UNCERTAINTY PROPAGATION THROUGH MACHINE LEARNING MODELS OF HYDROLOGICAL AND HYDRODYNAMIC SYSTEMS

Euan Andrew Macdonald
Supervised by
Dr Enrico Tubaldi and Professor Edoardo Patelli

14/08/2025

Declaration of Authenticity and Author's Rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree. The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:
Euan Macdonald
Date:
14/08/2025

Previously Published Works

- Macdonald, E, Tubaldi, E, Ferguson, N and Patelli, E; 2024. “Scottish Trunk Road Fragility to Extreme Precipitation”. Scottish Roads Research Board Technical Report.

- MacDonald, E., Tubaldi, E. and Patelli, E., 2025. Robust storm surge forecasts for early warning system: a machine learning approach using Monte Carlo Bayesian model selection algorithm. Stochastic Environmental Research and Risk Assessment, p.106184.

- Macdonald, E, Tubaldi, E and E Patelli. 2025. “Machine Learning-Based Coastal Inundation Forecasting with Explicit Uncertainty Quantification.”
  Under Review at Coastal Engineering

Source Code Availability

The source code and example applications for surge, wave and inundation forecasting models are fully documented and accessible in the Bayesian Coastal Toolbox Git repository at:

https://github.com/emacd-domain/Bayesian_Coastal_Forecasting.git

or via ssh using the following address:

git@github.com:emacd-domain/Bayesian_Coastal_Forecasting.git

Acknowledgements

I would like to extend my sincere gratitude to my parents, whose continuous encouragement over many years provided me with curiosity to begin this work, and the drive to finish it.

I am also profoundly grateful to my supervisors for their guidance and insights for the duration of this research. I owe special thanks to Enrico for his rigorous standards and attention to detail, and to Edoardo for his relaxed demeanour and fresh perspectives.

Finally, I must acknowledge my wife, Kateryna, for her support, patience, and understanding, especially during the write up for this thesis. Her quiet confidence in me has been both a comfort and a motivator.

Abstract

This thesis explores the use of machine learning and uncertainty quantification methods to improve prediction and forecasting for extreme weather events, with a focus on storm surges, coastal inundation and road flooding. Machine learning-based approaches, such as Artificial Neural Networks combined with Bayesian model selection and Monte Carlo simulations, are employed to enhance the accuracy of storm surge forecasts by propagating input uncertainty and providing confidence intervals. The methodology is applied to a case study in Millport, Scotland, demonstrating improved predictive performance and computational efficiency, with Pearson correlation coefficient of 0.942 for 24-hour surge forecasts. In the context of coastal inundation, a framework is presented that incorporates aleatoric and epistemic uncertainties, with operational validation during Storm Ciara in the Firth of Clyde, showing its effectiveness in addressing complex coastal flood risks. Additionally, the thesis addresses the fragility of Scotland's trunk road network to disruption from precipitation events, and particularly the development of empirical fragility curves to quantify the vulnerability of transportation infrastructure. The analysis, based on data from Transport Scotland, SEPA, and NIMROD, provides insights into the potential impacts of extreme weather on critical infrastructure, with a focus on uncertainty at each stage of the forecasting process. The thesis concludes with reflections on the challenges and potential improvements to these methodologies for future climate resilience and infrastructure planning.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Organisations	
BODC	British Oceanographic Data Centre
C3S	Copernicus Climate Change Service
CEDA	Natural Environment Research Council's Data Repository for Atmospheric Science and Earth Observation, which holds a number of Met Office Datasets
CEH	Centre of Ecology and Hydrology
CESM	Community Earth System Model
CMEMS	Copernicus Marine Environment Monitoring Service
ECMWF	European Centre for Medium-range Weather Forecasts
IPCC	The Intergovernmental Panel on Climate Change
NCEP	US National Centers for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
NSLF	National Sea Level Facility
NTSLF	National Tidal and Sea Level Facility
SEPA	Scottish Environmental Protection Agency
SWEEP	The Southwest Partnership for Environmental and Economic Prosperity
TRL	Transport Research Laboratory
UKHO	UK Hydrographic Office
USACE	US Army Corps of Engineers

Technical	
ABMS	Adaptive Bayesian Model Selection Algorithm
ADCIRC	ADvanced CIRCulation Model
ANN	Artificial Neural Networks
API	Application Programme Interface
ARIMA-SVR	Autoregressive Integrated Moving Average-Support Vector Regression
BMS	Bayesian Model Selection
CD	Chart Datum
CFD	Computational Fluid Dynamics
cGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Networks
CVI	Coastal Vulnerability Index
DDF	Depth-Duration-Frequency
DESYCO	DEcision support SYstem for COastal climate change impact assessment
DIVA	Dynamic and Interactive Vulnerability Assessment
DT	Decision Tree
ERA-5	ECMWF (European Centre for Medium-range Weatehr Forecasts) Re-Analysis
ESTOFS	Extratropical Surge and Tide Operational Forecast System
FPV	Flood Potential Value
FPV	Flood Potential Value
FVCOM	Finite Volume Community Ocean Model

GMM	Gaussian Mixture Model
GN SS	Global Navigation Satellite Systems
GRNN	General Regression Neural Network
IDF	Intensity-Duration-Frequency
IDW	Inverse Distance Weighting
IFS	Integrated Forecasting System
KED	Kriging with External Drift Method
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MCBA	Monte Carlo Bayesian Averaging Algorithm
MLP	Multilayer Perceptron
MSLP	Mean Sea Level Pressure
OWWL	Operational coastal Wave and Water Level model
P-Box	Probability Box
P-Surge	Probabilistic Surge
PBL	Planetary Boundary Layer
PCA	Principal Component Analysis
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
RSFS	Robust ANN-based surge forecasting system
RWFS	Robust ANN-based wind wave forecasting system
SBL	Surface Boundary Layer
SI	Scatter Index
SoVI	Social Vulnerability Index
SVM	Support Vector Machines
SWAN	Spectral Wave Model
VC	Visual Crossing
VMD	Variational Mode Decomposition
WAM	WAve model for which WW3 was developed
WW3	NOAA (National Oceanic and Atmospheric Administration) wave model WAVEWATCH III
Xbeach	Numerical Model that Simulates Hydrodynamic and Morphodynamicprocesses on Coastlines
XBNH	Nonhydrostatic mode
XBSB	Surf-Beat mode
XGBoost	Extreme Gradient Boost

Other

ENSO	El Niño-Southern Oscillation climate phenomena
GIS	Geographical Information System
IRIS	Integrated Road Information System
NIMROD	Met Office's automated weather analysis and forecasting system
OC	Operating Company

List of Symbols

Symbols	Meaning	Units
a	feature uncertainty variance	
b	beach slope steepness	
C	consequences	
CD	wave celerity in deep water	ms-1
CT	wave celerity in transitional water	ms-1
d	depth of SWL	m
ɗ	probability distribution	
D	damage/disruption	
E_{ind}	matrix of independent random errors	
E_{corr}	matrix of correlated random errors	
Fa	adjustment factor	
H	Wave height	m
H0	Wave height at initial location	m
H1	Wave height at nearshore location	m
Ha	astronomic tidal height	m
Hmax	maximum water height	m
Hr	non-tidal residual height	m
Hs	significant wave height	m
Hu	wave runup	m
Hu,2%	wave runup of 2% exceedance probability	m
Hz	hazard	
I	intensity measure	
L	wavelength	m
Ł	Cholesky matrix	
M	number of neural networks	
MSLP	mean sea level pressure	Pa
N	neural network model	
N_F	number of failures	
N_o	number of observations	
PC	principal component	
ΔP_{local}	pressure difference between the target location and the maximum pressure in the domain	Pa
Q	overtopping discharge Volume	m3
R	risk	
Rc	freeboard	m
SWL	still water level	m
su	setup	m
sw	swash	m
U10	10m eastward wind velocity component	ms-1
V10	10m northward wind velocity component	ms-1
VG,D	deep water group velocity	ms-1

VG,T	transitional water group velocity	ms-1
X_T	realised value of forecasted variable	
$X_{forecast}$	biased forecast of forecasted variable	
$\tilde{X}_{forecast}$	unbiased forecast of forecasted variable	
y	observed target	
yp	predicted target	
yabms	abms averaged prediction	
ymcba	mcba averaged prediction	
$\hat{y}^*$	best network prediction	
Z(x,y)	sample dataset with features x and targets y	
α	regression slope	
β	regression intercept	
σ	standard deviation	
ε	error	
ɴ	normal distribution	
Σ	covariance matrix	

Operation	Meaning
E	expected value
P	probability
T	transpose
V	variance
|	given
∩	intersect
¯	mean

Chapter 1
Introduction

All types of flooding including pluvial, fluvial, and coastal have the potential to cause severe and widespread damages, affecting the environment, society, and economy (Rosenzweig, et al., 2018), (Feyen, Dankers, Bodis, Salamon, & Barredo, 2012), (Hinkela, et al., 2014). Socially, coastal flooding poses significant risks to human safety, often causing injuries, fatalities, and the displacement of communities. It can destroy homes, schools, and healthcare facilities, straining public services and leading to long-term social impacts on affected populations. Economically, the consequences are equally devastating, as flooding damages infrastructure such as roads, bridges, ports, and utilities, disrupting transportation and commerce (Winter, et al., 2016). Businesses may face prolonged closures, with immediate financial impacts including loss of trade and income, or damaged stock to longer term financial impacts such as repair costs and increase insurance premiums (Wedawatta, Bingunath, & Proverbs, 2014). The cumulative effect of these damages highlights the critical need for effective flood management and climate adaptation strategies to mitigate the impacts of flooding and inundation.

Understanding the extensive damage caused by different types of flooding sets the stage for a deeper examination of how climate change exacerbates these events. The Intergovernmental Panel on Climate Change have highlighted how increased global temperatures contribute to more intense storms, and that global mean sea level will rise by 0.43 to 0.84 meters by the end of this century, depending on future emissions scenarios (IPCC, Special Report on the Ocean and Cryosphere in a Changing Climate, 2022). As sea levels rise, the risk of coastal flooding increases, as higher sea levels will most likely result in more frequent and severe flooding events (Nicholls, Hinkel, Lincke, Suckall, & Tol, 2018). Additionally, more severe storms will result in more severe storm surges, which further contribute to coastal flooding (Emanuel, 2017). With respect to the future of Extratropical storms (ETS) in the UK, UK Climate Projections (UKCP) indicates small increases in mean winter wind speed (and small summer decreases), with the magnitude of change uncertain and small relative to natural variability (Met Office, UKCP18 Factsheet: Storms, 2023). Nonetheless, multiple model lines of evidence—including the UKCP Global ensemble—project an eastward extension of the North Atlantic storm track and an increase in winter storm numbers and intensity over the UK by late century, even as neighbouring regions see decrease.

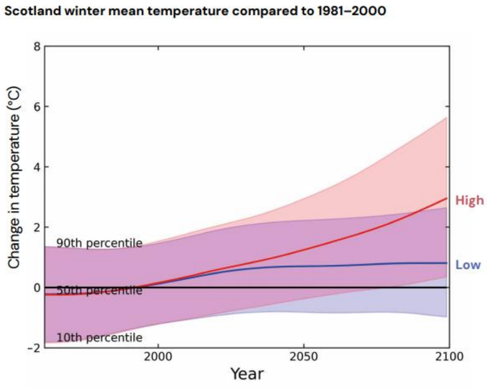


Figure 1: Adaption Scotland: Predicted Winter Temperature Changes, (Adaption Scotland, 2021).

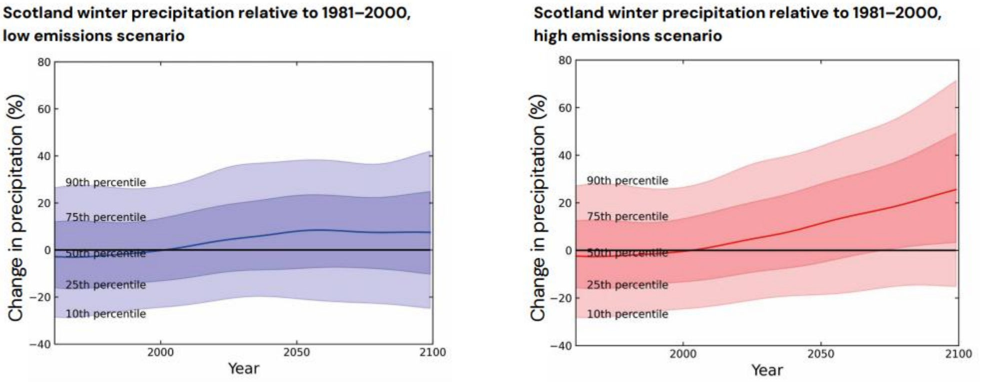


Figure 2: Adaption Scotland: Predicted Winter Rainfall Changes under different Emission Scenarios, (Adaption Scotland, 2021).

According to Adaption Scotland projections, as a result of climate change Scotland will experience warmer (Figure 1), wetter (Figure 2) winters, with more intense rainfall. By 2050, under a low emissions scenario, average winters are projected to be around 8% wetter and 1 degree warmer. By 2080, under a high emissions scenario average winters are projected to be around 19% wetter and 2.7 degrees warmer. According to the Centre of Ecology and Hydrology (CEH) report (Kay, Crooks, Davies, & Reynard, 2011), a potential increase in regional flood peak for different emission scenarios for the 2080 time horizon can also be expected. The average increase in peak flow for the ten main river basins in Scotland has been estimated as 7-20% for low emission, 10-24% for medium emission, and 13-31% for high emission scenarios (Kay, Crooks, Davies, & Reynard, 2011). Similarly, the anticipated rise in high intensity rainfall is likely to lead to more frequent road flooding and closures. With respect to projected sea level rise, Adaption Scotland estimate 0.5m to 0.8m around the Scottish coast (Figure 3) under high emission scenarios (Adaption Scotland, 2021).

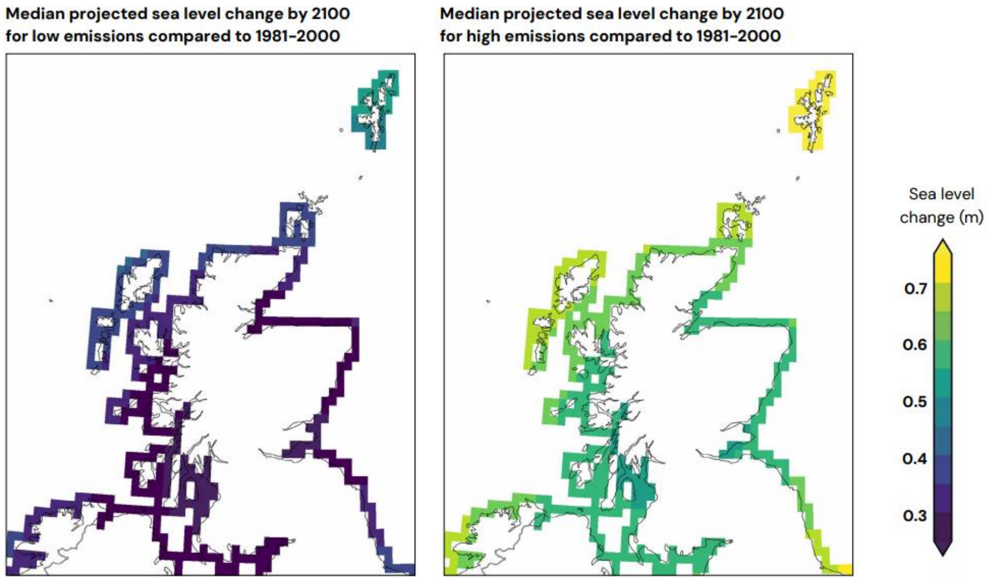


Figure 3: Adaption Scotland: Predicted Sea Level Rise under Different Emission Scenarios, (Adaption Scotland, 2021).

In terms of flood related consequences, between 1970 and 2019, water-related hazards accounted for 50% of all disasters and 45% of disaster-related deaths, according to the UNDRR's 2025 Global Assessment Report (UNDRR, 2025). Flooding remains the most economically damaging hazard, with current annual average global losses estimated at USD 388 billion, including USD 74.1 billion in infrastructure damage. These losses are projected to rise significantly due to climate change, reaching between USD 407–439 billion by 2050 depending on emission scenarios. Under the high-emission RCP8.5 pathway, infrastructure losses from riverine flooding alone could increase by up to 13%.

With the clear link between rising sea levels and increased flooding, the importance of effective flood management and adaptation strategies becomes even more apparent. These strategies are crucial for mitigating the risks and preparing for the projected increase in flood events. As we consider the vital strategies for flood management, it is equally important to recognize the role of stakeholders in implementing these measures. It is the role of responsible stakeholders to ensure safety, resilience, and economic sustainability through these possible outcomes and as such, decisions must be made regarding the prioritisation of resources and effective strategy implementation. For these decisions to be maximally informed, it is essential for the level of risk associated with each event to be quantified or qualified. We take the risk (R) associated with a hazardous event such as a coastal flood as the product of the probability of the hazard occurrence P(Hz), the probability of damage/disruption given the hazard P(D|Hz) (denoted as "fragility" or "vulnerability"), and the consequences of damage/disruption (C):

$$R = P(Hz) \times P(D|Hz) \times C \qquad \text{Eq. [1]}$$

The commonly applied formula introduced here encapsulates the complexities of risk assessment, combining hazard probability, vulnerability, and potential consequences into a comprehensive metric. It follows that the overall understanding of the risk involved with any particular scenario will be enhanced through improved understanding of any of the constituent components, the development of robust early warning systems and predictive models. These tools are essential for timely and effective emergency management, reducing overall vulnerability and mitigating potential impacts. For example, robust early warning tools that are capable of accurately predicting the hazard *Hz* with sufficient lead time can allow time for risk-mitigation and emergency-management measures. can be taken, which will reduce the vulnerability *P(D|Hz)* and/or consequences *C*. The components of Eq. [1] will be referenced throughout the thesis, when relevant to specific models.

Modern risk models integrate physical, economic, and social vulnerabilities to quantify exposure and potential damages across different sectors (Pollard, Spencer, & Jude, 2018). Because climate projections, storm behaviour, and human responses are inherently variable, uncertainty analysis—typically via probabilistic methods—replaces single-point forecasts with distributions, enabling stakeholders to plan across plausible best- and worst-case outcomes. These integrated, uncertainty-aware models are therefore essential for preparedness and response.

Accurately forecasting coastal flood events (particularly extremes) is difficult, and has multiple challenges such as local variability, a lack of knowledge around extreme events, data quality and assimilation issues, see Hawkes et al. (2008) for surge and Stockdon et al. (2023) for wave runup. A major step forward in this discipline has been the introduction and continuous development of computational fluid dynamics (CFD) coupled with a similar improvement in computational power and efficiency. Sophisticated hydrodynamic models such as ADCIRC (Luettich, Jr, Westerink, & Scheffner, 1992) now incorporate complex feedback loops, such as the interplay between storm surge, wave action, and rising sea levels, to provide more accurate predictions of flood hazards. The improvements in computational power and algorithms allow these models to run faster and accommodate finer spatial and temporal scales, enabling localised and event-specific forecasting.

Over the same time period, uncertainty analysis has also evolved, identifying model parameters, model physics inadequacies and initial conditions as potential sources of uncertainty, among others. Uncertainty methods in flood forecasting can be broadly classed as either statistical methods or ensemble methods. Statistical methods, acting as post-processors, estimate uncertainty by analysing past model residuals (the difference between the predicted and observed values) under the assumption that past uncertainty reflects future uncertainty. These methods, such as quantile regression, vary in complexity and face challenges in reliably representing uncertainty (Boelee, Lumbroso, Samuels, & Cloke, 2019). Ensemble models are created by combining different model runs, each representing a feasible configuration of the model structure, parameters, and forcing data, to quantify forecast uncertainty arising from limited knowledge of processes, parameters, or initial conditions (Boelee, Lumbroso, Samuels, & Cloke, 2019). The evolution of uncertainty analysis, especially in the realm of flood forecasting, has benefited greatly from advances in computational capabilities and machine learning techniques. These innovations enable more accurate predictions and efficient handling of complex data sets.

One of the drawbacks of conducting uncertainty analysis using high-fidelity modelling is that these models are definitionally complex and as such, computationally demanding. High-fidelity models often simulate intricate physical processes at fine spatial and temporal resolutions, requiring extensive computational resources to execute. This complexity means that conducting multiple simulations to explore a range of uncertainties, such as variations in input parameters or boundary conditions, can be prohibitively time-consuming and expensive. Additionally, the interdependencies between variables in high-fidelity models can amplify uncertainty, making it difficult to isolate and quantify the influence of individual factors (Stokes, et al., 2021). The incorporation of stochastic elements, necessary for capturing real-world variability, further complicates the process by increasing the dimensionality of the problem. As a result, generating reliable probabilistic outcomes or sensitivity analyses becomes a formidable task, often requiring significant computational efforts.

To overcome this challenge, innovative approaches such as surrogate modelling—which uses simplified models to replicate the outputs of more complex simulations—can significantly reduce computational costs. These surrogate models, often built using machine learning algorithms like Gaussian processes or neural networks, enable rapid exploration of input spaces. When combined with probabilistic methods such as Bayesian inference, they can also support rigorous uncertainty quantification by estimating the likelihood of different outcomes and updating beliefs based on observed data. This integration of statistical and computational tools allows for efficient and scalable analysis where direct simulation would be prohibitive. All of these modelling options are discussed in depth in Chapter 2-3. They could also be used to develop efficient and accurate empirical predictive models. These models excel in identifying subtle correlations and generating real-time flood forecasts, often outperforming traditional physics-based models and statistical methods. Moreover, machine learning algorithms are being used to enhance the accuracy of numerical models e.g. (Liu, Guoyou, & Kaige, 2019) for astronomic tide height and/or improve the computational efficiency e.g. (Shahabi & Tahvildari, 2024) for coastal water levels.

The fortuitous convergence of machine learning techniques, advanced computational capabilities, high-precision modelling, and uncertainty analysis has created a unique opportunity to address the challenge of flood-related risks intensified by a changing climate. This forms the central focus of this thesis: developing methods for propagating uncertainty through hydrological and hydrodynamic systems using machine learning models. This dissertation examines two key applications: forecasting coastal surge and inundation hazards and evaluating the vulnerability of road networks to pluvial flooding.

Given the limitations of conventional flood models in representing uncertainty and the limited adoption of ML-based surrogate models in real-world decision-making, this thesis positions itself within the context of increasingly complex flood forecasting needs and emphasizes a dual innovation: in maximising the precision of machine learning surrogate models, and the explicit quantification of input, forecast and model uncertainty through the coupling of artificial neural networks (ANNs) and Bayesian methods. Unlike many prior studies that apply machine learning superficially (Chapter 2-4), this work presents a unified framework that jointly addresses predictive accuracy and the propagation of both aleatoric and epistemic uncertainties. Furthermore, the research introduces a novel combination of Monte Carlo-based uncertainty and kriging metamodels to address vulnerability estimation in road networks — an area with previously limited empirical development. The thesis also outlines three case studies that validate its methodology with real-world environmental and infrastructure data, establishing its practical relevance and contribution beyond existing literature.

In this work, *surrogate modelling* refers to the development of computationally efficient approximations that emulate the behaviour of high-fidelity models, such as kriging metamodels used to derive fragility curves for infrastructure under extreme precipitation. *Probabilistic approaches* are employed to move beyond deterministic forecasts by generating ranges of possible outcomes, thereby capturing the inherent variability in storm surge and rainfall events. Central to this framework is the use of *Bayesian methods*, which facilitate model selection under uncertainty by updating beliefs based on data — specifically via an Adaptive Bayesian Model Selection (BMS) process enhanced with Monte Carlo sampling. These Bayesian tools not only quantify prediction intervals but also account for model structure, feature error, and observational uncertainty, enabling more robust forecasting and risk estimation.

The rest of the thesis is laid out as follows: Chapter 2 contains a six part literature review that covers the meteorological drivers behind the flood hazards; coastal inundation risk components, pluvial flood risk components, machine learning based forecasting, uncertainty analysis and finally a summary of the literature review findings. Chapter 3 is a description of the probabilistic processes that comprise the two novel and innovative branches of this thesis: Bayesian model selection algorithms for surge and coastal inundation forecasting and conditional probability processes for relating extreme precipitation to road failure likelihood. Chapter 4 is an adapted version of a journal paper titled “Robust Storm Surge Forecasts for Early Warning System: A Machine Learning Approach Using Monte Carlo Bayesian Model Selection Algorithm” that is currently under review in the journal of Stochastic Environmental Research and Risk Assessment. It describes the design and implementation of an uncertainty based storm surge forecasting model and its application in the Firth of Clyde. Chapter 5 is an adapted paper titled “Robust Localised Coastal Inundation Forecasting: A Deep Learning Approach” that has been submitted to the Coastal Engineering Journal. It describes the design and implementation of an uncertainty based surface wave characteristic forecasting model and its function as part of an uncertainty based coastal inundation forecasting framework. Chapter 6 is an adapted Scottish Roads Research Board funded report titled “Scottish Trunk Road Fragility to Extreme Precipitation”, which will form the basis of another journal publication. It describes the production of fragility curves relating the risk of disruption from road flooding to the intensity of precipitation using kriging metamodels and probability theory. Chapter 7 is an overarching conclusion of Chapters 3-6 and a discussion of the results along with recommendations for future research.

1-References

- Adaption Scotland. (2021). *Climate Projections for Scotland.* Scottish Government. Retrieved from https://www.adaptationscotland.org.uk/application/files/5716/1114/1258/Climate_projections_for_Scotland_summary_single_page_FINAL.pdf
- Boelee, L., Lumbroso, D., Samuels, P., & Cloke, H. (2019). Estimation of uncertainty in flood forecasts—A comparison of methods. *Journal of Flood Risk Management, 12*.
- Emanuel, K. (2017). Assessing the present and future probability of Hurricane Harvey's rainfall. *Proceedings of the National Academy of Sciences, 114(48)*, pg. 12681-12684.
- Feyen, L., Dankers, R., Bodis, K., Salamon, P., & Barredo, J. (2012). Fluvial flood risk in Europe in present and future climates. *Climatic Change, 112(1)*.
- Hinkela, J., Linckea, D., Vafeidisb, A., Perrettec, M., Nichollsd, R., Tole, R., . . . Levermannc, A. (2014). Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proceedings of the National Academy of Sciences, 111(9)*, 3292-9297.
- IPCC, I. (2022). Special Report on the Ocean and Cryosphere in a Changing Climate.
- Kay, A., Crooks, S., Davies, H., & Reynard, N. (2011). An assessment of the vulnerability of Scotland’s river catchments and coasts to the impacts of climate change: Work Package 1 Report. NERC/Centre for Ecology & Hydrology.
- Liu, J., Guoyou, S., & Kaige, Z. (2019). High-Precision Combined Tidal Forecasting Model. *Algorithms*. doi:10.3390/A12030065
- Luettich, Jr, R., Westerink, J., & Scheffner, N. (1992). ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts, and Estuaries. Report 1. *Theory and Methodology of ADCIRC-2DDI and ADCIRC-3DL.*
- Met Office. (2023). UKCP18 Factsheet: Storms.
- Nicholls, R. J., Hinkel, J., Lincke, D., Suckall, N., & Tol, R. S. (2018). Integrated assessment of global environmental change with a focus on rivers and coasts. *Handbook of Global Environmental Politics*, 307-322. Edward Elgar Publishing.

- Pollard, J., Spencer, T., & Jude, S. (2018). Big Data Approaches for coastal flood risk assessment and emergency response. *Wiley Interdisciplinary Reviews: Climate Change, 9(5)*.
- Rosenzweig, B., McPhillips, L., Chang, H., Cheng, C., Welty, C., Matsler, M., . . . Davidson, C. (2018). Pluvial flood risk and opportunities for resilience. *5(6)*. Wiley Interdisciplinary Reviews: Water.
- Shahabi, A., & Tahvildari, N. (2024). A deep-learning model for rapid spatiotemporal prediction of coastal water levels. *Coastal Engineering, 190*. doi:https://doi.org/10.1016/j.coastaleng.2024.104504
- Stokes, K., Poate, T., Masselink, G., King, E., Saulter, A., & Ely, N. (2021). Forecasting coastal overtopping at engineered and naturally defended coastlines. *Coastal Engineering, 164*. doi:https://doi.org/10.1016/j.coastaleng.2020.103827.
- UNDRR, U. (2025). *Global Assessment Report on Disaster Risk Reduction.* United Nations. Retrieved from https://www.undrr.org/gar
- Wedawatta, G., Bingunath, I., & Proverbs, D. (2014). Small businesses and flood impacts: The case of the 2009 flood event in Cockermouth. *Journal of Flood Risk Management, 7(1)*.
- Winter, M., Shearer, B., Palmer, D., Peeling, D., Harmer, C., & Sharpe, J. (2016). The economic impact of landslides and floods on the road network. *Procedia Engineering(143)*, 1425-1434.

Chapter 2

Literature Review

The literature for this thesis consists of six sections which align with the topics introduced in Chapter 1. Firstly, there is an overview of the meteorological processes that drive hydrodynamic and hydrological hazards, providing essential context for understanding the nature and drivers of these hazards. Secondly, there is a description of the hazard, vulnerability and consequences of coastal inundation along with the modelling options available to describe them. Thirdly there is a description of the hazard, vulnerability and consequences of pluvial flooding along with the modelling options available. Fourthly there is a review of the relevant metamodeling options that are available for forecasting and modelling these hazards. The fifth section contains an overview of uncertainties and uncertainty propagation. Finally, there is a literature review conclusion that encapsulates the insights gained from the reviewed literature, identifies existing gaps, and underscores opportunities for fresh and innovative contributions to the field of study.

2-1. Meteorological Drivers

Hydrological and hydrodynamic hazards, along with their associated vulnerabilities and consequences, pose significant risks to communities and infrastructure, especially in the face of climate change and extreme weather events. Prior to a detailed description of the hydrodynamic and hydrological hazards that are relevant to this thesis, it is valuable to first have an understanding of the systems that drive them. This foundation underscores the selection of inputs used in the models described in Chapters 4 and Chapter 5.

2-1.1. Boundary Layer Meteorology

Hydrodynamic and hydrological hazards are driven by complex physical processes involving the transfer of energy, matter, and momentum within the Earth's atmosphere. The exchanges in energy, matter and momentum originate in and are dominated by the planetary boundary layer (PBL), the region between the lower atmosphere and the Earth’s surface (Stull, 1988). Dramatic short distance surface changes in surface are associated with dramatic short distance changes in the properties of the PBL (Rohli & Li, 2021). The lowest 10% of the PBL is known as the surface boundary layer (SBL) or Prandtl layer. This layer is directly influenced by the Earth's surface and plays a critical role in the exchange of energy, matter, and momentum between the surface and the atmosphere. Within the SBL, these exchanges remain constant with height at any specific moment, meaning that the rate of flow of energy, matter, and momentum does not vary with altitude. However, this does not imply temporal constancy, as the fluxes are highly dynamic, changing significantly from minute to minute or day to day due to shifts in surface conditions like temperature and wind.

Two key features of boundary layers, including the SBL, are worth noting. First, substantial exchanges of energy, matter, and momentum occur between the surface and the air above. Momentum exchanges, in particular, have a significant influence on atmospheric circulation, affecting wind velocity (Garratt, 1992). Second, although the rate of change in temperature, humidity, and fluid velocity remains consistent in the vertical direction within the SBL, sharp variations in these properties occur within the boundary layer. These variations take the form of both spatial and temporal fluctuations, meaning that the properties change not only with time but also across different locations within the layer. These fluctuations are driven by turbulence, surface roughness, and thermal conditions, all of which play a role in shaping local weather and climate near the surface (Rohli & Li, 2021).

Modelling atmospheric motion is conventionally done through equations governing momentum, energy and moisture. These are the Reynold's Equations of Motion, describing wind speed changes and accounting for advection, convection, pressure gradients, friction and the Coriolis effect; the Thermodynamic Energy equation describing the combined internal and kinetic energy of air molecules driving large scale atmospheric motion; the Moisture Conservation Equation which describes advection, condensation and evaporation; The continuity Equation describing changes in air density due to advection and velocity; the Equation of State for an Ideal Gas, which links pressure, density temperature and serving as a unifying constraint for the other equations (Holton & Hakim, 2012). Numerical weather models are usually parameterised by solar radiation, aerosol concentration, atmospheric absorption, cloud and surface absorption and reflection, condensation and evaporation, turbulence, sensible heat flux and rain as well as topography and surface roughness. These models have become increasingly sophisticated, leveraging high-resolution data and computational grids to simulate these dynamic interactions over vast areas.

2-1.2. Winds and Pressure Gradient

Wind is a key component of weather, acting as a transporter of heat, moisture, and other substances, as well as momentum. Wind arises from pressure differences in the atmosphere, which are ultimately caused by the unequal heating of the Earth's surface by the Sun. The atmosphere works as a dynamic system, redistributing energy from the equator, where there is an energy surplus, to the poles, which experience an energy deficit (Rohli & Li, 2021). Atmospheric circulation, and thus wind speed and direction, is governed by five forces, according to Newton's Second Law: the pressure gradient force, Coriolis effect, friction, centripetal acceleration, and gravity. These forces combine to shape wind patterns, contributing to the dynamic movement of air across the planet.

Winds are driven by the pressure gradient force, which, according to Newton's Second Law, causes air to move from areas of high pressure to areas of low pressure. These systems cause weather patterns to shift across space and time. Once winds form, they are deflected due to the Coriolis effect, a consequence of Earth's rotation. In the Northern Hemisphere, winds are deflected to the right, while in the Southern Hemisphere, they are deflected to the left. The strength of this deflection depends on both wind speed and latitude. Friction near the Earth's surface slows the winds, reducing the impact of the Coriolis effect. Additionally, centripetal acceleration plays a role in curved wind flows, as seen in cyclones, while gravitational acceleration affects vertical air movements. These forces together shape the speed, direction, and nature of wind flow at various scales.

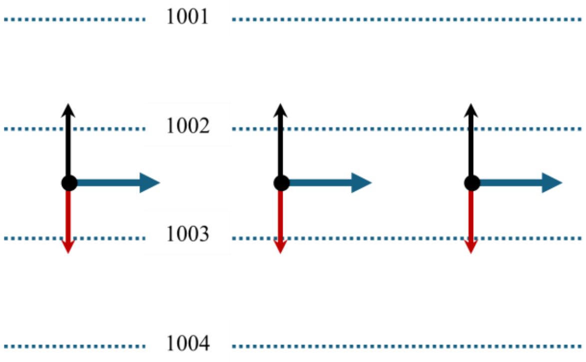


Figure 4: Balanced Coriolis force and pressure force causing Geostrophic flow with wind direction parallel to isobars. Pressure force (black arrows) tangential to isobars (dots) with units of millibar, wind direction (blue arrows) and Coriolis force (red arrows).

When the pressure gradient force and Coriolis effect are balanced Figure 4, the resulting air flow moves parallel to the isobars (lines of constant pressure) at a steady speed, rather than directly from high to low pressure. This is known as Geostrophic flow (Garratt, 1992). This type of flow is most commonly observed in the upper atmosphere or open oceans, where friction from the Earth's surface is minimal.

Near the surface, friction slows the wind, disrupting the balance between the pressure gradient and Coriolis forces. Geostrophic flow explains the large-scale movements of winds like the jet stream and also governs ocean currents.

When the pressure gradient force and Coriolis effect are imbalanced, a curvature is observed in wind flow that is explained by centripetal acceleration (Rohli & Li, 2021). When this centripetal acceleration is less than the pressure gradient force this creates cyclonic flow Figure 5, when the centripetal acceleration is greater than the pressure gradient force this creates anti-cyclonic flow, Figure 6.

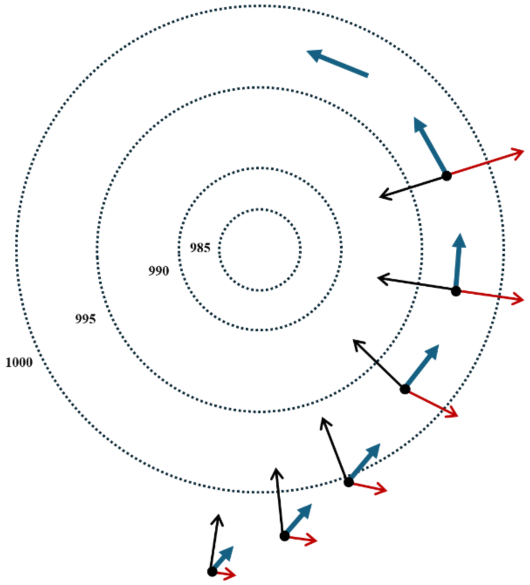


Figure 5: Cyclonic flow around a low pressure centre in the northern hemisphere showing pressure force (black arrows) tangential to isobars (dots) with units of millibar, wind direction (blue arrows) and Coriolis force (red arrows).

In the Northern Hemisphere, winds around anticyclones and ridges (areas of high pressure) follow a clockwise flow. This flow is classified as supergeostrophic when wind speeds exceed those of geostrophic flow. This happens because the Coriolis force must be stronger than the pressure gradient force to maintain the curved, clockwise path, which requires faster winds. In contrast, winds around cyclones and troughs (areas of low pressure) move counterclockwise, and if the pressure gradient force is unchanged, the flow becomes subgeostrophic — slower than geostrophic winds. Here, the Coriolis force must be weaker than the pressure gradient force to sustain the counterclockwise curve, which is achieved by slowing down the wind.

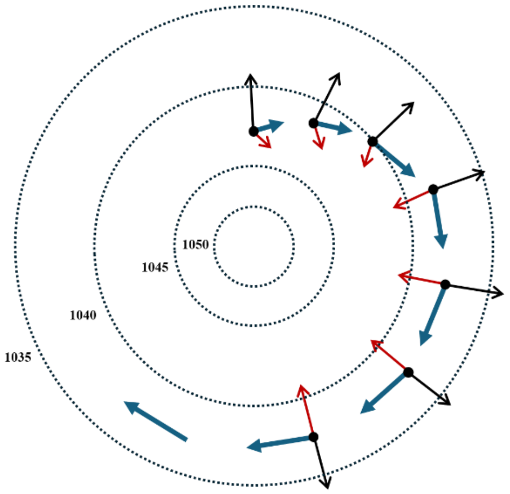


Figure 6: Anti-cyclonic flow around a high pressure centre in the northern hemisphere showing pressure force (black arrows) tangential to isobars (dots) with units of millibar, wind direction (blue arrows) and Coriolis force (red arrows).

For cyclones, the forcing—defined as the rate of generation or dissipation of kinetic energy—increases markedly with intensity, from about 2.5 cm^2/s^3 in weaker systems to roughly 12 cm^2/s^3 in stronger ones, with maximum values reaching 16 cm^2/s^3. Correspondingly, the median kinetic energy over their lifetime ranges from 10^{18} to 10^{19} J (Akperov, Golitsyn, & Semenov, 2024). Cyclones tend to reach their maximum intensity and wind speed about 30–36 hours after formation, and as their depth increases, forcing, kinetic energy, size, and wind speed all rise, though wind speed growth slows for the most intense storms. Anticyclones follow a similar general trend, but the increases are more modest: forcing values lie between 1 and 3 cm^2/s^3 (maximum about 3.2 cm^2/s^3) and kinetic energy is of the same order of magnitude as in cyclones. However, for anticyclones, size and wind speed growth tend to plateau earlier, and beyond certain depths these quantities may even decline. Anticyclones also show a less symmetric life cycle, with maxima often reached on the second or third day after formation rather than in a sharp early peak.

Rather than isolated, circular systems, real atmospheric flow in the upper troposphere is shaped by large-scale Rossby waves — undulating, west-to-east meanders in the jet stream, Figure 7. Ridges (high-pressure areas) form equatorward of wave crests, where warm tropical air rises, increasing the height of pressure surfaces. Troughs (low-pressure areas) form poleward of wave troughs, where cold polar air sinks, lowering the height of pressure surfaces. These undulating Rossby waves play a crucial role in weather patterns and energy redistribution between the tropics and poles.

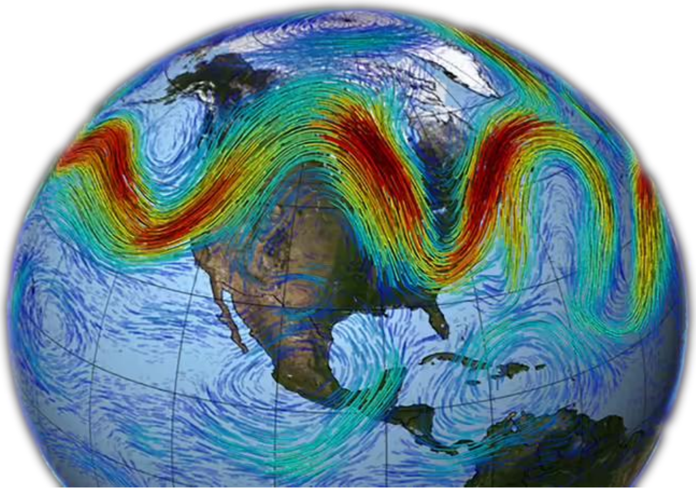


Figure 7: Interaction of jet stream and Rossby wave. Credit: NASA/GSFC

2-1.3. Sea Atmosphere Interactions

The ocean’s large heat capacity lets it store and mix vast amounts of thermal energy. Its surface mixed layer—analogous to the atmospheric PBL—contains over 30× the heat of the atmosphere, so ocean–atmosphere coupling strongly shapes climate. The ocean absorbs shortwave solar radiation and emits longwave radiation, setting the global energy balance, while sea-surface temperature (SST) modulates atmospheric stability: warm SSTs favour unstable conditions and vigorous upward turbulent fluxes of sensible and latent heat; stabilized boundary layers suppress vertical mixing and favour horizontal transfers (Rohli & Li, 2021). Wind stress acting on the mixed layer drives Ekman currents—a spiral-shaped shear flow in which velocity weakens and turns with depth—and produces a net Ekman transport 90° to the right of the wind in the Northern Hemisphere (left in the Southern), largely independent of the details of vertical mixing (Price, Weller, & Schudlich, 1987). This wind–Coriolis balance links the atmospheric wind field to upper-ocean convergence/divergence, thereby organizing upwelling and downwelling that shape warm/cool pools and, in turn, SST and climate variability. Observations show this wind-driven transport can be strongly surface-trapped within the upper few tens of meters, emphasizing how mixed-layer depth and stratification control air–sea exchange and the efficiency of wind forcing.

A key part of ocean circulation and climate variability is the role of oceanic waves, specifically Rossby waves and Kelvin waves. Rossby waves are large, slow-moving waves that propagate westward across ocean basins due to the combined effects of the Coriolis force and changes in Earth's vorticity with latitude (known as the β-effect). Unlike atmospheric Rossby waves, which move relatively quickly, oceanic Rossby waves are much slower, often taking months to years to cross an ocean basin. These waves impact sea surface height and the depth of the thermocline, which influences the distribution of heat in the ocean. This, in turn, affects climate phenomena such as El Niño-Southern Oscillation (ENSO), where shifts in the thermocline depth play a crucial role in ocean-atmosphere feedbacks (Kim & Kim, 2002).

Kelvin waves, on the other hand, are fast-moving waves that travel along the equator or along coastal boundaries (Wang B. , 2002). Unlike Rossby waves, which propagate westward, Kelvin waves move eastward with speeds much faster than Rossby waves. They are often triggered by changes in wind stress, particularly during ENSO events. For example, during an El Niño event, weakened trade winds allow downwelling Kelvin waves to propagate eastward, deepening the thermocline in the eastern Pacific and causing surface warming. This warming alters atmospheric convection and wind patterns, creating feedback loops that reinforce El Niño conditions. Kelvin waves can also travel along coastlines, influencing coastal upwelling and downwelling processes.

The Walker Circulation is a key component of tropical ocean–atmosphere interaction, driving the trade winds that help generate and modulate both Rossby and Kelvin waves. In its normal state, strong easterly trade winds push warm surface waters westward across the equatorial Pacific, causing upwelling in the east and rising motion with deep convection in the west, balanced by eastward return flow aloft. Changes in the Walker Circulation, such as its weakening during El Niño events, alter wind stress patterns, triggering Kelvin waves and modifying the propagation of Rossby waves, thereby influencing thermocline depth and sea surface temperature distribution (Kim & Kim, 2002).

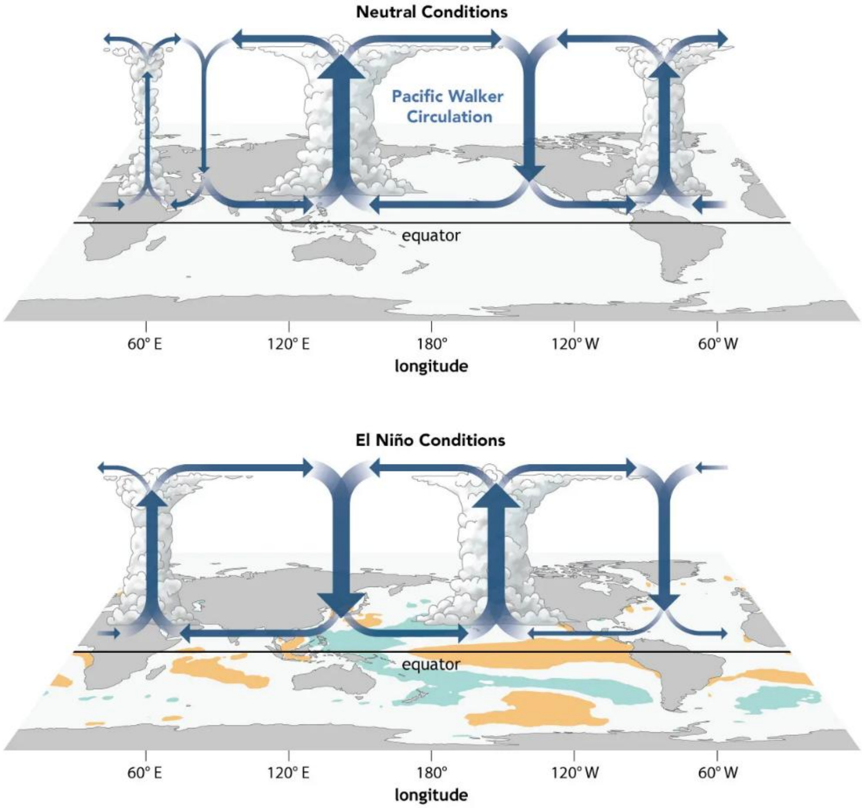


Figure 8: Changes in atmospheric circulation over the equator with ENSO. The change in heat distribution in the ocean is shown with increases in orange and decreases in blue. Illustration by NOAA

Together, Rossby waves and Kelvin waves act as key drivers of ocean-atmosphere interactions, with Rossby waves contributing to the slow redistribution of heat and Kelvin waves enabling rapid adjustments to wind stress and pressure anomalies. These waves influence the depth of the thermocline, regulate oceanic heat transport, and contribute to the evolution of major climate phenomena like ENSO. The dynamic interaction of these waves with the atmosphere highlights the ocean's critical role in regulating global weather, climate, and energy balance.

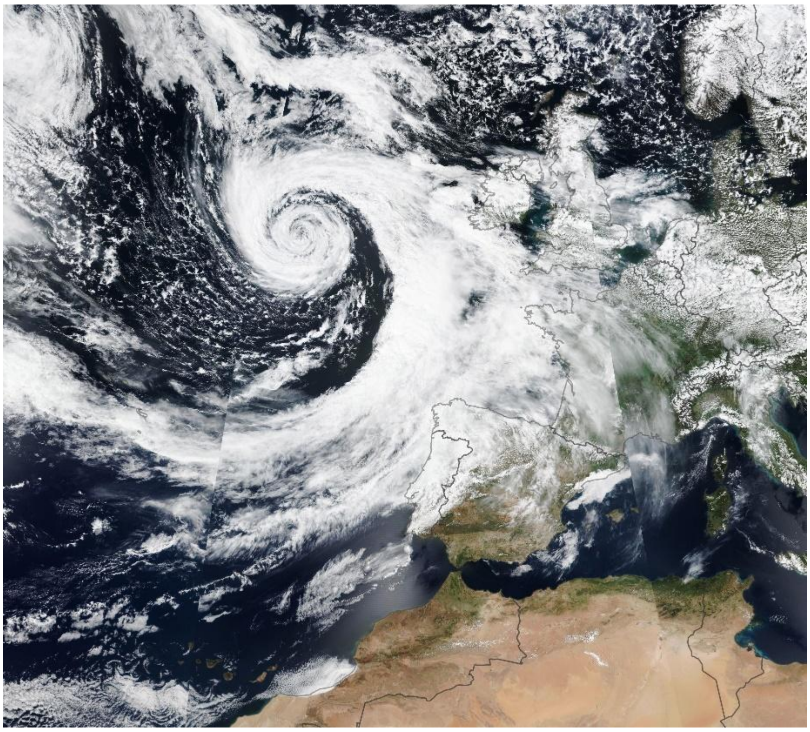


Figure 9: A large cyclonic storm front approaching the United Kingdom on May 7, 2019. The image was acquired by the Visible Infrared Imaging Radiometer Suite (VIIRS) on the Suomi NPP satellite. NASA Earth Observatory by Kasha Patel

2-1.4. Precipitation

Under unstable conditions, rising air cools until it reaches the lifting condensation level, where the temperature matches the dew point or frost point, causing the air to become saturated with a relative humidity of 100%. Further cooling results in condensation, freezing, or deposition of water vapor into liquid or ice, forming clouds. These processes rely on condensation and freezing nuclei, tiny solid particles in the air that provide surfaces for water molecules to attach. Concentrations of these nuclei are higher levels over land than oceans (Rohli & Li, 2021). Coastal zones exhibit sharp gradients in aerosol concentrations, leading to variations in cloud cover, pollution, and airborne particles like pollen.

After a cloud has formed and grows, it has the potential to produce precipitation, which occurs through two primary mechanisms. The first is the collision-coalescence process, which dominates in clouds with temperatures above freezing. In this process, larger water droplets fall faster than smaller ones due to differences in terminal velocity. As larger droplets fall, they collide and merge with smaller droplets, eventually growing large enough to overcome the cloud's buoyant forces and fall as rain. This process, also known as the "warm cloud model", is relatively inefficient, as droplets often avoid collision.

The second mechanism is the Wegener-Bergeron-Findeisen process, also called the "three-phase process", which occurs in clouds containing ice, liquid water, and water vapor (Rohli & Li, 2021). This process is most active in clouds where temperatures are below freezing, such as cirrus, altostratus, and cumulonimbus clouds. Although water droplets can remain in liquid form at subfreezing temperatures (as supercooled droplets), the key difference in this process is the lower saturation vapor pressure near ice compared to liquid water at the same temperature. This difference causes water vapor to deposit directly onto ice crystals, growing them into larger snowflakes. As vapor deposits on the ice, nearby supercooled droplets evaporate, replenishing the water vapor. This cycle of evaporation and deposition continues until the ice crystals become large enough to fall as snow. If the air beneath the cloud is warm enough, the snowflakes may melt into rain, or if conditions fluctuate, they may refreeze as sleet or freezing rain.

Global precipitation patterns are shaped by large-scale circulation systems, ocean–atmosphere interactions, and land–sea contrasts. Climatologically, the highest annual precipitation totals occur near the Intertropical Convergence Zone (ITCZ), monsoon regions, and midlatitude storm tracks, while subtropical high-pressure belts are comparatively dry (Lee & Biasutti, 2014). Seasonally, tropical rainfall shifts north and south with the solar cycle, and midlatitude precipitation maxima align with winter storm track activity.

Regional variability is strongly influenced by phenomena such as ENSO, which affects not only mean rainfall but also extremes. A global analysis of over 7,000 stations found that ENSO impacts are often asymmetric, with some regions responding strongly to only one ENSO phase (Sun, Renard, Thyer, Westra, & Lang, 2015). For example, El Niño tends to enhance boreal winter precipitation in southwestern North America, while La Niña can bring wetter conditions to South Africa in summer. Such variability affects design-relevant extremes, such as 10- or 100-year return period rainfall events, and must be considered in infrastructure planning.

Precipitation variability is also projected to increase in a warming climate. Model ensembles show that seasonal-mean precipitation variability rises by 4–5% K^{-1} over land, with the largest relative increases in the ITCZ, monsoon regions, and high latitudes (Pendergrass, Knutti, Lehner, Deser, & Sanderson, 2017). This increase links the risk of extreme wet events with drought risk, as heightened variability expands the range of possible outcomes.

Extreme precipitation events—defined statistically (e.g., exceeding the 99th percentile) or by fixed thresholds—pose major societal risks through flooding, landslides, and infrastructure damage. Their impacts are amplified in densely populated and urbanized regions, and they can trigger cascading hazards when combined with other factors such as storm surge (Gimeno, et al., 2022).

Observed trends show increases in the intensity of the most extreme rainfall in many regions, even where average precipitation has declined. These changes are consistent with the Clausius–Clapeyron relationship, whereby atmospheric moisture-holding capacity increases by ~7% K^{-1}. However, actual changes also depend on circulation shifts and moisture transport mechanisms. Moisture convergence through atmospheric rivers, monsoonal flows, and tropical cyclones plays a critical role in triggering extremes.

Future projections indicate that extreme precipitation will intensify in most wet regions, particularly in high-latitude and tropical areas with abundant moisture. ENSO events are likely to continue to modulate the timing and severity of such extremes, with some regions experiencing disproportionate impacts from one ENSO phase (Sun, Renard, Thyer, Westra, & Lang, 2015). Compounding effects—such as simultaneous coastal flooding from heavy rain and storm surge—are expected to become more common, further complicating risk management.

2-2. Coastal Overtopping

Overtopping hazard refers to the likelihood of water exceeding the height of a flood protection structure, potentially causing loss of life or damage to surrounding areas. The projected rise in overtopping by up to fifty times by the century's end suggests a dramatic increase in the frequency and severity of these events, necessitating significant investments in coastal defences and emergency response capabilities (Almar, et al., 2021).

Figure 10: Saltcoats, UK. 11th Nov, 2022. Winds with speeds up to 60 mph caused high tides and sea waves of 10 metres and more lashed the coast at Saltcoats, Ayrshire, Scotland, UK causing Scotrail to cancel trains on the coast line after 1.00pm. Credit: Findlay/Alamy

2-2.1. Coastal Overtopping Hazard and Modelling

The factors that affect the volume of overtopping Q are hydrodynamic and structural. The hydrodynamic elements are presented in Figure 11. The maximum height that the water reaches Hmax is the still water level (SWL) depth d, and the wave runup Hu. The still water level comprises the astronomic tide Ha and the non-tidal residual Hr and so the maximum water level can be expressed as:

$$\boldsymbol{H_{max} = H_a + H_r + H_u} \qquad \text{Eq. [2]}$$

The structural factors are structure specific, e.g. breakwaters (Franco, De Gerloni, & Van der Meer, 1995) and seawalls (Allsop, 2005) and specific consideration for each structure is outside the scope of this thesis. As such this section will deal with general case hydrodynamic hazard prediction. The linear, additive equation itself is structured to provide a straightforward, additive estimation of the highest potential water level by summing up the individual components and doesn't inherently restrict the components from influencing each other in real-world dynamics. Modelling the processes that lead to overtopping, such as wave breaking, wave run-up, and interaction with coastal defences, requires sophisticated numerical approaches.

