Maynooth University
National University of Ireland Maynooth

Bayesian generalised additive models for quantifying sea-level change: Methods and Software

A dissertation submitted for the degree of
Doctor of Philosophy

By:

Maeve Upton

Under the supervision of:

Prof. Andrew C. Parnell

Dr. Niamh Cahill

Hamilton Institute, & ICARUS
National University of Ireland Maynooth
Ollscoil na hÉireann, Má Nuad

August 2023

For my parents, Margaret and Gerard and,

In memory of Teresa Upton, my beloved Granny Tess

Declaration

I hereby declare that I have produced this manuscript without the prohibited assistance of any third parties and without making use of aids other than those specified.

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

Maeve Upton.

Maynooth, Ireland,

August 2023.

Sponsor

This work was supported by a by A4 (Aigéin, Aeráid, agus athrú Atlantaigh) project which is funded by the Marine Institute (grant: PBA/CC/18/01).

Collaborations

Andrew C. Parnell: As my supervisor, Professor Parnell (Maynooth University) supervised and collaborated on the work presented in all chapters.

Niamh Cahill: As my supervisor, Dr. Cahill (Maynooth University) supervised and collaborated on the work presented in all chapters.

Andrew Kemp: Dr. Kemp (Tufts University) contributed to the understanding of the data used throughout this thesis and collaborated on the work in each chapter.

Gerard McCarthy: Dr McCarthy (Maynooth University) assisted with Chapter 3 and Chapter 5 providing insight into the physical phenomena influencing sea-level change.

Erica Ashe: Dr. Ashe (Rutgers University) contributed in Chapter 3 by providing insight into alternative modelling strategies.

Christopher Piecuch: Dr Piecuch (Woods Hole Oceanographic Institution) contributed to the interpretation of the results and the physical processes driving these changes in Chapter 5.

Zoe Roseby: Dr Roseby (Trinity College Dublin) contributed to the collection and understanding of the data used in the real-world application in Chapter 5.

Fermin Alvarez: Fermin (Trinity College Dublin) contributed to the collection and understanding of the data used in the real-world application in Chapter 5.

Robin Edwards: As Fermin's and Zoe's supervisor, Professor Edwards (Trinity College Dublin) collaborated on Chapter 5 by assisting with the collection and understanding the data.

Publications

The chapters contained in this thesis are either published or submitted to peer-reviewed journals. Chapter 1 and Chapter 2 are based on published work from Upton et al. (2023a). Chapter 3 has been submitted to the journal *Journal of the the Royal Statistical Society, Series C (Applied Statistics)*. Chapter 4 has been submitted to the *R journal*. Chapter 5 is in preparation to be submitted to a journal.

Peer-reviewed published chapters:

- Upton, M., Cahill, N. & Parnell, A. C. Statistical modelling for relative sea-level data *Encyclopedia of Quaternary Science, 3rd Edition* (2023). https://doi.org/10.1016/B978-0-323-99931-1.00025-8

Submitted articles (under review):

- Upton, M. C., Parnell, A. C., Kemp A. C., Ashe, E., McCarthy, G. D., & Cahill, N. (2023). A noisy-input generalised additive model for relative sea-level change along the Atlantic coast of North America *Journal of the Royal Statistical Society, Series C (Applied Statistics)*. *arXiv* preprint: https://arxiv.org/abs/2301.09556

- Upton, M. C., Parnell, A. C., & Cahill, N. (2023). reslr: An R package for relative sea level modelling *R-Journal*. *arXiv* preprint: https://arxiv.org/abs/2306.10847

Articles in preparation:

- Upton, M. C., Parnell, A. C., Kemp A. C., Piecuch, C., McCarthy, G. & Cahill, N. (2023). An investigation of regional relative sea-level change for the coastline of the North Atlantic, focusing on the influence of dynamical sea-level change.

Contents

Abstract

Rising sea levels pose significant risks to coastal regions worldwide, and the 2021 Intergovernmental Panel on Climate Change AR6 report emphasised that rates of sea-level rise are the fastest in at least the last 3000 years. To understand historical sea-level trends at regional and local scales, it is crucial to analyse the drivers of sea-level change and their potential impacts. The influence of these different drivers interact at a range of spatial (global, regional, local level) and temporal (annual to millennia) scales. The development of a statistical model that seeks to estimate a number of these characteristics would be of immeasurable value to the sea level and climate impact communities. These characteristics would include: exhibiting flexibility in time and space; having the capability to examine the separate drivers; and taking account of uncertainty.

The aim of our project is to develop statistical models to examine historic sea-level changes for North America's Atlantic coast and extend to the North Atlantic region, incorporating Ireland's coastline. For our models, we utilise sea-level proxies and tide gauge data which provide relative sea level estimates with uncertainty. Proxy data can reconstruct sea-level variations over the late Holocene, spanning the last 2000 years, providing a valuable pre-anthropogenic context for understanding historical relative sea-level changes. We study a range of statistical models used to examine relative sea-level data accounting for uncertainty and varying in space and time. The statistical approaches employed range from simple linear regressions to advanced Bayesian Generalised Additive Models (GAMs), which allow separate components of sea-level change to be modelled individually and efficiently and for smooth rates of change to be calculated.

Our most advanced models are built in a Bayesian framework which allows for external prior information to constrain the evolution of sea-level change over space and time. To investigate the drivers of sea-level change, we use flexible and extended GAMs and effectively account for the uncertainty associated with proxy data using the noisy input uncertainty method. Through the integration of statistical models, proxy data, and tide gauge measurements, our findings reveal a significant rise in current sea levels along North America's Atlantic coast, reaching the highest point in at least the last 15 centuries. The GAMs exhibit a remarkable capability to examine various drivers of relative sea level change, including geological processes (e.g. glacial isostatic adjustment; GIA), local factors, and barystatic influences. Our models provide evidence that GIA primarily drove relative sea-level change along North America's Atlantic coast until the 20th century when a notable rise in the rate of sea-level rise became apparent.

We present the open-source `reslr` package, which serves as a valuable resource for the sea level community, offering a diverse range of statistical approaches. This R package enables Bayesian modeling of relative sea level data, providing a unified framework for loading data, fitting models, and summarising results. By incorporating various statistical models, it offers flexibility and versatility in sea level analysis. Notably, `reslr` takes into account measurement errors associated with relative sea-level data in multiple dimensions, enhancing the accuracy and reliability of the modelling process. With `reslr`, researchers and practitioners can explore and compare different statistical methodologies for a comprehensive understanding of historical sea-level changes, their uncertainties and importantly, the rate of change of these sea-level variations.

One critical driver of sea-level change is ocean dynamics, commonly referred to as dynamic sea-level change. Our statistical methodologies offer valuable insights into dynamic sea-level changes over the last 2,000 years, using both proxy records and tide gauges at a regional level. To investigate the dynamic sea-level component along the North Atlantic coastline, we employ an extended noisy input GAM, effectively decomposing the relative sea-level signal. In our investigation, we focus on two key components of dynamic sea-level change in the North Atlantic: the vertical (Atlantic Meridional Overturning Circulation - AMOC) circulation and

the quasi-horizontal circulation, involving surface-enhanced currents and gyres. Our results highlight a decline in the AMOC over the studied period of 2,000 years with an unprecedented rate of decrease similar to previous studies. Additionally, the quasi-horizontal circulation exhibits increased variability during the same timeframe with a notably difference north and south of Cape Hatteras, USA. This comprehensive analysis sheds light on the complex dynamics driving sea-level changes in the North Atlantic region, contributing to a better understanding of the factors influencing sea-level variations. Our approach places the present alterations in ocean circulation patterns within the extended context of a 2,000-year timeframe. However, the interpretability of these changes is constrained by the resolution of the proxy data.

Acknowledgements

First and foremost, I extend my gratitude to my supervisors, Prof. Andrew Parnell and Dr. Niamh Cahill, whose unwavering support has been instrumental throughout my PhD journey. I am truly thankful for their kindness, generosity, and patience, especially when I sent them endless pages and documents of questions! Andrew and Niamh consistently mentored me on all things research, allowing me to grow as a researcher and become independent and resilient. Their dedication and belief in my potential undoubtedly shaped my PhD journey, and I am genuinely honoured to have had their guidance as an integral part of my academic journey. As supervisors they have demonstrated the importance of compassion and constant support and as individuals, they are true inspirations and exemplary role models, influencing all my future endeavors.

I wish to express my sincere appreciation to the extensive community of faculty and fellow students at the Hamilton Institute, ICARUS, and the Department of Mathematics & Statistics at Maynooth University. Their collaborative endeavours have cultivated an inclusive, constructive, and gratifying learning environment that I am genuinely thankful for. In particular, thank you to Estevão, Hannah, Emma, Alessandra, Alan, Daire, Niamh, Mateus, Anna, Kevin, Eleni, Leah, Cormac, YC, Bill, Dara, Amit, Gabriel, Conor, Nahia, Paddy, Aoife, André, Akash, Ahmed and Shauna who patiently answered all my statistic questions and cheered me up everyday in the Hamilton!

Thank you to Dr. Gerard McCarthy for applying for our funding and forming the A4 team. I am grateful to have been apart of the A4 team. This research group challenged and inspire my work on a weekly basis and was an important support

group throughout my PhD. In particular, thank you to Catherine, André, Levke, Zoe, Fermin, Patrick, Sam H and Sam D. Thanks to the Línte na Farraige team for showing me that science and art go hand in hand.

Thank you to all my collaborators who helped me understand the importance of my work. As previously mention, Dr Gerard McCarthy has been a vital part of my PhD and has assisted in many aspects of my research and I am grateful for his kindness. Dr. Andy Kemp, a fantastic person who was always willing to help at a moments notice, I really appreciated his generosity and guidance. To Dr Erica Ashe who was a fantastic support to my research (and Charlie for being a cute doggo), thank you. Dr Chris Piecuch, whose enthusiasm, extensive knowledge and work ethic inspires me each day (and of course Penny and Lance for also being cute doggos). Prof Robin Edwards, thank you for your insights into all things sea level and for supporting me in Ballina!

I was fortunate to take part in fieldwork in Prince Edward Island in Canada, and I can confidently say that it significantly boosted my understanding by the end of my third year. It helped me place my research in a larger context and see the bigger picture. The PEI crew of Andy, Chris, Hannah, Robin, Andrea, Kelly, Fermin, Zoe and Emmanuel created a wonderful atmosphere and are a fantastic group of individuals. To all the people I have met at various conferences over the past four years and the sea-level community as a whole, I am truly honoured to have worked with you and to have met such a dynamic and enthusiastic group of people. At each conference and online meeting, I was inspired by your dedication and your commitment to top quality science.

A massive thank you to all my friends who showed my that my life wasn't just my PhD! Special mention to the Physics Peeps, the Fizzy Luas Boeurs, the Irish dancing ladies, the PwC crew and the Mayo/Sligo gals who make me laugh every second of the day. Thank you to my partner in crime, Aisling for encouraging me to take the PhD leap! Thank you to Madeleine and family for your generosity. Thank you to Carmel and Alison for your constant support. Thank you to Annmarie for your generosity and thoughtfulness, especially when putting up with rants and fashion shows in Glasnevin for many years. A huge thank you to my long suffering

housemate Mark O' Sullivan who painstakingly proof read this thesis. I don't ever think I will be able to repay you!

I would like to express my heartfelt gratitude to my parents, Margaret and Gerard, for their unwavering love, support, and sacrifices that have helped me get over the line time and time again. Their constant encouragement, guidance, and belief in my abilities have been crucial throughout this journey. Thank you for being my biggest supporters and laughing, fighting, listening, chatting and crying with me. Thank you to my brothers, Rory (and Caragh) and Cillian, for constantly keeping me on my toes and always keeping me grounded. Thank you to my Granny Tess and Granny Kenny, two strong independent women - true inspirations. Thank you to all my aunties, uncles, cousins and extended family as it has taken a village to raise me! To Darcy, Puca Dubh and Benji, thank you for being the best doggos.

My one and Eoghanie, I am truly blessed to have you in my life. Thank you for being my partner, my adventure buddy and my rock.

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation

Climate change is defined as a change in our climate over time due to natural variability or anthropogenic influences (IPCC, 2007). The latest IPCC AR6 report emphasises that human activities, particularly the release of greenhouse gases, are responsible for global warming (IPCC, 2023). As a consequence, our climate has experienced the most rapid rise in global surface temperature over the past 50 years compared to any other 50-year period in the last 2000 years (IPCC, 2023). With rising global temperatures, every region in the world has experienced an increase in extreme weather events such as heatwaves, heavy precipitation, droughts and tropical cyclones, which has resulted in agricultural damage, ecological losses and increased human mortality (Masson-Delmotte et al., 2021). Thus, climate change is the overarching challenge of the 21st century that encompasses various domains, including science, economics, health, sociology, and politics (Maslin, 2019).

Sea level change is a visual and tangible example of climate change and it has been shown, with "high confidence", that global mean sea level (GMSL) is rising and accelerating (Oppenheimer et al., 2019). Rates of GMSL rise between 2006 and 2015 were 3.6 mm/yr, which is 2.5 times the rate recorded between 1901 and 1990

(Masson-Delmotte et al., 2021). Currently, more than 800 million people globally (Reimann et al., 2023) are living in the low-elevation coastal zone (less than 10m above sea level; Hauer et al., 2021) with this number predicted to rise to over 1 billion in the future (Neumann et al., 2015; Reimann et al., 2023). With this in mind, the impacts of rising sea levels could have devastating consequences for coastal communities, for example increased coastal erosion (e.g., Douglas et al., 2000), coastal flooding (e.g., Finkl, 2013; Spalding et al., 2014) or storm damage due to extreme sea levels (Tebaldi et al., 2021). The future of our coastlines is unknown, yet understanding what has happened in the past is crucial for predicting future sea-level rise.

Our understanding of recent global sea-level rise has improved dramatically over the last thirty years with the advent of satellite data. In the late 1990s, the launch of the TOPEX/Poseidon satellite ushered in the modern satellite era, enabling direct measurements of global sea surface height to high degrees of accuracy (Legeais et al., 2018; Biferno et al., 2023). Prior to the modern satellite era, instrumental data collected by tide gauges provided the primary source of information regarding sea-level changes. The earliest continuous sea-level observations can be traced back to the late 1600s (Wöppelmann et al., 2006), while the advent of automatic tide gauge records commenced in the 19th century (Woodworth et al., 2011). Tide gauge records offer vital information on coastal sea level changes relative to the ground, although they exhibit a spatial bias towards northern hemisphere regions (Cazenave et al., 2022). The instrumental data from satellites and tide gauges primarily capture changes within the relatively recent anthropogenic periods, limiting our understanding of long-term sea-level variations (i.e. prior to the 1700s).

To address the temporal limitations of instrumental data, researchers turn to the use of proxies. Proxies are observable features, whether physical, biological, or chemical, that enable estimates of indirectly measurable variables (e.g., Murray-Wallace and Woodroffe, 2014). Over the past four decades, significant advancements in proxy data collection and processing techniques have greatly improved our understanding of past relative sea-level changes from the Common Era (CE: years between 0 and 2000) (e.g., Van De Plassche, 1986; Shennan et al., 2015). However there are limitations when reconstructing Common Era sea-level changes

using proxy records, in particular measurement uncertainties for the observations (e.g., Barlow et al., 2013). Consequently, specific statistical modelling strategies are necessary to investigate the spatio-temporal evolution of sea-level change while accounting for associated uncertainties. Recently, the variety of statistical models available to the sea-level community has grown (e.g., Kopp et al., 2009; Church and White, 2011; Cahill et al., 2015a; Dangendorf et al., 2017; Kopp et al., 2016; Walker et al., 2021). With these statistical advancements and data sources, it has been shown the rate of sea-level rise today is at its fastest for the last three millennia (Masson-Delmotte et al., 2021).

Sea-level change is not equal at every location and many regions deviate drastically from the GMSL (Cazenave and Llovel, 2010). As a result, researchers are interested in investigating physical processes impacting sea-level change at different spatial scales (Kopp et al., 2015; Brown et al., 2016). The increased availability of data and progress made in statistical modelling has allowed researchers to examine different physical drivers impacting global, regional and local changes in sea level over the Common Era (e.g., Shennan et al., 2006; Walker et al., 2021). Yet, these statistical models are impacted by computational limitations which have affected their wide-spread usability (Ashe et al., 2019). Overcoming these challenges will allow further insight into the drivers of sea-level change.

The outputs derived from statistical models in sea-level research not only deepen our understanding of sea-level fluctuations but also empower policymakers and stakeholders to develop adaptive strategies, make informed policy decisions, and implement effective mitigation measures to address the challenges posed by sea-level rise (IPCC, 2023). By combining scientific knowledge with proactive policy-making and resilient strategies, we can work towards a sustainable and secure future in the face of sea-level changes (O'Brien et al., 2012).

1.2 Relevant Statistical Techniques

Statistical models have been used to examine various aspects of our climate and are vital tools for understanding the changes that occur in our climate over various timescales and locations (e.g., Von Storch and Zwiers, 2002; Flato and Marotzke,

2013). For sea-level research, a variety of models have been developed using both parametric and non-parametric techniques. The outputs of these models can assist the sea-level community when examining changes in sea level, rates of these changes and potential causes for these changes in time and space.

In this section, we discuss the statistical modelling techniques that are used by the sea-level research community, as well as mentioning the extensions that we have implemented. Our discussion here is mostly descriptive, with the full technical specification provided in later chapters. In general we aim to estimate the true sea level which we define as $f(\boldsymbol{x}, t)$ where $\boldsymbol{x}$ represents space and t time, though much of this section is concerned with time only for which we drop the dependence on location. For each statistical model in this thesis, a Bayesian hierarchical framework is employed and Markov Chain Monte Carlo (MCMC) simulations are used to provide estimates for unknown parameters with associated uncertainties. The tool used to conduct MCMC simulations is Just Another Gibbs Sampler (JAGS; Plummer, 2003), which is implemented in R using the `rjags` package (Plummer et al., 2016). In the final components of this section, we provide a brief insight into the two uncertainty methods we employ and provide a rationale for their selection.

1.2.1 Linear Regression

A linear regression provides a line of best fit that, in sea level research, provides a rate of change of sea level over time. Mathematically, it is given by:

$$f(t) = \alpha + \beta t \tag{1.1}$$

where α is the intercept and β is the slope. Previous studies undertaken by the sea-level community have used this technique to assess the rate of relative sea level change over the past 4,000 years, however, measurement uncertainties (which are common to proxy records) were not incorporated (Shennan and Horton, 2002; Engelhart et al., 2009). In Chapter 4, we use a temporal linear regression method to estimate such rates of change in relative sea level (RSL). The simplistic interpretation of the results of this technique have contributed to its popularity. However, the rigidity of the assumption of linearity may be unrealistic when examining long-term historical trends in RSL change.

1.2.2 Change Point Models

The change point (CP) model is a modification of the linear regression that employs piecewise linear segments and determines the location or timing of trend changes in the data (Carlin et al., 1992; Cahill et al., 2015b). The multiple CP model, $f(t)$ is described as:

$$f(t) = \begin{cases} \alpha_1 + \beta_j(t - \lambda_1) \text{ when } j = 1, 2, \\ \alpha_{j-1} + \beta_j(t - \lambda_{j-1}), \text{ when } j = 3, ..., m + 1 \end{cases} \tag{1.2}$$

where α_j is the expected value of the response at the jth CP. λ_j is the time at which the CP occurs with the prior restriction that $\lambda_1 < \lambda_2 < ... < \lambda_m$ and m is the number of CPs (Cahill et al., 2015b). β_1 and β_{m+1} are the slopes before and after the first and last CP respectively. β_j for $j = 2 \ldots, m$ are the slopes between the $(j-1)$th and jth CP. β_1 and β_{m+1} are estimated as part of the model and are free parameters given prior distributions (Carlin et al., 1992). The remaining β_j are deterministically calculated since the function is continuous and the neighbouring linear segments must join together (Cahill et al., 2015b) such that:

$$\beta_j = \frac{\alpha_j - \alpha_{j-1}}{\lambda_j - \lambda_{j-1}}, j = 2, ..., m. \tag{1.3}$$

Change point modelling has been used in various sea level studies. For instance, the timing and magnitude of recent accelerated sea-level rise in North Carolina used proxy records and a change point modelling strategy (Kemp et al., 2009). In addition, change point modelling was used to examine instrumental records. One such example being Hogarth et al. (2020) where tide gauge records since 1958 for the British Isles were extended through data archaeology, and then change point models were applied to achieve more consistent estimates of sea level rise. Building on this work, Hogarth et al. (2021) further extended the technique to analyse changes in mean sea level around Great Britain over the past two centuries.

In Chapter 4, we explore the implementation of CP modelling within the `reslr` package. This technique allows for the estimation of abrupt changes in RSL. However, a drawback of many CP modelling techniques is the need to determine the number of change points in advance.

1.2.3 Gaussian Process

Rasmussen and Williams (2006) define a Gaussian process (GP) as a generalised multivariate Gaussian distribution where each individual random variable is normally distributed and their collective distribution also follows a Gaussian distribution. The GP can be described in the following way:

$$f(t) \sim GP(\mu(t), k(t, t')) \tag{1.4}$$

where $\mu(t)$ is the mean function and $k(t, t')$ is the covariance function. The covariance function indicates the relationship between outcome variables. The proximity of input variables (t and t') strengthens the association between their respective outcomes, whereas increasing distance between the input variables results in a weaker relationship (Rasmussen and Williams, 2006).

This modelling strategy has been used extensively in the sea-level community to examine the evolution of sea-level change in time and space and examining potential underlying processes causing sea level change (Kopp et al., 2009; Kopp, 2013; Kopp et al., 2016; Kemp et al., 2018; Walker et al., 2021). In Chapter 3, we discuss the use of GPs when examining the drivers of RSL change and highlight the requirement for a more efficient model as the computational burden associated with GPs grows in proportion with the cube of the number of data points.

In Chapter 4, we demonstrate an extension to the Gaussian Process called the Integrated Gaussian Process (IGP) which allows for the derivative of the curve to be directly estimated, providing an estimate for rate of change of RSL (Holsclaw et al., 2013). Previous research has demonstrated its success for a range of single site proxy records (e.g. Hawkes et al., 2016; Kemp et al., 2017; Shaw et al., 2018; Dean et al., 2019; Stearns et al., 2023; Kirby et al., 2023), however, a more efficient strategy is required for spatio-temporal modelling.

1.2.4 Splines

Splines are versatile mathematical tools used in various applications, including interpolation and data smoothing. In our research, we specifically examine two types of splines: B-splines (de Boor, 1978; Dierckx, 1995) and P-splines (Eilers and Marx,

1996). B-splines construct smooth curves by connecting piecewise polynomials at specific points called knots along the horizontal axis. Any degree of polynomial can be used provided an appropriate number of knots are used (de Boor, 1978; Dierckx, 1995; Cox, 2006). Mathematically, B-splines are described by:

$$f(t) = \sum_{k=1}^{K} b_k(t)\beta_k \tag{1.5}$$

where $b_k(t)$ is the spline basis function and β_k is the spline coefficient.

P-splines by contrast aim to overcome the difficulty of choosing the number of knots (Eilers and Marx, 1996) by constraining the smoothness of the spline via the spline coefficients. The penalty is constructed based on differences (of order d) between spline coefficients. A parameter applied to the penalty term controls the trade-off between model accuracy and smoothness. In Chapter 4, we demonstrate the P-spline technique when examining RSL change in time for a single location and the corresponding rate of change.

In our research, we extend the B-spline approach to account for both space and time. We form B-spline basis functions for each individual covariate; time, longitude and latitude (Wood, 2017a). In order to capture the variability over time and space, a tensor product is used to combine these individual basis functions (Wood, 2006). We use a spatio-temporal spline in Chapter 4 to examine sea level over multiple locations and time periods.

1.2.5 Generalised Additive Models

Generalised additive models are generalised linear models with linear predictors that depend "linearly on the sum of smooth functions of the predictor variable" (GAMs; Wood, 2017a, p. 161). It is constructed using a basis expansion for each smoother and an associated variable which penalises the function's smoothness as described in Hastie and Tibshirani (1990). GAMs allow for a combination of smoother functions along with random effects which allows for models to be carefully specified relevant to the problem at hand.

GAMs allow us to identify variations in sea-level across different spatial and temporal scales, instead of focusing solely on specific underlying physical processes.

The model's components have the ability to capture multiple underlying processes which can vary depending on the specific location and timeframe under study. These processes may occur simultaneously and interact in complex ways, potentially obscuring or amplifying the influence of other factors.

We use GAMs in Chapter 3 to understand the physical processes driving RSL change over the last 3,000 years. In Chapter 4, the `reslr` package enables the paleo sea-level community to use this technique for future sites. Chapter 5 demonstrates the capabilities of this technique on large datasets. Also, in Chapter 5 we demonstrate the extendable nature of this model which allows for a comparison between RSL changes at a regional level across the North Atlantic.

1.2.6 Uncertainty Methods

In our analysis, we use a combination of data sources from instrumental and proxy records. The data sourced from proxies contain bivariate uncertainties resulting from measurement error associated with the response variable (RSL) and with the input variable (time). In this section, we describe the two uncertainty methods employed in our research and provide a rationale for their selection.

1.2.6.1 Errors-in-Variables Uncertainty Method

The Errors-in-variables (EIV) method, proposed by Fuller (1987) and formulated by Dey et al. (2000), considers the input variable to be measured as an error-prone substitute, thereby accounting for potential errors in the input variable. This is in contrast to the conventional assumption of error-free input variables. The EIV model assumes that the input variable time, $\tilde{t}$, is given:

$$\tilde{t} = t + \epsilon_t \tag{1.6}$$

where t is the true unknown values for the input variable and $\epsilon \sim \mathbb{N}(0, s_t^2)$ with s_t^2 is the known measurement error variance associated with the input variable. Whilst the method can be applied to any problem with input noise, we apply it specifically to the problem of input time error when examining relative sea level changes. The EIV method is applied to linear regression, change point and integrated Gaussian process models presented in Chapter 4. Whilst this technique directly accounts

for the uncertain input process, the extra statistical modelling steps can lead to substantially increased computation time.

1.2.6.2 Noisy Input Uncertainty Method

An alternative to the EIV method is the Noisy Input method which also allows for data where the input and output measurements are corrupted by noise (McHutchon and Rasmussen, 2011). The Noisy Input (NI) method was originally introduced within a Gaussian Process model, yet, we demonstrate its capabilities for other process models. The NI method assumes the response variable y to be a noisy measurement $\tilde{y}$ given by:

$$\tilde{y} = y + \epsilon_y \tag{1.7}$$

where y is the true response and the error term is given by $\epsilon_y \sim \mathbb{N}(0, s_y^2)$ with s_y being the known measurement standard deviation of the response. The input measurements are assumed to be a noisy estimate, $\tilde{t}$ of the true time value t given as:

$$\tilde{t} = t + \epsilon_t \tag{1.8}$$

where the error term is given by $\epsilon_t \sim \mathbb{N}(0, s_t^2)$ and s_t is the known measurement error associated with the input variable. As a result, a function for the response variable with input noise is formed in the following way:

$$y = f(\mathbf{x}, \tilde{t} + \epsilon_t) + \epsilon_y \tag{1.9}$$

McHutchon and Rasmussen (2011) describes a three step process to account for these uncertainties: (1) the model is fitted without input noise; (2) the slope of the posterior mean is calculated and a corrective variance term is calculated; (3) the model is re-fitted with this corrective variance term added as an additional model error term, in other words, inflating output variance to overcome these noisy inputs.

The advantage of the NI approach is that it maintains a consistent level of noise regardless of whether the measurement is treated as an input or an output. This ensures that all the data contributes to informing the input noise variance, thereby aligning with output expectations and reducing the potential for overfitting (McHutchon

and Rasmussen, 2011). For this reason, we use this approach in our new modelling approach as described in Chapter 3 and demonstrated in Chapter 4 and Chapter 5.

1.3 Thesis Outline

The remainder of this thesis is organised as follows. In Chapter 2, we provide a detailed definition of relative sea level and describe the physical processes that cause relative sea level to vary in time and space. In addition, we provide insight into the data collection techniques used by the sea-level community. This chapter is based on the work presented in Upton et al. (2023a).

In Chapter 3, we address the main challenges encountered by the sea-level community, including dealing with data complexities and uncertainties, the confounding nature of the different relative sea-level components, and computational inefficiencies. We propose an additively decomposed Bayesian spatio-temporal model that has the capability to identify different drivers of relative sea-level change at regional and local levels. The model estimates multiple components of the relative sea-level field using a combination of splines and random effects, and accounts for the bivariate uncertainties associated with the proxy input data using the noisy-input uncertainty method. The outputs of our model provide insight into the dominant factors influencing RSL change along the densely populated Atlantic coast of North America over the past 3,000 years. Notably, our results emphasise that RSL has been increasing at an unprecedented rate during the 20th century.

Chapter 4 delves into a comprehensive discussion on the implementation of our R package, called `reslr`. We provide a thorough explanation of the design choices that govern the package architecture and enhance the interpretability of its key features. The `reslr` package facilitates the examination of relative sea-level change by employing a variety of Bayesian statistical models, while also accounting for measurement errors inherent in data derived from proxy records. To ensure usability for a wide range of users, regardless of their level of experience with R, we have curated a minimal set of functions. We illustrate the capabilities of the package through two case studies. The first case study showcases the temporal

evolution of RSL at a single location, including the rate of RSL change. The second case study expands upon this analysis by incorporating multiple locations, demonstrating how the package can effectively decompose the RSL signal.

In Chapter 5, we discuss the impact of ocean dynamics on RSL, which is referred to as dynamic sea-level change. We focus on ocean circulation patterns in the North Atlantic region due to their significant relevance in the field of global climate dynamics. In particular, we examine two components: (1) the Atlantic meridional overturning circulation (AMOC) and (2) the quasi-horizontal circulation pattern consisting of gyres and other surface-enhanced currents. We use proxy records and instrumental data from tide gauges to understand how dynamic sea level evolved in the North Atlantic over the last 2,000 years. Applying an extended noisy input generalised additive model, we explore the changes that occur in these two components of ocean circulation, focusing on two sub-regions: - the east west sub-region of the North Atlantic - and the south north sub-region of Cape Hatteras, North Carolina, USA. Our results demonstrate that the AMOC has been gradually declining, however, since the late 1850s, the rate of decline has dramatically increased. Over the past 2,000 years, quasi-horizontal circulation patterns have displayed substantial fluctuations, highlighting distinct dynamic sea-level variabilities on either side of Cape Hatteras, USA. In both scenarios, the resolution of our proxy records has posed limitations on the interpretation of our results, especially when considering shorter timeframes.

Finally, in Chapter 6, we provide a comprehensive conclusion to the thesis, summarising the key findings and insights obtained throughout our research. Additionally, we identify and discuss potential avenues for future research. All proposed methods in this thesis were implemented using the R (R Core Team, 2021) software and are accessible on the author's Github[1] via three public repositories. The repositories `NI-GAM`, `reslr`, and `RSL-NorthAtlantic` are related to Chapters 3, 4, and 5, respectively. R scripts are made available such that the analyses and plots presented throughout this thesis are reproducible. In addition, all datasets are available, either through R packages, which are presented in the R scripts, or files in the aforementioned repositories.

[1] https://github.com/maeveupton

CHAPTER 2

Introduction to Relative Sea-Level Change

Throughout this thesis, we examine relative sea-level changes using a variety of statistical models and data sources. In this chapter, we describe the physical processes that can cause these changes which result in different responses on spatial and temporal scales. In addition, we provide insight into the different data collection techniques and their associated uncertainties. The reader can refer back to this chapter to obtain definitions of geological terms which are used in the statistical models presented in later chapters.

2.1 Relative Sea Level (RSL)

The sea surface is constantly changing with some perturbations caused by the wind producing waves or storms, gravitational forces resulting in tides or atmospheric influences giving rise to ocean currents (Murray-Wallace and Woodroffe, 2014). Mean sea level is the average height of the sea surface, measured relative to the centre of the Earth, over time and space (Church et al., 2013). On a global scale, it is known as global mean sea level (GMSL). Relative sea level (RSL) is the height of

the ocean surface measured with respect to the Earth's sediment surface as shown in Figure 2.1 (Church et al., 2013).

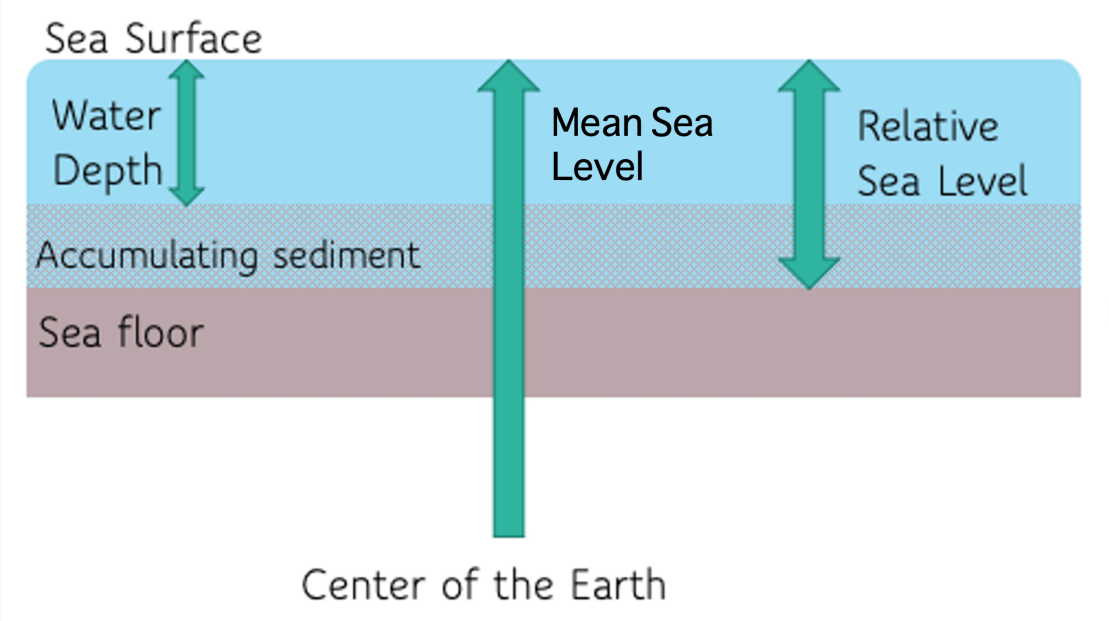


Figure 2.1: A schematic diagram for the relative sea level definition.

2.2 Drivers of RSL Change

RSL is the resultant effect of physical processes that can alter the height of the land and/or sea surface. The importance of specific processes varies markedly across space and through time giving rise to a complex and evolving pattern of RSL change (e.g. Warrick et al., 1993; Milne et al., 2009).

Shennan and Horton (2002) identified factors that can influence RSL changes ($\Delta\xi_{rsl}(\tau,\psi)$) over time ($\tau$) and space ($\psi$). Building upon their conceptual definition, we have expanded the framework to incorporate additional components relevant to our research:

$$\begin{aligned}\Delta\xi_{rsl}(\tau,\psi) = \Delta\xi_{bar}(\tau,\psi) + \Delta\xi_{thermo}(\tau,\psi) + \Delta\xi_{iso}(\tau,\psi) + \\ \Delta\xi_{dyn}(\tau,\psi) + \Delta\xi_{local}(\tau,\psi).\end{aligned} \tag{2.1}$$

$\Delta\xi_{bar}(\tau,\psi)$ demonstrates barystatic influences and in previous literature (e.g. Fairbridge, 1961; Lambeck et al., 2004) the term "eustatic" was commonly employed to

encompass barystatic influences. However, Gregory et al. (2019) discussed the diverse interpretations associated with the term "eustatic" and its subsequent omission from recent IPCC reports and from this thesis (e.g., Church et al., 2001; IPCC, 2007; Church et al., 2013). $\Delta\xi_{thermo}(\tau, \psi)$ represents thermosteric changes and $\Delta\xi_{iso}(\tau, \psi)$ represents isostatic adjustment. $\Delta\xi_{dyn}(\tau, \psi)$ represents dynamic sea-level and $\Delta\xi_{local}(\tau, \psi)$ represents local changes that can be impacted by factors such as tidal range changes and sediment compaction. Figure 2.2 provides visualised representation of these components adapted from Shugar et al. (2014).

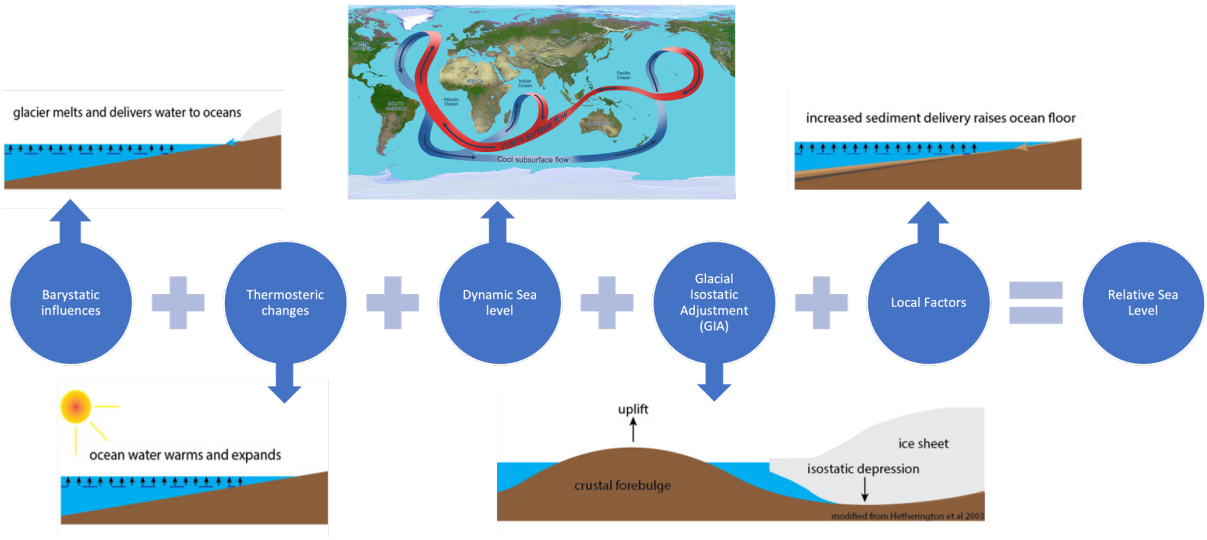


Figure 2.2: A list of the various physical processes that can impact relative sea level change (Shugar et al., 2014; Dunbar, 2010).

These processes can occur concurrently and in varying directions, which can obscure or magnify the contributions from other factors. In the upcoming sections, we describe these mechanisms, commencing with global and regional-scale processes, followed by local-scale processes. We emphasise the significant processes identified in our research. In our more complex statistical model, our objective is to deconstruct the relative sea-level (RSL) signal into components that represent these diverse processes. However, given their intricacy, we often perform this decomposition at an aggregated level. For instance, isostatic adjustment may have global ramifications, while local factors are subject to site-specific variations.

2.2.1 Barystatic Changes

Barystatic changes are influenced by the transfer of mass between land-based ice and oceans, where the growth or decay of ice causes global mean sea level to rise or fall (Gregory et al., 2019). The transfer of mass between land-based ice and the ocean drives RSL change. Ice melt returns mass to the ocean as liquid water, which causes a rise in global mean sea level (Leuliette, 2015). This contribution can vary in magnitude (and indeed sign) across all timescales, but is common to all locations. In addition, barystatic sea-level changes may result from anthropogenic influences, for example land-water storage, where water is retained on land by dams, reduces the transfer of water to the oceans, whereas, ground water pumping processes can accelerate the water transfer from the land to ocean (Church et al., 2013).

In Chapter 3, we highlight the impacts of barystatic changes along the Atlantic coast of North America and how they evolved over the past 3000 years on a regional scale. In Chapter 5, we extend the spatial scale to incorporate sites along the North Atlantic coastline and investigate how barystatic changes varied over a larger region.

2.2.2 Thermosteric Changes

Thermosteric contributions are influenced by changes in global temperature creating density variations within our oceans resulting in sea-level changes (Antonov et al., 2005). The magnitude of this contribution may differ across various timescales, but it remains consistent across all locations. With rising global temperatures, 90% of this additional heat is absorbed by our oceans (Masson-Delmotte et al., 2021). Changes in temperature alter the density of ocean water resulting in sea-level change. When water warms it becomes less dense and rises, while when water cools it becomes more dense and falls. Since water can move horizontally this thermosteric contribution can drive a change in global mean sea level (Grinsted, 2015).

Chapter 3 sheds light on the consequences of thermosteric shifts along the Atlantic coast of North America, delving into their historical evolution over the past 3000

years within a regional context. Expanding the investigation in Chapter 5, we extend the spatial scale to encompass sites along the North Atlantic coastline, aiming to uncover the varied dynamics of thermosteric changes across a broader and more expansive region.

2.2.3 Isostatic Adjustment

Isostatic adjustment is the gradual response of the Earth, the gravitational field and the oceans in response to changes in surface loads, for example ice sheets (Sleep, 2002). This phenomenon is particularly relevant in the context of glacial isostatic adjustment (GIA), which specifically deals with the response of the Earth's crust to the formation and melting of ice sheets (Whitehouse, 2018). The process displays both temporal and spatial variations attributed to the gradual formation and melting of ice sheets, as well as proximity to the ice sheet.

Throughout the Last Glacial Maximum, which occurred approximately 25,000 years ago, ice sheets covered significant portions of the Earth's surface (Peltier, 1999). The weight of these ice sheets caused the land beneath them to sink, a process known as subsidence, while simultaneously leading to the uplift of outlying regions, forming what is called a peripheral forebulge (PB; Rovere et al., 2016).

During deglaciation, the isostatic response was reversed and previously subsided areas began to uplift, resulting in a fall in RSL (Peltier, 1999). At the same time, the PB started to collapse and retreat, contributing to a rise in RSL. Additionally, as the mass of the melted ice was redistributed from being concentrated on land to being dispersed in the ocean, the geoid (the shape of the Earth's gravitational field) was altered (Fjeldskaar, 1994). The aspects of GIA are visually represented in Figure 2.3 as sourced from Whitehouse et al. (2021). Crucially, the process of GIA continues for several thousand years after deglaciation and is ongoing today with a characteristic spatial pattern (Engelhart et al., 2009).

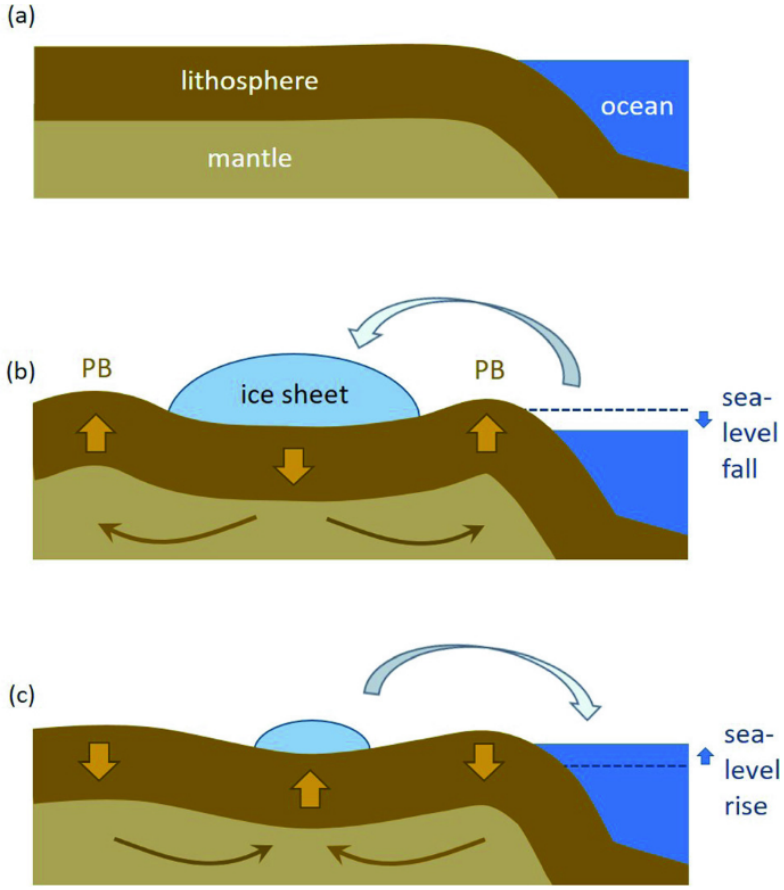


Figure 2.3: A schematic diagram of the physical process of glacial isostatic adjustment sourced from Whitehouse et al. (2021).(a) Equilibrium state of the Earth without any ice sheet. (b) The ice sheet causes land underneath to subside and uplifted regions form peripheral forebulges (PB) resulting in a fall in sea level. (c) After the ice sheet melts, sinking areas begin uplifting and the peripheral forebulges fall resulting in a rise sea level.

Physical models, known as Earth-ice models (e.g. Peltier, 2004; Caron et al., 2018), predict these broad global spatial GIA patterns, although rates of change at specific sites are sensitive to choices of Earth parameters such as mantle viscosity and ice histories (Peltier, 2004). These patterns are confirmed using empirical data in the form of coarse-resolution RSL reconstructions (e.g. Vacchi et al., 2016), measurements by permanent global positioning system (GPS) stations (e.g. Ivins et al., 2013), and tide gauges (e.g. Davis and Mitrovica, 1996).

2.2.4 Spatial Fingerprint of Sea-Level Change from Ice Sheets

While melting of land-based ice raises global mean sea level it also results in a characteristic spatial fingerprint of sea-level change with locations proximal to the melting ice sheet experiencing less rise (or even a fall) than the global mean, while distal locations experience rise in excess of the global mean (Jeon et al., 2021). The difficulty in identifying spatial fingerprints for specific locations, such as Greenland in the North Atlantic region, stems from the limited availability of data spanning multiple decades (Coulson et al., 2022). As a result, we do not focus on the impacts of this process in this thesis.

2.2.5 Dynamic Sea-Level Change

Ocean circulation systems transport water, heat, dissolved gases and nutrients around the globe and are vital component of our climate system (Pörtner et al., 2019). Dynamic sea-level change arises from the redistribution of ocean mass caused by shifts in prevailing patterns and strength of atmospheric and oceanic circulation (e.g., Lowe and Gregory, 2006). It refers to mean sea-level changes above the geoid, which represents the theoretical shape of the Earth's surface influenced solely by gravity, and is a consequence of ocean dynamics (Gregory et al., 2019). This process causes RSL to vary at non-linear rates among regions on decadal to multi-century timescales (Suzuki and Tatebe, 2020).

The ocean circulation patterns around the globe are complex and possess a variety of component that influence dynamic sea-level changes. In this research, we focus on the North Atlantic Ocean which is a vital regulator for our global climate (Rhines et al., 2008). The North Atlantic possesses active deep water formation sites which allow vertical mixing of waters resulting the transfer of heat, nutrients, dissolved gasses and other vital components throughout the ocean layers (Rahmstorf, 2006). The extent of the ocean circulation system in the North Atlantic is beyond this research. Instead, we focus on two key components that impact dynamic sea-level changes. The first is the Atlantic Meridional Overturning Circulation (AMOC) which is a system of currents that redistributes heat from the

equatorial regions towards the poles, as demonstrated in Figure 2.4, and is driven by deep water formation sites in the North Atlantic (Kuhlbrodt et al., 2007).

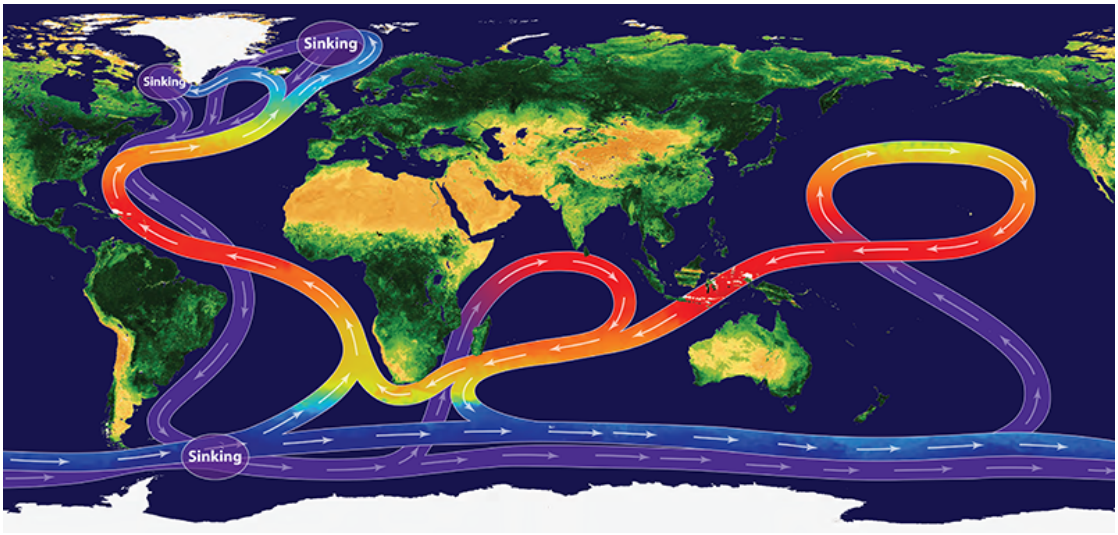


Figure 2.4: Global ocean circulation pattern known as the "great ocean conveyor" is crucial for transporting water, heat, nutrients and dissolved gases (Broecker, 1991). It is moves warm salty surface water northwards and returns cold dense water to the deep ocean (Sourced from National Oceanography Centre: NOC, 2014).

The second type of oceanic circulation that we focus on is quasi-horizontal circulation patterns. These quasi-horizontal circulation patterns encompass large-scale surface currents influenced by the gyres and other surface-enhanced currents. Gyres, circular rotating systems of ocean currents, are prone to forming in major ocean basins due to the combined effects of wind patterns and the Coriolis effect, which arises from the Earth's rotation (Law et al., 2010). While, surface-enhanced currents refer to oceanic currents that occur primarily in the upper layer of the ocean and are influenced by wind patterns, temperature differences, and Coriolis effect. Examples of surface-enhanced currents in the North Atlantic include the Gulf Stream, the North Atlantic Drift, and the Canary Current. The interaction of these gyres and surface-enhanced currents plays a crucial role in regulating the transfer of heat and energy within the North Atlantic Ocean and the adjacent regions as demonstrated in Figure 2.5.

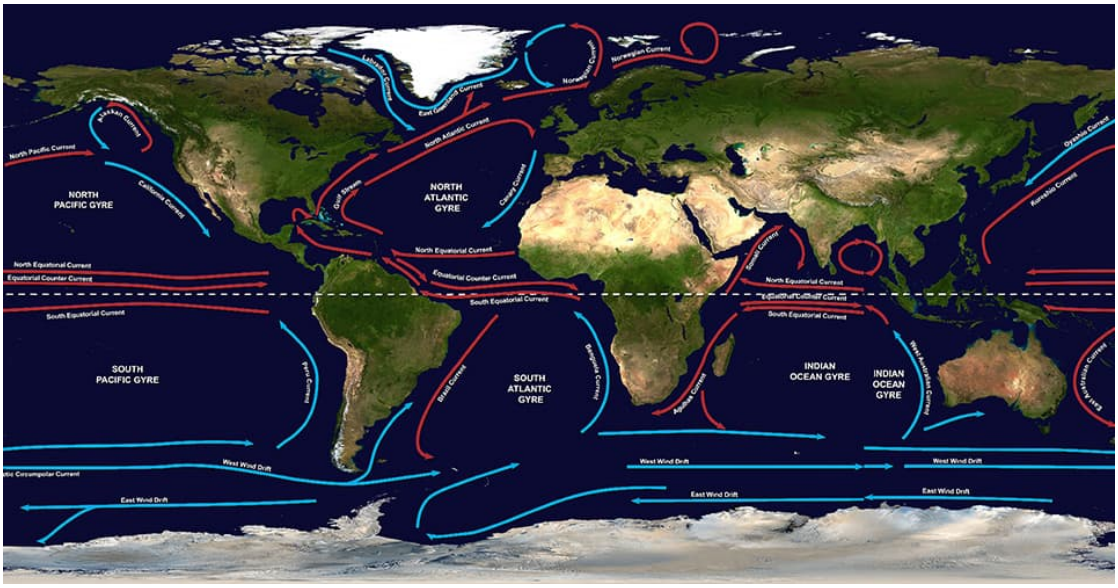


Figure 2.5: A map of the world's five major gyres (circular rotating systems of ocean currents) and surface-enhanced currents (e.g. Gulf Stream) influence global climate patterns and marine ecosystems through the transport water, heat, nutrients and dissolved gases (Sourced from National Oceanic and Atmospheric Administration: NOAA, 2018)

In Chapter 5, we discuss dynamic sea-level changes in conjunction with these oceanic circulation patterns in the North Atlantic. We use our statistical modelling approach to address the impacts of changing ocean currents on various locations along coastline of the North Atlantic and the Atlantic coast of North America over the past 2,000 years.

2.2.6 Sediment Compaction

Changes in sediment volumes can impact the height of the Earth's surface through the process of sediment compaction (Horton et al., 2018). Sediment compaction refers to a variety of natural processes that occur during and after sediment deposition in low-lying coastal areas, leading to a decrease in sediment volumes and land-level lowering (i.e. subsidence) (Kaye and Barghoorn, 1964; Allen, 2000; Brain, 2016). This subsidence is the result of mechanical compression processes that occur over time and as a result of stress, reducing the pore space within the sediment and increasing its bulk density (Brain et al., 2011).

Sediment compaction processes are important for developing effective coastal risk management strategies as they can trigger rapid and extensive alterations in coastal geomorphology resulting in RSL changes at a local level (Horton et al., 2018). The rates at which compaction induces RSL change differs based on spatial and temporal factors in response to a combination of natural processes and human activities that accelerate these processes (Brain, 2016).

In Chapter 3, sediment compaction is captured in a non-linear local component which identifies site -specific RSL changes along the Atlantic coast of North America. It is important to note that in many cases, reconstructions of historic RSL are not primarily aimed at understanding local-scale changes. Therefore, estimating the local-scale component is typically a way to isolate the contribution of processes occurring at larger spatial scales.

2.2.7 Tidal Range Changes

Tidal range is the difference between high and low water marks and changes in tidal ranges are considered to be site specific (Dipper, 2022). Ocean depths or changes in coastline can influence sedimentation in coastal areas which in turn impacts their tidal ranges over centuries to millennia (Hill et al., 2011; Hall et al., 2013; Walker et al., 2021). In addition, human activities, including dredging, deforestation or loss of wetlands, can impact tidal range processes on shorter timescales (Mawdsley et al., 2015).

As previously mentioned, the focus of Chapter 3 is to decompose the RSL signal into different components that vary in time and space and so will be explored further there. Local-scale RSL changes are usually framed as being the component that is site specific. More often than not, RSL reconstructions are not generated with the purpose of understanding local-scale change, and indeed sites are selected to attempt to minimise the influence of local-scale factors. Therefore estimating the local-scale component is most commonly a means to distill the contribution from processes acting at larger spatial scales. Changes in tidal range are an example of local-scale processes and the geomorphology of some sites (such as, a narrow and dynamic connection to the open ocean) renders some more susceptible to this change than others.

2.3 Data Collection Techniques

Three primary methods are used to investigate changes in RSL over time and space. Recent changes are captured through instrumental records from satellites and tide gauges, while proxy records reconstruct longer-term changes prior to human influence. The data sources include instrumental records from satellites (covering the previous 30 years) and tide gauges (covering the last 150 years), as well as proxy data extending beyond the instrumental timeframe. This thesis focuses specifically on RSL changes during the late Holocene era, spanning the past 3,000 years which requires proxy and tide gauge data.

2.3.1 Instrumental Data

One method for measuring the height of the ocean is using satellite altimetry. This technique involves using radar to accurately and precisely measure the height of the ocean with near global coverage (Church et al., 2013). However, this method is limited to data collected from 1992 onwards, since the launching of the TOPEX/-Poseidon satellites (Church et al., 2013). In Figure 2.6 we demonstrate the temporal range of the satellite data for examining sea level changes (Upton et al., 2023a). As a result, alternative data collection techniques will be the focus of this thesis.

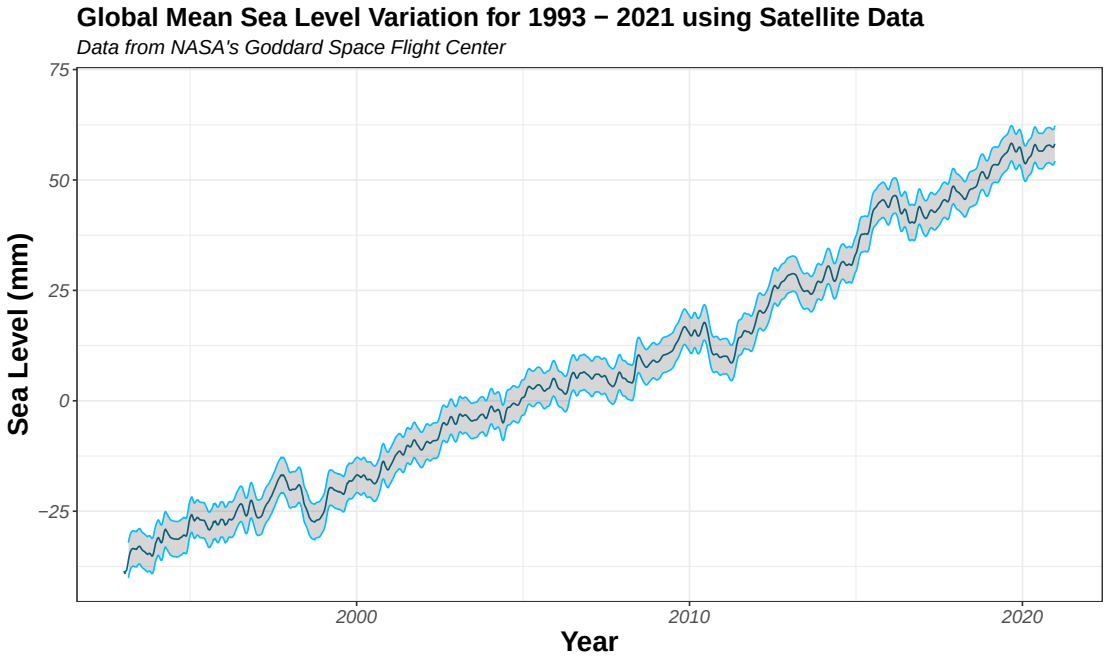


Figure 2.6: Global mean sea level variation over the past 30 years based on satellite data sourced from NASA's Goddard Space Flight Center (Data sourced from NASA: Biferno et al., 2023)

The next data collection technique is tide gauges. Tide gauges are devices fixed to the land and regularly measure the height of the adjacent sea surface at regular intervals, for example every hour or at higher frequencies (Pugh and Woodworth, 2014a). The Permanent Service for Mean Sea Level (PSMSL) compiles monthly and annual tide gauge data for approximately 1,500 stations globally (PSMSL; Woodworth and Player, 2003; Holgate et al., 2013). Figure 2.7 displays the highly uneven distribution of the global network of tide gauges across space and throughout time (Upton et al., 2023a). With the earliest tide gauge records beginning in the late 17th or early 18th centuries in northwestern Europe, the tide gauge records are temporally restricted and an additional data collection technique is required to extend these records further back in time (Wöppelmann et al., 2006).

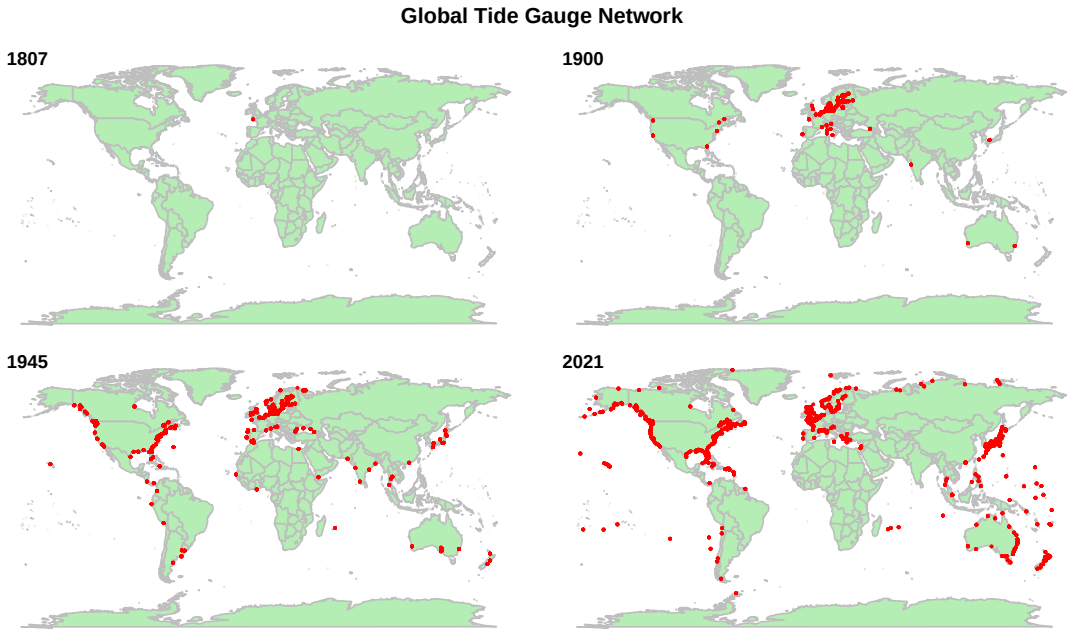


Figure 2.7: Location of the tide gauge stations from 1807 to 2021 using data from Permanent Service for Mean Sea Level online database (Data sourced from PSMSL: Woodworth and Player, 2003; Holgate et al., 2013)

In this thesis, we use tide gauge data to extend the proxy records to capture recent RSL changes using data from the online Permanent Service for Mean Sea Level database (PSMSL; Woodworth and Player, 2003; Holgate et al., 2013), focusing on sites along the Atlantic coast of North America in Chapter 3 and in Chapter 5 we extend our records to incorporate tide gauges along the coastline of the North Atlantic. In Chapter 4 we demonstrate the capabilities of the `reslr` package which automatically downloads the user's required tide gauge from the PSMSL online database.

2.3.2 Proxy Records

In the late 1970s, a number of projects were coordinated under the International Geological Correlation Programme with the aim of defining a strategy for collecting indicators of paleo sea level (Van De Plassche, 1986; Brooks and Edwards, 2006). Findings from these early studies provided a clear methodology for developing records of RSL change (Edwards, 2005). These indicators are classified

as sea-level proxies, which can be physical, biological, or chemical in nature, and possess a systematic and quantifiable relationship with sea level during their formation, for example, in situ sediments, fossil organisms, morphological features, and archaeological remains (Horton et al., 2018). Proxies for sea level can be collected from diverse environments that display vertical zonation allowing for meaningful interpretation, such as salt marshes in mid to high latitudes (e.g., Stearns et al., 2023) or mangrove sediments in low latitudes (e.g., Khan et al., 2022).

Proxy records were used to reconstruct historic RSL before the availability of instrumental records and can be obtained from two main sources: sea level index points and continuous cores. In this thesis, we focus on proxy records arising from continuous cores for sites along the coastline of the North Atlantic.

2.3.2.1 Sea Level Index Point

A sea-level index point (SLI or SLIP) is a single sample with a single proxy observation at a single location and representing a single historic date (e.g., Brooks and Edwards, 2006). SLIPs possess uncertainty in both the age and the reconstructed RSL elevation. The vertical uncertainty associated with the reconstructed RSL elevation arises from the local tidal range where the sample is collected, whereas the age uncertainty of the sample results from the dating process (Khan et al., 2019). SLIPs have been collected globally in many different environments and have a number of different uses when examining paleo sea-level change (e.g., Shennan and Horton, 2002; Horton et al., 2007). One advantage of SLIPs is that they can be used in testing and constraining physicals models such as GIA models (e.g. Shennan et al., 2002; Peltier, 1998). As a result, the focus on SLIPs is to obtain long term temporal records, i.e. dating back over 10,000 years, however, these types of records are extremely difficult to collect on a global scale (Shennan et al., 2006). For this reason, we do not use SLIPs in our analysis.

2.3.2.2 Continuous Cores

For our research, we use proxy records from continuous cores to reconstruct historic RSL prior to instrumental records (e.g. Kemp et al., 2009). Continuous cores involve collecting sediment cores (vertical columns of sediment obtained from the

Earth's sediment surface, with the oldest materials located at the lowest depths and the most recent materials found at the highest levels, as shown in Figure 2.8) from coastal or marine environments, offering a near continuous record of sediment deposition over time (e.g. Van De Plassche, 1986). Samples of sediment are recovered in a single core and interrogated in subsequent laboratory analysis to determine sample age and the elevation at which it accumulated relative to tidal datums (a defined stable zero level for sea level (Pugh, 2004)). These cores, known as geological tide gauges or near continuous proxy reconstructions, contain multiple proxy observations and numerous historic dates (e.g. Barlow et al., 2013). By analysing the sediment layers, researchers can identify changes in sea level based on variations in sediment composition, microfossil assemblages, or other indicators (Shennan et al., 2015). This allows for a detailed reconstruction of paleo sea level changes.

Figure 2.8: An example of continuous core of salt marsh sediment from Prince Edward Island, Canada. The sediment core is collected using a hand held Russian Peat corer.

To estimate the age of a sample, a history of sediment accumulation is developed

by directly dating a subset of levels (depths) within a core typically using radiocarbon measurements (Törnqvist et al., 2015). In addition, dating the shallowest (i.e. most recent) part of the core can be accomplished by identifying trends and events of a known age in down-core profiles of elemental abundance, isotopic ratios, activity, and pollen assemblages, which can be reliably attributed to historical causes (Marshall, 2015). These directly dated levels provide the input for a statistical age-depth model, which estimates the age of every sample (often but not always around 1-cm thick) in the core with uncertainty. Several statistical models are available to perform this step such as the Bchron (Parnell et al., 2008), Bacon (Blaauw and Christen, 2011), or Rplum (Aquino-López et al., 2018) packages in R, and comparisons between the models indicate that the output has minimal dependence on the specific model used (e.g. Wright et al., 2017).

Reconstructing the height of RSL requires a sea-level proxy. One example of a sea-level proxy is distinctive plant communities found in salt marshes and mangrove environments that have adapted to saltwater inundation within narrow elevation ranges (Redfield, 1972). By employing reasoning by analogy, the spatial distribution of these plant species in present-day salt marshes can be used to interpret their preserved counterparts found within core samples, allowing for paleomarsh elevation (elevation with respect to tidal elevation at the time of formation) to be reconstructed (Kemp and Telford, 2015).

Another example of a sea-level proxy is the remains of microfossils, which are single celled organisms (also known as foraminifera) that are preserved in salt-marsh sediment (e.g. Kemp et al., 2011). These form distinctive assemblages (as described by the relative abundance of species) with a strong relationship to tidal elevation as they are sensitive to duration and frequency of tidal inundation (e.g. Horton et al., 2018). There are two advantages to using microfossils to reconstruct RSL: (1) assemblages may occupy narrow elevations zones, which enables them to produce more precise reconstructions than proxies with broader tolerances, such as plants; (2) the large number of individuals in a microfossil assemblage (possibly several hundred in a single sample), makes them suitable for quantitative analysis (e.g. Kemp et al., 2009, 2013).

Typically, RSL reconstructions using microfossils rely on the application of transfer functions. A transfer function provides quantitative estimates of past environmental conditions using empirically-derived equations and paleontological data (Sachs et al., 1977). The transfer function is calibrated using the relationship between micro-fossil families and tidal elevation using a dataset representative of the modern environment (Kemp and Telford, 2015). Transfer functions may use Frequentist (Sachs et al., 1977; Horton and Edwards, 2006; Kemp et al., 2011) or Bayesian approaches (Cahill et al., 2016) which provide an estimate for paleo-marsh elevation with an associated uncertainty. This estimate for paleo-marsh elevation is given in units of a standardised water level index (SWLI) which is utilised to standardise the elevation of the modern day marsh across all sites of varying tidal ranges in the study (Wright et al., 2011). Next, this estimate is converted from SWLI units to paleo-marsh elevation in the core sample. From this value, the RSL for each core sample is reconstructed using the paleo-marsh elevation, surface altitude and the depth of core sample (Marshall, 2015).

The final proxy RSL record is completed by appending the age of each core sample provided by the age-depth models with the associated RSL reconstructed using sea-level proxies. The proxy RSL record contains stratigraphically-ordered data points of age and RSL with corresponding 1 sigma uncertainties for both age and RSL. The vertical uncertainty of sea level indicators, which varies among different proxies and is influenced by tidal amplitude, is associated with the precision of the measurements (Barlow et al., 2013). The age uncertainty, on the other hand, is inherent to the dating method, usually radiocarbon dating, and varies due to fluctuations in atmospheric radiocarbon concentration (Edwards, 2007). In Figure 2.9, we present a schematic illustration summarising the formation of the proxy record. The boxes represent the uncertainty and the dots are the midpoints of the boxes. In this thesis, proxy records from the coastline of the North Atlantic will be the focus.

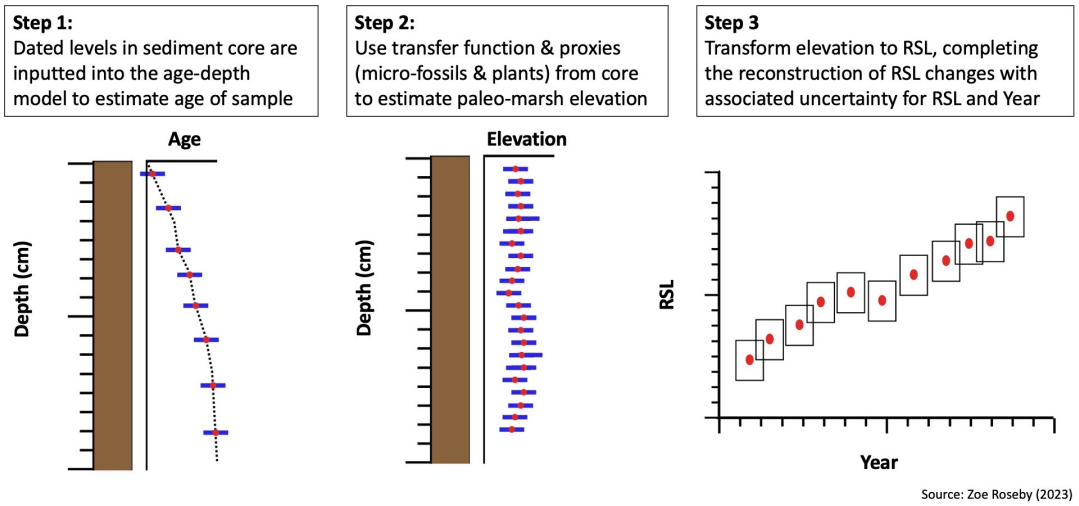


Figure 2.9: A diagram illustrating how to reconstruct relative sea-level using a continuous core of sediment. In step 1, the age and associated uncertainty of the sample is obtained by dating levels in the sediment core using a dating technique, e.g. radiocarbon dating, and inputting this result into an age-depth model. In step 2, sea level proxies from the core (i.e. plants and micro-fossils) and a transfer function are used to estimate paleo-marsh elevation with uncertainty. In step 3, the paleo-marsh elevation is converted to RSL and combined with the sample ages to form the proxy records used reconstruct historic RSL changes. The boxes represent the uncertainty and the red dots are the midpoint of the boxes.

2.4 Summary

In this chapter, we defined RSL and highlighted the underlying physical processes driving changes across temporal and spatial scales. We discussed various RSL data sources; satellite measurements, tide gauges and proxy records, and evaluated their inherent advantages and disadvantages. We demonstrated the temporal limitations of satellite data, leading to its exclusion from this thesis. We examined the instrumental records collected by tide gauges and highlighted its uses in the upcoming chapters. Also, we carefully examined proxy record sources, including SLIP and continuous cores. Ultimately, we selected continuous cores as their observations are nearly uninterrupted over time, which aligned with our research goals. In summary, this chapter established the research groundwork, defining key concepts, and assessing data sources. Subsequent chapters expand on this foundation, exploring statistical models and software advances to investigate physical drivers of RSL changes and their implications at various regional levels.

CHAPTER 3

A Noisy-Input Generalised Additive Model for Relative Sea-Level Change along the Atlantic Coast of North America

We propose a Bayesian, noisy-input, spatial-temporal generalised additive model to examine regional relative sea-level (RSL) changes over time. The model provides probabilistic estimates of component drivers of regional RSL change via the combination of a univariate spline capturing a common regional signal over time, random slopes and intercepts capturing site-specific (local), long-term linear trends and a spatial-temporal spline capturing residual, non-linear, local variations. Proxy and instrumental records of RSL and corresponding measurement errors inform the model and a noisy-input method accounts for proxy temporal uncertainties. Results focus on the decomposition of RSL over the past 3000 years along the Atlantic coast of North America. R code and data for NI-GAM implementation is available at https://github.com/maeveupton/NI-GAM.

3.1 Introduction

The Intergovernmental Panel for Climate Change (IPCC) in 2021 reported with “high confidence" that global mean rates of sea-level rise increased from approximately 1.3mm/yr between 1901 and 1971 to 3.7mm/yr between 2006 and 2018, with a further increase in rates predicted for the remainder of the 21st century (Masson-Delmotte et al., 2021, p. 5). In contextualising the socio-economic risk that this sea-level rise poses for coastal communities, it is necessary to place historic and predicted changes in a longer term (pre-anthropogenic) context and to recognise that local sea level can diverge sharply from the global average.

Relative Sea Level (RSL) is the height of the ocean surface at any given location and time, measured relative to the adjacent land (Church et al., 2013). Direct measurements of RSL (typically considered to be high accuracy and with low uncertainty) are made by a network of coastal tide gauges whose spatial distribution is highly uneven and whose temporal duration is typically limited to the past $\sim$ 100 years or less (Church and White, 2011). Understanding RSL before tide-gauge measurements began requires proxies (physical, biological, or chemical features with an "observable and systematic relationship to tidal elevation"; Horton et al., 2018) that are preserved in dated geological archives such as coastal sediment (e.g. Gehrels, 1994) or corals (e.g. Meltzner et al., 2017). For the past 3,000 years (a period in Earth’s history called the late Holocene), it is possible to generate near-continuous proxy RSL reconstructions which overlap tide-gauge measurements (Kemp et al., 2013). The suite of late Holocene RSL proxy reconstructions is growing, but their global distribution is highly uneven (Ashe et al., 2019). However, the Atlantic coast of North America has a relatively large number of datasets (Figure 3.1) generated from sediment that accumulated in salt-marsh (e.g. Kemp et al., 2018) and mangrove environments (e.g. Khan et al., 2022). We therefore focus on this region to develop a new statistical model for quantifying patterns, rates, associated uncertainties and possible causes of late Holocene RSL change from a combination of proxy reconstructions and tide-gauge measurements concurrently.

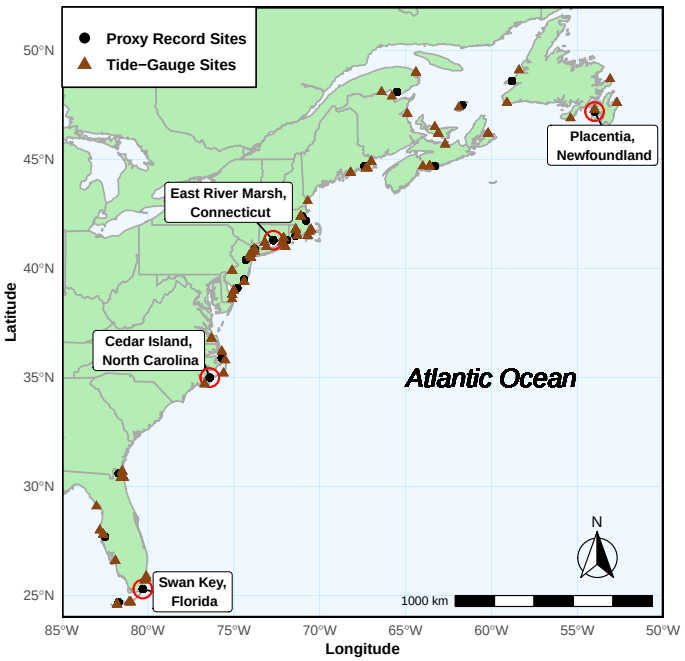


Figure 3.1: Location of the 66 tide gauge sites and 21 proxy record data sites along the Atlantic coast of North America with four proxy record sites chosen as case studies to present results of our model.

A general discussion on how proxy records are developed is provided in Section 3.2 and we point the reader in the direction of Shennan et al. (2015) for a more detailed account of the methodologies employed by the paleo sea-level community. In this paper, we focus on analysing published data arising from proxy RSL reconstructions. The proxy records contain RSL estimates throughout time for different locations, specifically along the Atlantic coast of North America, and have associated bivariate uncertainties, i.e. uncertainty in time and vertical uncertainty in RSL.

Tide gauges and proxy records can only capture RSL, which is the net outcome of a complex combination of physical processes operating on characteristic temporal (years to millennia) and spatial (site-specific to global) scales. These physical processes often act simultaneously and serve to reinforce or mask one another; they can change both the height of the sea-surface and that of the land differently through time and across space (Khan et al., 2022). Consequently, RSL

measurements can display a rich variety of spatio-temporal patterns. A principal goal of sea-level research is to interrogate these patterns to identify and quantify the contribution from specific physical processes including (but not limited to) the multi-millennial and regional response of the solid Earth to de-glaciation, decadal to centennial redistribution of ocean water by changing currents, and the recent (multi-decadal) acceleration of global average sea-level rise in response to a warming climate (Church et al., 2013). This goal requires a means to decompose the site-specific RSL signal at each locality in the network into contributions at different temporal and spatial scales, while accounting for uncertainties in the underlying data.

The most widely used tool for decomposing late Holocene RSL is a model developed by Kopp et al. (2016, hereon K16) and its various extensions (Kemp et al., 2018; Walker et al., 2021). The K16 model decomposes RSL into three categories: (1) a non-linear signal common to all records in the dataset being analysed (termed global, irrespective of the geographic range of input data); (2) a regional signal characterized by a linear rate of change over the past $\sim$ 2000 years and (3) a local (site-specific) signal that operates in a non-linear fashion. Rather than representing specific physical processes, these categories serve to represent groups of processes that operate at similar spatial and temporal scales informed by the data. K16 employs Gaussian Process (GPs) for each component and, due to the associated computational burden which grows in proportion with the cube of the number of data points, relies on a maximum likelihood approach to estimate and fix model hyperparameters. These modelling decisions, that aim to reduce the computational burden of GPs, can impact uncertainty quantification (Ashe et al., 2019). In this paper we aim to propose an alternative method for estimating these complex, interdependent components, that improves uncertainty quantification whilst remaining computational feasible.

