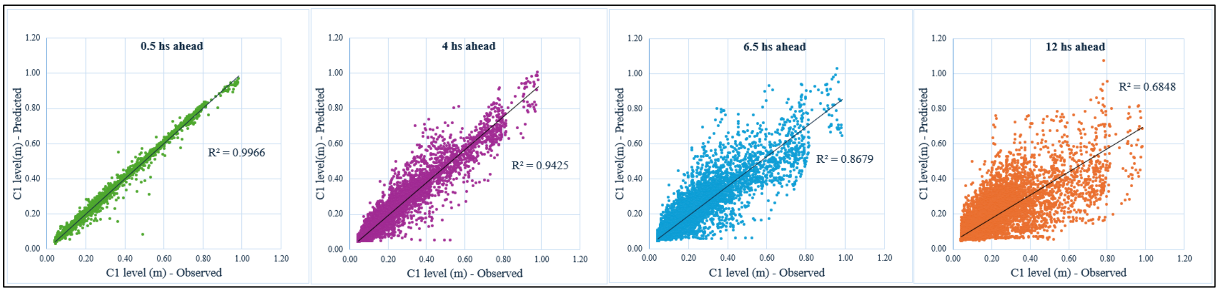


Figure 5-7 – R-squared values for multi-step-ahead LSTM model predictions using the IMERG-Cranford testing dataset

Figure 5.8 presents the corresponding variance plots for the results of the LSTM model with the Northolt gauge data which, in comparison with the results in Figure 5.7, there are similar patterns and range of variance. Both figures also show that along the plots there is a tendency to underestimate the predictions at greater lead times.

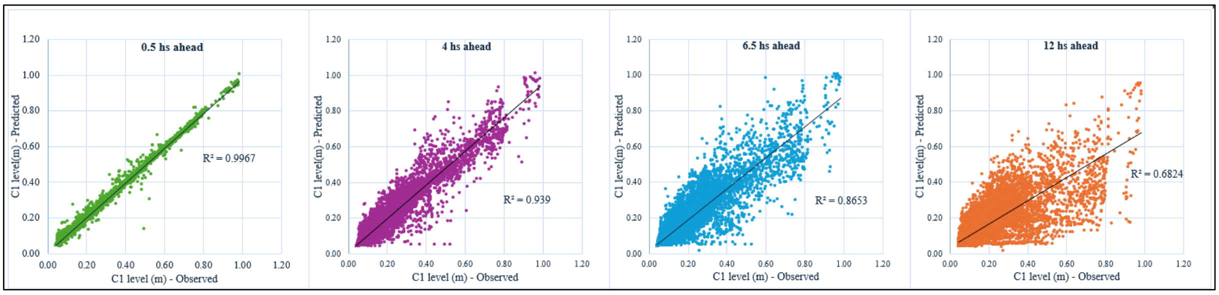


Figure 5-8 – R-squared values for multi-step-ahead LSTM model predictions using the Northolt gauge testing dataset.

Based on the R^2 values extracted from the plots, the model demonstrates a strong ability to capture underlying data trends across all lead times. This ability is particularly evident for lead times up to 6.5 hours, with R^2 values of 0.8679 and 0.8653 for the IMERG-Cranford and Northolt gauge data, respectively. Even at a 12-hour lead time, the model continues to produce predictions within an acceptable range of variance, with R^2 values of 0.6849 for IMERG-Cranford and 0.6824 for Northolt gauge data (Almikaeel et al., 2024). On comparing overall performances, the consistently higher R^2 values and lower RMSE values for the LSTM model using the IMERG-Cranford data indicate that this dataset outperforms the Northolt rain-gauge data for stream level prediction at Cranford.

Next, the assessment focused on event-based predictions and the model is tested with each rainfall dataset for predicting the water levels at Cranford during the February 2022 rainfall event as presented in Table 5.4.

Table 5.4 - Performance metrics of the LSTM model level predictions for February 2022. Rating thresholds: Very good (green), Good (blue), Satisfactory (orange).

RMSE (mm)		***MAE (mm)***		***Lead-time***	
Northolt gauge	**IMERG-Cranford**	**Northolt gauge**	**IMERG-Cranford**	***Hours***	***Steps ahead***
8.65	**5.69**	4.96	**3.90**	*0.5*	*1*
9.65	**8.99**	5.77	**5.71**	*1.0*	*2*
12.57	**12.46**	7.25	7.85	*1.5*	*3*
16.79	**15.92**	9.04	9.73	*2.0*	*4*
20.65	**19.29**	11.13	11.84	*2.5*	*5*
24.48	**23.18**	13.29	14.19	*3.0*	*6*
27.98	**26.76**	15.57	16.57	*3.5*	*7*
31.40	**30.53**	17.67	18.84	*4.0*	*8*
34.66	**34.09**	19.69	20.89	*4.5*	*9*
37.63	**37.30**	21.96	23.08	*5.0*	*10*
40.52	**40.49**	24.02	25.51	*5.5*	*11*
43.16	43.54	26.03	27.86	*6.0*	*12*
45.81	46.66	27.97	30.06	*6.5*	*13*
48.14	49.66	29.76	32.07	*7.0*	*14*
50.69	52.55	31.73	33.95	*7.5*	*15*
53.36	55.58	33.66	35.88	*8.0*	*16*
56.04	58.48	35.30	37.73	*8.5*	*17*
59.21	61.36	37.35	39.48	*9.0*	*18*
62.46	64.28	39.36	41.14	*9.5*	*19*
65.38	67.21	41.22	42.80	*10.0*	*20*
68.43	70.35	43.28	44.37	*10.5*	*21*
71.18	73.43	45.22	45.93	*11.0*	*22*
73.86	76.47	47.15	47.26	*11.5*	*23*
76.29	79.37	48.91	**48.45**	*12.0*	*24*
43.29	***43.90***	***26.55***	***27.71***	***Overall***	

The prediction metrics in table 5.4 show a comparable performance of the model with both datasets with very similar RMSE and MAE values beyond the 1-hour lead time. Nonetheless, LSTM model performance is considered very good on predictions with lead times up to 5 hours (RSR < 0.5), remaining good for lead times up to 6.0 hours (RSR < 0.6). Furthermore, as displayed by the plots in Figure 5.9, the LSTM model is able to follow the rising and descend of the true level curve on predictions with lead times up to 12 hours using the IMERG-Cranford data.

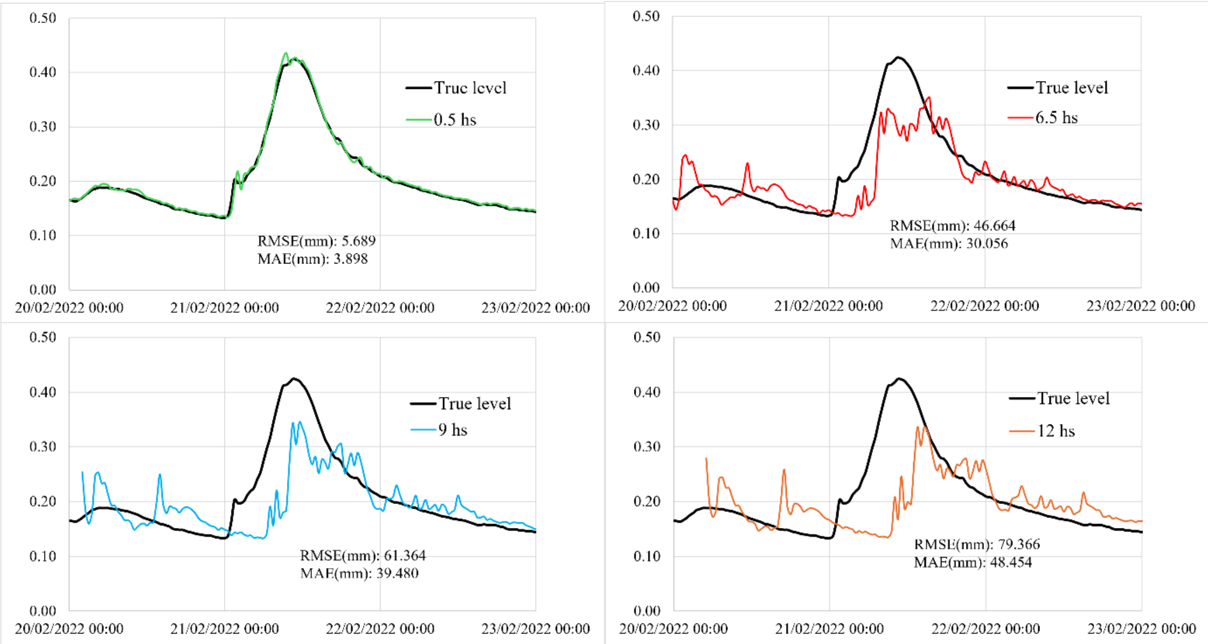


Figure 5-9 - LSTM model performance for different lead times in predictions of the February 2022 event with IMERG-Cranford data

For a comparison of the model's performance with each dataset, Figure 5.10 shows the plots of the predictions with the Northolt gauge data.

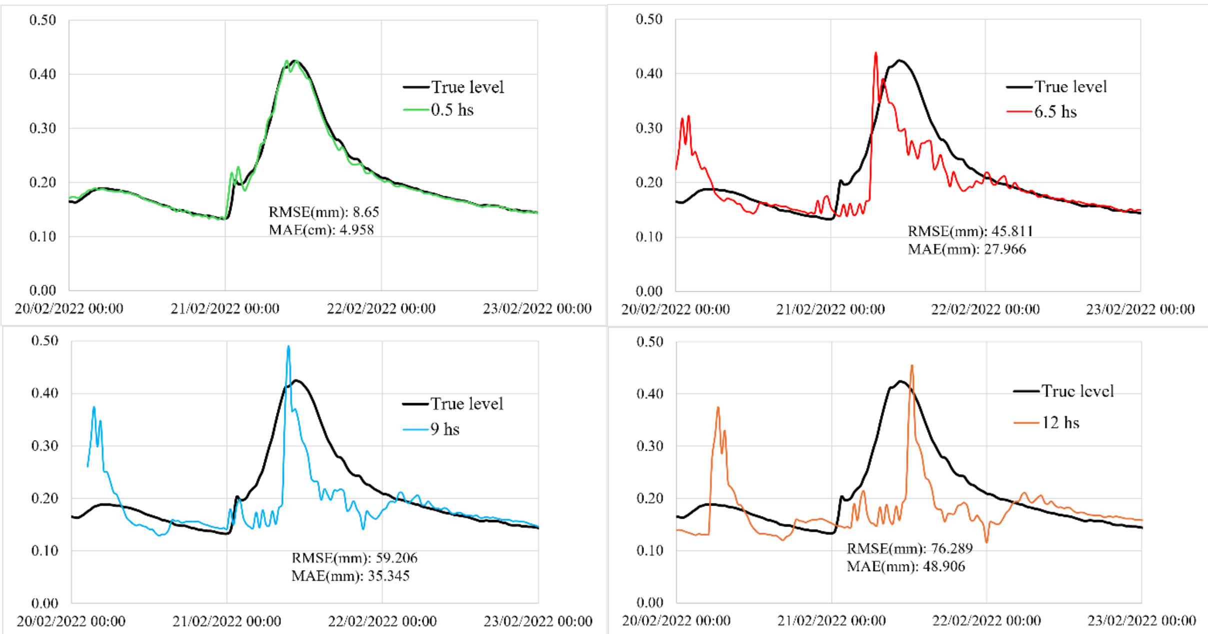


Figure 5-10 - LSTM performance for different lead times in predictions of the February 2022 event with Northolt gauge data

Observing the plots in Figures 5.9 and 5.10 is noted that the model significantly overestimates the long-term predictions of water level at the start of the event with both datasets. This initial overestimation can be related to past recent rainfall or previous level variations as the model receives inputs of data going back at least eight weeks before the event. However, the model's struggle to replicate the true level curve with the rain-gauge data is distinctive and extends beyond the early stages of the event. This is particularly concerning due to the extreme values and sharp changes, and the underestimation of the duration of the high stream level period which negatively impacts its application in risk forecasts.

In comparison, the model with the IMERG-Cranford data produced predictions that are generally closer to the true water level curve and accurately captures the duration of raised stream levels. Overall, although errors are more evenly distributed throughout the event period, there is a significant underestimation of the peak values and a delay on estimating the receding stage. However, the underestimation of peak values is a common issue in most hydrological time series models due to their regressive nature relying on past values for predictions (Nourani et al., 2022). Consequently, the LSTMs and other machine learning models, can struggle with sudden shifts in the data and might smooth out these changes (Kratzert et al., 2018) leading to the underestimation of peaks. In contrast, although the predictions using the rain-gauge data deviate more from the actual water level curve, there is a slight overestimation of the peaks. This behaviour is possibly related to the punctual representation the precipitation data or the dataset's temporal unit upscaling and will be further explored on the evaluation of the model's predictions with other rainfall events.

Next, the metrics of the prediction with the LSTM model for October 2022 event is presented in Table 5.5. The RMSE and MAE values for different lead times indicate that the model using IMERG data continues to perform at a comparable or superior level to the model using rain-gauge data. This time, however, the predictions using IMERG data exhibit consistently smaller absolute errors (MAE) compared to those using rain-gauge data, as highlighted by 'bold' values in the table.

Table 5.5 - Performance metrics of the LSTM model predictions for October 2022. Rating thresholds: Very good (green), Good (blue), Satisfactory (orange).

RMSE (mm)		*MAE (mm)*		*Lead-time*	
Northolt gauge	**IMERG-Cranford**	**Northolt gauge**	**IMERG-Cranford**	***Hours***	***Steps ahead***
13.38	19.45	6.74	**6.63**	*0.5*	*1*
20.25	28.63	8.89	8.92	*1.0*	*2*
29.58	33.88	11.57	11.71	*1.5*	*3*
36.43	39.45	14.89	14.95	*2.0*	*4*
41.68	45.48	18.27	18.44	*2.5*	*5*
46.55	50.08	22.01	**21.60**	*3.0*	*6*
50.15	54.07	25.24	**24.65**	*3.5*	*7*
53.24	58.95	28.31	**27.44**	*4.0*	*8*
56.69	63.71	31.04	**30.06**	*4.5*	*9*
61.52	67.93	34.07	**32.99**	*5.0*	*10*
67.07	72.76	37.47	**36.23**	*5.5*	*11*
72.41	77.61	40.90	**39.23**	*6.0*	*12*
77.40	81.87	43.81	**42.14**	*6.5*	*13*
82.42	86.76	46.60	**45.23**	*7.0*	*14*
88.16	92.70	49.49	**48.68**	*7.5*	*15*
95.21	98.76	52.80	**52.31**	*8.0*	*16*
103.17	104.19	56.57	**55.93**	*8.5*	*17*
111.02	**109.72**	60.55	**59.42**	*9.0*	*18*
118.58	**115.47**	64.52	**62.87**	*9.5*	*19*
126.06	**120.68**	68.48	**66.21**	*10.0*	*20*
133.21	**126.28**	72.35	**69.45**	*10.5*	*21*
139.79	**132.09**	76.29	**72.67**	*11.0*	*22*
146.23	**137.55**	80.27	**75.71**	*11.5*	*23*
152.60	**142.92**	84.26	**78.82**	*12.0*	*24*
80.12	***81.71***	***43.14***	***41.76***	***Overall***	

The LSTM model also underestimates the peaks of stream level with the IMERG-Cranford data during the October 2022 event, as shown in Figure 5.11.

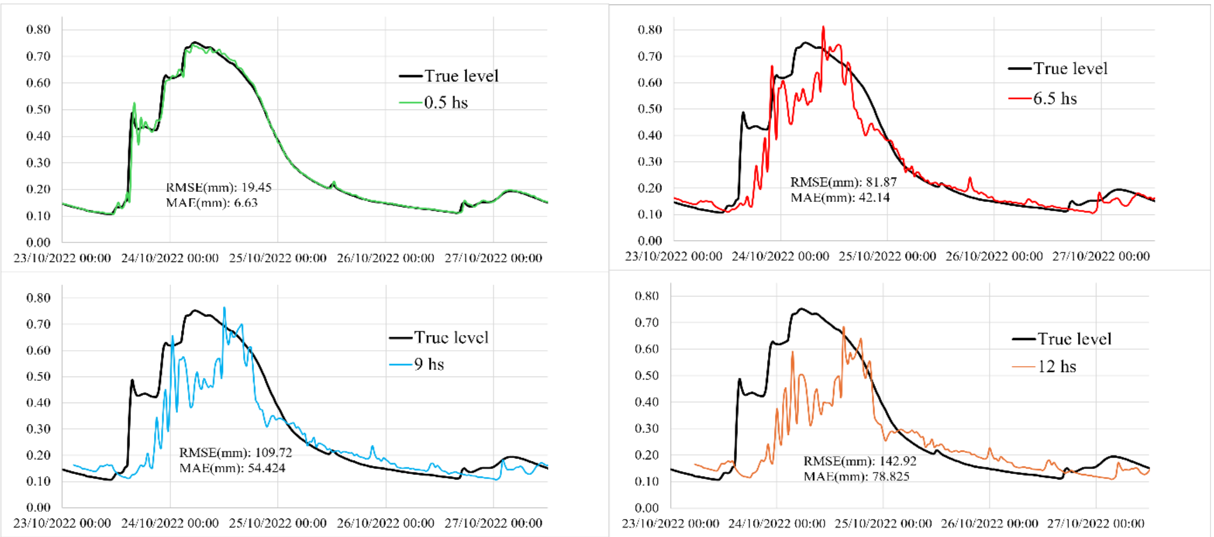


Figure 5-11 - LSTM performance for different lead times in predictions of the October 2022 event with IMERG-Cranford data.

Figure 5.12 shows that the tendency of the model to overestimate peak levels with the rain gauge data persists for this event.

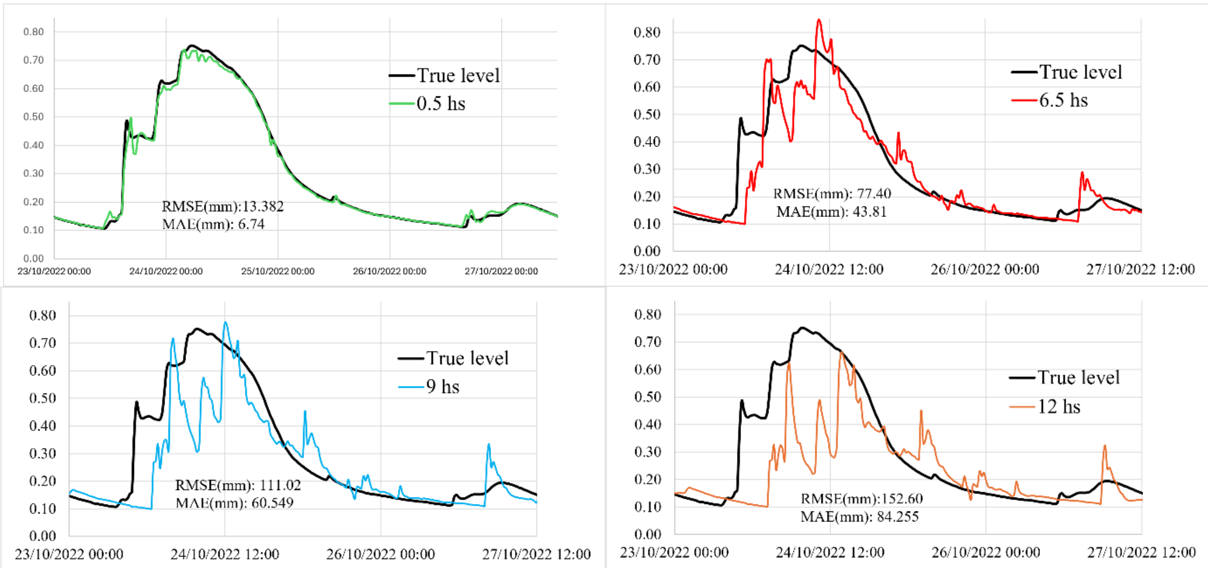


Figure 5-12 - LSTM performance for different lead times in predictions of the October 2022 event with Northolt gauge data.

Observing the plots in Figure 5.11 and 5.12, it is evident that the predictions using IMERG data align more closely with the true stream level curve and more effectively represent the receding stage.

Additionally, based on the predictions from the February 2022 event, there are three clear differences on the model's performance:

1. The errors' magnitude is significantly higher for October 2022. Yet, the performance is 'very good' (RSR < 0.5) on predictions with up to 8.5 hours lead time and 'good' for up 10.5 hours lead time, with both rainfall sources. This represents an increase of 3.5 hours and 4.5 hours in lead time of predictions with the same ratings in February 2022.
2. The predictions using IMERG-Cranford data for October 2022 show a slight overestimation of peak levels on 6.5 hours and 9 hours lead times. Therefore, there was an improvement of the model's performance between events with the same data sources. This is particular important for flooding risk detection considering that water levels during the October 2022 event rose by twice as much compared to the earlier February 2022.
3. The predictions for October 2022 using the rain-gauge data show peaks of overestimation in the receding stage of the water level which cannot be explained by unseen data in the plots which indicates to poor model performance with the data.

So far, these differences seem to indicate that the model's performance is influenced by the characteristics of the event and the source of the data. As such, this will be further explored along the analysis of the performance metrics for the third event which is presented in Table 5.6.

The metrics for the October 2023 events show that errors on predictions are larger across all lead time with the rain-gauge dataset, and that the differences of RMSE values between the predictions with each datasets increased. Therefore, indicating greater ability of the model for predictions with the IMERG-Cranford data which is confirmed by predictions considered 'very good' and 'good' with lead time up to 6.0 hours and 7.5 hours, respectively. In comparison, the predictions with the rain-gauge data were 'very good' only up to 5 hours and not considered satisfactory on prediction with lead times beyond 6.5 hours.

Table 5.6 - Performance metrics of the LSTM model predictions for October 2023. Rating thresholds: Very good (green), Good (blue), Satisfactory (orange).

RMSE (mm)		*MAE (mm)*		*Lead-time*
Northolt gauge	**IMERG-Cranford**	**Northolt gauge**	**IMERG-Cranford**	***Hours***
13.18	18.78	**8.38**	10.40	*0.5*
21.71	25.24	13.85	**13.31**	*1.0*
31.09	31.17	20.83	**17.99**	*1.5*
42.70	**39.22**	28.88	**23.24**	*2.0*
55.38	**46.59**	37.67	**29.04**	*2.5*
67.91	**54.06**	46.66	**35.61**	*3.0*
81.09	**62.50**	56.02	**42.68**	*3.5*
95.11	**72.41**	65.50	**49.79**	*4.0*
109.68	**82.67**	75.76	**56.83**	*4.5*
123.98	**92.60**	85.94	**63.14**	*5.0*
138.23	**102.49**	96.14	**69.32**	*5.5*
152.10	**112.13**	106.39	**75.40**	*6.0*
165.02	**122.17**	115.99	**81.95**	*6.5*
177.36	**132.21**	125.14	**88.32**	*7.0*
189.36	**141.46**	134.27	**94.63**	*7.5*
200.43	**150.43**	142.66	**100.77**	*8.0*
210.50	**158.73**	150.44	**106.51**	*8.5*
219.55	**166.39**	157.51	**112.08**	*9.0*
227.83	**173.97**	163.88	**117.51**	*9.5*
235.31	**181.00**	169.46	**122.50**	*10.0*
242.09	**187.85**	174.59	**127.54**	*10.5*
247.83	**194.48**	178.59	**132.27**	*11.0*
252.91	**200.64**	181.50	**136.44**	*11.5*
257.48	**206.24**	184.19	**140.20**	*12.0*
148.24	***114.81***	***105.01***	***76.98***	***Overall***

From the plots of predictions with the IMERG-Cranford data, in Figure 5.13, is noted that the model's errors increase significantly around mid-point of the event, at certain points oscillating between overestimation and underestimation. The peaks are evident on the plots for greater lead times, which is assumed to be the source of the difference between the RMSE values from both datasets. The remaining predictions show much smaller errors, in agreement with the smaller difference observed in MAE values.

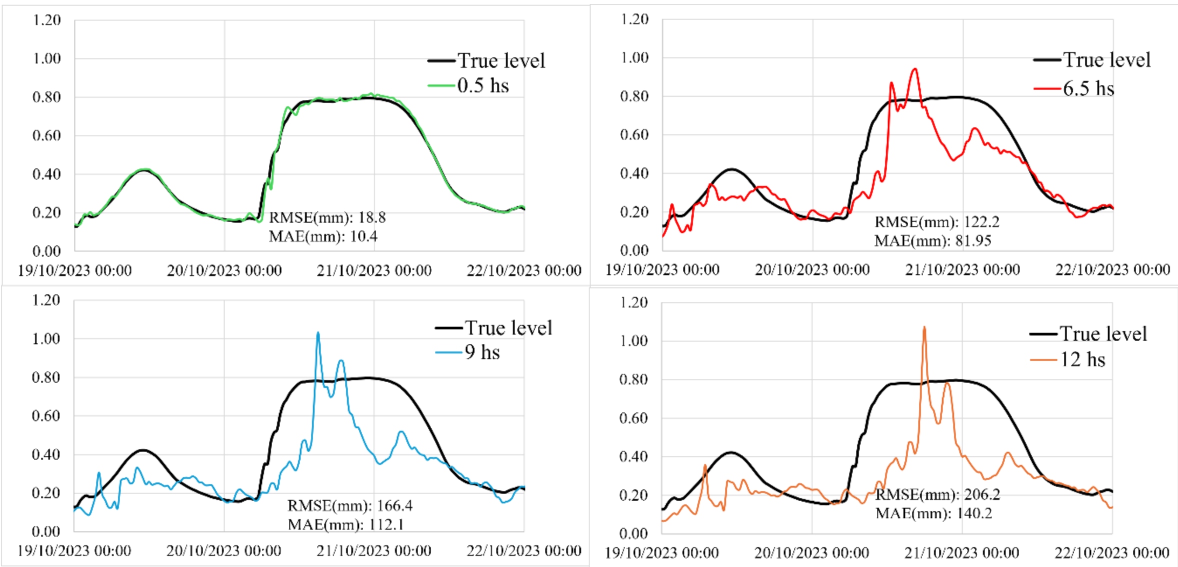


Figure 5-13 - LSTM performance for different lead times in predictions of the October 2023 event with IMERG-Cranford data.

The model also succeed in replicating the rising and receding stages of the water level with the IMERG-Cranford data, with a lightly inclination towards overestimation, which favours safety risk estimation.

In contrast, Figure 5.14 show that the predictions with the gauge data display a severe underestimation of values at increasing lead times, particularly in the first stage of the event.

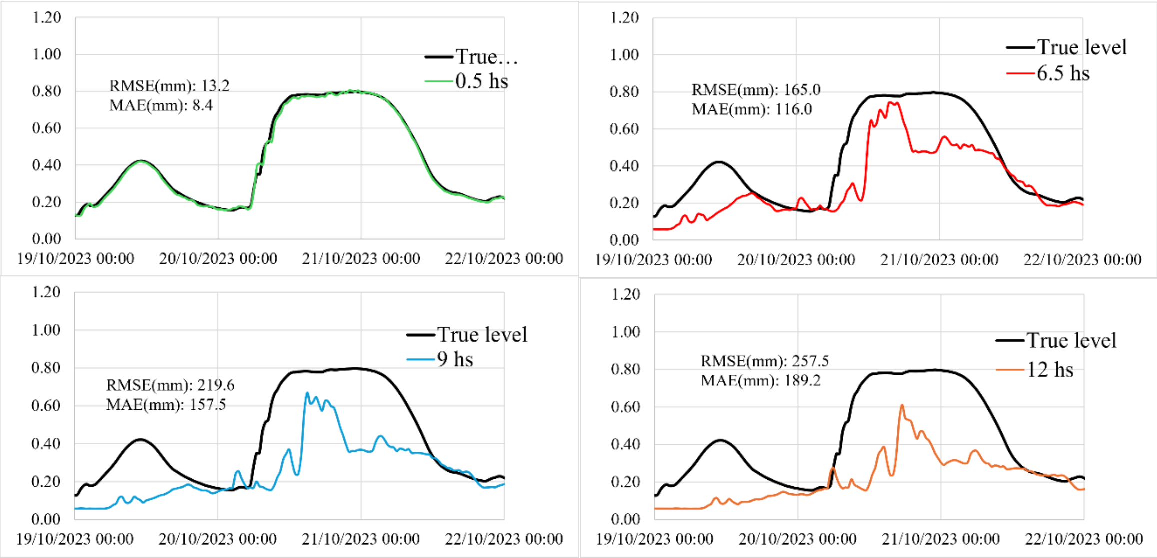


Figure 5-14 - LSTM performance for different lead times in predictions of the October 2023 event with Northolt gauge data

Considering the predictions of the three events with the IMERG-Cranford data, the results indicate a more consistence performance of the model with the IMERG-Cranford data providing best overall predictions particularly for greater lead times. More importantly, while the model's performed pattern was kept while in contrast the accuracy of the predictions with rain-gauge data seems to decrease as the severity of the event increases. The superior performance of the model with the IMERG-Cranford data in comparison with the data from the Northolt gauge, is likely to be related to the large precipitation area represented in the IMERG unit dataset (10 km x 10 km). Whereas the gauge captures a punctual information that only represents the surrounding precipitation and loses representativeness as the distance from the gauge increases (Kidd et al., 2017).

However, this apparent advantage of the local IMERG data over the local gauge data for stream level predictions, was detected based on the historical datasets and does not account for the latency of the IMERG products which hinder its application in real-time. Nevertheless, it indicates that the properties and quality the IMERG datasets are adequate and valuable for stream-level predictions with the LSTM model.

The next section show the results obtained from this study’s attempt to overcome the issue that the IMERG data latency represents for real-time applications, by exploring the application of data from areas of the IMERG grid distant from the catchment.