These processes are highly sensitive to changes in parameterisation, making calibration and validation of models particularly challenging. Additionally, the variability in coastal defence structures further complicates predictions of overtopping and subsequent inland flooding.

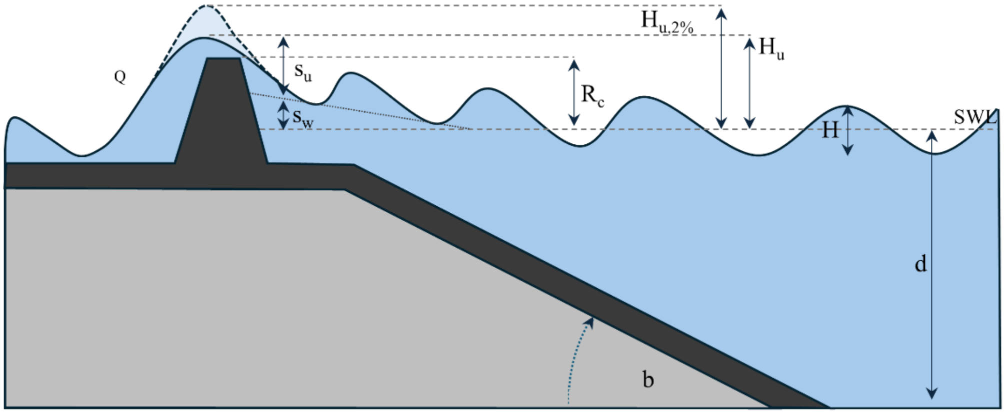


Figure 11: Schematic showing constituent elements of overtopping mechanism

To effectively address these complexities, innovative tools such as the SWEEP-OWWL model have been introduced. Developed by Stokes et al (2021) , this advanced overtopping forecasting system is specifically designed for the southwest coast of England and employs a high-resolution Delft 3D hydrodynamic model at a 1 km scale, driven by Met Office forecasts. The model outputs are processed through nearshore hydrodynamic equations and adapted using location-specific coastal profiles and defence characteristics. Demonstrating strong performance, the model achieved a 100% success rate in predicting extreme overtopping events during winter 2018 and an 80% overall accuracy (Stokes, et al., 2021). The hydrodynamic model runs in 2.5 hours using 8 cores with parallel computation, while the overtopping calculations take 16 minutes on a single core. The authors identify several sources of potential error, including the initial Delft-3D model, Met Office forecasts, run-up and overtopping equations, and beach profiles and geometries. However, due to the complexity of isolating the contributions of each error source, the individual impact of each aspect on the total error has not been assessed. As a result, robust confidence bounds for each forecasted runup and overtopping prediction could not be established.

Stockdon et al. (2023) presents a total water level and coastal change modelling framework that explores the complexities of nearshore uncertainty caused by morphological changes. This framework relies on wave characteristics from the Nearshore Wave Prediction System (van der Westhuysen, et al., 2013) and surge forecasts from the Extratropical Surge and Tide Operational Forecast System (ESTOFS) (Funakoshi, Feyen, Aikman, Van der Westhuysen, & Tolman, 2013) and the National Hurricane Center’s Probabilistic Surge (P-Surge) model (Taylor & Glahn, 2008). The authors assert that in their model, the uncertainties of wave and water level components are well-understood due to ongoing near-real-time validation efforts conducted for the U.S. East Coast (Stockdon, et al., 2023). However, this claim should be viewed critically, as it may not hold true for regions lacking similarly extensive and robust modelling systems. Each of the constituent terms in Eq. [2]**Error! Reference source not found.** will be described in the following sub sections, along with an overview of their modelling options and their associated uncertainties.

2-2.1.1. Astronomic Tide

Astronomic tides are the periodic rise and fall of sea levels caused primarily by the gravitational interactions between Earth, the Moon, and the Sun. The Moon's gravitational pull generates a bulge in Earth's oceans on the side facing the Moon, while a second bulge forms on the opposite side due to the centrifugal force from Earth's rotation around the Earth-Moon barycentre (Parker, 2007). Simultaneously, the Sun's gravitational influence contributes to tidal effects, either enhancing them during spring tides—when the Sun, Moon, and Earth align—or reducing them during neap tides, when the Sun and Moon are at right angles relative to Earth. These tidal movements occur predictably based on the celestial positions of the Moon and Sun, with variations influenced by factors like Earth's rotation, the Moon's orbital eccentricity, and local coastal topography.

Astronomic tide is most often predicted using harmonic analysis. Harmonic tide prediction is based on the observation that the tide at any given time can be represented as the sum of several sinusoidal components, each with its own amplitude, frequency, and phase. These components, known as tidal constituents, correspond to different astronomical cycles (Parker, 2007). For example, the M2 constituent, which has a period of approximately 12.42 hours, represents the principal lunar semidiurnal tide, while the S2 constituent represents the solar semidiurnal tide with a period of 12 hours. By performing harmonic analysis on historical tide data, the characteristics of each tidal constituent can be determined. The predictive model is created by summing the contributions of all significant tidal constituents. Because each constituent corresponds to a specific astronomical phenomenon, their combined effects produce the observed tide at any location. The accuracy of this prediction depends on the number and accuracy of the constituents included in the model. Generally, more constituents lead to a more precise prediction, as the model better captures the complexity of the tidal behaviour. Harmonic models were traditionally fit with Fourier analysis, which requires complete, regularly spaced timeseries. More recently harmonic analysis is applied using least squares fitting (Zhang, Yin, Wang, & Hu, 2017).

Nonharmonic models, such as time and height difference techniques, monthly mean analysis or mean cycle analysis are used for tidal predictions when harmonic analysis is not practical (Parker, 2007). While these models are less accurate than harmonic analysis, they require less data, making them more suitable for certain stations.

Errors in tide predictions often stem from the limitations of traditional harmonic analysis methods, particularly when dealing with incomplete or irregular data sets (Parker, 2007). Gaps or irregularities in the data, shorter time series, and misalignment of synodic periods can all reduce the accuracy of traditional harmonic predictions. Uncertainty analysis for tidal predictions can be done through Kalman filter or fuzzy logic forecasting (Zhang, Yin, Wang, & Hu, 2017). Kalman filtering is an optimal recursive estimation technique that uses indirect, inaccurate and uncertain observations to estimate the uncertainty of a given parameter and is applicable for short term tidal (Abubakar, Mahmud, Tang, Hussaini, & Md Yusuf, 2019).

2-2.1.2. Non-tidal Residual

Non-tidal residual refers to the component of coastal water movement that remains after subtracting the effects of tidal forces. It describes anomalous seal level changes combining atmospheric effects and the influence of tide surge interactions (Fernández-Montblanc, et al., 2019). Storm surges are abnormal changes in the non-tidal residual caused by intense storms. The behaviour of storms is influenced by oceanic Rossby and Kelvin waves. Rossby waves are the large-scale dynamical response of the ocean to atmospheric conditions (Chelton & Schlax, 1996). Their behaviour is responsible for pressure and windspeed variations that can drive storm surges, particularly in mid-latitudes. Kelvin waves, on the other hand, are a type of gravity wave that can rapidly propagate sea level changes along coastlines, contributing to the intensity and reach of storm surges (Wang B. , Kelvin Waves, 2002).

The field of storm surge modelling and forecasting has benefitted greatly from the development of high-fidelity numerical models providing detailed description of hydrodynamic processes to generate accurate and reliable storm surge responses, see e.g. (Taflanidis, Jia, Kennedy, & Smith, 2013). For surge forecasting, an extensive review was conducted by Al Kajbaf and Bensi (2020), grouping existing hydrodynamic modelling types as either high fidelity physics-based models, often with a substantial computational cost, or low-fidelity models with an associated drop in accuracy but with computational costs that are better suited to forecasting applications. Common high fidelity models include FVCOM (Chen, 2003) and ADCIRC (Luettich, Westerink, & Scheffner, 1992). Significant attention has been given in recent years to landfall location, where a large error in the forecasted landfall location can lead to a large error in surge height prediction (Kohno, et al., 2018). Numerical weather prediction models have become more popular in recent years for use as surge forecast inputs with the draw back that storm intensity is often underpredicted for storms that are either too weak or very strong (Kohno, et al., 2018). Storm surges are characterised by high variability and contribute the largest source of uncertainty in predicting sea level elevation (Mel, Viero, Carniello, Defina, & D'Alpaos, 2014).

Some studies have explored the nature of the uncertainty associated with different surge prediction methods such as surge response functions (Taylor, Irish, Udoh, Bilskie, & Hagen, 2015), with particular attention given to the lead time of the surge height predictions (Mel & Lionello, 2016). Multi-scenario storm surge forecasts are becoming increasingly popular, considering the variability in various parameters such as size, location of landfall and intensity (Kohno, et al., 2018).

The overall uncertainty in storm surge prediction is multi-faceted. From a modelling perspective, errors arise from simplifying assumptions, parameterisation, local topographic and bathymetric effects, and initial conditions. Additionally, uncertainties surrounding storm characteristics such as intensity, track, and speed of a storms which come from weather models carry their own uncertainties that propagate to the final surge height estimate. Similarly, variability in atmospheric pressure and wind speed also contribute to uncertainty in forecasting the meteorological conditions that drive storm surges. In a study into the interrelationship between the effects of combined random errors and bias in numerical weather prediction models and in surge models, Resio et al. (Resio, Powell, Cialone, Das, & Westerink, 2017) found that surge model bias can play a dominant role in distorting forecast probabilities.

2-2.1.3. Run-up

Wave run-up refers to the phenomenon where waves travel up the shore, beyond the usual high-water line, driven by factors like wave height, tidal conditions, and coastal topography. It is the distance or height that waves reach when they move up the beach or coastal surface, often influenced by the type of waves, the slope of the beach, and the presence of obstacles like dunes or seawalls. Run up can be considered as the sum of swash *Sw* and setup *Su*. Figure 11. Swash refers to the movement of water onto the beach after a wave breaks, involving an uprush as water moves up the beach and a backwash as it flows back toward the ocean. This process plays a crucial role in sediment transport along the shore. Setup, on the other hand, is the temporary elevation of the mean water level at the coast due to the impact of storm waves. This occurs when breaking waves push water shoreward, raising the local sea level above the normal conditions and can lead to coastal flooding, especially during severe weather events. The setup component is primarily generated by the gradient in radiation stresses—depth-integrated, phase-averaged momentum fluxes within the wave field—that occur as waves shoal and break in the surf zone. As waves dissipate energy through breaking, the reduction in radiation stress is balanced by an increase in mean water level at the shore, driving setup. Recent nonlinear analytical solutions for higher-order Stokes waves (Gao, et al., 2023) show that wave steepness and nonlinearity significantly influence the magnitude and spatial distribution of radiation stresses, thereby modifying setup potential. This process is especially significant during storm conditions, where high, long-period waves create large cross-shore radiation stress gradients, elevating coastal water levels and exacerbating overtopping and flooding hazards. Wave breaking can substantially influence the overtopping process by dissipating wave energy and altering run-up characteristics. The type of breaking—spilling, plunging, or surging—affects the spatial distribution and magnitude of overtopping discharges, and therefore should be considered in hazard assessments (Altomare & Gironella, 2024). Overtopping volumes also exhibit strong correlations with incident wave type, wavelength, and nearshore transformation processes such as shoaling and wave reflection. These processes influence wave height at the structure toe and modify the effective momentum transfer to the defence, thereby changing overtopping rates and inland flooding potential.

Run up is a critical parameter in coastal engineering as it determines the extent of wave action inland and the potential for flooding, erosion, or overtopping of structures. Basic runup models such as the Stockdon model (Stockdon, Holman, Howd, & Sallenger Jr, 2006) consider wave height, wave period and beach slope, more advanced models may account for the nonlinear interactions between waves and the beach, considering the effects of wave breaking, dissipation, and swash oscillations (da Silva, Coco, Garnier, & Klein, 2020). These models require more complex parameters, such as the wave energy dissipation rate, water depth variations, and detailed beach profiles, which can better represent the range of real-world conditions that influence runup but also introduce additional complexity in both the formulation and the application of the models. For a thorough review of runup models see da Silva et al (2020). Typically, runup formulas for irregular waves are expressed in terms of statistical measures, such as the runup exceeded by 2% of the waves $\boldsymbol{H_{runup,2}}$.

To avoid implying simple linear additivity (e.g. Eq. [2]), it is noteworthy that wave–coast interactions are intrinsically nonlinear and can shift both the mean and the tails of the total-water-level distribution. Laboratory and theory for long, irregular waves on a plane beach show that nonlinearity—quantified by a breaking parameter that scales shoreline-velocity variance to run-up variance—leaves the shoreline-velocity statistics essentially unchanged but alters the run-up displacement statistics (Denissenko, Didenkulova, Pelinovsky, & Pearson, 2011). Extremes follow a Rayleigh-type law for narrowband forcing, the displacement distribution becomes non-Gaussian, and positive tails are enhanced while negative tails are suppressed. Crucially, the mean sea level at the shoreline rises with increasing nonlinearity (wave set-up), so any given tide or surge level rides on a higher local baseline than linear superposition would suggest, increasing exceedance probabilities and prolonging inundation when defences are overtopped. These effects are consistent with nonlinear shallow-water theory (which predicts coincidence of extreme run-up measures in linear vs. nonlinear treatments under far-field forcing) and are confirmed experimentally; the skewness and kurtosis of run-up mirror those of the incident field and can be used diagnostically for flood risk. Operationally, this argues for total-water-level frameworks that embed state-dependent set-up and swash statistics.

Typical sources of uncertainty for wave runup arise from uncertainty in the deep water wave characteristics, simplifying assumptions in modelling equations and changes in morphology and errors in bathymetry (Rutten, Torres-Freyermuth, & Puleo, 2021). The inherent randomness of waves introduces intrinsic uncertainty, leading to significant variability in runup predictions across dissipative, intermediate, and reflective beach profiles (Rutten, Torres-Freyermuth, & Puleo, 2021). In their total water level and coastal change modelling framework, Stockdon et al. (2023) posited that static representation of coastal morphology is the largest source of uncertainty in empirical runup parameterisation (Stockdon, et al., 2023).

2-2.1.4. Wind-waves

Since deep water wave characteristics are a major source of uncertainty in wave runup models, it is essential to account for both the modelling of deep water wind waves and their associated uncertainty.

Surface waves are waves that travel along the interface between two different mediums, typically between water and air, and are primarily generated by the wind's friction on the ocean's surface. As wind blows over the water, it transfers energy to the surface, creating ripples that grow into larger waves as the wind continues to exert force. The size and strength of surface waves are influenced by factors such as wind speed, duration, and the distance over which the wind blows, known as fetch.

As waves approach the shore, their behaviour changes due to the interaction with the seafloor and the effects of decreasing water depth (USACE, 2003). As wave enter shallow water they begin shoaling, where waves slow down as the water depth decreases. This reduction in speed leads to a decrease in wavelength while the wave height increases to conserve energy. Additionally, refraction occurs because different parts of a wave encounter varying depths, causing the wavefront to bend. When waves steepen to the point where their height exceeds the wave break index γ, they become unstable and break. This value is determined empirically and is site specific. $\gamma = 0.78$ is commonly used as an initial estimate for many engineering practises (USACE, 2003). Multiple studies find breaker indices to be a function of beach slope and incident wave steepness. A comprehensive list of these studies can be found in Part 2 Section 4-II of the Coastal Engineering manual (USACE, 2003). Overtopping can compromise the effectiveness of these structures and lead to significant social, economic, and environmental consequences.

Similar to surge forecasting, surface-wave modelling has benefitted greatly from the development of high-fidelity numerical models and from improvements in computational efficiency. Many open-source and commercial numerical tools are available for simulating waves (e.g. see Thomas and Dwarakish (2015) for a detailed review), with the most common being SWAN (Booij, Holthuijsen, & Ris, 1996) and WAM (NOAA, 2019) for nearshore applications and WAVEWATCH III for deep water (Tolman, 2009). State of the art physics-based numerical wave models are defined by multiple physical formulations and multi component input data, typically concerning the bathymetry, meteorological conditions (e.g. wind speed and direction) and characteristic ocean parameters (e.g. sea temperature, tides and currents).

Calibrating surface wave models is a challenging task, an identifying the model parameters that provide accurate simulation of storm and wave conditions is conventionally difficult given the number of source terms, the various formulations and their respective parameters, and the spatial variability of the significance of these elements around a complex coastline. No established and broadly agreed methodology has been established yet for model fitting and validation, and it is common in wave modelling literature to establish the most sensitive parameters and then use a trial-and-error method to identify the best combination of tuneable parameters (Venugopal, 2015).

For forecasting, the quality of wind fields that drive wave models significantly affects wave predictions (Bitner-Gregersen, et al., 2022). Differences in wind input, for example, from ECMWF's ERA-5 (C3S C. C., 2017), can lead to notable discrepancies compared with higher resolution local wind products of 0.5-1m in Hs during storm events (Wang, Yang, Wu, & Grear, 2018). Additional significant sources of uncertainty arise from modelling wave-current interactions as well as the spatial and temporal resolution of wave models. Wave models can be based on structured and unstructured meshes, with unstructured, finer grid models being more suited to coastal regions due to their improved ability to capture the effects of complex bathymetry. Unstructured grids also provide more realistic description of flow patterns and are better suited to represent the propagation of tidal and surface waves (Chen, 2003). Assimilating satellite altimeter data into wave models can help correct inaccuracies in forecasts (Bitner-Gregersen, et al., 2022). Data assimilation is particularly useful for operational forecasting but introduces uncertainty due to the limitations of satellite return times and the spatial sparsity of observations. For a comprehensive review of wave modelling uncertainties see Bitner-Gregersen et asl. (2022).

2-2.2. Coastal Overtopping Vulnerability and Modelling

Overtopping vulnerabilities depend on structural characteristics (e.g., crest height, slope, and structure permeability) and the constituent parts of maximum observable water level, *Hmax* (Koosheh, Etemad-Shahidi, Cartwright, Tomlinson, & van Gent, 2021). A major vulnerability of coastal defences to overtopping is their design limitations with most coastal defences built with specific design parameters that have been derived from historical data. However, with the growing impacts of climate change, such as sea-level rise and the increased frequency and intensity of extreme weather events, the original design thresholds may be exceeded more frequently (Ferreira, Cardona, Jóia Santos, & Tenedório, 2021).

Another vulnerability lies in the physical structure and materials of coastal defences. Older structures may suffer from material degradation due to weathering, saltwater corrosion, or structural fatigue, making them more susceptible to overtopping. Over time, even small overtopping events can weaken the foundation or lead to scouring at the base of seawalls, undermining the structure's stability. In areas with frequent wave action, the cumulative effect of overtopping can result in cracks, fractures, or collapse of the defence system, leaving coastal areas exposed to flooding and erosion. Additionally, the increased frequency intensity of storm winds and extreme waves will accelerate the rate of damage to coastal defence structures (Ferreira, Cardona, Jóia Santos, & Tenedório, 2021). Vulnerability modelling must account for extreme weather events, which are expected to increase with climate change. Using scenario-based approaches, researchers can evaluate how storm surge frequency, wave height, and tidal action interact to exacerbate overtopping vulnerability. Methods like Monte Carlo simulations and stochastic models allow for probabilistic estimates of extreme event impacts (Koosheh, Etemad-Shahidi, Cartwright, Tomlinson, & van Gent, 2021).

Vulnerability of individual structures can be assessed through regular inspection of coastal defences, especially in the event of an extreme event. This might involve a combination of visual inspections, structural testing, hydrodynamic modelling, and vulnerability analysis. This process ensures the defence system remains functional under normal and extreme conditions.

Multiple indexes are used to quantify vulnerability, such as the Coastal Vulnerability Index (CVI) which considers environmental and physical factors like elevation, distance from flooded areas, and geomorphological features (Hamid, et al., 2019). Vulnerability can be assessed on a regional or global scale by tools like the Dynamic and Interactive Vulnerability Assessment (DIVA) (Hinkel & Klein, DINAS-COAST: developing a method and a tool for dynamic and interactive vulnerability assessment, 2003). DIVA integrates coastal databases, social and natural subsystem knowledge, and a graphical interface for scenario testing. It simulates the impacts of sea-level rise and storm surges on natural and human systems, supporting decision-making on coastal adaptation measures.

2-2.3. Coastal Overtopping Consequences and Modelling

The consequences of overtopping can be severe and far-reaching, with increasingly severe consequences on coastal communities and economies, often resulting in substantial financial burdens. One immediate impact is the flooding of low-lying coastal areas, where overtopped water can flood residential, commercial, and industrial areas, leading to property damage, displacement of communities, and loss of livelihoods, for example in Costa da Caparica, Portugal, about 30% of the region is at moderate to very high risk of experiencing significant overtopping and flooding events (Ferreira, Cardona, Jóia Santos, & Tenedório, 2021). Flooding of agricultural land can also reduce crop yields and result in economic losses for farming communities. Overtopping can overwhelm drainage systems and contaminate freshwater supplies with saltwater, compromising local water resources.

The socio-economic impacts include substantial property damage and the disruption of economic activities (Bouma, François, Schram, & Verbeke, 2009). These disruptions can particularly affect tourism-dependent locales, where the aesthetic and utility damage from overtopping can deter visitors, cutting deeply into local revenues. Furthermore, the direct impact on infrastructure—such as roads, bridges, and utilities—can be extensive, requiring millions, if not billions, of dollars in repairs and upgrades to withstand future events. Moreover, the costs associated with managing these risks are not just limited to immediate repairs and recovery. Long-term economic impacts include decreased property values in vulnerable areas, increased insurance premiums, and the potential relocation of populations and businesses to safer areas. These shifts can lead to a decreased tax base for local governments, which, coupled with the increased need for emergency services, risk management, and infrastructure reinforcement, creates a challenging economic cycle that can be difficult to break without substantial investment and planning.

Environmental consequences are equally significant. When overtopping occurs, it can transport large amounts of sediment, pollutants, and debris inland, affecting local ecosystems and water quality. Wetlands, which act as natural buffers against flooding, may be inundated with saltwater, leading to habitat loss and affecting local biodiversity. Coastal erosion can accelerate in areas where overtopping causes scouring and structural damage, leading to loss of beaches and natural dune systems that provide protection to inland areas. This erosion further reduces the effectiveness of natural and built defences, creating a feedback loop that increases vulnerability to future overtopping events. Social impacts are also significant, particularly for coastal communities. Overtopping can result in the displacement of residents from flooded homes, often requiring emergency evacuations. Vulnerable populations, such as the elderly, children, and low-income communities, are disproportionately affected due to their limited capacity to relocate or recover. The economic cost of recovery can be substantial, as communities face expenses related to infrastructure repair, property loss, and the restoration of public services. Prolonged disruptions to transport, energy, and communication networks can affect daily life and hinder emergency response efforts. Over time, the psychological impacts of repeated flooding, displacement, and financial stress can affect the mental health and well-being of affected communities.

The maximum overtopping volume from individual wave events can be several times larger than the mean discharge, making pedestrian and vehicle safety a critical concern. (Koosheh, Etemad-Shahidi, Cartwright, Tomlinson, & van Gent, 2021). Methods for economically valuing human life in the context of coastal hazards, essential for conducting Cost-Benefit Analyses of coastal defence measures, include the Human Capital Method and Contingent Valuation Method (Bouma, François, Schram, & Verbeke, 2009). The Human Capital Method estimates the economic value of a life based on potential future earnings, either in gross terms or net terms, which considers the financial impact on others. The Contingent Valuation Method, use surveys to determine how much individuals are willing to pay to reduce their risk of death, directly assessing monetary value from individual preferences.

DESYCO employs a weighted multi-criteria analysis and integrates data from physical and socioeconomic scenarios, offering an effective means for assessing climate-induced vulnerabilities and consequences (Critto, et al., 2011). Alternatively there is the Social Vulnerability Index (SoVI) for disaster management (Flanagan, Gregory, Hallisey, Heitgerd, &

Lewis, 2011). The SoVI focuses on socioeconomic factors, such as income, housing, and population density.

Hinkel et al. (2014) estimates that, absent additional adaptation, 25–123 cm of global mean sea-level rise could lead by 2100 to 0.2–4.6% of the world's population being flooded annually, with expected annual losses (EAL) of 0.3–9.3% of global GDP (Hinkel, et al., 2014); such magnitudes motivate widespread adaptation. In terms of exposure, the value of assets situated below today's 100-year water level may reach US$21–210 trillion under a high-emissions pathway by 2100, so even low-frequency overtopping can translate into very large damages. Methodologically, EAL is computed by integrating damages across the full extreme water-level probability distribution up to a 1-in-10,000 event, with depth–damage represented by a logistic function in which roughly 50% of exposed assets are lost at 1 m inundation—assumptions that render the translation from surge depth and exceedance probability to expected loss explicit and replicable. Adaptation strategy emerges as the dominant control on outcomes: systematically raising dikes with rising sea level and affluence reduces impacts by about two to three orders of magnitude relative to no upgrade, though it entails global annual capital and maintenance costs on the order of US$12–71 billion by 2100 (Hinkel, et al., 2014). Modelled losses are more sensitive to the protection strategy than to differences among climate, socioeconomic, or topographic inputs, and the literature cautions that widespread protection can amplify catastrophic consequences when defences fail—precisely the context in which overtopping becomes most consequential.

2-3. Pluvial Flooding

Pluvial flooding occurs when intense, short-duration rainfall overwhelms natural drainage systems, stormwater infrastructure, or impermeable surfaces, leading to water accumulation in urban areas. Unlike fluvial flooding (which results from river overflow) or coastal flooding (which results from storm surges), pluvial flooding can happen anywhere, even in areas far from rivers or coastlines. Road infrastructure is particularly vulnerable to pluvial flooding due to its reliance on effective drainage systems and its critical role in transport, emergency response, and economic activity

2-3.1. Pluvial Flood Hazard and Modelling

The hazard of pluvial flooding is driven primarily by extreme rainfall events, which are becoming more frequent and intense in the UK due to climate change (Kay, Crooks, Davies, & Reynard, 2011). When heavy rainfall occurs over a short period, the capacity of drainage systems, such as gutters, storm drains, and culverts, is often exceeded. The hazard element of precipitation has been researched extensively with common practise to deploy depth-duration-frequency (DDF) or intensity-duration-frequency (IDF) curves. A DDF curve shows the expected height of rainfall for a given duration and frequency (Vesuviano, 2022). The duration refers to the length of time that the rainfall event lasts, while the frequency refers to the probability of the rainfall event occurring in a given year. The intensity is the amount of rainfall that falls during the event, typically measured in millimetres per hour. IDF curves are useful for a variety of applications, including designing drainage systems and flood control measures. DDF and IDF curves are available or can be easily derived for any area in Scotland. Flood hazards can also be estimated, but this often requires the development of hydraulic models to transform the information on the precipitation into hydraulic parameters such as e.g. flood height and velocity.

In their comprehensive review paper, Shejule and Pekkat (2022) characterise rainfall prediction as a multi-scale, data-intensive problem that blends statistical, numerical weather prediction (NWP), and machine-learning (ML) approaches, often fused with radar/satellite nowcasting in operations. Reviews document that national services—including the UK Met Office—deploy short-lead systems such as NIMROD (Met Office, 1 km Resolution UK Composite Rainfall Data from the Met Office Nimrod System, 2003) alongside global and mesoscale NWP, reflecting the need to couple extrapolation nowcasts with model guidance for the 0–6 h window. Forecast skill is strongly conditioned by resolution, lead time, and terrain: finer horizontal grids generally raise the probability of detection, critical success index, and equitable threat score (Shejule & Pekkat, 2022), but performance can vary by region and terrain, while skill typically degrades with lead time as false alarms rise and both probability of detection and critical success index fall. Additional predictors and preprocessing improve performance: precipitable water vapour derived from global navigation satellite systems enhances nowcasting in coastal and mountainous settings, and hybrid machine-learning methods with signal-processing steps such as wavelet transforms, singular spectrum analysis, and ensemble empirical mode decomposition reduce root-mean-square error and mean absolute error and better capture extremes. Critically for coastal risk work, the review notes that rainfall forecasts with adequate lead time are mandatory inputs to hydrologic models for flood forecasting and that operational hydro-meteorological systems already integrate numerical weather prediction and remote sensing for short-fuse guidance—providing the pluvial and fluvial forcings that can be coupled with tide–surge and coastal inundation models in compound-flood frameworks e.g. (Luettich, Jr, Westerink, & Scheffner, 1992).

Hindcasting rainfall using interpolation of point data involves analysing past weather events by filling in gaps between observed data points to reconstruct historical rainfall patterns across a larger area. This is done by applying interpolation methods such as kriging, inverse distance weighting, or spline interpolation, which estimate rainfall values at unsampled locations based on the spatial relationships and patterns of nearby measured values. Multiple authors have investigated the comparison between basic mathematical interpolation methods such as Theissen polygons and IDW and geostatistical interpolation methods such as Kriging, and the conclusion is that for low-density networks of rain gauges, Kriging outperforms the alternatives for daily, monthly and annual rainfall (Goovaerts, 2000), (Mair, 2011). The interpolated rainfall data creates a continuous spatial field, allowing meteorologists to evaluate and enhance predictive models and understand past weather dynamics more comprehensively. This method is particularly useful in areas with sparse observation networks, providing a more detailed and accurate historical record of rainfall distribution and intensity.

Kriging, also known as Gaussian process regression, represents a full family of geostatistical interpolation techniques (Bhattacharjee, 2019). They are used to predict the values of a random field at unobserved locations based on observations of the field at nearby locations. The method assumes that the random field being modelled is a realisation of a spatially continuous stochastic process, and the relationships between the values of the random field at different locations are described by a covariance function or semivariogram (Bhattacharjee, 2019). The semivariogram is a plot of the variance of the difference between pairs of points as a function of the distance between them. The semivariogram provides information about the spatial structure of the variable being studied, such as its range and the degree of spatial dependence, as well as providing a measure of uncertainty based on the kriging variance. This uncertainty is derived from the variogram and the spatial configuration of the known points relative to the prediction location.

Point measurements from gauges can be combined with areal products from weather radar, satellite retrievals and atmospheric reanalyses, with optional auxiliary covariates (elevation, topography, or climatology). These heterogeneous inputs are amalgamated using three broad families of methods : initial-field correction, which builds a prior rain field from one source (e.g., satellite or reanalysis) and incrementally corrects it toward gauges under an error-minimising criterion; interpolation with auxiliary information, which treats gauges as the primary variable and folds in radar/satellite/reanalysis as covariates within geostatistical or regression frameworks (co-kriging, kriging with external drift, generalized additive or geographically weighted regression); and optimal matching, which finds weights or a cost-minimising solution across sources (Bayesian model averaging, probability-density matching, variational methods). Effective merging must also reconcile spatial and temporal scale differences, blend distinct error structures, and—ideally—report uncertainty alongside the fused rainfall field; accuracy gains are most pronounced where gauges are sparse.

2-3.2. Pluvial Flood Vulnerability and Modelling

Flooding can affect infrastructure in various ways, including physical damage to roads, bridges, and buildings, as well as disruptions to transportation services and as such it is very difficult to quantify the overall costs. A significant effort was made by the European Commission through the development of global flood depth-damage functions for many typologies of infrastructure assets (Huizinga, De Moel, & Szewczyk, 2017). The study concluded that between 4-18% of flood infrastructure damage was attributed to roads with urban areas being more affected than rural areas. The impermeable surfaces used for roads and in urban areas mean does not allow rainwater to infiltrate the ground and instead accumulates on the surface.

In addition to direct physical damage, flooding often leads to indirect costs, such as business interruptions, loss of productivity, and increased transportation costs. These indirect costs are often more challenging to measure and quantify compared to direct physical damages. Analysis of these have been specifically carried out in a study focused on the vulnerability of roads in Scotland (Winter, et al., 2016). Quite surprisingly, there is a general lack of models for describing the fragility (or vulnerability) of roads.

Empirical fragility curves derive from observed data from historical events or experiments, making them data-driven and potentially reflective of real-world performance. However, this method requires a substantial amount of reliable, quality-assured data. In instances where this data is incomplete or unavailable, numerical methods can fill data gaps and model complex systems and interactions. Though, they are highly sensitive to model assumptions and demand a detailed understanding of physical processes and sophisticated computational tools, making them resource-intensive. A comprehensive comparison of empirical and numeric flood fragility and vulnerability methods is given by Galasso et al., (Galasso, Pregnolato, & Parisi, 2021).

By comparison, expert judgment offers a swift estimate in resource-limited scenarios and can integrate holistic knowledge. However, it heavily relies on the individual's expertise, making results susceptible to personal biases and lacking precision in quantifying damage states compared to the other two methods. An example of cross disciplinary expert judgement applied to multiple hazard transport fragility for roads subjected to debris flow is given in (Argyroudis, Kaynia, Mitoulis, & Winter, 2019).

An empirical flood depth/disruption curve for road flooding was developed by Pregnolato et al. by fitting a relationship between the depth of standing water and vehicle speed (Pregnolato, Ford, Wilkinson, & Dawson, 2017). This relationship was fit to data points that were a combination on of experimental study, road safety literature and from expert opinion. This relationship was then incorporating this into existing transport models to produce better estimates of flood induced delays.

Fragility analysis is an important tool for analysing the susceptibility of a system or structure to failure or damage under different conditions. In disaster models, fragility curves are frequently employed to establish the likelihood of surpassing a specific damage level based on the intensity of the hazard that triggers the disaster. Fragility curves play a crucial role in evaluating and mitigating the risk of road damage caused by flooding, offering valuable insights for infrastructure planning, design, and emergency response strategies. These curves can be broadly classified into three main categories, depending on the approach employed for deriving them: empirical, numerical, expert judgment, or combinations thereof.

2-3.3. Pluvial Flood Consequences and Modelling

For rainfall-driven flood risk, multiple datasets can be amalgamated to quantify both hazard and economic loss in ways that are directly useful for policy. On the hazard side, daily precipitation from atmospheric reanalyses (for example, ERA5) can be combined with subnational economic accounts to estimate how different aspects of rainfall—greater totals, more “wet days,” and rarer extreme-intensity days—affect growth. Using a panel of 1,554 regions over four decades, fixed-effects regressions show that increases in the number of wet days and in extreme daily rainfall measurably reduce economic growth, with particularly strong effects in high-income economies and in services and manufacturing —evidence that translates hydrometeorological conditions into macroeconomic consequences (Kotz, Levermann, & Wenz, 2022). On the loss side, event-level amalgamation links observed damages with probabilistic event attribution: insured flood losses and economy-wide drought losses are paired with the “fraction of attributable risk” for each event to estimate the share of cost due to anthropogenic influence (attributable cost = FAR × observed cost). Applied to New Zealand’s 2007–2017 decade, this bottom-up merging yields about NZ$140 million of insured rainfall-flood damages and roughly NZ$800 million of drought losses attributable to human-caused climate change, while explicitly noting that insured losses understate total economic impacts by excluding uninsured assets and indirect production losses (Frame, et al., 2020). In practice, these two strands are complementary: reanalysis-economy panels quantify how shifts in the distribution of daily rainfall depress growth, while attribution-plus-accounts convert specific extreme events into attributable monetary losses. Together they provide a transferable template for coastal and inland risk studies: combine high-frequency hydro-meteorological data with observed losses and coherent attribution to generate decision-relevant economic evidence.

2-4. Machine Learning Based Forecasting

Machine learning models offer significant advantages in hydrological and hydrodynamic modelling due to their proficiency in capturing complex, non-linear relationships. Machine learning models excel at pattern recognition, which allows them to predict hydrological and hydrodynamic hazards with high accuracy and greater speed when compared to and used alongside traditional methods. The following subsection contains a literature review of machine learning models specifically developed for forecasting the hazards described in the previous sections. At the end of section 2-4, Table 1 is a summary of all the strengths and weaknesses of each of the common machine learning modelling options

2-4.1. Coastal Overtopping Hazard

Elbisy (2023) explores various machine learning techniques including artificial neural networks and support vector machines to predict wave overtopping discharges at coastal structures (Elbisy, 2023). The study utilises the EurOtop database and assesses the predictive performance using statistical features, where the general regression neural network (GRNN) showed highly accurate results. Sensitivity analysis was performed to determine the significance of each predictive variable and the optimal validation method for GRNN, revealing that leave-one-out validation method slightly outperformed cross-validation.

den Bieman, van Gent, and van den Boogaard (2021) present an advanced model using XGBoost, for predicting mean wave overtopping discharge (den Bieman, van Gent, & van den Boogaard, 2021). They improved the model's accuracy through feature selection, hyperparameter tuning, and adding new physical model data to the training set. The XGBoost model demonstrated significant error reduction compared to previous neural network models and empirical formulas, especially with normally incident waves.

Habib et al. (2022) provides a systematic review of machine learning algorithms applied to the prediction of wave overtopping characteristics at coastal flood defences (Habib, O'Sullivan, & Salauddin, 2022). The review highlighted decision trees and artificial neural networks as popular methods and compared these machine learning -based predictions with empirical models. It noted that while machine learning models offer quick and economically viable predictions, they require further refinement, especially in handling complex geometrical configurations of coastal structures.

2-2.4.1. Astronomic Tide

Several tidal prediction studies have demonstrated the enhancement of traditional tidal prediction models with machine learning methodologies. Liu et al. (2019), introduces a combined tidal forecasting model that integrates harmonic analysis with an Autoregressive Integrated Moving Average-Support Vector Regression (ARIMA-SVR) (Liu, Guoyou, & Kaige, 2019). This model enhances tidal prediction accuracy by incorporating non-astronomical factors alongside traditional astronomical influences, effectively capturing a broader range of influences on tidal levels. However, the model demands substantial data preprocessing and is computationally intensive, which may limit its use in real-time applications. Fei et al. (2023) integrates physics-based models and machine learning techniques such as LSTM neural networks and XGBoost (Fei, Haoxuan, & Liang, 2023). This integrated approach leverages detailed physical modelling and adaptive machine learning to offer more accurate and nuanced predictions of water levels influenced by both upstream discharge and astronomical tides. Similar to the work of Liu et al. (2019), the model's complexity and reliance on diverse and extensive data sets could pose operational challenges, particularly in calibration and validation processes. Ban et al. (2023) demonstrates a cutting-edge method using Variational Mode Decomposition combined with LSTM neural networks (Ban, Shen, Lu, Liu, & Pan, 2023). This approach focuses on decomposing tidal data into simpler components for more manageable and accurate predictions. In general there is great potential for hybrid models to enhance tidal prediction capabilities albeit with increased demands on data quality and computational resources.

2-2.4.2. Non-tidal Residual

Different types of machine learning processes have been applied to surge forecasting. French et al. (2017) presents a comprehensive study on the enhancement of storm surge forecasting using a hybrid approach that integrates Artificial Neural Networks (ANNs) with traditional hydrodynamic models (French, Mawdsley, Fujiyama, & Achuthan, 2017). The authors focus on the specific challenges of predicting extreme storm surge water levels, especially for ports located in estuarine areas such as Immingham in the UK, which are typically inadequately served by existing regional tide-surge models due to their low spatial resolution. The ANN is configured to simulate tidal surge by learning from metocean data, including observations of surge at distant tide gauges, wind and atmospheric pressure data, combined with the astronomical tide (French, Mawdsley, Fujiyama, & Achuthan, 2017). This approach allows for the prediction of flood extent and damage potential across the port area with an accuracy comparable or better than the UK national tide-surge model, especially for short-term forecasts (6 to 24 hours ahead). Convolutional Neural Networks (CNNs) are neural networks designed to process image data and can be used to analyse satellite and radar data to detect and track storms, as well as to extract features that can be used as inputs to other models. Multilayer Perceptron (MLP) Neural Networks are feedforward neural networks that can be used for storm forecasting by analysing various meteorological variables, such as wind speed, atmospheric pressure, and temperature. Recurrent Neural Networks (RNNs) are designed to process sequential data, making them well-suited for weather forecasting tasks that involve time-series data (Salman, Heryadi, Abdurahman, & Suparta, 2018).

Tiggeloven et al. (2021) explores the use of four architecture of deep learning methods (ANN, CNN, LSTM, and Convolutional LSTM) to predict the surge component of sea-level variability based on local atmospheric conditions (Tiggeloven, 2021). The models were constructed using global tide station data and showed the best performance in the mid-latitudes. The LSTM model generally outperformed the other models, but adding more predictor variables improved performance, though with increased computation time. The study found that the deep learning models could be useful for predicting extreme sea levels or emergency response.

2-2.4.3. Run-up

In the study by Durap et al. (2023), a comprehensive analysis of nine different machine learning methods is employed to predict wave runup, emphasizing a broad array of coastal parameters (Durap, 2023). The paper notes the limitations of linear models due to their inability to capture the complexities of the dataset and underscores the effectiveness of non-linear models. It highlights the importance of integrating factors such as significant wave height, peak wave period, and foreshore beach slope, which significantly affect wave runup predictions. This research underscores the potential of machine learning to aid in the design of resilient coastal structures and the management of coastal risks, emphasizing the necessity of managing feature selection and model complexity to improve prediction accuracy.

Naeini and Snaiki (2024) introduces a physics-informed machine learning model that efficiently simulates time-dependent wave runup (Naeini & Snaiki, 2024). Their approach combines the computational efficiency of the XBeach model's Surfbeat (XBSB) mode with the accuracy of its Nonhydrostatic (XBNH) mode using a conditional generative adversarial network (cGAN). This method enhances the model's predictive capability by incorporating physics-based knowledge from XBSB, showcasing its potential to significantly enhance coastal risk assessment and management by providing a robust tool for accurately predicting complex wave runup dynamics.

Kim and Lee (2024) focus on the interpretability of machine learning models, evaluating the predictive accuracy of models such as XGBoost and Support Vector Machines (SVM) against traditional empirical formulas for wave runup (Kim & Lee, 2024). Their findings indicate that while machine learning models, particularly XGBoost, provide superior prediction accuracy, they also face challenges when applied to data outside the training range. The study emphasizes the influence of factors like beach slope and wave steepness on predictions and advocates for the use of interpretable machine learning to understand how these models derive their predictions.

A common theme across these studies is the enhanced predictive accuracy of machine learning models over traditional empirical methods in predicting wave runup, highlighting the significance of incorporating advanced computational techniques in coastal engineering. There is also a shared emphasis on the need for understanding the internal workings of machine learning models and selecting appropriate features to enhance model reliability and applicability. Furthermore, the integration of physics-based models with machine learning techniques, as exemplified in the study by Saviz Naeini and Snaiki, illustrates a growing trend in coastal engineering research to blend traditional knowledge with new computational methods for better prediction and management of coastal hazards.

2-2.4.4. Wind-waves

Huang et al. (2022) explores the use of Convolutional Neural Networks (CNNs) for regional wave prediction, employing historical wind and wave data for model input (Huang, Jing, Chen, Zhang, & Liu, 2022). This study highlights the CNN's capability to capture spatial dependencies within the data, which significantly enhances prediction accuracy for regional wave heights. The adaptability of the model across different computational regions underscores its utility in operational settings, boasting superior computational efficiency and accuracy compared to traditional SWAN simulations.

Yeganeh-Bakhtiary et al. (2023) compare the efficacy of semi-empirical methods with Decision Tree (DT) models, particularly focusing on the M5p Decision Tree algorithm, for wave prediction (Yeganeh-Bakhtiary, EyvazOghli, Shabakhty, & Abolfathi, 2023). Utilising wind speed and altimeter data from satellites to train the models, the study demonstrates that the DT-based model significantly outperforms semi-empirical methods. The robustness of the M5p model over large spatial and temporal scales makes it a viable option for wave prediction.

Zhang et al. (2023) introduces an innovative approach to improve the prediction of local significant wave height by incorporating the memory of past winds into a Support Vector Regression model (Zhang, et al., 2023). This method leverages the residual impact of historical winds to significantly enhance prediction accuracy. The research emphasizes the importance of considering the directional and temporal aspects of wind, along with the physical geography of the measurement site, to refine prediction models.

Across these studies, several common themes emerge, including the integration of historical data to bolster predictive accuracy and the advantage of machine learning techniques over traditional methods. These approaches offer robust frameworks for handling complex, non-linear relationships inherent in meteorological data, providing superior predictive performance and computational efficiency. The distinction between regional and local prediction models is also evident, with each type serving critical, context-specific forecasting needs.

2-4.2. Pluvial Flood Hazard

Machine learning processes are also commonly applied to pluvial flood modelling. Forecasting future pluvial flooding events harnesses real-time or near-real-time data to predict imminent flood risks. Zhao et al. (2022) present a supervised learning model that integrates precipitable water vapor from the Global Navigation Satellite System and meteorological parameters to forecast rainfall accurately and promptly, essential for timely flood interventions. Similarly, Hofmann and Schüttrumpf (2021) explore the potential of a deep convolutional generative adversarial network, "floodGAN," which performs real-time image-to-image translation from rainfall data to flood inundation maps, drastically reducing prediction times suitable for early warning systems.

Further enhancing real-time capabilities, Safaei-Moghadam et al. (2024) introduce a hybrid machine learning model that not only predicts urban pluvial flood risk using a graph-based flood spreading model but also incorporates real-time data from Waze, enhancing the model's accuracy and responsiveness. Meanwhile, Barrera-Animas et al. (2022) focus on utilising LSTM-based networks to forecast rainfall using time-series data, proving effective in handling multivariate data and providing accurate hourly rainfall forecasts necessary for predicting potential flood events.

On the other hand, hindcasting involves analysing past events to refine and calibrate models for better accuracy in future forecasts. Boudevillain et al. (2016) exemplify this approach by utilizing the Kriging with External Drift (KED) method for high-resolution rainfall re-analysis based on radar and rain gauge data merging. This method significantly improves quantitative precipitation estimates, which are crucial for understanding and modelling past major rain events that could lead to flash floods. By providing a more accurate historical record, these enhanced estimates improve the calibration of models used for future forecasting.

These studies collectively underscore the importance of integrating various data sources and employing advanced machine learning techniques to enhance the predictive accuracy of flood forecasting and hindcasting models. By balancing the immediate needs of disaster response with the long-term requirements of planning and risk management, these approaches demonstrate a robust strategy to manage both immediate and future flooding risks effectively. Through forecasting and hindcasting, the methodologies not only address real-time applications but also ensure that lessons learned from past events significantly inform and improve future predictions.

The distinction between forecasting and hindcasting in these studies highlights their complementary roles. Forecasting approaches are designed to be dynamic, providing real-time or near-real-time data to address ongoing or imminent weather events that could result in flooding. These models are valued for their speed and responsiveness, incorporating live data feeds and advanced machine learning techniques to offer actionable insights quickly in emergency scenarios. Hindcasting, meanwhile, focuses on leveraging historical data to refine and calibrate predictive models. This process not only helps adjust the models to be more reflective of real-world conditions but also supports planning and mitigation strategies based on past flood events. This is the desirable outcome for the pluvial flood vulnerability objective in Chapter 6 of this thesis.

Model Type	Strengths	Limitations
CNN / ConvLSTM (Qin, Su, Chu, Zhang, & Song, 2023); (Tiggeloven, 2021); (Huang, Jing, Chen, Zhang, & Liu, 2022); (Hofmann & Holger, 2021)	exploit spatial structure in gridded forcings and capture coherent patterns; learn space–time dependencies when using convolutional recurrent layers; ingest multi-variable fields such as wind, pressure, and precipitation consistently; can detect storm fronts, jets, and mesoscale features relevant to surge; transferable across nearby regions with careful re-training; strong performance for regional pattern recognition and extreme-event mapping.	data- and computation-intensive to train and validate; require consistent grids, careful pre-processing, and quality control of reanalysis/radar inputs; risk of overfitting when events are few; limited extrapolation beyond the training climatology; reduced interpretability without post-hoc explanation tools; long training times and storage requirements.
LSTM / GRU (Qin, Su, Chu, Zhang, & Song, 2023); (Tiggeloven, 2021); (Fei, Haoxuan, & Liang, 2023); (Ban, Shen, Lu, Liu, & Pan, 2023); (Barrera-Animas, et al., 2022)	represent temporal dependence and long-lag memory in non-tidal residuals; support multi-step forecasting across lead times; integrate exogenous predictors such as atmospheric pressure, wind, and antecedent water level; handle nonlinearity and hysteresis in surge build-up and decay; well suited to sequence-to-sequence prediction.	data-hungry and sensitive to hyper-parameter choices; training instability if sequences are short or highly intermittent; explanations of learned dynamics are limited; may smooth sharp peaks without tailored loss functions; performance degrades under distribution shift if events differ from training cases.
MLP (Qin, Su, Chu, Zhang, & Song, 2023); (Tiggeloven, 2021); (Elbisy, 2023); (Habib, O’Sullivan, & Salauddin, 2022)	flexible nonlinear approximators for tabular predictors; fast, batched inference suitable for operational pipelines; straightforward regularisation and early-stopping; easy to combine with posterior weighting and Monte Carlo sampling.	no inherent temporal memory without explicit lag features; require careful feature engineering and scaling; prone to overfitting on small datasets; limited interpretability compared with tree-based models; extrapolation beyond training range is uncertain.
Gradient Boosting (Zounemat-Kermani, Batelaan, Fijani, & Hinkelmann, 2021); (den Bieman, van Gent, & van den Boogaard, 2021); (Fei, Haoxuan, & Liang, 2023); (Kim & Lee, 2024)	strong accuracy on structured/tabular data; capture nonlinear interactions and threshold effects; built-in regularisation and robust handling of heterogeneous predictors; useful variable-importance diagnostics; competitive with moderate sample sizes.	not sequence-aware by default; sensitive to hyper-parameters and learning-rate schedules; piecewise-constant predictions can under-represent sharp extremes; limited physical interpretability; extrapolation outside observed ranges is unreliable.
Random Forest/ Decision Tree (Xu & Liang, 2021); (Habib, O’Sullivan, & Salauddin, 2022); (Yeganeh-Bakhtiary, EyvazOghli, Shabakhty, & Abolfathi, 2023)	good out-of-the-box accuracy with minimal tuning; handle nonlinearities, interactions, and outliers; provide variable-importance rankings for feature screening; bagging reduces variance and improves stability on small datasets.	no explicit temporal memory; tendency to under-predict extreme peaks due to averaging; can be less accurate than boosting on complex targets; model size can be large; step-like responses hinder smooth sensitivity analysis.
SVR (Xu & Liang, 2021); (Liu, Guoyou, & Kaige, 2019); (Kim & Lee, 2024); (Zhang, et al., 2023)	effective with modest sample sizes; kernel functions capture nonlinear relationships; robust to outliers through margin-based loss; well suited to short-lead tabular prediction.	performance highly dependent on kernel choice and hyper-parameters; computational cost grows with dataset size; lacks native probabilistic outputs and requires extra calibration for uncertainty; no built-in treatment of temporal dependence.

Model Type	Strengths	Limitations
Gaussian Process Regression (Xu & Liang, 2021)	produces predictive distributions with calibrated uncertainty; sample-efficient under small to moderate data; flexible covariance kernels encode smoothness and correlation structures; naturally handles observation noise; supports inclusion of prior mean functions.	cubic computational scaling with the number of training samples; kernel selection and hyper-parameter optimisation are delicate; extrapolation beyond the support of the data is weak; stationary kernels can miss non-stationary dynamics unless extended.
Hybrid physics + ML (French, Mawdsley, Fujiyama, & Achuthan, 2017); (Qin, Su, Chu, Zhang, & Song, 2023); (Naeini & Snaiki, 2024)	leverage physical model structure for phase and boundary constraints while correcting bias with data-driven components; can reduce high-performance-computing cost via emulation or bias-correction; improved face validity and stakeholder acceptance; align with data assimilation workflows.	integration is complex and requires consistent forcing and data assimilation choices; two coupled systems to maintain and validate; potential for compensating errors between components; alignment of spatial and temporal grids is mandatory.
Ensembles (Zounemat-Kermani, Batelaan, Fijani, & Hinkelmann, 2021)	improves generalisation by combining diverse learners; reduces variance and hedges against model misspecification; provides spread that can be converted to uncertainty with calibration; often more robust to distribution shift than single models.	higher computational and operational complexity; risk of data leakage if stacking is not carefully partitioned; interpretability decreases as members increase; uncertainty spread requires calibration to achieve nominal coverage.
Bayesian model selection (Zhu, Yang, & Ren, 2023); (Xu & Liang, 2021); (Salighehdar, Ye, Liu, Ionut, & Blumberg, 2017); (Oparanji, Sheu, Bankhead, Austin, & Patelli, 2017)	principled posterior weighting of candidate models; coherent prediction intervals and credible intervals; ability to incorporate prior information; separates model, feature, and forecast uncertainties in a single framework; well suited to decision-making under uncertainty.	additional computational cost for posterior estimation; sensitivity to likelihood formulation and prior assumptions; implementation complexity increases with the number and diversity of candidate models; requires careful validation of probabilistic calibration.

Table 1: Machine Learning strengths and limitations of different machine learning modelling types.

2-5. Uncertainty Analysis

In the realm of uncertainty analysis, uncertainties are categorized into two primary types: aleatory and epistemic (Wang, et al., 2019). Aleatory uncertainty, also known as inherent or stochastic uncertainty, encompasses the natural variability inherent in physical processes and systems. This type of uncertainty cannot be reduced by acquiring more information but can be managed through design and operational strategies. Epistemic uncertainty, on the other hand, arises from incomplete knowledge or understanding of the system being studied. This type of uncertainty is reducible as more information becomes available, through improved data acquisition, enhanced modelling techniques, or broader scientific understanding.

Aleatory uncertainty in hydrodynamic and hydrological contexts includes several subtypes. For example, in riverine systems, it could encompass the randomness in seasonal rainfall, snowmelt patterns, or the occurrence of unpredictable meteorological events like sudden atmospheric depressions. In coastal engineering, it includes the variability in wave patterns, storm frequencies, or sea-level rises. Such uncertainties are critical in the planning and construction of flood defences, dams, levees, and coastal barriers. They represent the uncontrollable aspects of nature that engineers must design for, ensuring structures are resilient against a wide range of natural conditions and events. These uncertainties are handled through conservative design principles, probabilistic risk assessment, and contingency planning. This uncertainty is often referred to as Type A uncertainty in standards set by organizations such as the International Organization for Standardization (ISO), where it is characterized by its inability to be reduced through further data collection or analysis. Aleatory uncertainty represents the randomness or inherent variability that exists in natural systems, irrespective of the amount of data collected.