Our new spatio-temporal statistical approach to modelling RSL change uses Generalised Additive Models (GAMs). A GAM is a generalised linear model where "the linear predictor depends linearly on a sum of smooth functions of the predictor variable" (GAMs; Wood, 2017a, p. 161). GAMs flexibly model non-linear relationships using smooth functions (most commonly splines) and can reduce

computational complexity when compared with GPs (for example those used in K16) as they do not require large matrix inversions (Hastie and Tibshirani, 1990). We place our model in a Bayesian framework which allows for the estimation of parameters conditioned on the RSL data with full accounting for, and propagation of, uncertainty. Similar to K16, our model partitions the total RSL signal into components that characterize distinctive spatial and temporal scales, which (to varying degrees) are associated with specific physical processes. These components are: (1) a regional component, a non-linear signal common to all sites along the Atlantic coast of North America, and equivalent to the global term in K16; (2) a linear local component which contains unstructured random effects and is comparable the regional linear term in K16; and (3) a non-linear local component, which is site-specific and varies smoothly in space and time. Similar to K16, any variation not addressed by the model is captured by a residual term.

Since the data points at each site have a bivariate error structure, and the decomposition required involves differing structures, the simple application of default GAMs does not work in our case study. Previous methods, such as (Cahill et al., 2015a), provides guidance for how to model RSL with bivariate uncertainty. We follow K16 in accounting for the time error using the Noisy-Input uncertainty method of McHutchon and Rasmussen (2011). This method inflates the residual variance by a corrective term to compensate for noisy-input measurements using a smooth process. Whilst the original paper uses the method exclusively for GPs we extend the approach to spline terms. The RSL error is captured via a standard measurement error term added to the residual variance.

The structure of our paper is as follows. Section 3.2 addresses the proxy records and tide-gauge data used in our analysis. Section 3.3 describes the main physical processes driving RSL changes and Section 3.4 discusses the previous modelling strategies employed by Kopp et al. (2016). Section 3.5 gives a detailed description of our statistical model with different splines representing each driver of RSL change and introduces the noisy-input method. The model validations are shown in Section 3.6 and the results for different drivers of RSL change and their associated rates are presented in Section 3.7. Section 3.8 provides concluding remarks for our approach for the Atlantic coast of North America.

3.2 Data

We use a combination of instrumental data from tide gauges and proxy records. This section discusses the different sea-level data sets, including their collection methods and their associated uncertainties.

3.2.1 Tide-Gauge Data

Tide gauges are fixed to the land and regularly measure (for example hourly or to higher frequency) the height of the adjacent sea surface (Pugh and Woodworth, 2014a). For understanding RSL change, these observations are usually expressed as annual averages and held in the database maintained by the Permanent Service for Mean Sea Level (PSMSL; Woodworth and Player, 2003; Holgate et al., 2013). The ∼1500 stations in this global network display highly uneven distribution of data across space and through time and in addition individual records may have temporal gaps (Church and White, 2011).The earliest tide gauges records began in the late 17th or early 18th centuries in northwestern Europe (Wöppelmann et al., 2006). Along the Atlantic coast of North America, the longest tide-gauge record in the PSMSL database is The Battery from New York City (since 1856 CE) (Holgate et al., 2013). Annual tide-gauge data from the PSMSL are treated as having fixed and known ages without uncertainty in elevation measurements (Holgate et al., 2013).

In our analysis, we use 66 tide-gauge sites along the Atlantic coast of North America (Figure 3.1). Tide gauges meeting at least one of the following criteria were included in our analysis; (1) record length exceeding 150 years; (2) the nearest tide gauge to proxy site; (3) within 1 degrees distance to a proxy site and longer than 20 years (Kopp et al., 2016; Walker et al., 2021). The addition of tide-gauge data supplements the long-term proxy records and provides additional insight into recent changes in RSL. Annual data for each tide gauge were downloaded from the PSMSL and expressed in meters relative to the average over 2000-2018 CE. This time window captures variability resulting from the 19-year cycle in astronomical tides (Pugh and Woodworth, 2014b) and serves to make proxy and tide-gauge data comparable since the sediment cores used to develop proxy reconstructions were recovered since ∼ 2000 CE. In addition, we further average tide-gauge data

by decade to increase comparability with proxy reconstructions that are developed from 1 cm thick slices of core sediment which accumulated over a period of several years (depending on sedimentation rate) and are therefore inherently time averaged. At this step we include an uncertainty in the tide-gauge data (± 5 years for age and ± 1σ for RSL). See Appendix 3.A for additional information.

3.2.2 Proxy Records

Proxy-based reconstructions provide estimates of pre-anthropogenic RSL (Kemp et al., 2013). On the Atlantic coast of North America, these near-continuous proxy-based reconstructions are generated using buried sequences of salt-marsh (at mid to high latitudes; Gehrels et al., 2020) or mangrove (low latitudes; Khan et al., 2022) sediment. Samples of this sediment are recovered in a core (a column of sediment extracted from the ground, where the oldest material is at the bottom and the youngest material is at the top) and interrogated in subsequent laboratory analysis to determine the age of the sample and the tidal elevation (height above a tide level) at which it accumulated (Horton and Edwards, 2006).

A history of sediment accumulation provides estimates of sample ages by directly dating a subset of depths in the sediment core, typically using radiocarbon measurements (Törnqvist et al., 2015). In addition, the shallowest (i.e. most recent) part of the core can be dated by recognising historic pollution and land use changes of known age in down-core profiles of elemental abundance, isotopic activity and isotopic ratios (Marshall, 2015). These directly dated levels in the core are the input (i.e. age of sediment sample) for a statistical age-depth model (e.g. the Bchron (Parnell et al., 2008), Bacon (Blaauw and Christen, 2011), or Rplum (Aquino-López et al., 2018) packages in R). These age-depth models (irrespective of their specific similarities and differences) estimate the age of every 1 cm thick sediment sample in the core with uncertainty. Comparisons indicate that sediment accumulation histories have little dependence on the specific age-depth model used (Wright et al., 2017).

A sea-level proxy is required to reconstruct RSL. A sea-level proxy is any physical, biological or chemical feature with an observable and systematic relationship to tidal elevation (Shennan et al., 2015). Salt marshes and mangrove environments

are vegetated by distinctive plant communities that are adapted to inundation by salt water, resulting in distinct and narrow elevation ranges (Redfield, 1972). This distribution makes salt-marsh vegetation a valuable sea-level proxy. Through reasoning by analogy, the observable distribution of plants in modern salt marshes enables interpretation of their analogous counterparts preserved in core material (Kemp and Telford, 2015). In this way, the paleo-marsh elevation (elevation with respect to tidal elevation at the time of formation) is reconstructed. Another sea-level proxy preserved in salt-marsh sediment is the remains of micro-fossils (e.g., foraminifera) that form distinctive assemblages with a strong relationship to elevation (Edwards and Wright, 2015). When using micro-fossils to reconstruct RSL a transfer function is required which relates the abundance of specific micro-fossil families to tidal elevation using a dataset that is representative to the modern environment (Kemp and Telford, 2015). There are various transfer functions available using Frequentist (Sachs et al., 1977; Horton and Edwards, 2006; Kemp et al., 2011) and Bayesian approaches (Cahill et al., 2016), which all estimate paleo-marsh elevation with uncertainty. The age of each core sample with a corresponding paleo-marsh elevation reconstruction is provided by the age-depth model. Resulting in a single proxy RSL record comprised of stratigraphically-ordered data points of age (with 1 sigma uncertainty) and RSL (with 1 sigma uncertainty) as shown in Figure 3.2.

In Figure 3.2, we illustrate the proxy RSL records derived from the analysis. Each box represents the proxy sea-level reconstruction for a specific core sample, with the associated uncertainty clearly depicted in grey. The midpoint of the box is denoted by a black dot, and larger boxes indicate greater uncertainty associated with that particular core sample (Upton et al., 2023a). The vertical uncertainty is directly tied to the precision of the sea level indicator or proxy, such as micro-fossils or vegetation, utilized in the sea-level reconstruction. This precision varies among different proxies and is influenced by tidal amplitude (Edwards and Wright, 2015). Furthermore, the horizontal uncertainty is inherent to the dating approach employed. As previously mentioned, in techniques like radiocarbon dating, variations in atmospheric radiocarbon concentration lead to differences in the magnitude of uncertainty (Kemp and Telford, 2015).

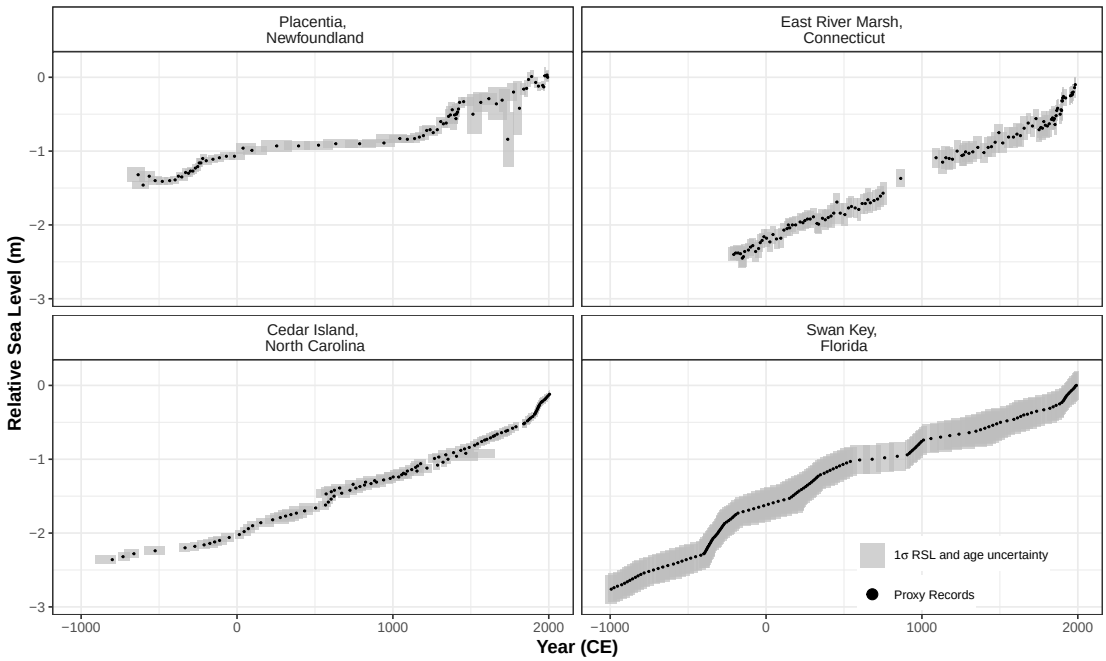


Figure 3.2: Proxy records from four proxy sites along the Atlantic coast of North America used as illustrative case studies. The y-axis is relative sea level (RSL) in meters, where 0m is present sea level and negative values indicate RSL below present. Each proxy record observation consists of paired age and RSL estimate at the corresponding site. The black dot represents the midpoint of the proxy sea-level reconstruction and the grey boxes of 1 standard deviation represent vertical and horizontal (temporal) uncertainty.

We analyze 21 RSL proxy records (totaling 1731 data points) located on the Atlantic coast of North America from the Florida Keys, USA to Newfoundland, Canada (Figure 3.1). There are 66 tide-gauge records that meet our criteria for inclusion (Figure 3.1). The spatial scope of our analysis is restricted to this coastline because it has (by a considerable margin) the greatest concentration of available records. Results presented in this paper are generated from all the proxy and tide-gauge records, yet we present four of these sites (Placentia Newfoundland Canada, East River Marsh Connecticut USA, Cedar Island North Carolina USA, and Swan Key Florida) as illustrative case studies throughout the remainder of the manuscript as shown in Figure 3.2. The four sites were selected to provide diversity of location and therefore the processes causing RSL change during the past $\sim$ 3000 years.

3.3 Drivers of RSL Change

Spatio-temporal models recognise sea-level variability characteristic of different spatial and temporal scales rather than from specific processes. Each component estimated in the model may capture several contributing processes depending on the location and time interval under examination. These processes may act simultaneously and in directions that mask or exaggerate contributions from other drivers.

Transfer of mass between land-based ice and the ocean drives RSL change. Ice melt/growth returns or removes mass to the ocean as liquid water which causes a rise/fall in global mean sea level (this process is termed barystatic) (Gregory et al., 2019). This contribution varies in magnitude across timescales, but is common to all locations. In addition, changes in global temperatures alter the density of ocean water resulting in a sea-level change (rise/fall when water warms/cools becoming less/more dense); this process is known as a thermosteric contribution (Grinsted, 2015). The global term in the K16 model and the regional component in our model attempts to capture influences from these processes.

Along the Atlantic coast of North America, the principal driver of RSL change during the pre-industrial late Holocene is glacial isostatic adjustment (GIA) (Roy and Peltier, 2015). GIA is the response of the Earth, the gravitational field, and the ocean to the growth or decay of ice sheets (Whitehouse, 2018). GIA can be reasonably approximated as a linear contribution through time on this relatively short timescale, but with considerably variability along the coast (Engelhart et al., 2009). There are a family of physical models known as Earth-ice models which use a representation of the physical Earth structure (such as lithospheric thickness and properties such as mantle viscosity) to predict changes in GIA that occur through loading and unloading of ice, and provide estimates of GIA rates. One such example of an Earth-ice physical model is the ICE5G VM2-90 (Peltier, 2004). It is important to recognise that other processes (e.g., tectonically-driven vertical land motion) can mimic the linear trend of GIA. However, along the passive margin of the Atlantic coast of North America these non-GIA drivers are likely modest in magnitude (Kopp et al., 2015). As a result, the linear local component in our

model and the linear regional term in the K16 may capture contributions from processes other than GIA that drive RSL changes.

There are processes with a spatially-coherent structure where the signal is shared by some but not all sites (Stammer et al., 2013). One such process that can cause RSL to vary on decadal to multi-century timescales is the redistribution of existing ocean mass by shifts in prevailing patterns and strength of atmospheric and oceanic circulation (termed dynamic sea-level change) (Gregory et al., 2019). Dynamic sea level varies by site, but the magnitude of the difference from one site to the next is too small to be detected using proxy data due to the resolution. Some processes (e.g. sediment compaction: which can impact the height of the solid Earth surface with changes in sediment volumes for each site (Horton et al., 2018)) can drive RSL changes that are site-specific. Consequently, contributions from these processes lack spatial coherence and display an unpredictable spatial structure. Therefore, site-specific RSL changes can vary markedly across closely-spaced sites. More often than not, RSL proxy reconstructions are not generated with the goal of understanding site-specific processes (Walker et al., 2021). It remains important to quantify this component as a means to distill the contribution from processes acting at larger spatial scales. In our model the structured (common to some, but not all sites) and unstructured (unique to one site) RSL variability on century timescales is captured by the non-linear, local component.

3.4 Previous Statistical Models for RSL Change

In this section, we review previous work on modelling RSL change, focusing in particular on K16. The model was further extended in Kemp et al. (2018) and Walker et al. (2021), here, we focus on the simpler K16 model. We first review the structure of this model, which decomposes RSL into component parts before discussing how the model might be fitted to the data and the potential influence of optimising hyperparameters using maximum likelihood. K16 forms the basis upon which we build our new approach in Section 3.5.

The RSL measurements are recorded in units of height; with meters used by default. In cases where the scale of the change is relatively small we use cm or mm

instead for some plots and discussion in the text. We write $y_{ij} = y(\mathbf{x}_j, t_{ij})$ for the RSL height at location $\mathbf{x}_j$ (latitude and longitude) and time t_{ij}. These observations arise from the proxy records and tide gauges with j indexing the data site and i the observation. For the resolution of the data the time is expressed in years CE. The K16 model can be written as:

$$y_{ij} = f(\mathbf{x}_j, t_{ij}) + w(\mathbf{x}_j, t_{ij}) + y_0(\mathbf{x}_j) + \epsilon^y_{ij} \tag{3.1}$$

where f is the full RSL spatio-temporal field, w is a white noise process representing sub-decadal trends unexplained by the data due to resolution of data, y_0 is a site-specific spatially variable vertical offset, and ϵ^y_{ij} is residual error. In K16 all the structured terms above are given Gaussian Process prior distributions with stationary covariance functions.

A key complication is that the times t_{ij} associated with the proxy records are observed with uncertainty. Thus the observed values $\tilde{t}_{ij}$ have measurement error, defined as:

$$\tilde{t}_{ij} = t_{ij} + \epsilon^t_{ij}. \tag{3.2}$$

Usually ϵ^t_{ij} is assumed iid normally distributed with known variance, though in reality the age-depth model through which the ages are estimated often provides skewed distributions. A previous attempt at resolving this issue can be found in Cahill et al. (2015a), though across large, multi-site datasets the imposition of this assumption is believed to have minor effects on the outcome of the model (as shown in Parnell and Gehrels, 2015).

3.4.1 Decomposing the RSL Field f

For the RSL process defined above as f, K16 use a spatio-temporal empirical Bayesian hierarchical model to partition the influence of the components into global, regional and local scales. The fields that make up f are, as above, given stationary GP priors that can vary in time and space as controlled by the covariance functions (Ashe et al., 2019). The standard decomposition of f is written:

$$f(\mathbf{x}_j, t_{ij}) = c(t_{ij}) + g(\mathbf{x}_j)(t_{ij} - t_0) + l(\mathbf{x}_j, t_{ij}) \tag{3.3}$$

where $c(t_{ij})$ is term the global term, the temporal non-linear signal common across all sites, designed to capture changes such as barystatic sea level rise and thermosteric changes. $g(\mathbf{x}_j)$ is a spatially varying term that captures slower processes such as long-term land level change (GIA) and vertical land motion driven by plate tectonics. The g term is multiplied by time t differenced from a reference point t_0 to form temporally linear field. Unlike the other components in K16, g is given a univariate normal prior distribution with the mean centred on the value obtained from a Earth-ice physical model (ICE5G VM2-90; Peltier, 2004) which estimates the GIA rate. $l(\mathbf{x_j}, t_{ij})$ is the local spatio-temporal field that describes factors such as dynamic sea level change, sediment compaction and tidal regimes. These terms are explained in more detail in K16.

Without strong prior information, it is difficult to separate out the magnitudes of the components. Thus in K16, the hyperparameters are first obtained by maximising the likelihood of the model conditioned on the observations but constrained to two timescale hyperparameters for the non-linear terms. The model is then re-fitted using these hyperparameters to estimate the components of the fields in an empirically Bayesian framework. In our approach, we aim to avoid the empirical Bayesian approach of fixing hyperparameters by placing informed priors on the model components. However, model shortcuts are unavoidable due to the complexity of the decomposition and the innate confounding of many of the key terms.

3.5 A New Approach Based on Generalised Additive Models

In this section we outline a new approach to evaluating the different drivers of spatio-temporal RSL using proxy records and tide gauge data. With careful choices of the prior distributions of the hyperparameters, we aim to recover the components of RSL change through the standard tools of Bayesian inference. Subsequently, we estimate of rates of RSL change at sites along the Atlantic coast of North America. We build our model inspired by the standard decomposition of the RSL field f as described above. Our approach contains four main differences compared to the

Kopp et al. (2016) (and subsequent) models:

1. We focus on the high quality sites along North America's Atlantic coast and aim to produce a regional RSL curve. Thus we avoid making statements about global sea-level change.

2. We use splines instead of GPs to avoid the computationally challenging inversion of the GP covariance matrices. The model, at its simplest, thus falls under the standard generalised additive modelling paradigm.

3. We fit the model in two stages to maximise the regional variability which would otherwise be confounded with the local structure. This allows us to perform a more complete posterior analysis of the model hyperparameters which might otherwise been fixed in K16.

4. We remove the spatial structure on the linear effect g in K16 and replace it with a univariate random effect on the slope. For the proxy records we use a prior mean for the slope that is informed by the data before 1800 CE (i.e. the pre-industrial time period; Neukom et al., 2019). This change is helpful because we have found the estimated values of the GIA rate from the Earth - ice physical model (e.g. Peltier, 2004; Argus et al., 2014; Caron et al., 2018) do not match the observed data well for the proxy record time period. For the tide gauge records the prior mean of the slope is taken from a physical Earth-ice model (ICE5G VM2-90; Peltier, 2004) with uncertainty taken from Engelhart et al. (2009). We refer back to this modelling choice in Section 3.8.

Below we outline the full posterior distribution of the model to highlight the assumed conditional independences, then outline each term and its structure. The temporal uncertainty in the data causes difficulties in fitting the model in one step, and we resort to McHutchon and Rasmussen (2011)'s noisy-input method to account for this uncertainty. We then discuss the prior distributions assumed for the hyperparameters, and the computational details of our model. In Section 3.7, we showcase the successful implementation of our model.

3.5.1 Model Notation

We now provide a full outline of our notation for reference:

- y_{ij} is an RSL observation in meters with $i = 1, ...n_j$ observations at site j with $j = 1, ..., m$ sites. We vectorise the full set of observations as $\mathbf{y}$ and the observations for each site as $\mathbf{y}_j$.
- t_{ij} are the ages of each RSL observation, indexed and vectorised as above. We represent age in years of the Common Era (CE).
- $\mathbf{x}_j$ is the 2-vector of a latitude and longitude pair for each site j.
- $z_{\mathbf{x}}$ is an index vector for the data sites that converts each site into a label. Thus $z_{\mathbf{x}_j} = j$.
- $f(\mathbf{x}_j, t_{ij}) = f_{ij}$ is the mean sea-level process at site j and time t_{ij}. We write $f(\mathbf{x}, t)$ as the mean process for a generic location and time, and continue with this notation below for brevity
- $r(t)$ is the regional component at time t.
- $l(\mathbf{x}, t)$ is the non-linear local component at location x at time t.
- $g(z_{\mathbf{x}})$ is the linear local component at location x.
- $h(z_{\mathbf{x}})$ is a site-specific vertical offset component at location x.
- $b_r(t)$ and $b_l(\mathbf{x}, t)$ are sets of known b-spline basis functions corresponding to the regional and local components respectively.
- $\boldsymbol{m}^g$ and $\boldsymbol{s}^g$ are the mean and standard deviation parameters respectively for the linear local correction component. These are site specific and so each is a vector of length m.
- $\boldsymbol{\beta}^r, \boldsymbol{\beta}^l$ are the spline regression coefficient vectors of the regional and local components respectively. $\boldsymbol{\beta}^r$ is of length k_r and $\boldsymbol{\beta}^l$ is of length k_l where k_r and k_l are the number of knots associated with each term.

- $\boldsymbol{\beta}^g, \boldsymbol{\beta}^h$ are parameter vectors, each of length m, containing the random effect coefficients for each site. $\boldsymbol{\beta}^g$ is the slope parameter for each site and $\boldsymbol{\beta}^h$ is the intercept parameter for each site.
- σ_r and σ_l are the smoothness parameters associated with the regional and local spline terms respectively.
- σ_h is the standard deviation of the site-specific offset.
- $s_{y_{ij}}$ is the known standard deviation of the RSL data point ij.
- $s_{t_{ij}}$ is the known standard deviation of the age of data point ij.
- σ is a residual standard deviation parameter to capture any remaining variability in y.

3.5.2 Posterior Distribution

The joint posterior distribution of our Bayesian hierarchical model is shown below:

$$\underbrace{p(\sigma^2, \boldsymbol{\beta}^r, \boldsymbol{\beta}^l, \boldsymbol{\beta}^g, \boldsymbol{\beta}^h, \sigma_r^2, \sigma_l^2, \sigma_h^2 | \mathbf{y}, \boldsymbol{b}_r, \boldsymbol{b}_l, m_g, s_g^2, \mathbf{s}_y^2, \mathbf{s}_t^2)}_{\text{posterior}} \propto$$

$$\underbrace{p(\mathbf{y}|\boldsymbol{f}, \sigma^2, \mathbf{s}_y^2, \mathbf{s}_t^2)}_{\text{likelihood}} \times \underbrace{p(\sigma^2)}_{\text{prior on error variance}} \times$$

$$\underbrace{p(\boldsymbol{\beta}^r|\sigma_r^2)}_{\text{prior on regional parameters}} \times \underbrace{p(\sigma_r^2)}_{\text{prior on regional smoothness parameter}} \times$$

$$\underbrace{p(\boldsymbol{\beta}^h|\sigma_h^2)}_{\text{prior on site-specific vertical offset parameters}} \times \underbrace{p(\sigma_h^2)}_{\text{prior on variance site-specific vertical offset parameters}}$$

$$\times \underbrace{p(\boldsymbol{\beta}^g|\boldsymbol{m}_g, \boldsymbol{s}_g^2)}_{\text{prior on linear local parameters}} \times$$

$$\underbrace{p(\boldsymbol{\beta}^l|\sigma_l^2)}_{\text{prior on non-linear local parameters}} \times \underbrace{p(\sigma_l^2)}_{\text{prior on non-linear local smoothness parameter}} \times$$

The likelihood $p(\mathbf{y}|\boldsymbol{f}, \sigma^2, \mathbf{s}_y^2, \mathbf{s}_t^2)$ can be deconstructed thus:

$$p(\mathbf{y}|\boldsymbol{f}, \sigma^2, \mathbf{s}_y^2, \mathbf{s}_t^2) = \prod_{j=1}^{m} \prod_{i=1}^{n_j} \mathcal{N}(y_{ij}|f_{ij}, \sigma^2 + s_{y_{ij}}^2 + s_{t_{ij}}^2) \tag{3.4}$$

3.5.3 A Fully Specified Generalised Additive Model for Decomposing the RSL Field

Our version of the decomposition of the mean sea level field can be written as:

$$f(\mathbf{x}, t) = r(t) + g(z_{\mathbf{x}}) + h(z_{\mathbf{x}}) + l(\mathbf{x}, t) + \epsilon_{\mathbf{y}} \tag{3.5}$$

All terms are as defined above: $r(t)$ is the regional component. $g(z_{\mathbf{x}})$ is the linear local component represented by a random effect with $z_{\mathbf{x}}$ representing each data site. $h(z_{\mathbf{x}})$ is the spatial vertical offset for each data site. $l(\mathbf{x}, t)$ is the non-linear local component. We represent $r(t)$ using a spline:

$$r(t) = \sum_{s=1}^{k_r} b_{r_s}(t)\beta_s^r \tag{3.6}$$

where β_s^r is the s^{th} spline coefficient, k_r is the number of knots and $b_{r_s}(t)$ is the s^{th} spline basis function at time t.

The linear local component, $g(z_{\mathbf{x}})$, is an unstructured random effect for each site which is formulated as:

$$g(z_{\mathbf{x}_j}) = \beta_j^g t \tag{3.7}$$

where β_j^g is a slope parameter specific for each site j. This specification is in contrast to K16 where the linear effect, g, varies smoothly in space and is informed through the prior by GIA model-derived values. We found such a restriction to adversely affect model performance due the lack of agreement between the data and the provided GIA values, and the wide variation in values between proximal sites (Engelhart et al., 2009).

The site-specific vertical offset h is a random effect used to capture vertical shifts associated with measurement variability between sites and is formulated as:

$$h(z_{\mathbf{x}_j}) = \beta_j^h \tag{3.8}$$

where β_j^h contains the random effect coefficients for site j.

The non-linear local component $l(\mathbf{x}, t)$ is described with a spatio-temporal spline function formulated by:

$$l(\mathbf{x}, t) = \sum_{s=1}^{k_l} b_{l_s}(\mathbf{x}, t)\beta_s^l \tag{3.9}$$

where β_s^l is the s^{th} spline coefficient, k_l is the number of knots and $b_{l_s}(\mathbf{x}, t)$ is the s^{th} spline basis function at time t and location $\mathbf{x}$.

We use B-splines (de Boor, 1978) for both the regional and local terms. Our B-splines are constructed as piece-wise polynomials which join together at equidistant knots such that the first derivatives are equal (Eilers and Marx, 1996). For the regional term we use cubic B-splines as we are interested in the behaviour of the first derivatives. We can simply calculate these by differentiating the cubic B-splines and multiplying with the posterior spline parameters to provide a posterior distribution for the derivative. However for the non-linear local component, which requires a tensor product to capture the variability over time and space (represented with longitude and latitude) so that the individual covariates are combined product-wise (Wood, 2006). We use a simpler and faster quadratic polynomial basis for the non-linear local component. Many other basis function types and options are available (see, e.g. Dierckx, 1995; Wood, 2017a) but we believe our approach balances both parsimony and computational efficiency for our application area.

3.5.4 Noisy-Input Uncertainty Method

Our data is corrupted with measurement error in the RSL values and that arising from the temporal uncertainty associated with radiocarbon dating the fossil layers of sediment. McHutchon and Rasmussen (2011) describe an assumption for GPs which avoids the need for complex errors-in-variables models (e.g. Dey et al., 2000; Cahill et al., 2015a) and instead adds an extra measurement uncertainty on the response variable. We adapt this noisy-input (NI) approach for our RSL GAM which we now term an NI-GAM.

The response variable y is assumed to be a noisy measurement with the true output

given as $\tilde{y}$:

$$y = \tilde{y} + \epsilon_y \tag{3.10}$$

where the error term is given by $\epsilon_y \sim \mathbb{N}(0, s_y^2)$ with s_y being the known measurement standard deviation of the RSL data. Similarly, for the input measurements, t is assumed to be a noisy estimate of the true time value $\tilde{t}$:

$$t = \tilde{t} + \epsilon_t \tag{3.11}$$

with the error term given by $\epsilon_t \sim \mathbb{N}(0, \tilde{s}_t^2)$ where $\tilde{s}_t$ is the known standard deviation of the date obtained from the age-depth model (described in Section 3.2). As a result, a function for the response variable is formed in the following way:

$$y = f(\mathbf{x}, \tilde{t} + \epsilon_t) + \epsilon_y \tag{3.12}$$

Following McHutchon and Rasmussen (2011) we can use a Taylor expansion about the latent state $\tilde{t}$ to obtain:

$$f(\mathbf{x}, \tilde{t} + \epsilon_t) = f(\mathbf{x}, \tilde{t}) + \epsilon_t^T \frac{\partial f(\mathbf{x}, \tilde{t})}{\partial \tilde{t}} + \cdots \approx f(\mathbf{x}, t) + \epsilon_t^T \frac{\partial f(\mathbf{x}, t)}{\partial t} \tag{3.13}$$

Thus the error in t can be approximated by an increase in the measurement error proportional to the derivative of f. McHutchon and Rasmussen (2011) calculate the derivative of the mean of the GP function, given as vector $\partial_{\bar{f}}$ for the first order case and $\Delta_{\bar{f}}$ for a D-dimensional matrix.

Analogously for our NI-GAM setting, the first order terms are expanded to form a linear model with input noise:

$$y = f(\mathbf{x}, t) + \epsilon_t^T \partial_{\bar{f}} + \epsilon_y \tag{3.14}$$

The derivative of the posterior mean for f is obtained using a two-step method. First the model is fitted ignoring the input uncertainty and then the slope of the posterior mean is calculated. From this, a corrective variance term can be calculated, which we write as s_t^2. We use this as an additional model error term in our subsequent full model fit.

Intuitively, the input noise impacts the gradient of the function mapping input to output and the input noise variance is related to the output by the square of

the posterior mean function's gradient (McHutchon and Rasmussen, 2011). As a result, the corrupted input measurements influence a rapidly changing output value more than an output value that remains constant. The advantage of this method is that the noise remains the same whether the measurement is considered an input or output measurement, and so all the data informs the input noise variance ensuring the output dimensions are met, reducing the chance of over-fitting.

3.5.5 Prior Distributions

Within the process level each component is given a prior distribution. Our prior for the spline coefficients of the regional component, β_s^r is:

$$\beta_s^r \sim \mathbb{N}(0, \sigma_r^2) \tag{3.15}$$

where σ_r is the standard deviation of the spline coefficient and fundamentally controls the smoothness of the model fit.

Our prior for the linear local component for the proxy records is:

$$\beta_j^g \sim \mathbb{N}(m_{g_j}, s_{g_j}^2) \tag{3.16}$$

where m_{g_j} and $s_{g_j}^2$ are the empirically estimated rate and associated variance for the data set obtained from the time period prior to 1800 CE (Neukom et al., 2019). For the tide-gauge records, we obtain m_{g_j} from a physical model (ICE5G VM2-90; Peltier, 2004) and $s_{g_j}^2$ from previous studies (Engelhart et al., 2009).

Our prior distribution for the site-specific vertical offset is:

$$\beta_j^h \sim \mathbb{N}(0, \sigma_h^2) \tag{3.17}$$

where σ_h^2 is the variance of the random intercept across data sites.

Our prior on the spline coefficient for the non-linear local component is given as:

$$\beta_s^l \sim \mathbb{N}(0, \sigma_l^2) \tag{3.18}$$

where σ_l^2 is the variance of the spline coefficients over space and time. This parameter fundamentally controls the smoothness of the local non-linear effect.

The remaining hyperparameters of the model include σ_r^2, σ_h^2 and σ_l^2. The standard deviation parameter σ_h represents the variability in the site specific vertical shift after taking account of the local linear trend. As it is measured in meters it is more interpretable in a physical context and so we place an informative prior here. The vertical shifts can be quite variable with some sites sitting many meters above or below others. From revisiting the publications associated with our data (see Appendix 3.A for the full list), shifted values spanning more than 5m seem unlikely. As a result, we specify the standard deviation to have a Cauchy distribution with mode 2.5m but with a wide scale of a further 2m (Gelman, 2006). For the variability of the spline coefficients across the knots, i.e. σ_r^2 and σ_l^2, we expect considerably smaller variation but we have less information, thus we use a truncated Cauchy distribution centred on zero and with scale value 1.

3.5.6 Model Fitting and Computational Details

In previous sections, we have described our Bayesian hierarchical model structure using GAMs and the manner in which we account for uncertainty. In this section, we address how to fit our NI-GAM model and the decisions that influenced our model fitting strategy. We are constrained because we have to fit the model twice as part of the noisy-input uncertainty method, described in Section 3.5.4. We also found that a single model fit yielded poor convergence due to the unavoidable confounding between the regional, vertical shift, linear, and non-linear local terms. Thus, we use the two-stage NI process to our advantage by fitting a slightly reduced model at the first stage to estimate the regional term, and using the posterior as strong prior information in the second stage to provide the estimate of the non-linear local term. Our approach has some similarities to that of cut feedback or modularised Bayesian models (Plummer, 2015), but we do not explore these avenues further here.

The two steps of our model fit are:

1. We first fit a simplified version of our process level model where we replace $f(\mathbf{x}, t)$ with $f^*(\mathbf{x}, t)$, defined as:

$$f^*(\mathbf{x}, t) = r(t) + g(z_{\mathbf{x}}) + h(z_{\mathbf{x}}) \tag{3.19}$$

This removes the non-linear local component and so avoids the confounding issue. From this model fit we calculate the first derivative of the posterior mean of $f^*(\mathbf{x}, t)$. The resulting slope estimate for each observation provides a corrective variance term, $s^2_{t_{ij}} = \tilde{s}^2_{t_{ij}} \partial^2_{\tilde{f}^*_{ij}}$. This term is added to the other model error variances for the fit in the second stage.

2. In the second step we fit the complete process model as defined in Section 3.5.4. The only changes being: (1) the addition of the new noisy-input measurement variance term; (2) the prior distribution on the regional spline terms now being $\beta^r_s \sim \mathbb{N}(m^r_s, (s^r_s)^2)$ where m^r_s and s^r_s are estimated in the first model run; and (3) the prior distribution on the vertical offset term being $\beta^h_s \sim \mathbb{N}(m^h_s, (s^h_s)^2)$ where as above m^h_s and s^h_s are estimated in the first model run.

In effect the second model fitting stage simply becomes a means by which the full error uncertainty is accounted for and the residuals are decomposed into a pure error and a non-linear local space-time effect. Whilst all our subsequent results are presented based on the second model fit, this stage is essentially only useful for providing interpretation of the model error and the degree to which local factors drive deviations from the main regional effect.

At each stage our models are written using the Just Another Gibbs Sample (JAGS Plummer, 2003) software, which in turn is based on the on Spiegelhalter et al. (2002). The JAGS language uses Gibbs sampling and the Markov Chain Monte Carlo (MCMC) algorithm to draw samples from the posterior distribution of the unknown parameters. We implement our approach using the `rjags` package in R (Plummer et al., 2016). For our models, we used 5000 iterations with a burn-in value of 1000, thinning at 5 and 3 chains. Convergence diagnostics for the parameters are investigated using the coda package (Plummer et al., 2006) and the ShinyStan app which provides an interactive visualization tool for investigating model convergence (Gabry and Goodrich, 2017). All convergence diagnostics were checked and ensured to be satisfactory before the model results were interpreted. The code and data for our model can be found here.

3.6 Model Validation

To assess the robustness of our model, we employed various techniques, including a 10-fold cross-validation (10-CV), a residual analysis, a GAM sensitivity study and a convergence diagnostics inspection, all detailed in this section. The outcomes from these analyses affirm the appropriateness of our model setup and validate the accuracy of our model fits.

3.6.1 10-Fold Cross Validation

We run the 10-CV exclusively for the 21 proxy sites since the tide gauge records are short in duration and would provide relatively weak information about model performance. We present the site-specific results for only the four case study sites with the remaining sites shown in Appendix 3.B. We evaluate the model performance based on out of sample empirical coverage and the Root Mean Squared Error (RMSE). The prediction intervals are created using posterior predictive simulations with the full error structure, i.e. $\hat{y}_{ij} \sim N(\hat{f}_{ij}, \sigma^2_{y_{ij}} + \sigma^2_{t_{ij}} + \sigma^2)$. The empirical coverage provides the percentage of occasions that the true RSL observation is within the model prediction interval (PI) for RSL. The RMSE provides insight into prediction performance in the same units as the response (meters).

The 10-CV for our full model using the 21 proxy sites obtained overall empirical coverage of 99.1% with the 95% prediction interval and 78.6% with the 50% prediction interval. These are satisfactory for a model fitted to complex data such as ours, especially with the addition of the model error. The conservative coverage values are likely a consequence of accounting for the observed measurement errors in the estimation of the prediction intervals. The RMSE for the 21 proxy sites is 0.14 m. An average out of sample error of 14 cm is reasonable given the scale and variety of the data set.

Table 3.1 provides a site-specific insight into the empirical coverage for the model and the size of the prediction intervals. Three out of the four sites have a coverage of 100% due to the large prediction intervals arising from the bivariate uncertainties associated with the proxy data (Figure 3.2). Based on the RMSE, the best fitting case study site is Cedar Island in North Carolina, where the RMSE is 6cm (Table

3.1). At the other end of the spectrum is Swan Key Florida, where the RMSE is larger at 19 cm (Table 3.1). Figure 3.3 presents the true RSL observations versus the model-based RSL point estimates with 95% prediction intervals at each site. The size of the prediction interval corresponds to the bi-variate uncertainties inherent in the proxy records, with Swan Key Florida exemplifying pronounced uncertainties, as evident in the large grey uncertainty boxes in Figure 3.2. To reiterate, the results presented here are used as case studies, while the model is run on the full dataset of 21 proxy sites and remaining sites are presented in Figure 3.B.6.

Site Name	95% Empirical Coverage	95% Average PI width	50% Empirical Coverage	50% Average PI width	RMSE (m)
Placentia, Newfoundland	0.97	0.36	0.59	0.12	0.11
East River Marsh, Connecticut	1.00	0.52	0.77	0.18	0.13
Cedar Island, North Carolina	1.00	0.26	0.78	0.09	0.06
Swan Key, Florida	1.00	0.77	1.00	0.26	0.19

Table 3.1: Empirical coverage from the 10 fold cross validation and the corresponding size of the prediction intervals (PI) used for model validation for our 4 chosen sites.

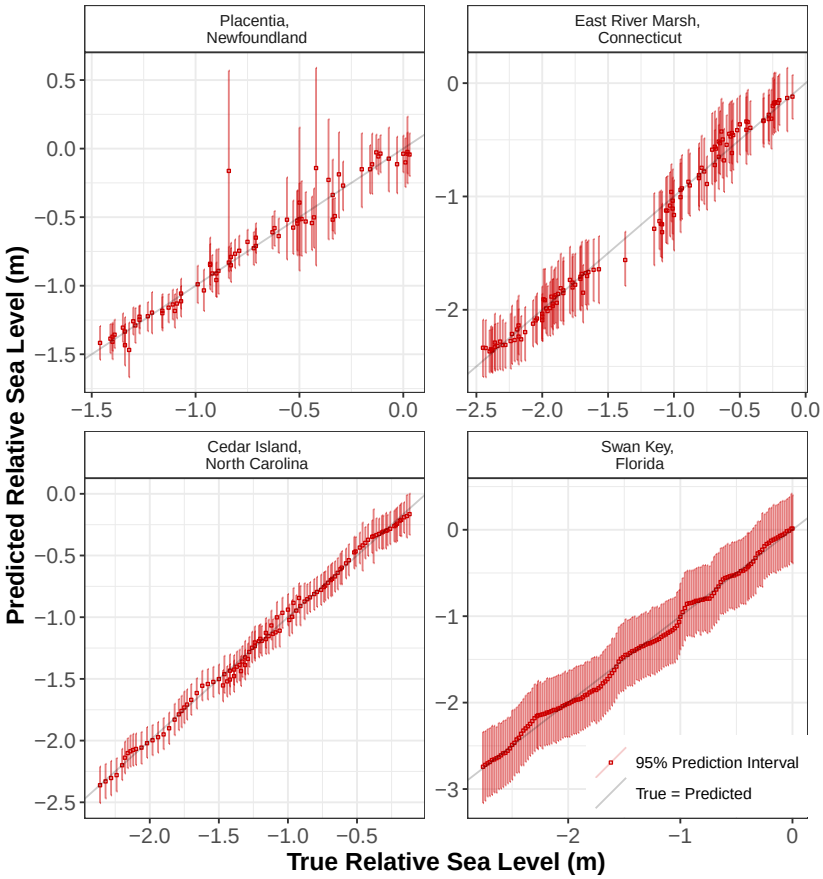


Figure 3.3: True vs Predicted RSL values for our case study sites at Placentia Newfoundland, East River Marsh Connecticut, Cedar Island North Carolina and Swan Key Florida using 10-fold cross validation (CV). The axis scales, expressed in meters, vary for each site. The predicted means are the red points with a vertical 95% prediction interval for each point. The identity line is shown in grey.

3.6.2 Residual Analysis

To evaluate the efficacy of our models, we analyse the residuals, representing the differences between observed and predicted values, through residual plots (Baddeley et al., 2005). In Figure 3.4, our residual plot provides insights into model performance, aiding in the identification of patterns, trends, or systematic errors in the data. Specifically, Figure 3.4 (a) reveals a discernible trend, indicating that the model has not fully captured underlying patterns in the data. In contrast, Figure 3.4 (b) illustrates the second model run, which eliminates the discernible trend, presenting residuals that are random and evenly spread around the hori-

zontal axis. This analysis justifies the adoption of a two-step modeling process, as outlined in section 3.5.6.

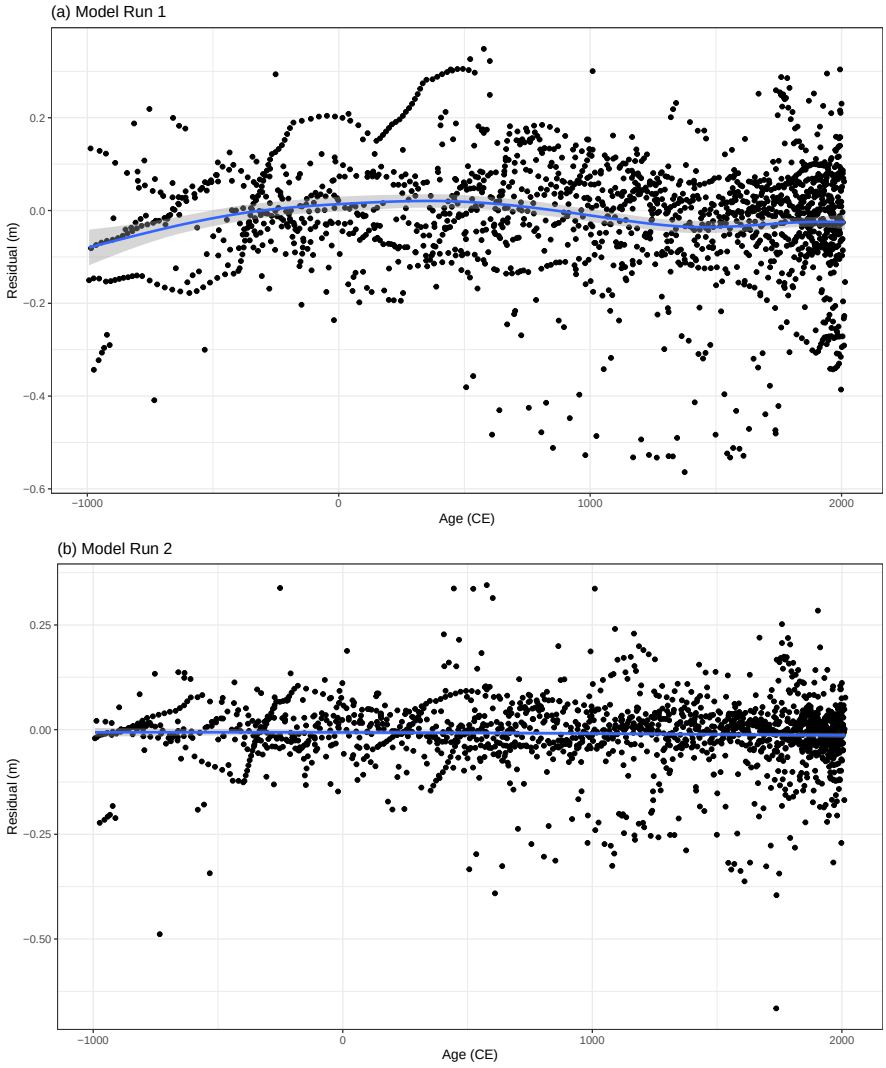


Figure 3.4: Residual analysis for the noisy input generalised additive model highlighting. (a) Model run 1 represents the first step of our modelling approach and the residuals possess evidence for an underlying trend. (b) Model run 2 presents the second step of our model which removed the underlying trend, highlighting the requirement of the two step modelling approach.

3.6.3 GAM Sensitivity Study

As mentioned in Section 3.5.3, we use B-splines for our regional and non-linear local terms. In this section, we utilise a GAM sensitivity study to highlight the

range of basis functions settings examined and how our model choices provided satisfactory model fits as well as computational efficiencies.

Initially, we assessed the capabilities of the `mgcv` package developed by Wood and Wood (2015), which facilitates the fitting of various smoother functions and generalized additive models. However, this package lacks provisions for Bayesian analysis or the incorporation of uncertainties associated with input and output variables. Subsequently, we explored the `jagam` package by Wood (2016), leveraging the functionalities of `mgcv` while automatically generating corresponding JAGS model code, enabling Gibbs sampling. Nevertheless, challenges emerged when incorporating informed priors for the linear local component, leading to issues with model fits at certain locations. Consequently, to enhance flexibility in the model structure, we developed our own basis functions and JAGS scripts.

Following a comprehensive examination, we scrutinized both the individual components and the overall fit of the model. Our residual analysis underscored the necessity for diverse levels of components and multiple model runs to effectively address uncertainty and capture inherent patterns in the data. Commencing with the most basic model involving a spline in time, we systematically progressed through a spline in space-time, eventually incorporating random effect terms. It is crucial to highlight that the structure and smoothness of basis functions associated with our regional (spline in time) and non-linear local components (spline in space-time) play a pivotal role in enhancing the performance and interpretability of statistical models.

Regarding the regional component, we conducted comparisons with B-splines (Hastie and Tibshirani, 1986) and P-splines (Eilers and Marx, 1996). Within our model, we observed that incorporating the P-spline addressed the knot selection issue; however, it introduced an additional parameter that affected the convergence of the complex NI-GAM. Consequently, we concentrated on B-splines and tuning the number of knots and validating our decision with 10-fold CV as previously mentioned. Figure 3.5 displays the overall model fit for the NI-GAM, showcasing a subset of sensitivity tests conducted to explore the interplay between model smoothness and efficiency focusing on the regional component. The plot reveals

that the optimal fit for the data is achieved by selecting a range of knots between 13 to 33 knots for the regional component, effectively capturing the underlying data variability a configuration adopted in our paper.

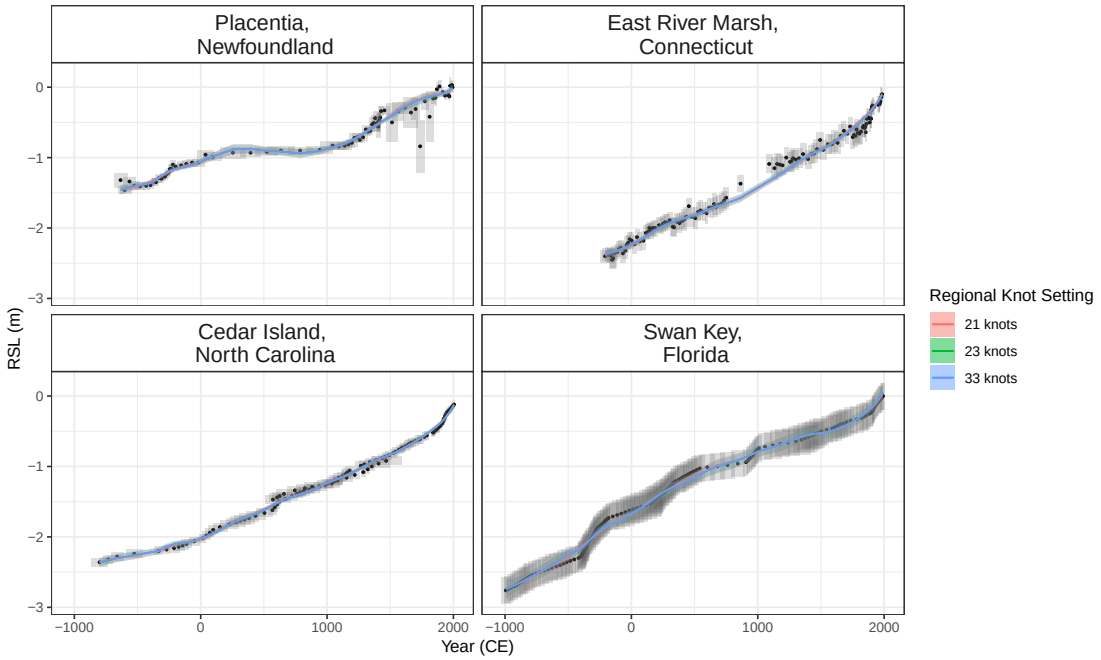


Figure 3.5: Total model fit for the noisy input generalised additive model using a range of knot values for the regional component. The grey boxes represent the bi-variate uncertainties associated with the proxy records and the black dot is the midpoint of the box. The coloured lines with corresponding 95% credible intervals represent a subset of sensitivity tests used to find the optimum regional knot setting.

In our exploration of the regional component's sensitivity, achieving the optimal balance in model fit, addressing identifiability challenges, and pinpointing the optimum number of knots required a careful assessment of both the total model fit and the corresponding component plot. In Figure 3.6, we illustrate that the optimal knot setting for the regional component has been identified as 23 knots. This choice proves more effective, significantly reducing uncertainty compared to the alternative with 21 knots, and effectively balancing the variability among the different components of the NI-GAM. In instances where one component exhibits less flexibility, the remaining component tends to become more variable. The application of the 10-fold cross-validation (CV) technique facilitated the determination of

this delicate balance among the various components.

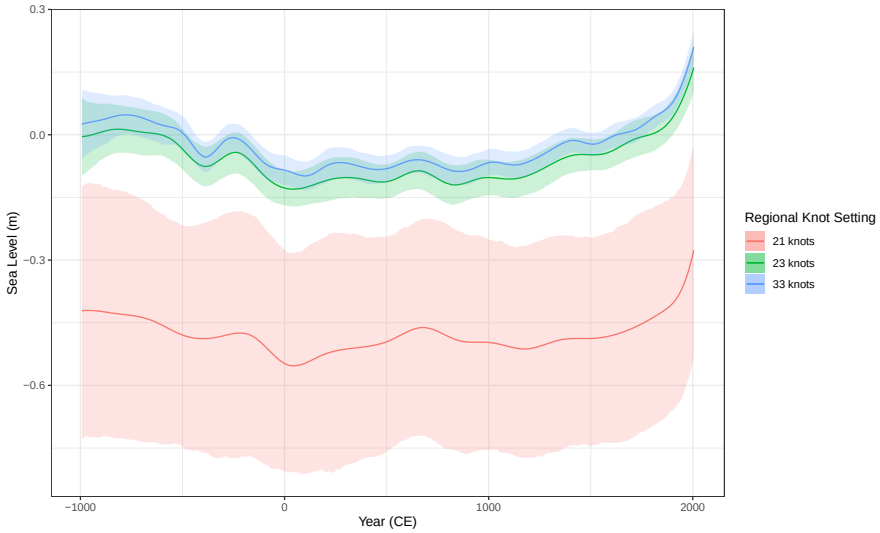


Figure 3.6: The regional component for the noisy input generalised additive model with the coloured lines and corresponding 95% credible intervals highlighted a subset of the knot settings examined during our sensitivity analysis.

A similar GAM sensitivity analysis was carried out for the non-linear local component. Figure 3.7 displays the overall model fit for the NI-GAM, showcasing a subset of sensitivity tests conducted to explore the interplay between model smoothness and efficiency focusing on the non-linear local component.

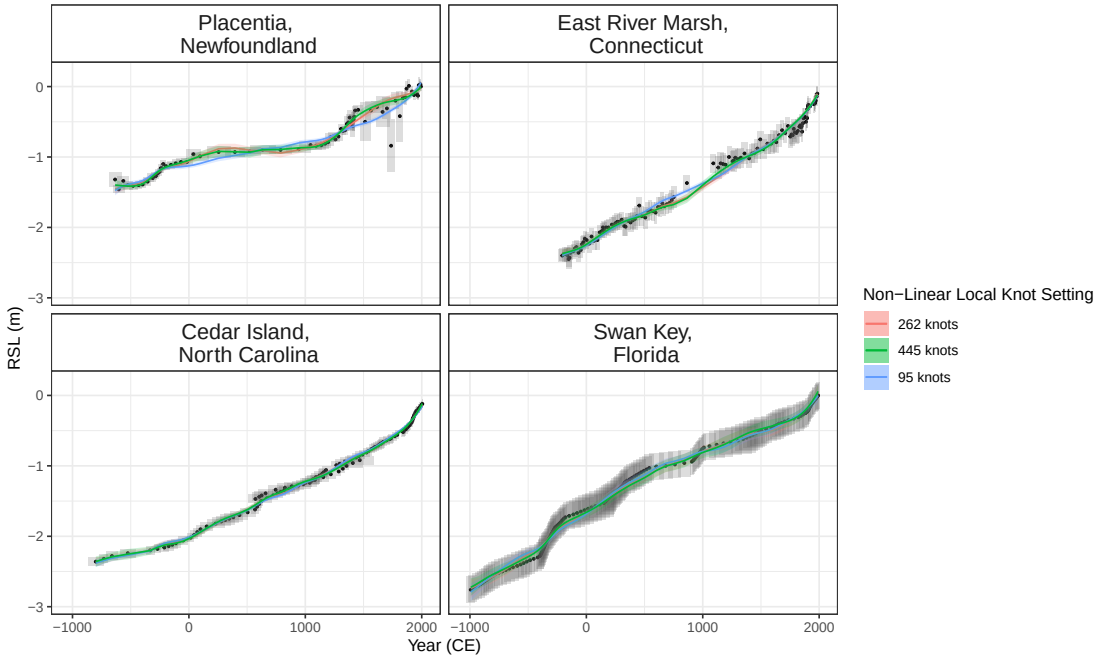


Figure 3.7: Total model fit for the noisy input generalised additive model using a range of knot values for the non-linear local component. The grey boxes represent the bi-variate uncertainties associated with the proxy records and the black dot is the midpoint of the box. The coloured lines with corresponding 95% credible intervals represent a subset of sensitivity tests used to find the optimum non-linear local knot setting.

Figure 3.8 demonstrates a subset of the knot settings examined for the non-linear local component and the impact of these settings on the component and the overall fit of the model. As a result, we found a simpler and faster quadratic polynomial basis, with 262 knots, the most approachable choice using our GAM sensitivity analysis and the 10-fold CV.

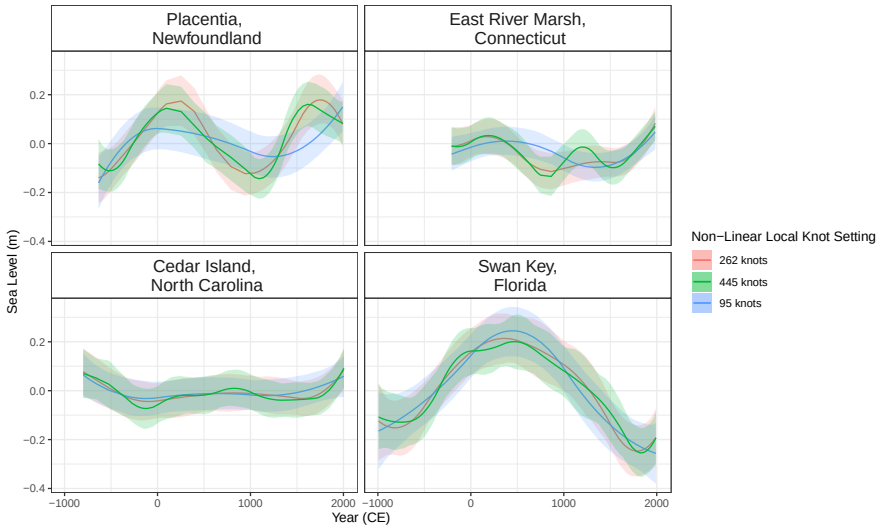


Figure 3.8: The non-linear local component for the noisy input generalised additive model with the coloured lines and corresponding 95% credible intervals highlighted a subset of the knot settings examined during our sensitivity analysis.

In this GAM sensitivity study, we have demonstrated how basis functions serve as fundamental components, playing a crucial role in representing the relationship between predictors and the response variable in a flexible and non-linear manner. We have presented a subset of our sensitivity study, which helps address the modeling decisions we made to ensure the smoothness of the basis functions. We complemented this sensitivity study with the 10-fold CV techniques, as previously mentioned, to assess the goodness-of-fit for different numbers of knots and determine the optimal number of knots that balance model complexity and efficiency.

3.6.4 Convergence Diagnostics

In this section, we use convergence diagnostics, specifically employing the ShinyStan app (Gabry and Goodrich, 2017), to ensure the well-convergence of our model, signifying that sampled values accurately represent the posterior distribution of the parameters. We present the results for one hyperparameter, σ_h, representing the variability of the site-specific vertical offset component as described in Section 3.5.5. The model convergence plots for the remaining hyperparameters, σ_l and σ_r, are provided in Section 3.2.2.1.

Model convergence in the context of Bayesian statistical analysis, particularly with JAGS, refers to the process by which the Markov chain generated by the sampler reaches a stable and stationary distribution. To assess model convergence in JAGS, we use several diagnostic tools which include:

- Gelman-Rubin Statistic (R-hat): This diagnostic compares the variance within multiple chains to the variance between chains (Gelman and Rubin, 1992a). An R-hat close to 1 indicates convergence.
- Trace Plots: Examining the trace plots of parameters across iterations can provide visual cues about convergence. A stable and stationary trace indicates good convergence.
- Autocorrelation Plots: Monitoring the autocorrelation of the chain can reveal how quickly the chain forgets previous values. Rapid decay in autocorrelation indicates better convergence.
- Effective Sample Size: This metric estimates the number of independent samples, considering autocorrelation. A higher effective sample size is desirable.
- Density plots: Visual tools for assessing the convergence and distribution of sampled parameters. These plots display the estimated probability density function of the parameter values.

Figure 3.9 demonstrates the results of a comprehensive range of model convergence tests, as described earlier, utilizing the output of the ShinyStan app (Gabry and Goodrich, 2017). The Rhat value of 1 signifies convergence, and the large n_eff (Effective Sample Size) value indicates robust convergence. The Kernel Density Estimate visualises the distribution of the hyperparameter σ_h. The autocorrelation plot exhibits a rapid decay, and the traceplot, involving 3 chains, displays a stable and stationary trace, affirming good convergence. These results collectively demonstrate that our model has converged, with the Markov chain sufficiently exploring the posterior distribution, rendering the results reliable for statistical inference.

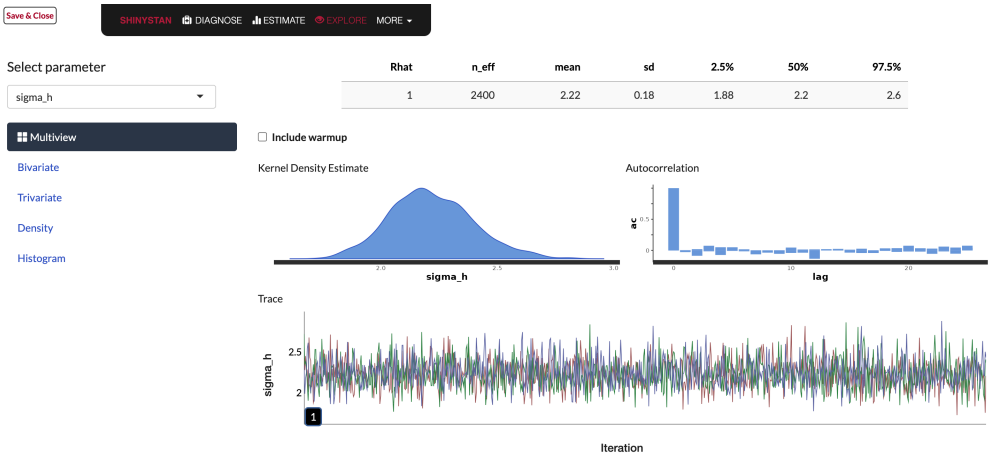


Figure 3.9: Convergence tests for the hyperparameter σ_h, which captures variability in the site-specific vertical offset component of the noisy input generalized additive model, were conducted using the ShinyStan app (Gabry and Goodrich, 2017). The Gelman-Rubin Statistic (Rhat) value is 1, indicating convergence, and the large Effective Sample Size (n_eff) value further supports robust model convergence. Visualisations, including the Kernel Density Estimate and autocorrelation plot, demonstrate a well-converged model. The traceplot, based on 3 chains, exhibits a stable and stationary trace, confirming the model's good convergence.

3.7 Results

In this section we present the results from our Bayesian hierarchical RSL model. We consider the full model fit and its decomposition into the different components of RSL, i.e., regional component, linear local component and non-linear local component.

3.7.1 Full Model Fit and Rate of Change

The full model fit for the four case study sites are shown in Figure 3.10 (results from all 21 proxy sites are included in Appendix 3.B). The model demonstrates how the RSL varies in time at each site. Overall, the model fits the data well. The 95% credible intervals for Swan Key Florida are larger due to the relatively large observation uncertainties at this site and the fit is notably smoother than the others. The data in Placentia, Newfoundland experiences additional variability in the observations compared with the other sites and this is reflected in a more variable total model fit.

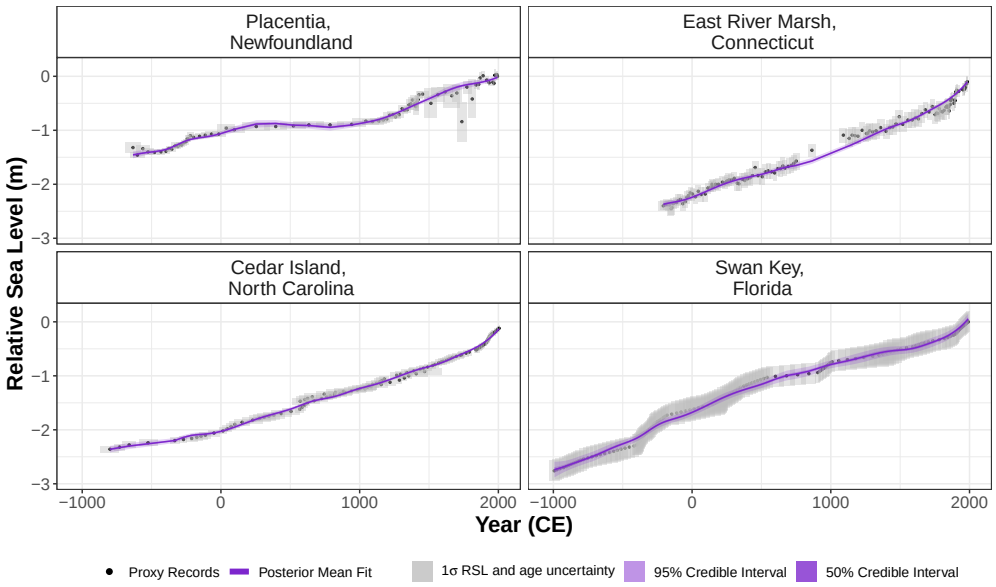


Figure 3.10: The noisy-input generalised additive model (NI-GAM) fit for four selected sites along the Atlantic coast of North America. The four sites include: Placentia, Newfoundland, Canada; East River Marsh, Connecticut, USA; Cedar Island, North Carolina, USA; and Swan Key, Florida, USA. The black dots and grey boxes represent the midpoint and associated uncertainty, respectively, for each proxy record. The solid purple line represents the mean of the model fit with a 95% credible interval denoted by shading.

Figure 3.11 shows the site-specific rates of change for the case study locations calculated as described in Section 3.5.3. The remaining sites are shown in Appendix 3.B. Late Holocene rates of RSL change display century to multi-century scale variability around a stable mean at each site until the 19th and 20th centuries since when the rate of rise appears unprecedented. Rates fluctuate throughout the last 2000 years but remain below 1.5mm/yr until the late 1800s in East River Marsh, Connecticut, and the early to mid 1900s in Cedar Island, North Carolina and Swan Key, Florida. The late 20th and 21st century rates at these sites are unprecedented in the last 2000 years with the most recent rates of change being 3.06 ± 0.3, 2.9 ± 0.5 and 2.9 ± 0.7 mm per year in East River Marsh Connecticut, Cedar Island North Carolina and Swan Key Florida respectively. Placentia Newfoundland does not appear to experience the same uptick in rates that the other sites do with the most recent rate being 1.21 ± 0.4 mm per year.

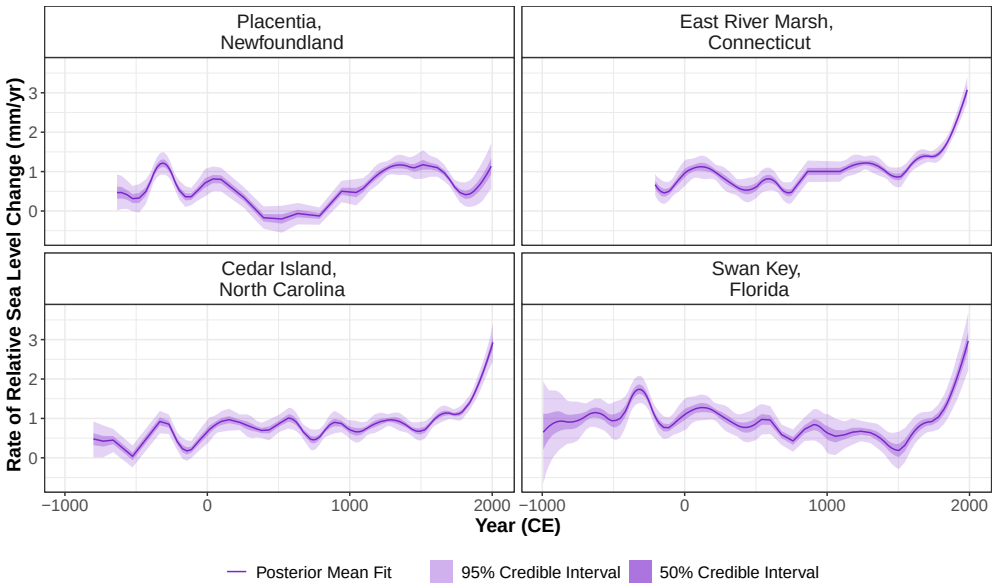


Figure 3.11: Rate of relative sea change found by taking the first derivative of the total model fit for four sites along the Atlantic coast of North America. The mean of the fit is the solid purple line with the dark shaded area being the 50% credible interval and the light shaded area being the 95% credible interval.

3.7.2 Examining the Decomposition of RSL

The RSL process level f consists of the regional component, the linear local component, the site-specific vertical offset, and a non-linear local component, all as described in Section 3.5. Figure 3.12 illustrates the decomposition in our case study sites and provides an insight into how the components of RSL have varied over time for the Atlantic coast of North America by demonstrating the scale of the different components and how they interact over time. The total posterior model is obtained by the sum of these separate components as illustrated by the purple line in Figure 3.12.

It is evident that the dominant driver of RSL change for these four sites until the late 1800s is the linear local component. After this interval, regional variability along the Atlantic coast of North America appears to take over and the total RSL trends at each site tend to reflect the RSL rise seen in the regional component. The non-linear local component is picking up the remaining variability and demonstrates that non-linear local effects on RSL variability are more apparent in

Placentia and Swan Key compared to Cedar Island and East River Marsh.

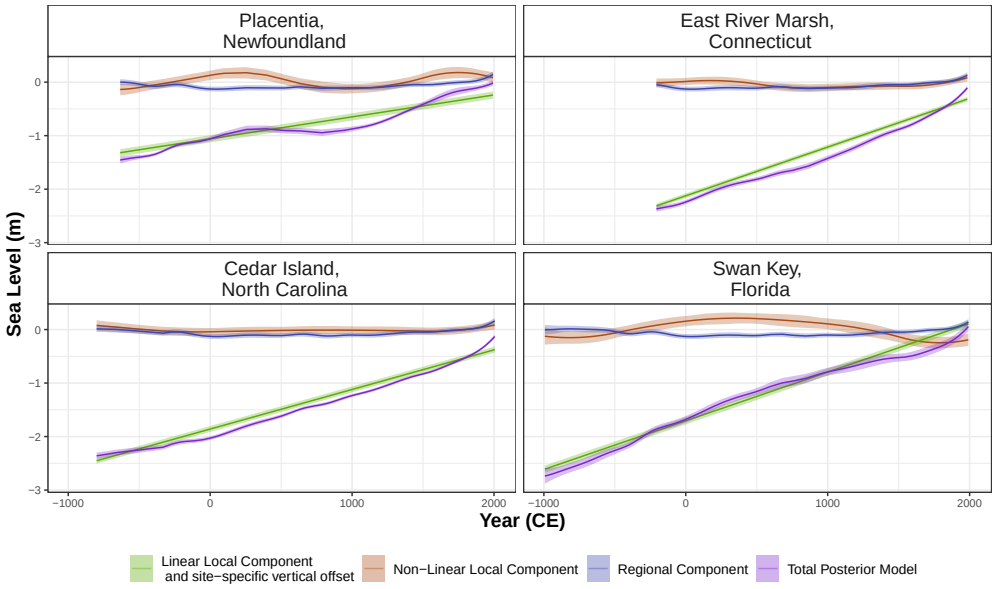


Figure 3.12: The decomposition of the relative sea level process level for the four sites, with solid lines indicating means and shaded areas 95% posterior credible intervals. The blue curve represents the regional component. The brown curve represents the non-linear local component. The green line represents the site-specific vertical offset plus the linear local component. The purple line is the sum of all three components and represents the full noisy-input generalised additive model fit.

Figure 3.13(a) shows the regional component (common to all sites) for the 21 proxy sites and 66 tide gauge sites along the Atlantic coast of North America. As a reminder, the regional component is represented with a spline in time which is common across all sites. Prior to 0 CE, sea level fluctuated from 0.01 m to -0.13 m. From 0 CE to 1200 CE, variability of sea level reduced ranging from -0.12 m to -0.09 m. Following 1200 CE, a sharp increase in sea level can be seen with brief periods of stability from 1410 CE to 1560 CE and from 1800 CE to 1840 CE. After 1800 CE, sea levels are consistently rising and the most dramatic increase can been seen from the mid-1800s until the present day. Figure 3.B.1 in Appendix 3.B demonstrates the underlying behaviour of the posterior samples for the regional component. Figure 3.13(b) shows the rate of change for the regional component along the Atlantic coast of North America. Rates fluctuate around 0 CE between -990 CE and 1800 CE after which a continuous increase can be see from 1800 CE

onwards. The rate from the late 20th century is unprecedented when compared with the last 3000 years and is estimated to be 1.8 ± 0.5 mm per year. This 20th regional sea-level rate depicts patterns over multi-decadal to centennial timescales due to the resolution limits and natural time-averaging of proxy reconstructions and decadally averaged of tide-gauge data.

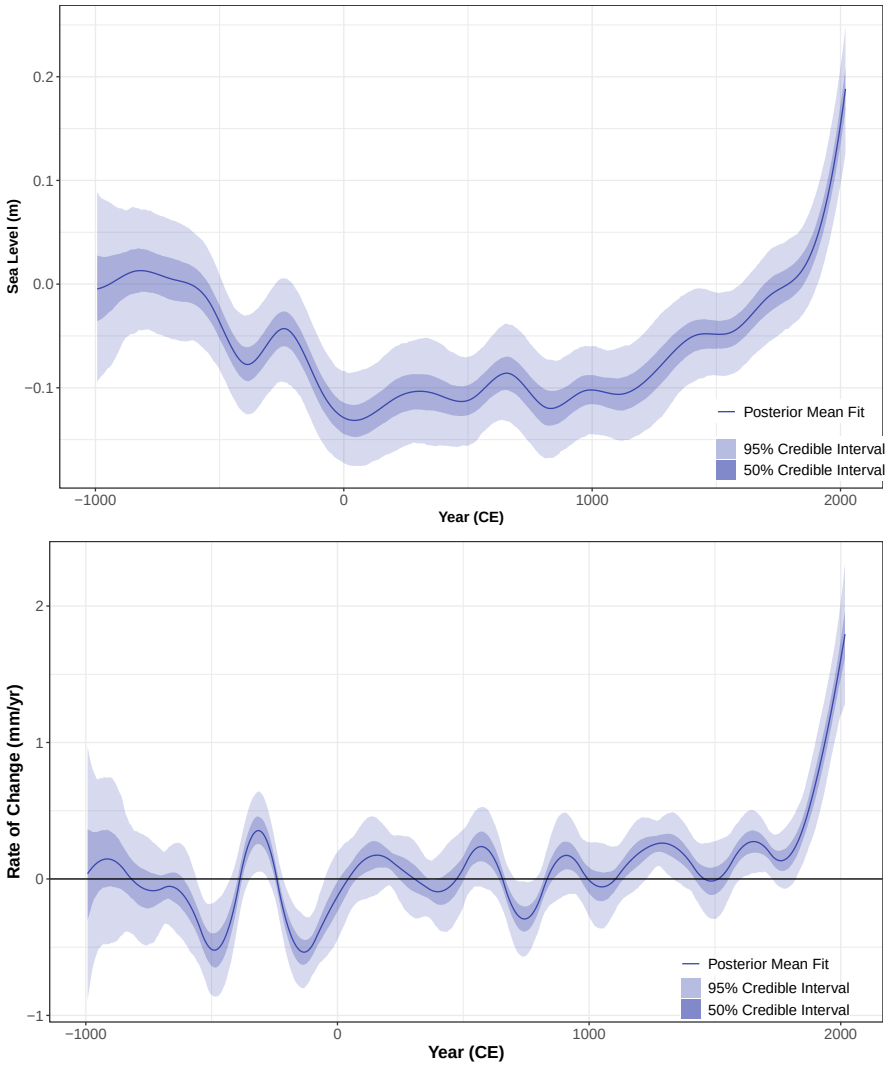


Figure 3.13: The noisy-input generalised additive model (NI-GAM) results for (a) the regional component and (b) the regional rate of change component. (a) The regional component mean model fit represented with a solid line and the shading indicating the 50% credible interval in dark blue and 95% credible interval in light blue. The y-axis is the sea level in m with the x-axis representing the time across the last 3000 years for the Atlantic coast of North America. (b) Rate of Change for the regional component for the Atlantic coast of North America with the solid line representing the mean of the fit, the dark blue shaded area representing the 50% credible interval and the light blue shaded area representing the 95% credible interval. The y-axis is the instantaneous rate of change of regional sea level in mm per year.

The linear local component is represented with a random slope effect as described in Section 3.5. As stated in Section 3.3, this parameter removes a long term variation driven principally (but perhaps not exclusively) by GIA. Table 3.2 compares

our empirically estimated values of this parameter for the proxy record sites in mm per year prior to 1800 CE. It is clear that there is wide variability in these values between sites with areas such as Swan Key Florida and East River Marsh Connecticut experiencing rates of 0.91 mm/yr whereas Placentia Newfoundland has just 0.41 mm/yr. To show a comparison with physical model-based GIA rates, Table 3.2 presents values obtained from the ICE5G - VM2-90 Earth-ice model (Peltier, 2004). It is evident that the data-driven rates and the GIA-model rates differ, with Swan Key Florida experiencing the greatest difference of 0.8mm/yr. Whereas, East River Marsh Connecticut has similar rates with a difference of only 0.05mm/yr.

	Empirical Rate prior to 1800 CE [mm/yr]	ICE5G-VM2-90 Earth-ice GIA rate [mm/yr] (Peltier, 2004)
Placentia, Newfoundland	0.41	0.21
East River Marsh, Connecticut	0.91	0.96
Cedar Island, North Carolina	0.74	0.69
Swan Key, Florida	0.91	0.11

Table 3.2: Linear local component for our four sites along the Atlantic coast of North America given in mm per year. The empirical rates represent the rate estimated from the data prior to 1800 CE, which is used to inform the priors for the linear local component (Neukom et al., 2019). ICE5G-VM2-90 Earth-ice GIA rate is from an Earth-ice physical model (Peltier, 2004).

Figure 3.14 shows our non-linear local component that represents the spatially structured behaviour specific to each site. There are clearly different patterns of non-linear local sea-level change, which is to be expected given that the common source of variation across all sites has been captured by the regional component. Placentia Newfoundland and Swan Key Florida show non-linear local variations in sea level ranging from 0.19 m to -0.12 m and 0.21 m to -0.25 m respectively. On the other hand, Cedar Island North Carolina and East River Marsh Connecticut do not experience this level of variability with sea levels in the non-linear local component fluctuating close to zero.

