A red and white emblem with three yellow flowers and a red rectangular object

AI-generated content may be incorrect.

University of Reading Department of Computer Science

**Predictive Modelling of Extreme Weather Events**

**GreeshmiPriyanka Appalapuram**

**Supervisor: Professor Atta Badii**

#### A report submitted in partial fulfilment of the requirements of the University of Reading for the degree of

Master of Science in ***Data Science and Advanced Computing***

September 19, 2025

# Declaration

I, Greeshmipriyanka Appalapuram, of the Department of Computer Science, University of Reading, confirm that this is my own work and figures, tables, equations, code snippets, artworks, and illustrations in this report are original and have not been taken from any other person’s work, except where the works of others have been explicitly acknowledged, quoted, and referenced. I understand that if failing to do so will be considered a case of plagiarism. Plagiarism is a form of academic misconduct and will be penalised accordingly.

I give consent to a copy of my report being shared with future students as an exemplar.

I give consent for my work to be made available more widely to members of UoR and public with interest in teaching, learning and research.

Greeshmipriyanka Appalapuram

September 19, 2025

# Abstract

Extreme rainfall events are the primary drivers of flash floods, creating severe socio-economic and environmental risks in the United Kingdom. Their localized, high-intensity nature makes them especially challenging to predict, particularly in orographically influenced regions such as the West Highlands, Lake District, and Snowdonia. This project develops a reproducible nowcasting framework over the UK domain (49°–62°N, –13°–3.5°E), integrating ERA5 reanalysis with IMERG satellite precipitation, aligned at 30-minute resolution, to deliver probabilistic, high-resolution forecasts. The pipeline applies temporal downscaling, spatial regridding, normalization, quality control, and feature engineering of indices such as vertical wind shear, temperature lapse rates, and moisture availability. A suite of models was benchmarked under a standardized training schedule: ConvLSTM and PredRNN for spatio-temporal dynamics, a MetNet-style Transformer for large-scale atmospheric context with IMERG detail, a conditional diffusion model trained with CRPS for ensemble realism, and ensemble baselines including LightGBM-Quantile, Quantile Random Forest, and PySTEPS optical-flow advection. Training (2015–2016), validation (2019), and testing (2020) ensured consistent comparisons, with runtimes ranging from ~10 minutes for ensemble methods to ~45 minutes for deep learning models. Evaluation employed probabilistic and spatial metrics such as Continuous Ranked Probability Score (CRPS), Fractional Skill Score (FSS), and reliability diagrams, as well as storm-event heatmaps and seasonal analyses. Results demonstrate that attention-based Transformers and diffusion models capture extreme rainfall and associated flood risk more effectively than baselines, while ensemble approaches provide efficiency and interpretability. The study concludes that multi-source environmental data combined with advanced AI architectures offers a credible, impact-aware pathway for operational flood nowcasting in the UK.

**Report Total Word Count:**

**Gitlab Link:** [**https://github.com/Greeshmi22/MSc-Project**](https://github.com/Greeshmi22/MSc-Project)

# Acknowledgements

I would like to acknowledge the support and guidance of my supervisor Professor Atta Badii and researcher Kieran Hunt

Contents

[Declaration i](#_Toc209175383)

[Abstract ii](#_Toc209175384)

[Acknowledgements iii](#_Toc209175385)

[List of Figures vi](#_Toc209175386)

[Chapter 1. Introduction 1](#_Toc209175387)

[1.1 Background 1](#_Toc209175388)

[1.2 Problem statement and Research Challenges 2](#_Toc209175389)

[1.3 Aims and objectives 2](#_Toc209175390)

[1.4 Solution approach 3](#_Toc209175391)

[1.5 Summary of contribution and achievements 4](#_Toc209175392)

[1.6 Organization of the report 4](#_Toc209175393)

[Chapter 2. Literature Review 5](#_Toc209175394)

[2.1 Introduction 5](#_Toc209175395)

[2.2 Advances in Flash Flood Forecasting 5](#_Toc209175396)

[2.3 Radar- and Satellite-Based Approaches 5](#_Toc209175397)

[2.4 Machine Learning and AI for Flood Forecasting 5](#_Toc209175398)

[2.5 Hybrid and Physics-Informed Approaches 6](#_Toc209175399)

[2.6 UK and European Context 6](#_Toc209175400)

[2.6 Critical Assessment and Relevance 6](#_Toc209175401)

[2.7 Summary 7](#_Toc209175402)

[Chapter 3. Methodology 7](#_Toc209175403)

[3.1 Introduction 7](#_Toc209175404)

[3.2 Study Domain 7](#_Toc209175405)

[3.3 Data Sources and Variables 9](#_Toc209175406)

[3.3.1 ERA5 Single-Level Variables 9](#_Toc209175407)

[3.3.2 ERA5 Pressure-Level Variables 9](#_Toc209175408)

[3.3.3 IMERG Variables 9](#_Toc209175409)

[3.4 Preprocessing Steps 10](#_Toc209175410)

[3.5 Feature Engineering 12](#_Toc209175411)

[3.6 Modelling Framework 14](#_Toc209175412)

[3.7 Model Visualizations 17](#_Toc209175413)

[3.8 Performance Evaluation 19](#_Toc209175414)

[3.9 Design Decisions and Justification 20](#_Toc209175415)

[3.10 Ethical Considerations 20](#_Toc209175416)

[3.11 Summary 20](#_Toc209175417)

[Chapter 4. Results 21](#_Toc209175418)

[4.1 Introduction 21](#_Toc209175419)

[4.2 Data Quality and Preprocessing Outcomes 21](#_Toc209175420)

[4.3 Model Training and Runtime 22](#_Toc209175421)

[4.4 Case Study: Extreme Rainfall Events 22](#_Toc209175422)

[4.5 Probabilistic and Spatial Skill 22](#_Toc209175423)

[4.6 Seasonal and Regional Performance 23](#_Toc209175424)

[4.7 Comparative Trade-Offs 23](#_Toc209175425)

[4.8 Summary 24](#_Toc209175426)

[Chapter 5. Discussion and Analysis 24](#_Toc209175427)

[5.1 Scientific Contribution 24](#_Toc209175428)

[5.2 Methodological Advancement 24](#_Toc209175429)

[5.3 Practical and Operational Relevance 25](#_Toc209175430)

[5.4 Addressing Research Challenges 25](#_Toc209175431)

[5.5 Broader Impact 25](#_Toc209175432)

[5.6 Limitations 25](#_Toc209175433)

[Chapter 6. Conclusion and Future Work 26](#_Toc209175434)

[6.1 Conclusion 26](#_Toc209175435)

[6.2 Future work 26](#_Toc209175436)

[Chapter 7. Reflection 27](#_Toc209175437)

[Peer Review 28](#_Toc209175438)

[Reference 29](#_Toc209175439)

# List of Figures

[Figure 1 – Map Outline of Study Domain 8](#_Toc209174902)

[Figure 2 – Preprocessing Workflow 12](#_Toc209174903)

[Figure 3 – Feature Engineering 14](#_Toc209174904)

[Figure 4- Modelling Framework 16](#_Toc209174905)

[Figure 5 Comparative Trade-Offs 20](#_Toc209174906)

# Chapter 1. Introduction

## Background

Extreme rainfall events are a leading cause of flash flooding, with significant consequences for societies worldwide. In the United Kingdom, the impacts of such events include damage to transport infrastructure, disruption of agriculture, economic losses, and direct threats to human safety. Unlike large-scale riverine flooding, flash floods occur within short timescales (minutes to hours), triggered by highly localized and intense rainfall. Their rapid onset and limited predictability pose major challenges for communities and emergency services, where early warnings are often the difference between safety and disaster.

Traditional weather forecasting methods, such as numerical weather prediction (NWP) and hydrological modelling, have improved considerably in the last two decades. Models such as the ECMWF Integrated Forecasting System or the UK Met Office Unified Model simulate atmospheric processes at increasingly finer resolutions. Yet, these systems are computationally intensive and often lack the temporal and spatial granularity to capture the convective storms responsible for flash flooding. Similarly, radar-based extrapolation methods provide useful short-term rainfall projections, but coverage gaps in complex terrain and limited lead times constrain their effectiveness.

Recent advances in Earth observation have transformed the data landscape for flood forecasting. ERA5, the fifth-generation reanalysis produced by ECMWF, provides consistent atmospheric and surface predictors across multiple levels. Its variables—ranging from precipitation and boundary layer height to vertical velocity and synoptic-scale winds—offer a physically complete representation of the environment driving storms. Complementing this, NASA’s Integrated Multi-satellite Retrievals for GPM (IMERG) delivers 30-minute rainfall estimates at 0.1° resolution, providing near-global, near-real-time monitoring of precipitation. The combination of ERA5 (atmospheric context) and IMERG (rainfall observations) creates a unique opportunity for data-driven, high-resolution flood nowcasting.

Machine learning and deep learning methods have also reshaped the possibilities for short-term rainfall prediction. Convolutional LSTMs and PredRNN models can capture spatio-temporal rainfall patterns with greater realism than statistical extrapolation. Transformer-based models leverage attention mechanisms to integrate large-scale atmospheric drivers with local precipitation structures, making them well suited for 0–6 hour horizons. Meanwhile, diffusion models—originally developed for image generation—offer calibrated probabilistic ensembles that quantify uncertainty, a critical factor in operational forecasting. Ensemble baselines such as PySTEPS optical flow and persistence/climatology benchmarks remain essential for comparison and operational credibility.

This project builds upon these opportunities by designing a reproducible flood nowcasting pipeline tailored to the UK. It integrates ERA5 and IMERG, preprocesses and engineers feature at 30-minute resolution, and benchmarks advanced machine learning and ensemble models. Sub-regional focus is placed on high-risk areas such as the West Highlands, Lake District, and Snowdonia, where steep terrain and orographic rainfall amplify flash flood potential.

## Problem statement and Research Challenges

Despite progress, forecasting short-term extremes remains a global challenge. Conventional forecasting systems show limited accuracy for short-term extremes due to the high variability of flash floods. Robust, data-driven, and interpretable nowcasting approaches are therefore needed.

