Hamilton Institute

Maynooth University
National University of Ireland Maynooth

Statistical and Machine Learning Models for Multivariate Sensor Data with Application to Environmental Monitoring

A dissertation submitted for the degree of
Doctor of Philosophy

By:

Amin Shoari Nejad

Under the supervision of:

Prof. Andrew C. Parnell
Dr. Gerard D. McCarthy

Hamilton Institute
Maynooth University
Maynooth, Co. Kildare, Ireland

November 2023

Abstract

The increased availability and use of sensor data in environmental monitoring has led to a vastly increased demand for tools that can process, analyze and report on important environmental events in real or near-real time. These data have special characteristics, being high dimensional, recorded in space and time, and with potential missing values / sparsity issues. This thesis concerns itself with the building of statistical and machine learning methods for data of this type. Statistical methods can play an important role in inferring valuable features of the latent data-generating processes of the sensor data, while properly taking into account the uncertainty. Machine learning methods can play a significant role by offering an automatic framework to unearth complex patterns underlying the data, being scalable and dealing with challenges such as sparsity in an innovative way.

This dissertation contributes to our understanding of how to monitor the environment, focusing on Dublin Bay as a proof of concept. It examines sea-level rise, water turbidity, and how to manage large environmental datasets that change over time and space with missing values. Firstly, it updates us on Dublin's sea level record to produce a processed data product from 1938 to 2016. It utilizes a new statistical approach to make better estimates of average sea levels and finds that sea levels have been rising more quickly in recent times. Then it investigates how human activities, such as dredging and dumping, affect water turbidity in Dublin Bay. It employs a new statistical model called VARICH to model the variability in water turbidity over two years and at different locations, finding that weather conditions like wind speed significantly impact most locations. Meanwhile, dredging operations show lower impacts, and dumping operations have a significant impact only at greater water depths. Lastly, the dissertation addresses the challenge of missing values in large-scale environmental datasets by proposing two innovative models for multivariate spatio-temporal forecasting that offer competitive performance without the need for imputation. These models are a transformer-based model, SERT, and a simpler, interpretable model, SST-ANN.

Altogether, the chapters of this dissertation provide new insights into environmental trends and offer novel methods for analysing environmental data. This is all taking place against the backdrop of significant global issues such as climate change, loss of

biodiversity, and increasing pollution. This thesis stands as an example of how data analysis is advancing in the field of environmental monitoring.

Declaration

I affirm that I have created this manuscript independently, without the unauthorized help of third parties and have only used the resources that were explicitly allowed.

The thesis work was conducted from August 2019 to August 2023 under the supervision of Professor Andrew C. Parnell and Dr. Gerard D. McCarthy in the Hamilton Institute, National University of Ireland Maynooth.

Amin Shoari Nejad,
Maynooth, Ireland,
November 2023

Acknowledgements

Above all, I am deeply grateful to my parents, family, and friends for their constant support and patience over the course of my educational journey. Their exceptional understanding and motivation have been crucial in enabling me to complete my studies successfully.

I am also thankful to Andrew, Gerard and Rocio for their mentorship during my doctoral studies. Their precious advice, steadfast support, and patience have been instrumental. Their profound knowledge and experience have consistently inspired me throughout my scholarly work and everyday life.

Furthermore, I am thankful to the wider community of staff and peers at the Hamilton Institute and the Department of Mathematics & Statistics for creating a welcoming, productive, and enjoyable academic atmosphere.

Collaborations

In preparing this thesis, I collaborated with several authors and would like to clarify the role of each below. However, I want to emphasize that all other work presented in this thesis is my own.

Andrew Parnell: Prof. Parnell served as my supervisor, providing guidance and oversight for all aspects of this thesis.

Gerard McCarthy: Dr. McCarthy served as my co-supervisor. Similar to Professor Parnell, he provided supervision and direction for the entirety of this work.

Rocio Rodríguez: Dr. Rodríguez played an advisory role during the final year of my PhD. She provided valuable suggestions for the direction of my last PhD paper presented in Chapter 4. She also contributed to the editing of the manuscript.

Anthony Grey and Brian Kelleher: Dr. Grey and Dr. Kelleher were my primary collaborators in the Predict project, introduced in Chapter 1. They provided most of the datasets analysed in this thesis and offered consultancy on technical issues related to the chemical behavior of the environmental phenomenon studied. Furthermore, they contributed to the editing of the papers.

Robert Devoy: Prof. Devoy provided valuable input on the sea-level perspective for my first PhD paper, presented in Chapter 2.

Alice Greene: Ms. Greene contributed a dataset that was crucial for the research presented in my first PhD paper.

Peter Thorne: Prof. Thorne offered valuable insights from a climate change perspective for my first PhD paper.

Publications

Of the three chapters in this thesis that cover novel contributions, one has been published, one has been accepted for publication, and the other has been submitted, peer-reviewed, revised, and is currently awaiting the journal's decision.

Peer-reviewed journal article:

- Shoari Nejad, Amin, Parnell, Andrew C., Greene, Alice, Thorne, Peter, Kelleher, Brian P., Devoy, Robert JN., McCarthy, Gerard, A newly reconciled dataset for identifying sea level rise and variability in Dublin Bay. *Ocean Science*, 18(2), 511–522, (2022). Publisher: Copernicus GmbH.

Peer-reviewed and accepted (waiting for publication):

- Shoari Nejad, Amin, McCarthy, Gerard D., Kelleher, Brian, Grey, Anthony, Parnell, Andrew, Vector Time Series Modelling of Turbidity in Dublin Bay. *Journal of Applied Statistics*, (2023).

Peer-reviewed and revised (waiting for decision):

- Shoari Nejad, Amin, Alaiz-Rodríguez, Rocío, McCarthy, Gerard D., Kelleher, Brian, Grey, Anthony, Parnell, Andrew, SERT: A Transfomer Based Model for Spatio-Temporal Sensor Data with Missing Values for Environmental Monitoring. *Computers and Geosciences*, (2023).

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

My thesis focuses on developing analytical approaches for analysing multidimensional data relevant to the study of coastal transformation, with a particular emphasis on Dublin Bay as a proof of concept. Accordingly, I will discuss the challenges I encountered in analysing environmental data pertaining to coastal transformation and the strategies I employed to address them.

Coastal regions worldwide represent a dynamic interface where the land meets the sea, providing diverse ecosystems and vital resources, as well as offering habitation and recreational space for a significant portion of the global population. Nonetheless, these vital zones are presently undergoing unprecedented transformation due to various human and natural influences [12]. Factors such as sea level rise, alterations in water turbidity, and shifts in other water quality variables have a consequential impact on the structure and function of coastal environments [94].

The study of sea level rise is indispensable in understanding coastal transformation, as this natural phenomenon exacerbates coastal erosion, inundation, and saltwater intrusion [96]. These effects often lead to the displacement of human populations and detrimental impacts on biodiversity, thereby having significant socio-economic and ecological implications [95]. Current evidence suggests that human-induced climate change is accelerating the rate of sea-level rise, making this area of study ever more crucial [119].

Water turbidity, often influenced by land-based activities such as deforestation and urbanisation, is another critical factor in coastal transformation. Higher turbidity can lead to significant alterations in light penetration, impeding photosynthetic processes and altering species composition in aquatic ecosystems [34]. Further, turbidity can also influence the nature of sediment deposition, thereby affecting the morphodynamics of coastal systems [131].

Moreover, changes in other water quality variables, such as temperature, pH, dissolved

oxygen, and nutrient concentrations, can impact the biological and physical processes in coastal areas. Such alterations can lead to coral bleaching, harmful algal blooms, and shifts in species distribution, which can have extensive ecological and economic implications [44].

Thus the study of coastal transformation is critical for establishing effective adaptive and mitigation strategies, providing the capacity to predict, respond to, and manage these changes. A comprehensive understanding of the complex interaction of human and natural influences on coastal transformation offers a foundation for sustainable coastal management practices and policies [126]. By studying the factors driving coastal transformation, including sea level rise, changes in water turbidity, and shifts in other water quality variables, we can better understand and predict future changes, mitigate adverse impacts, and exploit potential opportunities. As coastal areas continue to be hotspots for population and economic growth, ensuring their resilience becomes increasingly urgent [94]. With the continued advancement in scientific research and understanding of these areas, we are better equipped to face the challenges and opportunities that the future may bring to our coastal regions.

1.2 Introduction to the Predict Project

My PhD was completed as part of the "Predict" project (`https://www.dcu.ie/predict`) which integrates multidisciplinary geoscientific data to monitor and predict coastal change, with a proof of concept in Dublin Bay. The primary objective of the project is to coordinate a comprehensive program of coastal ocean observations. The goal is to integrate these datasets to generate models that can be used to predict environmental change. These models will contribute to future planning in a diversity of areas such as coastal mapping, flooding prediction, marine habitats and fisheries, climate change, environmental protection and policy. The project aims to provide an experimental proof-of-concept in Dublin Bay that can be extrapolated to a range of environments.

My main role in the project was to analyse in-situ data and build statistical and machine learning models to predict the coastal change and understand its potential drivers. In this chapter, I will introduce the data I worked with and the statistical and machine learning concepts that I used to analyse the data. The details of my data analyses and my developed methods are explained in Chapters 2-4. I undertook a three-stage approach to analysing the data collected and generated through the Predict project. In the first stage I analysed sea level data from tide gauges situated around the Dublin coastline. In the second stage I analysed data from buoys recording turbidity measurements in the bay. In the third stage, I analysed various environmental variables' data using advanced machine learning techniques. Below, I will provide an overview of the topics covered in the thesis

and their importance.

1.2.1 Sea Level Analysis

Sea level measurement is a critical aspect of oceanographic studies, climate change research, coastal engineering, and navigational safety. Changes in sea level can lead to coastal erosion, increased flooding, and the loss of habitat for plants, fish, and birds. Monitoring and understanding sea level changes can help in the development of mitigation strategies, adaptation policies, and the effective management of coastal resources. Traditionally, one of the most widely used tools for this purpose is the tide gauge. A tide gauge, shown in Figure 1.1, constituting an integral element of contemporary water level monitoring stations, is equipped with sensors designed to perpetually measure the elevation of the surrounding water level.

Figure 1.1: A Coastal Tidal Gauge Station Equipped with Solar-Powered Telemetry - Instrumented to collect high-precision tidal data, this station provides vital information for monitoring sea-level changes and coastal management.

Tide gauge data serves as a significant measure of the development and effects of global climate change, encompassing an array of local and regional events associated with decadal climate fluctuations, tides, storm surges, tsunamis, swells, and other coastal activities. These data are employed to confirm ocean models and to identify inaccuracies and deviations in satellite altimetry. The historical records of tide gauge data enable scientists to calculate the global rise in sea levels since the industrial revolution, acting

as a robust sign of both the course and the consequences of human-induced climate change [3]. The importance of this information is manifest in various coastal operations such as ensuring safe navigation, endorsing solid engineering practices, and contributing to habitat maintenance and conservation [4]. However, it is important to note that tide gauges are prone to malfunctioning, which could result in erroneous measurements. Hence, it is important to conduct data quality checks before using the collected data for further analyses.

In Chapter 2, I explain my work on collating sea level tide gauge data, conducting data quality checks using data visualisations and statistical models, correcting for measurement errors in the data, and finally analysing the patterns of sea level data fluctuations and estimating the rate of sea level rise in Dublin Bay.

1.2.2 Turbidity Analysis

Turbidity refers to the cloudiness or haziness of a fluid caused by large numbers of individual particles that are generally invisible to the naked eye. This physical property, typical in liquids such as water, is an important parameter for environmental science.

The phenomenon of turbidity is primarily attributed to suspended solids, colloidal particles, and dissolved colored material that scatter and absorb light rather than transmitting it. These materials can include silt, clay, microorganisms, organic and inorganic matter, and other microscopic substances [34].

Measurement of turbidity is conducted through several standardized methods that commonly involve the use of a nephelometer or turbidimeter. A nephelometer measures the intensity of scattered light at a specific angle to the incident light beam, while a turbidimeter measures the attenuation of a light beam passed through the liquid. The results are typically expressed in nephelometric turbidity units (NTU) or formazin turbidity units (FTU) [80].

Turbidity is critical to study for various reasons. In environmental contexts, it provides insight into the concentration of particulate matter in natural waters, which can have implications for aquatic life and ecosystem health. Elevated turbidity levels can hinder the growth of aquatic plants by reducing sunlight penetration and can affect the gill function in fish and other aquatic organisms. In water treatment and supply, turbidity is used as an indicator of the effectiveness of filtration processes, with higher turbidity often indicating the presence of pathogens or contaminants [120].

Turbidity is a complex field with important real-world applications. It's crucial to have precise and consistent ways to measure it in order to protect the environment and public health. Keeping turbidity in check is key to preserving the equilibrium of water-based habitats and ensuring the safety of our water resources. Thus, persistent study and

observation of turbidity are imperative for managing these varied and critical issues.

Chapter 3 delves into my work on analysing turbidity level data collected by multiple buoys in Dublin Bay. It explains the statistical models I developed to estimate the effects of dredging and dumping operations, wind speed, and the interaction between turbidity levels at different locations in the bay.

1.2.3 Environmental Data Analysis

Environmental variables such as salinity, water temperature, dissolved oxygen, precipitation, and wind speed are fundamental components that shape the dynamics and health of both aquatic and terrestrial ecosystems. Collectively, these factors play a critical role in regulating biological processes, chemical reactions, and physical properties of the environment.

- **Salinity**: The concentration of dissolved salts in water, often measured in parts per thousand (ppt), is a critical water quality indicator [88]. It affects the distribution of aquatic organisms and plays a key role in the functioning of coastal and marine ecosystems [116].
- **Water Temperature**: A vital physical property, water temperature influences metabolic rates of organisms, solubility of gases (including oxygen), and water density. Changes in water temperature can have far-reaching impacts on biological activities and ecosystem structure [16].
- **Dissolved Oxygen**: The concentration of oxygen dissolved in water is essential for the respiration of most aquatic organisms. Variations in dissolved oxygen levels can cause stress or mortality in fish and other aerobic organisms and can influence nutrient cycles [101].
- **Precipitation**: As a primary source of freshwater, precipitation affects soil moisture, groundwater recharge, and river flow. It impacts the distribution of biomes and influences patterns of plant growth and animal behavior [138].
- **Wind Speed**: Wind plays a multifaceted role in environmental processes. It influences weather patterns, water currents, and wave actions. Wind speed affects the rate of evaporation, heat transfer, and the dispersal of pollutants and seeds [53, 6].

These environmental variables interact in complex ways to influence ecosystem processes and functions. For example, water temperature and salinity can affect dissolved oxygen levels, which in turn influence the biological activity within a water body. Similarly, wind speed can impact both evaporation rates and precipitation patterns, thereby affecting water availability and quality.

Understanding these environmental variables and their interactions is paramount in the context of climate change, natural resource management, conservation, and public health. They provide vital information for modeling and predicting ecological responses to changes in environmental conditions, assisting in the planning and implementation of mitigation and adaptation strategies. The cumulative study of these variables facilitates informed decision-making that contributes to sustainable development and environmental stewardship. Whether in the context of agriculture, urban planning, fisheries management, or water resource management, a comprehensive understanding of these environmental factors is instrumental in shaping a resilient and sustainable future.

In the Predict Project, we collected data on the aforementioned environmental variables through environmental monitoring buoys placed in Dublin Bay. Figure 1.2 shows an image of one of the buoys being prepared for deployment in the water.

Figure 1.2: Dublin Bay data buoy.

In Chapter 4, I will introduce the details of environmental buoy dataset, which collects data on many variables, including those previously mentioned. I will also detail the methodology I used to analyse the data. Specifically, this chapter will delve into the machine learning approach I developed for efficient predictive modeling, even when faced with significant gaps in the sensor data.

1.3 Relevant Statistical and Machine Learning Methods

Throughout this thesis, I employ an array of statistical and machine learning methods, with the overarching goal of augmenting and enhancing the existing methodologies. Below I provide a succinct exposition of each method, together with its fundamental characteristics. In the later chapters I provide more complete descriptions including algebraic descriptions of the fitted models.

1.3.1 Statistical Methods

The use of statistical methods in the study of coastal transformation and environmental monitoring has been a cornerstone of research in these fields [89]. Regression analysis, time series analysis, spatial statistics, and multivariate analysis are among the most widely used statistical techniques in this area (for example see [139, 55]).

Regression and time series models allow researchers to understand and forecast the dynamic behavior of coastal systems [8]. Such models identify relationships between dependent and independent variables, and help to understand trends and cycles in data collected over time.

Spatial statistic models enable researchers to study and interpret spatial patterns and variations in coastal transformation phenomena, including sea-level rise and water turbidity changes. Multivariate analysis helps to understand the complex interrelationships among different environmental variables affecting coastal systems [123].

Linear Regression Analysis

In the initial stages of my research, I employed multiple linear regression modelling [7] to analyse in-situ measurements of sea level (Chapter 2). The multiple linear regression model facilitates the modeling of the relationship between a response variable and multiple predictors. The rationale behind adopting this model is to rectify the bias in mean sea level measurements by utilising low water and lunar activities as predictors within a specific timeframe, while leveraging the relationship established during another period of time, in which the mean sea level measurements were deemed unbiased. Moreover, after correcting for bias in the sea level measurements, I employed a multiple linear regression to estimate the rate of sea level rise in Dublin Bay.

Change-point Regression Analysis

The change-point regression model [25] characterises a time series by dividing it into segments with linear behavior and determines the points at which shifts in data trends transpire. Within the change-point model, estimates are derived for the change point(s),

representing the timing(s) of these shifts. The motivation behind employing this model is to identify the timeframe in which a decline in the congruity between the sea level measurements at Dublin port and those from neighboring locations occurs, potentially attributable to sensor malfunctioning that introduces bias into the measurement process.

Autoregression Analysis

Autoregression (AR) is a statistical model extensively used in time series analysis where a variable is predicted using its own past values [107]. An important extension to the AR model is the Autoregressive Conditional Heteroskedasticity (ARCH) framework, introduced by [49], to model and forecast volatility in financial returns. The ARCH model assumes that the variance of the current error term or the current volatility is a function of the actual sizes of the error terms in previous periods. It effectively addresses the common stylised fact of heteroskedasticity in many real-world time series data.

Vector Autoregression (VAR) is an extension of the AR model utilised predominantly in econometrics for forecasting multivariate time series data [86]. The underlying premise of VAR is the principle that each variable in the system is a noisy linear function of the past lags of itself and the past lags of the other variables in the system [115]. This multifaceted approach enables the VAR model to encapsulate the dynamic interrelationships among the variables, thus providing a comprehensive tool for multivariate time series analysis.

In my research, I integrated the structure of the ARCH model with that of the VAR model to effectively model spatiotemporal time series of turbidity. The primary objective was to estimate the impacts of dredging and dumping operations, along with the influence of wind speed, on turbidity levels.

Bayesian Framework

Where statistical models are employed in this thesis, they are fitted using a Bayesian inference framework. The Bayesian approach stands as a cornerstone in statistical modeling, primarily due to its holistic methodology of incorporating prior knowledge and updating it with observed data. Mathematically, Bayes' theorem can be expressed as:

$$P(\theta|Y) = \frac{P(\theta) \times P(Y|\theta)}{P(Y)}$$

where θ are parameters and Y are data, either of which could be discrete or continuous. Since this thesis primarily focuses on continuous parameters and data, the following definitions pertain to that scenario:

- $P(\theta|Y)$ is the posterior density of the parameters given the data (our updated belief after observing data).

- $P(Y|\theta)$ is the likelihood function (proportional to the probability density function of observing data Y conditioned on parameters θ).
- $P(\theta)$ is the prior probability density function (reflects our initial belief about θ before observing data).
- $P(Y)$ is the marginal likelihood, defined as $P(Y) = \int P(\theta)P(Y \mid \theta)d\theta$ (probability of observing data Y marginalised over the prior distribution of θ).

The final term, $P(Y)$, can usually be ignored as a proportionality constant because it does not depend on θ through marginalisation. As a result, we can write:

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta).$$

In the Predict project's context, the Bayesian approach was vital. The prior knowledge, potentially based on historical data or expert opinions, was updated using fresh in-situ data. For example, in Chapter 3, I explain how we used the prior knowledge about the range of turbidity values to infer the missing observations in the turbidity time series.

Inference using MCMC In the Bayesian framework, a major challenge is computing the posterior distribution, particularly in complex models. This difficulty arises because the denominator $P(Y)$ in Bayes' theorem involves an integral over the entire parameter space, which is often analytically intractable due to the model's complexity. To address this, Markov Chain Monte Carlo (MCMC) methods, such as the Metropolis-Hastings algorithm, are commonly used. These methods are advantageous as they do not require the calculation of the normalizing constant in the posterior distribution. Instead, they operate by ensuring the posterior can be evaluated up to a constant of proportionality.

MCMC works by constructing a Markov chain with a stationary distribution that matches the target posterior distribution. After a warm-up period, and conditional on some assumptions [21, Chapter 1], we can obtain a series of samples that are from the true posterior distribution. By using these samples, we can infer properties about the distribution, such as its mean, variance, or other moments [67].

A common MCMC method is the Metropolis-Hastings algorithm, which is popular for its ability to approximate the posterior distribution by generating a sequence of sample values from a probability density function for which direct sampling is difficult. However, Metropolis-Hastings has some limitations such as the struggle with sampling from correlated parameters [108]. More advanced MCMC techniques like Gibbs sampling and Hamiltonian Monte Carlo (HMC) address some of the limitations of the basic Metropolis-Hastings.

Gibbs Sampling is particularly useful in scenarios where sampling from the full joint distribution is challenging, but sampling from conditional distributions is feasible. It works by sequentially sampling each variable in the distribution while holding all other variables fixed [58]. This method is highly efficient in high-dimensional spaces where the conditional distributions are easier to handle. HMC, on the other hand, introduces concepts from physics to improve the sampling efficiency. It utilizes the Hamiltonian dynamics to propose new states in the Markov chain, which allows it to move through the parameter space more efficiently than traditional random-walk strategies. By incorporating information about the gradient of the log-posterior, HMC can take larger steps in the parameter space and explore it more thoroughly, making it particularly effective for complex models with high-dimensional parameters [13]. This leads to faster convergence and better exploration of the posterior distribution, especially in cases where other MCMC methods struggle.

In my work, I relied on two packages in the R programming language, namely: JAGS [102], which uses Gibbs sampling, and Stan [26], which uses HMC, to conduct MCMC and infer the parameters of my models. This enabled me to fit a wider range of statistical models to the data, with relatively less time spent tuning algorithms to ensure computational feasibility.

Convergence Assessment in MCMC Methods Assessing convergence in MCMC is crucial to ensure the reliability of the posterior distribution estimates. Convergence diagnostics involve checking whether the Markov chain has reached its stationary distribution. Common strategies include analyzing multiple chains with different starting points and monitoring within- and between-chain variances, seeking well-mixed chains as evidence of convergence. The Gelman-Rubin diagnostic, also known as the $\hat{R}$ value, is a popular method for this purpose. It compares the variance between multiple chains to the variance within each chain for an estimand θ. Specifically, $\hat{R}$ is calculated as $\sqrt{\frac{\hat{V}}{W}}$, where $\hat{V}$ is the marginal posterior variance of the estimand, given by $\hat{V} = \frac{N-1}{N}W + \frac{B}{N}$, with W being the within-chain variance and B representing the between-chain variance. Here, B is calculated as $\frac{N}{M-1}\sum_{m=1}^{M}(\bar{\theta}_m - \bar{\theta})^2$, and W as $\frac{1}{M(N-1)}\sum_{m=1}^{M}\sum_{n=1}^{N}(\theta_{nm} - \bar{\theta}_m)^2$, where θ_{nm} is the nth draw from the mth chain, $\bar{\theta}_m$ is the mean of the mth chain, $\bar{\theta}$ is the mean of the means of the chains, M is the number of chains, and N is the number of draws per chain [59, p. 282-285].

An important concern regarding the quality of MCMC samples is the impact of serial correlation, which can lead to an overestimation of the sample size and, consequently, the precision of estimates. To address this, the concept of effective sample size (ESS) is used, which adjusts the number of drawn samples by accounting for correlation. ESS is estimated as $\text{ESS} = \frac{MN}{1+2\sum_{t=1}^{T}\hat{\rho}_t}$, where ρ_t is the autocorrelation at lag t. To calculate ρ_t, we

first compute the total variance using the variance estimate from the marginal posterior variance ($\hat{V}$). Then, for each lag t, we compute the variogram V_t using the formula $V_t = \frac{1}{M(N-t)} \sum_{m=1}^{M} \sum_{n=t+1}^{N} (\theta_{nm} - \theta_{n-t,j})^2$, and estimate ρ_t as $\hat{\rho}_t = 1 - V_t/2\hat{V}$. The sum of autocorrelations is then computed, summing from lag 0 until the sum of estimates for two successive lags becomes negative. This sum is used in the ESS formula, providing a more accurate measure of the number of independent draws from the posterior. In general, lower $\hat{R}$ values (close to 1) and higher ESS values are good indications of convergence and efficiency of MCMC methods. For many purposes, $\hat{R}$ values below 1.1 and ESS of 100 are considered sufficient [59, p. 286-288].