5.3 LSTM model performance with precipitation data from remote IMERG pixels

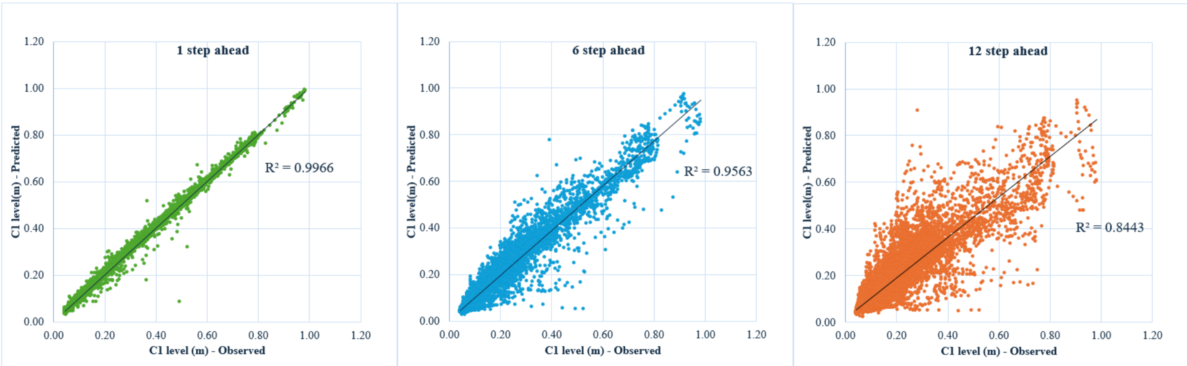
This section presents the LSTM model's performance in predicting water levels at Cranford using precipitation data collected from the selected remote area of IMERG pixels over the Atlantic Ocean, as detailed in Section 4.3.

First, the results in Table 5.7 show the metrics of the model’s prediction using data from two groups of pixels in the arrangement as stablished in Table 4.3, Section 4.3.

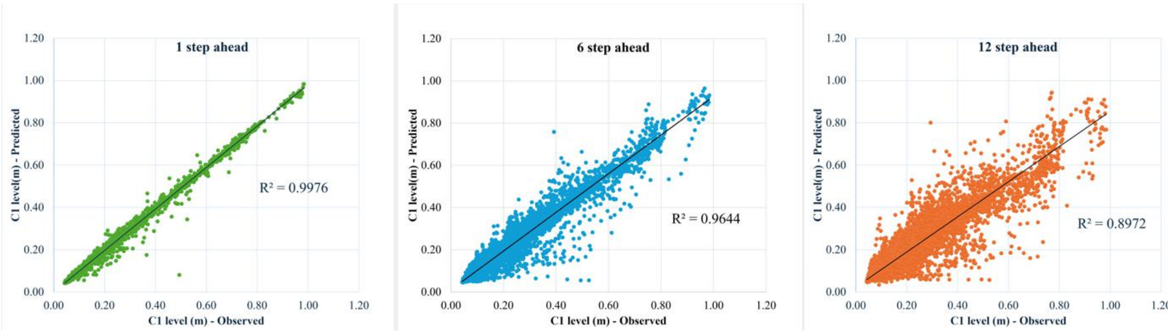
Table 5.7 - Performance metrics of LSTM model predictions using data from five IMERG pixels arranged horizontally or vertically in the Atlantic monitoring region.

RMSE (mm)		***MAE (mm)***		
Horizontal pixel arrangement	**Vertical pixel arrangement**	**Horizontal pixel arrangement**	**Vertical pixel arrangement**	***Steps-ahead***
5.99	**5.73**	2.78	**2.48**	*1*
7.89	9.39	3.78	**3.17**	*2*
10.68	**9.61**	4.94	**4.12**	*3*
13.86	**12.39**	6.19	**5.38**	*4*
17.34	**15.35**	7.44	**6.83**	*5*
20.73	**18.34**	8.61	**8.17**	*6*
24.20	**21.25**	9.89	**9.59**	*7*
27.66	**24.18**	11.19	**10.96**	*8*
30.98	**26.92**	12.52	**12.14**	*9*
34.04	**29.54**	13.72	**13.38**	*10*
36.76	**32.06**	14.69	**14.65**	*11*
39.28	**34.29**	**15.72**	15.88	*12*
22.45	***19.92***	***9.29***	***8.90***	***Overall***

The performance metrics of predictions reveal a ‘very good’ performance of the model with both pixel’s arrangements. However, the consistently smaller errors across all lead times indicate that the LSTM model captures the influence of precipitation on stream level more effectively using data from pixels in the vertical arrangement.



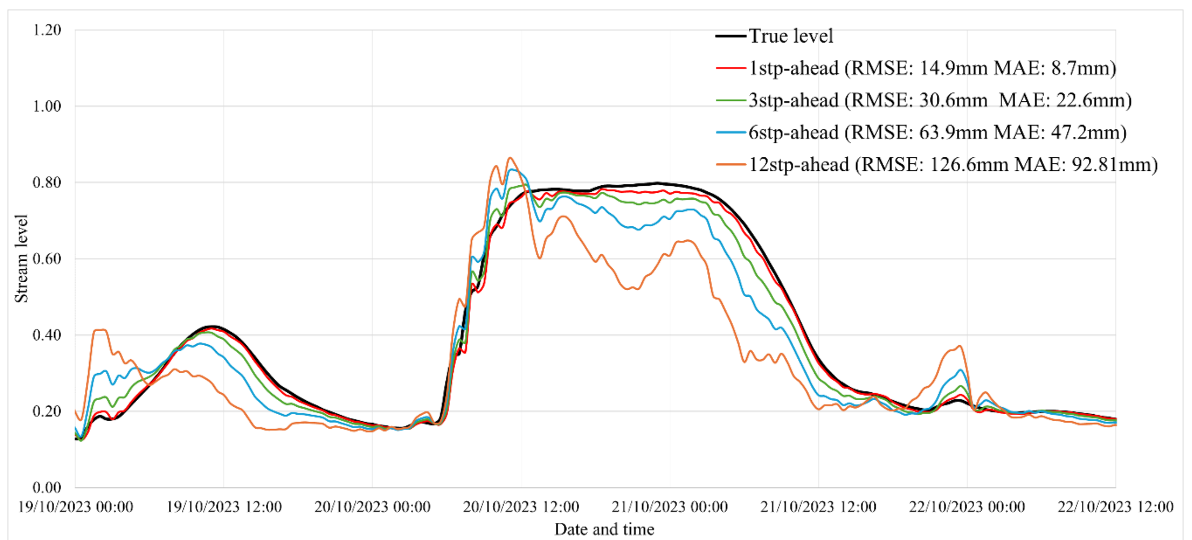
(a)



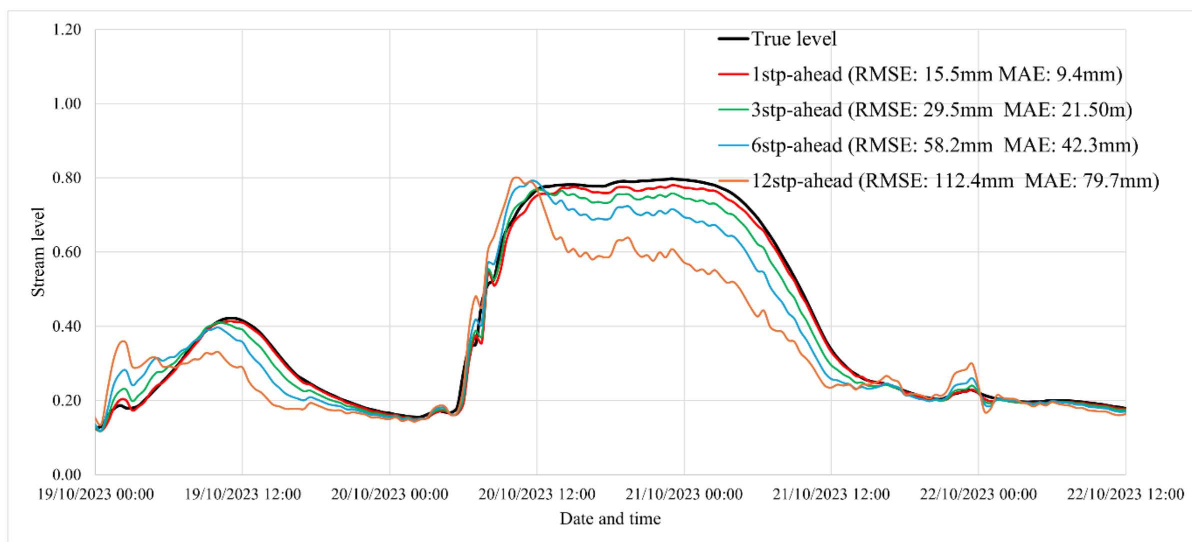
(b)

Figure 5-15 - R-squared values for step-ahead predictions of the LSTM model using data from five adjacent IMERG pixels aligned horizontally (a) and vertically (b)

This is confirmed in Figure 5.15, showing the variance plots of predicted and true level values with respective R-square values, and reinforced on the performances on event-based predictions for October 2023, as shown in Figure 5.16.



(a)



(b)

***Figure 5-16** - Event-based prediction of stream level with LSTM model using data from five adjacent IMERG pixels aligned horizontally (a) and vertically (b)*

Furthermore, a comparison of the metrics in Figure 5.16 with those in Table 5.6 shows that the LSTM model performed better in predicting the October 2023 event using data from both sets of remote pixels than it did using data from the IMERG-Cranford pixel or the Northolt gauge station. This is a promising outcome for the research, as the improved performance was achieved using data from pixels located over 650 km away from the catchment. This result may be attributed to the larger area of rainfall represented by the group of five pixels, covering approximately 500 km², or possibly due to the slightly higher average correlation between these pixels and the stream level at Cranford.

Regardless of the reason, the vertical arrangement appears to be the most suitable for modelling and was selected for testing the remaining high-correlation pixels within the focus area of the IMERG data for C1.

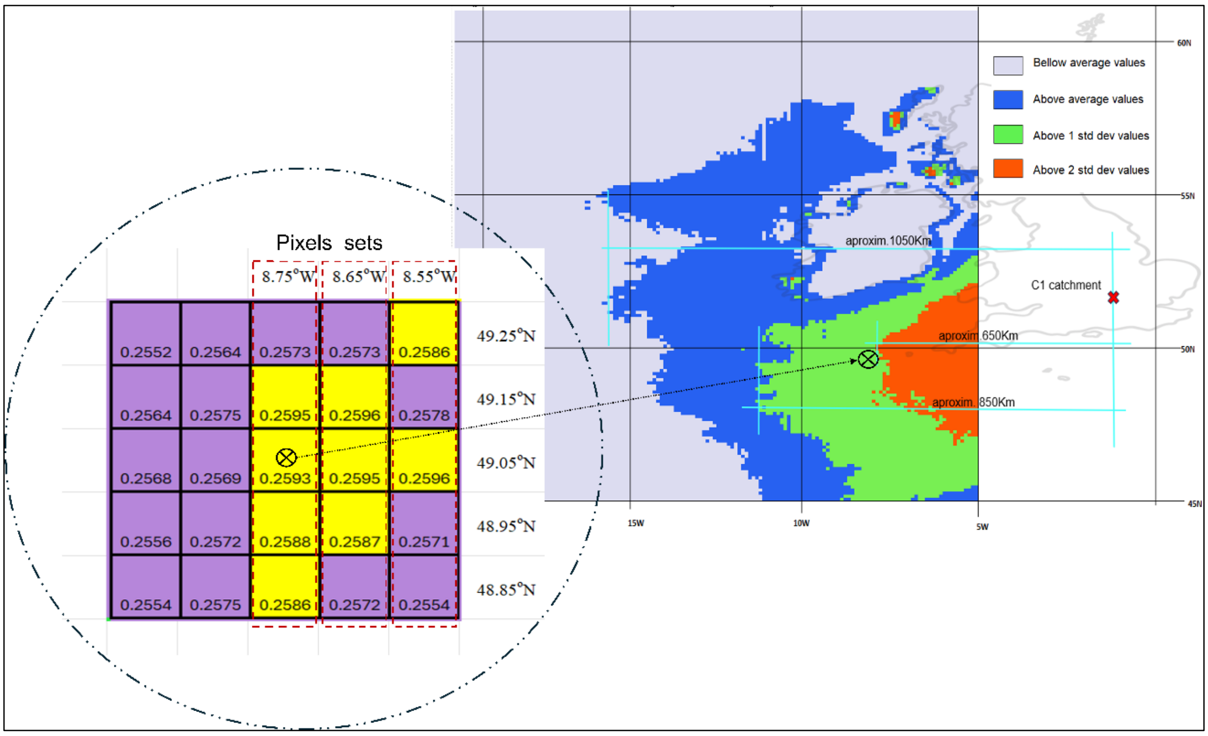


Figure 5-17 - Pixels-group 8.75, 8.65 and 8.55 and approximated position in the study area.

The final LSTM models were trained to predict stream level at C1 using precipitation estimates retrieved from three groups of five IMERG pixels each. These groups are detailed in Figure 5.17 and identified as pixels-group 8.75, 8.65, and 8.55, based on the common coordinate of the pixels in the vertical arrangement.

Table 5.8 presents the performance metrics of the LSTM models using data from the three groups of pixels, which confirm that the models were well-trained and capable of accurately predicting stream levels in the testing subset. This is supported by a similar and consistent performance with the three datasets, rated as ‘very good’ and ‘good’ up to 14 and 17 hours lead time, respectively.

Table 5.8 - LSTM model performance metrics - data from three groups of remote IMERG pixels. Rating thresholds: Very good (green), Good (blue), Satisfactory (orange).

Minimum Lead-time (hours)	*Steps ahead*	*RMSE (mm)*			*MAE (mm)*		
		8.75	**8.65**	**8.55**	**8.75**	**8.65**	**8.55**
6.5	*1*	6.41	6.42	6.96	2.97	2.81	3.40
7.0	*2*	9.89	9.89	10.27	3.70	3.81	4.13
7.5	*3*	13.15	13.01	13.39	4.83	4.97	4.98
8.0	*4*	16.14	16.13	16.63	6.07	6.16	6.08
8.5	*5*	19.16	19.29	19.70	7.26	7.32	7.13
9.0	*6*	22.11	22.34	22.68	8.41	8.48	8.19
9.5	*7*	25.06	25.44	25.71	9.41	9.67	9.39
10.0	*8*	28.00	28.52	28.96	10.41	10.88	10.73
10.5	*9*	31.196	31.68	32.30	11.67	12.14	12.13
11.0	*10*	34.37	34.75	35.47	12.88	13.41	13.47
11.5	*11*	37.54	37.80	38.46	14.13	14.76	14.79
12.0	*12*	40.66	40.81	41.48	15.38	16.06	16.13
12.5	*13*	43.78	43.75	44.59	16.67	17.34	17.50
13.0	*14*	46.59	46.59	47.68	17.85	18.56	19.07
13.5	*15*	49.27	49.34	50.47	18.97	19.87	20.42
14.0	*16*	51.94	52.07	53.13	20.08	21.13	21.71
14.5	*17*	55.92	54.72	55.77	21.84	22.35	22.96
15.0	*18*	56.92	57.28	58.41	22.35	23.55	24.19
15.5	*19*	59.19	59.73	60.94	23.49	24.75	25.20
16.0	*20*	61.39	62.02	63.41	24.62	25.74	26.33
16.5	*21*	63.47	64.15	65.79	25.75	26.69	27.48
17.0	*22*	65.42	66.12	67.94	26.79	27.62	28.63
17.5	*23*	67.20	67.92	69.93	27.81	28.51	29.91
18.0	*24*	68.89	69.61	71.73	28.85	29.30	31.00
Overall		***40.57***	***40.81***	***41.74***	***15.92***	***16.50***	***16.87***

This level of performance and consistency is confirmed in the variance plots of predictions and target values, as shown in Figure 5.18.

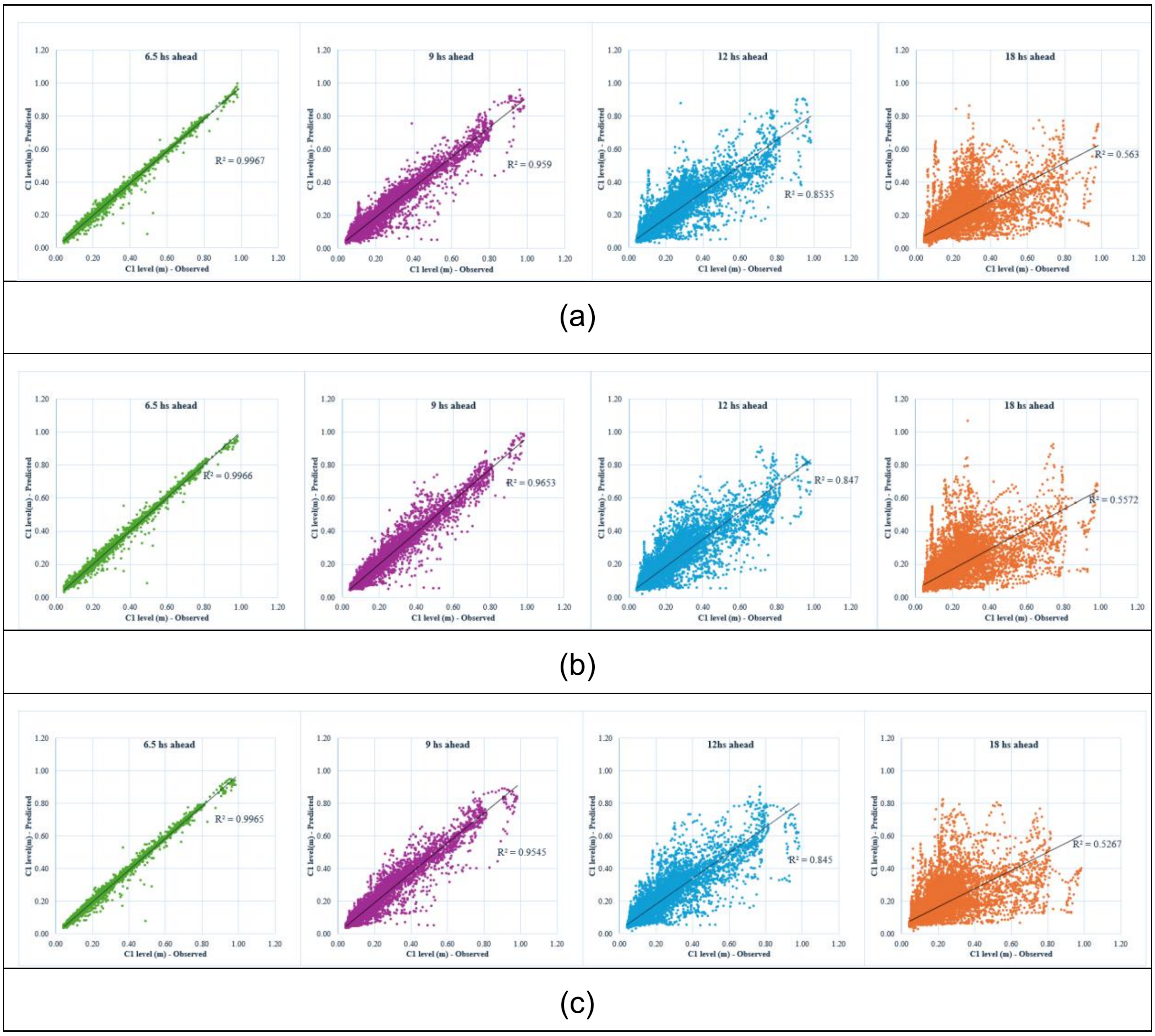


Figure 5-18 - R-squared values for step-ahead predictions of the LSTM model using data from the three sets of five adjacent pixels selected for the final tests on predictions with the IMERG precipitation data: 8.75(a), 8.65(b) and 8.55(c)

However, the LSTM model was able to get closer to the true level values at Cranford in predictions using data from the pixels group 8.75, which is the farthest from the C1 catchment among the three selected groups. These outcomes confirm the relevance and adequacy of the method applied for optimizing the IMERG data in the previous section (4.3), focusing on the historical relationship between the datasets rather than just its spatial position.

The most significant differences between the performance with each of the pixel's group is noted in event-based predictions during the rainfall event of October 2022 seen together in Figure 5.19.

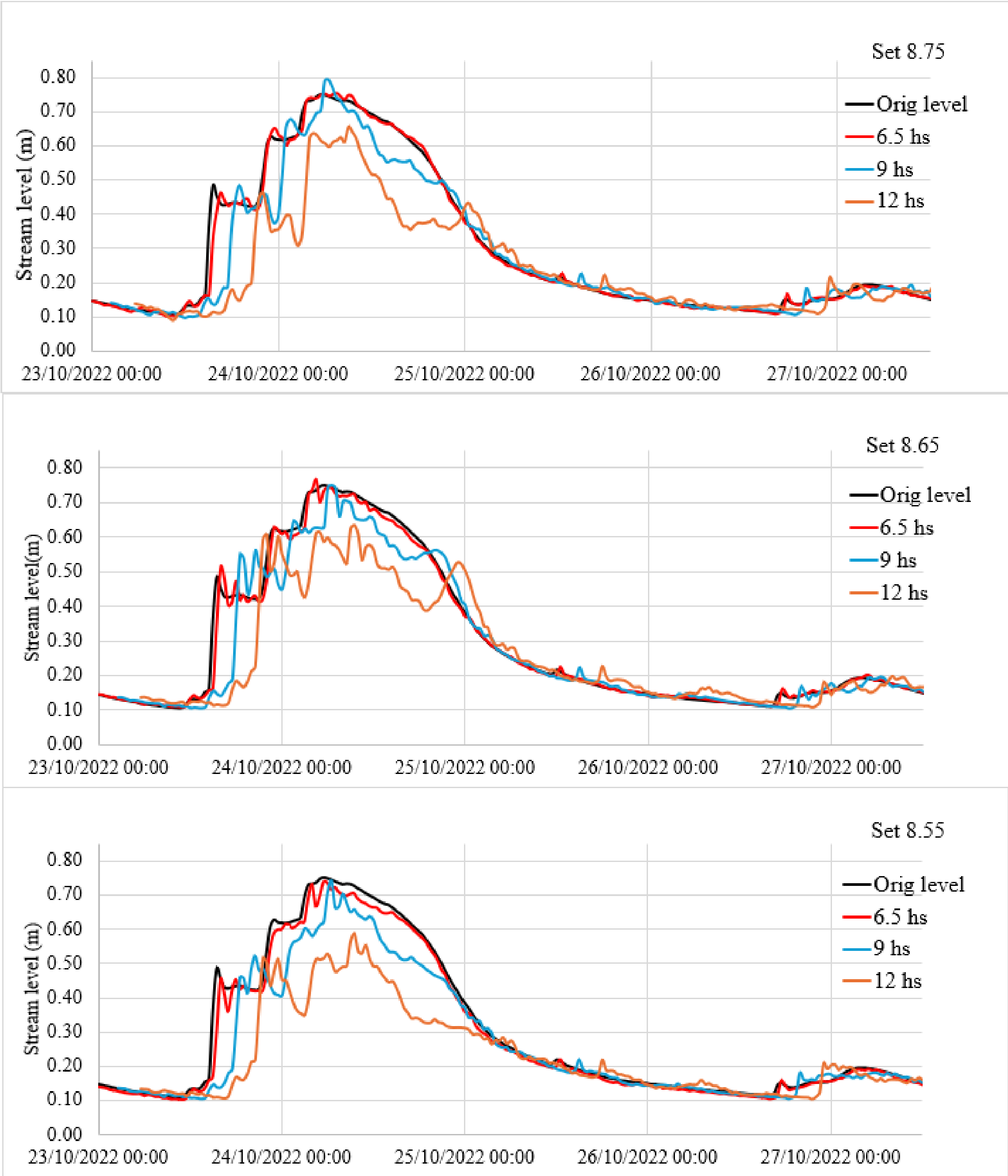


Figure 5-19 - True level at Cranford and LSTM model predictions at multiple lead times using data from pixel groups 8.75, 8.65, and 8.55 during the October 2022 rainfall event

For a better appraisal of the performances, Figure 5.20 shows a comparison of the curves and metrics performance of each IMERG pixels-group on different lead times.

On comparing these plots and metrics there is no clear winner between the groups, all rated as 'very good' performance on predictions at any lead times.

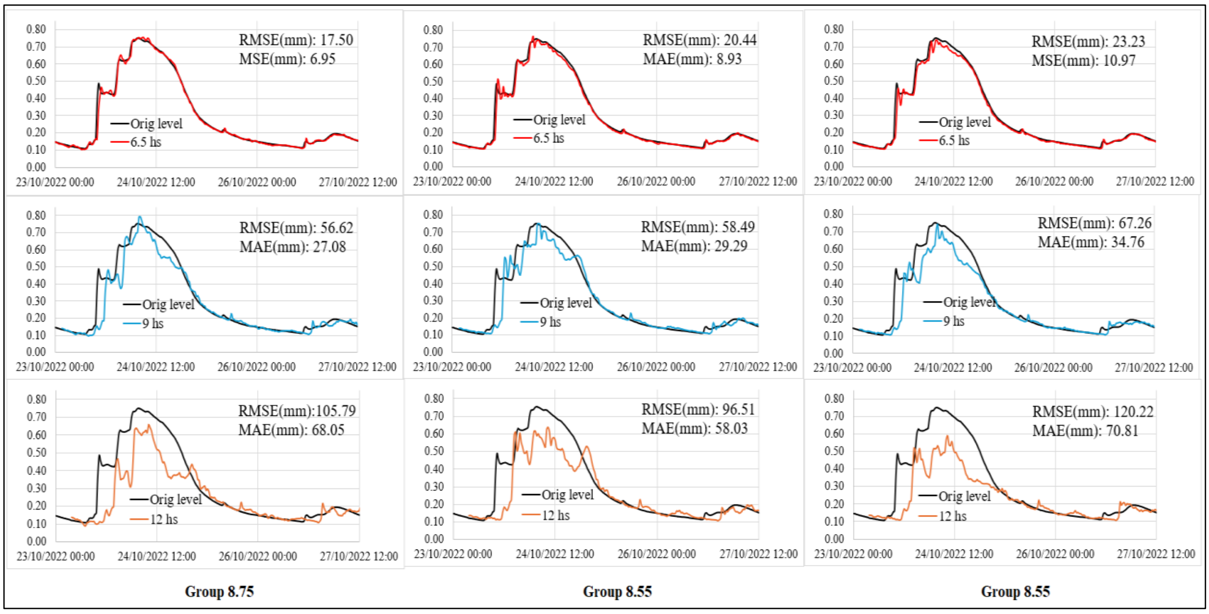


Figure 5-20 - LSTM model predictions against true level at multiple lead times with data from pixel groups 8.75, 8.65, and 8.55 during the October 2022 rainfall event

Looking back the results presented in section 5.2 , it is clear that predicting the C1 catchment's response to the October 2022 rainfall event was particularly challenging. In comparison, the LSTM models using data from the remote pixel groups produced predictions with only slightly higher errors than those using local rainfall data. For example, the model using pixel group 8.75 showed an average increase of 20 mm in error across the 24 time-steps for both RMSE and MAE metrics. Given the additional 6 hours of lead time gained from the spatial distance between the pixels and the catchment, this error increase is deemed acceptable.

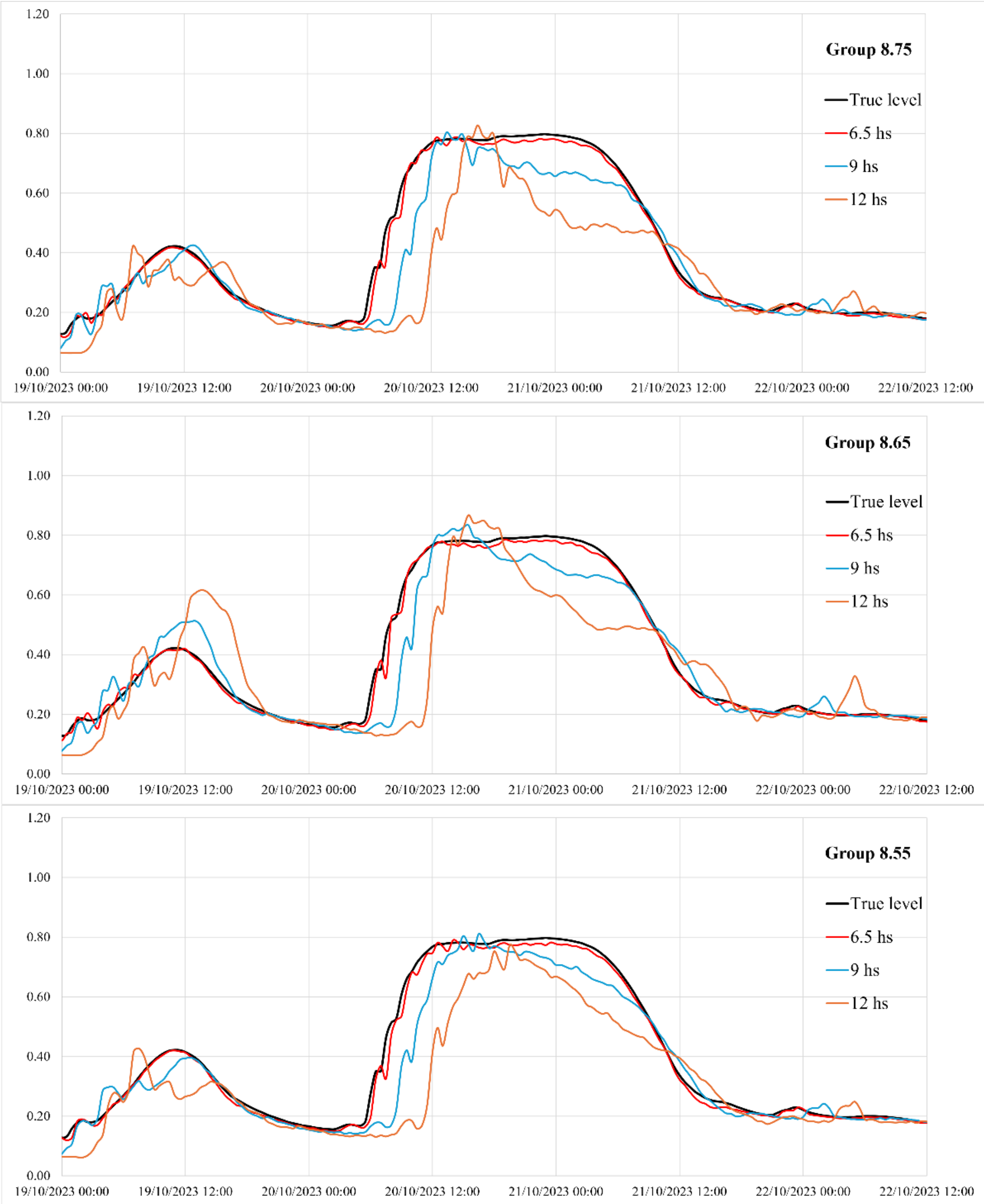


Figure 5-21 - True level at Cranford and LSTM model predictions at multiple lead times using data from pixel groups 8.75, 8.65, and 8.55 during the October 2023 rainfall event.