Epistemic uncertainty, also considered Type B uncertainty, includes data uncertainty, model uncertainty, and uncertainties due to limitations in our understanding of climatic variability. Data uncertainty arises from inaccuracies in data measurements, data collection methods, or processing techniques. Model uncertainty stems from the assumptions, simplifications, and mathematical formulations used in wave prediction models. These models may not fully capture the complex dynamics of marine environments or may rely on inadequate physical representations. Uncertainties due to natural climatic variability, which are considered a part of model uncertainty, involve the changes in climate patterns that are not accurately captured by historical data or adequately predicted by current models. Unlike aleatory uncertainty, epistemic uncertainties are reducible as improvements in technology, methodologies, and scientific insight allow for better data collection and more accurate modelling.

2-5.1. Uncertainties in Machine Learning

Within the modelling blocks that drive modern hydrological and hydrodynamic forecasts, uncertainty stems from data (noise, missing data, sampling bias/imbalance) and model (structure and parameters) (Fakour, Mosleh, & Ramezani, 2024). In the context of neural networks, uncertainties can also arise from the complexities involved in optimisation. Yu et al. (2020) discussed in detail the challenges of hyperparameter optimization in neural networks, including problems with dimensionality and the high cost of evaluating different combinations of hyperparameters (Yu & Hong, 2020). This leads to variations in performance and output accuracy, introducing uncertainty in selecting the best performing ANN. Yu et al. (2020) also discussed various solutions, such as using Bayesian optimization, gradient-based methods, and evolutionary algorithms for hyperparameter optimization. Fakour, Mosleh & Ramezani (2024) emphasize that uncertainty metrics should be evaluated directly (per-sample and dataset-level) and that deep-learning-centric techniques e.g. dropout, ensembles, conformal wrappers, often outperform naive “confidence” methods such as softmax probability as naïve scores are often overconfident, especially under distribution shift or class imbalance (Fakour, Mosleh, & Ramezani, 2024).

Kingston et al. (2008) used Bayesian model selection for water resources modelling, finding that it was an objective method for accurately selecting the optimal complexity of an ANN model when used in conjunction with the Bayesian training procedure (Kingston, Maier, & Lambert, 2008). Additionally, despite sharing the same architecture and training process, ANNs can produce different models due to the random initialization of their weights during training (Oparanji, Sheu, Bankhead, Austin, & Patelli, 2017). This leads to variations in performance and output accuracy, introducing uncertainty in selecting the best performing ANN. This can be mitigated by using optimized weight initialization methods and seeding to ensure reproducibility. Usually, multiple ANN models are trained in searching for the optimal one (Kim, Pan, & Mase, 2019) and cross-validation tests are commonly used to select the best performing ANN (Tolo, et al., 2018), whereas the other ANNs are discarded. However, this approach does not consider the potential noise and imprecision in the validation data or the ANN's performance on unseen data, and it does not fully exploit the results of the training process in developing an optimal model. Oparanji et al. (2017) proposed a Bayesian model averaging algorithm to provide an averaged prediction across a set of networks with each network weighted given its likelihood of being correct. The prediction obtained according to this approach is more accurate than that obtained with a single ANN. Moreover, the variation across the network set provides an estimate of the epistemic uncertainty in the neural network modelling approach. Generally, Bayesian inference and model averaging (e.g., Gaussian processes, Bayesian neural nets) are well established methods to carry posterior uncertainty over hypotheses into predictive uncertainty (Hüllermeier & Waegeman, 2021).

2-5.2. Uncertainties in Forecasting

Naturally, uncertainty in machine-learning forecasting can be classed as epistemic and aleatoric. Epistemic (model) uncertainty arising from limited knowledge or misspecification, and aleatoric (data/measurement) uncertainty arising from irreducible noise. Credible operational systems must return full predictive distributions, not just points (Wang, et al., 2019).

In hydrodynamic and flood-forecast settings, rigorous treatment of uncertainty typically proceeds via Bayesian inference: prior beliefs about parameters (e.g., roughness) are updated with observations using likelihood functions that reflect real error behaviour—often heteroscedastic, autocorrelated, and non-Gaussian (Pinheiro, Naghettini, & Palmier, 2019). This is often followed by Markov-chain Monte Carlo to sample the posterior and propagate uncertainty to predictive distributions of stages, discharges, and inundation. This approach cautions that adopting simplistic Gaussian, independent-error assumptions can bias posterior estimates and understate predictive spread. An alternative is generalized likelihoods that admit skewness and heavy tails improve realism and yield probabilistic flood maps suitable for decision-making under risk.

For the purposes of this thesis, we will use the terms feature error and forecast error to represent two critical sources of uncertainty in predictive modelling. We take feature error as the error that arises from the limitations in the available data used to describe the variability of key phenomena, such as surge heights in storm surge modelling. This type of error can result from sparse or biased datasets, measurement inaccuracies, or an inability to fully capture the complexity of physical processes. If the features used in a model do not sufficiently represent real-world variability, the model's predictive power is inherently constrained. On the other hand, forecast error stems from the uncertainty in weather predictions, which serve as inputs for models that rely on meteorological conditions. Since weather forecasts are based on numerical models that approximate atmospheric dynamics, they are subject to errors due to incomplete initial conditions, model assumptions, and the chaotic nature of weather systems. As forecast errors propagate through predictive models, they can significantly affect the reliability of downstream predictions.

2-6. Literature Review Conclusions

This literature review has provided a solid foundation for understanding the dynamics of meteorological processes, hydrodynamic and hydrological hazards, their associated vulnerabilities and consequences, and the application of metamodelling techniques. The synthesis of prior work underscores the critical nature of coastal and pluvial flooding under climate change. Throughout, advanced numerical models remain central, with machine-learning methods deployed to reduce computational burden and, in some cases, improve predictive skill. A consistent theme is forecast uncertainty: high-fidelity models struggle to characterise it without substantial computational resources. Against this backdrop, Chapters 3–5 develop a coastal inundation forecasting framework that explicitly quantifies both epistemic and aleatoric uncertainty.

The forecasting framework advances established practice by combining operational efficiency with calibrated uncertainty. Surrogate models replace repeated runs of high-fidelity hydrodynamic models and produce prediction intervals rather than single-value forecasts. Posterior-weighted averaging and Monte Carlo propagation quantify forecast, feature, and model-structure uncertainty, addressing persistent gaps in extreme-condition validation and comparability (Qin, Su, Chu, Zhang, & Song, Machine Learning for Storm Surge Prediction: A Review, 2023) (Zhu, Yang, & Ren, 2023) (Xu & Liang, 2021). Time-dependent error correlation is handled explicitly via covariance-aware sampling, improving multi-lead performance when atmospheric errors are serially correlated. Collectively, this provides an efficient and defensible alternative to both purely high-fidelity workflows and deterministic surrogate models.

In parallel, the specific vulnerabilities associated with different hydrological and hydrodynamic hazards have been examined, with emphasis on critical infrastructure such as coastal defences and road networks. A gap in the literature relating to road fragility under extreme precipitation is identified; Chapter 6 addresses this with an uncertainty-aware treatment.

The vulnerability analysis moves beyond region-average depth–damage or generic fragility assumptions by constructing uncertainty-based fragility relationships tailored to road networks. Empirical observations are fused with surrogate metamodels to estimate site-level intensity, and uncertainty is propagated to yield calibrated probability intervals for conditional failure across the intensity range. The framework distinguishes feature uncertainty (e.g., intensity estimation and coverage) and model-form uncertainty (fragility function choice), producing estimates that are both sharp and reliable for decision-making. In doing so, it addresses the identified gap in road fragility under extreme precipitation and provides operational thresholds to support warning and response.

2-References

- Abubakar, A., Mahmud, M., Tang, K., Hussaini, A., & Md Yusuf, N. H. (2019). A Review of modelling approaches on tidal analysis and prediction. *International Archives of the*

Photogrammetry, Remote Sensing and Spatial Information Sciences. doi:10.5194/isprs-archives-XLII-4-W16-23-2019

- Akperov, M., Golitsyn, G., & Semenov, V. (2024). Energy of Cyclones and Anticyclones during their Development. *Doklady Earth Sciences*. doi:https://doi.org/10.1134/S1028334X24603468
- Al Kajbaf, A., & Bensi, M. (2020). Application of surrogate models in estimation of storm surge: A comparative assessment. *Applied Soft Computing, 91*, 106184.
- Allsop, W. B. (2005). Wave overtopping at vertical and steep seawalls. *Proceedings of the Institution of Civil Engineers-Maritime Engineering, 158, 3*, 103-114. Thomas Telford Ltd.
- Almar, R., Ranasinghe, R., Bergsma, E., Diaz, H., Melet, A., Papa, F., . . . Kestenare, E. (2021). A global analysis of extreme coastal water levels with implications for potential coastal overtopping. *Nature communications, 12*(1), 3775.
- Altomare, C., & Gironella, X. (2024). Characterization of Overtopping Volumes from Focused Wave Groups over Smooth Dikes with an Emerged Toe: Insights from Physical Model Tests. *Journal of Marine Science and Engineering, 12*(7), 1143.
- Argyroudis, S., Kaynia, A., Mitoulis, S., & Winter, M. (2019). Fragility of transport assets exposed to multiple hazards: State-of-the-art review toward infrastructural resilience. *Reliability Engineering & System Safety*.
- Ban, W., Shen, L., Lu, F., Liu, X., & Pan, Y. (2023). Research on Long-Term Tidal-Height-Prediction-Based Decomposition Algorithms and Machine Learning Models. *Remote Sensing, 15*(12). doi:10.3390/rs15123045
- Barrera-Animas, A., Oyedele, Y., Bilal, M., Akinosho, T., Delgado, J., & Akanbi, L. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications, 7*.
- Bhattacharjee, S. S. (2019). Semantic Kriging for Spatio-temporal Prediction / [internet resource]. *1*.
- Bitner-Gregersen, E., Waseda, T., Parunov, J., Yim, S., Hirdaris, S., Ma, N., & Guedes Soares, C. (2022). Uncertainties in long-term wave modelling. *Marine Structures, 84*. doi:https://doi.org/10.1016/j.marstruc.2022.103217
- Booij, N., Holthuijsen, L. H., & Ris, R. C. (1996). THE" SWAN" WAVE MODEL FOR SHALLOW WATER. *Coastal engineering*, 668-676.
- Boudevillain, B., Delrieu, G., Wijbrans, A., & Confoland, A. (2016). A high-resolution rainfall re-analysis based on radar–raingauge merging in the Cévennes-Vivarais region, France. *Journal of Hydrology*.
- Bouma, J., François, D., Schram, A., & Verbeke, T. (2009). Assessing socio-economic impacts of wave overtopping: An institutional perspective. *Coastal Engineering, 56*, 203-209.
- C3S, C. C. (2017). ERA5: Fifth generations of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS). Retrieved July 2021, from https://cds.climate.copernicus.eu/cdsapp#!/home
- Chelton, D. B., & Schlax, M. G. (1996). Global observations of oceanic Rossby waves. *Science*, 234-238.
- Chen, C. L. (2003). An unstructured grid, finite-volume, three-dimensional, primitive equations ocean model: application to coastal ocean and estuaries. *Journal of atmospheric and oceanic technology, 20(1)*, 159-186.
- Critto, A., Torresan, S., Zabeo, A., Rizzi, J., Gallina, V., Giove, S., & Marcomini, A. (2011). DESYCO: a Decision Support System for the assessment of coastal flooding and erosion risk due to climate change. *EGU 2011.*

- da Silva, P. G., Coco, G., Garnier, R., & Klein, A. (2020). On the prediction of runup, setup and swash on beaches. *Earth-Science Reviews, 204*. doi:https://doi.org/10.1016/j.earscirev.2020.103148.
- den Bieman, J., van Gent, M., & van den Boogaard, H. (2021). Wave Overtopping Predictions Using an Advanced Machine Learning Technique. *Coastal Engineering, 166*. doi:10.1016/j.coastaleng.2020.103830
- Denissenko, P., Didenkulova, I., Pelinovsky, E., & Pearson, J. (2011). Influence of the nonlinearity on statistical characteristics of long wave runup. *Nonlinear Processes in Geophysics, 18*, 967–975. doi:https://doi.org/10.5194/npg-18-967-2011
- Durap, A. (2023). A Comparative Analysis of Machine Learning Algorithms for Predicting Wave Runup. *Anthropocene Coasts, 6*(1). doi:10.1007/s44218-023-00033-7
- Elbisy, M. (2023). Machine Learning Techniques for Estimating Wave-Overtopping Discharges at Coastal Structures. *Ocean Engineering, 273*. doi:10.1016/j.oceaneng.2023.113972
- Fakour, F., Mosleh, A., & Ramezani, R. (2024). A Structured Review of Literature on Uncertainty in Machine .
- Fei, K., Haoxuan, D., & Liang, G. (2023). Accurate Water Level Predictions in a Tidal Reach: Integration of Physics-Based and Machine Learning Approaches. *Journal of Hydrology, 622*. doi:10.1016/j.jhydrol.2023.129705
- Fernández-Montblanc, T., Vousdoukas, M., Ciavola, P., Voukouvalas, E., Mentaschi, L., Breyiannis, G., . . . Salamon, P. (2019). Towards robust pan-European storm surge forecasting. *Ocean Modelling, 133*, 129-144.
- Ferreira, J., Cardona, F., Jóia Santos, C., & Tenedório, J. (2021). Hazards, Vulnerability, and Risk Analysis on Wave Overtopping and Coastal Flooding in Low-Lying Coastal Areas: The Case of Costa da Caparica, Portugal. *Water, 13*(2). doi:https://doi.org/10.3390/w13020237
- Flanagan, B., Gregory, E., Hallisey, E., Heitgerd, J., & Lewis, B. (2011). A social vulnerability index for disaster management. *Journal of homeland security and emergency management, 8*(1).
- Frame, D., Rosier, S., Noy, I., Harrington, L., Carey-Smith, T., Sparrow, S., . . . Dean, S. (2020). Climate-change attribution and the economic costs of extreme weather events: damages from extreme rainfall and drought. *Climate Change, 162*, 781–797. doi:https://doi.org/10.1007/s10584-020-02729-y
- Franco, L., De Gerloni, M., & Van der Meer, J. (1995). Wave overtopping on vertical and composite breakwaters. *Coastal Engineering*, 1030-1045.
- French, J., Mawdsley, R., Fujiyama, T., & Achuthan, K. (2017). Combining machine learning with computational hydrodynamics for prediction of tidal surge inundation at estuarine ports. *Procedia IUTAM, 25*, 28-35.
- Funakoshi, Y., Feyen, J., Aikman, F., Van der Westhuysen, A., & Tolman, H. (2013). The Extratropical Surge and Tide Operational Forecast System (ESTOFS) Atlantic implementation and skill assessment. *NOAA technical report NOS CS , 32*. Coast Survey Development Laboratory (U.S.). Retrieved from https://repository.library.noaa.gov/view/noaa/16917
- Galasso, C., Pregnolato, M., & Parisi, F. (2021). A model taxonomy for flood fragility and vulnerability assessment of buildings. *International Journal of Disaster Risk Reduction, 53*.
- Gao, X., Ma, X., Li, P., Yuan, F., Wu, Y., & Dong, G. (2023). Nonlinear analytical solution for radiation stress of higher-order Stokes. *Ocean Engineering, 286*.
- Garratt, J. (1992). *The Atmospheric Boundary Layer.* Cambridge: Cambridge University Press.
- Gimeno, L., Sorí, R., Vázquez, M., Stojanovic, M., Algarra, I., Eiras-Barca, J., . . . Nieto, R. (2022). Extreme precipitation events. *WIREs Water*. doi:https://doi.org/10.1002/wat2.1611

- Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology, 228*, pp.113-129.
- Habib, M., O'Sullivan, J., & Salauddin, M. (2022). Prediction of Wave Overtopping Characteristics at Coastal Flood Defences Using Machine Learning Algorithms: A Systematic Rreview. *IOP Conference Series: Earth and Environmental Science.* doi:10.1088/1755-1315/1072/1/012003
- Hamid, A., Din, A., Yusof, N., Abdullah, N., Omar, A., & Abdul Khanan, M. (2019). Coastal vulnerability index development: a review. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42*, 229-235.
- Hinkel, J., & Klein, R. (2003). DINAS-COAST: developing a method and a tool for dynamic and interactive vulnerability assessment. *Loicz Newsletter, 27*, 1-4.
- Hinkel, J., Lincke, D., Vafeidis, A., Perrette, M., Nicholls, R., Tol, R., . . . Levermann, A. (2014). Coastal flood damage and adaptation costs under 21st-century sea-level rise. *Proceedings of the National Academy of Sciences, 111*(9), 3292–3297. doi:https://doi.org/10.1073/pnas.1222469111
- Hofmann, J., & Holger, S. (2021). Floodgan: Using Deep Adversarial Learning to Predict Pluvial Flooding in Real Time. *Water, 13*(16). doi:10.3390/w13162255
- Holton, J., & Hakim, G. (2012). *An Introduction to Dynamic Meteorology* (5 ed.). Amsterdam: Academic Press/Elsevier.
- Huang, L., Jing, Y., Chen, H., Zhang, L., & Liu, Y. (2022). A Regional Wind Wave Prediction Surrogate Model Based on CNN Deep Learning Network. *Applied Ocean Research, 126*. doi:10.1016/j.apor.2022.103287
- Huizinga, J., De Moel, H., & Szewczyk, W. (2017). Global flood depth-damage functions: Methodology and the database with guidelines. Seville: European Comission.
- Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 457-506. doi:https://doi.org/10.1007/s10994-021-05946-3
- Kay, A., Crooks, S., Davies, H., & Reynard, N. (2011). An assessment of the vulnerability of Scotland's river catchments and coasts to the impacts of climate change: Work Package 1 Report. NERC/Centre for Ecology & Hydrology.
- Kim, K., & Kim, Y. (2002). Mechanism of Kelvin and Rossby waves during ENSO events. *Meteorology and Atmospheric Physics, 81, 3*, 169-189.
- Kim, S., Pan, S., & Mase, H. (2019). Artificial neural network-based storm surge forecast model: Practical application to Sakai Minato, Japan. *Applied Ocean Research, 91*, 101871.
- Kim, T., & Lee, W. (2024). Prediction of Wave Runup on Beaches Using Interpretable Machine Learning. *Ocean Engineering, 297*. doi:10.1016/j.oceaneng.2024.116918
- Kingston, G. B., Maier, H. R., & Lambert, M. F. (2008). Bayesian model selection applied to artificial neural networks used for water resources modeling. *Water Resources Research, 44(4).*
- Kohno, N., Dube, S. K., Entel, M., Fakhruddin, S. H., Greenslade, D., Leroux, M. D., . . . Thuy, N. B. (2018). Recent progress in storm surge forecasting. *Tropical Cyclone Research and Review, 7(2)*, 128-139.
- Koosheh, A., Etemad-Shahidi, A., Cartwright, N., Tomlinson, R., & van Gent, M. (2021). Individual wave overtopping at coastal structures: A critical review and the existing challenges. *Applied Ocean Research, 106*. doi:https://doi.org/10.1016/j.apor.2020.102476
- Kotz, M., Levermann, A., & Wenz, L. (2022). The effect of rainfall changes on economic production. *Nature, 601*, 223–227. doi:https://doi.org/10.1038/s41586-021-04283-8
- Lee, D., & Biasutti, M. (2014). Climatology and variability of precipitation in the Twentieth-Century Reanalysis. *Journal of Climate, 27*(15), 5964–5981. doi:https://doi.org/10.1175/JCLI-D-13-00630.1

- Liu, J., Guoyou, S., & Kaige, Z. (2019). High-Precision Combined Tidal Forecasting Model. *Algorithms*. doi:10.3390/A12030065
- Luettich, Jr, R., Westerink, J., & Scheffner, N. (1992). ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts, and Estuaries. Report 1. *Theory and Methodology of ADCIRC-2DDI and ADCIRC-3DL*.
- Luettich, R., Westerink, J., & Scheffner, N. (1992). ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts, and Estuaries. Report 1. Theory and Methodology of ADCIRC-2DDI and ADCIRC-3DL.
- Mair, A. a. (2011). Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *Journal of Hydrologic Engineering, 16*, p.371.
- Mel, R., & Lionello, P. (2016). Probabilistic dressing of a storm surge prediction in the Adriatic Sea.
- Mel, R., Viero, D. P., Carniello, L., Defina, A., & D'Alpaos, L. (2014). Simplified methods for real-time prediction of storm surge uncertainty: The city of Venice case study. *Advances in Water Resources, 71*, 177-185.
- Met Office. (2003). 1 km Resolution UK Composite Rainfall Data from the Met Office Nimrod System. NCAS British Atmospheric Data Centre. Retrieved March 2023, from https://catalogue.ceda.ac.uk/uuid/27dd6ffba67f667a18c62de5c3456350
- Naeini, S., & Snaiki, R. (2024). A Physics-Informed Machine Learning Model for Time-Dependent Wave Runup Prediction. *Ocean Engineering, 295*. doi:10.1016/j.oceaneng.2024.116986
- NOAA, E. (2019). *WAVEWATCH III 6.07*. Retrieved November 2024, from https://github.com/NOAA-EMC/WW3
- Oparanji, U., Sheu, R. J., Bankhead, M., Austin, J., & Patelli, E. (2017). Robust artificial neural network for reliability and sensitivity analyses of complex non-linear systems. *Neural Networks, 96*, 80-90.
- Parker, B. (2007). Tidal analysis and prediction. NOAA NOS Center for Operational Oceanographic Products and Services. doi:http://dx.doi.org/10.25607/OBP-191
- Pendergrass, A., Knutti, R., Lehner, F., Deser, C., & Sanderson, B. (2017). Precipitation variability increases in a warmer climate. *Scientific Reports*. doi:https://doi.org/10.1038/s41598-017-17966-y
- Pinheiro, V., Naghettini, M., & Palmier, L. (2019). Uncertainty estimation in hydrodynamic modeling using Bayesian techniques. *Revista Brasileira de Recursos Hídricos*. doi:https://doi.org/10.1590/2318-0331.241920180110
- Pregnolato, M., Ford, A., Wilkinson, S., & Dawson, R. (2017). The impact of flooding on road transport: A depth-disruption function. 67-81. Transport Research- Part D: Transport and Environment.
- Price, J., Weller, R., & Schudlich, R. (1987). Wind-Driven Ocean Currents and Ekman Transport. *Science, 238*, 1534-1538.
- Qin, Y., Su, C., Chu, D., Zhang, J., & Song, J. (2023). Machine Learning for Storm Surge Prediction: A Review. *Journal of Marine Science and Engineering, 11*(9), 1729. doi:https://doi.org/10.3390/jmse11091729
- Resio, T. D., Powell, N. J., Cialone, M. A., Das, H. S., & Westerink, J. J. (2017). Quantifying impacts of forecast uncertainties on predicted storm surges. *Natural Hazards, 88*(3), 1423-1449.
- Rohli, R., & Li, C. (2021). Meteorology for Coastal Scientists. *1*. Cham: Springer International Publishing AG. doi:10.1007/978-3-030-73093-2
- Rutten, J., Torres-Freyermuth, A., & Puleo, J. (2021). Uncertainty in runup predictions on natural beaches using XBeach nonhydrostatic. *Coastal Engineering, 166*. doi:https://doi.org/10.1016/j.coastaleng.2021.103869.

- Safaei-Moghadam, A., Hosseinzadeh, A., & Minsker, B. (2024). Predicting real-time roadway pluvial flood risk: A hybrid machine learning approach coupling a graph-based flood spreading model, historical vulnerabilities, and Waze data. *Journal of Hydrology*.
- Salighehdar, A., Ye, Z., Liu, M., Ionut, F., & Blumberg, A. F. (2017). Ensemble-Based Storm Surge Forecasting Models. *Weather and Forecasting, 32(5)*, 1921-1936.
- Salman, A. G., Heryadi, Y., Abdurahman, E., & Suparta, W. (2018). Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting. *Procedia Computer Science, 135*, 89-98.
- Shejule, P., & Pekkat, S. (2022). A systematic quantitative review on the performance of some of the recent short-term rainfall forecasting techniques. *Journal of Water and Climate Change, 13*(8), 3004–3026. doi:https://doi.org/10.2166/wcc.2022.302
- Stockdon, H. F., Holman, R. A., Howd, P. A., & Sallenger Jr, A. H. (2006). Empirical parameterization of setup, swash, and runup. *Coastal Engineering, 53*, 573-588.
- Stockdon, H., Long, J., Palmsten, M., Van der Westhuysen, A., Doran, K., & Snell, R. (2023). Operational forecasts of wave-driven water levels and coastal hazards for US Gulf and Atlantic coasts. *Communications Earth & Environment, 4*(1). doi:https://doi.org/10.1038/s43247-023-00817-2
- Stokes, K., Poate, T., Masselink, G., King, E., Saulter, A., & Ely, N. (2021). Forecasting coastal overtopping at engineered and naturally defended coastlines. *Coastal Engineering, 164*. doi:https://doi.org/10.1016/j.coastaleng.2020.103827.
- Stull, R. (1988). An Introduction to Boundary Layer Meteorology.
- Sun, X., Renard, B., Thyer, M., Westra, S., & Lang, M. (2015). A global analysis of the asymmetric effect of ENSO on extreme precipitation. *Journal of Hydrology, 530*, 51-65. doi:https://doi.org/10.1016/j.jhydrol.2015.09.016
- Taflanidis, A. A., Jia, G., Kennedy, A. B., & Smith, J. M. (2013). Implementation/optimization of moving least squares response surfaces for approximation of hurricane/storm surge and wave responses. *Natural Hazards, 66(2)*, pp.955-983.
- Taylor, A., & Glahn, B. (2008). Probablistic guidance for hurricane storm surge. .
- Taylor, N. R., Irish, J. L., Udoh, I. E., Bilskie, M. V., & Hagen, S. C. (2015). Development and uncertainty quantification of hurricane surge response functions for hazard assessment in coastal bays. *Nat Hazards, 77*, 1103-1123.
- Thomas, T., & Dwarakish, G. (2015). Numerical Wave Modelling – A Review. *Aquatic Procedia, 4*. doi:https://doi.org/10.1016/j.aqpro.2015.02.059
- Tiggeloven, T. C. (2021). Exploring deep learning capabilities for surge predictions in coastal areas. *Scientific reports(1), 11*, 1-15.
- Tolman, H. (2009). User manual and system documentation of WAVEWATCH III TM version 3.14. *Technical note, MMAB contribution, 220*.
- Tolo, S., Tian, X., Bausch, N., Becerra, V., Santosh, T. V., Vinod, G., & Patelli, E. (2018). Robust on-line diagnosis tool for the early accident detection in nuclear power plants. *Reliability Engineering & System Safety, 186*, 110-119.
- USACE, U. (2003). *Coastal Engineering Manual (CEM)* (Vol. Part 2).
- van der Westhuysen, A., Padilla-Hernandez, R., Santos, P., Gibbs, A., Gaer, D., Nicolini, T., . . . Tolman, H. (2013). Development of the Nearshore Wave Prediction System. NOAA/NCEP/EMC.
- Vesuviano, G. (2022). *The FEH22 rainfall depthduration-frequency (DDF).* UK Centre for Ecology and Hydrology.
- Wang, B. (2002). Kelvin waves. Encyclopedia of atmospheric sciences.
- Wang, B. (2002). Kelvin Waves. In *Encyclopedia of atmospheric sciences* (p. 1062).
- Wang, B., Lu, J., Luo, H., Li, T., Zheng, Y., & Zhang, G. (2019). Deep Uncertainty Quantification: A Machine Learning Approach for Weather Forecasting. *Proceedings of the*

25th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD '19). doi:https://doi.org/10.1145/3292500.3330704

- Wang, T., Yang, Z., Wu, W., & Grear, M. (2018). A sensitivity analysis of the wind forcing effect on the accuracy of large-wave hindcasting. *Journal of Marine Science and Engineering, 6*(4), 146. doi:https://doi.org/10.3390/jmse6040146
- Winter, M., Shearer, B., Palmer, D., Peeling, D., Harmer, C., & Sharpe, J. (2016). The economic impact of landslides and floods on the road network. *Procedia Engineering(143)*, 1425-1434.
- Xu, T., & Liang, F. (2021). Machine learning for hydrologic sciences: An introductory overview. *WIREs Water, 8*(5). doi:https://doi.org/10.1002/wat2.1533
- Yeganeh-Bakhtiary, A., EyvazOghli, H., Shabakhty, N., & Abolfathi, S. (2023). Machine Learning Prediction of Wave Characteristics: Comparison Between Semi-Empirical Approaches and DT Model. *Ocean Engineering, 286*. doi:10.1016/j.oceaneng.2023.115583
- Yu, T., & Hong, Z. (2020). Hyper-parameter optimization: A review of algorithms and applications.
- Zhang, S., Yang, Z., Zhang, Y., Zhao, S., Wu, J., Wang, C., . . . Li, S. (2023). Improved Prediction of Local Significant Wave Height by Considering the Memory of Past Winds. *Water Resources Research, 59*(8). doi:10.1029/2023WR034974
- Zhang, Z., Yin, J., Wang, N., & Hu, J. (2017). A precise tidal prediction mechanism based on the combination of harmonic analysis and adaptive network-based fuzzy inference system model. *Acta Oceanologica Sinica, 36(11)*, 94-105. doi:https://doi.org/10.1007/s13131-017-1140-x
- Zhao, Q., Liu, Y., Yao, W., & Yao, Y. (2022). Hourly Rainfall Forecast Model Using Supervised Learning Algorithm. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1-9. doi:10.1109/TGRS.2021.3054582
- Zhu, J., Yang, M., & Ren, Z. (2023). Machine Learning in Environmental Research: Common Pitfalls and Best Practices. *Environmental Science & Technology, 57*(46). doi:https://doi.org/10.1021/acs.est.3c00026
- Zounemat-Kermani, M., Batelaan, O., Fijani, E., & Hinkelmann, R. (2021). Ensembles of machine learning algorithms for hydrological modeling. *Journal of Hydrology*. doi:https://doi.org/10.1016/j.jhydrol.2021.126266

Chapter 3
Probabilistic Processes

The central idea of this thesis is to develop methods for the propagation of uncertainty through hydrological and hydrodynamic models. One of the novel and innovative aspects of this thesis is the development of a Monte Carlo Bayesian Averaging Algorithm and its application to surge and coastal inundation forecasting problems. The development of these methods is central to the two journal papers associated with this thesis. The second novel aspect is the development of uncertainty based fragility curves for relating road disruption events to extreme precipitation. The methodology for this process was developed for a Scottish Roads Research Board technical report.

Accurate forecasting of hydrological and coastal hazards requires more than just precise numerical outputs—it requires a quantifiable understanding of uncertainty in those predictions. This is especially critical when predictions feed into risk frameworks, where uncertainty in the hazard term directly influences downstream decisions. Within the context of the fundamental risk equation introduced in Chapter 1, this section focuses on improving the estimation of $P(Hz)$, the probability of the hazard occurring. To address this, we implement a probabilistic modelling framework built upon Bayesian Model Selection (BMS). These algorithms offer a principled approach to uncertainty quantification by considering the full distribution of model outcomes rather than relying on single deterministic outputs. In particular, we develop and apply an Adaptive Bayesian Model Selection (ABMS) method for ensemble predictions and extend it using Monte Carlo simulations to incorporate aleatoric and epistemic uncertainty. This facilitates not only more reliable surge forecasting, but also robust estimation of prediction intervals critical for early warning systems.

3-1. Bayesian Model Selection Algorithms

Bayesian model selection (BMS) algorithms such as that proposed by Oparanji et al. (2017), are powerful tools for evaluating the likelihood of competing model predictions. BMS algorithms offer several advantages over traditional ‘best’ model selection methods, particularly in the context of probabilistic modelling and decision-making under uncertainty. Unlike traditional model selection methods that rely solely on point estimates, Bayesian approaches consider the full probability distribution of model parameters, leading to a more comprehensive assessment of model uncertainty. They quantify uncertainty in model selection, providing posterior probabilities that indicate the relative plausibility of competing models rather than relying on binary accept-or-reject decisions. This probabilistic framework allows for better integration of uncertainty into subsequent predictions and decision-making processes. Additionally, Bayesian model selection facilitates model averaging, where predictions are weighted based on the probability of different models being correct, leading to more robust and accurate forecasting. Furthermore, Bayesian methods can be applied even when data is sparse or noisy, as they systematically update beliefs as more evidence becomes available. The justification for model selection is well made in Chapter.

3-1.1. Adaptive Bayesian Model Selection Algorithm (ABMS)

Traditional Bayesian averaging methodologies for ensemble surge forecasting e.g. (Salighehdar, Ye, Liu, Ionut, & Blumberg, 2017) explore the use of Bayesian model selection techniques based on approximating the posterior probability of a particular model to either identify the 'best' network in the set or to average the predictions of the individual models within the set. This differs from this methodology in that the network architecture of the ANN is assumed fixed, and the ANN is trained multiple times producing a set of different performing networks. By so doing, the posterior probability for each network prediction in the set can be thought of as the degree of belief that its given prediction is 'true', and the variance across the set quantifies the level of uncertainty of the model itself (Oparanji, Sheu, Bankhead, Austin, & Patelli, 2017).

During the training of a neural network model for complex problems, the ANN weights will most likely be optimised to local minima meaning that the likelihood that the model will arrive at the global optimal is limited. Hence, instead of making predictions with a single network, a group of networks is used to make predictions with different models performing to different degrees for different parts of the target range. The Bayesian model selection process uses the posterior probability calculated using probability points to identify the network that is most likely to be correct. The process then uses the variance of all the predictions to create a prediction interval and adjustment factor that form the robust prediction.

The Bayesian model selection process uses the posterior probability calculated using probability points to identify the best trained network for a given observation. The process then uses the variance of all the predictions to create a prediction interval and adjustment factor that form the robust prediction.

A set of M neural network models N_k (k=1,2,…, M), each with a different seed, is trained over a dataset $D_{train}(x, y)$, where x represents the feature vector and y the target vector (here the target vector reduce to the scalar value of the surge). A second, independent dataset $D_{eval} = (x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)$ is defined with the same structure over n observations. For an unseen feature vector, each of the M ANNs generates response $y_k = N_k(x)$. Using Bayes' theory, the empirical posterior probability for the kth-network predicted response y_k given the evaluation data, $P(N_k, y_k|D_{eval})$, can be expressed as:

$$\boldsymbol{P(N_k, y_k|D_{eval}) = \frac{P(D_{eval}|N_k) \bullet P(N_k, y_k)}{P(D_{eval})}} \qquad \text{Eq. [3]}$$

where $P(N_k, y_k)$ denotes the prior probability (i.e. assumed probability distribution before evidence) of the N_k model, $P(D_{eval}|N_k)$ denotes the likelihood term (i.e. the probability for the sample data given the kth-network predicted response), and $P(D_{eval})$ is the evidence, which can be expressed as:

$$\boldsymbol{P(D_{eval}) = \sum_{k=1}^{M} P(D_{eval}|N_k) \bullet P(N_k, y_k)} \qquad \text{Eq. [4]}$$

Since the ANNs only differ for the seed number used to initialise the weights, there is no difference in terms of the individual ANN credibility and therefore the same prior probability is assigned to the various networks, i.e. $P(N_k) = 1/\mathrm{M}$. For any N_k the true target value y in the evaluation dataset can be written as $y = y_k + \varepsilon_k$, with ε_k as the corresponding prediction error. It is assumed that the error ε_k follows a normal distribution with zero mean and variance σ_k^2, i.e. $\varepsilon_k = N(0, \sigma_k^2)$. To support this assumption, it is desirable to select networks with minimal bias along the length of the target range. The variance of network N_k can be estimated from the maximum likelihood estimation for n independent observations (Kleinbaum & Klein, 2010):

$$\boldsymbol{\sigma_k^2 = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_{i,k}^2} \qquad \text{Eq. [5]}$$

Where σ_k^2 represents the intrinsic variability in the prediction errors of N_k, and assumes that the model structure is appropriate and that all relevant variables have been included. The likelihood function for true target response y given Nk can be approximated as:

$$\boldsymbol{P(y|N_k) \approx \frac{1}{\sqrt{2\pi\sigma_k^2}} \cdot e^{\frac{-(y-y_k)^2}{2\sigma_k^2}}} \qquad \text{Eq. [6]}$$

Heteroscedastic variance can also be used by calculating the variance associated with different magnitudes of prediction by splitting the outputs into different bins along the target range. In this case, Eq. [5] is adapted so that all parameters become bin dependent e.g. σ_k^2 is replaced by lth bin variance $\sigma_{l,k}^2$.

Denoting y_r as the prediction of the network with the highest posterior probability in the set given the evaluation data, i.e.:

$$\boldsymbol{r = \underset{k}{\mathrm{argmax}}\{P(N_k, y_k \mid D_{eval})\}, k = 1, 2, \dots, M} \qquad \text{Eq. [7]}$$

the Bayesian averaged prediction y_{abms}, is expressed as:

$$\boldsymbol{y_{abms} = y_r + \sum_{k=1}^{M} P(N_k, y_k \mid D_{eval}) \cdot F_k} \qquad \text{Eq. [8]}$$

The second term is a weighted adjustment factor, with F_k denoting the difference between network response y_k, and best network response y_r:

$$\boldsymbol{F_k = y_k - y_r} \qquad \text{Eq. [9]}$$

The variance of the robust response is then evaluated as:

$$\boldsymbol{V(y_{abms}) = \textstyle\sum_{k=1}^{M} P(N_k, y_k \mid D_{eval}) \cdot (y_k - y_{abms})^2} \qquad \text{Eq. [10]}$$

The main idea presented here is that the importance of each network's information is proportional to the quality of the prediction which varies across the range of target response values. The upper and lower bounds of the, e.g., 95% prediction interval $(\underline{y}_{abms}, \overline{y}_{abms})$, complete the robust prediction:

$$\boldsymbol{\overline{y}_{abms} = y_{abms} + 1.96\sqrt{V(y_{abms})}} \qquad \text{Eq. [11]}$$

$$\boldsymbol{\underline{y}_{abms} = y_{abms} - 1.96\sqrt{V(y_{abms})}} \qquad \text{Eq. [12]}$$

As stated above, the posterior probability of each network in the set is thought of as the degree of belief that its given prediction is 'true', thereby quantifying the level of uncertainty of the ensemble (Oparanji, Sheu, Bankhead, Austin, & Patelli, 2017). This paper will refer to this ensemble variance as network uncertainty. Unlike the approach by Oparanji et al. (2017), who derive posterior probabilities directly from the training data, the methodology presented here utilises evaluation data independent from the training data. This approach ensures that the probabilities reflect the model's performance on unseen data, providing a more accurate representation of uncertainty in real operational scenarios. Additionally, this methodology allows for the integration of both feature uncertainty, (i.e. inadequate feature selection and incomplete coverage of features across all physically possible scenarios) and network uncertainty. Relying solely on one type of uncertainty could lead to an underestimate in surge prediction uncertainty, leading to narrower prediction intervals than appropriate for real world, real time applications. By taking this alternative approach and deriving feature uncertainty and network uncertainty from distinct sources, the assumption of independence can be maintained when combining these uncertainties. By assuming that feature uncertainty is independent of network uncertainty, Eq. [10]

$$\boldsymbol{V(y_{abms}) = \sum_{k=1}^{M} P(N_k, y_k \mid D_{eval}) \cdot (y_k - y_{abms})^2} \qquad \text{Eq. [10]}$$

is adjusted to include the feature uncertainty for the kth network, a_k:

$$\boldsymbol{V(y_{abms}) = \sum_{k=1}^{M} P(N_k, y_k | D_{eval}) \cdot \{(y_k - y_{abms})^2 + a_k\}} \qquad \text{Eq. [13]}$$

In this model, a_k represents the error variance of the trained model when evaluating its own training data. This error is assessed in percentile bins across the target range, anticipating more feature error at the extremes where the data is sparser, and the targets are more challenging to predict. a_k indicates the strength of relationship between the features and targets. Specifically, when a_k is lower, it suggests a stronger and more predictable connection between the input features and the output targets. This indicates that the model is effectively learning from the training data and is capable of accurately predicting or estimating the targets based on the inputs it receives. In practical terms, the assumption of independence means that systematic error captured by the training errors from D_{train} does not influence the random variability of the prediction errors captured by σ_k^2 from D_{eval}.

3-1.2. Monte Carlo Bayesian Averaging Algorithm (MCBA)

While the ABMS method is used to quantify the epistemic uncertainty raised from the ANN models, it does not account for the forecast uncertainty. Here the ABMS algorithm is extended to include the effect of this source of uncertainty with a Monte Carlo based approach. Monte Carlo sampling is a powerful tool for simulation and estimating the behaviour of complex systems and propagating aleatoric uncertainty from input parameters. By generating a large number of realisations of input parameters and evaluating the response of the model under investigation, this approach can provide accurate, flexible, and robust estimates of statistical quantities.

The number of Monte Carlo samples used varies depending on the problem context, representing a trade-off between computational cost and the stability of the resulting uncertainty estimates—larger sample sizes generally yield more reliable predictions but require greater computational resources. To optimize efficiency with respect to ANNs, all Monte Carlo samples can be evaluated in batch by the ANN in a single function call. This design dramatically reduces computation time by avoiding repeated network invocations. Memory usage scales linearly with the number of samples and features, making the method scalable and practical for operational forecasting systems.

is trained on reanalysis data and run using forecast data. This is to allow the expected uncertainty in the forecast to be propagated to the surge prediction uncertainty. To quantify the error between the reanalysis data and forecast data, the IFS data must be mapped to the same domain as the reanalysis data. This is done by normalising the IFS forecast inputs and multiplying them with the principal component coefficients for the reanalysis data. The 'true' reanalysis value of the zth forecast variable $X_{z,T}$ can be expressed as the sum of an unbiased forecasted variable $\tilde{X}_{z,forecast}$ and a zero mean forecast error $\varepsilon(0,\sigma)$:

$$\boldsymbol{X_{z,T} = \tilde{X}_{z,forecast} + \varepsilon(0,\sigma)} \qquad \text{Eq. [14]}$$

The expected value of $X_{z,T}$, $E[X_{z,T}]$, can be expressed in terms of the expected value of the biased forecast $E[X_{z,forecast}]$, adjusted with regression coefficients α_z and β_z:

$$\boldsymbol{E[PC_{z,T}] = \alpha_z \cdot E[PC_{z,forecast}] + \beta_z} \qquad \text{Eq. [15]}$$

Since the expected value of an unbiased forecast and the expected value of the true value are necessarily equal, $\tilde{X}_{z,forecast}$ can substituted in Eq. [14] for the biased forecast with applied regression coefficients:

$$\boldsymbol{X_{z,T} = \alpha_z \cdot X_{z,forecast} + \beta_z + \varepsilon(0,\sigma_z)} \qquad \text{Eq. [16]}$$

Hence $PC_{k,R}$ can be expressed in terms of the forecast with an error that can be propagated through the model via Monte Carlo sampling. This process is applied to Z forecast features and a set of error distributions $\epsilon = \{\varepsilon(0,\sigma_1),\dots,\varepsilon(0,\sigma_z)\dots,\varepsilon(0,\sigma_Z)\}$ is obtained. N Monte Carlo samples are produced with jth Monte Carlo feature vector x_j expressed as:

$$\boldsymbol{x_j = x + \epsilon_j} \qquad \text{Eq. [17]}$$

where ϵ_j collects the feature specific realisations from the error distribution set. For each jth sample, N_k generates a response:

$$\boldsymbol{y_{j,k} = N_k(x_j)\ for\ k = 1,2,\dots,M} \qquad \text{Eq. [18]}$$

A Bayesian averaged prediction $y_{abms,j}$ can be determined for each sample by applying Eq. [7] and Eq. [8]. The MCBA averaged prediction y_{mcba} is then taken as the expected value of all $y_{abms,j}$ across N Monte Carlo samples:

$$\boldsymbol{y_{mcba} = \frac{1}{N}\sum_{j=1}^{N}(y_{abms,j})} \qquad \text{Eq. [19]}$$

The variance of the MCBA robust response $V(y_{mcba})$ is then evaluated as:

$$\boldsymbol{V(y_{mcba}) = \frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{M} P(N_k, y_{j,k}|D_{eval}) \cdot (y_{j,k} - y_{mcba})^2} \qquad \text{Eq. [20]}$$

Similar to Eq. [13], $V(y_{mcba})$ can be expanded to include feature uncertainty, a_k as:

$$V(y_{mcba}) = \frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{M} P\left(N_k, y_{j,k}\middle|D_{eval}\right) \cdot \left\{\left(y_{j,k} - y_{mcba}\right)^2 + a_k\right\} \qquad \text{Eq. [21]}$$

Hence the feature, network and forecast uncertainty are propagated to the final prediction interval. The upper and lower bound of this 95% prediction interval:

$$\overline{y}_{mcba} = y_{mcba} + 1.96\sqrt{V(y_{mcba})} \qquad \text{Eq. [22]}$$

$$\underline{y}_{mcba} = y_{mcba} - 1.96\sqrt{V(y_{mcba})} \qquad \text{Eq. [23]}$$

3-1.3. Time Dependent Uncertainties

In the event where the forecasting model has a lag between the feature and target, as is the case with wind waves where the height of a wave is dependent on wind action across the lifetime of the wave, the error structure must be adjusted to account for temporally and spatial correlated errors. Hence the value of the kth feature at time t, $X_{k,t}$ can be expressed in terms of the corresponding forecast $X_{forecast,k,t}$, using regression coefficients $\alpha_{k,t}$ and $\beta_{k,t}$ with associated zero mean random error $\varepsilon_{k,t}(0,\sigma)$:

$$X_{k,t} = \alpha_{k,t} \cdot X_{forecast,k,t} + \beta_{k,t} + \varepsilon_{k,t}(0,\sigma) \qquad \text{Eq. [24]}$$

In addition to the shifting lag structure, consideration must be given to the error correlation structure. It is shown in Chapter 5, that there are strong correlations between the errors of successive lead times of the same features. Hence an assumption of independence for the randomly generated variable may not be adequate for forecasting. As such, it is necessary to correlate the randomly generated forecast errors using Cholesky decomposition. Let the relationships between time dependent principal component errors be described by covariance matrix Σ:

$$\Sigma = \begin{bmatrix} var(\varepsilon_{1,1}) & \cdots & cov(\varepsilon_{1,1}, \varepsilon_{k,t}) \\ \vdots & \ddots & \vdots \\ cov(\varepsilon_{k,t}, \varepsilon_{1,1}) & \cdots & var(\varepsilon_{k,t}) \end{bmatrix} \qquad \text{Eq. [25]}$$

If covariance matrix Σ is positive definite then it can be expressed through Cholesky decomposition as the product of a lower triangular matrix or Cholesky matrix L and its transpose Ł^T:

$$\Sigma = \text{Ł}\text{Ł}^T \qquad \text{Eq. [26]}$$

A matrix E_{ind} consisting of columns of independently generated random errors ε that follow a given distribution $\text{ɗ}(0,1)$, with zero mean and unit standard deviation, is created with one column for each principal component k at timestep t:

$$E_{ind} = \left[\varepsilon_{1,1}, \varepsilon_{1,2}, \ldots, \varepsilon_{k,t}\right] \sim \text{ɗ}(0,1) \qquad \text{Eq. [27]}$$

This error matrix is converted to a correlated error matrix E_{corr} through matrix multiplication of the Cholesky matrix and E_{ind}:

$$E_{corr} = \text{Ł}E_{ind} \qquad \text{Eq. [28]}$$

Where E_{corr} have zero mean and standard deviations equal to the square root of the diagonal of Σ, whilst exhibiting cross-feature and temporal correlation structures. Hence, the forecast inputs can be transformed to the range suitable to make unbiased predictions, while providing a time dependent forecast error to be used by the MCBA algorithm.

3-2. Uncertainty Based Fragility Curves

Having established a method for quantifying the probability of hazard occurrence, $P(Hz)$, the next step in risk estimation is to assess how that hazard translates into actual damage or disruption. In the risk equation Eq. [1], this corresponds to the conditional probability of damage or disruption given the hazard, P(D|Hz), which represents the vulnerability—often expressed through fragility curves. The following section addresses this second component by presenting a probabilistic method for constructing uncertainty-based fragility curves, which describe the likelihood of system failure as a function of hazard intensity (e.g., precipitation depth or flood level). The approach combines empirical observations with Monte Carlo sampling to represent uncertainty in the intensity measure, and applies surrogate modelling techniques, such as kriging, to estimate spatial variability where direct measurements are not available. This enables a structured assessment of how vulnerable a system is under a range of extreme event conditions, completing the link between hazard forecasting and consequence evaluation within the broader risk framework.

The process for developing a fragility curve for assessing the vulnerability of a critical structure to failure, involves analysing the statistical relationship between the intensity measure I and the performance of the asset (whether it fails or not). It is commonplace to group I into bins, and the conditional probability of failure for each bin is calculated by dividing the number of failures in that bin by the total number of events. Fragility is a concept used in reliability engineering to describe the probability of failure of a system or component, given a specific set of conditions or events. It is also known as the conditional failure probability or the conditional probability of non-performance. In reliability engineering, a system or component is considered to have failed if it is unable to perform its intended function or meets some other specified criterion for failure. The probability is a rule that assigns a number between 0 and 1 to a given event, with zero being non chance occurrence and 1 being a certain change of occurrence. The probability of failure is denoted as $P(F)$:

$$\boldsymbol{P(F) = \frac{N_F}{N_o}} \qquad \text{Eq. [29]}$$

Where N_F is the number of failures and N_o is the number of observations. The probability of an event of given intensity x occurring is denoted as $P(I = x)$:

$$\boldsymbol{P(I = x) = \frac{N_{I=x}}{N_o}} \qquad \text{Eq. [30]}$$

Where $N_{I=x}$ is the number of events with a given intensity x. The conditional probability of failure is calculated by considering the probability of failure given certain conditions. The conditional probability of failure given a certain intensity level $P(F|I = x)$, is described mathematically as, (Rubinstein, 2017):

$$\boldsymbol{P(F|I = x) = \frac{P(F \cap I = x)}{P(I = x)}} \qquad \text{Eq. [31]}$$

Where $P(F \cap I = x)$ is the intersection between failure events and events of a given intensity x:

$$\boldsymbol{P(F \cap I = x) = \frac{N_{F.I=x}}{N_o}} \qquad \text{Eq. [32]}$$

Where $N_{F,I=x}$ is the number of failures when I = x. Substituting Eq. [32] and Eq. [30] into Eq. [31] and eliminating N_o yields:

$$\boldsymbol{P(F|I=x)=\frac{N_{F,I=x}}{N_{I=x}}} \qquad \text{Eq. [33]}$$

When this formula is applied to discretised intensity level values, a set of points describing the conditional probability of failure can be produced, Figure 12.

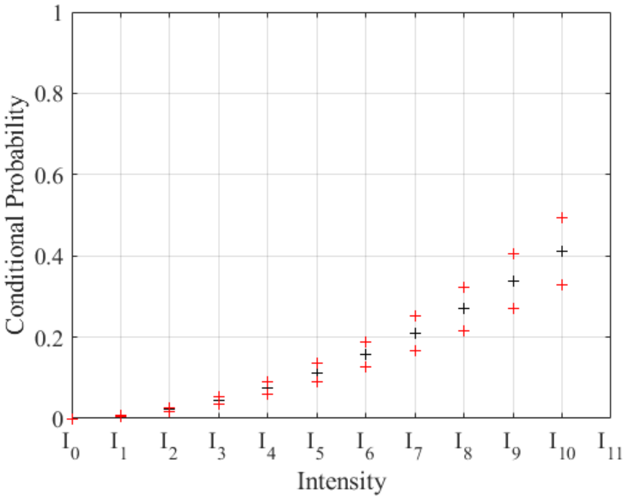


Figure 12: Conditional probabilities for arbitrary intensity measure with regularly spaced intensities showing upper and lower limit of ±20%.

It is likely that the true intensity measure value at the location of a failure event is unknown. To this end it is necessary to obtain an estimate of the intensity measure with a given level of uncertainty on the estimate. Monte Carlo methods are a set of computational techniques used to simulate and analyse complex systems or processes. They model the behaviour of a system by extracting random samples from a probability distribution that represents the input parameters of the system being modelled. The true value of intensity measure I can be expressed as the sum of expected value intensity measure $E[I]$ and zero mean error $\varepsilon(0,\sigma)$:

$$\boldsymbol{I=E[I]+\varepsilon(0,\sigma)} \qquad \text{Eq. [34]}$$

With the Monte Carlo method, n random samples are drawn from this error distribution for each point and applied to the location and timestep mean prediction. Each Monte Carlo sample I_{MC} is then treated in turn and the conditional probability of failure for the n-th Monte Carlo sample $P(F|I_{MC,n}=x)$ by modifying **Error! Reference source not found.** to:

$$\boldsymbol{P(F|I_{MC,n}=x)=\frac{N_{F,I_{MC,n}=x}}{N_{I_{MC,n}=x}}} \qquad \text{Eq. [35]}$$

With the expected value of the conditional probability for a given intensity over n Monte Carlo samples equal to the mean:

$$\boldsymbol{E[P(F|I_{MC}=x)]=\frac{1}{n}\sum_{m=1}^{n}P(F|I_{MC,m}=x)} \qquad \text{Eq. [36]}$$

And variance:

$$\boldsymbol{Var\big(P(F|I_{MC}=x)\big) = E[(P(F|I_{MC}=x) - E[P(F|I_{MC}=x)])^2]} \qquad \text{Eq. [37]}$$

By generating a large number of random samples, the Monte Carlo method can provide a comprehensive picture of the possible outcomes of the system. By considering the normal distribution of probability values about the expected conditional probability value, a confidence interval can be defined by 95% z-score with upper and lower bound:

$$\boldsymbol{upper_{95\%} = E[P(F|I_{MC}=x)] + 1.96\sqrt{\frac{Var\big(P(F|I_{MC}=x)\big)}{n}}} \qquad \text{Eq. [38]}$$

$$\boldsymbol{lower_{95\%} = E[P(F|I_{MC}=x)] - 1.96\sqrt{\frac{Var\big(P(F|I_{MC}=x)\big)}{n}}} \qquad \text{Eq. [39]}$$

By utilising the upper and lower limits of the 95% confidence interval for each intensity measure, an empirical step plot can be constructed, which describes the empirical fragility curve for the given intensity measure, an example is shown in Figure 13: Example empirical fragility curve. Selecting a specific intensity value along the x-axis allows for the determination of a 95% probability interval, while choosing a probability value along the y-axis provides upper and lower limits of intensity.