1.3.2 Machine Learning Methods

Despite the usefulness of statistical approaches, they have certain limitations. They often assume linear relationships and stationary processes, which might not be representative of complex and dynamic coastal systems. Furthermore, developing statistical models, which involves carefully thinking about the data generating processes, can be very challenging, especially when dealing with problems that include multi-scale and non-linear interactions. These interactions are often intrinsic to the functioning and response of these systems. On the other hand, machine learning (ML) models can aid in deciphering the complex patterns underlying the data automatically. Recent advancements in computational power and data availability have facilitated the use of ML methods in coastal transformation and environmental monitoring research. ML techniques such as decision trees, random forests, neural networks, and deep learning algorithms have been employed to handle high-dimensional, non-linear, and complex datasets [41].

These methods have shown potential in predicting coastal transformation phenomena [61]. For instance, deep learning models have been used to analyse satellite imagery for the detection and prediction of coastlines and shorelines which are subject to changes due to factors such as sea level rise [32]. Deep learning models have demonstrated significant potential in predicting sea-level rise and analysing its impacts [50].

ML methods are known to be adept at handling complex and non-linear relationships. However, they are often seen as 'black boxes' due to their lack of transparency and interpretability. The accuracy of ML models depends largely on the quality and quantity of the input data, making them prone to issues like overfitting and biases in the presence of noisy data.

Artificial Neural Networks

Artificial Neural Networks (ANNs) constitute a category of machine learning models inspired by biological neural networks and are particularly effective in addressing non-

linear and complex problems [112]. ANNs employ interconnected layers of nodes, or "neurons", that process and transmit information, much like neurons in a biological brain. Each neuron in the network takes a set of inputs, applies a weighted function to these inputs, and produces an output. ANNs can be expressed mathematically as a series of functions, typically linear combinations followed by non-linear activations. In its simplest form, the output of a neuron y can be described as a function f of the weighted sum of its inputs x, plus a bias b. Mathematically, this is often represented as:

$$y = f\left(\sum_i (w_i \cdot x_i) + b\right)$$

Here, w_i represents the weights, which determine the strength and sign of the input, and b is the bias term. The function f is a non-linear activation function like the sigmoid, hyperbolic tangent, or ReLU function, which introduces non-linear properties to the model.

In a multi-layered ANN, the output of one layer of neurons becomes the input for the next, allowing the network to learn complex patterns:

$$y = f_n(...f_2(f_1(\mathbf{X} \cdot \mathbf{W}_1 + \mathbf{b}_1) \cdot \mathbf{W}_2 + \mathbf{b}_2)... \cdot \mathbf{W}_n + \mathbf{b}_n)$$

Here, $\mathbf{X}$ is the input vector, $\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_n$ are the weight matrices for each layer, $\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n$ are the biases, and $f_1, f_2, ..., f_n$ are the activation functions for each layer respectively.

Through a process called "training", ANNs adjust their weights and biases to minimize the difference between the actual output and the desired output for a set of training data. This process typically involves a method known as backpropagation [110], combined with an optimization algorithm like gradient descent [109]. The network learns by iteratively updating its weights to reduce the error, improving its predictions over time.

ANNs in general, and deep learning models in particular, have demonstrated superior capabilities in diverse fields such as image and speech recognition and natural language processing [81]. However, despite their strengths, ANNs can sometimes suffer from issues such as overfitting and lack of interpretability. My research made use of an ANN to learn from a large dataset of multivariate time series to do spatiotemporal forecasting and extended it to provide interpretable insights into the importance of the predictors (Chapter 4).

Transformers

The transformer is a type of deep learning model, which is composed of artificial neural networks with many layers of processing units, designed to learn representations of data

with multiple levels of abstraction. Introduced in the paper "Attention is All You Need" by [128] in 2017, this architecture revolutionized the field of Natural Language Processing (NLP) by utilizing a novel "multi-head self-attention" mechanism. This mechanism allows the model to capture dependencies between words in a sentence without being restricted by their positional distance, a substantial advantage over previous models that relied on recurrence and convolutions.

The core of the transformer consists of multi-head self-attention mechanisms and feed-forward networks. These components work together to enable the model to process different parts of an input sequence simultaneously, which not only improves efficiency but also effectively handles long-range dependencies in sequences. Because of this parallelization capability, transformers have laid the groundwork for state-of-the-art models like GPT-3 from OpenAI and BERT from Google [22, 38]. These models excel in various NLP tasks such as translation, question-answering, summarization, and language generation, showcasing remarkable advancements in the field.

Transformers are composed of two main parts: an encoder and a decoder. The encoder is used to process the input sequence, while the decoder is employed in generating an output sequence, making the transformer architecture particularly well-suited for sequence-to-sequence tasks like machine translation. However, in certain applications such as time series prediction or text classification, only the encoder part may be utilized.

The self-attention mechanism is a defining feature of the transformer, as it enables each item in the input sequence to weigh the relevance of other parts of the sequence dynamically. This characteristic allows the transformer to focus on different elements of the input, ensuring that the model captures dependencies no matter how far apart they are in the sequence.

In addition to its applications in NLP, the transformer structure has been adapted for use in other domains, such as spatiotemporal forecasting [64]. By leveraging the transformer's architecture, new deep learning models can be developed to forecast complex patterns over both space and time, underscoring the transformer's versatility and power in handling a wide array of sequence-based problems.

Application of Transformers to Time Series Data

One of the traditional challenges with time series data is the capture of long-term dependencies. While recurrent neural networks (RNNs) and their variants, like long short-term memory networks (LSTMs), have been used for this purpose, they often face challenges in capturing very long-term dependencies [113, 104].

Transformers, with their self-attention mechanism, overcome this challenge. They can give varying attention scores to different time points in the past, making them adept at

understanding both recent and older historical data's relevance in predicting future occurrences. It is important to note, however, that this capability incurs a computational cost. The self-attention mechanism in transformers has a time and space complexity that is quadratic in the length of the input sequence [77]. Additionally, unlike RNNs, transformers are limited by the fixed length of their input sequence. They can learn long-range dependencies within this sequence, but their capability to do so is bounded by this length. This limitation can be mitigated by increasing the length of the sequence, though this comes at the cost of increased computational demand. Despite these, the computations can be efficiently parallelized on modern GPUs. This is a significant advantage over recurrent methods, which typically require sequential processing.

For environmental time series data, as in the Predict project, transformers' capabilities are invaluable. Environmental factors often have intricate relationships that play out over varying time scales. For instance, a sudden change in one factor might influence another factor hours or even days later. Transformers can capture these relationships efficiently.

The Predict project dealt with multivariate environmental time series data, where multiple environmental factors are monitored over time. Transformers are particularly well-suited for this kind of data. Given their architecture, transformers can not only capture temporal relationships (how one factor evolves over time) but also cross-sectional relationships (how different factors influence each other at a given time). This dual capability ensures that predictions made are based on a holistic understanding of the environment, accounting for the interplay of various factors. In the project, this would mean that while predicting coastal changes, the transformer model would consider the combined influence of factors like sea level, turbidity, temperature, and others, understanding how a change in one might ripple through and influence others.

1.3.3 Summary

In summary, both statistical and machine learning methods have significant roles in studying coastal transformation and environmental monitoring. While statistical methods provide robustness and interpretability, ML methods offer flexibility and adaptability in handling complex, large high-dimensional data. Therefore, the choice between these methods should be guided by the specific objectives of the research, the nature of the dataset, and the resources available. An integrative approach, utilizing both statistical and machine learning methods, can potentially offer the most comprehensive understanding of coastal transformations. Such an approach can take advantage of the interpretability and robustness of statistical models while harnessing the power of machine learning to handle high dimensional data and complex relationships. Thus, employing a combination of these techniques may result in more effective monitoring and prediction of coastal transforma-

tions, helping to inform more sustainable management practices and policy decisions in the face of changing climate and human impacts.

1.4 Contributions of this thesis

This thesis makes contributions to the fields of environmental data analysis, statistical modelling, and spatio-temporal forecasting, particularly concerning the marine environment of Dublin Bay, Ireland. The thesis's contributions are articulated through three published/submitted papers, each of which introduces new models, methods, or applications.

The first significant contribution of this thesis involves the comprehensive processing and analysis of tide gauge data for Dublin Port. This includes the collation of data from multiple sources, rigorous quality checks, and bias detection and correction via change point and multiple linear regression models. Through this rigorous methodology, an accurate estimation of the rate of sea level rise in Dublin Bay is achieved, which represents a critical contribution to our understanding of local sea level changes under climate change scenarios.

The second contribution lies in the statistical analysis and understanding of another critical environmental variable, turbidity in Dublin Bay. Here, this thesis introduces innovative statistical models, namely Vector Auto-Regressive Conditional Heteroskedasticity (VARCH) and Vector Auto-Regressive Integrated Conditional Heteroskedasticity (VARICH). These models present a novel way to measure the effects of multiple time series on each other while taking into account known volatility changes in the time series. The Bayesian modelling, combined with VAR and ARCH structures, makes the VARICH model a flexible tool for modelling a wide array of real-world random processes where spatial and temporal aspects play significant roles. The models also provide a means for quantifying uncertainty in both fixed effects and posterior predictions of the time series, even while imputing missing values. The main findings on the estimated effects of dredging and dumping operations on turbidity levels in Dublin Bay provide valuable insights for environmental management and policy-making.

The third and final contribution of this thesis lies in the development of two novel models for spatio-temporal forecasting: SERT and SST-ANN. SERT, based on the transformer paradigm, and SST-ANN, an ANN model combined with triplet encoding, provide new tools for this complex task. Additionally, the thesis introduces the use of STraTS [125], a model initially designed for clinical time series classification, and extends it for use in spatio-temporal forecasting. All these methods are particularly advantageous because they do not require data aggregation or missing value imputation, thus circumventing problems introduced by such techniques. Evaluations of these models on simulated and

real-world environmental datasets from Dublin Bay demonstrated their robust performance, with SERT outperforming other models in 7-hour ahead forecasting for most variables. Furthermore, SST-ANN was used to interpret the model's predictions, providing an additional tool for understanding and leveraging spatio-temporal forecasts.

In conclusion, this thesis significantly advances our understanding of Dublin Bay's marine environment and provides new statistical and machine learning tools that can be applied to similar spatio-temporal datasets. By doing so, it contributes to improving our ability to respond to environmental changes and challenges in Dublin Bay and other similar marine environments.

2

A newly reconciled data set for identifying sea level rise and variability in Dublin Bay

We provide an updated sea level dataset for Dublin for the period 1938–2016 at yearly resolution. Using a newly collated sea level record for Dublin Port, as well as two nearby tide gauges at Arklow and Howth Harbour, we perform data quality checks and calibration of the Dublin Port record by adjusting the biased high water level measurements that affect the overall calculation of mean sea level (MSL). To correct these MSL values, we use a novel Bayesian linear regression that includes the mean low water values as a predictor in the model. We validate the re-created MSL dataset and show its consistency with other nearby tide gauge datasets. Using our new corrected dataset, we estimate a rate of sea level rise of 1.1 mm yr^{-1} during 1953–2016 (95 % credible interval from 0.6 to 1.6 mm yr^{-1}), and a rate of 7 mm yr^{-1} during 1997–2016 (95 % credible interval from 5 to 8.8 mm yr^{-1}). The overall sea level rise is in line with expected trends, but large multidecadal variability has led to higher rates of rise in recent years.

2.1 Introduction

The global mean sea level (MSL) is rising due to anthropogenic climate change [40, 87], which is defined as the alteration of the Earth's climate system due to human activities. Understanding regional sea level trends is crucial for local and regional adaptation and the development of effective climate action plans. In Ireland, Dublin is the largest city, with a population of approximately 1.42 million [31], and is situated at the mouth of the river Liffey on the Irish Sea coast. Dublin also has the Republic of Ireland's longest tide gauge record (from 1938 onwards): the publicly available record for Dublin Port (also called Dublin North Wall). Understanding changes in mean sea level in Dublin is key for the protection of Ireland's largest city and, from a national perspective, for understanding long-term sea level rise (SLR) in Ireland [5].

The sea level around Ireland rose rapidly after the last glacial maximum 20 000 years ago, cutting Ireland off as an island 16 000 years ago [48]. Regionally, sea levels in Ireland had stabilised by the 20th century, after which sea levels began to rise again due to anthropogenically induced warming [87]. The importance of climate warming and SLR in Ireland has been emphasised by a number of authors: [39] discussed the risks of extreme climatic events and the ways in which Ireland should be prepared for them, while [24] state that satellite observations, which are associated with the open ocean rate, show sea levels rising around Ireland at a rate of 2–3 $\mathrm{mm\,yr^{-1}}$, in line with global averages for the early 21st century. However, the raw tide gauge record at Dublin Port shows a rate of sea level rise of 0.3 $\mathrm{mm\,yr^{-1}}$ from 1938 to 2000 [5] – much lower than the global average.

A number of authors have investigated trends in the Dublin Port tide gauge data prior to the year 2000 and have found similarly low rates of change. [27] investigated the Dublin Port record using tide gauge measurements, and reported a rising trend of 0.6 $\mathrm{mm\,yr^{-1}}$ before 1961 and a falling trend of $-0.3\,\mathrm{mm\,yr^{-1}}$ from then until 1980. [133, 134] estimated trends of 0.17 $\mathrm{mm\,yr^{-1}}$ ($\pm\,0.35$) from 1938 to 1988, and 0.23 $\mathrm{mm\,yr^{-1}}$ from 1938 to 1996.

In stark contrast to these low rates of SLR, the recently published climate change action plan for 2019–2024 by Dublin City Council [36] reports an SLR of 6–7 $\mathrm{mm\,yr^{-1}}$ between the years 2000 and 2016. This rate is approximately double that of the global mean sea level rise [93] and is particularly surprising given that the earlier rates of rise in Dublin were much lower than the global mean sea level rise over similar periods [33].

The goal of this paper is to further investigate the sea level trend in Dublin Port through careful assembly and quality control of the available data and by comparing those data with sea level records collected from nearby tide gauges. We find that the datums of Dublin's available sea level measurements are not aligned with those of nearby tide gauges and thus need further consideration. In addition, we find problems with the mean high water measurements, which indicate a drift over time. We use a Bayesian multivariate

linear regression to adjust for the drift and, after removing atmospheric factors, we find that the sea level record, at least for the 21st century, matches other local tide gauges to a far higher degree. This allows us to estimate more reliable measurements of sea level rise for the urban area of Dublin Port.

The rest of this paper is organised as follows. Section 2 explains how the sea level dataset for Dublin Port is reconciled from various sources. Section 3 discusses the quality check and calibration procedures done on the reconciled dataset. Section 4 discusses SLR rates at Dublin Port. Finally, some issues and suggestions pertaining to SLR analysis at Dublin Port and the important findings of this study are discussed and summarised in Section 5.

2.2 Data collation for Dublin Port

We compiled mean high and low water, mean tide level and, where available, mean sea level for Dublin Port from 1938 to 2018 from the following sources:

(1) Annual high and low water from [133, 103] for the period 1938–2001 from annual tabulations made by the Dublin Port Authority (hereafter the Port Authority annual dataset). Mean tide level (MTL) was calculated by averaging the mean high and low waters. Data from 1938 to 1977 are relative to Port Datum, which is 0.436 m above the Ordnance Survey Datum Dublin (Poolbeg Datum), and data from 1978 are relative to the lowest astronomical tide (LAT) which is 20 cm above Poolbeg Datum.

(2) Monthly values of mean high water (MHW) and mean low water (MLW) for the period of 1987–2001 (hereafter the Port Authority monthly dataset), which were digitised as part of this study. Quality control measures for the digitisation included automatic comparison and checking of the calculated and recorded mean levels. Comparison of the overlapping Port Authority datasets showed a mean difference of ≤ 1 mm during the 15 overlapping years (1987–2001). Data are reported relative to LAT.

(3) High-frequency (10 min) data supplied by the Permanent Service for Mean Sea Level [103, 69] for the period 2002–2009. These data were provided to the PSMSL by the harbourmaster in Dublin Port following a change in responsible authority in 2001 (hereafter the harbourmaster dataset). The data have a low vertical resolution of 0.1 m and are reported relative to LAT.

(4) High-frequency (5 min) data for the period 2007–2018 from the Irish National Tide Gauge Network (NTGN), which is maintained by the Irish Marine Institute (here-

Table 2.1: Details of the datasets collated to form a complete sea level record for Dublin Port.

Dataset name	Duration	Sampling frequency	Variables	Datum	Provider
Port Authority	1938–2001	Annual	MLW-MTL-MHW	LAT	PSMSL
Port Authority	1987–2001	Monthly	MLW-MTL-MHW	LAT	PSMSL
Harbourmaster	2002–2009	10 min	All calculable	LAT	PSMSL
NTGN	2007–2018	5 min	All calculable	ODM and LAT	Marine Institute
Greene	1968–2015	Twice daily	MHW	LAT	This study

after the NTGN dataset) [72]. Data are available relative to Ordnance Datum Malin (ODM) and LAT. All NTGN data are defined relative to ODM and LAT.

(5) High water levels for the period 1968–2015. These were digitised as part of an unpublished MSc thesis of Alice Greene (hereafter the Greene dataset) and are published here for the first time. The Greene dataset for the period 1968–1982 was transcribed from photographs of tidal charts from which the high water values can be read. During 1983–2003, the data were recorded in the format of hard-copy tidal charts. The hard copies consisted of three large A3 books. The remaining data from 2003–2013 from which high waters could be derived were in digital format.

With particular reference to the Greene dataset, prior to the availability of digital data in 2003, the high water values for each day were extracted from the tidal charts. This was completed by the generation of tables for each year, with two available cells for each day. These values were read off and input into the designated cell. The data from the period 1968–1976 were converted from feet and inches to metres. To locate the two high tides, for each day, the highest value was extracted. The second high tide occurred between 12 and 13 h before or after the highest value; therefore this value was extracted using the time component of the dataset. In the case where the time window went into the next calendar day, only a single high tide was recorded for that day. A summary of the datasets is shown in Table 2.1.

Difficulties in merging the Dublin Port datasets arose from differing datum definitions. For both Port Authority datasets, the tabulated annual and monthly data are relative to the same datum. These data have the same source and therefore agreement is expected. Three years of overlap exist between the harbourmaster dataset and the NTGN dataset from 2007–2009. The harbourmaster dataset is relative to LAT datum; the NTGN data are relative to ODM, with a value of 2.599 m between these datums. There is found to be a systematic underestimation of 0.044 m in the MHW and MLW values in the harbourmaster dataset, presumably due to its lower vertical resolution of 0.1 m, determined from the overlap with the NTGN data.

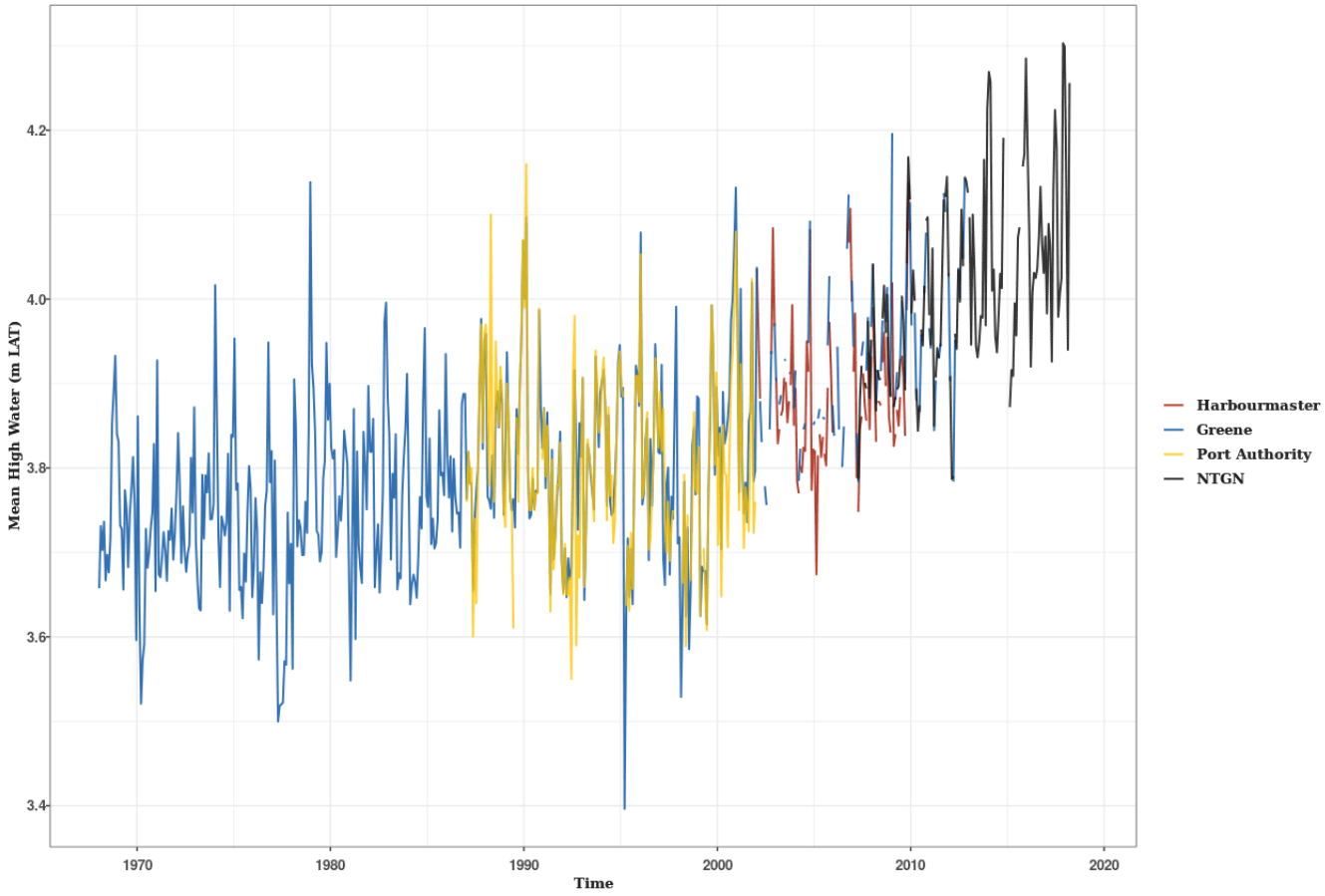

Figure 2.1: Monthly MHW in Dublin from the Port Authority, Greene, harbourmaster and NTGN datasets. A high level of agreement is found between the records, indicating consistent datum definition, with only a small adjustment of 0.008 m to the Greene dataset required for complete reconciliation.

While no overlaps exist between the harbourmaster dataset and the Port Authority dataset, the Greene dataset overlaps the Port Authority, harbourmaster and NTGN datasets. Figure 2.1 shows monthly MHW from the Port Authority, harbourmaster, NTGN and Greene datasets, each expressed with respect to LAT (note that our monthly data are limited to 1968 onward). There is a high level of agreement between the data, indicative of consistent datums. We find a residual 0.008 m difference between the Greene dataset and the other two datasets (Port Authority monthly dataset and the NTGN dataset). We thus add 8 mm to the Greene dataset as the final datum adjustment.

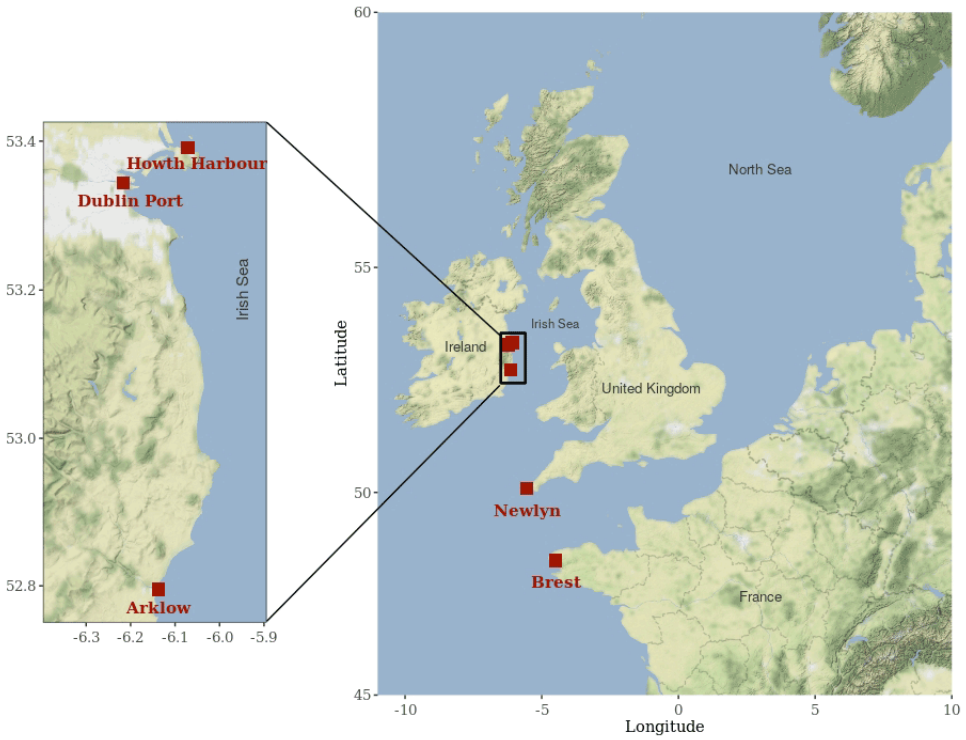

Figure 2.2: Locations referred to in this study: Dublin Port, Howth Harbour, Arklow, Brest and Newlyn.