Moving to the October 2023 event, similar performance patterns are observed in the LSTM model predictions with data from the three pixels groups. As shown in Figure 5.21, the model performs slightly better with that data from groups 8.75 and 8.65 groups, demonstrating a better assimilation of the relationship between rainfall and stream level compared to the 8.55 group, despite the latter being closer to the catchment.

Focusing on the predictions with particular lead times, those shown in Figure 5.22 are significantly better that the ones for October 2022 in Figure 5.20, which is in line with the previous performances observed for both events with models using local rainfall data. When comparing the predictions for the same event using local rainfall data (Northolt gauge and IMERG-Cranford pixel), errors on predictions for the October 2023 event with data from the pixels-group 8.75 and 8.65 are significant smaller. For prediction up the 12 hours lead times the predictions still very close or within the 'good' performance threshold.

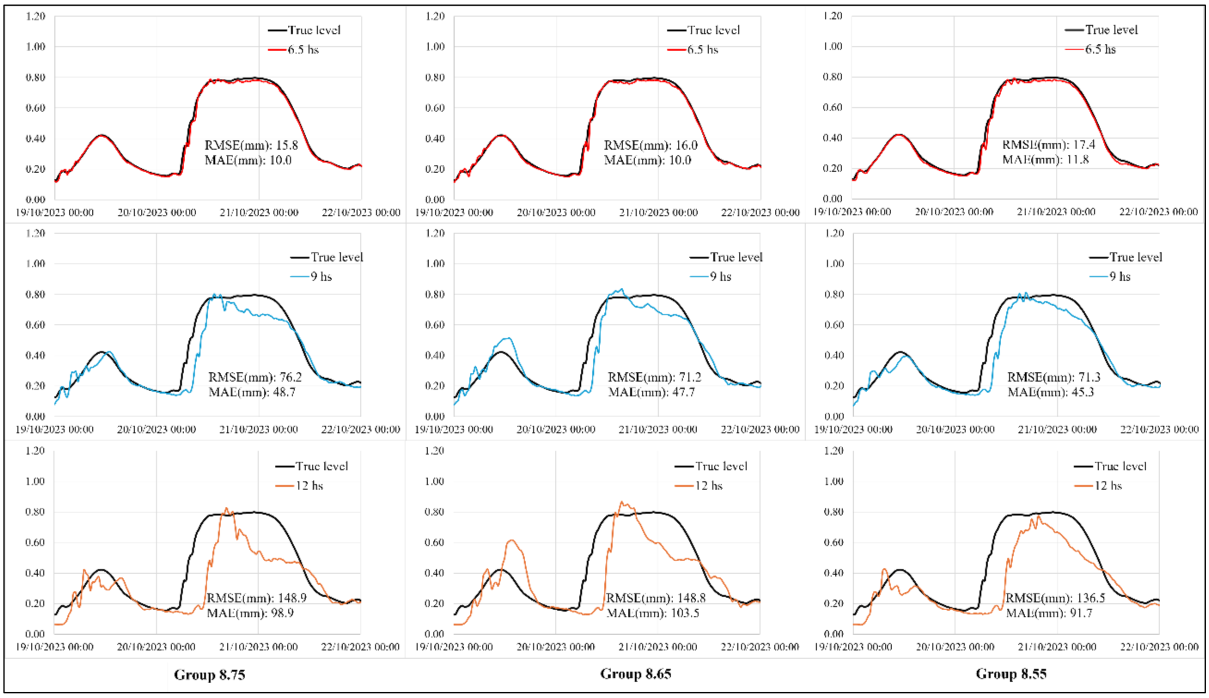
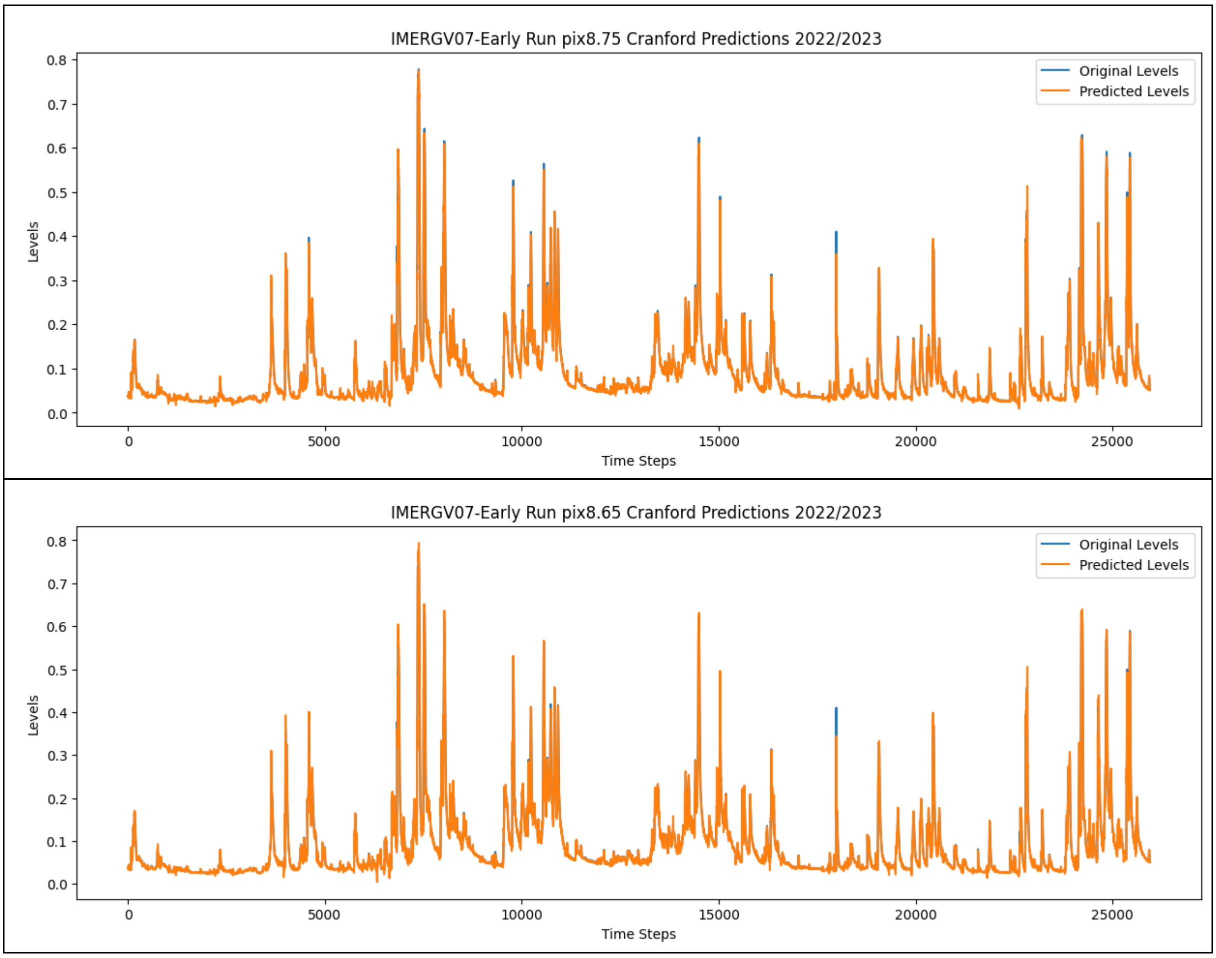


Figure 5-22 - LSTM model predictions against true level at multiple lead times with data from pixel groups 8.75, 8.65, and 8.55 during the October 2023 rainfall event.

Up to this point, the application of the IMERG data for stream level predictions has been very successful, as lead times beyond 6 hours were achieved with high accuracy. Yet, these performances were achieved the IMERG Final Run product replaced by the IMERG Early Run product in the test of the model for real-time predictions.

5.4 LSTM model performance with precipitation data from remote pixels of the NRT IMERG data

The results of the most revolutionary and impacting aspects of this research are presented in this section. While the previous sections built the foundations that support the IMERG data application in stream level predictions, this section presents the performance of the IMERG data in real time predictions. The LSTM models previously trained with the historical data of the IMERG Final Run product, use the near real time (NRT) precipitation data from the IMERG Early Run product for predictions.



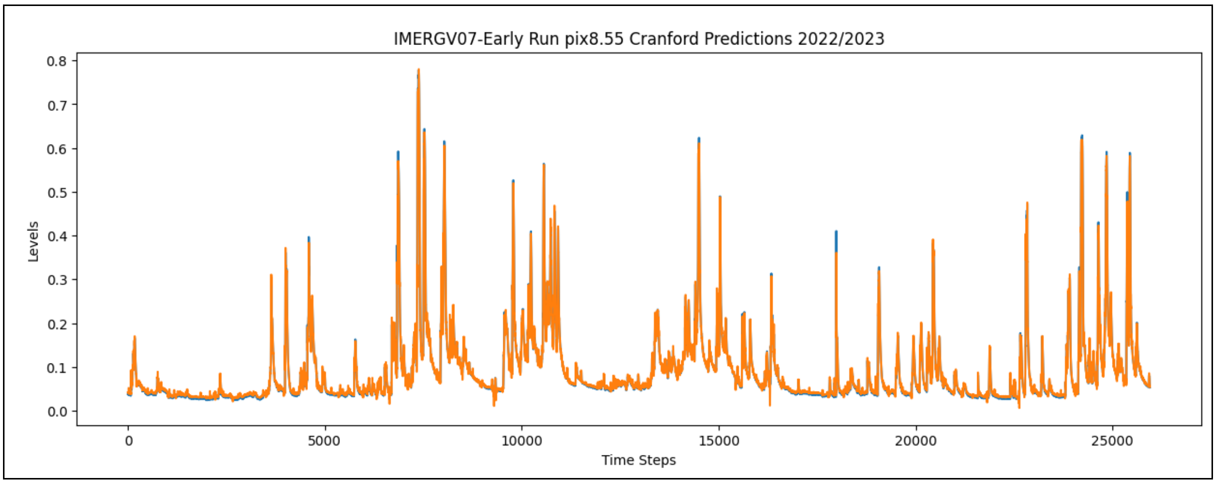


Figure 5.23 – LSTM level predictions with NRT data from pixels-group 8.75, 8.65 and 8.55.

The overall predictions of the LSTM model using data from pixel groups 8.75, 8.65, and 8.55 of the NRT dataset are shown in the plots in Figure 5.23, revealing the model' small tendency to overestimate with data from pixel group 8.55 and underestimate with data from pixel group 8.75.

Table 5.9 - Performance metrics for LSTM model on real-time predictions of stream level at Cranford with the IMERG Early Run product data. Rating thresholds: Very good (green), Good (blue), Satisfactory (orange).

Minimum lead-time (hours)	*RMSE (mm)*			*MAE (mm)*		
	8.75	**8.65**	**8.55**	**8.75**	**8.65**	**8.55**
6.5	7.31	6.92	8.55	3.20	3.11	4.98
7.0	10.91	10.57	11.47	3.98	3.94	5.57
7.5	14.35	14.00	14.63	5.22	5.07	6.67
8.0	17.60	17.40	18.08	6.53	6.45	8.04
8.5	20.90	20.98	21.91	7.91	8.11	9.53
9.0	24.15	24.65	25.67	9.20	9.96	10.90
9.5	27.40	28.21	29.21	10.36	11.56	12.24
10.0	30.64	31.81	32.78	11.52	13.11	13.69
10.5	34.18	35.56	36.34	12.94	14.85	15.28
11.0	37.69	39.20	39.64	14.35	16.60	16.62
11.5	41.21	43.04	42.74	15.80	18.26	17.87
12.0	44.70	46.76	45.87	17.27	19.86	19.23
12.5	48.16	50.57	48.91	18.77	21.39	20.47
13.0	51.34	54.23	52.04	20.13	23.01	21.86
13.5	54.34	57.70	55.14	21.44	24.59	23.07
14.0	57.32	60.93	58.21	22.75	26.11	24.40
14.5	60.20	64.05	61.34	24.08	27.72	25.73
15.0	62.90	67.18	64.27	25.41	29.17	27.06

15.5	65.39	70.22	67.11	26.70	30.60	28.46
16.0	67.77	73.07	69.67	27.96	31.97	29.88
16.5	70.03	75.70	72.05	29.22	33.13	31.27
17.0	72.15	78.07	74.29	30.40	34.27	32.65
17.5	74.06	80.21	76.31	31.54	35.43	33.95
18.0	75.85	82.15	78.06	32.70	36.48	35.03
Overall	***44.61***	***47.22***	***46.01***	***17.89***	***20.20***	***19.77***

The model's performance metrics in Table 5.9 indicates that the 8.55 pixel-group is no longer the worst performer, although the 8.75 pixel-group remains the best. More importantly, the performance on predictions is considered 'very good' on lead times up to 13.5 hours, 'good' on lead times up to 16 hours and still satisfactory on lead times of 18 hours. This demonstrates that the model fits the testing data well, despite being trained with another IMERG product.

Next, Figure 5.23 illustrates the event-based performance for October 2022 using NRT IMERG data. During this event the water level at C1 rose by approximately 600 mm in 13 hours, nearly five times the pre-event level as previously shown in Table 4.6

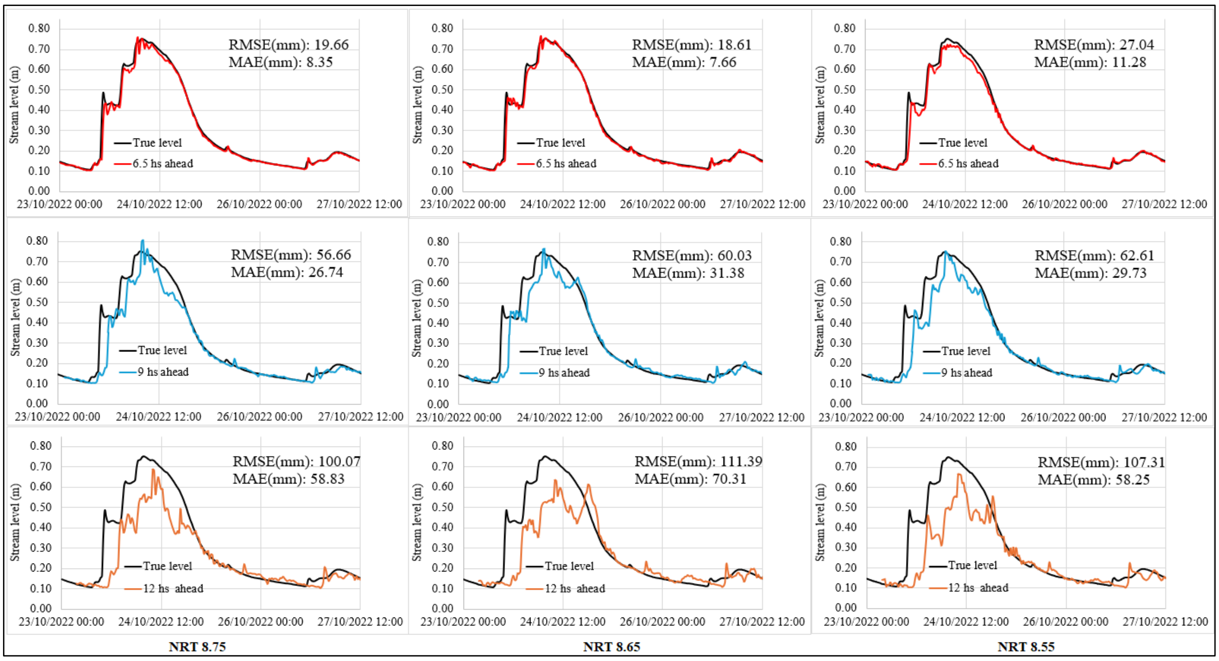


Figure 5-23 - LSTM model predictions against true level at multiple lead times with NRT IMERG data from pixels-group 8.75, 8.65, and 8.55 during the October 2022 rainfall event

Notably, the model accurately predicted both the timing and magnitude of the water level peak up to 9 hours in advance, regardless of the pixels' group used. This was also reflected in the RSR values rating the model's performance as 'very good' for all lead times and being particularly low on predictions up 9hs ahead, with RSR values that are less than 0.3 of the true level observations.

Similarly, during the October 2023 event, the graphs in Figure 5.24 demonstrate that the model had acceptable accuracy and correctly estimated the timing and magnitude of the water level peak on lead times up to 12 hours. The model also maintained a 'very good' performance on all predictions of the event up to 9hs lead times.

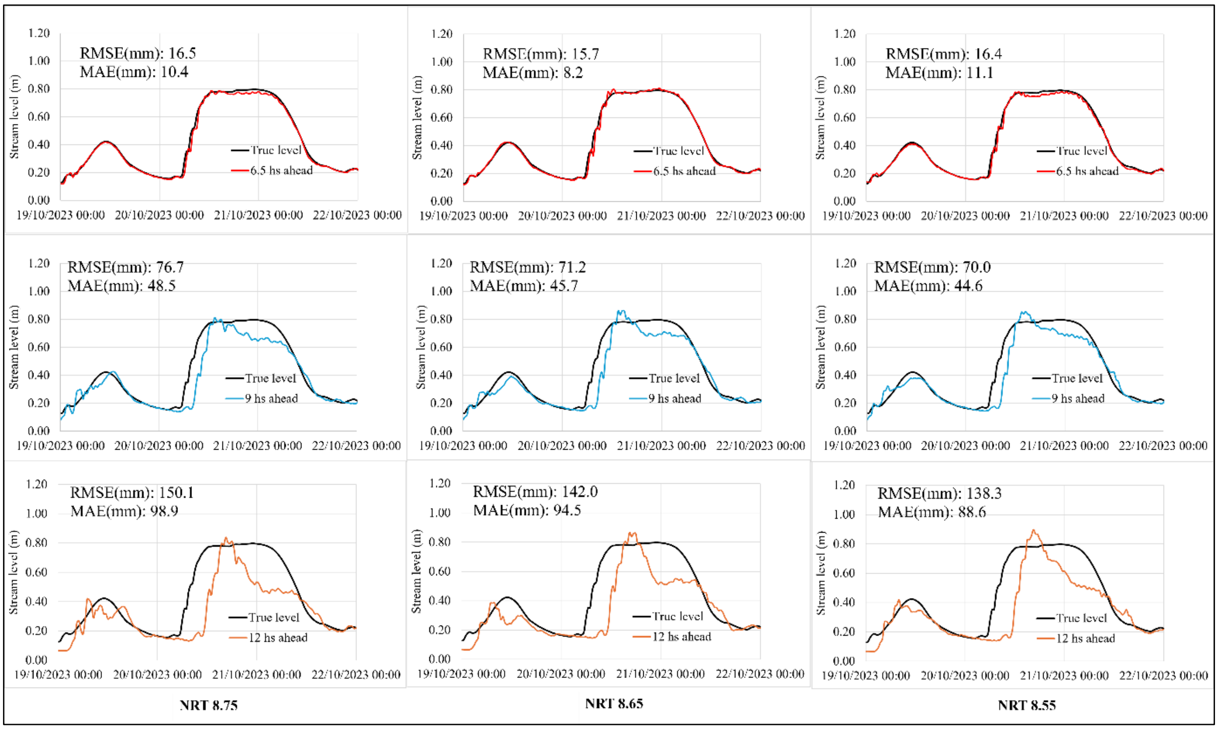


Figure 5-24 - LSTM model predictions against true level at multiple lead times with data from pixels-group 8.75, 8.65, and 8.55 during the October 2023 rainfall event.

Next, the average performances of the LSTM model with each of the datasets applied in this study are summarised in Table 10. This summary reveals that, in the overall and in the October 2023 event's predictions, the LSTM model performed slightly better with data from the IMERG pixels groups and at similar level with all remaining datasets including the NRT data.

Table 5.10 - Average performance metrics of the LSTM models using several data sources on predictions of stream level up to 24 time-steps ahead

Data source		Metrics	Overall	October 2022	October 2023
Local rainfall data	Northolt RAF rain-gauge	RMSE	43.3	80.1	148.2
		MAE	26.6	43.1	105.0
	IMERG-Cranford pixel	RMSE	43.9	81.7	114.8
		MAE	27.7	41.7	77.0
IMERG V07 Final Run product	pixels-group 8.75	RMSE	40.6	105.6	141.9
		MAE	15.9	62.3	96.4
	pixels-group 8.65	RMSE	40.8	102.2	145.8
		MAE	16.5	61.56	104.5
	pixels-group 8.55	RMSE	41.7	121.9	128.4
		MAE	16.9	71.70	92.0
IMERG V07 Early Run product	NRT pixels-group 8.75	RMSE	44.6	100.3	143.0
		MAE	17.9	61.6	96.2
	NRT pixels-group 8.65	RMSE	47.2	116.9	138.7
		MAE	20.2	72.9	95.8
	NRT pixels-group 8.55	RMSE	46.0	112.8	134.2
		MAE	19.8	62.1	87.8

The model's worst average performance with the NRT IMERG data is observed in the predictions for the October 2022 event. However, these still within the threshold considered as a 'very good' model performance (RSR < 0.5).

Furthermore, considering the LSTM model's plots of the errors' distribution for each dataset in Figure 5.25, the performance level can be considered equivalent in all predictions either with the historical or the NRT data from the remote IMERG pixels. This is an impressive result given that the IMERG Early Run is a simpler product in comparison with the IMERG Final Run product used in the model's training.

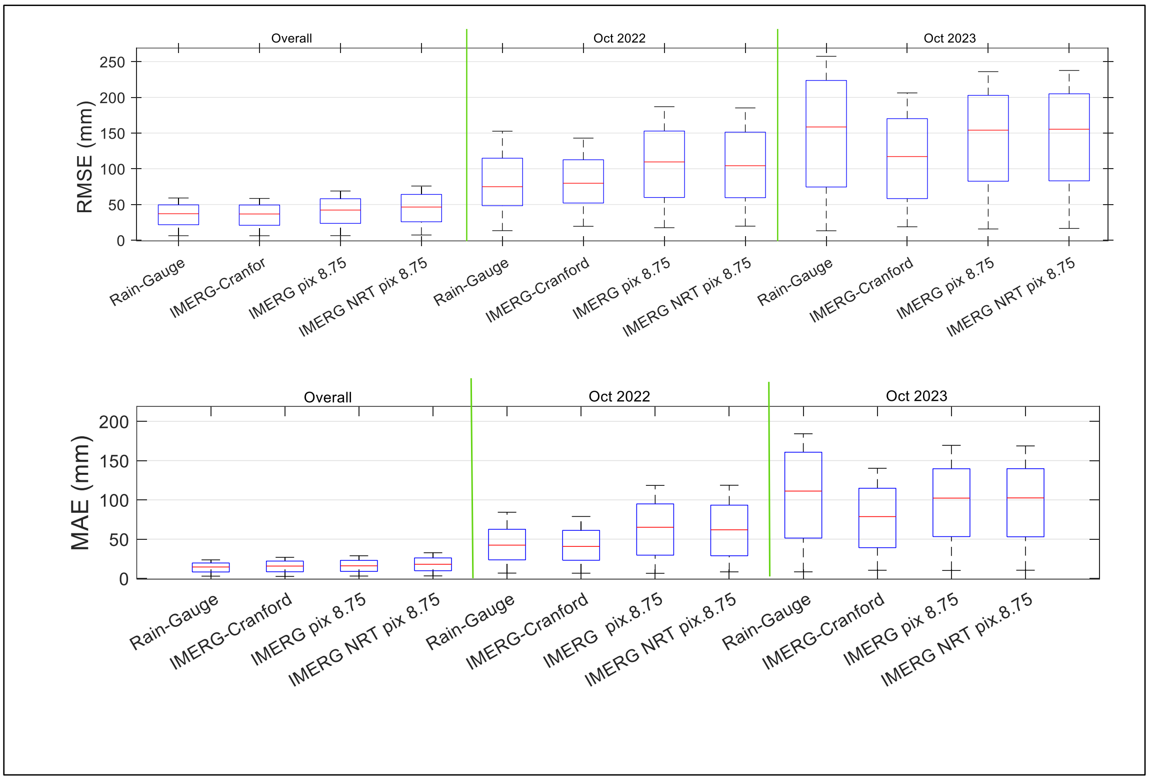


Figure 5-25 – LSTM model errors distribution for overall and event-based predictions.

The significance of these results can be assessed by comparing them with the outcomes of a study conducted in the River Pinn catchment, which is very similar to the River Crane and located only 10 km away from Cranford. The study, conducted by Piadeh et al. (2023a), aimed for real-time flooding forecasts using event-based data and produced predictions with 2 to 3 hours lead time and RMSE values between 63 mm and 100 mm. In contrast, the present study provided forecasts for the Cranford site with 9 hours lead time and RMSE values around 70 mm, which is a great achievement.

Finally, the model's ability to multi-step forecasting can be analysed from the changing rates observed in the RMSE values along the 24-time steps displayed in Figure 5.26.

shows the rates on how much RMSE values increase or decrease between predictions at consecutive steps ahead. The plots show a pronounced drop in the RMSE changing rates within few time steps ahead, and its stabilization at greater lead times demonstrating of the model's good ability for predicting at long range. Regarding the different model inputs, the plots representing the NRT data show a continuous and smooth curve that contrasts with the more volatile and abrupt changing rates in the curves representing the other datasets. This indicates a greater ability of the model for predictions with the NRT dataset in both rainfall events and can explains the higher accuracy of the model with the NRT IMERG data in comparison with the local data for predictions with greater lead times.

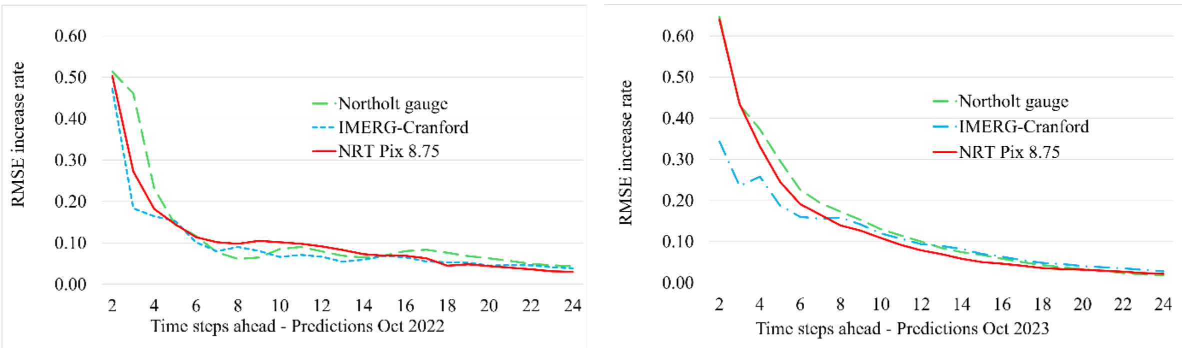


Figure 5-26 –RMSE values changing rates at increasing lead times of predictions during two rainfall events.

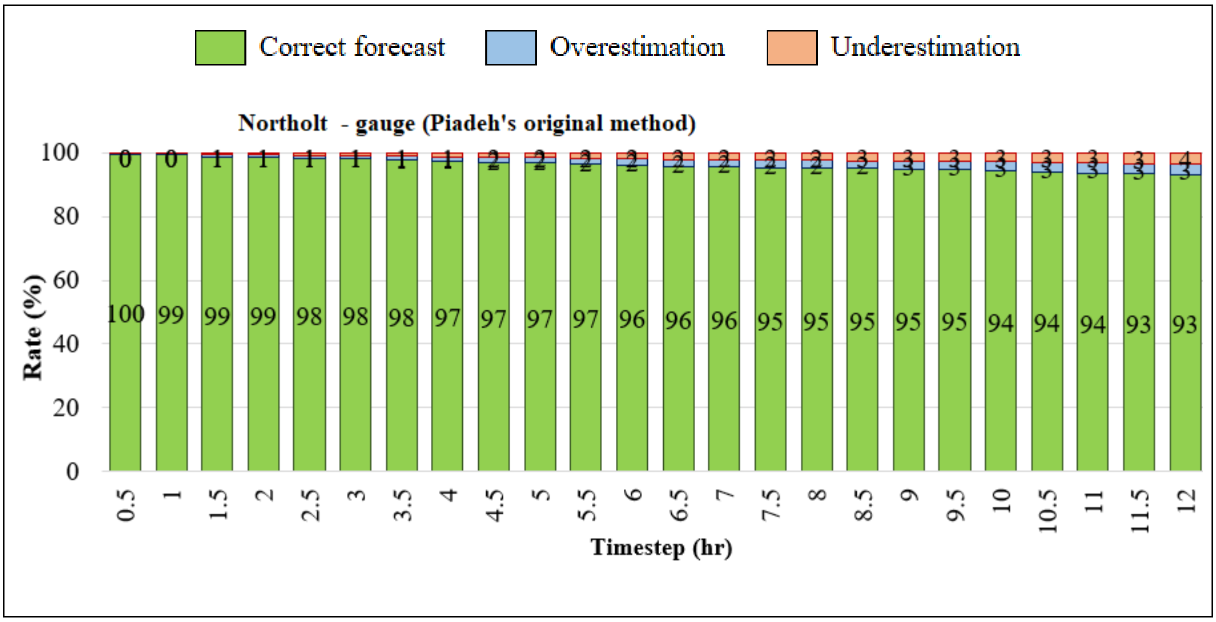
Yet, although the model demonstrated a good ability to accurately predict stream level several steps ahead, the is a significant difference between the event-based forecasts and the observed true level. The 12 hours forecasts were produced with errors in the range of 100mm and 150mm, proportionally minimal to the rise of stream level in each event but significant to distinguish between a flooding and non-flooding risk event.