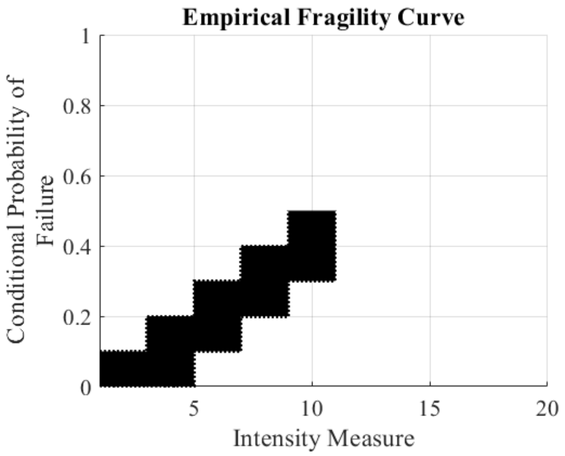


Figure 13: Example empirical fragility curve.

Additionally, given that an estimate may be desirable for unseen intensities, whilst conceding that the distribution to which these intensities belong is unknown, a probability box (P-Box) is constructed. For this the best-fit for several cumulative distribution functions is fitted through the upper and lower boundary values as shown in Figure 14. For this process it is necessary to establish a theoretical maximum rainfall intensity that guarantees failure. In the example shown, 5 two-parameter cumulative distributions (normal, logistic, lognormal, Weibull and extreme value) were fitted to the points.

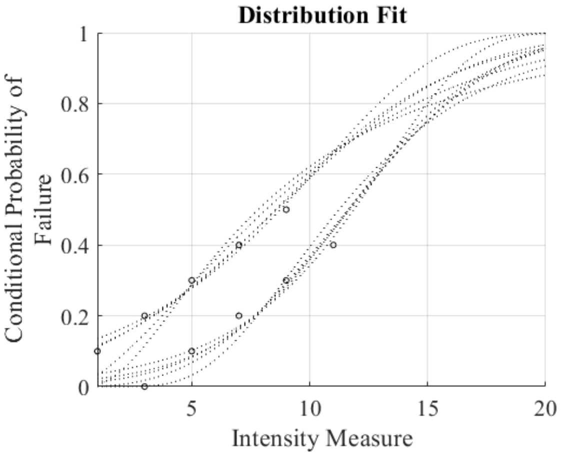


Figure 14: Example distribution fit to dummy empirical conditional probability values.

From this, a probability box is defined (Figure 15) by taking the maximum probabilities of the lower limit and the minimum probabilities of the upper limit. The idea here is that since the shape of the distribution is unknown, a probability box that describes all possible values is presented in order to inform the possible shape of the fragility curve in the unknown region, which can only be truly known through the acquisition of additional data.

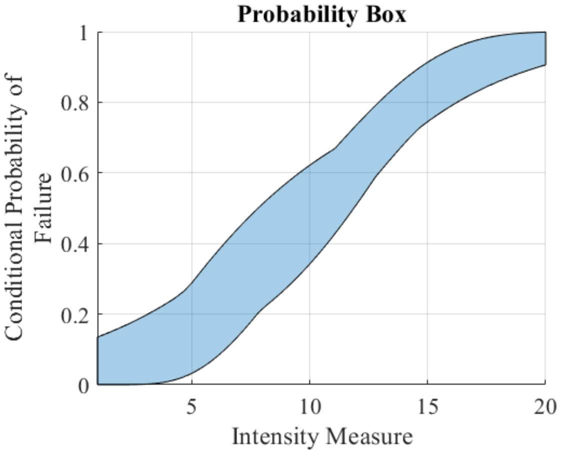


Figure 15: Example probability box for dummy distribution fits.

3-3. Error metrics

Various metrics should be employed to form a comprehensive evaluation of a surrogate model's performance. The R2 value, represented by Eq. [42], determines how well the predictions align with the S observations by expressing the proportion of the unexplained variance compared to the total variance. To compute the error approximation in the same unit as the observations, the normalized error metrics of root mean squared error (RMSE), as described by Eq. [41], and mean absolute error (MAE), specified by Eq. [40], are often used. RMSE assigns a larger weight to larger errors, whereas MAE gives equal weight to all errors. Additionally, the Bias, represented by Eq. [43] is considered as modelled time series data frequently exhibit systematic bias (Jackson, 2019). The scatter index (SI), represented by Eq. [44], is a normalized measure of error that provides more precise and accurate information about the accuracy of a numerical simulation than RMSE (Mentaschi, 2013). Eq. [45] is Pearson's correlation coefficient (CC) which is a measure of the strength and direction of a linear relationship between observed and predicted values. This provides insights into how well the predictions match the actual data.

$$MAE = \frac{\sum_{i=1}^{S} \left| y_i - y_{p,i} \right|}{S} \qquad \text{Eq. [40]}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{S} \left(y_i - y_{p,i} \right)^2}{S}} \qquad \text{Eq. [41]}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{S} \left(y_i - f(y_{p,i}) \right)^2}{\sum_{i=1}^{s} (y_i - \bar{y})^2} \qquad \text{Eq. [42]}$$

$$Bias = \frac{\sum_{i=1}^{s} (y_{p,i} - y_i)}{S} \qquad \text{Eq. [43]}$$

$$SI = \sqrt{\frac{\sum_{i=1}^{S} \left((y_i - \bar{y}) - (y_{p,i} - \bar{y}_{p,i}) \right)^2}{\sum_{i=1}^{S} y_i^2}} \qquad \text{Eq. [44]}$$

$$CC = \frac{\sum_{i=1}^{s} (y_i - \bar{y})(y_{p,i} - \bar{y}_p)}{\sqrt{\sum_{i=1}^{s} (y_i - \bar{y})^2 \sum_{i=1}^{s} (y_{p,i} - \bar{y}_p)^2}} \qquad \text{Eq. [45]}$$

In each case, yi and $y_{p,i}$ denotes the i -th observation and robust prediction values respectively, and ȳ and $\bar{y}_p$ are the mean value of the observations and predicted values. In this dissertation, the above metrics are used to give a complete comparison of model performance.

3-References

- Jackson, E. R. (2019). Introductory overview: Error metrics for hydrologic modelling–A review of common practices and an open source library to facilitate use and adoption. *Environmental modelling & software, 119*, 32-48.
- Kleinbaum, D. G., & Klein, M. (2010). Maximum Likelihood Techniques: An Overview. *Logistic Regression*, 103-127. Springer.
- Mentaschi, L. B. (2013). Problems in RMSE-based wave model validations. *Ocean Modelling(72)*, 53-58.

- Oparanji, U., Sheu, R. J., Bankhead, M., Austin, J., & Patelli, E. (2017). Robust artificial neural network for reliability and sensitivity analyses of complex non-linear systems. *Neural Networks, 96*, 80-90.
- Rubinstein, R. a. (2017). Simulation and the Monte Carlo method [internet resource]. *Third edition*. Wiley series in probability and statistics.
- Salighehdar, A., Ye, Z., Liu, M., Ionut, F., & Blumberg, A. F. (2017). Ensemble-Based Storm Surge Forecasting Models. *Weather and Forecasting, 32(5)*, 1921-1936.

Chapter 4

Surge Forecasting under Uncertainty

This chapter details the application of the ABMS and MCBA methodology to a surge forecasting problem. It is a reworking of journal paper titled “Robust Storm Surge Forecasts for Early Warning System: A Machine Learning Approach Using Monte Carlo Bayesian Model Selection Algorithm” submitted as part of this PhD work to the journal of Stochastic Environmental Research and Risk Assessment.

Within the risk framework introduced in Chapter 1, the hazard term P(Hz)—the probability that surge conditions capable of driving coastal inundation will occur—is central to credible early warning and downstream risk estimation. In mid– and high-latitude settings, projected increases in the intensity of the most severe extratropical storms heighten exposure to extreme water levels, so sharpening P(Hz) directly improves decisions that limit disruption and loss across coastal communities and critical infrastructure.

Chapter 2 highlighted why surge forecasting is hard: uncertainty accumulates from the atmosphere (track, timing, intensity), hydrodynamic model structure and parameters, and local morphology; even state-of-the-art systems (e.g., ADCIRC) face large spread near landfall and in complex estuaries. Multi-scenario approaches help, but bias and random errors in the weather and surge models can distort forecast probabilities if uncertainty is not handled explicitly.

Chapter 3 therefore developed a probabilistic ML framework for P(Hz). An ensemble of fixed-architecture ANNs is trained and combined using Adaptive Bayesian Model Selection (ABMS) so that predictions are weighted by posterior model probabilities, capturing epistemic (model) uncertainty. We then propagate aleatoric (forecast-input) uncertainty via a Monte Carlo Bayesian Averaging (MCBA) layer that samples debiased meteorological drivers and carries their error through to surge predictions; time-dependent correlations in forecast errors are preserved using covariance-aware sampling (Cholesky), yielding calibrated prediction intervals across lead times. Together these elements provide operational surge forecasts with transparent attribution of feature, network, and forecast uncertainty.

In this chapter we apply that framework to a UK case study, using reanalysis for training and operational forecasts for evaluation to demonstrate reliability of intervals and robustness under real-time conditions. The result is a defensible estimate of P(Hz) that can be integrated with vulnerability and consequence models from Chapter 1’s risk equation to support coastal warning, response, and planning.

4-1. Robust ANN-based Surge Forecasting System

An increase in the intensity and frequency of storms goes hand in hand with an increase in the intensity of storm surges, which contribute to coastal flooding (Emanuel, 2017). In addition, warmer oceans are providing more energy to fuel these storms, making them more dangerous and unpredictable (IPCC, Special Report on the Ocean and Cryosphere in a Changing Climate, 2019). In order to enable stakeholders to take optimal risk-mitigation decisions, early warning systems must account for these uncertainties in the predictions.

The robust ANN-based surge forecasting system (RSFS) schematic is shown in Figure 16, describing the steps involved in converting a 24-hour weather forecast to a 24-hour surge height prediction. The pre-processing and input uncertainty evaluation described in this Chapter are written in Python as modular functions for ease of operation. A Long Short-Term Memory (LSTM) model ensemble is presented in this chapter as a robust surge early warning system. The early warning system utilises a Bayesian Model Selection algorithm, which assigns probabilistic weights to different LSTM models based on their predictive performance and inherent uncertainty. The early warning system considers model, feature, and forecast uncertainty.

The steps required for ANN Set prediction, model averaging, combining uncertainties and producing the robust surge height prediction are detailed in the next sub-sections. This approach not only enhances the robustness of surge height forecasts but also provides a probabilistic framework that quantifies uncertainty, offering more reliable insights for decision-making.

4-1.1. Input Selection

An important issue in developing surrogate models for storm surge prediction is the choice of the relevant set of input variables, and the characterisation and propagation of the uncertainty inherent to these. A careful choice of the input variables can simplify the modelling process and reduce the computational burden, making it more efficient and less prone to overfitting. Second, it can help to identify the most important variables and their relationships, leading to more interpretable models and better insights into the underlying system.

The behaviour of storms is influenced by Rossby and Kelvin waves. Rossby waves are the large-scale dynamical response of the ocean to atmospheric conditions (Chelton & Schlax, 1996). Their behaviour is responsible for pressure and windspeed variations that can drive storm surges, particularly in mid-latitudes. Kelvin waves, on the other hand, are a type of gravity wave that can rapidly propagate sea level changes along coastlines, contributing to the intensity and reach of storm surges (Wang B. , Kelvin Waves, 2002). In this framework, we consider the wind velocity components in the eastward and northward directions at 10 meters above the ground, referred to as U10 and V10 respectively, and mean sea level pressure (MSLP) across the input grid. Additionally, the U10, V10, and the pressure difference at the target location, along with the current surge level, are included in the analysis.

The selected predictors mirror what is available operationally from ECMWF (ERA-5 for training; IFS for forecasting) and target the mechanisms that drive surge in the UK: pressure gradients and low-level winds modulated by baroclinic Rossby/Kelvin wave dynamics (Sabatino, Murray, Hills, Speirs, & Heath, 2016). Gridded MSLP and 10-m wind components (U10, V10) capture the large-scale forcing over a 1000 km footprint; principal components then compress and decorrelate these fields to keep training tractable while retaining ~95% of variance. Local ΔP (site minus domain maximum), local U10/V10, and the contemporaneous surge level add the immediate, site-specific signal needed for a 24-h forecast. Together, this balances physical relevance, data availability, and computational cost for an operational pipeline.

On tide–surge interaction. Where dependence between tide and skew surge is documented, tide height (and phase) should be included explicitly. For Millport, The Department for Environment, Food and Rural Affairs (DEFRA) concluded that tide is independent of surge for the Firth of Clyde (DEFRA, 2005). As such, tide was omitted here without degrading skill. We therefore treat tide–surge coupling as site-contingent: include it where tests show dependence; omit where independence holds.

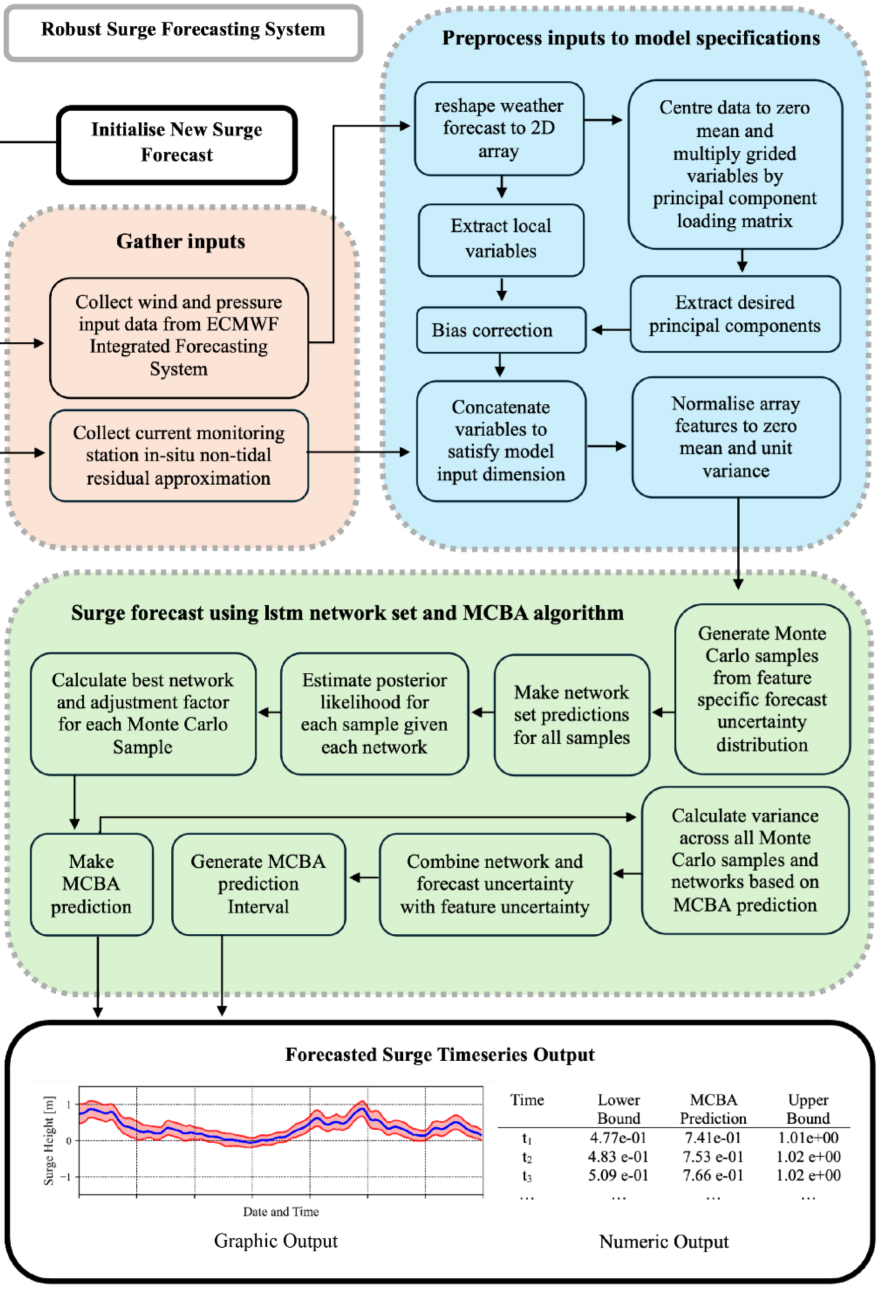


Figure 16: Operational schematic of robust surge forecasting system

On local bathymetry. The Clyde is bathymetrically complex, and bathymetry clearly influences surge transformation nearshore. In this single-site design we exploit the fact that historical gauge surges already embed local morphology, so the network learns an effective site response without ingesting static bathymetric grids. This choice keeps the model lightweight but limits portability: transferring to new sites (or resolving spatial gradients within an estuary) would benefit from adding static descriptors (e.g., depth, shelf slope, coastal orientation) or training a multi-site model that conditions on morphology.

Machine learning studies based on observational data are limited due to the sparsity of in situ data and the lack of large and well- structured datasets that are suitable for training machine learning models (Qin, Su, Chu, Zhang, & Song, A review of application of machine learning in storm surge problems, 2023). This issue can be overcome by using reanalysis data sets which have increased temporal and spatial coverage compared to in-situ measurements. The size of area over which the inputs are considered, or ‘footprint’, is significant (Tiggeloven, 2021). A larger input area can improve surge height predictions with the cost of additional computational demand for model training and run time.

The early warning surge system is trained on reanalysis data and provides real time prediction on forecast data. More specifically, first reanalysis data is used so that forecast errors are not introduced to the models during training, providing the best opportunity to find patterns between the surge response and the ‘true’ wind and pressure reanalysis data. Thereafter, the (weather) *forecast error* is characterised through comparison of historical forecast data and its concurrent reanalysis data. Then, the surge prediction, i.e. the model output, is obtained using wind speed and pressure forecasts, whilst accounting for the corresponding amount of forecast error. Through this methodology the uncertainty in the weather forecast is leveraged and added to *network uncertainty* and *feature uncertainty* to estimate the final uncertainty in the surge height prediction.

In this framework, reanalysis data for model development is taken from ERA-5 with the developed early warning system driven by European Centre for Medium-Range Weather Forecasts (ECMWF) operational forecasts. ERA-5 (C3S C. , 2017) is the fifth edition of the ECMWF atmospheric reanalysis of the global climate, providing hourly estimates of meteorological variables on a global 30km grid. This is used for model training and validation. The operational forecasts are made by the ECMWF Integrated Forecasting System (IFS) which is a global numerical model supported by a sophisticated data assimilation system that estimates the likely evolution of the weather (ECMWF, 2024).

4-1.1.1. Input Dimension Reduction

Input grids for surge models can cover extensive footprints and many of the most popular modelling options consider time-dependent variables (see Section 4-2.1.), which can imply vast input arrays with large numbers of dimensions. Input grid dimensionality can be reduced using principal component analysis (PCA). PCA is a mathematical method used for reducing the dimensionality of large data sets by extracting the most important features that explain the variability of the data. PCA has been widely and successfully applied to help understand and interpret large, spatially extensive climate datasets (Reusch, Alley, & Hewitson, 2005).

PCA is based on space transformation where the original data set is linearly transformed onto a new coordinate system represented by a set of orthogonal axes created to capture the maximum variance of the data in each subsequent dimension. The original data can then be represented in this new coordinate system using a smaller number of components able to preserve the most significant features of the original data. The result is a compressed representation of the data that captures the most important information. The principal components that explain the desired variability are kept and those deemed to represent an insufficient amount of variability are discarded. A second advantage of PCA is that principal components are independent of one another, and this removes correlation in the inputs. This is useful since highly correlated data reduces the distinctiveness of data representation and can impede ANN model training and result in models that struggle to generalise (Mohamad-Saleh & Hoyle, 2008). A downside to PCA is that principal components are just mathematical constructs that represent variance in the data and do not necessarily have an inherently physical explanation, meaning that there can be issues with feature interpretability (Reusch, Alley, & Hewitson, 2005).

The number of principal components to be selected for the model fitting is a trade-off between computational demand and preserved input variation. The principal components are normalised by subtracting the mean and dividing by the standard deviation to eliminate scale issues with ANN training. The outputs of principal component analysis include the principal components (also referred as the score), the amount of variation explained by each principal component (used to determine the desired number of principal components) and the principal component coefficients (used to transform the input data to the principal component domain and back).

4-1.1.2. Input Uncertainty Characterisation

As mentioned in Section 4-2.1, the early warning surge system is trained on reanalysis data and run using forecast data. This is to allow the expected uncertainty in the forecast to be propagated to the surge prediction uncertainty. To quantify the error between the reanalysis data and forecast data, the IFS data must be mapped to the same domain as the reanalysis data, as outlined in Section 4-2.2. This is done by normalising the IFS forecast inputs and multiplying them with the principal component coefficients for the reanalysis data. The ‘true’ reanalysis value of the k^{th} principal component $\boldsymbol{PC_{k,R}}$ can be expressed in terms of biased forecast $\boldsymbol{PC_{k,forecast}}$ with an error $\boldsymbol{\varepsilon(0,\sigma)}$ by applying Eq. [14] to Eq. [16]:

$$\boldsymbol{PC_{k,R} = \alpha_k \cdot PC_{k,forecast} + \beta_k + \varepsilon(0, \sigma_k)} \qquad \text{Eq. [46]}$$

Hence $\boldsymbol{PC_{k,R}}$ can be expressed in terms of the original forecast with an error that can be propagated through the model with Monte Carlo sampling.

In this framework we do not model ERA-5 error explicitly. Instead, we treat ERA-5 as the operational “truth” for training and map IFS forecasts into the ERA-5 principal-component (PC) space before prediction. IFS fields are normalised and projected with the ERA-5 loading matrix so that each IFS PC aligns with its ERA-5 counterpart; a linear bias correction is then fitted between concurrent IFS and ERA-5 PCs (here using year-2020 pairs). The de-biased residuals—whose distributions are logistic with heavier tails and heteroscedastic across the feature range—are sampled (bin-wise) and propagated through the network via MCBA. Thus, differences between IFS and ERA-5 are quantified and carried forward, while any ERA-5 representation error relative to the gauge is implicitly absorbed during training and reflected in the learned mapping and in the feature/network components of the prediction interval.

Because the system is anchored to ERA-5, systematic ERA-5 biases (e.g., increasing wind-speed error at high percentiles over the Atlantic (Campos, Gramcianinov, de Camargo, & da Silva Dias, 2022), may leak into the trained relationship, especially in the tail. This is visible in our ERA-5-driven tests and is one reason why feature + network uncertainty explains most interval width for the full series, with forecast uncertainty becoming most important for extremes under IFS forcing. A full reanalysis-error model (e.g., fitting ERA-5 to independent scatterometer/buoy data and sampling that error separately) was not implemented here and is left as future work.

Meteorological resolution and representativeness (ERA-5 at ~30 km). ERA-5 provides hourly global fields on a ~30 km grid. At this resolution sharp near-surface pressure gradients and narrow wind maxima associated with fronts, sting-jets, and secondary lows can be partially smoothed, reducing the amplitude of local forcing that matters for coastal surge. In our pipeline this "representation error" is handled implicitly in two ways.

Firstly, by training the network on ERA-5 (so any ERA-5–vs–gauge mismatch is absorbed into the learned mapping and appears as feature/network uncertainty), and secondly by projecting IFS forecasts into the ERA-5 PC space and sampling their de-biased residuals with a heteroscedastic, heavy-tailed error model (logistic by PC and by bin), which carries resolution and model-physics differences from forecast to prediction via MCBA. Together these steps explain why most interval width in the full series is attributable to feature and network uncertainty, while forecast uncertainty becomes decisive in the tail.

It cannot be claimed that ERA-5 fully resolves the most intense, small-scale wind features in UK extratropical storms; and smoothing is unavoidable at ~30 km and likely contributes to residual under-coverage for rare extremes.

4-1.2. ANN Model Architecture

A Bidirectional Long Short Term Memory (LSTM) architecture is used in the surge forecasting framework. Details about LSTM structure are provided in Appendix A-2. A Bidirectional LSTM is a configuration of LSTM where data is processed both in a forward and backward direction, effectively learning from sequences in a way that considers both past and future data. This dual-direction processing provides a more complete understanding of the sequence, enhancing performance in tasks where the context from both before and after a data point is important.

As concluded in Chapter 2's literature review, Long Short-Term Memory (LSTM) networks are the most suitable class for this problem because they capture long-range temporal dependencies in surge response to evolving meteorological drivers while remaining data-efficient for hourly, gauge-based targets. Accordingly, this chapter adopts a Bidirectional LSTM ensemble, with full gate-equation details provided in Appendix A-2. In brief, LSTMs augment a recurrent cell with input/forget/output gates to regulate information flow and mitigate vanishing-gradient issues; the bidirectional configuration conditions each prediction on past and (within-window) future context to improve sequence representation.

Alternative deep architectures (e.g., convolutional encoders, hybrid CNN–RNNs, or related designs explored in recent storm-surge ML studies) are not benchmarked here; a formal architecture search lies outside the scope of this thesis and is identified as future work in Section 4-4. We note that comparative avenues and broader model classes are active in the literature, to which readers are referred for context.

In this framework a Bidirectional LSTM network is designed that uses meteorological variable principal components n_{pca} from the computational grid. This network also uses additional inputs, including the pressure difference between the target location and the maximum pressure in the domain $\Delta\boldsymbol{P}_{\boldsymbol{local}}$, as well as the U10 and V10 wind components at the target location, $\boldsymbol{U10}_{\boldsymbol{local}}$ and $\boldsymbol{V10}_{\boldsymbol{local}}$. All inputs are normalised to have zero mean and unit standard deviation. They are considered across a 49-hour window, spanning 48-hour prior to the prediction hour and the prediction hour. Given that the forecast lead time is 24 hours this means the previous 24 hours are considered along with the present hour and the 24-hour forecast. In addition, the current and previous 24 hours surge heights are considered. For areas where there is a dependency between surge height and tide, it may be desirable to include tide height as well (Williams, Horsburgh, Williams, & Proctor, 2016). This model structure is shown in Figure 17.

Given the large number of trainable parameters in the model, the dense layer is accompanied by a dropout layer with a 0.1 dropout rate. Dropout layers uncouple a random group of weights between layers, preventing all neurons in a layer from synchronously optimizing their weights. This decorrelates the weights by preventing all the neurons from converging to the same targets (Labach, Salehinejad, & Valaee, 2019). Initial tests also revealed that using kernel, bias and activity regularizes reduced overfitting and provided better a trade-off in variance and bias during the training process especially for predictions in the extreme range. The dense layer contains 100 nodes.

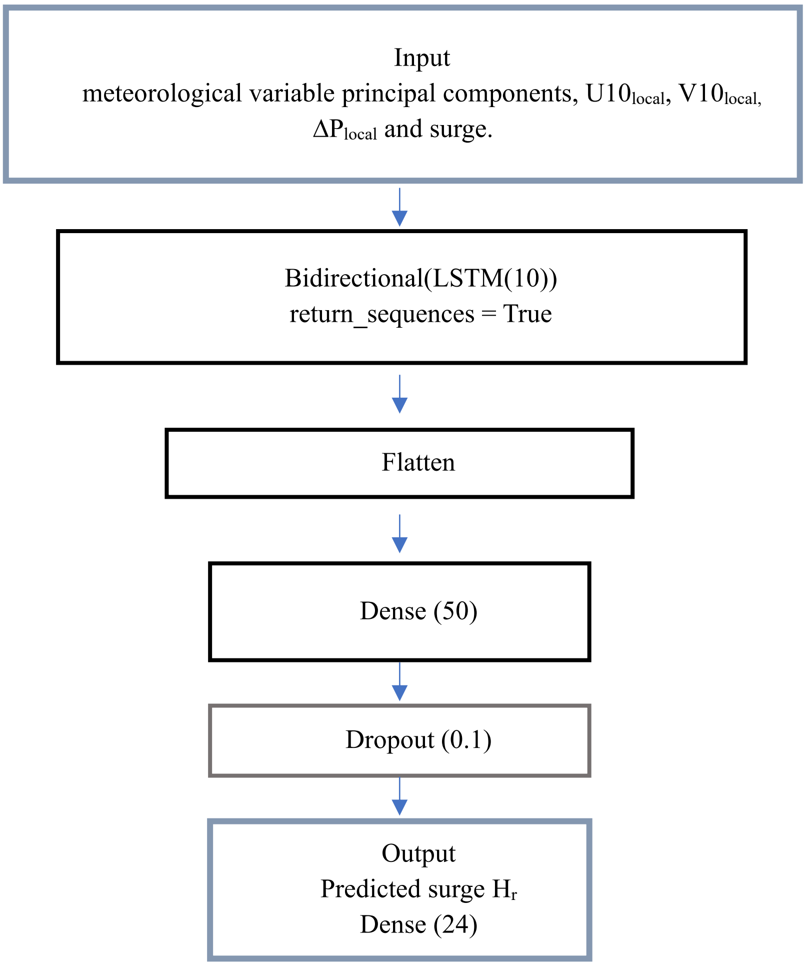


Figure 17: Architecture of LSTM based surge forecasting network.

4-2. Model Application

In this section, the forecasting model described in Section 4-2 is applied with the Bayesian Model Selection algorithms described in 3-1.1. and 3.1.2. and validated for the *Firth of Clyde* basin in southwest Scotland. This is an area that is prone to coastal inundation due to its complex bathymetry and exposure to Atlantic generated weather systems (Sabatino, Murray, Hills, Speirs, & Heath, 2016). Model predictions are compared to Millport surge observations in order to validate the model using ERA5 inputs. The model is then adapted to accommodate forecast inputs and its operational performance is assessed against Millport surge observations. All pre-processing tools, model building tools and operational functions are contained in the Bayesian Coastal Forecasting Toolbox (Macdonald, 2024).

4-2.1 Case study – Firth of Clyde

The Firth of Clyde Basin was identified by the SEPA Coastal Hazard Mapping Study (SEPA, 2015) as a vulnerable coastal area in Scotland. Along the east coastline of the Firth of Clyde, the A78 trunk road connects the ferry port of Stranraer to Glasgow and the rest of Scotland. This road link is critical for trade and connecting isolated towns and villages with emergency services and is prone to annual road closures due to overtopping. This makes the area suitable for an early warning surge forecasting system to inform responsible authorities of the likelihood of occurrence of extreme surges with sufficient warning time for them to take meaningful risk and impact mitigation actions. Figure 18 illustrates the area of interest along with the trunk road network and relevant locations.

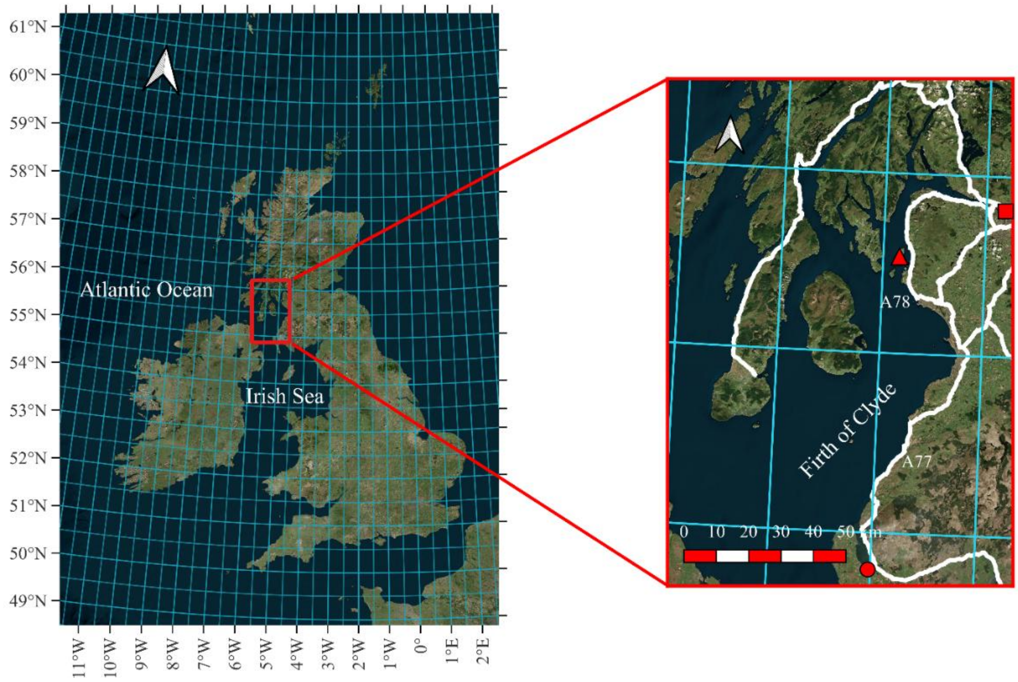


Figure 18: map of the UK (on the left) and case study location (on the right) of the Firth of Clyde showing Millport (triangle), Stranraer (circle), Glasgow (square) and trunk road network (white). The A77 and A78 link have been specifically highlighted.

The Transport Research Laboratory (TRL) report (Milne, et al., 2017) examined surge events for Millport between 1995 and 2013 stated that flooding is likely to occur on the A78 if the flood potential value (FPV) exceeded 5.3m Chart Datum (CD). The report defines the FPV as:

$$\boldsymbol{FPV = H_a + 2H_r} \qquad \text{Eq. [47]}$$

where $\boldsymbol{H_a}$ is the astronomical tide and $\boldsymbol{H_r}$ is the non-tidal residual (storm surge). This formulation was based on the assumptions that the height of waves breaking at the coast during a storm are at least as high as the storm surge and that larger surges will have higher waves breaking at the shore. Figure 19 shows the Chart Datum tidal height against surge height at Millport for the years 1980 to 2019. In this plot, the extreme value boundary (estimated as 0.75m) is established as the minimum surge height value required for the FPV value to exceed the high risk threshold.

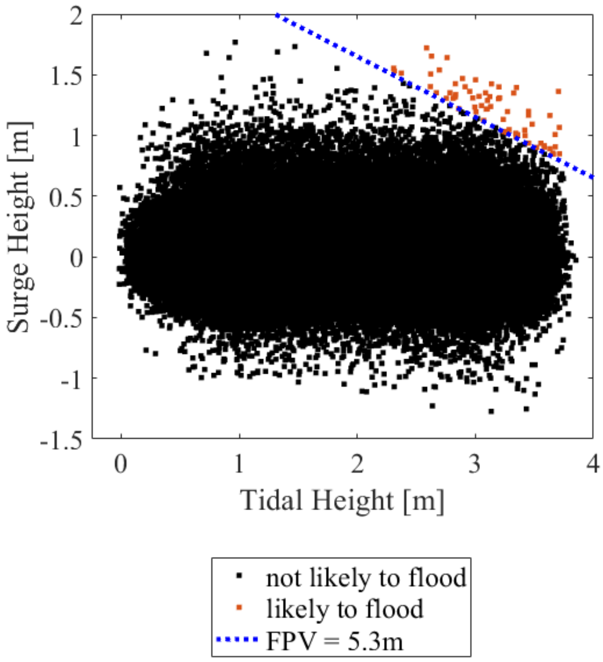


Figure 19: Plot of tidal height over the Chart Datum against surge height for Millport for 1980 to 2019 showing flood potential value (FPV) threshold and flood likely observations.

While Eq. [47] is an adequate approximation to establish an extreme surge height threshold, more comprehensive wave height and overtopping models exist, e.g. (Pullen, et al., 2018), that explicitly consider the water level, wave characteristics, and defence structure characteristics along with their associated uncertainty. The proposed methodology is integrated into a larger and detailed overtopping framework in the next Chapter of this dissertation.

4-2.2. Model Definition

This **case** study uses a 1000 km footprint around the target location of Millport in the Firth of Clyde as shown in Figure 20. To reduce operating computational demand, the spatial resolution of the domain reduces with distance to the target location: 400 – 1000 km has 1 degree resolution, 200-400 km has 0.5 degree resolution and < 200 km has 0.25 degree resolution. Locations that lie east are removed as they do not affect the target location. There are 434 locations over water in the domain. Considering mean sea level pressure, U10 and V10 wind speed components, this generates 1302 inputs for every timestep. ERA-5 data was obtained for the years 1980 to 2020.

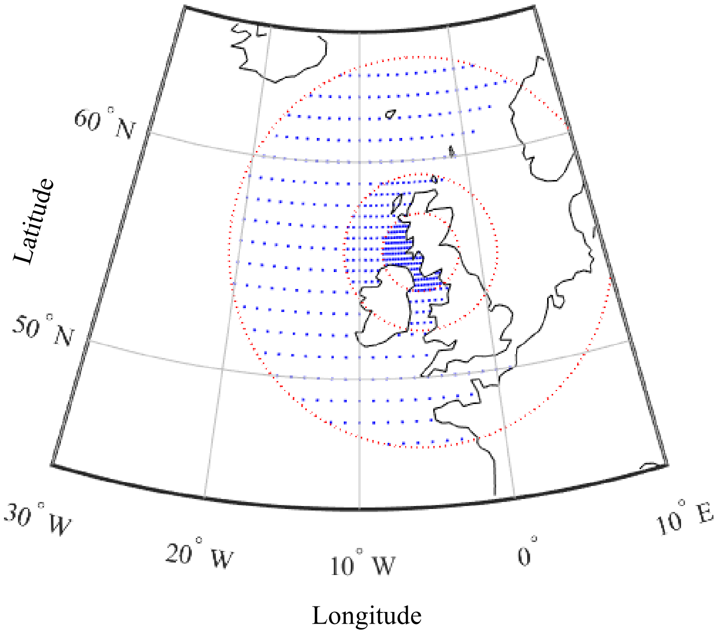


Figure 20: Map of the UK computational domain centred around the Firth of Clyde, ERA-5 variable locations (blue), and 200 km, 400 km and 1000 km boundary (red).

If 49 timesteps (the previous 24-hours + present hour + 24-hour forecast at 1 hour interval) are used to make each 24-hour prediction, this increases the number of inputs to 63,798 (1302 inputs at 49 timesteps). This leads to unmanageable network training times especially since multiple networks needs to be trained for the ABMS network set. To facilitate faster training by reducing the dimensionality of the problem and to de-correlate the input data which prevents overfitting, the inputs are subjected to principal component analysis. The cumulative percentage of variation explained for the principal components is shown in Figure 21.

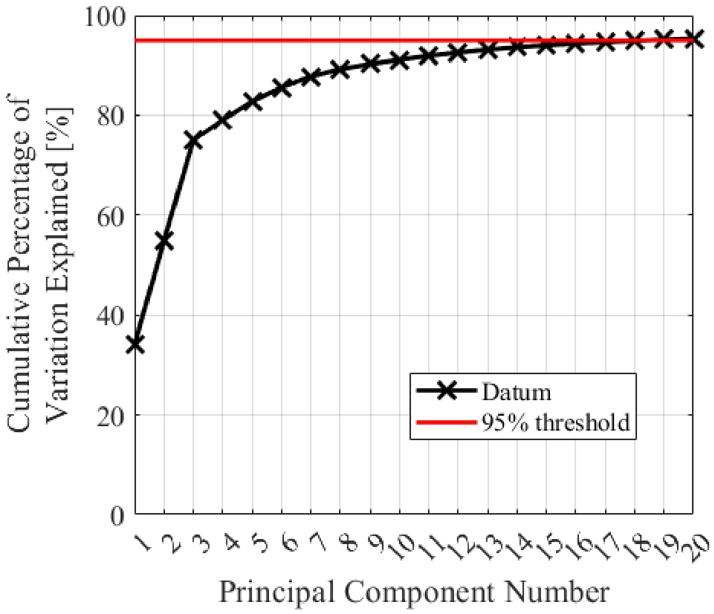


Figure 21: Plots of PCA component number against the cumulative sum of variation explained by the first principal components.

Selecting the desired number of principal components is a trade-off between the quantity of preserved input information and the amount of computational time required for training and running the model. 18 principal components that described 95% of the feature variation were used in this case study.

In addition to the 18 principal components, 3 local meteorological variables are added: the difference between the mean sea level pressure at Millport and the maximum pressure in the domain, U10 and V10. The location of these variables is -5° Latitude, 55.5° Longitude.

Non-tidal residual data was obtained for Millport for the years 1980 to 2023 from the British Oceanographic Data Centre National Tidal and Sea Level Facility (NTSLF) (BODC, 1980-2023). For this series, any values missing or deemed questionable by NTSLF and were removed from the analysis.

Hence the 18 principal components, 3 local variables and surge level at Millport combine to make the total number of features 22. The inputs are considered with a 1 hour timestep spanning a 49 hour window, 24 hours either side of the present hour. The output is a 24 hour forecast with one hour resolution. To preserve the input shape for the model, surge values for hours 25-49 are padded with zeros.

240000 inputs were divided for training, validation and testing with a 60/20/20 % split. The split was made chronologically to ensure independence for the groups. These percentage splits do not align precisely with the years spanning 1980 to 2020 owing to gaps in the Millport surge timeseries. Due to the low frequency of extreme events and high frequency of non-extreme events, the target distribution is highly imbalanced. To address this imbalance, the 24 hour lead time normalised surge heights are binned. 126 unique values are registered when a bin width of 0.1 is used, with counts between 1 (for the most extreme values) and 10000 (for the most common residual value, ~0 m). 1000 values are selected for each unique normalised surge height: When the bin count is greater than 1000, 1000 values are randomly sampled; when the bin count is less than 1000, the values are repeated to bring the bin count to 1000. Balancing the dataset prevents the high frequency, non-threatening surge heights from dominating the training process.

The Adam algorithm was selected for network training given its suitability for problems with large numbers of parameters and sparse data (Kingma & Ba, 2014). Additionally, it is easily implemented and computationally efficient. The loss function and metric are both *MAE* (Eq. [40]). Networks were continuously trained with various random seeds until 8 networks, each with an absolute bias of less than 0.1m for values above 0.75m, were produced, with the architecture shown in Figure 17. Each network was trained with a learning rate of 0.001 and early stopping criteria of 3 epochs was applied to validation loss to prevent overfitting. Each model training epoch took approximately 90 seconds using a CPU, with models taking roughly 25 epochs to train before early stopping criteria was met. Individual model's training error were evaluated in 0.25m bins across the full target range.

The posterior probabilities of the unseen data are inferred from a gaussian mixture model (GMM) fitted to the posteriors of the 40000 point validation dataset using Eq. [3] to Eq. [6]. An example for the 1st network is given in Figure 22. The surge height targets were divided into 10 equally populated bins to ensure a reasonable minimum population for fitting each mixture model.

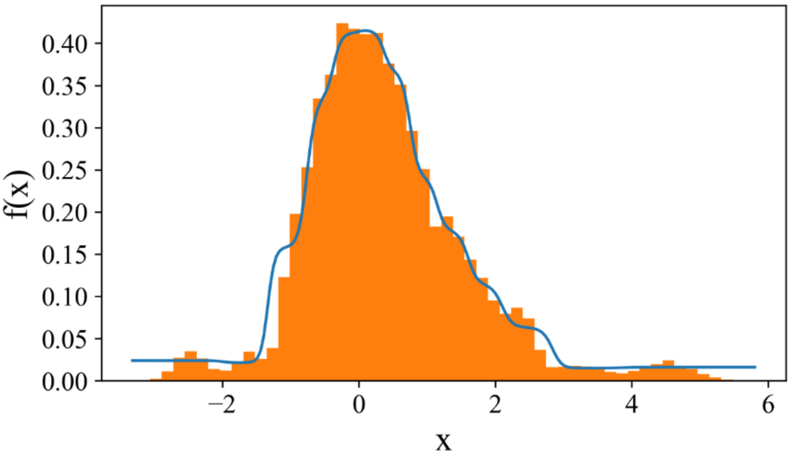


Figure 22: Gaussian Mixture Model for the prediction from the first neural network. X is surge height normalised by subtracting the mean and dividing by the standard deviation [dimensionless], f(x) is probability density.

4-2.3. ABMS Validation with ERA-5 Inputs

A comparison between 40000 observations from the test dataset and 24-hour predictions at Millport has been used to validate the ABMS surge forecast model. Box plots for the test data for each of these trained networks are shown in Figure 23 for the full range of values and in Figure 24 for the most extreme values, i.e. surge > 0.75m. The results establish good agreement for both the full range and the extreme values, demonstrating that the model inputs selected are able to describe the surge level with a good degree.

The benefits of using the ABMS algorithm are apparent as for the full series predictions, the ABMS averaged prediction outperforms every individual network in the set across all metrics (i.e. highest values for R^2, CC, lower values for MAE and SI and Bias values around 0). Similarly, for extreme predictions the ABMS averaged prediction outperforms every individual network in the set across all metrics, except for bias where a single network has a better bias value by 0.02m but with huge variability. For both the full series and the extreme values, the best individual network result for each metric is compared to the ABMS averaged prediction in Table 2.

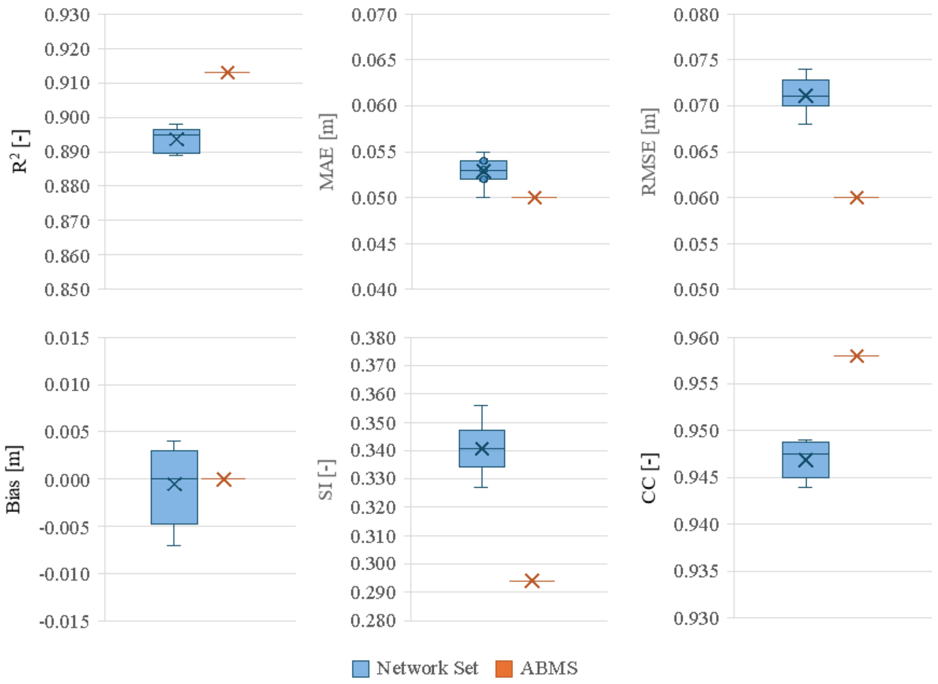


Figure 23: Box plots of network set performance compared with ABMS result for the full test data set.

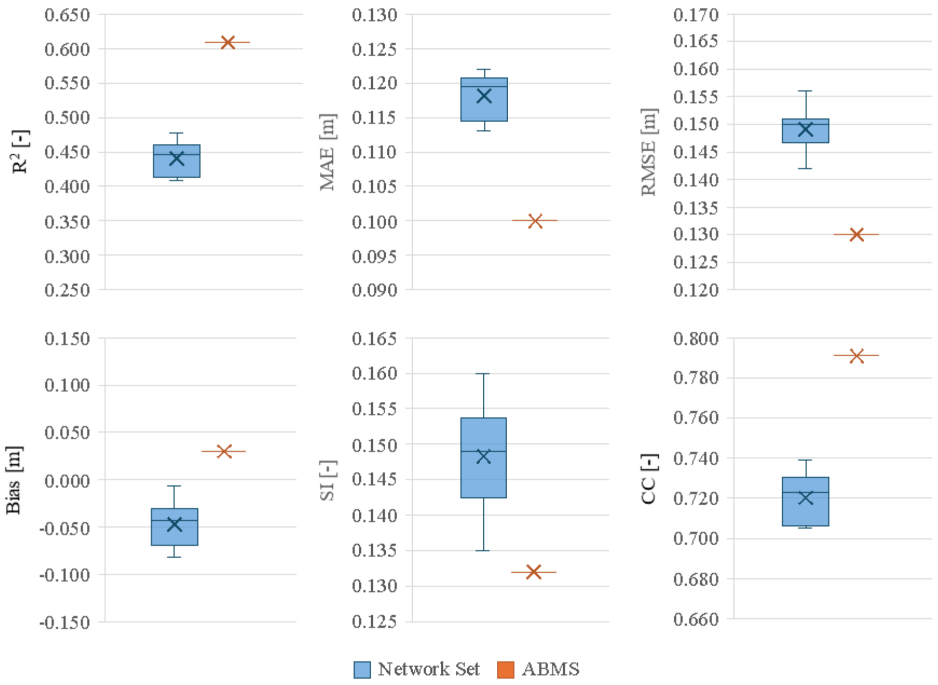


Figure 24: Box plots of network set performance compared with ABMS result for the extreme surge observations in test data set.

24 Hour Lead Time Surge Prediction Method Metric Comparison

Metric	Units	Best Network within the Set		ABMS	
		Full series	Extreme Surge (> 0.75m)	Full series	Extreme Surge (> 0.75m)
R^2	-	0.898	0.478	0.913	0.609
MAE	m	0.05	0.11	0.05	0.10
RMSE	m	0.07	0.14	0.06	0.13
BIAS	m	0.00	0.01	0.00	0.03
SI	-	0.327	0.135	0.294	0.132
CC	-	0.949	0.739	0.958	0.791

Table 2: 24 hour forecast metric comparison of ABMS averaged prediction and best net prediction within the set.

The increase in R^2 recorded in Table 2 shows that the ABMS method enhances the amount of observed variation explained by the model, compared with any individual ANN model within the set, both for the full data set and for extreme values. Likewise, the increase in CC shows that the ABMS surge predictions have a stronger relationship with the observed surge data than any individual model within the set. Figure 25 shows the performance of the 24-hour forecast surge height for Millport using the ABMS model, and a good agreement between the model predictions and the observed surge height can be seen. In Figure 25, the high density regions track line of $y=x$ which aligns with the bias metric of 0.00m for the full range and 0.03m for extreme values.

The drop in R^2 and the slightly higher MAE/RMSE for extreme surges are not an artefact of using a different time period—both figures evaluate the same test split—but result from conditioning the metrics on a small set of large-magnitude events (surge > 0.75 m). With far fewer samples, the variance of the "truth" in this restricted subset is lower and noise is proportionally larger, so R^2 degrades even when bias remains near zero, while MAE/RMSE increase relative to the full series (consistent with the discussion around Table 2). Two mechanisms dominate: first, extremes are rare despite bin-balancing during training, so they exert weaker constraints on the network and inflate ensemble spread; adaptive bootstrap model selection reduces, but cannot eliminate, this variability compared with single models (see the box-plots). Second, driver representativeness and phase errors matter more at peaks: training on ERA5 (~30 km) smooths sharp pressure gradients and jet features that govern peak surge, and forecast–reanalysis residuals are heavy-tailed and heteroscedastic, so errors grow toward feature extremes; moreover, small timing offsets are penalised more harshly in hour-by-hour scoring during peaks. These effects raise error for large events without implying a mean directional bias (cf. Fig. 20). In short, the lower R^2 for extremes reflects properties of the extreme-value subset itself—rarity, reduced variance, and greater sensitivity to representativeness and phase—not a change in evaluation period, which is why we report both full-series and extreme-only metrics.

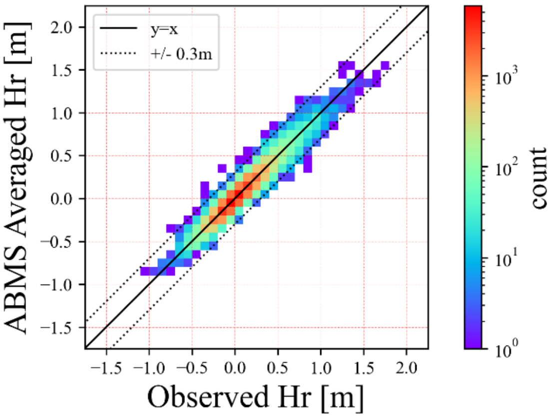


Figure 25: ABMS Predicted Surge Height Against Observed for Millport.

To evaluate the contribution of different errors to the prediction interval, three test cases are considered: a single network with its own *feature uncertainty*, ABMS ensemble excluding *feature uncertainty* Eq. [10] and ABMS ensemble including *feature uncertainty,* Eq. [13]. The robust predictions (expected value with 95% prediction interval) for these three cases are shown in Figure 26, Figure 27 and Figure 28 respectively.

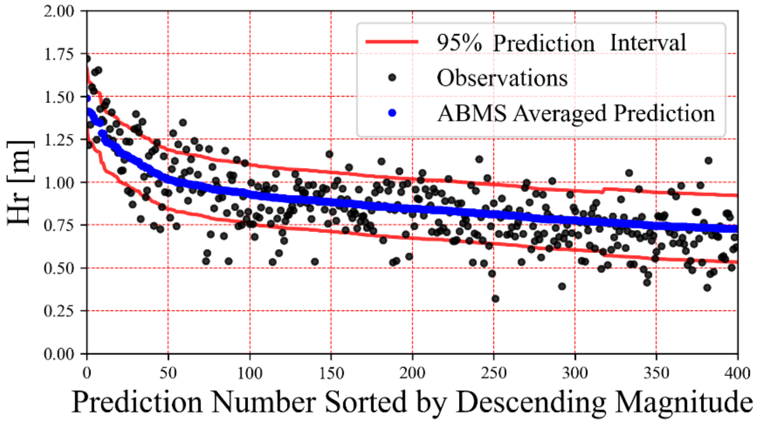


Figure 26: Single ANN predictions including *feature uncertainty* for 400 largest predictions sorted by descending magnitude.

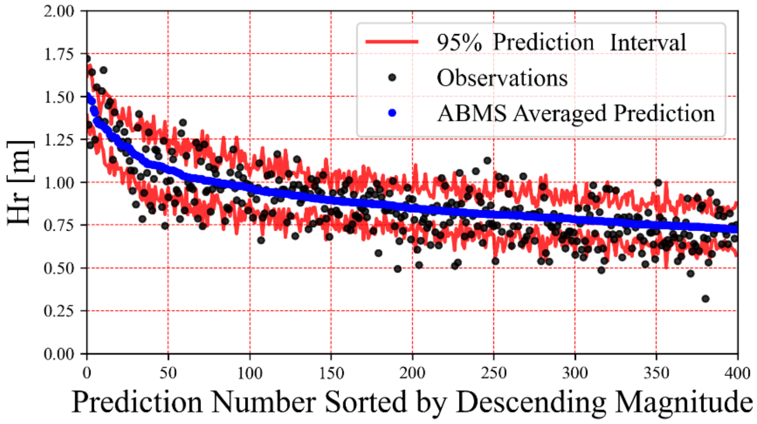


Figure 27: ABMS predictions excluding *feature uncertainty* applied for 400 largest predictions sorted by descending magnitude.