2.3 Reconciliation of Dublin Port against nearby tide gauges

We now use our newly merged dataset (hereafter the merged Dublin Port dataset) including monthly MLW, MTL, MSL and MHW measurements. In order to check the reliability of the merged Dublin Port dataset, we compare with two different nearby tide gauges (maximum distance 60 km) at Howth Harbour and Arklow, and two other tide gauges at Newlyn in the UK and Brest in France. The Arklow and Howth Harbour datasets are derived from bubbler gauges. The Newlyn and Brest data are gauges with a long history of use in sea level studies [18, 136]. Figure 2.2 shows the locations of the tide gauges. Table 2.2 provides the details of the datasets.

For the Arklow and Howth Harbour datasets, we first aggregate the values up to daily and monthly level for MLW, MSL and MHW. During our pre-processing, we compared MLW, MSL and MHW at Dublin Port with those at Arklow and Howth Harbour. We noticed that the other sites exhibited low levels of agreement with Dublin Port after 2017, so we restricted our analysis to use only data up to the end of 2016.

Figure 2.3 and Table 2.3 demonstrate that rates of SLR in MSL and MHW are significantly higher in Dublin Port than in Arklow or Howth Harbour. A possible cause is

Table 2.2: Details of the four datasets used for comparison with Dublin Port.

Dataset	Duration	Sampling frequency	Provider
Arklow	2003–2018	15 min	Office of Public Works
Howth Harbour	2007–2018	6 min	Marine Institute
Brest	1938–2016	Annual	PSMSL
Newlyn	1938–2016	Annual	PSMSL

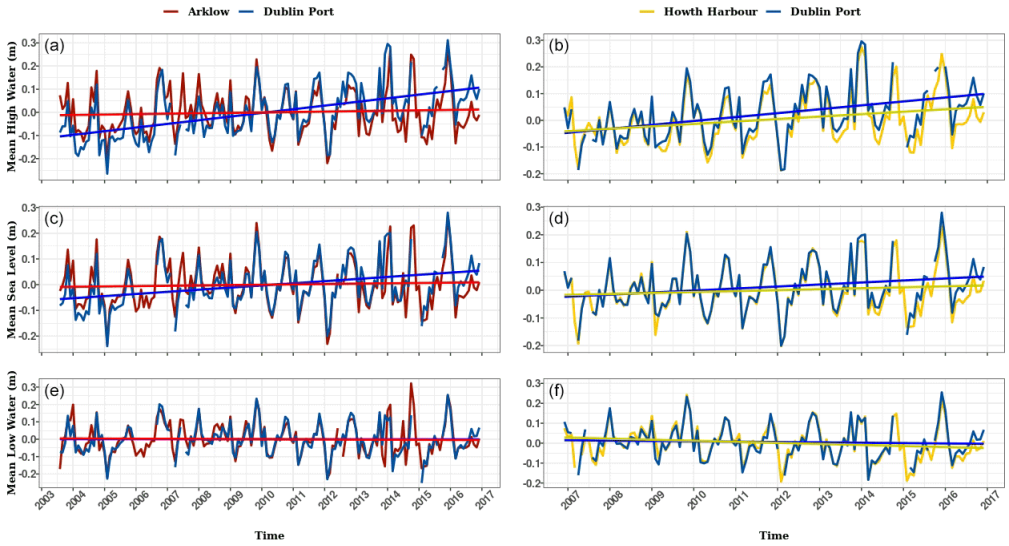

Figure 2.3: Comparison of MHW, MSL and MLW monthly values for Dublin Port against those for Arklow **(a, c, e)** and Howth Harbour **(b, d, f)**, with linear trends shown for each. The MLW linear trends for Dublin Port and Arklow are in good agreement, as are those for Dublin Port and Howth Harbour. However, this is not the case with the linear trends for MSL and MHW when comparing Dublin Port to the two other locations.

the malfunction of the Dublin Port tide gauge in measuring the high water levels, due to drift. A Druck pressure transducer was used at Dublin Port [92], and this has the potential to exhibit drift proportional to the height of the water column. Accuracies of pressure sensors are reported to be proportional to full water column depth. While drift need not necessarily be proportional to full water column depth, this is a possibility that is supported by further analysis (see next section). Table 2.4 provides the details of the tide gauges installed at Dublin Port from 1938 to the present.

From Figure 2.3, we can see that the MLW values at Dublin port are highly similar to those from nearby gauges, so we use these as a baseline to correct the MSL values. To do this, we create a regression model that estimates MSL given MLW from older Dublin Port measurements. We then use the predictions from this model to estimate MSL at Dublin Port for the more recent period.

Table 2.3: Differences between rates of SLR in MLW, MSL and MHW for Dublin Port and those for Arklow or Howth Harbour. Large values with small standard errors indicate a significantly higher rate at Dublin Port.

Locations	MLW ($\mathrm{mm\,yr^{-1}}$)	MSL ($\mathrm{mm\,yr^{-1}}$)	MHW ($\mathrm{mm\,yr^{-1}}$)
Dublin Port – Arklow	0.6 ($\pm$2.8)	6.9 ($\pm$2.6)	14 ($\pm$2.7)
Dublin Port – Howth Harbour	1.5 ($\pm$4.0)	3.8 ($\pm$3.9)	5.8 ($\pm$4.2)

Table 2.4: Details of the tide gauges installed at Dublin Port.

Gauge type	Start	End	Reference
Float gauge	1938	2001	Original chart supplied by Dublin City Council indicates the float gauge in operation.
Druck pressure transducer	2001	2016	This sensor began to malfunction in 2017 and was replaced in 2018.
OTT bubbler	2018	–	New sensor installed in 2018. Data not used in this study.

To find the period of time over which to train the regression, we use a change point model [25] that takes the absolute difference between MSL values of Dublin Port and Newlyn in the UK as the inputs. Details of the change point model are discussed in the Appendix. The model suggests that there is a change point in the agreement between Dublin and Newlyn in 1976. Furthermore, Figure 2.4 shows the comparison of MSL data from Dublin Port (not yet corrected for the bias demonstrated in Figure 2.3) with that of Newlyn and Brest in France. These two are selected due to the relative completeness and integrity of their records and their proximity to Dublin. According to the figure, there is strong agreement between stations for MSL in the period 1938–1976. After 1976, the level of agreement deteriorates, which is consistent with the change point model result. Note that this does not necessarily mean that the weaker agreement post-1976 is related to data quality. However, based on these observations, we do not have any evidence supporting the quality of data from 1976 onwards, and so we limited our modelling approach to data from the preceding period.

We correct the bias in the MSL values at Dublin Port using a Bayesian multivariate linear regression. Our model consists of an intercept, a term to account for a linear effect of MLW, a harmonic function with a period of 18.6 years to model the lunar nodal cycle, and a period of 4.4 years to account for the 8.85-year cycle of lunar perigee [65, 135]. The two latter terms will primarily contribute to the astronomical tidal component of MLW.

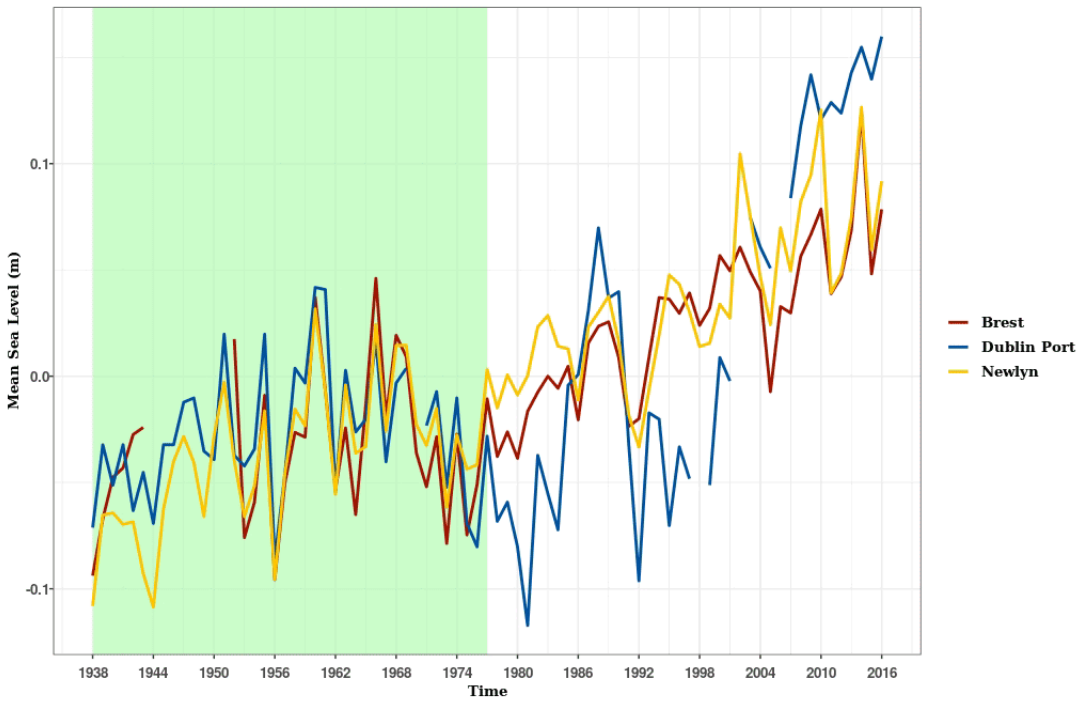

Figure 2.4: Yearly MSL values for Dublin Port, Newlyn and Brest. The green area shows our chosen time period, during which there is good agreement between Dublin Port and the other sites.

The model formulation is as follows:

$$
\begin{aligned}
(\mathrm{MSL})_t &\sim \text{Normal}\left(\mu_t, \sigma^2\right) \\
\mu_t &= \beta_0 + \beta_1(\mathrm{MLW})_t + \beta_2 \cos\left(\omega_1 t\right) + \beta_3 \sin\left(\omega_1 t\right) \\
&\quad + \beta_4 \cos\left(\omega_2 t\right) + \beta_5 \sin\left(\omega_2 t\right) \\
&\text{with} \quad \omega_1 = \frac{2\pi}{18.61},\ \omega_2 = \frac{2\pi}{4.4},
\end{aligned}
$$

where $(\mathrm{MSL})_t$ is the MSL in year t, μ_t is the mean, σ^2 is the residual variance, β_0 is the intercept, β_1 is the MLW coefficient, β_2 and β_3 are the amplitudes of the cosine and sine functions of the 18.6-year lunar nodal modulation respectively, and β_4 and β_5 are the amplitudes of the cosine and sine functions of the 4.4-year modulation respectively.

We fitted the model using the JAGS software [102] and R [105], and used three Markov chain Monte Carlo chains (2000 iterations per chain with 1000 as burn-in, and a thinning value of 1). Convergence was assessed using the R-hat diagnostic [20, 60]. All R-hat values associated with β and σ were close to 1, so the model was assumed to be sampling from the posterior distribution. The minimum ESS across all the parameters was 790. The new estimated MSL and the original values, together with the yearly MSL values of

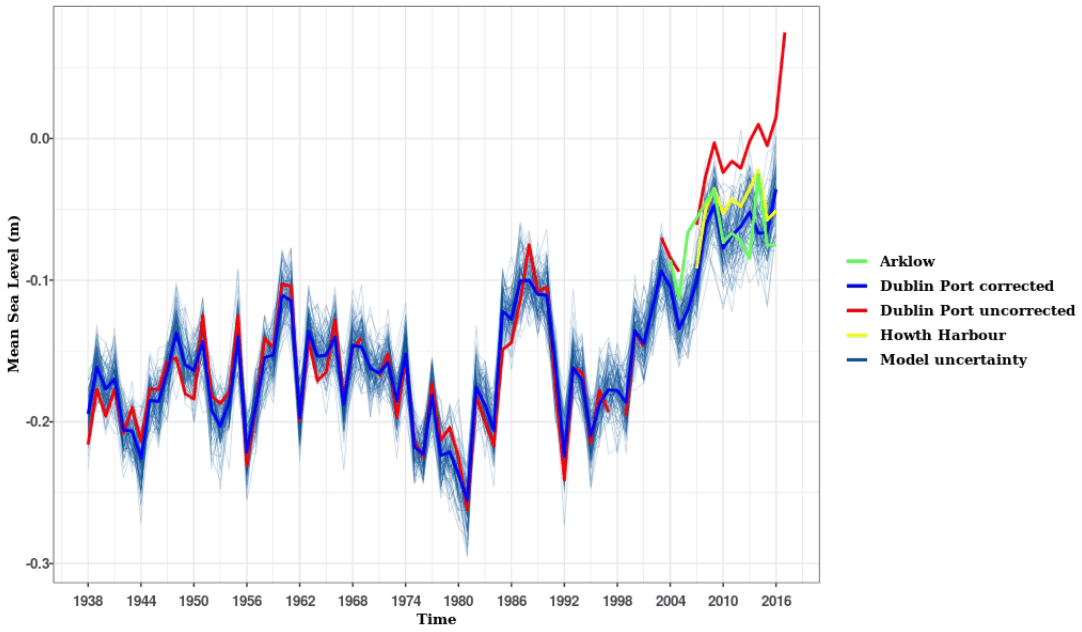

Figure 2.5: The uncorrected and corrected yearly MSL values of Dublin Port, with yearly MSL values of Arklow and Howth Harbour for comparison. The newly corrected Dublin Port MSL values lie much closer to the neighbouring tide gauges. The faded lines in the background show posterior samples of MSL from the model and an indication of model uncertainty.

Arklow and Howth Harbour, are shown in Figure 2.5. This figure demonstrates that the newly modelled Dublin Port yearly MSL data (posterior predictive mean from the model) are changed only slightly between 1938 and 2001. After 2001, we can see a clear gap between the old (red) and new (blue) versions, with the new corrected data exhibiting superior agreement with the Arklow and Howth Harbour records. We note again here that these adjacent records were not used in the creation of the new Dublin Port data so are an independent validation of our approach. The larger adjustment in Figure 2.5 after 2001 seems to support the discrepancies shown in Figure 2.3 for the higher water levels and provides further justification for our change point calibration approach.

2.4 Rates of sea level rise at Dublin Port and nearby gauges

We now use the corrected data from Dublin Port to calculate rates of sea level rise. We use the yearly MSL data from Brest and Newlyn for comparison. The MSL data from Brest is missing between 1944 and 1952, so we decided to limit our SLR rate estimations to 1953–2016, during which the data for all three sites were complete. We initially removed the inverse barometer and wind effects on sea level at each site by regressing MSL on

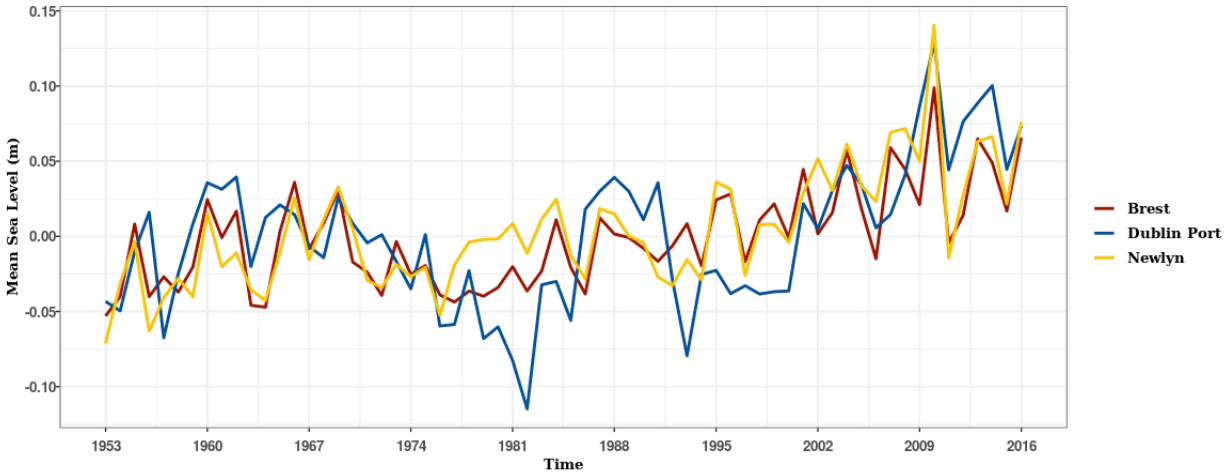

Figure 2.6: New yearly MSL values of Dublin Port and yearly MSL values of Brest and Newlyn, with atmospheric effects removed, between 1953 and 2016.

the mentioned variables and using the residuals for subsequent steps, as suggested in [52] and [42](we omitted this step in the previous section due to a lack of atmospheric data during 1938–1948). While it's possible to directly include these variables in the models for estimating SLR rates, we opted for a more common approach in oceanography literature, knowing that the results would not significantly differ. Atmospheric data are accessed via the RNCEP package [78] in the R programming language, which accesses the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) and Department of Energy Reanalysis I and II datasets [74, 75]. Figure 2.6 shows the atmospherically corrected MSL data from Dublin Port, Brest and Newlyn superimposed for comparison.

To calculate the rates of SLR, as before, we use a Bayesian multivariate linear regression, including an intercept and a linear trend. The model is fitted in JAGS with the same settings and convergence requirements as previously described. We write the model as

$$(\text{MSL})_t \sim \text{Normal}\left(\mu_t, \sigma'^2_t + \sigma^2\right)$$
$$\mu_t = \beta_0 + \beta_1 t,$$

where $(\text{MSL})_t$ is the MSL at time t, μ_t is the mean, σ'^2_t is the posterior predictive variance at time t extracted from the posterior distribution of the calibration model to account for the uncertainty in modelling the MSL introduced in the previous section, σ^2 is the residual variance, β_0 is the intercept and β_1 is the rate in $\text{mm}\,\text{yr}^{-1}$. We use the same approach (but without the fixed measurement error) for estimating the rates of rise at

Table 2.5: Estimated rates of SLR ($\mathrm{mm\,yr^{-1}}$) at Dublin Port, Newlyn and Brest, with 95 % credible intervals for the time periods of interest.

Location	1953–2016	1975–1985	1986–1996	1997–2016
Dublin Port	1.1 (0.6, 1.6)	−2.9 (−9, 2.9)	−8.6 (−13.4, −3.7)	7 (5.0, 8.8)
Newlyn	1.3 (1.0, 1.6)	3.9 (0.7, 6.9)	1.6 (−3.4, 6.8)	3.1 (0.7, 5.5)
Brest	0.9 (0.6, 1.2)	2.5 (0.0, 5.0)	3.1 (−0.2, 6.4)	2.3 (0.3, 4.3)

Brest and Newlyn.

The estimated rates with their associated 95 % posterior credible intervals are given in Table 2.5, which shows that, between 1953 and 2016, the rate of SLR at Dublin Port has a mean estimate of $1.1\,\mathrm{mm\,yr^{-1}}$, consistent with the estimated rate of $0.9\,\mathrm{mm\,yr^{-1}}$ at Brest and that of $1.3\,\mathrm{mm\,yr^{-1}}$ at Newlyn. However in more recent years, specifically between 1997 and 2016, Dublin Port has experienced a greater SLR of $7\,\mathrm{mm\,yr^{-1}}$, larger than the SLRs of $2.3\,\mathrm{mm\,yr^{-1}}$ at Brest and $3.1\,\mathrm{mm\,yr^{-1}}$ at Newlyn. Figure 2.6 also suggests that sea level in Dublin Port has experienced larger decadal fluctuations and is not as secular as the sea level at the two other locations.

2.5 Discussion and conclusions

Taken over the full time period of observations, 1953–2016, the estimated sea level rise of $1.1\,\mathrm{mm\,yr^{-1}}$ in Dublin is consistent with those in Brest and Newlyn, both of which are located on the western European coastline. The rates of rise for earlier periods (i.e. pre-1953) are less than $1.1\,\mathrm{mm\,yr^{-1}}$ [27, 133] and are consistent with the findings here, and were lower because the decades of larger sea level rise and variability (1980s, 2000s) were not included in the trend estimation. Elsewhere in Ireland, [98] investigated tide gauge records in Malin Head (1958–1998) and Belfast harbour (1918–2002), where they reported substantial annual variation for both sites, with overall negative trends of $-0.2\,\mathrm{mm\,yr^{-1}}$ for Belfast and $-0.16\,\mathrm{mm\,yr^{-1}}$ for Malin Head. Both Belfast and Malin Head, located in the north of the country, are in regions experiencing uplift due to glacial isostatic adjustment (GIA). This process involves the Earth's surface slowly rebounding or adjusting in response to changes in the weight of ice sheets and glaciers that cover it, which in turn reduces the relative sea level rise in these areas. However, Dublin is situated in a region where glacial isostatic uplift is neutral. Therefore, the long-term effects of post-glacial land motion should be negligible there. As a result, greater consistency between local sea level rise rates and the global average is expected [17].

More surprising is the large decadal variability revealed. This study has found a rate of sea level rise for Dublin of $-2.9\,\mathrm{mm\,yr^{-1}}$ for the period 1975–1985, followed by another negative sea level trend in the next decade ($-8.6\,\mathrm{mm\,yr^{-1}}$ during 1986–1996) and a rise

of 7 mm yr^{-1} for the period 1997–2016. Similar patterns of decadal variability in sea level to those in Dublin were also noted in Belfast by [97] and are linked to decadal variation of the North Atlantic Oscillation. This would seem a likely explanation for similar patterns in Dublin. However, a full investigation of the causes of decadal variability in Dublin sea level remains to be performed.

Comparisons of the MLW, MSL and MHW recordings at Dublin Port suggest a possible issue with the observation of high water levels. Our model recreated the MSL and showed that there is good agreement between the observed MSL and the modelled MSL for the period 1938–2001. However, after 2001 there is considerable divergence. Our analysis shows that the modelled MSL is more consistent with the data collected by the nearby tide gauges and also at the farther sites in Newlyn and Brest. This suggests that the malfunction probably started during or after the year 2002. We would consider the Howth Harbour sea level record, alongside the modelled MSL data created in our study, as the more reliable dataset for a future analysis of sea level in Dublin Bay compared to the data collected at Dublin Port.

Dublin City Council have recently increased the coastal defences in Dublin, allowing for between 40 and 65 cm of mean sea level rise [100]. Projections of sea level rise for Dublin, based on UKCP18 [54], depend heavily on greenhouse gas emissions trajectories. By 2100, the Dublin mean sea level is projected to rise by 0.6 m at the 50th percentile (1.0 m at the 95th percentile) under an RCP8.5 scenario and by 0.3 m at the 50th percentile level (0.6 m at the 95th percentile level) under an RCP2.6 scenario. These projections do not simulate the decadal scale variability reported here, similar to many other decadal climate phenomena. Understanding the origin and duration of the decadal fluctuations of mean sea level in Dublin is crucial for the preparation and defence of Ireland's capital city in the coming decades.

To sum up, we have collated multiple sources of tide gauge data for Dublin Port, and subsequently corrected them for bias in the MHW level. We have then shown that these corrected MSL measurements agree with both Howth Harbour and Arklow to a far higher degree than the raw data. A longer-term comparison with Brest and Newlyn also indicates overall agreement. There remains a difference during the 1970s and the 1980s, where a large cyclic disparity in Dublin contrasts with the other two records. Our final adjusted dataset estimated the rate of SLR at Dublin Port to be 1.1 mm yr^{-1} between 1953 and 2016, and 7 mm yr^{-1} between 1997 and 2016.

Appendix

To identify the period of data to use in calibrating the Dublin Port MSL and MLW values (see Figure 2.4), we use a linear regression change point model [25]. The model we use can be formulated as follows:

$$\begin{aligned} y_t &\sim N\left(\mu_t, \sigma_t^2\right), \\ \mu_t &= \alpha_{J(t)}, \\ \sigma_t^2 &= \eta_{J(t)}, \\ J(t) &= 1 + u\left(t - t_c\right), \end{aligned}$$

where y_t is the absolute difference between the measured mean sea levels at Dublin Port and Newlyn in year t ($t = 1, 2, \ldots, T$). We assume y_t to be normally distributed with mean μ_t and variance σ_t^2. The mean and variance are set to α_1 and η_1 respectively, if $t < t_c$; and α_2 for the mean and η_2 for the variance otherwise, and t_c is the time of the change point. The function $u(t)$ is the unit step function. We used the following prior distributions:

$$\begin{aligned} \alpha_1 &\sim \text{Normal}(0, 1), \\ \alpha_2 &\sim \text{Normal}(0, 1,) \\ \eta_1 &\sim \text{Uniform}(0, 1), \\ \eta_2 &\sim \text{Uniform}(0, 1), \\ t_c &\sim \text{Uniform}\left(1938, 2016\right), \end{aligned}$$

which are considered to be uninformative, knowing that the range of the data is between 0 and 0.12. Figure 2.7 shows the absolute difference between the MSLs of Dublin Port and Newlyn. The vertical red line indicates the posterior mean of the change-point estimate t_c.