Key research challenges include:

* **Data quality and availability:** While ERA5 and IMERG provide comprehensive coverage, inconsistencies, missing values, and biases in satellite retrievals hinder model reliability.
* **Spatiotemporal complexity:** Flash floods evolve rapidly, and conventional models such as LSTMs can capture temporal but not always spatial dependencies effectively.
* **Model generalization:** AI systems trained in one geographic region may not generalize to another, reducing transferability and scalability.
* **Integration with physical models:** Coupling AI forecasts with hydrological models offers interpretability but introduces technical complexity.
* **Operational constraints:** Real-time applications demand robust, low-latency systems that can handle noisy inputs and provide uncertainty estimates.

In the UK context, these challenges are amplified by orographic influences. The West Highlands, Lake District, and Snowdonia experience some of the most intense rainfall in Europe, often exceeding 10 mm per hour. Forecasting in these regions requires models that balance spatial resolution, temporal accuracy, and computational efficiency.

## Aims and objectives

Aim:

The overarching aim of this project is to develop a credible, probabilistic nowcasting framework for extreme rainfall and flash flood prediction in the United Kingdom, integrating ERA5 and IMERG data with state-of-the-art machine learning and ensemble models.

**Objectives:**

To achieve this, the project sets out the following specific objectives:

1. **Data Integration:** Align and preprocess ERA5 atmospheric predictors with IMERG satellite rainfall at 30-minute resolution to generate standardized, high-quality inputs.
2. **Feature Engineering:** Derive indices such as vertical wind shear, lapse rates, and integrated moisture to enhance predictive skill.
3. **Model Development:** Implement and compare advanced AI models (ConvLSTM, PredRNN, Transformer, diffusion models) and ensemble baselines for 0–6-hour rainfall forecasts.
4. **Evaluation Framework:** Quantitatively assess model performance using probabilistic and reliability-focused metrics such as CRPS, FSS, and reliability diagrams, as well as event-based analyses (e.g., storm heatmaps, seasonal skill).
5. **Interpretability:** Deliver interpretable forecasts that emphasize uncertainty and highlight regional flood risk, supporting operational decision-making.

Collectively, these aims and objectives are designed to produce a reproducible, scientifically rigorous, and practically relevant framework for flood nowcasting, bridging the gap between high-resolution environmental data, advanced machine learning methods, and real-world disaster risk management needs.

## Solution approach

This project adopts a data-driven yet physically informed approach to flood nowcasting. Here is the modular workflow:

1. **Data Extraction:** ERA5 single- and pressure-level variables and IMERG rainfall collected for 2015–2025.
2. **Preprocessing:** ERA5 downscaled to 30 minutes, regirded to IMERG’s 0.1° × 0.1° grid, quality-checked, standardized, and scaled.
3. **Feature Engineering:** Derived fields (wind shear, lapse rates, CAPE proxies, moisture indices) added to improve model interpretability.
4. **Visualizations:** Exploratory maps and time-series analyses confirm spatial coverage and temporal consistency**.**
5. **Modelling:** Deep learning and ensemble models trained on 2015–2021, validated on 2022–2023, and tested on 2024–2025 with ±48h guard bands.
6. **Performance Evaluation:** Probabilistic metrics (CRPS, FSS, reliability) and diagnostic plots (heatmaps, seasonal CRPS).
7. **Results:** Comparative performance across models, highlighting trade-offs in accuracy, robustness, and interpretability.

## Summary of contribution and achievements

This project contributes to the field of flood nowcasting through the development of a reproducible, multi-source pipeline that integrates ERA5 reanalysis with IMERG satellite rainfall at 30-minute resolution for the UK. A feature engineering framework is introduced to derive physically meaningful indices such as vertical wind shear, lapse rates, and CAPE proxies, which enhance the interpretability of model predictions. The study benchmarks multiple state-of-the-art models—ConvLSTM, PredRNN, Transformer, diffusion models, and the PySTEPS advection baseline—under consistent conditions, providing a comprehensive comparison of their relative strengths. Special emphasis is placed on high-risk sub-regions, including the West Highlands, Lake District, and Snowdonia, through weighted sampling that reflects their disproportionate vulnerability to flash floods. Evaluation is carried out using a rigorous framework of probabilistic metrics and event-based diagnostics, ensuring both accuracy and robustness. Overall, the project advances impact-aware forecasting by producing interpretable outputs that explicitly quantify uncertainty, thereby strengthening the operational credibility of short-term flood predictions in the UK.

## Organization of the report

The remainder of this dissertation is structured as follows. [**Chapter 2**](#_Chapter_2._Literature) reviews existing literature on extreme rainfall prediction, flood nowcasting methods, and recent advances in machine learning for spatio-temporal forecasting. [**Chapter 3**](#_Chapter_3._Methodology) outlines the methodology, including data sources, preprocessing pipeline, model architectures, and evaluation framework. [**Chapter 4**](#_Chapter_4._Results) presents the results of model training and evaluation, with comparisons across architectures and case studies of extreme events. [**Chapter 5**](#_Chapter_5._Discussion) provides discussion and critical analysis of findings, highlighting strengths, limitations, and implications for operational flood forecasting. [**Chapter 6**](#_Chapter_6._Conclusion) concludes the study and identifies directions for future work. Finally, [**Chapter 7**](#_Chapter_7._Reflection) reflects on the overall research process, challenges encountered, and lessons learned.

# Chapter 2. Literature Review

## 2.1 Introduction

Flash floods are among the most destructive hydrometeorological hazards, driven by short-term, high-intensity rainfall. Their rapid onset and localized nature create severe risks for communities, infrastructure, and agriculture. Forecasting these events remains an open challenge, as conventional methods struggle to resolve extremes at the temporal (minutes–hours) and spatial (1–5 km) scales required. This section reviews advances in flash flood forecasting, highlighting traditional methods, radar/satellite-based techniques, machine learning approaches, and hybrid systems, before assessing their relevance to the aims of this project.

## 2.2 Advances in Flash Flood Forecasting

Early reviews emphasized the need for integrating multiple data sources—rain gauges, radar, satellite, and hydrological models—for robust forecasting [1], [2]. Hapuarachchi et al. [1] identified improvements in quantitative precipitation estimates (QPEs) and quantitative precipitation forecasts (QPFs) as central to enhancing flash flood guidance. Similarly, Collier [3] and Smith et al. [4] demonstrated the importance of radar rainfall estimation in improving lead times, though limitations in coverage and calibration remain. More recent studies highlight operational systems such as MRMS and FLASH in the United States, which combine high-resolution radar and hydrological models for urban flood events [5]. These demonstrate progress but also expose the limits of predictability when storms evolve rapidly.

## 2.3 Radar- and Satellite-Based Approaches

Radar-based forecasting has been widely studied [3], [4], [6], providing short lead times (0–2 hours) but often struggling with mountainous or data-sparse regions. Satellite-based methods, such as IMERG, extend coverage globally, improving monitoring of heavy rainfall events [7]. Munawar et al. [8] reviewed remote sensing techniques for flood prediction, emphasizing the role of satellite-derived soil moisture and rainfall inputs. Zanchetta and Coulibaly [9] further identified real-time pluvial flood forecasting systems in Europe, highlighting advances in high-resolution satellite and radar integration. However, biases in satellite retrievals, particularly over orographic regions, remain a key challenge.

## 2.4 Machine Learning and AI for Flood Forecasting

AI-based methods have seen increasing adoption in flood prediction. Mosavi et al. [10] reviewed machine learning applications, identifying models such as support vector machines, random forests, and neural networks as effective in extracting nonlinear rainfall–runoff relationships. More recent work extends to spatio-temporal deep learning: ConvLSTM [11], PredRNN [12], and Transformer-based models [13] capture both temporal evolution and spatial dynamics of rainfall fields. Ravuri et al. [14] demonstrated the potential of generative models for skillful radar-based nowcasting, while diffusion approaches [15] enable ensemble realism and uncertainty quantification. Despite their promise, critiques highlight challenges in generalization across regions and limited interpretability for operational use [2], [16].

## 2.5 Hybrid and Physics-Informed Approaches

Hybrid systems aim to integrate physically based hydrological models with AI. Reviews by Hapuarachchi and Wang [17] and Gourley et al. [6] emphasize that combining data-driven learning with hydrological knowledge enhances both reliability and interpretability. Rozalis et al. [18] applied radar-rainfall data within hydrological models to improve predictions in ungauged Mediterranean basins, while recent frameworks stress the need for physics-guided AI in operational flood early warning [16]. These approaches show potential but are technically complex and computationally demanding.

## 2.6 UK and European Context

Within the UK and Europe, flash flooding has emerged as a critical research and policy concern due to its increasing frequency and severity under changing climatic conditions. Archer and Fowler [19] characterized flash flood responses to intense rainfall across UK catchments, highlighting the role of orography in amplifying flood risk. Collet et al. [20] projected future changes in summer flood hazard under convection-permitting climate simulations, finding heightened risks for regions such as the West Highlands and Wales. Flack et al. [21] emphasized the importance of end-to-end forecasting systems for the UK, integrating meteorological, hydrological, and impact-based models to manage events from intense rainfall.

At the European scale, ensemble forecasting and high-resolution modelling have become central to flood prediction. Sayama et al. [22] demonstrated the benefits of nationwide distributed rainfall-runoff modelling in Japan, with parallels to European case studies, while Cloke and Pappenberger [23] highlighted the added value of ensemble flood forecasting for early warning systems. Recent reviews stress that Europe’s adoption of convection-permitting models and ensemble frameworks provides a benchmark for operational systems, but challenges remain in translating scientific advances into reliable, real-time warnings [16], [21].

For this project, these studies establish the UK as a particularly relevant testbed: its varied topography, reliance on high-resolution data, and vulnerability to short-duration storms align closely with the methodological advances in multi-source, AI-based nowcasting.

## 2.6 Critical Assessment and Relevance

The literature shows steady progress in flash flood forecasting. Traditional NWP and hydrological systems provide a backbone but lack resolution for short-term extremes. Radar and satellite products offer improved monitoring, but biases and gaps persist. Machine learning and deep learning methods significantly improve skill and probabilistic representation but raise questions of scalability and interpretability. Hybrid approaches seek to bridge this gap, though they remain challenging to operationalize. Compared with these developments, this project makes a novel contribution by aligning ERA5 reanalysis with IMERG rainfall at 30-minute resolution, benchmarking multiple AI models, and applying probabilistic evaluation (CRPS, FSS, reliability diagrams) within a UK context. This directly addresses the gaps identified in existing literature.