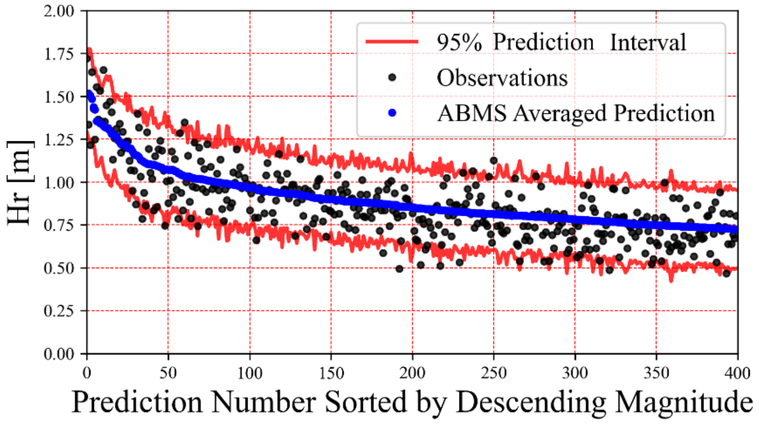


Figure 28: ABMS predictions including *feature uncertainty* for 400 largest predictions sorted by descending magnitude.

In each case, the predictions run through the centre of the test data which is reflected in the low bias described in Table 2. The percentage of observations that fall inside each of these prediction intervals for both the full series and the extreme surges is recorded in Table 3. The ABMS with feature uncertainty prediction (Figure 28) interval is less erratic than the ABMS model omitting feature uncertainty (Figure 27). A steady prediction interval provides consistent and reliable measures of uncertainty across different estimates of similar magnitudes, indicating that the level of variability in the data or model predictions is relatively constant.

24 Hour Leadtime Surge Prediction Method Interval Comparison

	Percentage of Values in 95% Prediction Interval [%]			Average Width of 95% Prediction Interval [m]		
Error Type	Single Model Inc. feature uncertainty	ABMS Exc. feature uncertainty	ABMS Inc. feature uncertainty	Single Model Inc. feature uncertainty	ABMS Exc. feature uncertainty	ABMS Inc. feature uncertainty
Full Series	92.6	64.1	97.3	0.25	0.11	0.28
Extreme Surge (> 0.75m)	76.2	68.7	92.9	0.35	0.28	0.48

Table 3: Prediction interval results and average width for three test cases.

Table 3 explores the effect that different errors have on the size of the prediction interval. The single model case with feature uncertainty falls 2.4% short of the desirable 95% of points for the full series, while falling 18.8% short for extreme surges. The ABMS prediction excluding feature uncertainty falls 30.9% and 26.3% short of the desired 95% for full series and extreme surges. This means that for the network structure described in this chapter, *network uncertainty* alone is not sufficient to build a usable prediction interval. The ABMS prediction including feature uncertainty meets the desired 95% of observations captured by the prediction interval and falls 2.1% short for extreme surges. Comparing the ABMS with feature uncertainty to a single network with feature uncertainty, the biggest difference comes for extreme surges.

There are several factors that would contribute to this. Firstly, the selected features and model architecture do not adequately describe the target space giving rise to significant feature uncertainty, and this is more significant for extreme surges, given the increased MAE and RMSE for extreme predictions compared with the full series (Figure 23 and Figure 24). Extreme surges are inherently rare and thus underrepresented in data sets. This scarcity of data complicates the ability of machine learning algorithms to discern patterns between features and targets. Additionally, extreme surges are influenced by complex atmospheric and oceanographic phenomena that can vary greatly in intensity and behaviour, making them more challenging to model accurately. Moreover, since the models are training on ERA5 data, its ability to describe the target space is constrained by the limitations of ERA5. ERA5 wind speed errors in the Atlantic Ocean tend to increase with wind speed intensity percentiles (Campos, Gramcianinov, de Camargo, & da Silva Dias, 2022). Furthermore, each model's size and structure are likely suboptimal for capturing the intricate patterns that drive surge responses, partly due to the compromise between model size and computational demands and because a fully comprehensive optimisation study was beyond the scope of this investigation.

The improvement in prediction interval for extreme surges vs full series surges, using the ABMS with feature uncertainty method compared to a single model with feature uncertainty, indicates that there is significantly greater variance in predictions across the set for extreme predictions than for non-extreme predictions. This aligns with the challenges of limited data availability for extreme surges and the complexities of machine learning modelling mentioned earlier. It highlights the benefits of using an ensemble of machine learning models that incorporate *network uncertainty* over a single model that does not, particularly in the context of forecasting extreme surges. As discussed earlier, *network uncertainty* by itself does not provide a robust basis for constructing a reliable prediction interval. However, when considered with feature uncertainty, it becomes sufficiently robust to be usable.

The improved prediction interval of the ABMS method when applied with feature uncertainty, alongside the clear improvements in ABMS predictions across multiple metrics demonstrated by Figure 23, Figure 24 and Table 2, make a very strong case for the adoption of the ABMS method for surge prediction compared to using a single machine learning model, especially when faced with data constraints. The performance of the ABMS method with feature uncertainty across the full range of test values is shown in Figure 29.

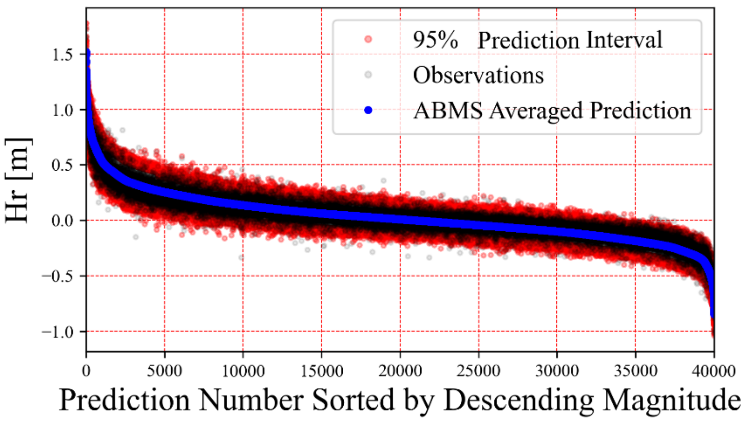


Figure 29: ABMS predictions including model error applied for all 40000 test predictions sorted by descending magnitude.

Errors appear larger under residual-by-residual verification because the target distribution is densely concentrated near 0 m (Figure 25). Even after bin-balancing during training, most verification pairs cluster around small residuals, so residual-wise scoring disproportionately penalises small phase offsets and moderate amplitude under/overshoots that occur frequently near zero. This inflates root-mean-square error and widens apparent under-coverage relative to event-wise peak (maximum) comparisons. The pattern is explicit in Table 2: coverage shortfalls are largest when input variability is not represented (adaptive bootstrap model selection, ABMS, without feature uncertainty: −30.9% for the full series and −26.3% for extremes) and shrink as uncertainty sources are added (single model plus feature uncertainty: −2.4% full, −18.8% extremes; ABMS plus feature uncertainty: approximately on-target for the full series and −2.1% for extremes). Several mechanisms drive these outcomes. First, meteorological representativeness: training uses ERA5 at ~30 km resolution and operations map IFS forecasts into the ERA5 principal-component space. At this scale sharp pressure gradients and narrow wind jets within extratropical cyclones are partially smoothed, reducing the local forcing that generates peak surge; the de-biased IFS→ERA5 residuals are heavy-tailed and heteroscedastic, so hourly errors grow toward feature extremes and concentrate in high-impact conditions, which necessitates wider intervals for extremes and leads to under-coverage when feature or forecast uncertainty is omitted. Second, pre-processing and spatial footprint: inputs are aggregated via principal components over a ~1000 km domain with coarser sampling at range; while efficient for synoptic drivers, this dampens sub-mesoscale contributions that matter for local peaks, making residual-wise errors look larger than event-maximum errors. Third, timing (phase) mismatch: hourly verification is sensitive to small lags between meteorological drivers and the gauge response, whereas event-maximum comparisons are less phase-sensitive and typically yield lower apparent error. For interpretation, the stricter residual-by-residual view emphasises day-to-day skill near zero and highlights where additional uncertainty is required to calibrate the tail. In ERA5 tests, combining feature and network uncertainty via ABMS stabilises intervals; under operational forcing, adding forecast uncertainty through Monte-Carlo bias adjustment is decisive for extremes while leaving mean predictions nearly unchanged, consistent with Figures 28–30 and Table 4. Event-wise peak metrics can be added in future work, but are not the focus here.

4-2.4. Robust surge forecasting system validation with IFS forecasts

To validate the surge prediction in realistic situations, one day IFS weather forecast data was obtained for the period January 2020-December 2023. To use the MCBA algorithm, the *forecast uncertainty* must be quantified and any possible bias between the IFS and ERA-5 removed. Therefore, the data for 2020 was used to quantify the uncertainty associated with the forecast by comparing them against the ERA-5 measurements for the same period. The forecasts for wind speed and direction were converted into U10 and V10 components. Thereafter, the wind and pressure inputs were transformed to the same domain as the reanalysis principal components by applying the loading matrix from the PCA process used on the original ERA-5 dataset. These transformed inputs were then normalised. To remove the potential bias, a linear regression model was applied to the principal components of the IFS forecast and its equivalent ERA-5 component using Eq. [14] - Eq. [17]. This process is shown for the first principal component in Figure 30.

The unbiased error histograms follow a logistic distribution as shown by the first principal component bias corrected error histogram in Figure 31, with heavier tails indicating that larger errors are more frequent compared to those in a normal distribution. The magnitude of the forecast error varies with the magnitude of the principal components and increases from the centre to both ends of the feature range. This relationship is demonstrated by the binned logistic fit standard deviations shown in Figure 32. As such, errors are assessed across 10 evenly distributed bins that span the range of each feature, creating a more versatile error structure. To guarantee that the errors applied within each bin are unbiased, the mean error for each bin is calculated and incorporated into the predictions. 250 Monte Carlo samples are used for each prediction.

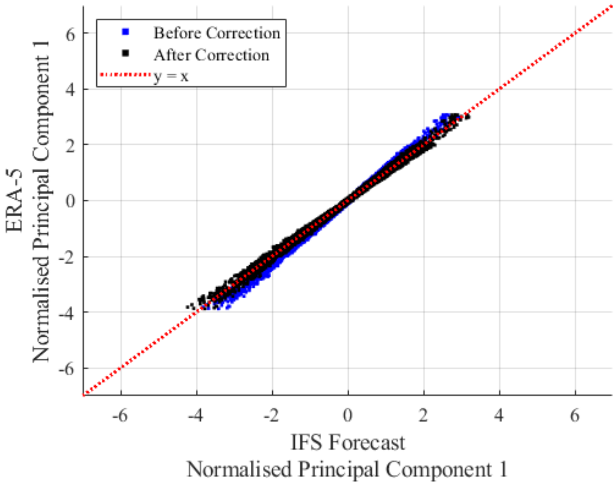


Figure 30: Bias correction for IFS Variable Principal Component 1 before (blue) and after (black) correction.

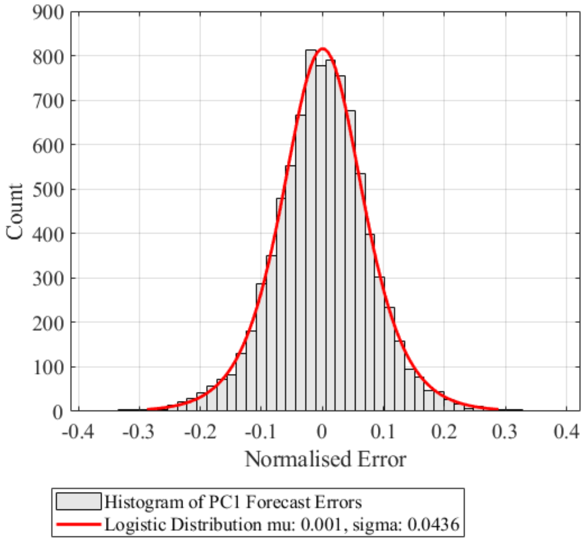


Figure 31: Forecast error histogram and logistic distribution fit for first principal component.

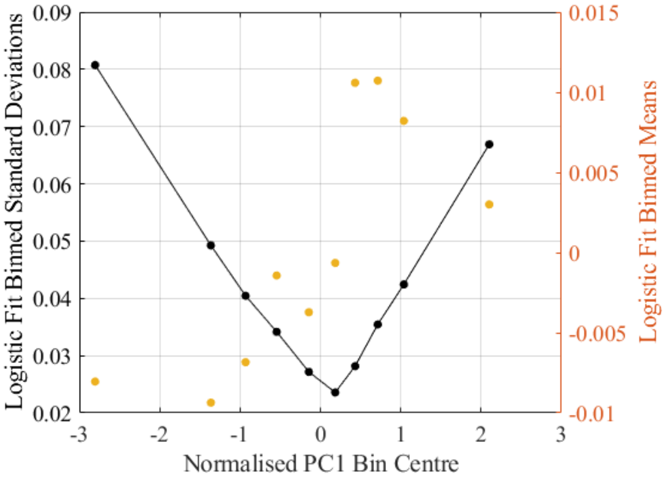


Figure 32: Logistic fit standard deviations (left y-axis) and means (right y-axis) for 10 equally populated bins along the range of the first principal component.

Two cases are evaluated here, ABMS with *feature uncertainty* and MCBA with *feature uncertainty*. For clarity, the MCBA method is the development of ABMS that considers *forecast uncertainty* and so a comparison of the two methods will highlight the significance of *forecast uncertainty* to the final prediction interval. IFS forecasts are obtained for years 2021 to 2023. The years selected here lie outside the training, validation, and previous testing period. The results of these cases for are recorded in Table 4. The volume of data available for extreme surges is low and the spread of predictions is high relative to the range of values observed. As such the calculated R^2 value for this range is not meaningful and has been omitted from the table.

Table 5 metrics indicate strong metrics show a strong alignment between the model predictions and in situ observations. The minimal differences in average predictions between the two methods are expected, as the primary distinction between them lies in the incorporation of unbiased forecast error. This effect of this distinction on the prediction interval is shown in Table 5.

24 Hour Lead Time Operational Surge Prediction Method Metric Comparison

Metric	Units	ABMS		MCBA	
		Full series	Extreme Surge (> 0.75m)	Full series	Extreme Surge (> 0.75m)
R^2	-	0.882	-	0.878	
MAE	m	0.05	0.10	0.06	0.10
RMSE	m	0.07	0.13	0.07	0.12
BIAS	m	-0.02	0.04	-0.02	0.03
SI	-	0.332	0.148	0.337	0.146
CC	-	0.943	0.170	0.942	0.173

Table 4: 24 Hour Surge Prediction comparison for the results 2021-23 of the MCBA and ABMS algorithm with feature uncertainty.

24 Hour Lead Time Operational Prediction Interval Comparison

	Percentage of Values in 95% Prediction Interval [%]		Average Width of 95% Prediction Interval [m]	
Error Type	ABMS Inc. feature uncertainty	MCBA Inc. feature uncertainty	ABMS Inc. feature uncertainty	MCBA Inc. feature uncertainty
Full Series	95.6	96.6	0.29	0.30
Extreme Surge (> 0.75m)	90.5	94.6	0.46	0.48

Table 5: Prediction interval results and average width comparison for ABMS with feature uncertainty and MCBA with feature uncertainty.

The prediction intervals for both methods cover the expected range of data through the full series. Since this holds true for both the ABMS and MCBA methods, it suggests that most of the prediction uncertainty arises from *feature uncertainty*, and *network uncertainty*, with *forecast uncertainty* playing a smaller role. This may be due to the way the models are set up; for example, the simplification inherent in principal component analysis can introduce errors by reducing complex patterns to simpler forms during training. Uncertainty that might latterly be introduced as forecasting error is instead embedded during training, manifesting as *network uncertainty* and potentially explaining variation between models in the set. Additionally, forecasting errors might be less noticeable because they are averaged over a large input domain before pre-processing. However, the ABMS method falls short by 4.5% of the targeted quantity for extreme surge events, whereas the MCBA method is only 0.4% off. Although forecasting error appears to be unnecessary for low-risk surge events, it is significant in constructing satisfactory prediction intervals for extreme surges. The MCBA with feature uncertainty robust predictions for extreme surges are shown in Figure 33, with the full series shown in Figure 34.

Four uncertainty configurations are evaluated: network only (ABMS without feature uncertainty), feature only (single ANN with feature uncertainty), network + feature (ABMS with feature uncertainty), and (iv) network + feature + forecast (MCBA with feature uncertainty). Network spread alone was unusable, with about 31% under-coverage on the full series and 26% on extremes—intervals were erratic and failed to reflect data variability. Adding feature uncertainty largely fixed this for routine conditions (full-series shortfall ≈2.4%) but still left the tail under-covered by ~18.8%. Combining network + feature via ABMS met the 95% target for the full series and cut extreme under-coverage to ~2.1% in ERA5 tests, with far steadier interval widths. Finally, introducing forecast uncertainty with MCBA barely changed mean predictions but was decisive for extremes under operational IFS forcing, improving the tail shortfall from ~4.5% (ABMS) to ~0.4%, with typical operational intervals around 0.48 m. In short: feature uncertainty drives most interval width; network variance alone is inadequate; and explicit forecast uncertainty is needed to calibrate extreme probabilities in real time.

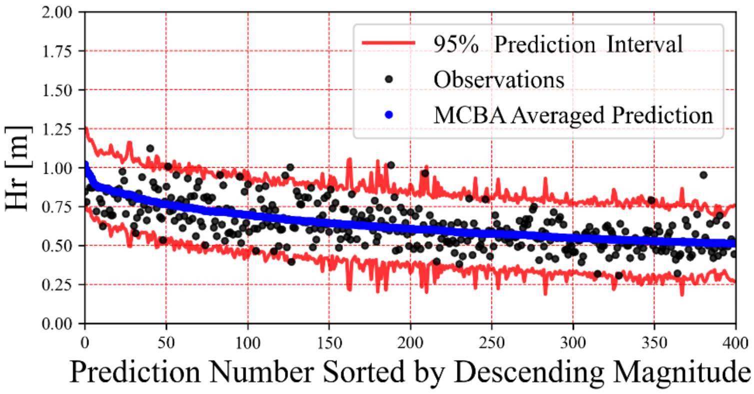


Figure 33: 2021-2023 24 hour lead time surge predictions. Predictions made using MCBA algorithm with feature uncertainty and IFS forecasts showing 400 largest predictions sorted by descending magnitude.

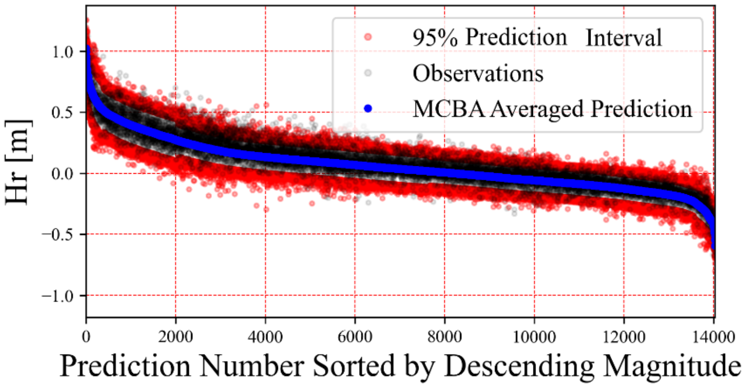


Figure 34: 2021-2023 24 hour lead time surge predictions. Predictions made using MCBA algorithm with *feature uncertainty* and IFS forecasts showing all predictions sorted by descending magnitude.

The MCBA with feature uncertainty has prediction intervals that typically span 48cm. Figure 33 illustrates that values lying outside the prediction interval generally differ by only a few centimetres. The efficacy of the model ensemble and the prediction interval is deemed satisfactory. This supports the conclusion that the model assumptions are sufficient for characterising surge *forecast uncertainty*.

Figure 35 shows the MCBA with feature uncertainty is able to predict surge height with a good agreement with observed data from Millport with data from 2021 till 2023. The high density regions track the identity line (y=x) which aligns with the bias metric of -0.02m for the full range and 0.03m for extreme values recorded in Table 5. The CC for this time period is 0.942 with a RMSE of 0.07m.

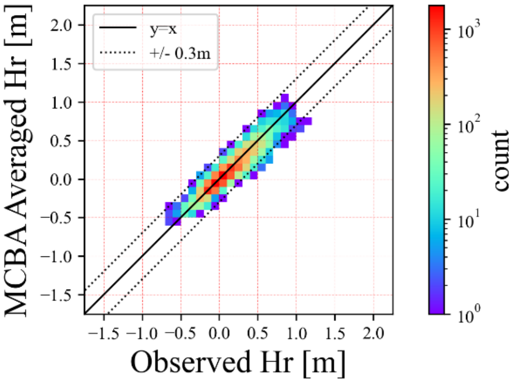


Figure 35: Validation of the MCBA with feature uncertainty: Millport for 2021-23.

For comparison, the NTSLF NEMO tide-surge model provides forecasts ranging from 0-6 hours, covering the period from 2020 to 2024, with the data organised in monthly files. Data from 2021 to 2023 was specifically analysed for Millport, with the monthly values consolidated into a composite three-year value for RMSE and CC, with each month weighted according to its length. For the 0-6 hour forecasts, the physics-based model NTSLF achieves a CC of 0.927 and an RMSE of 0.09 meters. Comparatively, the MCBA surge forecasting framework shows better CC and lower RMSE for its 24-hour lead time forecasts. The MCBA with feature uncertainty method offers a robust prediction interval that takes into account all the uncertainties (from data, model and forecast). It is structured modularly, allowing for straightforward updates to the model as network design, data availability, or forecast accuracy improves. These updates can then be reflected in both the accuracy and the prediction interval of the model.

A strict like-for-like comparison of prediction intervals is not possible because the NTSLF product is deterministic in our setup and evaluated with point-error metrics (correlation, RMSE), whereas the RSFS (MCBA with feature uncertainty) returns calibrated 95% intervals. Operationally (2021–2023), RSFS intervals have a typical width of ≈0.48 m, achieve near-nominal coverage on the full series, and reduce the extreme-tail shortfall to ~0.4 percentage points (versus ~4.5 points when forecast uncertainty is omitted in ABMS). This indicates that most interval width derives from feature + network uncertainty, while explicit forecast uncertainty is crucial to calibrate extremes.

RSFS delivers on-demand intervals in seconds per prediction, which the NTSLF configuration here does not. Consequently, we compare accuracy directly (RSFS at 24 h outperforms NTSLF at 0–6 h at Millport) and report RSFS calibration separately. A fair future benchmark would construct empirical NTSLF intervals by conditioning residuals on lead-time and magnitude classes over 2021–2023, then compare coverage and width against RSFS.

The narrower “extreme” range in 2021–2023 reflects realised storminess rather than shrunken model uncertainty. That period lacked the largest surges seen in the longer ERA5/gauge archive (1980–2020) and the 2020 forecast–reanalysis pairing. Interval behaviour remains consistent—typical width ≈0.48 m, near-nominal full-series coverage, and tighter tail calibration with MCBA (extreme shortfall ~0.4% vs ~4.5% for ABMS). On the broader archive, intervals widen as expected and ABMS extreme shortfall is ~2.1%, indicating readiness for larger events.

4-2.5. Application of the surge forecasting system

The robust surge forecasting system uses the MCBA method with feature uncertainty. The operational capability of the model is shown by applying the method to 24-hour lead time surge height predictions between the 27th October and 4th November 2022 and between 18th and 25th September 2023 as shown in Figure 36 and Figure 37, respectively. Shown here are 2 weeks of 24-hour forecasts. In practise 24 predictions are made, one for each hour up to and including 24 hours. The RSFS expected value (EV) prediction tracks the observed surge height, and the prediction interval is both time periods meaningful and usable. The computational time required to obtain one prediction is in the order of seconds and therefore allowing real-time predictions.

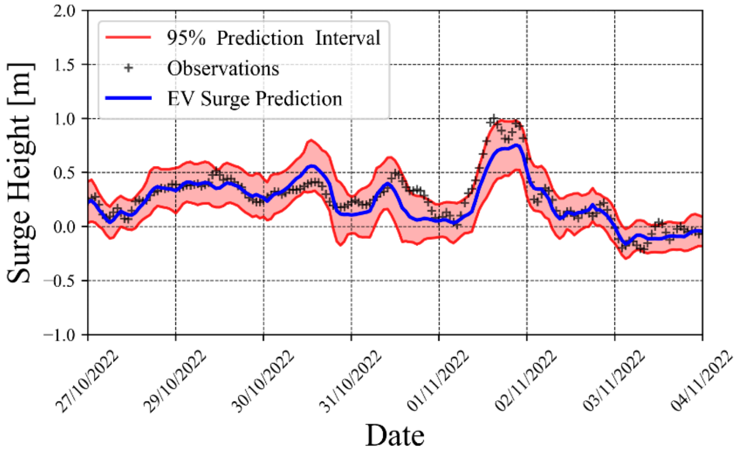


Figure 36: 24 Hour RSFS 24-hour predictions for the week spanning 27/10/2022 to 04/11/2022.

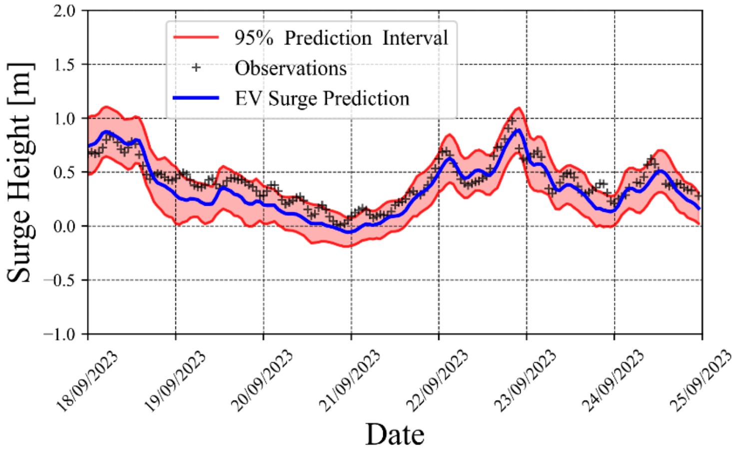


Figure 37: 24 Hour RSFS 24-hour predictions for the week spanning 18/09/2023 to 25/09/2023.

4-3. Limitation and potential improvements

While the purpose of the chapter was to demonstrate the viability of the MCBA process for surge height predictions, several improvements to the model structure can be made. Further analysis on the architecture selection of the ANN and hyper parameter tuning may lead to an improvement in the individual models within the set and an improvement in the MCBA prediction accuracy. Additionally, since the MCBA method can be applied to models with differing architectures that use the same inputs, the unique abilities of different sized architectures to capture specific surge patterns can and should be exploited to improve the averaged prediction. The predicative capability of the surge can be further improved by increasing the input grid beyond the dimensions used in this case study. A larger input domain would better capture the atmospheric processes that generate surges. Investigation should be conducted to evaluate the amount of uncertainty that arises from changing the number of selected principal components.

These changes directly impact the computational cost of the analysis during the training of the model, but it might offer the opportunity for a longer time prediction. For instance, the medium-range IFS forecast extends up to two-weeks, and the methodology can readily be extended to accommodate this prediction length albeit with further consideration given to the propagation of uncertainty since an extension of the forecast lead time would strongly depend on the reliability of the atmospheric forcing. An interesting area of further work would be to investigate the spatial and temporal dependencies of meteorological forecast errors and assess their impact on surge predictions. In addition, specific storm characteristic inputs such as radius of maximum wind speed or land fall location can be easily incorporated into the model framework to improve the prediction capability of the model, or the proposed methodology can be applied to existing neural network-based models or ensemble models that have uncertain inputs.

4-4. Surge Forecasting under Uncertainty Conclusions

This chapter has illustrated the development and validation of a robust 24-hour surge forecasting machine learning based model. This builds and improves on current methodologies. More specifically, Adaptive Bayesian Model Selection approach has been used to improve on predictions than can be achieved with a single neural network whilst providing a usable and reliable prediction interval that considers both model uncertainty and attribute uncertainty. Thereafter a Monte Carlo procedure has been integrated to include the effect of *forecast uncertainty* on the surge prediction interval. The results show that for 24-hour lead times, model, attribute and forecast uncertainties are crucial for accurately characterising surge prediction intervals, particularly for extreme surges.

The robust surge forecasting system has been successfully deployed and validated at Millport within the Clyde basin in Scotland. Operational performance has been shown through model application using weather forecast provided by ECMWF's IFS, demonstrating that the proposed approach is able to perform robust surge predictions with an error boundary approximately of 48 cm wide which is both meaningful and usable. The methodology presented offers a straightforward and effective way to construct and train a robust storm surge model. Around the UK, the National Tidal and Sea Level Facility (NTSLF) provides a storm surge model with good spatial coverage. However, this model does not capture all the uncertainties that this approach addresses due to the computational burden of running physics-based models. Given that ERA-5 and IFS are global datasets, this methodology can, in principle, be applied to any location using historical surge height timeseries or surge height reanalysis data series. While most of the coastline is not covered by tide gauges, reanalysis data sets provide a valuable alternative. This framework can be readily modified to accommodate the additional uncertainty introduced by using surge reanalysis datasets.

4-References

- BODC, B. (1980-2023). Retrieved January 2023, from https://www.bodc.ac.uk/data/hosted_data_systems/sea_level/uk_tide_gauge_network/
- C3S, C. (2017). *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate*. Retrieved Dec 2022, from https://cds.climate.copernicus.eu/cdsapp#!/home
- Campos, R., Gramcianinov, C., de Camargo, R., & da Silva Dias, P. (2022). Assessment and Calibration of ERA5 Severe Winds in the Atlantic Ocean Using Satellite Data. *Remote Sensing, 14*(19). doi:https://doi.org/10.3390/rs14194918
- Chelton, D. B., & Schlax, M. G. (1996). Global observations of oceanic Rossby waves. *Science*, 234-238.
- DEFRA, E. (2005). Joint Probability: Dependence Mapping and Best Practice: Technical report on dependence mapping. R&D Technical Report FD2308/TR1.
- ECMWF, E.-R. (2024). Integrated Forecasting System (IFS). Reading.
- Emanuel, K. (2017). Assessing the present and future probability of Hurricane Harvey's rainfall. *Proceedings of the National Academy of Sciences, 114(48)*, pg. 12681-12684.
- IPCC, I. (2019). Special Report on the Ocean and Cryosphere in a Changing Climate.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization.
- Labach, A., Salehinejad, H., & Valaee, S. (2019). Survey of dropout methods for deep neural networks. *arXiv preprint arXiv:1904.13310*.
- Macdonald, E. (2024). Bayesian Coastal Forecasting Toolbox. Retrieved from https://github.com/emacd-domain/Bayesian_Coastal_Forecasting

- Milne, F., Winter, M., Reeves, S., Knappett, J., Dawson, S., Dawson, A., . . . Brown, M. (2017). *Assessing the risks to infrastructure from coastal storms in a changing climate: Project Report PPR800.* Wokingham: Transport Research Laboratory.
- Mohamad-Saleh, J., & Hoyle, B. S. (2008). Improved neural network performance using principal component analysis on Matlab. *International journal of the computer, the internet and Management, 16(2)*, 1-8.
- Pullen, T., Liu, Y., Otinar Morillas, P., Wyncoll, D., Malde, S., & Gouldby, B. (2018). A generic and practical wave overtopping model that includes uncertainty. *In Proceedings of the Institution of Civil Engineers-Maritime Engineering, 171*, 109-120.
- Qin, Y., Su, C., Chu, D., Zhang, J., & Song, J. (2023). A review of application of machine learning in storm surge problems. *Journal of Marine Science and Engineering, 11*(9).
- Reusch, D. B., Alley, R., & Hewitson, B. C. (2005). Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data. *Polar Geography, 29(3)*, 188-212.
- Sabatino, A., Murray, R., Hills, A., Speirs, D., & Heath, M. (2016). Modelling sea level surges in the Firth of Clyde, a fjordic embayment in south-west Scotland. *Natural Hazards, 84*, 1601–1623.
- SEPA, S. (2015). Flood Modelling Guidance for Responsible Authorities Version 1.1. 59.
- Tiggeloven, T. C. (2021). Exploring deep learning capabilities for surge predictions in coastal areas. *Scientific reports(1), 11*, 1-15.
- Wang, B. (2002). Kelvin Waves. In *Encyclopedia of atmospheric sciences* (p. 1062).
- Williams, J., Horsburgh, K., Williams, J., & Proctor, R. (2016). Tide and skew surge independence: New insights for flood risk. *Geophysical Research Letters, 43*(12), 6410-6417.

Chapter 5

Surface Wave and Coastal Inundation Forecasting under Uncertainty

This chapter details the application of the MCBA methodology to a wind wave and coastal inundation forecasting problem. It is a reworking of journal paper titled "Machine Learning-Based Coastal Inundation Forecasting with Explicit Uncertainty Quantification" submitted as part of this PhD work to the journal of Coastal Engineering.

Within the risk framework introduced in Chapter 1, the hazard term P(Hz)—the probability that still water level plus wave effects will reach inundation-relevant levels—is central to credible early warning and downstream risk estimation. In mid- and high-latitude settings, the most severe extratropical storms can expose coasts to jointly high surge and waves; sharpening P(Hz) directly improves decisions that limit disruption and loss across communities and critical infrastructure. In this chapter, inundation is operationally defined through a site-specific threshold on total (tide + surge + waves), consistent with the overtopping/FPV formulation used later for the Firth of Clyde case study.

Chapter 2 highlighted why wave and inundation forecasting is hard: observations and asset data are sparse or outdated; waves integrate wind forcing along fetch with time lags; nearshore transformation and run-up depend on local morphology; and physics-based chains, while powerful, can be computationally demanding and sensitive to parameterisation—making it difficult to communicate uncertainty at the cadence needed for early warning. These challenges motivate a transparent, data-efficient uncertainty treatment alongside point predictions.

Chapter 3 therefore developed a probabilistic ML framework that we now generalise from surge to waves and inundation. We couple the Robust Surge Forecasting System (RSFS) to a Robust Wind-Wave Forecasting System (RWFS) and propagate feature, network, and forecast uncertainty via MCBA. Time-dependent forecast-error correlations are preserved using a covariance-aware (Cholesky) sampler, yielding calibrated prediction intervals across lead times. Nearshore transformation (linear theory) and an empirical run-up step then produce a time-varying inundation indicator with uncertainty.

In this chapter we apply the framework to the Firth of Clyde: we validate wave characteristics against CMEMS and an Irvine waverider dataset, then run the full chain on Storm Ciara (Feb 2020) to demonstrate 24-h lead-time exceedance forecasts relative to an operational threshold.

5-1. Inundation Forecasting Framework

The inundation framework has been designed to enable a robust forecast to be made using and easily accessible weather forecasts. The framework uses basic inputs with uncertainties considered for all inputs and models and propagated through to the final inundation forecast. All pre-processing tools, model building tools, and model algorithms are available in the Bayesian Coastal Forecasting Toolbox git repository (Macdonald, Bayesian Coastal Forecasting Toolbox, 2024). The flowchart for this process is detailed in Figure 38.

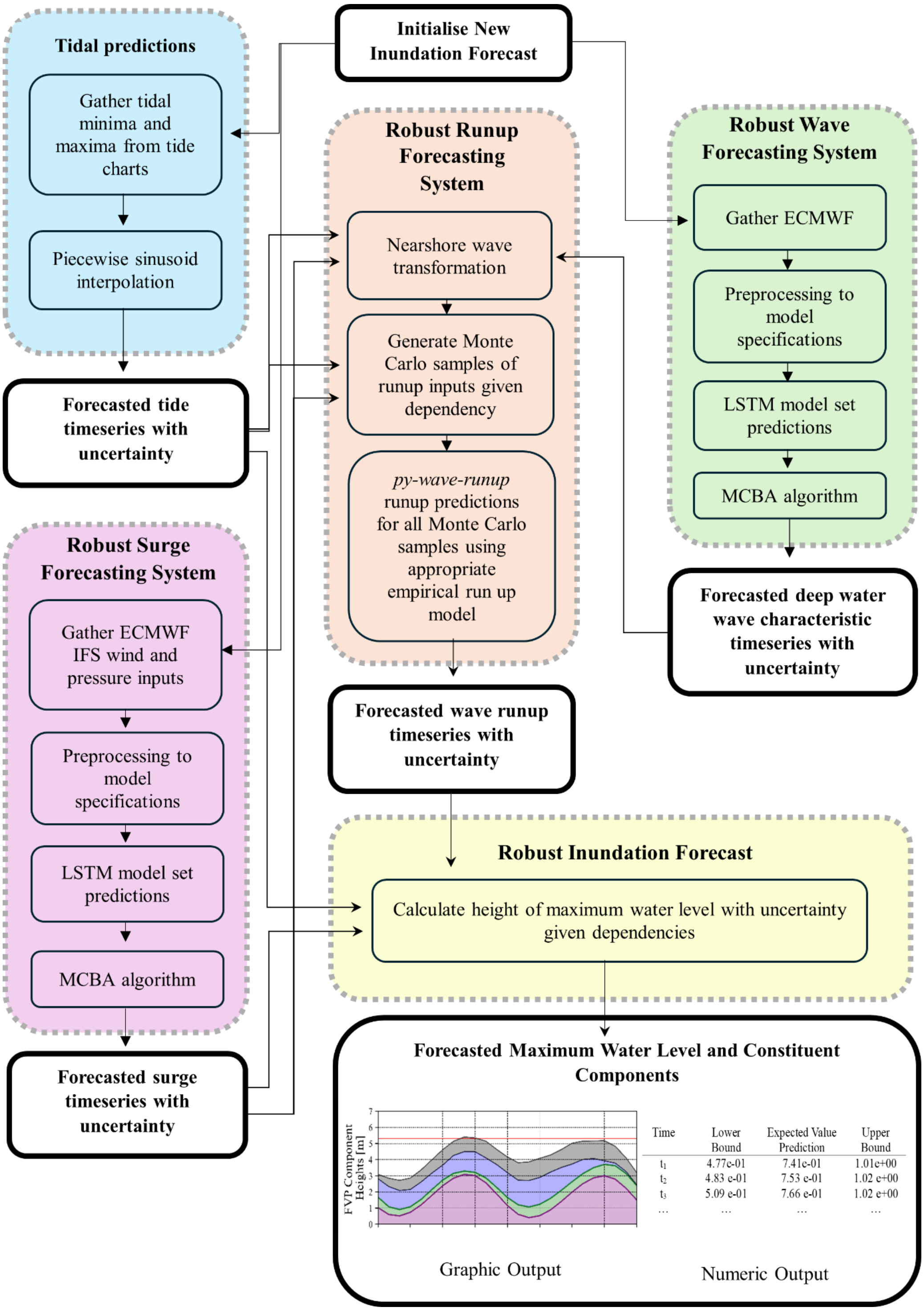
The overarching inundation framework takes a 24-hour wind and pressure forecast and makes an averaged prediction with uncertainty for surge height using the robust storm surge forecasting system (RSFS) (Macdonald, Tubaldi, & Patelli, 2025), with consideration given to the dependency of surge height on tide where appropriate. The non-tidal residual prediction $\boldsymbol{H_r}$ is added to the tidal prediction $\boldsymbol{H_a}$ to provide a Still Water Level (SWL). Deep water wave predictions with uncertainty are made using a robust wind wave forecasting system (RWFS) (detailed in subsection 5-2.3) and the surface wave characteristics are transformed to nearshore wave characteristic using linear wave theory. These transformed wave characteristics are passed to a runup model to give a nearshore runup prediction $\boldsymbol{H_u}$ with uncertainty. The runup uncertainty and SWL uncertainty are then added in quadrature to provide final inundation forecast. The constituent parts of this framework are outlined in the following sub-sections.

5-1.1. Astronomic Tide

Tidal time series are generated in this model through piecewise sinusoidal interpolation of tide chart maxima and minima. The accuracy of this method is evaluated by analysing the errors involved. For locations not covered by tidal charts, harmonic analysis can be applied if there is a nearby tide recording station, or a global tidal model can be used otherwise. For a review of global tidal models see (Stammer, et al., 2014) The uncertainties must be assessed individually in each case.

5-1.2. Surge Forecasting

The Robust Surge Forecasting System (RSFS) (Chapter 4) is an implementation of the Monte Carlo Bayesian Averaging algorithm (MCBA) applied to a set of surge forecasting artificial neural networks (ANN). These networks are bidirectional Long Short Term Memory (LSTM) networks designed to capture temporal dependencies in sequential data, both forwards and backwards. The MCBA algorithm is a methodology for characterising and propagating three types of uncertainty: feature uncertainty, the uncertainty associated with feature selection for the problem; network uncertainty, the uncertainty introduced by the random starting weights and training process of a neural network; and forecast uncertainty, the uncertainty introduced to surge forecast by the weather forecast. Monte Carlo simulation uses random sampling to estimate the range of probable outcomes for a system under uncertainty. Each Monte Carlo sample represents a credible variant of the meteorological forecast that drives the surge model.



Time	Lower Bound	Expected Value Prediction	Upper Bound
t_1	4.77e-01	7.41e-01	1.01e+00
t_2	4.83 e-01	7.53 e-01	1.02 e+00
t_3	5.09 e-01	7.66 e-01	1.02 e+00
…	…	…	…

Figure 38: Flowchart of coastal inundation framework.

5-1.3. Surface-wave Forecasting

As wave growth is dependent on wind speed, wind duration and fetch, the wave that an observer sees on the coast will have been forced by wind action that tracks backward along its path to its origin. In order to fully capture the physical processes affecting a single wave, the wind velocity profile at all points along the travelled path of the wave must be ideally known.

However, this is not practical for surface-waves. The alternative here is to consider the lag of maximum cross-correlation between the waves at the target location and the wind action at a secondary, distant location. The lag at maximum cross-correlation is the time delay when one time series is most strongly correlated with another. An example is shown in Figure 39 (left) with maximally correlated ranges indicated by dotted semi circles extending away from 0 hr target. For a 0 hr wave prediction, all inputs are nowcasts. If the +12 hour value was desired for the same target location, the values will be shifted to preserve the maximally correlated lag structure, Figure 39 (centre). Here the 12 hr prediction inputs are a combination of nowcasts and forecasts.

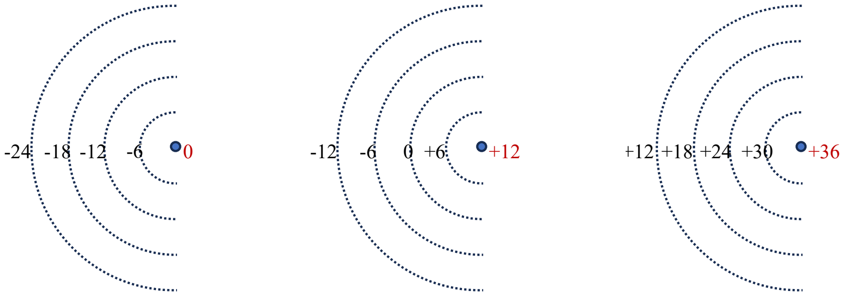


Figure 39: relative lag structure for shifting forecast lead times.

If the 36 hr prediction is desirable, the lags are shifted again preserving the maximally correlated lag structure and all inputs are forecasts, Figure 39 (right). The proportion of nowcast and forecast is dependent on the maximal cross correlation structure of the input grid and the required lead time of the wave forecast.

5-1.3.1. Input Selection and Dimension Reduction

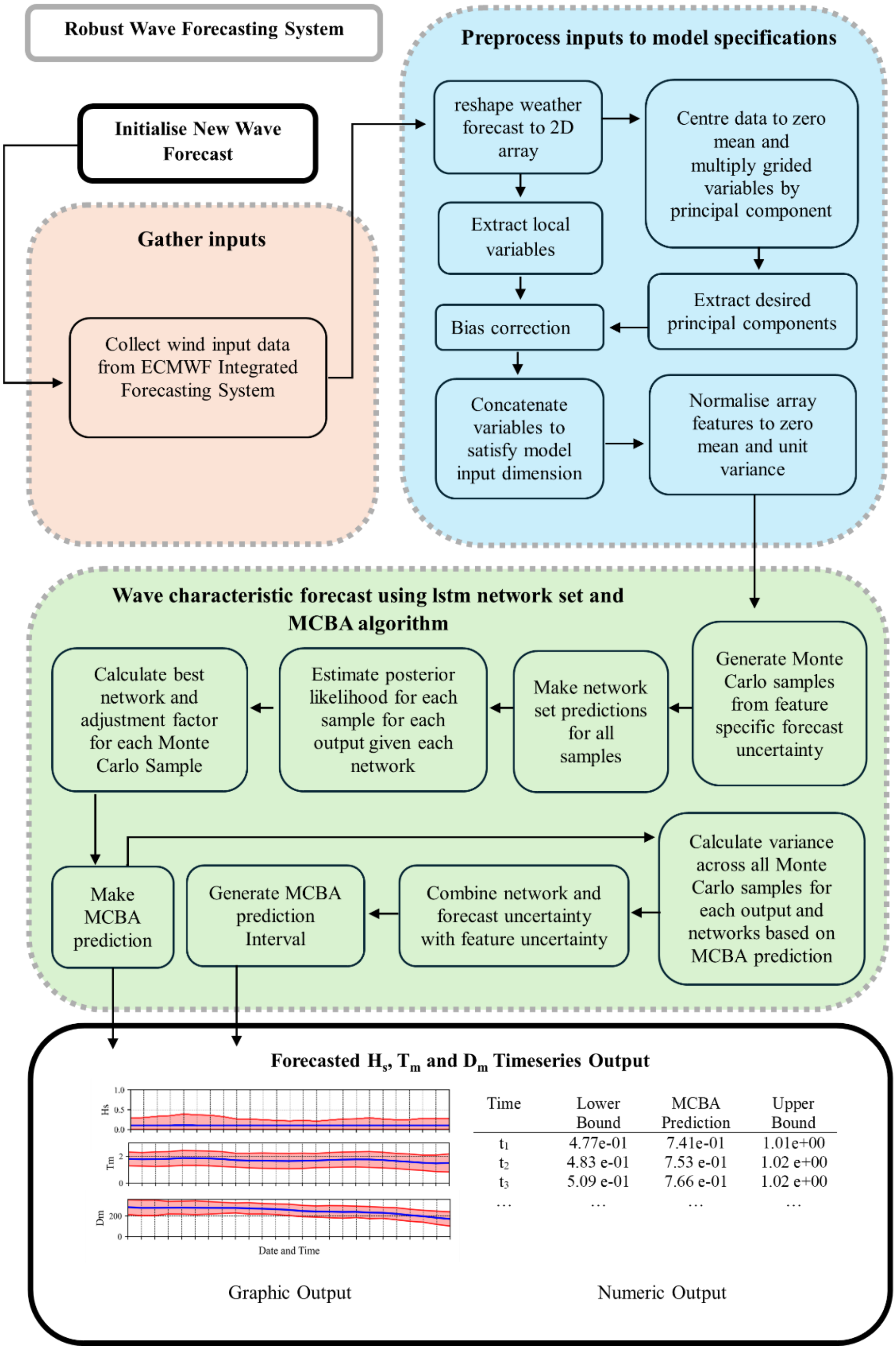
As the wave height is influenced by winds over the entire length of its path, it is necessary to consider a wide area for the prediction of surface waves. To decrease the computational burden during runtime, the wind inputs are simplified using principal component analysis (PCA). PCA operates by transforming a dataset into a new coordinate system through a linear transformation, where each orthogonal axis in this system aims to capture maximum variance in the data, dimension by dimension. This process allows the original dataset to be represented with fewer components while still retaining the most critical features. Consequently, PCA provides a more compact form of the data, emphasizing the most significant information. The most informative principal components are retained, while those contributing minimal variability are discarded. Another benefit of PCA is that it generates components that are independent, eliminating correlations among inputs, which enhances the uniqueness of the data representation and improves the training and generalization of Artificial Neural Network (ANN) models. A general limitation of PCA is that the principal components derived from different input types may not have physically interpretable features, potentially complicating the understanding of the transformed features (Reusch, Alley, & Hewitson, 2005). In this case, all the inputs are wind components and have the same units. The components (as a linear sum) are therefore also in ms-1 and so have a physical meaning.

5-1.3.2. Input Uncertainty Characterisation

As mentioned in Section 4-2.1, the early warning surge system is trained on reanalysis data and run using forecast data. This is to allow the expected uncertainty in the forecast to be propagated to the surge prediction uncertainty. To quantify the error between the reanalysis data and forecast data, the IFS data must be mapped to the same domain as the reanalysis data, as outlined in Section 4-2.2. This is done by normalising the IFS forecast inputs and multiplying them with the principal component coefficients for the reanalysis data. The 'true' reanalysis value of the kth principal component $\boldsymbol{PC_{k,R}}$ can be expressed in terms of biased forecast $\boldsymbol{PC_{k,forecast}}$ with an error $\boldsymbol{\varepsilon(0,\sigma)}$ by applying Eq. [14] - Eq. [17].

5-1.3.3. Robust ANN-based Wave Forecasting System

The robust ANN-based wave forecasting system (RWFS) schematic is shown in Figure 40. It describes the steps involved in converting 24-hour wind forecast to a 24-hour significant wave height, mean wave period and mean wave direction forecast. The pre-processing and input uncertainty evaluation described above are written in Python as modular functions for ease of operation. The MCBA algorithm is given in Chapter 3-1.2. This methodology is developed further by considering time dependent uncertainties in the wind forecast detailed in Chapter 3-1.3.



Time	Lower Bound	MCBA Prediction	Upper Bound
t_1	4.77e-01	7.41e-01	1.01e+00
t_2	4.83 e-01	7.53 e-01	1.02 e+00
t_3	5.09 e-01	7.66 e-01	1.02 e+00
…	…	…	…

Figure 40: Schematic of the proposed robust wave forecasting system

The MCBA method (Macdonald, et al., 2024) works by exploiting the variation in predictions made by a set of ANNs that have all been trained on the same data, see Oparanji et al. (2017) . In this ensemble the wind-wave forecasting models have the architecture shown in Figure 41. The model architecture type is bidirectional Long-Short Term Memory (LSTM). LSTM networks are a type of recurrent neural network architecture that are effective for sequence prediction tasks, such as time series forecasting. The bidirectional node enhances the standard LSTM architecture by allowing the network to have access to information from both past and future contexts in the input sequence. All three wave characteristics are predicted by different branches of the same model to decrease the number of model evaluations required to make each set of predictions. Due to the large number of trainable parameters in the model, a dense layer is added after the Bidirectional LSTM with a 0.1 dropout rate. Dropout layers randomly disconnect a subset of weights between layers, preventing all neurons in a layer from simultaneously optimizing their weights. tf represents the number of forecasted timesteps, and nfeat indicates the number of features. The optimal architecture along with the optimal number of wave models in the ensemble is out with the scope of this investigation and is not established here.

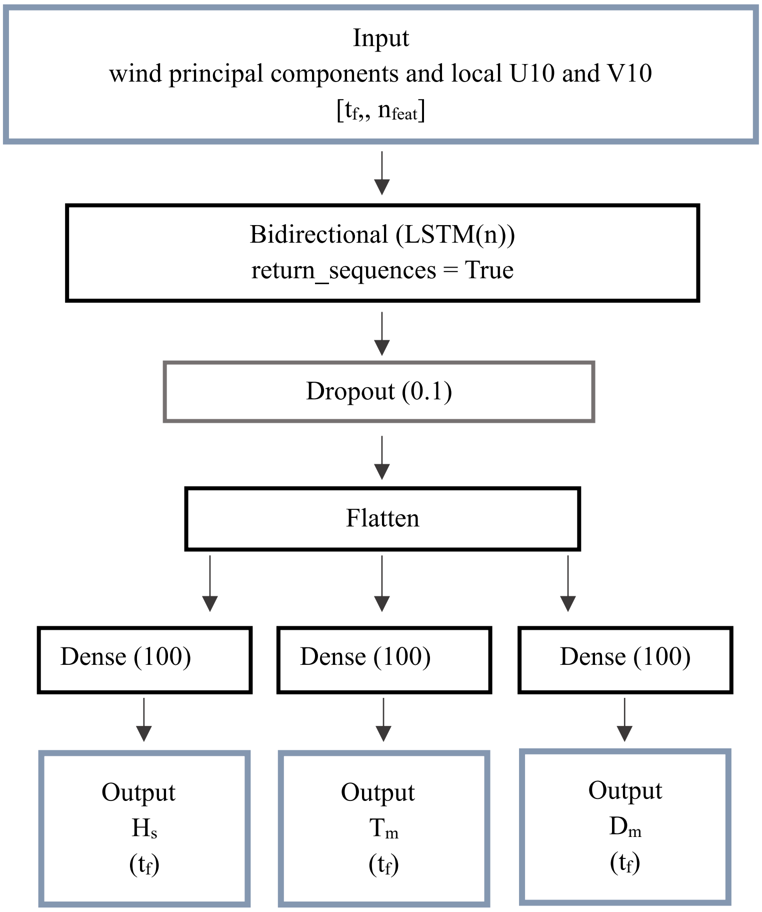


Figure 41: Wind-wave characteristic forecasting LSTM network architecture.

5-1.4. Nearshore Transformation

Nearshore wave transformations are done using linear wave theory. Water celerity refers to the speed at which waves propagate through a body of water. Firstly, an initial wavelength is estimated for both the initial and nearshore location using the deep water wavelength approximation:

$$\mathbf{L} = \frac{\mathbf{g}\boldsymbol{T_p}^2}{\mathbf{2\pi}} \qquad \text{Eq. [48]}$$

Given that the model outputs mean wave period, $\boldsymbol{T_m}$, a conversion is made using the approximation $\boldsymbol{T_p} \approx \mathbf{1.1} \times \boldsymbol{T_m}$. From here an estimate is made as to the depth of the water using the classifications in Table 6.

Water Depth Classification	Boundary Ratios		
Deep	0.5 ≤	Length/Depth	
Transitional	0.05 ≤	Length/Depth	< 0.5
Shallow		Length/Depth	< 0.05

Table 6: Water depth classification for Airy wave theory and associated threshold wavelength to water depth ratios.