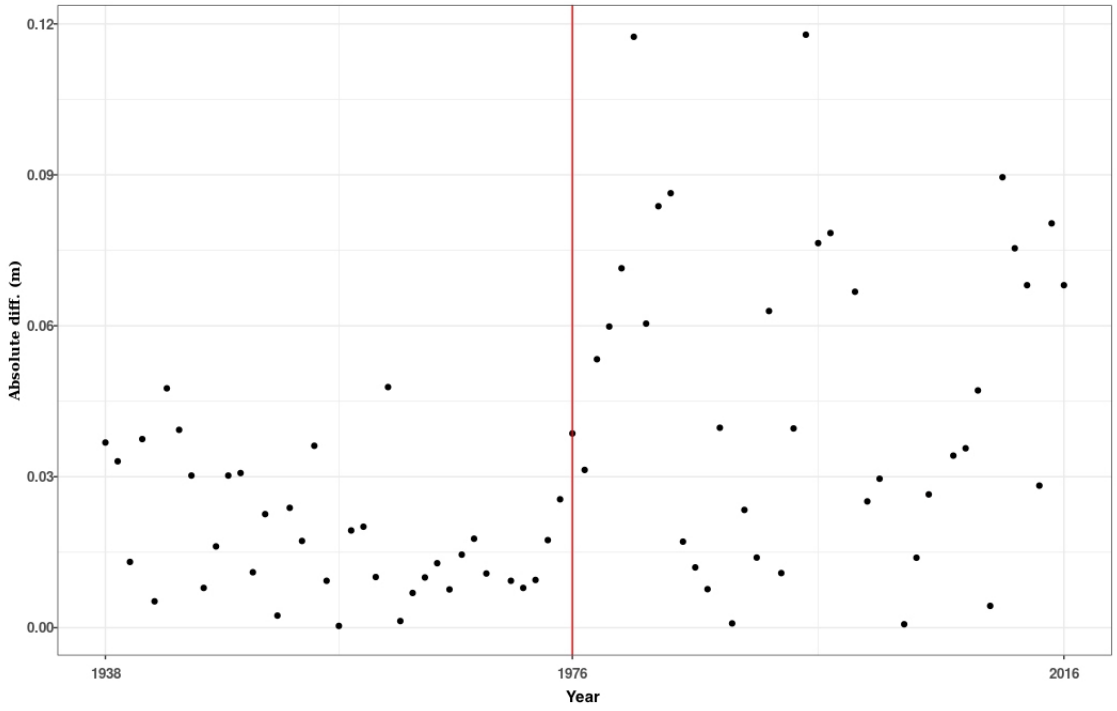

Figure 2.7: The absolute difference in MSL between Dublin Port and Newlyn from 1938 to 2016. The mean posterior estimated change-point time given by the model is indicated by the vertical red line at the year 1976.

3

Vector Time Series Modelling of Turbidity in Dublin Bay

Turbidity is commonly monitored as an important water quality index. Human activities, such as dredging and dumping operations, can disrupt turbidity levels and should be monitored and analyzed for possible effects. In this paper, we model the variations of turbidity in Dublin Bay over space and time to investigate the effects of dumping and dredging while controlling for the effect of wind speed as a common atmospheric effect. We develop a Vector Auto-Regressive Integrated Conditional Heteroskedasticity (VARICH) approach to modelling the dynamical behaviour of turbidity over different locations and at different water depths. We use daily values of turbidity during the years 2017-2018 to fit the model. We show that the results of our fitted model are in line with the observed data and that the uncertainties, measured through Bayesian credible intervals, are well calibrated. Furthermore, we show that, overall, the daily effects of dredging and dumping on turbidity are negligible when compared to the effects of wind speed.

3.1 Introduction

Studying the variables affecting turbidity is of importance in maintaining coastal ecosystem health. Turbidity is an index for water clarity which measures how suspended solids in water hinder the transmission of light [35]. There are many sources of suspended solids including: phytoplankton; particles from coastal erosion; re-suspended bed sediments; organic detritus from streams; and excessive algae growth [19]. Variability in water turbidity influences the transportation dynamics and distribution of nutrients, contaminants, and biological production [124, 79, 43, 56, 57]. Water turbidity is an important habitat factor in many estuarine systems, and changes in it can have a significant impact on management decisions such as the dredging of ports and canals [14].

Our goal in this paper is to evaluate the variations of turbidity in Dublin Bay explained by dredging and dumping operations when controlling for the effect of wind speed, which is an important atmospheric contributor. Dublin has a long history of difficult access for ships to the port area due to sandbanks at the mouth of the port [47]. To solve this problem regular dredging operations have been carried over decades to remove unwanted waste as well as dangerous accumulations of sediments from areas that ships use when entering the port. The excavated materials from the dredging operations are dumped at a more remote location in the bay.

There are relatively few studies focusing on water turbidity in Dublin bay. In one example, [19] used frequentist statistical tests to show that turbidity can be strongly influenced by vessel activity in Dublin bay using data collected from a single location. By contrast, we take a broader approach and look at multiple measuring sites simultaneously corresponding to both the sites where sediment is dumped and dredged, whilst considering issues of turbidity down the water column. We develop a Vector Auto Regressive Integrated Conditional Heteroskedasticity (VARICH) model to control for the spatio-temporal structure using turbidity data measured by five buoys installed at different locations in the bay.

We fit and compare four different models using the turbidity data. The data has many missing values and big gaps for some periods. To fit the models we follow a Bayesian framework to appropriately handle the missingness and infer the parameters of the models using Hamiltonian Monte Carlo (HMC). The purpose of our study is to estimate the effect of covariates on turbidity rather than provide forecasts. The models we fit combine the well-studied approaches Auto Regressive Conditional Heteroskedasticity (ARCH) and Vector Autoregression (VAR). When combined they form a Vector ARCH (or VARCH) model, which we adapt into an integrated model which we name VARICH. A full discussion of these approaches is given below. We show that ARCH-type models perform better for modelling turbidity compared to VAR models that do not account for the het-

eroscedasticity, and in particular our extended model has the best performance of all. We also show that the daily effects of dredging and dumping on turbidity are negligible in comparison to that of wind speed.

We organise our paper as follows. In Section 3.2, we describe the data we use in our study. In Section 3.3, we give a brief introduction to spatiotemporal modelling. In Section 3.4, we explain our modelling framework. In Section 3.5, we discuss our findings including plots of the model outputs. We summarise the paper in Section 3.6 by considering the strengths and weaknesses of our approach and potential areas for future research.

3.2 Data description

Water turbidity levels are measured in Nephelometric Turbidity Units (NTU) which calculate the amount of light reflected through a set of suspended particles. Our dataset contains measurements of water turbidity in NTU at five different locations, four of which take measurements at a single depth and are located throughout the channel from the River Liffey towards Dublin Bay where dredging takes place. The fifth buoy takes measurements at three different levels of the water column and is located approximately 10 kilometres away from Dublin port at the location where the sediments are dumped. Figure 3.1 shows the locations of the buoys in the bay.

Turbidity measurements are recorded every 15 minutes by the buoys, but for our analysis we aggregated the raw data into daily averages. This allowed us to focus on the impact of dredging whilst removed short term fluctuations (e.g. that of tides) or the instantaneous impact of vessels arriving or leaving from the port. The aggregation resulted in a total of 488 daily observations per buoy from 31/08/2017 to 31/12/2018. However there are some periods with missing data which seems to be due to equipment failure (e.g. discharged batteries) and gives rise to data gaps when working with our sensor data. A plot of the raw data with missing values is provided in Figure 3.2. Additionally, we use wind speed data measured at Dublin Airport, provided by [91] for the same period as turbidity data, to control for its effect on turbidity.

3.3 Spatio-temporal models

There are two common approaches to modelling the spatiotemporal structure of data (see e.g. [132], for a review). One approach involves building a full covariance matrix for each point in space-time and using a multivariate distribution to account for the data generating process. A second approach is to use a multivariate time series model to account for the evolution of a spatial process. The first approach requires matrix operations to be run on a large covariance matrix, and so the second is a useful simplification and commonly

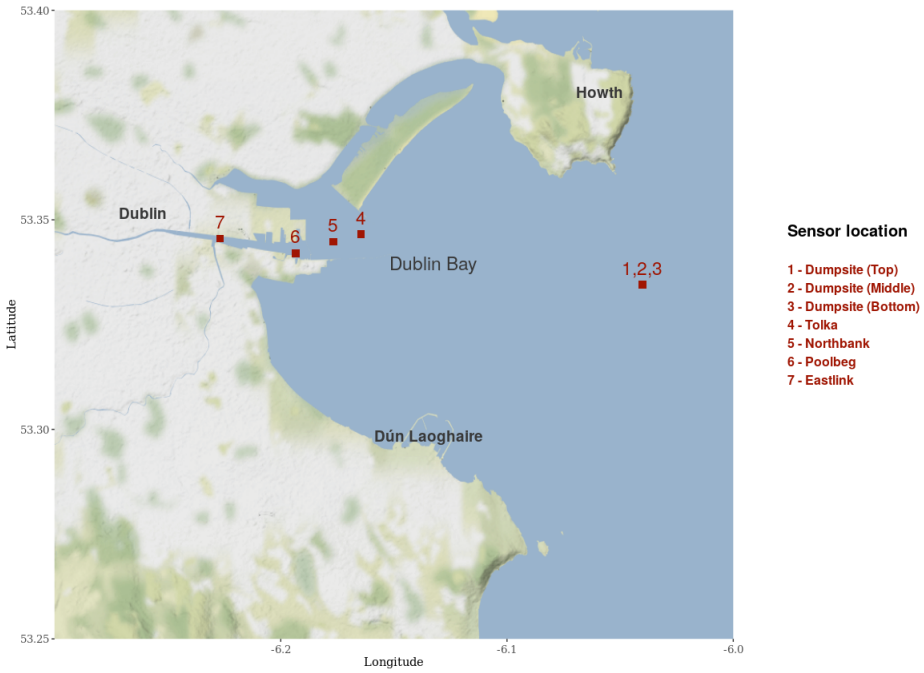

Figure 3.1: Buoys measuring turbidity in Dublin Bay. We use the same numbering scheme when referring to each site throughout the paper. Buoys 4 to 7 are potential dredging sites, whilst the sediment is dumped at the dumpsite.

used in the applied literature [111]. We similarly found the second approach more suitable for our study in terms of computational efficiency and interpretability due to our data being time rich and space poor. Thus we focus on separable space-time models.

As mentioned above, spatio-temporal models are a class that are used to model the evolution of a spatial process. Such processes can be continuous in time, but here we focus on cases where time is discrete and the process is given by $\{Y_t(\mathbf{s}) : \mathbf{s} \in D_s; t = 0, 1, \ldots\}$. The joint distribution is commonly decomposed using a Markov assumption to give an autoregressive likelihood of the form $p\left(Y_t(\mathbf{s})|Y_{t-1}(\mathbf{s}), \ldots, Y_0(\mathbf{s})\right) = p\left(Y_t(\mathbf{s})|Y_{t-1}(\mathbf{s})\right)$. When the model error is assumed additive (and commonly Gaussian) the model can be written as:

$$\mathbf{Y}_t = \mathbf{\Phi}\mathbf{Y}_{t-1} + \epsilon_t \tag{3.1}$$

where $\mathbf{Y}_t$ is a vector of the process values at time t, $\mathbf{\Phi}$ is the autoregressive coefficient matrix, and ϵ_t is a vector of spatially white noise processes. Typically the noise processes are assumed to be independent in time [132] and the autoregressive coefficient matrix is assumed to be stationary. Such models are known as Vector Autoregressive (VAR) models, originally introduced by [115] and widely used in macroeconomics, causal inference, and forecasting [76, 9, 111]. One limitation of VAR models is their inability to model the

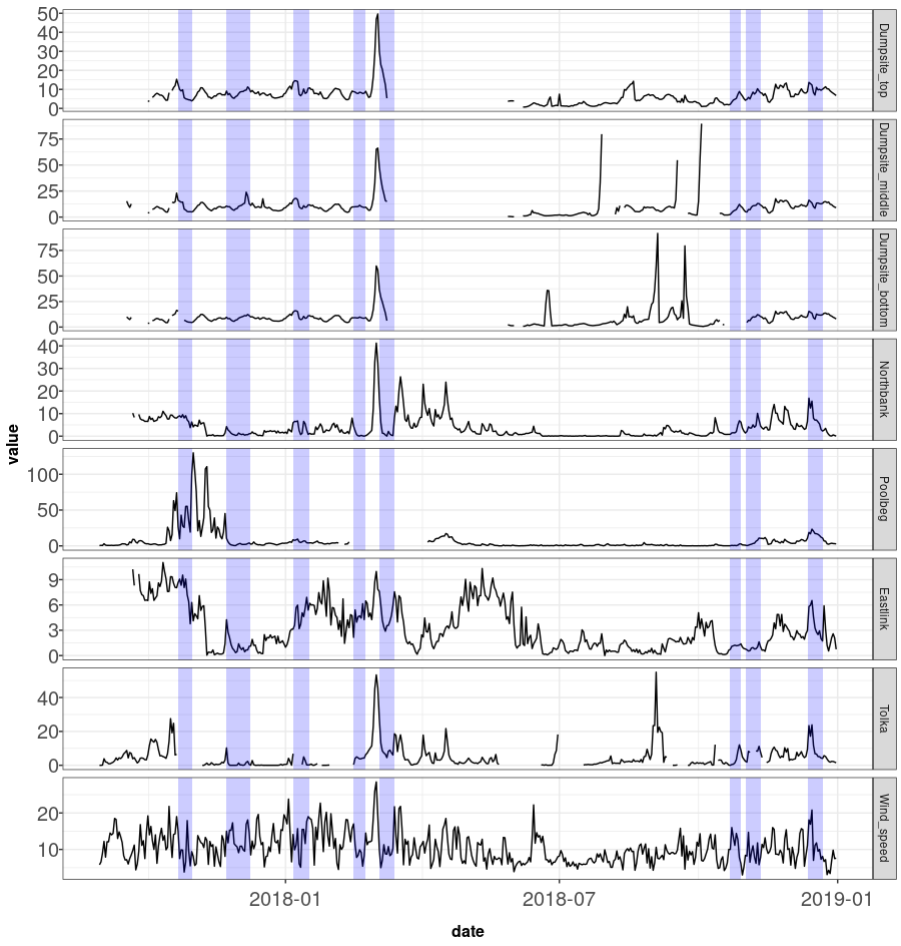

Figure 3.2: Daily measurements of turbidity (NTU) at Tolka, Eastlink, Poolbeg, Northbank, and various depths of the dumpsite, alongside wind speed (knots) measurements, from August 31, 2017, to December 31, 2018. The highlighted regions indicate the periods during which dredging and dumping operations occurred.

heteroscedasticity of the data. To overcome this limitation, it is possible to relax the independence assumption on the noise processes and model their temporal dependence. A very well known approach to model temporal dependence of the noise process is the autoregressive conditional heteroskedasticity (ARCH) model [49] in which the variance of the process is defined as follows:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 \tag{3.2}$$

where ${\sigma_\mathbf{t}}^2$ represent the vector of a diagonal covariance matrix applied to ϵ_t, α_0, and α_i are the parameters of the model, and ϵ_{t-i} are the lagged residuals. The ARCH model is further generalised as the GARCH model [15] which is widely used in finance to model the volatility of financial time series [15]. Furthermore, they have been extended to multivariate time series by considering the covariance matrix of the noise processes and have been used to model non-stationary heteroscedastic data in the spatiotemporal setting (see e.g. [70, 99]). In the next section we explain some variations of these models, including the extended model (VARICH) that we use in our study to model turbidity in Dublin bay and infer the effects of dumping, dredging and wind speed on turbidity levels.

3.4 Modelling procedure

In this section we describe the general modelling framework that we follow to build a spatio-temporal model that describes the response of turbidity to a variety of environmental factors. We then provide specific variations on this template to create four different models which we use for fitting on the data. We denote Y_t as an S-vector of turbidity measurements at time t where S is the number of locations (or equivalently buoys), $s = 1, 2, ..., S$ represent the locations and times t ($t = 1, 2, ..., T$). We write the model hierarchically as:

$$Y_t = M_t + \epsilon_t \tag{3.3}$$

$$\epsilon_t = \Phi \epsilon_{t-1} + \eta_t \tag{3.4}$$

$$\eta_t \sim MVN\left(\mathbf{0}, \Sigma_t\right) \tag{3.5}$$

$$M_t = A + \sum_{j=1}^{P} X_{jt} \circ \beta_j \quad (3.6)$$

where Σ_t is the variance-covariance matrix at time t. A is an intercept vector, X_{jt} is an S-vector of covariate values associated with covariate $j = 1, \ldots, P$, β_j is an S-vector of fixed effects associated with covariate j. We use $\circ$ to denote the Hadamard product.

The four different structures we consider for fitting the model involve specifying structures for the autoregressive coefficient matrix Φ and the covariance matrix Σ_t. We specify prior distributions associated with these models in the section below following their definition.

Model 1 An ARCH structure with varying $\Sigma_t = diag\{\sigma^2_{t,1}, ..., \sigma^2_{t,s}\}$ and:

$$\sigma^2_{t,s} = \theta_{1,s} + \theta_{2,s}\eta^2_{t-1,s} \quad (3.7)$$

with $\Phi = \text{diag}\{\phi_1, ..., \phi_s\}$ being an $S \times S$ diagonal matrix of autocorrelation parameters.

Model 2 A VAR model with a fixed time-invariant covariance matrix given an inverse-Wishart $\mathcal{IW}$ prior:

$$\Sigma \sim \mathcal{IW}^{-1}(\Psi, \nu) \quad (3.8)$$

with ν and Ψ as fixed hyper-parameters (we use $\nu = 14$ and $\Psi = I$ in our example), and where now Φ is a full matrix:

$$\Phi = \begin{bmatrix} \phi_{1,1} & \cdots & \phi_{1,s} \\ \vdots & \ddots & \vdots \\ \phi_{s,1} & \cdots & \phi_{s,s} \end{bmatrix}.$$

Model 3 A VARCH model which has both the full Φ from model 2 and the time-varying error covariance matrix of model 1.

Model 4 A VARICH integrated model that uses differencing to allow for a non-stationary mean process and also uses the time-varying error covariance matrix:

$$Y'_t = Y_t - M_t \quad (3.9)$$

$$\tilde{Y}_t = Y'_t - Y'_{t-1} \quad (3.10)$$

$$\tilde{Y}_t = \Phi \tilde{Y}_{t-1} + \eta_t \tag{3.11}$$

where Y'_t represents the covariate-corrected data, $\tilde{Y}_t$ corresponds to first differences, and everything else is similar to Model 3.

The posterior distribution for model 4 can be written out in full as follows:

$$p\left(A, \beta, \theta, \Phi \mid Y_{1:T}\right) \propto \left(\prod_{t=3}^{T} p\left(Y_t \mid Y_{t-1}; A, \beta, \theta, \Phi\right)\right) \times p(A)p(\beta)p(\theta)p(\Phi) \tag{3.12}$$

To complete the model we need to specify prior distributions for all parameters. We aim to use informative priors for those where we have some degree of information, and use weakly informative and non-informative priors for the remainder. In the below we outline our prior specification for the most complex of the models we fit, model 4, though identical priors were used in the simpler models which corresponds to setting some of the parameter values to zero in a nested model structure.

Our covariates contained in $X_{j,t}$ consist of values associated with dumping and dredging (binary yes/no knowing that the operations happened in the same day, dreding at the dredging sites and dumping at the dumpsite), and wind speed (knots). It is helpful, for prior specification, to consider the regression parameters β in terms of their individual scalar components $[\beta_{dredge/dump,s}, \beta_{wind,s}]$ at site s. The full set of priors we used for these values is:

$$\beta_{dredge/dump,s} \sim N\left(0, 5^2\right)$$

$$\beta_{wind,s} \sim N\left(0, 5^2\right)$$

For the Φ matrix we focus most of the prior mass in the range (-1,1), including the stationary region that is concentrated in this Euclidean space, though non-stationarity can be found if the data are indicative of such phenomena. We thus use:

$$\phi_{ij} \sim N\left(0, 0.5^2\right)$$

For the remaining parameters we set:

$$\begin{aligned} A_s &\sim N(0, 5^2) \\ \theta_{1,s} &\sim TN_0(0, 1) \\ \theta_{2,s} &\sim Beta(1, 5) \end{aligned}$$

where TN_a refers to the truncated normal distribution with minimum value a. All these are expected to be weakly informative, guiding the model towards sensible values whilst letting the data provide the majority of the information . Turbidity in our dataset ranges between 0-130 (NTU), and we use a log transformation to mitigate the skewness of the data, so the prior values chosen for A, β_{dredge}, β_{dump}, are considered to be uninformative with respect to this range. The same is true for β_{wind} knowing that wind speed can reach as high as 70 knots during storms.

As a final remark on priors we note that many of the turbidity values across sites are missing. We assume that these values are missing at random (MAR; [84]) and impute them as part of the model fitting step by treating them as parameters to be estimated. When using the likelihood given above we found that we struggled to produce a posterior with finite variance so we added the extra prior constraint $y_{missing} \sim TN^5_{-5}(0, 3)$, a truncated normal distribution between -5 and 5, which seemed to stabilise the missing value estimates.

In summary, model 1 provides a baseline univariate autoregressive model with time changing variance. A more basic constant variance model was also attempted but not shown here due to poor performance. Model 2 tests whether a richer full rank vector mean structure improves the fit at the expense of the changing variance. Model 3 combines both the full vector autoregression with the time changing variance. Finally, model 4 introduces a difference in the latent parameters to capture any potential non-stationarity in the mean. This is because covariate effects may persist over time and contribute to non-stationarity in the process. Below we fit each of these models to the data described in Section 3.2, and use a combination of posterior predictive distributions, information criteria, and plots of the posterior distributions of the parameters to determine the optimal models which we use for interpreting our findings.

3.5 Results

In this section, we report the results of fitting the models described in Section 3.4 to the turbidity data described in the previous section. We summarise the estimated effects of dredging and dumping operations and account for the wind speed effect by including the daily wind speed measured in knots. We compare the different models according to their fit to the data, and interpret the best fitting model with a view to obtaining a better understanding of turbidity behaviour in Dublin bay.

3.5.1 Model fitting and comparison

We fit the models using R [105] and the Stan modelling framework [117]. This approach uses Hamiltonian Monte Carlo to update all parameters simultaneously and aims

to rapidly converge to the posterior distribution. Through repeated fitting of the models we found that using 1000 iterations, with a warm-up period of 200 iterations, produced acceptable results. We used a MacBook Air equipped with an M1 chip, an 8-core CPU, 8GB of RAM, and 256GB of SSD storage; the computation time was 12 minutes. The $\hat{R}$ values were all close to the target value of 1, and the minimum ESS across all the model parameters was 111, indicating sufficient mixing and convergence of the MCMC chains. Details of the estimated parameters are provided in the Appendix A.

To compare between the models we use the Widely Applicable Information Criterion (WAIC, [130]) and the Leave-One-Out Information Criterion (LOOIC, [129]) which penalise the likelihood of the model fit based on the complexity of the model. These two information criteria have the added advantage of being easily implemented in R and providing an uncertainty estimate on the value itself. In addition, we also compared VARICH to a frequentist VAR model, details of which are provided in Appendix B.

Figure 3.3 shows the estimated WAIC and LOOIC values for the four models. The VARICH and VARCH models have the lowest WAIC and LOOIC values indicating better fits. However whilst the mean value of LOOIC for the VARICH model is slightly lower there is no clear difference between them. The VARICH model has no extra complexity compared to VARCH, i.e. there are no extra parameters to estimate. Furthermore we computed the posterior distribution of the spectral radius of Φ for both models, and calculated the probability that these were greater than 1 to provide an estimate of the proportion of samples that were outside the stationary region. VARICH gave 0 compared to 0.035 for VARCH, which indicates that the VARICH model seems to have removed some of the non-stationarity present in the VARCH formulation. We thus use the VARICH model to create our further results.

To further assess the goodness of fit for the VARICH model, we have plotted the model's posterior predictions against the actual data values, as depicted in Figure 3.4. The expected values of the fit and the observed values are shown with solid lines coloured in red and blue respectively and the 95% credible intervals are shown with grey bands. As mentioned in Section 3.2, the dataset has missing periods which are imputed for each location by the model during the fitting process. The vector autoregressive part of the VARICH model allows for drawing information for each site using the available information from the other sites which specifically helps regulate the uncertainty for the missing periods. As expected, the uncertainty during high volatility periods grows as expected through the dynamic structure applied to the variance.