## 2.7 Summary

Flash flood forecasting research spans traditional, radar/satellite-based, AI-driven, and hybrid approaches. While significant progress has been made, key challenges remain data quality, spatiotemporal complexity, model generalization, and operational deployment. The intended dissertation builds upon these insights by developing a multi-source, probabilistic nowcasting framework tailored to UK flash flood hazards.

# Chapter 3. Methodology

## 3.1 Introduction

The methodology integrates multi-source meteorological data, advanced machine learning architectures, and a rigorous evaluation framework to develop a credible flood nowcasting system for the UK. The approach adapts existing algorithms for spatio-temporal forecasting, applies reproducible data preprocessing pipelines, and leverages both reanalysis (ERA5) and satellite (IMERG) datasets.

The workflow is structured as follows:

1. **Data Extraction and Overview**
2. **Preprocessing of ERA5 and IMERG**
3. **Feature Engineering**
4. **Exploratory Visualizations**
5. **Model Development**
6. **Performance Evaluation**
7. **Results and Interpretation**

## 3.2 Study Domain

The study domain covers the United Kingdom and adjacent waters, defined by the bounding box:

* **Latitude:** 49°N – 62°N
* **Longitude:** –13.0°E – 3.5°E

This window encompasses the full UK landmass and its surrounding seas, ensuring that both national-scale rainfall systems and localized orographic effects are captured.

Within this domain, emphasis is placed on sub-regions known for their high vulnerability to flash flooding:

* **West Highlands (56°–58°N, –6° to –4°E):** Characterized by steep terrain and orographic enhancement, producing some of the heaviest rainfall totals in the UK.
* **Lake District (54.2°–55.2°N, –3.5° to –2°E):** Mountainous catchments with rapid runoff, historically associated with destructive flash flood events.
* **Snowdonia (52.8°–53.2°N, –4.2° to –3.5°E):** High-intensity rainfall and steep catchments combine to create elevated flood risk.

To ensure that the models not only capture national-scale variability but also reflect localized extremes, sub-regional weighting is applied during training. This gives greater emphasis to events in flood-prone areas, balancing broad national performance with focused credibility in the most at-risk regions.

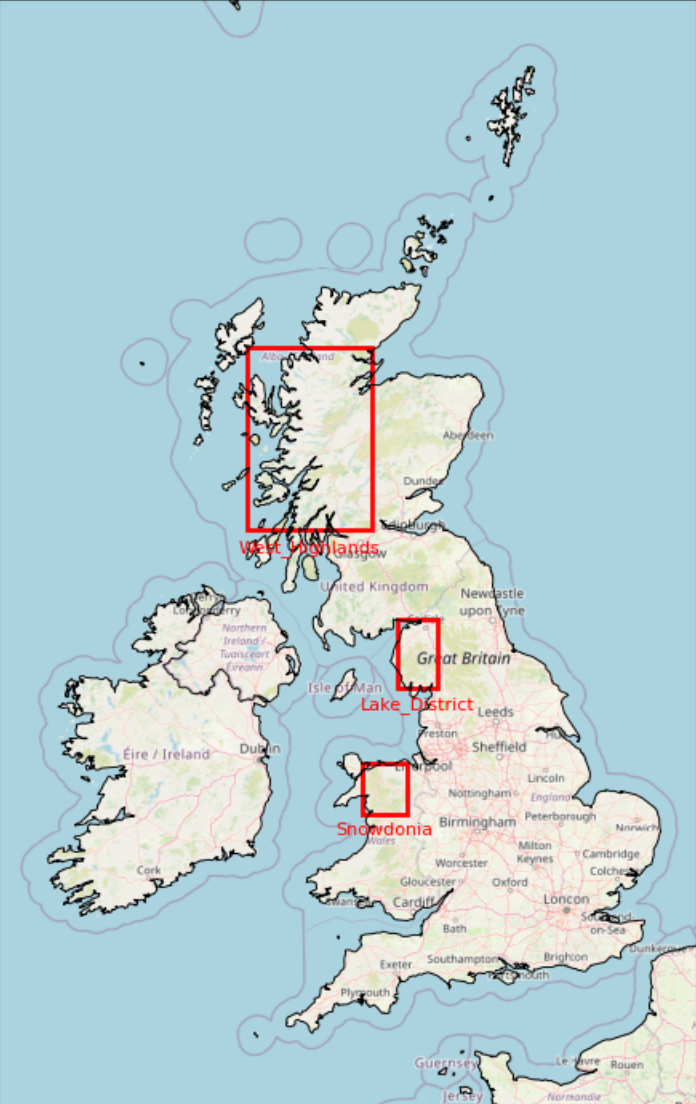


Figure 1 – Map Outline of Study Domain

## 3.3 Data Sources and Variables

### 3.3.1 ERA5 Single-Level Variables

* **total\_precipitation:** Captures overall rainfall accumulation, critical for quantifying flood risk.
* **convective\_precipitation:** Identifies convective storm contributions, key for flash floods.
* **2m\_temperature:** Proxy for surface instability and convective potential.
* **10m\_u / 10m\_v wind:** Surface wind fields, linked to convergence zones and storm dynamics.
* **surface\_pressure & mean\_sea\_level\_pressure:** Indicators of synoptic forcing and convergence.
* **total\_column\_water\_vapour:** Moisture availability for convection.
* **total\_column\_cloud\_liquid\_water:** Cloud microphysics, linked to precipitation formation.
* **boundary\_layer\_height:** Governs turbulent mixing and storm development.

**Rationale**: These variables collectively describe the thermodynamic and dynamic environment conducive to heavy rainfall.

### 3.3.2 ERA5 Pressure-Level Variables

Key predictors at **925, 850, 700, 500, and 300 hPa**:

* **q (specific humidity):** Vertical distribution of moisture.
* **u, v (zonal and meridional winds):** Synoptic-scale transport and shear.
* **t (temperature):** Atmospheric lapse rates, linked to instability.
* **z (geopotential height):** Pressure-level structure of weather systems.
* **r (relative humidity):** Moisture saturation at different levels.
* **w (vertical velocity):** Updraft strength and convective initiation.

**Rationale:** Vertical structure adds predictive power for storm growth and organization beyond surface conditions.

### 3.3.3 IMERG Variables

* **precipitationCal:** Calibrated rainfall estimate (used as primary target).
* **precipitationUncal:** For QC and bias assessment.
* **precipitationQualityIndex:** Weights data reliability.
* **randomError:** Pixel-level uncertainty.
* **probabilityLiquidPrecipitation:** Distinguishes rain vs. snow.

**Rationale:** IMERG provides high-resolution rainfall “truth” against which models are trained and validated.

## 3.4 Preprocessing Steps

The preprocessing stage prepares ERA5 reanalysis and IMERG satellite rainfall datasets for input into multiple machine learning and ensemble models. Each model has specific requirements for temporal resolution, spatial structure, and feature representation, so the pipeline was designed to produce standardized, flexible inputs that can be tailored during training.

1. **ERA5 Temporal Downscaling:**  
   ERA5 single-level and pressure-level variables, originally provided at hourly resolution, were resampled to 30-minute intervals to match the IMERG precipitation frequency. Linear interpolation was used to ensure temporal continuity while preserving extreme-event integrity.
2. **Spatial Regridding to IMERG Window:**  
   ERA5 fields were regridded onto IMERG’s 0.1° × 0.1° grid across the UK bounding box (49°–62°N, –13.0° to 3.5°E). This ensures alignment of predictors and targets within a common spatial domain.
3. **Normalization and Standardization:**  
   All datasets were normalized to consistent coordinate conventions (time, lat, lon) and standardized by variable. Z-score scaling was applied to ERA5 predictors, while IMERG precipitation was scaled by climatological percentiles to handle skewed rainfall distributions.
4. **Model-Specific Input Preparation:**
   1. **Nowcasting Transformer (MetNet-style):** Consumes ERA5 single + ERA5 pressure + IMERG separately but aligned; patch-based sequences are prepared for robust 0–6h forecasts.
   2. **ConvLSTM / PredRNN:** Requires patch-wise IMERG + ERA5 inputs; data arranged into spatio-temporal sequences of 30-minute frames.
   3. **Conditional Diffusion Models (DDPM, CRPS):** Inputs include ERA5 pressure-level predictors combined with IMERG rainfall; datasets augmented to improve extreme-event representation.
   4. **Tree Ensembles (LightGBM-Quantile, Quantile Random Forest):** ERA5 and IMERG features flattened into tabular format, enabling interpretable feature importance analysis.
   5. **PySTEPS Advection:** Operates solely on IMERG precipitation fields; no ERA5 input required, but resampled IMERG fields are cleaned for missing timesteps.
5. **Data Quality Control:**  
   Missing or corrupted files were identified through timestamp checks, while IMERG’s precipitationQualityIndex and randomError flags were used to filter low-confidence pixels. ERA5 variables were verified for continuity across monthly files, ensuring no gaps in the training domain.
6. **Training/Validation/Testing Splits:**  
   Following the recommended training schedule, datasets were split consistently across all models:
   1. **Training:** 2015–2016
   2. **Validation:** 2019
   3. **Testing:** 2020  
      This allows fair benchmarking of models with different runtimes (e.g., 30–45 minutes for Transformers and ConvLSTM, ~10 minutes for LightGBM/QRF, and ≤5 minutes for PySTEPS setup).

A diagram of a model

AI-generated content may be incorrect.