Therefore, the predictions of the LSTM model were also tested as input in a flooding risk categorization tool that was originally designed to use predictions from a E-Narx model. The results of this test are presented in the next section.

5.5 Assessment of model and predictions with an alternative forecasting method

This section presents the performance assessment of the real-time predictions generated by this study, on its application in the event-based decision support model described at the end of section 4. The forecasts obtained with this new method are compared with the true level at Cranford and categorized as: TP, TN, FN, or FP following the definition presented in section 3.2.1.

As all models use the water level at Cranford as stream level dataset, the results presented are distinguished in the graphs only by the source of the rainfall data used in the method. The first graph in Figure 5.27 depicts the forecasts obtained with the original DS platform structure using rainfall data from the Northolt gauge, while the second graph show the forecasts from the version the DS platform adapted for this study, using the LSTM model in place of the E-NARX, with the same dataset.



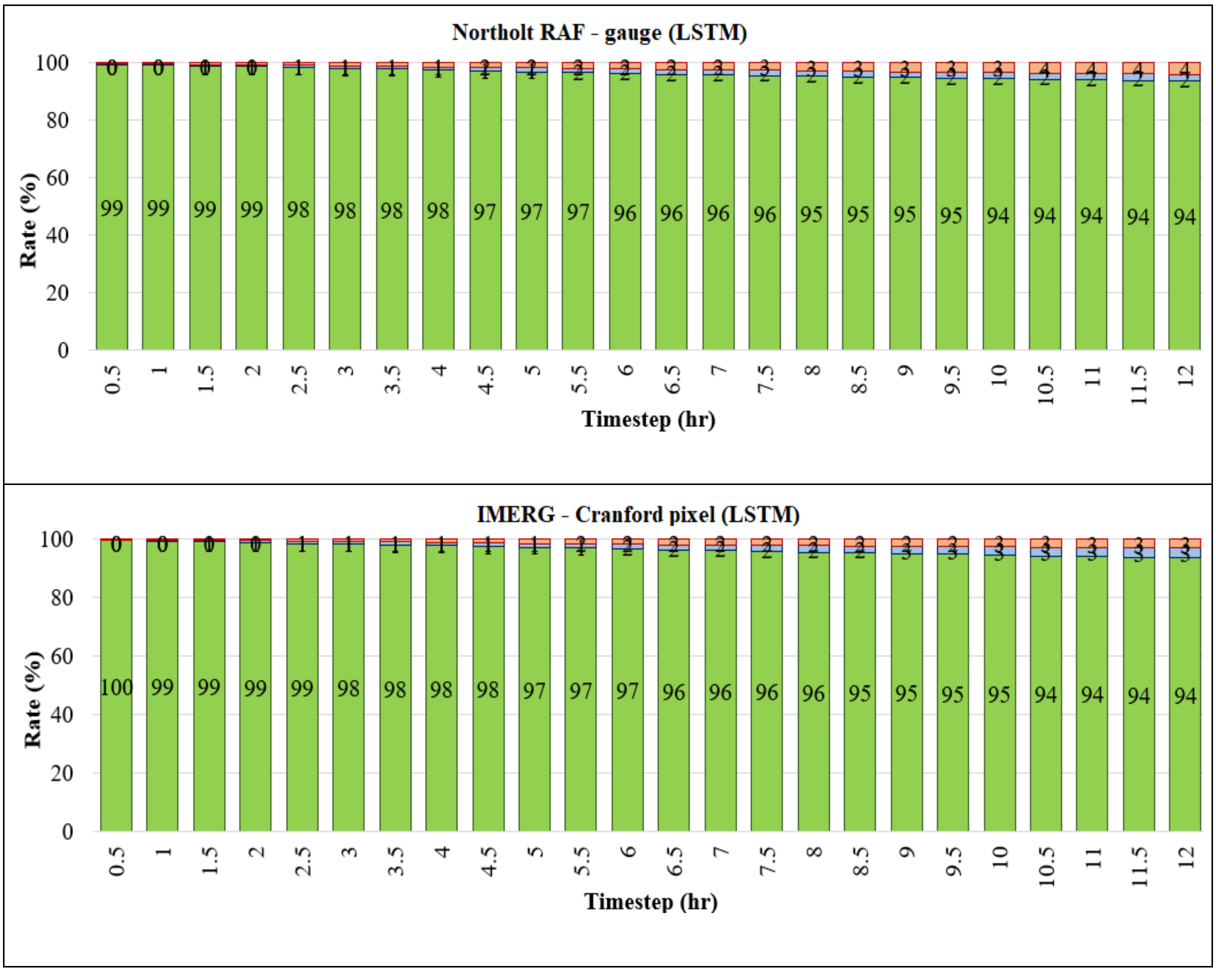
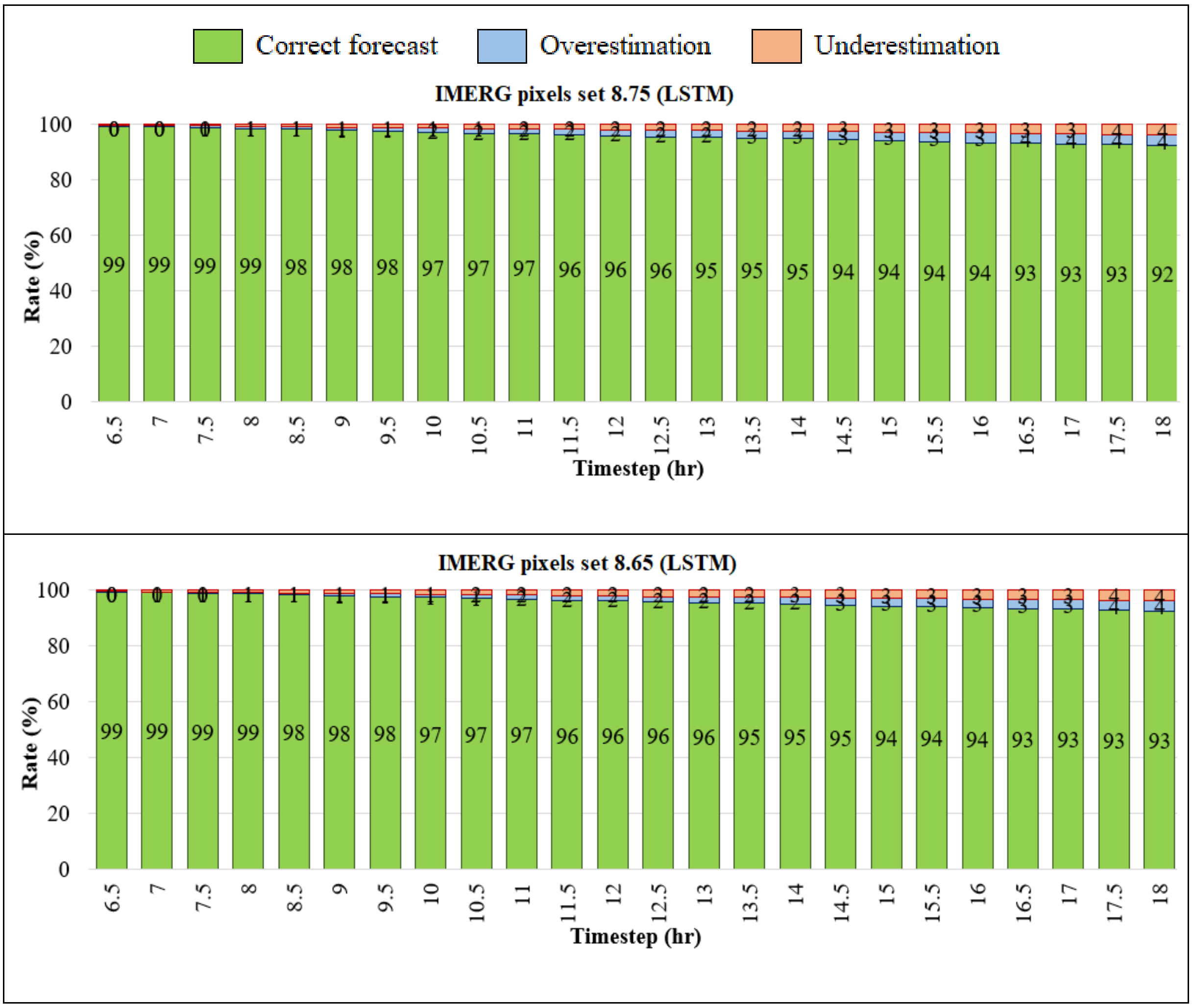


Figure 5-27 - Rates of targets hit, overestimation or underestimation with the original method and the adapted method using precipitation data from the local sources.

The graph's green bars reflect the percentage of outcomes forecasting a flooding or non-flooding event accurately, which can be interpreted as a measurement of the overall accuracy of method on classification (Ireland et al., 2015). Respectively, the orange and blue bars represent the missed targets due to overestimation or underestimation of forecasts. The third graph exhibit the forecasts of the adapted model with the IMERG-Cranford pixel precipitation data.

Therefore, the first two graphs in Figure 5.27 allow for a direct comparison between the performance of the original and adapted methods. The results show that, while both methods perform well initially, the adapted method outperforms the original for longer lead times. This trend continues when the Northolt gauge rainfall data is replaced with IMERG-Cranford pixel data, as shown in the third graph, demonstrating the effectiveness of the LSTM model and the reliability of local IMERG precipitation estimates for forecasting flood events using this method.

Next, Figure 5.28 depicts the forecasts of the adapted method using the IMERG precipitation data from the sets of remote pixels. The accuracy is still comparable with the ones observed in Figure 5.27 for all sets and errors are fairly balanced between the underestimation and overestimation of forecasts.



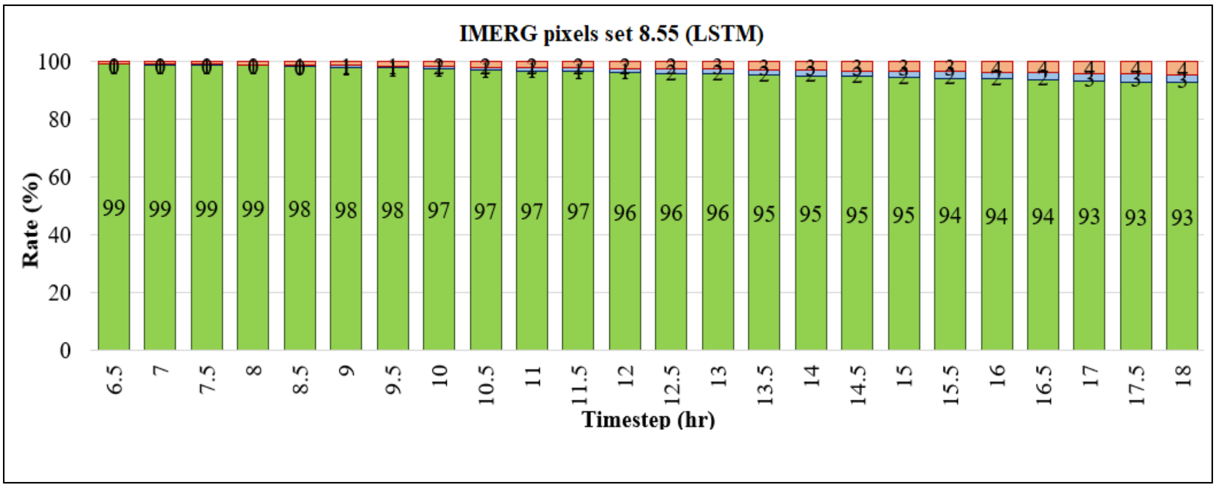


Figure 5-28 - Rates of targets hit, overestimation or underestimation with the adapted method using precipitation data from the remote IMERG pixels sets

However, while the method tends to underestimate rather than overestimate forecasts using data from sets 8.55 and 8.65, with data from set 8.75, this tendency changes to overestimation rather than underestimation, and error's work towards greater safety on flood risk predictions.

The analysis of the method's outcomes also looks into the key indicators of performance and respective formulas previously presented in Table 3.1.

These indicators are used to evaluate how well the method distinguish two classes. Figure 5.28 shows the performance indicators of both methods using rain-gauge data and the adapted method using each of the IMERG-based datasets. It also considers the delay on the production of the IMERG data and the lead time due to geographical position of the remote IMERG pixels.

The best performances are associated with the values at the top of the range of each indicator, with top for each lead time highlighted in yellow in Figure 5.29. This clearly demonstrates the advantages of the LSTM model, particularly for longer lead times.

Lead Step	LSTM - IMERG V07 Cranford pixel						
4 hs data delay							
0.5	99.53	0.98	98.79	99.64	2.40	98.19	99.52
1	99.31	0.97	98.08	99.49	2.20	97.35	99.29
1.5	99.13	0.96	97.46	99.38	2.09	96.69	99.10
2	98.89	0.95	96.40	99.26	1.96	95.74	98.84
2.5	98.63	0.94	95.57	99.09	1.86	94.78	98.56
3	98.37	0.93	94.46	98.95	1.77	93.75	98.26
3.5	98.09	0.92	93.54	98.77	1.69	92.71	97.94
4	97.84	0.90	92.13	98.69	1.62	91.70	97.63

Lead time	Piadeh's method - Northolt gauge							LSTM - IMERG V07 Cranford pixel							LSTM - Northolt gauge						
hours	ACC	MCC	TPR	TNR	DP	F1-sc	CKR	ACC	MCC	TPR	TNR	DP	F1-sc	CKR	ACC	MCC	TPR	TNR	DP	F1-sc	CKR
0.5	99.58	0.98	98.13	99.80	2.43	98.37	99.57	97.53	0.89	90.45	98.59	1.55	90.48	97.25	99.47	0.98	96.96	99.84	2.37	97.93	99.45
1	99.29	0.97	97.31	99.59	2.17	97.28	99.27	97.24	0.88	89.13	98.45	1.50	89.35	96.90	99.24	0.97	96.23	99.68	2.15	97.03	99.20
1.5	98.58	0.94	93.85	99.28	1.83	94.48	98.48	96.99	0.87	88.00	98.33	1.45	88.36	96.57	98.99	0.95	95.19	99.55	2.01	96.06	98.93
2	98.61	0.94	94.84	99.17	1.84	94.65	98.52	96.68	0.85	86.83	98.15	1.40	87.17	96.18	98.76	0.94	94.13	99.45	1.91	95.17	98.68
2.5	98.28	0.92	93.18	99.04	1.74	93.36	98.14	96.43	0.84	85.84	98.01	1.36	86.19	95.84	98.47	0.93	92.90	99.30	1.80	94.04	98.35
3	97.98	0.91	91.97	98.88	1.66	92.22	97.80	96.22	0.83	85.09	97.88	1.33	85.38	95.56	98.10	0.92	91.53	99.09	1.69	92.61	97.93
3.5	97.63	0.89	90.71	98.66	1.58	90.85	97.37	95.91	0.82	84.27	97.65	1.29	84.26	95.16	97.81	0.90	90.34	98.92	1.62	91.45	97.57
4	97.35	0.88	89.46	98.52	1.52	89.74	97.02	95.63	0.81	83.28	97.47	1.26	83.17	94.77	97.50	0.89	89.21	98.74	1.55	90.26	97.20
4.5	96.86	0.86	87.58	98.24	1.43	87.85	96.40	95.33	0.79	82.64	97.22	1.22	82.11	94.36	97.13	0.87	87.67	98.54	1.48	88.79	96.72
5	96.81	0.86	87.67	98.17	1.42	87.71	96.35	95.08	0.78	81.96	97.04	1.20	81.22	94.03	96.83	0.86	86.63	98.36	1.43	87.66	96.35
5.5	96.50	0.85	86.57	97.98	1.38	86.53	95.95	94.81	0.77	81.10	96.86	1.17	80.23	93.65	96.58	0.85	85.38	98.26	1.39	86.65	96.02
6	96.17	0.83	84.98	97.84	1.33	85.21	95.50	94.55	0.76	80.33	96.67	1.14	79.27	93.27	96.33	0.84	84.14	98.15	1.35	85.62	95.68
6.5	95.75	0.81	82.62	97.70	1.27	83.45	94.90	94.31	0.75	79.51	96.51	1.12	78.38	92.91	96.05	0.82	82.71	98.04	1.31	84.45	95.28
7	95.70	0.81	83.04	97.59	1.27	83.37	94.85	94.09	0.74	78.72	96.39	1.10	77.58	92.60	95.85	0.81	81.50	97.99	1.29	83.60	95.00
7.5	95.32	0.79	81.21	97.42	1.22	81.82	94.30	93.84	0.73	77.77	96.24	1.08	76.62	92.21	95.63	0.80	80.04	97.95	1.26	82.62	94.67
8	95.27	0.79	81.54	97.32	1.22	81.73	94.25	93.62	0.72	76.91	96.12	1.06	75.79	91.88	95.35	0.79	78.41	97.87	1.23	81.40	94.26
8.5	95.00	0.78	80.15	97.21	1.18	80.61	93.85								95.11	0.78	76.93	97.83	1.20	80.34	93.90
9	94.54	0.76	77.77	97.05	1.14	78.72	93.17								94.87	0.76	75.65	97.74	1.17	79.29	93.54
9.5	94.73	0.77	79.40	97.02	1.16	79.65	93.48								94.62	0.75	74.46	97.63	1.15	78.23	93.16
10	94.36	0.75	78.14	96.78	1.12	78.25	92.93								94.38	0.74	73.20	97.54	1.12	77.19	92.80
10.5	93.83	0.73	75.65	96.54	1.07	76.10	92.12								94.17	0.73	71.88	97.49	1.10	76.18	92.46
11	93.56	0.71	74.57	96.39	1.04	75.02	91.69								94.00	0.72	70.95	97.44	1.09	75.44	92.20
11.5	93.28	0.70	73.49	96.24	1.02	73.96	91.27								93.82	0.71	69.94	97.38	1.07	74.60	91.91
12	93.01	0.69	72.41	96.08	1.00	72.89	90.83								93.60	0.70	68.68	97.32	1.05	73.60	91.57
12.5	92.73	0.68	71.33	95.93	0.97	71.82	90.40														
13	92.46	0.66	70.25	95.77	0.95	70.75	89.95														
13.5	92.19	0.65	69.17	95.62	0.93	69.67	89.51														
14	91.91	0.64	68.09	95.46	0.91	68.60	89.06														

Lead time	LSTM - IMERG V07 pixels set 8.75							LSTM - IMERG V07 pixels set 8.65							LSTM - IMERG V07 pixels set 8.55						
hours	ACC	MCC	TPR	TNR	DP	F1-sc	CKR	ACC	MCC	TPR	TNR	DP	F1-sc	CKR	ACC	MCC	TPR	TNR	DP	F1-sc	CKR
0.5																					
1																					
1.5																					
2																					
2.5																					
3																					
3.5																					
4																					
4.5																					
5																					
5.5																					
6																					
6.5	99.46	0.98	98.37	99.63	2.32	97.95	99.45	99.45	0.98	96.91	99.83	2.35	97.85	99.43	99.19	0.96	94.35	99.91	2.36	96.80	99.14
7	99.21	0.97	97.53	99.46	2.13	96.97	99.19	99.22	0.97	95.85	99.72	2.16	96.97	99.19	99.01	0.96	93.65	99.81	2.14	96.07	98.94
7.5	98.97	0.95	96.40	99.35	1.99	96.03	98.92	98.96	0.95	95.21	99.52	1.99	95.98	98.91	98.86	0.95	92.96	99.74	2.04	95.49	98.78
8	98.61	0.94	95.17	99.12	1.85	94.67	98.53	98.71	0.94	94.16	99.38	1.88	94.97	98.62	98.66	0.94	91.93	99.66	1.95	94.68	98.55
8.5	98.27	0.92	94.02	98.90	1.74	93.37	98.14	98.40	0.93	92.70	99.25	1.78	93.76	98.27	98.37	0.93	90.89	99.49	1.81	93.54	98.21
9	97.89	0.91	92.72	98.66	1.64	91.95	97.71	98.06	0.91	91.42	99.05	1.68	92.44	97.87	98.08	0.91	89.88	99.30	1.71	92.38	97.87
9.5	97.55	0.89	91.53	98.45	1.56	90.65	97.30	97.77	0.90	90.41	98.87	1.61	91.31	97.52	97.80	0.90	88.64	99.17	1.64	91.28	97.53
10	97.24	0.88	90.43	98.26	1.50	89.48	96.92	97.40	0.88	89.06	98.64	1.53	89.88	97.07	97.50	0.89	87.41	99.00	1.56	90.07	97.16
10.5	96.91	0.86	89.02	98.09	1.44	88.20	96.50	97.03	0.87	87.85	98.39	1.46	88.46	96.61	97.17	0.87	86.04	98.83	1.50	88.74	96.74
11	96.63	0.85	87.72	97.96	1.40	87.10	96.13	96.74	0.85	86.48	98.27	1.41	87.32	96.24	96.86	0.86	84.91	98.64	1.44	87.54	96.35
11.5	96.41	0.84	87.16	97.79	1.37	86.31	95.86	96.44	0.84	85.29	98.11	1.37	86.16	95.85	96.58	0.85	83.75	98.50	1.39	86.42	95.98
12	95.95	0.82	85.82	97.46	1.30	84.61	95.25	96.17	0.83	84.23	97.95	1.33	85.10	95.48	96.33	0.83	82.69	98.37	1.36	85.40	95.64
12.5	95.62	0.81	84.58	97.27	1.26	83.37	94.80	95.84	0.81	82.82	97.78	1.28	83.78	95.02	96.07	0.82	81.50	98.25	1.32	84.34	95.28
13	95.48	0.80	84.03	97.18	1.25	82.83	94.61	95.57	0.80	81.96	97.60	1.25	82.75	94.65	95.73	0.81	79.97	98.08	1.27	82.93	94.79
13.5	95.12	0.79	82.73	96.97	1.21	81.50	94.11	95.28	0.79	81.34	97.36	1.22	81.73	94.26	95.42	0.79	78.72	97.91	1.23	81.68	94.36
14	94.80	0.77	81.41	96.80	1.17	80.27	93.65	95.02	0.78	80.46	97.19	1.19	80.74	93.89	95.12	0.78	77.33	97.77	1.20	80.44	93.92
14.5	94.46	0.76	80.08	96.60	1.13	78.95	93.14	94.67	0.76	79.22	96.97	1.15	79.41	93.38	94.81	0.76	76.11	97.60	1.16	79.21	93.47
15	94.15	0.74	78.96	96.41	1.10	77.78	92.68	94.37	0.75	78.45	96.75	1.12	78.35	92.96	94.56	0.75	74.92	97.49	1.14	78.15	93.09
15.5	93.80	0.73	77.61	96.22	1.07	76.48	92.15	94.02	0.74	77.48	96.49	1.09	77.09	92.45	94.27	0.74	73.25	97.40	1.11	76.83	92.64
16	93.51	0.72	76.51	96.05	1.05	75.38	91.71	93.73	0.72	76.33	96.33	1.06	75.97	92.00	93.99	0.72	71.64	97.33	1.08	75.58	92.20
16.5	93.23	0.70	75.19	95.92	1.02	74.25	91.26	93.47	0.71	75.12	96.20	1.04	74.91	91.59	93.64	0.71	69.96	97.18	1.05	74.08	91.66
17	92.97	0.69	73.69	95.85	1.00	73.12	90.83	93.16	0.70	73.95	96.03	1.01	73.74	91.11	93.33	0.69	68.57	97.03	1.02	72.75	91.18
17.5	92.71	0.68	72.23	95.76	0.98	71.99	90.39	92.87	0.68	72.70	95.88	0.99	72.59	90.65	93.01	0.68	67.38	96.84	0.99	71.45	90.68
18	92.43	0.66	70.64	95.68	0.95	70.78	89.92	92.61	0.67	71.57	95.74	0.97	71.53	90.22	92.69	0.66	65.62	96.72	0.97	69.96	90.15

Highest values per indicator per lead time

Precipitation data collected locally

Precipitation data collected from remote locations

Figure 5-29 - Performance indicators for the original method and the method with the LSTM model with each of the IMERG dataset.

Furthermore, observing the forecasts' indicators in Figure 5.29, at the 6.5hs lead time common threshold, the performance with data from the remote sets of pixels is significantly better than with the local data. In fact, the forecasts with the IMERG pixels with 6.5hs lead time predictions outperform also the forecasts over 1h lead time with the local rainfall data, demonstrating an impressive level of performance of model with the IMERG remote data.

Comparing the indicators of forecasts with the remote rainfall data, the difference between pixels' sets is minimal across all lead times. Also, although the pixels set 8.55 most frequently presents the best values, these refer to indicators that treat underestimation and overestimation equally (ACC, MCC, DP and CKR), or focused on the correct predictions of non-flooding events (TNR). Indictors of the correct estimation of stream overflow (flood) had the highest score in most lead times associated with set 8.65 (F1-score) and 8.75 (TPR) which is particularly important for early risk detection and triggering early warning.

However, Figure 5.27 also illustrates the difficulties for using the local IMERG data from the Cranford pixel on real-time predictions. As the data is available 4hs after the satellite observation, only the 8 time-steps ahead predictions are useful for real-time applications with a significant lower accuracy in comparison with the other datasets.

A more efficient alternative is offered by taking advantage of the best methods and datasets for each lead time. Therefore, the IMERG data can be used for the detection of flood risks from rainfall over the Atlantic producing forecasts with the LSTM model with at least 4.5hs ahead depending on the pixel location and the rainfall events speed forward. For better accuracy on short lead time forecasts the rain-gauge data should be used with Piadeh's original method and with the LSTM model between 1.5 and 4hs lead time.

Additionally, to mitigate the repercussions on flooding management is important to make the end user aware of the errors and uncertainties in the predictions, so it can be is considered on the planning and design of flooding strategies.

6 Summary and conclusions

This chapter presents an overall summary of this research work, the main conclusions, important contributions and significant limitations, and briefly discusses its applications and possible future developments.

6.1 Summary

This PhD research investigated the hypothesis that the lead time for real-time predictions of riverine flooding risks could be extended by combining the capabilities of data-driven deep learning technologies with the wide coverage of SPPs datasets. A deep learning LSTM algorithm was used to capture the patterns in the relationship between rainfall and stream level time series datasets, which vary across space and time. The IMERG products, selected as the most appropriated of the SSPs for real-time applications, were used for detecting the patterns in rainfall events that will determine its effects on stream water level in the catchment. While the IMERG products extensive coverage allows for these events to be detected early the simplification of hydrological modelling through DL techniques, allows for a fast-modelling process with minimal data requirements. Therefore, the combination of IMERG and LSTM modelling can be useful to determine well in advance if a particular rainfall event represents a risk of flooding for a certain catchment.

However, the data driven approach requires that the properties, limitations, and requirements of the datasets for modelling are well understood. Also, as an innovative method several aspects are yet to be explored. Therefore, for the successful application of this theory in real world flooding forecasting, several questions must be previously answered. Follow a summary of these questions, the aspects addressed, and the research work carried out in search for answers.

For data suitability:

- Can a satellite-based precipitation products be an adequate source of rainfall data for riverine flooding risks predictions? Are they suitable for deep learning modelling?

  To answer this question, precipitation data from the IMERG V07 SPPs was scrutinized, tested, and validated for stream level predictions. For validation, the performance of the IMERG V07 data was compared against those of rain-gauges data, which are a well-established source of rainfall data for flooding forecasts.

  The results presented in Section 5 shows a similar performance of the LSTM model with both rainfall sources. More importantly, the model's performance with the local IMERG V07 data was considered very good in event-based predictions with lead times extending at least up to 5 hours ahead. This indicates not only that the IMERG V07 products are an adequate source of rainfall data for predicting riverine flooding risks but also a serious competitor for rain-gauges data.

For modelling capabilities:

- Can a deep learning model, by analysing long periods of precipitation and stream level data, identify and learn the spatiotemporal relationship between remote rainfall events and their subsequent impact on stream water levels? Can it be done when significant spatial distance is involved, thus enabling early flood risks detection?

  The response to this question, came through the completion of two of the main objectives of this research. It started by the identification of an optimum region in the IMERG grid containing highly correlated precipitation data with stream level variations in the catchment. For increased lead time, this optimum region must allow for the detection of rainfall events at least four hours before it reaches the catchment. To enhance the model's ability for capturing the spatiotemporal relationships, the data from the optimum region was in tested different arrangements of modelling inputs.

  The completion of these tasks revealed that the LSTM model's performance improves with the expansion of the area of the rainfall represented in the datasets and with the application of data collected from pixels along a vertical arrangement in the IMERG grid.

  These insights guided the selection of three groups of pixels in the remote areas of the IMERG grid to provide the data of used for testing the methodology. The model performance with this data was considered very good in the overall predictions, for up to 15 hours lead time, and in the event-based predictions, for up to 12 hours lead time. These results show a significant increase in the lead times while keeping a similar level of accuracy of the predictions with the precipitation data from the local IMERG pixel (IMERG-Cranford) and rain-gauge (Northolt). Such performance is remarkable when accounting for the 650Km distance between the region of IMERG data collection and the C1 catchment boundaries, indicating a notable capability of the LSTM model to detect the rainfall-stream level relationship at great spatial and temporal variabilities.

For real-time applications:

- How do the deep learning models, trained with historical data, perform with real-time datasets in the predictions of stream level? Can this performance be compared with already stablished modelling technologies using rain-gauge datasets?

  The first part of the question was answered by replacing the historical data from the testing datasets with the real-time stream level readings and NRT data from the IMERG Early Run product on the inputs of the LSTM trained model.

  Remarkably the change of datasets had a minimal impact on the quality of the model's predictions maintaining an overall performance at similar levels of the ones observed with the rain gauges data. The results confirm that the model trained with historical data was able to recognize this patterns in the real-time data and produce accurate predictions.