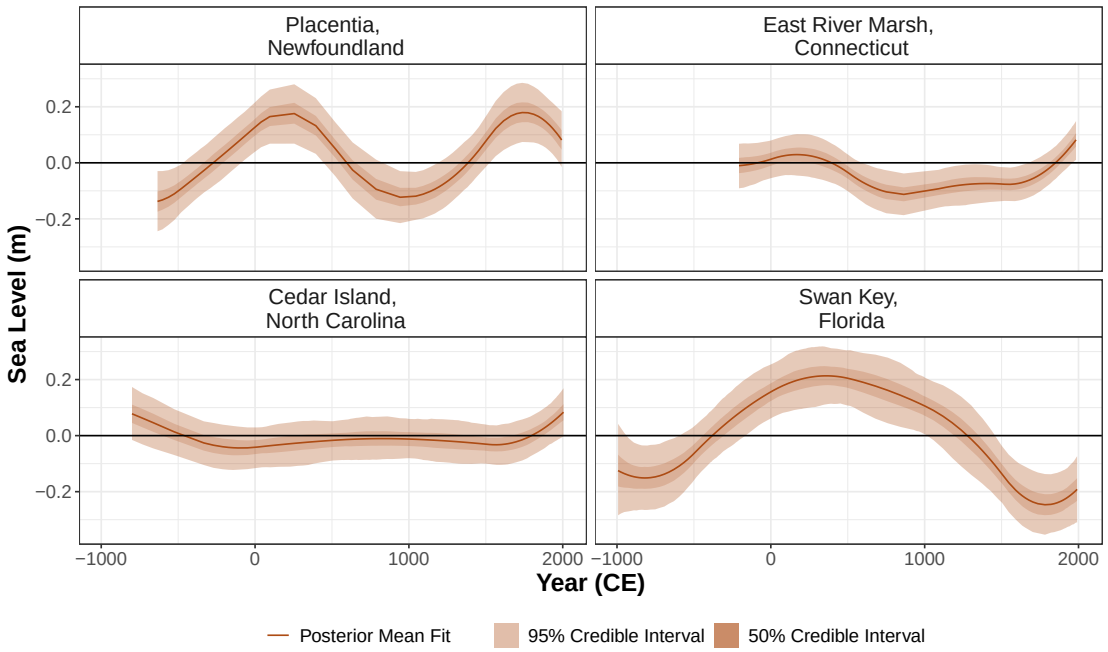


Figure 3.14: The non-linear local component for our four sites along the Atlantic coast of North America. The y-axis represents sea level in meters. The brown solid line represents the mean of the model fit with the 50 % credible interval in dark brown shading and 95% credible interval in the light brown shading.

3.8 Discussion

RSL change is the net result of multiple physical processes within the oceans, atmosphere, and solid Earth, that can alter the height of the land and/or sea surface (Church et al., 2001). The importance of specific processes varies markedly across space and through time, giving rise to a complex and evolving pattern of RSL change. Tide-gauge data and proxy records contain information about these processes and the sea-level community requires advanced statistical tools to decompose the net RSL signal into contributions from physical processes while accounting for uncertainties in the underlying data. The need for RSL decomposition motivated current modeling strategies such as the K16 model and our approach presented here.

Our approach provides a more computationally efficient method for decomposing the RSL signal. The process level of our model utilises a spatio-temporal field decomposed into: a regional component; a linear local component; and a non-linear

local component. In contrast to K16 which uses GPs, we use splines to examine the different drivers of RSL. This is due to the computational complexity associated with the likelihood computation for a Gaussian Processes being of $O(n^3)$ where n is the number of data points. In contrast, the likelihood computation for the equivalent spline with pre-computed basis functions is just $O(n)$ (Wood, 2017a). Thus our model can be fitted quicker than K16, allowing for further checks on the performance of our model. The model validations presented in Section 3.6 highlight this and we are confident that the NI-GAM is effectively capturing the different components of RSL along the Atlantic coast of North America. In addition, the construction of our GAM using spline basis functions and random effects allows for easy interpretability without the need for covariance matrices and correlation functions (Porcu et al., 2021). We can efficiently model late Holocene RSL changes along the Atlantic coast of North America and the interpretability of GAMs allows for these changes to be easily examined (Figure 3.10).

Our approach attempts to deviate from the Empirical Bayesian framework as implemented by K16 and related models. Piecuch et al. (2017) demonstrated that Empirical Bayesian methodologies can underestimate uncertainty when examining historic sea-level change along the Atlantic coast of North America. However, we recognise the difficulty of a fully Bayesian approach due to the confounding nature of the regional, linear local, and non-linear local components. Instead we opted to take advantage of the two-step fitting required by our use of the noisy-input method to take account of age errors. The first step of the modeling procedure obtains posterior distributions for the regional component and the site-specific vertical offset. The second step uses the resulting posterior estimates and uncertainties to inform the priors for the remaining linear local and non-linear local components, and the extra measurement variance contribution from the age uncertainties. The first step can be thought of as estimating the main component of our model: the regional RSL curve, with the second step designed to decompose the residuals and ensure the uncertainty is properly calibrated. Our modelling strategy avoids fixing process model parameters, and severe confounding that would occur were we to fit the model in one step.

Considering the individual RSL components, pre-anthropogenic (before 1800 CE;

Neukom et al., 2019) RSL change along the Atlantic coast of North America is dominated by the linear local component which is principally capturing the contribution from ongoing GIA. However, there are some notable differences between the empirically-estimated rates obtained from our models and the GIA rates obtained from the ICE5G-VM2-90 Earth-ice physical model (Table 3.2). There are several possible explanations for these discrepancies. First, a single Earth-ice model generates GIA predictions from a specific representation of the solid Earth (e.g., mantle viscosity and lithospheric thickness parameters) and history of deglaciation. It is unlikely that any single Earth-ice model will perfectly estimate GIA at all places and all times because the parameters are uncertain and may vary by location (Roy and Peltier, 2015). In particular, locations close to the margins of former ice sheets (such as Newfoundland) may exhibit particularly pronounced differences in GIA estimated by different Earth-ice models. Systematic difference between RSL predicted by specific Earth-ice models and proxy reconstructions on Holocene timescales is well documented in eastern North America (Vacchi et al., 2018) and elsewhere (Shennan et al., 2018). One such example is Placentia Newfoundland where our empirically-estimated rate and the Earth-ice physical model GIA rate differ by 0.1 mm/yr (Table 3.2). Second, physical Earth-ice models only estimate the contribution from GIA, while the empirical approach captures contributions from other processes such as vertical land motion from tectonic process which may also be a linear driver of RSL change on the timescales under consideration. Although these non-GIA processes may be small on the passive margin of the Atlantic coast of North America, they are also unlikely to be zero at all sites. For example, Khan et al. (2022) identified an anomalously high rate of rise at Swan Key Florida compared to nearby Snipe Key Florida and proposed that dissolution of the carbonate bedrock beneath the site resulted in an additional approximately linear component of RSL rise. This is highlighted in Table 3.2 where Swan Key Florida exhibits a large difference between the empirical rate of 0.91 mm/yr compared with the ICE5G-VM2-90 GIA rate of 0.11 mm/yr.

After $\sim$ 1900 CE the regional component dominates and we see the regional rate of change increase markedly from 0.7 ± 0.5 mm/yr in 1902 to 1.8 ± 0.5 mm/yr at the end of the 20^{th} century (Figure 3.13(b)). This change is the result of anthropogenic

forcing of the climate system (Neukom et al., 2019), which drove sea-level rise through thermosteric and barystatic processes (Frederikse et al., 2020). Our estimate of regional sea-level rise during the 20^{th} century represents trends sustained on multi-decadal to centennial timescales because of the natural time-averaging and resolution limits of the proxy reconstructions and our decadal average of tide gauge measurements. Despite our analysis being limited to the Atlantic coast of North America, our estimated rate is comparable to century-scale estimates generated using only tide gauge data (Hay et al., 2015; Frederikse et al., 2020) and the K16 statistical model (Kopp et al., 2016; Walker et al., 2022).

The diverse trends captured by the non-linear local component, as shown in Figure 3.14, highlight the important influence site-specific processes can have on the RSL. At Placentia Newfoundland the non-linear local component experiences large fluctuations with maximum peaks reaching values of $\sim$ 0.19m at around 250 CE and 1775 CE and minimum troughs of -0.12 m at around -450 CE and 1050 CE. This is a particularly pronounced degree of variability. The original study of the site by Kemp et al. (2018) recognised that the geomorphology at Placentia rendered it sensitive to site-specific RSL change due to the position of the salt marsh. The salt marsh is separated from the open ocean by a narrow inlet which is likely prone to opening and closing of the dynamic sediment barrier. In contrast, the East River Marsh record was generated exclusively through sediment in direct contact with bedrock to negate the potential influence of sediment compression as a driver of RSL change (Kemp et al., 2015). This contrast is reflected in our estimate of the non-linear local component where variability is present with a slight increase in sea level followed by a fall at around 100 CE and from 650 CE onwards an increase. Thus the component is non-zero due to the presence of other processes that can affect individual sites or groups of sites (e.g., dynamic sea level change Kemp et al., 2015).

There are a number of potential extensions to the NI-GAM model which have not been addressed in our paper. A future aim is to extend NI-GAM further to larger regions, e.g. North Atlantic, or potentially to examine global RSL trends. This poses a challenge as the network of proxy records and tide gauges is non-uniformly spread and biased to coastal regions in the Northern hemisphere. Previous at-

tempts to resolve this spatial bias have used a variety of techniques, (e.g. Jevrejeva et al., 2008; Wenzel and Schröter, 2010; Church and White, 2011; Hay et al., 2015; Dangendorf et al., 2017; Berrett et al., 2020), yet have mostly focused on instrumental data from tide gauges and satellites. Models like K16 and its extensions (e.g. Khan et al., 2017; Kemp et al., 2018; Walker et al., 2021) created a global component which may give insight into the changes in sea level common across many sites. Therefore, our model would require more components and further solutions to additional potential confounding issues. Yet, NI-GAM is an extendable modeling approach due the flexible structures of spline-GAMs and the Bayesian framework which allows for the inclusion of informed priors from future RSL analyses. Our modelling strategy is of course not limited to RSL changes. Rather it has the potential to be expanded to other areas of research that require the decomposition of a signal into different components that vary in time and space with complex measurement errors. One such example would be investigating historic temperature trends at a local and regional level to gauge the components that alter temperature spatially and temporally.

Appendix

3.A Data

Table 3.A.1 provides a list of all the proxy record sites used along the Atlantic coast of North America in our model and Figure 3.A.1 represents the proxy data associated with the 21 proxy record sites used in our model with the grey boxes the 1 σ uncertainty in the age and RSL value and the black dots the midpoint of the uncertainty box. Table 3.A.1 gives the references associated with each proxy data site and more information in regard to data collection can be sourced here. In addition, Table 3.A.1 contains the GIA rate used to inform the prior for the linear local component calculated using a linear regression for the data prior to 1800 CE for each site (Neukom et al., 2019). This provides an estimate for the rebounding effect of the tectonic plate after a glacier melts (Whitehouse, 2018). In previous models, physical GIA models are used to inform the prior for the linear local component. Peltier (2004) developed the ICE5G - VM2-90 Earth - ice which provides a GIA rate for each site. We carried out comparison between these techniques however, the data driven GIA rates were our preferred choice for the proxy records.

Table 3.A.2 gives the reference name associated with each tide-gauge data site and its' location along North America's Atlantic coast using PSMSL database (Holgate et al., 2013). For each tide gauge site, the associated GIA rate is provided which is obtained using the ICE5G - VM2-90 Earth - ice developed by Peltier (2004). The uncertainty associated with these values is selected to be 0.3 mm per year based on study carried out by Engelhart et al. (2009).

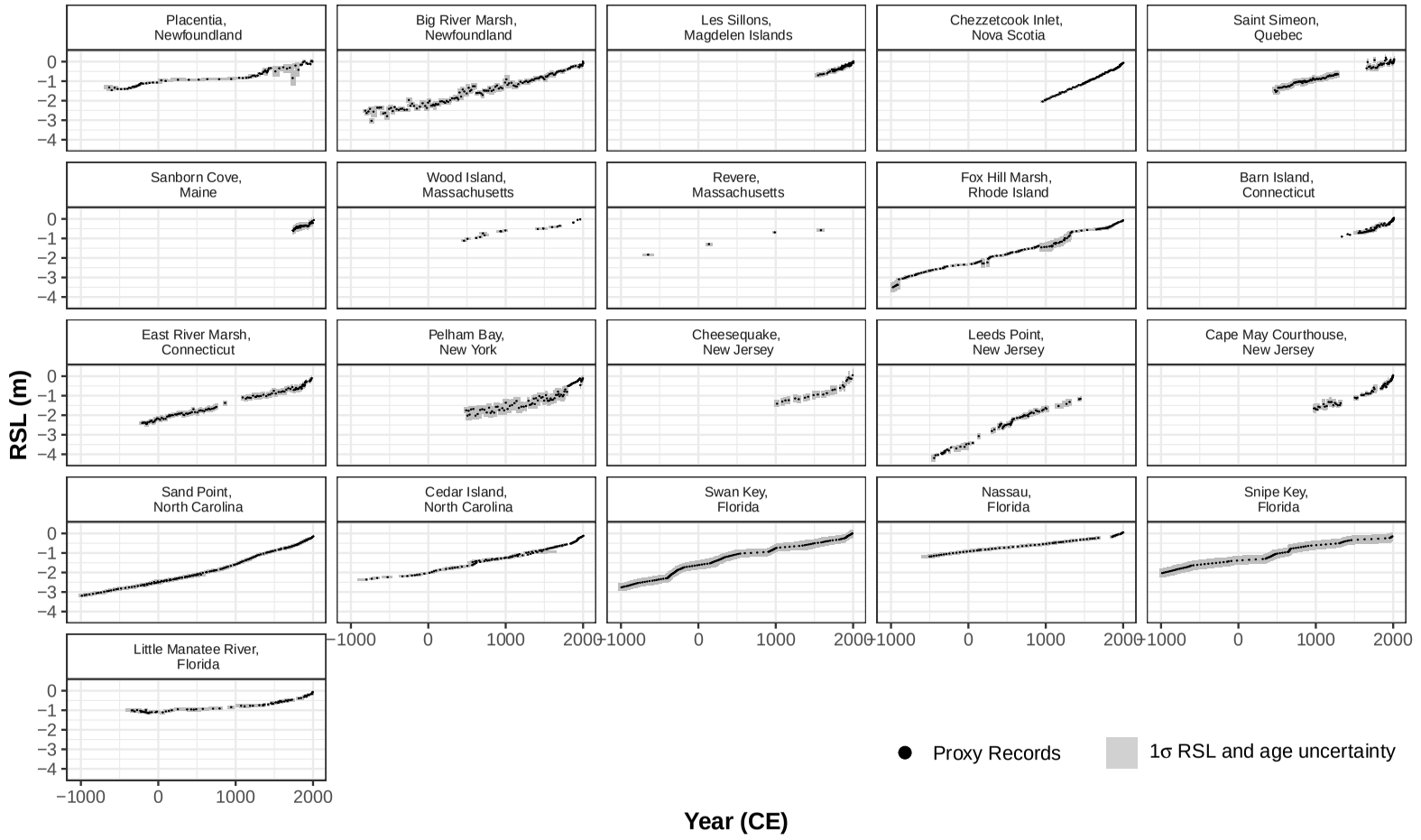


Figure 3.A.1: Proxy records for 21 sites along Atlantic coast of North America. The grey boxes represents the 1 σ uncertainty in RSL and age. The black points represents the midpoint of the uncertainty boxes which we use as the input of our data.

Reference	Site Name	Empirical rate prior to 1800 CE (mm/yr)	ICE5G GIA rate (mm/yr) (Peltier, 2004)
Donnelly et al. (2004); Gehrels et al. (2020)	Barn Island, Connecticut	1.00	0.99
Kemp et al. (2015); Stearns and Engelhart (2017)	East River Marsh, Connecticut	0.91	0.96
Kemp et al. (2014)	Nassau, Florida	0.42	0.28
Donnelly (2006)	Revere, Massachusetts	0.58	0.40
Kemp et al. (2011)	Wood Island, Massachusetts	0.52	0.40
Kemp et al. (2011, 2017)	Sand Point, North Carolina	0.97	0.69
Kemp et al. (2011, 2017)	Cedar Island, North Carolina	0.74	0.69
Kemp et al. (2013); Cahill et al. (2016)	Cape May Courthouse, New Jersey	1.19	1.24
Kemp et al. (2013); Cahill et al. (2016)	Leeds Point, New Jersey	1.69	1.41
Barnett et al. (2017)	Les Sillons, Magdelen Islands	1.23	2.20
Gerlach et al. (2017)	Little Manatee River, Florida	0.28	0.14
Kemp et al. (2018)	Big River Marsh, Newfoundland	0.87	0.60
Kemp et al. (2018)	Placentia, Newfoundland	0.41	0.21
Barnett et al. (2019)	Saint Simeon, Quebec	0.93	2.33
Gehrels et al. (2020)	Chezzetcook Inlet, Nova Scotia	1.76	0.63
Gehrels et al. (2020)	Sanborn Cove, Maine	3.23	0.08
Khan et al. (2022)	Snipe Key, Florida	0.66	0.13
Khan et al. (2022)	Swan Key, Florida	0.90	0.11
Walker et al. (2021)	Cheesequake, New Jersey	0.85	1.31
Kemp et al. (2017); Stearns and Engelhart (2017)	Pelham Bay, New York	0.8	1.31
Stearns and Engelhart (2017)	Fox Hill Marsh, Rhode Island	1.00	1.07

Table 3.A.1: The 21 data sites used in our model and the associated reference for each location. The Site Name are a combination of the site-specific name and the corresponding state. In addition, a comparison is made between the GIA rates we use from the data and GIA rates from physical models such as the ICE5G - VM2-90 Earth-ice model by Peltier (2004).

Longitude	Latitude	Site Name	ICE5G-VM2-90 GIA rate (mm/yr) (Peltier, 2004)
-54.00	47.30	ARGENTIA	0.21
-74.40	39.40	ATLANTICCITY	1.41
-68.20	44.40	BARHARBOR,FRENCHMANBAY,ME	-0.11
-76.70	34.70	BEAUFORT,NORTHCAROLINA	0.61
-63.60	44.70	BEDFORDINSTITUTE	0.49
-65.80	47.90	BELLEDUNE	1.98
-74.10	40.60	BERGENPOINT,STATENIS.	1.31
-53.10	48.70	BONAVISTA	0.36
-71.10	42.40	BOSTON	0.40
-64.00	44.70	BOUTILIERPOINT	0.49
-73.20	41.20	BRIDGEPORT	0.96
-70.60	41.70	BUZZARDSBAY	1.07
-61.90	47.40	CAPAUXMEULES	2.20
-75.60	35.20	CAPEHATTERAS,NORTHCAROLINA	0.69
-75.00	39.00	CAPEMAY	1.24
-83.00	29.10	CEDARKEYI	0.24
-63.10	46.20	CHARLOTTETOWN	1.67
-82.80	28.00	CLEARWATERBEACH	0.25
-67.20	44.60	CUTLER	0.08
-67.30	44.60	CUTLERII	0.08
-66.40	48.10	DALHOUSIE	2.12
-75.70	36.20	DUCKPIEROUTSIDE	0.67
-67.00	44.90	EASTPORT	0.08
-81.50	30.70	FERNANDINABEACH	0.42
-81.90	26.60	FORTMYERS	0.13
-70.70	43.10	FORTPOINT,NEWCASTLEISLAND	-0.30
-63.60	44.70	HALIFAX	0.49
-80.10	25.90	HAULOVERPIER	0.11
-75.10	38.60	INDIANRIVERINLET	1.13
-81.60	30.40	JACKSONVILLE	0.28
-81.00	24.70	KEYCOLONYBEACH	0.07
-81.80	24.60	KEYWEST	0.13
-73.80	40.80	KINGSPOINT,NEWYORK	1.31
-58.40	49.10	LARKHARBOUR	-0.82
-75.10	38.80	LEWES(BREAKWATERHARBOR)	1.13
-64.90	47.10	LOWERESCUMINAC	2.24
-81.00	24.70	MARATHONSHORES	0.07
-81.40	30.40	MAYPORT	0.42
-81.40	30.40	MAYPORT(BARPILOTSDOCK),FLORIDA	0.42
-80.10	25.80	MIAMIBEACH	0.11
-72.00	41.00	MONTAUK	0.99
-76.70	34.70	MOREHEADCITY	0.61
-72.10	41.40	NEWLONDON	0.99
-71.30	41.50	NEWPORT	1.07
-73.80	40.90	NEWROCHELLE	1.31
-74.00	40.70	NEWYORK(THEBATTERY)	1.31
-60.20	46.20	NORTHSYDNEY	1.51
-75.50	35.80	OREGONINLETMARINA,NORTHCAROLINA	0.91
-75.10	39.90	PHILADELPHIA(PIER9N)	1.24
-62.70	45.70	PICTOU	0.93
-72.20	41.20	PLUMISLAND	0.99
-59.10	47.60	PORTAUXBASQUES	1.19
-73.10	41.00	PORTJEFFERSON	0.96
-76.30	36.80	PORTSMOUTH(NORFOLKNAVYYARD)	0.67
-71.40	41.80	PROVIDENCE(STATEPIER)	1.07
-64.40	49.00	RIVIERE-AU-RENARD	1.12
-63.30	46.50	RUSTICO	1.67
-70.50	41.80	SANDWICHMARINA,CAPECODCANALENTRANCE	1.20
-74.00	40.50	SANDYHOOK	1.31
-52.70	47.60	ST.JOHN'S,NFLD.	0.41
-82.60	27.80	ST.PETERSBURG	0.26
-55.40	46.90	STLAWRENCE	0.59
-81.10	24.70	VACAKEY	0.07
-80.20	25.70	VIRGINIAKEY,FL	0.11
-73.80	40.80	WILLETSPOINT	1.31
-70.70	41.50	WOODSHOLE(OCEAN.INST.)	1.07

Table 3.A.2: The 66 tide-gauge data sites and their geographical coordinates used in our model from Holgate et al. (PSMSL 2013). Also, the GIA rate for each tide gauge site is provided using the ICE5G - VM2-90 Earth - ice physical model developed by Peltier (2004).

3.B Model Results and Validations for the Full Dataset

In this section, we will present the results from our full dataset of 21 proxy sites and 66 tide gauges. In addition, we present the results from the 10-fold cross validation using the 21 proxy sites.

3.B.1 Results for full dataset

The model is run using 21 proxy sites and 66 tide gauge sites, yet we present the results of the proxy record sites as their long temporal trend provide insight into long term changes in RSL along the Atlantic coast of North America. Figure 3.B.2 provides the total model fit for the 21 proxy sites along the Atlantic coast of North America and Figure 3.B.3 provides the rates of change for the corresponding 21 proxy sites. Figure 3.B.4 provides the decomposition of the NI-GAM into the total model fit and the three components; regional, linear local component with the site-specific vertical offset and non-linear local component for the 21 proxy sites. Figure 3.B.1 presents the regional component of the NI-GAM. The grey lines represent 10 randomly chosen posterior samples showing the underlying behaviour of the posterior for the regional component. Figure 3.B.5 provides the non-linear local component for each proxy record data site.

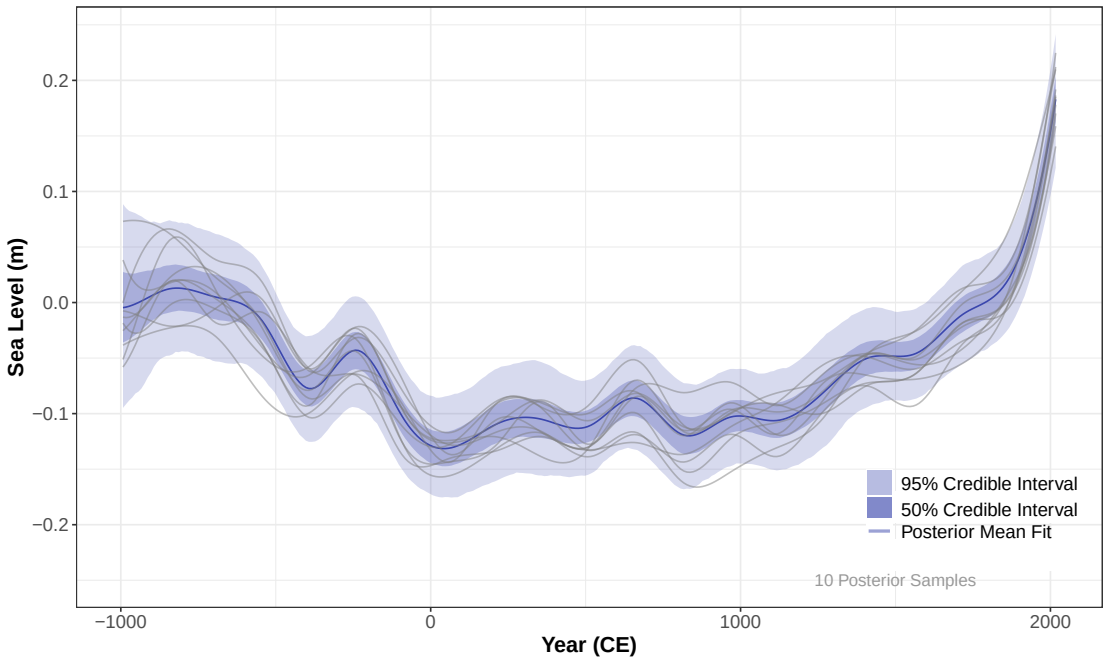


Figure 3.B.1: Regional component of the noisy-input generalised additive model using 21 proxy sites and 66 tide gauge sites along the Atlantic coast of North America. The dark blue line highlights the mean posterior model fit and the dark blue shading indicated the 50% credible interval and the lighter blue shading is the 95% credible interval. The grey lines represent 10 randomly chosen samples to demonstrate the underlying behaviour of the posterior.

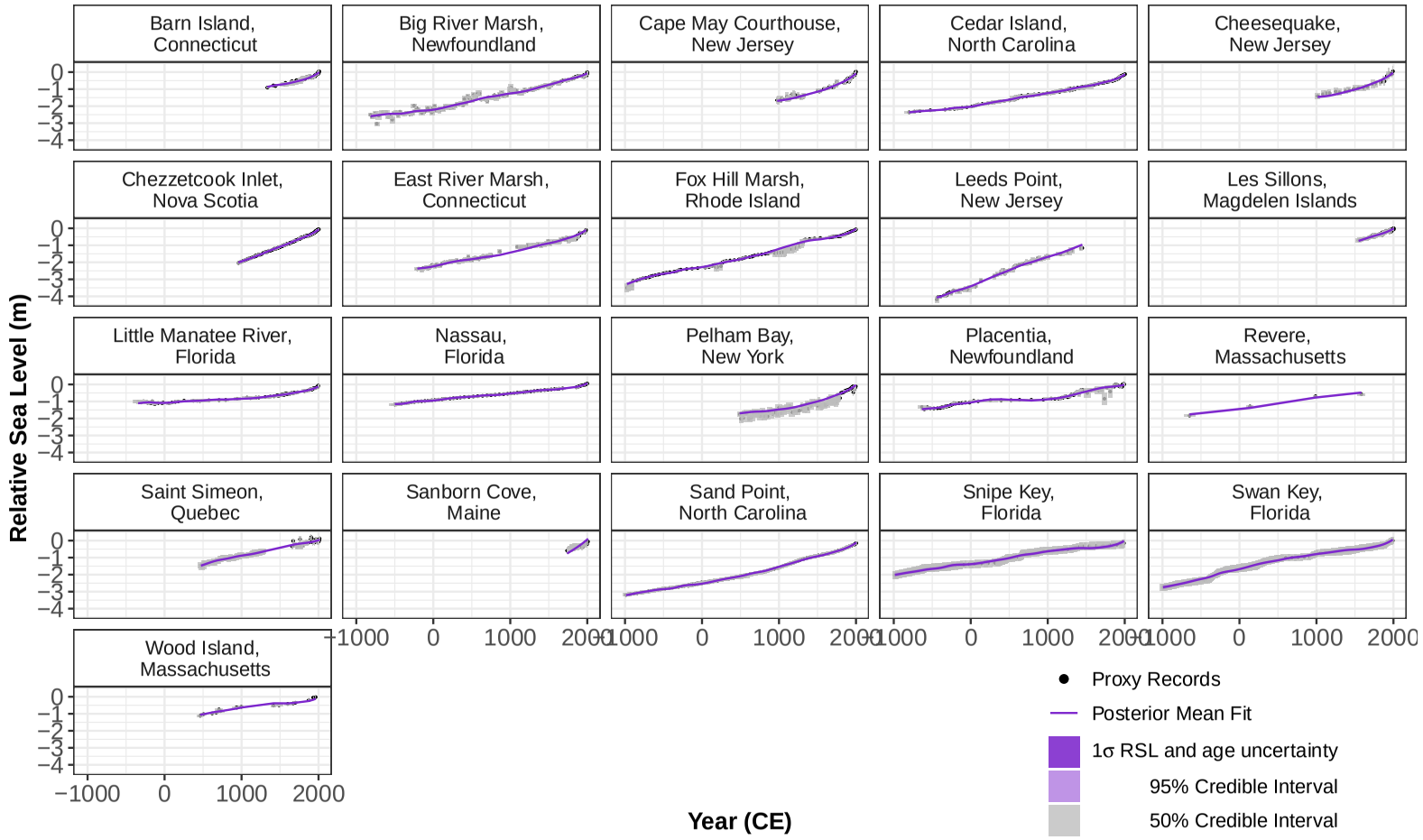


Figure 3.B.2: The Noisy-Input Generalized Additive Model (NI-GAM) is applied to 21 proxy sites along the Atlantic coast of North America, where the black dots and grey boxes signify the midpoint and associated uncertainty, respectively, for each proxy record. The solid purple line represents the mean of the model fit, accompanied by a shaded 95% credible interval.

Figure 3.B.3: Rate of relative sea change found by taking the first derivative of the total model fit for 21 proxy sites along the Atlantic coast of North America. The mean of the fit is the solid purple line with the dark shaded area being the 50% credible interval and the light shaded area being the 95% credible interval.

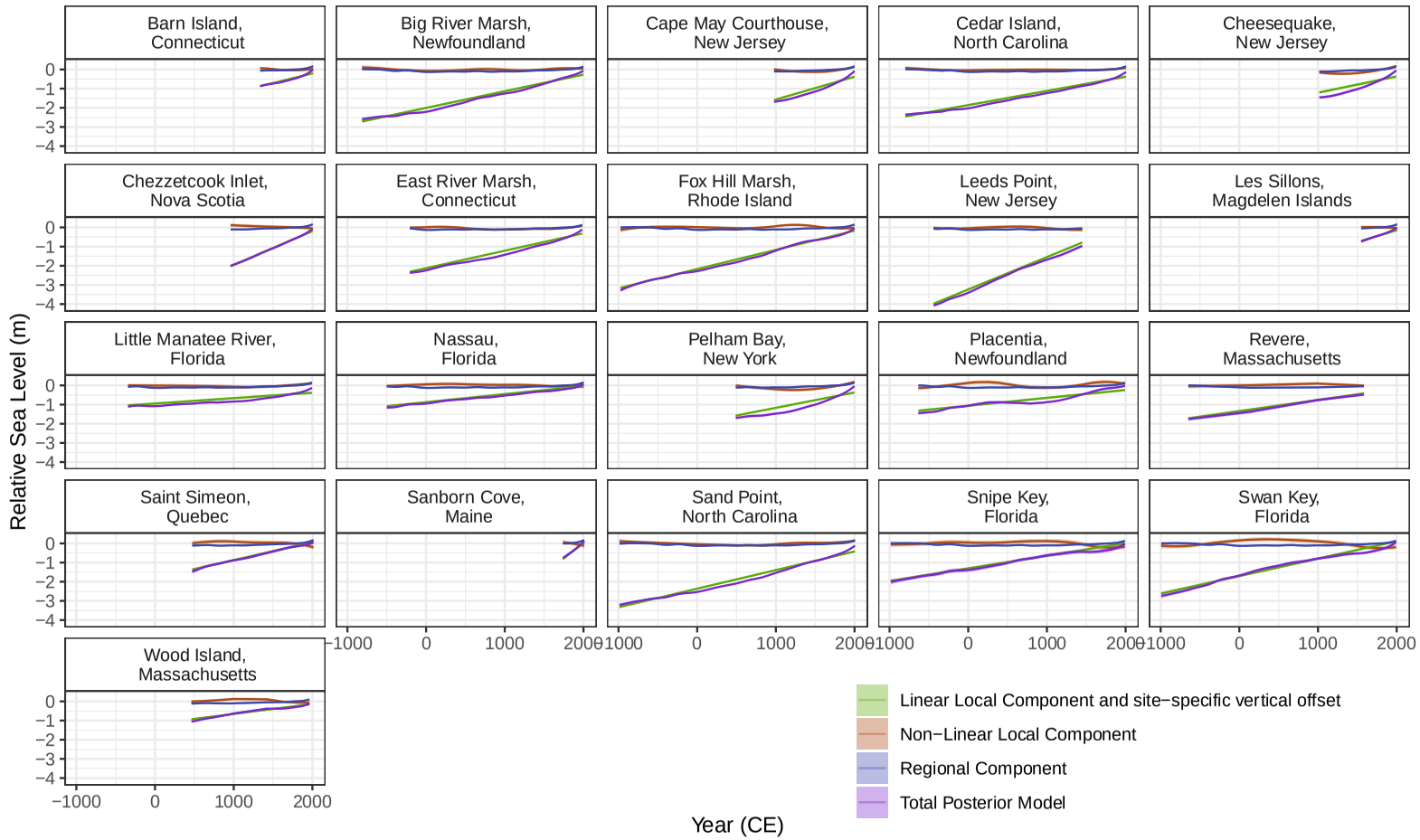


Figure 3.B.4: All components of the NI-GAM for the 21 proxy sites along the Atlantic coast of North America. The regional component is in blue with a 95% credible interval. The linear local component and the site-specific vertical offset are green with a 95% credible interval. The non-linear local component is brown with a 95% credible interval. The total posterior model fit is purple with 95% credible interval. The x-axis is in years and y axis is in meters.

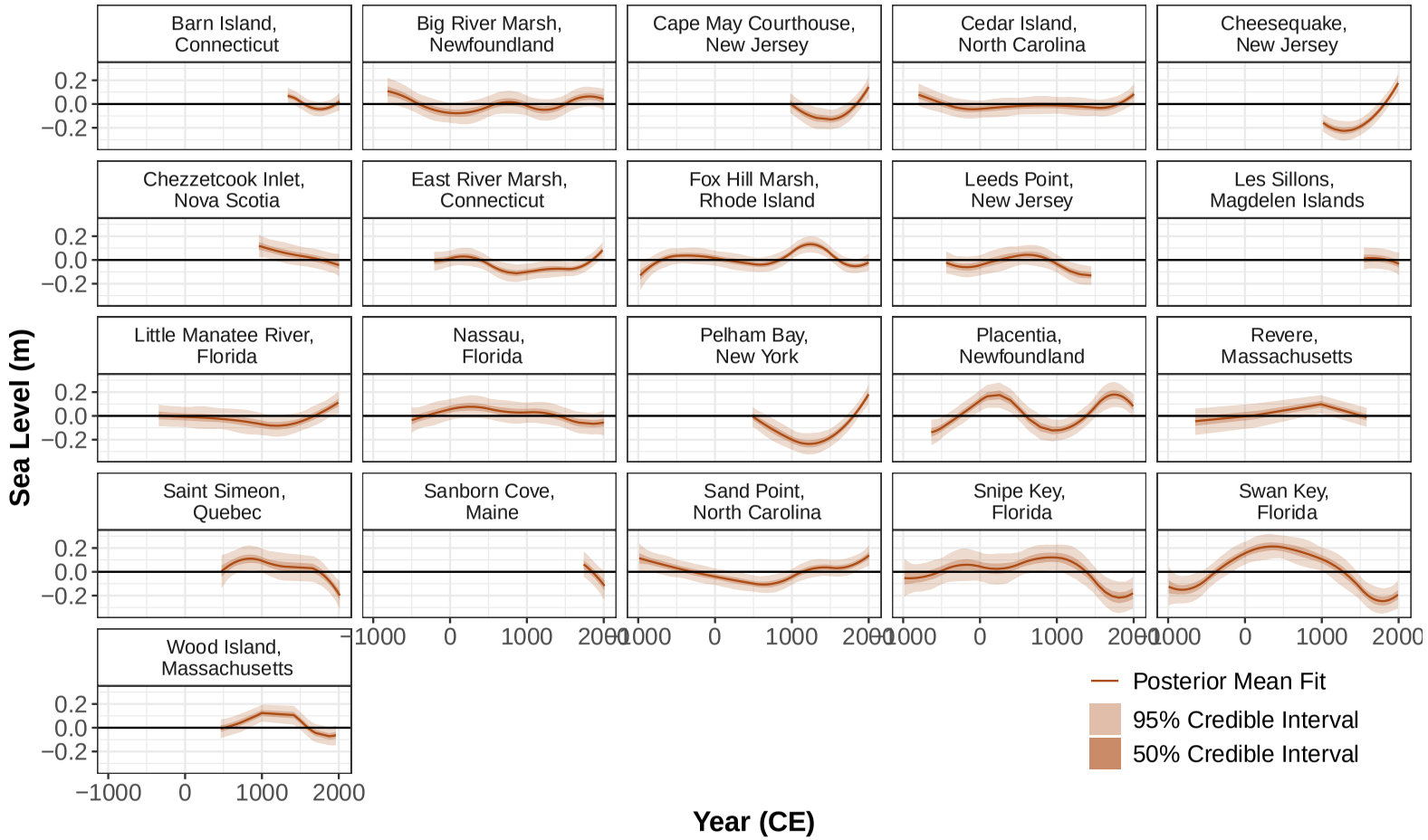


Figure 3.B.5: The non-linear local component for all proxy sites along the Atlantic coast of North America. The y-axis represents sea level in meters. The brown solid line represents the mean of the model fit with the 50 % credible interval in dark brown shading and 95% credible interval in the light brown shading.

3.B.2 Model Validation

Model validations using 10-fold cross validation was undertaken using the data set from the 21 proxy sites along the Atlantic coast of North America. The tide gauge data was not used in the 10-fold cross validation as many sites had fewer than 10 data points. Figure 3.B.6 provides the true versus predicted RSL for the 21 proxy sites using 10 fold cross validation. In table 3.B.1, the empirical coverage of the model for all 21 sites along the Atlantic coast of North America is examined. The empirical coverage indicates the percentage of times the true observation lies within the prediction interval. A comparison is made between the 95% empirical coverage and the 50% empirical coverage. It is evident that the prediction intervals for our model are large resulting in 100% coverage in many sites. This is due to the large size of the prediction intervals resulting from the large bivariate uncertainty that arises from the proxy records. Also included in Table 3.B.1, it the root mean square error (RMSE) for the 21 data sites along the Atlantic coast of North America which gives an insight into the prediction errors.

Site Name	Empirical 95% Coverage	95% PI width	Empirical 50% Coverage	50% PI width	RMSE(m)
Barn Island, Connecticut	1.00	0.31	0.59	0.11	0.08
Big River Marsh, Newfoundland	0.96	0.46	0.63	0.16	0.12
Cape May Courthouse, New Jersey	0.99	0.53	0.81	0.18	0.14
Cedar Island, North Carolina	1.00	0.26	0.78	0.09	0.06
Cheesequake, New Jersey	1.00	0.81	0.78	0.28	0.21
Chezzetcook Inlet, Nova Scotia	1.00	0.26	0.95	0.09	0.07
East River Marsh, Connecticut	1.00	0.52	0.77	0.18	0.13
Fox Hill Marsh, Rhode Island	0.98	0.40	0.61	0.14	0.13
Leeds Point, New Jersey	1.00	0.48	0.64	0.17	0.12
Les Sillons, Magdelen Islands	1.00	0.38	0.87	0.13	0.10
Little Manatee River, Florida	1.00	0.30	0.84	0.10	0.07
Nassau, Florida	1.00	0.30	1.00	0.10	0.07
Pelham Bay, New York	1.00	0.70	0.51	0.24	0.19
Placentia, Newfoundland	0.97	0.36	0.59	0.12	0.11
Revere, Massachusetts	0.50	0.34	0.00	0.12	0.07
Saint Simeon, Quebec	1.00	0.64	0.92	0.22	0.16
Sanborn Cove, Maine	1.00	0.72	0.50	0.25	0.18
Sand Point, North Carolina	1.00	0.33	0.96	0.11	0.08
Snipe Key, Florida	1.00	0.93	1.00	0.32	0.23
Swan Key, Florida	1.00	0.77	1.00	0.26	0.19
Wood Island, Massachusetts	0.78	0.27	0.28	0.09	0.06

Table 3.B.1: Empirical 95% coverage for the 21 data sites along the Atlantic coast of North America with the associated prediction interval(PI). As a comparison, the prediction intervals are reduced to 50% intervals and the empirical coverage for the 50% is presented. The root mean square error (RMSE) is included in meters

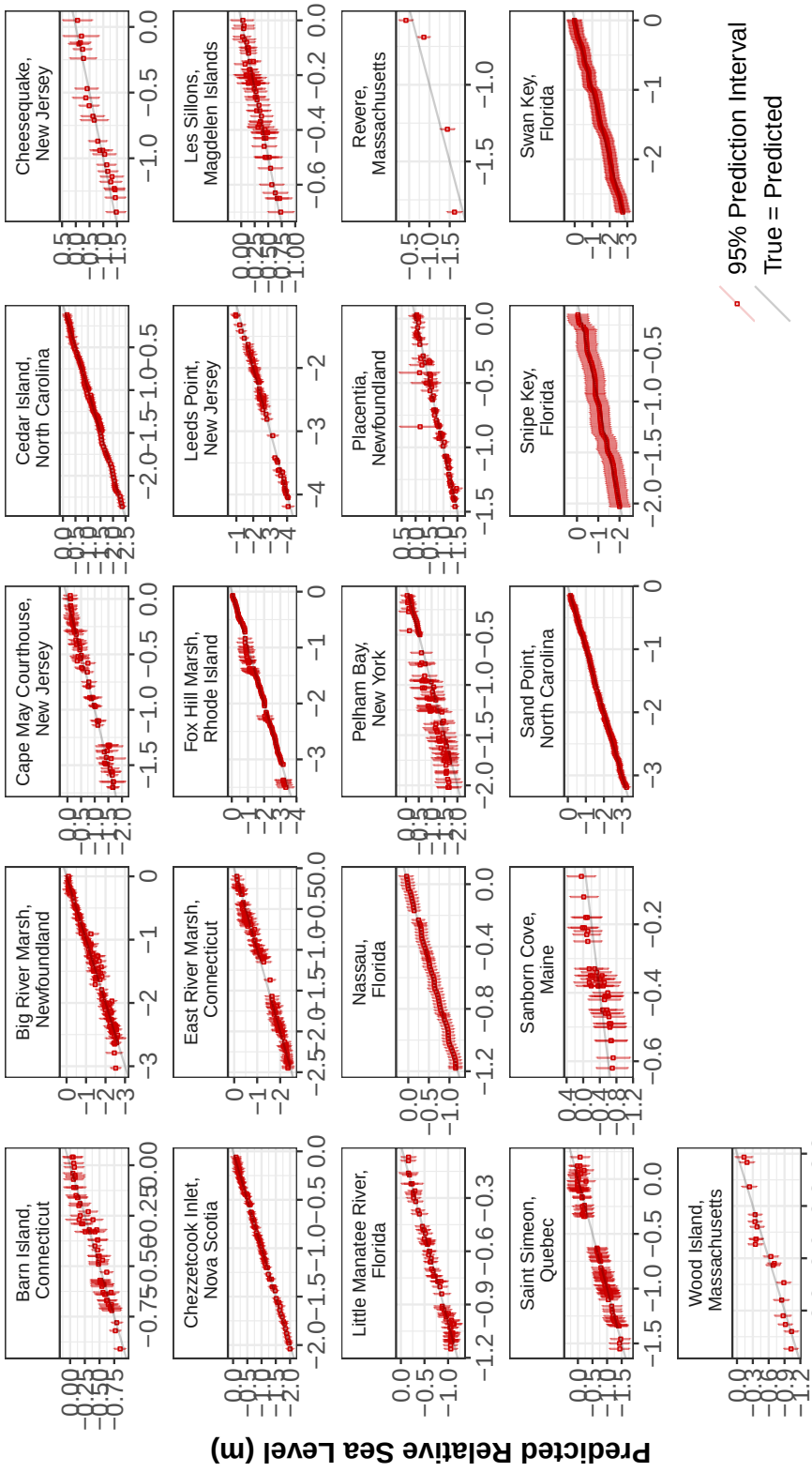


Figure 3.B.6: True versus Predicted RSL for all the 21 proxy sites along the Atlantic coast of North America using 10 fold cross validation. The grey line indicating the identity line.

3.2.2.1 Convergence Diagnostics

We assess the convergence of our model through various tests, employing the ShinyStan app (Gabry and Goodrich, 2017). The findings for one hyperparameter, σ_h, are detailed in Section 3.6.4, and we extend this analysis to encompass the remaining hyperparameters, σ_l and σ_r, in this section. These hyperparameters account for the variability in the spline coefficients governing the non-linear local and regional components, as discussed in Section 3.5.5.

Figure 3.2.7: Convergence tests for the hyperparameter σ_r, which captures variability in the site-specific vertical offset component of the noisy input generalized additive model, were conducted using the ShinyStan app (Gabry and Goodrich, 2017). The Gelman-Rubin Statistic (Rhat) value is 1, indicating convergence, and the large Effective Sample Size (n_eff) value further supports robust model convergence. Visualisations, including the Kernel Density Estimate and autocorrelation plot, demonstrate a well-converged model. The traceplot, based on 3 chains, exhibits a stable and stationary trace, confirming the model's good convergence.

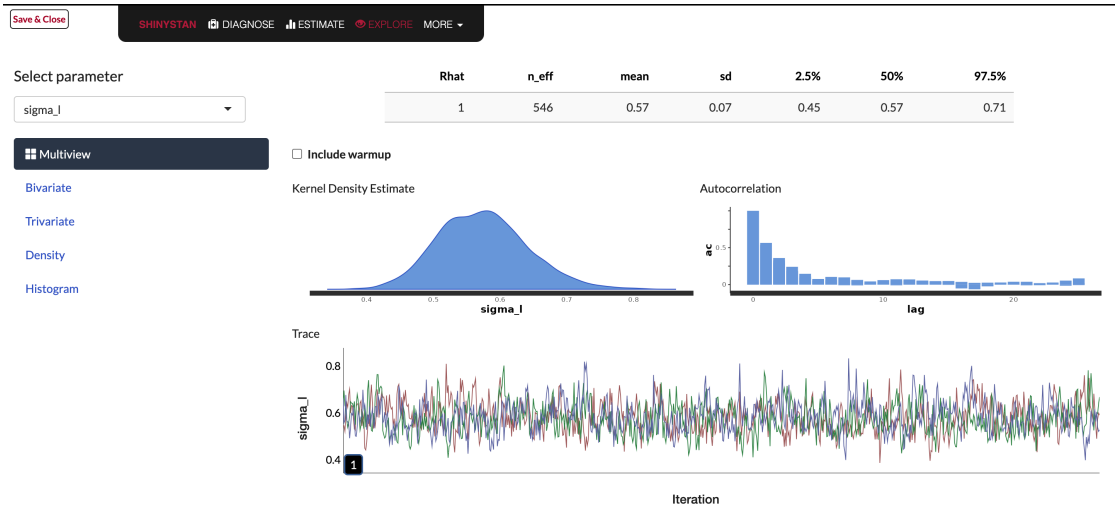


Figure 3.2.8: Convergence tests for the hyperparameter σ_l, which captures variability in the site-specific vertical offset component of the noisy input generalized additive model, were conducted using the ShinyStan app (Gabry and Goodrich, 2017). The Gelman-Rubin Statistic (Rhat) value is 1, indicating convergence, and the large Effective Sample Size (n_eff) value further supports robust model convergence. Visualisations, including the Kernel Density Estimate and autocorrelation plot, demonstrate a well-converged model. The traceplot, based on 3 chains, exhibits a stable and stationary trace, confirming the model's good convergence.

CHAPTER 4

reslr: An R Package for Relative Sea-Level Modelling

We present `reslr`, an R package to perform Bayesian modelling of relative sea level data. We include a variety of different statistical models previously proposed in the literature, with a unifying framework for loading data, fitting models, and summarising the results. Relative sea-level data often contain measurement error in multiple dimensions and so our package allows for these to be included in the statistical models. When plotting the output sea level curves, the focus is often on comparing rates of change, and so our package allows for computation of the derivative of sea level curves with appropriate consideration of the uncertainty. We provide a large example dataset from the Atlantic coast of North America and show some of the results that might be obtained from our package. An R package that implements our approach is available on the Comprehensive R Archive Network at https://cran.r-project.org/web/packages/reslr or on GitHub at https://github.com/maeveupton/reslr.

4.1 Introduction

Understanding the rates and spatial patterns of Relative Sea-Level (RSL) change across various timescales, spanning from decades to millennia, poses a significant challenge. The task involves analysing sparse and noisy proxy and/or instrumental data sources that often have large measurement uncertainties. To address these complexities and provide robust assessments, statistical models play a pivotal role and have become indispensable in the task of quantifying RSL changes (as examined by Cahill et al., 2015a; Khan et al., 2015) and in the evaluation of temporal and spatial variability (e.g. Kopp, 2013; Kopp et al., 2016; Kemp et al., 2018; Walker et al., 2021). To that end the paleo sea-level community would benefit from a comprehensive toolset capable of analysing the historical evolution of sea-level changes across different times and locations. This motivated us to create the `reslr` package, which is available on the Comprehensive R Archive Network at `https://cran.r-project.org/web/packages/reslr` or on GitHub at `https://github.com/maeveupton/reslr`. Our package includes a suite of statistical models appropriate for modelling the complexity of sea-level over time and space while accounting for sea-level data uncertainties and remaining computationally tractable. The output of our package provides insight into temporal and spatial sea-level variability and rates of sea-level change.

The `reslr` package includes a comprehensive dataset of proxy RSL reconstructions for 21 locations along the Atlantic coast of North America (Kemp et al., 2013). These reconstructions rely heavily on dated geological archives obtained from coastal sediments (e.g. Gehrels, 1994) or corals (e.g. Meltzner et al., 2017). Moreover, users of the package have the option to incorporate instrumental sea-level data sourced from the Permanent Service Mean Sea Level online database, which provides annual RSL measurements for approximately 1,500 tide gauge stations worldwide (Holgate et al., 2013). By offering this diverse range of data sources, the `reslr` package caters to the needs of researchers seeking to explore and analyse sea-level variations across different locations and time periods.

The `reslr` package offers a range of statistical models which include: linear regression (Ashe et al., 2019), change point models (Cahill et al., 2015b), integrated

Gaussian process (IGP) models (Cahill et al., 2015a), temporal splines (de Boor, 1978), spatio-temporal splines (Simpson, 2018) and generalised additive models (GAM) (Upton et al., 2023c). In all cases, a Bayesian framework is employed, facilitating the estimation of unknown parameters based on the RSL data while fully accounting for the associated uncertainties. The `reslr` package enables researchers to gain comprehensive insights into sea-level variations, leveraging the flexibility and robustness of these statistical models.

When it comes to addressing measurement uncertainty in proxy records, the `reslr` package offers two distinct approaches. The first approach involves employing the Errors-in-Variables (EIV) method, which takes into account the inherent uncertainties in the input variables (Dey et al., 2000). This method acknowledges that the input variables are not error-free and incorporates this knowledge in the analysis. The second approach offered by the package is the Noisy Input (NI) uncertainty method. This method tackles uncertainty by inflating the output noise variance with a corrective term that is directly linked to the input noise variance (McHutchon and Rasmussen, 2011). By considering the level of uncertainty in the input, this approach provides a more accurate representation of the overall uncertainty in the results. Both the EIV and NI methods have their respective advantages, and the `reslr` package recommends the most suitable uncertainty method based on the statistical model being employed. This ensures that researchers can select the appropriate approach to effectively address measurement uncertainties within their specific analysis context.

For each model, the `reslr` package generates informative plots illustrating the model-based estimates of RSL. In the case of more complex models like the IGP, splines, and GAMs, the resulting plots not only provide RSL estimates but also offer insights into the rates of RSL change. Of particular significance to the paleo-sea level community, the GAM model provides estimates for separate components that represent potential drivers of RSL change. This feature enables comparisons between different components and contributes to a more comprehensive understanding of the factors influencing RSL fluctuations (Upton et al., 2023c). These visual representations enable researchers to gain a clearer understanding of when and where RSL changes occurred, including the magnitude of their temporal vari-

ations. Moreover, the package grants users access to the posterior samples used to generate the plots, providing the option to delve deeper into the underlying statistical distributions and uncertainties associated with the estimated RSL changes. The combination of these outputs serves as a valuable resource for researchers, aiding in the investigation and interpretation of RSL dynamics across various spatial and temporal contexts.

The `reslr` package is uniquely tailored to address the challenges inherent in analysing historic RSL changes using proxy records. Its design, characterised by minimal functions and a user-friendly interface, draws inspiration from other packages such as `simmr`, which integrates mathematical equations within a Bayesian framework (Govan et al., 2023), and `mgcv`, renowned for its diverse statistical modeling options (Wood and Wood, 2015). While packages like `oce` offer intriguing approaches to studying oceanographic data (Kelley and Richards, 2014), they lack the advanced statistical modeling choices provided by `reslr`. Crucially, there are currently no competing R packages that match the breadth of capabilities offered by `reslr`. This uniqueness stems from its capability to provide users with a selection of Bayesian statistical models and account for bi-variate uncertainty, crucial in sea level research.

Our paper has the following structure. Firstly, we introduce the example dataset provided within the package, which serves as the foundation for the examples presented throughout the paper. We also provide insight into additional data sources. Secondly, we offer an overview of the statistical models available in the package, providing necessary background information. Next, we explore the uncertainty methods employed within these statistical models. Following this, we provide a detailed description of the functionality of the `reslr` package, outlining the diverse outputs and plots accessible to users. Finally, we conclude with important remarks and discuss potential future extensions for the package's advancement. Whilst this paper is just a summary of the features of `reslr`, a more complete vignette containing examples of the full functionality of the package is available at `https://maeveupton.github.io/reslr/`.

4.2 Data and Models

4.2.1 Data Sources

Proxy sea-level data are vital sources of information for examining historic changes in RSL prior to the instrumental data period. A proxy refers to a characteristic that can be observed and used to estimate a variable of interest, which cannot be measured directly, and can be of physical, biological or chemical nature (e.g., Gornitz, 2009). In sea-level studies, the proxy data can be sources from microorganisms such as foraminifera (e.g., Edwards and Wright, 2015), geochemical measurements (e.g., Marshall, 2015), or vegetation that have accumulated in the tidal realm (e.g., Kemp and Telford, 2015). The data sets we use have had their proxy measurements transformed into sea level using various techniques which are beyond the scope of our paper (e.g., Gehrels, 1994; Shennan et al., 2015; Kemp et al., 2018). In the `reslr` package, we provide an example proxy dataset which contains 21 proxy sea-level records (See Appendix 4.A) from the Atlantic coast of North America as used in Upton et al. (2023c).

Within the context of sea-level analysis, instrumental data plays an important role by providing direct measurements obtained from tide gauges and satellites (although the latter are currently not incorporated into the `reslr` package). To enhance the versatility of the package, we have implemented a feature that allows users to download annual tide-gauge data from the PSMSL Level online database and store it in a temporary file, making it readily available when needed (PSMSL, 2023 Holgate et al., 2013; Woodworth and Player, 2003).

To ensure the comparability of the tide-gauge data with proxy records, we apply two processing steps. Firstly, the tide-gauge data in the PSMSL database is given in millimetres relative to a revised local reference datum (a coordinate system which defines the zero level for sea level measurements Pugh and Woodworth (2014a)). Within `reslr`, we transform the data by removing 7000 mm to revert the tide-gauge data into the observed reference frame and convert the RSL to metres following the guidance from the PSMSL website as described in Aarup et al. (2006). The second processing step involves averaging the tide-gauge data to equation with the resolution of the temporal resolution of the more recent proxy

data. However, we provide flexibility for users to adjust this averaging period according to the specific characteristics of their data.

4.2.2 Statistical Models

Within the `reslr` package, a Bayesian hierarchical framework is employed for each statistical modelling technique. Markov Chain Monte Carlo (MCMC) simulations are carried out using the Just Another Gibbs Sampler (JAGS) tool (Plummer, 2003) and implemented using the `rjags` package in R (Plummer et al., 2016). Other tools, such as Stan (Carpenter et al., 2017), are utilized for MCMC simulations. However, our preference for JAGS is based on its user-friendly interface, computational efficiency, and flexible model design capabilities.

Mathematically, the data level for every statistical model is described as:

$$y = f(\mathbf{x}, t) + \epsilon_y \tag{4.1}$$

where y is the response data (RSL in metres). $f(\mathbf{x}, t)$ is the process mean that depends on location $\mathbf{x}$ and time t. ϵ_y is the error term given by $\epsilon_y \sim \mathbb{N}(0, \sigma_y^2 + s_y^2)$, where σ_y^2 is the residual variance and s_y the known measurement error associated with RSL. In Table 4.2.1 we provide a list of all the possible options for f within the `reslr` package. Since some of the models we fit do not vary over space (they apply to a single site or treat a set of sites as identical) we use $f(t)$ rather than $f(\mathbf{x}, t)$ to denote the process model.

When using proxy RSL data, measurement error is also present in the input variable (time) due to the dating technique used. For the input measurements, $\tilde{t}$ is assumed to be a noisy estimate of the true time value t:

$$\tilde{t} = t + \epsilon_t \tag{4.2}$$

with the error term given by $\epsilon_t \sim \mathbb{N}(0, s_t^2)$ where s_t is the known measurement error associated with time.

We use two methods to account for the time measurement uncertainty. The first is the Errors-in-variables (EIV) method which assumes that the input variable, e.g. time, is measured as an error-prone substitute and models it directly (Dey et al.,

2000). The second uncertainty method is the Noisy Input method. This method fits an initial model and uses the derivative of the mean of f to calculate a corrective variance term. Then, the model is re-run with this additional corrective variance term allowing for the input noise variation to be learned from the complete outputs of the model (McHutchon and Rasmussen, 2011). Within the `reslr` package, the EIV method is used for the linear regression, change point and IGP model and the temporal spline, the spatio-temporal spline and the GAM use the Noisy Input uncertainty method. In general, the EIV method tend to be slower but models the uncertain input process directly, whilst the NI method is faster but requires the model to be fitted twice.

In Table 4.2.1, we present a range of statistical modelling techniques for $f(\mathbf{x}, t)$, the component of our approach, available in the `reslr` package. While our seven distinct models are implemented using JAGS software, users are not permitted to input their own JAGS models. This restriction is in place because defining each model output and its corresponding prior values would entail additional work for the user. Consequently, we do not offer an option for user-defined JAGS models. Below we discuss each technique and provide insight into the potential uses of these techniques for the paleo-environmental community.

Statistical Model	Model Information	`model_type` code
Errors in variables simple linear regression	A straight line of best fit taking account any age and measurement errors in the RSL values using the method of Cahill et al. (2015b)	`"eiv_slr_t"`
Errors in variables change point model	An extension of the linear regression modelling process. It uses piece-wise linear sections and estimates where/when trend changes occur in the data (Cahill et al., 2015b)	`"eiv_cp_t"`
Errors in variables integrated Gaussian process	A non-linear fit that utilities a Gaussian process prior on the rate of sea-level change that is then integrated (Cahill et al., 2015a).	`"eiv_igp_t"`
Noisy Input spline in time	A non-linear fit using regression splines as mentioned Upton et al. (2023c)	`"ni_spline_t"`
Noisy Input spline in space and time	A non-linear fit for a set of sites across a region using the method of Upton et al. (2023c).	`"ni_spline_st"`
Noisy Input Generalised Additive model for the decomposition of the RSL signal	A non-linear fit for a set of sites across a region and provides a decomposition of the signal into regional, local linear and non-linear local components. This full model is as described in Upton et al. (2023c).	`"ni_gam_decomp"`

Table 4.2.1: List of all statistical models available in the `reslr` package. We provide a short description and the relevant literature for each model. The `model_type` code column represents the text input the user should use when implementing their preferred modelling technique.

4.2.3 EIV Linear Regression

The EIV linear regression model given by:

$$f(t) = \alpha + \beta t \tag{4.3}$$

where α is the intercept, β is the slope and t is the time. Earlier studies, for example Shennan and Horton (2002) and Engelhart et al. (2009), employed linear regression when evaluating the rate of RSL change over the past 4000 years. The `reslr` package implements a temporal linear regression as its simplicity is popular for approximate estimates of linear rates of RSL change. However, linearity assumptions for RSL change are often unrealistic when examining long-term historical trends.

4.2.4 EIV Change Point Model

The EIV change point (CP) model, an extension of the linear regression model, assumes the RSL process is piecewise linear and estimates when trend changes occur in the data (Carlin et al., 1992; Cahill et al., 2015b). Mathematically, the multiple CP model, $f(t)$ is described as:

$$f(t) = \begin{cases} \alpha_1 + \beta_j(t - \lambda_1) \text{ when } j = 1, 2, \\ \alpha_{j-1} + \beta_j(t - \lambda_{j-1}), \text{ when } j = 3, ..., m + 1 \end{cases} \tag{4.4}$$

where α_j is the expected value of the response at the jth CP. λ_j is the time at which the CP occurs with the prior restriction that $\lambda_1 < \lambda_2 < ... < \lambda_m$ and m is the number of CPs (Cahill et al., 2015b). In the `reslr` package, the user can select m to be 1, 2 or 3 CPs. β_1 and β_{m+1} are the slopes before and after the first and last CP respectively. β_j for $j = 2 \ldots, m$ are the slopes between the $(j-1)$th and jth CP. β_1 and β_{m+1} are estimated as part of the model and are free parameters given prior distributions (Carlin et al., 1992). The remaining β_j are deterministically calculated since the function is continuous and the neighbouring linear segments must join together (Cahill et al., 2015b) such that:

$$\beta_j = \frac{\alpha_j - \alpha_{j-1}}{\lambda_j - \lambda_{j-1}}, j = 2, ..., m. \tag{4.5}$$

This technique has be used in different aspects of the sea-level literature. For example, Kemp et al. (2009) determined the magnitude and the timing of recent accelerated sea-level rise using change point models in North Carolina, USA. Brain et al. (2012) used the CP method to examine the impacts of sediment compaction on reconstructing recent sea-level rise in the United Kingdom. Hogarth et al.

(2020) used CP models to obtain more consistent estimates of sea-level rise since 1958 for the British Isles. The main advantage of the CP model is its' ability to identify sudden changes in RSL. However, the number of change points must be specified by the user.

4.2.5 Integrated Gaussian Process

An Integrated Gaussian process (IGP) is a modelling strategy that has been used extensively by the sea-level community when examining the temporal evolution of sea level change (e.g. Cahill et al., 2015a; Hawkes et al., 2016; Kemp et al., 2017; Shaw et al., 2018; Dean et al., 2019; Stearns et al., 2023; Kirby et al., 2023).

The IGP uses Gaussian Process (GP) to directly estimate the rate of change of the response (Holsclaw et al., 2013). In order to extract the original $f(t)$ we integrate $p(t)$:

$$f(t) = \alpha + \int_0^t p(u)du \tag{4.6}$$

where α is the intercept and is the rate of change, $p(t) = \frac{df}{dt}$, described as:

$$p(t) \sim GP(\mu(t), k(t, t')) \tag{4.7}$$

with t time and $\mu(t)$ the mean function and $k(t, t')$ is the covariance function. The covariance function provides insight into the relationship between the outcome variables, i.e. if input variable, t and t' are in close proximity, the corresponding outcomes will be more correlated and vice versa (Rasmussen and Williams, 2006). It is written as (Cahill et al., 2015a):

$$k(t, t') = \nu^2 \rho^{(t-t')^2} \tag{4.8}$$

where ρ is the correlation parameter and ν^2 is the variance of the rate process.

The technique, described by Cahill et al. (2015a), offers insights into examining rates of Relative Sea Level (RSL) change using proxy records from a single location. Apart from the IGP model, the `reslr` package does not rely on GP methods. We acknowledge that the use of GP modelling has gained considerable traction within the sea-level research community, particularly for investigating the spatio-temporal evolution of sea-level changes, as evidenced by notable studies (e.g., Kopp et al.,

2009; Kopp, 2013; Kopp et al., 2016; Kemp et al., 2018; Walker et al., 2021). Nevertheless, in the context of the `reslr` package, we have intentionally opted for computationally efficient alternatives- splines and GAMs - as detailed in our prior work (Upton et al., 2023c). These methods offer practical and effective approaches to analysing sea-level data, accommodating the complexities of spatio-temporal dynamics while ensuring computational tractability.

4.2.6 Temporal Spline

Splines are mathematical tools used in a wide range of settings from interpolation to data smoothing. There are a variety of different splines available, yet in this research we focus on B-splines (de Boor, 1978; Dierckx, 1995) and P-splines (Eilers and Marx, 1996). Mathematically, B-splines are described in the following way:

$$f(t) = \sum_{k=1}^{K} b_k(t)\beta_k \tag{4.9}$$

where $b_k(t)$ is the spline basis function and β_k is the spline coefficient.

Following on from B-splines, Eilers and Marx (1996) describe a method to overcome the difficulty of choosing the correct number of knots by developing the penalised spline or P-splines. Penalised differences in the spline coefficients control the smoothness of the spline based on differences (of order d) of the spline coefficients. The first order differences are written as:

$$\Delta\beta_k = \beta_k - \beta_{k-1} \tag{4.10}$$

The spline coefficient will be centered on the previous value with a smoothness parameter σ_β^2:

$$\Delta\beta_k \sim \mathcal{N}(0, \sigma_\beta^2) \tag{4.11}$$

In our package, P-splines are used for the NI spline in time and the extendable nature of these splines allows for different components to be examined within the GAM which is described below.

4.2.7 Spatio-Temporal Spline

We use a spatio-temporal spline to examine RSL evolving over time at multiple locations. We include a tensor product to capture the variability over time and

space (represented with longitude and latitude). For each individual covariate, time (t) and longitude (x_1) and latitude (x_2), we construct a B-spline basis (Wood, 2017b). These basis functions are combined product-wise in the following way (Wood, 2006):

$$f(t, x_1, x_2) = \sum_{h=1}^{H} \sum_{i=1}^{I} \sum_{j=1}^{J} b_h(t) b_i(x_1) b_j(x_2) \beta_{hij} \tag{4.12}$$

where β_{hij} is the spline coefficient. H is the number of knots for $b_h(t)$ the spline basis function in time t. I is the number of knots for $b_i(x_1)$ the spline basis function for longitude. J is the number of knots for $b_j(x_2)$ the spline basis functions for latitude values. The prior for the spline coefficient is given as:

$$\beta_{hij} \sim \mathcal{N}(0, \sigma_\beta^2) \tag{4.13}$$

where σ_β^2 is the smoothness parameter for the spatio-temporal spline. The `reslr` package uses B-splines for the NI spline in space and time allowing for multiple sites to be examined. The advantage of the tensor B-spline approach is that the basis functions are simple to construct, each depending on only one input variable. However the number of parameters to estimate does increase considerably.

4.2.8 Generalised Additive Models

Generalised additive models are an extension of generalised linear models that use a basis expansion and a smoothing penalty to create linear predictors that are dependent on the sum of smooth functions of the predictor variable (GAMs; Wood, 2017a). The model developed by Upton et al. (2023c) uses splines and random effects to create a spatio-temporal relative sea level field. It identifies variations of sea-level at different spatial and temporal scales, encompassing multiple underlying processes and avoiding a focus on specific physical processes. The decomposition of this mean relative sea level field can be written as:

$$f(\mathbf{x}, t) = r(t) + g(z_\mathbf{x}) + h(z_\mathbf{x}) + l(\mathbf{x}, t) \tag{4.14}$$

where $r(t)$ is the regional component at time t represented with a spline in time. $g(z_\mathbf{x})$ is the linear local component at location x represented by a random effect with $z_\mathbf{x}$ representing each data site. $h(z_\mathbf{x})$ is the spatial vertical offset for each data

site. $l(\mathbf{x}, t)$ is the non-linear local component represented with a spline in space time.

The regional component ($r(t)$) represents temporal processes that are common to all locations, including barystatic and thermosteric contributions, where the former is caused by the transfer of mass between land-based ice and oceans (Gregory et al., 2019) and the latter is influenced by changes in global temperature creating density variations within the oceans (Grinsted, 2015). It is described using a spline in time:

$$r(t) = \sum_{s=1}^{k_r} b_{r_s}(t)\beta_s^r \tag{4.15}$$

where β_s^r is the s^{th} spline coefficient, k_r is the number of knots and $b_{r_s}(t)$ is the s^{th} spline basis function at time t. The prior for the spline coefficients of the regional component β_s^r are:

$$\beta_s^r \sim \mathcal{N}(0, \sigma_r^2) \tag{4.16}$$

where the smoothness of the model fit is controlled by σ_r is the standard deviation of the spline coefficient.

The linear local component ($g(z_x)$) of the sea level model aims to capture linear trends present in the relative sea level signal. One such cause is glacial isostatic adjustment (GIA), which is a response of the Earth, the gravitational field, and the ocean to changes in the size of ice sheets (Whitehouse, 2018). On relatively short timescales, it is approximated to be linear through time with spatial variability along the Atlantic coastline of North America (Engelhart et al., 2009). Mathematically, the linear local component is an unstructured random effect for each site which is formulated as:

$$g(z_{\mathbf{x}_j}) = \beta_j^g t \tag{4.17}$$

where β_j^g is a slope parameter specific for each site j. The prior for the linear local component is given by:

$$\beta_j^g \sim \mathcal{N}(m_{g_j}, s_{g_j}^2) \tag{4.18}$$

where m_{g_j} and $s_{g_j}^2$ are the empirically estimated rate and associated variance (refer to Upton et al. (2023c) for a detailed description).

The site-specific vertical offset h is a random effect used to capture vertical shifts associated with measurement variability between sites and is formulated as:

$$h(z_{\mathbf{x}_j}) = \beta_j^h \tag{4.19}$$

where β_j^h contains the random effect coefficients for site j. The prior for the site-specific vertical offset β_j^h is given as:

$$\beta_j^h \sim \mathcal{N}(0, \sigma_h^2) \tag{4.20}$$

where σ_h^2 is the variance of the random intercept across all data sites.

The non-linear local component ($l(\mathbf{x}, t)$) captures structured and unstructured RSL variability on century timescales, including dynamic sea-level changes (atmospheric and oceanic circulation patterns (Gregory et al., 2019)) and site-specific processes (e.g. sediment compaction affecting solid Earth's surface (Horton et al., 2018)). It is described using a spatio-temporal spline function formed using a tensor product and is formulated as:

$$l(\mathbf{x}, t) = \sum_{s=1}^{k_l} b_{l_s}(\mathbf{x}, t)\beta_s^l \tag{4.21}$$

where β_s^l is the s^{th} spline coefficient, k_l is the number of knots and $b_{l_s}(\mathbf{x}, t)$ is the s^{th} spline basis function at time t and location $\mathbf{x}$. The prior for the spline coefficient β_s^l is given as:

$$\beta_s^l \sim \mathcal{N}(0, \sigma_l^2) \tag{4.22}$$

where σ_l^2 is the variance of the spline coefficients over space and time.

As described in Upton et al. (2023c), B-splines are used for both the regional and local terms as this model structure balances both model usability and computational efficiency for examining proxy-based sea level reconstructions on a regional to local scale. B-splines also allow for easier prior elicitation of the smoothness parameters since they directly control the variability of the spline weights in the model.

4.3 Implementation

Within the package, we keep the number of functions to a minimum to ensure accessibility for users R experience. We run the statistical models using an MCMC algorithm and include a summary function to obtain a high level insight into the outputs. We use S3 classes to access the summary, print and plot commands. The package has functions to plot the input data and resulting model fits using `ggplot2` (Wickham, 2016). The user has access to all the underlying information used to create these plots allowing these visualisations to be re-created. In addition, the functions within the package are extendable allowing advanced users access to more complex outputs.

In this section, we provide insight into the example dataset and additional data sources using tide-gauge data within the `reslr` package. A discussion is provided into each function using two separate case studies; a single location and multiple locations. In the first case study, we demonstrate the Noisy Input temporal spline (`model_type = "ni_spline_t"`) which is an example modelling strategy for a single location. In the second case study, we examine multiple locations using the Noisy Input GAM decomposition (`model_type = "ni_gam_decomp"`).

4.3.1 Example Dataset

We include a dataset used as an example called `NAACproxydata`. The full dataset with names of the locations and associated literature is in the Appendix 4.A. The `NAACproxydata` is a data frame with 1715 rows and 8 columns which include:

- Region: Region name
- Site: Site name
- Latitude: Latitude of the site
- Longitude: Longitude of the site
- RSL: Relative Sea level in metres
- RSL_err: 1 standard deviation error associated with relative sea level measured in metres
- Age: Age in years Common Era (CE)
- Age_err: 1 standard deviation error associated with the age in years CE

4.3.2 Including Tide Gauge Data

The tide-gauge data available to users can be obtained from the PSMSL online database (Holgate et al., 2013; Woodworth and Player, 2003) through the `reslr` package. To ensure compatibility with the proxy records, several processing steps are performed within the package, as discussed earlier.

When incorporating tide-gauge data, users have three methods to select their preferred tide gauge(s). The first option is to provide a list of tide-gauge names from the PSMSL database, allowing users the freedom to select any tide gauge available. The second option is to automatically identify the nearest tide gauge to the proxy location that has more than 20 years of observations. This option proves particularly useful when examining proxy records and extending the temporal range to capture recent changes in RSL.

The final option enables the selection of all tide gauges within a 1-degree radius (latitude and longitude) of the proxy location, provided they have more than 20 years of observations. This option grants users access to a wide array of tide gauges within a larger geographic area. Moreover, users can combine the first option with either the second or the third option, allowing for the freedom to choose specific tide gauges while incorporating the nearest tide gauge or multiple tide gauges.

All the values mentioned in this paragraph are arguments that can be adjusted within the function, giving users flexibility in customizing their data selection process according to their specific requirements.

4.3.3 Case Study for 1 Location

In the following sections, we use one site, Cedar Island North Carolina USA (Kemp et al., 2011, 2017), from the example dataset, `NAACproxydata`:

```
CedarIslandNC <- reslr::NAACproxydata %>%
  dplyr::filter(Site == "Cedar Island")

glimpse(CedarIslandNC)

#> Rows: 104
```

```
#> Columns: 8
#> $ Region    <chr> "North Carolina", "North Carolina", "North Carolina", "North
#> $ Site      <chr> "Cedar Island", "Cedar Island", "Cedar Island", "Cedar Islan
#> $ Latitude  <dbl> 34.971, 34.971, 34.971, 34.971, 34.971, 34.971, 34.971, 34.9
#> $ Longitude <dbl> -76.38, -76.38, -76.38, -76.38, -76.38, -76.38, -76.38, -76.
#> $ RSL       <dbl> -0.12, -0.14, -0.16, -0.18, -0.19, -0.21, -0.22, -0.23, -0.2
#> $ Age       <dbl> 2005, 1996, 1988, 1979, 1974, 1963, 1957, 1951, 1941, 1937,
#> $ Age_err   <dbl> 2.25, 2.00, 5.00, 5.75, 5.50, 5.50, 7.00, 7.75, 7.75, 8.00,
#> $ RSL_err   <dbl> 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06,
```

For a single location such as this case study, we recommend using an EIV IGP or a NI spline in time as they demonstrate how the RSL changed over time for this example site called Cedar Island. In this example, tide gauge data is not included but it is an option available to the user if they require. The next example will demonstrate a more complex analysis with the inclusion of tide gauges.

After selecting the data site from the example dataset, we use the `reslr_load` function to process the data prior to running the statistical model and it has a number of different settings that the user can alter depending on the model choice. One such setting is the `prediction_grid_res` option. This provides the resolution at which predictions of RSL and RSL rates are made and subsequently plotted. We set the default at 40 years and if a finer grid is required, the user can alter the setting for `prediction_grid_res`. The `reslr_load` function includes additional settings to include tide gauge data and linear rates which will be discussed in the next case study. For the single site case study, we demonstrate the `reslr_load` function:

```
CedarIslandNC_input <- reslr_load(data = CedarIslandNC)
```

The output of this function is a list of two dataframes called `data` and `data_grid`. The `data` dataframe is the inputted data with an additional column called `data_type_id` which distinguishes proxy records from tide gauge data. The `data_grid` is a dataframe that is evenly spaced in time based on the `prediction_grid_res` value chosen by the user and is used to create the plots:

```
glimpse(CedarIslandNC_input$data_grid)
```

```
#> Rows: 57
#> Columns: 5
#> Groups: SiteName [1]
#> $ Longitude    <dbl> -76.38, -76.38, -76.38, -76.38, -76.38, -76.38, -76.38, -
#> $ Latitude     <dbl> 34.971, 34.971, 34.971, 34.971, 34.971, 34.971, 34.971, 3
#> $ SiteName     <fct> "Cedar Island,\n North Carolina", "Cedar Island,\n North
#> $ data_type_id <fct> ProxyRecord, ProxyRecord, ProxyRecord, ProxyRecord, Proxy
#> $ Age          <dbl> -865.25, -750.00, -700.00, -650.00, -600.00, -550.00, -50
```

A brief insight into the outputs of the `reslr_input` function can be obtained using the print function which provides the number of observations and the sources of the data as shown below:

```
print(CedarIslandNC_input)
```

```
#> This is a valid reslr input object with 104 observations and  1 site(s).
#> There are  1  proxy site(s) and  0  tide gauge site(s).
#> The age units are; Common Era.
#> Decadally averaged tide gauge data was not included.
#> It is recommended for the ni_gam_decomp model
#> The linear_rate or linear_rate_err was not included.
#> It is required for the ni_gam_decomp model
```

The next step is using the plot function to plot the raw data, shown in Figure 4.3.1, using the following:

```
plot(CedarIslandNC_input,
     plot_caption = FALSE)
```

4.3.3.1 Noisy Input Spline in time

The NI spline in time (`model_type = "ni_spline_t"`) examines how the response variable, RSL, varies in time. While the EIV-IGP method is commonly used in the sea-level community, we demonstrate that the NI spline in time is a superior alternative for future research. Unlike the Gaussian process, which has a computational complexity that grows exponentially with the number of data points, the

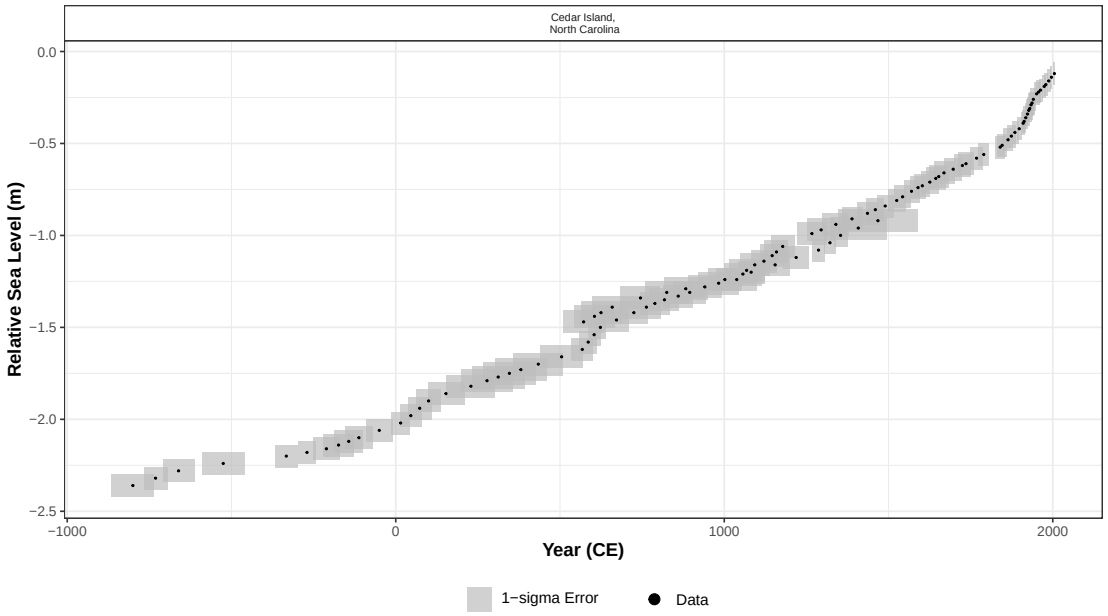


Figure 4.3.1: A plot of the raw data for our example site Cedar Island North Carolina (Kemp et al., 2011, 2017). The x-axis is time in years in the Common Era (CE) and the y-axis is relative sea level in metres. The grey boxes are 1 standard deviation vertical and horizontal (temporal) uncertainty. The black dots are the midpoints of the uncertainty boxes.

spline equivalent uses pre-computed basis functions, resulting in a more efficient computation (Wood, 2017a).