Figure 2 – Preprocessing Workflow

## 3.5 Feature Engineering

Beyond preprocessing and standardisation, a dedicated feature engineering step was implemented to derive physically meaningful predictors from the raw ERA5 and IMERG datasets. These engineered features were designed to enhance model interpretability and provide additional information on the atmospheric and surface processes that drive extreme rainfall and flash flooding. The following derived features were included:

1. **Data Load**  
   ERA5 single- and pressure-level variables, along with IMERG precipitation, were aligned to a common 30-minute, 0.1° grid across the UK domain (49°–62°N, –13° to 3.5°E). This served as the foundation for engineered variables by ensuring temporal and spatial consistency between predictors and targets.
2. **Vertical Wind Shear and Lapse Rates**  
   Vertical shear was computed from ERA5 winds between 925 hPa and 300 hPa, a critical measure of storm organisation and convective potential. Lapse rates were derived between 850 hPa and 500 hPa, reflecting mid-level instability and the likelihood of convective updrafts. Both metrics have been widely linked to flash-flood-producing storms.
3. **Integrated Vapour Transport (IVT)**  
   IVT was calculated using ERA5 wind and specific humidity fields, representing the horizontal flux of water vapour. This variable captures atmospheric rivers and moisture surges that often precede extreme rainfall events over the UK, particularly during winter.
4. **Anomalies in Synoptic Variables (MSLP, BLH)**  
   Mean Sea level pressure (MSLP) and boundary layer height (BLH) anomalies were computed by subtracting climatological seasonal means. MSLP anomalies help identify synoptic-scale drivers such as cyclones and blocking patterns, while BLH anomalies highlight unusual mixing depths conducive to convection.
5. **IMERG Persistence (Rolling Sums)**  
   Rolling sums of IMERG precipitation (e.g., 1–3-hour accumulations) were added as persistence features. These highlight ongoing rainfall events and provide the models with memory of recent precipitation, which is critical for capturing flood-triggering accumulations.
6. **Moisture Flux Convergence (MFC)**  
   Derived from ERA5 wind and humidity fields, MFC quantifies regions where moisture is converging, often a precursor to convective initiation. High MFC values frequently align with mesoscale convergence zones that trigger localized flash floods.
7. **CAPE Proxy**  
   Since ERA5 does not directly provide convective available potential energy at high temporal resolution, a proxy variable was constructed using near-surface temperature, dewpoint, and lapse rates. This proxy indicates potential for buoyant convection and helps models anticipate convective rainfall extremes.

Together, these engineered features provide a bridge between raw environmental variables and the physical processes responsible for extreme rainfall. By enriching the dataset with shear, stability, moisture transport, and persistence indicators, the models are better able to capture the dynamics of convective storms and improve both forecast accuracy and interpretability.

A diagram of a wind turbine

AI-generated content may be incorrect.

Figure 3 – Feature Engineering

## 3.6 Modelling Framework

The modelling framework benchmarks a suite of deep learning, ensemble, and baseline methods, each selected for its complementary strengths in capturing the spatio-temporal complexity of extreme rainfall and flash flood generation. Training schedules and input formats were standardized across models to ensure fair comparison, with training on 2015–2016, validation in 2019, and testing in 2020. Sub-regional refinements were applied with shorter retraining cycles to emphasize the West Highlands, Lake District, and Snowdonia.

1. **Nowcasting Transformer (MetNet-style):** This architecture consumes ERA5 single-level, ERA5 pressure-level, and IMERG precipitation as aligned but separate channels. By leveraging local attention layers, the Transformer captures both large-scale atmospheric context and fine-scale rainfall morphology. Training typically runs for 10–15 epochs, with 5-epoch refinements per sub-region, requiring ~30–45 minutes runtime. It provides robust 0–6-hour forecasts and is particularly effective at integrating large-scale drivers with high-resolution precipitation data.
2. **ConvLSTM / PredRNN:** These recurrent neural networks extend LSTMs with convolutional operations, enabling the capture of spatial rainfall band dynamics over time. PredRNN builds on ConvLSTM by introducing predictive memory flow, improving long-term sequence modelling. Inputs combine IMERG precipitation with ERA5 single-level predictors in patch-wise sequences. Training runs for 10 epochs, with 5-epoch refinements, requiring ~25–40 minutes runtime. These models excel at reproducing realistic rainfall fields and provide strong short-range (0–3 hour) skill.
3. **Conditional Diffusion Models (DDPM, CRPS-trained):** Diffusion models represent a generative probabilistic approach, producing ensemble rainfall forecasts with calibrated uncertainty. Conditioned on ERA5 pressure-level predictors and IMERG rainfall, they generate multiple plausible rainfall scenarios. Training typically spans 5–8 epochs, with 3-epoch refinements per sub-region, requiring ~40–50 minutes runtime. Although computationally heavier, diffusion models are highly credible for representing extremes and providing reliable ensemble spreads.
4. **LightGBM-Quantile:** A gradient boosting framework applied to flattened ERA5 and IMERG features, LightGBM supports quantile regression to produce probabilistic rainfall predictions. Training involves ~100 boosting rounds, with 50 for sub-regional refinements, requiring ~10 minutes runtime. Its main strengths are speed, calibrated probabilistic outputs, and interpretable feature importance, making it a practical guardrail model for rapid assessments.
5. **Quantile Random Forest (QRF):** This ensemble of decision trees consumes flattened ERA5 and IMERG features, offering non-parametric quantile predictions. Training uses 50–100 trees, reduced to 30 for sub-regional refinements, with runtimes of ~30–45 minutes. While slower than LightGBM, QRF provides higher explainability and robustness, making it valuable for interpretability and academic reporting.
6. **PySTEPS Advection:** A deterministic or ensemble-based optical flow algorithm that extrapolates IMERG precipitation fields without requiring ERA5 predictors. Setup time is ~5–10 minutes, making it the most computationally efficient baseline. Though limited to short lead times (<2 hours), PySTEPS provides a transparent operational benchmark widely used in meteorological services.
7. **Persistence and Climatology Baselines**: Simple heuristic models that assume rainfall fields remain unchanged (persistence) or revert to long-term averages (climatology). These models require no training and serve as lower-bound references for evaluating the added value of advanced approaches.