  To address the second part of the question, the LSTM model was compared with an E-NARX model for producing predictions used in platform for the classification of flooding risk events. With rain-gauge datasets, the LSTM model outperformed the E-NARX model for lead times of 1.5 hours or more. However, for predictions with lead times of 6.5 hours or more, the LSTM model using NRT IMERG data outperformed any predictions made with rain-gauge data with significantly higher accuracy, regardless of the model used. It is worth highlighting that these lead times are a minimum of 6.5 hours, as the study considers weather events moving towards the catchment at a high forward speed (60 km/h). For events moving at an average speed (30 km/h), the same predictions can be provided with twice the lead time.

Additionally, the robustness of the method combining IMERG data and the LSTM model for predicting stream levels for catchment C1, was further validated by similar performance on the predictions for catchments C3 and C2 which are displayed in Appendixes D and E, respectively.

6.2 Conclusions

The extensive research and findings summarised in the previous section confirm the successful achievement of this thesis' three main objectives and allows for the following conclusions to be drawn::

I. Satellite-based precipitation estimates from the IMERG products, are adequate sources of rainfall data for predicting stream level variations due to rainfall incidence. Additionally, these products can be used for real-time prediction with deep learning models, provided their latency is considered..

II. The LSTM algorithm was able to detect the spatiotemporal patterns between stream water level variations and the characteristics of a rainfall events represented in the IMERG data. This ability allows for the real-time prediction of flooding risks from rainfall events that can be detected at great distances from the catchments through the NRT IMERG precipitation estimates. As such, the lead time on flooding risk forecasting can be increased through the early detection of extreme events along the extended range of precipitation monitoring area.

III. The IMERG and LSTM model integration method was proven robust by its consistent provision of accurate stream level predictions for different rainfall events and distinct catchments scenarios. Moreover, the model maintained very good performance in real-time predictions using the NRT IMERG product, despite its expected lower accuracy compared to the Final Run IMERG data used in the model's training.

6.3 Contributions

The following important contributions were made through and along the development of this methodology:

i. The application of this methodology in the UK effectively extended the lead time of real-time flood risk forecasts from three hours, as reported in previous studies, to at least nine hours while keeping very good accuracy
ii. By applying a deep learning modelling technology and effectively selecting the data collection area, this study enabled the application of the IMERG products in real-time forecasting tasks.

iii. Along the development of the methodology, this thesis contributed with the identification of the most common paths of rainfall events across the Atlantic Ocean affecting water levels in three different UK catchments over the past 20 years.

iv. Given the extensive amount of work already completed, the proposed methodology can be readily expanded to include all catchments with naturally flowing rivers across the UK, thereby extending its application to forecasting event-based flooding risks over a significantly larger portion of the country.

v. The study also identified a region in the Atlantic Ocean where IMERG data correlates with both catchments, C1 and C3, which are located on opposite sides of England. This suggests the existence of a common area in the Atlantic for simultaneously monitoring flooding hazards for several UK catchments.

vi. To regions affected by extreme weather originating from ungauged areas, this works offers an alternative method for long range detection of flooding hazards and earlier flooding risks forecasts.

6.4 Limitations and future work

Although the ability of increasing lead time with SPPs is an exciting achievement, this study only proved it functional under certain conditions and considerations:

- SPPs have limited real-time application due to the long latency of the products, which is at least four hours for the IMERG V07 Early Run. Therefore, the proposed methodology is recommended for detecting flooding risks from long-duration precipitation events, but not applicable for detecting the short and intense rainfall events that cause flash flooding.

- The methodology was proved successful for predictions of water level in rivers with natural flow regime and requires long periods of data regularly recorded. Other circumstances were not investigated.
- Before applying the methodology to a new region, the IMERG products datasets corresponding to that regions must be thoroughly assessed. The assessment is required to identify and avoid potential areas of issues in the IMERG datasets related to the study area geolocation.

Nonetheless, future work can address some of these limitations as well as expand the application of the methodology. For example:

- While this study used over 20 years of data to train the LSTM model, shorter periods of data can be tested to determine the minimum period required for the desired accuracy on predictions.
- While due to limited computational resources this study tested data from limited selection of the IMERG pixels. The data collection area should be expanded to test model performances with data at increasing distances from the catchment potentially reaching even longer lead-times with reasonable accuracy.
- On practical applications, the model in this study processes over 200,000 time-steps of input data, covering two years of precipitation estimates from five IMERG pixels, in under 10 seconds running on a GPU NVIDIA GeForce RTX 4090. Therefore, considering that real-time predictions may use only a few weeks of data (less than 1000 time-steps per pixel location), with powerful

computing resources and easy access to NRT IMERG data, monitoring the wider areas of the IMERG grid can becomes an easy task.

Therefore, a system can be developed to monitor these areas and, once a hazardous event is detected, automatically generate predictions of flooding risks in the catchments. As the rainfall event progresses, the predictions can be updated with the new information provided in NRT IMERG at every 30 minutes, thus informing the catchment in real-time. If the weather system changes, for better or worse, the predictions will follow, and the level of alertness adjusted accordingly until the weather system reaches a perimeter within the range of more reliable land-based sources of rainfall data.

Author’s personal note:

More importantly, these suggestions for future work represent only a fraction of the wide range of research opportunities and practical applications made possible by this thesis, which I hope will contribute meaningfully to the advancement of flood forecasting and mitigation strategies.

7 References:

Abdollahipour, A., Ahmadi, H. and Aminnejad, B. (2021) ‘Evaluating the hydrological utility of satellite-based rainfall products using neural network models over the Ghare Ghieh River basin, Iran’, Journal of Water and Climate Change |, 12, p. 2021.

Adamowski, J., & Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. Journal of Hydrology, 390(1-2), 85-91.

Adane, G.B.; Hirpa, B.A.; Gebru, B.M.; Song, C.; Lee, W.-K. (2021) Integrating Satellite Rainfall Estimates with Hydrological Water Balance Model: Rainfall-Runoff Modeling in Awash River Basin, Ethiopia. Water 2021, 13, 800.

Admojo, D. D., Tebakari, T., & Miyamoto, M. (2018). Evaluation of a satellite-based rainfall product for a runoff simulation of a flood event: A case study. Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering). 74. I_73-I_78. 10.2208/jscejhe.74.I_73.

Ahmed, E., Al Janabi, F., Zhang, J., Yang, W., Saddique, N., & Krebs, P. (2020). Hydrologic assessment of TRMM and GPM-based precipitation products in transboundary river catchment (Chenab River, Pakistan). Water, 12(7), 1902.

Al-Areeq, A. M., Al-Zahrani, M. A., & Sharif, H. O. (2022). Assessment of the performance of satellite rainfall products over Makkah watershed using a physically based hydrologic model. *Applied Water Science*, *12*(11), 246.

Allen-Dumas, M. R., Xu, H., Kurte, K. R., & Rastogi, D. (2021). Toward Urban Water Security: Broadening the Use of Machine Learning Methods for Mitigating Urban Water Hazards. Frontiers in Water. Frontiers Media S.A.

Almikaeel, W., Šoltész, A., Čubanová, L., & Baroková, D. (2024). Hydro-informer: a deep learning model for accurate water level and flood predictions. Natural Hazards, 1-21.

Alpaydin, E. (2020). Introduction to machine learning. MIT press.

Andrieu, H., Creutin, J. D., Delrieu, G., Faure, D. (1997). Use of a weather radar for the hydrology of a mountainous area. Part I: Radar measurement interpretation. Journal of Hydrology, 193(1-4), 1-25.

Antwi-Agyakwa KT, Afenyo MK, Angnuureng DB (2023). Know to Predict, Forecast to Warn: A Review of Flood Risk Prediction Tools. Water. 2023; 15(3):427.

Arnaud, P., & Lavabre, J. (2002). Coupled rainfall model and discharge model for flood frequency estimation. Water Resources Research, 38(6), 11-1.

Arnell, N. W., Freeman, A., Kay, A. L., Rudd, A. C., & Lowe, J. A. (2021). Indicators of climate risk in the UK at different levels of warming. *Environmental Research Communications*, *3*(9), 095005.kendon

Ashley, R., Clemens, F., Tait, S, Schellart, A. (2008). Climate change and the implications for modelling the quality of flow in combined sewers. 11th

International Conference on Urban Drainage, Edinburgh, Scotland, UK, 2008 Climate, Edinburgh, Scotland, UK.

Badrzadeh , H., Sarukkalige, R., & Jayawardena, A. W. (2013). Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting. Journal of Hydrology, 507, 75-85.

Bates, P. D., Savage, J., Wing, O., Quinn, N., Sampson, C., Neal, J., and Smith, A. (2023). Flood HA climate-conditioned catastrophe risk model for UK flooding [Supplementary material], Nat. Hazards Earth Syst. Sci., 23, 891–908,

Behrangi, A., Hsu, K. L., Imam, B., Sorooshian, S., Huffman, G. J., & Kuligowski, R. J. (2009). PERSIANN-MSA: A precipitation estimation method from satellite-based multispectral analysis. Journal of Hydrometeorology, 10(6), 1414-1429.

Berne, A., & Krajewski, W. F. (2013). Radar for hydrology: Unfulfilled promise or unrecognized potential?. Advances in Water Resources, 51, 357-366.

Bitew, M. M., Gebremichael, M., Ghebremichael, L. T., & Bayissa, Y. A. (2012). Evaluation of high-resolution satellite rainfall products through streamflow simulation in a hydrological modeling of a small mountainous watershed in Ethiopia. Journal of Hydrometeorology, 13(1), 338–350.

Black, A., Peskett, L., MacDonald, A., Young, A., Spray, C., Ball, T., ... & Werritty, A. (2021). Natural flood management, lag time and catchment scale: Results from an empirical nested catchment study. *Journal of Flood Risk Management*, *14*(3), e12717.

Boonyuen, K., Kaewprapha, P., Weesakul, U., & Srivihok, P. (2019). Convolutional Neural Network Inception-v3: A Machine Learning Approach for Leveling Short-Range Rainfall Forecast Model from Satellite Image. Advances in Swarm Intelligence, 11656, 105-115.

Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. arXiv preprint arXiv:1809.03006.

Boyd, E. H. (2021). The crucial role of nature-based solutions in addressing the climate crisis. Climate expo, environment agency, UK. Accessed on 21.06.2021 at: https://www.gov.uk/government/speeches/the-crucial-role-of-nature-based-solutions-in-addressing-the-climate-crisis

Campolo, M., Andreussi, P., & Soldati, A. (1999). River flood forecasting with a neural network model. Water resources research, 35(4), 1191-1197.

Brown JD. 2009. The coefficient of determination. JALT Testing & Evaluation SIG Newsletter. 7 (1) March 2003 (p. 14 - 16). (online) Available at: https://teval.jalt.org/test/PDF/Brown16.pdf

Camus, P., Haigh, I. D., Quinn, N., Wahl, T., Benson, T., Gouldby, B., ... & Nadal-Caraballo, N. C. (2024). Tracking the spatial footprints of extreme storm surges around the coastline of the UK and Ireland. Weather and Climate Extremes, 44, 100662

Carbone, G., Garofalo, G., Piro,P., (2014). Decentralized Real Time Control in Combined Sewer System by Using Smart Objects, Procedia Engineering, Elsevier.

Chan, F.K.S., Griffiths, J.A., Hoggitt, D., Xu, S., Zhu, F., Tang, Y., Xu, Y., Thorne, C.R., (2018). “Sponge City” in China—A breakthrough of planning and flood risk management in the urban context, Land Use Policy, Elsevier.

Chen Y (2015) A New Methodology of Spatial Cross-Correlation Analysis. PLoS ONE 10(5): e0126158.

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data mining for the internet of things: literature review and challenges. International Journal of Distributed Sensor Networks, 11(8), 431047.

Cheng, L., & Aghakouchak, A. (2014). Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. Scientific Reports, 4(1), 7093.

Chen, J., Li, Y., Zhang, C., Tian, Y., & Guo, Z. (2023). Urban flooding prediction method based on the combination of LSTM neural network and numerical model. *International Journal of Environmental Research and Public Health*, *20*(2), 1043.

Chicco D, Warrens MJ, Jurman G. 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science 7:e623

Cho, M., Kim, C., Jung, K., & Jung, H. (2022). Water Level Prediction Model Applying a Long Short-Term Memory (LSTM)–Gated Recurrent Unit (GRU) Method for Flood Prediction. Water (Switzerland), 14(14).

Committee of climate change (2017). The UK risk assessment 2017 evidence report (online). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/584281/uk-climate-change-risk-assess-2017.pdf; [accessed on 23.06.2018]

DEFRA (2023). Hydrology Data Explorer. Department for Environment Food and Rural Affairs – DEFRA (online). Available at: https://environment.data.gov.uk/hydrology/explore. [accessed on 18.05.2023]

Demuzere, M., Orru, K., Heidrich, O., Olazabal, E., Geneletti, D., Orru, H., Bhave, A.G., Mittal, N., Feliu, E., Faehnle, M., (2014). Mitigating and adapting to climate change: Multi-functional and multi-scale assessment of green urban infrastructure, Journal of Environmental Management, Elsevier

Deng, H., Chen, W., & Huang, G. (2022). Deep insight into daily runoff forecasting based on a CNN-LSTM model. Natural Hazards, 113(3), 1675–1696.

Department for business, energy and industrial strategy (2021). Industrial Decarbonisation Strategy, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/970229/Industrial_Decarbonisation_Strategy_March_2021.pdf; accessed on 17.05.2021

Dewitte, S.; Cornelis, J.P.; Müller, R.; Munteanu, A. (2021). Artificial Intelligence Revolutionises Weather Forecast, Climate Monitoring and Decadal Prediction', Remote Sensing 2021, Vol. 13, Page 3209, 13(16), p. 3209.

Ding, Y., Zhu, Y., Feng, J., Zhang, P., & Cheng, Z. (2020). Interpretable spatio-temporal attention LSTM model for flood forecasting. Neurocomputing, 403, 348-359.

Dong, W., & Yang, Q. (2019). Data-driven solution for optimal pumping units scheduling of smart water conservancy. *IEEE Internet of Things Journal*, *7*(3), 1919-1926.

Ehsani, M. R., Zarei, A., Gupta, H. V., Barnard, K., Lyons, E., & Behrangi, A. (2022). NowCasting-Nets: Representation learning to mitigate latency gap of satellite precipitation products using convolutional and recurrent neural networks. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-21.

Elkhrachy, I. (2022). Flash Flood Water Depth Estimation Using SAR Images, Digital Elevation Models, and Machine Learning Algorithms" Remote Sensing 14, no. 3: 440.

Environment Agency (2018). Estimating the economic costs of the 2015 to 2016 winter floods. Environment Agency, January 2018, section 3.

Environment Agency (2023). Flooding in England: A National Assessment of Flood Risk; Environment Agency, Bristol, UK

Environment Agency (2023a). Catchment data explorer. Environment Agency. (online) . Available at : https://environment.data.gov.uk/hydrology/explore [accessed on 24.08.2023]

EPA (2015). Green Infrastructure for Stormwater Control: Gauging Its. Effectiveness with Community Partners. United States Environmental Protection Agency (online). Available at: https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100NE3S.TXT [accessed on 18.04.2021]

Esri. (2016). *Emerging hotspots* [Webpage]. ArcGIS Desktop. https://desktop.arcgis.com/en/arcmap/10.3/tools/space-time-pattern-mining-toolbox/emerginghotspots.htm [accessed on 15.09.2024]

ESW (2023). European Severe Storms Laboratory (online). Available at: https://eswd.eu/cgi-bin/eswd.cgi [accessed on 19.05.2024]

Farahani, G. (2020). Feature Selection Based on Cross-Correlation for the Intrusion Detection System. Security and Communication Networks, 2020(1), 8875404.

Fordjour, G. K., & Kalyanapu, A. J. (2024). Implementing a GRU Neural Network for Flood Prediction in Ashland City, Tennessee. arXiv preprint arXiv:2405.10375.

Foroughnia, F., Alfieri, S. M., Menenti, M., & Lindenbergh, R. (2022). Evaluation of SAR and optical data for flood delineation using supervised and unsupervised classification. Remote Sensing, 14(15), 3718.

François, B., Schlef, K. E., Wi, S., & Brown, C. M. (2019). Design considerations for riverine floods in a changing climate–a review. Journal of Hydrology, 574, 557-573.

Fritsch, J. M., Houze Jr, R. A., Adler, R., Bluestein, H., Bosart, L., Brown, J., ... & Zrnic, D. (1998). Quantitative precipitation forecasting: Report of the eighth prospectus development team, US Weather Research Program. Bulletin of the American Meteorological society, 79(2), 285-299.

Funk, C., Peterson, P., Landsfeld, M. et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci Data 2, 150066 (2015).

Funk, C., Peterson, P., Landsfeld, M. et al., (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci Data 2, 150066 (2015).

Gao, X., Zhu, Q., Yang, Z., & Wang, H. (2018). Evaluation and hydrological application of CMADS against TRMM 3B42V7, PERSIANN-CDR, NCEP-CFSR, and gauge-based datasets in Xiang River Basin of China. *Water*, *10*(9), 1225.

Garcia-Aristizabal A, E. Bucchignani, E. Palazzi, D. D'Onofrio, P. Gasparini, and W. Marzocchi (2015). Analysis of non-stationary climate-related extreme events considering climate change scenarios: an application for multi-hazard assessment in the Dar es Salaam region, Tanzania. Natural Hazards, 75:289-320.

Ghorbanzadeh, O., Rostamzadeh, H., Blaschke, T., (2018). A new GIS-based data mining technique using an adaptive neuro-fuzzy inference system (ANFIS) and k-fold cross-validation approach for land subsidence susceptibility mapping. Nat Hazards 94, 497–517.

Golmohammadi, J., Xie, Y., Gupta, J., Farhadloo, M., Li, Y., Cai, J., Detor, S. ,Roh, A., Shekhar, S. (2018). An introduction to spatial data mining. The Geographic Information Science & Technology Body of Knowledge. 4th Quarter 2020 Edition. Chapter: AM-97

GOV.UK (2021). Counting the cost of flooding. UK Government News, 25 February,2021. https://www.gov.uk/government/news/counting-the-costs-of-flooding

GOV.UK (2021). Counting the cost of flooding. UK Government News, 25 February,2021. https://www.gov.uk/government/news/counting-the-costs-of-flooding

Green, A. C., Kilsby, C., & Bárdossy, A. (2024). Assessing rainfall radar errors with an inverse stochastic modelling framework. Available at SSRN 4478761.

Guo, Z., Leitao, J. P., Simões, N. E., & Moosavi, V. (2021). Data‐driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks. *Journal of Flood Risk Management*, *14*(1), e12684.

Hafizi, H., & Sorman, A. A. (2022). Assessment of 13 gridded precipitation datasets for hydrological modeling in a mountainous basin. Atmosphere, 13(1), 143.

Hamza, A., Anjum, M. N., Masud Cheema, M. J., Chen, X., Afzal, A., Azam, M., ... & Gulakhmadov, A. (2020). Assessment of IMERG-V06, TRMM-3B42V7, SM2RAIN-ASCAT, and PERSIANN-CDR precipitation products over the Hindu Kush Mountains of Pakistan, South Asia. Remote Sensing, 12(23), 3871.

Hayder, G., Solihin, M. I., & Najwa, M. R. N. (2022). Multi-step-ahead prediction of river flow using NARX neural networks and deep learning LSTM. H2Open Journal, 5(1), 42–59. https://doi.org/10.2166/h2oj.2022.134

HC 71 (2024). Resilience to flooding. House of Commons Committee of Public Accounts (online). Available at: https://publications.parliament.uk/pa/cm5804/cmselect/cmpubacc/71/summary.html [accessed on 10.02.2024]

Hinge, G., Hamouda, M. A., Long, D., & Mohamed, M. M. (2022). Hydrologic utility of satellite precipitation products in flood prediction: A meta-data analysis and lessons learnt. *Journal of Hydrology*, *612*, 128103.

Hirsch, R.M. (2011). A Perspective on Nonstationarity and Water Management; JAWRA Journal of the American Water Resources Association, 47: 436-446.

Hochreiter, S. (1997). Long Short-term Memory. Neural Computation MIT-Press.

Hsu, K., Gao, X., Sorooshian, S., & Gupta, H. V. (1997). Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks, Journal of Applied Meteorology, 36(9), 1176-1190. Retrieved Jan 24, 2022, from https://journals.ametsoc.org/view/journals/apme/36/9/1520-0450_1997_036_1176_pefrsi_2.0.co_2.xml

Huffman, G. J., Bolvin, D. T., Joyce, R., Kelley, O.A, Nelkin, E. J., Tan, J., Watters, D.C., West, B.J. (2023). Integrated Multi-satellitE Retrievals for GPM (IMERG) technical documentation. Nasa/Gsfc. Retrieved Sep 24, 2024, from https://gpm.nasa.gov/sites/default/files/2023-07/IMERG_TechnicalDocumentation_final_230713.pdf

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., ... & Xie, P. (2018). NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD) Version, 4.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K. L., Joyce, R. J., Kidd, C., ... & Xie, P. (2020). Integrated multi-satellite retrievals for the global precipitation measurement (GPM) mission (IMERG). Satellite precipitation measurement: Volume 1, 343-353.

Hur, J., Raghavan, S. V., Nguyen, N. S., & Liong, S. Y. (2016). Evaluation of high-resolution satellite rainfall data over Singapore. *Procedia Engineering*, *154*, 158-167.

Ireland, G.; Volpi, M.; Petropoulos, G.P. (2015). Examining the Capability of Supervised Machine Learning Classifiers in Extracting Flooded Areas from Landsat TM Imagery: A Case Study from a Mediterranean Flood. Remote Sensing, 7, 3372-3399.

Jackson, J. (2002). Data Mining; A Conceptual Overview. Communications of the Association for Information Systems, 8, pp-pp.

Jain, S. K., Mani, P., Jain, S. K., Prakash, P., Singh, V. P., Tullos, D., ... & Dimri, A. P. (2018). A Brief review of flood forecasting techniques and their applications. International journal of river basin management, 16(3), 329-344.

JBA Risk Management Ltd. (2021). Modelling the impact of spending on defense maintenance on flood losses. ABI and Flood Re report (online). Available at: https://www.abi.org.uk/globalassets/files/publications/public/flooding/modelling-the-impact-of-spending-on-defence-maintenance.pdf

Jiang, L., & Bauer-Gottwein, P. (2019). How do GPM IMERG precipitation estimates perform as hydrological model forcing? Evaluation for 300 catchments across Mainland China. *Journal of Hydrology*, *572*, 486-500.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. Journal of Hydrometeorology, 5(3), 487–503.

Kao, I. F., Zhou, Y., Chang, L. C., & Chang, F. J. (2020). Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. Journal of Hydrology, 583, 124631.

Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., Williams, J., Sparks, T., & West, F. (2023). State of the UK Climate 2022. International Journal of Climatology, 43(S1), 1–82.

Kidd, C., & Levizzani, V. (2011). Status of satellite precipitation retrievals. Hydrology and Earth System Sciences, 15(4), 1109-1116.

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, How Much of the Earth's Surface Is Covered by Rain Gauges?. Bulletin of the American Meteorological Society, 98(1), 69-78.

Kidd, C., & Levizzani, V. (2022). Satellite rainfall estimation. In Rainfall (pp. 135-170). Elsevier.

Kim, G., & Barros, A. P. (2001). Quantitative flood forecasting using multisensor data and neural networks. Journal of Hydrology, 246(1-4), 45-62.

Kim, H. I., Han, K. Y., & Lee, J. Y. (2020). Prediction of Urban Flood Extent by LSTM Model and Logistic Regression. KSCE Journal of Civil and Environmental Engineering Research, 40(3), 273–283.

Konrad, C. P. (2003). Effects of urban development on floods.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences, 22(11), 6005-6022.

Kumar V, Sharma KV, Caloiero T, Mehta DJ, Singh K. (2023). Comprehensive Overview of Flood Modeling Approaches: A Review of Recent Advances. Hydrology. 2023; 10(7):141.

Kumar, S., & Mohbey K. (2019). A review on big data based parallel and distributed approaches of pattern mining; Journal of King Saud University - Computer and Information Sciences -In press.

Latif, S. D., & Ahmed, A. N. (2023). A review of deep learning and machine learning techniques for hydrological inflow forecasting. Environment, Development and Sustainability, 25(11), 12189-12216.

Le, M. H., Lakshmi, V., Bolten, J., & Du Bui, D. (2020). Adequacy of satellite-derived precipitation estimate for hydrological modeling in Vietnam basins. Journal of Hydrology, 586, 124820.

Lee, W. K., & Tuan Resdi, T. A. (2016). Simultaneous hydrological prediction at multiple gauging stations using the NARX network for Kemaman catchment, Terengganu, Malaysia. Hydrological Sciences Journal, 61(16), 2930-2945.

Lee, Y., Hall, D., Stewart, J., & Govett, M. (2019). Machine Learning for Targeted Assimilation of Satellite Data. Machine Learning and Knowledge Discovery in Databases, 11053, 53-68.nvironment Agenc

Leinonen, J., Hamann, U., Sideris, I. V., & Germann, U. (2023). Thunderstorm Nowcasting With Deep Learning: A Multi-Hazard Data Fusion Model. Geophysical Research Letters, 50(8), e2022GL101626.

Lennon, M., Scott, M., O’Neill E. (2014). Urban Design and Adapting to Flood Risk: The Role of Green Infrastructure; Journal of Urban Design, Dublin, Ireland.

Li, D., Liu, Y., & Chen, C. (2021). MSDM v1. 0: A machine learning model for precipitation nowcasting over eastern China using multisource data. Geoscientific Model Development, 14(6), 4019-4034.

Li J (2017) Assessing the accuracy of predictive models for numerical data: Not r nor r , why not? Then what? PLoS ONE 12(8): e0183250.

Li, X., Xu, W., Ren, M., Jiang, Y., & Fu, G. (2022). Hybrid CNN-LSTM models for river flow prediction. Water Supply, 22(5), 4902-4919.

Linardos, V., Drakaki, M., Tzionas, P., & Karnavas, Y. L. (2022). Machine learning in disaster management: recent developments in methods and applications. *Machine Learning and Knowledge Extraction*, *4*(2).

Liu, Z. (2023). Accuracy of satellite precipitation products in data-scarce Inner Tibetan Plateau comprehensively evaluated using a novel ground observation network. Journal of Hydrology: Regional Studies, 47, 101405.

Li, Y., Pang, B., Ren, M., Shi, S., Peng, D., Zhu, Z., & Zuo, D. (2022). Evaluation of performance of three satellite-derived precipitation products in capturing extreme precipitation events over Beijing, China. *Remote Sensing*, *14*(11), 2698.

Maggioni, V., Meyers, P. C., & Robinson, M. D. (2016). A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) era. *Journal of Hydrometeorology*, *17*(4), 1101-1117.

Mancipe, N., Buchberger, S., Suidan, M. (2014). SITES - A novel approach for controlling combine sewer overflows; Science Direct, Elsevier, Cincinnati, US.

Medina, Y., & Muñoz, E. (2020). Estimation of Annual Maximum and Minimum Flow Trends in a Data-Scarce Basin. Case Study of the Allipén River Watershed, Chile. Water, 12(1), 162. https://doi.org/10.3390/w12010162

Merizalde, M. J., Muñoz, P., Corzo, G., Muñoz, D. F., Samaniego, E., & Célleri, R. (2023). Integrating geographic data and the SCS-CN method with LSTM networks for enhanced runoff forecasting in a complex mountain basin. Frontiers in Water, 5, 1233899.

Met Office (2019). What are weather radar? Met Office (online). Available at: https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/weather-science/radar_a4_1_02.pdf. [accessed on 10.10.2021]

Met Office (2023a). Air mass types. Met Office (online). Available at: https://www.metoffice.gov.uk/weather/learn-about/weather/atmosphere/air-masses/types [accessed on 10.07.2023]

Met Office (2023b). Tropical cyclone facts. Met Office (online). Available at: https://www.metoffice.gov.uk/research/weather/tropical-cyclones/facts [accessed on 15.11.2023]

Met Office (2024). Past weather events. Met office (online). Available at: https://www.metoffice.gov.uk/weather/learn-about/past-uk-weather-events [accessed on 06.06.2024]

Miau, S., & Hung, W. H. (2020). River flooding forecasting and anomaly detection based on deep learning. Ieee Access, 8, 198384-198402.

Moges, D. M., Kmoch, A., & Uuemaa, E. (2022). Application of satellite and reanalysis precipitation products for hydrological modeling in the data-scarce Porijõgi catchment, Estonia. *Journal of Hydrology: Regional Studies*, *41*, 101070.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), 885-900.

Mosavi, A., Ozturk, P., & Chau, K. -w. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, *10*(11), 1536.

Mounce, S. R., Shepherd, W., Sailor, G., Shucksmith, J., & Saul, A. J. (2014). Predicting combined sewer overflows chamber depth using artificial neural networks with rainfall radar data. Water Science and Technology, 69(6), 1326-1333.

Muñoz, P., Corzo Perez, G. A., Solomatine, D. P., Feyen, J., & Célleri, R. (2024). Use of Near-Real-Time Satellite Precipitation Data and Machine Learning to Improve Extreme Runoff Modeling. Advanced Hydroinformatics: Machine Learning and Optimization for Water Resources, 149-175.

Murphy, JM, Sexton, DM, Jenkins, GJ, Booth, BB, Brown, CC, Clark, RT, Collins, M, Harris, GR, Kendon, EJ, Betts, RA, Brown, SJ, Humphrey, KA, McCarthy, MP, McDonald, RE, Stephens, A, Wallace, C, Warren, R, Wilby, R and Wood, RA (2009) UK Climate Projections Science Report: Climate Change projections. Working Paper. Met Office Hadley Centre, Exeter. Available at: https://research-portal.uea.ac.uk/en/publications/uk-climate-projections-science-report-climate-change-projections [accessed on 17 September 2021]

Nanda, T., Sahoo, B., Beria, H., & Chatterjee, C. (2016). A wavelet-based non-linear autoregressive with exogenous inputs (WNARX) dynamic neural network model for real-time flood forecasting using satellite-based rainfall products. Journal of Hydrology, 539, 57-73.