The wavelength is then iterated over using the depths, classifications and their accompanying formula until a wave length solution reaches convergence with its given water depth classification. The formula for wavelength in transitional waters is given as:

$$\mathbf{L_T} = \frac{\mathbf{g}\boldsymbol{T_p}^2}{\mathbf{2\pi}} \mathbf{tanh}\left(\frac{\mathbf{2\pi d}}{\mathbf{L}}\right) \qquad \text{Eq. [49]}$$

Based on these classifications, the wave celerity for deep water $\boldsymbol{C_D}$ and transitional water CT is calculated by:

$$\mathbf{C_D} = \frac{\mathbf{g}\boldsymbol{T_p}}{\mathbf{2\pi}} \qquad \text{Eq. [50]}$$

$$\mathbf{C_T} = \frac{\mathbf{g}\boldsymbol{T_p}}{\mathbf{2\pi}} \mathbf{tanh}\left(\frac{\mathbf{2\pi d}}{\mathbf{L}}\right) \qquad \text{Eq. [51]}$$

Group velocity refers to the velocity at which energy is transmitted by a group of waves or a wave packet, rather than the velocity of the individual waves within the packet. In other words, it describes how quickly the shape or envelope of a group of waves moves through a medium. The group velocity in deep water $\mathbf{V_{G,D}}$ and transitional water $\mathbf{V_{G,T}}$ are given by:

$$\mathbf{V_{G,D}} = \frac{\mathbf{g}\boldsymbol{T_p}}{\mathbf{4\pi}} \qquad \text{Eq. [52]}$$

$$\mathbf{V_{G,T}} = \frac{\mathbf{1}}{\mathbf{2}}\left[\mathbf{1} + \frac{\frac{\mathbf{4\pi d}}{\mathbf{L}}}{\mathbf{sinh}\left(\frac{\mathbf{4\pi d}}{\mathbf{L}}\right)}\right] \cdot \mathbf{C_T} \qquad \text{Eq. [53]}$$

And shoaling coefficient KS is the square root of the ratio of the group velocity of the region being left and the region being entered. For deep to transitional water this becomes:

$$\mathbf{K_S} = \sqrt{\frac{\mathbf{V_{G,D}}}{\mathbf{V_{G,T}}}} \qquad \text{Eq. [54]}$$

The refraction coefficient for water waves $\boldsymbol{K_R}$ describes how much the direction of the waves is altered when they pass from one depth to another. It is determined by the ratio of the wave speeds in the two depths by:

$$\mathbf{K_R} = \sqrt{\frac{\cos(\theta_0)}{\cos(\theta_1)}} \qquad \text{Eq. [55]}$$

where $\boldsymbol{\theta_0}$ and $\boldsymbol{\theta_1}$ are the angles of the wave relative to the shoreline in their respective location. Since $\boldsymbol{\theta_0}$ is known, $\boldsymbol{\theta_1}$, can be calculated using Snells law:

$$\frac{\sin(\theta_1)}{\sin(\theta_0)} = \frac{C_1}{C_0} \qquad \text{Eq. [56]}$$

Nearshore wave height $\boldsymbol{H_1}$ can then be expressed as the product of initial depth wave height $\boldsymbol{H_0}$, refraction coefficient $\boldsymbol{K_R}$ and shoaling coefficient $\boldsymbol{K_S}$:

$$\mathbf{H_1 = H_0 \cdot K_S \cdot K_R} \qquad \text{Eq. [57]}$$

5-1.5. Run-up

Run up estimations are done using the python package py-wave-runup (Leaman, Beuzen, & Goldstein, 2020). Py-wave-runup is a Python module that contains a suit of published empirical wave runup models including Stockdon et al. This package has been selected for its convenience and flexibility to the user in selecting individual or ensemble models that suit their particular requirements.

5-2. Robust Wave Forecasting System for Firth of Clyde Basin

This sub-section describes the implementation of the robust wave forecasting system (RWFS) of the inundation model outlined in Section 5-2 to the Firth of Clyde basin. A map of the Firth of Clyde basin is shown in Figure 42. The model is validated in 10m deep water at Irvine against an in-situ waverider buoy data for a 3 month data set in 2001. Due to a lack of real wave data in the region, the wave characteristic targets are taken from the Atlantic-European North West Shelf- Wave Physics Reanalysis of the E.U. Copernicus Marine Service Information (CMEMS). The target location of this model is chosen as the closest grid cell to the location where the Irvine waverider buoy is deployed.

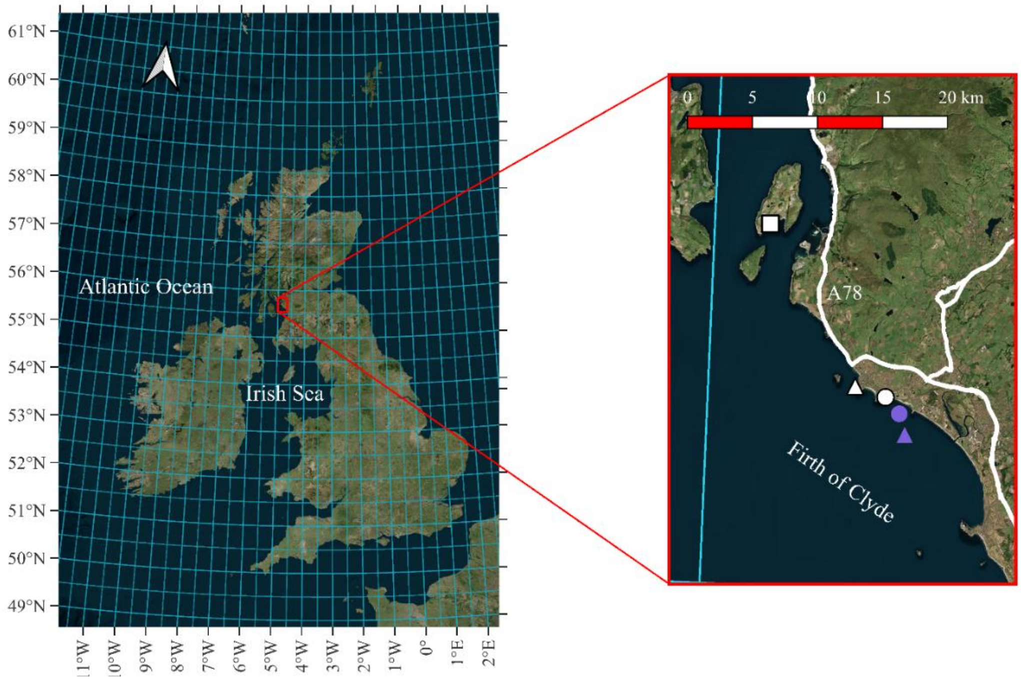


Figure 42: Model domain showing the Firth of Clyde basin in the Southwest coast of Scotland. Significant locations include Millport (white square), Irvine Wave Rider Buoy (purple circle), CMEMS reanalysis grid target cell (purple triangle), Saltcoats Harbour (white circle) and Ardrossan Harbour (white triangle).

5-2.1. Model Domain

For hydrodynamic models, it is common to use reanalysis data (output from a modern high fidelity model driven by historical observations) for training machine learning models with meteorological inputs, as this offers greater coverage compared to in-situ measurements. The size of the input area, also known as the "footprint," plays a significant role in quality of the predictions. However, larger input areas usually increase the computational demand for model training and run time. To capture the effect of fetch on wave heights at this location, a 1000 km boundary is drawn from the north entry between Northern Ireland and Southwest Scotland and southern boundary between Southeast Ireland and Southwest Wales as shown in Figure 43. This boundary defines the area over which the wind is expected to influence wave growth at the target location.

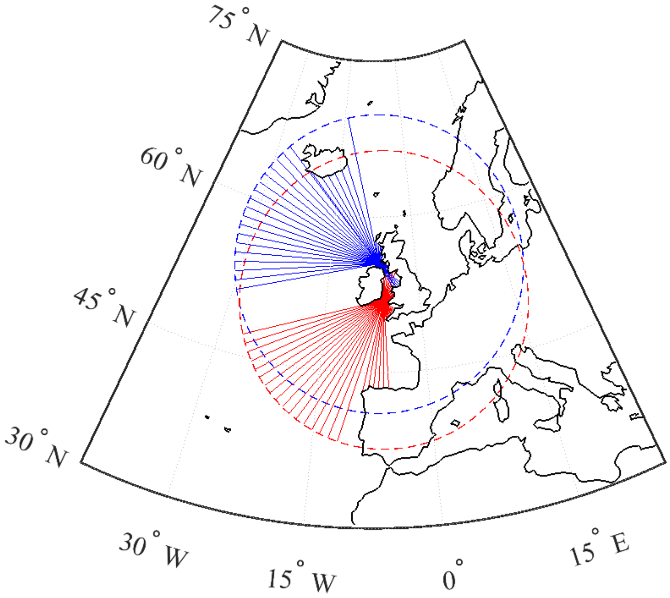


Figure 43: Map showing 1000-km fetch boundary to the north (blue dash) and south (red dash). The 'spokes' indicate the fetch lengths from the respective boundary to the north and south entry point.

From here the fetch directions are calculated by drawing lines from the origin on the boundary to the fetch boundary, stopping at the first intersection point with the coast. The area bound by these fetch lengths constitutes the input grid, shown in Figure 44. +u the RWFS is trained on ERA5 reanalysis data (C3S, 2017). ERA5 reanalysis data has a latitude/longitude grid of 0.25 degrees but to manage the computational demand of input processing, and to ensure that overarching inundation forecasting framework can be run quickly with standard PC processing power, the resolution of points within the fetch area is strategically varied, with a higher concentration of points nearer the target location and a lower concentration closer to the fetch boundaries. Areas within 250 km of the Firth of Clyde are described by 0.25 deg resolution. Areas within 250-500 km of the Firth of Clyde are described by 0.5 deg resolution. Areas beyond 500km are entered with 1 deg resolution. This creates an array with 680 locations. ERA5 wind data is stored as 10m eastward and northward wind components, U10 and V10.

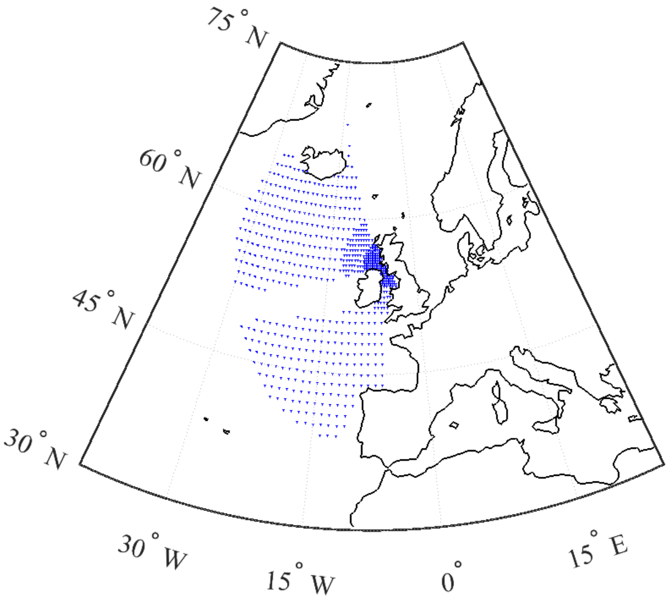


Figure 44: Map showing ERA5 grid input locations corresponding to the fetch area in Figure 43.

To establish the lag associated with peak cross-correlation (discussed in section 3.3.1) between wind magnitude at each point of the grid and Hs at the target location, the eastward and northward wind components., U and V respectively, are converted to wind magnitude and directions for all grid locations. The cross correlation is calculated with a lag range from -48 hours to + 48 hours. The hour at which the cross correlation between the wind magnitude and Hs is maximal is the lag of peak cross-correlation and is calculated for every location on the grid. An example at 52o latitude, -15o longitude is shown in Figure 45, with the lag of peak-cross correlation at -10 hours.

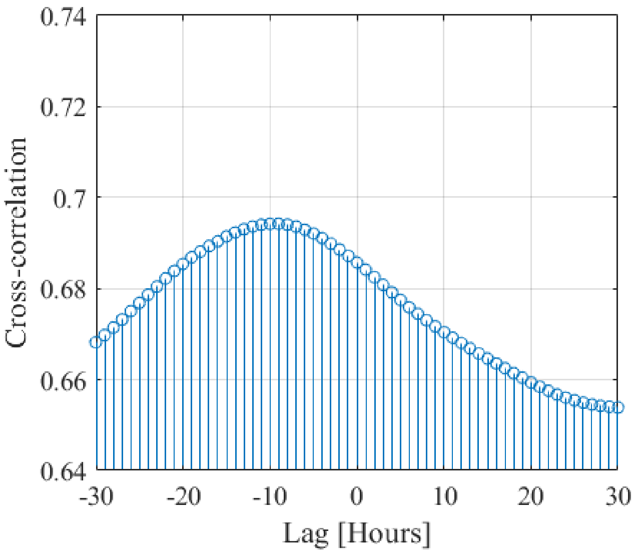


Figure 45: Lag against cross-correlation for wind magnitude events pointing towards the target location and Hs at the location.

The lag of peak cross-correlation is extracted for every location, and a second-order polynomial surface is fitted to the points using maximum likelihood estimation to ensure a continuous lag structure. The second order polynomial has an R2 value of 0.96 meaning that the majority of the lag variation is described by the surface. The equation for the fitted surface is given in Appendix C. The contour plot of this process is shown in Figure 46.

New lags are calculated by evaluating the surface at each latitude and longitude from the input grid and rounded to the nearest whole number to align with the hourly input time steps. For each time step the respective lag for the Hs value is applied to the U10 and V10 wind component. This creates an input set for each Hs, with a single value provided at each location that is maximally correlated with the target value. This methodology maximises the important information of each grid location to the target, whilst minimising the computational burden.

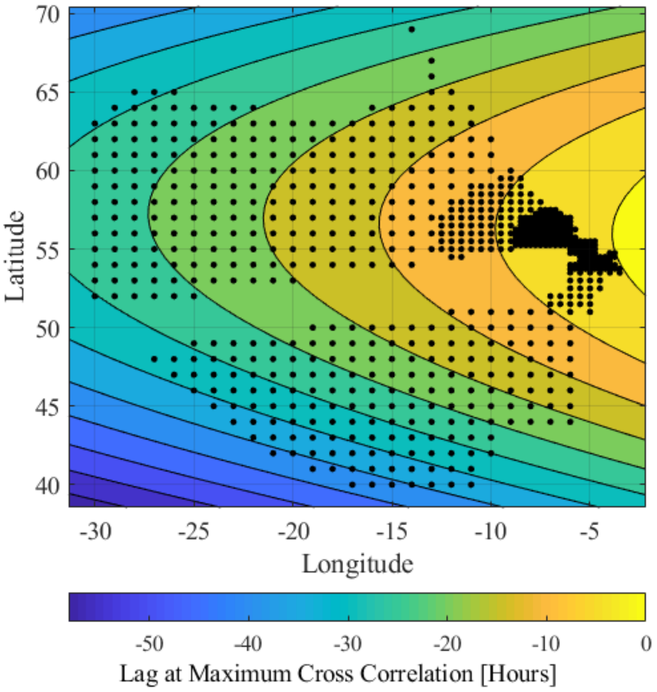


Figure 46: Surface fit to lags of maximum correlation.

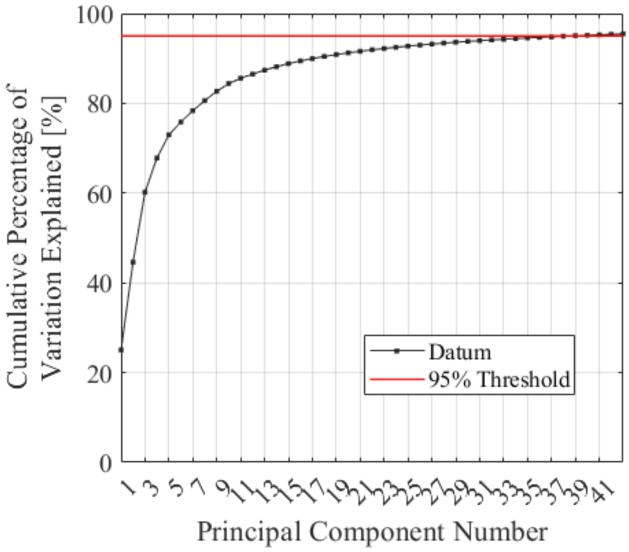


Figure 47: Cumulative percentage of wind variable variation explained by principal components.

Thereafter, the lag-adjusted U10 and V10 values have PCA applied and the principal components that describes 95% of the input variation is chosen – here corresponding to 42 principal components. The cumulative sum of variation explained for these principal components is shown in Figure 47.

In addition to these principal components, the U10 and V10 wind components for the input grid cell next to the target location are included as extra variables, Ulocal and Vlocal, making the total number of features 44. ERA5 data and CMEMS data was gathered for the years 1980 to 2022. The CMEMS data was interpolated from three hourly to hourly resolution to match up with the input resolution of ERA5. From the CMEMS data the significant wave height Hs, mean wave period Tm and mean wave direction Dm are all forecasted. The initial design matrix comprises data from Jan 1980 to Dec 2019 with the minimum wave height threshold set at 10cm. This resulted in 1.85 ×105 observations in total with 1.6×105 observations allocated to training and validation and 2.5×104 observations for testing. The features and targets are normalised and reshaped with a 37-hour sliding window to produce a final design matrix, described in Table 7.

Dimension	Description	Total	Breakdown
1	Observations	1.85×10^5	1.65×10^5 Training and Validation, 2.5×10^4 Testing
2	Timesteps	37	-12 to +24 hour forecast in 1hr steps
3	Variables	47	44 features: 42 principal components, U_{local}, V_{local} 3 targets: H_s, T_m, D_m

Table 7: Training and testing design matrix: dimensions and descriptions.

5-2.2. Model Design and Training

The architecture depicted in Figure 41 was used to train 8 networks, utilising an 80/20 split of the previously described training and validation data. Each network is designed to produce output variables that must remain unbiased across the entire range of the variable. Additionally, the output must be unbiased for variable values when the concurrent significant wave height is above 1.5 meters. The Adam algorithm (Kingma & Ba, 2014) is selected for network training given its suitability for problems with large numbers of parameters and sparse data. Additionally, it is easily implemented and computationally efficient. The objective function is the minimisation of the mean absolute error across all normalised targets. Each network has been trained for 100 epochs with a learning rate of 10-3. Early stopping is enabled to prevent overfitting with a patience value of 5. Section 4.5 shows that 8 models of the prescribed architecture produces a stable output and the agreement for individual models between the forecast and observed wave characteristics are very good, however questions over ensemble set size and optimal architecture require further investigation and ought to be addressed at a later stage. The feature error for each network is recorded for each output.

5-2.3. Time Dependent Forecasting Errors

The model was originally trained using ERA-5 wind data, while operational forecasts are driven by ECMWF IFS forecasts. The forecast uncertainty is quantified using Eq. [14] to Eq. [17] with ERA5 and IFS data for the year 2020. Thereafter the model is validated using IFS data for 2021 and 2022.

IFS Forecast and nowcast windspeed data is downloaded using the Visual Crossing weather API (Visual Crossing, 2023) at the locations shown in Figure 44. The wind speed and direction values are converted into U and V wind component grids. Individual grid location U10 and V10 components are bias corrected, normalised, concatenated to match the inputs structure of the ERA5 training data and multiplied by the loading coefficients of the ERA5 principal components. The desired number of principal components are extracted from the output and scaled to match the range of ERA5 values. This process is repeated for each forecast lead time with respect to the evolving lag structure described in sub-section 5-3.3.1, such that each lead time's inputs comprise the desired proportion of nowcast and forecast. For a given lead time, if the lag of maximal cross-correlation is positive the wind speed is taken from the forecast while if it is less than or equal to zero, nowcast data is used. Hence, 25 sets of inputs are created, for 0 to 24 hour lead time. Each input set is normalised to zero mean and unit variance. Figure 48 illustrates the change in error standard deviation for both the 0-hour and 24-hour lead time cases. The increased scatter observed in the 24-hour lead time plot highlights the additional uncertainty introduced with increased lead time.

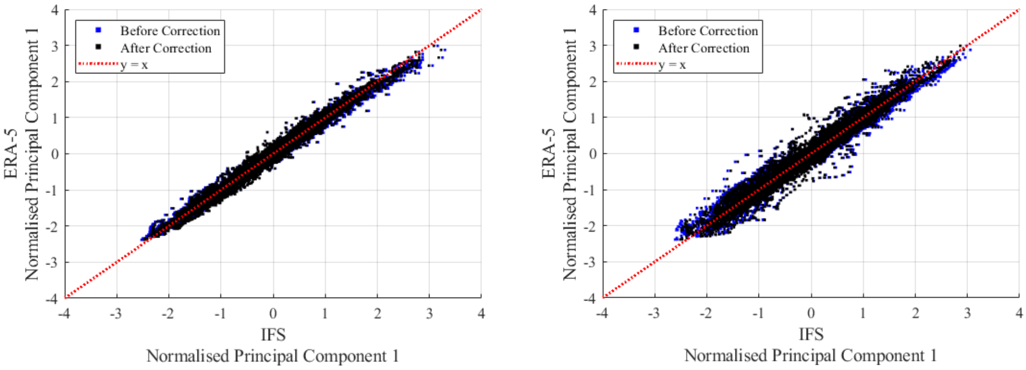


Figure 48: Visual-crossing 0 hour lead time first principal component (left), 24 hour lead time first principal component (right) for 2020.

To understand the impact of time dependent errors, two scenarios are considered: 1) when the errors are independent, and 2) when the errors show temporal dependency and dependencies between principal components. The second assertion about the error structure is that a significant portion of the error exhibits temporal dependencies. Forecasting models often rely on initial conditions or past data that are inherently uncertain. As errors in the initial conditions propagate over time, they can influence subsequent predictions, leading to temporal correlations. Additionally, spatial correlations arise because many physical systems modelled in forecasts are interconnected - errors in one region can affect neighbouring areas. In this instance, given that a lag structure is imposed on the input grid, each forecast lead time has an associated spatiotemporal dimension which will necessarily contain correlated errors. The error for the first principal component follows a logistic distribution and is shown in Figure 49.

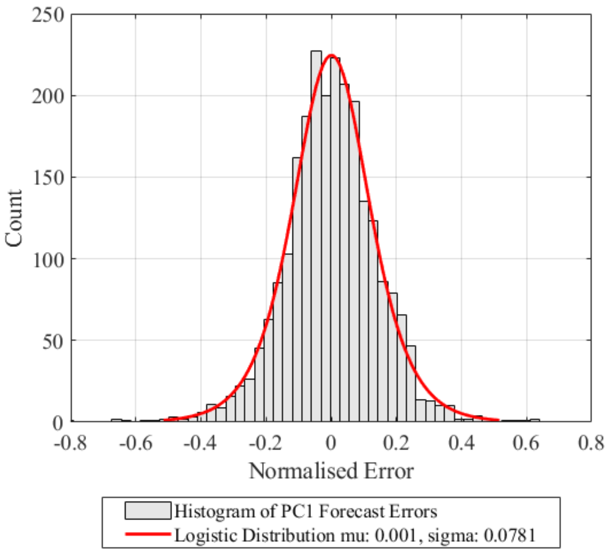


Figure 49 : Forecast error histogram and logistic distribution fit for first principal component.

The temporal correlations of these errors are shown in Figure 50. As the lead time increases, the proportion of forecast that contributes to the prediction increases and by extension the contribution to each principal component increases. This can be seen

through the colour gradation in Figure 50 as adjacent 3 hour timesteps become more correlated as the lead time increases.

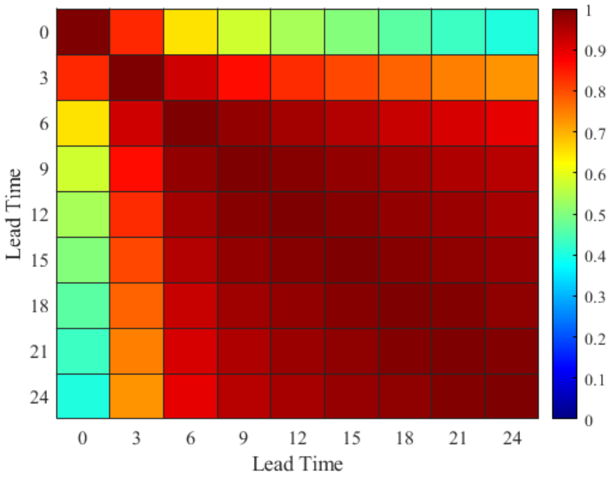


Figure 50 0 – 24 hour lead time error correlation for 1st principal component for 2020 in 3 hour timesteps. The colour bar describes Pearsons correlation coefficient

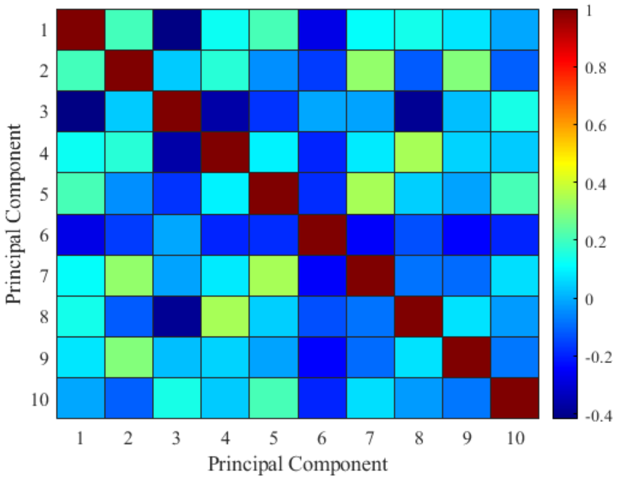


Figure 51 Error correlation for all features for 0 hour lead time for first 10 principal components in 2020. The colour bar describes Pearsons correlation coefficient.

Figure 51 shows the correlations between errors of different principal components for 0 hour lead time. It demonstrates a range of moderate negative and positive correlations between principal component errors. A covariance matrix is calculated on the time dependent errors across the year 2020. A small value (1e-6) is added to zeros term cells in the covariance matrix to make the matrix positive definite. Thereafter, Cholesky decomposition is performed, and a set of correlated errors can be generated using Eq. [25] - Eq. [28]

5-2.4. Time Dependent MCBA Wave Model Validation

The MCBA algorithm is applied to the IFS forecast data for the years 2021 to 2023. This is tested against CMEMS data for the same time period, using the independent and dependent error cases for the 22m depth location in Irvine Bay indicated by the purple triangle in Figure 42. 250 Mote Carlo samples are used for the forecast uncertainty. The predictions are made prior to nearshore transformation equations are applied. The results are grouped by variable and are shown for 24 hour lead time.

5-4.4.1. Significant Wave Height

The 24-hour significant wave height forecasts are shown in Figure 52. There is a good agreement between the forecasted and observed Hs for 24-hour lead time for both time dependent errors (CC = 0.918) and independent errors (CC = 0.922). All results for both error structures for the entire data set and for extreme values only (Hs > 1.5m) are shown in Table 8. There is no meaningful R2 value for Hs > 1.5m owing to the spread of predictions relative to the range of values observed. As such the values has been omitted from the table.

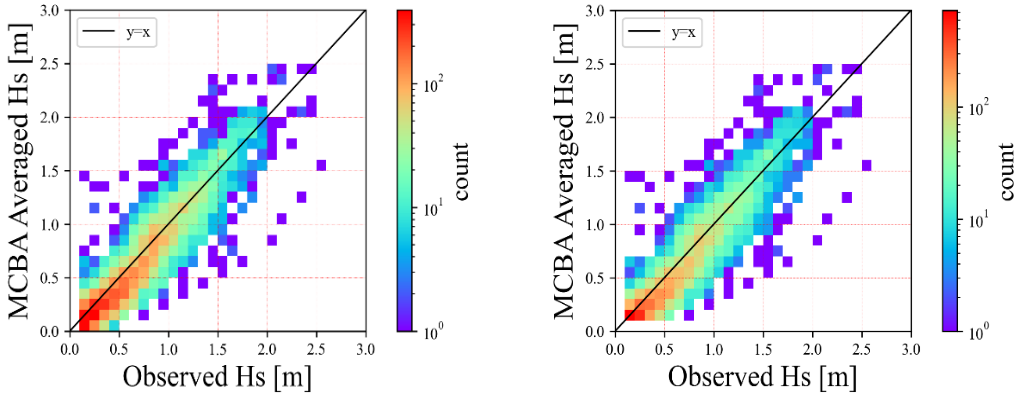


Figure 52: Significant wave height prediction against CMEMS test set for time dependent errors (left) and independent errors (right).

The occasions where there is a large discrepancy between forecasted and observed significant wave heights for the CMEMS test set correspond with large errors in the local wind forecast as shown in Figure 53. Large underpredictions are associated with large U10 forecast wind component underestimates, Figure 53 (left). Large overpredictions correspond with large V10 overestimates, Figure 53 (right).

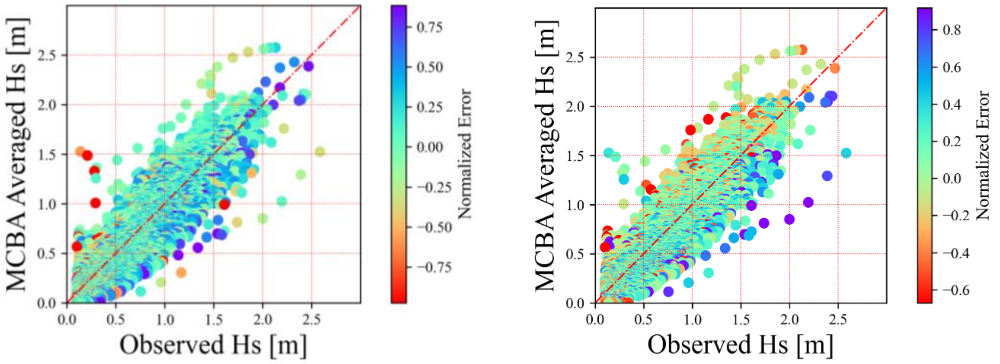


Figure 53: Significant wave height prediction against CMEMS test set for time dependent errors, with local U10 (left) and local V10 error (right). Normalised error is ERA local wind minus VC local wind.

Table 8 also demonstrates that there are no significant differences in the metrics R2, MAE, RMSE, SI or Bias between the two error structures. The small differences in these metrics are attributed to the Monte Carlo sample size of 250, as this value tends to infinity, the mean responses should converge to the same values. The significant differences are the percentage of observations inside the confidence interval and the boundary size. This is visually explored in the ranked significant wave height robust predictions shown in Figure 54.

	Ecorr		Eind	
Metric	Full Series	Hs > 1.5m	Full Series	Hs > 1.5m
R2 [-]	0.809	-	0.812	-
MAE [m]	0.137	0.217	0.136	0.204
RMSE [m]	0.185	0.289	0.184	0.269
Bias [m]	-0.046	0.146	-0.060	0.113
SI [-]	0.239	0.148	0.235	0.138
CC [-]	0.918	0.521	0.922	0.544
Inside 95% Confidence [%]	95.6	93.6	86.3	88.7
Boundary Size [m]	0.73	1.08	0.51	0.79

Table 8: Metric results for MCBA significant wave height averaged prediction for CMEMS test set.

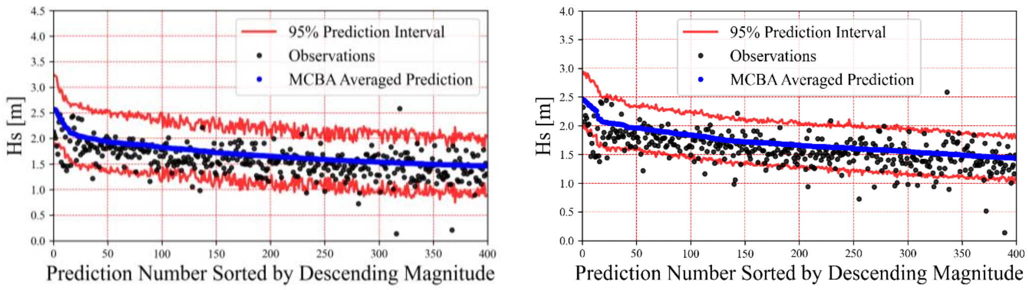


Figure 54: significant wave height robust prediction sorted from highest to lowest for highest 400 observations for time dependent errors (left) and independent errors (right).

It is clear from Figure 54 that not considering the correlation in the error structure leads to an under estimate of the prediction interval size that leads to the interval capturing 8.7% short of the desired 95% of observations for the full series, and 6.3% short for values over 1.5m. On the other hand, the time dependent error prediction interval captures 95.5% of all observations, and 1.4% short for values over 1.5m. Many of the other values fall just outside of the prediction interval and would likely be within the interval if a larger number of Monte Carlo samples is selected. In both cases the predictions run through the centre of the test data which is reflected in the low bias.

5-4.4.2. Mean Wave Period

The 24-hour mean wave period (Tm) forecasts are shown in Figure 55. There is a good agreement between the forecasted and observed Tm for 24 hour lead time for both time dependent errors (left) and independent errors (right). This is reflected in the strong CC values for time dependent errors (CC=0.918) and independent errors (CC=0.922) respectively. The Tm prediction metrics are recorded in Table 9.

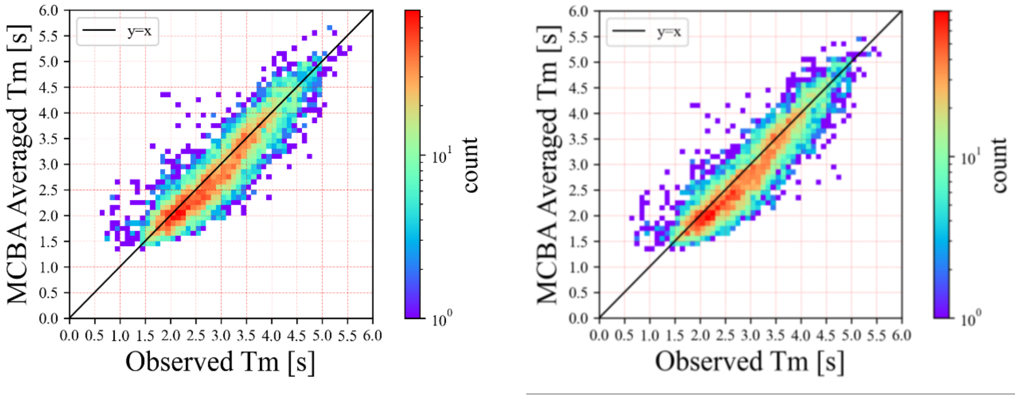


Figure 55: Mean wave period prediction against CMEMS test set for time dependent errors (left) and independent errors (right).

As for the prediction of the wave height, the occasions where there is a large discrepancy between forecasted and observed mean wave period for the CMEMS test set for both cases large errors in the local wind forecast. Similar to Hs, Table 9 shows no significant differences in R2, MAE, RMSE, SI, Bias or CC between the two error structures, with the small differences attributed to Monte Carlo sample size. Again, similar to Hs, there is a significant difference in the percentage of observations inside the prediction and the boundary size. The correlated error structure prediction interval captures 96.1% of observations for both the full observation set, and the cases where the accompanying Hs measurement is greater than 1.5m. The independent error structure falls short of the desired 95% of observations by 10.4% in both cases. As with Hs time correlated error structures are required to effectively characterise the prediction uncertainty. This is visually explored in the ranked mean wave period robust predictions shown in Figure 56.

	Ecorr		Eind	
Metric	Full Series	Hs > 1.5m	Full Series	Hs > 1.5m
R2 [-]	0.819	0.817	0.813	0.811
MAE [s]	0.257	0.258	0.263	0.264
RMSE [s]	0.341	0.341	0.346	0.347
Bias [s]	-0.094	-0.094	-0.127	-0.127
SI [-]	0.110	0.110	0.109	0.109
CC [-]	0.918	0.917	0.922	0.921
Inside 95% Confidence [%]	96.1	96.1	84.6	84.6
Boundary Size [s]	1.5	1.5	0.93	0.93

Table 9: Metric results for MCBA mean wave period averaged prediction for CMEMS test set.

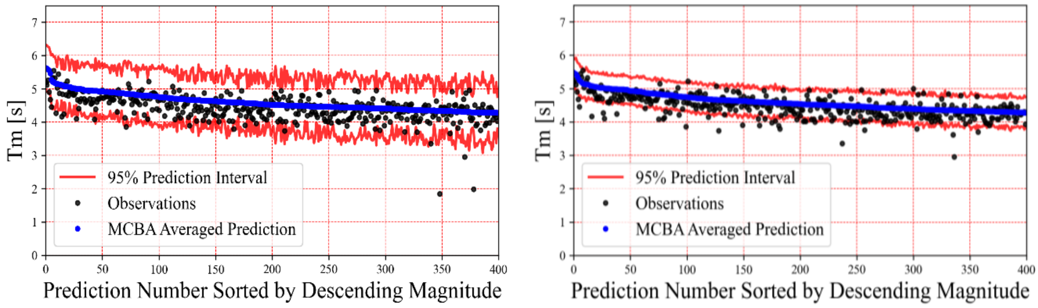


Figure 56: Mean wave period robust prediction sorted from highest to lowest for 400 largest Hs observations for time dependent errors (left) and independent errors (right).

In both cases the predictions run through the centre of the observations which is reflected in the low bias score recorded in Table 9.

5-4.4.2. Mean Wave Direction

The 24 hour mean wave direction forecasts are shown for both error structures in Figure 57. There is a good agreement between the forecasted and observed Dm for 24 hour lead time for both time dependent errors (left) and independent errors (right). The Dm prediction metrics are recorded in Table 10.

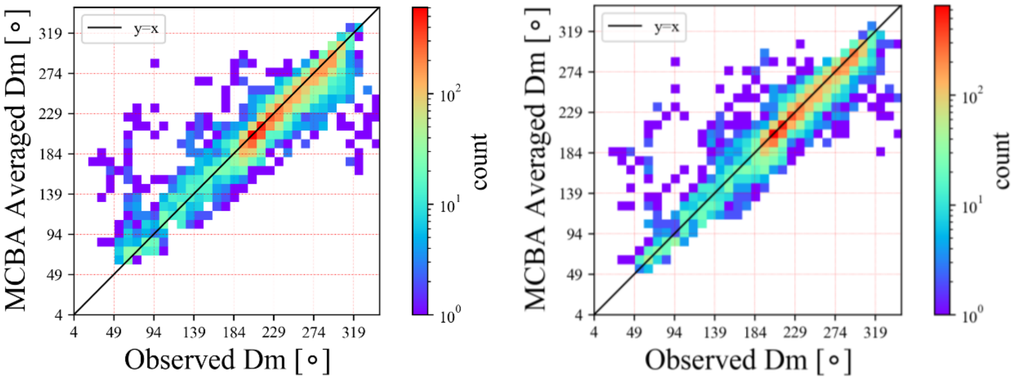


Figure 57: mean wave direction prediction against CMEMS test set for time dependent errors (left) and independent errors (right).

Like Hs and Tm, Table 10 shows only small differences in R2, MAE, RMSE, SI, Bias and CC between the two error structures for Dm. Unlike Hs and Tm, independent error prediction interval for Dm does a much better job of capturing the observations -1.2% and 2.9% short for the desired 95% for the full series and Hs values greater than 1.5m compared with ~10% for the other variables. It appears that the time dependent error structure is less critical to characterising Dm uncertainty than is the case for Tm or Hs.

	E_{corr}		E_{ind}	
Metric	Full Series	Hs > 1.5m	Full Series	Hs > 1.5m
R^2 [-]	0.884	0.878	0.908	0.859
MAE [°]	10	4	9	4
RMSE [°]	18	6	16	7
Bias [°]	-4	-3	-3	-3
SI [-]	0.076	0.024	0.068	0.027
CC [-]	0.943	0.954	0.954	0.943
Inside 95% Confidence [%]	96.7	96.5	93.8	92.1
Boundary Size [°]	67	24	46	19

Table 10: Metric results for MCBA mean wave direction averaged prediction for CMEMS test set.

The time dependent error prediction interval exceeds the desired percentage. Interestingly, the interval is much tighter for extreme values than for the full series meaning that there is less variance across the Dm prediction set when the waves are higher.

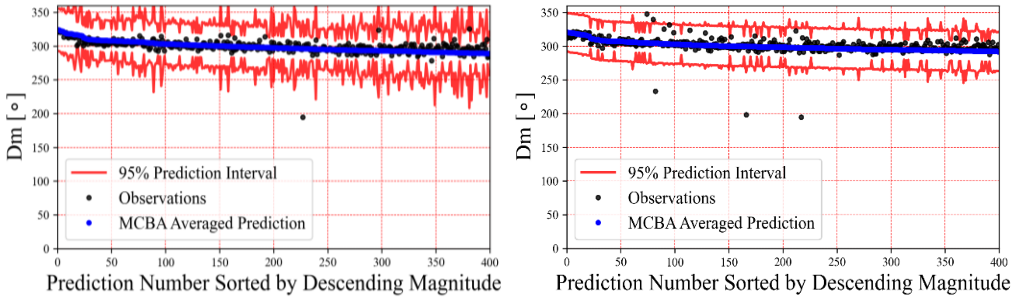


Figure 58: mean wave direction robust prediction sorted from highest to lowest for 400 largest Hs observations for time dependent errors (left) and independent errors (right).

Figure 58 shows the ranked mean wave period robust predictions for the test data set. As with Hs and Tm, the predictions run through the centre of the observations which is reflected in the low bias score recorded in Table 10. It is not obvious in this case whether the time dependent errors is advantageous over independent errors, as the independent error case almost satisfies the desired percentage of observations and is much more regular relative to the width of the prediction interval.

5-2.5. Surface Wave Forecasting Model Validation with Nearshore Transformation

The nearshore transformation model is validated using in situ significant wave height data gathered from a waverider buoy at Irvine, positioned at 51.617 decimal latitude, -4.767 decimal longitude at an approximate depth of 10m, from 15th December 2000 to 20th March 2001. By contrast the validation in subsection 4.5 do not utilise nearshore transformation equations are applied to reanalysis data. The period for which there is a test wave height dataset in the Firth of Clyde does not overlap with the available IFS forecast. As such, ERA5 data is used to make the predictions. This data was omitted from the training data set described in subsection 3.1.

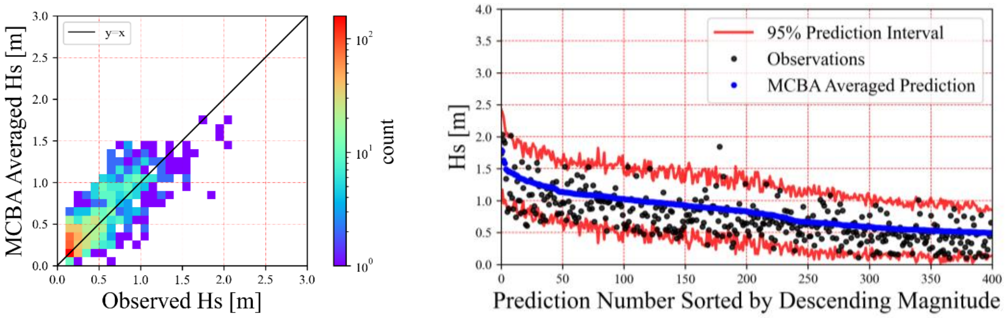


Figure 59: significant wave height averaged prediction against Irvine wave height data (left). MCBA significant wave height robust prediction sorted from highest to lowest for highest 400 observations in test set sorted by descending magnitude (right).

The left plot of Figure 59 shows good agreement (CC=0.865) between the forecasted significant wave height and observed Irvine Wave Buoy significant wave height considering that the model was trained on reanalysis data. There is a clear tendency for the MCBA method to overestimate wave heights. The metrics describing this data are recorded in Table 11. The MCBA method has a bias of 9cm for all significant wave height predictions and 15cm bias for forecasted wave heights above 1.0m. This threshold is dropped compared to the previous analysis to account for the small number of extreme values observed. The right image of Figure 59 shows the confidence interval capturing 88% of values, with a large number of additional values falling just outside the interval. The model has a tendency to over predict for 0.6m – 1.2m, which is likely due to discrepancies between the CMEMS data and the real data. There is no meaningful R2 values for the larger significant wave height bins owing to spread of predictions relative to the range of observations so this has been omitted from Table 11.

The nearshore step uses linear refraction–shoaling (Airy theory with depth-appropriate phase and group speeds, Snell refraction, and a shoaling factor) evaluated at still water level (tide + surge) depth, yielding a transparent, seconds-fast transform that integrates with MCBA so offshore uncertainty propagates to coastal Hs with minimal cost. Against the Irvine buoy (10 m) it shows useful skill (CC = 0.865, small positive bias); 95% coverage is ~88% with a tendency to overpredict in the 0.6–1.2 m band, consistent with reanalysis–in situ representativeness limits that a linear step cannot fully correct. The approach's advantages are speed, traceability, and parsimony (requiring only local depth and shoreline orientation), while its linear assumptions omit explicit treatment of depth-limited breaking, bottom friction, wave–current interaction, nonlinear transfers, diffraction/shadowing, and infragravity, which can bias extremes in complex settings.

By contrast, SWAN-class spectral models (SWAN, 2006) resolve the full directional–frequency spectrum with wind input, dissipation, and nonlinear interactions, offering higher physics fidelity over complex bathymetry and within surf zones but at minutes to tens of minutes per run—costly for large Monte Carlo ensembles. For the Clyde, the linear method is fit-for-purpose for rapid, uncertainty-aware warning; where morphology or breaking dominates, a hybrid is recommended—retain ANN+MCBA offshore and benchmark or bias-correct nearshore HnH_nHn with a stationary SWAN transect or small 2-D nest, comparing coverage, width, and runtime.

Because sustained, real-time buoy observations were not available for the Firth of Clyde during the study period, the wave model was trained on the long, internally consistent CMEMS reanalysis (1980–2019/22) and evaluated operationally against forecasts, with a short, non-contemporaneous Irvine waverider record (Dec 2000–Mar 2001) used only as a sanity check. This choice exploits record length and completeness but introduces nearshore representativeness differences when compared to in-situ data; in practice we see good point skill with a small positive bias and sub-nominal 95% coverage (~88%), including slight overprediction around 0.6–1.2 m. To handle this, we report scaled R2R^2R2 and, where local in-situ data become available, recommend an offline calibration that maps the reanalysis-trained output to the buoy distribution—linear de-biasing if differences are roughly uniform, or quantile mapping when bias varies across the range—applied to the full MCBA ensemble so the mean shifts without collapsing interval width. Upstream, operational IFS winds are already mapped into the ERA-5 PC space with an annual de-bias, so forecast-input bias is handled consistently; the calibration described here aligns the wave side to any future buoy measurements while preserving the uncertainty propagation that underpins the inundation forecasts.

Metric	Full Series	Hs > 1.0m
R^2 [-]	0.730^{*}	-
MAE [m]	0.131	0.258
RMSE [m]	0.174	0.313
Bias [m]	0.09	0.15
SI [-]	0.390	0.247
CC [-]	0.865	0.492
Inside 95% Confidence [%]	91.5	88.393
Boundary Size [m]	0.518	1.008

Table 11: Metric results for MCBA significant wave height averaged prediction for Irvine Wave rider Buoy test set. *Models are evaluated using scaled R2 due to the CMEMS propensity to overestimate in situ data at low wave heights for this location. Scaled R2 removes the influence of differing means and variances between the in situ and reanalysis data to give a more fair comparison of the level of agreement between the forecast and the real data.

Improvements are possible for regions that contain more in-situ data by correcting the reanalysis data using methods such as bias correction or quantile mapping, the latter is especially useful when discrepancies vary across the range of values, such as when larger differences occur along the range of the targets as is the case here.

5-3. Overtopping Forecasting Framework During Storm Ciara

The SEPA Coastal Hazard Mapping Study (SEPA, 2015) mapped several vulnerable areas along the Scottish coastline, one of which is the Firth of Clyde basin. This area is prone to coastal inundation due to its complex bathymetry and exposure to Atlantic generated weather systems. Along the coastline, the A78 trunk road connecting the ferry port of Stranraer to Glasgow and the rest of Scotland, is prone to annual road closures due to overtopping. This makes the area suitable for an early warning inundation forecasting system to inform responsible authorities of high-risk flood events in enough time for them to take meaningful action.

We aim to evaluate whether our model can accurately predict the overtopping event that led to the closure of the A78 road on February 9, 2020, with a 24-hour lead time. This case study focuses on the inundation caused by Storm Ciara from February 8 to February 10, 2020. The storm resulted in the closure of the northbound A78 between Largs and Skelmorlie in Ayrshire due to flooding. It also caused cancellations of rail services between Kilwinning and Largs, as well as between Kilwinning and Ardrossan, due to severe weather conditions. Reports from network rail also confirmed waves overtopping the sea wall at Saltcoats on the East coast of the Firth of Clyde at 9.59am on the 09/02/2020 that caused a cancellation in services (The National, 2020). The Transport Research Laboratory (TRL) report (Milne, 2017) analysed surge events in Millport from 1995 to 2013. It indicated that flooding is probable on the A78 if the flood potential value (FPV) exceeds 5.3m Chart Datum (CD). The report defines FPV as **Error! Reference source not found.**. This equation is based on the assumption that the height of breaking waves during a storm is at least equal to the storm surge height, with larger surges resulting in higher breaking waves at the shore. Given that the overtopping framework presented in this chapter explicitly accounts for wave characteristics, we can reformulate Eq. [47] with the total water level equation fromEq. [2]:

$$\boldsymbol{FPV = H_a + H_r + H_u} \qquad \text{Eq. [58]}$$

Therefore, we are using **Error! Reference source not found.** to define the likelihood of inundation. To do so, the uncertainty evaluation process outlined in Section 3.3 was repeated using all data for 2020 to 2022 for uncertainty analysis, removing February 2020 from the analysis to ensure independence for the case study test. The same approach is also applied to the surge forecasting model ensuring independence for that element as well. UK Hydrographic Office maximum and minimum tidal heights and times are taken for Ardrossan Harbour for February 2020 using tidetimes (TideTimes, 2024).

These maxima and minima are converted to an hourly timeseries through piecewise sinusoidal interpolation since tides are inherently sinusoidal. 7-day tidal forecast are also available using the UKHO Easy Tide function (UKHO, n.d.). Surge forecasts are made using the surge model ensemble (Macdonald, Tubaldi, & Patelli, 2025) and wind wave forecasts are made using the enhanced methodology proposed in this chapter. Nearshore wave transformations are made using the equations laid out in sub-section 3.4. Since the water level is at the coast, the nearshore wave transformation equations are performed to the combined depth of tide and surge and the runup equations are not applied. This is in part because they are not necessary for the TRL inundation formulation. However, there is scope here to develop bespoke overtopping model that consider the local bathymetry, obstacles, roughness and coastal defence structures but that is deemed to be out with the scope of this investigation. The operational process for inundation forecasting follows the flow diagram in Figure 38. The 24 Hour FPV forecasts are presented for storm Ciara in 6 hour steps from 08/02/2020 04:00 to 09/02/2020 10:00 in Figure 60 and Figure 61.

The top plots of Figure 60 show the top of the 95% prediction interval pass over the FPV flood threshold at 08/02/2020 11:00 indicating a low chance of flooding with 7 hours lead time at the next high tide. The middle plot of the same figure then shows this timestamp to no longer be a flood risk with a shorter lead time. The middle plot of Figure 60 forecasts likely flooding at 09/02/2020 10:00 with 24 hour lead time as TRL threshold is midway through the grey boundary. As the forecasts develop, the full FPV 95% confidence interval crosses the TRL threshold, see the bottom plot of Figure 60 for 18 hour lead time and the top, middle and bottom plot of Figure 60 for 12, 6 and 0 hour lead time respectively. These all indicate a likelihood of flooding that is > 97.5%.

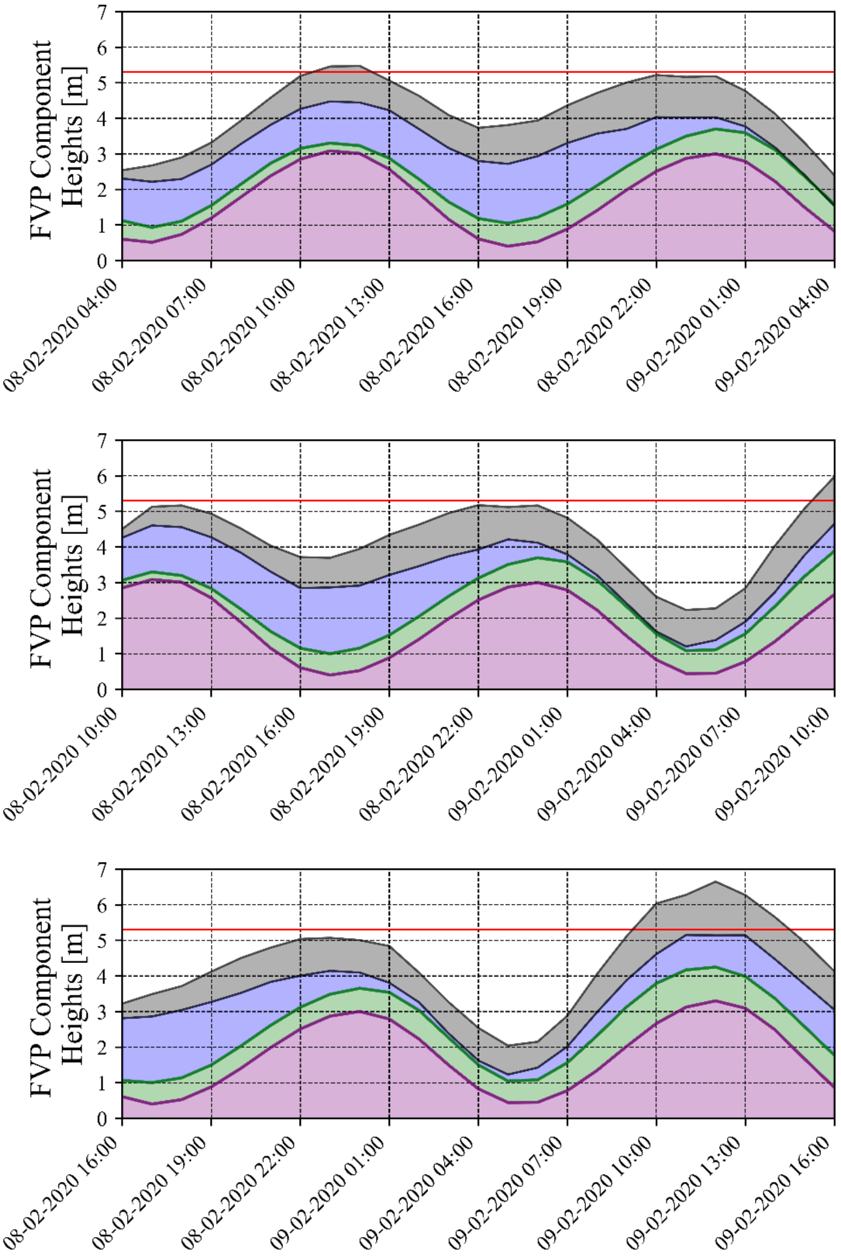


Figure 60: 24 hour FPV forecast made at 08/02/2020 04:00 (top), 08/02/2020 10:00 (middle) and 08/02/2020 16:00 (bottom). FPV elements are shown cumulatively, with astronomical tide (red), non-tide residual (green), wind waves (blue), FPV 95% confidence interval (grey) and TFL report likely inundation threshold (red line).