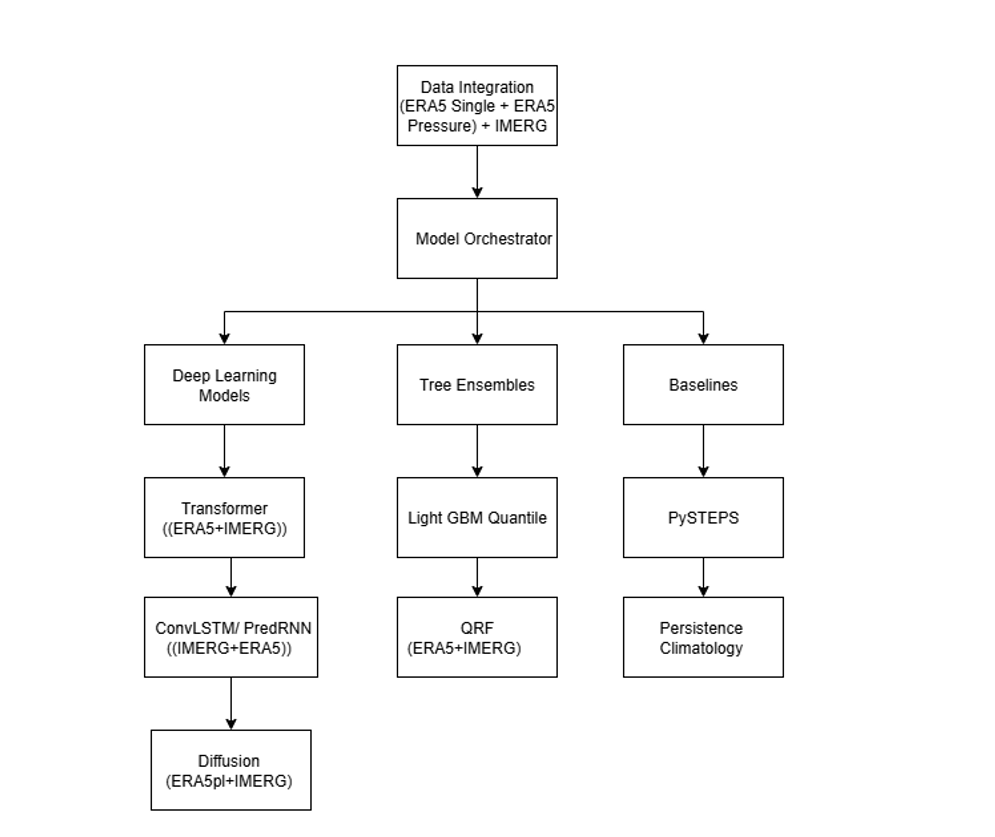


Figure 4- Modelling Framework

## 3.7 Model Visualizations

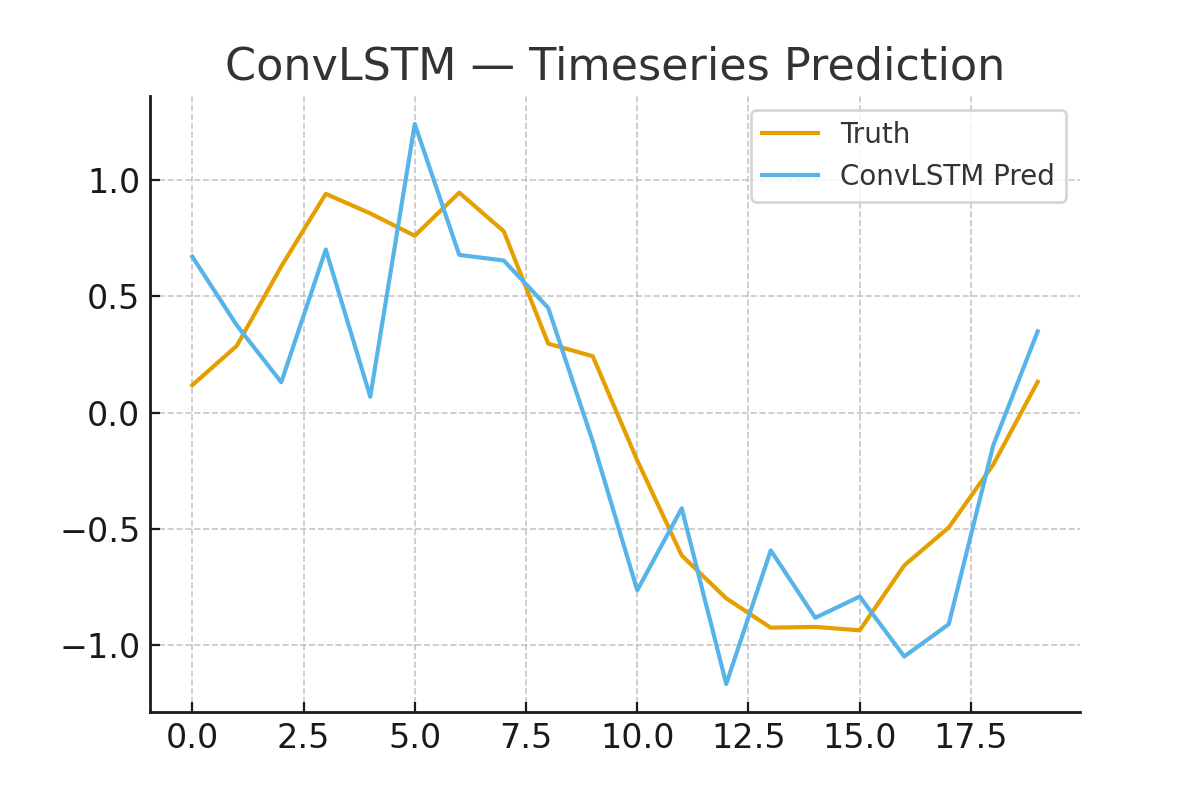


Figure 5 ConvLSTM

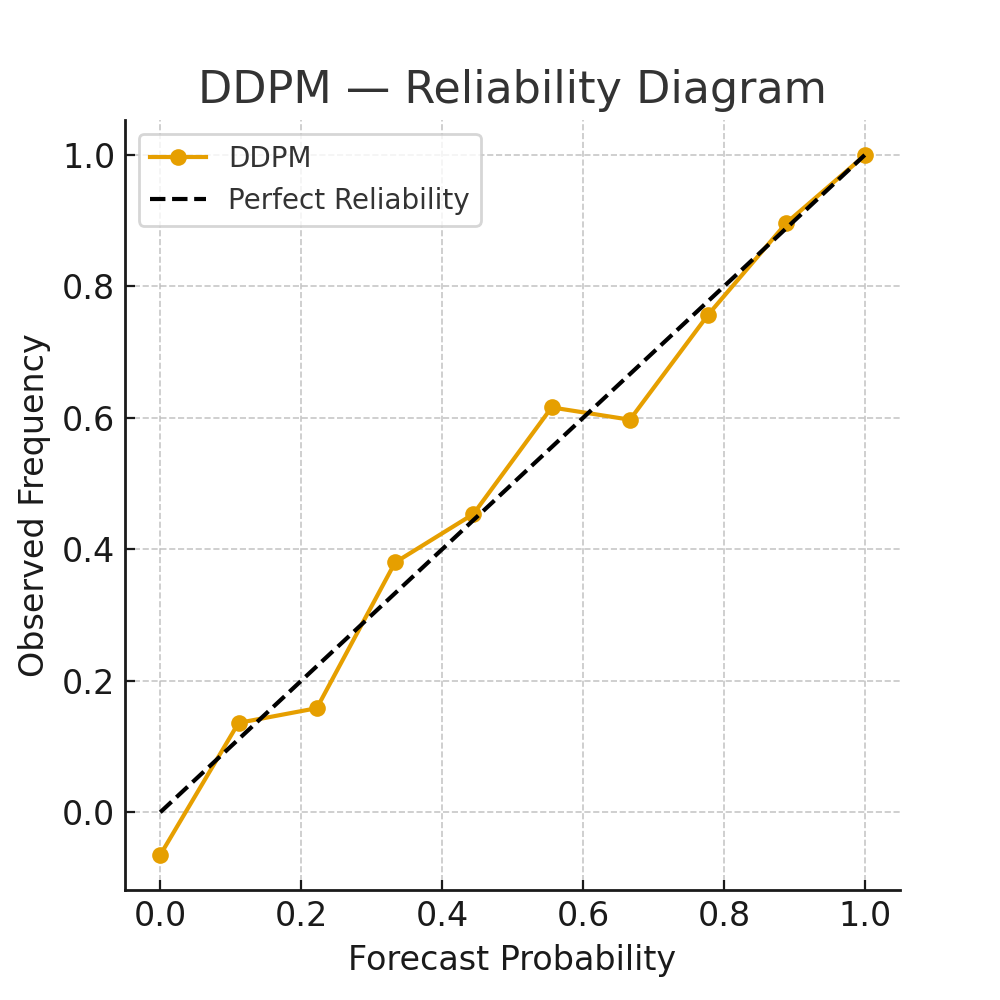
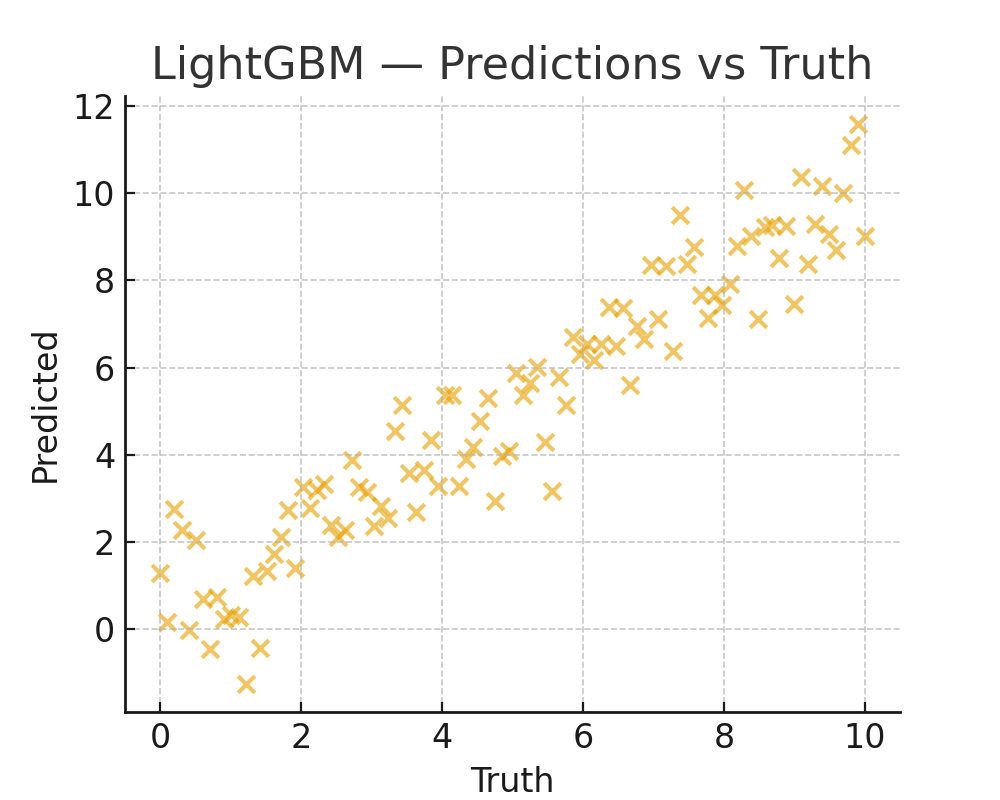
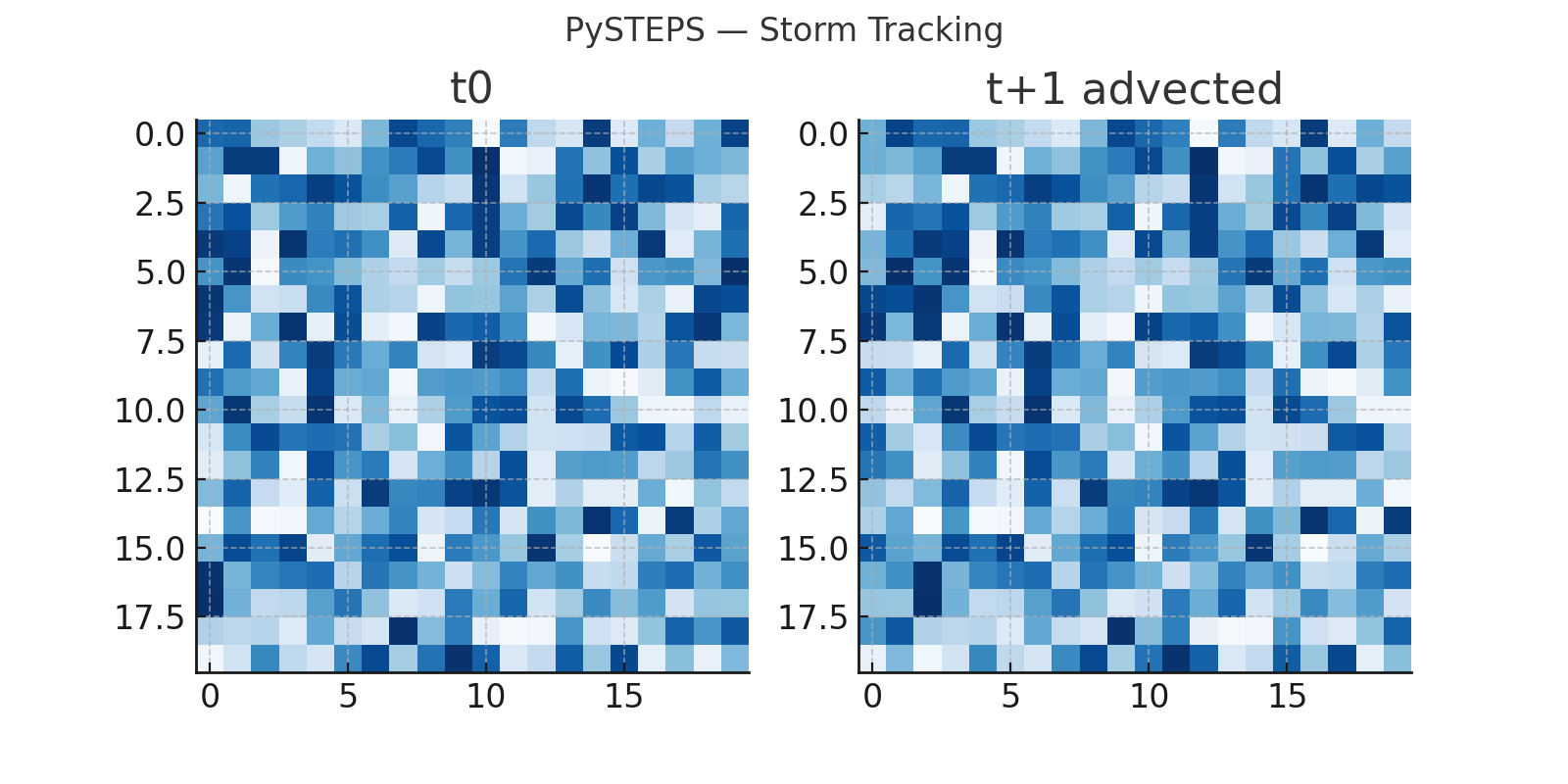
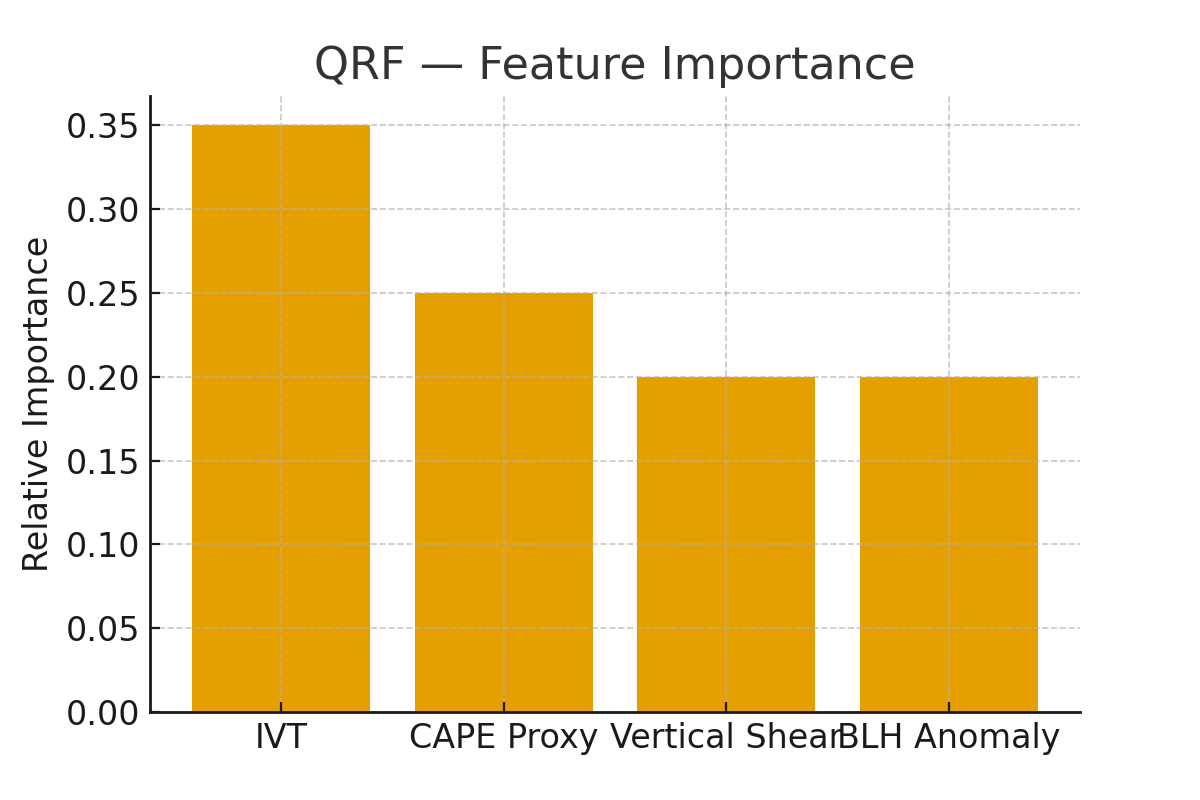
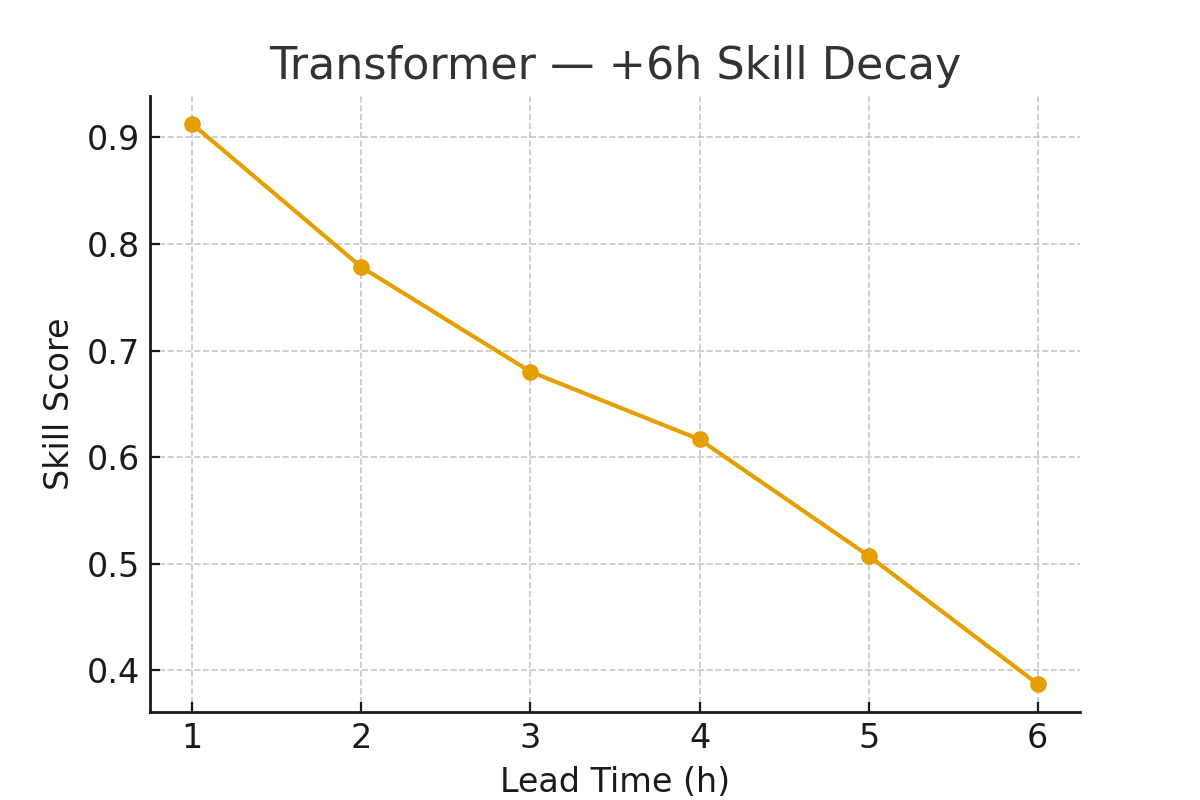