Figure 3.5 shows the posterior predictive distributions from the VARICH model against the true values with vertical lines indicating the 80% uncertainty intervals. On average the intervals cover 79% of the data. The figure shows that the model can successfully retrieve the true values of the turbidity in the dataset with well-calibrated uncertainty

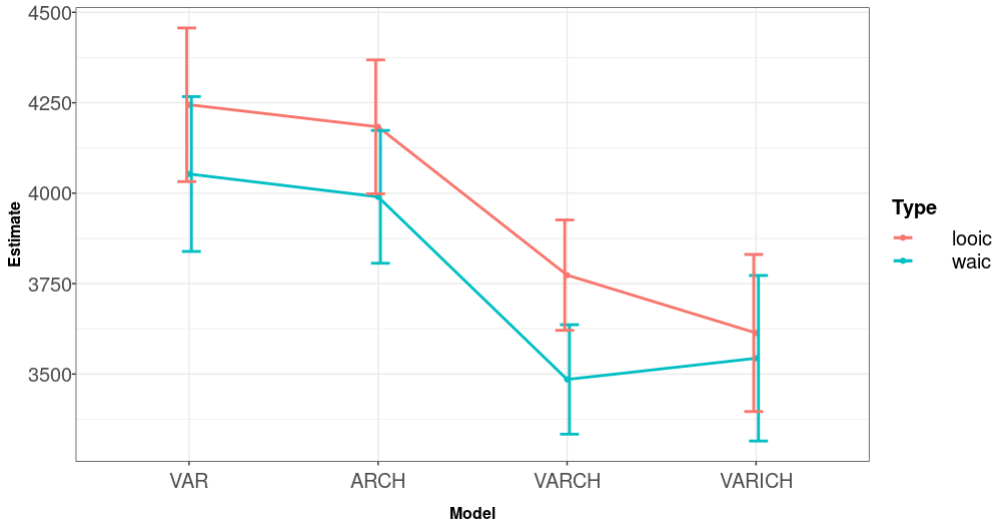

Figure 3.3: WAIC and LOOIC values for the four fitted models with their associated standard errors.

estimation at the dumpsite and the dredging sites respectively.

3.5.2 Effects of covariates on turbidity

To determine the degree to which dumping and dredging operations affect turbidity, we evaluate the posterior distribution of the fixed effects β. Figure 3.6 shows the expected value of the dumping and dredging effects respectively with their 95% credible intervals for different locations. Most effects are observed to have an expected value close to zero with less than %95 probability of deviating from zero.

Figure 3.7 shows the wind speed effects for the 7 buoys. These are measured in log(NTU) per knot and these wind effects can be more clearly identified than the effects of dredging and dumping. The values are reasonably consistent but with greater uncertainty at the lower positions in the dumping buoy, and a far smaller effect at Eastlink, again likely due to its position in the bay. By contrast, the Tolka buoy seems most influenced by wind and is the site that is most far out to sea. The Tolka buoy is situated within the confines of the estuary walls, adjacent to North Bull wall. This area of the estuary is relatively shallow and at low tide is exposed to the wind. The positive effect of wind can be explained by its ability to generate waves and surface currents in water bodies. When these waves and currents are strong enough, they can stir up sediments lying on the bottom, resuspending them into the water column. This process increases the concentration of suspended particles, ultimately leading to higher turbidity.

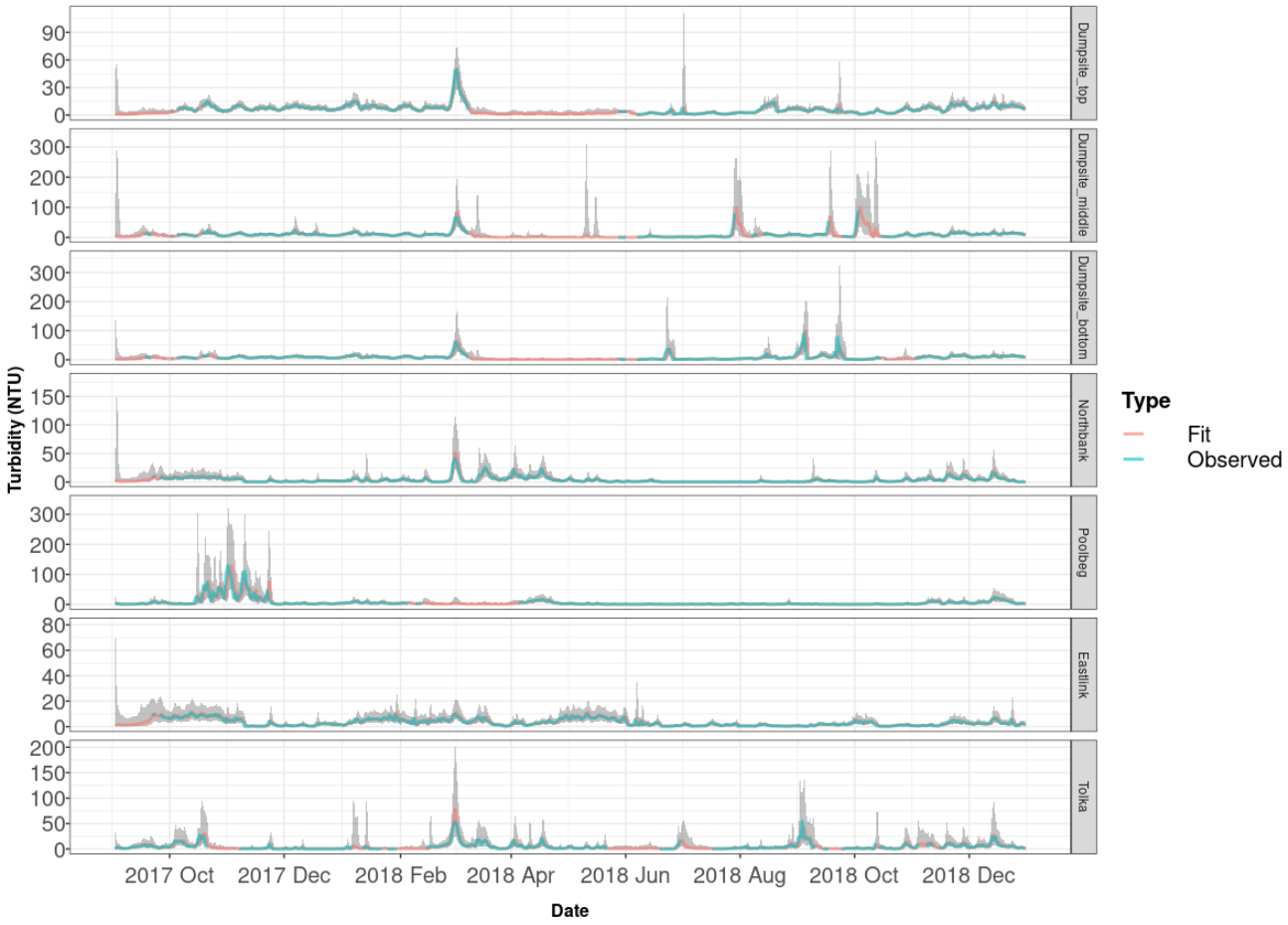

Figure 3.4: Posterior prediction from the VARICH model vs observed values of turbidity over time for the 7 buoys as labelled. Note the differing vertical axis heights. The shaded periods indicate 95% credible intervals.

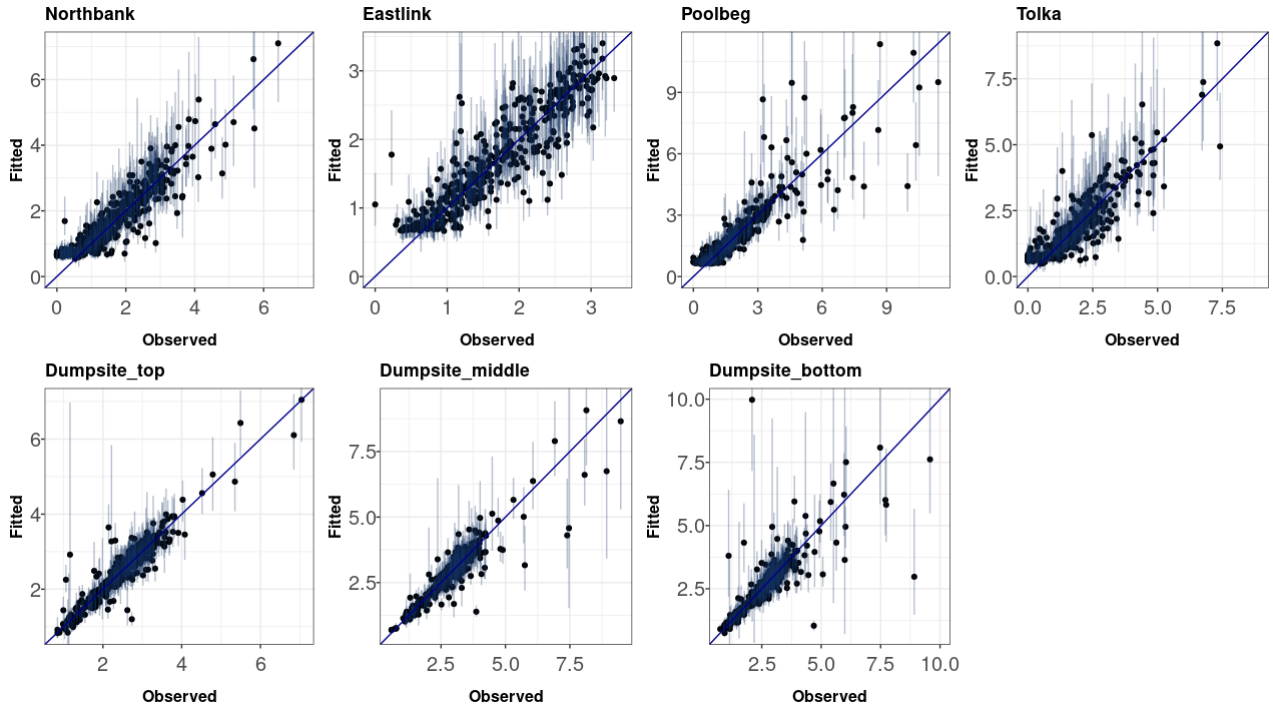

Figure 3.5: Fitted values from the VARICH model versus observed values of turbidity (log scale) at different sites. The vertical bars indicate the 80% uncertainty intervals which provide evidence of the coverage properties of the model.

3.5.3 Influence of the autoregressive component

As a final part of the analysis, we examine the autoregressive coefficients from the VARICH model. Figure 3.8 shows the posterior coefficients of Φ where we have separated out the diagonal values which indicate the influence of the time series on itself from the off-diagonal elements which show the influence of one site on another. The numbering of the sites is as shown in Figure 3.1.

Of the diagonal elements, the dumpsite (middle) seems to have the most dependence after accounting for the integration component. The other sites have values close to zero after accounting for uncertainty. Of the off-diagonal elements, some of these are well away from zero and provide for interesting, if not entirely straightforward, interpretation. Φ_{42} is the largest, corresponding to the relationship between buoy 4 (Tolka) and dumpsite (middle), which should perhaps be read in conjunction with their joint time series behaviour as shown in Figure 3.4. Many of the other off-diagonal elements show similar clear non-zero effect sizes though they are considerably smaller than Φ_{42}. These values provide evidence of cross site learning in the time series model.

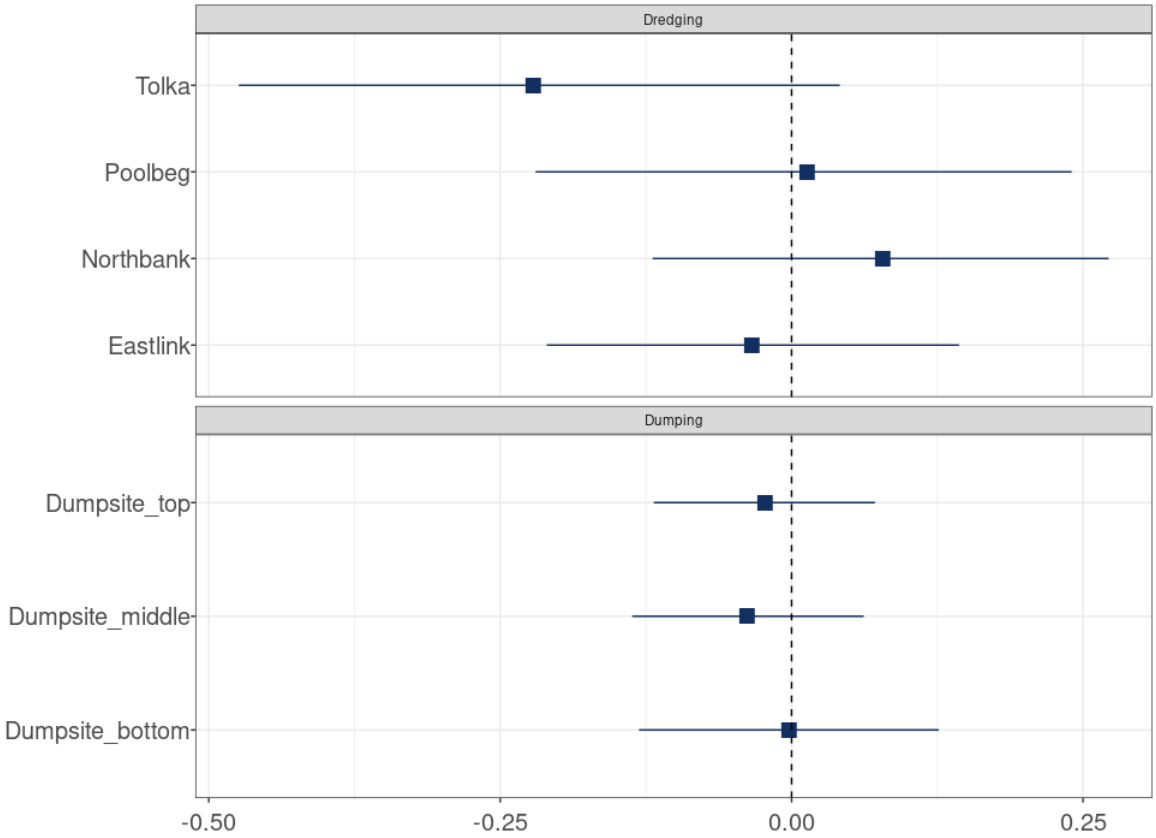

Figure 3.6: Dumping and dredging effects (log(NTU)/day) at different locations with the 95% credible interval.

3.6 Conclusions

We have introduced a set of models for understanding the behaviour of turbidity in Dublin bay. Both the VARCH and the VARICH models introduced in Section 3.4 allow for measuring the effects of multivariate time series on each other, whilst taking account of the known volatility changes in the time series. However, the VARICH model had slightly better performance. The combination of Bayesian modelling, VAR and ARCH structures makes the VARICH model a useful tool for flexible modelling of a wide range of real world random processes in which spatial and temporal aspects are playing major roles. Furthermore, the Bayesian approach allows for uncertainty quantification of both the fixed effects and the posterior predictions of the time series, whilst simultaneously imputing the missing values within the series.

Our main finding has been that the dumping and dredging operations have minimal effect on the turbidity levels, which seem to be more affected by wind speed and previous values of the series. We thus suggest that, at an aggregate daily level, there is minimal

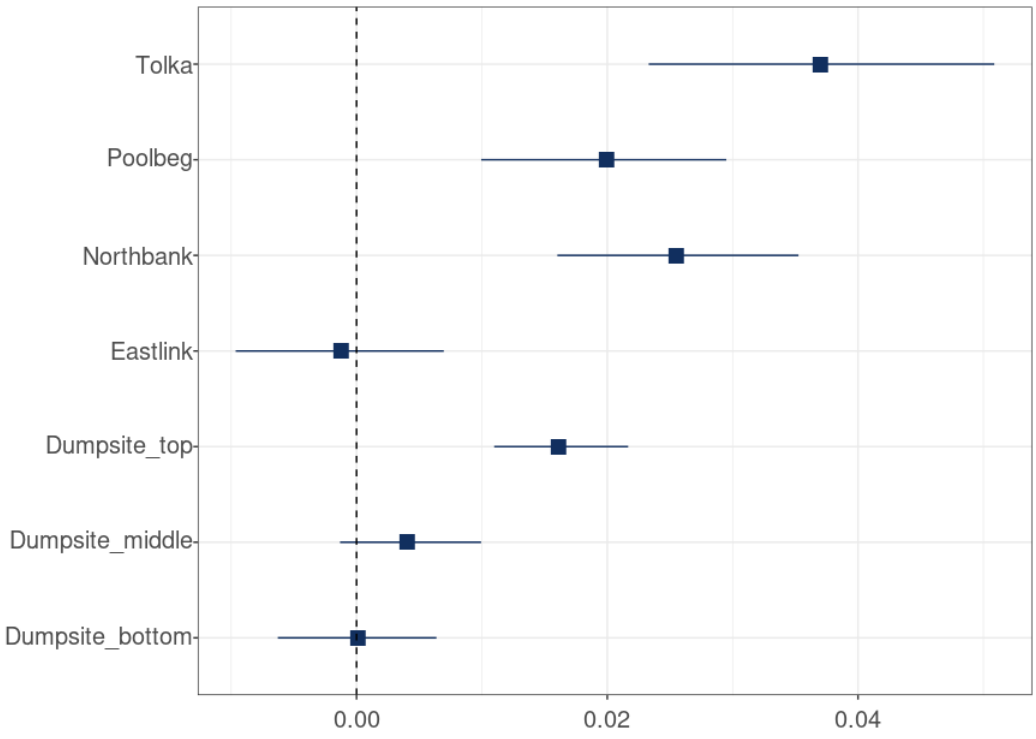

Figure 3.7: Effect of wind speed (log(NTU)/Knot) at different locations and depths with the 95% credible interval.

effect of the operations on the turbidity levels in Dublin bay. The models we produced seem to fit the data well and the results make physical sense according to the location of the buoys in the bay. A longer time series and a more complete record would add further weight to our conclusions.

Our model fitting technique of using HMC appeared to converge efficiently and quickly on a standard laptop, taking around 12 minutes to reach R-hat values below the common standard of 1.1 whilst requiring only 3 chains of 1000 posterior draws (with 200 removed during the warm-up phase). However, for larger data sets it may be that users need to increase the number of draws. For very large data sets the HMC technique may prove infeasible and so other methods such as MultiBUGS [62] might be more appropriate. Other computational difficulties may occur should the model structure be made more complex. Interesting extensions of our approach might involve looking at time-varying behaviour of the coefficients, or multiple lags or long memory of the multivariate time series itself.

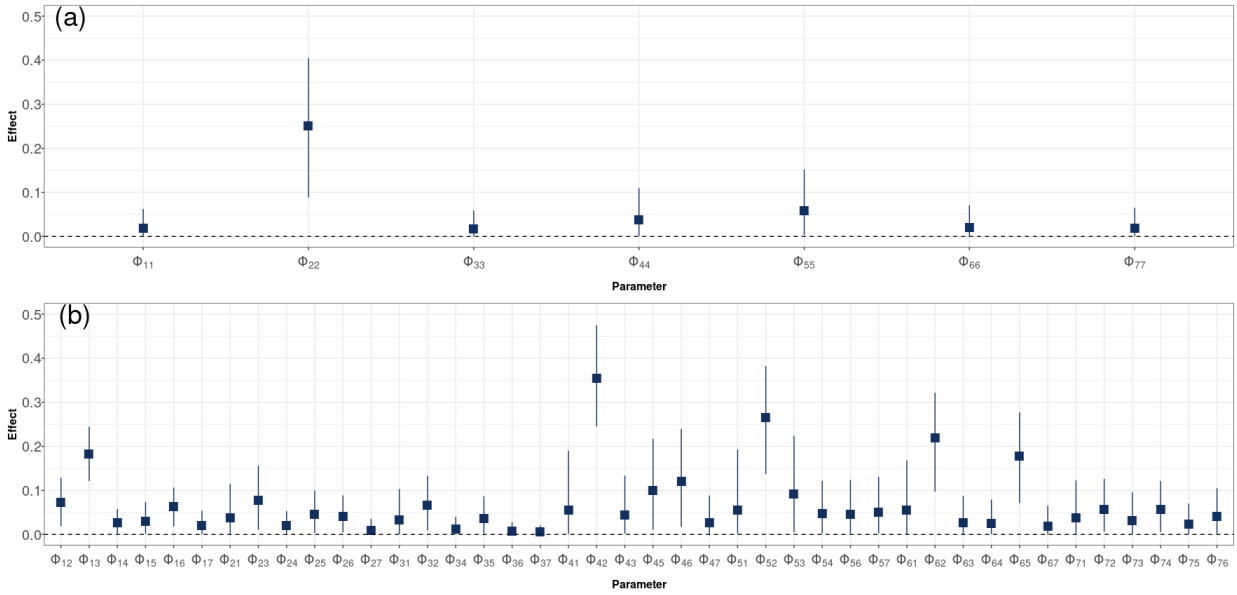

Figure 3.8: Coefficients of the Φ matrix with their 95% credible interval. Diagonal values are shown in the top panel (a) and off-diagonal values are shown in (b). The two subscripts indicate the parent and child relationship respectively, so that Φ_{12} for example is the degree to which buoy 2 influences the time series of buoy 1. The numbers of the buoys follow the labelling defined in Figure 3.1.