For this model type, the `reslr_mcmc` function is used to implement the MCMC simulation using JAGS and the model type setting is selected to be `model_type = "ni_spline_t"`.

```
res_ni_spline_t <- reslr_mcmc(input_data = CedarIslandNC_input,
                              model_type = "ni_spline_t")
```

The output of the `reslr_mcmc` function is a list that stores the JAGS model run, the input dataframe and the dataframes for plotting the results. The user can set the size of the credible intervals by changing the `CI` setting in this function; the current default is `CI = 0.95`. In addition, the user can alter the number of iterations which will be required if the model is not converging.

To obtain a brief insight into the outputs of the `reslr_mcmc` function, the user can use the print function which provides the number of iterations and the model type:

```
print(res_ni_spline_t)
```

```
#> This is a valid reslr output object with 104 observations and  1 site(s).
#> There are  1  proxy site(s) and  0  tide gauge site(s).
#> The age units are; Common Era.
#> The model used was the Noisy Input Spline in time model.
#> The input data has been run via reslr_mcmc and has
produced 3000 iterations over 3 MCMC chains.
```

The convergence of the MCMC algorithm can be examined for the `"ni_spline_"` model using the `summary` function and ensures the scale reduction factor (R-hat) is close to 1 (Gelman and Rubin, 1992b; Gelman et al., 2013). If the model run has converged, the package will print: "No convergence issues detected". If the package prints: "Convergence issues detected, a longer run is necessary". The user is required to update the `reslr_mcmc` function with additional iterations as described above. The `summary` function provides insight into the parameter estimates from the model using the following:

```
summary(res_ni_spline_t)
```

```
#> No convergence issues detected.
```

```
#> # A tibble: 2 x 7
#>   variable      mean      sd     mad       q5    q95  rhat
#>   <chr>        <num>   <num>   <num>    <num>  <num> <num>
#> 1 sigma_beta 2.10    0.682   0.545   1.30     3.32    1.00
#> 2 sigma_y    0.00626 0.00474 0.00462 0.000521 0.0153  1.00
```

For the parameter estimates, "sigma_beta" acts as a smoothness parameter controlling the penalisation of the splines coefficients for the spline in time model and "sigma_y" represents the data model variation. These are σ_y and σ_β as described in Section 4.2.6.

The final results from the "ni_spline_t" model can be illustrated using the `plot` function and the corresponding dataframes are stored in the `res_ni_spline_t` object called `output_dataframes` as a named list element. Figure 4.3.2 demonstrates the posterior model fit for our example site using:

```
plot(res_ni_spline_t,
     plot_type = "model_fit_plot",
     plot_caption = FALSE)
```

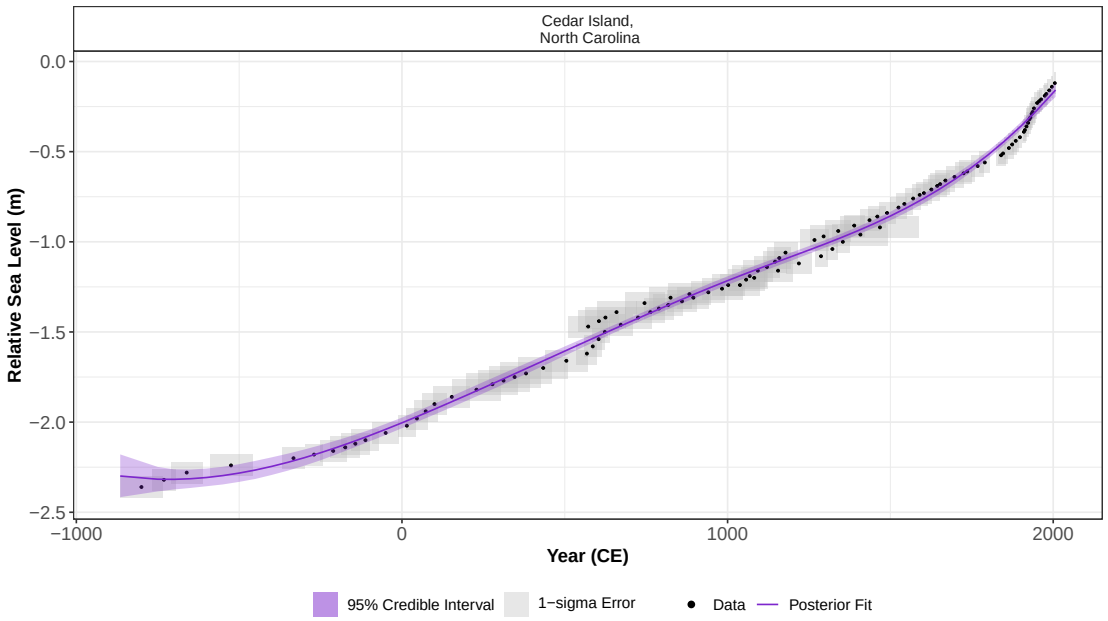


Figure 4.3.2: The plot of the noisy input spline in time model fit for our example site, Cedar Island, North Carolina (Kemp et al., 2011, 2017). The x-axis is time in years in the Common Era (CE) and the y-axis is relative sea level in metres. The grey boxes are 1 standard deviation vertical and horizontal (temporal) uncertainty. The black dots are the midpoints of the uncertainty boxes. The solid purple line represents the mean of the model fit with a 95% credible interval denoted by shading.

In Figure 4.3.3 the rate of change of this posterior model fit is presented and can be viewed using:

```
plot(res_ni_spline_t,
     plot_type = "rate_plot",
     plot_caption = FALSE)
```

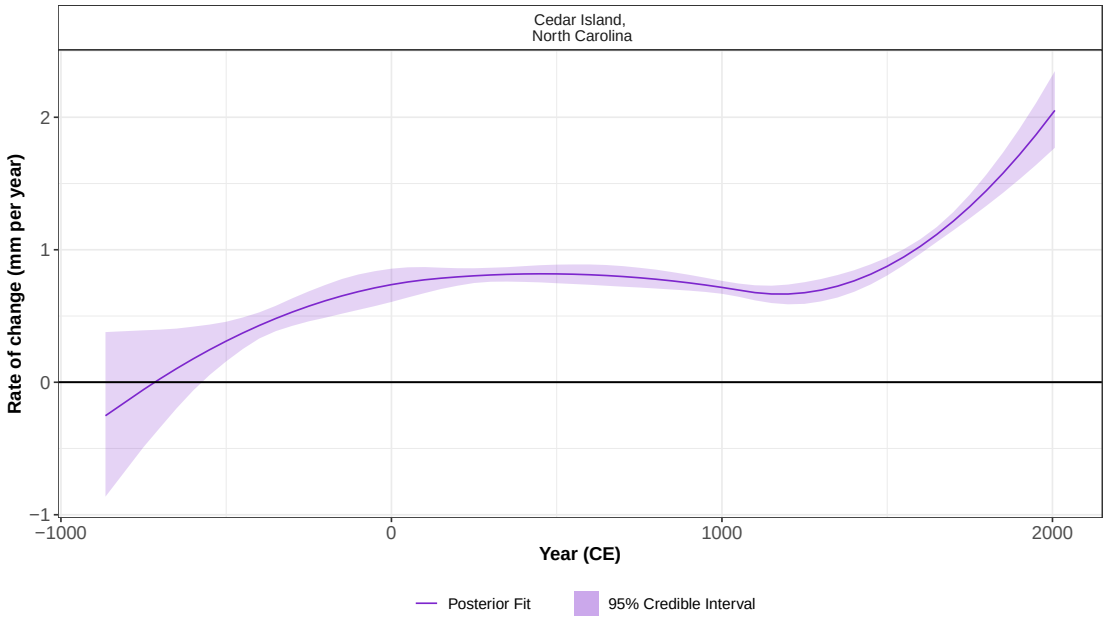


Figure 4.3.3: The rate of change model fit using noisy input spline in time model for our example site Cedar Island, North Carolina (Kemp et al., 2011, 2017). The rate is calculated by taking the first derivative of the total model fit. The x-axis is time in years in the Common Era (CE) and the y-axis is the instantaneous rate of change of sea level in mm per year. The solid purple line represents the mean of the model fit with a 95% credible interval denoted by shading. There is a black horizontal line which is the zero rate of change for this site.

4.3.4 Case Study for Multiple Sites

The sea-level research community is commonly interested in temporal and spatial variations in RSL. To cater to this interest, the `reslr` package offers two models for spatio-temporal modeling. The first model is a noisy-input spline that accounts for noise in both time and space, providing a robust representation of RSL dynamics. The second model, a more intricate option, is the noisy input GAM. Mathematical details concerning the Noisy Input GAM can be found in Upton et al. (2023c). In our upcoming example, we will focus on this model as it empowers users to explore the decomposition of the RSL signal over time and space, unraveling valuable insights into the underlying dynamics.

4.3.4.1 Noisy Input Generalised Additive Model for Decomposition of Response Signal

We demonstrate the functions settings required for the NI GAM. This model requires an adequate number of sites to perform the decomposition and the minimum sites required will depend on the signal in the data. In this example, we use nine sites from the example dataset, `NAACproxydata`, which are selected in the following manner:

```
multi_site <- reslr::NAACproxydata %>%
  dplyr::filter(Site %in% c("Cedar Island","Nassau",
                            "East River Marsh", "Swan Key",
                            "Placentia",
                            "Pelham Bay","Fox Hill Marsh",
                            "Snipe Key","Big River Marsh"))
```

Next, the `reslr_load` function is required for the preparation of input data for the NI GAM, which necessitates additional information not required by earlier models. Firstly, the statistical model relies on an estimate of the "linear local rate" and its associated uncertainty. By setting `include_linear_rate = TRUE`, the package incorporates this rate, which is assumed to stem from physical processes like Glacial Isostatic Adjustment (GIA) (refer to Section 4.2.8 for more information). Users have the flexibility to include their preferred linear rate values as additional columns (`linear_rate` and `linear_rate_err`) in the input dataframe. If these values are not provided, the package automatically calculates them using the available data.

Secondly, users are encouraged to include tide gauge data by setting `include_tide_gauge = TRUE`. As discussed previously, users need to make a decision regarding the inclusion of the closest tide gauge (`TG_minimum_dist_proxy = TRUE`), selecting specific tide gauges by providing a list of names (`list_preferred_TGs = c("ARGENTIA")`), or including all tide gauges within a one-degree proximity of the proxy site (`all_TG_1deg = TRUE`). Additionally, the tide gauge data requires values for the `linear_rate` and `linear_rate_err` columns, which are calculated using a physical model known as an Earth-Ice model called the ICE-5G (VM2) Model (Peltier, 2004) with an un-

certainty value of 0.3 mm/year (Engelhart et al., 2009), both provided within the `reslr` package.

Thirdly, the tide gauge data is averaged over a decade to equate with the resolution of proxy records. If necessary, users can adjust the size of the averaging window to accommodate varying sediment accumulation rates. For example, a longer sediment accumulation rate would result in a larger average, such as 20 years. The default setting for `sediment_average_TG` is 10 years, which we will use in our example.

The final setting of the `reslr_load` function is `prediction_grid_res`, allowing users to modify the resolution of the output plots. The default setting of 50 years serves as a starting point, but users have the flexibility to explore alternative options. For our example, we will utilize nine proxy sites and select all tide gauges within a one-degree range of our proxy site, maximizing the number of data points to demonstrate the capabilities of our package. The specific settings employed are described below:

```
multi_site_input <- reslr_load(
  data = multi_site,
  include_tide_gauge = TRUE,
  include_linear_rate = TRUE,
  TG_minimum_dist_proxy = TRUE,
  all_TG_1deg = TRUE)
```

Similar to the previous example, the output of this function is a list of two dataframes called `data` and `data_grid`. The `data` dataframe is the inputted data with additional columns for the data_type_id which will contain "ProxyRecord" and "TideGaugeData". The `data_grid` is a dataframe that is evenly spaced in time based on the `prediction_grid_res` value chosen by the user and is used to create the plots. In this example, we have 9 proxy sites and 26 tide gauges and the data can be accessed by:

```
glimpse(multi_site_input$data)
```

```
#> Rows: 1,130
#> Columns: 14
```

```
#> $ Region          <chr> "Florida", "Florida", "Florida", "Florida", "Florida",
#> $ Site            <chr> "Nassau", "Nassau", "Nassau", "Nassau", "Nassau", "Nas
#> $ LongLat         <chr> "30.6_-81.7", "30.6_-81.7", "30.6_-81.7", "30.6_-81.7"
#> $ Latitude        <dbl> 30.6, 30.6, 30.6, 30.6, 30.6, 30.6, 30.6, 30.6, 30.6,
#> $ Longitude       <dbl> -81.7, -81.7, -81.7, -81.7, -81.7, -81.7, -81.7, -81.7
#> $ RSL             <dbl> 0.05, 0.03, 0.01, -0.01, -0.03, -0.05, -0.07, -0.09, -
#> $ Age             <dbl> 2002, 1990, 1980, 1974, 1964, 1936, 1920, 1906, 1896,
#> $ Age_err         <dbl> 4.25, 5.50, 4.25, 4.50, 9.50, 10.75, 8.75, 9.75, 9.50,
#> $ RSL_err         <dbl> 0.07, 0.07, 0.07, 0.07, 0.07, 0.07, 0.07, 0.07, 0.07,
#> $ SiteName        <fct> "Nassau,\n Florida", "Nassau,\n Florida", "Nassau,\n F
#> $ data_type_id    <fct> ProxyRecord, ProxyRecord, ProxyRecord, ProxyRecord, Pr
#> $ linear_rate     <dbl> 0.417923, 0.417923, 0.417923, 0.417923, 0.417923, 0.41
#> $ linear_rate_err <dbl> 0.002958023, 0.002958023, 0.002958023, 0.002958023, 0.
#> $ ICE5_GIA_slope  <dbl> 0.278517, 0.278517, 0.278517, 0.278517, 0.278517, 0.27
```

A brief insight into the outputs of the `reslr_input` function, e.g. number of observations and number of locations, can be obtained using the print function shown below:

```
print(multi_site_input)
```

```
#> This is a valid reslr input object with 1130 observations and  35 site(s).
#> There are  9  proxy site(s) and  26  tide gauge site(s).
#> The age units are; Common Era.
#> Decadally averaged tide gauge data included by the package.
#> The linear_rate and linear_rate_err has been included.
```

A plot of the raw data can be created using `plot` function with an option to plot the tide gauges and the proxy records together or have present separate plots for each data source. Figure 4.3.4 demonstrates the resulting plot for the proxy records only using the following function:

```
plot(x = multi_site_input,
     plot_proxy_records = TRUE,
     plot_tide_gauges = FALSE,
     plot_caption = FALSE)
```

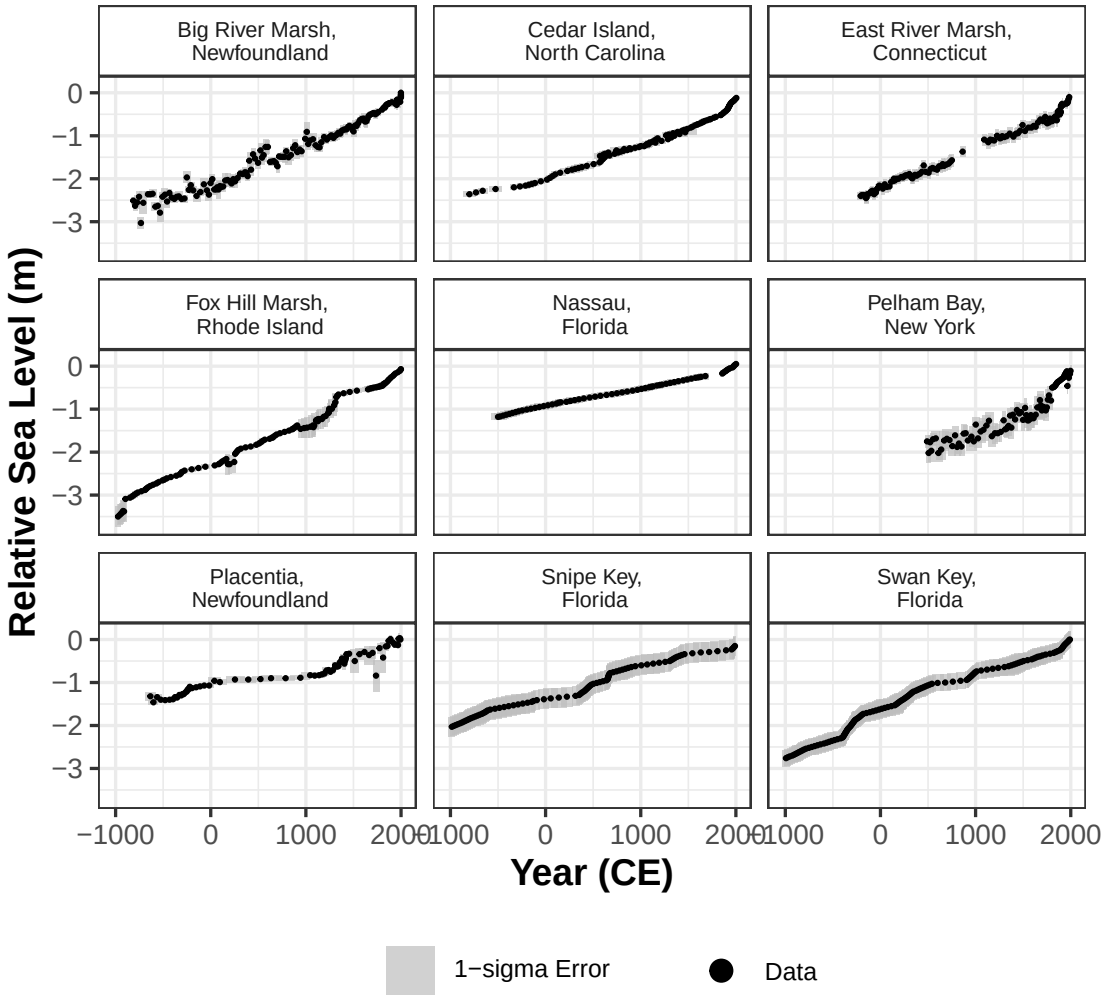


Figure 4.3.4: A plot of the raw data for our nine example sites along the Atlantic coast of North America. The x-axis is time in years in the Common Era (CE) and the y-axis is relative sea level in metres. The grey boxes are 1 standard deviation vertical and horizontal (temporal) uncertainty. The black dots are the midpoints of the uncertainty boxes. The separate sites will appear in separate windows on the plot.

This model type should use `model_type = "ni_gam_decomp"` in the `reslr_mcmc` function and the MCMC simulation settings can be altered to ensure convergence.

```
res_ni_gam_decomp <- reslr_mcmc(
  input_data =  multi_site_input,
```

```
  model_type = "ni_gam_decomp"
)
```

The output of the `reslr_mcmc` function is a list that stores the JAGS model run, the input dataframe and the dataframes for plotting the results. Identical to the other model processes, the convergence of the MCMC algorithm is examined and the parameter estimates from the model can be investigated using the following:

```
summary(res_ni_gam_decomp)
```

```
#> No convergence issues detected.
```

```
#> # A tibble: 4 x 7
#>   variable       mean      sd     mad     q5    q95  rhat
#>   <chr>         <num>   <num>   <num>  <num>  <num> <num>
#> 1 sigma_beta_h 1.76   0.250   0.236   1.39   2.20    1.00
#> 2 sigma_beta_r 0.282  0.0533  0.0501  0.208  0.381   1.00
#> 3 sigma_beta_l 0.971  0.150   0.145   0.747  1.25    1.00
#> 4 sigma_y      0.0142 0.00110 0.00113 0.0125 0.0161  1.00
```

For the parameter estimates, we provide the standard deviation associated with each component of the NI GAM decomposition. Specifically, `"sigma_beta_r"` represents the standard deviation of the spline coefficient for the regional component, `"sigma_beta_l"` represents the standard deviation of the spline coefficient for the non-linear local component, `"sigma_beta_h"` denotes the standard deviation of the site-specific vertical offset component, and `"sigma_y"` indicates the data model variation. These names correspond to the algebraic components described in Section 4.2.8.

One of the key advantages of the NI GAM approach is its ability to decompose regional RSL change into separate components. The results from the `ni_gam_decomp` model can be visualised using the `plot` function, which generates individual plots for each component. Additionally, all components, except for the linear local component, have corresponding rate plots. Users can access the data used to create

each plot in the `res_ni_gam_decomp` object as separate dataframes for each component.

In our example, we demonstrate the rate of change for the total model fit in Figure 4.3.5. This figure illustrates the rate of change at each site, which is useful to understand the variations of the relative sea-level signal, i.e. $f(\mathbf{x}, t)$. To plot the rate of change, users can employ the following method:

```
plot(res_ni_gam_decomp,
     plot_type = "rate_plot",
     plot_caption = FALSE)
```

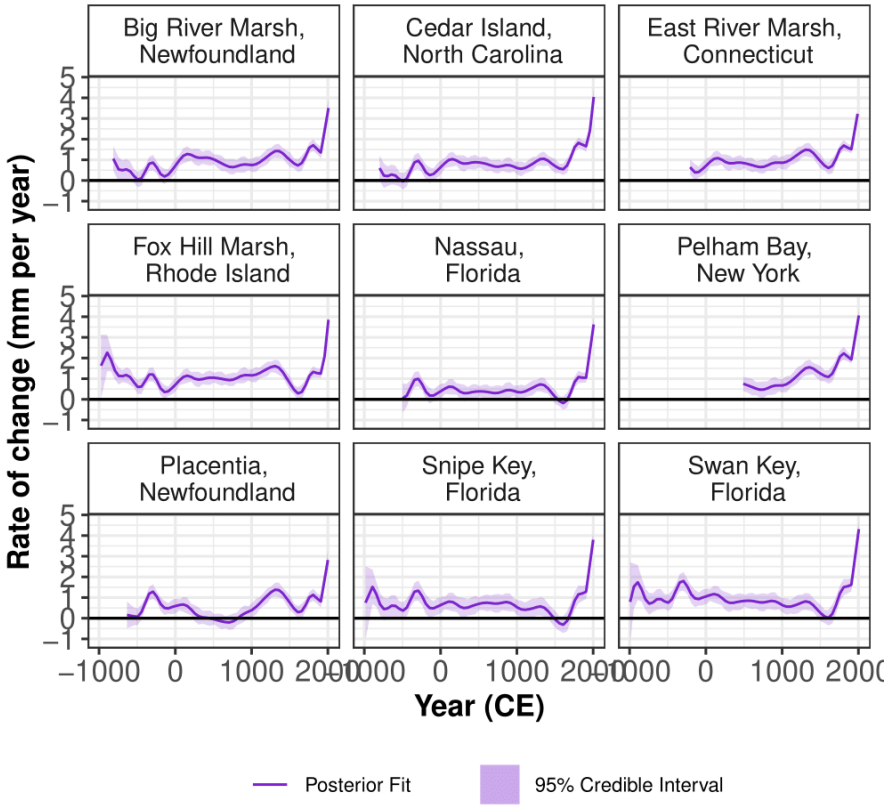


Figure 4.3.5: The rate of change for the total model fit for the noisy input generalised additive model for sites along the Atlantic coast of North America. It is calculated by finding the derivative of the total model fit. The solid purple line is the mean rate of change fit and the shading denotes 95% credible interval for each site along the Atlantic coast of North America. The x-axis is time in years in the Common Era (CE) and the y-axis is rate of change in mm per year.

The regional component ($r(t)$) captures the mean of RSL change along the Atlantic coast of North America. The associated rate of change of the regional component, as seen in Figure 4.3.6, provides an important visual insight into the rate at which this trend varied over the past 3000 years. It is accessed by:

```
plot(res_ni_gam_decomp,
     plot_type = "regional_rate_plot",
     plot_caption = FALSE)
```

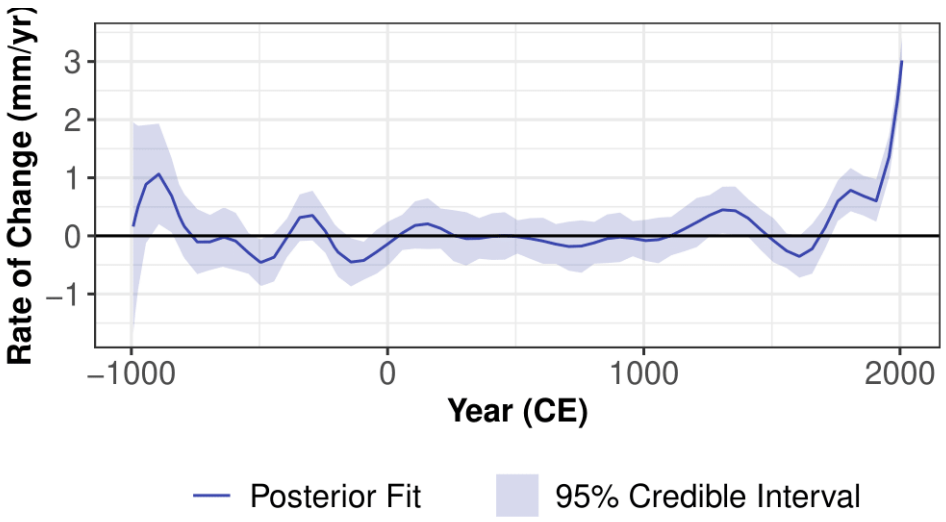


Figure 4.3.6: The rate of change for the regional component of the noisy input generalised additive model for the nine proxy sites and the eleven tide gauges along the Atlantic coast of North America. It is calculated by finding the derivative of the regional component fit. The solid blue line is the mean rate of change fit and the shading denotes 95% credible interval. The x-axis is time in years in the Common Era (CE) and the y-axis is rate of change in mm per year.

4.4 Summary

In this paper, we have presented an overview of the `reslr` package and discussed its various features and design decisions. Our goal was to address the specific needs of the paleo sea-level community and provide an efficient and flexible R package that caters to different types of source data, whilst maintaining a simple workflow that does not require the user to learn too many different functions.

Through two case studies, we demonstrated the simplicity and accessibility of the package. The first case study examined a single site using the NI spline in time. Our results showed that the `reslr` package can provide RSL estimates and associated rate of change values over time for a single location. In the second case study, we showcased the capabilities of the `reslr` package when analysing data from multiple locations. We highlighted its flexibility, allowing for the decomposition of the relative sea-level signal into different components. Additionally, we presented a comprehensive method for incorporating tide-gauge data, which can help to provide valuable insights into recent changes in RSL not captured by proxy records.

There are a number of limitations with our package. One notable constraint arises from the inherent complexity of the NI GAM, as demonstrated in our second case study, which demands a substantial volume of data. This becomes a challenge due to the sparsity often observed in proxy records. Despite offering seven models for diverse users, advanced users are unable to customise the priors in the Bayesian framework. Addressing this would empower those seeking more control in Bayesian modeling with the `reslr` package. Acknowledging and refining these limitations will contribute to the ongoing development and improvement of the `reslr` package.

There are several potential extensions for the `reslr` package. One possibility is to include additional statistical models, such as machine learning techniques, to accommodate larger datasets. Another improvement could be the integration of other instrumental data sources, such as satellite data, enabling the examination of other variables related to climate change. Overall, the `reslr` package offers a powerful toolkit for the paleo sea-level community, and we anticipate that it will continue to evolve and expand its capability to meet the evolving needs of researchers in this field.

Appendix

4.A Example Dataset

The `reslr` package contains a dataset used as an example called `NAACproxydata`. This dataset contains proxy records from the Atlantic coast of North America as used in Upton et al. (2023). The 21 different proxy data sites and the references for each data source can be found in Table 4.A.1.

Site Name	Reference
Barn Island, Connecticut	Donnelly et al. (2004); Gehrels et al. (2020)
Big River Marsh, Newfoundland	Kemp et al. (2018)
Cape May Courthouse, New Jersey	Kemp et al. (2013); Cahill et al. (2016)
Cedar Island, North Carolina	Kemp et al. (2011, 2017)
Cheesequake, New Jersey	Walker et al. (2021)
Chezzetcook Inlet, Nova Scotia	Gehrels et al. (2020)
East River Marsh, Connecticut	Kemp et al. (2015); Stearns et al. (2023)
Fox Hill Marsh, Rhode Island	Stearns et al. (2023)
Leeds Point, New Jersey	Kemp et al. (2013); Cahill et al. (2016)
Les Sillons, Magdelen Islands	Barnett et al. (2017)
Little Manatee River, Florida	Gerlach et al. (2017)
Nassau, Florida	Kemp et al. (2014)
Pelham Bay, New York	Kemp et al. (2017); Stearns and Engelhart (2017)
Placentia, Newfoundland	Kemp et al. (2018)
Revere, Massachusetts	Donnelly (2006)
Saint Simeon, Quebec	Barnett et al. (2017)
Sanborn Cove, Maine	Gehrels et al. (2020)
Sand Point, North Carolina	Kemp et al. (2011, 2017)
Snipe Key, Florida	Khan et al. (2022)
Swan Key, Florida	Khan et al. (2022)
Wood Island, Massachusetts	Kemp et al. (2011)

Table 4.A.1: A list of names of all the sites available in the example dataset within the `reslr` package. For each site we include the reference in the literature to the source of the data.

4.B Vignettes for the reslr package

The reslr package includes three distinct vignettes to aid users in effectively utilising its functionalities. The first vignette, titled the Quick Start Guide, offers users a concise overview of the key functions within the reslr package. The second vignette, known as reslr, serves as the primary vignette, providing comprehensive and detailed instructions for each function and configuration available in the package. Lastly, the Advanced vignette is designed to enable users to delve deeper into specific functions, offering a more thorough exploration of their capabilities. In this thesis, we have included the complete set of associated vignettes and they can also be accessed at https://maeveupton.github.io/reslr/news/index.html.

4.B.1 Main Vignette for reslr package

reslr: Statistical Models for examining Relative Sea Level Change in R

Maeve Upton, Andrew Parnell & Niamh Cahill

4.B.1.1 Introduction

If you require fast instructions, check out the reslr: Quick start.

The reslr package is specifically developed for Bayesian modeling of relative sea-level data. It offers a diverse selection of statistical models, including linear regression, change-point regression, integrated Gaussian process regression, splines, and generalized additive models. One notable feature is the incorporation of measurement uncertainty in multiple dimensions, which is crucial when analyzing relative sea-level data. The package provides a unified framework for data loading, model fitting, and summarising changes in relative sea level (RSL) over time and space. The generated plots depict sea level curves and corresponding rates of change, taking into account the associated uncertainty.

4.B.1.2 Modelling Options

There are a number of modelling options available to the user:

Statistical Model	Model Complexity & Information	model_type code
Errors in variables simple linear regression	Low: A straight line of best fit taking account of any age and measurement errors in the RSL values using the method of Cahill et al (2015)	**"eiv_slr_t"**

Statistical Model	Model Complexity & Information	`model_type` code
Errors in variables change point model	Medium: An extension of the linear regression modelling process. It uses piece-wise linear sections and estimates where/when trend changes occur in the data (Cahill et al. 2015).	**"eiv_cp_t"**
Errors in variables integrated Gaussian Process	High: A non linear fit that utilities a Gaussian process prior on the rate of sea-level change that is then integrated (Cahill et al. 2015).	**"eiv_igp_t"**
Noisy Input spline in time	High: A non-linear fit using regression splines using the method of Upton et al (2023).	**"ni_spline_t"**
Noisy Input spline in space and time	High: A non-linear fit for a set of sites across a region using the method of Upton et al (2023).	**"ni_spline_st"**
Noisy Input Generalised Additive model for the decomposition of the RSL signal	Extreme: A non-linear fit for a set of sites across a region and provides a decomposition of the signal into regional, local-linear (commonly GIA) and local non-linear components. Again this full model is as described in Upton et al (2023).	**"ni_gam_decomp"**

As presented in this table, users can choose from seven statistical models with varying complexities. The simplest, labeled "Low," is the errors-in-variables simple linear regression, suitable for examining linear trends in sea level data but limited to linear trends and a single location. The errors-in-variables change point model, categorized as "Medium" complexity, requires users to specify 1, 2, or 3 change points, making it effective for identifying abrupt changes and estimating change rates but limited to a few change points and a single location.

Moving to higher complexity, the errors-in-variables integrated Gaussian Process is labeled "High" and offers advantages over previous models. It is flexible in time, providing sea level trends and associated rates of change, though confined to a single location with long computational run times. For improved computational efficiency, the noisy input spline in time maintains advantages of the errors-in-variables integrated Gaussian process but is restricted to a single location.

To explore multiple locations in time and space, the noisy input spline in space and time, labeled "High" complexity, captures sea level trends and rates for many locations simultaneously. The final approach, the noisy input generalised additive model for decomposing the RSL signal, is classified as "Extreme" complexity. Users need to provide additional information on the physical processes driving RSL changes in space and time, offering insights into these underlying mechanisms.

For all of the above models the user is able to quantify and visualise changes of RSL and rates of change with associated uncertainties. Indeed a full posterior distribution ensemble of values is available in the output of the functions. For the decomposed full model, **"ni_gam_decomp"**, the user is able to access the posterior probability distributions of the individual components.

4.B.1.3 Installation of the `reslr` package

Installation of the reslr package

The `reslr` package uses the JAGS (Just Another Gibbs Sampler) software to run the models. Before installing `reslr`, visit the JAGS website and download and install JAGS for your operating system.

Next, start Rstudio and find the window with the command prompt (the symbol `>`). Type

```
install.packages("reslr")
```

It may ask you to pick your nearest CRAN mirror (the nearest site which hosts R packages). You will then see some activity on the screen as the `reslr` package and the other packages it uses are downloaded. The final line should then read:

```
package 'reslr' successfully unpacked and MD5 sums checked
```

You then need to load the package.

```
library(reslr)
```

This will load the `reslr` package and all the associated packages. You'll need to type the `library(reslr)` command every time you start R. If you have problems, visit the Issues page and leave a message to tell us what went wrong.

4.B.1.4 Considerations before running reslr

Prior to running the `reslr` package, there are a few points to consider.

4.B.1.5 Installating JAGS software

In this package, the models are written using Just Another Gibbs Sample (JAGS) which uses Gibbs sampling and Markov Chain Monte Carlo (MCMC) algorithm to draw samples from the posterior distribution of the unknown parameters. To download the JAGS package use this link.

4.B.1.6 Working with scripts

The best way to use the `reslr` package is by creating scripts. A script can be created in Rstudio by clicking `File > New File > Rscript`. This opens a text window which allows commands to be typed in order and saved. The command can be sent to the command prompt (which Rstudio calls the Console) by highlighting the command and clicking Run (or going to Code > Run Lines). There are also keyboard shortcuts to speed up the process. We strongly recommend you learn to run R via scripts.

4.B.1.7 Inputting User's data

`reslr` can handle three different types of data structure. It is important to note that varying the number of data sites will require different statistical modelling strategy:

- A single site. This may occur when you have data for only one individual data site. In the case of a single site, we recommend using a temporal model, for example EIV Integrated Gaussian Process or NI spline in time.
- Multiple sites. This may occur if you have a dataset which has multiple different data sites. In this situation, the user must use a spatial temporal model, for example NI spline in space time.
- Multiple group of sites with different drivers of change. This may occur if you have multiple locations are interested in investigating how the regional, linear local and non-linear local components vary. In this case, the NI GAM decomposition is recommended.

The user must ensure that the input data is a dataframe. For a single site or multiple sites only one dataframe should be given to the package, i.e. combined all sites into one dataframe, with the following columns names:

Region	Site	Latitude	Longitude	RSL	Age	Age_err	RSL_err	linear_rate	linear_rate_err
"Leeds Point"	"New Jersey"	41.33	-71.86	-0.91	1336.00	16.50	0.04	1.69	0.03
"Leeds Point"	"New Jersey"	41.33	-71.86	-0.78	1413.00	13.50	0.04	1.69	0.03
...	...	...	...	...	...	...	...	...	...
"Cedar Island"	"North Carolina"	40.33	-72.8	-0.63	1570.00	43.00	0.05	0.74	0.01
"Cedar Island"	"North Carolina"	40.33	-72.8	-0.53	1655.00	13.50	0.05	0.74	0.01

- Site is the name of the site in question, e.g. Leeds Point.
- Region is the area in which is was collected, e.g. New Jersey. To note, in the package, the Site and the Region columns will be combined to form a factor called the `SiteName` which results in an output, e.g.

  ```
  ''Leeds Point,\n New Jersey''
  ```

- Age is the year of the data in Common era ("CE") or Before Present ("BP"). If calibrated in the BP form the package will convert the data into the Common Era form.

- Age_err is the 1 standard deviation (σ) Age uncertainty associated with the Age.
- RSL is the Relative Sea Level measured in meters.
- RSL_err is the 1 standard deviation (σ) Relative Sea level uncertainty associated with the RSL measured in meters.
- Longitude is the longitude of the site in degrees (in decimal, range: -180 to 180). It is important to note that if there are slightly different longitude values for all the observations in site this will lead to errors. Therefore, select one longitude value for each site.
- Latitude is the Latitude of the site in degrees (degree in decimal, range -90 to 90). Similarly, it is important to note that if there are slightly different latitude values for all the observations in site this will lead to errors. Therefore, select one latitude value for each site.
- **Optional** linear_rate is a column that contains the linear rate in mm per year associated with that site, arising from processes such as glacial isostatic adjustment (GIA). *This is not a requirement.* For the NI GAM decomposition model, the user has the option of using their own rate, otherwise the package will calculate it using the data. It is important to note that if the linear_rate is provided without the linear_rate_err column this will lead to issues, i.e. both columns need to be provided
- **Optional** linear_rate_err is a column that contains associated uncertainty for the linear rate in mm per year. *This is not a requirement.* For the NI GAM decomposition model, the user has the option of using their value, otherwise the package will calculate it using the data. It is important to note that if the linear_rate_err is provided without the linear_rate column this will lead to issues, i.e. both columns need to be provided.

4.B.1.8 Tide Gauge Data

There is an option in the `reslr` package to include tide gauge data as an additional source of data which we recommend when using the `model_type = "ni_gam_decomp"`. The package will extract tide gauge data from the PSMSL website. The data is downloaded from this website and stored in a temporary directory.

The tide gauges undergo a number of processing steps within the package. Firstly, certain tide gauges have been flagged by the PSMSL website and we remove these locations. Secondly, the tide gauge data in the PSMSL database is given in millimetres relative to a revised local reference datum (a coordinate system which defines the zero level for sea level measurements (Pugh et al., 2014)). We transform the data by removing 7000 mm to revert the tide gauge data into the observed reference frame and convert the RSL to metres following the PSMSL guidance as described in Aarup et al. 2006 . Lastly, the tide gauge data is averaged over a decade to make it comparable with sedimentation rates associated proxy records sedimentation rates. The user can alter the size of the average if required when accumulation rates for the sediment in the proxy record is estimated to have a higher or lower accumulation rate, e.g. longer sediment

accumulation rate result in a larger window average of 20 years.

Within the `reslr_load` function, the user has three options to choose from:

1. Provide a list of the preferred tide gauges from the PSMSL website, ensuring spelling, capitalisation and spacing is exactly the same as the website. Note, the package will not work if error in spelling occurs. In addition, certain tide gauges have been flagged by the PSMSL website and are not included in this package and will return an error if selected. This is done by giving a list to the `list_preferred_TGs = c("ARGENTIA","MAYPORT")` option in the `reslr_mcmc` function
2. The nearest tide gauge to proxy site based on minimum distance in kilometers, which is done by setting `TG_minimum_dist_proxy = TRUE`.
3. Any tide gauge within 1 degree from the proxy site, which is done by setting `all_TG_1deg = TRUE`.

The user can select a combination of option 1 and option 2 or option 1 and option 3 which allows for additional tide gauge data to be included. The final output is a data frame which contains an additional column, called `data_type_id`, identifying the data source "ProxyRecord" or "TideGaugeData" depending on the observation in question.

4.B.1.9 Glacial Isostatic Adjustment (GIA)

For the NI GAM decomposition, the statistical model requires an estimate for the local linear rate arising from processes such as GIA and associated uncertainty for this rate each site. According to Whitehouse (2018), GIA represents the Earth's reaction to the growth or melting of ice sheets, including the gravitational field and ocean. GIA can be approximated as a linear contribution over a short timescale, but with variable effects along the coast (Engelhart et al., 2009). Earth-ice models, which incorporate the physical structure of the Earth to predict GIA changes due to ice loading and unloading, can provide estimates of GIA rates. There are a range of Earth-ice models with one such example being the ICE5G VM2-90 (Peltier, 2004). It should be noted that other processes, such as tectonic vertical land motion, can mimic the linear trend of GIA. Therefore, the linear local component within the NI GAM decomposition may account for contributions from processes other than GIA that drive changes in relative sea level. These are included as additional columns, `linear_rate` and `linear_rate_err`, in the input dataframe provided by the user.

If the GIA rate for the proxy site is not provided then package will automatically calculate these rates using the data provided and we do not estimates the rates from any Earth-ice physical model. The user can source their own rate estimates as previously mentioned. Two examples of

GIA rate sources (not limited to) include Prof. Peltier's webpage and the associated publication (Peltier, 2004) or the Caron et al. 2018 publication and data.

Important to note, the tide gauge data require values for the `linear_rate` and `linear_rate_err` columns. This is calculated using ICE-5G (VM2 L90) Earth ice model (Peltier et al. 2004) with an uncertainty value of 0.3 mm/year from Engelhart et al. 2009.

4.B.1.10 Example Data Set

The `reslr` package possesses a large dataset used as an example called `NAACproxydata`. This dataset contains proxy records from the Atlantic coast of North America as used in Upton et al 2023 along with tide gauge data which will be discussed below. The 21 different proxy data sites and the references for each data source can be found in the following table:

Site Name	Reference
Barn Island, Connecticut	Donnelly et al (2004), Gehrels et al (2020)
Big River Marsh, Newfoundland	Kemp et al (2018)
Cape May Courthouse, New Jersey	Kemp et al (2013), Cahill et al (2016)
Cedar Island, North Carolina	Kemp et al (2011), Kemp et al (2017)
Cheesequake, New Jersey	Walker et al (2021)
Chezzetcook Inlet, Nova Scotia	Gehrels et al (2020)
East River Marsh, Connecticut	Kemp et al (2015), Stearns et al (2023)
Fox Hill Marsh, Rhode Island	Stearns et al (2023)
Leeds Point, New Jersey	Kemp et al (2013), Cahill et al (2016)
Les Sillons, Magdelen Islands	Barnett et al (2017)
Little Manatee River, Florida	Gerlach et al (2017)
Nassau, Florida	Kemp et al (2014)
Pelham Bay, New York	Kemp et al (2017), Stearns et al (2017)
Placentia, Newfoundland	Kemp et al (2018)
Revere, Massachusetts	Donnelly et al (2006)
Saint Simeon, Quebec	Barnett et al (2017)
Sanborn Cove, Maine	Gehrels et al (2020)
Sand Point, North Carolina	Kemp et al (2011), Kemp et al (2017)
Snipe Key, Florida	Khan et al (2022)
Swan Key, Florida	Khan et al (2022)
Wood Island, Massachusetts	Kemp et al (2011)

The `NAACproxydata` is a data frame with 1715 rows and 8 columns which include:

- Region: All regions along Atlantic coast of North America
- Site: All sites along Atlantic coast of North America
- Latitude: Latitude of the data site
- Longitude: Longitude of the data site
- RSL: Relative Sea level in meters
- RSL_err: 1 standard deviation error associated with relative sea level measured in meters
- Age: Age in years common era (CE)
- Age_err: 1 standard deviation error associated with the Age in years CE

If you are interested in a specific site or multiple sites from the example dataset, then filter for that site prior to running the package, using the following method:

```
# For 1 site
data_1site <- reslr::NAACproxydata %>% dplyr::filter(Site == "Cedar Island")
# For multiple sites
data_multisite <- reslr::NAACproxydata %>% dplyr::filter(Site \%in\% c(
  "Snipe Key", "Cheesequake",
  "Placentia", "Leeds Point"
))
```

4.B.1.11 How to run reslr

The general structure for running `reslr` is as follows:

Step 0. Users have the option to use the provided database, accessed by using `reslr::NAACproxydata` or they can supply their own data. In the latter case, it is imperative for users to confirm that their data is in dataframe format and adhere to the template outlined in Section 4.B.1.7.

Step 1. Load in the data using `reslr_load`. The user can alter the following options:

- When analysing proxy RSL data, tide gauge data serves as an additional source that users may need to examine recent RSL changes. If tide gauge data is necessary, please update the argument to `include_tide_gauge = TRUE`. From there, users have three options as described in Section 4.B.1.8.
- If sedimentation accumulation rates for the proxy records are less than or greater than a decade the user can alter this size using `sediment_average_TG = 10`. The default setting is 10 years.
- If the user is interested in the linear rate, please update the argument to `include_linear_rate = TRUE`. More information about the linear_rate is given in Section 4.B.1.9.
- The user can select the resolution of the output by changing the value of `prediction_grid_res = 50`. The default setting is 50 years.

- The input_age_type argument is associated with the type of input age where the default is in Common Era as described in Section 4.B.1.7.

Step 1a. The print function provides a brief insight into the inputted data.

Step 2. Plot the raw data using plot and select whether to include tide gauges in the output plot. The user can update the title (plot_title) and axis labels (xlab,ylab). The captions (plot_captions) can be included on the plots which provides a summary of the number of proxy sites and tide gauge sites.

Step 3. Choose your preferred model type from the available list above and use the reslr_mcmc function to obtain the parameter estimates and the dataframes required for plotting the outputs. This function has a number of settings which allow the user to improve model diagnostics. In addition, this function allows the user to chose their preferred credible interval size, the default setting is CI = 0.95.

Step 3a. The print function provides a brief insight into the output of the reslr_mcmc function.

Step 4. Check the model converged and examine the results of the parameters with the summary function

Step 5. Visualise the results with plot and access the dataframes used to create the plots. The plot_type option allows users to print individual plots, for example the model fits ("model_fit_plot") and the rates ("rate_plot") separately. The captions (plot_captions) can be included on the plots which provides a summary of the model type, the number of proxy sites and tide gauge sites. The user can select to include the tide gauge (plot_tide_gauges) in the output plots.

4.B.1.12 Errors-in-Variables Simple Linear Regression ("eiv_slr_t")

The simplest model the reslr package can fit is a simple linear regression using the Errors-in-Variables method to account for the uncertainty associated with the proxy records, i.e. uncertainty associated with input (age) and the output (RSL). We would not recommend any model simpler than this (e.g. lm) as it will ignore some of the key uncertainties in the data.

This technique focuses on 1 site and is not recommended for multiple proxy sites together. Tide gauge data can be included to gain insight into recent changes in RSL, however, the user must investigate which tide gauge is suitable. As an example, we will filter the example dataset NAACproxydata to select one site to demonstrate the process:

```
# For 1 site
CedarIslandNC <- NAACproxydata %>% dplyr::filter(Site == "Cedar Island")
```

Step 1: Load in the data using the reslr_load function:

```
CedarIslandNC_input <- reslr_load(
  data = CedarIslandNC,
  include_tide_gauge = FALSE,
  include_linear_rate = FALSE,
  TG_minimum_dist_proxy = FALSE,
  list_preferred_TGs = NULL},
  all_TG_1deg = FALSE,
  prediction_grid_res = 50,
  input_age_type = "CE",
  sediment_average_TG = 10
)
```

In this function, the user can select to add tide gauge data and estimates for linear_rate, by changing `include_tide_gauge =TRUE` and `include_linear_rate =TRUE` respectfully. If `include_tide_gauge =TRUE` the user must decide if they require the closest tide gauge i.e. `TG_minimum_dist_proxy = TRUE`, or select specific tide gauge i.e. `list_preferred_TGs = c("ARGENTIA")`, or all tide gauges within 1 degree of the proxy site i.e. `all_TG_1deg = TRUE`. The default setting is `rolling_window_average = 10` which corresponds to sediment accumulation rates of the proxy records, yet the user has the ability to alter this sediment accumulation rate. Note that for a simple linear regression we recommend using the default settings as demonstrated in the above code chunk. The user can alter the resolution of the output plots using `prediction_grid_res` with the default set at 50 years.

The output of this function is a list of two dataframes called `data` and `data_grid`.

- The `data` dataframe is the inputted data with additional columns for the linear_rate, linear_rate_err and data_type_id which will contain two options, "ProxyRecord" or "TideGaugeData". It can be accessed by:

```
data <- CedarIslandNC_input$data
```

- The `data_grid` is a dataframe that is evenly spaced in time based on the `prediction_grid_res` value chosen by the user and is used to create the plots and is accessed by

```
data_grid <-CedarIslandNC_input$data_grid
```

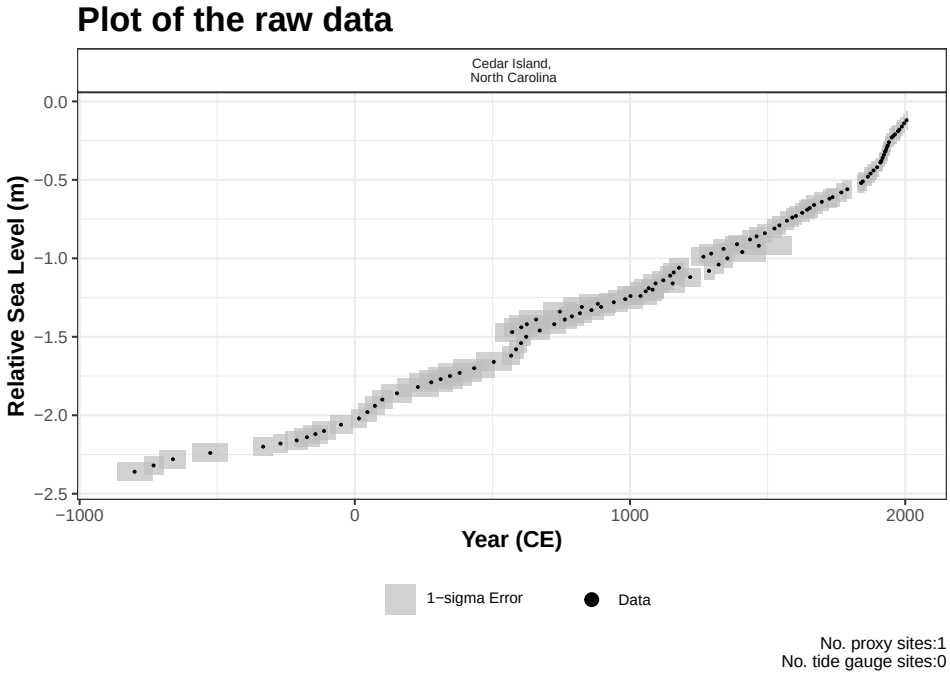
Step 1a: A brief insight into the outputs of the `reslr_input` function can be obtained using:

```
print(CedarIslandNC_input)
#> This is a valid reslr input object with 104 observations and  1 site(s).
#> There are  1  proxy site(s) and  0  tide gauge site(s).
```

```
#> The age units are; Common Era.
#> Decadally averaged tide gauge data was not included. It is recommended for the ni_gam_decomp model
#> The linear_rate or linear_rate_err was not included. It is required for the ni_gam_decomp model
```

Step 2: Plotting the data the raw data with:

```
plot(x = CedarIslandNC_input,
  title = "Plot of the raw data",
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)",
  plot_tide_gauges = FALSE,
  plot_proxy_records = TRUE,
  plot_caption = TRUE)
```



This will produce a plot of Age on the x-axis and Relative Sea Level on the y-axis in meters. Grey boxes represent the uncertainty associated with the vertical and horizontal uncertainty. The black data points are the midpoints of these uncertainty boxes. The following extra arguments can be used which allows the user to updated the titles and axis labels. The caption `plot_caption`, included by default, provides the number of proxy sites and tide gauge sites that will be used in the model and can be removed if required `plot_caption = FALSE`. In addition, the user can select to plot the additional tide gauge data, `plot_tide_gauge =TRUE`.

Step 3: To run the the model the following code is used:

```
res_eiv_slr_t <- reslr_mcmc(
  input_data = CedarIslandNC_input,
  model_type = "eiv_slr_t",
  CI = 0.95)
```

This command takes the input data and the user specifies the statistical model, i.e. a simple linear regression using the EIV uncertainty method ("eiv_slr_t"). The `CI` setting allows the user to set the credible intervals, the current default is `CI = 0.95`. The function tells `reslr` to store the output of the model run in an object called `res_eiv_slr_t`.

Step 3a: A brief insight into the outputs of the `reslr_output` function can be obtained using:

```
print(res_eiv_slr_t)
#> This is a valid reslr output object with 104 observations and  1 site(s).
#> There are  1  proxy site(s) and  0  tide gauge site(s).
#> The age units are; Common Era.
#> The model used was the Errors-in-Variables Simple Linear Regression model.
#> The input data has been run via reslr_mcmc and has produced 3000 iterations over 3 MCMC chains.
```

Step 4: The convergence of the algorithm is examined and he parameter estimates from the model can be investigated using the following:

```
summary(res_eiv_slr_t)
#> No convergence issues detected.
#> # A tibble: 3 x 7
#>   variable    mean      sd     mad      q5     q95  rhat
#>   <chr>      <num>   <num>   <num>   <num>   <num> <num>
#> 1 alpha    -1.99   0.0171  0.0178  -2.02   -1.96    1.00
#> 2 beta      0.824  0.0129  0.0127   0.802   0.844   1.00
#> 3 sigma_y   0.0665 0.00943 0.00943  0.0513  0.0820  1.00
```

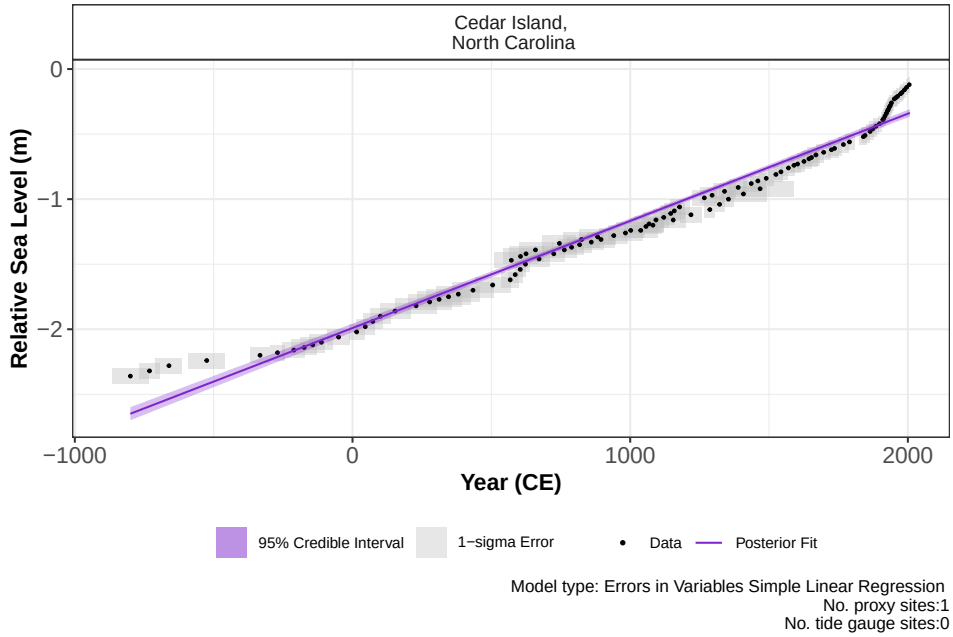
If the model run has the package will print: "No convergence issues detected". If the package prints: "Convergence issues detected, a longer run is necessary". The user is required to update the `reslr_mcmc` function with additional iterations in the following manner:

```
res_eiv_slr_t <- reslr_mcmc(
  input_data = CedarIslandNC_input,
  model_type = "eiv_slr_t",
  # Update these values
  n_iterations = 6000, # Number of iterations
  n_burnin = 1000, # Number of iterations to discard at the beginning
  n_thin = 4, # Reduces number of output samples to save memory and computation time
  n_chains = 3 # Number of Markov chains)
```

The output of this function allows to user to examine the parameter estimates. For the eiv_slr_t model, the parameters of interest are the intercept ("alpha"), the slope ("beta") and the residual standard deviation of the model ("sigma_y"). When using the eiv_slr_t model, an estimate of the of the rate of sea-level change can be obtained by examining the value of the slope, i.e."beta".

Step 5: The results from the eiv_slr_t model can be visualised using the following function:

```
plot(res_eiv_slr_t,
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)")
```



The output of this function is a graph of the input data, i.e. Age and RSL and associated uncertainty boxes, and the model fit with 95 % credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`.

To examine the data creating these plots the user types the following:

```
output_dataframes <- res_eiv_slr_t$output_dataframes
head(output_dataframes)
#>   Longitude Latitude    SiteName    data_type_id  Age      pred
#> 1   -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -800 -2.649079
#> 2   -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -750 -2.607902
#> 3   -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -700 -2.566724
#> 4   -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -650 -2.525546
#> 5   -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -600 -2.484369
#> 6  -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -550 -2.443191
#>         upr       lwr  CI
#> 1  -2.699450  -2.598455 95%
#> 2  -2.657003  -2.558537 95%
#> 3  -2.614578  -2.518346 95%
#> 4  -2.572246  -2.478260 95%
#> 5  -2.529913  -2.438299 95%
#> 6  -2.487382  -2.398293 95%
```

4.B.1.13 Errors-in-Variable Change Point Model ("eiv_cp_t")

The Errors-in-Variable Change Point model is an extension of the linear regression and allows the user to specify the number of change points required.

This technique focuses on 1 site and the maximum number of change points available to the user is 3. We do not recommended for multiple proxy sites together. Tide gauge data can be included to gain insight into recent changes in RSL, however, the user must investigate which tide gauge is most suitable. It is important to note that certain data sites will not work with 2 or 3 change points as there is no distinct changing points in the data. In this case, we recommend testing different number of change points and reviewing the resulting plots to confirm the correct number of change points is selected.

As an example, we will filter the example dataset `NAACproxydata` to select one site to demonstrate the process:

```
# For 1 site
CedarIslandNC <- reslr::NAACproxydata %>% dplyr::filter(Site == "Cedar Island")
```

Step 1: Load in the data using the `reslr_load` function:

```
CedarIslandNC_input <- reslr_load(
  data = CedarIslandNC,
  include_tide_gauge = FALSE,
  include_linear_rate = FALSE,
  TG_minimum_dist_proxy = FALSE,
  list_preferred_TGs = NULL,
  all_TG_1deg = FALSE,
  prediction_grid_res = 50,
  sediment_average_TG = 10)
```

In this function, the user can select to add tide gauge data and estimates for linear_rate, by changing `include_tide_gauge = TRUE` and `include_linear_rate = TRUE` respectfully. If `include_tide_gauge = TRUE` the user must decide if they require the closest tide gauge i.e. `TG_minimum_dist_proxy = TRUE`, or select specific tide gauge i.e. `list_preferred_TGs = c("ARGENTIA")`, or all tide gauges within 1 degree of the proxy site i.e. `all_TG_1deg = TRUE`. The default setting is `sediment_average_TG = 10` which corresponds to sediment accumulation rates of the proxy records, yet the user has the ability to alter this sediment accumulation rate by changing the size of the rolling window average.

Note that for a change point model, we recommend using the default settings as demonstrated in the above code chunk. The user can alter the resolution of the output plots using `prediction_grid_res` with the default set at 50 years. The output of this function is a list of two dataframes called `data` and `data_grid`. The `data` dataframe is the inputted data with additional column for the data_type_id which will contain, "ProxyRecord". It can be accessed by:

```
data <- CedarIslandNC_input$data
head(data)
#>      Region       Site Latitude Longitude   RSL  Age Age_err RSL_err
#> 1 North Carolina Cedar
                    Island   34.971    -76.38 -0.12 2005    2.25    0.06
#> 2 North Carolina Cedar
                    Island   34.971    -76.38 -0.14 1996    2.00    0.06
#> 3 North Carolina Cedar
                    Island   34.971    -76.38 -0.16 1988    5.00    0.06
#> 4 North Carolina Cedar
                    Island   34.971    -76.38 -0.18 1979    5.75    0.06
#> 5 North Carolina Cedar
                    Island   34.971    -76.38 -0.19 1974    5.50    0.06
#> 6 North Carolina Cedar
                    Island   34.971    -76.38 -0.21 1963    5.50    0.06
#>                         SiteName data_type_id
#> 1 Cedar Island,\n North Carolina  ProxyRecord
#> 2 Cedar Island,\n North Carolina  ProxyRecord
#> 3 Cedar Island,\n North Carolina  ProxyRecord
```

```
#> 4 Cedar Island,\n North Carolina  ProxyRecord
#> 5 Cedar Island,\n North Carolina  ProxyRecord
#> 6 Cedar Island,\n North Carolina  ProxyRecord
```

- The `data_grid` is a dataframe that is evenly spaced in time based on the `prediction_grid_res` value chosen by the user and is used to create the plots and is accessed by

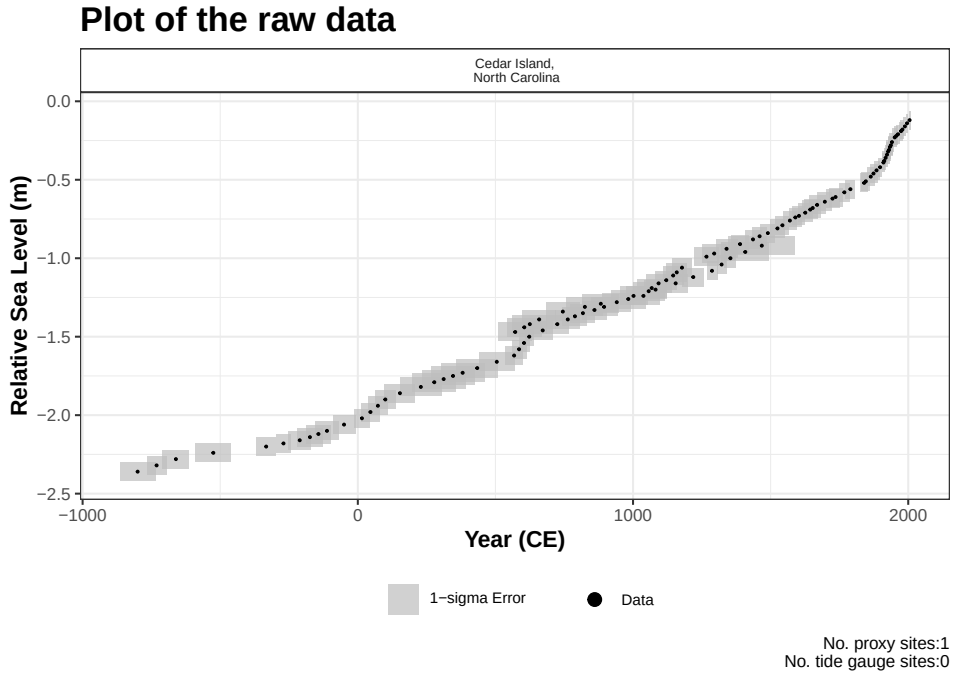
```
data_grid <- CedarIslandNC_input$data_grid
head(data_grid)
#> A tibble: 6 x 5
#> \# Groups:   SiteName [1]
#>   Longitude Latitude SiteName                          data_type_id   Age
#>       <dbl>    <dbl> <fct>                             <fct>        <dbl>
#> 1     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -800
#> 2     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -750
#> 3     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -700
#> 4     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -650
#> 5     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -600
#> 6     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -550
```

Step 1a: A brief insight into the outputs of the `reslr_input` function can be obtained using:

```
print(CedarIslandNC_input)
#> This is a valid reslr input object with 104 observations and  1 site(s).
#> There are  1  proxy site(s) and  0  tide gauge site(s).
#> The age units are; Common Era.
#> Decadally averaged tide gauge data was not included. It is recommended for the ni_gam_decomp model
#> The linear_rate or linear_rate_err was not included. It is required for the ni_gam_decomp model
```

Step 2: Plotting the data the raw data with:

```
plot(x = CedarIslandNC_input,
  title = "Plot of the raw data",
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)",
  plot_proxy_records = TRUE,
  plot_tide_gauges = FALSE)
```



This will produce a plot of Age on the x-axis and Relative Sea Level on the y-axis in meters. Grey boxes represent the uncertainty associated with the vertical and horizontal uncertainty. The black data points are the midpoints of these uncertainty boxes. The extra arguments can be used which allows the user to updated the titles and axis labels. The caption `plot_caption`, included by default, provides the number of proxy sites and tide gauge sites that will be used in the model and can be removed if required `plot_caption = FALSE`. The user can select to plot the additional tide gauge data, `plot_tide_gauge = TRUE`.

Step 3: Run the model using the following code and select the number of change points you require:

```
res_eiv_cp1_t <- reslr_mcmc(
  input_data = CedarIslandNC_input,
  model_type = "eiv_cp_t",
  n_cp = 1,
  CI =0.95)
```

If the user is interested in running 2 change points use method:

```
res_eiv_cp2_t <- reslr_mcmc(
```

```
  input_data = CedarIslandNC_input,
  model_type = "eiv_cp_t",
  n_cp = 2, # Updating the default setting to include an additional change point.
  CI =0.95)
```

The `CI` setting allows the user to set the credible intervals, the current default is `CI = 0.95`. Similar to the earlier model, the output object `res_eiv_cp1_t` stores the JAGS model run and should take a second to run.

Step 3a: A brief insight into the outputs of the `reslr_output` function can be obtained using:

```
print(res\_eiv\_cp1\_t)
#> This is a valid reslr output object with 104 observations and  1 site(s).
#> There are  1  proxy site(s) and  0  tide gauge site(s).
#> The age units are; Common Era.
#> The model used was the Errors-in-Variables Change Point model with 1 change point.
#> The input data has been run via reslr_mcmc and has produced 3000 iterations over 3 MCMC chains.
```

Step 4: The convergence of the algorithm is examined and the parameter estimates from the model can be investigated using the following:

```
summary(res_eiv_cp1_t)
#> No convergence issues detected.
#> A tibble: 5 x 7
#>   variable                mean     sd    mad        q5     q95  rhat
#>   <chr>       <num>  <num>  <num>     <num>   <num> <num>
#> 1 alpha -1.05   0.634  0.0679 -2.04     -0.559   1.00
#> 2 beta[1] 0.646  0.153  0.0249  0.354     0.766   1.00
#> 3 beta[2] 1.99   0.828  0.746   0.880     3.06    1.00
#> 4 Change Point in CE: 1266      0.774  0.0599  0.0518    1.85    1.00
#> 5 sigma_y   0.0156 0.0142 0.0109  0.000854  0.0442  1.00
```

If the model run has the package will print: "No convergence issues detected". If the package prints: "Convergence issues detected, a longer run is necessary". The user is required to update the `reslr_mcmc` function with additional iterations in the following manner:

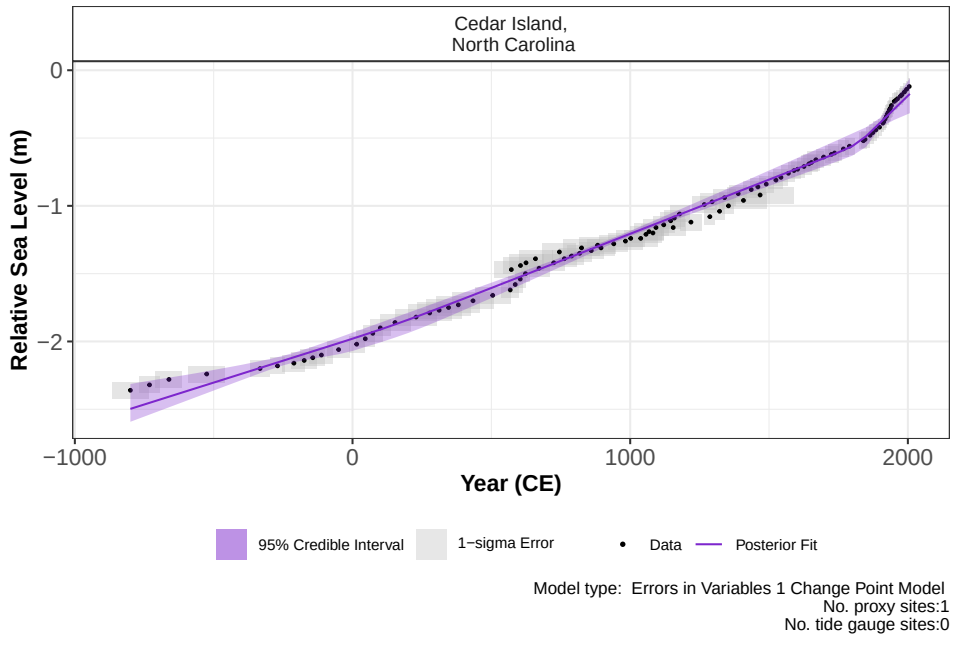
```
res_eiv_cp1_t <- reslr_mcmc(
  input_data = CedarIslandNC_input,
  model_type = "eiv_cp_t",
  # Update these values
```

```
  n_iterations = 6000, # Number of iterations
  n_burnin = 1000, # Number of iterations to discard at the beginning
  n_thin = 4, # Reduces number of output samples to save memory and computation time
  n_chains = 3 # Number of Markov chains)
```

For the eiv_cp_t model, the parameters of interest are the intercept (alpha), the slopes before the change point ("beta[1]") and after the change point ("beta[2]"), the year of the change point (Change Point) and "sigma_y" the variance of the model.

Step 5: The results from the EIV Change Point model can be illustrated using:

```
plot(res_eiv_cp1_t,
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)")
```



The output of this function is a graph of the input data, i.e. Age and RSL and associated uncertainty boxes, and the model fit with 95 % credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`.

To examine the data creating these plots the user types the following:

```
output_dataframes <- res_eiv_cp1_t$output_dataframes
head(output_dataframes)
#>   Longitude Latitude SiteName    data_type_id      Age      pred
#> 1    -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -800 -2.496999
#> 2    -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -750 -2.464701
#> 3    -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -700 -2.432402
#> 4    -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -650 -2.400104
#> 5    -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -600 -2.367806
#> 6    -76.38   34.971 Cedar Island,\n
                         North Carolina  ProxyRecord -550 -2.335507
#>         upr       lwr  CI
#> 1 -2.592305 -2.313382 95%
#> 2 -2.553800 -2.295996 95%
#> 3 -2.515442 -2.279266 95%
#> 4 -2.477038 -2.262920 95%
#> 5 -2.438731 -2.246754 95%
#> 6 -2.400375 -2.229156 95%
```

4.B.1.14 Errors-in-Variable Integrated Gaussian Process Model ("eiv_igp_t")

The EIV Integrated Gaussian Process model provides the underlying rate of the process directly from the model. Further reading on this modeling approach can be found here.

This technique focuses on 1 site and we do not recommended for multiple proxy sites together. Tide gauge data can be included to gain insight into recent changes in RSL, however, the user must investigate which tide gauge is suitable. As an example, we will filter the example dataset `NAACproxydata` to select one site to demonstrate the process:

```
# For 1 site
CedarIslandNC <- reslr::NAACproxydata %>% dplyr::filter(Site == "Cedar Island")
```

Step 1: Load in the data using the `reslr_load` function:

```
CedarIslandNC_input <- reslr_load(
  data = CedarIslandNC,
  include_tide_gauge = FALSE,
```

```
  include_linear_rate = FALSE,
  TG_minimum_dist_proxy = FALSE,
  list_preferred_TGs = NULL,
  all_TG_1deg = FALSE,
  prediction_grid_res = 50,
  sediment_average_TG = 10)
```

In this function, the user can select to add tide gauge data and estimates for linear_rate, by changing `include_tide_gauge = TRUE` and `include_linear_rate = TRUE` respectfully. If `include_tide_gauge = TRUE` the user must decide if they require the closest tide gauge i.e. `TG_minimum_dist_proxy = TRUE`, or select specific tide gauge i.e. `list_preferred_TGs = c("ARGENTIA")`, or all tide gauges within 1 degree of the proxy site i.e. `all_TG_1deg = TRUE`. The default setting is `sediment_average_TG = 10` which corresponds to sediment accumulation rates of the proxy records, yet the user has the ability to alter this sediment accumulation rate.

Note that for an IGP we recommend using the default settings as demonstrated in the above code chunk. The user can alter the resolution of the output plots using `prediction_grid_res` with the default set at 50 years. The output of this function is a list of two dataframes called `data` and `data_grid`. - The `data` dataframe is the inputted data with additional columns for the data_type_id which will contain "ProxyRecord". It can be accessed by:

```
data <- CedarIslandNC_input$data
```

- The `data_grid` is a dataframe that is evenly spaced in time based on the `prediction_grid_res` value chosen by the user and is used to create the plots and is accessed by

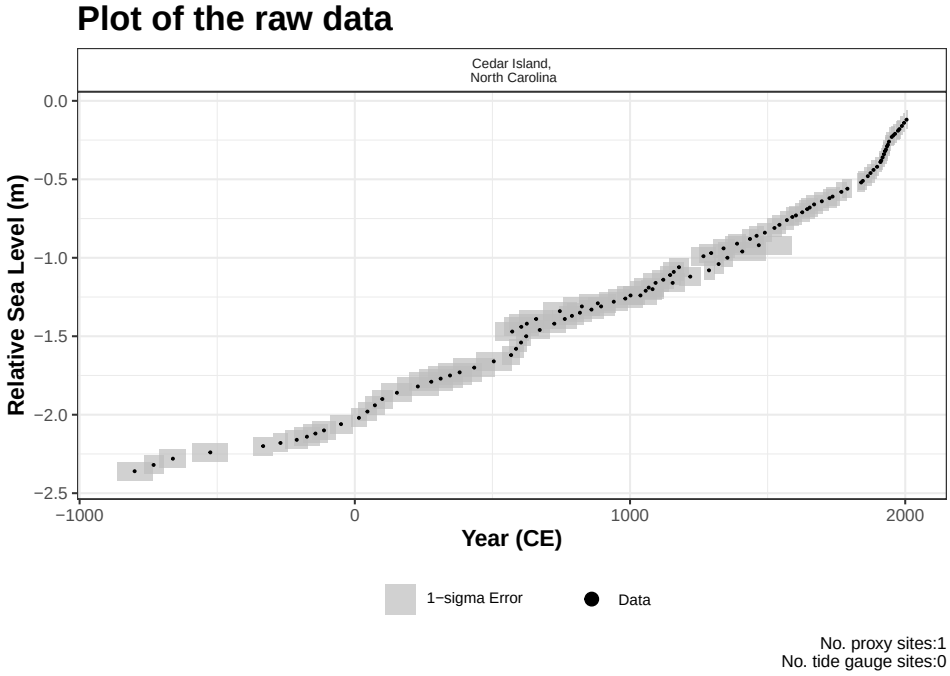
```
data_grid <-CedarIslandNC_input$data_grid
```

Step 1a: A brief insight into the outputs of the `reslr_input` function can be obtained using:

```
print(CedarIslandNC_input)
#> This is a valid reslr input object with 104 observations and  1 site(s).
#> There are  1  proxy site(s) and  0  tide gauge site(s).
#> The age units are; Common Era.
#> Decadally averaged tide gauge data was not included. It is recommended for the ni_gam_decomp model
#> The linear_rate or linear_rate_err was not included. It is required for the ni_gam_decomp model
```

Step 2: Plotting the data the raw data with:

```
plot(
  x = CedarIslandNC_input,
  title = "Plot of the raw data",
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)",
  plot_proxy_records = TRUE,
  plot_tide_gauges = FALSE)
```



This will produce a plot of Age on the x-axis and Relative Sea Level on the y-axis in meters. Grey boxes represent the uncertainty associated with the vertical and horizontal uncertainty. The black data points are the midpoints of these uncertainty boxes. The extra arguments can be used which allows the user to updated the titles and axis labels. The caption `plot_caption`, included by default, provides the number of proxy sites and tide gauge sites that will be used in the model and can be removed if required `plot_caption = FALSE`. In addition, the user can select to plot the additional tide gauge data, `plot_tide_gauge = TRUE`.

Step 3: To run the eiv_igp_t model the following function should be used:

```
res_eiv_igp_t <- reslr_mcmc(
  input_data = CedarIslandNC_input,
  model_type = "eiv_igp_t",
  CI =0.95
)
```

This command takes the input data and the user specifies the statistical model, i.e. an integrated Gaussian process using the EIV uncertainty method ("eiv_slr_t"). It tells `reslr` to store the output of the model run in an object called `res_eiv_igp_t`. The `CI` setting allows the user to set the credible intervals, the current default is `CI = 0.95`. The computational run time for this model is approximately 14 minutes.

Step 3a: A brief insight into the outputs of the `reslr_output` function can be obtained using:

```
print(res_eiv_igp_t)
```

Step 4: The convergence of the algorithm is examined and he parameter estimates from the model can be investigated using the following:

```
summary(res_eiv_igp_t)
```

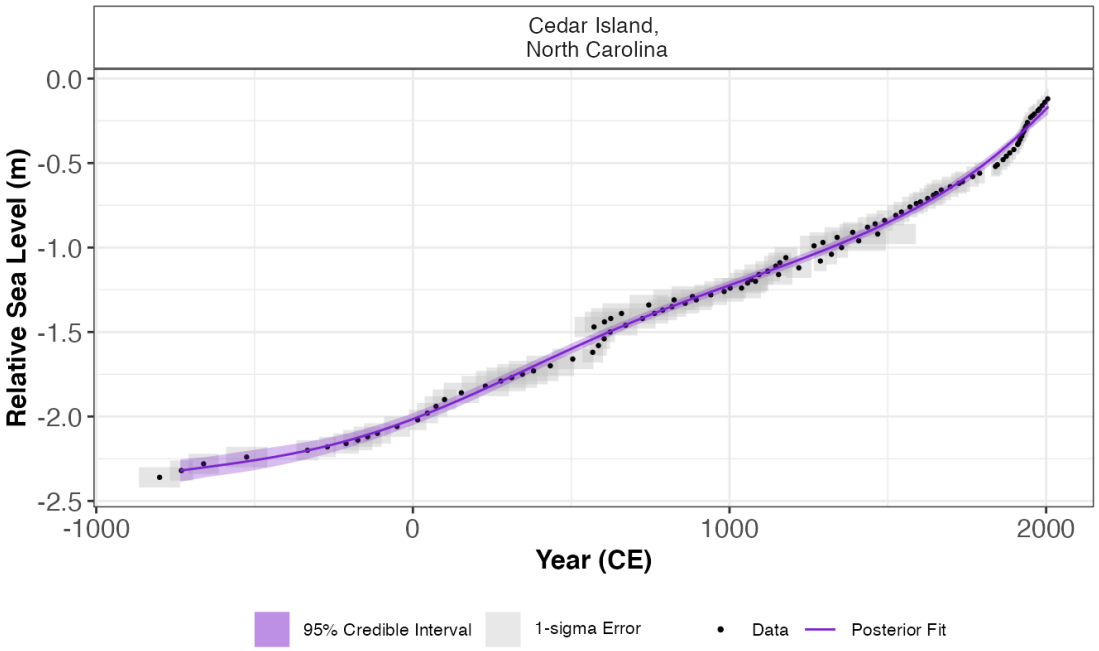
If the model run has the package will print: "No convergence issues detected". If the package prints: "Convergence issues detected, a longer run is necessary". The user is required to update the `reslr_mcmc` function with additional iterations in the following manner:

```
res_eiv_igp_t <- reslr_mcmc(
  input_data = CedarIslandNC_input,
  model_type = "eiv_igp_t",
  # Update these values
  n_iterations = 6000, # Number of iterations
  n_burnin = 1000, # Number of iterations to discard at the beginning
  n_thin = 4, # Reduces number of output samples to save memory and computation time
  n_chains = 3 # Number of Markov chains)
```

For the parameter estimates, the length scale parameter, "rho" is the correlation parameter and "nu" is the standard deviation of the rate process. "sigma_y" is the variation of the model.