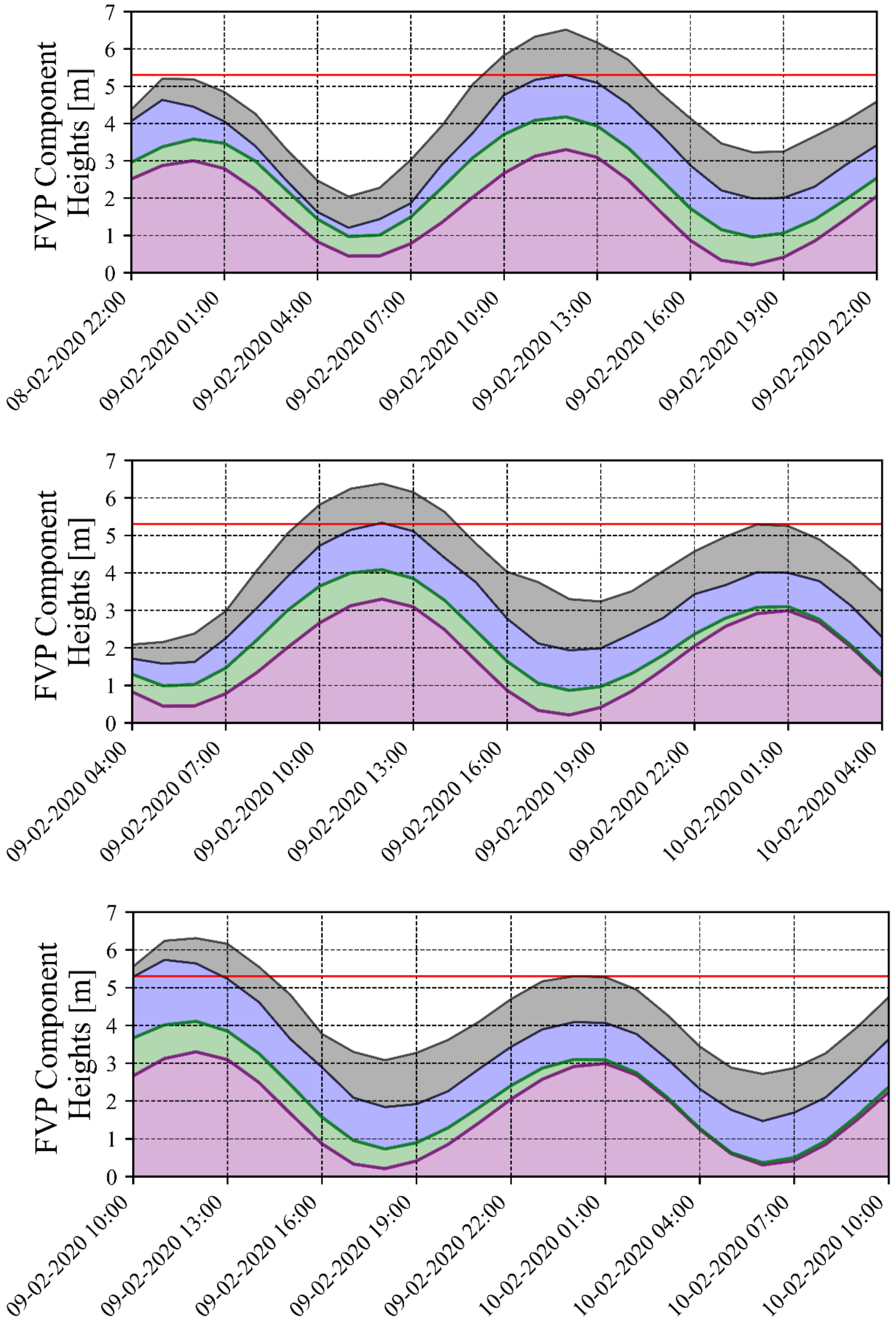


Figure 61: 24 hour FPV forecast made at 08/02/2020 22:00 (top), 09/02/2020 04:00 (middle) and 09/02/2020 10:00 (bottom). FPV elements are shown cumulatively, with astronomical tide (red), non-tide residual (green), wind waves (blue), FPV 95% confidence interval (grey) and TFL report likely inundation threshold (red line).

5-4. Overtopping Forecasting Framework Conclusions

This chapter presented a robust early-warning inundation tool that incorporates both aleatoric and epistemic uncertainty analysis throughout the entire forecasting process.

As part of this inundation framework, a wind-wave forecasting model was presented that evaluated the significance of time-dependent error structures concluding that effective characterisation of the wind-wave forecast uncertainty requires time dependent errors. Performance of the wind wave model was evaluated using forecasts provided by IFS and validated against CMEMS data. It has proven effective in predicting wind waves, as shown by the strong correlation between forecasted and observed values for significant wave height, mean wave period, and mean wave direction for the years 2021 and 2022. This methodology offers a reliable and computationally efficient way of forecasting wind wave characteristics with associated uncertainty. Thereafter the proposed methodology was validated at Irvine against using in situ data from a wave rider buoy in 2001.

Operational validation was demonstrated through the prediction of an historical overtopping in the Firth of Clyde basin during storm Ciara in February 2020. It has been shown that the proposed tool can successfully forecast inundation with a 24-hour lead time by providing detailed time varying information on the constituent parts of the forecast. To further validate and improve the proposed approach, future research should focus on extensive testing across diverse coastal areas, alongside comparisons with existing wave models and in situ measurements.

One of the advantages of this method over more regional inundation forecasting models is that it is tailored to a specific location and as such could be set up to monitor the flood risk for a piece of critical infrastructure. A location-specific model can be fine-tuned using historical data and local environmental factors (e.g., wind patterns, seabed topography, tidal influences) to improve prediction accuracy. This enhances the model's reliability in forecasting. In addition, the model is fast and required limited computational resources allowing to be deployed for the prediction of multiple strategic locations. This could provide accurate early warnings of potential hazards, allowing for better preparedness and mitigation strategies to protect infrastructure. Finally, given that ERA-5 and IFS are global datasets, the proposed framework can theoretically be applied to any location worldwide using historical time series or reanalysis data.

While the inputs are globally available, transferability is site-limited by coastal setting. Performance will depend on bathymetry and coastline geometry because the present nearshore step uses linear refraction–shoaling rather than a spectral nearshore model; depth-limited breaking, bottom friction, wave–current interaction, diffraction and complex shadowing are not represented explicitly, so strongly barred, reefed or headland-dominated coasts may see biased extremes even when bulk skill is acceptable.

Exposure and local wind–wave climatology also matter: the fetch geometry, lag structure and PCA loadings are fitted to each domain, and must be re-estimated when the wind footprint or sheltering differs (e.g., semi-enclosed basins vs open coasts).

In addition, representativeness of training/validation data can limit calibration near shore (e.g., CMEMS vs buoy), so local in-situ observations—where available—should be used to de-bias forecasts (linear or quantile mapping) without collapsing MCBA ensemble spread.

Finally, although time-dependent forecast uncertainty is propagated through MCBA, its size and correlation structure are location- and lead-time-specific and should be re-estimated during deployment.

In practice, to adapt for new regions it is desirable to rebuild the lag surface and PCA on local ERA-5/IFS, verify the linear nearshore step (and, where morphology is complex, hybridise with a short SWAN transect or alternative), and apply local de-biasing to align with any buoy or gauge data. This would preserve the framework's operational advantages while acknowledging morphology, geometry and wind-climate constraints on portability.

5-References

- C3S, C. (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Retrieved Dec 2022, from https://cds.climate.copernicus.eu/cdsapp#!/home
- CMEMS, E. (n.d.). Atlantic- European North West Shelf- Wave Physics Reanalysis. doi:https://doi.org/10.48670/moi-00060
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization.
- Leaman, C., Beuzen, T., & Goldstein, E. (2020). py-wave-runup:v0.1.10. doi:https://doi.org/10.5281/zenodo.2667464
- Macdonald, E. (2024). Bayesian Coastal Forecasting Toolbox. Retrieved from https://github.com/emacd-domain/Bayesian_Coastal_Forecasting
- Macdonald, E., Tubaldi, E., & Patelli, E. (2025). Robust Storm Surge Forecasts for Early Warning System: A Machine Learning Approach Using Adaptive Monte Carlo Bayesian Model Selection Algorithm.
- Milne, F. W. (2017). Assessing the risks to infrastructure from coastal storms in a changing climate: Project Report PPR800. Wokingham: Transport Research Laboratory.
- Oparanji, U., Sheu, R. J., Bankhead, M., Austin, J., & Patelli, E. (2017). Robust artificial neural network for reliability and sensitivity analyses of complex non-linear systems. Neural Networks, 96, 80-90.
- Reusch, D. B., Alley, R., & Hewitson, B. C. (2005). Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data. Polar Geography, 29(3), 188-212.
- SEPA, S. (2015). Flood Modelling Guidance for Responsible Authorities Version 1.1. 59.
- Stammer, D., Ray, R., Andersen, O., Arbic, B., Bosch, W., Carrere, L., . . . Erofeeva, S. (2014). Accuracy assessment of global barotropic ocean tide models. Reviews of Geophysics, 53(3), 243-282.
- SWAN. (2006). SWAN Technical Documentation. Delft University of Technology.
- The National. (2020, 2 9). Retrieved 10 10, 2024, from https://www.thenational.scot/news/18222609.storm-ciara-chaos-trains-ferries-brought-halt/
- TideTimes. (2024). Retrieved from https://www.tidetimes.co.uk/
- UKHO, U. (n.d.). Admiralty EasyTide. Retrieved 11 2024, from https://easytide.admiralty.co.uk/
- Visual Crossing. (2023, Jan). Visual Crossing Weather. Retrieved from https://www.visualcrossing.com/

Chapter 6

Road Fragility to Extreme Precipitation

This chapter is a reworking of a technical report titled 'Scottish Trunk Road Fragility to Extreme Precipitation', funded by the Scottish Roads Research Board.

Within the risk framework introduced in Chapter 1, the vulnerability term P(D|Hz) — the conditional probability of damage or disruption given a hazard — is central to understanding the resilience of transport infrastructure. For road networks, this represents the likelihood that a given precipitation event will result in flooding and consequent loss of functionality. Such losses can be severe: road flooding disrupts mobility, delays emergency response, interrupts supply chains, and imposes substantial direct repair costs alongside wider economic impacts. Prolonged closures on trunk routes can isolate communities, reduce economic productivity, and increase accident risk due to detours on unsuitable roads.

While the hazard term P(Hz) has benefitted from advances in extreme rainfall forecasting, the vulnerability of roads to these events has often been represented with deterministic thresholds, ignoring the uncertainty inherent in both hazard measurement and disruption reporting. This limits the reliability of network-scale risk assessments and constrains the ability to prioritise adaptation measures.

Chapter 3 introduced a probabilistic methodology for constructing uncertainty-based fragility curves — functional relationships between hazard intensity (e.g., precipitation depth) and the probability of disruption. This approach combines empirical disruption records with Monte Carlo sampling of hazard intensities to capture measurement and sampling uncertainty, and applies kriging metamodels to estimate vulnerability in unmonitored locations. By treating both the hazard input and the fragility relationship as probabilistic, the resulting curves provide a robust representation of P(D|Hz), suitable for integration into full risk assessments.

In the context of the Scottish trunk road network, these fragility curves quantify how the probability of flooding-related disruption increases with rainfall intensity, and express the uncertainty in that relationship. This enables decision-makers to evaluate vulnerability consistently across the network, even where historical disruption data are sparse, and to directly link rainfall forecasts to probabilistic risk estimates for infrastructure planning and emergency management.

6-1. Requirement for Road Fragility Curves to Extreme Precipitation

The Scottish trunk road network is a system of major roads in Scotland that are managed by Transport Scotland. The trunk road network includes some of the most important and busiest roads in Scotland, such as motorways, dual carriageways, and major single carriageway roads. These roads provide key connections between towns and cities, as well as access to ports, airports, and other important destinations. The Scottish trunk road network consists of over 3,500 miles of road and is Scottish Ministers' single biggest asset (Transport Scotland, The Trunk Road Network, Overview 2014). It has a gross asset value of over £20.8 billion and represents 6% of the total Scottish road network. It carries 35% of all traffic and 60% of heavy goods vehicles.

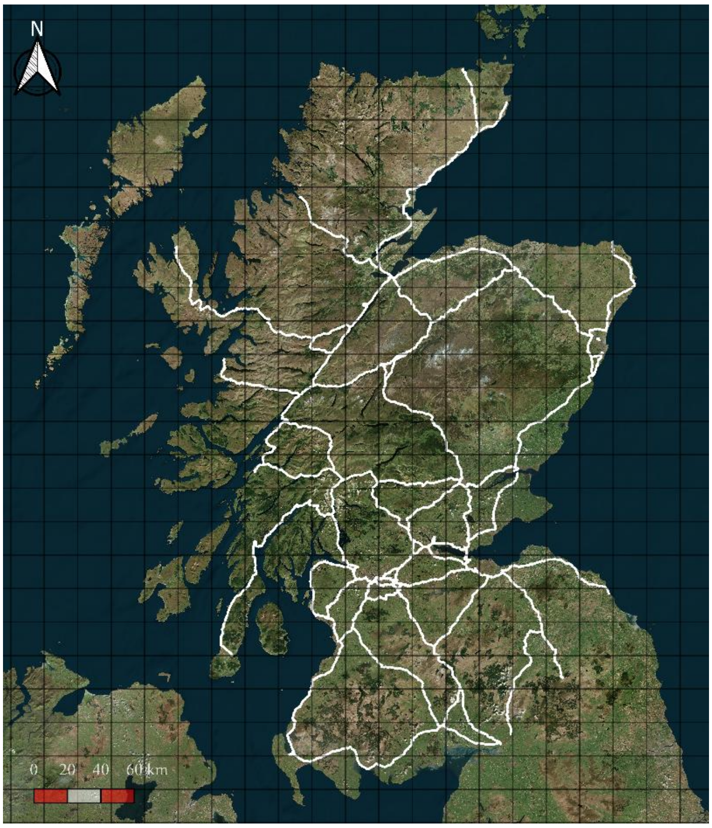


Figure 62: Map of Scottish trunk road network.

6-2. Disruption Events

Transport Scotland has developed an asset management systems to gather information on trunk road assets and aid decision making. Collaborative efforts with the Welsh Government led to the procurement of a single contract for an Integrated Road Information System (IRIS) in 2012. IRIS is a fully integrated Geographical Information System (GIS) map-based asset management system that links data on conditions, inventory, accidents, structures, drainage and maintenance. Transport Scotland currently uses IRIS functions along with other core systems to manage and maintain the trunk road network (Transport Scotland, Asset Management Systems and Data 2014).

As of August 2022, there are two companies that maintain the trunk road network in four regions of Scotland, known as the Operating Companies (OCs):

- BEAR Scotland - responsible for the South East and North West units
- Amey - responsible for the North East and South West units

Each OC is responsible for maintaining and improving the trunk roads in its area, including routine maintenance, winter maintenance, emergency response, and major projects (Transport Scotland, The Trunk Road Network, Operating Companies 2014). They work closely with Transport Scotland to ensure that the trunk road network is safe, efficient, and well-maintained. In the event of a road failure and the resulting call out of the operating company, the call out is logged in the IRIS Management of Incidents system (Transport Scotland, Asset Management Systems and Data 2014). It is worth highlighting that the true number of incidents that occur on the trunk road network is likely to be higher than is recorded in the IRIS database since the database only contains the events that are responded to by the operating companies. A further consideration is that the likelihood of attending an incident is unlikely to be uniform across all incident magnitudes as operating companies will prioritise the most severe events and so the probability that these will be represented in the IRIS database is higher. The variation in attendance probability with incident severity is not known. Table 12 provides the breakdown of the flood related events at the time of analysis spanning the period January 2015 to December 2021. The location of the 506 disruption events is shown in Figure 63.

Stage	Data Points
Iris Database	4898
Events that caused Disruption	569
Events that caused Disruption in window 01-01-2015 to 01-01-2021	506

Table 12: Breakdown of flood related events in the IRIS database.

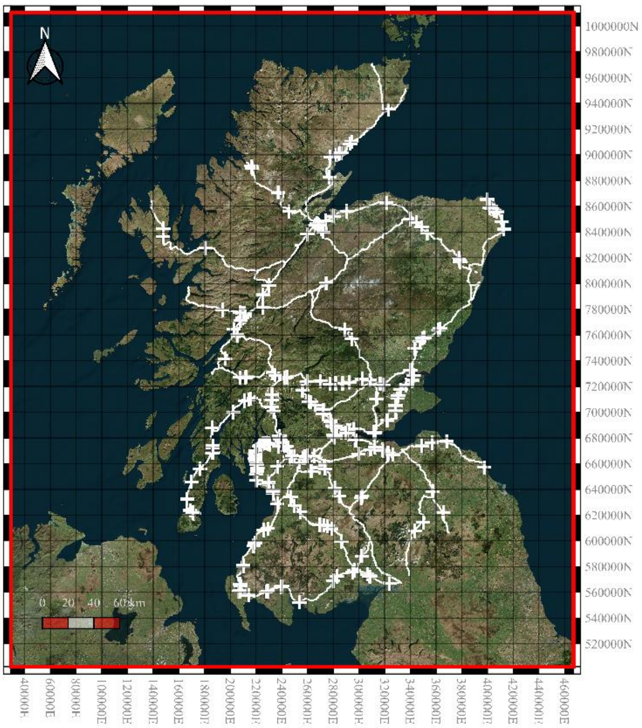


Figure 63: Location of events recorded as disruption caused in IRIS database and considered in this dissertation.

The disruption database records events that triggered an operating-company response. Events that were not observed, not reported, or did not meet the response threshold may therefore be missing, implying that the empirical disruption rate is a lower bound on the true failure probability. Because response rates ρ (the probability that a true disruption is logged) are not available for our study period or network, a correction is not applied, and any attempt to assign one is arbitrary. Throughout this chapter, the estimated fragility functions should be interpreted as:

$$\boldsymbol{P(F|I) = P(F\ logged\ |I)} \qquad \text{Eq. [59]}$$

which is a conservative estimate of $P(F\ true \mid I)$. Consequently, all risk metrics derived from Eq. [33] Eq. [59] should be treated as conservative lower bounds on disruption probability, pending availability of response-rate data to permit a formal bias correction.

6-3. Rainfall Estimation Under Uncertainty

What is required for the application of the methodology is a spatially and temporally continuous, uncertainty-quantified rainfall estimate at each disruption location and time step, to feed into the fragility framework outlined in Chapter 3. In order to build the fragility curves, a rainfall timeseries estimation must be made at each of the 506 event locations recorded in Table 12, spanning the length of the database from January 2015 to December 2021. To achieve this we have two primary sources of rainfall information: point rainfall data from SEPA weather stations (SEPA n.d.) and 1km Met Office Rain Radar Data from the NIMROD System (Met Office 2003); along with and latitude, longitude and altitude data (googleapis n.d.).

As outlined in the literature review

6-3.1. SEPA Point Rainfall Data

Point rainfall data was gathered for the stated time period for all locations shown in Figure 64 on a map of Scotland. The OS Easting and Northing of the SEPA and radar rainfall data were stated with each dataset. The altitude variables were obtained by using the Google Elevation API.

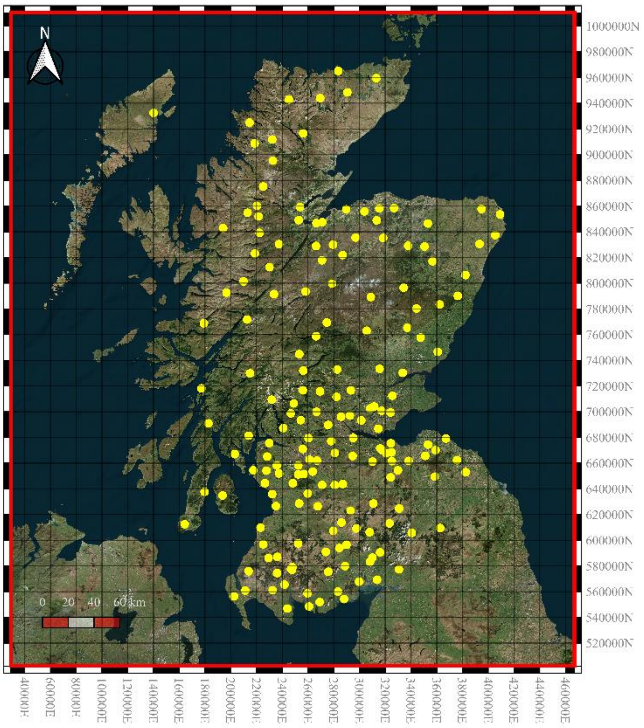


Figure 64: Location of SEPA weather stations.

Accompanying the data was a quality code for each value. Bad values were removed from the analysis, and good values accepted. The remainder of the unchecked rainfall values were evaluated through a simple process of comparison with concurrent radar estimations for an encompassing area. Thereafter, a 5km box was drawn around each station location and the maximum and minimum radar value across that area were recorded along with the radar measurement at each location. Any unchecked point radar values that fell out with the radar range for the 5km box were indexed. Next, errors between the radar measurements and the station observations were calculated for checked rain gauge measurements. Any unchecked value that fell outside the checked error margin for the given intensity at that location was considered anomalous. Any station reading that was out with the radar range for the surrounding 5km and was flagged as an anomaly was considered suspicious and removed from the dataset. For timesteps that did not have a corresponding radar measurement, the values were checked against the nearest rain gauge stations to evaluate if the measurement was reasonable. The procedure was composed to ensure that observations would only be removed if there was compelling evidence to do so. The numbers of values flagged by SEPA are recorded in Table 13. The number of values flagged by the data quality procedure outlined above are recorded in Table 14.

	Count	% of total
Checked = 0	9,702,084	53.05
Unchecked = 1	7,966,444	43.56
Suspect = 2	619,136	3.39
Total	18,287,664	100

Table 13: Data mask SEPA rainfall stations for the period 01/01/2015 – 31/12/2021.

Treatment of Unchecked Sepa Values

	Count	% of total
Not Suspect	7,952,281	99.82
Suspect	14,163	0.18
Total	7,966,444	100

Table 14: Data Quality Check of Unchecked Values

6-3.2. 1km NIMROD System Radar Rainfall Data

The 1km NIMROD product is a UK wide composite of quality controlled and corrected surface precipitation values from the UK's network of C-band radars. Values are evaluated at 5 minute resolution on a Cartesian National Grid. The data has undergone extensive processing to correct for various sources of radar error such as noise, attenuation and **(Stone, Harrison and Standing 2008)**. The 1km gridded composite C-Band radar images of precipitation are calibrated with rain gauges. For full details see the NIMROD Radar Processing document (Harrison, Driscoll and Kitchen 1998).

At each point in the grid the product uses the highest quality and resolution data available and as such the quality of the composite radar product is dependent on both the quality and resolution of the source data. This in turn is dependent on distance from the nearest radar site (Stone, Harrison and Standing 2008). For Scotland these radar stations are recorded in Table 15.

Name	Latitude	Longitude	Area
Munduff Hill	56°12’53”N	003°18’38”W	Fife
Hill of Dudwick	57°25’51”N	002°02’10”W	Aberdeenshire
Holehead	56°01’06”N	004°13’08”W	Stirling
Corse Hill	55°41’28”N	004°13’53”W	Strathclyde
Drium-a-Starraig	58°12’40”N	006°10’59”W	Isle of Lewis

Table 15: Name and location of NIMROD radar stations in Scotland.

The decrease of grided radar data quality decreasing with distance is demonstrated here in Table 15. The correlation between the SEPA hourly rainfall observations and NIMROD radar estimates significantly decreases as the distance from the location to the nearest radar station increases. This is not a surprising result, and it follows that the value of radar information as a correlated variable for the external variable kriging methods is location dependent.

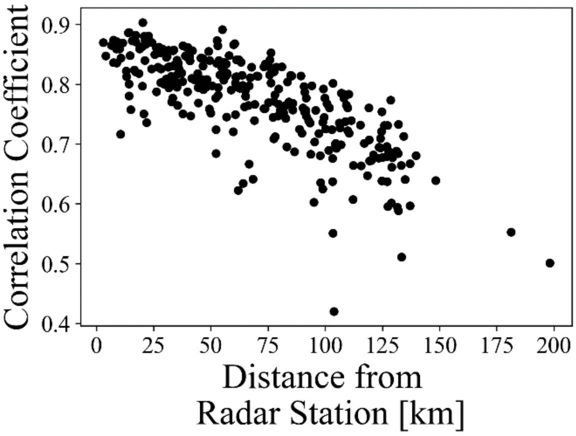


Figure 65: Correlation between hourly SEPA rainfall station observations and radar hourly rainfall estimates for the same location, against distance from the nearest radar station.

NIMROD 1km data was downloaded from CEDA (Met Office 2003) for the years 2015 to 2021. Files were cropped to the area of Scotland and hour average rainfall maps averaged from all available 5 minute resolution files for any given hour. To turn rainfall rates into mm/hr, NIMROD rates were divided by 32 to adjust values from the NIMROD system data saving standard which were integer precipitation rates in units of (mm/hr)*32. Figure 66 shows an example rainfall map from NIMROD for the 01/01/2015 averaged across the period 00:00 to 00:55. Table 16 shows the number of 5 minute NIMROD files present within each hour of the analysis.

	Year								
File Count	2015	2016	2017	2018	2019	2020	2021	total	Total [%]
0	295	121	131	129	148	283	20	1127	1.837
1	8	10	19	3	4	2	1	47	0.076
2	4	4	3	0	1	2	1	15	0.024
3	7	7	5	5	3	2	0	29	0.047
4	1	6	5	0	1	1	2	16	0.026

5	8	2	2	2	4	3	4	25	0.041
6	3	4	9	3	0	3	2	24	0.039
7	5	4	5	1	0	1	3	19	0.031
8	1	6	10	3	0	2	1	23	0.038
9	2	5	5	6	2	2	1	23	0.038
10	6	8	14	3	1	5	1	38	0.062
11	22	14	39	29	8	5	6	123	0.200
12	8398	8593	8513	8576	8588	8473	8718	59859	97.541
total	8760	8784	8760	8760	8760	8784	8760	61368	100

Table 16: Number of NIMROD 1km 5 minute resolution files present within each hour time step between 00:00 01/01/2015 and 23:55 31/12/2021. Total counts for each file number across the full timespan are recorded along with the percentage of the total number of hours observed across the full timespan.

Table 16 shows that over 98% of the individual hours modelled by NIMROD 1km radar have at least one measurement present, with over 97.5% having 12 x 5-minute increments present. In this analysis only hours where all 5-minute files were present for each hour were considered, any hours where one file was missing was discarded.

It is recognised that hourly aggregation smooths short, intense bursts typical of convective cells. The implication is a conservative bias: very high 5–15 min peaks may be diluted when averaged to an hour, potentially shifting the estimated fragility so that higher hourly totals are required to reach a given disruption probability. Two features of our workflow mitigate this. First, by anchoring to gauges and using Kriging with External Drift (KED), extremes observed at gauges are preserved at those points and propagated coherently in space rather than being uniformly damped. Second, the kriging prediction variance increases in locations and hours where the radar–gauge information is sparse or inconsistent; this larger uncertainty is carried forward in the Monte Carlo step of the fragility analysis, widening the P(D|I) intervals where short-duration peaks are most likely to be under-resolved.

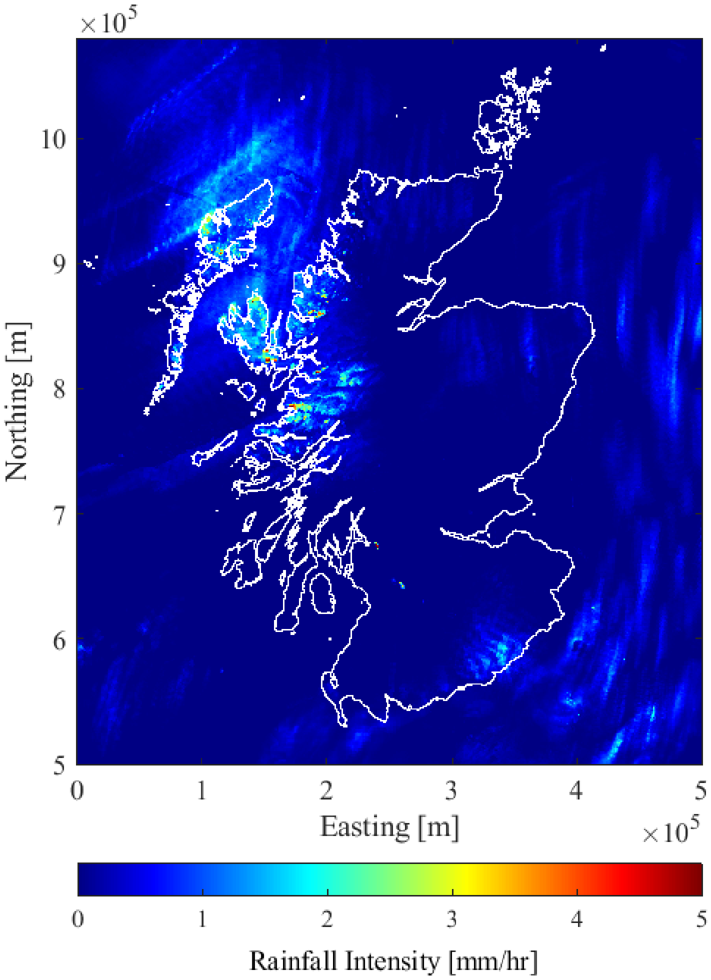


Figure 66: NIMROD 1km radar for Scotland at 01/01/2015 averaged between 00:00 and 00:55.

6-3.3. Rainfall Timeseries Estimation

Kriging is a natural fit for rainfall prediction in this study. In sparse gauge networks—typical of Scotland—it consistently outperforms simple interpolators such as Thiessen polygons or inverse-distance weighting. It also returns a location-specific uncertainty (the kriging variance) derived from the semivariogram. That allows us to propagate rainfall uncertainty into the fragility term P(D|I) within the probabilistic framework set out in Chapter 3. To this end, a total of 6 different kriging applications were performed and contrasted using 6 test locations described above. These are ordinary kriging of SEPA point data with latitude and longitude (OKLL), universal kriging of SEPA point data with latitude and longitude (UKLL), universal kriging of SEPA point data with latitude, longitude and radar (UKLLR), kriging of SEPA point data with altitude as an external drift variable (KEDA), kriging of SEPA point data with latitude, longitude and radar as external drift variables (KEDLLR), and regression kriging of SEPA point data with latitude, longitude and radar as regression variables (RKLLR) The different kriging flavours are described in Appendix A.1.2-A.1.6. The test locations were simulated for the year

2021. Two metrics were used to assess the quality of each kriging process: mean absolute error (MAE) and correlation coefficient (CC). Definitions are stated at the end of Chapter 3.

As demonstrated by Figure 65, radar skill in our domain decays rapidly with range and is further degraded by terrain blockage, bright-band contamination and storm morphology. The resulting bias between radar and rain gauge station is therefore non-stationary—it varies across space and time—so a single (or even coarsely stratified) pre-correction applied to the radar field cannot remove the local errors that matter at road scale. In practice, that leaves two choices. The first is a domain-wide bias fix: it improves the mean but leaves substantial local residual error, forcing us to carry a large, conservative uncertainty everywhere and risking misplacement or damping of heavy-rain maxima that drive disruption. The second is a spatially aware correction — such as KED — which anchors estimates to gauges while letting radar supply the spatial structure. KED locally debiases the field (nearby gauges pull the estimate toward the correct level), naturally down-weights radar where it is least reliable, and preserves gauge-observed extremes, which is critical for estimating vulnerability. Crucially, KED returns a location-specific prediction variance, giving us a principled, pixel-wise uncertainty that we propagate directly into the fragility analysis $P(D|I)$.

An interesting result of the kriging methodologies is that different methods work to different degrees for different locations, as shown in Table 17. OKLL is the only model that does not outperform radar estimations alone, indicating kriging's usefulness in improving rainfall estimations compared with radar and justifies its selection as a methodology. Broadly speaking, inclusion of NIMROD data improved the accuracy of rainfall predictions with the exception of Tobermory for which UKLL performed best. A comparison of Tobermory and Laxdale, which have similar proximities to other point rainfall stations, demonstrates the influence that radar station proximity has on improving hourly rainfall predictions.

As a side point, if new point rainfall stations are to be established, preference should be made for regions where there is not adequate radar cover as this would have the most significant improvement on nation-wide rainfall interpolation. Similarly, new radar stations should be deployed for regions that are not well covered by point rainfall stations for the same reason.

Since altitude as an external drift variable did not improve the estimations compared to UKLL for any metric for any location, there have been no further estimations using altitude. While there has been a considerable amount of research into the significance of altitude for rainfall interpolation over longer timespans such as monthly and annual rainfall, this did not seem to be the case for hourly rainfall in this investigation. However, this observation is inconclusive due to the small sample size and short time period of observation. This is an interesting area for future investigation.

Station Name	**Distance to Nearest SEPA station [km]**	**Distance to Nearest NIMROD station [km]**	**Kriging Type**											
			Ordinary (Lat, Lon) OK_{LL}		**Universal** (Lat, Lon) UK_{LL}		**External Drift** (Altitude) KED_{A}		**External Drift** (Radar) KED_{LLR}		**Regression** (Radar) RK_{LLR}		**Radar Alone**	
			CC	MAE* [mm]	CC	MAE* [mm]	CC	MAE* [mm]	CC	MAE* [mm]	CC	MAE * [mm]	CC	MAE * [mm]
Newton of Falkland	5.1	8.8	0.85	0.69	0.85	0.69	0.85	0.69	0.80	0.61	0.88	0.66	0.61	1.76
Esslemont House	13.4	8.7	0.74	0.87	0.76	0.84	0.74	0.88	0.89	0.54	0.89	0.57	0.88	0.64
Laxdale	34.3	56.8	0.38	1.53	0.55	1.34	0.51	1.52	0.72	1.09	0.72	1.09	0.72	1.19
Drummore of Cantray	5.2	118.4	0.77	0.98	0.78	0.96	0.78	0.96	0.81	0.78	0.72	0.93	0.67	1.04
Townfoot (Glencaple)	14.0	121.1	0.79	0.96	0.79	0.92	0.79	0.95	0.78	0.94	0.82	0.99	0.63	1.55
Tobermory	32.5	133.3	0.34	1.62	0.57	1.21	0.49	1.41	0.52	1.28	0.47	1.35	0.33	1.65
mean	-	-	**0.65**	**1.11**	**0.72**	**0.99**	**0.69**	**1.07**	**0.75**	**0.87**	**0.75**	**0.93**	**0.64**	**1.31**

*MAE values were calculated for all observed hourly rainfall values above 1mm.

The mean result across all stations is displayed with the best performing method at each location and overall underlined.

Table 17: Kriging prediction metric results for all test point rainfall locations for 2021.

KEDLLR has the joint highest mean CC across all locations and the best mean MAE across all test locations and thus is selected to estimate the rainfall intensity at the event locations for January 2015 to December 2021. An average CC of 0.75 for KEDLLR demonstrates very good correlation between the observed and predicted hourly rainfall.

Regarding spatial scale, radar provides an areal estimate (pixel size ~5 km), whereas gauges are point measurements. A 5 km average can be lower than peak point intensities within the pixel, especially for small convective cores. However, the asset at risk is a road segment with kilometre-scale spatial footprint and drainage contributing areas that are not strictly point-like; for network screening and planning, an areal intensity is therefore a reasonable proxy for the load experienced by the system. Importantly, our KED product is evaluated at the road location: gauges locally debias the radar drift, so where nearby gauges exist the estimate reflects point-scale behaviour; where gauges are distant, the radar drift dominates and the model reports larger prediction variance, which again is propagated into the fragility curves and results in more conservative uncertainty bands.

6-4. Scottish Trunk Road Fragility Curves

After processing hourly data, SEPA and Radar intensities have been also aggregated into 3 hour, 6 hour, 12 hour and 24 hour periods with an hour timestep and kriging is carried out on each timestep. Hence 5 datasets were produced, one for each time window with its associated variance estimation.

The methodology outlined in 3-2 is followed for the data generated by the process outlined in 6-3, with the intensity measure used as mm of rainfall per hour. To ensure there was a sufficient population of points for the analysis, the intensity measures were divided into bins. For example, the probability of failure due to 10-12 mm of hourly rainfall was determined by dividing the number of failures recorded for that intensity in the IRIS database by the total instances of 10-12 mm rainfall observed at locations where failures occurred. Given that the IRIS database did not contain the time that the event occurred but only the date of the response time, the intensity measures considered were the maximum observed for the intensity measure in the window spanning 24 hours either side of midnight on the event date. The theoretical maximum used is the highest observed UK value of rainfall for each duration is incorporated as a point, with a corresponding disruption probability of 1. This assumes that if the largest rainfall ever recorded in the UK were to occur on a road section, failure would be inevitable. The UK maximums are recorded in Table 8.

	UK Maximum [mm]
Hourly	92
3 Hour	178
6 Hour	187*
12 Hour	204*
24 Hour	238

Table 18: Met Office UK rainfall records.

*Linearly interpolated value

The results are recorded for peak, 3 hour cumulative, 6 hour cumulative, 12 hour cumulative and 24 hour cumulative rainfall. For the peak rainfall, the peak event in the 24 hours before and the day of the recorded event is selected since the database did not contain a time stamp for the disruption event.

For the cumulative windows, the totals were calculated with a sliding one hour timestep and then the number of events was divided by the size of the window to ensure that all the extreme events were captured, and the results were not distorted by an arbitrary starting position. As with the peak rainfall the largest aggregate rainfall in the 24 hours before and the day of the recorded event was selected.

Each section contains four figures: a histogram showing the distribution of magnitudes associated with each event (including the 95% upper and lower confidence bounds calculated from uncertainty curve based on the distance from the event location to the nearest radar station), an empirical disruption probability curve, the fitting of multiple distributions through the upper and lower estimates and a probability box from the maximum and minimum values associated with the group of distributions.

6-4.1. 1 Hour (Peak)

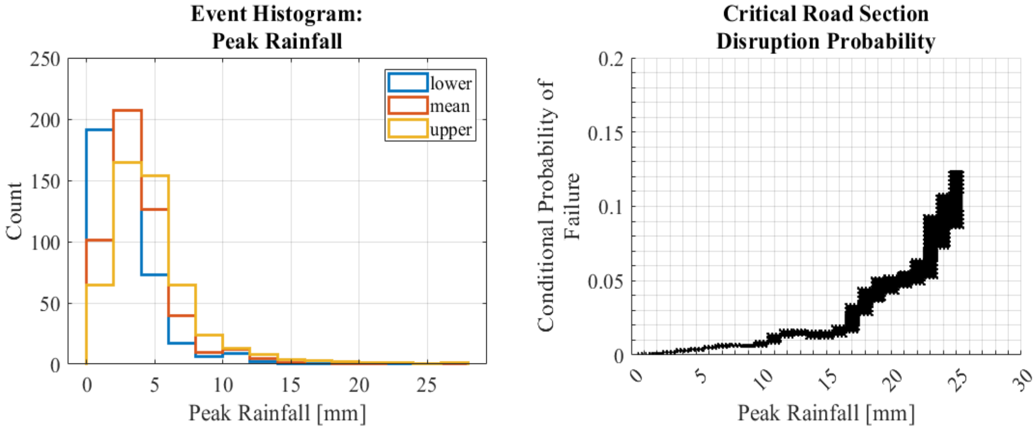


Figure 67: Histogram of mean and lower and upper 95% confidence bounds for peak predictions associated with the 506 disruption events (left). Empirical Disruption/Conditional Probability Curve for Peak Rainfall (right).

The one hour fragility curve (right plot of Figure 67)—conditional disruption probability versus peak hourly rainfall—shows the expected pattern of near-zero probabilities at low intensities, a sharp rise through the mid-range, and wider uncertainty in the upper tail where observations are sparse. Its sensitivity is driven mainly by how rainfall is estimated and paired to events. Because intensity is derived from gauge-anchored fields (with radar as drift) at an hourly resolution, any systematic under- or over-estimation of rainfall acts like a horizontal shift of the curve; this matters most in the steep mid-section, where small changes in intensity translate into noticeable changes in disruption probability. Hourly aggregation, chosen to match the gauge record and operational timescales, inevitably smooths short convective bursts, which can dilute true peaks and bias the curve slightly to the right—i.e., higher hourly totals are needed to reach the same probability. Spatial averaging from ~5-km radar pixels can have a similar effect for compact storms; where gauges are distant or radar skill is weak, the estimation variance increases and is carried forward, so uncertainty bands widen rather than conveying false precision. The stepwise appearance reflects binning of intensities: narrower bins expose within-bin gradients but raise variance, especially in the upper tail, while wider bins stabilize estimates at the cost of smoothing detail. Event-pairing choices (e.g., the temporal window used to associate rainfall with a recorded disruption) can modestly flatten the curve if a disruption is matched to a nearby but non-causal peak; conversely, tightening the window steepens the mid-range at the risk of missing causal rainfall where timing is uncertain. Finally, because the database contains only operator-responded events, the entire curve should be read as a conservative lower bound on true disruption probability; unknown reporting rates would primarily produce a vertical uplift of the curve rather than altering its shape. In short, the curve is most robust at low intensities and where gauge control is strong, most sensitive to small horizontal biases around the steep mid-range, and most uncertain in the upper tail due to data sparsity—sensitivities that are made explicit by propagating the rainfall-estimation variance into the probability bands.

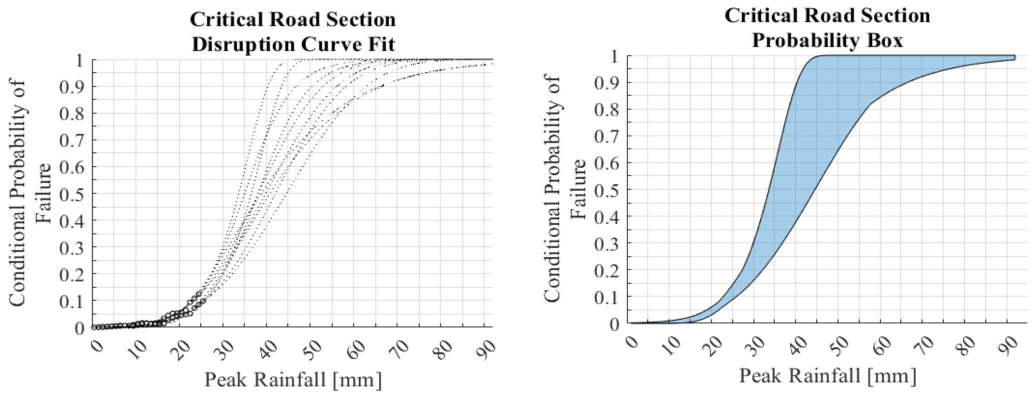


Figure 68: Cumulative distributions fitted to empirical conditional probability values for peak rainfall including UK maximum point (left). Probability box formed by the maximum and minimum values of all fitted distributions.

The probability box (Figure 68, right) wraps an imprecise fragility curve: it shows the range of disruption probabilities that are consistent with the data and modelling assumptions once measurement and model uncertainty are propagated. Because the empirical record covers only about one-fifth of the intensity range (roughly the low end), the envelope is tight where you have data support—probabilities remain near zero and the bounds are close together below ~20 mm—and then widens rapidly once you move into the extrapolation regime. The broad band between roughly 25–60 mm is dominated by epistemic uncertainty (limited or no observations), so small changes in how rainfall is estimated or paired could shift the curve substantially there. The lower bound rising steeply reflects plausible "early-failure" behaviour; the upper bound's slower rise allows for more robust sections that fail only at higher intensities. At the very high end the band narrows again as both bounds approach one, which is driven by the monotonic constraints of the fragility model rather than dense data. Practically, interpret intensities within the observed range as well-constrained and treat probabilities in the mid-range as envelopes for decision-making (plan for both early- and late-failure scenarios). Tightening this p-box would require more events at higher intensities (or pooling across sites/seasons), finer-scale rainfall information, and/or stronger structural priors.

6-4.2. 3 Hour Cumulative

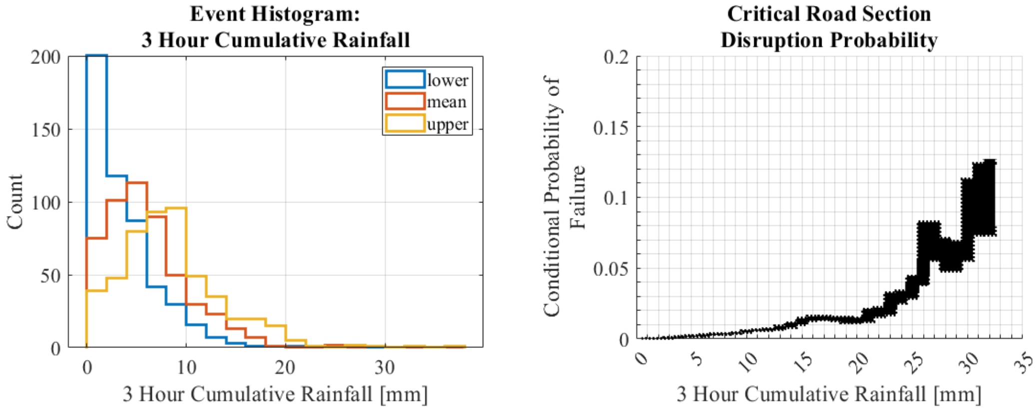


Figure 69: Histogram of mean and lower and upper 95% confidence bounds for 3 hour cumulative predictions associated with the 506 disruption events (left). Empirical Disruption/Conditional Probability Curve for 3 hour cumulative rainfall (right).

The 3-hour empirical curve (Figure 69, right) retains the same overall shape as the 1-hour—flat at low totals and then a stepped rise—but the transition is smoother and shifted to higher cumulative amounts (first noticeable lift ~12–15 mm over 3 h, with the largest steps appearing beyond ~25–30 mm). Because totals are accumulated over three hours, small convective spikes that drove sharper steps in the 1-hour curve are damped, so the mid-range slope is less abrupt and the staircase appears more coherent. Peak probabilities top out at a similar level (~0.15), but are reached at higher cumulative rainfall, indicating that disruption likelihood integrates sustained rainfall rather than short isolated bursts.

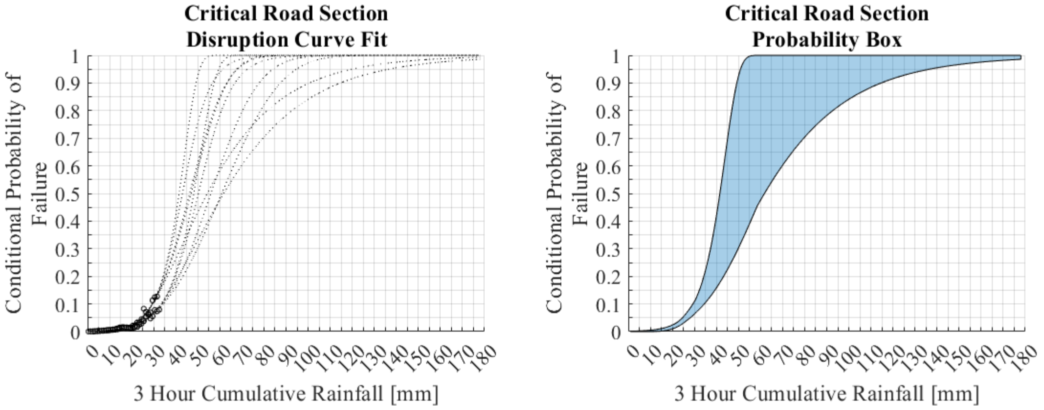


Figure 70: Cumulative distributions fitted to empirical conditional probability values for 3 hour cumulative rainfall including UK maximum point (left). Probability box formed by the maximum and minimum values of all fitted distributions.

The 3-hour p-box (Figure 70, right) is tighter at very low totals—reflecting strong agreement that failure is unlikely under modest 3-hour accumulations—but it widens over a broader intensity span, with the envelope extending well past 100 mm. Compared with the 1-hour p-box, the lower bound rises more gradually and the upper bound approaches unity more slowly, signalling greater imprecision about when prolonged rainfall tips sections into failure. This wider middle band is consistent with extrapolation beyond the empirically supported range and with additional smoothing from temporal aggregation; effectively, the 3-hour view trades short-term sharpness for a more integrated signal of stress, but pays with a larger uncertainty envelope across the mid–high totals.

6-4.3. 6 Hour Cumulative

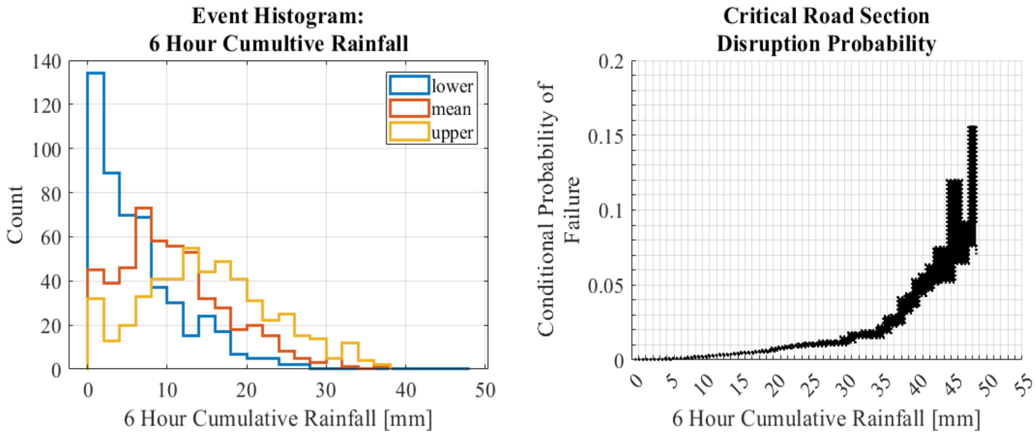


Figure 71: Histogram of mean and lower and upper 95% confidence bounds for 6 hour cumulative predictions associated with the 506 disruption events (left). Empirical Disruption/Conditional Probability Curve for 6 hour cumulative rainfall (right).

At 6 hours the empirical disruption curve (Figure 71, right) shifts decisively to higher cumulative totals, with the first discernible lift only in the low-20s (mm) and most of the probability mass appearing beyond ~40–50 mm. The staircase is more compressed toward the right-hand side than in the 1-hour case: rather than many small steps across the mid-range, there's a delayed but concentrated rise near the upper end, peaking at similar probabilities (~0.15–0.17). This pattern points to a section that tolerates intermittent bursts but becomes vulnerable when rainfall is sustained long enough to deplete storage and overwhelm drainage; in other words, the "dose" over several hours, not short spikes, is the governing load.

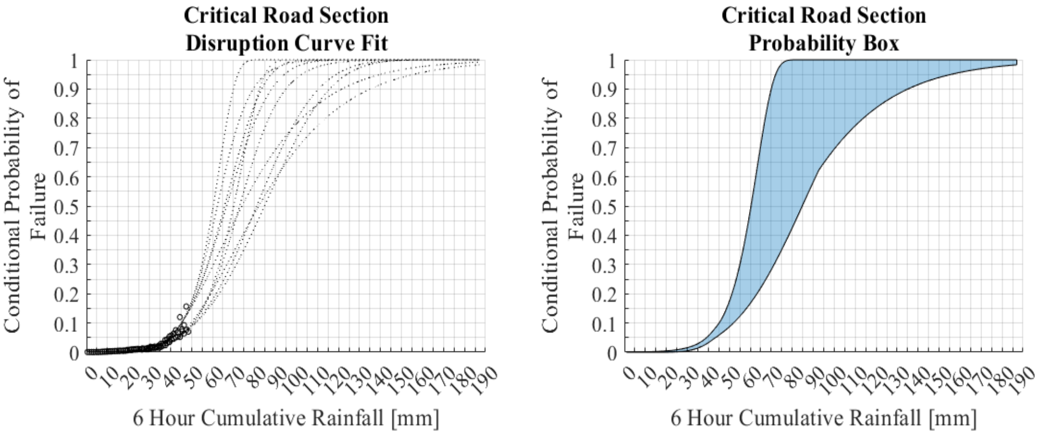


Figure 72: Cumulative distributions fitted to empirical conditional probability values for 6 hour cumulative rainfall including UK maximum point (left). Probability box formed by the maximum and minimum values of all fitted distributions.

The 6-hour envelope (Figure 72, right) is tight at very small totals, then widens over a much broader intensity span than the 1-hour view, with a pronounced asymmetry: the lower bound surges toward high probabilities around ~35–45 mm, while the upper bound climbs more gradually well past 100 mm. This indicates substantial imprecision about the endurance threshold under prolonged wetting—some plausible histories lead to early failure once accumulations pass ~40 mm, while others remain resilient until much larger totals. Relative to

1-hour, the band is displaced to the right and remains wide for longer, reflecting both the limited empirical support at high 6-hour totals and the greater scenario diversity in how sustained rainfall translates to disruption.

6-4.4. 12 Hour Cumulative

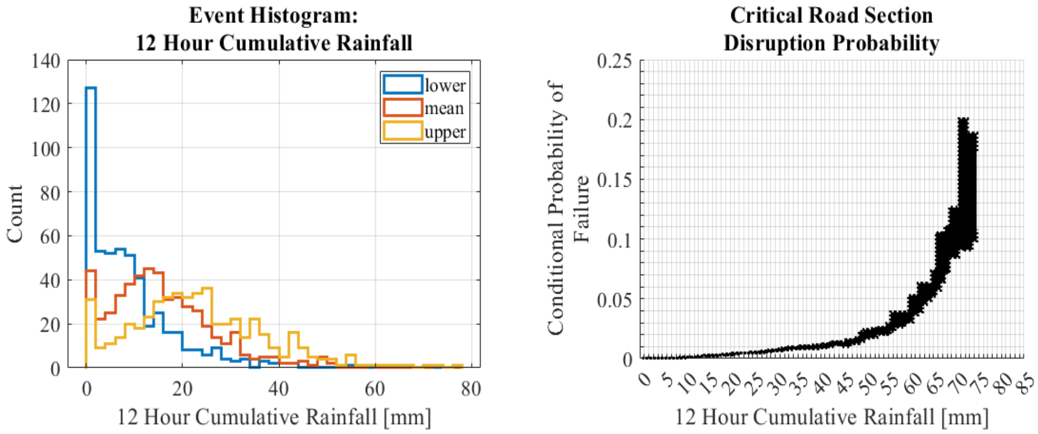


Figure 73: Histogram of mean and lower and upper 95% confidence bounds for 12 hour cumulative predictions associated with the 506 disruption events (left). Empirical Disruption/Conditional Probability Curve for 12 hour cumulative rainfall (right).

At 12 hours the empirical curve (Figure 73, right) shifts furthest to the right and behaves more like a threshold: probabilities stay negligible until accumulations approach ~60–70 mm, then rise sharply with most of the staircase compressed between ~80 and 90 mm, peaking a little higher (~0.20) than the shorter windows. Compared with the 1-hour curve, short spikes play almost no role; disruption likelihood is dominated by sustained loading, consistent with storage depletion, rising groundwater and drainage back-up acting over many hours. The steeper late rise and tighter clustering of steps indicate that once a cumulative endurance level is reached the section transitions quickly from low to appreciable failure probability.