Appendix A

Table 3.1: Summary statistics of estimated parameters, including their mean, median, standard deviation, 5th percentile, 95th percentile, the R-hat statistic, and ESS.

variable	mean	median	std	q5	q95	$\hat{R}$	ESS
a_1	-0.15	-0.15	4.23	-6.99	7.08	1.01	912
a_2	1.46	2.46	5.16	-8.05	8.78	1.01	270
a_3	0.11	0.11	3.73	-6.05	6.30	1.00	1074
a_4	-0.38	-0.48	3.44	-6.17	5.62	1.00	3104
a_5	0.45	0.65	4.79	-7.76	8.13	1.00	1374
a_6	-0.07	-0.15	3.53	-5.98	6.09	1.00	3009
a_7	0.14	0.06	4.06	-6.57	7.04	1.00	1744
$\theta_{1,1}$	0.03	0.03	0.00	0.03	0.04	1.00	1889
$\theta_{1,2}$	0.05	0.05	0.01	0.04	0.05	1.01	1396
$\theta_{1,3}$	0.06	0.06	0.01	0.05	0.07	1.00	2324

Table 3.1: Summary statistics of estimated parameters, including their mean, median, standard deviation, 5th percentile, 95th percentile, the R-hat statistic, and ESS.

variable	mean	median	std	q5	q95	$\hat{R}$	ESS
$\theta_{1,4}$	0.22	0.22	0.02	0.18	0.26	1.00	1882
$\theta_{1,5}$	0.14	0.14	0.02	0.12	0.17	1.00	1903
$\theta_{1,6}$	0.15	0.15	0.02	0.13	0.18	1.00	2089
$\theta_{1,7}$	0.11	0.11	0.01	0.10	0.13	1.00	2829
$\theta_{2,1}$	0.53	0.53	0.12	0.34	0.72	1.00	2116
$\theta_{2,2}$	0.66	0.67	0.09	0.51	0.81	1.00	1416
$\theta_{2,3}$	0.81	0.82	0.07	0.68	0.91	1.00	2456
$\theta_{2,4}$	0.41	0.41	0.09	0.26	0.57	1.00	1724
$\theta_{2,5}$	0.29	0.29	0.09	0.14	0.45	1.00	1887
$\theta_{2,6}$	0.36	0.36	0.09	0.21	0.53	1.00	2634
$\theta_{2,7}$	0.28	0.27	0.07	0.17	0.39	1.00	3362
$\phi_{1,1}$	0.02	0.01	0.02	0.00	0.05	1.00	2366
$\phi_{2,1}$	0.04	0.03	0.03	0.00	0.10	1.00	2258
$\phi_{3,1}$	0.03	0.02	0.03	0.00	0.09	1.00	1576
$\phi_{4,1}$	0.06	0.04	0.05	0.00	0.16	1.00	1807
$\phi_{5,1}$	0.06	0.04	0.05	0.00	0.15	1.00	1542
$\phi_{6,1}$	0.05	0.04	0.04	0.00	0.14	1.00	1973
$\phi_{7,1}$	0.04	0.03	0.03	0.00	0.10	1.00	1921
$\phi_{1,2}$	0.07	0.07	0.03	0.03	0.12	1.00	2178
$\phi_{2,2}$	0.25	0.25	0.08	0.11	0.38	1.03	166
$\phi_{3,2}$	0.07	0.07	0.03	0.01	0.12	1.03	111
$\phi_{4,2}$	0.35	0.35	0.06	0.26	0.45	1.00	1692
$\phi_{5,2}$	0.27	0.27	0.06	0.16	0.36	1.01	679
$\phi_{6,2}$	0.22	0.22	0.06	0.12	0.30	1.01	1003
$\phi_{7,2}$	0.06	0.05	0.03	0.01	0.11	1.00	1937
$\phi_{1,3}$	0.18	0.18	0.03	0.13	0.23	1.00	2533
$\phi_{2,3}$	0.08	0.08	0.04	0.02	0.14	1.01	2027
$\phi_{3,3}$	0.02	0.01	0.02	0.00	0.05	1.00	2029
$\phi_{4,3}$	0.04	0.04	0.04	0.00	0.11	1.00	1718
$\phi_{5,3}$	0.09	0.08	0.06	0.01	0.20	1.01	476
$\phi_{6,3}$	0.03	0.02	0.02	0.00	0.07	1.00	2042
$\phi_{7,3}$	0.03	0.02	0.02	0.00	0.08	1.00	2203
$\phi_{1,4}$	0.03	0.02	0.01	0.00	0.05	1.00	1554

Table 3.1: Summary statistics of estimated parameters, including their mean, median, standard deviation, 5th percentile, 95th percentile, the R-hat statistic, and ESS.

variable	mean	median	std	q5	q95	$\hat{R}$	ESS
$\phi_{2,4}$	0.02	0.02	0.01	0.00	0.05	1.00	1332
$\phi_{3,4}$	0.01	0.01	0.01	0.00	0.03	1.00	1794
$\phi_{4,4}$	0.04	0.03	0.03	0.00	0.09	1.00	1305
$\phi_{5,4}$	0.05	0.04	0.03	0.00	0.11	1.02	284
$\phi_{6,4}$	0.03	0.02	0.02	0.00	0.07	1.00	2077
$\phi_{7,4}$	0.06	0.06	0.03	0.01	0.11	1.01	1025
$\phi_{1,5}$	0.03	0.03	0.02	0.00	0.07	1.00	1755
$\phi_{2,5}$	0.05	0.04	0.03	0.01	0.09	1.00	2014
$\phi_{3,5}$	0.04	0.03	0.02	0.00	0.08	1.00	2049
$\phi_{4,5}$	0.10	0.10	0.05	0.02	0.19	1.00	2076
$\phi_{5,5}$	0.06	0.05	0.04	0.01	0.14	1.00	1495
$\phi_{6,5}$	0.18	0.18	0.05	0.09	0.26	1.00	2257
$\phi_{7,5}$	0.02	0.02	0.02	0.00	0.06	1.00	1749
$\phi_{1,6}$	0.06	0.06	0.02	0.03	0.10	1.00	1888
$\phi_{2,6}$	0.04	0.04	0.02	0.01	0.08	1.00	1830
$\phi_{3,6}$	0.01	0.01	0.01	0.00	0.02	1.00	1940
$\phi_{4,6}$	0.12	0.12	0.06	0.03	0.21	1.00	1493
$\phi_{5,6}$	0.05	0.04	0.03	0.00	0.11	1.00	2063
$\phi_{6,6}$	0.02	0.02	0.02	0.00	0.06	1.00	2501
$\phi_{7,6}$	0.04	0.04	0.03	0.00	0.09	1.00	1774
$\phi_{1,7}$	0.02	0.02	0.01	0.00	0.05	1.00	1802
$\phi_{2,7}$	0.01	0.01	0.01	0.00	0.03	1.00	1829
$\phi_{3,7}$	0.01	0.00	0.01	0.00	0.02	1.00	1780
$\phi_{4,7}$	0.03	0.02	0.02	0.00	0.07	1.00	2184
$\phi_{5,7}$	0.05	0.04	0.04	0.00	0.12	1.00	1681
$\phi_{6,7}$	0.02	0.01	0.02	0.00	0.05	1.00	1921
$\phi_{7,7}$	0.02	0.01	0.02	0.00	0.05	1.00	1753
$\beta_{wind,1}$	0.02	0.02	0.00	0.01	0.02	1.00	2471
$\beta_{wind,2}$	0.00	0.00	0.00	-0.00	0.01	1.00	2439
$\beta_{wind,3}$	0.00	0.00	0.00	-0.01	0.01	1.00	2293
$\beta_{wind,4}$	0.04	0.04	0.01	0.03	0.05	1.00	2316
$\beta_{wind,5}$	0.03	0.03	0.00	0.02	0.03	1.00	2068
$\beta_{wind,6}$	0.02	0.02	0.00	0.01	0.03	1.00	2495

Table 3.1: Summary statistics of estimated parameters, including their mean, median, standard deviation, 5th percentile, 95th percentile, the R-hat statistic, and ESS.

variable	mean	median	std	q5	q95	$\hat{R}$	ESS
$\beta_{wind,7}$	-0.00	-0.00	0.00	-0.01	0.01	1.00	2638
$\beta_{dredge/dump,1}$	-0.02	-0.02	0.05	-0.10	0.06	1.00	3097
$\beta_{dredge/dump,2}$	-0.04	-0.04	0.05	-0.12	0.04	1.01	2898
$\beta_{dredge/dump,3}$	-0.00	-0.00	0.06	-0.11	0.10	1.00	4022
$\beta_{dredge/dump,4}$	-0.22	-0.22	0.13	-0.43	-0.01	1.00	2909
$\beta_{dredge/dump,5}$	0.08	0.08	0.10	-0.09	0.24	1.00	2594
$\beta_{dredge/dump,6}$	0.01	0.01	0.12	-0.18	0.21	1.00	3365
$\beta_{dredge/dump,7}$	-0.03	-0.03	0.09	-0.18	0.11	1.00	2474

Appendix B

To fit a frequentist VAR (Vector Autoregression) model to our data, we utilized the 'vars' package in R. This package does not accommodate missing values in the data, and to the best of our knowledge, there is no package that can naturally handle missing data without resorting to imputation techniques. Therefore, we first addressed this issue by imputing the missing values using the forward-fill method, whereby we carried the last observed values forward to replace any missingness. After fitting the model, we calculated the root mean squared error (RMSE) for one-step-ahead predictions. We then compared these results with those obtained from the VARICH model, as detailed in Table 3.2. Please note that calculating the WAIC and LOOIC criteria to compare the two models, as was done in Figure 3.3, is not possible because these calculations require evaluating the log-likelihood over the predictive posterior distribution, which is unavailable in the frequentist paradigm. As presented in Table 3.2, the VARICH model's RMSE values are lower for all sites except Poolbeg, where both models performed the worst, possibly due to the highest volatility among the sites.

Table 3.2: Comparison of RMSE values: VARICH (Bayesian) vs VAR (Frequentist).

Variable	**VARICH**	**VAR**
Buoy 1 Top	1.89	2.08
Buoy 1 Middle	5.14	8.31
Buoy 1 Bottom	5.69	6.65
Tolka	4.26	4.30
Northbank	2.40	2.45
Poolbeg	8.94	8.36
Eastlink	1.19	1.22

4

SERT: A Transfomer Based Model for Spatio-Temporal Sensor Data with Missing Values for Environmental Monitoring

Environmental monitoring is crucial to our understanding of climate change, biodiversity loss and pollution. The availability of large-scale spatio-temporal data from sources such as sensors and satellites allows us to develop sophisticated models for forecasting and understanding key drivers. However, the data collected from sensors often contain missing values due to faulty equipment or maintenance issues. The missing values rarely occur simultaneously leading to data that are multivariate misaligned sparse time series. We propose two models that are capable of performing multivariate spatio-temporal forecasting while handling missing data naturally without the need for imputation. The first model is a transformer-based model, which we name ***SERT*** *(**S**patio-temporal **E**ncoder **R**epresentations from **T**ransformers). The second is a simpler model named* ***SST-ANN*** *(**S**parse **S**patio-**T**emporal **A**rtificial **N**eural **N**etwork) which is capable of providing interpretable results. We conduct extensive experiments on two different datasets for multivariate spatio-temporal forecasting and show that our models have competitive or superior performance to those at the state-of-the-art.*

4.1 Introduction

The importance of spatio-temporal forecasting has increased significantly in recent years due to the availability of large-scale spatio-temporal data from various sources such as sensors and satellites [66]. Spatio-temporal forecasting involves predicting how data vary over space and time, which is critical for a wide range of applications such as water quality forecasting [37]. A common approach to modelling spatio-temporal data is to use a multivariate time series structure, where each time series is associated with a variable at a specific location [132].

Spatio-temporal data often contain missing values which is a common problem in environmental monitoring, and can be caused by sensor failure, malfunction or communication problems (for example see Figure 4.1). A common remedy for forecasting with missing data is to impute the missing values using a variety of methods such as mean substitution, interpolation, or advanced techniques like multiple imputation [127, 118]. However, these methods are not always effective as they may introduce biases or fail to accurately capture the underlying patterns in the data. For instance, mean substitution can oversimplify the data's complexity, leading to a loss of variability and potentially masking important signals. Advanced techniques like multiple imputation, while more sophisticated, might not always be suitable for all types of data or missingness patterns. They may lead to models that are overfitted to the imputed values.

We propose a new model that is capable of performing multivariate spatio-temporal forecasting named SERT (**S**patio-temporal **E**ncoder **R**epresentations from **T**ransformers). Our model is an extension of the well-known transformer architecture that has shown remarkable success in natural language processing and also image analysis [38, 45]. SERT is designed to capture the complex joint temporal and spatial dependencies among the input variables. An important feature of our proposed model which differentiates it from the many other available methods is its ability to handle missing data more naturally without requiring any missing value imputation.

In addition to the SERT model, we introduce an interpretable simplified version that provides insights into the underlying factors that drive the predicted values, which can assist in decision and policy-making. Our proposed simplified model is named SST-ANN (**S**parse **S**patio-**T**emporal **A**rtificial **N**eural **N**etwork) and removes the transformer layers from the SERT structure. Despite being less accurate than SERT, it is capable of providing insightful results with faster computation time while similarly being able to handle missing values. Depending on the complexity of the problem, the required accuracy and the available computational resources, the user can fit both SERT and SST-ANN, using the former to provide more accurate forecasts and the latter to forecast and gain insights about how the results were obtained.

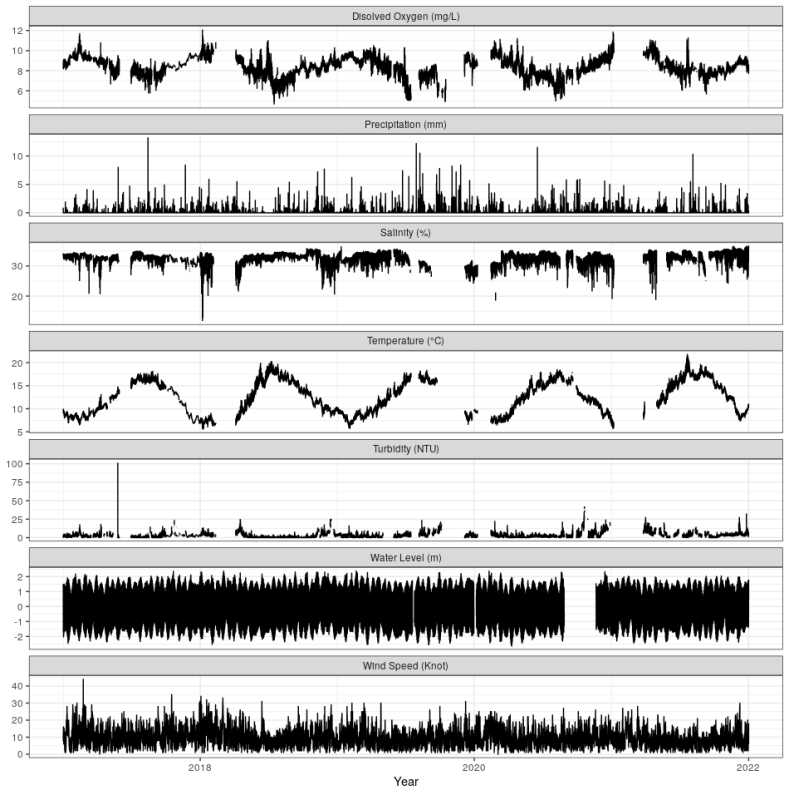

Figure 4.1: An example of our environmental monitoring dataset; a multivariate environmental time-series with missing values.

To evaluate the performance of our proposed models, we conducted extensive experiments on two different datasets for multivariate spatio-temporal forecasting. We fitted our models to a simulated dataset to assess their ability to function under different levels of sparsity. We also evaluated the performance of the models on a real-world dataset, including missing values, of environmental variables in Dublin Bay for 7 hour ahead forecasting. Our experimental results show that our models are competitive with state-of-the-art models for multivariate spatio-temporal forecasting.

Our paper is organized as follows. In Section 4.2, we provide a brief overview of related work on models developed for analysing sequential data in general and spatio-temporal forecasting applied to environmental monitoring in particular. In Section 4.3, we describe the proposed SERT and SST-ANN models in detail. In Section 4.4, we present the experimental results and analysis. Finally, in Section 4.5, we conclude the paper and discuss future directions of research.

4.2 Related Work

In this section we provide a brief overview of the recent developments in deep learning models for sequential data analysis and spatio-temporal models for environmental monitoring, and also methods for handling missing data and adding interpretability to deep learning models applied to time series data.

4.2.1 Deep Learning Models for Sequential Data

Recurrent Neural Networks (RNNs) have been one of the most popular deep learning models for sequential data [82]. However, RNNs suffer from the vanishing gradient problem which makes them unable to learn long-term dependencies in the data. To address this problem, Long Short-Term Memory (LSTM) networks [68] and Gated Recurrent Unit (GRU) networks [29] were introduced. These models have been applied to various tasks such as machine translation [11], speech recognition [63], and time series forecasting [83].

More recently a new type of deep learning model named transformers was introduced [128]. Transformers are based on the self-attention mechanism which enables them to learn the dependencies between the input and output sequences. Transformers are comprised of an encoder and a decoder network. The encoder network is responsible for learning the representation of the input sequence, while the decoder network is responsible for generating the output sequence based on the learned representation. Models developed on the transformer architecture include BERT [38], which uses only the encoder part, and GPT [106] which uses only the decoder part. These models have been applied to various tasks in natural language processing such as question answering [38], text classification [121], and text summarization [85]. Overall, transformers have proven themselves to be more effective than recurrent based models in many applications, especially in natural language processing.

4.2.2 Deep Learning Models for Spatio-Temporal Data

The application of deep learning models for spatio-temporal forecasting is not new. For example, [142] used an LSTM to forecast daily land surface temperature, [137] developed a deep learning model that utilizes convolutional long short-term memory as the building block for predicting sea surface temperature fields, and [90] introduced MCxM, a deep learning approach specifically designed for emergency pollution forecasting. However, the application of transformers to spatio-temporal forecasting is relatively new and challenging because transformers have been mainly developed in the field of natural language processing (NLP). Nonetheless researchers were inspired by the success of transformers in NLP and started adapting them to spatio-temporal forecasting which can be formulated

as a sequence-to-sequence problem, where the input is a sequence of historical observations of multiple variables at different locations, and the output is a sequence of future predictions of the same variables at the same locations. A common approach for sequence-to-sequence modeling is to use an encoder-decoder architecture, where an encoder network maps the input sequence into a latent representation, and a decoder network generates the output sequence from the latent representation [122]. [64] used this idea to develop a new model called Spacetimeformer and applied it to traffic prediction and weather forecasting. However, to the best of our knowledge, the application of transformers to environmental monitoring is limited to the recent work by [140] who used a transformer-based model for hourly $PM_{2.5}$ forecasting in Los Angeles.

4.2.3 Addressing Missing Values in Modelling

As mentioned in the introduction, a major challenge in spatio-temporal forecasting in the environmental monitoring context is dealing with missing values. A common approach for dealing with the missing values is imputation [127] before conducting any analysis. In time series modelling care needs to be taken to avoid introducing bias; last observation carried forward is a common approach. An alternative used in the literature is that of a Bayesian framework which enables defining a prior distribution over the missing values so that they can be inferred with the other unobserved parameters when fitting the models. For example, in Chapter3, we proposed a novel approach (called VARICH) for spatio-temporal modelling of turbidity data with many missing values. However, this approach is computationally expensive and requires a large number of samples from the posterior distribution to obtain acceptable results and thus is not suitable for large spatio-temporal datasets.

To address the missing data problem in time series, [71] proposed a novel approach to encode multivariate time series using set functions and introduced a new model called SeFT for classifying time series with irregularly sampled clinical data. More recently, [125] proposed a new transformer based model called STraTS that represents each observation as a triplet of the form (time, variable name, value). As opposed to SeFT, STraTS uses a learnable positional encoding and a Continuous Value Embedding (CVE) scheme that is a one-to-many feed-forward network. STraTS was developed to perform multivariate time series forecasting during the pre-training phase and classification as the final task on irregularly sampled clinical data. It has been shown to have higher accuracy than its predecessor, the SeFT model, when applied to the classification of clinical time series. Both SeFT and STraTS can handle missing values without requiring any imputation. To the best of our knowledge, none of these novel methods have been applied to environmental monitoring challenges; we adapt STraTS to a spatio-temporal setting.

4.2.4 Interpretability of Deep Learning Models

Deep learning models are often considered as black-box models because they are hard to interpret [23]. However, in many applications, it is important to understand the model's decision making process [46]. Some authors have proposed methods to help interpret the results of deep learning models. For example, SeFT uses the attention mechanism in its architecture and the authors of the work showed that the attention weights can be used to gain insights into the importance of input data, including multiple variables. Similarly, [28] used a transformer-based model to forecast floods and employed attention maps extracted from the attention layers of the transformer model for interpretation. [125], inspired by [30] and [141], took a different approach and proposed an interpretable version of the STraTS model, called STraTS-I, which uses an almost identical structure to their STraTS model but instead uses encoded inputs directly to the output layer as opposed to STraTS that uses the contextualized inputs to the output layer. Their approach allows for the calculation of a contribution score for each input observation towards the prediction, achieved by multiplying the encoded input, attention weights, and output layer weights. This modification aims to compensate accuracy for interpretability while both models have similar computational complexity. We follow a similar simplification routine in the creation of our SST-ANN approach explained in Section 4.3.3.

4.3 Proposed Methods

In this section we first define the problem followed by the details of the general model architectures that we use to build SERT and SST-ANN to address the problem. We then introduce a modification for encoding location information in the models' input data. Finally, we describe the masked loss function that we use for training the models.

4.3.1 Problem Definition

We have a dataset $D = \{(T_j, Y_j, M_j)\}_{j=1}^{J}$ where T_j is a multivariate time series consisting of triples (time, variable, value) written as $\{(t_i^j, f_i^j, v_i^j), i = 1, \ldots, N\}$. Here, N can be thought of as analogous to the maximum number of input tokens in NLP models, where each token corresponds to a triplet in our context. $Y_j = \{y_1^j, ..., y_k^j\}$ is the output vector including the values of k variables for a future time horizon for which we want the model to forecast. M_j is a binary vector indicating whether each of the elements in Y_j is observed for sample j in the dataset. M_j is used in the loss function (see Section 4.3.5 for details) for masking the unobserved values in the forecast window. The goal is to learn a model G that maps T_j to Y_j, i.e., $G(T_j) = Y_j$, without imputing the missing values in D or aligning the time series.

A schematic of a sample from the dataset is shown in Figure 4.2. According to the figure, it is evident that the original input data, delineated by blue dashed border lines and in a wide format, is converted into a long format, indicated by T_j. This conversion is followed by the removal of missing values. It is important to note that in this long format, removing missing values is equivalent to ignoring or deleting individual missing cells in the original wide dataset, rather than discarding entire rows due to a single missing value. Consequently, SERT can be considered a model that operates effectively with long format datasets, enabling training on all observed values. This approach avoids the loss of any observations, which would otherwise be necessary with row removal or the imputation of missing values in the wide format.

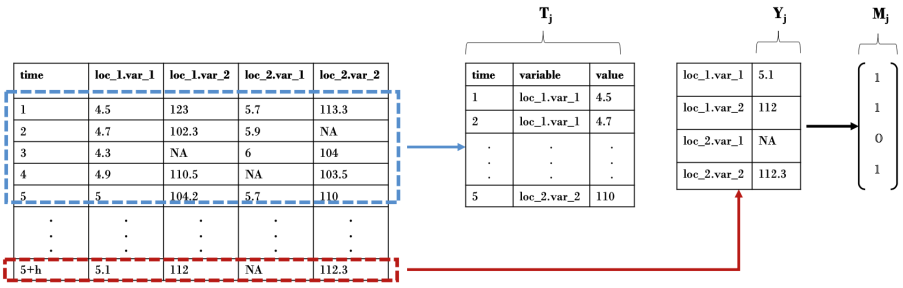

Figure 4.2: An example of a sample of a spatiotemporal dataset to be used for training our proposed models where h is the desired forecast horizon.

4.3.2 SERT

We first describe the data encoding scheme and then the model architecture including an encoder network and a linear layer. The schematic diagram of the model is shown in Figure 4.3. We will slightly modify this structure to show an alternative approach for encoding the location information in Section 4.3.4.

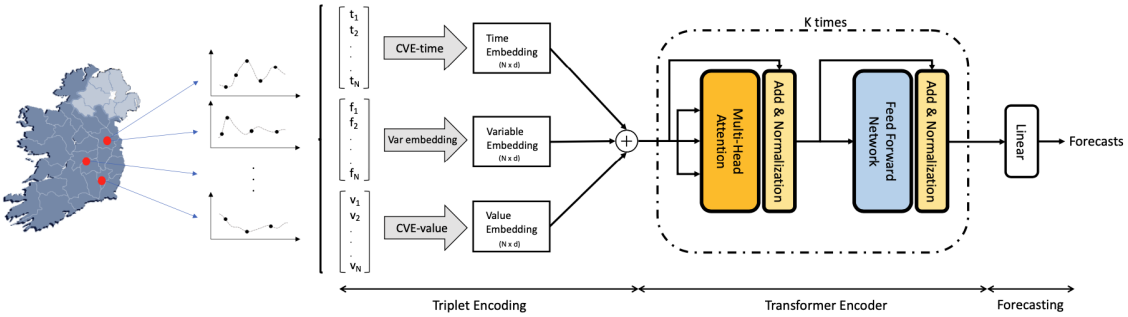

Figure 4.3: Schematic diagram of the SERT model.

Data Encoding Scheme The input data to our model is a multivariate time series dataset T that consists of N time series, each of which is associated with a time series variable f which is a sequence of observation values v's. Accordingly, an individual data point i is represented as a triplet (t_i, f_i, v_i) where t_i is time, f_i is the variable name and v_i is the measured value of the data point. To use the triplet in the model, we encode each component into an embedding and then add the embeddings together. Let $e_i^f \in R^d$ be the embedding of the variable name f_i which can be encoded similar to words using a lookup table, $e_i^t \in R^d$ be the embedding of the time index t which can be encoded using a continuous value embedding (CVE) scheme which is a one-to-many feed forward neural network [125] and $e_i^v \in R^d$ be the embedding of the value v_i which can also be encoded using the CVE. The embedding of the triplet i is then defined as $e_i = e_i^f + e_i^t + e_i^v$. The size of the embedding vector d is a hyperparameter of the model.

Encoder Network Similar to the well-known BERT model [38], the main component of our model is the encoder part of the transformer model introduced by [128]. Since transformers have become very common, we omit the details of the architecture and refer the reader to [128] for the full description. Intuitively, we can think of the encoder network as layers that take the triplet embeddings of the input data and transform them into contextualized embeddings that capture the long-range dependencies within a time series as well as cross dependencies between different time series.

Linear Layer After obtaining the contextualised embeddings of the input data using the encoder network, we then flatten the embeddings and apply a linear layer to them to generate the predictions. The linear layer is a feed-forward network with a single hidden layer and a ReLU activation function.

4.3.3 SST-ANN

The SST-ANN model is a simplified version of the SERT model that consists of only the triplet encoding and a linear layer to the output with no transformer structure in between. SST-ANN first encodes the input data using the triplet encoding scheme and then uses the embeddings as the input to single layer feed forward network to generate the predictions. Since there is no transformer structure in between, the SST-ANN model is much faster than the SERT model and, using the embeddings and the weights of the linear layer, we can compute a contribution score for each observation to the final prediction. This is useful for interpretability and variable importance analysis. More formally the output of the model can be expressed as follows:

$$\hat{y}_k = \sum_{i=1}^{N} c_i + b, \text{ with } c_i = W_{ik}^T \cdot e_i, \tag{4.1}$$

where $\hat{y}_k$ is the prediction of variable k, c_i is the contribution of the triplet (t_i, f_i, v_i) to the prediction, N is the number of observations in the input sample, b is the bias term, e_i are the embeddings of the triplet and W_{ik} is the vector of output weights associated with the embedding e_i and the target variable k.