NASA GES DISC (2023). Goddard Earth Sciences Data and Information Services Center (online). Available at: https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGHH_07/summary [accessed on 16 April 2023]

NASA GSFC (2019). Diagram of the GPM satellite constellation as of early 2019. National Aeronautics and Space Administration, Goddard Space Flight Center (online). Available at: https://gpm.nasa.gov/missions/GPM/constellation [accessed on 26 September 2021]

Naser, M. Z., & Alavi, A. H. (2021). Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences. Archit. Struct. Constr.(2021).

National Audit Office (2023). Resilience to flooding summary (PDF), November 2023, p. 5-7.

Nemerow, N.L., Agardi F.J., Sullivan, P., Salvato, J.A.(2009). Water, Wastewater, Soil and Ground Treatment and Remediation; 6th edn, John Wiley & Sons, Inc., Hoboken, New Jersey.

Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., ... & Matias, Y. (2022). Flood forecasting with machine learning models in an operational framework. Hydrology and Earth System Sciences, 26(15), 4013-4032.

Nhangumbe, M., Nascetti, A., Georganos, S., & Ban, Y. (2023). Supervised and unsupervised machine learning approaches using Sentinel data for flood mapping and damage assessment in Mozambique. Remote Sensing Applications: Society and Environment, 32, 101015.

Noor, H. M., Ndzi, D., Yang, G., & Safar, N. Z. M. (2017). Rainfall-based river flow prediction using NARX in Malaysia. Proceedings - 2017 IEEE 13th International Colloquium on Signal Processing and Its Applications, CSPA 2017, 67–72.

Nourani, V., Gökçekuş, H., & Gichamo, T. (2021). Ensemble data-driven rainfall-runoff modeling using multi-source satellite and gauge rainfall data input fusion. Earth Science Informatics, 14(4), 1787-1808.

Nourani, V., Khodkar, K., & Gebremichael, M. (2022). Uncertainty assessment of LSTM based groundwater level predictions. Hydrological Sciences Journal, 67(5), 773–790. https://doi.org/10.1080/02626667.2022.2046755

Nystuen, J. A. (1999). Relative Performance of Automatic Rain Gauges under Different Rainfall Conditions.

Ochoa-Rodriguez, S., Wang, L. P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., ... & ten Veldhuis, M. C. (2015). Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. Journal of Hydrology, 531, 389-407.

Ozdogan, M. (2017). Time Series Analysis of Remotely sensed data & accuracy assessment. European Space Agency - ESA (online). Available at: https://eo4society.esa.int/wp-content/uploads/2020/08/Time-series_Ozdogan_2017.pdf

Panagoulia, D., Tsekouras, G. J., & Kousiouris, G. J. G. N. J. (2017). A multi-stage methodology for selecting input variables in ANN forecasting of river flows. Glob. Nest J, 19, 49-57

Paradis, E., O'Brien, B., Nimmon, L., Bandiera, G., & Martimianakis, M. A. (2016). Design: Selection of data collection methods. Journal of graduate medical education, 8(2), 263-264.

Park, Yoon Soo & Konge, Lars & Artino, Anthony. (2019). The Positivism Paradigm of Research. Academic Medicine. 95. 1. 10.1097/ACM.0000000000003093.

Peng, J.; Liu, T.; Huang, Y.; Ling, Y.; Li, Z.; Bao, A.; Chen, X.; Kurban, A.; De Maeyer, P. Satellite-Based Precipitation Datasets Evaluation Using Gauge

Observation and Hydrological Modeling in a Typical Arid Land Watershed of Central Asia. Remote Sens. 2021, 13, 221.

Perez, R., Cebecauer, T., Suri, M., (2013). Solar energy forecasting and resource assessment. Academic press, Pages 21-48, ISBN 9780123971777,

Pham, M. K., Nguyen, P. L., Vu, V. H., Truong, T. N., Vo-Van, H., & Ngo-Duc, T. (2024). A data-driven approach for high accurate spatiotemporal precipitation estimation. Neural Computing and Applications, 1-20.

Piadeh, F., Behzadian, K., & Alani, A. M. (2022). A critical review of real-time modelling of flood forecasting in urban drainage systems. Journal of Hydrology, 607, 127476.

Piadeh, F., Behzadian, K., Chen, A. S., Campos, L. C., Rizzuto, J. P., & Kapelan, Z. (2023a). Event-based decision support algorithm for real-time flood forecasting in urban drainage systems using machine learning modelling. Environmental Modelling & Software, 167, 105772.

Piadeh, F., Behzadian, K., Chen, A. S., Kapelan, Z., Rizzuto, J. P., & Campos, L. C. (2023b). Enhancing urban flood forecasting in drainage systems using dynamic ensemble-based data mining. *Water Research*, *247*, 120791.

Pitt, Sir M. (2008). Learning lessons from the 2007 floods (The Pitt Review). Cabinet Office, London. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20100702215619/http://ar chive.cabinetoffice.gov.uk/pittreview/thepittreview/final_report.html (accessed on 29 October 2021)

Pradhan, R. K., Markonis, Y., Godoy, M. R. V., Villalba-Pradas, A., Andreadis, K. M., Nikolopoulos, E. I., ... & Hanel, M. (2022). Review of GPM IMERG performance: A global perspective. Remote Sensing of Environment, 268, 112754.

Rachdane, M., Khalki, E. M. E., Saidi, M. E., Nehmadou, M., Ahbari, A., & Tramblay, Y. (2022). Comparison of high-resolution satellite precipitation products in sub-saharan Morocco. Water, 14(20), 3336.

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. Scientific Reports, 14(1), 6086.

Rajoub, B. (2020). Supervised and unsupervised learning. In Biomedical signal processing and artificial intelligence in healthcare (pp. 51-89). Academic Press.

Rjeily, Y. A., Abbas, O., Sadek, M., Shahrour, I., & Chehade, F. H. (2017). Flood forecasting within urban drainage systems using NARX neural network. Water Science and Technology, 76(9), 2401–2412.

Saghafian, B., Farazjoo, H., Bozorgy, B. (2008). Flood Intensification due to Changes in Land Use. Water Resource Management 22, 1051–1067 (2008).

Schaefer, A. M., Udluft, S., & Zimmermann, H. G. (2008). Learning long-term dependencies with recurrent neural networks. Neurocomputing, 71(13-15), 2481-2488.

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. E. (2013). A survey of forecast error measures. World applied sciences journal, 24(24), 171-176.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. Anesthesia & analgesia, 126(5), 1763-1768.

Serra, Y. L. (2018). Precipitation measurements from the Tropical Moored Array: A review and look ahead. *Quarterly Journal of the Royal Meteorological Society*, *144*, 221-234.

Setti, S., Yumnam, K., Rathinasamy, M., & Agarwal, A. (2023). Assessment of satellite precipitation products at different time scales over a cyclone prone coastal river basin in India. Journal of Water and Climate Change, 14(1), 38-65.

Singh, J., Knapp, H. V., Arnold, J. G., & Demissie, M. (2005). Hydrological modeling of the Iroquois river watershed using HSPF and SWAT 1. JAWRA Journal of the American Water Resources Association, 41(2), 343-360.

Sokol, Z., Szturc, J., Orellana-Alvear, J., Popova, J., Jurczyk, A., & Célleri, R. (2021). The role of weather radar in rainfall estimation and its application in meteorological and hydrological modelling—A review. Remote Sensing, 13(3), 351.

Song, H. J., Lim, B., & Joo, S. (2019). Evaluation of rainfall forecasts with heavy rain types in the high-resolution unified model over South Korea. *Weather and Forecasting*, *34*(5), 1277-1293.

Speight, L., and Krupska, K. (2021). Understanding the impact of climate change on inland flood risk in the UK. Weather, 76(10), 330-331.

Su, J., Li, X., Ren, W., Lü, H., & Zheng, D. (2021). How reliable are the satellite-based precipitation estimations in guiding hydrological modelling in South China?. Journal of Hydrology, 602, 126705.

Sungmin, O., Foelsche, U., Kirchengast, G., Fuchsberger, J., Tan, J., & Petersen, W. A. (2017). Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria. Hydrology and Earth System Sciences, 21(12), 6559–6572.

Talchabhadel, R., Shah, S., & Aryal, B. (2022). Evaluation of the spatiotemporal distribution of precipitation using 28 precipitation Indices and 4 IMERG datasets over Nepal. *Remote Sensing*, *14*(23), 5954.

Tam, T. H., Abd Rahman, M. Z., Harun, S., Hanapi, M. N., & Kaoje, I. U. (2019). Application of satellite rainfall products for flood inundation modelling in Kelantan River Basin, Malaysia. Hydrology, 6(4), 95.

Tang, G., Clark, M. P., Papalexiou, S. M., Ma, Z., & Hong, Y. (2020). Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. Remote sensing of environment, 240, 111697.

Tang, X., Yin, Z., Qin, G., Guo, L., & Li, H. (2021). Integration of satellite precipitation data and deep learning for improving flash flood simulation in a poor-gauged mountainous catchment. Remote Sensing, 13(24), 5083.

Tanim, A. H., McRae, C. B., Tavakol-Davani, H., & Goharian, E. (2022). Flood detection in urban areas using satellite imagery and machine learning. Water, 14(7), 1140.

Tavakol-Davani H, Burian SJ, Devkota J, & Apul D (2016). Performance and Cost-Based Comparison of Green and Gray Infrastructure to Control Combined Sewer Overflows. Journal of Sustainable Water in the Built Environment, 2(2), 04015009–1–12. 10.1061/JSWBAY.0000805

Thirumalaiah, K., & Deo, M. C. (1998). River stage forecasting using artificial neural networks. Journal of Hydrologic Engineering, 3(1), 26-32.

Tilford KA, Sene K, Chatterton JB, Whitlow C. 2003. Flood forecasting–real time modelling. Defra/Environment Agency R&D Technical Report W5C-013/5/TR, Environment Agency, Bristol; 246.

Tohka, J., & Van Gils, M. (2021). Evaluation of machine learning algorithms for health and wellness applications: A tutorial. Computers in Biology and Medicine, 132, 104324.

Tripathy, S. S., Vittal, H., Karmakar, S., & Ghosh, S. (2020). Flood risk forecasting at weather to medium range incorporating weather model, topography, socio-economic information and land use exposure. Advances in Water Resources, 146.

UNFCCC (2016). Taking Action. United Nations - Sustainable Development Goals. Retrieved in 21 Jan 2024 from: https://www.un.org/sustainabledevelopment/climate-action/ParisAgreement

Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. Artificial Intelligence Review, 53(8), 5929-5955.

Wang, J. (2017). Combined sewer overflows impact on water quality and environmental ecosystem in the Harlem River; American Geophysical Union, US

Van, S. P., Le, H. M., Thanh, D. V., Dang, T. D., Loc, H. H., & Anh, D. T. (2020). Deep learning convolutional neural network in rainfall–runoff modelling. Journal of Hydroinformatics, 22(3), 541-561.

Wang, C., Tang, G., Han, Z., Guo, X., & Hong, Y. (2018). Global intercomparison and regional evaluation of GPM IMERG Version-03, Version-04 and its latest Version-05 precipitation products: Similarity, difference and improvements. Journal of hydrology, 564, 342-356.

Wang, L., Cui, S., Li, Y., Huang, H., Manandhar, B., Nitivattananon, V., Fang, X., & Huang, W. (2022). A review of the flood management: from flood control to flood resilience. In Heliyon (Vol. 8, Issue 11). Elsevier Ltd.

Wang, Z., Zhong, R., Lai, C., & Chen, J. (2017). Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility. Atmospheric Research, 196, 151-163.

Wardah, T. et al. (2008) Use of geostationary meteorological satellite images in convective rain estimation for flash-flood forecasting, Journal of Hydrology, 356(3–4), pp. 283–298.

Watters, D. C., Huffman, G. J., Gatlin, P. N., Kirstetter, P. E., Bolvin, D. T., Joyce, R., ... & Wolff, D. B. (2025). A Diagnosis of Oceanic Precipitation in IMERG-GMI. Journal of Hydrometeorology, 26(4), 459-480.

Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G., and Roberts, N. M. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall, Rev. Geophys., 52, 522– 555.

Willems, P. (2012) Impacts of Climate Change on Rainfall Extremes and Urban Drainage Systems, Water Intelligence (online).

WMO (2020) – Observation components of the Global Observing System. World Meteorological Organization (online). Available at: https://library.wmo.int/records/item/68835-state-of-the-global-climate-2023

WMO (2023a). The Global Climate 2011-2020 – A decade of accelerating climate change. World Meteorological Organization Report, WMO-No.1338 (online). Available at: https://library.wmo.int/records/item/68585-the-global-climate-2011-2020

WMO (2023b). Guide to Meteorological Instruments and Methods of Observation, Volume I – Measurement of Meteorological Variables. World Meteorological Organization, WMO - No.8 (online). Available at: https://library.wmo.int/idurl/4/68695

WMO (2024). State of the Global Climate 2023. World Meteorological Organization Report, WMO-No.1347. https://library.wmo.int/records/item/68835-state-of-the-global-climate-2023

Wu, S. J., Yeh, K. C., Ho, C. H., & Yang, S. H. (2016). Modeling probabilistic lag time equation in a watershed based on uncertainties in rainfall, hydraulic and geographical factors. Hydrology Research, 47(6), 1116-1141.

Xiang, Z., Nengcheng, C., Chuli, H., & Xiaoting, P. (2018). Spatio-temporal distribution of three kinds of flash droughts over agricultural land in China from 1983 to 2015. *Advances in Earth Science*, *33*(10), 1048.

Xiang, Z., Yan, J., & Demir, I. (2020). A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water resources research, 56(1), e2019WR025326.

Xiao, S., Xia, J., & Zou, L. (2020). Evaluation of multi-satellite precipitation products and their ability in capturing the characteristics of extreme climate events over the Yangtze River Basin, China. *Water*, *12*(4), 1179.

Xu, W., Jiang, Y., Zhang, X., Li, Y., Zhang, R., & Fu, G. (2020). Using long short-term memory networks for river flow prediction. Hydrology Research, 51(6), 1358-1376.

Yeditha, P. K., Kasi, V., Rathinasamy, M., & Agarwal, A. (2020). Forecasting of extreme flood events using different satellite precipitation products and wavelet-based machine learning methods. Chaos: An Interdisciplinary Journal of Nonlinear Science, 30(6)..

Yeditha, P. K., Rathinasamy, M., Neelamsetty, S. S., Bhattacharya, B., & Agarwal, A. (2022). Investigation of satellite precipitation product driven rainfall-runoff model using deep learning approaches in two different catchments of India. Journal of Hydroinformatics, 24(1), 16-37.

Yuan, F., Wang, B., Shi, C., Cui, W., Zhao, C., Liu, Y., ... & Yang, X. (2018). Evaluation of hydrological utility of IMERG Final run V05 and TMPA 3B42V7 satellite precipitation products in the Yellow River source region, China. Journal of hydrology, 567, 696-711.

Yves, A. R., Abbas, O., Sadek, M., Shahrour, I., & Fadi, H. C. (2017). Flood forecasting within urban drainage systems using NARX neural network. Water Science and Technology, 76(9), 2401-2412.

Zhang, D., Lindholm, G., & Ratnaweera, H. (2018). Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. Journal of hydrology, 556, 409-418.

Zhang, Y., Stensrud, D. J., & Zhang, F. (2019). Simultaneous Assimilation of Radar and All-Sky Satellite Infrared Radiance Observations for Convection-Allowing Ensemble Analysis and Prediction of Severe Thunderstorms; Monthly Weather Review, 147(12), 4389-4409. Retrieved Jan 1, 2022, from https://journals.ametsoc.org/view/journals/mwre/147/12/mwr-d-19-0163.1.xml

Zhang, Y., & Thorburn, P. J. (2022). Handling missing data in near real-time environmental monitoring: A system and a review of selected methods. Future Generation Computer Systems, 128, 63-72.

Zhou, Z. H. (2003). Three perspectives of data mining. Z.-H. Zhou, Artificial Intelligence, vol. 143, no. 1, pp. 139-146, 2003.

Zhu, S., Wei, J., Zhang, H., Xu, Y., & Qin, H. (2023). Spatiotemporal deep learning rainfall-runoff forecasting combined with remote sensing precipitation products in large scale basins. Journal of Hydrology, 616, 128727.

8 Appendices

Appendix A – Summary of accessed literature

Table A1 - Overview of background information from literature review

Subject	Method	Key findings	Key points and relevant gaps	Reference
Flooding intensification and land use	Trend analysis and runoff modelling	Changes in land use has considerable impact on flood peak discharge	Urban areas face increased susceptibility to flooding	Saghafian et al. (2008)
Urban drainage	Review of current methods for assessing future changes in urban rainfall extremes and their impact in urban drainage systems	In spite of significant advances there are still many limitations in our understanding of precipitation, under climate change, to design and operate urban drainage.	Most big metropolitan areas are predicted to become increasingly vulnerable to heavy rainfall, land use and drainage design.	Willems et al., (2012)
Climate change in UK	Observation of weather and climate. Generation of climate models.	Total rainfall from extremely wet days increased by around 17% in a decade in UK	Increase in intensity and severity of convective and frontal storms	Murphy et al., (2010)
Climate related extreme events	Bayesian framework analysis and testing of nonstationarities in extreme events	For temperature and precipitation data a NRT trend was the only model to perform better than the stationary one.	The current evidence is that hydroclimatic system is non-stationary GAP: is necessary to apply methods that allow testing in non-stationary conditions	Garcia-Aristizabal et al., (2015)
Climate time-series data	Bayesian framework	Substantial evidence shows that the climate is non-stationary, possibly due to anthropogenic climate change	GAP: There is a need for statistical models that explicitly allow for non-stationarity	Cheng and Aghakouchak (2014)
Climate-related nonstationarity	Empirical approach using long-term hydrologic records	Nonstationarity highlights the importance of the continuity of hydrologic records and the need for	Nonstationarity is due to a variety of reasons: urbanisation, ground-water development, climate change, etc.	Hirsch, (2011)

		repeated analysis of the data as the time series grow	GAP: a need to rethink planning and operations approach	
Urban flooding	Critical examination of three alternatives in response to urban flooding	This paper advances an alternative design perspective that advocates 'working with' as opposed to 'dominating' or 'adapting to' nature.	Solutions based on historical data do not work for the new conditions imposed by climate change	Lennon et al., (2014)
Combined Sewer System real time control	SWMM modelling of smart gates operation in combined sewers	The use of smart gates provides positive effects on the overall hydraulic performance of the drainage system	Areas experiencing heavy rainfall can transfer some of the excess volume to other areas where full capacity has not yet been reached using smart gates	Carbone et al., (2014)
Comparison green/gray scenario on CSO control	Long- term continuous hydrologic and hydraulic simulations and lifecycle cost analysis techniques	Combining uniform RWH with grey facilities improves performance on CSO control.	Further optimization studies on RWH system characteristics may improve the performance of the RWH and reduce the costs	Tavakol-Davani et al., (2016)
Environmental Engineering	Environmental Engineering: Water, Wastewater, Soil and Groundwater	It covers historical development of the concept of sustainability, contemporary debates about how to achieve it, and obstacles and the prospects for overcoming them.	Water conservation strategies can give a simple and cost-effective way of decreasing hydraulic loads	Nemerow et al., (2009)
CSO control	Intercept, temporarily detain and release rainfall-runoff during and after a storm event.	Synchronised inhale then exhale strategy (SITES) can meet the required performance standards at much lower cost than conventional concrete facilities	Aims to use a network of green and grey elements to collect, retain and release storm water on the collection system at a chosen time.	Mancipe et al., (2014)
Flood risk management	"Sponge City" concept explanation	To achieve better flood control, more effective development and implementation of	Use the benefits of Low Impact Development (LID) to expand water infiltration and storage during rainstorm	Chan et al., (2018)

		land-use guidance and assessment tools are recommended	events and reduce CO2 level	
CSO reduction	Laboratory analysis of water samples	Water contamination in the Harlem River is directly impacted residents and community who lives along the river, especially in the Bronx Section	Green wall, green roof, and wetland had been used to reduce stormwater runoff & CSOs in the Bronx River	Wang, J. (2017)
AI	Non-dominated sorting generic algorithm applied to search Pareto-optimal solutions	Derived intelligent urban flood drainage system can serve as reliable and efficient operational strategies for urban flood management and flood risk mitigation	Prediction-based AI simulation to produce superior results, i.e., smarter, more robust, and cost-effective solutions in urban water systems	Dong & Yang, (2019)
ANN	ANN model trained with pseudo-inverse rule	Used artificial neural networks (ANN) as an alternative to hydraulic models and capable of predicting combined sewer overflow depth with less than 5% error for predictions more than 1 hour ahead for unseen data	Rainfall intensity data from other sources can be used to improve on this approach	Mounce et al., (2014)
AI	Big data and pattern mining	Traditional algorithms for pattern extraction usually present some issues and difficulties during large-scale data processing. Parallel and distributed mining have arisen to solve this problem.	Because of their capacity to find regression patterns across different sets of data, AI may also assist with big data challenges	Kumar and Mohbey., (2019)
Flood forecasting	NARX neural network-based tool	Trained NARX performed well on minor and severe storm events, making a robust and efficient tool for flood forecast	Develop a flood forecasting system capable of linking variations in water depth in manholes to rainfall intensities	Yves et al., (2017)
Satellite images and rainfall	CNN inception-v3 model for forecast levelling	The training from scratch of inception-v3 with areas of	Succeeded in finding the relationship between satellite	Boonyue et al., (2019)

	of daily rainfall	satellite image got highest accuracy from previous work	images and daily rainfall data by using convolutional neural networks up to 3 days ahead of the event	
Machine learning	Review of machine learning techniques applied to aspects of hydrological cycle and mitigation of water hazards	Many aspects of urban water security and hazard modelling are still underrepresented as ML problems.	A review of machine learning methods for mitigation of urban water hazards that describes numerous studies related to rainfall predictions	Allen-Dumas et al., (2021)
Satellite weather forecast	Ensemble analysis and forecast using satellite data associated and radar data compared with available observations	The assimilation of both BT and Vr observations yields the best ensemble forecasts, providing better confidence, accuracy, and lead times on prediction of midlevel mesocyclones	Satellites BT has " the potential to increase the forecast and warning lead times of severe weather events compared with radar observations	Zhang et al. (2019)
Satellite images and rainfall	PERSIANN and CMORPH	Both methods proved to be better using satellite information compared with use of other sources like radar or ground measurements	Succeed to predict extreme weather	Joyce et al., (2004); Hsu et al.,(1999); Behrangi et al., (2009)
Extreme flood events forecasting	Satellite data as input on machine learning models	Machine learning methods applied to meteorological satellite images have potential to forecast extreme flood events with longer lead times	Prediction of extreme flood events or flash floods and identification of flood risk areas	Yeditha et al., (2020); Wardah et al., (2008)
Satellite based rainfall products	Satellite data in conjunction with AI technology Applied to monitor river and stream water levels	Bias correction of satellite rainfall estimates significantly improves model simulations. Satellite rainfall products can support rapid flood forecasting required for early warning systems and is a	GAPS: *On AI-based application for flood prediction, more research is needed to verify models * Future research should use rainfall data from multiple satellite sources	Tam et al., (2019); Adane et al., (2021); Nourani et al., (2021); Abdollahipour et al., (2021); Bitew et al., (2012);

		promising option for accurate modelling of rainfall-runoff.	*Need to downscale the satellite original resolution and apply deep-learning convolutional neural networks	
Flash flood	Satellite datasets and digital surface models linked to machine learning	Recommends integration of various training datasets and machine learning algorithms for effective forecasting of water depth from satellite data	Able to detect region's water depth to prioritise victim rescue and dispatch resources	Ghorbanzadeh et al., (2018);
Data mining technique – land subsidence assessment	GIS, neuro-fuzzy inference system and k-fold cross validation	The choice of training dataset and the membership functions significantly affects the results.	GAP: future research needs to integrate various training datasets and machine learning algorithms	Elkhrachy I., (2022)
Real-time flood forecasting	Ensembled machine learning, data mining and decision tree techniques	Accurate classification of flood/non-flood events. Rainfall duration and intensity have greater impact in accuracy than other characteristics	High hit rate for 3hs forecasting. GAP: the model uses only rain-gauge based rainfall data	Piadeh et al., (2023)

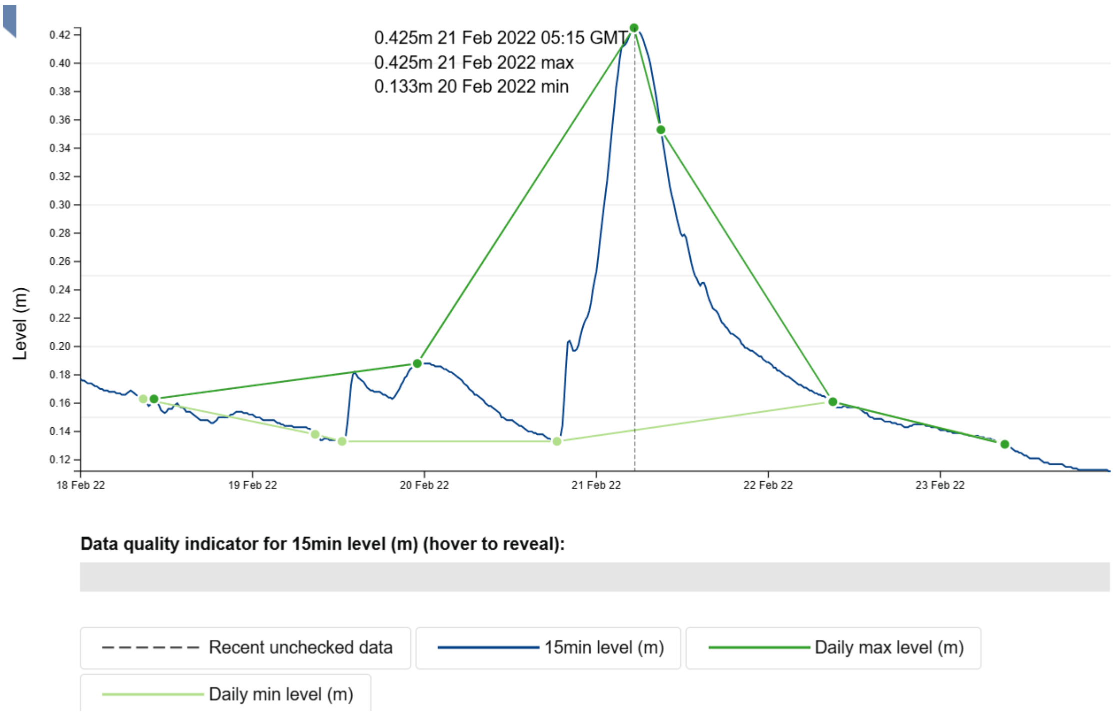
Appendix B – Dataset's scrutiny additional material

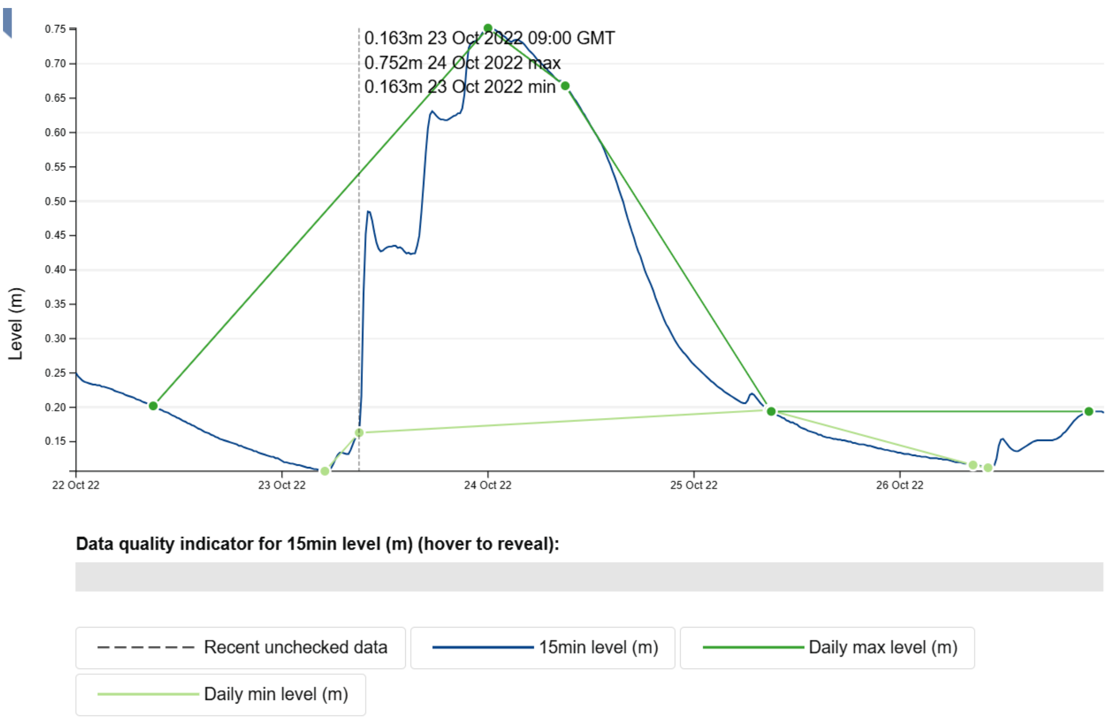
Links for access to the videos produced for scrutiny of the IMERG dataset in image format. The videos allow for better visualization of the anomalous areas and data discrepancies discussed in the main text: https://youtu.be/mGg5DSyI4Ro