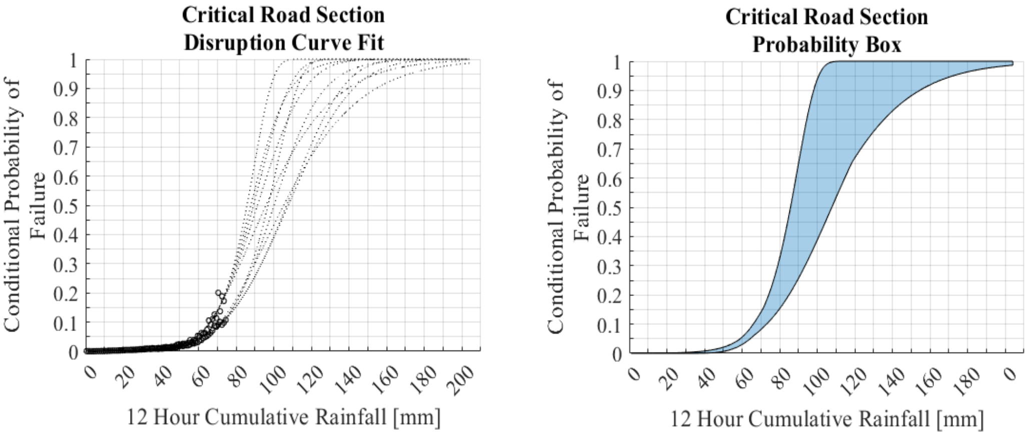


Figure 74: Cumulative distributions fitted to empirical conditional probability values for 12 hour cumulative rainfall including UK maximum point (left). Probability box formed by the maximum and minimum values of all fitted distributions.

The envelope (Figure 74, right) remains narrow through moderate totals and then opens abruptly in the 70–110 mm band, reflecting limited data support and genuine ambiguity about the endurance threshold under prolonged wetting. Relative to 1-hour, the lower bound "switches on" later but climbs more decisively, while the upper bound increases more gradually toward unity across a long tail—signalling a wider range of plausible behaviours once accumulations are very high. The result is a right-shifted, asymmetric p-box: tight at low totals, broad across the endurance range, then slowly narrowing again at the extreme end as monotonic constraints drive both bounds toward one.

6-4.5. 24 Hour Cumulative

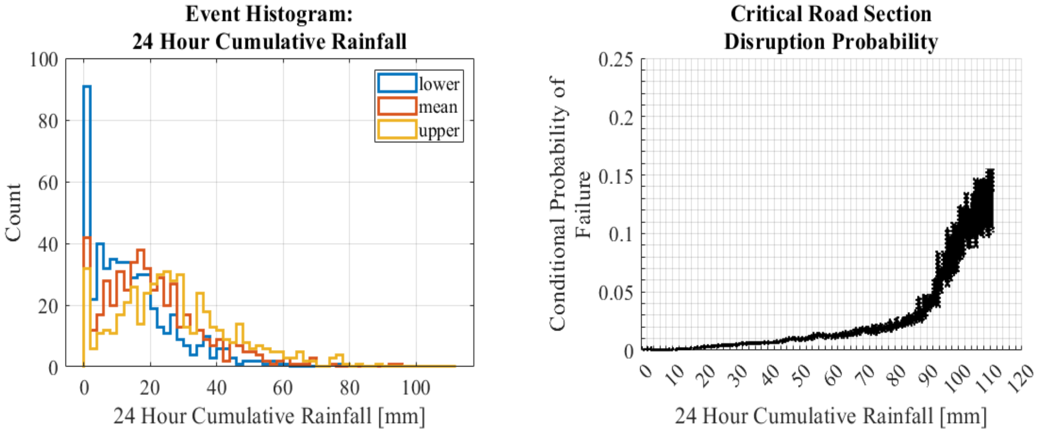


Figure 75: Histogram of mean and lower and upper 95% confidence bounds for 24 hour cumulative predictions associated with the 506 disruption events (left). Empirical Disruption/Conditional Probability Curve for 24 hour cumulative rainfall (right).

Over 24 hours the empirical curve (Figure 75, right) shifts furthest right and behaves most like a cumulative-load threshold: probabilities stay negligible until totals approach ~90–100 mm, then climb over a compact band to peak values (~0.15) around 105–115 mm. Compared with the 1-hour curve, short spikes are irrelevant; vulnerability is governed by day-long wetting that saturates verges and cuttings, fills storage, and sustains inflow to drains. The staircase is compressed near the upper end, signalling that once a 24-hour accumulation level is breached the section moves quickly from little risk to appreciable disruption likelihood.

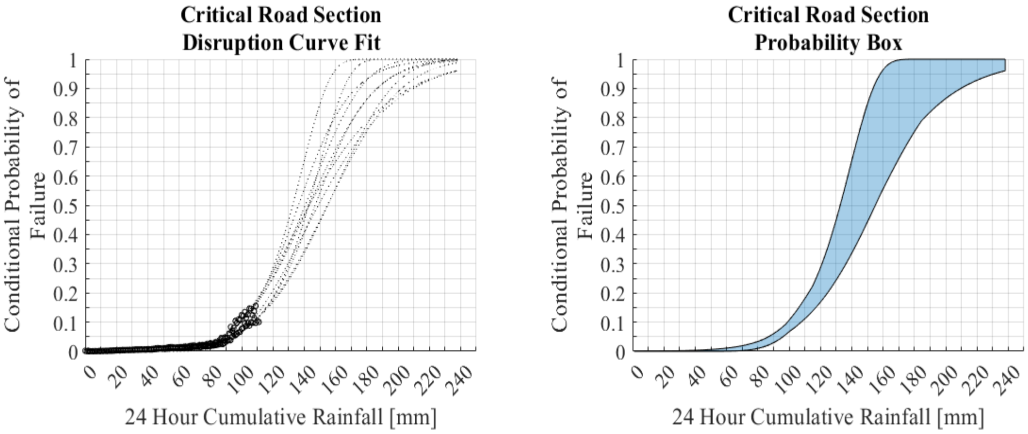


Figure 76: Cumulative distributions fitted to empirical conditional probability values for 24 hour cumulative rainfall including UK maximum point (left). Probability box formed by the maximum and minimum values of all fitted distributions.

The 24-hour envelope (Figure 76, right) is narrow at modest totals, then opens broadly across roughly 120–200 mm before tapering as both bounds approach one. Relative to the 1-hour p-box it is displaced to higher totals and remains wide for longer, reflecting sparse empirical support at very large day-long accumulations and greater scenario diversity in how prolonged saturation translates to failure. Practically, treat probabilities below ~80–90 mm as well constrained, and use the mid-range band as a planning envelope for endurance-type events, where plausible outcomes span earlier versus later loss of service.

6-5. Road Fragility to Extreme Precipitation Conclusions

This chapter has described the methodology for building empirical fragility curves for critical road sections belonging to the Scottish Trunk Road network, relating the conditional probability of road failure to the intensity of the rainfall event. Using this methodology, fragility curves have been built by considering different aggregate rainfall observations. The fragility curves are presented with a 95% confidence interval to account for the uncertainty that is present at various stages of the analysis. The fragility curves will provide a vital quantitative tool for network risk analysis and will aid in identifying parts of the network that are failing at a greater frequency than their design specifications.

With respect to the empirical fragility curves, as the accumulation window lengthens, the disruption curve shifts steadily to the right: first discernible lift moves from ~8–12 mm (1 h) to ~12–15 mm (3 h), ~20–30 mm (6 h), ~60–70 mm (12 h), and ~90–100 mm (24 h). The staircase becomes progressively more compressed toward the upper end—short-window curves show several mid-range steps (spiky, convective influence), whereas long-window curves exhibit a delayed but rapid rise once an endurance threshold is reached. Peak probabilities are similar to slightly higher at longer windows, indicating vulnerability is governed more by sustained load than isolated peaks.

With respect to the probability boxes, each envelope is tight at low totals, then the whole p-box migrates rightward with window length and the **widest part** moves to larger accumulations: ~25–60 mm (1 h), broader through ~30–90 mm (3 h), ~35–110 mm (6 h), ~70–110 mm (12 h), and ~120–200 mm (24 h). The lower bound "switch-on" is progressively delayed, while the upper bound climbs more gradually over a longer span—so mid-range imprecision increases with duration (reflecting sparse data and more possible rainfall histories), before narrowing again as both bounds approach one at extreme totals. In short: with longer windows, the p-boxes shift right and widen across the endurance range, capturing growing uncertainty about when prolonged wetting tips the section into failure.

While this work fills a critical gap in the quantification of the risk of the Scottish Trunk Road network to extreme precipitation, there are several areas for potential improvement. The analysis is carried out over a 7 year timespan, which may be too short a time period to sufficiently represent the problem or underlying uncertainties. The longer the timespan over which the analysis is considered, the more likely the conditional probability of failure estimates will converge to the true values, and this will also lead to a tightening of the confidence interval. Additionally, the range of rainfall values observed will increase and so the full structure of the fragility curves will become clearer. While distributions have been fitted to the estimated conditional probabilities, it should be noted that extrapolating too far beyond the fitted points is ill advised, and the true structure of the fragility curves in the most extreme regions is unknown.

It follows that this process should be repeated when more disruption information data has been gathered.

Furthermore, the analysis is based on the assumption that the IRIS database is representative of the true number of road failures. This is not the case since if multiple failures occur at once, the failures are prioritised by OC's in order of importance and so some road flood events that require attention may go without a response and will not be recorded. Hence the true value of the probability of failure is likely higher than the values that are presented within this chapter. An adjustment factor based on the response rate of the operating companies could be incorporated into the estimates to make the analysis more robust. Similarly, more precise timing information from the IRIS database would allow for the analysis windows to be tighter and may describe the disruption likelihood more precisely. An investigation into the nature and frequency of simultaneous events would also provide vital vulnerability information but is outside the scope of this investigation. Moreover, in order to estimate the average fragility of the whole network, the developed methodology should be extended by including also the cases of no failure for a given rainfall intensity.

A major area of improvement regards the simplistic approach to the cause of flooding. In essence it is assumed that flood events are caused by rainfall on the road section. This assumption does not take into consideration the relative position of the road to the surrounding landscape or the geometry of the road itself. It is likely that a number of flood events are caused by runoff from the surrounding landscape and so it may be beneficial to repeat this process using a block or catchment approach where each location is assigned a catchment and the total rainfall for the catchment, perhaps normalized by the area of the catchment, is considered instead of simply the rainfall at each location. This would lead to a more comprehensive understanding of the causes of rainfall related disruption.

Rainfall estimations could be improved through the application of cokriging. At the outset of this investigation, cokriging was ruled out given the increased computational time required to solve the cokriging matrices versus the available computational resource. With additional computing resource, cokriging or co-located cokriging could be applied to give more accurate rainfall estimations. The process variance would likely be smaller and hence the confidence intervals of the fragility curves would be tighter.

As a side note, the analysis considers only the risk of disruption from extreme precipitation. If the overall risk were to be evaluated this analysis would have to be extended to consider risk to life and injury from extreme events through analysis of road accident reports.

6-References

- googleapis. (n.d.). Retrieved Jan 2023, from https://developers.google.com/maps/documentation/elevation/start
- Harrison, D., Driscoll, S., & Kitchen, M. (1998). Improving precipitation estimates from weather radar using quality control and correction techniques. Forecasting Systems Meterological Office.
- Met Office. (2003). 1 km Resolution UK Composite Rainfall Data from the Met Office Nimrod System. NCAS British Atmospheric Data Centre. Retrieved March 2023, from https://catalogue.ceda.ac.uk/uuid/27dd6ffba67f667a18c62de5c3456350
- SEPA, S. E. (n.d.). Retrieved from https://www2.sepa.org.uk/rainfall
- Stone, D., Harrison, D., & Standing, R. (2008). Nimrod System Documentation Paper No.2: Nimrod format for image and model field files. British Atmospheric Data Centre (BADC). Retrieved January 2024, from https://data.ceda.ac.uk/badc/ukmo-nimrod/doc/

- Transport Scotland. (2014). Asset Management Systems and Data. Retrieved 02 16, 2023, from https://www.transport.gov.scot/publication/road-asset-management-plan-for-scottish-trunk-roads-january-2016/j408891-12/
- Transport Scotland. (2014). The Trunk Road Network, Operating Companies. Retrieved 02 16, 2023, from https://www.transport.gov.scot/transport-network/roads/the-trunk-road-network/#45651
- Transport Scotland. (2014). The Trunk Road Network, Overview. Retrieved 02 16, 2023, from https://www.transport.gov.scot/transport-network/roads/the-trunk-road-network/

Chapter 7

Conclusions and Discussion

This thesis began by framing the significant and multifaceted impacts of various types of flooding—including pluvial, fluvial, and coastal—on the environment, society, and economy. It highlighted the extensive social and economic devastation, such as injuries, fatalities, displacements, and destruction of critical infrastructure, which disrupt essential services and economic activities. This discussion underscored the pressing need for effective flood management and climate adaptation strategies to mitigate these impacts.

The thesis further delineated how climate change intensifies flooding scenarios through increased global temperatures, leading to more frequent and severe storms, increased rainfall and rising sea levels. The introduction emphasized the crucial role of stakeholders in implementing robust flood management strategies and highlighted the importance of advanced risk assessments facilitated by sophisticated hydrodynamic models and machine learning techniques. These tools are essential for accurately predicting and managing flood risks, though they also bring challenges, particularly in terms of the computational resources needed to address the inherent uncertainties in high-fidelity modelling.

The literature review built on this foundation by establishing a deep understanding of the meteorological processes, hydrodynamic and hydrological hazards, and their associated vulnerabilities, particularly to coastal and pluvial flooding exacerbated by climate change. It showcased the integration of advanced numerical models with machine learning techniques to reduce computational demands and improve predictive accuracy. The review also set the stage for the development of a fast, reliable, and computationally efficient coastal inundation forecasting model that effectively integrates both epistemic and aleatoric uncertainties, detailed Chapters 4 and 5.

Furthermore, the thesis explored specific vulnerabilities associated with different hydrodynamic and hydrological hazards, with a focus on critical infrastructure like coastal defences and road networks. It identified a significant gap in the literature regarding the fragility of roads to extreme precipitation events. This gap was addressed in Chapter 6, integrating machine learning processes and uncertainty-based analysis to offer novel insights and solutions.

7-1. Key Findings

This dissertation has demonstrated the successful development and validation of several advanced methodologies designed to enhance the prediction and management of extreme weather events, specifically focusing on storm surges, coastal inundation, and the fragility of critical infrastructure to extreme precipitation.

7-1.1. Surge Forecasting

Starting with storm surge forecasting, the research illustrates the application of a machine learning-based model enhanced by an Adaptive Bayesian Model Selection approach, which improves upon traditional models by addressing the feature, model and forecast uncertainty. This method was used to forecast 24-hour storm surge events, with validation performed in the Clyde Basin, Scotland. The integration of a Monte Carlo method into an previously established

Adaptive Bayesian Model Selection algorithm whilst also expanding the formulation to consider feature error led to the provision of meaningful prediction intervals around the surge predictions resulted in a model that not only offers more reliable forecasts but also highlights the importance of addressing uncertainty in storm surge prediction. The model achieved an error boundary of approximately 45 cm and demonstrated strong operational performance, suggesting that it could be applied in other coastal regions worldwide using historical surge height time series or surge reanalysis data such as ERA-5 and IFS.

7-1.2. Coastal Inundation Forecasting

The thesis then introduces an innovative approach to coastal inundation forecasting, with a focus on wind-wave dynamics and their integration into overtopping predictions. The framework developed in this dissertation accounts for both aleatoric (due to inherent randomness) and epistemic (due to model uncertainty) uncertainties throughout the forecasting process, ensuring that the predictions are not only accurate but also reliable in the face of uncertainty. The wind-wave forecasting model presented was successfully validated against observed data from the Firth of Clyde and demonstrated the significance of time-dependent error structures in accurately characterizing wind-wave forecast uncertainty. The model showed strong correlation with observed wave conditions (including significant wave height, mean wave period, and mean wave direction) for both 2021 and 2022, proving to be an effective and computationally efficient tool for predicting wind waves. This model's ability to account for environmental factors such as wind patterns, seabed topography, and tidal influences provides a significant advantage over more generalized, regional models. By incorporating time-varying uncertainty and tailoring predictions to specific locations, this method holds promise for early warning systems that can protect critical coastal infrastructure from potential flooding and overtopping events.

7-1.3. Road Fragility to Extreme Precipitation

The thesis presents empirical fragility curves for Scotland's trunk road network in relation to extreme precipitation events, providing valuable insights into the vulnerabilities of critical transportation infrastructure. By developing empirical fragility curves that relate road failure probability to rainfall intensity, this dissertation quantifies the risk of disruption from flooding, considering uncertainty from both the precipitation estimation and uncertainties associated from a lack of knowledge in the underlying distribution. The methodology employed highlights the importance of incorporating uncertainty at every stage of the analysis, providing a more robust and conservative estimation of failure probabilities.

7-2. Common Themes and Differences

Across all three themes—storm surge forecasting, coastal inundation prediction, and road infrastructure fragility—common threads emerge, particularly the need to integrate uncertainty at every stage of the modelling process. Each methodology emphasizes the importance of accounting for both input and model uncertainty, ensuring more reliable and actionable predictions in the face of complex and variable environmental factors. Whether forecasting the impact of storm surges, predicting coastal flooding due to overtopping, or assessing the vulnerability of road networks to extreme rainfall, the integration of machine learning, Bayesian model selection, and uncertainty propagation is key to improving the accuracy, robustness, and operational applicability of these models. These approaches offer a flexible, scalable framework that can be adapted to different locations, data sources, and environmental conditions, ultimately enhancing our ability to predict and mitigate the impacts of extreme weather on both coastal infrastructure and transportation networks.

For ungauged locations we propose an integrated regional–ML strategy. A regional hydrodynamic or reanalysis product supplies pseudo-observations and physically consistent covariates; site-level surrogates are then trained so nearby coasts share information but retain local effects. Once trained, these surrogates deliver fast, site-specific forecasts with quantified uncertainty even where gauges are absent.

7-3. Practical Implications

The research conducted on storm surge forecasting and coastal inundation using machine learning and advanced Bayesian techniques presents significant practical implications. Firstly, the development of a machine learning-based storm surge model, validated in the Clyde Basin, provides more reliable and precise predictions with meaningful prediction intervals. This model can be seamlessly integrated into existing early warning systems, thereby enhancing the capacity to provide timely and accurate alerts for storm surges. Such improvements in early warning systems are crucial for enhancing the preparedness and response strategies of coastal communities, potentially reducing the economic and human impact of coastal disasters.

Furthermore, the introduction of an innovative approach to coastal inundation forecasting, which incorporates wind-wave dynamics, adds a layer of sophistication to the prediction models. The ability of this model to account for both aleatoric and epistemic uncertainties ensures that the forecasts are not only accurate but also robust against the inherent unpredictability of natural systems. This model's validation against observed data from the Firth of Clyde and its strong correlation with real wave conditions demonstrate its operational effectiveness. By leveraging such advanced modelling techniques, coastal planners and infrastructure managers can significantly improve their strategies for managing and mitigating the risks associated with coastal overtopping.

These advancements in forecasting technology allow for the development of tailored, location-specific strategies that can protect critical coastal infrastructure. The models' capability to integrate environmental factors such as wind patterns, seabed topography, and tidal influences provides a significant advantage over more generalized, regional models. This specificity is particularly valuable for early warning systems that need to provide precise and actionable information to safeguard lives and properties in coastal regions around the world.

The trained surrogates are operationally light-weight: they run in 1-2 minutes per forecast on a standard workstation CPU, enabling hourly refresh for early warning. Where ungauged reaches require broader context, a hybrid pipeline can drive the surrogates with coarse regional physics to provide boundaries and scenarios while preserving speed.

The development of fragility curves for extreme precipitation provides a critical missing link between hazard forecasting and the consequences in the risk equation. By relating the probability of road failure directly to rainfall intensity, these fragility curves facilitate a more comprehensive numerical risk analysis of the entire road network. This enhanced methodological approach is pivotal for transportation planners and infrastructure managers as they develop strategies to mitigate risks associated with extreme weather events. The introduction of fragility curves significantly improves the accuracy of risk assessments and enhances the resilience of critical transportation infrastructure against the increasing frequency of extreme precipitation events attributed to climate change. This improved precision in risk analysis allows for more informed decision-making in infrastructure maintenance and emergency preparedness, ultimately contributing to safer and more reliable transportation networks. Moreover, the model developed in this dissertation can be adapted and extended to other countries, offering a valuable tool for global infrastructure risk management in the face of evolving climate challenges.

7-4. Limitations and Challenges

Despite the significant insights gained from this research, it is important to acknowledge the inherent limitations and challenges encountered throughout the research, which have impacted the findings and conclusions.

7-4.1. Surge Forecasting

The limitations in the surge forecasting model significantly affect its reliability and precision, including missing data at the Millport timeseries, a lack of extreme value surges, and a high dependency on the quality of forecasts. The incomplete data record at Millport introduces gaps in historical context, which could skew the model's learning phase and affect the accuracy of surge predictions. This missing data means the model might not capture the full range of possible surge scenarios, especially if data from critical events are lacking. Additionally, the rarity of extreme surge events presents a challenge. Without sufficient examples of such extremes, the model may not adequately learn to predict them or might underestimate their severity. Although physics-based models can compensate to some extent by integrating known physical processes, their effectiveness heavily relies on the accuracy of these theoretical inputs and sufficient data to validate them. Lastly, the model's performance is deeply contingent on the quality of input forecasts, including meteorological and oceanographic data. Inaccuracies in these forecasts can lead to errors in surge predictions, highlighting the need for high-quality, high-resolution forecast data. To mitigate these issues, strategies such as incorporating additional data sources, using ensemble methods to capture a broader range of surge events, and continuously updating and validating the model against new data can be employed to enhance its robustness and reliability.

Missing data can introduce bias as well as uncertainty, and in this domain the gaps are often not random. Sensors fail disproportionately during storms, so the very conditions that produce large surges and extreme waves are the ones most likely to be absent from the record. Such "missing-not-at-random" gaps can attenuate estimated relationships, narrow apparent uncertainty, and distort skill—typically biasing residual-wise metrics (which over-sample quiescent periods) differently from event-wise metrics. To mitigate this, the analysis should be explicit about gap-aware validation (leave-storm/leave-site splits), use multiple imputation with uncertainty for key inputs, and run sensitivity checks with inverse-probability reweighting so that extremes are not under-represented.

7-4.2. Coastal Inundation Forecasting

The coastal inundation model faces significant limitations that can affect its accuracy and reliability, particularly regarding data availability and dependency on forecast quality.

Firstly, a critical limitation is the lack of in situ buoy data, especially concerning measurements of extreme wave heights. This gap in data can severely impact the model's ability to predict and validate extreme coastal inundation events accurately. Extreme events, by nature, are rare and thus underrepresented in typical datasets, which challenges the model's learning algorithms to generalize and predict such occurrences accurately. Although physics-based models can be utilised to mitigate this limitation, their effectiveness hinges on their calibration.

These models must be specifically calibrated to extreme wave heights to compensate for the data deficiency. By integrating physical laws and relationships known to govern wave dynamics, a physics-based model can extrapolate beyond the available data to estimate the behaviour of wave heights during extreme conditions. However, the precision of these predictions largely depends on the accuracy and relevance of the physical parameters used in the model.

Secondly, there is a high dependency on the quality of the forecasts used as input for the coastal inundation model. The model's outputs are only as reliable as its inputs, which means that any inaccuracies or biases in the meteorological and oceanographic forecasts directly affect the model's predictions. This dependency underscores the need for high-resolution, high-quality forecast data. When the input forecasts are flawed, particularly regarding wave and weather conditions leading up to extreme events, the predictions from the inundation model will likely be compromised.

To address these limitations, it is essential to pursue a multi-faceted approach. Enhancing the integration of physics-based methods within the machine learning framework can provide a more robust predictive capability by leveraging the strengths of both methodologies. Additionally, efforts to improve the quality and resolution of forecast data are crucial. This might involve investing in better forecasting technology or developing partnerships with meteorological agencies to ensure access to the most accurate and up-to-date data available. Moreover, increasing the network of in situ measurement devices, such as buoys, would directly address the data gap, especially for capturing data during extreme conditions, thus providing a more substantial basis for model training and validation.

7-4.3. Road Fragility to Extreme Precipitation

The analysis conducted in the thesis offers valuable insights into the vulnerabilities of road networks to extreme precipitation events. However, several areas have been identified where improvements could enhance the robustness and accuracy of the findings.

Firstly, extending the observation period would be beneficial to capture a more comprehensive dataset. A longer duration of data collection would provide a broader spectrum of extreme weather events, which could improve the model’s ability to predict and understand the frequency and severity of road failures under different climatic conditions. Capturing data over multiple decades, for example, could encompass varying climate cycles and more extreme events, thereby enriching the dataset and enhancing the predictive power of the model.

Secondly, refining the rainfall estimations using advanced techniques such as cokriging could significantly improve the accuracy of the input data. Cokriging is a statistical technique that allows for the integration of multiple types of correlated spatial data to produce more precise interpolations. By applying cokriging, the methodology could leverage additional meteorological data points to refine the estimates of rainfall intensity, which is a critical factor in predicting road failures. This method would provide a more detailed and locally specific understanding of rainfall patterns, thereby improving the model's reliability in assessing the impact of precipitation on road infrastructure.

Additionally, the work acknowledges the limitations of relying solely on road failure data from the IRIS database. The data currently used may not fully represent the true extent of road failures as it likely omits unrecorded events. The reliance on this database suggests that the actual failure probabilities could be higher than reported. This discrepancy points to a need for more comprehensive data collection methods that can capture a more accurate picture of road failures. Expanding the data sources to include reports from local authorities, maintenance records, and perhaps even crowd-sourced data could provide a more complete dataset. This broader data collection approach would help in developing a more accurate and comprehensive understanding of the failure rates and the factors contributing to these failures, leading to more effective mitigation strategies.

By addressing these areas for improvement, the research can significantly advance the understanding and management of infrastructure vulnerability under extreme weather conditions, leading to more resilient and reliable transportation networks.

7.5. Future Research Directions

Looking ahead, the future work section of this research aims to build upon the foundational studies conducted while addressing the limitations and challenges encountered, to further refine methodologies and expand the scope of investigation for more comprehensive and impactful outcomes.

7-5.1. Surge Forecasting

To further advance surge research, future efforts could focus on expanding current methodologies to new and more complex geographic locations. Each area presents unique challenges due to its specific coastal configurations, bathymetry, and climatic conditions, making it essential to test and refine existing models across a variety of settings. Particularly, regions such as the Southeast coast of America and East Asia, which frequently experience violent storms like hurricanes and typhoons, would benefit from enhanced surge forecasting models. These areas suffer from severe meteorological events that can lead to devastating storm surges, and providing more accurate predictions could greatly improve preparedness and response strategies.

In addition to geographic expansion, incorporating more detailed storm characteristic inputs into surge models could significantly enhance their accuracy (Lockwood, Lin, Oppenheimer, & Lai, 2022). Inputs such as storm intensity, speed, direction, and pressure changes are crucial for a nuanced understanding of how surges develop and impact coastal areas. Moreover, integrating 3D spatial data and utilizing satellite altimetry could offer a more comprehensive view of sea surface heights and improve both the spatial resolution and temporal accuracy of the models (Srinivasan & Tsontos, 2023). Such data would provide clearer insights into the dynamic movements of water in three dimensions across affected regions.

Another critical area for improvement involves hyperparameter optimization in the modelling process. This involves fine-tuning the parameters of the machine learning algorithms to maximise model performance.

Pursuing these strategies would not only enhance the predictive capabilities of surge forecasting models but also contribute significantly to the safety and preparedness of vulnerable coastal populations worldwide. This continued innovation is vital for advancing scientific understanding and improving the resilience of communities against the impacts of extreme weather events.

7-5.2. Coastal Inundation Forecasting

To advance this coastal inundation framework, a comprehensive approach can be taken that focuses on both expanding the methodology to new areas and enhancing the technical sophistication of the models used. First, applying existing coastal inundation models to new and more complex locations would be instrumental. These locations should include areas with diverse coastal defence structures and complex bathymetries, which present unique challenges in terms of modelling and prediction accuracy. Such an expansion would not only test the robustness of current methodologies but also allow for refinement and adaptation to different environmental contexts.

Additionally, it would be beneficial to focus on regions that are prone to extensive flooding. These areas often face significant challenges in managing flood risks, and improved inundation models could lead to better preparedness and mitigation strategies. By specifically targeting these high-risk areas, research can directly contribute to enhancing community resilience and reducing the impacts of flooding.

Incorporating more sophisticated data inputs, such as 3D spatial data and satellite altimetry, into inundation models can also significantly improve their accuracy (Tapoglou, Forster, Dorrell, & Parsons, 2021). These technologies offer detailed insights into sea surface heights and coastal topographies, providing a richer dataset for model calibration and validation. Such high-resolution data can help in understanding the nuances of how water interacts with the land during flood events, which is critical for accurate prediction and planning.

Furthermore as with surge forecasting, focusing on hyperparameter optimisation is crucial for refining the performance of coastal inundation models (Domala, Lee, & Kim, 2022).

By expanding the scope of research to include these areas and focusing on technical enhancements, future coastal inundation studies can provide more precise forecasts and more effective risk management tools. This holistic approach not only pushes the boundaries of current scientific understanding but also significantly contributes to the practical capabilities of coastal communities to face and manage the challenges of inundation.

7-5.3. Road Fragility to Extreme Precipitation

Further research on fragility curves can significantly enhance the understanding and management of infrastructure vulnerability to extreme weather events. One valuable direction would be aggregating data from multiple countries, which would enrich the dataset and provide a more comprehensive understanding of how different road infrastructures react under various climatic conditions. This approach would allow for a more robust analysis and the development of universally applicable models or region-specific adjustments based on localized data patterns.

Another enhancement could involve dividing the research into different classes of roads, such as highways, arterial roads, and local streets. Each class may exhibit distinct behaviours under similar weather conditions due to varying construction standards, usage patterns, and maintenance practices. This stratification would enable more precise predictions and targeted mitigation strategies tailored to the specific vulnerabilities of each road type.

Integrating a catchment model e.g. (Vincendon, et al., 2016), could broaden the scope of the research to assess additional types of flooding, such as those resulting from river overflow or excessive runoff, which might affect road infrastructure differently than coastal surges. This integration would help in understanding the cumulative effects of multiple water sources, which is critical for comprehensive flood risk management.

Additionally, incorporating the coastal inundation model to evaluate disruptions caused by coastal floods would provide a dual perspective on how coastal events specifically impact road networks. This combined analysis could highlight areas where road infrastructure is particularly susceptible to coastal flooding, aiding in the prioritization of resilience-building efforts in these critical zones.

Extending the research to consider the implications of road failures on accidents and fatalities would also be crucial (Gissing, Opper, Tofa, Coates, & McAneney, 2019). This aspect would provide insights into the human and economic impacts of road disruptions, helping policymakers and emergency services to better focus resources and planning efforts in cases of emergencies, thereby enhancing overall community safety and response efficiency.

Finally, repeating the experiment using cokriging with radar data to maximize the prediction quality of precipitation events would likely improve the accuracy of the models (Usowicz, Lipiec, Łukowski, & Słomiński, 2021). Cokriging, by incorporating multiple sources of information, can offer a more precise estimation of rainfall patterns, especially in areas where traditional gauge data might be sparse or unreliable. This approach would enhance the model's reliability and ensure that the fragility curves reflect a more accurate risk landscape for road networks.

By pursuing these lines of research, the scientific community can offer more nuanced and actionable insights into the fragility of road networks, leading to better-informed decisions and more effective mitigation strategies against the increasing threat of extreme weather conditions.

7-References

- Domala, V., Lee, W., & Kim, T. (2022). Wave data prediction with optimized machine learning and deep learning techniques. *Journal of Computational Design and Engineering, 9*(3), 1107-1122.
- Gissing, A., Opper, S., Tofa, M., Coates, L., & McAneney, J. (2019). Influence of road characteristics on flood fatalities in Australia. *Environmental hazards, 18*(5), 434-445.
- Lockwood, J., Lin, N., Oppenheimer, M., & Lai, C. (2022). Using Neural Networks to Predict Hurricane Storm Surge and to Assess the Sensitivity of Surge to Storm Characteristics. *Journal of Geophysical Research. Atmospheres, 127*(24).
- Srinivasan, M., & Tsontos, V. (2023). Satellite altimetry for ocean and coastal applications: A review. *Remote Sensing, 15*(16).
- Tapoglou, E., Forster, R., Dorrell, R., & Parsons, D. (2021). Machine learning for satellite-based sea-state prediction in an offshore windfarm. *Ocean Engineering, 235*.
- Usowicz, B., Lipiec, J., Łukowski, M., & Słomiński, J. (2021). Improvement of spatial interpolation of precipitation distribution using cokriging incorporating rain-gauge and satellite (SMOS) soil moisture data. *Remote Sensing, 13*(5), 1039.
- Vincendon, B., Édouard, S., Dewaele, H., Ducrocq, V., Lespinas, F., Delrieu, G., & Anquetin, S. (2016). Modeling flash floods in southern France for road management purposes. *Journal of Hydrology, 541*, 190-205.

Appendix A: Artificial Neural Networks

Inspired by biological neural networks, Artificial Neural Networks (ANN's) are networks of simple interconnected processors called nodes that map the inputs or 'features' to the outputs or 'targets' as shown in Figure 77. The optimal number of layers and the number of nodes within each layers is problem specific. The amount of training time increases with the size of the network.

Appendix A-1: Multi Layer Perceptron

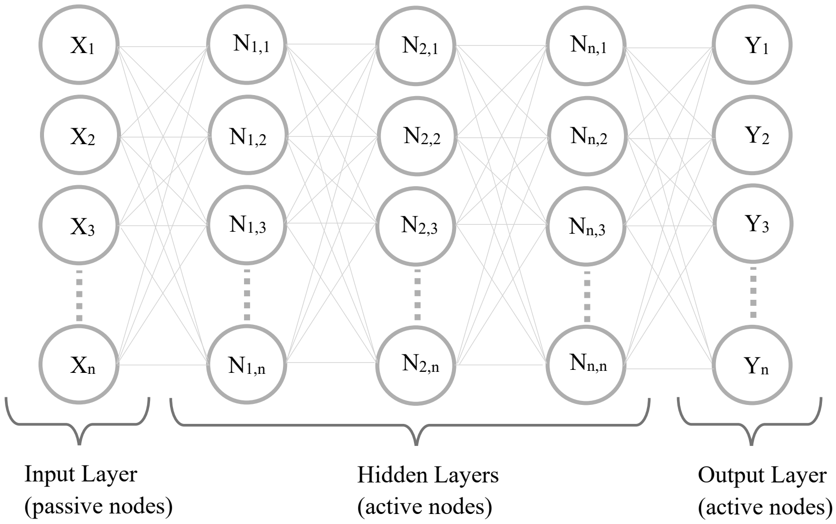


Figure 77: Basic Neural Network Architecture

In the case of artificial neural networks these nodes are referred to as neurons. An individual neuron is shown in Figure 78. These neurons comprise three parts: weights, transfer function and activation function. Input information is passed through a layer of weights and the net input is then combined and passed through an activation function. Weights are the parameters that the neural network learns during training. They represent the strength of the connection between neurons in adjacent layers. In a mathematical context, each weight modulates the input it receives, and the output is the sum of these weighted inputs. This output is then passed to the next layer. The goal of training is to adjust these weights to minimize the difference between the predicted output and the actual output (training data target). A bias is an additional parameter in a neural network that is used along with the weighted sum of inputs to each neuron. The bias allows the activation function to be shifted to the left or right, which can be critical for learning patterns more effectively. Essentially, biases help to fine-tune the output along with the weights, providing an extra degree of freedom. The selection of activation function is problem dependent and crucial since they introduce non-linear properties to the network's output. Without non-linearity, the neural network would behave just like a linear regression model, unable to capture more complex patterns in data.

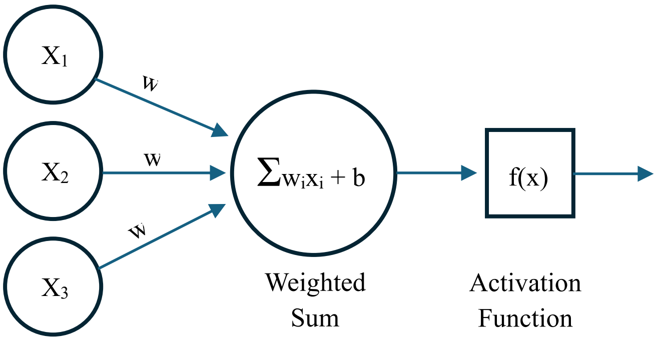


Figure 78: Deep Node Linking Inputs X_1, X_2 and X_3 to output y

Appendix A-2: Long Short Term Memory

A LSTM architecture is used in the surge forecasting framework. LSTMs are designed to remember information for prolonged periods. This capability is achieved through a sophisticated internal architecture featuring gates that regulate the flow of information (Le, et al. 2019) as shown in Figure 79. The forget, new memory, and output gates, denoted by numbers 1, 2 and 3, decide what information should be retained or discarded as the sequence progresses. As data flows through an LSTM, the new memory gate controls how much new information enters the memory cell, the forget gate determines what part of the existing memory to keep, and the output gate decides what part of the current cell state makes it to the output. The sigmoid functions used in the gates decide how much information to retain or let through by outputting values between 0 and 1. The *tanh* functions act as nonlinear transformation function to modify the information to be added to the cell state. Additionally, they process the cell state information to help determine what ultimately gets outputted from the cell, influencing both the next cell state in the sequence and the final predictions. These processes enable LSTM networks to retain long term dependencies effectively.

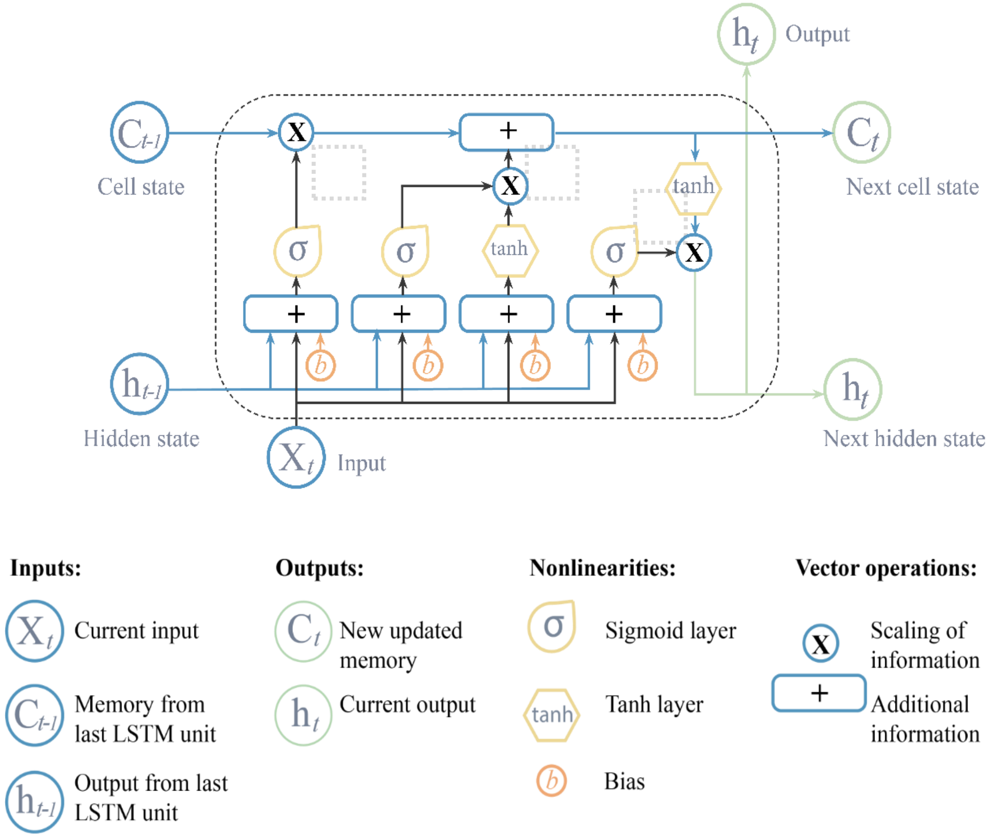


Figure 79. The structure of the Long Short-Term Memory (LSTM) cell. The figure has been redrawn and adapted from (Yan 2016). Forget gate is denoted by 1, new memory gate by 2 and output gate by 3.

Appendix B: Kriging

Before a description of the multiple kriging processes is presented, it is important to introduce the concepts that kriging utilises, mainly expected value, covariance and semivariance. The expected value of a random variable is the weighted average of the possible values that this function can take (Rubinstein, 2017). In the discrete case:

$$\boldsymbol{E}[\boldsymbol{X}] = \sum \boldsymbol{x}\,\boldsymbol{p}(\boldsymbol{x}) \qquad \text{Eq. [60]}$$

Where p(x) is the probability of x. Another useful quantity is the variance, which measures the spread or dispersion of the distribution from the expected value:

$$\boldsymbol{Var}(\boldsymbol{X}) = \boldsymbol{E}[(\boldsymbol{X} - \boldsymbol{E}[\boldsymbol{X}])\text{^}\boldsymbol{2}\,] \qquad \text{Eq. [61]}$$

For the normal distribution, the expected value is mean $\boldsymbol{\mu}$ and the variance is σ^2.

Semivariance is used to model the spatial autocorrelation of a random variable or attribute over a geographic region. The semivariance function, also referred to as the experimental variogram, is used to describe the degree of spatial dependence between pairs of observations as a function of distance or lag. In kriging, the semivariance is used to estimate the covariance structure of the underlying random process. The variogram is estimated by applying the Matheron's method, which in one dimension corresponds to the following equation (Oliver, 2015):

$$\widehat{\boldsymbol{\gamma}}(\boldsymbol{h}) = \frac{\boldsymbol{1}}{\boldsymbol{2m}(\boldsymbol{h})} \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}(\boldsymbol{h})} \{\boldsymbol{z}(\boldsymbol{x_i}) - \boldsymbol{z}(\boldsymbol{x_i} + \boldsymbol{h})\}^2 \qquad \text{Eq. [62]}$$

where $\widehat{\boldsymbol{\gamma}}(\boldsymbol{h})$ is the average semivariance between two points at distance h, $\boldsymbol{z}(\boldsymbol{x_i})$ and $\boldsymbol{z}(\boldsymbol{x_i}+\boldsymbol{h})$ are the observed values at location $\boldsymbol{x_i}$ and $\boldsymbol{x_i}+\boldsymbol{h}$. When these semivariances are calculated and plotted against there respective lag h, this constitutes the experimental variogram.

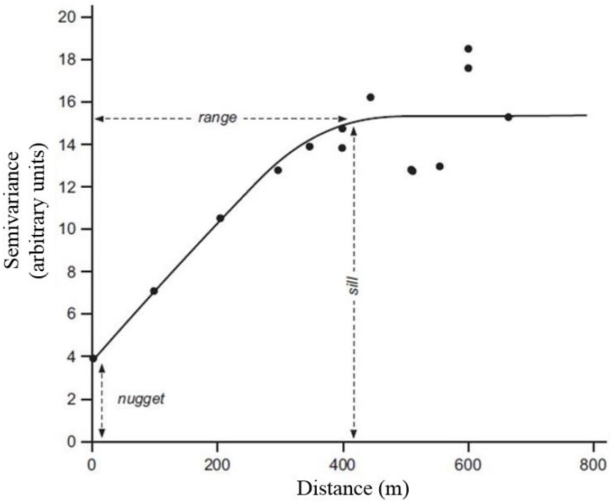


Figure 80: variogram diagram (aspetix, 2019).

The variogram in Figure 80 displays three key properties: range, sill, and nugget. The range is the maximum distance at which spatial correlation is present and beyond which it becomes negligible, while the sill is the level of variance observed at this range. The nugget, or non-zero intercept, indicates a discontinuity in variation, which can arise due to measurement errors or variations over distances shorter than the smallest sampling interval. Typically, the nugget variance is a small component of the overall variation (Oliver, 2015).

The kriging model's reliability is highly dependent the accurate fitting of the experimental variogram (Oliver, 2015). If the variogram describes the variation poorly, then the kriged predictions are also likely to be poor. Accuracy of the variogram depends on the following factors:

- Size of the sample.
- Number of lags used for estimation.
- Lag interval relative to the spatial scale of variation
- Marginal distribution of the variable (probability distribution of the separated behaviour of a single variable in a multivariant system).
- Anisotropy (property where a system exhibits different behaviour in different directions).

For a thorough analysis of the sensitivity of the semivariogram to these factors please refer to Basic Steps in Geostatistics: The Variogram and Kriging, Chapter 3.2 Factors Affecting the Reliability of Experimental Variograms (Oliver, 2015).

There are a number of different functions available for fitting the experimental semivariance points, the most common of which are spherical and exponential. Verworn and Haberlandt did an analysis on the spatial interpolation of hourly rainfall and the effect of adding additional information. They stressed the time consuming nature of applying individual variograms to each hourly timestep and suggested generating event specific variograms (Verworn, 2011). An alternative is to autofit a number of variograms and select the best performing using a suitable performance metric. The remainder of the subsection will describe 6 types of kriging: simple, ordinary, universal, kriging with and external drift and trend kriging

Appendix B-1. Simple Kriging

In simple kriging the mean of the variable is assumed to be known and constant across the entire study area. The kriging estimator then incorporates this information into the interpolation process. Simple kriging can provide more accurate estimates when the mean of the variable is well-known or easily estimated. In simple kriging, the variance of the estimator depends on the spatial autocorrelation and the estimation error.

Kriging estimates the unknown values at unobserved locations by minimizing the prediction error variance, subject to the constraint that the predictions are unbiased and consistent with the observed values (Oliver, 2015). Estimate $\boldsymbol{\hat{Z}(x_0)}$ for unknown location x_0 is calculated by:

$$\boldsymbol{\hat{Z}(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i) + \left\{1 - \sum_{i=1}^{N} \lambda_i\right\} \mu} \qquad \text{Eq. [63]}$$

where $\boldsymbol{\mu}$ is the constant mean value over the entire region of interest, $\boldsymbol{\lambda_i}$ are the weights. In simple kriging, the weights assigned to neighbouring data points are determined based on the spatial correlation or covariance structure of the data and are calculated to minimize the estimation error while honouring the assumed mean (Oliver, 2015). The simple kriging variance is given by:

$$\sigma_{SK}^2(x_0) = C(0) - \sum_{i=1}^{N} \lambda_i C(x_i, x_0) \quad \text{Eq. [64]}$$

where C(0) is the variance of the random process and $C(x_i, x_0)$ is the covariance between known location x_i and target location x_0 (Oliver, 2015).

Appendix B-2. Ordinary Kriging

In real world problems, the true mean and covariance function of the underlying random function Z(x) are unknown, making simple kriging unapplicable for most practical problems. Ordinary kriging does not assume a known constant mean, but assumes a quasi-stationary condition (varying mean but constant covariance) (Negreiros, 2010). Ordinary kriging estimates are robust even with moderate departures from stationarity conditions. The mean is assumed stationary in the local search neighbourhood i.e. a local mean at each point is calculated based on the neighbouring observations. This allows for greater flexibility in modelling the spatial autocorrelation of the variable and can be useful in cases where the mean varies across the study area. In ordinary kriging, the variance of the estimator also includes the estimation error associated with the local mean estimate. Estimate $\hat{Z}(x_0)$ for unknown location x_0 is calculated by:

$$\hat{Z}_{OK}(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i) \quad \text{Eq. [65]}$$

As with simple kriging, the predictions are unbiased. To ensure this the weights of ordinary kriging are constrained to sum to 1 (Oliver, 2015):

$$\sum_{i=1}^{N} \lambda_i = 1 \quad \text{Eq. [66]}$$

The variance of the predictions is given by:

$$\sigma_{OK}^2(x_0) = 2\sum_{i=1}^{N} \lambda_i \gamma(x_i - x_0) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \gamma(x_i - x_0) \quad \text{Eq. [67]}$$

Where the quantity $\gamma(x_i - x_0)$ is the semivariance of Z between known location x_i and target location x_0 whereas $\gamma(x_i - x_0)$ is the semivariance between the *i*-th and *j*-th sampling locations (Oliver, 2015).

Appendix B-3. Universal Kriging

Simple and ordinary kriging assume a stationarity or quasi-stationarity of the real-valued random function Z(x). But in reality the mean value of some spatial data cannot be assumed constant in general, since it will most likely depend on the absolute location of the sample. For example, the intensity of rainfall is higher on average in the west coast of Scotland compared to the east, and spatial variations are observed in the relationship between rainfall and altitude across the UK (Salles, 2001). Universal kriging is introduced as a method that splits the random function into a linear combination of non-stationary deterministic functions with a random residual function. The estimation of $\widehat{\boldsymbol{Z}}(\boldsymbol{x_0})$ is the same formulation as before:

$$\widehat{\boldsymbol{Z}}(\boldsymbol{x_0}) = \sum_{i=1}^{N} \boldsymbol{\lambda_i Z(x_i)} \qquad \text{Eq. [68]}$$

where the underlying random function can be expressed as the sum of non-random trend function $\boldsymbol{\psi(x)}$ and residual random function $\boldsymbol{Y(x)}$:

$$\boldsymbol{Z(x)} = \boldsymbol{\psi(x)} + \boldsymbol{Y(x)} \qquad \text{Eq. [69]}$$

Trend function $\boldsymbol{\psi(x)}$ can be evaluated using a regression model. Here the formula is presented for a linear regression model of latitude L_1 and longitude L_2:

$$\boldsymbol{\psi(x)} = \boldsymbol{a_0} + \boldsymbol{a_1 L_1(x)} + \boldsymbol{a_2 L_2(x)} \qquad \text{Eq. [70]}$$

Where a_0, a_1 *and* a_2 are the regression coefficients. Hence the full estimation term is expressed by substituting **Error! Reference source not found.** and **Error! Reference source not found.** into **Error! Reference source not found.**:

$$\psi(x) = a_0 + a_1 L_1(x) + a_2 L_2(x) \qquad \text{Eq. [71]}$$

Universal kriging is particularly useful when the trend component of the data is significant and needs to be accurately estimated to make reliable predictions or to perform spatial interpolation as it allows the trend component to be modelled explicitly as a function of the spatial coordinates and estimates both the trend and the residual components simultaneously. Wagner et al. found that for daily rainfall estimates, interpolation methods that use covariates outperform univariate interpolation methods (Wagner, 2012).

Appendix B-4. Kriging with External Drift

If additional variables exist that are linearly related to the target variable, it is possible to incorporate them into the kriging system to improve predictions. In this case the assumption of a constant expected value is replaced with the linear relationship between the target and correlated variable. The formulation of kriging with an external drift is similar to that of universal kriging; however, in the external drift model, the deterministic component (Equation 15) is assumed to be linearly related to a set of auxiliary variables. In the case of kriging with external drift. The estimation of $\widehat{\boldsymbol{Z}}(\boldsymbol{x_0})$ is:

$$\widehat{\boldsymbol{Z}}(\boldsymbol{x_0}) = \sum_{i=1}^{N} \boldsymbol{\lambda_i Z(x_i)} \qquad \text{Eq. [72]}$$

For:

$$\textstyle\sum_{i=1}^{N} \boldsymbol{\lambda_i Z(x_i) \cdot q_k(x_i)} = \boldsymbol{q_k(x_0)} \qquad \text{for k=1, …, p} \qquad \text{Eq. [73]}$$

Where $\boldsymbol{q_k}$ is the *k-th* predictor variable, *p* is the number of predictors and the other symbols have their usual meanings. The variance of the predictions is given by:

$$\boldsymbol{\sigma^2_{KED}(x_0) = C_0 + C_1 - c_0^T \cdot \lambda_0} \quad \text{Eq. [74]}$$

Where $\mathbf{C_0}$ and $\mathbf{C_1}$ are estimated parameters of the semi-variance function, and $\mathbf{c_0^T}$ is the extended vector of variances at the new location:

$$\begin{aligned}&\boldsymbol{c_0^T} = \{\boldsymbol{C(x_0, x_1)}, \dots \boldsymbol{C(x_0, x_n)}, \boldsymbol{q_1(x_0)}, \dots \boldsymbol{q_p(x_0)}\}^{\boldsymbol{T}};\\ &\boldsymbol{q_0(x_0) = 1}\end{aligned} \quad \text{Eq. [75]}$$

Appendix B-5. Regression Kriging

Regression kriging used is alternative to universal kriging but instead of the trend component being modelled explicitly and the trend and residual components simultaneously estimated, the trend component is evaluated using a regression model, such as linear or non-linear regression. Once the regression model is fitted, the residuals of the model, which represent the spatially correlated variation that cannot be explained by the covariates, are interpolated using ordinary kriging. The trend component and residual estimation are then combined to obtain the final predictions.

Regression kriging has several advantages over traditional kriging methods, including the ability to incorporate covariate information, which can improve the accuracy of the predictions, and the ability to quantify the relative importance of the covariates in predicting the variable of interest.

Appendix C: Surface fit to lag structure in Figure 46

Lag structure best fit
z: lag
x: longitude [º]
y: latitude [º]
$z = -9.23 + 6.5x + 2.7y - 0.03x^2 - 0.464xy - 3.32y^2$

Appendix References

- aspetix, 2019. *Introducing the variogram.* s.l.:s.n.
- Le, X., Ho, H., Lee, G. & Jung, S., 2019. Application of long short-term memory (LSTM) neural network for flood forecasting. s.l.:s.n.
- Negreiros, J. P. M. A. F. a. A. M., 2010. Geographical information systems principles of ordinary kriging interpolator. s.l.:s.n.
- Oliver, M. A. a. R. W., 2015. *Basic Steps in Geostatistics : The Variogram and Kriging.* 1 ed. s.l.:SpringerBriefs in Agriculture. Web..
- Rubinstein, R. a. K. D., 2017. *Simulation and the Monte Carlo method [internet resource].* Third edition ed. s.l.:Wiley series in probability and statistics.
- Salles, M. C. P. a. C. R., 2001. Spatial Variations in the Average Rainfall-Altitude Relationship in Great Britain: An Approach using Geographically Weighted Regression. s.l.:s.n.
- Verworn, A. a. H. U., 2011. Spatial interpolation of hourly rainfall–effect of additional information, variogram inference and storm properties. s.l.:s.n.
- Wagner, P. F. P. W. F. K. S. a. S. K., 2012. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions.. 464-465 ed. s.l.:s.n.
- Yan, S., 2016. Understanding LSTM and Its Diagrams., s.l.: s.n.