Using the contribution scores, we can define a variable importance index. We first calculate the average contribution value of all observations belonging to the same variable as the average contribution of that variable. This calculation can be performed for a single sample to gain insights into the importance of the predictor variable for a specific target prediction, or for multiple samples used in multiple predictions to obtain an overall understanding of the predictor variable's importance in general. Next, we compute the importance of each variable by normalizing the absolute value of the average contribution values for the variables. More formally, we can express this as:

$$I_k = \frac{|\bar{c}_k|}{\sum_{k=1}^{K} |\bar{c}_k|} \times 100 \tag{4.2}$$

where I_k is the importance (in percentage) and $\bar{c}_k$ is the average contribution value of the variable k.

4.3.4 Location Encoding

We consider two different approaches to encode the location information in the input data. The first approach is to encode the location information together with the variable name f_i in the triplet encoding scheme. For example, our naming scheme for the variables can be $f_i = \{location\}_i \cdot \{variable\}_i$ where $\{location\}_i$ is the location of the time series i and $\{variable\}_i$ is the variable name of the time series i (e.g. Tolka.Turbidity). This approach uses the exact same architecture explained in Section 4.3.2. The second approach is to encode the location information separately from the variable name. This way the input time series are all assumed to arise from the same location and the location embedding is concatenated to the contextualized embeddings before the linear layer is used for prediction. In this approach, we need to structure the dataset such that all the time series from the same location are grouped together. Formally, we can define the grouped dataset as $D' = \{(T_j^L, Y_j^L, M_j^L)\}_{j=1}^{J'}$ where $L \in S$ is the location and S is the set of all locations. This approach needs a minimal modification to the previously described architecture and its schematic diagram is shown in Figure 4.4.

The first approach is similar to how the Spacetimeformer model [64] encodes the

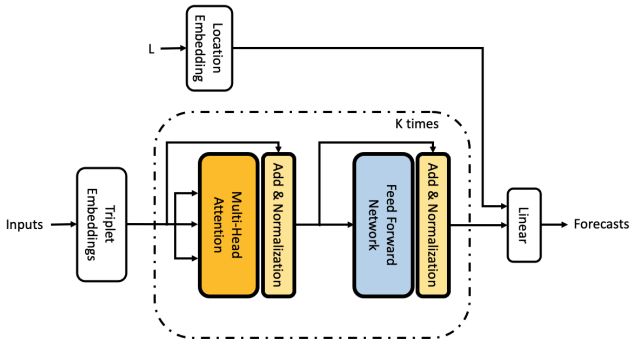

Figure 4.4: Schematic diagram of the SERT model with a separated location embedding layer.

location information while the second approach is similar to how the STraTS model [125] encodes the non-temporal information (patient demographics in their work).

4.3.5 Masked Loss Function

We use a masked Mean Squared Error (MSE) loss function to train our models. Masking in the loss function is used to handle the missing values in the output data. The masked MSE loss function is defined as follows:

$$\mathcal{L} = \frac{1}{J} \sum_{j=1}^{J} \sum_{k=1}^{K} m_k^j \left(\hat{\mathbf{y}}_k^j - \mathbf{y}_k^j \right)^2, \tag{4.3}$$

where $\mathbf{y}_k^j$ is the ground truth, $\hat{\mathbf{y}}_k^j$ is the predicted value, and m_k^j is the mask value of the target variable k in sample j.

4.4 Experiments

In this section, we will describe the experiments we conducted to evaluate the performance of our proposed models. We evaluated our models using both a simulated dataset and a real-world dataset. The primary objective of the simulation experiment is to investigate how the models perform under different sparsity levels. The real-world experiment aimed to serve as a proof of concept for the models' ability to conduct spatiotemporal multi-step ahead forecasting in real-world scenarios.

To assess the effectiveness of our models, we compared them with a baseline Naive forecaster model, which simply uses the present time observation as the next time step forecast. We also compare our models against the LSTM model and the STraTS model. Since the Naive forecaster and LSTM model cannot handle missing values, we first im-

puted these values using a forward filling method [127, p. 16].

4.4.1 Sparsity Analysis

We first simulated a dataset that consists of 16 time series each with 40,000 observations. We denote $Y_t \in R^{16}$ as the vector of observations at time t generated from the following process:

$$Y_t = 2 + 0.4Y_{t-1} + X_t + s_t,\ s_t \sim MVN(0, \Sigma) \tag{4.4}$$

where s_t is spatial random effect with mean zero and variance-covariance matrix Σ that is generated with:

$$\Sigma = U \cdot U^T,\ U \in R^{16\times 16},\ U_{i,j} \sim Uniform(-1, 1), \tag{4.5}$$

and $X_t \in R^{16}$ is the vector of temporal effects generated as:

$$\begin{aligned} X_t^T \quad = \quad & [10\sin(p_1 t), \cos(p_2 t), p_3 t, -p_3 t + 10\sin(p_1 t), \\ & 5\sin(p_2 t), 12\cos(p_2 t), 7\sin(p_2 t), 8\cos(p_2 t), \\ & 2\sin(p_2 t), 3\cos(p_2 t), 12\sin(p_2 t), 18\cos(p_2 t), \\ & 4\sin(p_2 t), 15\cos(p_2 t), 11\sin(p_2 t), 10\cos(p_2 t)] \end{aligned} \tag{4.6}$$

where $p_1 = 0.005$, $p_2 = 0.0005$ and $p_3 = 0.002$.

We use the first 37,000 time steps to train the models and the remaining 3,000 time steps to evaluate the performance of them. We used the model structure shown in Figure 4.3 to train our proposed models. We trained all models for 1 step ahead forecasting using the previous 10 time steps of observations. We consider five different sparsity levels to fit the models. Accordingly, we remove $n\%$ of the observations randomly for $n = \{0\%, 20\%, 40\%.60\%, 80\%\}$. We use the root mean squared error (RMSE) as the evaluation metric. The results are presented in Figure 4.5.

4.4.2 Real Dataset; Environmental Monitoring in Dublin bay, Ireland

The dataset includes hourly measurements from 2017-01-01 to 2021-12-31. An example of the data is shown in Figure 4.1. The locations of the data are shown in Figure 4.6. We use the data from the first four years to train the models and the last year to evaluate their performance. We train the models using the previous 10 hours of observations as the input and forecasting seven hours ahead. We use the same evaluation metric as in the simulated dataset. We tried both location encoding approaches (explained in Section

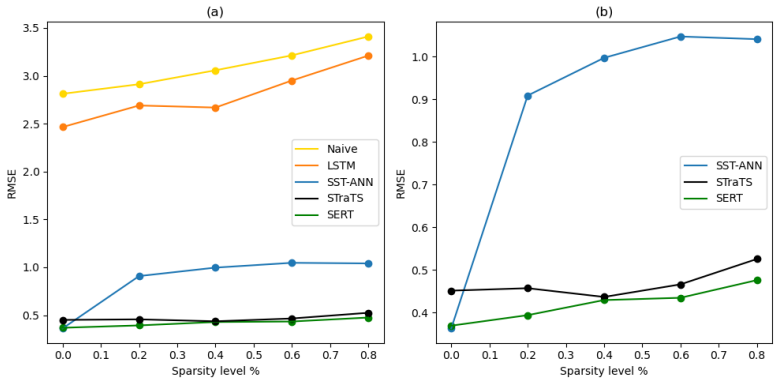

Figure 4.5: a) Performance of all the models used on the simulated dataset. b) Comparison of SST-ANN, STraTS, and SERT zoomed in.

4.3.4) with our proposed models on the real-world dataset and found that the second approach, depicted in Figure 4.4, performed better and here we only report the results of the superior approach. The results are presented in Table 4.1.

Table 4.1: RMSE of the models for 7 hour ahead forecasting of the 7 environmental variables in Dublin bay.

Model	Dissolved Oxygen	Precipitation	Salinity	Temperature	Turbidity	Water Level	Wind Speed
Naive	0.64	1.19	1.22	0.21	0.93	1.87	0.8
LSTM	0.65	0.89	0.8	0.41	0.81	0.5	0.82
STraTS	0.53	**0.88**	0.7	**0.18**	**0.72**	**0.38**	0.76
SERT (ours)	**0.49**	**0.88**	**0.67**	**0.18**	**0.72**	0.4	**0.73**
SST-ANN (ours)	0.51	**0.88**	0.73	0.45	0.88	0.63	0.79

Deep learning models have been developed utilizing Keras, with TensorFlow as the underlying framework. We utilized the same computational resources (a single NVIDIA P100 GPU) to train the models. Every model is trained using a batch size of 70, employing the Adam optimizer, and training halts if the loss remains stagnant for two consecutive epochs. The specifications and speed performance details of the models are reported in the Table 4.2.

Table 4.2: Computational specifications of the fitted models.

Model	Specifications	Sec/epoch
Naive	—	—
LSTM	parameters = 139844	2
STraTS	parameters = 183147, # of heads = 6, k = 6	230
SERT	parameters = 137795, # of heads = 6, k = 6	40
SST-ANN	parameters = 118983	9

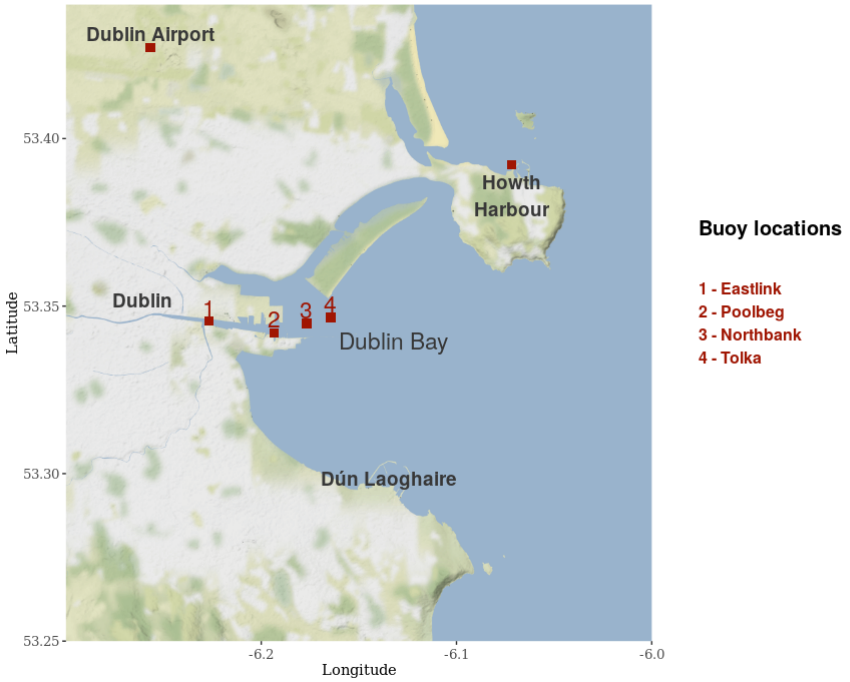

Figure 4.6: Buoys measuring various environmental variables in Dublin Bay. The 4 main environmental variables (Turbidity, Salinity, Dissolved Oxygen, Temperature) are available at the number buoys. The weather measurements (Rainfall, Wind Speed) come from Dublin Airport. The water level variable comes from Howth Harbour.

4.4.3 Interpretability of the SST-ANN Model

The results of the variable importance for the real-world dataset experiment are presented in Figure 4.7. In this example, we only consider the contributions of water level, temperature, wind speed and precipitation to predictions of turbidity, disolved oxygen and salinity, since we know that the former variables could affect the latter variables but not vice versa. However, one should take caution and not interpret these numbers as they would in a causal model. Instead, the application of variable importance in SST-ANN should be more closely compared to the concept of variable importance as used in a random forest model (for example, see [10]). According to the results, temperature is the most important variable in predicting the target variables. This can be explained by understanding that temperature indeed has a significant causal effect on them [114, 2, 1].

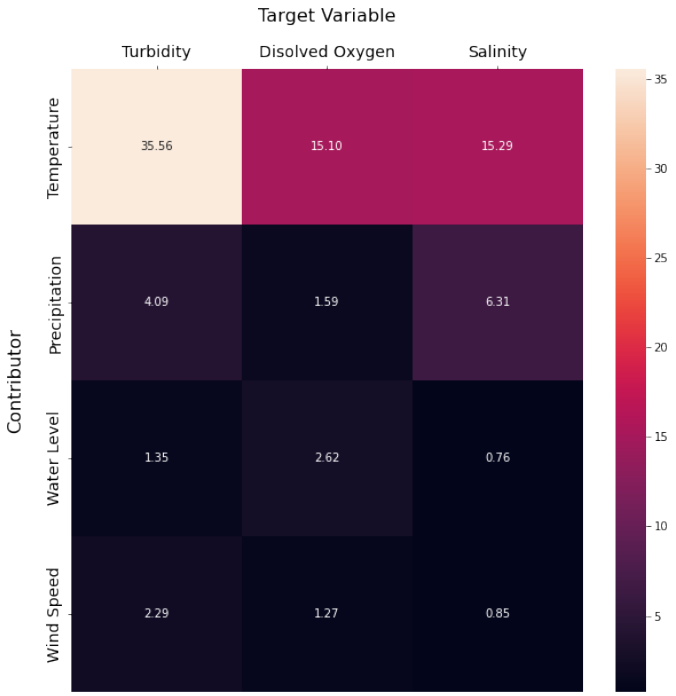

Figure 4.7: Variable importance of the selected variables in predicting the target variables.

4.5 Conclusion

In this paper, we proposed two novel models for spatio-temporal forecasting called SERT and SST-ANN. SERT is a transformer based model while SST-ANN is a simple ANN model combined with triplet encoding of the input data. Furthermore, we showed that STraTS, a model originally developed for sparse and irregularly sampled clinical time series classification, can be used for spatio-temporal forecasting, especially when missing values are present in the data. The proposed approaches do not require aggregation or missing value imputation techniques, and avoid the problems introduced by such methods. We evaluated the performance of the proposed models on a simulated dataset with varying levels of sparsity and showed that in general increasing sparsity has a negative effect in the performance of all the models, but SERT followed by STraTS and SST-ANN are more robust to the increase in sparsity. We also evaluated the performance of the proposed models on a real-world dataset of environmental variables in Dublin Bay, Ireland. The results indicate that SERT outperformed the other models in 7-hour ahead forecasting for 6 out of the 7 variables, with 3 of them being on par with STraTS while demonstrating significantly faster performance. We then showed how SST-ANN can be used to interpret the predictions of the model by calculating and using the contribution score of the input data to develop an importance index using the average contribution scores.

We introduced two different methods to encode the location information in our proposed models, including encoding the time series variable name with the location name simultaneously and encoding the location name separately. However, neither of these methods takes into account the distance between the locations, which is a limitation of our work. We believe future research should focus on incorporating this information into the models, as it has the potential to improve forecasting performance and be utilized for spatiotemporal interpolation tasks.

5
Concluding Remarks

The concluding section revisits key points and outcomes from Chapters 2-4, and outlines potential paths for future exploration.

5.1 Sea level rise in Dublin Bay

The sea level has long been a topic of concern due to its implications on coastal ecosystems, urban areas, and global climate patterns. I presented an updated sea level dataset for Dublin, covering the period from 1938 to 2016. By providing a more reliable record and understanding the variations in sea level rise, my research aims to offer insights that could be important for future urban planning and coastal defense measures in Dublin.

My main objective was to refine and update the sea level dataset for Dublin and estimate the rate of sea level rise using a more reliable dataset. Specifically, I aimed to rectify any inconsistencies in the Dublin Port sea level record by incorporating data from other reliable sources and applying calibration techniques. My approach could be divided into the following steps:

- **Data Collection:** I assembled comprehensive sea level records from five main sources: Dublin Port, Arklow, and Howth Harbour, Brest and Newlyn tide gauges. Having multiple sources of data is crucial for data quality checks and detecting anomalous periods.

- **Data Calibration:** One major challenge identified was the biased high water measurements, which skewed the mean sea level (MSL) calculations. To address this, I employed a Bayesian linear regression model. This model is particularly effective because it provides a probabilistic approach, allowing for the inclusion of prior knowledge and uncertainties in the data. Crucially, the Mean Low Water (MLW) values were integrated as a predictor in the model. This helped reconstruct the MSL and eliminate the bias introduced by the erroneous MHW values.

- **Validation:** Post-calibration, it was essential to validate the new dataset. I compared the re-created MSL dataset with other nearby tide gauge datasets to ensure its consistency and accuracy. This step was vital to ensure the general applicability and reliability of the dataset.

- **Analysis:** With the validated dataset in hand, I then delved into understanding the rates of sea level rise. I mainly focused on the following periods: 1953-2016 and 1997-2016. My analysis revealed significant insights into the multidecadal variability of sea level rise, particularly highlighting the increased rates in recent years.

In summary, I have provided a detailed framework for researchers studying sea level rise. This framework guides the aggregation of tide gauge data from multiple sources and recommends conducting data quality checks. These checks can be done by visualizing sea level time series at different levels, from low waters to high waters, and by employing statistical techniques such as change point modeling. I also addressed the removal of bias in the data using regression and emphasized the importance of properly accounting for uncertainty using a Bayesian framework. Furthermore, I highlighted the decadal variability in sea level and its rate of rise, which is almost twice the global average. This rate is alarming and has significant implications for policy makers.

5.2 Turbidity in Dublin Bay

Turbidity, a measure of water clarity, holds significance in preserving the health of coastal ecosystems. Factors contributing to turbidity range from natural phenomena, such as weather changes, to human-induced activities like dredging and dumping operations. Particularly in regions like Dublin Bay, these activities can have significant implications for water quality. Hence, understanding and predicting turbidity patterns becomes indispensable for effective coastal management and environmental conservation.

I focused on studying water turbidity levels in Dublin Bay, extracting measurements from five different locations across a span of two years (2017-2018). These measurements offer insights into how light is reflected through suspended particles in the water. Notably, four of the locations focused on a singular depth, mapping the channel from the River Liffey to Dublin Bay, a crucial passage for vessels and a prime site for dredging. The fifth monitoring point, positioned around 10 kilometers from Dublin port, is of particular interest as it is the designated site for sediment disposal and measures turbidity at three different water column depths.

In my journey to decipher the intricate interactions of variables affecting turbidity, I introduced the Vector Auto-Regressive Integrated Conditional Heteroskedasticity (VARICH) model. This dynamic spatiotemporal model is tailored to describe turbid-

ity's response to a multitude of environmental factors, most notably human activities like dredging and dumping.

The VARICH approach is able to measure the effects of multivariate time series on each other, while also accounting for known volatility changes in the series. Through this model, I aimed to offer a comprehensive, yet nuanced understanding of turbidity patterns, especially in relation to human activities and their potential environmental ramifications. I demonstrated that these activities have varying effects at different locations. While controlling for wind speed as the primary natural factor, none of these human-induced activities showed effects as pronounced as that of wind speed in Dublin Bay.

Upon fitting various models, including VAR, ARCH, VARCH, and VARICH, to the data, I compared their performances. I found that accounting for heteroskedasticity is crucial when modeling turbidity levels. Specifically, the VARICH model outperformed the others. By leveraging a Bayesian framework and incorporating informative priors, I demonstrated the model's utility in handling incomplete time series data—a common issue in environmental monitoring due to sensor failures.

My research endeavors to bridge the gap between data-driven analytics and environmental conservation. By employing the VARICH model, I not only offer a specialised analytical tool for understanding turbidity but also emphasise the importance of continuous monitoring and data analysis in safeguarding our coastal ecosystems.

As coastal regions globally grapple with the challenges of balancing development with conservation, my findings underscore the importance of informed decision-making. Through rigorous data analysis and modeling, we can better anticipate environmental impacts, ensuring that our coasts remain pristine for generations to come.

5.3 SERT

In the body of this thesis, a critical and often overlooked aspect of environmental science has been brought to the forefront: the intricacies and challenges associated with large-scale spatio-temporal data analysis. Traditional methodologies often fall short in addressing the complexities presented by this data, particularly when faced with multivariate misaligned sparse time series — a common occurrence resulting from inevitable issues such as equipment malfunctions or inconsistent data collection schedules.

I have developed two models, each embodying a nuanced approach to multivariate spatio-temporal anlysis. These methodologies represent practical and impactful advances in the realm of environmental monitoring. The models are crafted to handle missing data adeptly, bypassing the need for traditional imputation techniques that often introduce biases or noise.

SERT (Spatio-temporal Encoder Representations from Transformers): The

development of the SERT model marks a significant milestone. By harnessing the power of transformer architectures, known for their exceptional performance in language processing tasks, I adapted and extended this technology to suit the demands of spatio-temporal data analysis. This adaptation was not without its challenges, necessitating an in-depth exploration of both the transformer architecture and the specific demands of environmental data.

The resulting model is a testament to the versatility of the transformer architecture, demonstrating its applicability beyond its usual domains. SERT distinguishes itself by seamlessly accommodating data irregularities and gaps, factors that traditionally introduce challenges to the analysis.

SST-ANN (Sparse Spatio-Temporal Artificial Neural Network): Parallel to the development of SERT, I recognized the necessity for a model that, while simpler, did not compromise on the ability to handle the nuances of environmental data. The SST-ANN model was conceptualized and brought to fruition with this need in mind. It stands as a robust alternative to more complex models, requiring fewer computational resources and offering greater ease of interpretation.

Despite its simplicity, the SST-ANN still performs well in managing sparse, multivariate, and misaligned data. Its architecture, though less complex than SERT's, has been designed to tackle the specific challenges of environmental data. The model maintains its efficiency without incurring significant performance costs, making it a practical choice for real-world applications where resource constraints are a pressing concern.

The journey through the conceptualization, development, and optimization of these models was exhaustive. I engaged with various data scenarios, simulating real-world environmental monitoring challenges to test the models' robustness. Through this rigorous evaluation process, both models demonstrated superior performance in handling missing data and providing more accurate forecasts compared to existing methods not designed for such data.

This thesis does not conclude with the development of these models but opens a gateway for future exploration and innovation. The adaptability of the SERT and SST-ANN models offers numerous avenues for further research, from exploring different architectural tweaks to adapting these models for other data-intensive fields.

Moreover, the implications of this work are far-reaching, extending beyond the academic sphere and poised to influence on-the-ground environmental monitoring practices. By enhancing data accuracy and forecasting reliability, these models could serve as valuable tools for policymakers and environmental stakeholders, informing more effective and timely decision-making processes.

5.4 Future Research

The methodologies outlined in this thesis contribute to improving the depth and reliability of environmental data analysis. However, achieving reliability requires more than just enhancing computational models; it demands a careful approach to collecting and managing fundamental data. As we explore areas marked by uncertainty, the importance of our initial data integrity becomes crucial. Consequently, future research must prioritize refining data collection and data management procedures. This represents a significant step in uncovering the complex patterns and subtle dynamics inherent in environmental phenomena.

This thesis presents my contributions to the field via the development of statistical and machine learning models designed to analyze complex environmental data. These models represent initial steps, providing a foundation for further research and enhancement. They are intended to serve both as practical tools for current use and as guides for future studies, inspiring researchers to explore new areas. Future research could take several directions, a few of which I suggest below. These recommendations are not exhaustive but indicative of the expansive horizon that awaits scholarly exploration and inquiry.

The sea level analysis presented in Chapter 2, is a multi-stage approach including deterministic adjustments and model fittings which are performed sequentially. While this approach has been beneficial for understanding various aspects of the data, it has inherent limitations, particularly in the propagation of uncertainty. Each stage is treated somewhat independently, with uncertainties not fully carried over to subsequent stages. This can lead to underestimation of the overall uncertainty. An alternative and more holistic approach is data fusion, which involves the joint modeling of multiple data sets, taking into account various sources of uncertainty. This approach allows for a more comprehensive understanding of the data by considering all sources of information and their uncertainties in a unified model. One effective way to implement data fusion is through state space models. These models, as part of the data fusion approach, provide a flexible framework for accommodating various complexities of real-world data. They allow for the modeling of dynamic systems and can incorporate changes and trends over time, making them particularly suitable for time series analysis. By treating all data sources as part of a single model, these models naturally propagate uncertainty through all stages of the analysis, providing more accurate and reliable quantification of uncertainty (for example see [73]). However, such an approach comes at the cost of increased computational complexity and requires a more sophisticated modeling framework.

The VARICH model, as it stands, unveils several avenues for enhancement and refinement. A primary limitation in its current configuration is the oversight of spatial relationships between data collection points. Our initial applications involved closely-

knit sites, minimizing spatial discrepancies. However, as we broaden the scope to encompass more geographically dispersed locations, the model must adapt to acknowledge the increased likelihood of interaction and dependency among adjacent sites compared to those more distant. Integrating geographical data and the relative positioning of sites could significantly augment the model's precision. Moreover, the model's rigidity in handling exogenous variables could benefit from incorporating coefficients that adjust over time. This added flexibility could enhance the model's capability to accurately represent non-linear anomalies and trends, providing a more realistic view of environmental dynamics. Expanding the model's dimensions represents another frontier, enabling it to include different environmental variables and to recognize dependencies with greater historical depth through higher autoregressive orders. This multidimensional approach, however, comes at the cost of increased computational demands. To mitigate this, strategies aimed at boosting computational efficiency must be employed. Implementing sparsity within the autoregressive matrices or leveraging informative priors could be viable strategies. Additionally, reparameterization of the model can facilitate the convergence processes of algorithms like HMC, enhancing their speed and efficiency. Finally, a possible extension of the VARICH model could involve its evolution into a model capable of working with irregularly sampled time series data, possibly by transitioning from discrete to continuous time model development. Modeling approaches such as the utilization of Gaussian Processes (GPs) or P-splines could accommodate continuous time, but developing an approach that models the interaction of multivariate continuous time series could prove to be more challenging and requires further research.