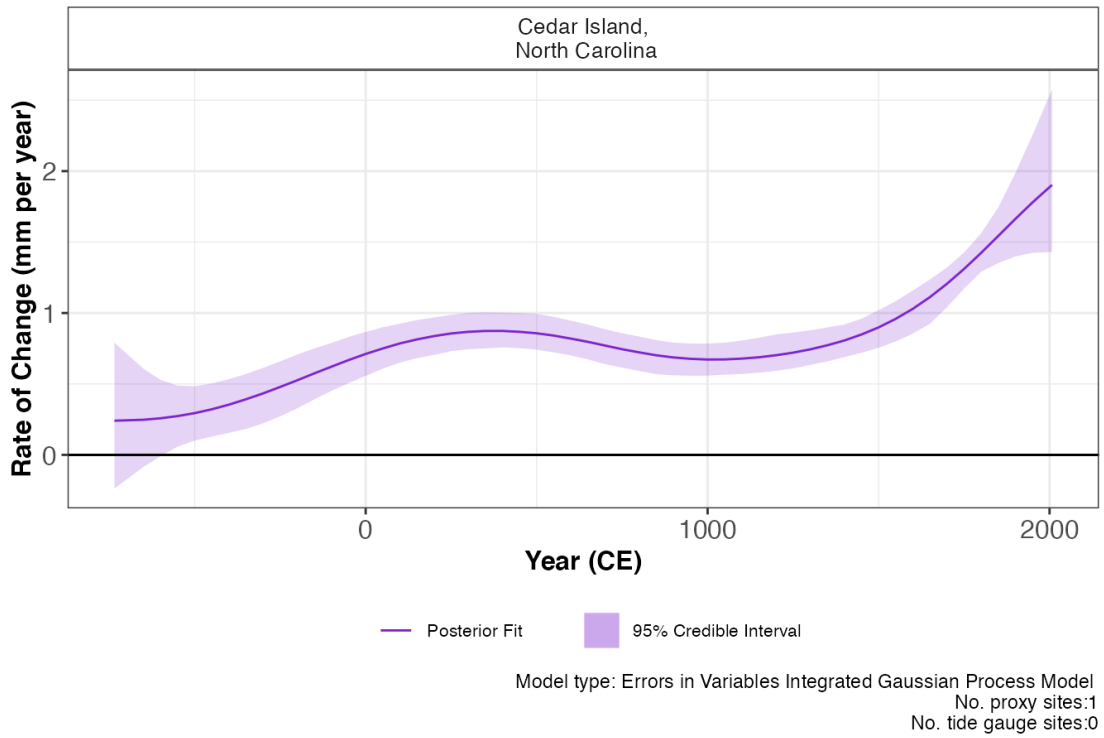
Step 5: The results from the EIV IGP model can be illustrated using:

```
plot(res_eiv_igp_t,
  plot_type = "model_fit_plot",
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)",
  plot_proxy_records = TRUE},
  plot_tide_gauges = FALSE)
```



The output of this function is a graph of the input data, i.e. Age and RSL and associated uncertainty boxes, and the model fit with 95 % credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`. In order to view the rate of change plot, the following setting should be used:

```
plot(res_eiv_igp_t,
  plot_type = "rate_plot",
  xlab = "Year (CE)",
  y_rate_lab = "Rate of Change (mm per year)"
)
```



This prints the plot of the rate of change with 95 % credible intervals. The caption provides the model type, the number of proxy sites and tide gauge sites that were used.

To examine the data creating these plots the user types the following:

```
output_dataframes <-res_eiv_igp_t$output_dataframes
```

4.B.1.15 Noisy input spline in time ("ni_spline_t")

An alternative method to examine how the response variable varies in time is using the Noisy input spline in time (ni_spline_t). It model can obtain results in more efficient computational run times when compared with the eiv_igp_t model.

This technique focuses on 1 site and we do not recommended for multiple proxy sites together. Tide gauge data can be used to gain insight into recent RSL changes. As an example, we will filter the example dataset `NAACproxydata` to select one site to demonstrate the process:

```
# For 1 site
CedarIslandNC <- reslr::NAACproxydata %>% dplyr::filter(Site == "Cedar Island")
```

Step 1: Load in the data using the `reslr_load` function:

```
CedarIslandNC_input <- reslr_load(
  data = CedarIslandNC,
  include_tide_gauge = FALSE,
  include_linear_rate = FALSE,
  TG_minimum_dist_proxy = FALSE,
  list_preferred_TGs = NULL,
  all_TG_1deg = FALSE,
  prediction_grid_res = 50,
  sediment_average_TG = 10
)
```

In this function, the user can select to add tide gauge data and estimates for linear_rate, by changing `include_tide_gauge = TRUE` and `include_linear_rate = TRUE` respectfully. If `include_tide_gauge = TRUE` the user must decide if they require the closest tide gauge i.e. `TG_minimum_dist_proxy = TRUE`, or select specific tide gauge i.e. `list_preferred_TGs = c("ARGENTIA")`, or all tide gauges within 1 degree of the proxy site i.e. `all_TG_1deg = TRUE`. The default setting is `sediment_average_TG = 10` which corresponds to sediment accumulation rates of the proxy records, yet the user has the ability to alter this sediment accumulation rate by changing the size of the rolling window average.

Note that for a spline in time, we recommend using the default settings as demonstrated in the above code chunk. The user can alter the resolution of the output plots using `prediction_grid_res` with the default set at 50 years. The output of this function is a list of two dataframes called `data` and `data_grid`. The `data` dataframe is the inputted data with additional columns for the data_type_id which will contain "ProxyRecord". It can be accessed by:

```
data <- CedarIslandNC_input$data
```

- The `data_grid` is a dataframe that is evenly spaced in time based on the `prediction_grid_res` value chosen by the user and is used to create the plots and is accessed by

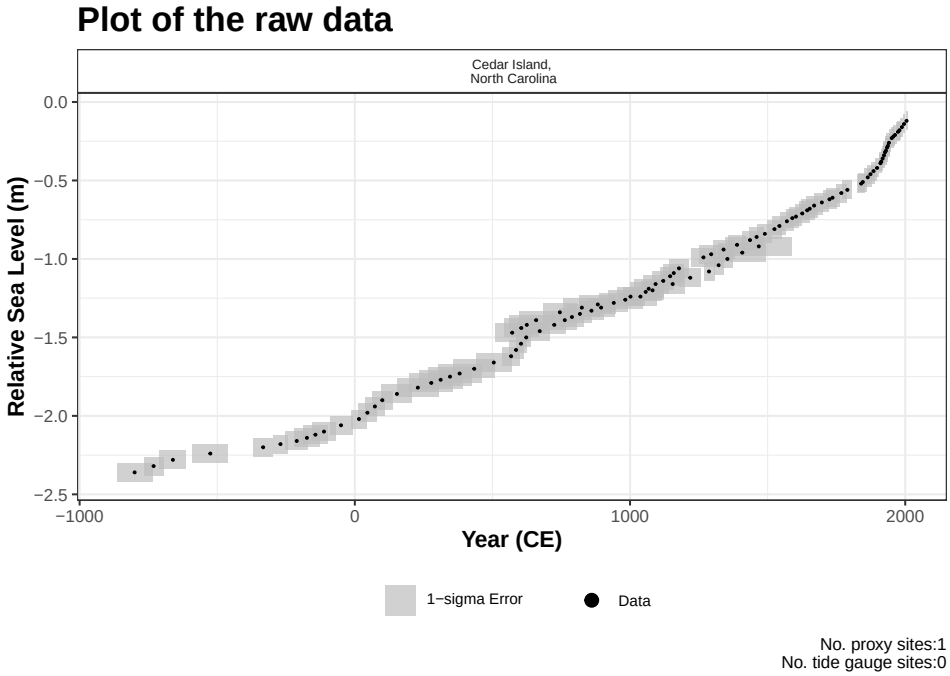
```
data_grid <-CedarIslandNC_input$data_grid
```

Step 1a: A brief insight into the outputs of the `reslr_input` function can be obtained using:

```
print(CedarIslandNC_input)
#> This is a valid reslr input object with 104 observations and  1 site(s).
#> There are  1  proxy site(s) and  0  tide gauge site(s).
#> The age units are; Common Era.
#> Decadally averaged tide gauge data was not included. It is recommended for the ni_gam_decomp model
#> The linear_rate or linear_rate_err was not included. It is required for the ni_gam_decomp model
```

Step 2: Plotting the data the raw data with:

```
plot(
  x = CedarIslandNC_input,
  title = "Plot of the raw data",
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)",
  plot_proxy_records = TRUE,
  plot_tide_gauges = FALSE)
```



This will produce a plot of Age on the x-axis and Relative Sea Level on the y-axis in meters. Grey boxes represent the uncertainty associated with the vertical and horizontal uncertainty. The black data points are the midpoints of these uncertainty boxes. The following extra arguments can be used which allows the user to updated the titles and axis labels. The caption `plot_caption`, included by default, provides the number of proxy sites and tide gauge sites that will be used in the model and can be removed if required `plot_caption = FALSE`. In addition, the user can select to plot the additional tide gauge data, `plot_tide_gauge = TRUE`.

Step 3: To run this model type use the following:

```
res_ni_spline_t <- reslr_mcmc(
  input_data = CedarIslandNC_input,
  model_type = "ni_spline_t",
  CI = 0.95)
```

The output object `res_ni_spline_t` stores the JAGS model run. The `CI` setting allows the user to set the credible intervals, the current default is `CI = 0.95`. Note that there will be two model runs printed in the console here but the output will be the same format as earlier models.

Step 3a: A brief insight into the outputs of the `reslr_output` function can be obtained using:

```
print(res_ni_spline_t)
#> This is a valid reslr output object with 104 observations and  1 site(s).
#> There are  1  proxy site(s) and  0  tide gauge site(s).
#> The age units are; Common Era.
#> The model used was the Noisy Input Spline in time model.
#> The input data has been run via reslr_mcmc and has produced 3000 iterations over 3 MCMC chains.
```

Step 4: The convergence of the algorithm is examined and he parameter estimates from the model can be investigated using the following:

```
summary(res_ni_spline_t)
#> No convergence issues detected.
#> A tibble: 2 x 7
#>   variable      mean      sd     mad       q5    q95  rhat
#>   <chr>     <num>   <num>   <num>    <num>  <num> <num>
#> 1 sigma_beta 2.10    0.699   0.551   1.29     3.43    1.00
#> 2 sigma_y    0.00618 0.00475 0.00456 0.000465 0.0152  1.00
```

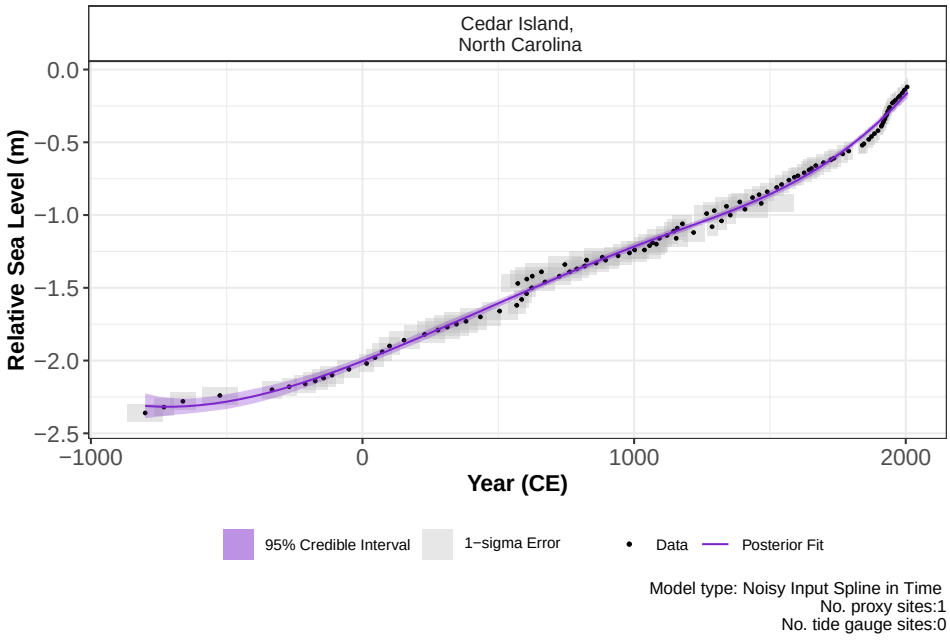
If the model run has the package will print: "No convergence issues detected". If the package prints: "Convergence issues detected, a longer run is necessary". The user is required to update the `reslr_mcmc` function with additional iterations in the following manner:

```
res_ni_spline_t <- reslr_mcmc(
  input_data = CedarIslandNC,
  model_type = "ni_spline_t",
  # Update these values
  n_iterations = 6000, # Number of iterations
  n_burnin = 1000}, # Number of iterations to discard at the beginning
  n_thin = 4, # Reduces number of output samples to save memory and computation time
  n_chains = 3 # Number of Markov chains)
```

For the parameter estimates, we can present the standard deviation associated with the NI spline time model. Where "sigma_beta" highlights the variation associated with the spline coefficient for the spline in time and "sigma_y" presenting the overall variation of the model.

Step 5: the results from the ni_spline_t model can be illustrated using:

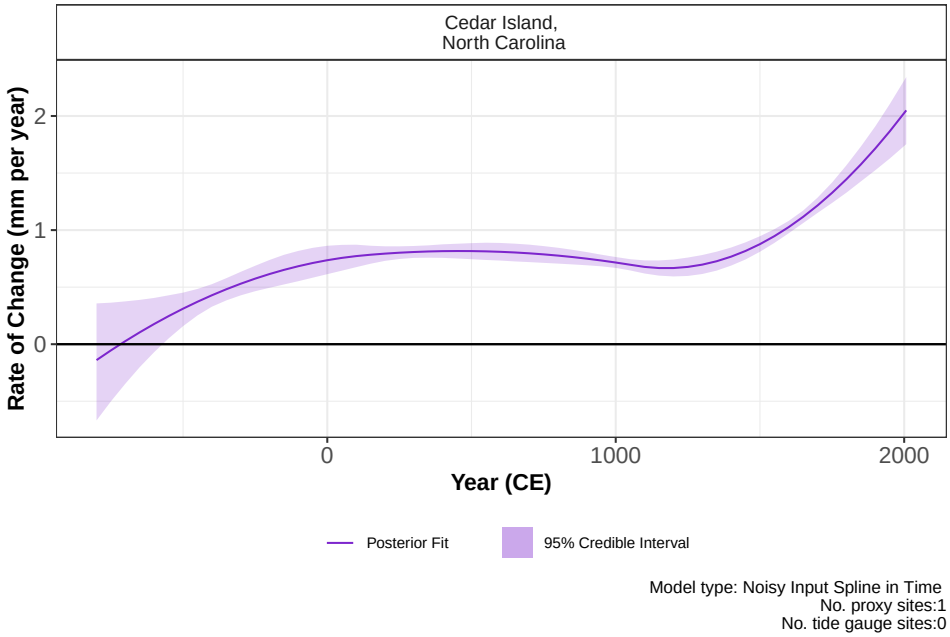
```
plot(res_ni_spline_t,
  plot_type = "model_fit_plot",
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)")
```



The output of this function is a graph of the input data, i.e. Age and RSL and associated uncertainty boxes, and the model fit with 95 % credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if

necessary with `plot_caption = FALSE`. In order to view the rate of change plot, the following setting should be used:

```
plot(res_ni_spline_t,
  plot_type = "rate_plot",
  xlab = "Year (CE)",
  y_rate_lab = "Rate of Change (mm per year)")
```



This prints the plot of the rate of change with 95 % credible intervals. Again, the caption provides the model type, number of proxy sites and tide gauge sites that were used.

To examine the data creating these plots the user types the following:

```
output_dataframes <-res_ni_spline_t$output_dataframes
head(output_dataframes)
#>   Longitude Latitude                       SiteName data_type_id  Age      pred
#> 1    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -800 -2.311847
#> 2    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -750 -2.316669
#> 3    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -700 -2.317327
#> 4    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -650 -2.314025
#> 5    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -600 -2.306967
#> 6    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -550 -2.296357
#>         upr       lwr   rate_pred    rate_upr  rate_lwr  CI
```

```
#> 1 -2.395677 -2.225187 -0.13944226 -0.66827226 0.3574160 95%
#> 2 -2.385091 -2.248401 -0.05412126 -0.50095360 0.3638846 95%
#> 3 -2.374171 -2.258137  0.02712015 -0.34603889 0.3755087 95%
#> 4 -2.365972 -2.261774  0.10428175 -0.20067569 0.3884338 95%
#> 5 -2.356117 -2.256532  0.17736354 -0.06926482 0.4055157 95%
#> 6 -2.344780 -2.245924  0.24636553  0.05057027 0.4283991 95%
```

4.B.1.16 Noisy input spline in space time ("ni_spline_st")

The Noisy input spline in space time examines changes in RSL over multiple locations and throughout time. For this model, a minimum of 2 proxy sites should be used and tide gauge data provides insight into recent changes in RSL if the user is requires_ As an example, we will filter the example dataset `NAACproxydata` to select two sites to demonstrate the process:

```
# For 2 site
multi_site <-reslr::NAACproxydata %>%
  dplyr::filter(Site %in% c("Cedar Island", "Nassau"))
```

Step 1: Load in the data using the `reslr_load` function:

```
multi_site_input <- reslr_load(
  data = multi_site,
  include_tide_gauge = FALSE,
  include_linear_rate = FALSE,
  TG_minimum_dist_proxy = FALSE,
  list_preferred_TGs = NULL,
  all_TG_1deg = FALSE,
  prediction_grid_res = 50,
  sediment_average_TG = 10)
```

In this function, the user can select to add tide gauge data and estimates for linear_rate, by changing `include_tide_gauge = TRUE` and `include_linear_rate = TRUE` respectfully. If `include_tide_gauge = TRUE` the user must decide if they require the closest tide gauge i.e. `TG_minimum_dist_proxy = TRUE`, or select specific tide gauge i.e. `list_preferred_TGs = c("ARGENTIA")`, or all tide gauges within 1 degree of the proxy site i.e. `all_TG_1deg = TRUE`. The default setting is `rolling_window_average = 10` which corresponds to sediment accumulation rates of the proxy records, yet the user has the ability to alter this sediment accumulation rate.

Note that for a spline in space time, we recommend using the default settings as demonstrated in the above code chunk or investigating the resulting plots if additional tide gauge data could provide insight into recent changes. The user can alter the resolution of the output plots using `prediction_grid_res` with the default set at 50 years. The output of this function is a list

of two dataframes called `data` and `data_grid`. The `data` dataframe is the inputted data with additional columns for the linear_rate, linear_rate_err and data_type_id which will contain two options, "ProxyRecord" or "TideGaugeData". It can be accessed by:

```
data <- multi_site_input$data
head(data)
#>    Region   Site Latitude Longitude   RSL  Age Age_err RSL_err
#> 1 Florida Nassau   30.587   -81.666  0.05 2002    4.25    0.07
#> 2 Florida Nassau   30.587   -81.666  0.03 1990    5.50    0.07
#> 3 Florida Nassau   30.587   -81.666  0.01 1980    4.25    0.07
#> 4 Florida Nassau   30.587   -81.666 -0.01 1974    4.50    0.07
#> 5 Florida Nassau   30.587   -81.666 -0.03 1964    9.50    0.07
#> 6 Florida Nassau   30.587   -81.666 -0.05 1936   10.75    0.07
#>            SiteName data_type_id
#> 1 Nassau,\n Florida  ProxyRecord
#> 2 Nassau,\n Florida  ProxyRecord
#> 3 Nassau,\n Florida  ProxyRecord
#> 4 Nassau,\n Florida  ProxyRecord
#> 5 Nassau,\n Florida  ProxyRecord
#> 6 Nassau,\n Florida  ProxyRecord
```

- The `data_grid` is a dataframe that is evenly spaced in time based on the `prediction_grid_res` value chosen by the user and is used to create the plots and is accessed by

```
data_grid<- multi_site_input$data_grid
head(data_grid)
#> A tibble: 6 x 5
#>   Longitude Latitude SiteName                          data_type_id   Age
#>       <dbl>    <dbl> <fct>                             <fct>        <dbl>
#> 1     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -800
#> 2     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -750
#> 3     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -700
#> 4     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -650
#> 5     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -600
#> 6     -76.4     35.0 "Cedar Island,\n North Carolina" ProxyRecord   -550
```

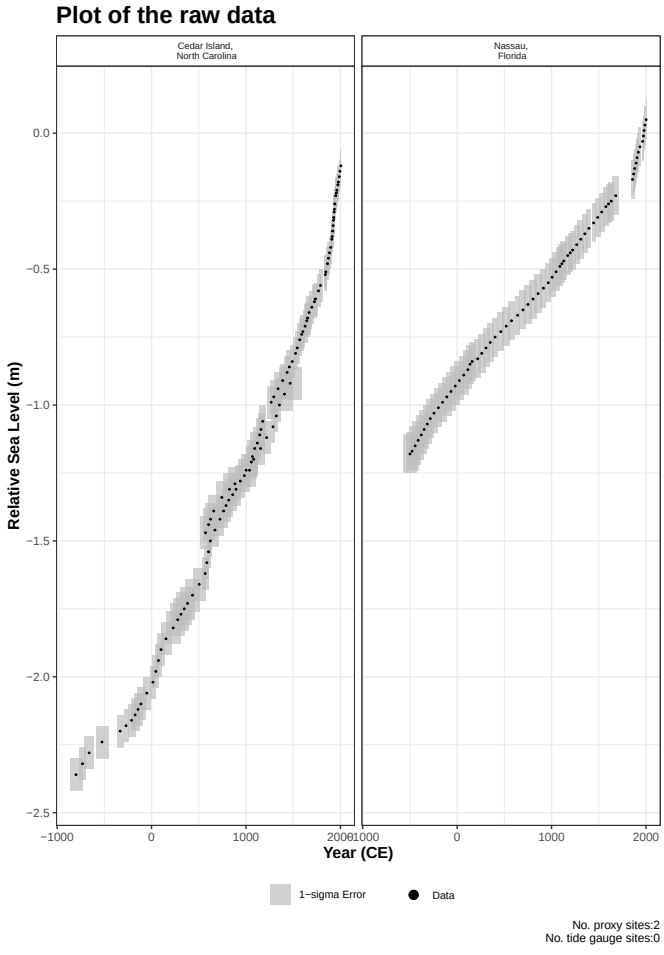
Step 1a: A brief insight into the outputs of the `reslr_input` function can be obtained using:

```
print(multi_site_input)
#> This is a valid reslr input object with 169 observations and  2 site(s).
#> There are  2  proxy site(s) and  0  tide gauge site(s).
```

```
#> The age units are; Common Era.
#> Decadally averaged tide gauge data was not included. It is recommended for the ni_gam_decomp model
#> The linear_rate or linear_rate_err was not included. It is required for the ni_gam_decomp model
```

Step 2: Plotting the data the raw data with:

```
plot(
  x = multi_site_input,
  title = "Plot of the raw data",
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)",
  plot_proxy_records =TRUE,
  plot_tide_gauges = FALSE)
```



This will produce a plot of Age on the x-axis and Relative Sea Level on the y-axis in meters. Grey boxes represent the uncertainty associated with the vertical and horizontal uncertainty. The black data points are the midpoints of these uncertainty boxes. The separate sites will appear in separate windows on the plot. The extra arguments can be used which allows the user to updated the titles and axis labels. The caption `plot_caption`, included by default, provides the number of proxy sites and tide gauge sites that will be used in the model and can be removed if required `plot_caption = FALSE`. In addition, the user can select to plot the additional tide gauge data, `plot_tide_gauge = TRUE`.

Step 3: Run the model for the two sites.

```
res_ni_spline_st <- reslr_mcmc(
  input_data = multi_site_input,
  model_type = "ni_spline_st",
  CI = 0.95)
```

The output object `jags_output.ni_spline_st` stores the JAGS model run. The `CI` setting allows the user to set the credible intervals, the current default is `CI = 0.95`. Note that additional computational run time is required for this model compared with the ni_spline_t.

Step 3a: A brief insight into the outputs of the `reslr_output` function can be obtained using:

```
print(res_ni_spline_st)
```

Step 4: The convergence of the algorithm is examined and he parameter estimates from the model can be investigated using the following:

```
summary(res_ni_spline_st)
```

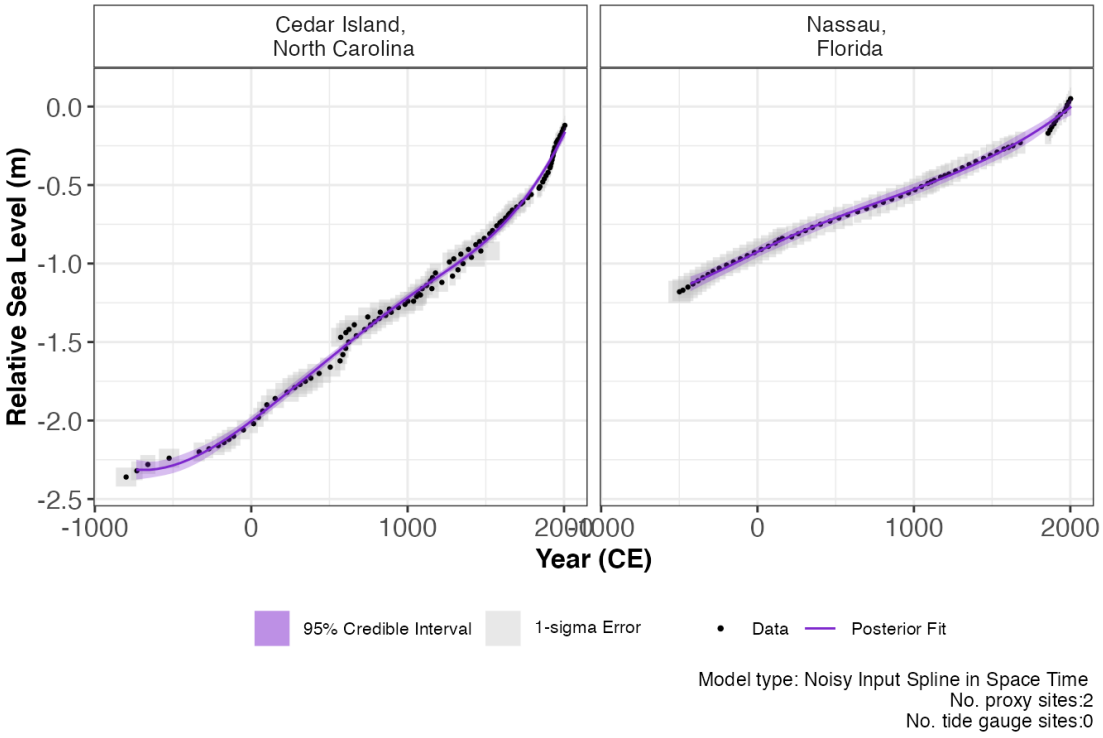
If the model run has the package will print: "No convergence issues detected". If the package prints: "Convergence issues detected, a longer run is necessary". The user is required to update the `reslr_mcmc` function with additional iterations in the following manner:

```
res_ni_spline_st <- reslr::reslr_mcmc(
  input_data = multi_site_input,
  model_type = "ni_spline_st",
  # Update these values
  n_iterations = 6000, # Number of iterations
  n_burnin = 1000, # Number of iterations to discard at the beginning
  n_thin = 4, # Reduces number of output samples to save memory and computation time
  n_chains = 3 # Number of Markov chains
)
```

For the parameter estimates, we can present the standard deviation associated with the NI spline space time model. Where "sigma_beta" highlights the variation associated with the spline coefficient of the spline in time and "sigma_y" presenting the overall variation.

Step 5: the results from the ni_spline_st model can be illustrated using:

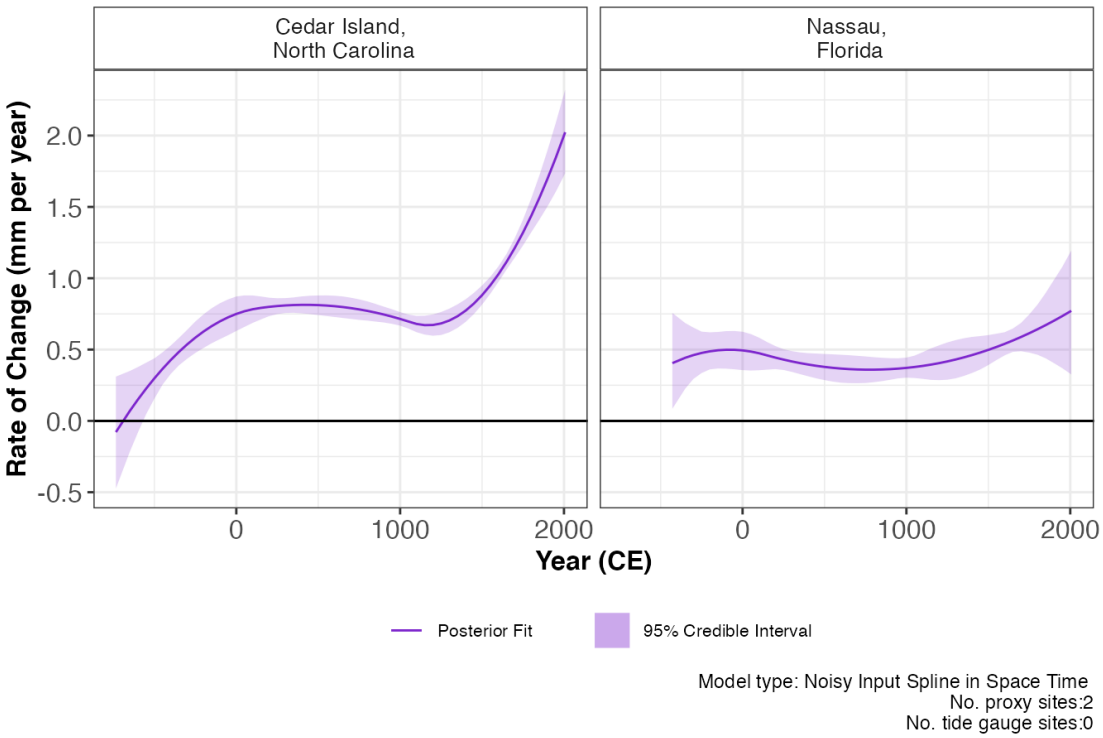
```
plot(res_ni_spline_st,
  plot_type = "model_fit_plot",
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)")
```



The output of this function is a graph of the input data, i.e. Age and RSL and associated uncertainty boxes, and the model fit with 95 % credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`. In order to view the rate of change plot, the following setting should be used:

```
plot(res_ni_spline_st,
  plot_type = "rate_plot",
```

```
xlab = "Year (CE)",
y_rate_lab = "Rate of Change (mm per year)")
```



This will print the plot of the rate of change with 95 % credible intervals. Again, the caption provides the model type, the number of proxy sites and tide gauge sites that were used.

To examine the data creating these plots the user types the following:

```
output_dataframes<- res_ni_spline_st$output_dataframes
```

4.B.1.17 Noisy Input Generalised Additive Model for decomposition of response signal ("ni_gam_decomp")

The Noisy Input Generalised Additive Model for the decomposition of the response signal (RSL). In the case of RSL, there are different drivers influence the changing RSL signal and these drivers vary in time and space. .The three main components of RSL change being examined using this

model type at a regional, local linear component and non-linear local component. A detailed description of this model can be found here.

There are a number of settings within the package that are important when using this model type. For the local linear component, GIA rate and associated uncertainty of the GIA rate must be provided prior to running. If the GIA rate is not provided for each location, then the `reslr` package will calculate it using the data and if this is not possible, the package will print an error message. Also, we recommend using tide gauge data averaged over a decade to match the accumulation rates of the proxy records, which is an additional argument in the function.

This model needs an adequate number of proxy sites to perform the decomposition and the minimum sites required will depend on the signal of the data. We found that in general we need a minimum of five proxy sites and at least five associated tide gauge sites. Also, we strongly recommend using tide gauge data for this model to obtain insight into recent changes in RSL. As an example, we will filter the example dataset `NAACproxydata` to select nine random sites to demonstrate the process:

```
# For 9 site
multi_9_sites <- reslr::NAACproxydata %>%
  dplyr::filter(Site %in% c(
    "Cedar Island", "Nassau", "Snipe Key",
    "Placentia", "Cape May Courthouse", "East River Marsh",
    "Fox Hill Marsh", "Swan Key", "Big River Marsh"))
```

Step 1: Load in the data using the `reslr_load` function:

```
multi_9_sites_input <- reslr_load(
  data = multi_9_sites,
  include_tide_gauge = TRUE,
  include_linear_rate = TRUE,
  TG_minimum_dist_proxy = FALSE,
  list_preferred_TGs = NULL,
  all_TG_1deg = TRUE,
  prediction_grid_res = 50,
  sediment_average_TG = 10)
```

In this function, the user can select to add tide gauge data and estimates for linear_rate, by changing `include_tide_gauge = TRUE` and `include_linear_rate = TRUE` respectfully. If `include_tide_gauge = TRUE` the user must decide if they require the closest tide gauge i.e. `TG_minimum_dist_proxy = TRUE`, or select specific tide gauge i.e. `list_preferred_TGs = c("ARGENTIA")`, or all tide gauges within 1 degree of the proxy site i.e. `all_TG_1deg = TRUE`. In this example, we use all tide gauges within 1 degree of the proxy site. The default setting is `sediment_average_TG`

`= 10` which corresponds to sediment accumulation rates of the proxy records, yet the user has the ability to alter this sediment accumulation rate.

Note that for this model, we recommend using the default settings as demonstrated in the above code chunk. If the user has not provided the linear rate and the associated linear rate uncertainty within the `linear_rate` and `linear_rate_err` column prior to running the package, the package to calculate it using the data.

The output of this function is a list of two dataframes called `data` and `data_grid`. The `data` dataframe is the inputted data with additional columns for the linear_rate, linear_rate_err and data_type_id which will contain two options, "ProxyRecord" or "TideGaugeData". It can be accessed by:

```
data <- multi_9_sites_input$data
```

- The `data_grid` is a dataframe that is evenly spaced in time based on the `prediction_grid_res` value chosen by the user and is used to create the plots and is accessed by

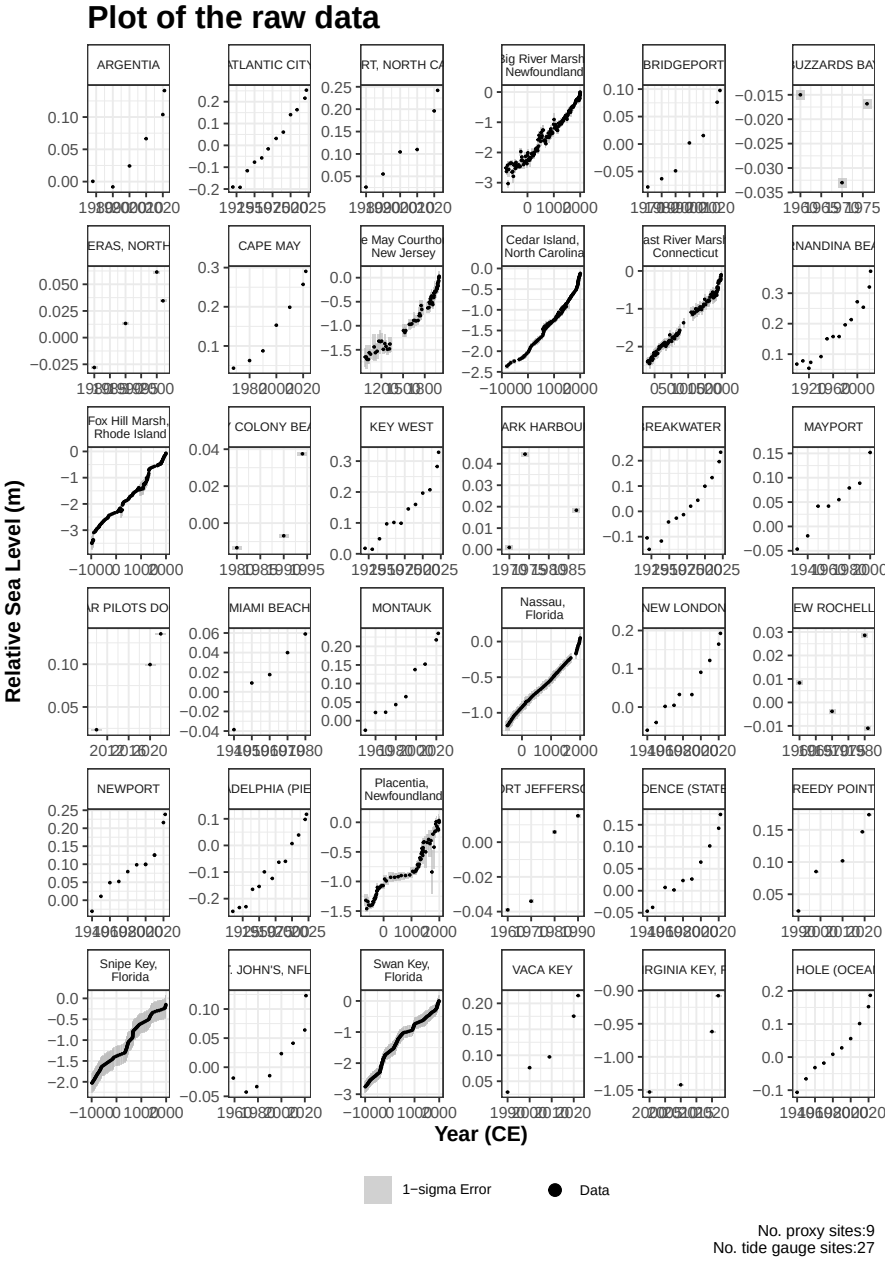
```
data_grid <- multi_9_sites_input$data_grid
```

Step 1a: A brief insight into the outputs of the `reslr_input` function can be obtained using:

```
print(multi_9_sites_input)
#> This is a valid reslr input object with 1124 observations and  36 site(s).
#> There are  9  proxy site(s) and  27  tide gauge site(s).
#> The age units are; Common Era.
#> Decadally averaged tide gauge data included by the package.
#> The linear_rate and linear_rate_err has been included.
```

Step 2: Plotting the data the raw data with:

```
plot(
  x = multi_9_sites_input,
  title = "Plot of the raw data",
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)",
  plot_proxy_records =TRUE},
  plot_tide_gauges =TRUE)
```



This will produce a plot of Age on the x-axis and Relative Sea Level on the y-axis in meters. Grey boxes represent the uncertainty associated with the vertical and horizontal uncertainty. The black data points are the midpoints of these uncertainty boxes. The separate sites will appear in separate windows on the plot. The extra arguments can be used which allows the user to updated the titles and axis labels. The caption `plot_caption`, included by default, provides the number of proxy sites and tide gauge sites that will be used in the model and can be removed

if required `plot_caption = FALSE`. In addition, the user can select to plot the additional tide gauge data, `plot_tide_gauge = TRUE`.

Step 3: Run the model

```
res_ni_gam_decomp <- reslr_mcmc(
  input_data = multi_9_sites_input,
  model_type = "ni_gam_decomp",
  CI = 0.95)
```

The output object `res_ni_gam_decomp` stores the JAGS model run. The `CI` setting allows the user to set the credible intervals, the current default is `CI = 0.95`. Note that there will be two model runs printed in the console here but the output will be the same format as earlier models.

Step 3a: A brief insight into the outputs of the `reslr_output` function can be obtained using:

```
print(res_ni_gam_decomp)
```

Step 4: The convergence of the algorithm is examined and he parameter estimates from the model can be investigated using the following:

```
summary(res_ni_gam_decomp)
```

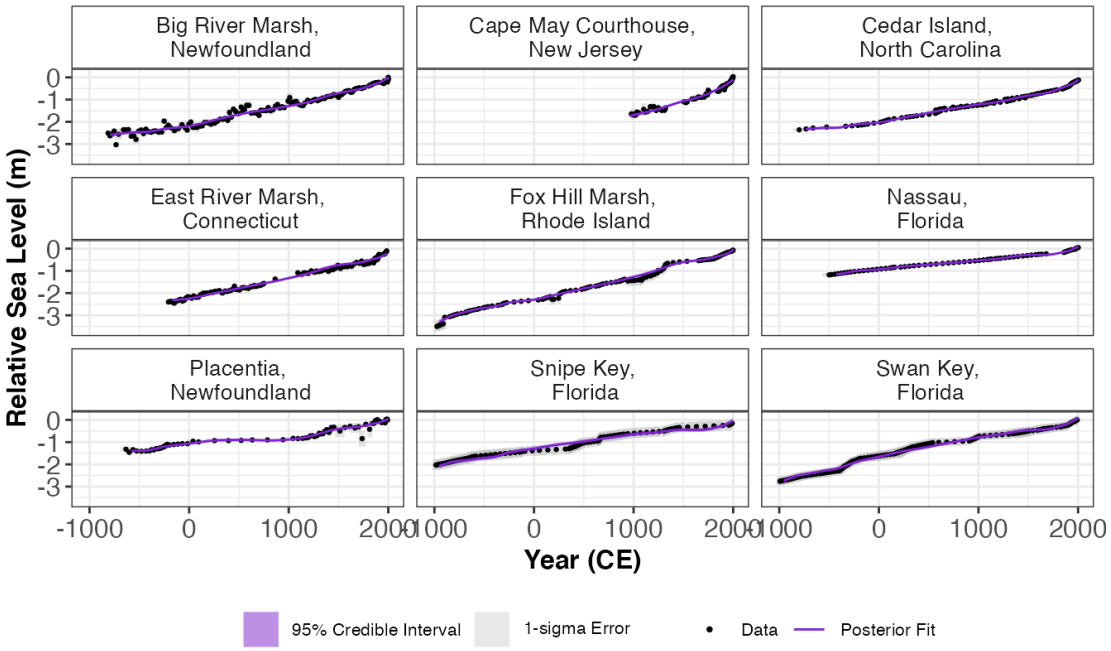
If the model run has the package will print: "No convergence issues detected". If the package prints: "Convergence issues detected, a longer run is necessary". The user is required to update the `reslr_mcmc` function with additional iterations in the following manner:

```
res_ni_gam_decomp <- reslr_mcmc(
  input_data = multi_9_sites_input,
  model_type = "ni_gam_decomp",
  # Update these values
  n_iterations = 6000, # Number of iterations
  n_burnin = 1000, # Number of iterations to discard at the beginning
  n_thin = 4, # Reduces number of output samples to save memory and computation time
  n_chains = 3 # Number of Markov chains
)
```

For the parameter estimates, we can present the standard deviation associated with each component of the NIGAM decomposition. This gives an insight into the variation caused by the different components with "sigma_r" representing the regional component, "sigma_l" highlighting the non-linear local component, "sigma_y" presenting the overall variation and "sigma_h" representing the site specific vertical offset.

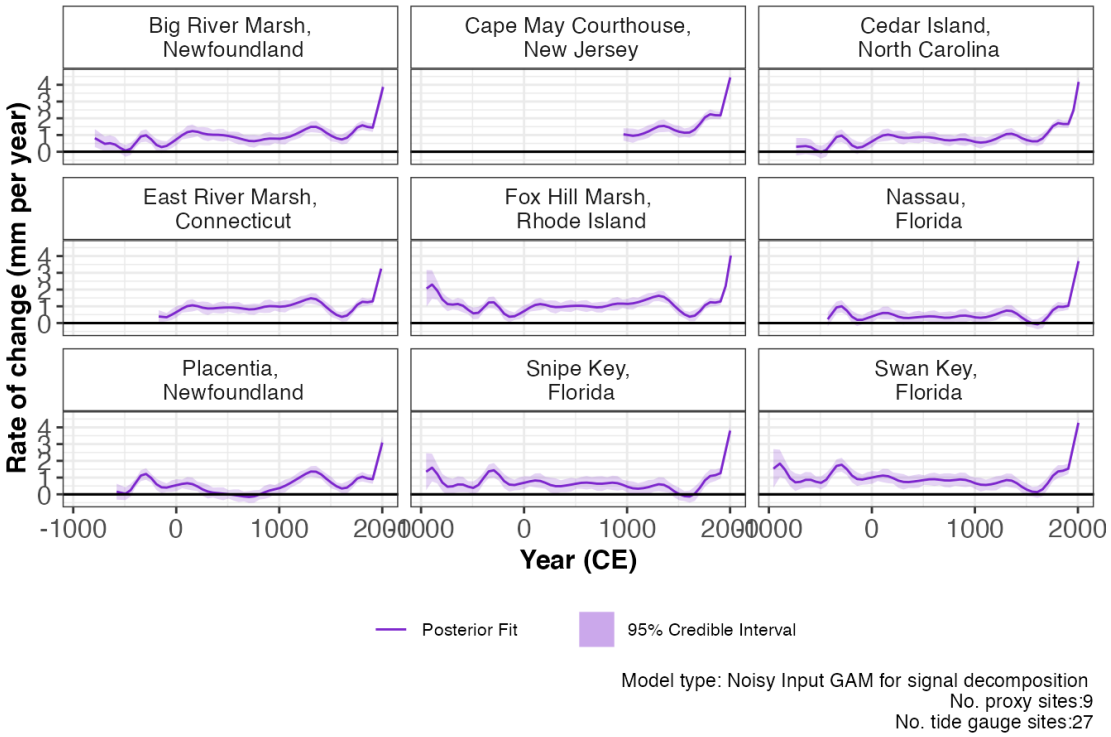
Step 5: The results from the `ni_gam_decomp` model can be illustrated with the option of excluding the tide gauge using:

```
plot(res_ni_gam_decomp,
  plot_type = "model_fit_plot",
  plot_tide_gauge = FALSE)
```



In addition, the user can select to plot the additional tide gauge data, `plot_tide_gauge = TRUE` in the plot. The output of this function is a graph of the input data, i.e. Age and RSL and associated uncertainty boxes, and the model fit with 95 % credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`. In order to view the rate of change plot, the following setting should be used:

```
plot(res_ni_gam_decomp,
  plot_type = "rate_plot")
```

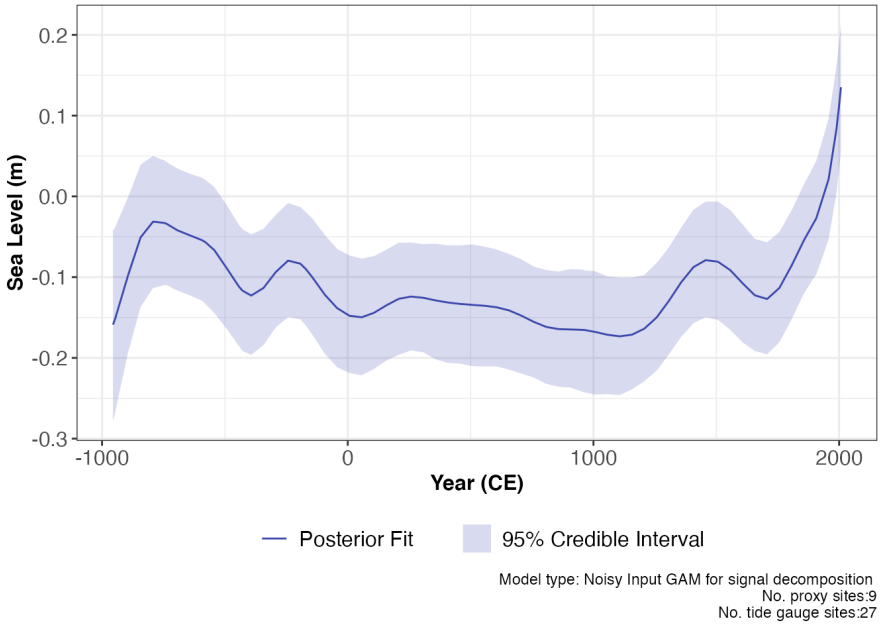


This will print the plot of the rate of change with 95 % credible intervals. The caption provides the model type, the number of proxy sites and tide gauge sites that were used. To examine the data creating the total model fit and the rate of change plot, the user can use:

```
total_model_fit_df <- res_ni_gam_decomp$output_dataframes$total_model_fit_df
```

There are separate settings to examine the plot of each component and its associated rate. To examine the regional component plot use:

```
plot(res_ni_gam_decomp, plot_type = "regional_plot")
```

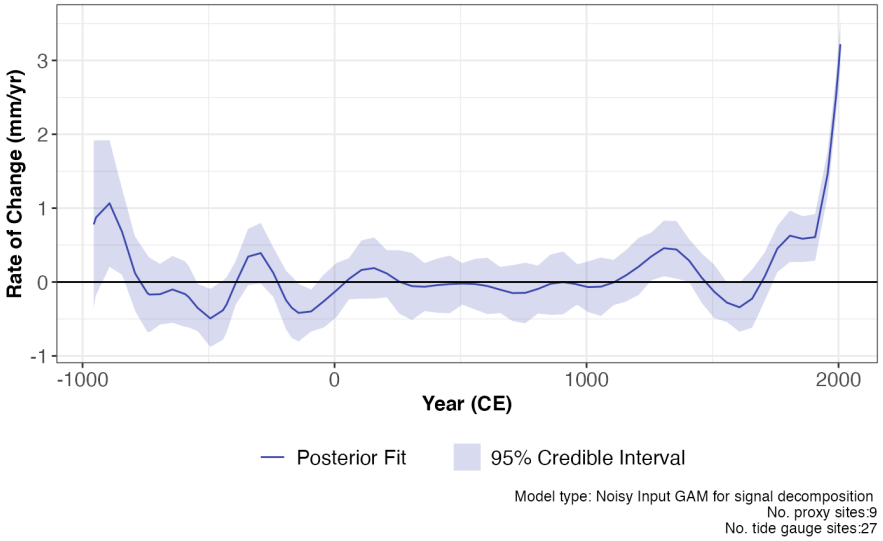


The regional component and the rate of change of the regional component is presented with 95% credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`. To examine the data creating the regional component plot and rate plot, the user can use:

```
regional_component_df <- res_ni_gam_decomp$output_dataframes$regional_component_df
```

The rate for the regional component can be accessed using:

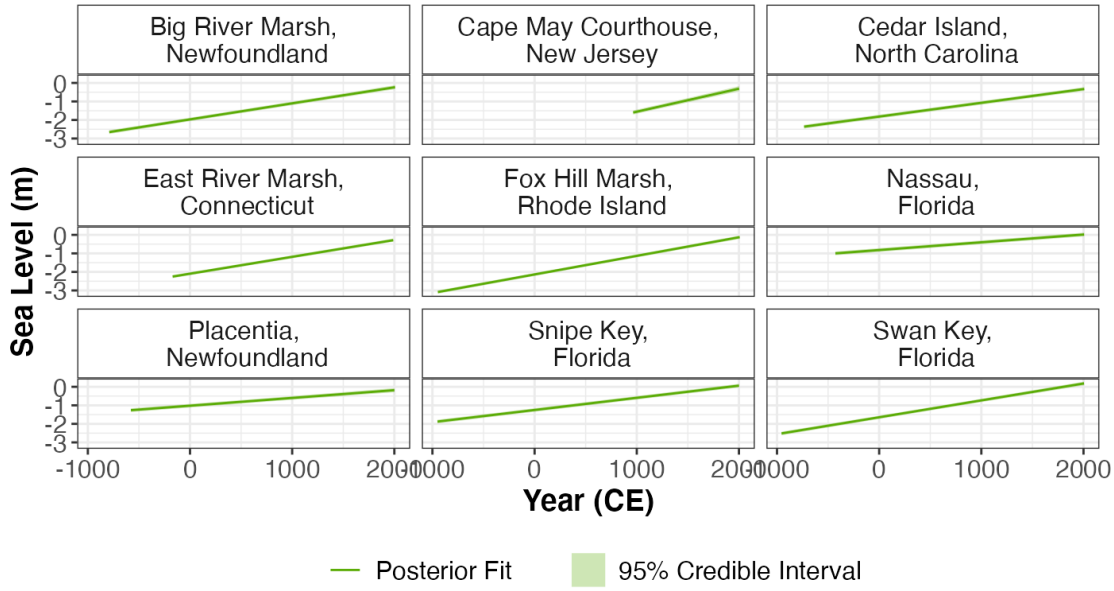
```
plot(res_ni_gam_decomp, plot_type = "regional_rate_plot")
```



Similarly, the rate of change of the regional component is presented with 95% credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`.

To examine the linear local component plot use:

```
plot(res_ni_gam_decomp, plot_type = "linear_local_plot")
```



The linear local component is plotted with 95% credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`. To examine the data creating the linear local component plot, the user can use:

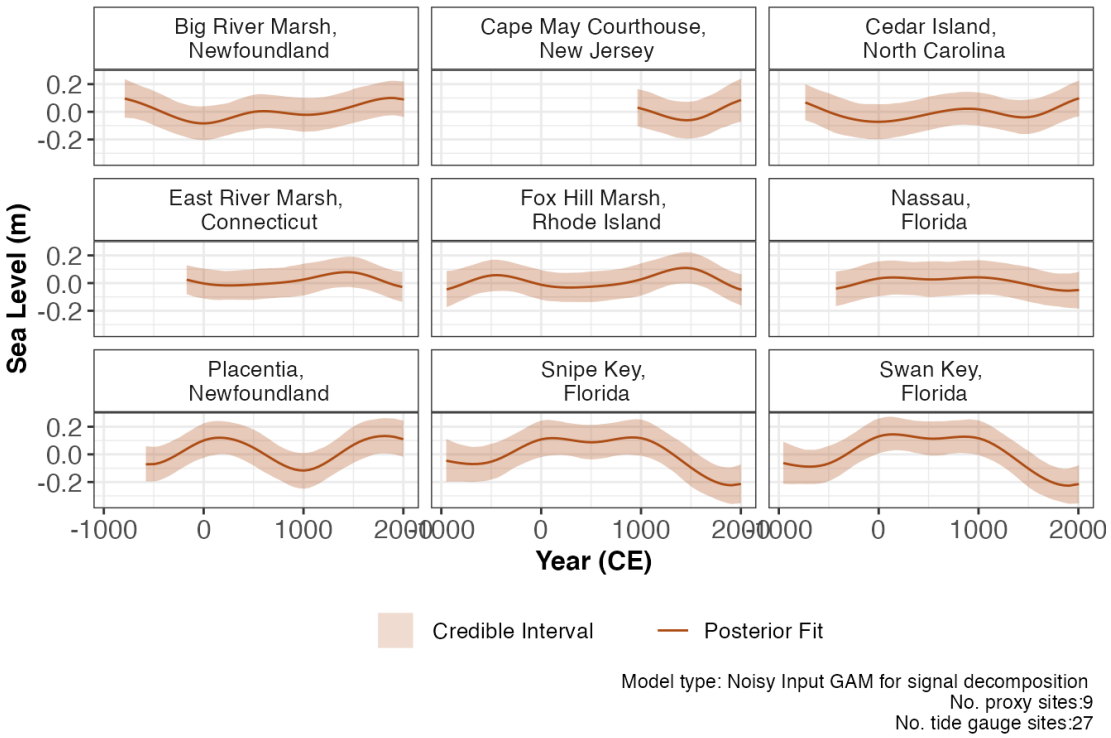
```
lin_loc_component_df <-res_ni_gam_decomp$output_dataframes$lin_loc_component_df
```

The associated linear local component rates for each location can be accessed by:

```
lin_loc_component_rates <- lin_loc_component_df %>%
  dplyr::group_by(SiteName) %>%
  dplyr::summarise(
    linear_rate = unique(linear_rate),
    linear_rate_err = unique(linear_rate_err))
```

To examine the non-linear local component plot use:

```
plot(res_ni_gam_decomp, plot_type = "non_linear_local_plot")
```

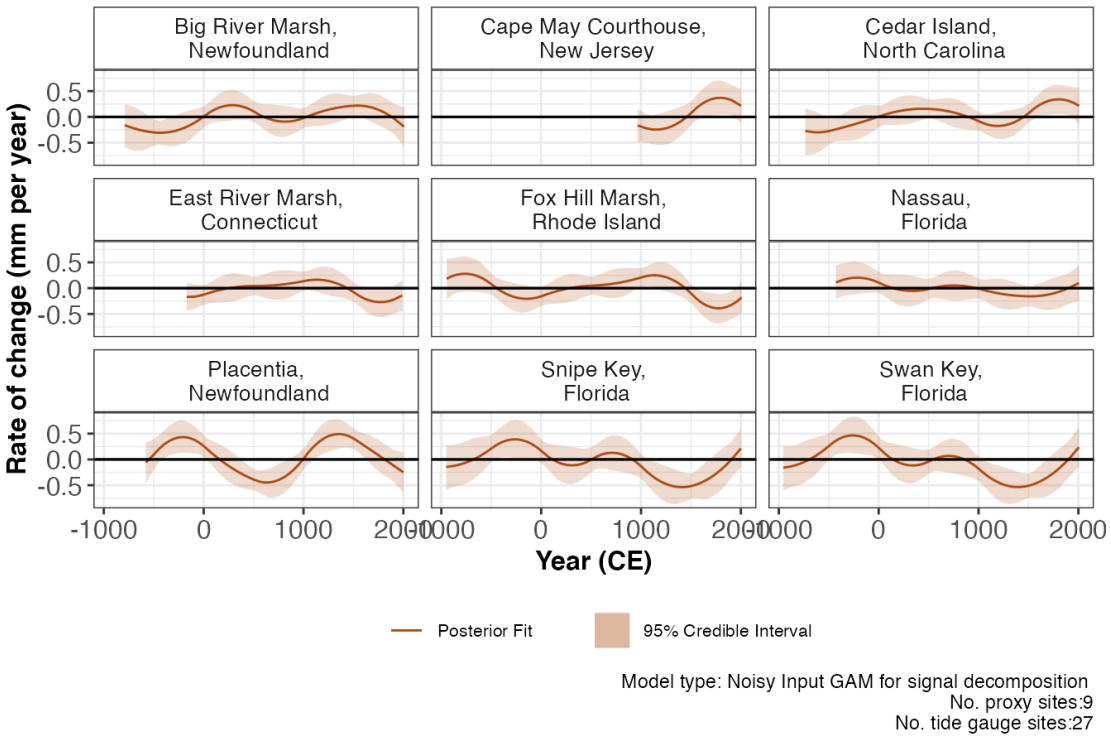


The non-linear local component is plotted with with 95% credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`. To examine the data creating the non-linear local component plot and rate plot, the user can use:

```
non_lin_loc_component_df <-res_ni_gam_decomp$output_dataframes$non_lin_loc_component_df
```

The plot of the rate of change for the non-linear local component use:

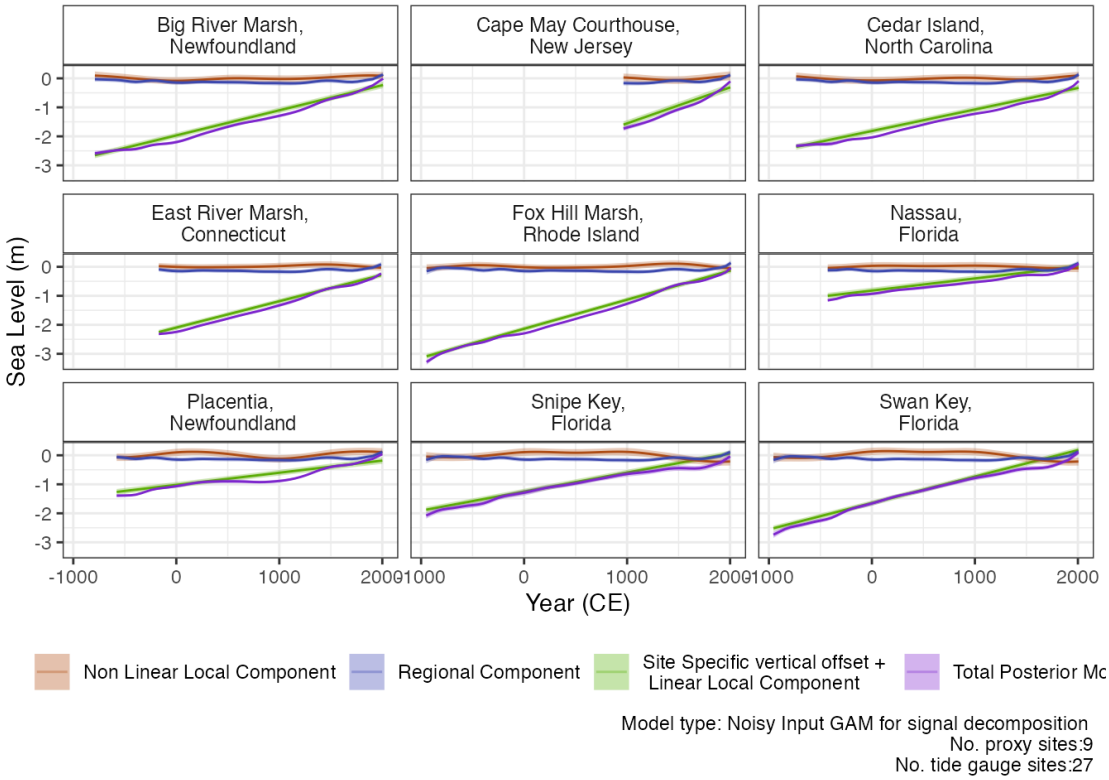
```
plot(res_ni_gam_decomp, plot_type = "non_linear_local_rate_plot")
```



The rate of change of the non-linear local component is plotted with with 95% credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`.

In order to examine how all components vary, the user can examine the plot using the following method:

```
plot(res_ni_gam_decomp, plot_type = "nigam_component_plot")
```



Each component is plotted with an 95% credible interval and this plot gives insight into the variability of the different components through time and at the different locations in question. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption =l FALSE`.

4.B.1.18 Appendix - suggested reading

For an introduction into statistical modelling for relative sea level change:

Upton, Maeve, Cahill, Niamh and Parnell, Andrew C. (2023), 'Statistical Modelling for Relative Sea-Level Data', Reference Module in Earth Systems and Environmental Sciences, Elsevier

For the maths on the original Change Point models:

Cahill, Niamh, Rahmstorf, Stefan and Parnell Andrew C. (2015), 'Change points of global temperature', Environmental Research Letters, 10(8), 084002

For the maths on the original EIV models:

Cahill, Niamh, Kemp, Andrew C , Horton, Benjamin P and Parnell, Andrew C (2015), 'Modeling

sea-level change using Errors-in-Variables integrated Gaussian Process 1', The Annals of Applied Statistics 9(2), 547–571

For the maths on the original NIGAM: Upton, Maeve, Parnell, Andrew C, Kemp, Andrew C , Ashe, Erica, McCarthy, Gerard and Cahill, Niamh (2023) 'A noisy-input generalised additive model for relative sea-level change along the Atlantic coast of North America'

For the background of GIA rates:

Whitehouse, Pippa L (2018), 'Glacial isostatic adjustment modelling: historical perspectives, recent advances, and future directions', Earth Surf. Dynam 6, 401–429.

Engelhart, Simon E., Benjamin P. Horton, Bruce C. Douglas, W. Richard Peltier and Torbj&oorn E. T́ornqvist (2009), 'Spatial variability of late Holocene and 20th century sea-level rise along the Atlantic coast of the United States', Geology 37(12), 1115–1118

Peltier, W.R (2004), 'Global Glacial Isostasy and the Surface of the Ice-Age Earth: The ICE-5G (VM2) Model and GRACE', Annual Review of Earth and Planetary Sciences 32, 111–149

For the background to tide gauge data:

Pugh, David, and Philip Woodworth. 2014. "Tidal Forces: Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes." In Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes, 36–59. Cambridge University Press.

Holgate, Simon J., Andrew Matthews, Philip L. Woodworth, Lesley J. Rickards, Mark E. Tamisiea, Elizabeth Bradshaw, Peter R. Foden, Kathleen M. Gordon, Svetlana Jevrejeva, and Jeff Pugh. 2013. "New Data Systems and Products at the Permanent Service for Mean Sea Level." Journal of Coastal Research 29 (3): 493–504.

Aarup, T., M. Merrifield, B. Pérez Gómez, I. Vassie, and P.Woodworth. 2006. "Manual on Sea-level Measurements and Interpretation, Volume IV : An update to 2006." Intergovernmental Oceanographic Commission of UNESCO 4. https://unesdoc.unesco.org/ark:/48223/pf0000147773

4.B.2 Quick Start Vignette for `reslr` package

reslr: quick start guide

Maeve Upton, Andrew Parnell & Niamh Cahill

4.B.2.1 Step 1: install `reslr`

Use:

```
install.packages("reslr")

library(reslr)
```

Note: The JAGS software is a requirement for this instruction sheet and refer back to main vignettes for more information.

4.B.2.2 Step 2: load in the data into `reslr`

There is a large example dataset included in the `reslr` package called `NAACproxydata`. In this example, we demonstrate how to include proxy record data which is stored in a csv file. This csv file of data can be found in the package and the `readr` function reads the csv file:

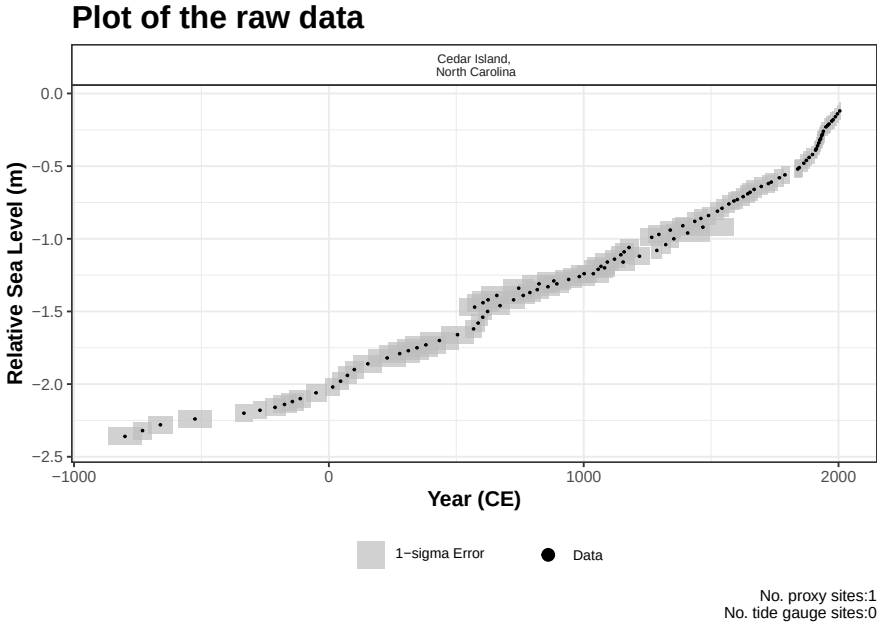
```
path_to_data <- system.file("extdata", "one_data_site_ex.csv", package = "reslr")
example_one_datasite <- read.csv(path_to_data)
```

Using the `reslr_load` function to read in the data into the `reslr` package:

```
example_one_site_input <- reslr_load(
  data = example_one_datasite)
```

4.B.2.3 Step 3: plot the data

```
plot(
  x = example_one_site_input,
  title = "Plot of the raw data",
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)",
  plot_tide_gauges = FALSE,
  plot_caption = TRUE)
```



4.B.2.4 Step 4: Run your statistical model and check convergence

Select your modelling technique from the modelling options available:

Statistical Model	Model Information	`model_type` code
Errors in variables simple linear regression	A straight line of best fit taking account of any age and measurement errors in the RSL values using the method of Cahill et al (2015). Use for single proxy site.	**"eiv_slr_t"**
Errors in variables change point model	An extension of the linear regression modelling process. It uses piece-wise linear sections and estimates where/when trend changes occur in the data (Cahill et al.2015).	**"eiv_cp_t"**
Errors in variables integrated Gaussian Process	A non linear fit that utilities a Gaussian process prior on the rate of sea-level change that is then integrated (Cahill et al.2015).	**"eiv_igp_t"**
Noisy Input spline in time	A non-linear fit using regression splines using the method of Upton et al (2023).	**"ni_spline_t"**
Noisy Input spline in space and time	A non-linear fit for a set of sites across a region using the method of Upton et al (2023).	**"ni_spline_st"**
Noisy Input Generalised Additive model for the decomposition of the RSL signal	A non-linear fit for a set of sites across a region and provides a decomposition of the signal into regional, local-linear (commonly GIA) and local non-linear components. Again this full model is as described in Upton et al (2023).	**"ni_gam_decomp"**

For this example, it is a single site and we are interested in how it varies over time select the Noisy Input spline in time. If it was multiple sites, we recommend using a spatial temporal model, i.e.Noisy Input spline in space and time, or for decomposing the signal, i.e. Noisy Input Generalised Additive model.

Once the model is chosen use the `reslr_mcmc` function to run it:

```
res_one_site_example <- reslr_mcmc(
  input_data = example_one_site_input,
  model_type = "ni_spline_t",
  CI = 0.95
)
```

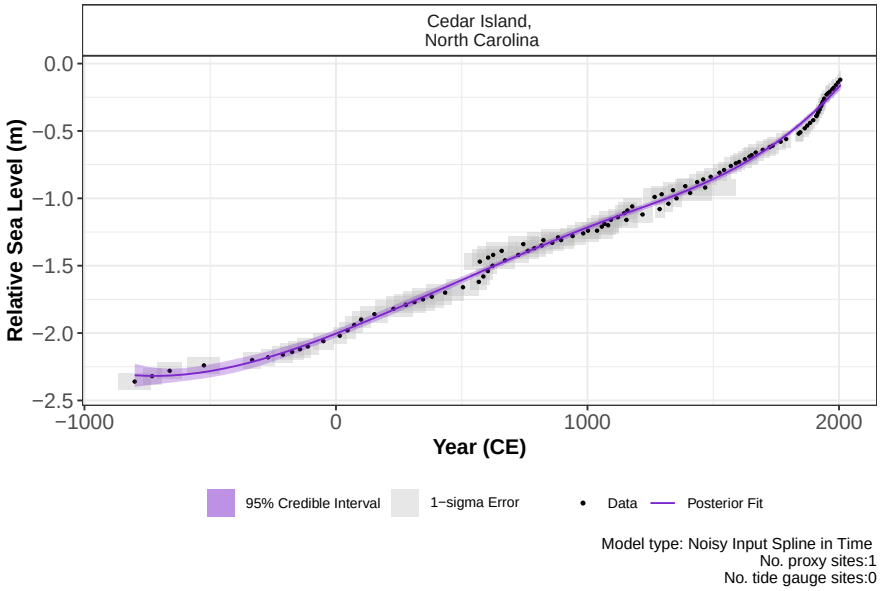
The convergence of the algorithm is examined and he parameter estimates from the model can be investigated using the following:

```
summary(res_one_site_example)
#> No convergence issues detected.
#> \# A tibble: 2 x 7
#>   variable      mean      sd     mad       q5    q95  rhat
#>   <chr>        <num>   <num>   <num>    <num>  <num> <num>
#> 1 sigma_beta 2.11    0.681   0.557   1.31     3.40    1.00
#> 2 sigma_y    0.00621 0.00465 0.00457 0.000478 0.0151  1.00
```

4.B.2.5 Step 5: Plot the results

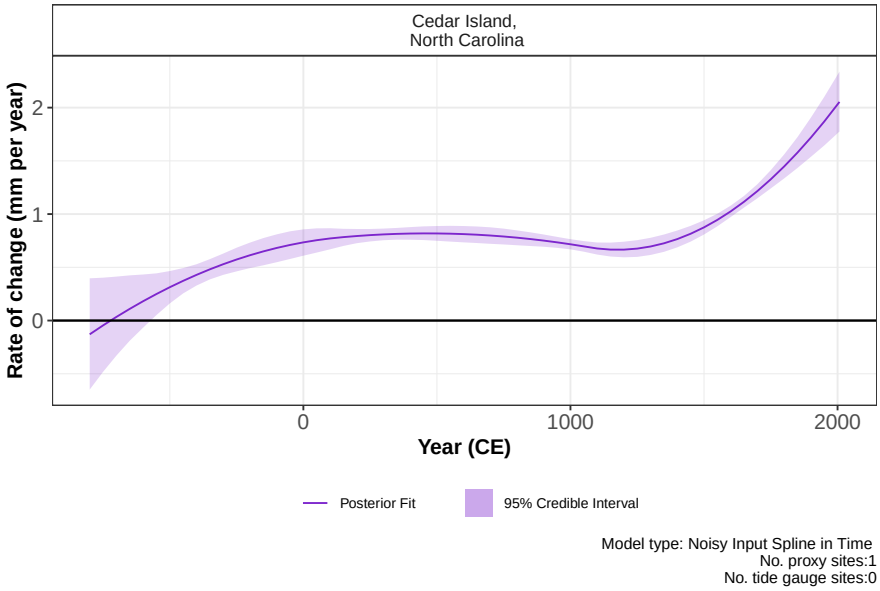
The model fit results can be visualised using the following function:

```
plot(res_one_site_example,
  xlab = "Year (CE)",
  ylab = "Relative Sea Level (m)",
  plot_type = "model_fit_plot")
```



For the rate of change plot use:

```
plot(res_one_site_example,
  plot_type = "rate_plot")
```



To examine the data creating these plots the user types the following:

```
output_dataframes <- res_one_site_example$output_dataframes
head(output_dataframes)
#>   Longitude Latitude                       SiteName data_type_id  Age      pred
#> 1    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -800 -2.313305
#> 2    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -750 -2.317669
#> 3    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -700 -2.317949
#> 4    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -650 -2.314343
#> 5    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -600 -2.307049
#> 6    -76.38   34.971 Cedar Island,\n North Carolina  ProxyRecord -550 -2.296264
#>         upr       lwr   rate_pred    rate_upr  rate_lwr  CI
#> 1 -2.400030 -2.227257 -0.12943764 -0.64875213 0.3968221 95%
#> 2 -2.387780 -2.249597 -0.04578351 -0.48234828 0.4027361 95%
#> 3 -2.376887 -2.260269  0.03391535 -0.32931190 0.4128568 95%
#> 4 -2.366038 -2.264015  0.10965871 -0.18955952 0.4243838 95%
#> 5 -2.354508 -2.259395  0.18144659 -0.06696884 0.4325678 95%
#> 6 -2.343160 -2.246958  0.24927898  0.05161619 0.4432910 95%
```

To examine the additional options in the `reslr` package, see the main vignette.

4.B.3 Advanced Vignette for `reslr` package

reslr: advanced

Maeve Upton, Andrew Parnell & Niamh Cahill

4.B.3.1 Introduction

In this document we present a range of advanced options that are available in the reslr package. To examine the all other options in the `reslr` package, see the main vignette.

If you have created other examples demonstrating `reslr` capabilities yourself, please contact me so I can include them here for other people to see.

Installing `reslr` and loading it with:

```
install.packages("reslr")
library(reslr)
```

4.B.3.2 Errors in Variables Integrated Gaussian Process with detrended data

The EIV Integrated Gaussian Process model provides the underlying rate of the process directly from the model. If the user requires a model applied to data that has been de-trended this is available in our package for a single site using the `model_type = "eiv_igp_t"`. The data is de-trended using the linear rate which can be provided by the user within the input dataframe along with the uncertainty associated with the rate. Or the linear rate is calculated within the package using the data and this rate is used to de-trend the data. In addition, the user will need to provide the estimated year of core collection.

This technique focuses on 1 site and we do not recommended for multiple proxy sites together. Tide gauge data can be included, however, the user must investigate which tide gauge is suitable and the tide gauge data will use the same linear_rate as the proxy records. As an example, we will filter the example dataset `NAACproxydata` to select one site to demonstrate the process:

```
# For 1 site
 CedarIslandNC<- NAACproxydata[NAACproxydata$ Site == "Cedar Island" ,]
```

Step 1: Load in the data using the `reslr_load` function:

```
 CedarIslandNC_input_detrend<-reslr_load(
 data =  CedarIslandNC,
 include_tide_gauge = FALSE,
 include_linear_rate = TRUE,
 TG_minimum_dist_proxy = FALSE,
 list_preferred_TGs = NULL,
 all_TG_1deg = FALSE,
 prediction_grid_res = 50,
 sediment_average_TG = 10,
 detrend_data = TRUE,
 core_col_year = 2010)
```

For this case, we update the setting `detrend_data =TRUE` and we provide the `core_col_year = 2010` corresponding to the year of the core collection. The linear_rate is required and is included by setting `include_linear_rate = TRUE`.