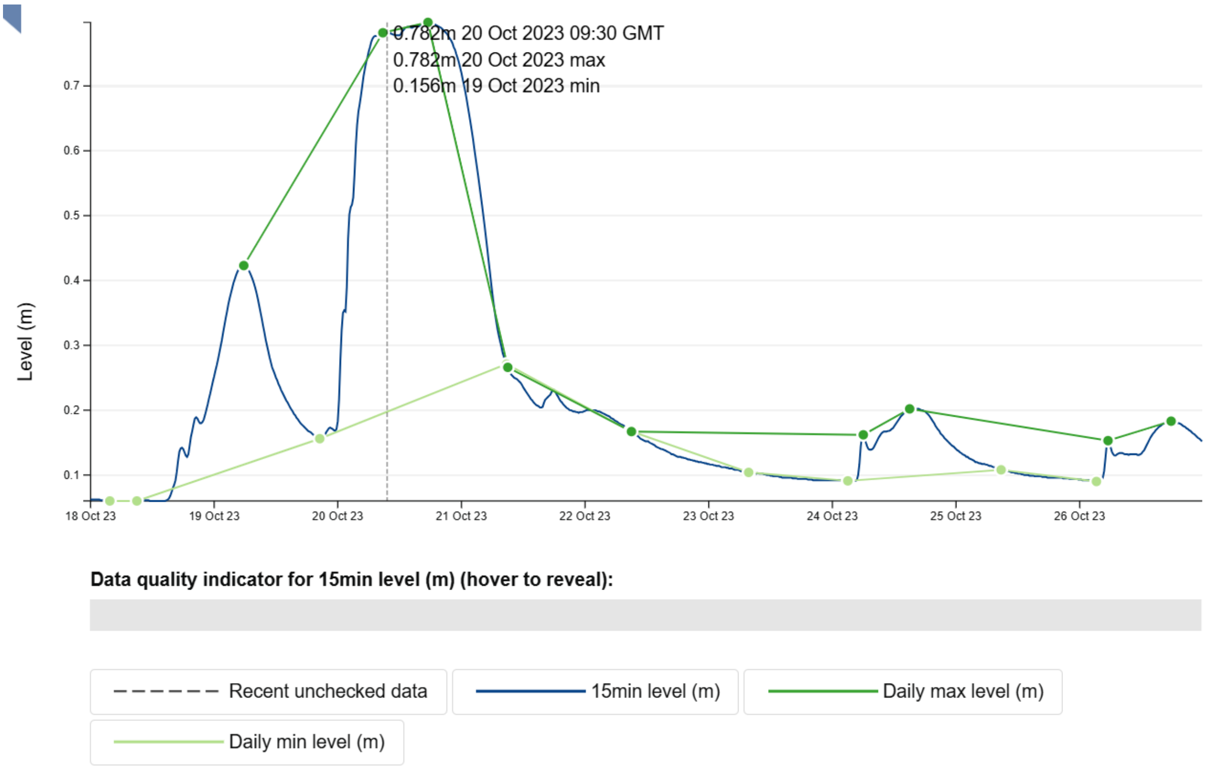
https://youtu.be/yILArU2Vmvo (recommended to watch with 2x playback speed)

Appendix C – Case study additional information

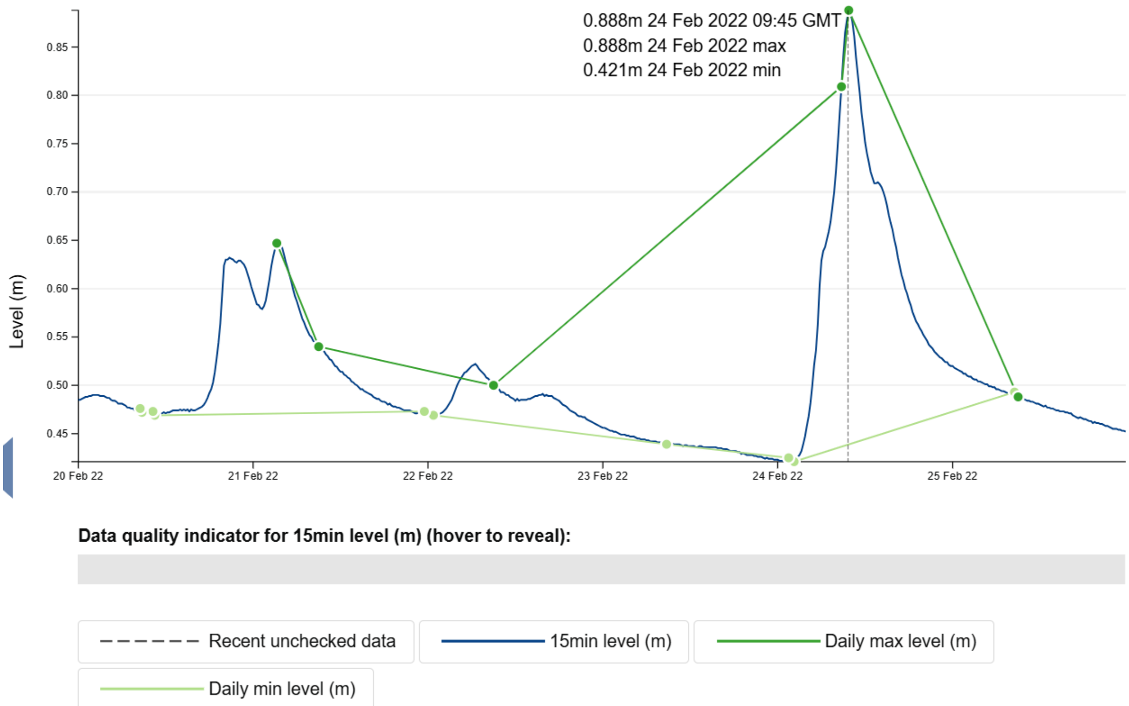
C1 – River Crane level records at Cranford Park station (C1) during the period of the selected rainfall events (Source: DEFRA 2023).

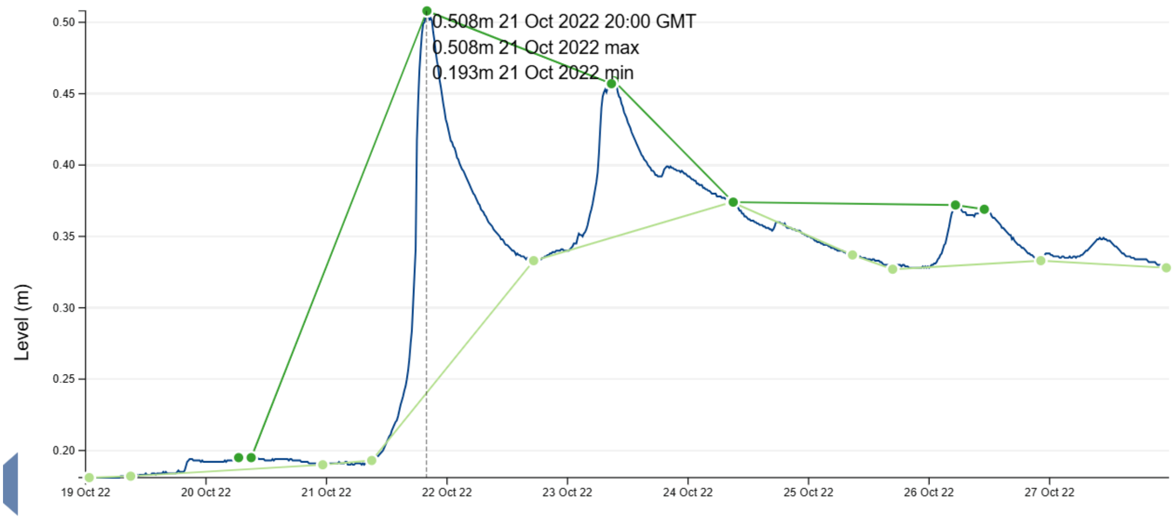






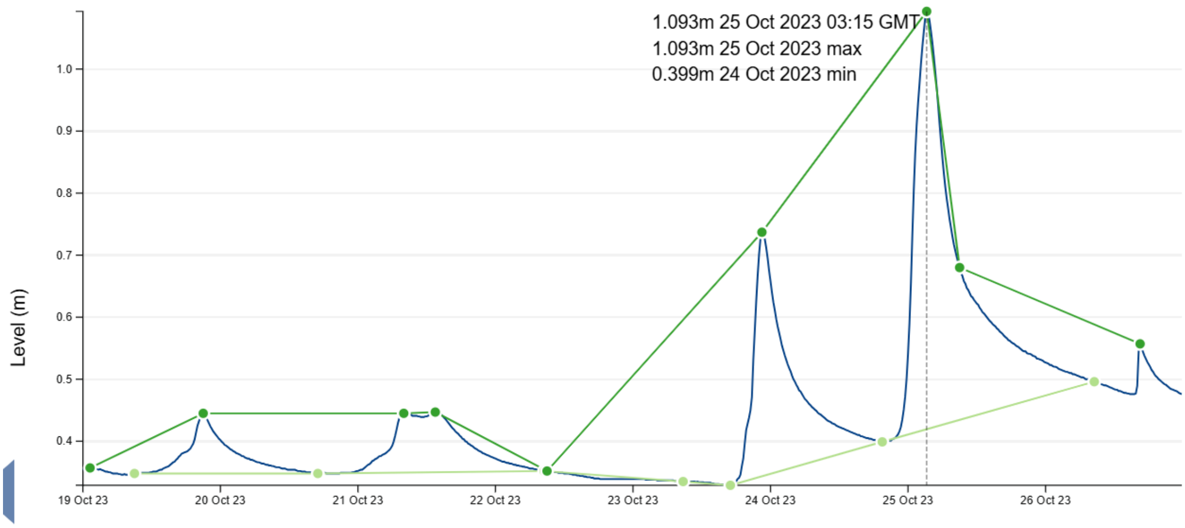
C2 - River Lew level records at Gribbleford Bridge station (C2) during the period of the selected rainfall events (Source: DEFRA 2023).





Data quality indicator for 15min level (m) (hover to reveal):

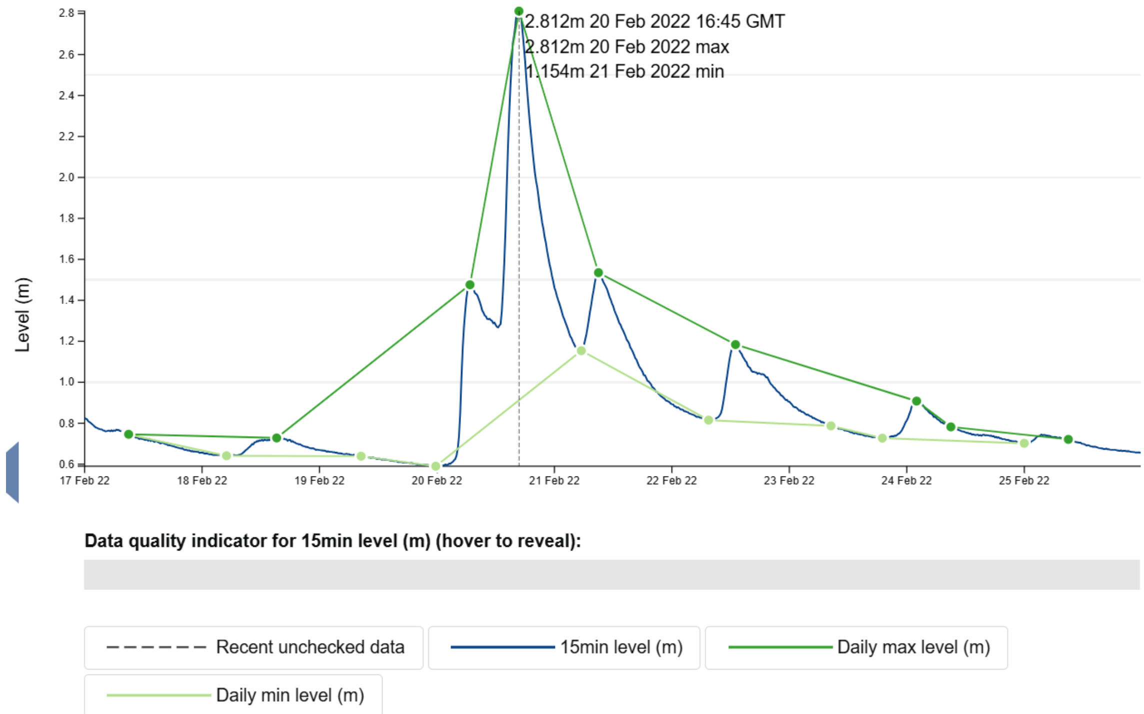
Recent unchecked data
15min level (m)
Daily max level (m)
Daily min level (m)

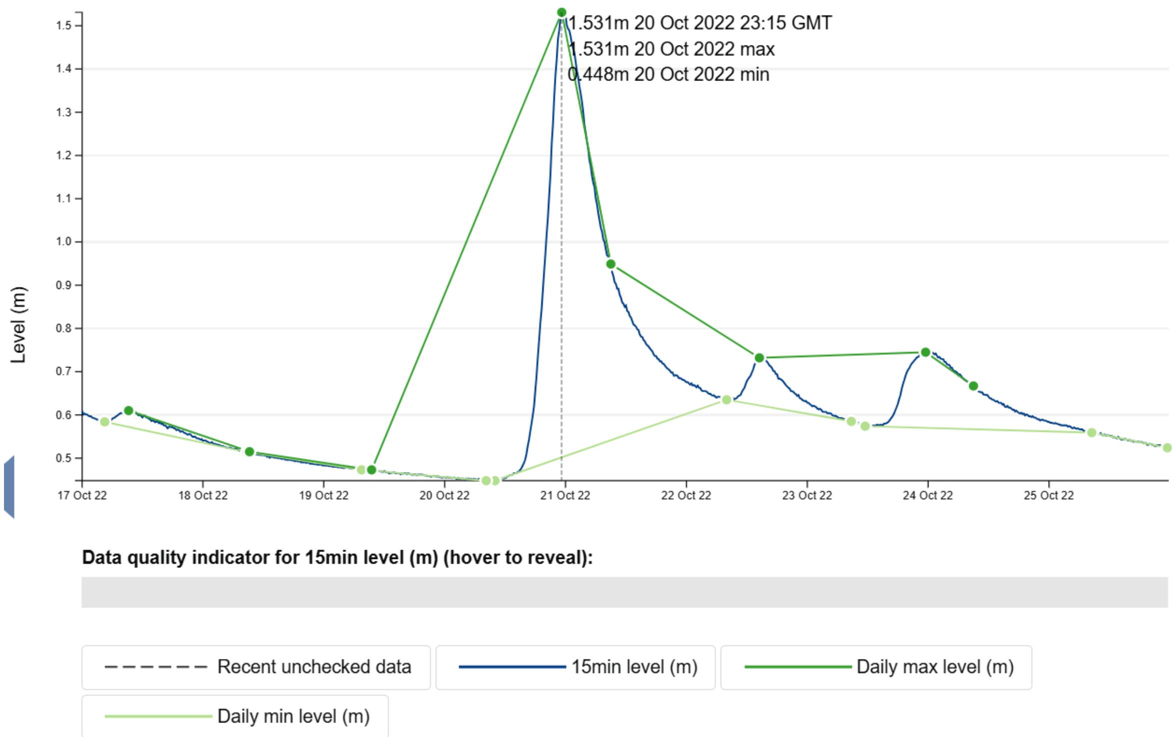


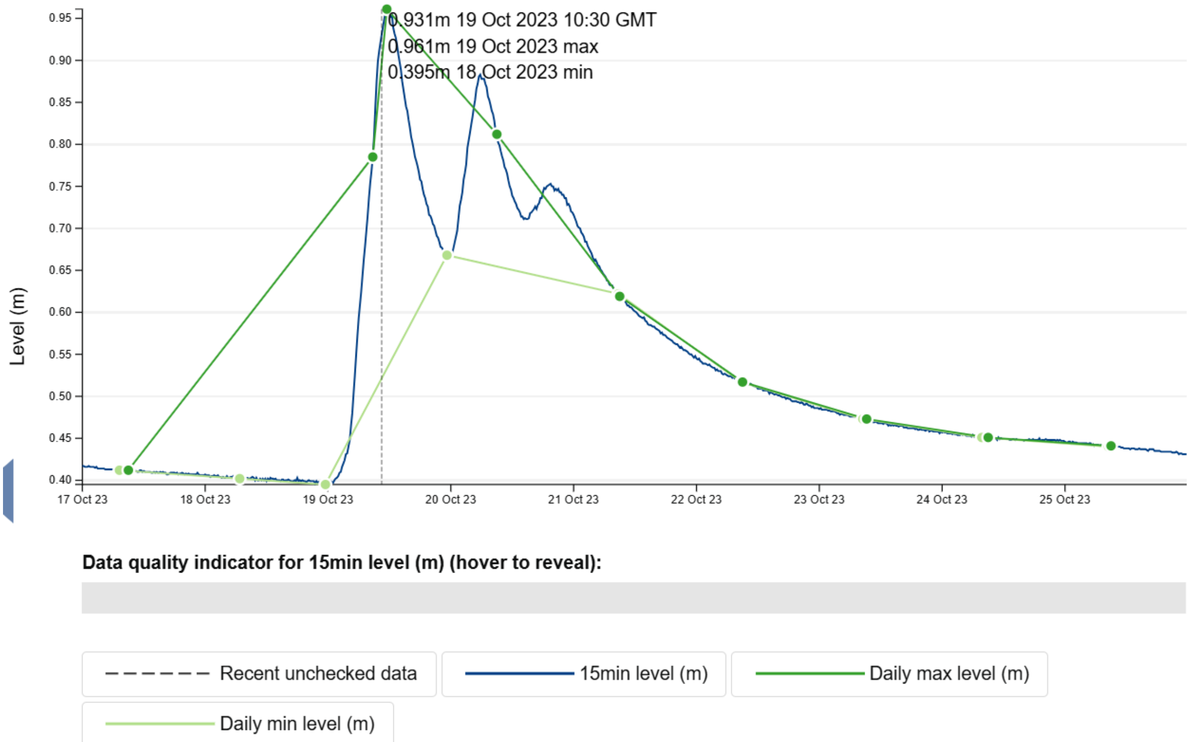
Data quality indicator for 15min level (m) (hover to reveal):

Recent unchecked data
15min level (m)
Daily max level (m)
Daily min level (m)

C3 – River Caldew level records at Cummersdale station (C3) during the period of the selected rainfall events (Source: DEFRA 2023).







Appendix D – Results for catchments C3

LSTM model predictions of river Caldew levels at Cummersdale (C3) using data from the closes rain gauge, at Catlowdy station, and IMERG data from five pixels located between the 54.70°N and 55.10°N latitudes spread along the 10.55°W longitudinal line. Given that the pixels group 10.55 is positioned at approximately 485Km from the C3 catchment, rainfall data is collected from the group with at least 4 hours lead time for an event with forward speed of 60Km/h.

Table D1 – Performance metrics of the LSTM model on predictions at C3 with gauge and IMERG precipitation data

C3	Catlowdy rain-gauge		IMERG-Group 10.55	
Lead-time	**RMSE (mm)**	**MAE (mm)**	**RMSE (mm)**	**MAE (mm)**
0.5hs	10.74	4.85		
1.0hs	15.72	6.24		
1.5hs	23.24	8.08		
2.0hs	30.87	10.05		
2.5hs	38.25	12.27		
3.0hs	45.12	14.43		
3.5hs	51.28	16.49		
4.0hs	56.85	18.26		
4.5hs	61.92	19.93	9.82	5.28
5.0hs	66.53	21.64	15.44	7.30
5.5hs	70.76	23.38	23.14	10.09
6.0hs	74.70	25.03	31.29	12.76
6.5hs	78.35	26.60	39.38	15.24
7.0hs	81.77	28.06	47.04	17.36
7.5hs	84.99	29.44	54.23	19.61
8.0hs	88.00	30.77	60.88	21.73
8.5hs	90.84	32.02	67.05	23.83
9.0hs	93.52	33.16	72.80	25.96
9.5hs	96.06	34.24	78.12	28.26
10.0hs	98.46	35.34	83.02	30.59
10.5hs	100.73	36.44	87.59	33.03
11.0hs	102.87	37.56	91.78	35.35
11.5hs	104.91	38.64	95.56	37.61
12.0hs	106.83	39.75	98.95	39.62
12.5hs			101.88	41.20
13.0hs			104.47	42.47
13.5hs			106.73	43.63
14.hs			108.73	44.85
14.5hs			110.52	45.84
15.0hs			112.14	46.57
15.5hs			113.67	47.55
16.0hs			115.11	48.51
Overall	**69.72**	**24.28**	**76.22**	**30.18**

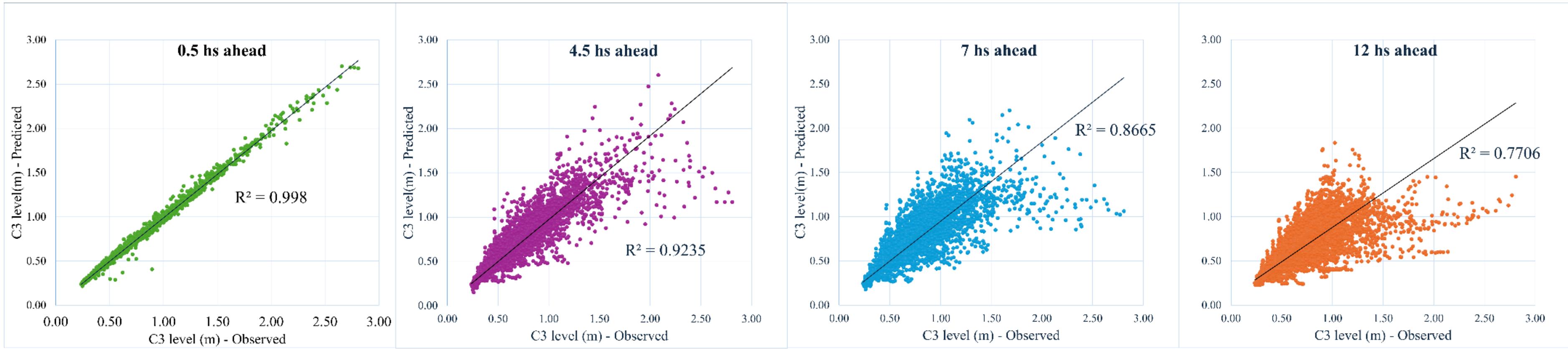


Figure D2 - R-squared values for multi-step-ahead LSTM model predictions using the Catlowdy rain-gauge dataset.

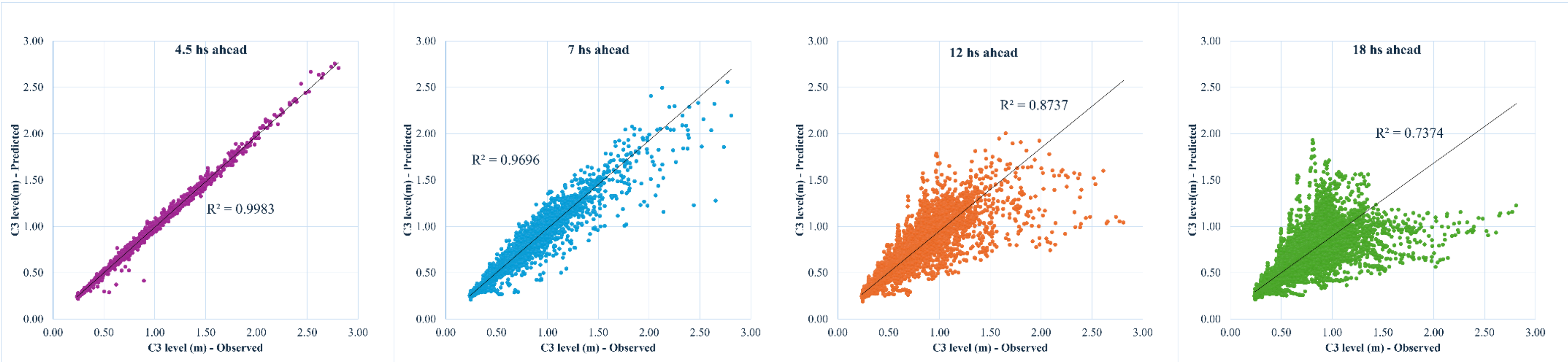


Figure D3 - R-squared values for multi-step-ahead LSTM model predictions using the IMERG 10.55 pixels-group dataset.

Table D4 – Performance metrics of the LSTM model on event-based predictions using rain-gauge and IMERG 10.55 group datasets

C3 - LSTM	*Febr 2022 Catlowdy gauge*		*October 2022 Catlowdy gauge*		*October 2023 Catlowdy gauge*		*Febr 2022 IMERG - Group 10.55*		*October 2022 IMERG - Group 10.55*		*October 2023 IMERG – Group 10.55*	
lead-time	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
0.5hs	45.41	28.74	8.10	6.35	12.53	8.88						
1.0hs	82.14	41.62	9.71	7.69	15.15	10.85						
1.5hs	132.33	61.67	12.61	9.74	19.03	13.36						
2.0hs	182.55	87.87	16.57	11.97	24.62	16.57						
2.5hs	229.67	113.54	21.54	14.67	32.16	21.15						
3.0hs	270.11	136.38	27.24	17.69	41.29	26.46						
3.5hs	303.04	153.60	32.90	20.87	51.21	32.24						
4.0hs	329.35	167.42	38.51	23.76	61.56	38.32						
4.5hs	351.49	180.43	44.06	26.57	72.01	44.29	30.85	24.80	6.88	5.28	16.79	12.54
5.0hs	370.96	194.81	48.74	28.97	82.55	50.57	62.57	39.41	10.72	7.74	22.10	15.74
5.5hs	387.88	208.39	52.77	31.15	92.86	57.10	114.57	61.39	15.31	10.32	29.05	21.17
6.0hs	401.27	219.62	56.34	33.31	102.62	63.56	166.33	86.67	20.59	13.54	36.99	27.12
6.5hs	411.85	228.87	59.10	35.19	111.88	70.09	215.42	111.55	26.04	16.83	44.92	33.36
7.0hs	420.48	238.82	61.16	36.81	120.51	76.49	259.14	135.72	31.40	20.40	52.60	39.40
7.5hs	427.98	247.49	62.88	38.33	128.35	82.33	297.35	157.96	36.46	23.69	60.04	45.60
8.0hs	433.73	253.72	64.08	39.71	135.40	87.69	329.49	177.98	41.05	26.73	66.93	51.39
8.5hs	437.26	257.56	64.91	41.03	141.78	92.94	356.28	195.96	45.30	29.90	73.45	56.93
9.0hs	438.70	260.98	65.28	42.29	147.62	97.75	378.27	211.51	48.82	32.80	79.02	62.24
9.5hs	439.06	266.15	65.79	43.59	152.91	102.22	395.47	222.84	51.52	35.31	84.12	67.08

10.0hs	439.32	274.50	66.33	44.85	157.63	106.36	407.43	231.05	54.03	38.05	89.10	71.21
10.5hs	440.88	282.12	66.95	46.11	161.70	110.10	415.34	237.97	56.41	40.78	93.82	74.99
11.0hs	443.25	289.97	67.75	47.55	165.19	113.48	420.37	244.40	58.90	43.71	98.40	78.25
11.5hs	446.85	297.32	68.74	49.05	168.24	116.28	423.78	250.94	61.49	46.80	102.33	81.36
12.0hs	450.37	303.46	69.98	50.76	170.80	118.56	426.02	257.41	64.05	49.71	106.03	84.74
12.5hs							427.42	262.81	66.13	52.20	109.05	87.96
13.0hs							428.82	267.33	67.89	54.06	111.50	89.68
13.5hs							430.93	274.43	69.34	55.47	113.38	90.81
14.hs							433.21	281.50	71.12	57.09	114.90	91.76
14.5hs							435.53	288.66	73.31	59.05	115.94	92.78
15.0hs							438.79	295.97	75.56	61.10	116.83	94.29
15.5hs							443.34	303.69	78.08	63.43	117.41	95.73
16.0hs							449.06	312.07	80.84	66.02	117.91	96.85
Overall	**346.497**	**199.7948**	**48.00101**	**31.16684**	**133.3675**	**84.05199**	**341.0742**	**205.5844**	**50.46868**	**37.91737**	**82.19185**	**65.12395**

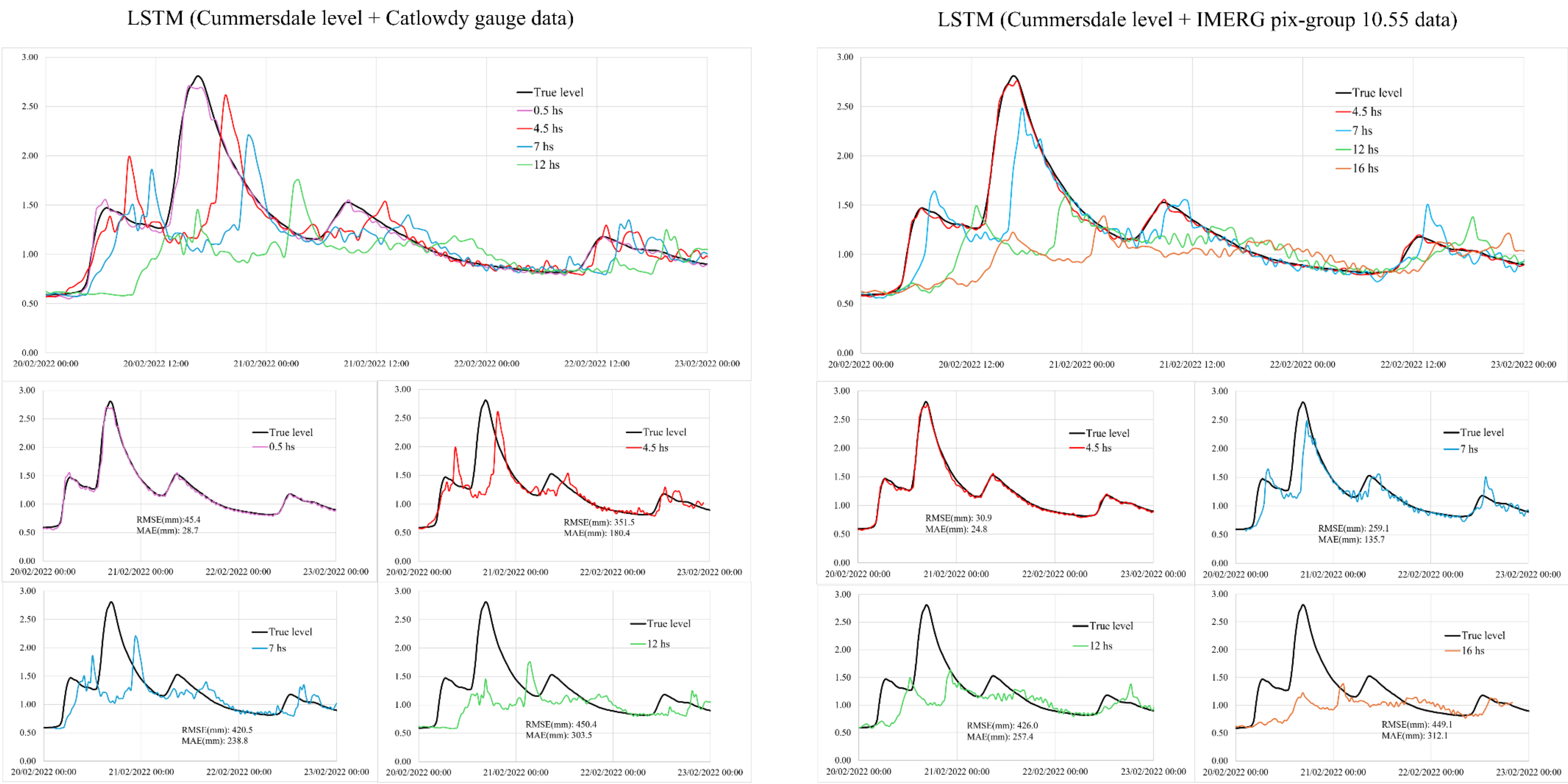


Figure D5 – Graphical comparison of true level and model predictions at C3 for several lead times during the February 2022 event.