Further development of the SERT model might involve replacing the transformer structure with a more efficient one, rather than entirely eliminating it as the SST-ANN does. Convolutional layers could be a viable alternative. Like attention layers, these seek to aggregate information; however, unlike attention layers, they do not attend to every input token, thereby reducing computational demands. Moreover, enhancing SERT to offer prediction with confidence intervals could present significant advantages. This enhancement can be achieved through model-agnostic methods like conformal prediction [51] or by modifying the model's structure and loss function to predict data distribution parameters rather than mere point predictions. Another potential refinement involves incorporating a geographical information encoding scheme, facilitating spatiotemporal interpolation. This could be accomplished by devising an efficient method for encoding geographical coordinates (i.e., longitude and latitude), thereby equipping the model to handle such data and rendering it particularly appropriate for point process data.

Bibliography

[1] Dissolved oxygen and water. https://www.usgs.gov/special-topics/water-science-school/science/dissolved-oxygen-and-water#:~:text=Cold%20water%20can%20hold%20more,oxygen%20concentration%20is%20often%20lower. Accessed: 2024-01-24.

[2] Temperature and salinity. https://www.jsg.utexas.edu/helper/files/WuLi_all_327G.pdf. Accessed: 2024-01-24.

[3] Tide gauge sea level data — climate data guide. https://climatedataguide.ucar.edu/climate-data/tide-gauge-sea-level-data. Accessed: 2023-08-09.

[4] What is a tide gauge? - NOAA's national ocean service. https://oceanservice.noaa.gov/facts/tide-gauge.html. Accessed: 2023-08-09.

[5] Dublin city council, Dublin coastal flooding protection project, report no. 9M2793. Technical report, 2005.

[6] Y. Abbassi, H. Ahmadikia, and E. Baniasadi. Impact of wind speed on urban heat and pollution islands. *Urban Climate*, 44:101200, 2022.

[7] L. S. Aiken, S. G. West, and S. C. Pitts. Multiple linear regression. *Handbook of Psychology*. 481–507, 2003.

[8] M. Ali, R. Prasad, Y. Xiang, and R. C. Deo. Near real-time significant wave height forecasting with hybridized multiple linear regression algorithms. *Renewable and Sustainable Energy Reviews*, 132:110003, 2020.

[9] A. A. Alola, G. Uzuner, and S. S. Akadiri. Modeling tourism and fear nexus in G4 countries. *Current Issues in Tourism*, 24(10):1333–1339, 2021.

[10] K. J. Archer and R. V. Kimes. Empirical characterization of random forest variable importance measures. *Computational statistics & data analysis*, 52(4):2249–2260, 2008.

[11] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint:1409.0473*, 2014.

[12] E. B. Barbier, S. D. Hacker, C. Kennedy, E. W. Koch, A. C. Stier, and B. R. Silliman. The value of estuarine and coastal ecosystem services. *Ecological Monographs*, 81(2):169–193, 2011.

[13] M. Betancourt. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint:1701.02434*, 2017.

[14] A. J. Bever, M. L. MacWilliams, and D. K. Fullerton. Influence of an observed decadal decline in wind speed on turbidity in the San Francisco Estuary. *Estuaries and Coasts*, 41(7):1943–1967, 2018.

[15] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[16] L. Bonacina, F. Fasano, V. Mezzanotte, and R. Fornaroli. Effects of water temperature on freshwater macroinvertebrates: a systematic review. *Biological Reviews*, 98(1):191–221, 2023.

[17] S. Bradley, G. Milne, I. Shennan, and R. Edwards. An improved glacial isostatic adjustment model for the British Isles. *Journal of Quaternary Science*, 26(5):541–552, 2011.

[18] E. Bradshaw, P. Woodworth, A. Hibbert, L. Bradley, D. Pugh, C. Fane, and R. Bingley. A century of sea level measurements at Newlyn, Southwest England. *Marine Geodesy*, 39(2):115–140, 2016.

[19] C. Briciu-Burghina, T. Sullivan, J. Chapman, and F. Regan. Continuous high-frequency monitoring of estuarine water quality as a decision support tool: a Dublin Port case study. *Environmental Monitoring and Assessment*, 189(9):5561–5580, 2014.

[20] S. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

[21] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.

[22] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[23] V. Buhrmester, D. Münch, and M. Arens. Analysis of explainers of black box deep neural networks for computer vision: A survey. *Machine Learning and Knowledge Extraction*, 3(4):966–989, 2021.

[24] W. Camaro Garcia, N. Dwyer, F. Barrett, A. Berry, M. Cronin, C. Cusack, S. Gallagher, J. Gault, M. Gill, E. Gleeson, J. Hanley, P. Kane, K. Lambkin, R. Lawlor, K. Lydon, K. Lyons, D. Martin, G. McCarthy, E. McGovern, C. Murphy, G. Nolan, C. Nugent, B. O'Dwyer, J. Ovadnevaite, C. Quinlan, M. Saunders,

J. Silke, G. Smith, R. Thomas, S. Walsh, G. Westbrook, W. Eoin, and R. Wilkes. Climate Status Report for Ireland. Technical report, EPA, 2021.

[25] B. P. Carlin, A. E. Gelfand, and A. F. Smith. Hierarchical Bayesian analysis of changepoint problems. *Journal of the Royal Statistical Society: series C (Applied Statistics)*, 41(2):389–405, 1992.

[26] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76, 2017.

[27] R. Carter. Sea-level changes in Northern Ireland. *Proceedings of the Geologists' Association*, 93(1):7–23, 1982.

[28] M. Castangia, L. M. M. Grajales, A. Aliberti, C. Rossi, A. Macii, E. Macii, and E. Patti. Transformer neural networks for interpretable flood forecasting. *Environmental Modelling & Software*, 160:105581, 2023.

[29] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *ArXiv preprint:1409.1259*, 2014.

[30] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in Neural Information Processing Systems*, 29, 2016.

[31] CSO. Central statistics office: Population and migration estimates april 2020. `https://www.cso.ie/en/csolatestnews/pressreleases/2020pressreleases/pressstatementpopulationandmigrationestimatesapril2020/`, 2020. (Accessed: 19.04.2021).

[32] K. B. Dang, K. C. Vu, H. Nguyen, D. A. Nguyen, T. D. L. Nguyen, T. P. N. Pham, T. L. Giang, H. D. Nguyen, T. H. Do, et al. Application of deep learning models to detect coastlines and shorelines. *Journal of Environmental Management*, 320:115732, 2022.

[33] S. Dangendorf, M. Marcos, G. Wöppelmann, C. Conrad, T. Frederikse, and R. Riva. Reassessment of 20th century global mean sea level rise. *Proceedings of the National Academy of Sciences*, 114(23):5946–5951, 2017.

[34] R. Davies-Colley and D. Smith. Turbidity suspended sediment, and water clarity: a review. *JAWRA Journal of the American Water Resources Association*, 37(5):1085–1101, 2001.

[35] R. Davies-Colley and D. Smith. Turbidity, suspended sediment, and water clarity: a review. *Journal of the American Water Resources Association*, 37(5):1085–1101, 2001.

[36] DCC. Dublin city council: Climate change action plan. `https://www.dublincity.ie/sites/default/files/content/WaterWasteEnvironment/Waste/Documents/2019DCCClimateChangeActionPlan.pdf`, 2017. (accessed: 20.03.2020).

[37] L. Deng, K. Chen, Z. Liu, B. Wu, Z. Chen, and S. He. Spatiotemporal variation evaluation of water quality in middle and lower Han River, China. *Scientific Reports*, 12(1):1–14, 2022.

[38] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint:1810.04805*, 2018.

[39] R. Devoy. Coastal vulnerability and the implications of sea-level rise for Ireland. *Journal of Coastal Research*, 24(2(242)):325–341, 2008.

[40] R. J. Devoy. Sea-level rise: causes, impacts, and scenarios for change. In *Coastal and Marine Hazards, Risks, and Disasters*. Elsevier. 197–241, 2015.

[41] D. Dhall, R. Kaur, and M. Juneja. Machine learning: a review of the algorithms and its applications. *Proceedings of ICRIC 2019: Recent Innovations in Computing*. 47–63, 2020.

[42] S. T. Diabaté, D. Swingedouw, J. J. M. Hirschi, A. Duchez, P. J. Leadbitter, I. D. Haigh, and G. D. McCarthy. Western boundary circulation and coastal sea-level variability in northern hemisphere oceans. *Ocean Science Discussions*. 1-34, 2021.

[43] M. Díez-Minguito and H. E. de Swart. Relationships between chlorophyll-a and suspended sediment concentration in a high-nutrient load estuary: An observational and idealized modeling approach. *Journal of Geophysical Research: Oceans*, 125(3), 2020.

[44] S. C. Doney, M. Ruckelshaus, J. Emmett Duffy, J. P. Barry, F. Chan, C. A. English, H. M. Galindo, J. M. Grebmeier, A. B. Hollowed, N. Knowlton, et al. Climate change impacts on marine ecosystems. *Annual Review of Marine Science*, 4:11–37, 2012.

[45] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16

words: Transformers for image recognition at scale. *arXiv preprint:2010.11929*, 2020.

[46] M. Du, N. Liu, and X. Hu. Techniques for interpretable machine learning. *Communications of the ACM*, 63(1):68–77, 2019.

[47] Dublin Port Company. Report: Maintenance dredging campaign. Technical report, 2016.

[48] R. Edwards and K. Craven. Relative sea-level change around the Irish coast. In *Advances in Irish Quaternary Studies*. 2017.

[49] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*. 987–1007, 1982.

[50] J. H. Faghmous and V. Kumar. A big data guide to understanding climate change: The case for theory-guided data science. *Big data*, 2(3):155–163, 2014.

[51] M. Fontana, G. Zeni, and S. Vantini. Conformal prediction: a unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.

[52] T. Frederikse, K. Simon, C. A. Katsman, and R. Riva. The sea-level budget along the northwest Atlantic coast: GIA, mass changes, and large-scale ocean dynamics. *Journal of Geophysical Research: Oceans*, 122(7):5486–5501, 2017.

[53] L.-T. Fu. Effect of the local wind reduction zone on seed dispersal from a single shrub element on sparsely vegetated land. *AoB Plants*, 13(4):plab025, 2021.

[54] F. Fung, M. Palmer, T. Howard, J. Lowe, P. Maisey, and J. Mitchell. UKCP18 factsheet: Sea level rise and storm surge. Technical report, Met Office Hadley Centre, Exeter, 2018.

[55] A. Gamble and M. Babbar-Sebens. On the use of multivariate statistical methods for combining in-stream monitoring data and spatial analysis to characterize water quality conditions in the White River Basin, Indiana, USA. *Environmental Monitoring and Assessment*, 184:845–875, 2012.

[56] N. K. Ganju, J. M. Testa, S. E. Suttles, and A. L. Aretxabaleta. Spatiotemporal variability of light attenuation and net ecosystem metabolism in a back-barrier estuary. *Ocean Science*, 16(3):593–614, 2020.

[57] J. Ge, R. Torres, C. Chen, J. Liu, Y. Xu, R. Bellerby, F. Shen, J. Bruggeman, and P. Ding. Influence of suspended sediment front on nutrients and phytoplankton dynamics off the Changjiang Estuary: A FVCOM-ERSEM coupled model experiment. *Journal of Marine Systems*, 204:103292, 2020.

[58] A. E. Gelfand. Gibbs sampling. *Journal of the American Statistical Association*, 95(452):1300–1304, 2000.

[59] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3 edition, 2013.

[60] A. Gelman and D. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

[61] E. B. Goldstein, G. Coco, and N. G. Plant. A review of machine learning applications to coastal sediment transport and morphodynamics. *Earth Science Reviews*, 194:97–108, 2019.

[62] R. J. Goudie, R. M. Turner, D. De Angelis, and A. Thomas. Multibugs: A parallel implementation of the bugs modelling framework for faster Bayesian inference. *Journal of Statistical Software*, 95, 2020.

[63] A. Graves, A.-r. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 6645–6649,2013.

[64] J. Grigsby, Z. Wang, and Y. Qi. Long-range transformers for dynamic spatiotemporal forecasting. *arXiv preprint:2109.12218*, 2021.

[65] I. D. Haigh, M. Eliot, and C. Pattiaratchi. Global influences of the 18.61 year nodal cycle and 8.85 year cycle of lunar perigee on high tidal levels. *Journal of Geophysical Research: Oceans*, 116(C6), 2011.

[66] A. Hamdi, K. Shaban, A. Erradi, A. Mohamed, S. K. Rumi, and F. D. Salim. Spatiotemporal data mining: a survey on challenges and open problems. *Artificial Intelligence Review*. 1–48, 2022.

[67] S. M. Harrington, V. Wishingrad, and R. C. Thomson. Properties of Markov chain Monte Carlo performance across many empirical alignments. *Molecular Biology and Evolution*, 38(4):1627–1640, 2021.

[68] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[69] S. Holgate, A. Matthews, P. Woodworth, L. Rickards, M. Tamisiea, E. Bradshaw, P. Foden, K. Gordon, S. Jevrejeva, and J. Pugh. New data systems and products at the permanent service for mean sea level. *Journal of Coastal Research*, 29(3):493 – 504, 2013.

[70] S. Hølleland and H. A. Karlsen. A stationary spatio-temporal GARCH model. *Journal of Time Series Analysis*, 41(2):177–209, 2020.

[71] M. Horn, M. Moor, C. Bock, B. Rieck, and K. Borgwardt. Set functions for time series. In *International Conference on Machine Learning*. PMLR. 4353–4363, 2020.

[72] IMI. Irish marine institute website. `https://erddap.marine.ie/erddap/tabledap/IrishNationalTideGaugeNetwork.html`, 2019. (Accessed: 20.03.2020).

[73] S. Johnson, S. Heaps, K. Wilson, and D. Wilkinson. A Bayesian spatio-temporal model for short-term forecasting of precipitation fields. *Environmetrics*, 2023.

[74] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77:437–471, 1996.

[75] M. Kanamitsu, W. Ebisuzaki, J. Woollen, S.-K. Yang, J. Hnilo, M. Fiorino, and G. Potter. NCEP–DOE AMIP-II reanalysis (r-2). *Bulletin of the American Meteorological Society*, 83:1631–1643, 2002.

[76] S. Karlsson. Forecasting with Bayesian vector autoregression. *Handbook of Economic Forecasting*, 2:791–897, 2013.

[77] F. D. Keles, P. M. Wijewardena, and C. Hegde. On the computational complexity of self-attention. In *International Conference on Algorithmic Learning Theory*. PMLR. 597–619, 2023.

[78] M. U. Kemp, E. Emiel van Loon, J. Shamoun-Baranes, and W. Bouten. RNCEP: global weather and climate data at your fingertips. *Methods in Ecology and Evolution*, (3):65–70, 2012.

[79] M. J. Kennish. *Ecology of estuaries: anthropogenic effects*, volume 1. CRC press, 1991.

[80] B. G. Kitchener, J. Wainwright, and A. J. Parsons. A review of the principles of turbidity measurement. *Progress in Physical Geography*, 41(5):620–642, 2017.

[81] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[82] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao. Independently recurrent neural network (indrnn): Building a longer and deeper rnn. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5457–5466, 2018.

[83] B. Lim and S. Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194):20200209, 2021.

[84] R. J. Little and D. B. Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

[85] Y. Liu and M. Lapata. Text summarization with pretrained encoders. *arXiv preprint:1908.08345*, 2019.

[86] H. Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

[87] V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Z. (eds.). IPCC, 2021: Summary for policymakers. in: Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change. Technical report, Cambridge University Press, 2021.

[88] R. B. McCleskey, C. A. Cravotta III, M. P. Miller, F. Tillman, P. Stackelberg, K. J. Knierim, and D. R. Wise. Salinity and total dissolved solids measurements for natural waters: An overview and a new salinity method based on specific conductance and water type. *Applied Geochemistry*. 105684, 2023.

[89] T. L. McDonald. Review of environmental monitoring methods: survey designs. *Environmental Monitoring and Assessment*, 85:277–292, 2003.

[90] M. Mendil, S. Leirens, P. Armand, and C. Duchenne. Hazardous atmospheric dispersion in urban areas: A deep learning approach for emergency pollution forecast. *Environmental Modelling & Software*, 152:105387, 2022.

[91] Met Éireann. Dublin wind data. `https://cli.fusio.net/cli/climate_data/webdata/dly532.csv`, 2022. (accessed: 13.06.2022).

[92] J. Murphy, G. Sutton, C. O'Mahony, and P. Woodworth. Scoping study to assess the status of Ireland's tide gauge infrastructure and outline current and future requirements. Technical report, 2003.

[93] R. Nerem, B. Beckley, J. Fasullo, B. Hamlington, D. Masters, and G. Mitchum. Climate change driven accelerated sea level rise detected in the altimeter era. *Proceedings of the National Academy of Sciences*, 115(9):2022–2025, 2018.

[94] B. Neumann, A. T. Vafeidis, J. Zimmermann, and R. J. Nicholls. Future coastal population growth and exposure to sea-level rise and coastal flooding-a global assessment. *PloS one*, 10(3):e0118571, 2015.

[95] R. Nicholls, P. Wong, V. Burkett, J. Codignotto, and J. Hay. Coastal systems and low-lying areas. In *Climate Change 2007: Impacts, Adaptation and Vulnerability.* Cambridge University Press: Cambridge, UK. 315–356, 2007.

[96] R. J. Nicholls and A. Cazenave. Sea-level rise and its impact on coastal zones. *Science*, 328(5985):1517–1520, 2010.

[97] J. Orford, , and J. Murdy. Presence and possible cause of periodicities in 20th-century extreme coastal surge: Belfast harbour, northern Ireland. *Global and Planetary Change*, 133:254–262, 2015.

[98] J. Orford, J. Murdy, and R. Freel. Developing constraints on the relative sea-level curve for the northeast of Ireland from the mid-holocene to the present day. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 364(1841):857–866, 2006.

[99] P. Otto, W. Schmid, and R. Garthoff. Generalised spatial and spatiotemporal autoregressive conditional heteroscedasticity. *Spatial Statistics*, 26:125–145, 2018.

[100] G. O'Connell. Interview on pat kenny newstalk: Is catastrophic Dublin flooding inevitable? `https://www.newstalk.com/podcasts/highlights-from-the-pat-kenny-show/catastrophic-dublin-flooding-inevitable`, 2019. (accessed: 19.04.2021).

[101] D. R. Piatka, J. J. Venkiteswaran, B. Uniyal, R. Kaule, B. Gilfedder, and J. A. Barth. Dissolved oxygen isotope modelling refines metabolic state estimates of stream ecosystems with different land use background. *Scientific Reports*, 12(1):10204, 2022.

[102] M. Plummer et al. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vienna, Austria. 124: 1–10, 2003.

[103] PSMSL. Permanent service for mean sea level (PSMSL), tide gauge data. `http://www.psmsl.org/data/obtaining`, 2020. (accessed: 22.03.2020).

[104] C. Qin, L. Chen, Z. Cai, M. Liu, and L. Jin. Long short-term memory with activation on gradient. *Neural Networks*, 164:135–145, 2023.

[105] R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing. Technical report, Vienna, Austria, 2022.

[106] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, et al. Improving language understanding by generative pre-training. 2018.

[107] M. Regis, P. Serra, and E. R. van den Heuvel. Random autoregressive models: A structured overview. *Econometric Reviews*, 41(2):207–230, 2022.

[108] C. P. Robert, V. Elvira, N. Tawn, and C. Wu. Accelerating MCMC algorithms. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(5):e1435, 2018.

[109] S. Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint:1609.04747*, 2016.

[110] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[111] E. Santos-Fernandez, J. M. Ver Hoef, E. E. Peterson, J. McGree, D. J. Isaak, and K. Mengersen. Bayesian spatio-temporal models for stream networks. *Computational Statistics & Data Analysis*, 170:107446, 2022.

[112] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[113] A. Sherstinsky. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404:132306, 2020.

[114] M. Shi, J. Ma, and K. Zhang. The impact of water temperature on in-line turbidity detection. *Water*, 14(22):3720, 2022.

[115] C. A. Sims. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*. 1–48, 1980.

[116] K. Smyth and M. Elliott. Effects of changing salinity on the ecology of the marine environment. *Stressors in the Marine Environment*. 161–174, 2016.

[117] Stan Development Team. CmdStanR: an R interface to Stan. R package version: 0.4.0.9001, 2022.

[118] D. J. Stekhoven and P. Bühlmann. Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118, 2012.

[119] T. Stocker. *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press, 2014.

[120] N. Sulaiman, S. I. Zubairi, N. A. Sani, and Z. M. Kasim. The efficacy of treated water from water filtration machines for safe drinking water supply in Bandar Baru Bangi and Kajang, Selangor. *Journal of Food Quality*, 2020:1–9, 2020.

[121] C. Sun, X. Qiu, Y. Xu, and X. Huang. How to fine-tune bert for text classification? In *Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, Proceedings 18, 194–206, October 18–20, 2019*.

[122] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27, 2014.

[123] A. H. Tanim and E. Goharian. Developing a hybrid modeling and multivariate analysis framework for storm surge and runoff interactions in urban coastal flooding. *Journal of Hydrology*, 595:125670, 2021.

[124] T. Tian, A. Merico, J. Su, J. Staneva, K. Wiltshire, and K. Wirtz. Importance of resuspended sediment dynamics for the phytoplankton spring bloom in a coastal marine ecosystem. *Journal of Sea Research*, 62(4):214–228, 2009.

[125] S. Tipirneni and C. K. Reddy. Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16(6):1–17, 2022.

[126] R. K. Turner, D. Burgess, D. Hadley, E. Coombes, and N. Jackson. A cost–benefit appraisal of coastal managed realignment policy. *Global Environmental Change*, 17(3-4):397–407, 2007.

[127] S. Van Buuren. *Flexible imputation of missing data*. CRC press, 2018.

[128] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[129] A. Vehtari, A. Gelman, and J. Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432, 2017.

[130] S. Watanabe. A widely applicable Bayesian information criterion. *arXiv preprint:1208.6338*, 2012.

[131] P. L. Wiberg and C. R. Sherwood. Calculating wave-generated bottom orbital velocities from surface-wave parameters. *Computers & Geosciences*, 34(10):1243–1262, 2008.

[132] C. K. Wikle. Modern perspectives on statistics for spatio-temporal data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(1):86–98, 2015.

[133] P. Woodworth, S. Shaw, and D. Blackman. Secular trends in mean tidal range around the British Isles and along the adjacent European coastline. *Geophysical Journal International*, 104(3):593–609, 1991.

[134] P. Woodworth, M. Tsimplis, R. Flather, and I. Shennan. A review of the trends observed in British Isles mean sea level data measured by tide gauges. *Geophysical Journal International*, 136(3):651–670, 1999.

[135] P. L. Woodworth. A note on the nodal tide in sea level records. *Journal of Coastal Research*, 28(2):316–323, 2012.

[136] G. Wöppelmann, N. Pouvreau, and B. Simon. Brest sea level record: a time series construction back to the early eighteenth century. *Ocean Dynamics*, 56(5):487–497, 2006.

[137] C. Xiao, N. Chen, C. Hu, K. Wang, Z. Xu, Y. Cai, L. Xu, Z. Chen, and J. Gong. A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environmental Modelling & Software*, 120:104502, 2019.

[138] H. Yan, C. Liang, Z. Li, Z. Liu, B. Miao, C. He, and L. Sheng. Impact of precipitation patterns on biomass and species richness of annuals in a dry steppe. *PLoS One*, 10(4):e0125300, 2015.

[139] Y.-H. Yang, F. Zhou, H.-C. Guo, H. Sheng, H. Liu, X. Dao, and C.-J. He. Analysis of spatial and temporal water pollution patterns in lake Dianchi using multivariate statistical methods. *Environmental Monitoring and Assessment*, 170:407–416, 2010.

[140] M. Yu, A. Masrur, and C. Blaszczak-Boxe. Predicting hourly PM2.5 concentrations in wildfire-prone areas using a spatiotemporal transformer model. *Science of the Total Environment*, 860:160446, 2023.

[141] X. Zhang, B. Qian, S. Cao, Y. Li, H. Chen, Y. Zheng, and I. Davidson. INPREM: An interpretable and trustworthy predictive model for healthcare. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 450–460, 2020.

[142] X. Zhang, Q. Zhang, G. Zhang, Z. Nie, Z. Gui, and H. Que. A novel hybrid data-driven model for daily land surface temperature forecasting using long short-term memory neural network based on ensemble empirical mode decomposition. *International Journal of Environmental Research and Public Health*, 15(5):1032, 2018.