The output of this function is a list of two dataframes called `data` and `data_grid`. - The `data` dataframe is the inputted data with additional columns for the data_type_id which will contain "ProxyRecord". It can be accessed by:

```
data<- CedarIslandNC_input_detrend$data
```

- The `data_grid` is a dataframe that is evenly spaced in time based on the `prediction_grid_res` value chosen by the user and is used to create the plots and is accessed by

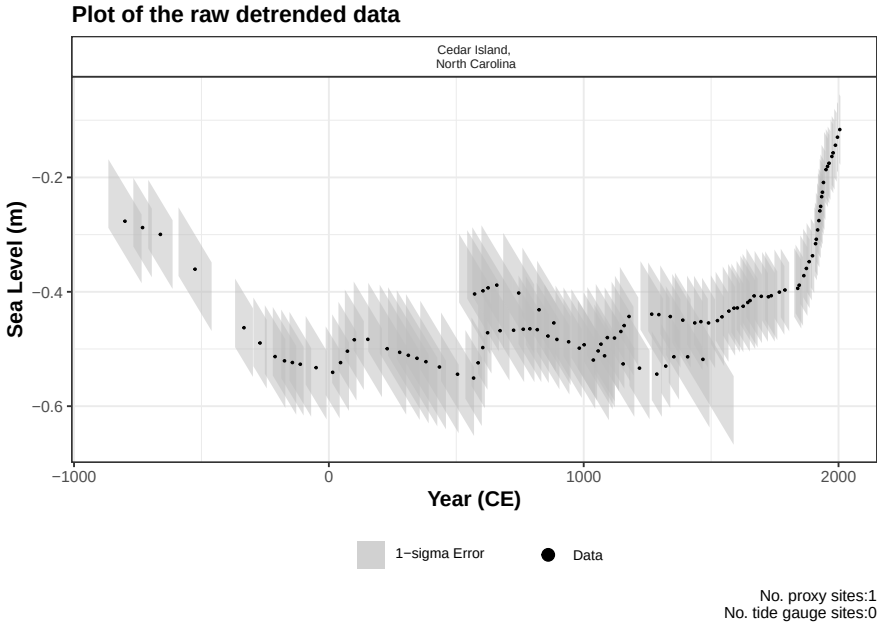
```
data_grid<- CedarIslandNC_input_detrend$data_grid
```

Step 1a: A brief insight into the outputs of the `reslr_input` function can be obtained using:

```
print(CedarIslandNC_input_detrend)
#> This is a valid reslr input object with 104 observations and  1 site(s).
#> There are  1  proxy site(s) and  0  tide gauge site(s).
#> The age units are; Common Era.
#> Decadally averaged tide gauge data was not included. It is recommended for the ni_gam_decomp model
#> The linear_rate and linear_rate_err has been included.
#> Data has been detrended.
```

Step 2: Plotting the data the raw data with:

```
plot(
 x =  CedarIslandNC_input_detrend,
 title = "Plot of the raw detrended data" ,
 xlab = "Year (CE)" ,
 ylab = "Sea Level (m)",
 plot_proxy_records = TRUE,
 plot_tide_gauges = FALSE)
```



This will produce a plot of Age on the x-axis and Sea Level on the y-axis in meters. Grey boxes represent the uncertainty associated with the vertical and horizontal uncertainty. The black data points are the midpoints of these uncertainty boxes. The extra arguments can be used which allows the user to updated the titles and axis labels. The caption `plot_caption`, included by default, provides the number of proxy sites and tide gauge sites that will be used in the model and can be removed if required `plot_caption = FALSE`. In addition, the user can select to plot the additional tide gauge data, `plot_tide_gauge = TRUE`.

Step 3: To run the eiv_igp_t model the following function should be used:

```
res_eiv_igp_t_detrend<-reslr_mcmc(
input_data = CedarIslandNC_input_detrend,
model_type = "eiv_igp_t",
CI = 0.95)
```

This command takes the input data and the user specifies the statistical model, i.e. an integrated Gaussian process using the EIV uncertainty method ("eiv_igp_t"). It tells `reslr` to store the output of the model run in an object called `res_eiv_igp_t`. The `CI` setting allows the user to set the credible intervals, the current default is `CI = 0.95`.

Step 3a: A brief insight into the outputs of the `reslr_output` function can be obtained using:

```
print(res_eiv_igp_t_detrend)
```

Step 4: The convergence of the algorithm is examined and he parameter estimates from the model can be investigated using the following:

```
summary(res_eiv_igp_t_detrend)
```

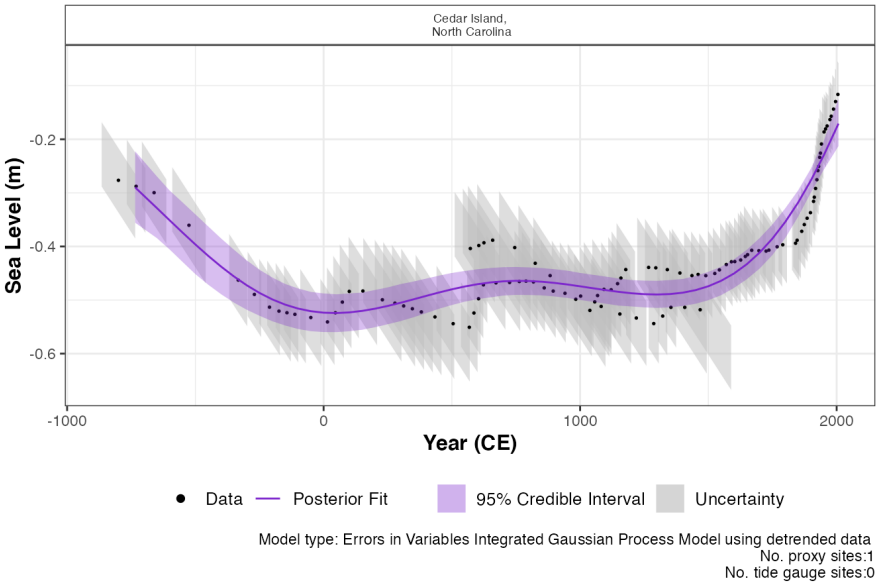
If the model run has the package will print: "No convergence issues detected". If the package prints: "Convergence issues detected, a longer run is necessary". The user is required to update the `reslr_mcmc` function with additional iterations in the following manner:

```
res_eiv_igp_t_detrend<-reslr_mcmc(
input_data =CedarIslandNC_input_detrend,
model_type ="eiv_igp_t",
 # Update these values
n_iterations = 6000,# Number of iterations
n_burnin =1000, # Number of iterations to discard at the beginning
n_thin = 4, # Reduces number of output samples to save memory and computation time
n_chains = 3# Number of Markov chains
)
```

For the parameter estimates, the length scale parameter, "rho" is the correlation parameter and "nu" is the standard deviation of the rate process. "sigma_y" is the variation of the model.

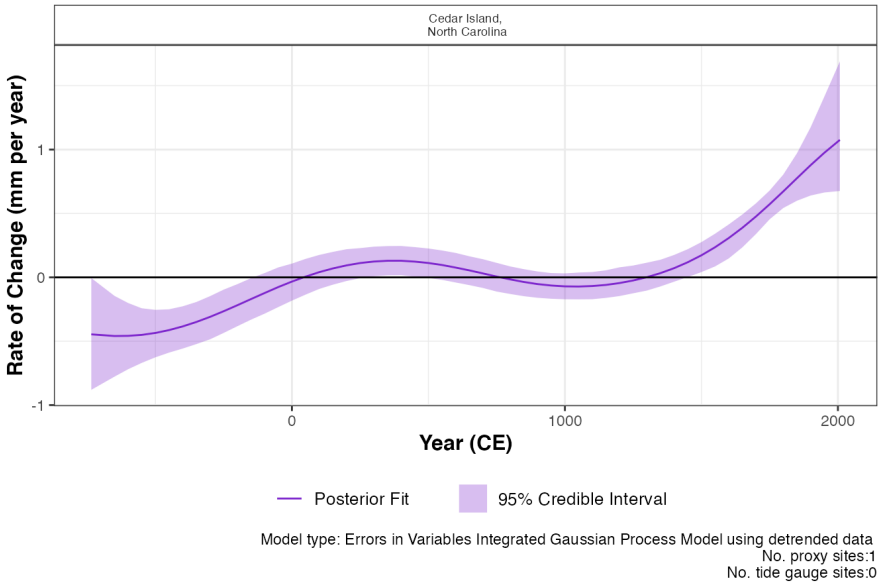
Step 5: The results from the EIV IGP model for the de-trended data can be illustrated using:

```
plot(res_eiv_igp_t_detrend,
 plot_type = "model_fit_plot",
 xlab = "Year (CE)",
 ylab = "Sea Level (m)",
 plot_proxy_records = TRUE,
 plot_tide_gauges = FALSE)
```



The output of this function is a graph of the input data, i.e. Age and SL and associated uncertainty boxes, and the model fit with 95 % credible interval. The caption provides the model type used and number of proxy sites and tide gauge sites used and can be removed if necessary with `plot_caption = FALSE`. In order to view the rate of change plot, the following setting should be used:

```
plot(res_eiv_igp_t_detrend,
 plot_type = "rate_plot" ,
 xlab = "Year (CE)" ,
 y_rate_lab = "Rate of Change (mm per year)")
```



This prints the plot of the rate of change with 95 % credible intervals. The caption provides the model type, the number of proxy sites and tide gauge sites that were used.

To examine the data creating these plots the user types the following:

```
output_dataframes<- res_eiv_igp_t_detrend$output_dataframes
```

4.B.3.3 Using a different input age unit

The default input age for the reslr package is in units of years in the Common Era. The user can select to input the age value in Before Present by updating the `input_age_type = "BP"` within the `reslr_load` function. Inside the package, the Age is converted into CE in order to run the models. The plots are then altered to account for the input age in BP and the x-axis is be reversed.

Load in the data using the `reslr_load` function:

```
CedarIslandNC_input_age_BP<-reslr_load(
data =  data_age_bp,
input_age_type = "BP")
```

The output of this function is a list of two dataframes called `data` and `data_grid`. Both dataframes will contain two new columns which identify the `Age_type = "BP"` and an additional column is included called `Age_BP` which contains the original input Age in units Before Present.If the user is recreating these plots, they are required to use the `Age_BP` column along the x-axis as it provides the original input age in units Before Present (BP) and the x-axis will need to be reversed.

4.B.3.4 Including Tide Gauge data

In this package, we have described three ways to select a tide gauge and in the example we showed how to include the closest tide gauge to the proxy site. Here we are demonstrating the other options available to the user.

```
# For 2 site
 multi_site<- NAACproxydata[NAACproxydata$ Site %in% c("Cedar Island", "Nassau"),]
```

If the user has a list of chosen tide gauges from PSMSL website, include them in a list in the `reslr_load` function as demonstrated:

```
 multi_site<-reslr_load(
 data =  multi_site,
 include_tide_gauge = TRUE,
 include_linear_rate = TRUE,
 TG_minimum_dist_proxy = FALSE,
  # There is no limit to the number of tide gauges provided in the list
 list_preferred_TGs = c(
    "ARGENTIA", "MAYPORT",
    "JACKSONVILLE", "LAKE WORTH PIER",
    "MAYPORT (BAR PILOTS DOCK), FLORIDA"
   ),
 all_TG_1deg = FALSE,
 prediction_grid_res = 50,
 sediment_average_TG = 10)
```

Here is how to plot this example:

```
plot(
 x = multi_site,
 title = "Plot of the raw data" ,
 xlab = "Year (CE)" ,
 ylab = "Relative Sea Level (m)",
 plot_tide_gauges = TRUE} ,
 plot_proxy_records = TRUE ,
 plot_caption = TRUE)
```

If the user requires all tide gauges within 1 degree of the proxy site, which is the equivalent of 111kms, this can be done by updating `all_TG_1deg=TRUE` in the following way:

```
 multi_site<-reslr_load(
 data =  multi_site,
 include_tide_gauge =TRUE,
 include_linear_rate = TRUE,
 TG_minimum_dist_proxy = FALSE,
 list_preferred_TGs = NULL,
 all_TG_1deg = TRUE ,
 prediction_grid_res = 50)
```

Here is how to plot this example:

```
plot(
 x = multi_site,
 title ="Plot of the raw data",
 xlab ="Year (CE)",
 ylab = "Relative Sea Level (m)",
 plot_tide_gauges =TRUE,
 plot_proxy_records = TRUE ,
 plot_caption =TRUE)
```

4.B.3.5 Plotting techniques

In the package, all plot labels for results, i.e. x and y labels and titles, can be updated in the following manner:

```
# Example
 final_plots<-plot(x =reslr_mcmc(CedarIslandNC, model_type =}"ni_spline_t"))
 final_plots$ plot_result
# Adding new title to the total model fit plot
 final_plots$plot_result+ggplot2::ggtitle("New Title Added as Example")
 final_plots$plot_result+ggplot2::xlab("New x axis label Added as Example")
 final_plots$plot_result+ggplot2::ylab("New y axis label Added as Example")
```

The user can access the dataframes creating these plots in order to recreate the output plots in their own style. To access the data frame used to create the raw plot and the output model fit plot use:

```
 raw_data<- CedarIslandNC_input_detrend$data
 output_data<- res_eiv_igp_t_detrend$output_dataframes
```

This dataframes contains the model fit under the column `pred` and the rate of change column is `rate_pred`.

For the `ni_gam_decomp` model, there are a separate dataframe for each component and the rate of change for each component has a corresponding dataframe. This ensures the decomposition of the RSL signal can be plotted with all four components together, using separate colours to identify the different drivers of RSL change.

4.B.3.6 Accessing the posterior samples

In the package the user has the ability access all the posterior sample for each unknown parameter in the following way:

```
# Example
 CedarIslandNC_input<-reslr_load(
 data = CedarIslandNC)
 res_eiv_slr_t<-
  reslr_mcmc(CedarIslandNC_input,
             model_type = "eiv_slr_t" )
# Accessing the slope of the EIV simple linear regression
 beta<- res_eiv_slr_t$noisy_model_run_output$BUGSoutput$sims.list$beta
```

4.B.3.7 Smoothing settings for splines

When constructing a spline piecewise polynomials join together at knots to form smooth curves. The number of knots for these models can be reduced to achieve a more smooth model fit or increased for more flexibility. In the `reslr` package, tests have been carried out to select the correct number of knots for the spline in time, spline in space time and the GAM. However, if the situation arise where the model fits require additional flexibility or require additional smoothing, the user can adjust these settings. To adjust the setting for the number of knots in the spline in time update the reslr function in the following way:

```
res_ni_sp_t<-
 reslr_mcmc(CedarIslandNC_input,
           model_type = "ni_spline_t",
           spline_nseg = NULL)
```

By updating the `spline_nseg` function, the number of knots we change. Note the minimum value for this setting is 2. The default for this setting is `NULL` as we use a rule of thumb calculation to identify the number of knots based on the range of observations. Similarly, for the spline in space time. The number of knots is updated using:

```
res_ni_sp_t<-
 reslr_mcmc(CedarIslandNC_input,
           model_type = "ni_spline_st" ,
           spline_nseg = NULL)
```

By updating the `spline_nseg` function, the number of knots we change. Note the minimum value for this setting is 2. The default for this setting is `NULL` as we use a rule of thumb calculation to identify the number of knots based on the range of observations.

When using the NI GAM approach, the number of knots for the separate components can be altered. To control the smoothness of the regional component (i.e. spline in time) the setting to vary is the `spline_nseg_t`. To vary the non-linear local component (i.e. spline in space time), the user can vary the `spline_nseg_st` setting. The following example will demonstrate these settings:

```
res_ni_sp_t<-
 reslr_mcmc(CedarIslandNC_input,
           model_type = "ni_gam_decomp",
           spline_nseg_t = 20,
           spline_nseg_st = 6)
```

The default settings have been tested using cross validation techniques to incorporate the require flexibility for both components.

4.B.3.8 Cross Validation tests for Spline and NI-GAM

In the previous section, we discussed altering the smoothness of the splines by varying the number of knots in the model. The user can test the validity of the model and the number of knots selected using this cross validation function with in the package. For example, the user can run different values for the knots and compare the model fits for each setting using the following cross validation function.

```
 data1site_example<- NAACproxydata[NAACproxydata$Site == "Cedar Island",]
# Cross Validation test
 cv<-cross_val_check(data = data1site_example,
                     model_type ="ni_spline_t",
                     n_iterations =1000,
                     n_burnin =100,
                     n_thin =5,
                     n_chains =2,
    # User the package to calculate the number of knots
                     spline_nseg =NULL,

     # n_fold allows the user to alter the cross validation, i.e. 3, 5, 10 fold
                     n_fold =3,
    #To reproducible results,seed stores the output of the random selection
                     seed =NULL,
    # Size of the credible intervals and prediction intervals
                     CI =0.95)
```

To alter the number of knots in the model, include the `spline_nseg` or `spline_nseg_t` or `spline_nseg_st` arguments in this function. The different meanings for these arguments are described in the section above. By altering the `CI` argument, the user can change the size of the prediction intervals.

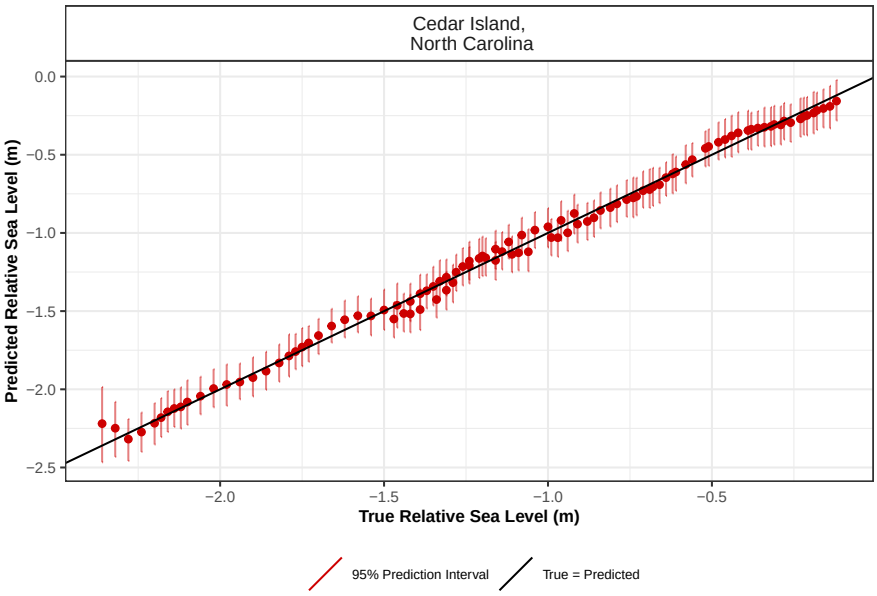
The `n_fold = 3` argument allows the user to change the number of folds used in the cross validation test. We recommend a minimum of 3 folds and the ideal number of folds is 5 or 10, however, using 10 folds will result in longer run times.

By updating the `model_type` argument, the user can undertake cross validation for the noisy input spline in time, the noisy input spline in space time or the noisy input generalised additive

model. This function will print a statement regarding model convergence. If the convergence issues are detected, the package will inform the user to increase the number of iterations in the `cross_val_check` function.

The outputs of the `cross_val_check` function are stored as a list. To examine the true versus predicted plot use the following command.

```
cv$true_pred_plot
```



This figure presents the true RSL value in metres versus the predicted RSL value in metres for each data site. The predicted means are the red dots and the vertical lines over each point are the prediction intervals.

To access the dataframe creating this plot use:

```
CV_model_df<- cv$CV_model_df
```

This dataframe contains the `true_RSL` values from the test set and the corresponding predicted values `pred_RSL`. In this dataframe, there are the prediction intervals `upr_PI` and `lwr_PI`. The

`SiteName` column identifies each site present in the dataframe and allows for site specific analysis of the model if required.

The model performance is examined using a variety of different test. The first is the out of sample empirical coverage. The empirical coverage provides the percentage of occasions that the true RSL observation is within the model prediction interval (PI) for RSL and is given by:

```
# Overall coverage
 total_empirical_coverage<- cv$total_coverage
 total_empirical_coverage
# Coverage by site
 coverage_by_site<- cv$coverage_by_site
 coverage_by_site
#> A tibble: 1 x 2}
#>   SiteName                          coverage_by_site
#>   <fct>                                        <dbl>
#> 1 "Cedar Island,\n North Carolina"                 1
# Size of the prediction intervals
 prediction_interval_size<- cv$prediction_interval_size
 prediction_interval_size
#> \# A tibble: 1 x 2
#>   SiteName                          PI_width
#>   <fct>                                <dbl>
#> 1 "Cedar Island,\n North Carolina"   -0.246
```

The prediction intervals are created using posterior predictive simulations with the full error structure, i.e. $\hat{y}_{ij} \sim N(\hat{f}_{ij}, \sigma^2_{y_{ij}} + \sigma^2_{t_{ij}} + \sigma^2)$ where $\sigma^2_{y_{ij}}$ is the observed uncertainty associated with the RSL and $\sigma^2_{t_{ij}}$ is calculated to account for the noisy input uncertainty associated with the input (time). The size of these prediction intervals for each site can be examined to understand how the coverage is calculated and using different prediction intervals, e.g. 95% versus 50%, the model validity can be further examined.

The remaining tools used to examine model performance within the `reslr` package are the Root Mean Squared Error (RMSE), mean error (ME) and mean absolute error (MAE). The RMSE provides insight into prediction performance in the same units as the response (meters). The RMSE can be examined by site or for each fold in the cross validation test. The overall mean error (ME) by finding the difference between the predicted observation and the true RSL observation. The mean absolute error (MAE) is calculated by taking the absolute value of the mean error, in turn, this provides the degree to which our model is biased. For each test, the recommendation is the lower the value the better the model fit and these values can be accessed by:

```
# Overall
 ME_MAE_RSME_overall<- cv$ME_MAE_RSME_overall
```

```
 ME_MAE_RSME_overall
#>         RSME        MAE           ME
#> 1 0.01269653 0.03528882 -0.001244997
# By fold and site
 ME_MAE_RSME_fold_site<- cv$ ME_MAE_RSME_fold_site
 ME_MAE_RSME_fold_site
#> A tibble: 3 x 5
#>   SiteName                         CV_fold_number   RSME    MAE      ME
#>   <fct>                            <fct>           <dbl>  <dbl>   <dbl>}
#> 1 "Cedar Island,\n North Carolina" 1              0.0634 0.0325 -0.0107
#> 2 "Cedar Island,\n North Carolina" 2              0.0643 0.0281 -0.0110
#> 3 "Cedar Island,\n North Carolina" 3              0.105  0.0451  0.0177
# By site
 ME_MAE_RSME_site<- cv$ME_MAE_RSME_site
 ME_MAE_RSME_site
#> A tibble: 1 x 4}
#>   SiteName                           RSME    MAE       ME
#>   <fct>                             <dbl>  <dbl>    <dbl>
#> 1 "Cedar Island,\n North Carolina" 0.0127 0.0353 -0.00124
# By fold
 ME_MAE_RSME_fold<- cv$ME_MAE_RSME_fold
 ME_MAE_RSME_fold
#> A tibble: 3 x 4
#>   CV_fold_number   RSME    MAE      ME
#>   <fct>           <dbl>  <dbl>   <dbl>
#> 1 1              0.0634 0.0325 -0.0107
#> 2 2              0.0643 0.0281 -0.0110
#> 3 3              0.105  0.0451  0.0177
```

For more information about the outputs to these cross validation tests for RSL data refer to Upton et al. 2023 paper.

CHAPTER 5

An investigation of ocean dynamics and relative sea-level change in the North Atlantic

Sea-level rise, in particular ocean dynamics, pose a devastating risk to coastal communities in the North Atlantic. We discuss the relative sea-level changes in the North Atlantic for the last 2,000 years using proxy records and tide gauges data. We employ a noisy input Bayesian generalised additive model to decompose the RSL signal in order to examine the dynamic sea level component along the coastlines of the North Atlantic. We investigate two components of dynamic sea-level change; the vertical Atlantic Meridional Overturning Circulation (AMOC) circulation and the quasi-horizontal circulation (surface-enhanced currents and gyres). Results show a decline in the AMOC with an unprecedented rate of decrease beginning in 1850 CE, aligning with previous studies. Quasi-horizontal circulation varies significantly over the last 2,000 years with an notably difference north and south of Cape Hatteras, USA. However for both components, the considerable bivariate uncertainties stemming from proxy records have led to substantial uncertainties within our outcomes, thereby reducing the resolution of our results. R code and data is available at https://github.com/maeveupton/RSL-NorthAtlantic.

5.1 Introduction

In 2019 the Intergovernmental Panel for Climate Change (IPCC) reported that the cryosphere (Earth's frozen areas) has been diminishing due to global warming (Pörtner et al., 2019). The average rate of ice loss from the Greenland ice sheet is estimated to be 278 ± 11 Gt/yr between 2006 and 2015 with the oceans having absorbed 89% of the excess heat, resulting in warming since 1970 (Pörtner et al., 2019; von Schuckmann et al., 2023). The ice loss has been widely reported as contributing to global mean sea-level rise (Church et al., 2011, 2013; Gregory et al., 2013; Bilbao et al., 2015). The population along the low-lying coastal zones (less than 10 metres above sea level) is predicted to rise to over one billion people by 2050 (Pörtner et al., 2019; IPCC, 2023).

Relative sea level (RSL) is the difference between the sea surface and the Earth's sediment surface and varies in time and space (e.g. Shennan et al., 2015). RSL change is impacted by a range of processes on different spatial (global, regional, local) and temporal (annual to millennial) scales. Such processes include, for example, geological processes, ocean dynamical processes, atmospheric processes and human activities (e.g. Bulgin et al., 2023). In particular, changes in ocean circulation, and ocean dynamical process, can have far-reaching consequences, impacting regional climates, sea surface temperatures, and even the modulation of global atmospheric circulation systems.

In this paper, we investigate the influence ocean dynamics can have on RSL changes, which is defined as dynamic sea-level change (Gregory et al., 2019), using proxy records and instrumental data from tide gauges. Proxy records are used to extend the temporal range prior to the instrumental era using physical, biological, or chemical features that are preserved in dated geological archives such as coastal sediment (e.g. Gehrels, 1994) or corals (e.g. Meltzner et al., 2017). For the instrumental data, we use tide gauge data to gain insights into coastal sea-level change, as satellite instrumental data, while providing high resolution, is limited to the last thirty years (Church and White, 2011). The combination of proxy records, tide gauges and appropriate statistical models prove useful when examining the pre-industrial evolution of dynamic sea- level changes along the coastlines of the

North Atlantic (Upton et al., 2023c).

Our analysis focuses on the North Atlantic region as it holds great significance in the realm of global climate dynamics, exerting a profound influence on both the region itself and the entire planet (e.g. Lozier et al., 1995). In this region, we are interested in investigating two main components of large-scale ocean circulation that impact dynamic sea level (Little et al., 2019). The two components include; (1) the deep overturning circulation known as the Atlantic meridional overturning circulation (AMOC) and (2) the quasi-horizontal circulation consisting of gyres (circular systems of currents formed due to Earth's rotation and wind patterns, Denny, 2008) and other surface-enhanced currents (near surface currents driven by wind patterns, temperature differences and Earth's rotation, Vallis, 2012). Some of the major physical forces influencing these ocean circulation patterns include: wind; thermohaline circulation; the Coriolis effect (caused by the Earth's rotation deflecting the direction of ocean currents to the right in the Northern Hemisphere or to the left in the Southern Hemisphere); gravitational forces; topography and continental boundaries (shape of the ocean basins); atmospheric and oceanic variability (e.g. North Atlantic Oscillation which is the North South pressure difference between low pressure in Iceland and high pressure at the Azores; Vallis, 2017).

The AMOC redistributes heat from south to north, influencing not only the North Atlantic climate but also affecting weather patterns across Europe, North America and the broader Northern Hemisphere (e.g. Frankignoul et al., 2013; Ionita et al., 2022; Bellomo et al., 2023). Previous studies have shown that weakening of the AMOC is linked to projected, and possibly ongoing, sea-level rise along the Atlantic coast of North America (Little et al., 2019), with the North Atlantic experiencing stronger rates of rise compared to the North Pacific (Körper et al., 2009).

Quasi-horizontal circulation in the North Atlantic is influenced by a combination of gyres (North Atlantic Subtropical Gyre and North Atlantic Subpolar Gyre) and surface-enhanced currents (e.g. the Gulf Stream, the North Atlantic Drift, and the Canary Current) (Little et al., 2019). It is a critical component of dynamic sea-level change and McCarthy et al. (2015) demonstrated how dynamic sea level decreases northwards due to the transitions between North Atlantic subtropical

and subpolar gyres using tide-gauge data along the western Atlantic seaboard. As a result, we seek to determine the level of influence the AMOC and quasi-horizontal circulation has had on pre-industrial dynamic sea level in the North Atlantic.

Understanding how dynamic sea level evolved in the North Atlantic over the last 2,000 years (a period in Earth's history known as the late Holocene or Common Era) is crucial and requires a flexible statistical modelling approach. The sea-level data collected using instrumental records and proxies are the net outcome of various physical processes that vary in time and space (Khan et al., 2022). Recently, statistical models have been used to decompose this RSL signal in order to examine the different drivers of sea-level change (Kopp et al., 2016; Walker et al., 2021; Upton et al., 2023c). In this paper, we use a noisy input generalised additive model (NI-GAM) approach, as developed by Upton et al. (2023c), focusing on regional changes in the North Atlantic. The NI-GAM is extended from Upton et al. (2023c) to possess four components: (1) a common component averaged over time for the full region, (2) a hierarchical regional component averaged over time for different regions, (3) a linear local component and (4) a non-linear local component. To retrieve and process instrumental RSL data, we use the `reslr` package which is an open source R software that allows the paleo sea-level community to examine RSL changes (Upton et al., 2023b). Using the extended NI-GAM, we decompose the RSL signal and focus on the influence of dynamic sea-level change in the North Atlantic region.

We undertake two strategies to investigate the two main components of the North Atlantic ocean circulation system: the AMOC and the quasi-horizontal circulation. First, we apply the extended NI-GAM (Upton et al., 2023c) to proxy sea-level records and tide gauge data and examine the difference between the regional component from the east and west coast of the North Atlantic in order to understand the evolution of the AMOC. For the second strategy, we apply the same extended NI-GAM (Upton et al., 2023c) to proxy sea-level records and tide gauges from the Atlantic coast of North America and investigate the regional component north and south of Cape Hatteras to examine quasi-horizontal circulation in the region. The results from our analysis address the reported long term changes in the intensity and stability of the AMOC and the quasi-horizontal circulation in

the North Atlantic (e.g. Rahmstorf et al., 2015; McCarthy et al., 2015; Thornalley et al., 2018; Caesar et al., 2021).

We structure our paper in the following manner. In Section 5.2, we provide background information on the ocean dynamics that influence sea-level changes in the North Atlantic. In Section 5.3, we discuss the data sources that we use in this research. In Section 5.4, we provide a detailed mathematical description for the NI-GAM focusing on the extension used in this analysis. The results for our analysis are presented in Section 5.5 and we discuss these results in Section 5.6.

5.2 Ocean Dynamics in the North Atlantic

Oceans are indispensable to our climate system, acting as crucial regulators and conveyors of water, heat, and nutrients across the globe (e.g. Pörtner et al., 2019). On interannual and decadal time scales, ocean dynamics are a crucial component of sea-level change which can impact marine life in our oceans and our coastlines (Di Lorenzo et al., 2008; Qiu and Chen, 2012; Minobe et al., 2017). On large spatial scales, ocean dynamics can vary and these variations have dramatically increased due to global warming resulting in non-uniform global sea-level changes (e.g., Yin et al., 2010; Slangen et al., 2014). As a result, Gregory et al. (2019) defines dynamic sea-level change as the mean sea-level changes above the geoid (theoretical surface that depicts the shape of the Earth's surface if influenced solely by gravity) caused by ocean dynamics.

The North Atlantic Ocean is crucial for global climate regulation (e.g. Muschitiello et al., 2019), marine biodiversity (e.g. Olafsson et al., 2021), carbon sequestration (e.g. Olafsson et al., 2021), and sea-level variability (Thompson, 1986; Lozier et al., 1995; Kopp, 2013). One key aspect of the North Atlantic Ocean's importance lies in the presence of active deep water formation sites. These sites facilitate the vertical mixing of waters, allowing for the transfer of heat, nutrients, and other vital components throughout the ocean layers (Rahmstorf, 2006). As a result, the ocean circulation system in the North Atlantic possesses a complex array of currents that can impact sea-level rise at a range of time scales and contribute to the delicate balance of the region's climate (Bjerknes, 1964).

In the upcoming sections, we focus on two important components of the North Atlantic Ocean circulation system; (1) the Atlantic Meridional Overturning Circulation (AMOC) which impacts vertical circulation in the North Atlantic and (2) quasi-horizontal circulation, which is driven by a combination of gyres (North Atlantic Subpolar Gyre and North Atlantic Subtropical Gyre) and surface-enhanced currents, for example the Gulf Stream. The aim of our paper is to investigate the evolution of dynamic sea level over the Common Era, impacted by these ocean circulation patterns, using proxy records and tide-gauge data.

5.2.1 Atlantic Meridional Overturning Circulation(AMOC)

The Atlantic Meridional Overturning Circulation (AMOC) is a system of currents that redistributes heat from the equatorial regions towards the poles (McCarthy et al., 2020). The deep water formation in the North Atlantic serves as a critical driver for large-scale oceanic circulation patterns, such as the AMOC. The AMOC significantly contributes to the relatively mild climate of northwestern Europe (Seager et al., 2002). It is responsible for maintaining temperatures approximately 3 °C warmer on the northwestern European seaboard compared to similar maritime climates on the eastern seaboard of North America (Rhines et al., 2008).

Changes in the intensity or stability of the AMOC can have far-reaching consequences, impacting regional climates, sea surface temperatures, and even the modulation of global atmospheric circulation systems. Recent literature has indicated that the AMOC is weakening (Rahmstorf et al., 2015; McCarthy et al., 2015; Thornalley et al., 2018; Thibodeau et al., 2018; Spooner et al., 2020; Caesar et al., 2021). However, these studies have been limited in time. We seek to extend this analysis to examine how dynamic sea-level variations are impacted by the AMOC over the Common Era.

Mathematically, the relationship between the AMOC and dynamic sea level in the North Atlantic can be given as:

$$Q_o \propto h_e - h_w \tag{5.1}$$

where Q_o is the water volume transported by the AMOC and h_e and h_w are the eastern-boundary sea level and western-boundary sea level respectively (Little et al., 2019). In other words, when water volume transported by AMOC increases, a simultaneous increase in the east-minus-west sea-level difference across the North Atlantic ocean basin is expected to be observed.

In our paper, we use a statistical model to decompose the RSL signal into four separate components: common; regional; linear local and; non-linear local components. When examining the dynamic sea-level change for the North Atlantic, we find the difference between the regional component for the east coast and the regional component for the west coast. This allows us to investigate the influence of the AMOC on our dynamic sea-level change for the Common Era.

5.2.2 Quasi-Horizontal Circulation

The quasi-horizontal circulation in the North Atlantic refers to the large-scale oceanic movement that occurs mainly along the surface of the ocean (Vallis, 2012). This circulation pattern is characterised by the presence of gyres and other surface-enhanced currents. Gyres are circular, rotating systems of ocean currents that tend to form in the major ocean basins due to the combined effects of wind patterns and the Coriolis effect (a result of Earth's rotation McCarthy et al., 2020). In the North Atlantic, there are several prominent gyres, including the North Atlantic Subtropical Gyre and the North Atlantic Subpolar Gyre (Little et al., 2019). The North Atlantic Subtropical Gyre is the larger and more stable of the two, driven by the trade winds blowing from east to west across the subtropical regions (Yin et al., 2010). The clockwise rotation of this gyre tends to accumulate warm surface waters in the central part of the North Atlantic (Rossby, 1999). The North Atlantic Subpolar Gyre, on the other hand, is located in the higher latitudes of the North Atlantic and experiences more variable wind patterns (Little et al., 2019). This gyre rotates cyclonically and tends to accumulate colder and fresher waters from the Arctic region (Rossby, 1999).

The North Atlantic experiences various other surface-enhanced currents, such as the Gulf Stream, the North Atlantic Drift, and the Canary Current (McCarthy et al., 2020). The Gulf Stream, in particular, is a significant and well-known ocean

current that flows along the eastern coast of North America, transporting warm water from the Gulf of Mexico towards the North Atlantic (Minobe et al., 2008). The interaction of these gyres and surface-enhanced currents plays a crucial role in regulating the transfer of heat and energy within the North Atlantic Ocean and the adjacent regions. The quasi-horizontal circulation in the North Atlantic exhibits significant differences between the northern and southern regions of Cape Hatteras in North Carolina, as it marks the boundary between the subtropical and subpolar gyres (McCarthy et al., 2015). These differences demonstrate the intricate interplay of oceanic currents and their impact on the climate, marine ecosystems, and coastal environments in the North Atlantic region.

McCarthy et al. (2015) describes a relationship between the transport along the intergyre boundary (region between the North Atlantic Subpolar Gyre and the North Atlantic Subtropical Gyre) and the dynamic sea level north and south of Cape Hatteras as:

$$\upsilon_{ig} \propto h_s - h_n \tag{5.2}$$

where υ_{ig} represents the water transported along the intergyre boundary. h_s is the dynamic sea level south of Cape Hatteras and h_n is dynamic sea level south of Cape Hatteras.

As previously mentioned, we decompose the RSL signal and investigate the evolution of dynamic sea-level change along the Atlantic coast of North America. By finding the difference between our regional component north and south of Cape Hatteras, we can address the influence of these quasi-horizontal circulations (i.e. gyres and surface-enhanced currents) on dynamic sea-level changes in this region for the Common Era.

5.3 Data

In this analysis, we focus on coastal locations around the North Atlantic. In this section, we discuss the sourced data from proxy and tide-gauge records along the Atlantic coast of North America and from the coast of Iceland and European sites as shown in Figure 5.3.1. In particular, we discuss the separate regions that will be examined using separate statistical models. The first region is an analysis of

all the data around the North Atlantic coastlines with a separate identifier for the east coast (i.e. North American sites) and the west coast (i.e. Europe and Iceland). The second region is an analysis of the Atlantic coast of North America with sites above Cape Hatteras deemed the north sub-region and sites below Cape Hatteras are designated as the south sub-region.

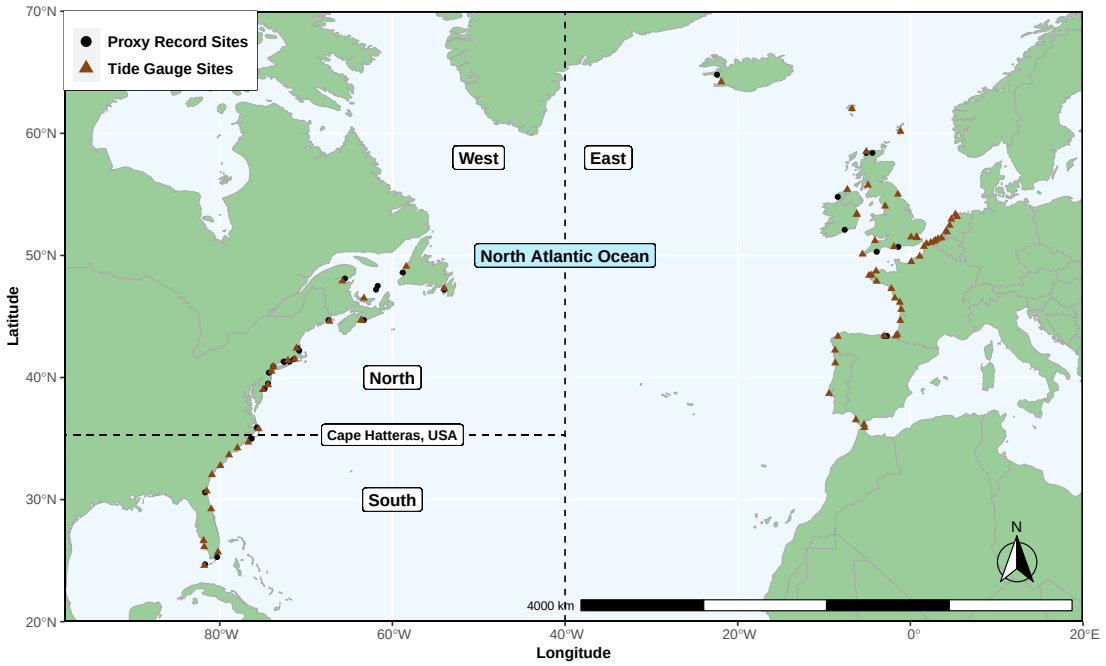


Figure 5.3.1: Location of the tide-gauge sites and proxy record sites along the North Atlantic coastline. The east North Atlantic section comprises all the sites to the east of 40°W, as indicated by the black dashed line, while the remaining sites belong to the west North Atlantic section. Cape Hatteras North Carolina, is highlighted with the black dashed line. Sites located above this line are categorised as part of the north Cape Hatteras sub-region, whereas those below the line fall under the classification of the south Cape Hatteras sub-region.

5.3.1 Proxy Sea-Level Records

Proxy sea-level datasets are vital sources of information for investigating the evolution of RSL changes over the Common Era. Near-continuous proxy RSL reconstructions can provide information on RSL changes prior to the instrumental period (e.g. Kemp et al., 2013). They are often obtained from sediment cores from salt marshes (at mid- to high latitudes; Gehrels et al., 2020) or mangrove sediment (low latitudes; Khan et al., 2022).

In these sediment cores, a subset of depths are dated using radiocarbon measurements or other dating methods. Additionally, the shallowest part of the core, representing the most recent sediment layers, can be dated by identifying historic pollution and land use changes of known age in down-core profiles. These changes are observed through elemental abundance, isotopic activity and isotopic ratios (Marshall, 2015). These directly dated levels in the core (i.e., age of sediment sample) form the input for a statistical age-depth model, e.g. the Bchron (Parnell et al., 2008), Bacon (Blaauw and Christen, 2011), or Rplum (Aquino-López et al., 2018) packages in R. These age-depth models (irrespective of their specific similarities and differences) estimate the age of every 1 cm thick sediment sample in the core with uncertainty.

To reconstruct RSL, a sea-level proxy is used. A proxy encompasses physical, biological, or chemical properties, that are directly measured and they provide valuable insights into an unmeasurable variable of interest, i.e., RSL (e.g., Gornitz, 2009). In the field of sea-level studies, proxy data can be obtained from various sources, including microorganisms like foraminifera (e.g., Edwards and Wright, 2015), geochemical measurements (e.g., Marshall, 2015), and the accumulation of vegetation within the tidal zone (e.g., Kemp and Telford, 2015). The information from these sources is preserved in dated geological archives, and it can be used to estimate paleo-marsh elevation (i.e., elevation with respect to height above a tide level at the time of formation) with uncertainty (Kemp et al., 2013). This estimation is achieved through either a transfer function approach, which establishes a relationship between the abundance of specific micro-fossil families and tidal elevation using data representative of the modern environment (Sachs et al., 1977; Horton and Edwards, 2006; Kemp et al., 2011; Cahill et al., 2016), or through reasoning by analogy. The reasoning by analogy method involves observing the distribution of plants in modern salt marshes and using this knowledge to interpret their analogous counterparts found in core material (Kemp and Telford, 2015). By employing these approaches, paleo-marsh elevation can be estimated along with associated uncertainty (e.g., Gehrels, 1994; Shennan et al., 2015; Kemp et al., 2018). The age of each core sample derived from an age-depth model and a paleo-marsh elevation reconstruction for the core are combined to provide a single

proxy RSL record comprised of age (with 1 sigma uncertainty) and RSL (with 1 sigma uncertainty). Figure 5.3.2 illustrates the proxy RSL record, displaying the associated 1-sigma uncertainty for both age and RSL, providing a visual representation.

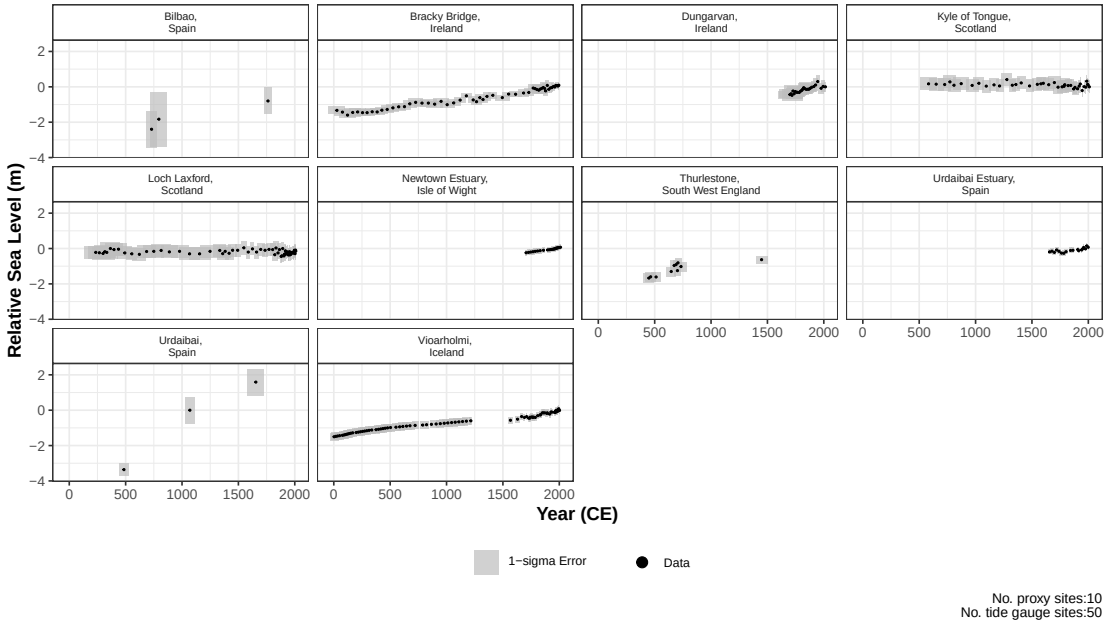


Figure 5.3.2: Proxy records for the east coast of the North Atlantic coastline. The y-axis is relative sea level (RSL) in metres and the x-axis is the years in units of Common Era. The black dot represents the midpoint of the proxy sea-level reconstruction and the grey boxes of 1 standard deviation represent vertical and horizontal (temporal) uncertainty.

5.3.2 Tide Gauge Data

Tide gauges provide direct regular (e.g. minute to hourly) measurements of RSL (Pugh and Woodworth, 2014b). The Permanent Service for Mean Sea Level online database provides tide-gauge records for $\sim$ 1500 stations (PSMSL, 2023: Holgate et al., 2013; Woodworth and Player, 2003). The `reslr` package enables processing of the tide-gauge data to ensure that it is comparable with relevant (i.e., nearby) proxy records (detailed description of the methodology is provided in Upton et al., 2023b). Through the `reslr` package, we select tide-gauge data with more than 20 years of observations situated closest to the proxy site. In addition, the `reslr` package enables users to choose supplementary tide gauges from the PSMSL online

database. We incorporate extra tide-gauge data along the European coastline, spanning an area from the Netherlands to Spain, under the stipulation that the temporal record of the tide gauges exceeds 50 years. This approach has provided us with an extensive collection of tide gauges within the North Atlantic region reducing the spatial bias present in the data. Similarly, we incorporate additional tide gauge data south of Cape Hatteras to overcome the spatial bias in this region.

For our analysis, we present two separate sub-regions in the North Atlantic which include: (1) east west of the North Atlantic coastline and (2) north south of Cape Hatteras North Carolina USA (Longitude: -75.5 and Latitude: 35.3). In the first dataset, we class all sites in Europe and Iceland as the east, as demonstrated in Figure 5.3.2, and the west sites include all proxy and tide-gauge sites on the Atlantic coast of North America as shown in Figure 5.7.1. For the second dataset, we focus on the Atlantic coast of North America and organise all the tide gauge and proxy sites north of Cape Hatteras in Figure 5.7.2 and south of Cape Hatteras in Figure 5.7.3. In the Appendix 5.7, we present a table of all the proxy data sites and tide gauge location and a corresponding reference to the source literature.

5.4 Statistical Methodology

We now describe the statistical approach to decomposing the RSL signal and investigate the contribution of dynamic sea-level change in the North Atlantic region. In particular, we are interested in the contribution of the two main components of dynamic sea-level change in this region which are the AMOC and quasi-horizontal circulation (Refer to Section 5.2 for detailed description of both). The goal is to use proxy records and tide-gauge data to examine the pre-industrial dynamic sea-level changes which are influenced by the AMOC and quasi-horizontal circulation in the North Atlantic. For this analysis, we require two separate strategies for examining these complex ocean circulation systems. The two strategies use the same statistical model but with different input data as described here:

1. We model the data sites along the coastline of the North Atlantic using an extended version of NI-GAM (Upton et al., 2023c). The extended version possesses an additional component, as compared to the model outlined in

Upton et al. (2023c), to capture the regional signals as well as a common signal along the North Atlantic coastline. We identify separate regional components for the east of the North Atlantic (Figure 5.3.2) and west of the North Atlantic (Figure 5.7.1) using a hierarchical prior on the spline coefficient. This east-west analysis allows us to investigate the influence of the AMOC on RSL changes for the region over the Common Era.

2. Similarly, we employ an extended variant of the NI-GAM model (Upton et al., 2023c) to characterise the data sites located along the Atlantic coast of North America. To reiterate, this extended model incorporates an additional component aimed at capturing the shared signal prevalent along the Atlantic coast of North America (common component). The regional component involves a hierarchical spline coefficient with distinct values for the north (refer to Figure 5.7.2) and south (refer to Figure 5.7.3) of Cape Hatteras, North Carolina, USA (Longitude: -75.5288, Latitude: 35.2505). Through the north-south analysis, we gain insights into the variations of RSL over the last 2,000 years resulting from quasi-horizontal ocean circulation.

5.4.1 Statistical Model

For both analyses, we use an extended version of Noisy Input Generalised Additive model (NI-GAM) as outlined in Upton et al. (2023c). The original NI-GAM uses a combination of splines and random effects to decompose the regional RSL signal into three separate components. In our extended NI-GAM, an additional component is included to represent the common signal across the region which we outline below. The model uses a Bayesian hierarchical framework with Markov Chain Monte Carlo (MCMC) simulations being undertaken using Just Another Gibbs Sampler (JAGS) tool (Plummer, 2003) and implemented using the `rjags` package in R (Plummer et al., 2016).

The data level for the NI-GAM is described as:

$$y = f(\mathbf{x}, t) + \epsilon_y \tag{5.3}$$

where y is the RSL measurement in metres. $f(\mathbf{x}, t)$ is the process underlying the RSL data that depends on location $\mathbf{x}$ and time t. ϵ_y is the error term given

by $\epsilon_y \sim \mathcal{N}(0, \sigma_y^2 + s_y^2)$, where σ_y^2 is the residual variance and s_y is the known measurement error associated with each RSL reconstructed data point.

To account for age uncertainty that arises when using proxy RSL records we use the Noisy Input (NI) uncertainty method (McHutchon and Rasmussen, 2011). For the input measurement, $\tilde{t}$ is assumed to be the a noisy estimate for the true value t given as:

$$\tilde{t} = t + \epsilon_t \tag{5.4}$$

where the error term is given by $\epsilon_t \sim \mathcal{N}(0, s_t^2)$ where s_t is the known age measurement error associated with that data point. The NI method is a three step process where (1) an initial model is fitted without age error, (2) the derivative of f is estimated using this model, and a corrective variance term is computed in units of y (as opposed to t). Finally in step (3), the model is re-run with this additional corrective variance term added to the ϵ_y variance. By employing this approach, our model is able to acquire knowledge of the input noise variation by studying the complete output it generates.

For the process level, the extended NI-GAM uses a spatio-temporal RSL field and includes an additional component written in the following manner:

$$f(\mathbf{x}, t) = c(t) + r(t, w_{\mathbf{x}}) + g(z_{\mathbf{x}}) + h(z_{\mathbf{x}}) + l(\mathbf{x}, t) + \epsilon_{\mathbf{y}} \tag{5.5}$$

where $f(\mathbf{x}, t)$ is the underlying RSL process for a generic location and time. $c(t)$ is the common component over time for all sites. $r(t, w_{\mathbf{x}})$ is the regional component at time t which varies depending on east or west of the North Atlantic and north or south of Cape Hatteras, with $w_{\mathbf{x}}$ identifying the section east or west or north or south. $g(z_{\mathbf{x}})$ is the linear local component represented by a random effect with $z_{\mathbf{x}}$ representing each data site. $h(z_{\mathbf{x}})$ is the spatial vertical offset for each data site. $l(\mathbf{x}, t)$ is the non-linear local component.

The common component, $c(t)$, is represented using a spline in time and represents physical processes that are common to all locations, including barystatic and thermosteric contributions, where the former is caused by the transfer of mass between land-based ice and oceans (Gregory et al., 2019) and the latter is influenced by changes in global temperature creating density variations within the oceans (Grin-

sted, 2015). For this component, we use cubic B-splines (piece-wise polynomials joined at equidistant knots with equal first derivatives de Boor, 1978; Eilers and Marx, 1996), described by:

$$c(t) = \sum_{s=1}^{k_c} b_{c_s}(t)\beta_s^c \tag{5.6}$$

where β_s^c is the s^{th} spline coefficient, k_c is the number of knots and $b_{c_s}(t)$ is the s^{th} spline basis function at time t. We define the prior for the spline coefficients of the common component β_s^c by:

$$\beta_s^c \sim \mathcal{N}(0, \sigma_c^2) \tag{5.7}$$

where σ_c is the standard deviation of the spline coefficients for the common component which controls the general smoothness of the spline. In addition, we calculate the rate of change of the common component by finding the derivative over time for each basis function, $b_{c_s}(t)$, using first principles.

The regional component, $r(t, w_{\mathbf{x}})$, represents physical processes that occur at specific spatial scales. One such example is dynamic sea-level changes that impact groups of sites differently, as previously mentioned in Section 5.2. Similar to the previous component, the regional component is mathematically described using a cubic B-spline in time. However in contrast, the regional component has a hierarchical structure which identifies different groups of sites ($w_{\mathbf{x}}$), for example east and west of the North Atlantic or north and south of Cape Hatteras in the following way:

$$r(t, w_{\mathbf{x}}) = \sum_{s=1}^{k_r} b_{r_s}(t, w_{\mathbf{x}})\beta_s^r \tag{5.8}$$

where β_s^r is the s^{th} spline coefficient with i identifying the different section, k_r is the number of knots and $b_{r_s}(t)$ is the s^{th} spline basis function at time t. We define the prior for the spline coefficients of the regional component β_s^r by:

$$\beta_s^r \sim \mathcal{N}(0, \sigma_r^2) \tag{5.9}$$

where σ_r is the standard deviation of the spline coefficients and controls the smoothness of the regional spline. For the regional component, there is a different σ_r depending on the group of sites in question, i.e. east and west or north

and south. As previously mentioned, we calculate the rate of change of the regional component by finding the derivative over time for each basis function, $b_{r_s}(t)$, using first principles.

The linear local component, $g(z_{\mathbf{x}})$, aims to capture linear trends present in the RSL signal. One such reason for the expected linearity is glacial isostatic adjustment (GIA), which is a response of the Earth, the gravitational field, and the ocean, to changes in the size of ice sheets (Whitehouse, 2018). On relatively short geological timescales, it is approximated to be linear through time with spatial variability along the Atlantic coast of North America (Engelhart et al., 2009). Mathematically, the linear local component is an unstructured random effect for each site which is formulated as:

$$g(z_{\mathbf{x}_j}) = \beta_j^g t \tag{5.10}$$

where β_j^g is a slope parameter specific for each site j. The prior for the linear local component is:

$$\beta_j^g \sim \mathcal{N}(m_{g_j}, s_{g_j}^2) \tag{5.11}$$

where m_{g_j} and $s_{g_j}^2$ are empirically estimated rate and associated variance values as described in detail in Upton et al. (2023c).

The site-specific vertical offset h is used to capture vertical shifts linked to variability between sites. It is formulated using is a random effect in the following manner:

$$h(z_{\mathbf{x}_j}) = \beta_j^h \tag{5.12}$$

where β_j^h contains the random effect coefficients for site j. The prior for the site-specific vertical offset is given by:

$$\beta_j^h \sim \mathcal{N}(0, \sigma_h^2) \tag{5.13}$$

where σ_h^2 is the variance of the random intercept across all data sites.

The non-linear local component $l(\mathbf{x}, t)$ captures RSL variability on century to millennia timescales which are site specific, for example sediment compaction (impacting the height of the Earth's surface: Brain et al., 2011; Horton et al., 2018) and tidal range changes (difference between high and low water marks impacted

by sedimentation, geomorphology of site and human activites: Hill et al., 2011; Hall et al., 2013; Mawdsley et al., 2015). It is described using a spatio-temporal spline function constructed with B-splines (de Boor, 1978) and a tensor product to capture the variability over time and space (represented by longitude and latitude) so that the individual covariates are combined product-wise (Wood, 2006). It is formulated by:

$$l(\mathbf{x}, t) = \sum_{s=1}^{k_l} b_{l_s}(\mathbf{x}, t)\beta_s^l \tag{5.14}$$

where β_s^l is the s^{th} spline coefficient, k_l is the number of knots and $b_{l_s}(\mathbf{x}, t)$ is the s^{th} spline basis function at time t and location $\mathbf{x}$. The prior for the spline coefficient β_s^l is given as:

$$\beta_s^l \sim \mathcal{N}(0, \sigma_l^2) \tag{5.15}$$

where σ_l^2 is the variance of the spline coefficients over space and time, and controls the smoothness of the space-time component. Also, we calculate the rate of change of the non-linear local component by finding the derivative over time and space for each basis function, $b_{l_s}(\mathbf{x}, t)$, using first principles.

5.5 Results

In this section, we provide the results for the extended NI-GAM for the east-west North Atlantic and the north-south Cape Hatteras analysis. For both analysis, we demonstrate how the RSL signal is decomposed and we provide a comparative analysis for the regional components. The remaining components for the decomposition of the RSL signal are documented in Appendix 5.7. These additional components include the total model fit, the common component, the linear local component, non-linear local component and associated rate of changes for each component and for each site. We then discuss how the dynamic sea level from our east-west analysis of the North Atlantic compares with other AMOC proxy reconstructions. Finally, we provide insight into the fluctuations of dynamic sea level north and south of Cape Hatteras driven by quasi-horizontal circulation from gyres and other surface-enhanced currents.

5.5.1 East West Analysis of the North Atlantic

In this section, we investigate the decomposition of the RSL signal for the North Atlantic. We present one component of the RSL decomposition, the regional component and the corresponding rates of change informed by our proxy records and tide-gauge data. We demonstrate how the difference between the east and west regional components can related to dynamic sea-level changes.

In Figure 5.5.1(a) we display the regional component for the east sites (i.e. east in blue) and west sites (i.e. west in purple) along the North Atlantic coastline. The west sub-region possesses a minimal variation, ranging between -0.004m ± 0.03m to 2mm ±0.04m, with a minor anomaly emerging between 1740 CE to 1770 CE, followed by a gradual decline until the present data. The east possesses a gradual increase from 0 CE to $\sim$ 840 CE. After 840 CE, we observe a gradual decline in the eastern trend until 1770 CE. Following 1770 CE, we see a short period of increase until 1893 CE following by a decrease until the present day.

The rates associated with these trends are presented in Figure 5.5.1(b). The rates of change associated with the west sub-region reflect the minimal fluctuations observed in 5.5.1(c). We observe a slight increase from 0mm/year between 1740CE and 1770 CE, following be a gradual decrease until today. The rate of change for the east sub-region reached its max value of 0.3mm/year ± 0.4mm/year in 573CE. The minimum rate of change observed in the east sub-region is the present day(2023) with rates of -0.2mm/year ± -0.6mm/year.

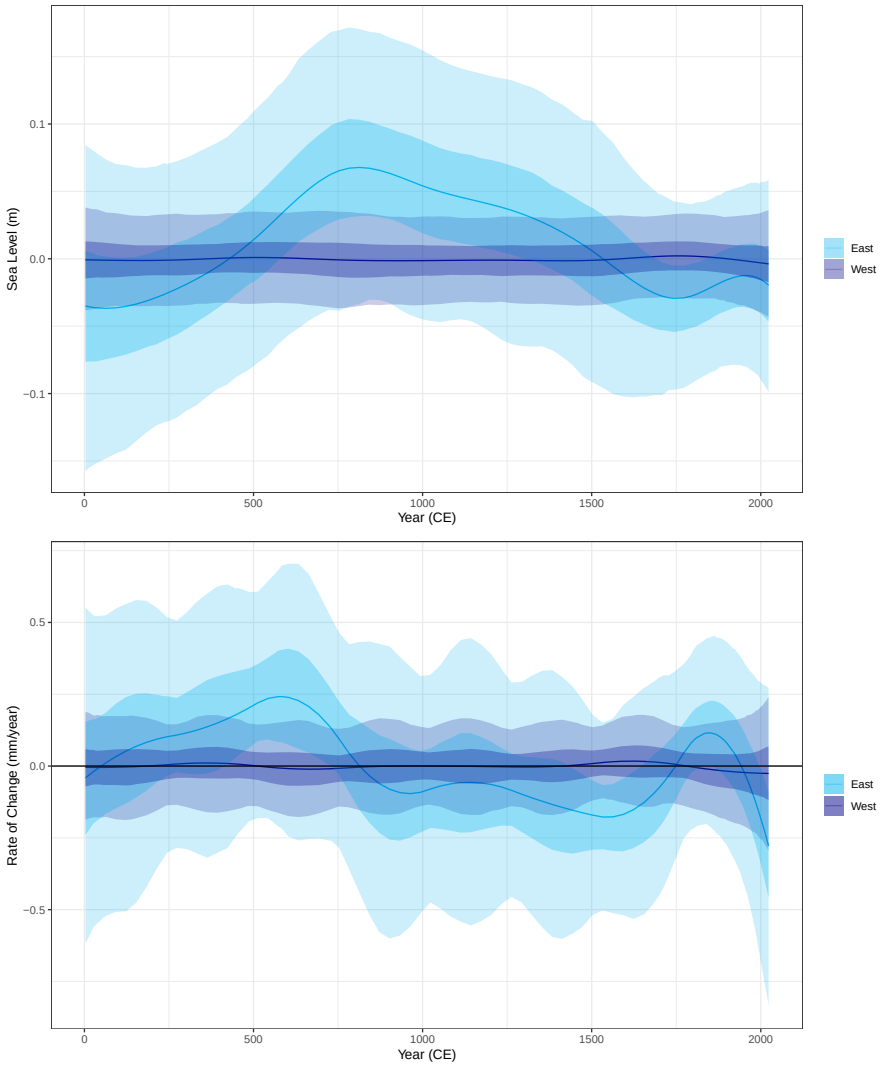


Figure 5.5.1: The extended noisy-input generalised additive model (NI-GAM) results for (a) the regional component and (b) the regional rate of change component. By using separate colours (blue for east and purple for west), we can compare the regional component for the east sites of the North Atlantic versus the west sites of the North Atlantic. (a) The regional component mean model fit represented with a solid line and the shading indicating the 50% and 95% credible intervals. The y-axis is the sea level in m with the x-axis representing the time across the last 3000 years for the North Atlantic region. (b) Rate of change for the regional component for the North Atlantic region with the solid line representing the mean of the fit and shaded area representing the 50% and 95% credible intervals. The y-axis is the instantaneous rate of change of sea level in mm per year.

In Section 5.2, we describe how the difference between the dynamic sea-level component from the easthern- and western- boundary of the North Atlantic basin is

proportional to the water transported by the AMOC (Little et al., 2019). Our models decompose the RSL signal and the regional component captures changes in dynamic sea level using a hierarchical temporal spline. Figure 5.5.2 represents the difference between east west regional component (a) and the associated rate of change of this difference (b). From 0 CE to 810 CE, the trend of the east-west difference exhibited a gradual increase. Subsequent to 810 CE, this trend shifted, leading to a descent until 1770 CE. Following this, there was a renewed ascent in the trend, which persisted until $\sim$ 1900 CE. However, from that point onwards until the present day in 2023, there was a significant decrease in the east-west difference trend. In Figure 5.5.2 (b) we present the associated rate of change of the east-west difference over the past 2,000 years. The rates displayed a gradual increase until $\sim$ 570 CE, succeeded by a decline until 920 CE. Between 920 CE and 1530 CE, there was a gradual decrease in the rates. Following 1530 CE until 1850 CE, there was an observed increase, followed by a subsequent downturn in the rates prior to 1850 CE. This pattern has persisted up to the present day, with rates of change for the disparity between the eastern and western regional components measuring at -0.2mm/year $\pm$ 0.6mm/year as of 2023.

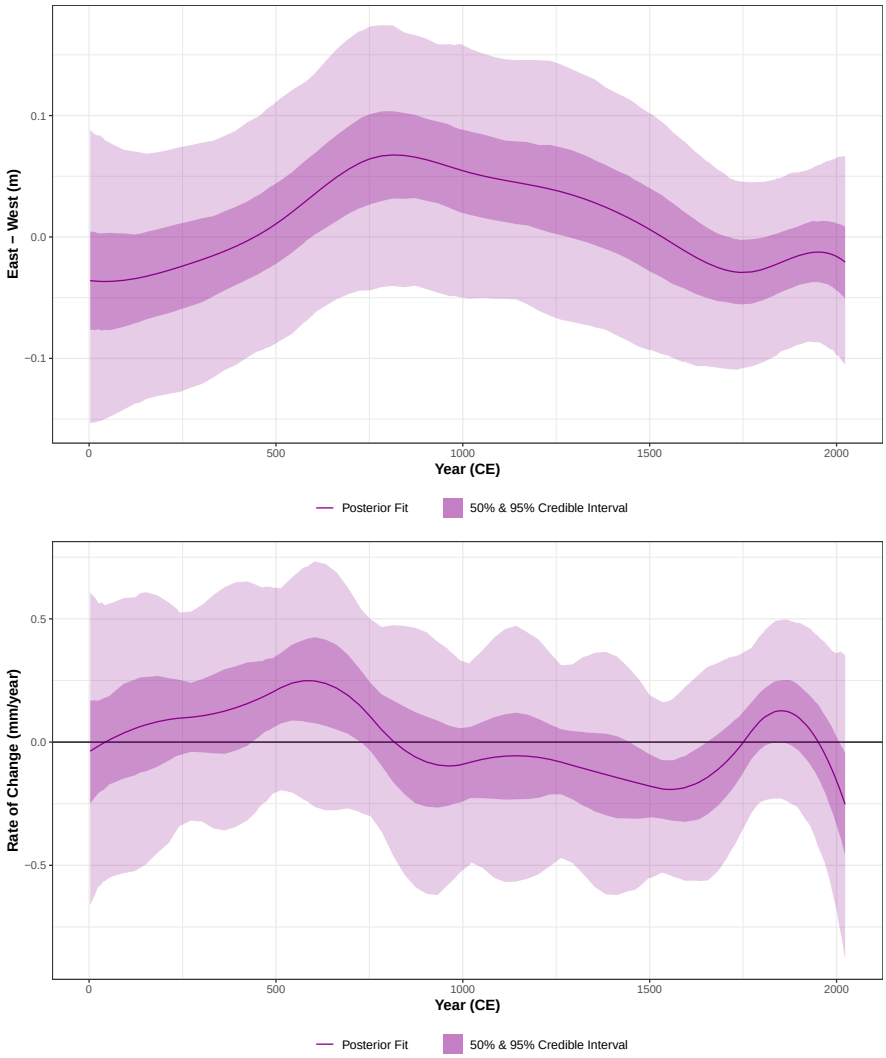


Figure 5.5.2: The difference between the regional component from noisy-input generalised additive model (NI-GAM) for the east and west coasts of the North Atlantic used to identify dynamic sea-level changes. (a) is the difference between the regional components and (b) is the associate rate of change for the difference between the regional components for east and west coasts of the North Atlantic. (a) For the difference between the regional components for east and west, the mean model fit is represented with a solid line and the shading indicating the 50% and 95% credible intervals. The y-axis is the sea level in m with the x-axis representing the time across the last 2,000 years for the North Atlantic region. (b) Rate of change for the difference between regional components for east and west of the North Atlantic region is represented with a solid line for the mean of the fit and shaded area representing the 50% and 95% credible intervals. The y-axis is the instantaneous rate of change of sea level in mm per year.

The results from our regional east-west difference plot (Figure 5.5.2) can be com-

pared with other approaches which have reconstructed the AMOC using with other forms of proxy and instrumental records. By comparing our result with these other trends, we seek to understand how dynamic sea-level estimates could represent the influence the AMOC is having on the North Atlantic region. Caesar et al. (2021) provided insight into the evolution of the AMOC over the last millennium using a range of reconstruction techniques and data sources. We focus on six of these different approaches which are comparable over our longer temporal (i.e. century to millennia). The first approach uses temperature differences between mean surface temperature and the subpolar gyre in the Northern Hemisphere as explained in Rahmstorf et al. (2015). The second includes proxy data using compound-specific δ^{15} N of deep-sea gorgonian corals described in Sherwood et al. (2011). The third study uses marine sediment records to understand the North Atlantic subpolar gyre(SPG) presented by Spooner et al. (2020). The fourth study uses δ^{18}O records in benthic foraminifera (single cell organisms) from sediment cores in the Laurentian Channel as developed by Thibodeau et al. (2018). In the final study, two proxies from marine cores of slit are used to reconstruct the AMOC and data collection process is documented in Thornalley et al. (2018).

In Figure 5.5.3, we compare the rate of change of our regional east-west difference with selected approaches from Caesar et al. (2021). The Rahmstorf et al. (2015) trend is presented in red and demonstrates how the AMOC fluctuates with a small range until a steep down trend is observed begin in the twentieth century and continuing to present day. However, this trend possess large uncertainty making it difficult to confirm the trend of the AMOC. Sherwood et al. (2011) results are presented in gold with the data points possess large bi-variate uncertainties which decrease over time. The additional trend beginning in 1926 CE displays a downward trend similar to the previous reconstruction however, the data is limited to the last 100 years with the exception of the bi-variately uncertain data points. Spooner et al. (2020) reconstruction is presented in green and examines the AMOC variability from 393 CE to 2013 CE. The dark green trend represents the Thibodeau et al. (2018) reconstruction which ranges from 708 CE to 1962 CE. The results using Thornalley et al. (2018) data present two separate trends with the blue result examining the changing AMOC from 380 CE to 1995 CE

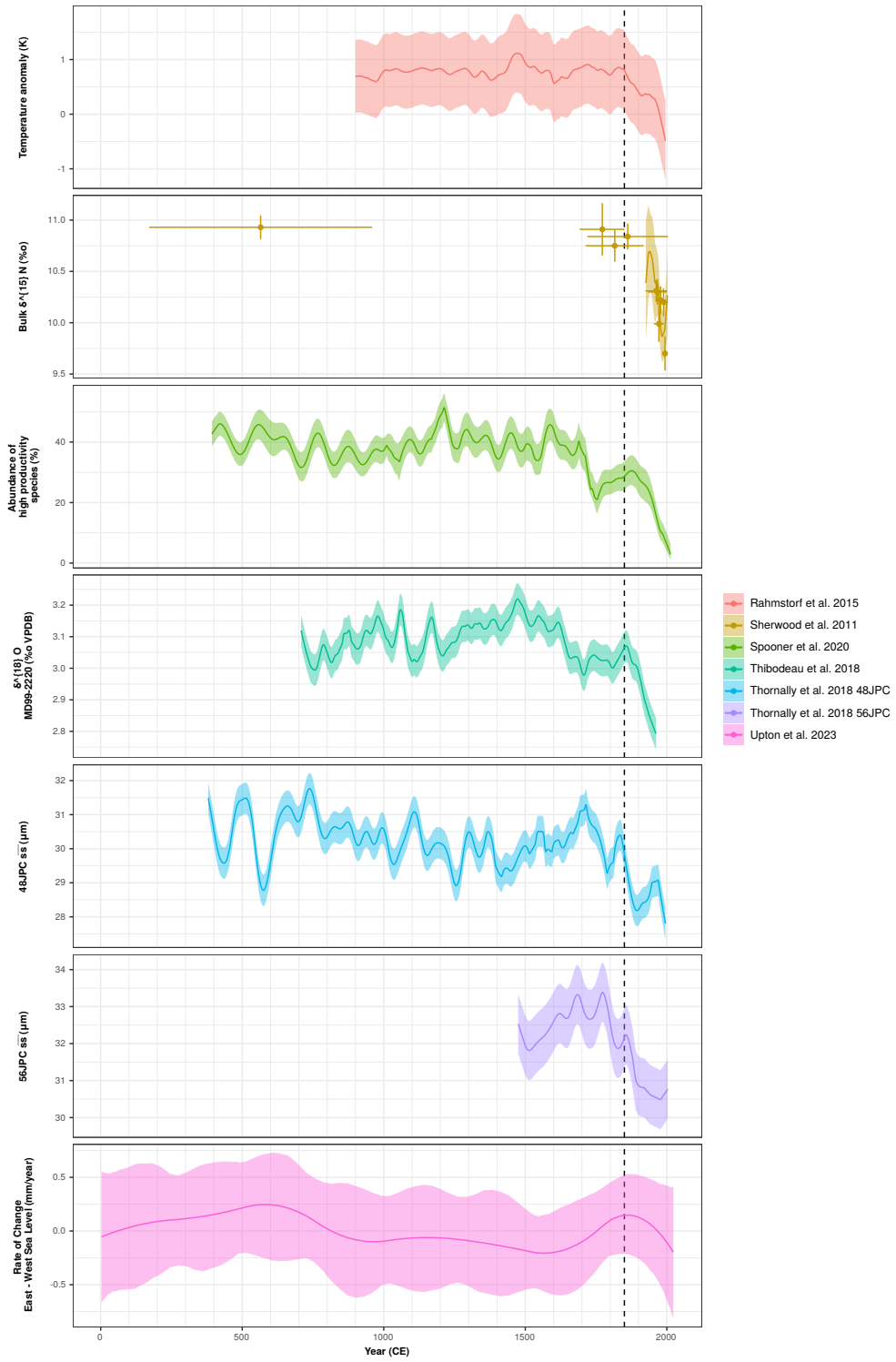


Figure 5.5.3: Comparing the rate of change for the difference in the regional component east and west of the North Atlantic (g) with a range of other Atlantic Meridional Overturning Circulation (AMOC) reconstructions as discussed in Caesar et al. (2021). The black vertical dashed line at 1850 CE highlights the beginning of the downward trend for our rate of change for the east west difference. For each AMOC reconstruction, a 50 year LOWESS filter is applied.(a) The sea surface temperature based proxy demonstrating the response of the temperature of the North Atlantic resulting from a change in Atlantic meridional heat transport (Rahmstorf et al., 2015). (b) Proxy data using compound-specific δ^{15} N of deep-sea gorgonian corals (Sherwood et al., 2011). (c) The relative abundance of *T.quinqueloba*(type of foraminifera) in marine sediment cores (Spooner et al., 2020). (d) δ^{18}O records in benthic foraminifera (single cell organisms) from sediment cores in the Laurentian Channel (Thibodeau et al., 2018). (e,f) Mean grain size of sortable-silt data *ss* from marine cores (Thornalley et al., 2018).

while the purple trend presents a shorter timeframe, from 1475 CE to 2003 CE. Both trends present significant amount of fluctuation. Our rate of change of the east-west dynamic sea level difference, presented in pink, shows the evolution of the AMOC over the past 2,000. The black vertical dashed line represents the year in which the rate of our east west difference began to decrease (1850 CE). It is evident from Figure 5.5.3 that the decline we observe coincides with Rahmstorf et al. (2015), Thornalley et al. (Core 56JPC: 2018) and Thibodeau et al. (2018). Whereas, Spooner et al. (2020), Sherwood et al. (2011) and Thornalley et al. (Core 48JPC: 2018) observe this decline at a later date. However, the precision of our findings introduces complexities when attempting a confident comparison with other AMOC reconstructions.

5.5.2 North South along the North American Atlantic coastline

We demonstrate the decomposition of the RSL signal for the Atlantic coast of North America using the extended NI-GAM approach and data sourced from proxies and tide gauges. We investigate the regional component north and south of Cape Hatteras (Longitude: -75.5288 and Latitude: 35.2505). As previously mentioned, Cape Hatteras in North Carolina holds significance as it is the boundary between the North Atlantic subpolar and subtropical gyre and is a location of importance for the Gulf Stream (McCarthy et al., 2020).

In Figure 5.5.4(a), we examine the variation of the regional component for the north(blue) and south(purple) sub-regions. It is evident that the sea level peaked in the south sub-region in 813CE with a value of 0.03m $\pm$ 0.04m and two minimums troughs in 15 CE and 1560CE with values of -0.04m $\pm$ 0.06m and -0.03m $\pm$ 0.05m respectively. Whereas, the north sub-region possess two maximum peaks in 3CE and 1560CE with sea level values of 0.02m $\pm$ 0.04m and 0.01 $\pm$ 0.05m. Figure 5.5.4(b) demonstrates the rate of change for the regional component with separate trends for the north and south sub-regions. It clear that the rates reflect the oscillatory behaviour represent in Figure 5.5.4(a). The southern rates begin positively, followed by a downward trend until around 1560 CE. Afterward, there is an increase until the present. On the other hand, the northern rates start

negatively and increase to a maximum of 0.07 mm/year ± 0.13 mm/year in 1208 CE, followed by a decrease until the present. Recent rates for the southern sub-region are slightly higher than those for the northern sub-region. The rate of change for the southern sub-region in 2022 is estimated to be 0.08 mm/year ± 0.15 mm/year, whereas the rate of change for the northern sub-region in 2022 is estimated to be -0.05 mm/year ± 0.11 mm/year.

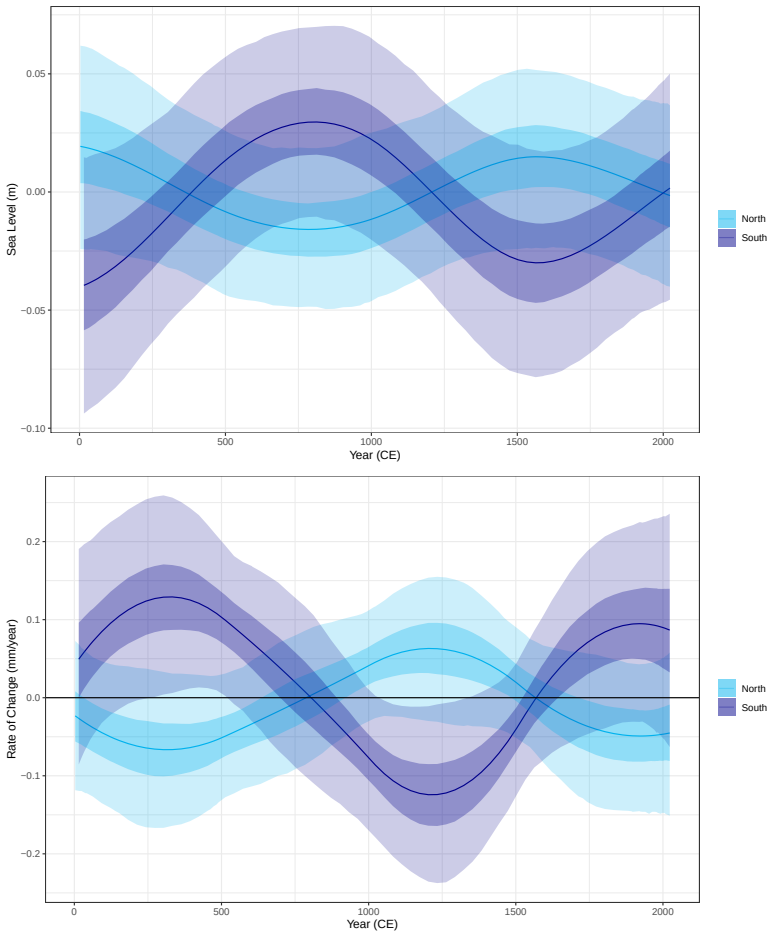


Figure 5.5.4: The extended noisy-input generalised additive model (NI-GAM) results for (a) the regional component and (b) the regional rate of change component. By using separate colours (blue for North and purple for South), we can compare the regional component for the northern sites versus the southern sites of the Atlantic coast of North America. (a) The regional component mean model fit represented with a solid line and the shading indicating the 50% and 95% credible intervals. The y-axis is the sea level in m with the x-axis representing the time across the last 3000 years for the North Atlantic region. (b) Rate of change for the regional component for the Atlantic coast of North America, north and south of Cape Hatteras, with the solid line representing the mean of the fit and shaded area representing the 50% and 95% credible intervals. The y-axis is the instantaneous rate of change of sea level in mm per year.