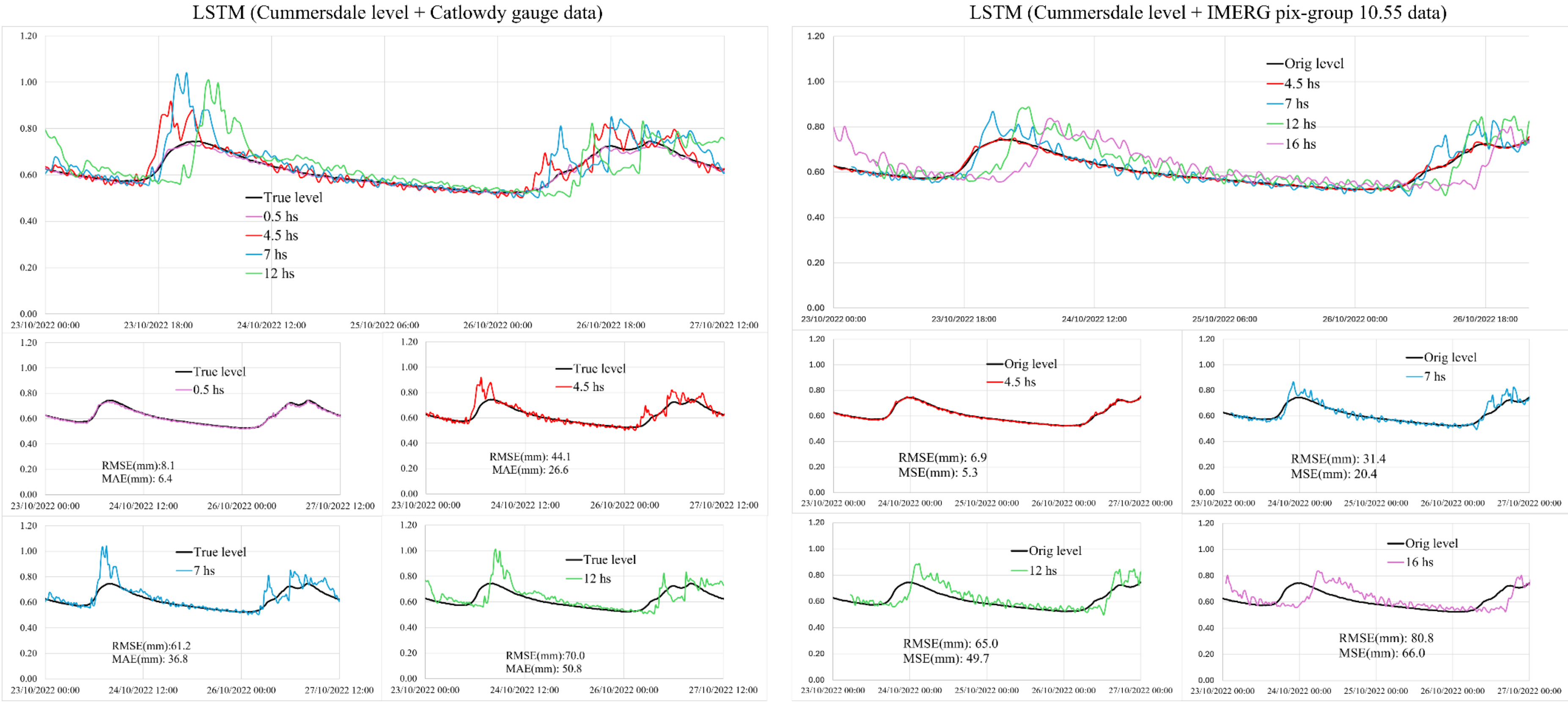


Figure D6 – Graphical comparison of true level and predictions at C3 for several lead times during the October 2022 event.

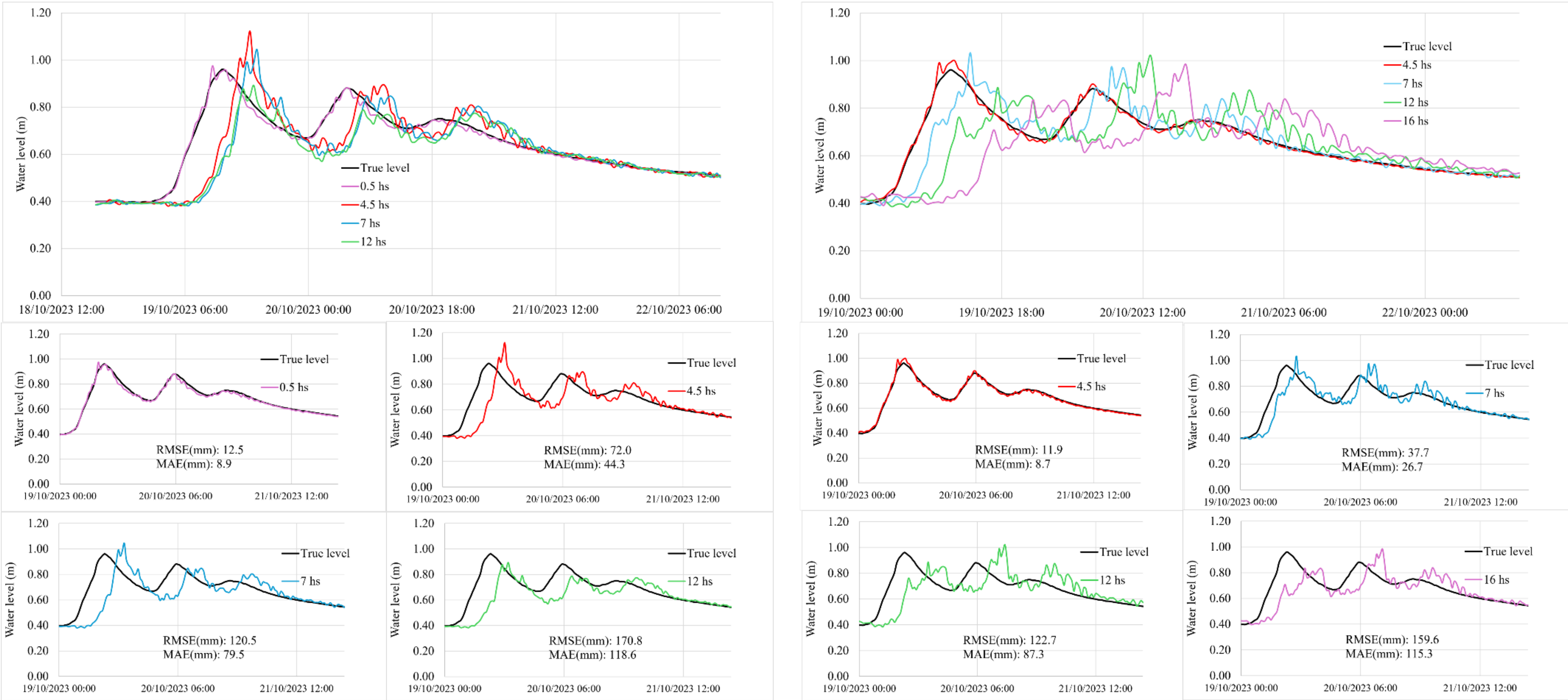


Figure C7 – Graphical comparison of true level and predictions at C3 for several lead times during the October 2023 event.

Appendix E – Results for catchments C2

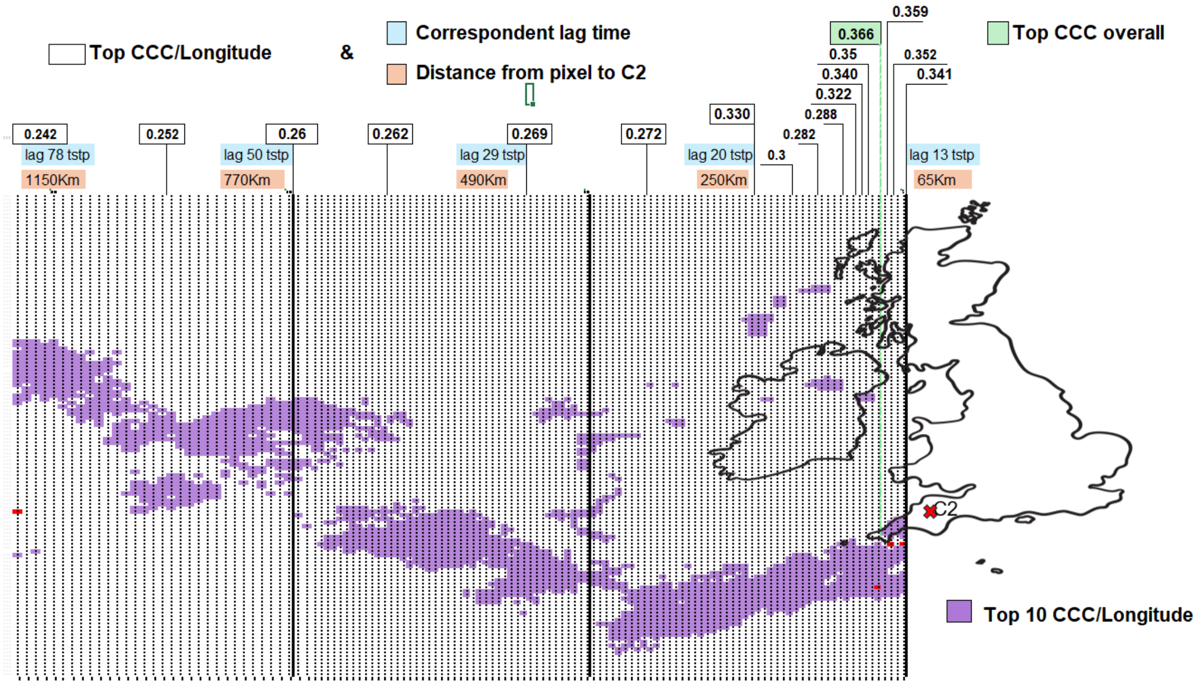


Figure E1 - Top 10% CCC values at increasing distances from the catchment C2.

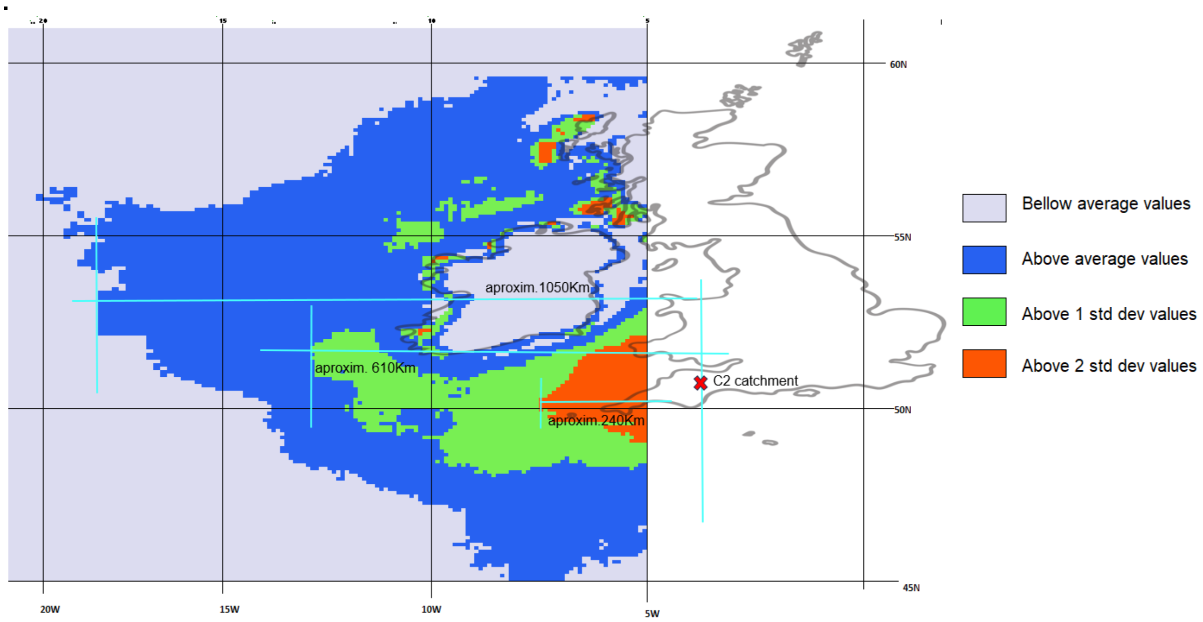


Figure E2 - Distribution of IMERG pixels across the monitoring region based on the average and standard deviation of CCC values with C2.

LSTM model predictions of river Lew levels at Gribbleford Bridge (C3) using data from the closes rain gauge, at Mary Tavy station, and IMERG data from five pixels located between the 54.50ºN and 54.95ºN latitudes spread along the 8.55ºW longitudinal line. Given that the pixels group 8.55 is positioned at approximately 533Km from the C2 catchment, rainfall data is collected from the group with at least 4.5 hours lead time for an event with forward speed of 60Km/h.

Table E3 – Performance metrics of the LSTM model on predictions at C2 with gauge and IMERG precipitation data

C2	**Mary Tavy rain-gauge**		**IMERG-Group 8.55**	
Lead-time	**RMSE (mm)**	**MAE (mm)**	**RMSE (mm)**	**MAE (mm)**
0.5hs	9.01	3.43		
1.0hs	14.44	4.95		
1.5hs	20.97	7.23		
2.0hs	27.63	10.22		
2.5hs	34.14	13.30		
3.0hs	40.09	15.93		
3.5hs	45.52	17.80		
4.0hs	50.65	19.29		
4.5hs	55.36	20.51	11.01	6.51
5.0hs	59.78	21.72	16.97	7.30
5.5hs	63.93	22.95	25.14	9.27
6.0hs	67.86	24.31	33.68	12.08
6.5hs	71.62	25.79	41.64	14.83
7.0hs	75.12	27.24	48.95	17.72
7.5hs	78.24	28.64	55.51	20.74
8.0hs	80.94	29.95	61.22	23.56
8.5hs	83.26	31.32	66.11	26.16
9.0hs	85.19	32.66	70.04	28.26
9.5hs	86.79	34.04	73.32	30.19
10.0hs	88.11	35.33	76.10	31.93
10.5hs	89.21	36.51	78.45	33.48
11.0hs	90.17	37.58	80.45	34.56
11.5hs	91.05	38.52	82.27	35.61

12.0hs	91.86	39.28	83.87	36.39
12.5hs			85.32	37.21
13.0hs			86.62	38.07
13.5hs			87.79	38.95
14.hs			88.90	39.88
14.5hs			89.93	40.86
15.0hs			90.86	41.79
15.5hs			91.66	42.46
16.0hs			92.40	43.08
Overall	**62.54**	**24.1**	**67.43**	**28.79**

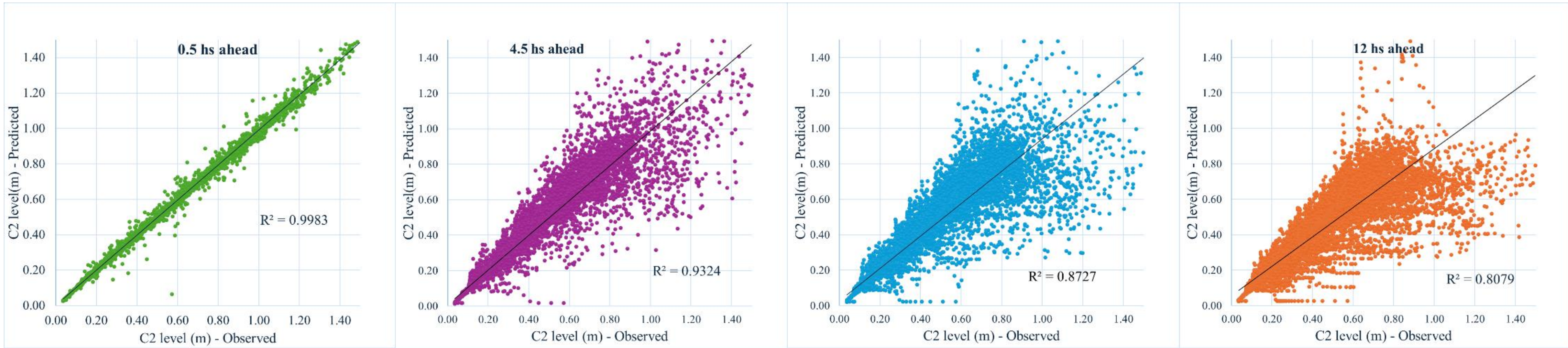


Figure E4 - R-squared values for multi-step-ahead LSTM model predictions using the Mary Tavy rain-gauge dataset.

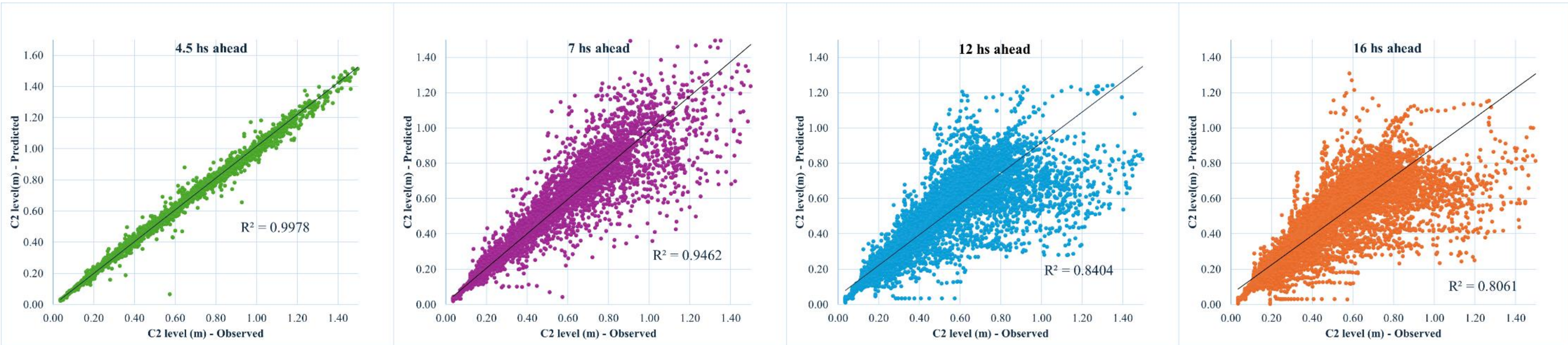


Figure E5 - R-squared values for multi-step-ahead LSTM model predictions using the IMERG 8.55 pixels-group dataset.

Table E6 – Performance metrics of the LSTM model on event-based predictions using rain-gauge and IMERG 8.55 group datasets

C2 - LSTM	October 2022 MaryT gauge		October 2023 MaryT gauge		October 2022 IMERG Group 8.55		October 2023 IMERG Group 8.55	
lead-time	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
0.5hs	4.69	2.68	3.35	2.18				
1.0hs	6.59	3.58	6.67	3.73				
1.5hs	10.51	5.41	11.33	6.26				
2.0hs	19.29	8.26	16.44	9.21				
2.5hs	26.57	11.18	21.76	12.36				
3.0hs	29.65	13.67	26.68	15.38				
3.5hs	30.65	15.50	31.02	18.22				
4.0hs	34.36	17.99	34.79	20.76				
4.5hs	40.73	20.88	37.73	22.65	10.59	8.19	10.77	9.22
5.0hs	47.87	23.59	39.94	24.15	11.57	8.24	13.71	9.82
5.5hs	54.38	26.18	41.55	25.39	14.08	8.93	18.99	11.53
6.0hs	59.80	28.54	42.82	26.58	17.60	11.00	24.92	14.34
6.5hs	64.45	31.02	44.12	27.79	21.00	13.10	29.88	16.86
7.0hs	68.36	33.33	45.45	29.07	25.00	15.75	34.09	19.53
7.5hs	71.66	35.22	46.64	30.25	29.41	18.75	37.78	22.04
8.0hs	74.31	36.92	47.53	31.26	33.53	21.81	40.93	24.29
8.5hs	76.05	38.23	48.14	32.07	37.47	24.96	43.71	26.62
9.0hs	77.83	39.78	48.56	32.70	40.99	28.06	45.74	28.24
9.5hs	79.13	41.26	49.02	33.27	44.62	31.16	47.61	29.81
10.0hs	79.17	42.14	49.57	33.88	48.28	34.08	49.50	31.94
10.5hs	79.23	42.95	49.97	34.42	51.66	36.76	51.23	33.94
11.0hs	79.30	43.72	50.21	34.84	54.52	38.70	52.59	35.47
11.5hs	79.28	44.44	50.33	35.11	57.26	40.46	54.05	37.16
12.0hs	79.35	45.02	50.35	35.42	59.56	41.72	55.25	38.68
12.5hs					61.73	42.95	56.55	40.24
13.0hs					63.60	44.13	57.80	41.77
13.5hs					65.26	45.33	59.00	43.22
14.hs					66.88	46.62	60.33	44.77
14.5hs					4.64	47.96	61.66	46.26
15.0hs					69.40	49.03	62.70	47.51
15.5hs					69.83	49.63	63.22	48.46
16.0hs					69.59	49.76	63.57	49.31
Overall	**53.05**	**27.15**	**37.25**	**24.04**	**42.84**	**31.55**	**45.65**	**31.29**

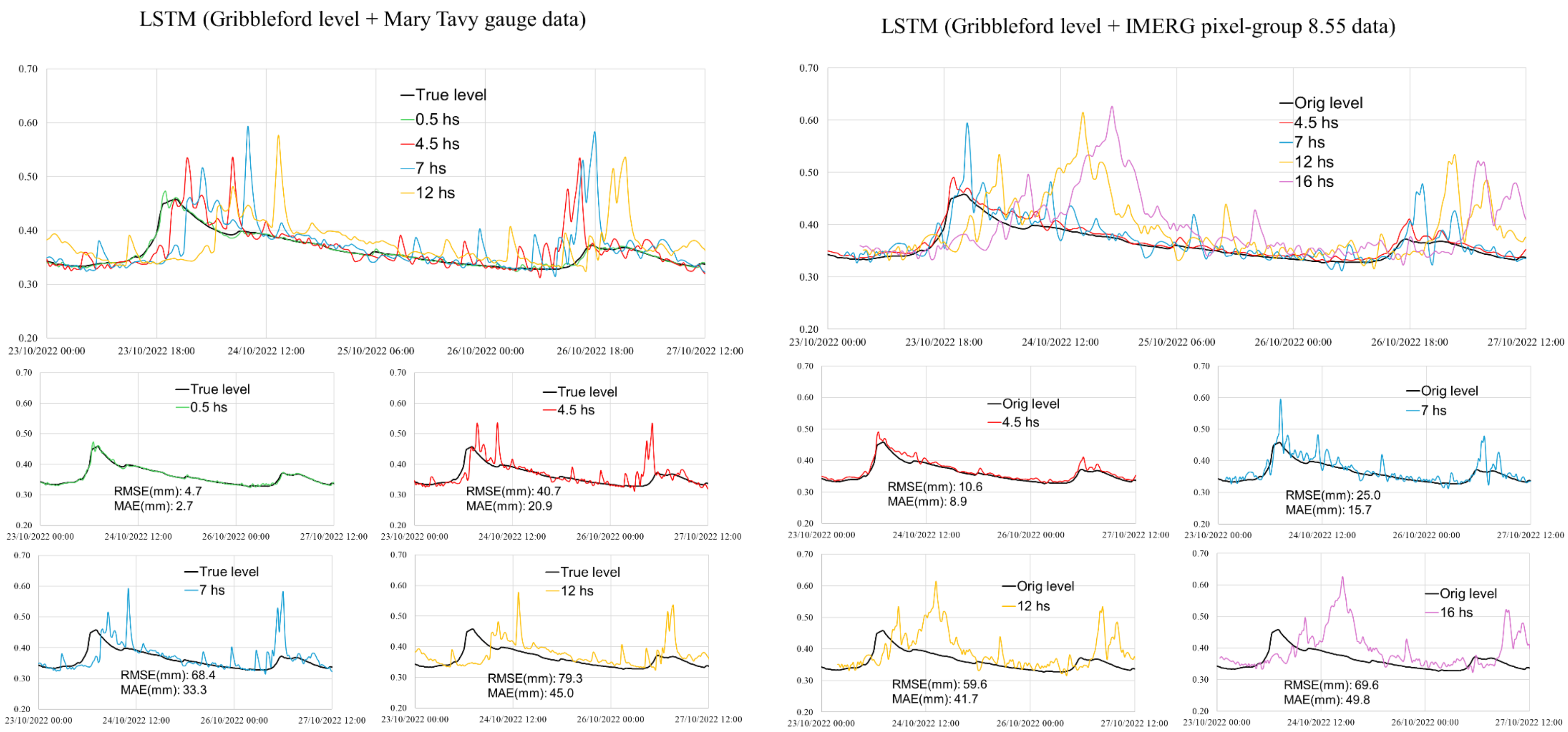


Figure E7 – Graphical comparison of true level and predictions at C2 for several lead times during the October 2022 event.

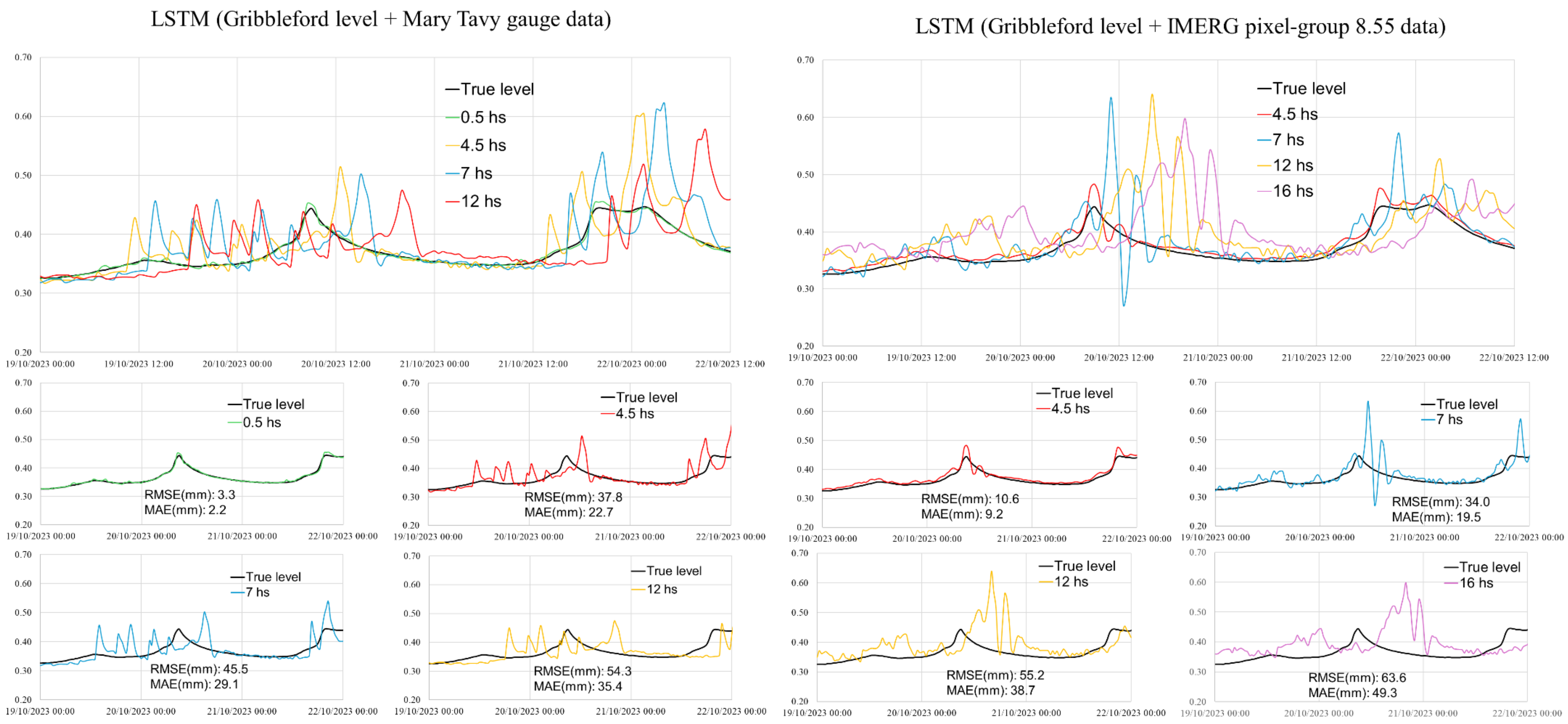


Figure E8 – Graphical comparison of true level and predictions at C2 for several lead times during the October 2023 event.