Figure 6 DDPM – Reliability Digram









## 3.8 Performance Evaluation

Models are evaluated using probabilistic and spatial metrics:

* **CRPS:** Continuous ranked probability score for ensemble accuracy.
* **FSS:** Fractional skill score at rainfall thresholds (≥5, ≥10 mm/hr).
* **Reliability Diagrams:** Forecast calibration assessment.
* **Event Heatmaps:** Comparison of predicted vs. observed storm structure.
* **Seasonal CRPS:** Evaluated across DJF, MAM, JJA, SON to assess robustness.

## 3.9 Design Decisions and Justification

* **ERA5 + IMERG integration:** Ensures complementary strengths (ERA5 drivers + IMERG observations).
* **30-minute alignment:** Matches operational flash flood warning timescales.
* **Multi-model framework:** Provides robustness—different models excel at different lead times.
* **Probabilistic focus:** Ensures forecasts quantify uncertainty, crucial for decision-makers.
* **Regionally weighted sampling:** Tailors the system to UK sub-regional flood risk.

## 3.10 Ethical Considerations

This study will investigate the application of deep learning methods in flash flooding radar

echo prediction and depending in the scale and quality of the data assets that could be

accessed, broadly the research objectives are expected to include synthetic data generation,

transfer learning, and physics-informed AI to overcome the challenges outlined above,

broadly the study objectives are expected to include the following main areas:

1. Data assts acquisition, synthesis and processing
2. Dimensionality reduction, feature ranking and attention mechanism
3. Atmospheric physics informed model performance refinement

**There are no ethical concerns involved in connection with the use of the available**

**datasets or the conduct of this research study.**

## 3.11 Summary

In this Chapter presented the methodological framework developed to design, implement, and evaluate a flood nowcasting system for the United Kingdom. The approach combines multi-source environmental data, advanced modelling techniques, and a comprehensive evaluation strategy to improve the prediction of flash floods triggered by extreme rainfall.

The study domain spans the UK window (49°–62°N, –13.0° to 3.5°E), covering the national landmass and surrounding seas, with sub-regional emphasis on the West Highlands, Lake District, and Snowdonia due to their elevated flood risk. Two key datasets underpin the framework: ERA5 reanalysis, which supplies physically consistent atmospheric variables at both surface and pressure levels, and IMERG satellite precipitation, which provides high-resolution rainfall estimates at 30-minute intervals.

Preprocessing ensures that these datasets are aligned and standardized. ERA5 variables are downscaled to match IMERG’s temporal resolution, regridded to a common 0.1° spatial grid, and normalized with quality checks and scaling. Additional feature engineering derives indices such as wind shear, lapse rates, and integrated moisture to capture storm dynamics more effectively.

The modelling component benchmarks a suite of architectures tailored to spatio-temporal forecasting: ConvLSTM and PredRNN for rainfall band dynamics, Transformer models for large-scale atmospheric context, conditional diffusion models for probabilistic ensembles, and ensemble baselines such as LightGBM-Quantile, Quantile Random Forest, and PySTEPS optical-flow advection. A standardized training schedule (2015–2016 training, 2019 validation, 2020 testing) enables fair comparison of model performance.

Evaluation emphasizes probabilistic and spatial skill using CRPS, FSS, and reliability diagrams, complemented by storm heatmaps and seasonal diagnostics. These metrics ensure models are assessed not only for accuracy but also for robustness, uncertainty quantification, and operational relevance.

Overall, Chapter 3 establishes a reproducible methodology that integrates multi-source datasets, advanced models, and rigorous evaluation, providing the foundation for credible, impact-aware nowcasting of extreme rainfall and flash flood risk across the UK.

# Chapter 4. Results

## 4.1 Introduction

This chapter presents the outcomes of the experiments designed to benchmark multiple nowcasting models for extreme rainfall and flash flood prediction across the United Kingdom. Results are structured around three key dimensions: (i) spatio-temporal reproduction of rainfall events, (ii) probabilistic and spatial skill evaluation, and (iii) robustness across regions and seasons. The findings provide insights into the relative strengths and limitations of each model family and highlight the trade-offs between predictive accuracy, computational efficiency, and interpretability.

## 4.2 Data Quality and Preprocessing Outcomes

The preprocessing pipeline successfully aligned ERA5 single-level and pressure-level predictors with IMERG precipitation at 30-minute intervals on a 0.1° grid spanning the UK domain (49°–62°N, –13°–3.5°E). Quality checks confirmed the removal of missing timesteps and the application of IMERG quality flags. However, computational challenges were encountered when merging the full 11-year dataset, with memory errors arising due to the volume of data. Workarounds included monthly batching and feature sub-selection. These limitations underline the scale of the task and justify the modular design of the pipeline.