Figure 5.5.5 represents the difference between south and north regional components (a) and the associated rate of change of this difference (b). It is clear that the south-north difference increased from 0CE to 802CE followed by a downward trend until $\sim$ 1580. After $\sim$ 1580, the south-north difference increase and it continues to increase to the present. The corresponding rate of change plot (b) demonstrates

how the rate ranged between -0.19mm/year and 0.2 mm/year over the past 2,000 years. The rates have evolved from the maximum in 324CE to the minimum in 1214CE followed by a gradually increase until $\sim$ 1920. After $\sim$ 1920, the rates begin to decrease and the current rate value is 0.14mm/year $\pm$ 0.18mm/year.

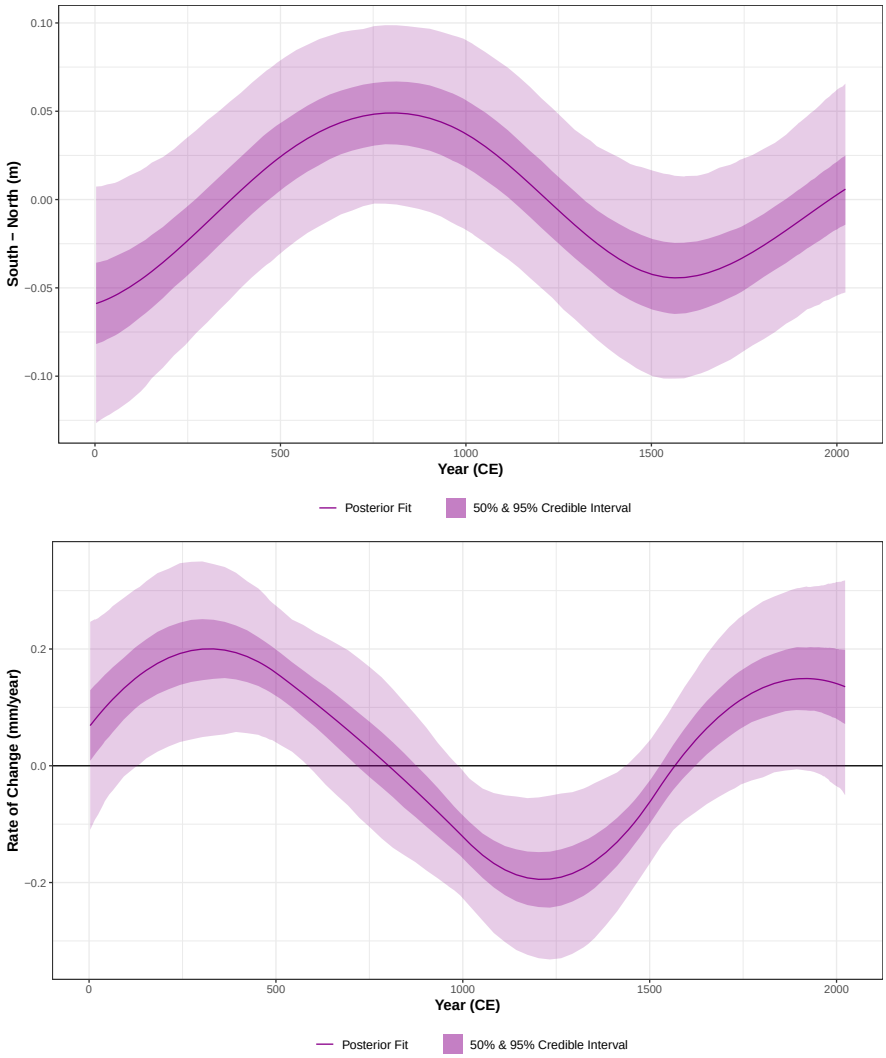


Figure 5.5.5: The difference between the regional component from noisy-input generalised additive model (NI-GAM) for the north and south of Cape Hatteras for the Atlantic coast of North America (a) the difference between the regional components and (b) is the associate rate of change for the difference between the regional components. (a) For the difference between the regional components for north and south, the mean model fit is represented with a solid line and the shading indicating the 50% and 95% credible intervals. The y-axis is the sea level in m with the x-axis representing the time across the last 2,000 years for the North Atlantic region. (b) Rate of change for the difference between regional components for north and south of Cape Hatteras North Carolina along the Atlantic coast of the North America is represented with a solid line for the mean of the fit and shaded area representing the 50% and 95% credible intervals. The y-axis is the instantaneous rate of change of sea level in mm per year.

5.6 Discussion

By analysing RSL changes across different temporal and spatial scales, we have obtained valuable insights into the intricate interplay between our oceans and coastlines. Through the decomposition of the RSL signal into its constituent components, we gain a deeper comprehension for factors driving these fluctuations. One such driver, which we have investigated in this paper, is dynamic sea level in the North Atlantic region. The data obtained by tide gauges and proxy records have provided us with insights into dynamic sea-level changes over the past 2,000 years along the coastline of the North Atlantic.

Our approach used a statistical model within a Bayesian framework to decompose the RSL signal into four components, achieved through a generalised additive model formed by a combination of random effect terms and splines. This statistical model was an extended version of the NI-GAM as described by Upton et al. (2023c), and accounted for the bi-variate uncertainty associated within the data using the Noisy Input uncertainty method (McHutchon and Rasmussen, 2011). The model possessed four separate components within the process level: common (temporal spline), regional (hierarchical temporal spline), linear-local (random effect) and non-linear local component (spatial temporal spline). In Section 3.7, we present one component of our RSL decomposition: the regional component. The remaining components of the decomposition were recorded in the Appendix 5.7.2. In order to estimate dynamic sea level changes, we carried out our extend NI-GAM on two different datasets, the North Atlantic east west data and north south of Cape Hatteras, North Carolina USA. We demonstrated the successful implementation of the models using 10-fold cross validation as demonstrated in Section 5.7.3. The east west dataset obtained an empirical coverage value of 99.3% while, the north south dataset yield an empirical coverage value of 99.0%. These values are conservative estimates, given the substantial bi-variate uncertainties linked with the proxy records.

We concentrated on the North Atlantic region as it holds important significance for climate regulation for the globe as it carries heat from the subtropics northwards (Talley, 2003; Roemmich and Wunsch, 1985; Biastoch et al., 2008). As the

North Atlantic ocean is characterised by complex and influential ocean dynamics (Marshall et al., 2001), we examined two key components of this ocean circulation system; the AMOC and quasi-horizontal circulation (gyres and surface-enhanced currents). Other factors that impact the North Atlantic ocean, including wind patterns and temperature gradients (Marshall et al., 2001), were not the focus of this paper, yet the interplay of these elements shapes the ocean's behavior and plays a crucial role in global climate systems (Delworth et al., 2017). We utilised the dynamic sea level, estimated using the extended NI-GAM and, by finding the difference between the east-west sub-region of the North Atlantic, we were able to investigate changes in the AMOC. A similar approach using the south-north sub-region allowed us to examine changes in quasi-horizontal circulation. The results from our models are summarised in Section 3.7.

To examine the relationship between the AMOC and dynamic sea level, we used the difference between the eastern and western boundary sea level as described by Little et al. (2019). Figure 5.5.2(a) demonstrated that the dynamic sea-level change on the eastern boundary has fluctuated while, the western boundary experienced minimal variation over the past 2,000 years. The response of sea level due to ocean circulation changes along the eastern boundary is not clear, with some models suggesting that sea level could fall (Little et al., 2019) and others suggesting it could rise (Yin et al., 2009). From our perspective, this observed trend could emanate from an underlying process, or more likely it arises due to an inherent bias in the data distribution on either side of the North Atlantic, which is biased heavily to the western boundary. Consequently, the variability linked to the western boundary exerts a notably greater influence on the common component, leading the regional component that pertains to the western sub-region to exhibit a tendency toward a relatively stable trajectory.

The difference between our east and west regional components and the associated rate of change is presented in Figure 5.5.2. The posterior model fit, shown in Figure 5.5.2 (a), illustrates the dynamic sea level trend's evolution over the past 2,000 years, reaching a peak in 810 CE, followed by a gradual decline until 1770 CE, a subsequent increase until 1900 CE, and a pronounced decline thereafter. However, the considerable uncertainty stemming from the inherent variability in proxy data

complicates the identification of dynamic sea-level changes, underscoring the need for additional data to comprehensively understand the influence of the AMOC on this region. Similarly, the rate of change presented in Figure 5.5.2(b) possesses a pronounced decline since the year 1850 CE, yet, the large uncertainties introduces challenges in identifying the significance of these variations and in turn the potential influence of water transported by the AMOC. As demonstrated in Figure 5.5.3, the rate of this downward trend is consistent with other studies which claim that the AMOC has been weakening (e.g., Rahmstorf et al., 2015; Caesar et al., 2018, 2021). However, our results exhibit large uncertainties due to the substantial uncertainty in the proxy data and the additional component has contributed to increased identifiability challenges. We would require additional data to overcome the North Atlantic data bias and external information to improve prior constraints which could improve the resolution of our results.

The term quasi-horizontal circulation describe a variety of components including gyres and surface enhanced currents, with Cape Hatteras in North Carolina marking the boundary for these components (Woodworth et al., 2011). The dynamic sea level south and north of Cape Hatteras is proportional to transport along the intergyre boundary (McCarthy et al., 2015). Using our extended NI-GAM and proxy RSL reconstructions, we investigated the evolution of the difference between our north and south regional components, along with the associated rate of change, over the past 2,000 years. In Figure 5.5.5, we have highlighted that the water transported over the intergyre boundary has fluctuated dramatically, with a strengthening of southward transport by the North Atlantic subtropical gyre until 802CE followed by weakening trend until 1580CE when the transport increased again. Since 1920, the rate of change of the southward transport has decreased. The observed posterior model fit in Figure 5.5.5(a) reveals an anti-correlation in dynamic sea level changes on either side of Cape Hatteras, consistent with findings in Diabaté et al. (2021); however, the presence of substantial uncertainties hinders definitive confirmation. In providing context to these findings, McCarthy et al. (2015) highlighted the association between dynamic sea-level changes north of Cape Hatteras and fluctuations in overturning circulation (Bingham and Hughes, 2009; Ezer, 2013), and changes south of Cape Hatteras with variations in the Gulf

Stream from Florida to Cape Hatteras. However, due to the limitations in resolution within our results, discerning the specifics of the underlying quasi-horizontal circulation patterns in the northern and southern regions becomes challenging, making it difficult to gain a clear understanding of the contributing factors. The dataset used in this analysis is significantly smaller when compare to the east-west analysis, yet south and north of Cape Hatteras possess similar data balance issues which can impact the temporal range of the models.

Using proxy sea-level reconstructions to inform our understanding of past circulation changes is an important advance, because it provides an independent, orthogonal data constraint on past ocean dynamics. In order to clarify current debates surrounding whether recent ocean circulation changes in the Atlantic which are viewed as extraordinary, our models would require a number of improvements. The first is the resolution of outputs. As previously mentioned in (Upton et al., 2023c), identifiability has been a constant challenge when decomposing the RSL signal. By including the extra component, we have attempted to separate barystatic influences and thermosteric contributions, which are common across all sites, from processes that may influence certain regions differently over time. However, it is difficult to be definite that the remaining variation in the regional component is due to dynamic sea-level changes. Other process like the spatial fingerprint of sea-level change from ice sheets is known to impact large areas such as Greenland, yet, the lack of data makes it difficult to quantify this influences (Long et al., 2012; Coulson et al., 2022). An addition of higher resolution data over larger spatial scales would improve certainty in our results, however, proxy reconstructions are difficult to collect and require numerous data processing procedures (e.g., Marshall, 2015). The outcomes derived from these enhanced models could contextualise present-day observations within the framework of the recent geological history. This approach would enable us to discern whether the alterations occurring in the Atlantic region today are indeed exceptional or follow patterns seen in the past.

In future research, the community would benefit from the development of a comprehensive database which incorporates a range of instrumental and proxy sources. The integration of satellite data would furnish high-resolution observations, thereby enhancing the precision of our model results, particularly for recent time periods

and broader spatial extents. This additional data would require a statistical model that could account for these varied data sources in an efficient manner. One such improvement would to replace our current MCMC algorithm, implemented using JAGS software (Plummer, 2003), with more efficient software package using e.g. variational Bayes or Integrated Nested Laplace Approximations (INLA: Rue et al., 2009). Another model improvement would be to use adaptive penalties for splines (Jullion and Lambert, 2007) or adaptive smoothness.

A final alternative would be to employ a multi-proxy analysis, akin to the approach undertaken by Croke et al. (2021). Their work effectively integrated a diverse range of palaeoclimate data to enhance comprehension of water security planning dynamics. As exemplified in Figure 5.5.3, an abundance of data sources spanning various temporal and spatial scales serve as proxies for investigating alterations in North Atlantic ocean circulation (Sherwood et al., 2011; Rahmstorf et al., 2015; Thibodeau et al., 2018; Thornalley et al., 2018; Spooner et al., 2020). The implementation of a multi-proxy analysis framework would facilitate a profound exploration of the nuances within the AMOC, quasi-horizontal circulation, and other global ocean circulation systems.

5.7 Appendix

5.7.1 Data Sources

This dataset contains proxy records from the coastline of the North Atlantic. The 33 different proxy data sites and the references for each data source can be found in Table 5.7.1. The 77 tide-gauge sites we used are listed in Table 5.7.2 and were obtained from the PSMSL database (Aarup et al., 2006; Holgate et al., 2013) using the *reslr* package.

Site Name	Reference
Barn Island, Connecticut	Donnelly et al. (2004); Gehrels et al. (2020)
Bassin, Magdelen Islands	Barnett et al. (2017)
Bracky Bridge, Ireland	Kirby et al. (2023)
Big River Marsh, Newfoundland	Kemp et al. (2018)
Bilbao, Spain	Leorri et al. (2012)
Cape May Courthouse, New Jersey	Kemp et al. (2013); Cahill et al. (2016)
Cedar Island, North Carolina	Kemp et al. (2011, 2017)
Cheesequake, New Jersey	Walker et al. (2021)
Chezzetcook Inlet, Nova Scotia	Gehrels et al. (2020)
Dungarvan, Ireland	Alvarez et al (In progress)
East River Marsh, Connecticut	Kemp et al. (2015); Stearns et al. (2023)
Fox Hill Marsh, Rhode Island	Stearns et al. (2023)
Leeds Point, New Jersey	Kemp et al. (2013); Cahill et al. (2016)
Hammock River, Connecticut	Van de Plassche (1991); Engelhart and Horton (2012)
Kyle of Tongue, Scotland	Barlow et al. (2014)
Les Sillons, Magdelen Islands	Barnett et al. (2017)
Little Manatee River, Florida	Gerlach et al. (2017)
Loch Laxford, Scotland	Barlow et al. (2014)
Nassau, Florida	Kemp et al. (2014)
Newtown Estuary, Isle of Wight	Long et al. (2014)
Pelham Bay, New York	Kemp et al. (2017); Stearns and Engelhart (2017)
Placentia, Newfoundland	Kemp et al. (2018)
Revere, Massachusetts	Donnelly (2006)
Saint Simeon, Quebec	Barnett et al. (2017)
Sanborn Cove, Maine	Gehrels et al. (2020)
Sand Hill Point, North Carolina	Kemp et al. (2011, 2017)
Sand Point, North Carolina	Kemp et al. (2011, 2017)
Snipe Key, Florida	Khan et al. (2022)
Swan Key, Florida	Khan et al. (2022)
Thurlestone, South West England	Gehrels et al. (2011)
Tump Point, North Carolina	Kemp et al. (2011, 2017)
Urdaibai, Spain	García-Artola et al. (2009); Leorri et al. (2012)
Urdaibai Estuary, Spain	García-Artola et al. (2009)
Vioarholmi, Iceland	Gehrels et al. (2006); Saher et al. (2015)
Wood Island, Massachusetts	Kemp et al. (2011)

Table 5.7.1: We provide the names of all the proxy sites along the coastline of the North Atlantic with are used in the east-west and north-south analysis. For each site we include the reference in the literature to the source of the data.

Site Name	Longitude	Latitude
ARCACHON-EYRAC	-1.16	44.67
ARGENTIA	-54.00	47.30
ATLANTIC CITY	-74.40	39.40
BEAUFORT, NORTH CAROLINA	-76.70	34.70
BEDFORD INSTITUTE	-63.60	44.70
BELLEDUNE	-65.80	47.90
BILBAO	-3.00	43.40
BOSTON	-71.10	42.40
BOUCAU	-1.51	43.53
BOULOGNE	1.58	50.73
BOURNEMOUTH	-1.90	50.70
BREST	-4.49	48.38
CADIZ II	-6.31	36.53
CALAIS	1.87	50.97
CAPE MAY	-75.00	39.00
CASCAIS	-9.42	38.68
CEUTA	-5.32	35.89
CHARLESTON I	-79.92	32.78
CONCARNEAU	-3.91	47.87
CUTLER II	-67.30	44.60
DAYTONA BEACH	-81.00	29.23
DEN HELDER	4.75	52.96
DEVONPORT	146.36	-41.18
DIEPPE	1.08	49.93
DUBLIN	-6.20	53.40
DUBLIN	-6.22	53.35
DUNKERQUE	2.37	51.05
FERNANDINA BEACH	-81.50	30.70
FORT MYERS	-81.87	26.65
FORT PULASKI	-80.90	32.03
GIBRALTAR	-5.36	36.15
HALIFAX	-63.60	44.70
HARLINGEN	5.41	53.18
HEYSHAM	-2.92	54.03
HOEK VAN HOLLAND	4.12	51.98
IJMUIDEN	4.55	52.46
ILFRACOMBE	-4.10	51.20
KEY WEST	-81.80	24.60
KINLOCHBERVIE	-5.10	58.50
LA CORUÑA II	-8.40	43.36
LA ROCHELLE-LA PALLICE	-1.22	46.16
LARK HARBOUR	-58.40	49.10
LE CONQUET	-4.78	48.36
LE HAVRE	0.11	49.48
LEIXOES	-8.70	41.18
LERWICK	-1.14	60.15
LES SABLES D OLONNE	-1.79	46.50
MAASSLUIS	4.25	51.92
MALIN HEAD	-7.30	55.40
MILLPORT	-4.91	55.75
NAPLES	-81.81	26.13
NEW LONDON	-72.10	41.40
NEW ROCHELLE	-73.80	40.90
NEWLYN	-5.54	50.10
NEWPORT	-71.30	41.50
NIEUWPOORT	2.73	51.15
NORTH SHIELDS	-1.44	55.01
OOSTENDE	2.92	51.23
OREGON INLET MARINA, NORTH CAROLINA	-75.50	35.80
PORT BLOC	-1.06	45.57
REYKJAVIK	-21.90	64.20
ROSCOFF	-3.97	48.72
RUSTICO	-63.30	46.50
SANDY HOOK	-74.00	40.50
SHEERNESS	0.74	51.45
SOUTHEND	0.72	51.51
SPRINGMAID PIER	-78.92	33.66
ST JEAN DE LUZ (SOCOA)	-1.68	43.40
ST. NAZAIRE	-2.20	47.27
TORSHAVN	-6.77	62.02
TOWER PIER	0.08	51.50
VIGO	-8.73	42.24
VIRGINIA KEY, FL	-80.20	25.70
VLISSINGEN	3.60	51.44
WEST-TERSCHELLING	5.22	53.36
WILMINGTON	-77.95	34.23
ZEEBRUGGE	3.20	51.35

Table 5.7.2: We provide the names of all the tide-gauge sites along the coastline of the North Atlantic with are used in the east-west and north-south analysis. For each site we include the longitude and latitude of the site as sourced from the online PSMSL database (Aarup et al., 2006; Holgate et al., 2013).

5.7.1.1 East and West of the North Atlantic

We present all the proxy data sites along the coastline of the North Atlantic. We identify the east coast as all sites in Europe and Iceland, as shown previously in Figure 5.3.2, which contains 10 proxy sites and 60 tide-gauge sites. The west side of the North Atlantic contains all the North American sites as shown in Figure 5.7.1 which contains 23 proxy sites and 51 tide-gauges sites. In Figures 5.3.2 and 5.7.1, the proxy sea level record features stratigraphically-ordered age and RSL data points with 1 sigma uncertainties for both, where vertical uncertainty, varying across proxies and influenced by tidal amplitude, reflects measurement precision (Barlow et al., 2013). Age uncertainty, inherent to the dating method (typically radiocarbon dating), fluctuates due to atmospheric radiocarbon concentration (Edwards, 2007). The boxes represent these uncertainties, varying with the core sample, and the dots mark their midpoints, with larger boxes indicating greater associated uncertainties for specific core samples.

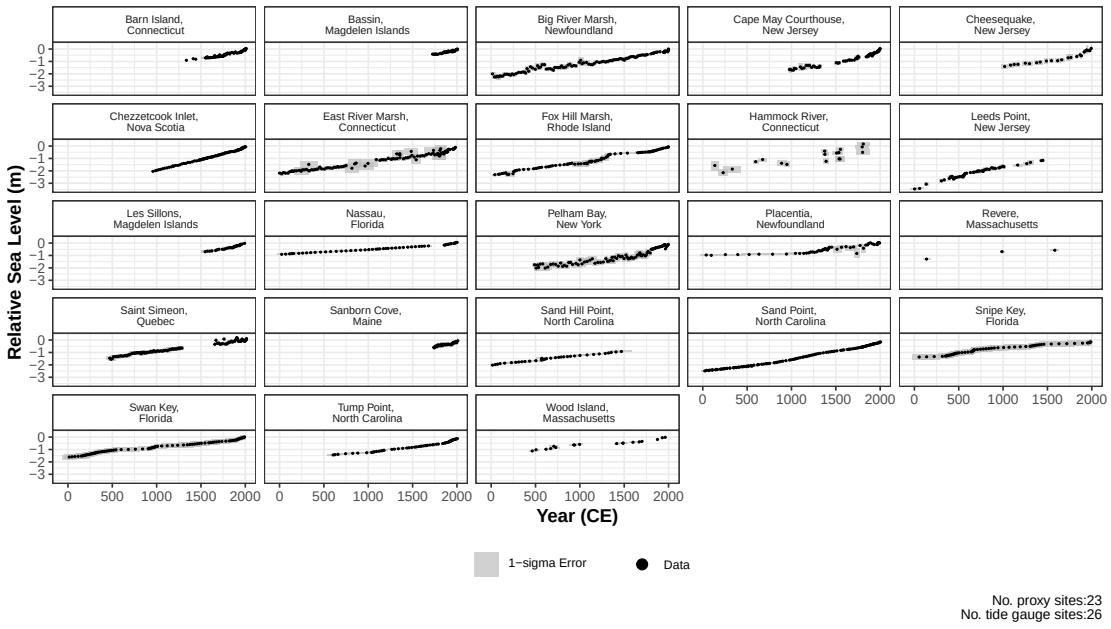


Figure 5.7.1: Proxy records for the west coast sites along the North Atlantic coastline. The y-axis is relative sea level (RSL) in metres and the x-axis is the years in units of Common Era. The black dot represents the midpoint of the proxy sea-level reconstruction and the grey boxes of 1 standard deviation represent vertical and horizontal (temporal) uncertainty.

5.7.1.2 North and South of Cape Hatteras, North Carolina, USA

In this analysis, we examine sites along the North American Atlantic coast and investigate what happens north and south of Cape Hatteras (Longitude: -75.5288 and Latitude: 35.2505). Figure 5.7.2 demonstrates the raw data from proxy records for sites north of Cape Hatteras. There are 18 proxy sites and 15 tide gauges sites for the north of Cape Hatteras region. Figure 5.7.3 demonstrates the raw data from proxy records for sites south of Cape Hatteras. There are 5 proxy sites and 11 tide gauges sites for the south of Cape Hatteras region. As mentioned previously, Figures 5.7.2 and 5.7.3, present the proxy sea level record which contains stratigraphically-ordered age and RSL data points with corresponding 1 sigma uncertainties for both. The vertical uncertainty, influenced by tidal amplitude and varying across proxies, denotes measurement precision (Barlow et al., 2013) and the age uncertainty, intrinsic to the dating method (typically radiocarbon dating), fluctuates due to atmospheric radiocarbon concentration (Edwards, 2007). Representing these uncertainties, the varying-sized boxes, contingent on core samples, are marked with dots at their midpoints, where larger boxes indicate greater associated uncertainties for specific core samples.

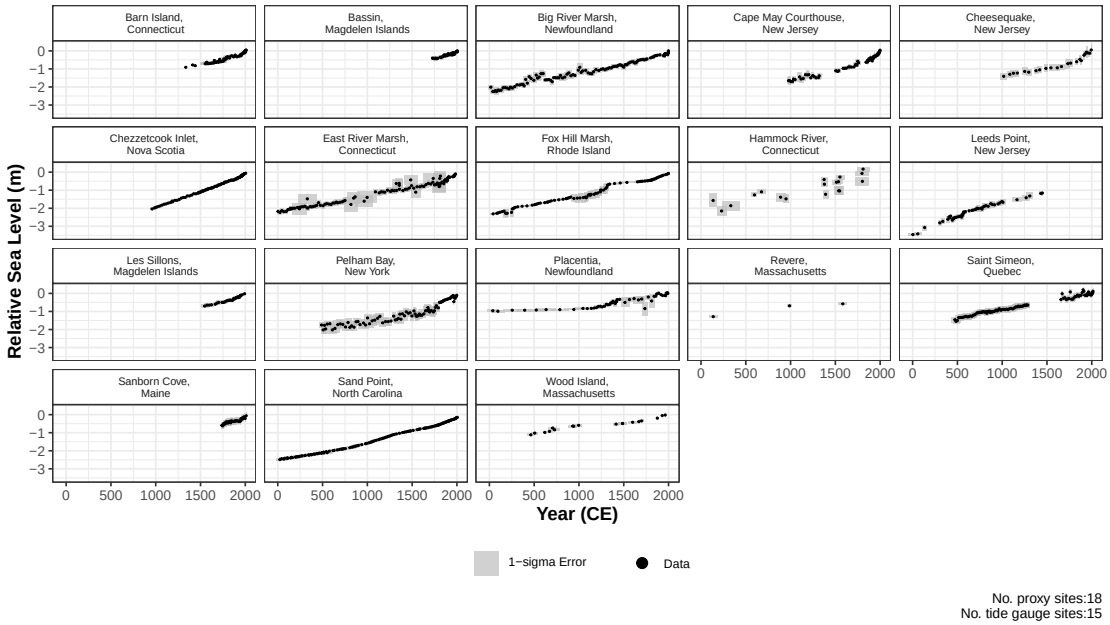


Figure 5.7.2: Raw data from proxy record sites that are located north of Cape Hatteras.The y-axis is relative sea level (RSL) in metres and the x-axis is the years in units of Common Era. The black dot represents the midpoint of the proxy sea-level reconstruction and the grey boxes of 1 standard deviation represent vertical and horizontal (temporal) uncertainty.

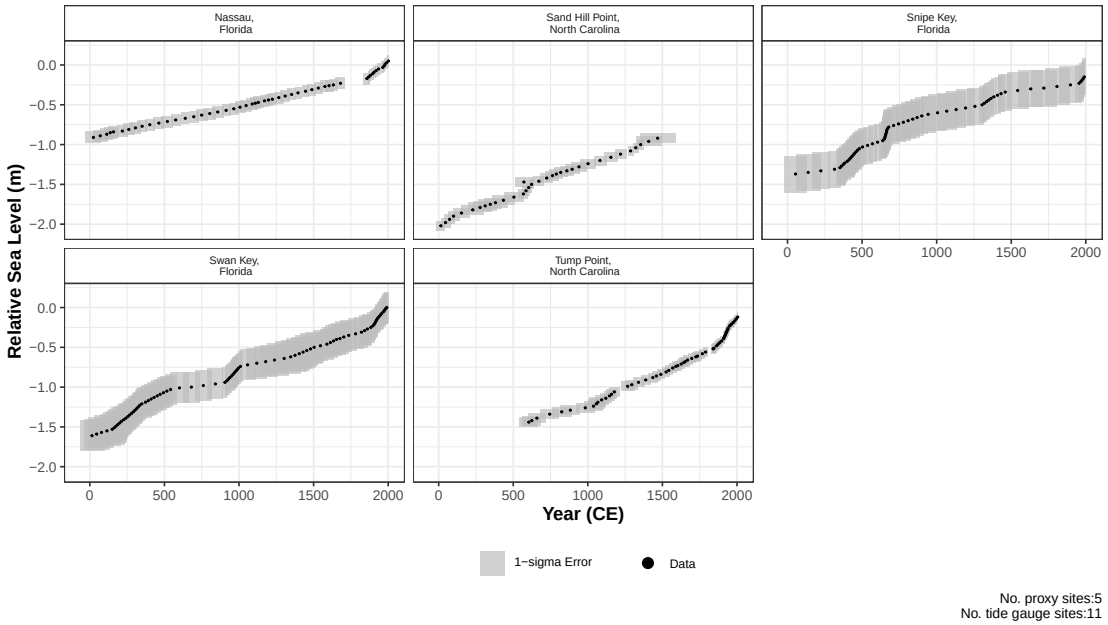


Figure 5.7.3: Raw data from proxy record sites that are located south of Cape Hatteras. The y-axis is relative sea level (RSL) in metres and the x-axis is the years in units of Common Era. The black dot represents the midpoint of the proxy sea-level reconstruction and the grey boxes of 1 standard deviation represent vertical and horizontal (temporal) uncertainty.

5.7.2 Decomposition of the RSL signal

In this section, we present the plots from the decomposition of the RSL signal using the extended NI-GAM for the East-West North Atlantic dataset and the North-South Cape Hatteras dataset.

5.7.2.1 East West North Atlantic

In Figure 5.7.4 we present the common component (a) and its associated rate of change (b) for 33 proxy sites and 111 tide gauge locations in the North Atlantic region. Figure 5.7.5 demonstrates the total model fit for each proxy site using the extended NI-GAM for the coastline of the North Atlantic. Figure 5.7.6 demonstrates the rate of change for the total model fit for each proxy site using the extended NI-GAM for the coastline of the North Atlantic. In Figure 5.7.7 we present the linear local component for our sites along the coastline of the North Atlantic. Figure 5.7.8 represents the non-linear local component for the 37 proxy

sites along the coastline of the North Atlantic and Figure 5.7.9 represents the corresponding rate of change for this component.

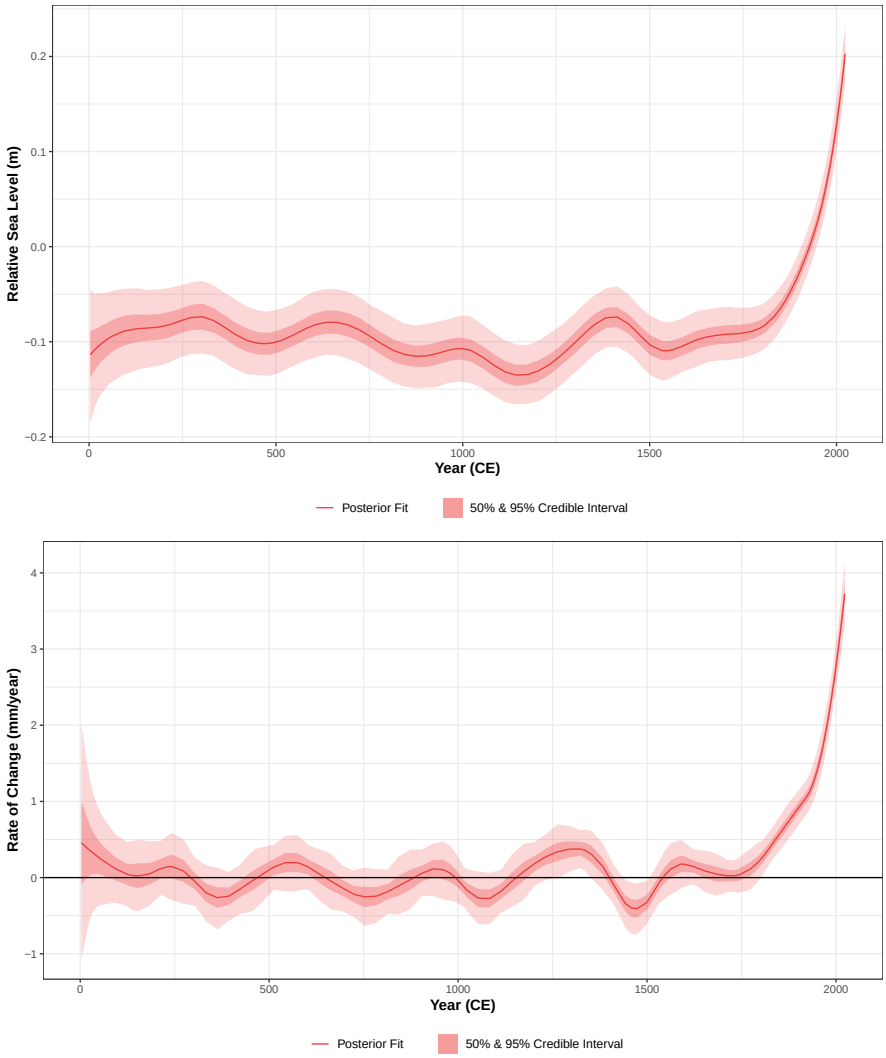


Figure 5.7.4: The extended noisy-input generalised additive model (NI-GAM) results for (a) the common component and (b) the common rate of change component. (a) The common component mean model fit represented with a solid line and the shading indicating the 50% and 95% credible intervals in light red. The y-axis is the sea level in m with the x-axis representing the time across the last 2,000 years for the North Atlantic region. (b) Rate of change for the common component for the North Atlantic region with the solid line representing the mean of the fit, the light red shaded area representing the 50% and 95% credible intervals. The y-axis is the instantaneous rate of change of sea level in mm per year.

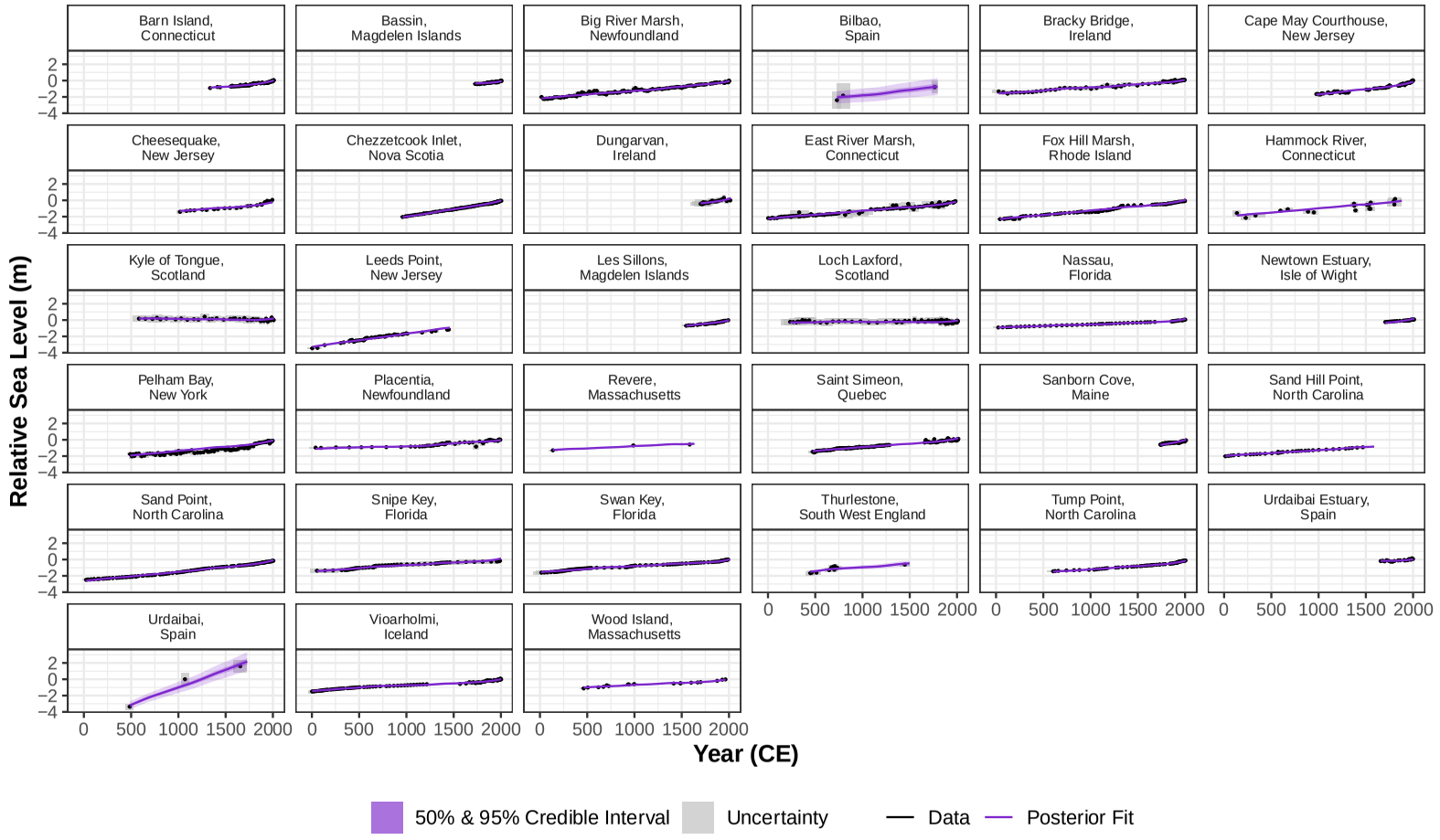


Figure 5.7.5: The extended noisy-input generalised additive model (NI-GAM) fit for 37 proxy sites along the coastline of the North Atlantic. The black dots and grey boxes represent the midpoint and associated uncertainty, respectively, for each proxy record. The solid purple line represents the mean of the model fit with a 50% and 95% credible intervals denoted by shading.

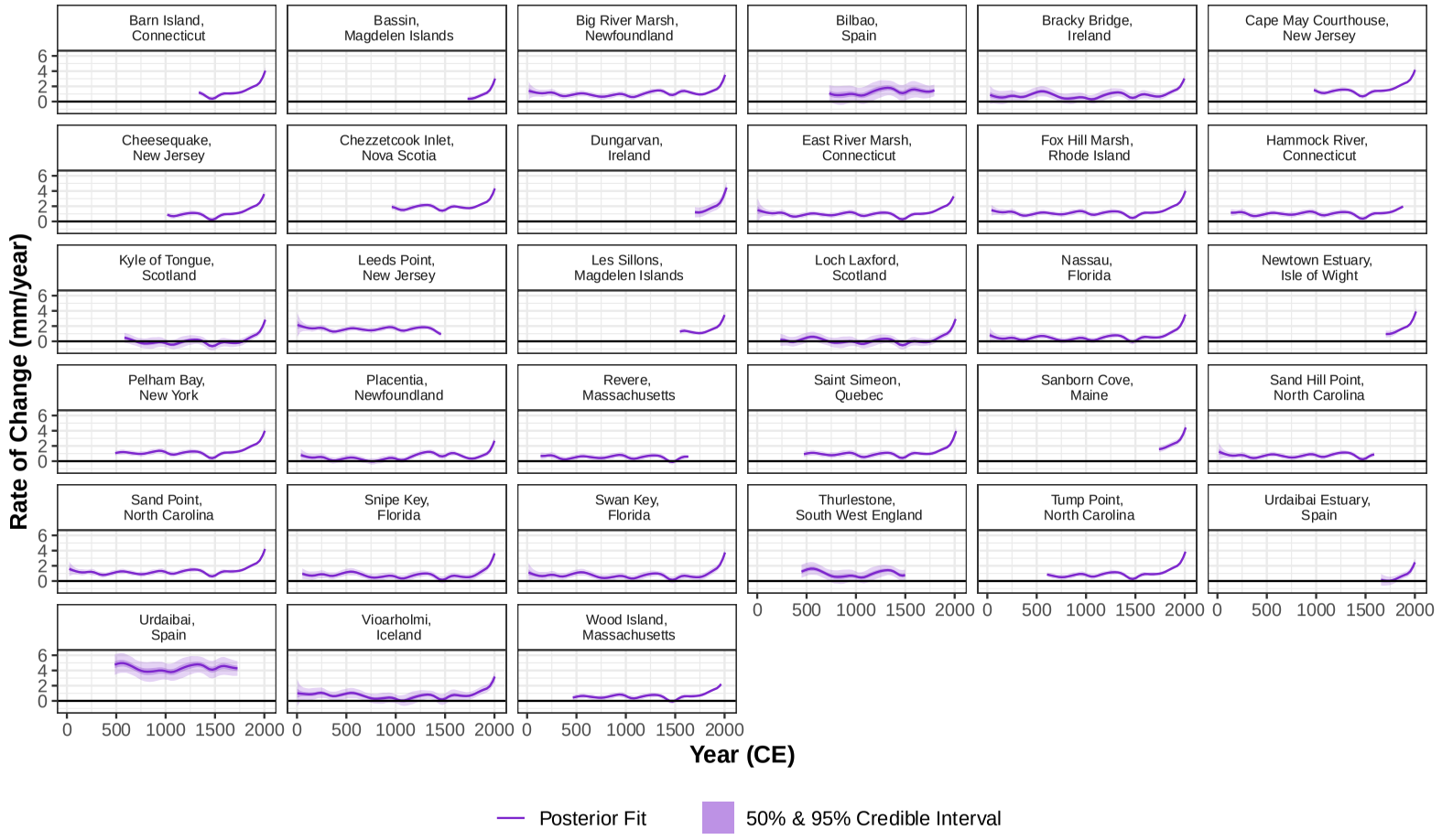


Figure 5.7.6: Rate of relative sea change found by taking the first derivative of the total model fit for 37 proxy sites along the coastline of the North Atlantic. The mean of the fit is the solid purple line with the light shaded area being the 50% and 95% credible intervals.

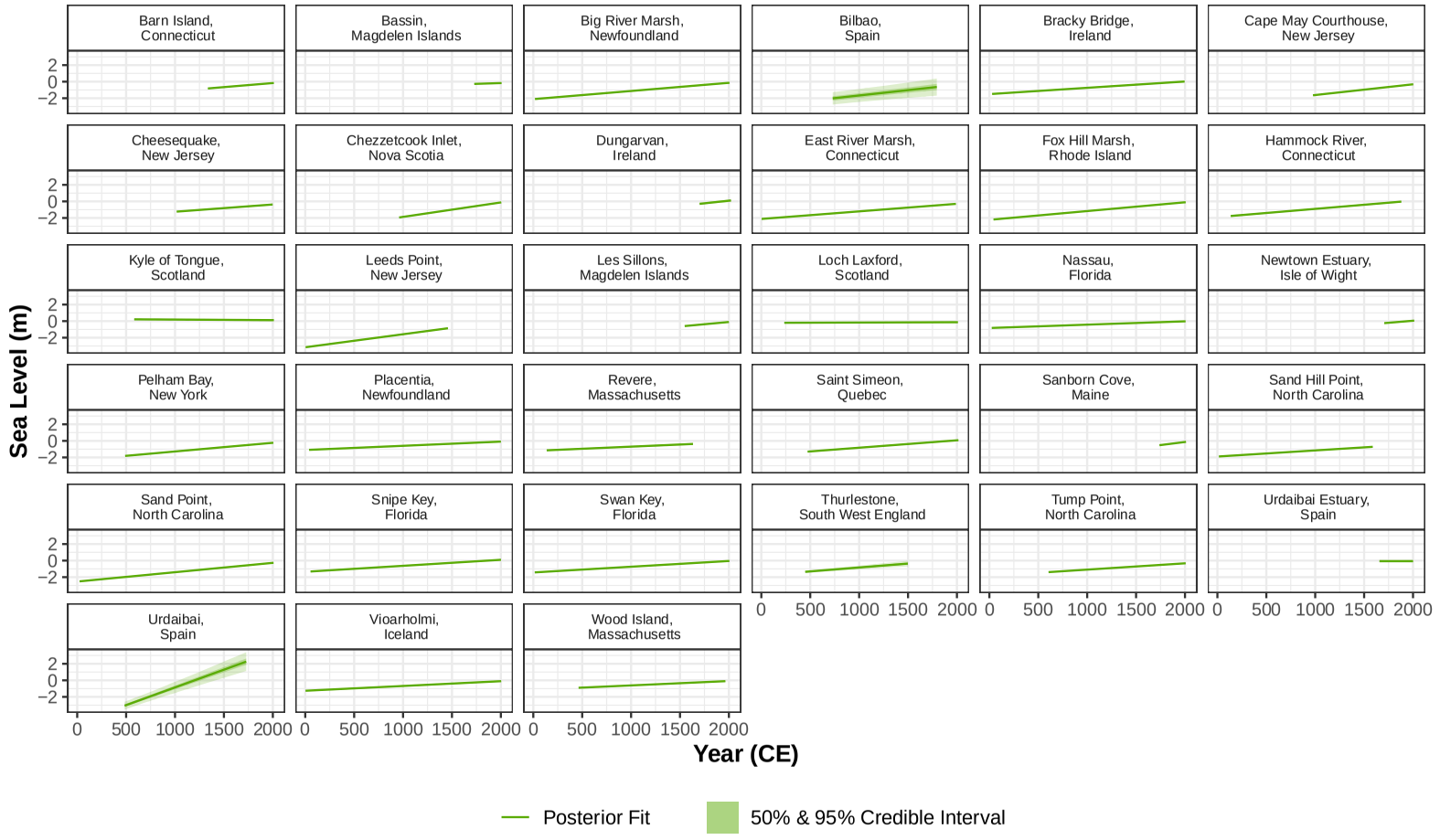


Figure 5.7.7: The linear local component of the extended noisy-input generalised additive model (NI-GAM) for 37 proxy sites along the coastline of the North Atlantic. The solid green line represents the mean of the model fit with a 50% and 95% credible intervals denoted by shading.

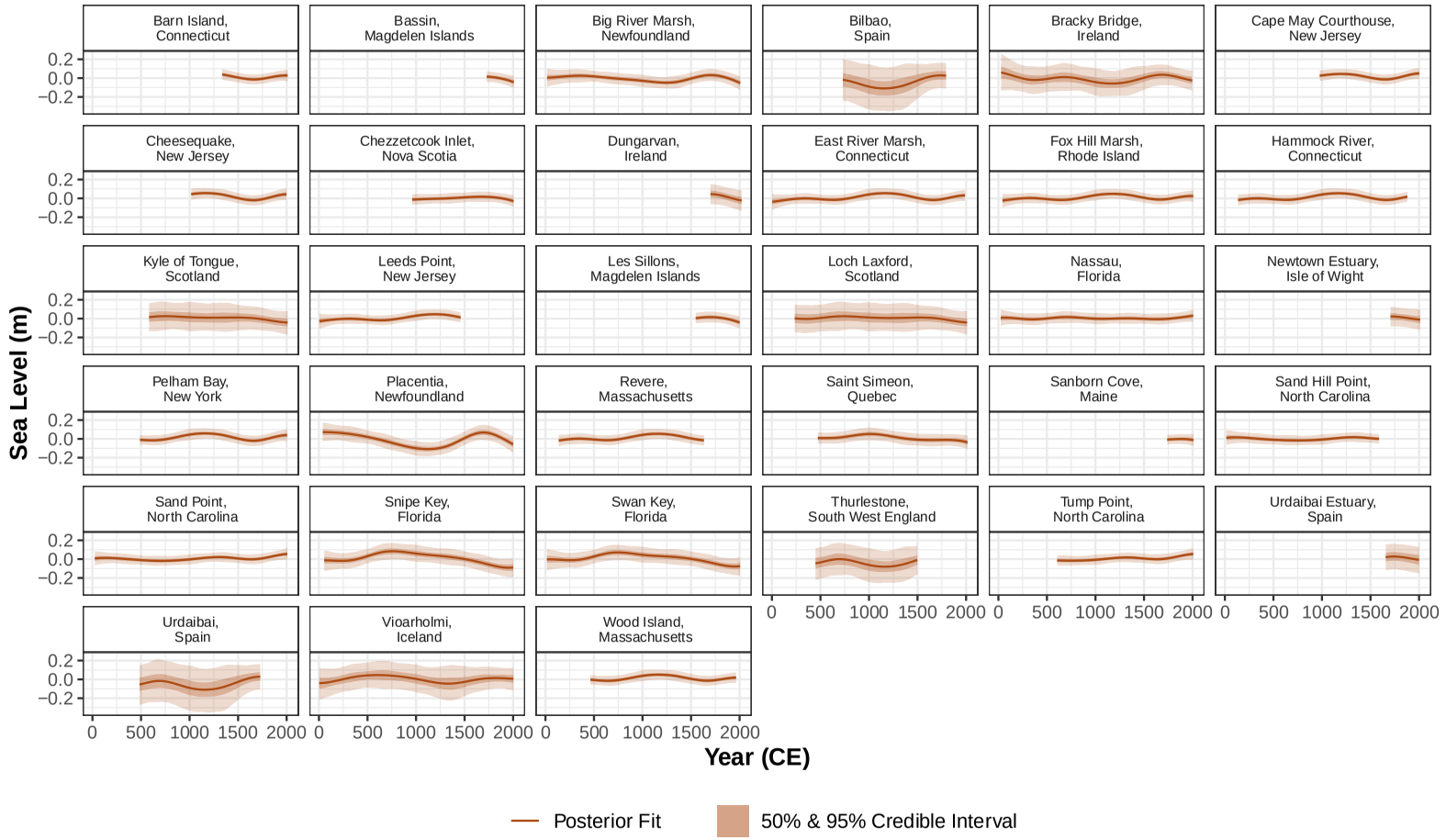


Figure 5.7.8: The non-linear local component of the extended noisy-input generalised additive model (NI-GAM) for 37 proxy sites along the coastline of the North Atlantic. The solid brown line represents the mean of the model fit with a 50% and 95% credible intervals denoted by shading.

Figure 5.7.9: The rate of change for the non-linear local component of the extended noisy-input generalised additive model (NI-GAM) for 37 proxy sites along the coastline of the North Atlantic. The solid brown line represents the rate of change of the model fit with a 50% and 95% credible intervals denoted by shading.

5.7.2.2 North South Cape Hatteras

In Figure 5.7.10 we present the common component (a) and its associated rate of change (b) for the Atlantic coast of North America using 23 proxy sites and 26 tide gauge sites. Figure 5.7.11 demonstrates the total model fit for each proxy site using the extended NI-GAM for the coastline of the North Atlantic. Figure 5.7.12 demonstrates the rate of change for the total model fit for each proxy site using the extended NI-GAM for the coastline of the North Atlantic. In Figure 5.7.13 we present the linear local component for our sites along the coastline of the North Atlantic. Figure 5.7.14 represents the non-linear local component for the 37 proxy sites along the coastline of the North Atlantic and Figure 5.7.15 represents the corresponding rate of change for this component.

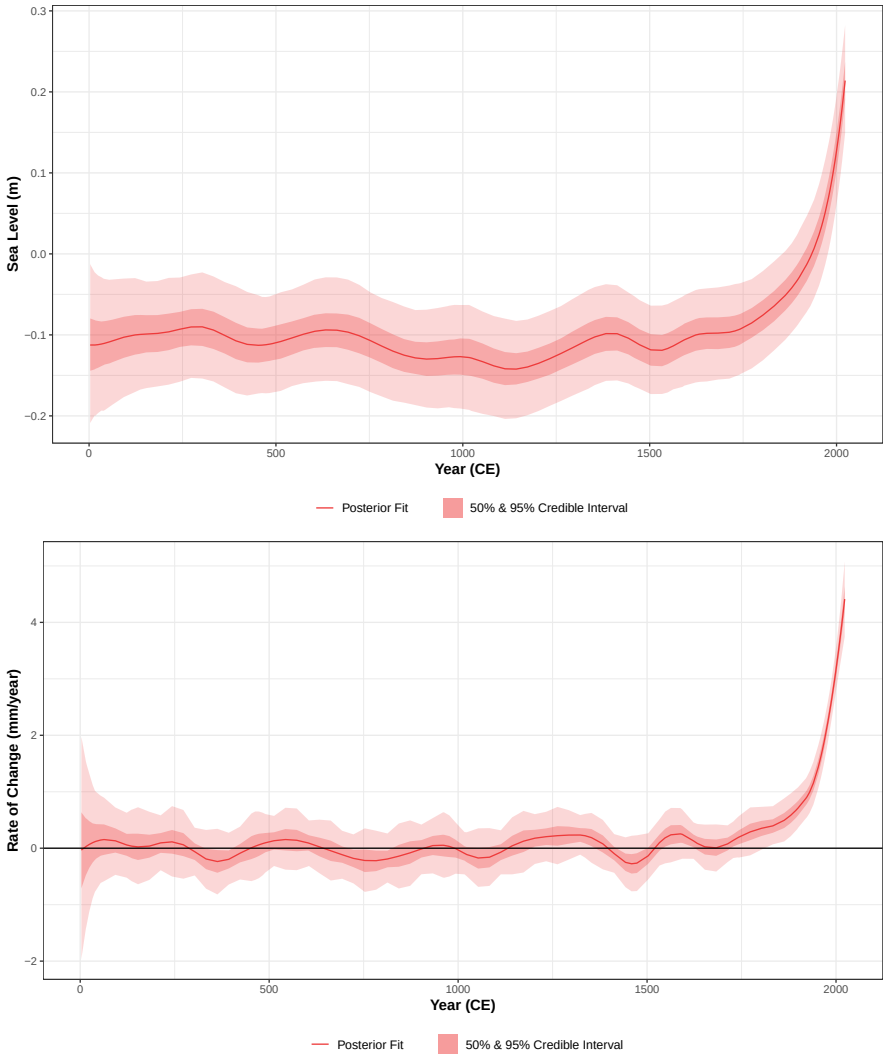


Figure 5.7.10: The extended noisy-input generalised additive model (NI-GAM) results for (a) the common component and (b) the common rate of change component for the Atlantic coast of North America. (a) The common component mean model fit represented with a solid line and the shading indicating the 50% and 95% credible intervals. The y-axis is the sea level in m with the x-axis representing the time across the last 3000 years for the North Atlantic region. (b) Rate of change for the common component for the North Atlantic region with the solid line representing the mean of the fit, the light red shaded area representing the 50% and 95% credible intervals. The y-axis is the instantaneous rate of change of sea level in mm per year.

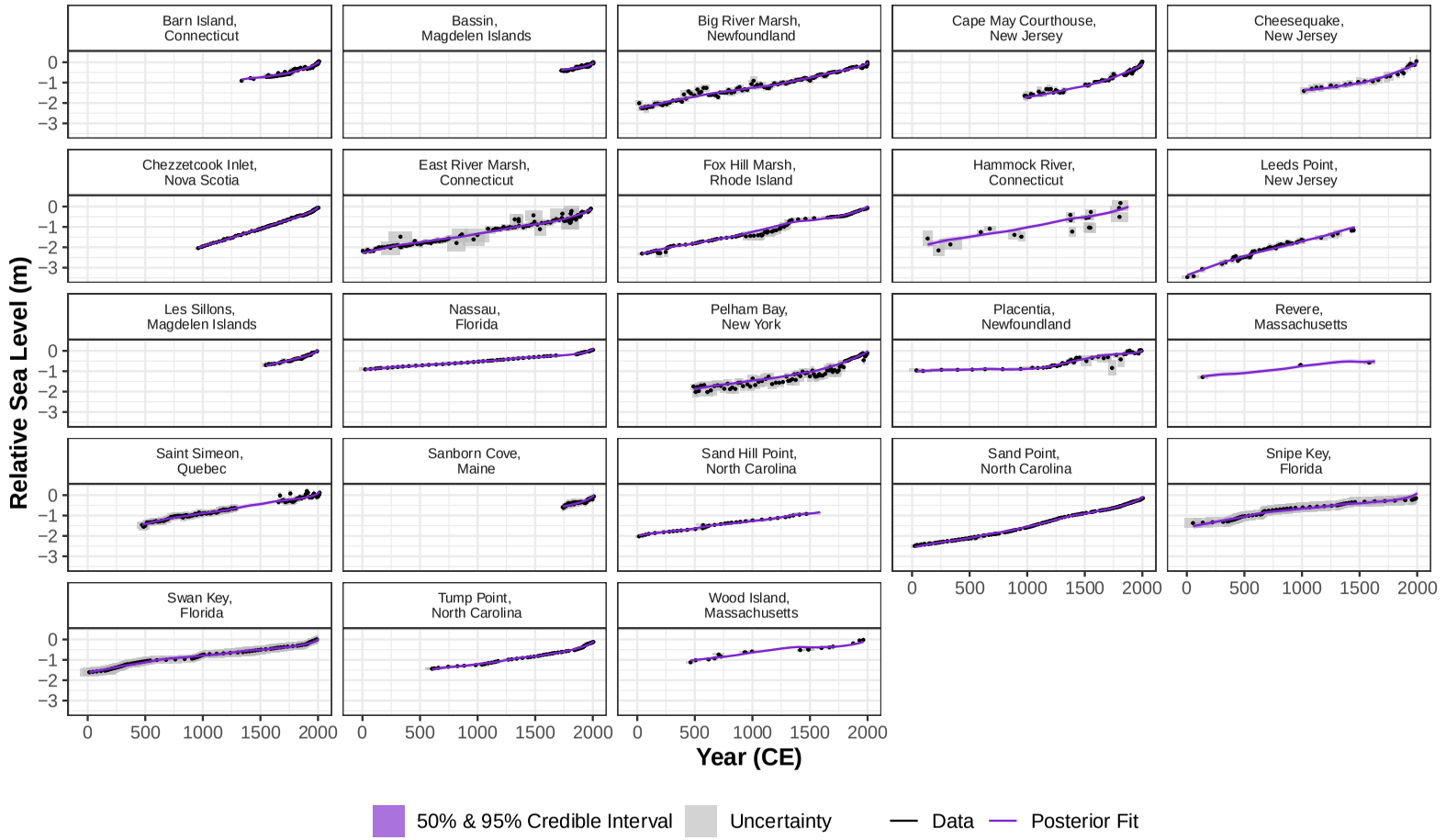


Figure 5.7.11: The extended noisy-input generalised additive model (NI-GAM) fit for the proxy sites along the Atlantic coastline of North America. The black dots and grey boxes represent the midpoint and associated uncertainty, respectively, for each proxy record. The solid purple line represents the mean of the model fit with a 50% and 95% credible intervals denoted by shading.

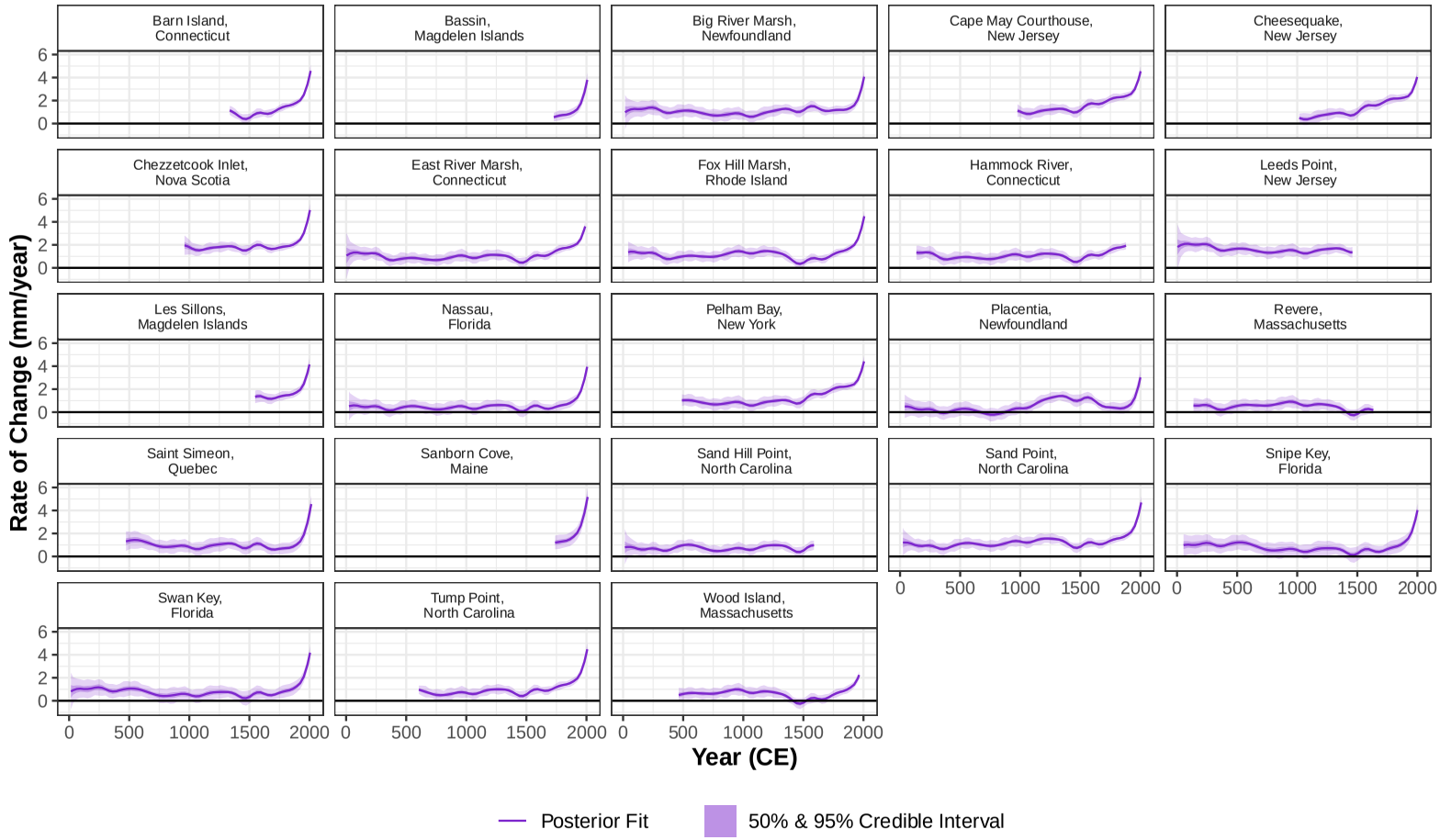


Figure 5.7.12: Rate of relative sea change found by taking the first derivative of the total model fit for the proxy sites along the Atlantic coastline of North America. The mean of the fit is the solid purple line with the light shaded area being the 50% and 95% credible intervals.

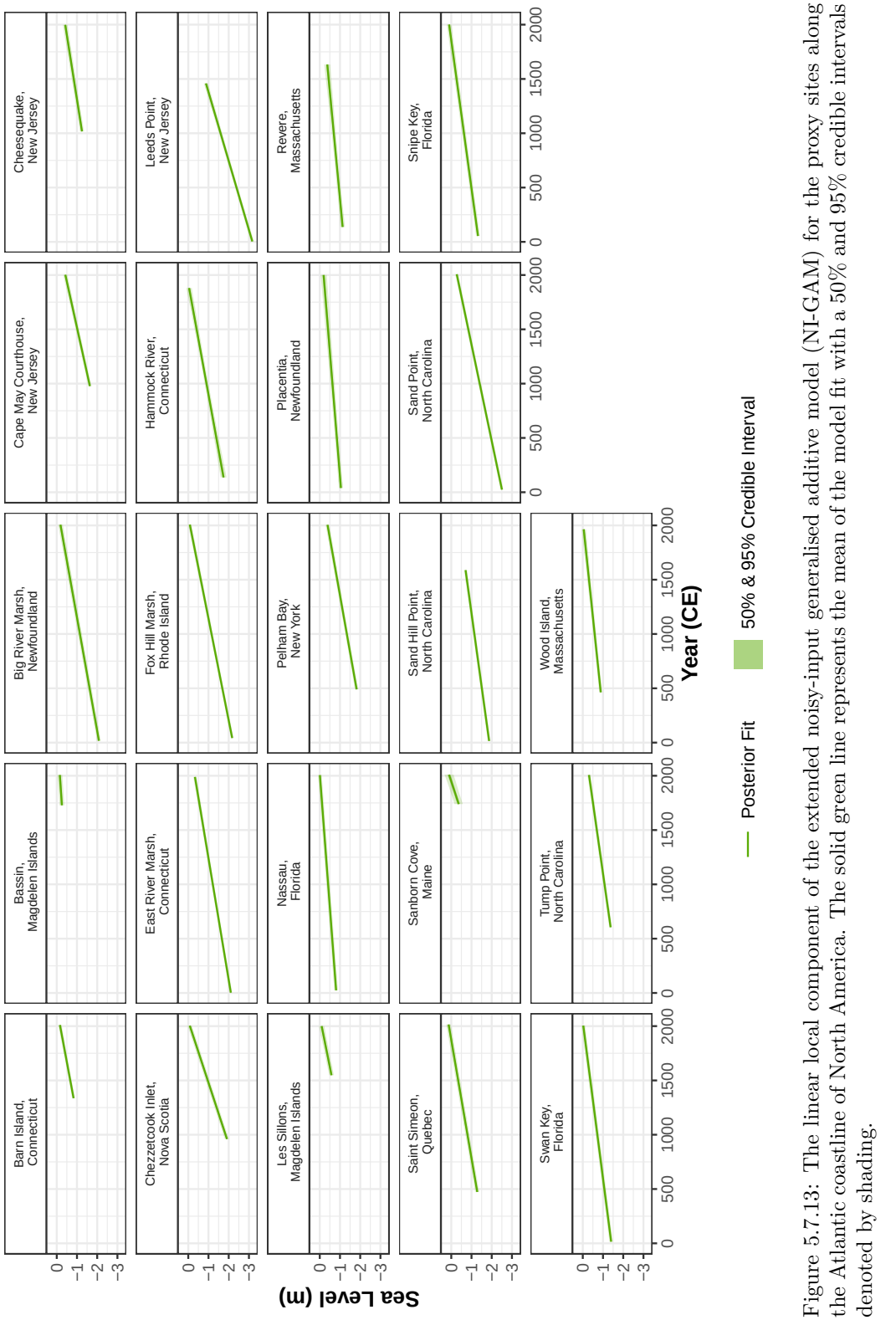


Figure 5.7.13: The linear local component of the extended noisy-input generalised additive model (NI-GAM) for the proxy sites along the Atlantic coastline of North America. The solid green line represents the mean of the model fit with a 50% and 95% credible intervals denoted by shading.

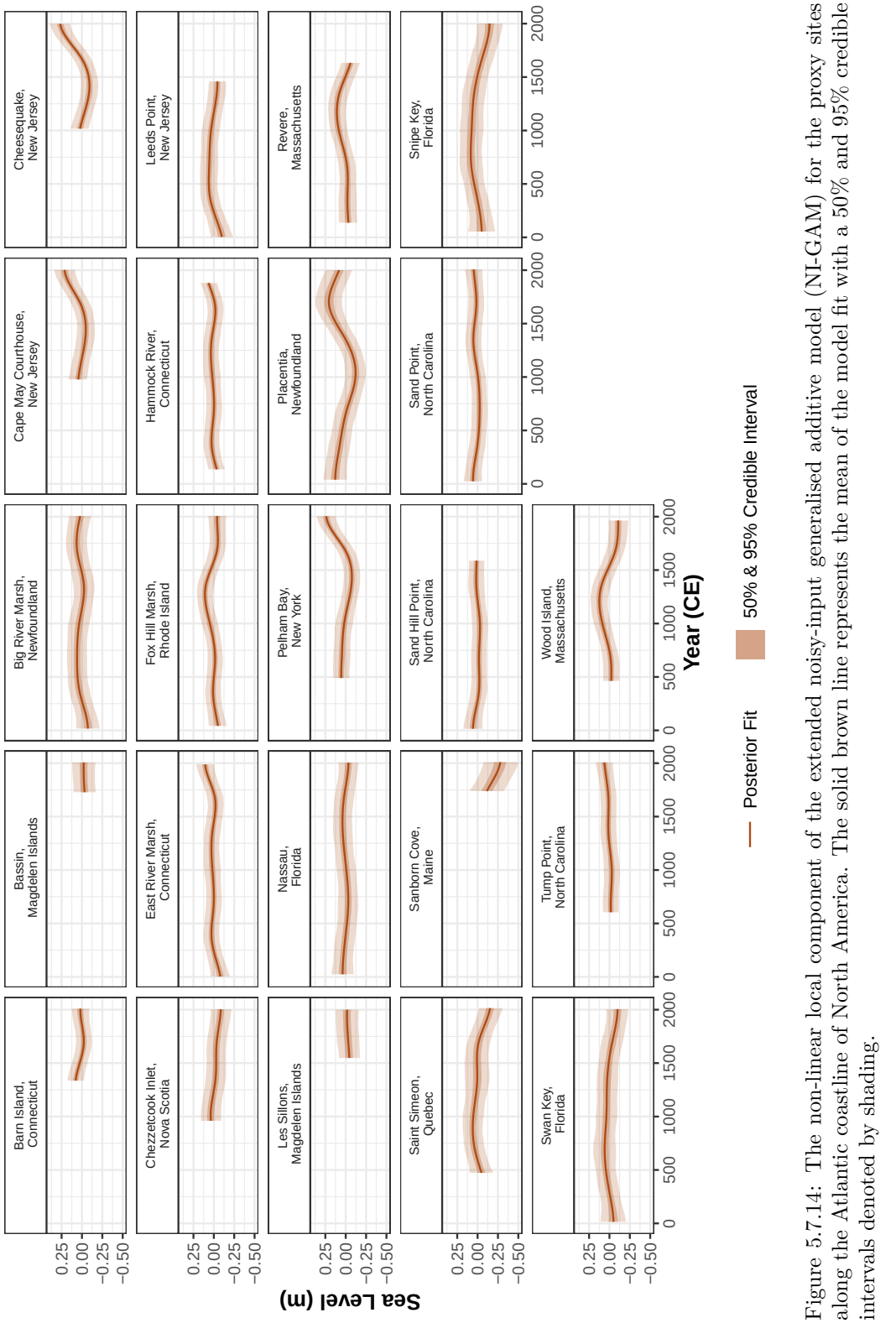


Figure 5.7.14: The non-linear local component of the extended noisy-input generalised additive model (NI-GAM) for the proxy sites along the Atlantic coastline of North America. The solid brown line represents the mean of the model fit with a 50% and 95% credible intervals denoted by shading.

Figure 5.7.15: The rate of change for the non-linear local component of the extended noisy-input generalised additive model (NI-GAM) for the proxy sites along the Atlantic coastline of North America. The solid brown line represents the rate of change of the model fit with a 50% and 95% credible intervals denoted by shading.

5.7.3 Model Validation

In this section we present the model validation tests which were carried out for both the east-west and north-south analysis. In both cases, we used 10-fold cross-validation (10-CV) the proxy sites and tide gauge records. We present the site-specific results for both east-west and north-south analysis and we evaluated the model performance based on out of sample empirical coverage and the Root Mean Squared Error (RMSE). The prediction intervals were created using posterior predictive simulations with the full error structure as described in Upton et al. (2023c). As a result, the size of the prediction intervals vary by data point which corresponds to the large and changeable bi-variate uncertainty associated with the core sample used in the proxy records as demonstrated in Figures 5.3.2, 5.7.1, 5.7.2 and 5.7.3. The empirical coverage provides the percentage of occasions that the true RSL observation is within the model prediction interval (PI) for RSL. The RMSE provides insight into prediction performance in the same units as the response (meters).

5.7.3.1 East West North Atlantic

For the east west analysis, the overall empirical coverage value is 99.3% and is a satisfactory result for our model fit as the input data which possess bivariate uncertainties. The conservative coverage values likely stems from taking into account the observed measurement errors when estimating the prediction intervals. The overall RSME for the east west analysis is 0.28m which is reasonable given the scale and variability of the data set. Figure 5.7.16 presents the true RSL observations versus the model-based RSL point estimates with 95% prediction intervals at each site. It is evident that locations like Loch Laxford and Kyle of Tongue in Scotland exhibit substantial prediction intervals, indicative of the inherent uncertainty variability in the proxy records associated with these sites. Furthermore, these particular sites display a heightened sensitivity to the physical process of GIA, significantly influencing the resulting linear local component of the NI-GAM. Table 5.7.3 provides a site-specific insight into the empirical coverage for the model, the RSME and the size of the prediction intervals. Many of our sites have a coverage of 100% due to the large prediction intervals arising from the

bivariate uncertainties associated with the proxy data.

Site Name	Prediction Interval Width (m)	Empirical Coverage	RSME(m)
Barn Island, Connecticut	0.5262	1.0000	0.0577
Bassin, Magdelen Islands	0.5969	1.0000	0.0164
Big River Marsh, Newfoundland	0.5742	0.9909	0.0555
Bracky Bridge, Ireland	0.8932	1.0000	0.0454
Cape May Courthouse, New Jersey	0.6796	1.0000	0.1842
Cheesequake, New Jersey	0.9105	1.0000	0.1053
Chezzetcook Inlet, Nova Scotia	0.4752	1.0000	0.0809
Dungarvan, Ireland	1.2180	1.0000	0.0130
East River Marsh, Connecticut	0.7126	1.0000	0.0106
Fox Hill Marsh, Rhode Island	0.5722	0.9438	0.2852
Hammock River, Connecticut	0.8497	0.7778	0.4008
Kyle of Tongue, Scotland	1.4575	1.0000	0.0313
Leeds Point, New Jersey	0.6071	0.9600	0.0747
Les Sillons, Magdelen Islands	0.5783	1.0000	0.0074
Loch Laxford, Scotland	1.4541	1.0000	0.0056
Nassau, Florida	0.4622	1.0000	0.0266
Newtown Estuary, Isle of Wight	0.7659	1.0000	0.0438
Pelham Bay, New York	0.8068	1.0000	0.8838
Placentia, Newfoundland	0.5605	0.9804	0.0300
Saint Simeon, Quebec	0.7350	1.0000	0.1369
Sanborn Cove, Maine	0.8456	1.0000	0.0491
Sand Hill Point, North Carolina	0.4011	1.0000	0.0098
Sand Point, North Carolina	0.4695	1.0000	0.0263
Snipe Key, Florida	0.9682	1.0000	0.1221
Swan Key, Florida	0.8381	1.0000	0.0057
Tump Point, North Carolina	0.4495	1.0000	0.0313
Urdaibai Estuary, Spain	0.7558	1.0000	0.0287
Vioarholmi, Iceland	0.8642	1.0000	0.0140
Wood Island, Massachusetts	0.4211	1.0000	0.0110
BREST	0.4932	1.0000	0.0022
HARLINGEN	0.5650	1.0000	0.0089
IJMUIDEN	0.5524	1.0000	0.0071
DEN HELDER	0.5426	1.0000	0.0106
SHEERNESS	0.5310	1.0000	0.0285
MAASSLUIS	0.5336	1.0000	0.0189
HOEK VAN HOLLAND	0.5513	1.0000	0.0359
NEWLYN	0.5959	1.0000	0.0001
DUBLIN	0.6037	1.0000	0.0046
VLISSINGEN	0.5412	1.0000	0.0002
CASCAIS	0.5272	1.0000	0.0023
NORTH SHIELDS	0.5768	1.0000	0.0295
ATLANTIC CITY	0.6133	1.0000	0.0181
HALIFAX	0.5687	1.0000	0.0148
FERNANDINA BEACH	0.5810	1.0000	0.0034
KEY WEST	0.6138	1.0000	0.0195

Table 5.7.3: The empirical coverage, prediction interval width and Root Mean Square Error (RMSE) by site for our east west analysis.

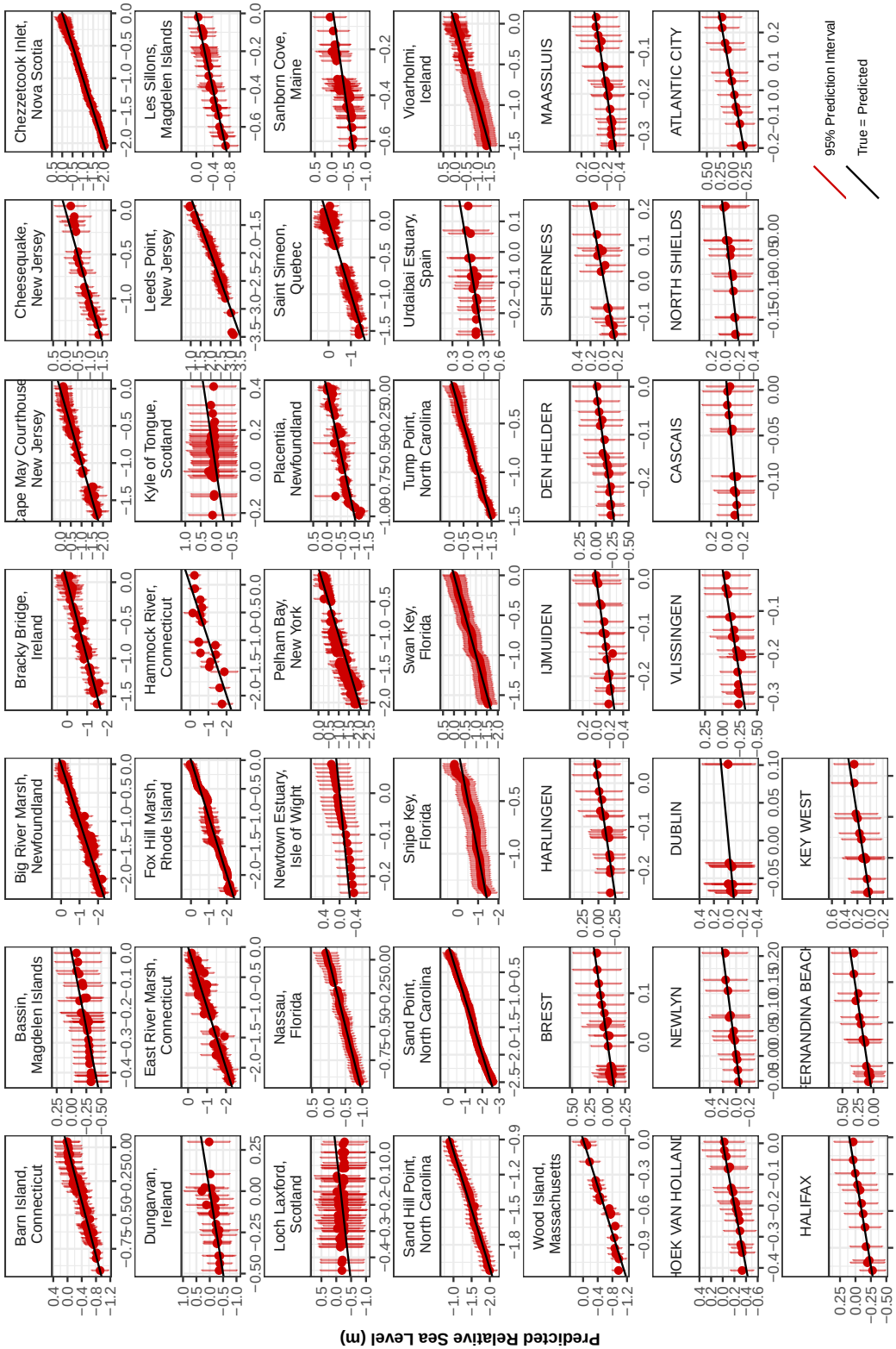


Figure 5.7.16: True versus predicted relative sea level for east west analysis using 10-fold cross validation (CV) using proxy and tide gauge data along the coastlines of the North Atlantic. The predicted means are the red points with a vertical 95% prediction interval for each point. The identity line is shown in black.