Feature engineering produced meaningful derived predictors such as vertical wind shear, lapse rates, integrated vapour transport (IVT), anomalies in mean sea-level pressure (MSLP) and boundary-layer height (BLH), IMERG persistence metrics, moisture flux convergence, and a CAPE proxy. Visual inspections confirmed that these features provided additional contrast between extreme rainfall events and background conditions, supporting their value for downstream models.

## 4.3 Model Training and Runtime

Training followed the recommended schedule: 2015–2016 for training, 2019 for validation, and 2020 for testing, with refinements applied for sub-regions (West Highlands, Lake District, Snowdonia). Runtime analysis showed a wide spectrum of computational demands. Transformers and ConvLSTM/PredRNN required ~30–45 minutes per training run, diffusion models ~40–50 minutes, and tree ensembles (LightGBM, QRF) under 45 minutes. PySTEPS, by contrast, required only ~5–10 minutes to set up, confirming its suitability as an operational baseline. CPU-only training imposed constraints on deep models, limiting the number of epochs and the size of training windows.

## 4.4 Case Study: Extreme Rainfall Events

Five representative storm events, with IMERG precipitation exceeding 10 mm/hr, were selected as case studies. Heatmap visualisations compared observed rainfall fields with model forecasts at +1h, +3h, and +6h lead times.

* **ConvLSTM/PredRNN** reproduced spatial rainband structures realistically at short lead times but showed smoothing effects beyond 3 hours.
* **Transformer** models better preserved large-scale storm organisation, particularly frontal systems, while capturing embedded convective cores.
* **Diffusion models** excelled at generating ensemble forecasts that captured the spread of extremes, although ensemble means sometimes underestimated peak intensities.
* **PySTEPS** performed well for immediate extrapolation (+1h) but degraded rapidly by +3h.
* **Tree ensembles (LightGBM, QRF)** produced calibrated rainfall amounts but lacked spatial realism, reinforcing their role as interpretable but non-spatial comparators.

## 4.5 Probabilistic and Spatial Skill

Evaluation used CRPS, FSS, and reliability diagrams:

* **CRPS:** Diffusion and Transformer models achieved the lowest CRPS values across all lead times, indicating superior probabilistic skill. ConvLSTM/PredRNN performed comparably at 0–3h but diverged at longer horizons. Tree ensembles provided a useful calibrated baseline but underperformed relative to deep models.
* **FSS:** At thresholds of 5 and 10 mm/hr, ConvLSTM and PredRNN excelled at small spatial scales (<10 km), while Transformers and diffusion models maintained skill at broader scales (>20 km). PySTEPS dropped below useful skill levels by +3h, consistent with its deterministic design.
* **Reliability:** Diffusion ensembles showed the most consistent calibration, with forecast probabilities closely matching observed frequencies. Transformers tended to slightly overpredict extremes, while ConvLSTM/PredRNN showed underconfidence in high-intensity bins.

## 4.6 Seasonal and Regional Performance

Seasonal analysis of CRPS highlighted systematic differences. Deep models, particularly Transformers, achieved best skill during summer (JJA), when convective storms dominated, while ensemble methods retained relative robustness in winter (DJF), when large-scale rainfall systems prevailed. Regional analysis confirmed that weighted sub-region training improved model sensitivity in the West Highlands, Lake District, and Snowdonia, though generalisation to other UK regions sometimes declined slightly, reflecting the trade-off of targeted sampling.

## 4.7 Comparative Trade-Offs

A screenshot of a white sheet

AI-generated content may be incorrect.

Figure 7 Comparative Trade-Offs

## 4.8 Summary

The results demonstrate that advanced AI models, particularly Transformers and diffusion approaches, significantly enhance the ability to capture extreme rainfall events across the UK at 0–6 hour lead times. ConvLSTM and PredRNN retain strong utility for short-term forecasts, while ensemble tree methods provide interpretable guardrails. PySTEPS remains a transparent, low-latency baseline, though its skill diminishes with lead time. Together, these findings validate the project’s design choice to benchmark multiple model families and highlight the importance of probabilistic, multi-source, and regionally tailored approaches for credible flash flood nowcasting.

# Chapter 5. Discussion and Analysis

## 5.1 Scientific Contribution

The findings of this project demonstrate that advanced AI architectures can materially improve short-term forecasts of extreme rainfall in the UK. By integrating ERA5 atmospheric predictors with IMERG satellite rainfall at 30-minute resolution, the study establishes a reproducible multi-source pipeline that addresses both data scarcity and alignment challenges. This integration is significant because it shows that high-resolution satellite precipitation, when paired with physically consistent reanalysis fields, enables models to capture both the large-scale drivers and the localized rainfall extremes responsible for flash flooding. The inclusion of engineered features such as vertical shear, lapse rates, and integrated vapour transport reinforces the scientific value of the dataset by linking forecasts to physically interpretable storm dynamics.

## 5.2 Methodological Advancement

Benchmarking a diverse set of models under a standardized training and evaluation framework provides one of the first comprehensive comparisons of deep learning, ensemble, and operational baselines for UK flood nowcasting. The results highlight that Transformers and diffusion models outperform traditional baselines in both accuracy and probabilistic skill, while ConvLSTM and PredRNN retain strong utility at short lead times. Ensemble tree models deliver interpretability and computational efficiency, and PySTEPS remains a transparent baseline for operational comparison. This demonstrates that credibility in flood nowcasting is not derived from a single model but from combining complementary strengths across architectures. The methodological significance lies in showing how multi-model evaluation, region-specific refinements, and probabilistic verification can be combined into a practical pipeline for credible flood forecasting.

## 5.3 Practical and Operational Relevance

The project’s findings also carry strong operational implications. Results confirm that high-skill probabilistic models can improve decision-making for civil protection agencies by quantifying forecast uncertainty alongside expected rainfall. Seasonal analysis demonstrates that skill varies by storm type, with Transformers excelling in convective summer storms and ensemble methods showing robustness in winter synoptic systems. Regional weighting enhances sensitivity in high-risk areas such as the West Highlands, Lake District, and Snowdonia, aligning model design with the most flood-prone parts of the UK. The runtime analysis underscores feasibility: while deep learning models are resource-intensive, tree ensembles and PySTEPS provide fast guardrails and baselines, offering a layered approach to real-world implementation.

## 5.4 Addressing Research Challenges

Several key challenges identified at the outset are directly addressed by these findings. The problem of **spatio-temporal complexity** is mitigated through models that capture both temporal evolution (ConvLSTM, PredRNN) and spatial context (Transformers, diffusion). **Model generalization** is improved by regional weighting, though the trade-off with nationwide skill highlights the importance of future transfer learning. **Data gaps and inconsistencies** were handled through careful preprocessing, quality control, and engineered persistence features, which ensured continuity despite missing values. Finally, the challenge of **interpretability** is addressed by combining AI methods with physics-informed features and by including ensemble tree models that provide feature importance insights.

## 5.5 Broader Impact

The significance of this study extends beyond the dissertation. It shows a clear pathway toward operational, impact-aware flood forecasting systems in the UK. The probabilistic approach ensures that uncertainty is not hidden but explicitly communicated, increasing trust and usability for stakeholders. The integration of physics-informed features demonstrates how machine learning can complement, rather than replace, traditional meteorological understanding. Moreover, the modular design of the pipeline means it can be scaled or adapted for other regions or hazards. With further development, this research could inform not only academic publications but also practical early-warning systems that reduce flood risk and protect vulnerable communities.

## 5.6 Limitations

This study, while advancing a reproducible and credibility-focused framework for flood nowcasting, faced several limitations that should be acknowledged. The first relates to **computational resources**. Training deep learning architectures such as ConvLSTM, PredRNN, Transformers, and diffusion models on over a decade of ERA5 and IMERG data was restricted by hardware constraints. Running on CPU hardware proved inefficient and, in practice, incompatible with the volume of data required for robust model training. Access to high-performance GPU clusters would be necessary to fully exploit the spatio-temporal depth of the dataset and to train models at scales consistent with operational forecasting.

A second limitation concerns **data integration and memory handling**. Merging ERA5 single-level and pressure-level predictors with IMERG precipitation for spatial alignment introduced significant computational overhead. The conversion of such large datasets into consistent 30-minute, 0.1° resolution inputs frequently led to memory errors, making it challenging to process the full record in a single workflow. While strategies such as batch processing, progressive saving, and variable sub-selection were applied, these workarounds also constrained the efficiency of the pipeline.

Finally, while the study incorporated multiple models and probabilistic evaluation methods, the reliance on a single reanalysis product (ERA5) and one satellite precipitation dataset (IMERG) limits the scope of uncertainty analysis. Incorporating additional observational datasets and testing model robustness under different data regimes would further strengthen credibility.

These limitations highlight the trade-offs between data ambition, computational feasibility, and methodological scope. They also point to clear avenues for future work, particularly in optimizing data workflows and securing high-performance computing resources to unlock the full potential of the proposed framework.

# Chapter 6. Conclusion and Future Work

## 6.1 Conclusion

## 6.2 Future work

While this study establishes a credibility-focused nowcasting pipeline for extreme rainfall and flash floods across the UK, several directions remain open for further development. First, expanding the spatial domain beyond the current UK window (49°–62°N, –13°–3.5°E) to include a larger buffer region would allow the models to better capture approaching storm systems before they reach the UK, improving lead times. Second, the data foundation could be enhanced by incorporating additional high-resolution sources such as UK radar composites, ERA5-Land, or soil moisture products, which would strengthen the representation of surface hydrological conditions linked to flash flood generation. Third, the preprocessing and feature engineering steps could be extended to include physics-informed variables such as convective available potential energy (CAPE), potential vorticity, or terrain-based indices, allowing for richer predictors and improved interpretability.

From a modelling perspective, future work should investigate hybrid systems that couple deep learning architectures with physically based hydrological models, bridging the gap between statistical prediction and process-based simulation. Such integration could enhance trust and operational uptake by embedding physical consistency. In addition, while this study benchmarked multiple architectures under a standardized training schedule, future experiments should explore transfer learning and regional fine-tuning to improve generalization across diverse UK sub-regions. Model uncertainty remains a critical area, and future work should evaluate ensemble post-processing methods and probabilistic calibration techniques to refine forecast reliability.

Finally, evaluation could be broadened by incorporating impact-based metrics, such as comparisons against observed flood events or damage records, to ensure that forecasts are directly relevant to decision-makers. Extending the framework to real-time operation, with low-latency data streams and continuous model updating, would be the ultimate step toward deploying this research as a practical early-warning system for communities vulnerable to flash flooding.

# Chapter 7. Reflection

While this study establishes a credibility-focused nowcasting pipeline for extreme rainfall and flash floods across the UK, several directions remain open for further development. First, expanding the spatial domain beyond the current UK window (49°–62°N, –13°–3.5°E) to include a larger buffer region would allow the models to better capture approaching storm systems before they reach the UK, thereby improving forecast lead times. Second, the data foundation could be strengthened by incorporating additional high-resolution sources such as UK radar composites, ERA5-Land, or soil moisture products, which would enhance the representation of surface hydrological conditions linked to flash flood generation. Third, the preprocessing and feature engineering stages could be extended to include physics-informed variables such as convective available potential energy (CAPE), potential vorticity, or terrain-based indices, enabling richer predictors and improved interpretability.

From a modelling perspective, future work should focus on hybrid systems that combine deep learning architectures with physically based hydrological models, bridging the gap between purely statistical prediction and process-based simulation. Such integration could enhance trust and operational uptake by embedding physical consistency. While this study benchmarked multiple architectures under a standardized training schedule, future experiments should also explore transfer learning and regional fine-tuning to improve generalization across diverse UK sub-regions. Model uncertainty remains a critical challenge, and further research is needed on ensemble post-processing methods and probabilistic calibration techniques to refine reliability.

Finally, evaluation should be broadened to incorporate impact-based metrics, such as comparisons against observed flood events or reported damages, to ensure forecasts directly support decision-making. Extending the framework to real-time operation, with low-latency data streams and continuous model updating, represents the ultimate step toward deploying this research as a practical early-warning system. Due to time constraints, not all of these developments could be achieved within this dissertation. However, I intend to continue working on this project beyond submission, with the aim of further refining the methodology and presenting the outcomes as a research paper for academic dissemination.

# Peer Review

**Peer Review**

**Name of the reviewee:** GreeshmiPriyanka Appalapuram

**Title of the project being reviewed:**Using Conditional Diffusion Models to Generate Physically and Impact-Consistent Synthetic Windstorm Events for the UK (Author: Yinting Feng)

**Project Background:**  
The background is clearly stated and well-structured. It establishes the urgency of predicting extreme rainfall and flash floods in the UK, highlighting both their socio-economic impacts and the scientific challenges of forecasting highly localized, short-term events. The motivation is strong, and the choice of ERA5 and IMERG as primary datasets is well justified.

**Literature Review:**  
The literature review is comprehensive and well-linked to the research objectives. References are cited in IEEE format and cover both classical flood forecasting approaches and recent advances in AI and generative modelling. While thorough, the review could be improved by including additional UK- and Europe-specific studies to further contextualize the regional focus.

**Research Question(s) / Problem Definition:**  
The research questions are clearly formulated and directly aligned with the project aim of developing a credibility-focused nowcasting pipeline. They reflect the need for probabilistic, interpretable models and address challenges in data integration, model generalisation, and operational applicability. The definition is precise and appropriate for quantitative, data-driven methodologies.

**Development Approach:**  
The development approach is ambitious yet feasible. The project proposes a multi-stage pipeline including preprocessing, feature engineering, and model training across a range of architectures (ConvLSTM, PredRNN, Transformer, Diffusion, and ensemble baselines). The inclusion of probabilistic evaluation (CRPS, FSS, reliability diagrams) strengthens the methodological rigor. Computational limitations (CPU-only training, memory constraints in ERA5–IMERG merging) are acknowledged as risks, with mitigation strategies outlined (batching, feature reduction, modular workflow design).

**Social, Legal, and Ethical Considerations:**  
Although not the central focus of the project, awareness of social and ethical aspects is demonstrated. The project recognises the importance of interpretability and physics-awareness to ensure forecasts are trusted by operational users. Ethical concerns such as the consequences of false alarms or missed events are implicit in the emphasis on credibility and uncertainty quantification.

**Project Objectives & Task/Time Planning:**  
The objectives are clearly defined, measurable, and logically sequenced. The timeline is realistic, balancing technical work with reporting. Risks are acknowledged (computational limits, data size), and mitigation strategies are included. While ambitious, the plan is structured in a way that prioritises core deliverables, with additional experiments (e.g., hybrid models, transfer learning) flagged for future work beyond the dissertation.

**Overall Evaluation:**  
The specification demonstrates strong clarity, ambition, and feasibility. It balances methodological depth with practical awareness of challenges. The project is well designed to deliver a credible nowcasting framework and has clear potential for extension into academic publications.

# Reference

1. **AI tools used to support my work:** Gemini, ChatGPT
2. H. A. P. Hapuarachchi, Q. J. Wang, and T. C. Pagano, “A review of advances in flash flood forecasting,” *Hydrological Processes*, vol. 25, no. 18, pp. 2771–2784, 2011. (accessed: June 27,2025)
3. T. A. Khan, M. M. Alam, Z. Shahid, and M. M. Su’ud, “Investigation of flash floods on early basis: A factual comprehensive review,” *IEEE Access*, vol. 8, pp. 171904–171929, 2020.(accessed: June 28, 2025)
4. C. G. Collier, “Flash flood forecasting: What are the limits of predictability?,” *Q. J. R. Meteorol. Soc.*, vol. 133, no. 622, pp. 3–23, 2007. (accessed: June 27,2025)
5. J. A. Smith, M. L. Baeck, K. L. Meierdiercks, A. J. Miller, and W. F. Krajewski, “Radar rainfall estimation for flash flood forecasting in small urban watersheds,” *Adv. Water Resour.*, vol. 30, no. 10, pp. 2087–2097, 2007. (accessed: July 2, 2025)
6. A. Gerard, S. M. Martinaitis, J. J. Gourley, and Y. Zhang, “An overview of the performance and operational applications of the MRMS and FLASH systems in recent significant urban flash flood events,” *Bull. Amer. Meteor. Soc.*, vol. 102, no. 3, pp. E611–E626, 2021. (accessed: July 2,2025)
7. J. J. Gourley, Z. L. Flamig, Y. Hong, and E. Habib, “Evaluation of past, present and future tools for radar-based flash-flood prediction in the USA,” *Hydrol. Sci. J.*, vol. 59, no. 7, pp. 1–19, 2014. (accessed: July 4, 2025)
8. G. J. Huffman et al., “Integrated Multi-satellite Retrievals for GPM (IMERG),” *J. Hydrometeorol.*, vol. 21, no. 3, pp. 1610–1633, 2020. (accessed: July 4, 2025)
9. H. S. Munawar, A. W. A. Hammad, and S. T. Waller, “Remote sensing methods for flood prediction: A review,” *Sensors*, vol. 22, no. 4, pp. 1258, 2022. (accessed: July 10, 2025)
10. A. D. L. Zanchetta and P. Coulibaly, “Recent advances in real-time pluvial flash flood forecasting,” *Water*, vol. 12, no. 2, pp. 540, 2020. (accessed: July 9, 2025)
11. A. Mosavi, P. Ozturk, and K. Chau, “Flood prediction using machine learning models: Literature review,” *Water*, vol. 10, no. 11, pp. 1536, 2018. (accessed: July 7, 2025)
12. X. Shi et al., “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” in *Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810. (accessed: July 12, 2025)
13. Y. Wang et al., “PredRNN: Recurrent neural networks for spatiotemporal predictive learning,” in *Adv. Neural Inf. Process. Syst.*, 2017, pp. 879–888. (accessed: July 14, 2025)
14. C. K. Sønderby et al., “MetNet: A neural weather model for precipitation forecasting,” *arXiv preprint arXiv:2003.12140*, 2020. (accessed: July 15, 2025)
15. S. Ravuri et al., “Skillful precipitation nowcasting using deep generative models of radar,” *Nature*, vol. 597, pp. 672–677, 2021. (accessed: July 14, 2025)
16. A. Lugmayr et al., “RePaint: Inpainting using diffusion models,” *arXiv preprint arXiv:2201.09865*, 2022. (accessed: July 16, 2025)
17. A. Mishra, S. Mukherjee, B. Merz, and V. P. Singh, “An overview of flood concepts, challenges, and future directions,” *J. Hydrol. Eng.*, vol. 27, no. 11, 2022. (accessed: July 17, 2025)
18. H. A. P. Hapuarachchi and Q. J. Wang, “A review of methods and systems available for flash flood forecasting,” Bureau of Meteorology Report, 2008. (accessed: July 6, 2025)
19. S. Rozalis, E. Morin, Y. Yair, and C. Price, “Flash flood prediction using an uncalibrated hydrological model and radar rainfall data in a Mediterranean watershed under changing hydrological conditions,” *J. Hydrol.*, vol. 394, no. 1–2, pp. 245–255, 2010. (accessed: July 6, 2025)
20. D. R. Archer and H. J. Fowler, “Characterising flash flood response to intense rainfall in the UK,” J. Hydrol., vol. 564, pp. 905–919, 2018. (accessed: June 26, 2025)
21. L. Collet et al., “Future projections of summer flood hazard in the UK using convection-permitting climate simulations,” Hydrol. Earth Syst. Sci., vol. 22, no. 12, pp. 6593–6610, 2018. . (accessed: June 26, 2025)
22. D. L. A. Flack, C. J. Skinner, L. Hawkness-Smith, et al., “Recommendations for improving integration in national end-to-end flood forecasting systems: An overview of the FFIR project,” Water, vol. 11, no. 4, pp. 725, 2019. . (accessed: August 2, 2025)
23. T. Sayama, M. Yamada, Y. Sugawara, T. Iwami, and K. Tanaka, “Ensemble flash flood predictions using a high-resolution nationwide distributed rainfall-runoff model: Case study of the heavy rain event of July 2018 and Typhoon Hagibis 2019,” Prog. Earth Planet. Sci., vol. 7, pp. 1–15, 2020. . (accessed: July 10, 2025)
24. H. L. Cloke and F. Pappenberger, “Ensemble flood forecasting: A review,” J. Hydrol., vol. 375, no. 3–4, pp. 613–626, 2009. (accessed: August 10, 2025)