5.7.3.2 North South Cape Hatteras

In the context of the north-south analysis, the overall empirical coverage stands at 99.0%, which is a satisfactory outcome considering the presence of bivariate uncertainties in the input data. The cautious coverage values likely arise from factoring in the observed measurement errors when estimating prediction intervals. The overall root mean square error (RSME) for the east-west analysis is 0.41m, a reasonable value given the scale and variability of the dataset. Figure 5.7.17 illustrates the comparison between actual relative sea level (RSL) observations and model-derived RSL point estimates, accompanied by 95% prediction intervals at each site. Evidently, Snipe Key in Florida and Swan Key in Florida exhibit wide prediction intervals, underscoring the substantial bi-variate uncertainties inherent in the proxy data. While, locations like Hammock River in Connecticut and Wood Island in Massachusetts demonstrate comparatively sparse data in relation to the other sites. Additionally, Table 5.7.4 provides a site-specific breakdown of empirical coverage, RSME values, and prediction interval sizes for the model. It's notable that numerous sites exhibit 100% coverage due to the wide prediction intervals resulting from the bivariate uncertainties associated with the proxy data.

Site Name	Prediction Interval Width	Empirical Coverage	RSME
Barn Island, Connecticut	0.3548	1.0000	0.0510
Bassin, Magdelen Islands	0.4311	1.0000	0.0327
Big River Marsh, Newfoundland	0.4852	0.9818	0.0609
Cape May Courthouse, New Jersey	0.5549	1.0000	0.1545
Cheesequake, New Jersey	0.8359	1.0000	0.0216
Chezzetcook Inlet, Nova Scotia	0.3006	1.0000	0.0028
East River Marsh, Connecticut	0.6377	1.0000	0.0041
Fox Hill Marsh, Rhode Island	0.4497	0.9551	0.3918
Hammock River, Connecticut	0.7975	0.7222	0.4568
Leeds Point, New Jersey	0.5416	0.9600	0.0355
Les Sillons, Magdelen Islands	0.4170	1.0000	0.0174
Nassau, Florida	0.3401	1.0000	0.0043
Pelham Bay, New York	0.7176	1.0000	1.0125
Placentia, Newfoundland	0.4476	1.0000	0.0938
Saint Simeon, Quebec	0.6525	1.0000	0.0593
Sanborn Cove, Maine	0.7453	1.0000	0.0031
Sand Hill Point, North Carolina	0.2902	1.0000	0.0043
Sand Point, North Carolina	0.3600	1.0000	0.0212
Snipe Key, Florida	0.9317	1.0000	0.1209
Swan Key, Florida	0.7808	1.0000	0.0310
Tump Point, North Carolina	0.3019	1.0000	0.0239
Wood Island, Massachusetts	0.3020	0.8889	0.0285
FERNANDINA BEACH	0.2660	1.0000	0.0097
KEY WEST	0.2666	1.0000	0.0114
HALIFAX	0.2723	1.0000	0.0167
ATLANTIC CITY	0.2741	1.0000	0.0053

Table 5.7.4: The empirical coverage, prediction interval width and Root Mean Square Error (RMSE) by site for our north-south analysis.

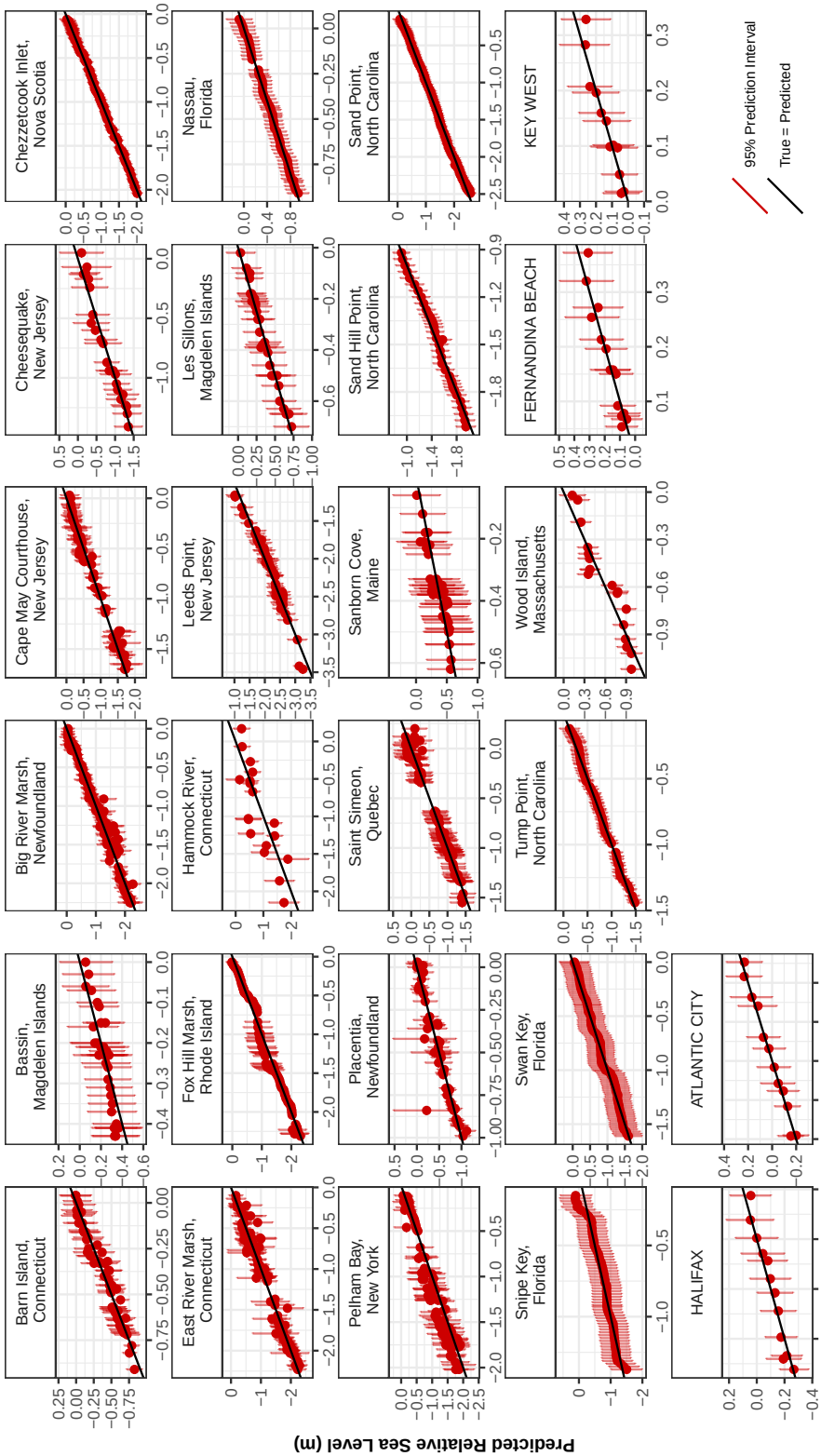


Figure 5.7.17: True versus predicted relative sea level for north south analysis using 10-fold cross validation (CV) using proxy and tide gauge data along the Atlantic coast of the North America. The predicted means are the red points with a vertical 95% prediction interval for each point. The identity line is shown in black.

CHAPTER 6

Conclusions

In this thesis, we have developed novel statistical modelling techniques to examine the physical drivers of RSL change that vary in time and space. We have demonstrated how our model allows users to decompose the RSL signal to examine the underlying physical drivers while accounting for bivariate uncertainties associated with the data. In this final chapter, we summarise the main findings from Chapters 3 to 5 and we address the limitations and potential directions for future research in this study area.

6.1 Chapter Summaries and Limitations

In Chapter 3, we discussed our approach to decomposing the overall RSL signal into contributions from individual physical processes which have differing degrees of significance in different locations and over time, resulting in a complex and ever-changing pattern of RSL change. To gain insights into these patterns, we used tide-gauge data, proxy records and advanced statistical tools to decompose the net RSL signal into contributions from physical processes while accounting for uncertainties in the underlying data. Our statistical model utilised a new generalised additive model containing a combination of splines and random effects within a Bayesian framework. This sophisticated model not only allowed us to discern the contributions of different physical processes to the net RSL signal but also facilitated the incorporation of uncertainties in both parameter estimates

and data. Previous modelling strategies (Kopp et al., 2016; Walker et al., 2021), relied on Gaussian Processes and exhibited computational complexities growing with the cube of data points. Our approach, with the equivalent spline and pre-computed basis functions, offered a far more efficient likelihood computation of just $O(n)$ (Wood, 2017a). As a result, our model could be fitted much quicker than the current state-of-the art approaches, granting us the opportunity to perform additional checks on the model's performance.

The findings from our NI-GAM analysis revealed distinct patterns of pre-anthropogenic (before 1800 CE; Neukom et al. (2019)) RSL change along the Atlantic coast of North America. Initially, the dominant factor was the linear local component, mainly representing the contribution from ongoing GIA. However, after approximately 1900 CE, the regional component took precedence, and we observed a significant increase in the regional rate of change, rising from 0.7 ± 0.5 mm/yr in 1902 to 1.8 ± 0.5 mm/yr at the end of the 20^{th} century. This shift was a consequence of anthropogenic forcing on the climate system (Neukom et al., 2019), leading to sea-level rise through thermosteric and barystatic processes (Frederikse et al., 2020). In addition, the diverse trends captured by the non-linear local component underscores the crucial influence that site-specific processes can exert on RSL changes.

We demonstrated the capabilities of our statistical modelling approach to examine physical drivers influencing RSL change in Chapter 3. As we decomposed the RSL signal, we faced several challenges relating to parameter identifiability and the confounding nature of the regional, linear local, and non-linear local components. To address these complexities efficiently, we opted for an Empirical Bayesian approach instead of a full Bayesian approach. We leveraged the two-step fitting procedure, as necessitated by the noisy input uncertainty method (McHutchon and Rasmussen, 2011), to accurately estimate the unknown parameters. This combination of methods proved crucial in overcoming the difficulties encountered and obtained reliable results in our analysis. In addition, it is essential to acknowledge that our model's scope is limited to regional changes, which sets it apart from prior research (Kopp, 2013; Kopp et al., 2016; Walker et al., 2021). To extend our analysis to capture global RSL changes, we would need to address the challenges posed by the non-

uniform spread and biased distribution of the available data. Previous efforts to address this spatial bias have predominantly focused on using instrumental data from tide gauges and satellites (e.g., Jevrejeva et al., 2008; Wenzel and Schröter, 2010; Church and White, 2011; Hay et al., 2015; Dangendorf et al., 2017; Berrett et al., 2020). The proxy reconstructions are limited to trends sustained on multi-decadal to centennial timescales due to the inherent time-averaging and resolution limits. While there is a growing global distribution of proxy records, accompanied by improvements in data resolution over time, it's crucial to highlight that the collection and analysis techniques required for such records demand extensive human resources and substantial funding. This highlights the ongoing challenges in expanding and enhancing our understanding of global RSL changes.

In Chapter 4, we offered an extensive overview of our R software package `reslr`, which allows for the estimation of rates and spatial patterns of RSL changes over a wide range of timescales (decades to millennia), while accounting for uncertainties associated with noisy and sparse input proxy and/or instrumental data sources. We discussed the range of statistical models available to the user (linear regression (e.g., Ashe et al., 2019), change point models (e.g., Cahill et al., 2015b), integrated Gaussian process (IGP) models (e.g., Cahill et al., 2015a), temporal splines (e.g., de Boor, 1978), spatio-temporal splines (e.g., Simpson, 2018) and generalised additive models (GAM) (e.g., Upton et al., 2023c). We provided insight into the Bayesian framework and the uncertainty methods used in the package which facilitated the estimation of unknown parameters based on the RSL data while fully accounting for the associated uncertainties. We demonstrate the accessible package design using two cases which used our unifying framework for loading data, fitting models and summarising results.

The `reslr` package offers extensive possibilities for the sea-level community and serves as a key contribution to the open-source approach required in the field of climate science. However, there are several limitations associated with our package. One such limitation, as mentioned in the earlier Chapter 3, is that the complexity of the NI-GAM necessitates a large amount of data, which becomes challenging due to the sparsity of proxy records. Another limitation is that the package's primary focus lies in investigating RSL changes from proxy records.

Although we presented a comprehensive method for incorporating tide-gauge data, which can provide valuable insights into recent RSL changes, the user is restricted to using tide-gauge data solely from the PSMSL website (Aarup et al., 2006). Enhancing the flexibility of data sources within the `reslr` package would address this limitation. Furthermore, the package encompasses seven different statistical models, making it comprehensive. However, since the `reslr` package targets the sea-level community, which possesses a wide range of R experience, this results in challenging decisions for the users. For advanced users, the `reslr` package has limitations, as it lacks the capability to allow for tuning of priors within the Bayesian framework to be altered, impacting customization options.

In Chapter 5 we used the extended NI-GAM to examine RSL change along the North Atlantic coastline. The extended NI-GAM remained within a Bayesian framework and decomposed the RSL signal into four components: a common (temporal spline), a regional (hierarchical temporal spline), a linear-local (random effect) and a non-linear local component (spatial temporal spline). This decomposition enhanced comprehension of RSL changes, notably, dynamic sea-level changes driven by ocean dynamics. The North Atlantic region was our focal point, given its vital role as a heat conveyor, transporting warm water from the subtropics northwards (Talley, 2003; Roemmich and Wunsch, 1985; Biastoch et al., 2008). Diverse factors contribute to circulation patterns in the North Atlantic ocean, yielding a complex array of currents (Marshall et al., 2001). Therefore, we targeted two key components: the AMOC and quasi-horizontal circulation (gyres and surface-enhanced currents). Using dynamic sea-level estimates derived from the extended NI-GAM model, along with proxy and tide gauge data, we calculated differences within the east-west sub-region of the North Atlantic. This allowed us to investigate changes in the AMOC. We applied a similar method to the Atlantic coast of North America, focusing on north and south of Cape Hatteras, which enabled us to explore shifts in quasi-horizontal circulation.

When examining the AMOC, dynamic sea-level changes on the eastern boundary fluctuated dramatically while, the western boundary showed minimal change over the past 2,000 years. Previous studies demonstrated that the dynamic sea-level changes along the eastern boundary remain unclear, with models suggesting a pos-

sible sea level fall (Little et al., 2019) or rise (Yin et al., 2009). This trend could have stemmed from an underlying process or, arose from inherent data bias on both sides of the North Atlantic, skewed towards the western boundary. Thus, the variability of the western boundary exerted a strong influence on the common component, resulting in a relatively stable regional component for the western sub-region. The difference between our east and west regional components offered insight into the water transported by the AMOC (Little et al., 2019), which demonstrated an AMOC decline since the late 1800s. The rate of this decline, since the year 1850 CE, marked the steepest drop in 2,000 years, aligning with arguments for a weakening AMOC (e.g, Rahmstorf et al., 2015; Caesar et al., 2018, 2021).

Regarding quasi-horizontal circulation, we studied dynamic sea-level shifts north and south of Cape Hatteras, a boundary for gyres and enhanced surface currents (Woodworth et al., 2011), and these shifts are proportional to water transported along this intergyre region (McCarthy et al., 2015). Water transported across the intergyre boundary fluctuated dramatically over 2,000 years, reflecting strengthening and weakening southward transport by the North Atlantic subtropical gyre. These fluctuations demonstrated an anti-correlation between dynamic sea level changes either side of Cape Hatteras which aligned with results previously presented in Diabaté et al. (2021). McCarthy et al. (2015) indicated that dynamic sea-level changes north of Cape Hatteras were associated with fluctuations in the overturning circulation (Bingham and Hughes, 2009; Ezer, 2013), while southern dynamic sea-level changes reflected variations in surface-enhanced currents, particularly the Gulf Stream from Florida to Cape Hatteras (McCarthy et al., 2015).

In Chapter 5, we demonstrated how proxy sea-level reconstructions, tide-gauge data, and the extended NI-GAM approach can be used to investigate changes in ocean circulation over the past 2,000 years in the North Atlantic. However, our approach possesses a number of limitations which have impacted the interpretation of our results. The first challenge resulted from substantial bivariate uncertainties linked to proxy records and data biases on both sides of the North Atlantic and along the Atlantic coast of North America, which influenced the resolution of our model outputs. The resolution limitations tied to our model outputs posed a challenge when attributing the observed trends solely to changes in AMOC for the

east-west sub-region and quasi-circulation patterns in the north-south sub-region. Other processes such as the spatial fingerprint of sea-level change from ice sheets, are known to impact substantial areas like Greenland, but limited data hinders the quantification of these influences Long et al. (2012); Coulson et al. (2022). Furthermore, tackling parameter identifiability was a challenge we faced while decomposing the RSL signal into distinct components, akin to Chapter 3. Incorporating higher-resolution data across broader spatial scales would improve certainty in our results, aiding in placing present-day observations within the context of recent geological history. This method would allow us to distinguish whether the changes observed in the current Atlantic region are indeed exceptional or follow historical patterns.

6.2 Future Work

In our research, we explored the temporal and spatial variations of relative sea level (RSL) and gained valuable insights into the diverse physical phenomena influencing these changes. Nevertheless, in this section we present opportunities for further extensions and advancements in this field of study.

In Chapter 3, our study showcased the effectiveness of the NI-GAM in identifying the different drivers of relative sea-level (RSL) change along the Atlantic coast of North America. Building on this, Chapter 5 expanded the spatial scope to investigate the North Atlantic region. However, despite these advancements, our models were restricted to regional RSL changes. Extending our modeling approach to incorporate global RSL changes holds great potential for the sea-level community. By encompassing a broader geographical scope, we can gain a more comprehensive understanding of the complex interplay between various drivers and factors influencing sea-level dynamics on a global scale.

The application of GAMs to decompose spatio-temporal signals into different components holds significant potential beyond sea-level research. For instance, employing GAMs to examine the components influencing temperature variations in time and space could prove highly valuable in other areas of research. By disentangling the different drivers and factors affecting temperature changes, GAMs can pro-

vide deeper insights into the complexities of climate dynamics and contribute to a better understanding of the mechanisms driving temperature fluctuations on both regional and global scales. The versatility of GAMs makes them a promising tool for various disciplines seeking to unravel complex spatio-temporal phenomena and uncover underlying patterns and drivers.

Another potential future avenue would be to incorporate other instrumental data sources, such as satellite data, which would allow for the investigation of other variables related to climate change. Satellite data provides high resolution data (Church and White, 2011) and the temporal range is growing each year (Vignudelli et al., 2019). Such an extension would not only enhance our knowledge of sea-level fluctuations but also contribute valuable insights into the broader implications of climate change and its impact on coastal regions worldwide.

In Chapter 4, we showcased the open-source `reslr` package, a powerful tool for Bayesian modeling of relative sea level data. While the package offers a diverse range of statistical models, there is room for improvement by incorporating additional statistical techniques, such as machine learning algorithms. For example, Camargo et al. (2023) utilised a neural network approach and a network detection approach to examine the regional drivers of sea-level change using satellite data on large spatial scales. This approach could provide the `reslr` user with an additional data source via satellites, in turn reducing the spatial bias when examining RSL changes over large spatial scales. Another illustration of the potential of machine learning is evident in endeavors to automatically model and forecast short-term regional coastal sea level changes, as elucidated by Nieves et al. (2021). Through the integration of machine learning techniques, `reslr` could accommodate diverse data sources, enhancing its capability to forecast forthcoming alterations in regional sea levels. This enhancement would broaden the package's applicability and further empower the sea-level community in their endeavors to understand and address the challenges posed by sea-level changes on a larger scale.

Chapter 5 addressed how ocean dynamics influenced RSL changes in the North Atlantic over the past 2,000, by examining dynamic sea-level changes using proxy records and tide-gauge data. Future research could benefit from an inclusive

database integrating instrumental, proxy sources, and satellite data to enhance model precision, especially for recent times and broader spatial extents. Incorporating this additional data would demand a statistical model adept at efficiently accommodating these diverse sources. A potential enhancement involves replacing our current MCMC algorithm, employed via JAGS software (Plummer, 2003), with more efficient alternatives like variational Bayes or Integrated Nested Laplace Approximations (INLA: Rue et al., 2009). An additional model improvement could entail incorporating adaptive penalties for splines, as proposed by Jullion and Lambert (2007), or adopting adaptive measures for controlling spline smoothness, which could better capture complex variations in the data.

In Chapter 5, we used a range of proxy sources from previous studies to place current trends in a wider temporal context (Sherwood et al., 2011; Thibodeau et al., 2018; Thornalley et al., 2018; Spooner et al., 2020; Caesar et al., 2021). An alternative strategy would be to integrate our database of proxy records and instrumental data with these additional proxy sources. This would allow for a multi-proxy analysis to be undertaken, providing a deeper understanding of the variations in the Atlantic Meridional Overturning Circulation (AMOC), quasi-horizontal circulation, and other global ocean circulation systems. This is similar to work undertaken by Croke et al. (2021) which incorporated a range of palaeoclimate data to understand water security planning. This approach offers the potential for a more thorough and detailed comprehension of the complex mechanisms driving dynamic sea-level changes. Such a database and modeling framework would enhance collaboration and knowledge-sharing among researchers, fostering a more holistic understanding of ocean dynamics and their broader implications for climate and coastal regions.

In future work, we can enhance our statistical approach by incorporating a combination of instrumental data from tide gauges and satellite measurements, along with proxy records. This integrated approach holds the potential to provide more accurate and robust projections for future sea-level changes at both regional and local levels. While the AR6 IPCC report currently presents projections for future sea level rise until 2300 (Oppenheimer et al., 2019), the online regional projection tools have limitations, notably in spatial coverage. For example, for Ireland, the existing tool is limited to Dublin (NASA and IPCC, 2021). By integrating our so-

phisticated models, we can extend the spatial projections to cover a broader area, providing valuable insights into sea-level changes at various locations along the North Atlantic coast and beyond. This extension would offer more comprehensive and localised projections, aiding coastal planning and decision-making efforts to address the challenges posed by future sea-level rise.

To conclude, all proposed methods are freely available at `https://github.com/maeveupton` in the repositories named `NI-GAM`, `reslr`, and `RSL-NorthAtlantic`, for Chapters 3, 4 and 5, respectively. This ensures every analysis presented in this thesis is reproducible and methodologies are available to interested practitioners.

Bibliography

Aarup, T., Merrifield, M., Pérez Gómez, B., Vassie, I., and Woodworth, P. (2006). Manual on Sea-level Measurements and Interpretation, Volume IV : An update to 2006. *Intergovernmental Oceanographic Commission of UNESCO*, 4. xxxii, 92, 219, 221, 248

Allen, J. R. (2000). Morphodynamics of Holocene salt marshes: a review sketch from the Atlantic and Southern North Sea coasts of Europe. *Quaternary Science Reviews*, 19(12):1155–1231. 20

Antonov, J. I., Levitus, S., and Boyer, T. P. (2005). Thermosteric sea level rise, 1955–2003. *Geophysical Research Letters*, 32(12). 15

Aquino-López, M. A., Blaauw, M., Christen, J. A., and Sanderson, N. K. (2018). Bayesian Analysis of 210 Pb Dating. *Journal of Agricultural, Biological and Environmental Statistics*, 23(3):317–333. 27, 36, 196

Argus, D., Peltier, W., Drummond, R., and Moore, A. (2014). The Antarctica component of postglacial rebound model ICE-6GC (VM5a) based on GPS positioning, exposure age dating of ice thicknesses, and relative sea level histories. *Geophysical Journal International*, 198:537–563. 43

Ashe, E. L., Cahill, N., Hay, C., Khan, N. S., Kemp, A., Engelhart, S. E., Horton, B. P., Parnell, A. C., and Kopp, R. E. (2019). Statistical modeling of rates and trends in Holocene relative sea level. *Quaternary Science Reviews*, 204:58–77. 3, 31, 33, 41, 89, 247

Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(5):617–666. 54

Barlow, N. L., Long, A. J., Saher, M. H., Gehrels, W. R., Garnett, M. H., and Scaife, R. G. (2014). Salt-marsh reconstructions of relative sea-level change in the North Atlantic during the last 2000 years. *Quaternary Science Reviews*, 99:1–16. 220

Barlow, N. L., Shennan, I., Long, A. J., Gehrels, W. R., Saher, M. H., Woodroffe, S. A., and Hillier, C. (2013). Salt marshes as late Holocene tide gauges. *Global and Planetary Change*, 106:90–110. 3, 26, 28, 222, 223

Barnett, R., Bernatchez, P., Garneau, M., Brain, M., Charman, D., Stephenson, D., Haley, S., and Sanderson, N. (2019). Late Holocene sea-level changes in eastern Québec and potential drivers. *Quaternary Science Reviews*, 203:151–169. 76

Barnett, R. L., Bernatchez, P., Garneau, M., and Juneau, M.-N. (2017). Reconstructing late Holocene relative sea-level changes at the Magdalen Islands (Gulf of St. Lawrence, Canada) using multi-proxy analyses. *Journal of Quaternary Science*, 32(3):380–395. 76, 118, 220

Bellomo, K., Meccia, V. L., D'Agostino, R., Fabiano, F., Larson, S. M., von Hardenberg, J., and Corti, S. (2023). Impacts of a weakened AMOC on precipitation over the Euro-Atlantic region in the EC-Earth3 climate model. *Climate Dynamics*. 189

Berrett, C., Christensen, W. F., Sain, S. R., Sandholtz, N., Coats, D. W., Tebaldi, C., and Lopes, H. F. (2020). Modeling sea-level processes on the U.S. Atlantic Coast. *Environmetrics*, 31(4). 73, 247

Biastoch, A., Böning, C. W., Getzlaff, J., Molines, J.-M., and Madec, G. (2008). Causes of Interannual–Decadal Variability in the Meridional Overturning Circulation of the Midlatitude North Atlantic Ocean. *Journal of Climate*, 21(24):6599 – 6615. 215, 248

Biferno, A., Shaftel, H., Callery, S., Jackson, R., Bailey, D., and Callery, S. (2023). Sea level: Global Climate Change. *NASA*. xvi, 2, 23

Bilbao, R. A. F., Gregory, J. M., and Bouttes, N. (2015). Analysis of the regional pattern of sea level change due to ocean dynamics and density change for 1993–2099 in observations and CMIP5 AOGCMs. *Climate Dynamics*, 45(9):2647–2666. 188

Bingham, R. J. and Hughes, C. W. (2009). Signature of the Atlantic meridional overturning circulation in sea level along the east coast of North America. *Geophysical Research Letters*, 36(2). 217, 249

Bjerknes, J. (1964). Atlantic Air-Sea Interaction. *Advances in Geophysics*, 10:1–82. 191

Blaauw, M. and Christen, J. A. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis*, 6(3):457 – 474. 27, 36, 196

Brain, M. J. (2016). Past, Present and Future Perspectives of Sediment Compaction as a Driver of Relative Sea Level and Coastal Change. *Current Climate Change Reports*, 2(3):75–85. 20, 21

Brain, M. J., Long, A. J., Petley, D. N., Horton, B. P., and Allison, R. J. (2011). Compression behaviour of minerogenic low energy intertidal sediments. *Sedimentary Geology*, 233(1-4):28–41. 20, 202

Brain, M. J., Long, A. J., Woodroffe, S. A., Petley, D. N., Milledge, D. G., and Parnell, A. C. (2012). Modelling the effects of sediment compaction on salt marsh reconstructions of recent sea-level rise. *Earth and Planetary Science Letters*, 345-348:180–193. 96

Broecker, W. S. (1991). The Great Ocean Conveyor. *Oceanography*, 4(2):79–89. xvi, 19

Brooks, A. and Edwards, R. (2006). The Development of a Sea-Level Database for Ireland. *Irish Journal of Earth Sciences*, 24:13–27. 24, 25

Brown, S., Nicholls, R. J., Lowe, J. A., and Hinkel, J. (2016). Spatial variations of sea-level rise and impacts: An application of DIVA. *Climatic Change*, 134(3):403–416. 3

Bulgin, C. E., Mecking, J. V., Harvey, B. J., Jevrejeva, S., McCarroll, N. F., Merchant, C. J., and Sinha, B. (2023). Dynamic sea-level changes and potential implications for storm surges in the UK: a storylines perspective. *Environmental Research Letters*, 18(4):044033. 188

Caesar, L., McCarthy, G. D., Thornalley, D. J. R., Cahill, N., and Rahmstorf, S. (2021). Current Atlantic Meridional Overturning Circulation weakest in last millennium. *Nature Geoscience*, 14(3):118–120. xxvi, 191, 192, 208, 209, 217, 249, 252

Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., and Saba, V. (2018). Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature*, 556(7700):191–196. 217, 249

Cahill, N., Kemp, A. C., Horton, B. P., and Parnell, A. C. (2015a). Modeling sea-level change using Errors-in-Variables integrated Gaussian Process 1. *The Annals of Applied Statistics*, 9(2):547–571. 3, 34, 41, 47, 89, 90, 95, 97, 247

Cahill, N., Kemp, A. C., Horton, B. P., and Parnell, A. C. (2016). A Bayesian hierarchical model for reconstructing relative sea level: from raw data to rates of change. *Climate of the Past*, 12(2):525–542. 28, 37, 76, 118, 196, 220

Cahill, N., Rahmstorf, S., and Parnell, A. C. (2015b). Change points of global temperature. *Environmental research letters*, 10(8):84002. 5, 89, 95, 96, 247

Camargo, C. M., Riva, R. E., Hermans, T. H., Schütt, E. M., Marcos, M., Hernandez-Carrasco, I., and Slangen, A. (2023). Regionalizing the sea-level budget with machine learning techniques. *Ocean Science*, 19(1):17–41. 251

Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992). Hierarchical Bayesian Analysis of Changepoint Problems. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(2):389–405. 5, 96

Caron, L., Ivins, E. R., Larour, E., Adhikari, S., Nilsson, J., and Blewitt, G. (2018). GIA Model Statistics for GRACE Hydrology, Cryosphere, and Ocean Science. *Geophysical Research Letters*, 45(5):2203–2212. 17, 43

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76. 93

Cazenave, A., Gouzenes, Y., Birol, F., Leger, F., Passaro, M., Calafat, F. M., Shaw, A., Nino, F., Legeais, J. F., Oelsmann, J., Restano, M., and Benveniste, J. (2022). Sea level along the world's coastlines can be measured by a network of virtual altimetry stations. *Communications Earth & Environment*, 3(1):117. 2

Cazenave, A. and Llovel, W. (2010). Contemporary Sea Level Rise. *Annual Review of Marine Science*, 2(1):145–173. 3

Church, J. A., Clark, P. U., Cazenave, A., Gregory, J. M., Jevrejeva, S., Levermann, A., Merrifield, M. A., Milne, G. A., Nerem, R. S., Nunn, P. D., Payne, A. J., Pfeffer, W. T., Stammer, D., and Unnikrishnan, A. S. (2013). Sea level change in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *IPCC 2013*, 13. 12, 13, 14, 15, 22, 31, 33, 188

Church, J. A., Gregory, J. M., Huybrechts, P., Kuhn, M., Lambeck, K., Nhuan, M. T., Qin, D., and Woodworth, P. L. (2001). Changes in Sea Level: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel. 14, 69

Church, J. A. and White, N. J. (2011). Sea-Level Rise from the Late 19th to the Early 21st Century. *Surveys in Geophysics*, 32(4):585–602. 3, 31, 35, 73, 188, 247, 251

Church, J. A., White, N. J., Konikow, L. F., Domingues, C. M., Cogley, J. G., Rignot, E., Gregory, J. M., van den Broeke, M. R., Monaghan, A. J., and Velicogna, I. (2011). Revisiting the Earth's sea-level and energy budgets from 1961 to 2008. *Geophysical Research Letters*, 38(18). 188

Coulson, S., Dangendorf, S., Mitrovica, J. X., Tamisiea, M. E., Pan, L., and Sandwell, D. T. (2022). A detection of the sea level fingerprint of Greenland Ice Sheet melt. *Science*, 377(6614):1550–1554. 18, 218, 250

Cox, M. G. (2006). Practical spline approximation. In *Topics in Numerical Analysis: Proceedings of the SERC Summer School, Lancaster, July 19–August 21, 1981*, pages 79–112. Springer. 7

Croke, J., Vítkovský, J., Hughes, K., Campbell, M., Amirnezhad-Mozhdehi, S., Parnell, A., Cahill, N., and Dalla Pozza, R. (2021). A palaeoclimate proxy database for water security planning in Queensland Australia. *Scientific Data*, 8(1):292. 219, 252

Dangendorf, S., Marcos, M., Wöppelmann, G., Conrad, C. P., Frederikse, T., and Riva, R. (2017). Reassessment of 20th century global mean sea level rise. *Proceedings of the National Academy of Sciences*, 114(23):5946–5951. 3, 73, 247

Davis, J. L. and Mitrovica, J. X. (1996). Glacial isostatic adjustment and the anomalous tide gauge record of eastern North America. *Nature*, 379(6563):331–333. 17

de Boor, C. (1978). A Practical Guide to Spline. *Applied Mathematical Sciences, New York: Springer, 1978*, 27. 6, 7, 47, 90, 98, 201, 203, 247

Dean, S., Horton, B. P., Evelpidou, N., Cahill, N., Spada, G., and Sivan, D. (2019). Can we detect centennial sea-level variations over the last three thousand years in Israeli archaeological records? *Quaternary Science Reviews*, 210:125–135. 6, 97

Delworth, T. L., Zeng, F., Zhang, L., Zhang, R., Vecchi, G. A., and Yang, X. (2017). The central role of ocean dynamics in connecting the North Atlantic Oscillation to the extratropical component of the Atlantic multidecadal oscillation. *Journal of Climate*, 30(10):3789–3805. 216

Denny, M. (2008). *How the Ocean Works: An Introduction to Oceanography*. Princeton University Press, Princeton. 189

Dey, D. K., Ghosh, S. K., and Mallick, B. K. (2000). *Generalized linear models: a Bayesian perspective*. CRC Press. 8, 47, 90, 93

Di Lorenzo, E., Schneider, N., Cobb, K. M., Franks, P., Chhak, K., Miller, A. J., McWilliams, J. C., Bograd, S. J., Arango, H., Curchitser, E., et al. (2008). North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophysical research letters*, 35(8). 191

Diabaté, S. T., Swingedouw, D., Hirschi, J. J.-M., Duchez, A., Leadbitter, P. J., Haigh, I. D., and McCarthy, G. D. (2021). Western boundary circulation and coastal sea-level variability in Northern Hemisphere oceans. *Ocean Science*, 17(5):1449–1471. 217, 249

Dierckx, P. (1995). *Curve and surface fitting with splines*. Oxford University Press. 6, 7, 47, 98

Dipper, F. (2022). Chapter 2 - The seawater environment and ecological adaptations. In Dipper, F., editor, *Elements of Marine Ecology (Fifth Edition)*, pages 37–151. Butterworth-Heinemann, fifth edition edition. 21

Donnelly, J. P. (2006). A Revised Late Holocene Sea-Level Record for Northern Massachusetts, USA. *Journal of Coastal Research*, 22(5):1051–1061. 76, 118, 220

Donnelly, J. P., Cleary, P., Newby, P., and Ettinger, R. (2004). Coupling instrumental and geological records of sea-level change: Evidence from southern New England of an increase in the rate of sea-level rise in the late 19th century. *Geophysical Research Letters*, 31(5). 76, 118, 220

Douglas, B., Kearney, M. S., and Leatherman, S. P. (2000). *Sea level rise: History and consequences*. Elsevier. 2

Dunbar, B. (2010). NASA study finds Atlantic "conveyor belt" not slowing. *NASA*. xvi, 14

Edwards, R. (2005). Sea levels: abrupt events and mechanisms of change. *Progress in Physical Geography*, 29:599–608. 24

Edwards, R. (2007). Sea levels: resolution and uncertainty. *Progress in Physical Geography: Earth and Environment*, 31(6):621–632. 28, 222, 223

Edwards, R. and Wright, A. (2015). *Foraminifera: Handbook of Sea-Level Research*, chapter 13, pages 191–217. John Wiley & Sons, Ltd. 37, 92, 196

Eilers, P. and Marx, B. (1996). Flexible Smoothing with B-splines and Penalities. *Statistical Science*, 11. 6, 7, 47, 56, 98, 201

Engelhart, S. E. and Horton, B. P. (2012). Holocene sea level database for the Atlantic coast of the United States. *Quaternary Science Reviews*, 54:12–25. Coastal Change during the Late Quaternary. 220

Engelhart, S. E., Horton, B. P., Douglas, B. C., Peltier, W. R., and Törnqvist, T. E. (2009). Spatial variability of late Holocene and 20th century sea-level rise along the Atlantic coast of the United States. *Geology*, 37(12):1115–1118. 4, 16, 39, 43, 46, 49, 74, 96, 100, 111, 202

Ezer, T. (2013). Sea level rise, spatially uneven and temporally unsteady: Why the US East Coast, the global tide gauge record, and the global altimeter data show different trends. *Geophysical Research Letters*, 40(20):5439–5444. 217, 249

Fairbridge, R. W. (1961). Eustatic changes in sea level. *Physics and Chemistry of the Earth*, 4:99–185. 13

Finkl, C. W. (2013). *Coastal hazards*. Springer. 2

Fjeldskaar, W. (1994). The amplitude and decay of the glacial forebulge in Fennoscandia. *Norsk Geologisk Tidsskrift*, 74(1):2–8. 16

Flato, G. and Marotzke, J. (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *IPCC 2013*, 13. 3

Frankignoul, C., Gastineau, G., and Kwon, Y.-O. (2013). The Influence of the AMOC Variability on the Atmosphere in CCSM3. *Journal of Climate*, 26(24):9774 – 9790. 189

Frederikse, T., Landerer, F., Caron, L., Adhikari, S., Parkes, D., Humphrey, V. W., Dangendorf, S., Hogarth, P., Zanna, L., Cheng, L., et al. (2020). The causes of sea-level rise since 1900. *Nature*, 584(7821):393–397. 72, 246

Fuller, W. (1987). *A Single Explanatory Variable*, chapter 1, pages 1–99. John Wiley & Sons, Ltd. 8

Gabry, J. and Goodrich, B. (2017). rstanarm: Bayesian applied regression modeling via Stan. xix, xxii, 51, 60, 61, 62, 86, 87

García-Artola, A., Cearreta, A., Leorri, E., Irabien, M. J., and Blake, W. H. (2009). Coastal marshes as geological archives of recent variations in sea level. *Gazette*, 47(3-4):109–112. 220

Gehrels, W. R. (1994). Determining Relative Sea-Level Change from Salt-Marsh Foraminifera and Plant Zones on the Coast of Maine, U.S.A. *Journal of Coastal Research*, 10(4):990–1009. 31, 89, 92, 188, 196

Gehrels, W. R., Dangendorf, S., Barlow, N. L. M., Saher, M. H., Long, A. J., Woodworth, P. L., Piecuch, C. G., and Berk, K. (2020). A Preindustrial Sea-Level Rise Hotspot Along the Atlantic Coast of North America. *Geophysical Research Letters*, 47(4). 36, 76, 118, 195, 220

Gehrels, W. R., Horton, B. P., Kemp, A. C., and Sivan, D. (2011). Two millennia of sea level data: the key to predicting change. *Eos, Transactions American Geophysical Union*, 92(35):289–290. 220

Gehrels, W. R., Marshall, W. A., Gehrels, M. J., Larsen, G., Kirby, J. R., Eiríksson, J., Heinemeier, J., and Shimmield, T. (2006). Rapid sea-level rise in the North Atlantic Ocean since the first half of the nineteenth century. *The Holocene*, 16(7):949–965. 220

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian analysis*, 1(3):515–534. 50

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press. 107

Gelman, A. and Rubin, D. B. (1992a). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472. 61

Gelman, A. and Rubin, D. B. (1992b). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472. 107

Gerlach, M. J., Engelhart, S. E., Kemp, A. C., Moyer, R. P., Smoak, J. M., Bernhardt, C. E., and Cahill, N. (2017). Reconstructing Common Era relative sea-level change on the Gulf Coast of Florida. *Marine Geology*, 390:254–269. 76, 118, 220

Gornitz, V. (2009). *Paleoclimate Proxies, An Introduction*, pages 716–721. Springer Netherlands. 92, 196

Govan, E., Jackson, A. L., Inger, R., Bearhop, S., and Parnell, A. C. (2023). simmr: A package for fitting stable isotope mixing models in r. 91

Gregory, J. M., Griffies, S. M., Hughes, C. W., Lowe, J. A., Church, J. A., Fukimori, I., Gomez, N., Kopp, R. E., Landerer, F., Cozannet, G. L., Ponte, R. M., Stammer, D., Tamisiea, M. E., and van de Wal, R. S. (2019). Concepts and Terminology for Sea Level: Mean, Variability and Change, Both Local and Global. *Surveys in Geophysics*, 40(6):1251–1289. 14, 15, 18, 39, 40, 100, 101, 188, 191, 200

Gregory, J. M., White, N. J., Church, J. A., Bierkens, M. F., Box, J. E., Van den Broeke, M. R., Cogley, J. G., Fettweis, X., Hanna, E., Huybrechts, P., et al. (2013). Twentieth-century global-mean sea level rise: Is the whole greater than the sum of the parts? *Journal of Climate*, 26(13):4476–4499. 188

Grinsted, A. (2015). Projected Change—Sea Level. *Second assessment of climate change for Baltic Sea basin*, pages 253–263. 15, 39, 100, 200

Hall, G. F., Hill, D. F., Horton, B. P., Engelhart, S. E., and Peltier, W. (2013). A high-resolution study of tides in the Delaware Bay: Past conditions and future scenarios. *Geophysical Research Letters*, 40(2):338–342. 21, 203

Hastie, T. and Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, 1(3):297–318. 56

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis. 7, 34

Hauer, M. E., Hardy, D., Kulp, S. A., Mueller, V., Wrathall, D. J., and Clark, P. U. (2021). Assessing population exposure to coastal flooding due to sea level rise. *Nature Communications*, 12(1):6900. 2

Hawkes, A. D., Kemp, A. C., Donnelly, J. P., Horton, B. P., Peltier, W. R., Cahill, N., Hill, D. F., Ashe, E., and Alexander, C. R. (2016). Relative sea-level change in northeastern Florida (USA) during the last 8.0 ka. *Quaternary Science Reviews*, 142:90–101. 6, 97

Hay, C. C., Morrow, E., Kopp, R. E., and Mitrovica, J. X. (2015). Probabilistic reanalysis of twentieth-century sea-level rise. *Nature*, 517(7535):481–484. 72, 73, 247

Hill, D. F., Griffiths, S. D., Peltier, W. R., Horton, B. P., and Törnqvist, T. E. (2011). High-resolution numerical modeling of tides in the western Atlantic, Gulf of Mexico, and Caribbean Sea during the Holocene. *Journal of Geophysical Research: Oceans*, 116(C10). 21, 203

Hogarth, P., Hughes, C., Williams, S., and Wilson, C. (2020). Improved and extended tide gauge records for the British Isles leading to more consistent estimates of sea level rise and acceleration since 1958. *Progress in Oceanography*, 184:102333. 5, 96

Hogarth, P., Pugh, D., Hughes, C., and Williams, S. (2021). Changes in mean sea level around Great Britain over the past 200 years. *Progress in Oceanography*, 192:102521. 5

Holgate, S. J., Matthews, A., Woodworth, P. L., Rickards, L. J., Tamisiea, M. E., Bradshaw, E., Foden, P. R., Gordon, K. M., Jevrejeva, S., and Pugh, J. (2013). New Data Systems and Products at the Permanent Service for Mean Sea Level. *Journal of Coastal Research*, 29(3):493 – 504. xvii, xxxi, xxxii, 23, 24, 35, 74, 77, 89, 92, 103, 197, 219, 221

Holsclaw, T., Sansó, B., Lee, H. K. H. L., Heitmann, K., Habib, S., Higdon, D., and Alam, U. (2013). Gaussian Process Modeling of Derivative Curves. *Technometrics*, 55(1):57–67. 6, 97

Horton, B. P., Culver, S. J., Harbattle, M. I. J., Larcombe, P., Milne, G. A., Morigi, C., Whittaker, J. E., and Woodroffe, S. A. (2007). Reconstructing Holocene sea-level change for the Central Great Barrier Reef (Australia) using subtidal foraminifera. *Journal of Foraminiferal Research*, 37:327–343. 25

Horton, B. P. and Edwards, R. J. (2006). Quantifying Holocene sea level change using intertidal foraminifera: lessons from the British Isles. *Departmental Papers (EES)*, page 50. 28, 36, 37, 196

Horton, B. P., Kopp, R. E., Garner, A. J., Hay, C. C., Khan, N. S., Roy, K., and Shaw, T. A. (2018). Mapping Sea-Level Change in Time, Space, and Probability. *Annual Review of Environment and Resources*, 43(1):481–521. 20, 21, 25, 27, 31, 40, 101, 202

Ionita, M., Nagavciuc, V., Scholz, P., and Dima, M. (2022). Long-term drought intensification over Europe driven by the weakening trend of the Atlantic Meridional Overturning Circulation. *Journal of Hydrology: Regional Studies*, 42:101176. 189

IPCC (2007). Climate change 2007: The physical science basis. Intergovernmental Panel On Climate Change. *Summary for Policy makers*. 1, 14

IPCC (2023). Summary for Policymakers. In: Climate Change 2023:Synthesis Report. A Report of the Intergovernmental Panel on Climate Change. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. *In press*, pages 1–36. 1, 3, 188

Ivins, E. R., James, T. S., Wahr, J., O. Schrama, E. J., Landerer, F. W., and Simon, K. M. (2013). Antarctic contribution to sea level rise observed by GRACE with improved GIA correction. *Journal of Geophysical Research: Solid Earth*, 118(6):3126–3141. 17

Jeon, T., Seo, K.-W., Kim, B.-H., Kim, J.-S., Chen, J., and Wilson, C. R. (2021). Sea level fingerprints and regional sea level change. *Earth and Planetary Science Letters*, 567:116985. 18

Jevrejeva, S., Moore, J. C., Grinsted, A., and Woodworth, P. L. (2008). Recent global sea level acceleration started over 200 years ago? *Geophysical Research Letters*, 35(8). 73, 247

Jullion, A. and Lambert, P. (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Computational Statistics & Data Analysis*, 51(5):2542–2558. 219, 252

Kaye, C. A. and Barghoorn, E. S. (1964). Late Quaternary sea-level change and crustal rise at Boston, Massachusetts, with notes on the autocompaction of peat. *Geological Society of America Bulletin*, 75(2):63–80. 20

Kelley, D. and Richards, C. (2014). oce: Analysis of oceanographic data. *R package version*, 1(0). 91

Kemp, A. C., Bernhardt, C. E., Horton, B. P., Kopp, R. E., Vane, C. H., Peltier, W. R., Hawkes, A. D., Donnelly, J. P., Parnell, A. C., and Cahill, N. (2014). Late Holocene sea- and land-level change on the U.S. southeastern Atlantic coast. *Marine Geology*, 357:90–100. 76, 118, 220

Kemp, A. C., Hawkes, A. D., Donnelly, J. P., Vane, C. H., Horton, B. P., Hill, T. D., Anisfeld, S. C., Parnell, A. C., and Cahill, N. (2015). Relative sea-level change in Connecticut (USA) during the last 2200 yrs. *Earth and Planetary Science Letters*, 428:217–229. 72, 76, 118, 220

Kemp, A. C., Hill, T. D., Vane, C. H., Cahill, N., Orton, P. M., Talke, S. A., Parnell, A. C., Sanborn, K., and Hartig, E. K. (2017). Relative sea-level trends in New York City during the past 1500 years. *The Holocene*, 27(8):1169–1186. xxii, xxiii, 6, 76, 97, 103, 106, 108, 109, 118, 220

Kemp, A. C., Horton, B. P., Culver, S. J., Corbett, D. R., van de Plassche, O., Gehrels, W. R., Douglas, B. C., and Parnell, A. C. (2009). Timing and

magnitude of recent accelerated sea-level rise (North Carolina, United States). *Geology*, 37(11):1035–1038. 5, 25, 27, 96

Kemp, A. C., Horton, B. P., Donnelly, J. P., Mann, M. E., Vermeer, M., and Rahmstorf, S. (2011). Climate related sea-level variations over the past two millennia. *Proceedings of the National Academy of Sciences*, 108(27):11017–11022. xxii, xxiii, 27, 28, 37, 76, 103, 106, 108, 109, 118, 196, 220

Kemp, A. C., Horton, B. P., Vane, C. H., Bernhardt, C. E., Corbett, D. R., Engelhart, S. E., Anisfeld, S. C., Parnell, A. C., and Cahill, N. (2013). Sea-level change during the last 2500 years in New Jersey, USA. *Quaternary Science Reviews*, 81:90–104. 27, 31, 36, 76, 89, 118, 195, 196, 220

Kemp, A. C. and Telford, R. J. (2015). *Transfer functions: Handbook of Sea-Level Research*, chapter 31, pages 470–499. John Wiley & Sons, Ltd. 27, 28, 37, 92, 196

Kemp, A. C., Wright, A. J., Edwards, R. J., Barnett, R. L., Brain, M. J., Kopp, R. E., Cahill, N., Horton, B. P., Charman, D. J., Hawkes, A. D., et al. (2018). Relative sea-level change in Newfoundland, Canada during the past 3000 years. *Quaternary Science Reviews*, 201:89–110. 6, 31, 33, 40, 72, 73, 76, 89, 92, 98, 118, 196, 220

Khan, N. S., Ashe, E., Horton, B. P., Dutton, A., Kopp, R. E., Brocard, G., Engelhart, S. E., Hill, D. F., Peltier, W., Vane, C. H., and Scatena, F. N. (2017). Drivers of Holocene sea-level change in the Caribbean. *Quaternary Science Reviews*, 155:13–36. 73

Khan, N. S., Ashe, E., Moyer, R. P., Kemp, A. C., Engelhart, S. E., Brain, M. J., Toth, L. T., Chappel, A., Christie, M., Kopp, R. E., and Horton, B. P. (2022). Relative sea-level change in South Florida during the past 5000 years. *Global and Planetary Change*, 216:103902. 25, 31, 32, 36, 71, 76, 118, 190, 195, 220

Khan, N. S., Ashe, E., Shaw, T. A., Vacchi, M., Walker, J., Peltier, W. R., Kopp, R. E., and Horton, B. P. (2015). Holocene Relative Sea-Level Changes from Near-, Intermediate-, and Far-Field Locations. *Current Climate Change Reports*, 1(4):247–262. 89

Khan, N. S., Horton, B. P., Engelhart, S., Rovere, A., Vacchi, M., Ashe, E. L., Törnqvist, T. E., Dutton, A., Hijma, M. P., and Shennan, I. (2019). Inception of a global atlas of sea levels since the Last Glacial Maximum. *Quaternary Science Reviews*, 220:359–371. 25

Kirby, J. R., Garrett, E., and Gehrels, W. R. (2023). Holocene relative sea-level changes in northwest Ireland: An empirical test for glacial isostatic adjustment models. *The Holocene*. 6, 97, 220

Kopp, R., Horton, B., Kemp, A., and Tebaldi, C. (2015). Past and future sea-level rise along the coast of North Carolina, USA. *Climatic Change*, 132(4):693–707. 3, 39

Kopp, R. E. (2013). Does the mid-Atlantic United States sea level acceleration hot spot reflect ocean dynamic variability? *Geophysical Research Letters*, 40(15):3981–3985. 6, 89, 98, 191, 246

Kopp, R. E., Kemp, A. C., Bittermann, K., Horton, B. P., Donnelly, J. P., Gehrels, W. R., Hay, C. C., Mitrovica, J. X., Morrow, E. D., and Rahmstorf, S. (2016). Temperature-driven global sea-level variability in the Common Era. *Proceedings of the National Academy of Sciences of the United States of America*, 113(11):E1434–E1441. 3, 6, 33, 34, 35, 43, 72, 89, 98, 190, 246

Kopp, R. E., Simons, F. J., Mitrovica, J. X., Maloof, A. C., and Oppenheimer, M. (2009). Probabilistic assessment of sea level during the last interglacial stage. *Nature*, 462(7275):863–867. 3, 6, 97

Körper, J., Spangehl, T., Cubasch, U., and Huebener, H. (2009). Decomposition of projected regional sea level rise in the North Atlantic and its relation to the AMOC. *Geophysical Research Letters*, 36(19). 189

Kuhlbrodt, T., Griesel, A., Montoya, M., Levermann, A., Hofmann, M., and Rahmstorf, S. (2007). On the driving processes of the Atlantic meridional overturning circulation. *Reviews of Geophysics*, 45(2). 19

Lambeck, K., Antonioli, F., Purcell, A., and Silenzi, S. (2004). Sea-level change along the Italian coast for the past 10,000yr. *Quaternary Science Reviews*, 23(14):1567–1598. 13

Law, K. L., Morét-Ferguson, S., Maximenko, N. A., Proskurowski, G., Peacock, E. E., Hafner, J., and Reddy, C. M. (2010). Plastic accumulation in the North Atlantic subtropical gyre. *Science*, 329(5996):1185–1188. 19

Legeais, J.-F., Ablain, M., Zawadzki, L., Zuo, H., Johannessen, J. A., Scharffenberg, M. G., Fenoglio-Marc, L., Fernandes, M. J., Andersen, O. B., Rudenko, S., Cipollini, P., Quartly, G. D., Passaro, M., Cazenave, A., and Benveniste, J. (2018). An improved and homogeneous altimeter sea level record from the ESA Climate Change Initiative. *Earth System Science Data*, 10(1):281–301. 2

Leorri, E., Cearreta, A., and Milne, G. (2012). Field observations and modelling of Holocene sea-level changes in the southern Bay of Biscay: implication for understanding current rates of relative sea-level change and vertical land motion along the Atlantic coast of SW Europe. *Quaternary Science Reviews*, 42:59–73. 220

Leuliette, E. W. (2015). The Balancing of the Sea-Level Budget. *Current Climate Change Reports*, 1(3):185–191. 15

Little, C. M., Hu, A., Hughes, C. W., McCarthy, G. D., Piecuch, C. G., Ponte, R. M., and Thomas, M. D. (2019). The Relationship Between U.S. East Coast Sea Level and the Atlantic Meridional Overturning Circulation: A Review. *Journal of Geophysical Research: Oceans*, 124(9):6435–6458. 189, 193, 206, 216, 249

Long, A., Barlow, N., Gehrels, W., Saher, M., Woodworth, P., Scaife, R., Brain, M., and Cahill, N. (2014). Contrasting records of sea-level change in the eastern and western North Atlantic during the last 300 years. *Earth and Planetary Science Letters*, 388:110–122. 220

Long, A. J., Woodroffe, S. A., Milne, G. A., Bryant, C. L., Simpson, M. J., and Wake, L. M. (2012). Relative sea-level change in Greenland during the last 700 yrs and ice sheet response to the Little Ice Age. *Earth and Planetary Science Letters*, 315:76–85. 218, 250

Lowe, J. A. and Gregory, J. M. (2006). Understanding projections of sea level rise in a Hadley Centre coupled climate model. *Journal of Geophysical Research: Oceans*, 111(C11). 18

Lozier, M., Owens, W., and Curry, R. G. (1995). The climatology of the North Atlantic. *Progress in Oceanography*, 36(1):1–44. 189, 191

Marshall, J., Johnson, H., and Goodman, J. (2001). A study of the interaction of the North Atlantic Oscillation with ocean circulation. *Journal of Climate*, 14(7):1399–1421. 216, 248

Marshall, W. (2015). *Chronohorizons: Handbook of Sea-Level Research*, chapter 25, pages 373–385. John Wiley & Sons, Ltd. 27, 28, 36, 92, 196, 218

Maslin, M. A. (2019). Climate change: essential knowledge for developing holistic solutions to our climate crisis. *Emerging Topics in Life Sciences*, 3(2):245–256. 1

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B. (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. *Cambridge University Press*. 1, 2, 3, 15, 31

Mawdsley, R. J., Haigh, I. D., and Wells, N. C. (2015). Global secular changes in different tidal high water, low water and range levels. *Earth's Future*, 3(2):66–81. 21, 203

McCarthy, G. D., Brown, P. J., Flagg, C. N., Goni, G., Houpert, L., Hughes, C. W., Hummels, R., Inall, M., Jochumsen, K., Larsen, K. M. H., Lherminier, P., Meinen, C. S., Moat, B. I., Rayner, D., Rhein, M., Roessler, A., Schmid, C., and Smeed, D. A. (2020). Sustainable Observations of the AMOC: Methodology and Technology. *Reviews of Geophysics*, 58(1):e2019RG000654. 192, 193, 210

McCarthy, G. D., Haigh, I. D., Hirschi, J. J. M., Grist, J. P., and Smeed, D. A. (2015). Ocean impact on decadal Atlantic climate variability revealed by sea-level observations. *Nature*, 521(7553):508–510. 189, 191, 192, 194, 217, 249

McHutchon, A. and Rasmussen, C. E. (2011). Gaussian Process training with input noise. *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*, pages 1–9. 9, 34, 43, 47, 48, 49, 90, 94, 200, 215, 246

Meltzner, A. J., Switzer, A. D., Horton, B. P., Ashe, E., Qiu, Q., Hill, D. F., Bradley, S. L., Kopp, R. E., Hill, E. M., Majewski, J. M., Natawidjaja, D. H., and Suwargadi, B. W. (2017). Half-metre sea-level fluctuations on centennial timescales from mid-Holocene corals of Southeast Asia. *Nature Communications*, 8(1):14387. 31, 89, 188

Milne, G. A., Gehrels, W. R., Hughes, C. W., and Tamisiea, M. E. (2009). Identifying the causes of sea-level change. *Nature Geoscience*, 2(7):471–478. 13

Minobe, S., Kuwano-Yoshida, A., Komori, N., Xie, S.-P., and Small, R. J. (2008). Influence of the Gulf Stream on the troposphere. *Nature*, 452(7184):206–209. 194

Minobe, S., Terada, M., Qiu, B., and Schneider, N. (2017). Western Boundary Sea Level: A Theory, Rule of Thumb, and Application to Climate Models. *Journal of Physical Oceanography*, 47(5):957 – 977. 191

Murray-Wallace, C. V. and Woodroffe, C. D. (2014). *Quaternary sea-level changes: a global perspective*. Cambridge University Press. 2, 12

Muschitiello, F., D'Andrea, W. J., Schmittner, A., Heaton, T. J., Balascio, N. L., deRoberts, N., Caffee, M. W., Woodruff, T. E., Welten, K. C., Skinner, L. C., Simon, M. H., and Dokken, T. M. (2019). Deep-water circulation changes lead North Atlantic climate during deglaciation. *Nature Communications*, 10(1):1272. 191

NASA and IPCC (2021). NASA and IPCC Sea Level Projection Tool. *NASA and IPCC*. 252

Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., Franke, J., Kaufman, D. S., Lücke, L., Rehfeld, K., Schurer, A., Zhu, F., Brönnimann, S., Hakim, G. J., Henley, B. J., Ljungqvist, F. C., McKay, N., Valler, V., von Gunten, L., and 2k Consortium, P. (2019). Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era. *Nature Geoscience*, 12(8):643–649. xxxi, 43, 49, 68, 71, 72, 74, 246

Neumann, B., Vafeidis, A. T., Zimmermann, J., and Nicholls, R. J. (2015). Future Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding - A Global Assessment. *PLOS ONE*, 10:1–34. 2

Nieves, V., Radin, C., and Camps-Valls, G. (2021). Predicting regional coastal sea level changes with machine learning. *Scientific Reports*, 11(1):7650. 251

NOAA (2018). What is a gyre? *NOAA's National Ocean Service Website*. xvi, 20

NOC (2014). Scientific background: Rapid climate change and the atlantic overturning circulation: scientific background. *National Oceanography Centre*. xvi, 19

Olafsson, J., Olafsdottir, S. R., Takahashi, T., Danielsen, M., and Arnarson, T. S. (2021). Enhancement of the North Atlantic CO_2 sink by Arctic Waters. *Biogeosciences*, 18(5):1689–1701. 191

Oppenheimer, M., Glavovic, B., Hinkel, J., van de Wal, R., Magnan, A., Abd Elgawad, A., Cai, R., Cifuentes, Jara, M., DeConto, R., Ghosh, T., Hay, J., Isla, F., Marzeion, B. Meyssignac, B., Sebesvari, Z., Portner, H., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegria, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., and Weyer, N. (2019). Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities. *In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, pages 321–445. 1, 252

O'Brien, K., Pelling, M., Patwardhan, A., Hallegatte, S., Maskrey, A., Oki, T., Oswald-Spring, U., Wilbanks, T., and Yanda, P. (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. *A*

Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC), pages 437–486. 3

Parnell, A. C. and Gehrels, W. R. (2015). *Using chronological models in late Holocene sea-level reconstructions from saltmarsh sediments*, chapter 32, pages 500–513. John Wiley & Sons, Ltd. 41

Parnell, A. C., Haslett, J., Allen, J. R., Buck, C. E., and Huntley, B. (2008). A flexible approach to assessing synchroneity of past events using Bayesian reconstructions of sedimentation history. *Quaternary Science Reviews*, 27(19-20):1872–1885. 27, 36, 196

Peltier, W. (1998). Postglacial variations in the level of the sea:implications for climate dynamics and solid earth geophysics. *Reviews of Geophysics*, 36:603 – 689. 25

Peltier, W. (1999). Global sea level rise and glacial isostatic adjustment. *Global and Planetary Change*, 20(2):93–123. 16

Peltier, W. (2004). Global Glacial Isostasy and the Surface of the Ice-Age Earth: The ICE-5G (VM2) Model and GRACE. *Annual Review of Earth and Planetary Sciences*, 32:111–149. xxxi, 17, 39, 42, 43, 49, 68, 74, 76, 77, 110

Piecuch, C. G., Huybers, P., and Tingley, M. P. (2017). Comparison of full and empirical Bayes approaches for inferring sea-level changes from tide-gauge data. *Journal of Geophysical Research: Oceans*, 122(3):2243 – 2258. 70

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, TUWien*. 4, 51, 93, 199, 219, 252

Plummer, M. (2015). Cuts in Bayesian graphical models. *Statistics and Computing*, 25(1):37–43. 50

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news*, 6(1):7–11. 51

Plummer, M., Stukalov, A., and Denwood, M. (2016). rjags: Bayesian graphical models using MCMC. *R package version*, 4(6). 4, 51, 93, 199

Porcu, E., Furrer, R., and Nychka, D. (2021). 30 Years of space-time covariance functions. *WIREs Computational Statistics*, 13(2):1512. 70

Pörtner, H. O., Roberts, D. C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., and Weyer, N. (2019). Summary for Policymakers: IPCC Special Report on the Ocean and Cryoshper in a Changing Climate. *IPCC*. 18, 188, 191

Pugh, D. (2004). *Changing sea levels: effects of tides, weather and climate*. Cambridge University Press. 26

Pugh, D. and Woodworth, P. (2014a). *Mean sea-level changes in time: Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes*, page 252–295. Cambridge University Press. 23, 35, 92

Pugh, D. and Woodworth, P. (2014b). *Tidal forces: Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes*, page 36–59. Cambridge University Press. 35, 197

Qiu, B. and Chen, S. (2012). Multidecadal sea level and gyre circulation variability in the northwestern tropical Pacific Ocean. *Journal of Physical Oceanography*, 42(1):193–206. 191

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 11

Rahmstorf, S. (2006). Thermohaline ocean circulation. *Encyclopedia of Quaternary Sciences*, 5. 18, 191

Rahmstorf, S., Box, J. E., Feulner, G., Mann, M. E., Robinson, A., Rutherford, S., and Schaffernicht, E. J. (2015). Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nature Climate Change*, 5(5):475–480. xxvi, 191, 192, 208, 209, 210, 217, 219, 249

Rasmussen, C. E. and Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. Technical report, MIT Press. 6, 97

Redfield, A. C. (1972). Development of a New England salt marsh. *Ecological monographs*, 42(2):201–237. 27, 37

Reimann, L., Vafeidis, A. T., and Honsel, L. E. (2023). Population development as a driver of coastal risk: Current trends and future pathways. *Cambridge Prisms: Coastal Futures*, 1:e14. 2

Rhines, P., Häkkinen, S., and Josey, S. A. (2008). Is oceanic heat transport significant in the climate system? *Arctic–Subarctic ocean fluxes: defining the role of the northern seas in climate*, pages 87–109. 18, 192

Roemmich, D. and Wunsch, C. (1985). Two transatlantic sections: Meridional circulation and heat flux in the subtropical North Atlantic Ocean. *Deep Sea Research Part A. Oceanographic Research Papers*, 32(6):619–664. 215, 248

Rossby, T. (1999). On gyre interactions. *Deep Sea Research Part II: Topical Studies in Oceanography*, 46(1):139–164. 193

Rovere, A., Stocchi, P., and Vacchi, M. (2016). Eustatic and relative sea level changes. *Current Climate Change Reports*, 2:221–231. 16

Roy, K. and Peltier, W. (2015). Glacial isostatic adjustment, relative sea level history and mantle viscosity: reconciling relative sea level model predictions for the U.S. East coast with geological constraints. *Geophysical Journal International*, 201(2):1156–1181. 39, 71

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2):319–392. 219, 252

Sachs, H. M., Webb III, T., and Clark, D. (1977). Paleoecological transfer functions. *Annual Review of Earth and Planetary Sciences*, 5(1):159–178. 28, 37, 196

Saher, M. H., Gehrels, W. R., Barlow, N. L., Long, A. J., Haigh, I. D., and Blaauw, M. (2015). Sea-level changes in Iceland and the influence of the North Atlantic Oscillation during the last half millennium. *Quaternary Science Reviews*, 108:23–36. 220

Seager, R., Battisti, D. S., Yin, J., Gordon, N., Naik, N., Clement, A. C., and Cane, M. A. (2002). Is the Gulf Stream responsible for Europe's mild winters? *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 128(586):2563–2586. 192

Shaw, T. A., Plater, A. J., Kirby, J. R., Roy, K., Holgate, S., Tutman, P., Cahill, N., and Horton, B. P. (2018). Tectonic influences on late Holocene relative sea levels from the central-eastern Adriatic coast of Croatia. *Quaternary Science Reviews*, 200:262–275. 6, 97

Shennan, I., Bradley, S., Milne, G., Brooks, A., Bassett, S., and Hamilton, S. (2006). Relative sea-level changes, glacial isostatic modelling and ice-sheet reconstructions from the British Isles since the Last Glacial Maximum. *Journal of Quaternary Science*, 21(6):585–599. 3, 25

Shennan, I., Bradley, S. L., and Edwards, R. (2018). Relative sea-level changes and crustal movements in Britain and Ireland since the Last Glacial Maximum. *Quaternary Science Reviews*, 188:143–159. 71

Shennan, I. and Horton, B. (2002). Holocene land- and sea-level changes in Great Britain. *Journal of Quaternary Science*, 17(5-6):511–526. 4, 13, 25, 96

Shennan, I., Long, A. J., and Horton, B. P. (2015). Handbook of Sea-Level Research. *John Wiley & Sons, Ltd.* 2, 26, 32, 36, 92, 188, 196

Shennan, I., Peltier, W. R., Drummond, R., and Horton, B. (2002). Global to local scale parameters determining relative sea-level changes and the post-glacial isostatic adjustment of Great Britain. *Journal of Quaternary Science*, 21:297–408. 25

Sherwood, O. A., Lehmann, M. F., Schubert, C. J., Scott, D. B., and McCarthy, M. D. (2011). Nutrient regime shift in the western North Atlantic indicated by compound-specific δ 15N of deep-sea gorgonian corals. *Proceedings of the National Academy of Sciences*, 108(3):1011–1015. xxvi, 208, 209, 210, 219, 252

Shugar, D. H., Walker, I. J., Lian, O. B., Eamer, J. B., Neudorf, C., McLaren, D., and Fedje, D. (2014). Post-glacial sea-level change along the Pacific coast of North America. *Quaternary Science Reviews*, 97:170–192. xvi, 14

Simpson, G. L. (2018). Modelling palaeoecological time series using generalised additive models. *Frontiers in Ecology and Evolution*, 6:149. 90, 247

Slangen, A., Carson, M., Katsman, C., Van de Wal, R., Köhl, A., Vermeersen, L., and Stammer, D. (2014). Projecting twenty-first century regional sea-level changes. *Climatic Change*, 124:317–332. 191

Sleep, N. H. (2002). Isostasy and Flexure of the Lithosphere. *Physics Today*, 55(10):57–59. 16

Spalding, M. D., Ruffo, S., Lacambra, C., Meliane, I., Hale, L. Z., Shepard, C. C., and Beck, M. W. (2014). The role of ecosystems in coastal protection: Adapting to climate change and coastal hazards. *Ocean & Coastal Management*, 90:50–57. 2

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639. 51

Spooner, P. T., Thornalley, D. J., Oppo, D. W., Fox, A. D., Radionovskaya, S., Rose, N. L., Mallett, R., Cooper, E., and Roberts, J. M. (2020). Exceptional 20th century ocean circulation in the northeast Atlantic. *Geophysical Research Letters*, 47(10). xxvi, 192, 208, 209, 210, 219, 252

Stammer, D., Cazenave, A., Ponte, R. M., and Tamisiea, M. E. (2013). Causes for Contemporary Regional Sea Level Changes. *Annual Review of Marine Science*, 5(1):21–46. 40

Stearns, R. B. and Engelhart, S. E. (2017). A High-Resolution Reconstruction of Late-Holocene Relative Sea Level in Rhode Island, USA. 76, 118, 220

Stearns, R. B., Engelhart, S. E., Kemp, A. C., Hill, T. D., Brain, M. J., and Corbett, D. R. (2023). Within-region replication of late Holocene relative sea-level change: An example from southern New England, United States. *Quaternary Science Reviews*, 300:107868. 6, 25, 97, 118, 220

Suzuki, T. and Tatebe, H. (2020). Future dynamic sea level change in the western subtropical North Pacific associated with ocean heat uptake and heat redistribution by ocean circulation under global warming. *Progress in Earth and Planetary Science*, 7(1):67. 18

Talley, L. D. (2003). Shallow, intermediate, and deep overturning components of the global heat budget. *Journal of Physical oceanography*, 33(3):530–560. 215, 248

Tebaldi, C., Ranasinghe, R., Vousdoukas, M., Rasmussen, D. J., Vega-Westhoff, B., Kirezci, E., Kopp, R. E., Sriver, R., and Mentaschi, L. (2021). Extreme sea levels at different global warming levels. *Nature Climate Change*, 11(9):746–751. 2

Thibodeau, B., Not, C., Zhu, J., Schmittner, A., Noone, D., Tabor, C., Zhang, J., and Liu, Z. (2018). Last century warming over the Canadian Atlantic shelves linked to weak Atlantic meridional overturning circulation. *Geophysical Research Letters*, 45(22):12–376. xxvi, 192, 208, 209, 210, 219, 252

Thompson, K. R. (1986). North Atlantic sea-level and circulation. *Geophysical Journal International*, 87(1):15–32. 191

Thornalley, D. J., Oppo, D. W., Ortega, P., Robson, J. I., Brierley, C. M., Davis, R., Hall, I. R., Moffa-Sanchez, P., Rose, N. L., Spooner, P. T., et al. (2018). Anomalously weak Labrador Sea convection and Atlantic overturning during the past 150 years. *Nature*, 556(7700):227–230. xxvi, 191, 192, 208, 209, 210, 219, 252

Törnqvist, T. E., Rosenheim, B. E., Hu, P., and Fernandez, A. B. (2015). *Radiocarbon dating and calibration: Handbook of Sea-Level Research*, chapter 23, pages 347–360. John Wiley & Sons, Ltd. 27, 36

Upton, M., Cahill, N., and Parnell, A. (2023a). Statistical modelling for relative sea-level data. In *Encyclopedia of Quaternary Science, 3rd Edition*. Elsevier. v, 10, 22, 23, 37

Upton, M., Parnell, A., and Cahill, N. (2023b). reslr: An R package for relative sea level modelling. *arXiv preprint arXiv:2306.10847*. 190, 197

Upton, M., Parnell, A., Kemp, A., Ashe, E., McCarthy, G., and Cahill, N. (2023c). A noisy-input generalised additive model for relative sea-level change along the Atlantic coast of North America. *arXiv preprint arXiv:2301.09556*. 90, 92, 95, 98, 99, 100, 101, 109, 189, 190, 198, 199, 202, 215, 218, 239, 247

Vacchi, M., Engelhart, S. E., Nikitina, D., Ashe, E. L., Peltier, W. R., Roy, K., Kopp, R. E., and Horton, B. P. (2018). Postglacial relative sea-level histories along the eastern Canadian coastline. *Quaternary Science Reviews*, 201:124–146. 71

Vacchi, M., Marriner, N., Morhange, C., Spada, G., Fontana, A., and Rovere, A. (2016). Multiproxy assessment of Holocene relative sea-level changes in the western Mediterranean: Sea-level variability and improvements in the definition of the isostatic signal. *Earth-Science Reviews*, 155:172–197. 17

Vallis, G. K. (2012). *Climate and the Oceans*. Princeton University Press. 189, 193

Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics*. Cambridge University Press. 189

Van De Plassche, O. (1986). *Sea-Level Research a manual for the collection and evaluation of data*. Geo Books. 2, 24, 26

Van de Plassche, O. (1991). Late Holocene sea-level fluctuations on the shore of Connecticut inferred from transgressive and regressive overlap boundaries in salt-marsh deposits. *Journal of Coastal Research*, pages 159–179. 220

Vignudelli, S., Birol, F., Benveniste, J., Fu, L.-L., Picot, N., Raynal, M., and Roinard, H. (2019). Satellite altimetry measurements of sea level in the coastal zone. *Surveys in geophysics*, 40:1319–1349. 251

von Schuckmann, K., Minière, A., Gues, F., Cuesta-Valero, F. J., Kirchengast, G., Adusumilli, S., Straneo, F., Ablain, M., Allan, R. P., Barker, P. M., Beltrami, H., Blazquez, A., Boyer, T., Cheng, L., Church, J., Desbruyeres, D., Dolman, H., Domingues, C. M., García-García, A., Giglio, D., Gilson, J. E., Gorfer, M., Haimberger, L., Hakuba, M. Z., Hendricks, S., Hosoda, S., Johnson, G. C., Killick, R., King, B., Kolodziejczyk, N., Korosov, A., Krinner, G., Kuusela, M., Landerer, F. W., Langer, M., Lavergne, T., Lawrence, I., Li, Y., Lyman, J., Marti, F., Marzeion, B., Mayer, M., MacDougall, A. H., McDougall, T., Monselesan, D. P., Nitzbon, J., Otosaka, I., Peng, J., Purkey, S., Roemmich, D., Sato, K., Sato, K., Savita, A., Schweiger, A., Shepherd, A., Seneviratne, S. I., Simons, L., Slater, D. A., Slater, T., Steiner, A. K., Suga, T., Szekely, T., Thiery, W., Timmermans, M.-L., Vanderkelen, I., Wjiffels, S. E., Wu, T., and Zemp, M. (2023). Heat stored in the Earth system 1960–2020: where does the energy go? *Earth System Science Data*, 15(4):1675–1709. 188

Von Storch, H. and Zwiers, F. W. (2002). *Statistical analysis in climate research*. Cambridge University Press. 3

Walker, J. S., Kopp, R. E., Little, C. M., and Horton, B. P. (2022). Timing of emergence of modern rates of sea-level rise by 1863. *Nature communications*, 13(1):1–8. 72

Walker, J. S., Kopp, R. E., Shaw, T. A., Cahill, N., Khan, N. S., Barber, D. C., Ashe, E. L., Brain, M. J., Clear, J. L., Corbett, D. R., and Horton, B. P. (2021). Common Era sea-level budgets along the U.S. Atlantic coast. *Nature Communications*, 12(1):1841. 3, 6, 21, 33, 35, 40, 73, 76, 89, 98, 118, 190, 220, 246

Warrick, R. A., Barrow, E. M., and Wigley, T. M. L. (1993). *Climate and Sea Level Change: Observations, Projections and Implications*. Cambridge University Press. 13

Wenzel, M. and Schröter, J. (2010). Reconstruction of regional mean sea level anomalies from tide gauges using neural networks. *Journal of Geophysical Research: Oceans*, 115(C8). 73, 247

Whitehouse, P., Milne, G., and Lambeck, K. (2021). *Glacial Isostatic Adjustment*, pages 383–413. Springer International Publishing. xvi, 16, 17

Whitehouse, P. L. (2018). Glacial isostatic adjustment modelling: historical perspectives, recent advances, and future directions. *Earth Surf. Dynam*, 6:401–429. 16, 39, 74, 100, 202

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. 102

Wood, S. and Wood, M. S. (2015). Package 'mgcv'. *R package version*, 1(29):729. 56, 91

Wood, S. N. (2006). Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models. *Biometrics*, 62(4):1025–1036. 7, 47, 99, 203

Wood, S. N. (2016). Just another gibbs additive modeller: interfacing jags and mgcv. *arXiv preprint arXiv:1602.02539*. 56

Wood, S. N. (2017a). *Generalized additive models: an introduction with R*. CRC Press. 7, 33, 47, 70, 99, 106, 246

Wood, S. N. (2017b). P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Statistics and Computing*, 27(4):985–989. 99

Woodworth, P. and Player, R. (2003). The Permanent Service for Mean Sea Level: An update to the 21st century. *Journal of Coastal Research*, 19:287–295. xvii, 23, 24, 35, 92, 103, 197

Woodworth, P. L., Gehrels, R. W., and Nerem, S. R. (2011). Nineteenth and Twentieth Century Changes in Sea Level. *Oceanography*, 24(2):80–93. 2, 217, 249

Wright, A. J., Edwards, R. J., and van de Plassche, O. (2011). Reassessing transfer-function performance in sea-level reconstruction based on benthic salt-marsh foraminifera from the Atlantic coast of NE North America. *Marine Micropaleontology*, 81(1-2):43–62. 28

Wright, A. J., Edwards, R. J., van de Plassche, O., Blaauw, M., Parnell, A. C., van der Borg, K., de Jong, A. F., Roe, H. M., Selby, K., and Black, S. (2017). Reconstructing the accumulation history of a saltmarsh sediment core: Which age-depth model is best? *Quaternary Geochronology*, 39:35–67. 27, 36

Wöppelmann, G., Zerbini, S., and Marcos, M. (2006). Tide gauges and Geodesy: a secular synergy illustrated by three present-day case studies. *Comptes Rendus Geoscience*, 338(14):980–991. La Terre observée depuis l'espace. 2, 23, 35

Yin, J., Griffies, S. M., and Stouffer, R. J. (2010). Spatial variability of sea level rise in twenty-first century projections. *Journal of Climate*, 23(17):4585–4607. 191, 193

Yin, J., Schlesinger, M. E., and Stouffer, R. J. (2009). Model projections of rapid sea-level rise on the northeast coast of the United States. *Nature Geoscience*, 2(4):262–266. 216, 249