

AN ABSTRACT OF THE DISSERTATION OF

Kai Parker for the degree of Doctor of Philosophy in Civil Engineering presented on September 6, 2018

Title: Climate Change Impacts to Estuarine Hydrodynamics: Tools, Methodologies, and Predictions

Abstract approved: _____

David F. Hill

Coastal communities in the U.S. Pacific Northwest (PNW) are rapidly engaging with the idea that hazards and environmental pressures are changing and may be characteristically different in the near future. This has led to a need for scientific knowledge and tools that can help coastal communities prepare and build resilience in the face of changing risk. This is especially true for coastal systems that experience multiple and complex hazard drivers. In these systems, extreme events can result from the sum of multiple drivers and it is unclear how the changing climate will impact the overall sum. This dissertation focuses on such a system, specifically estuaries, and provides key steps towards characterizing how the risk of flooding will change through time. This is broadly approached through developing tools, methodologies, and predictions for PNW estuaries, but the developed tools and approaches should be applicable to other regions as well.

Estuarine hydrodynamics are changing as a result of the changing climate. Our current best predictions for how the climate is evolving come from global climate models (GCMs).

However, climate models often exhibit bias when compared to observations. These biases can propagate through a modeling framework and lead to corresponding biases in predicted flood magnitude. Chapter 2 of this dissertation evaluates a tool, statistical bias correction (BC), for removing bias from model output. This chapter examines a variety of BC procedures as applied to GCM-forced wave model output. BC methodologies are generally applicable but waves are chosen for testing due to large biases (from poorly resolved GCM wind fields) and because waves are significant contributors to coastal flooding in the PNW. BC is shown to be an efficient and effective method for removing bias. Particular emphasis is placed on evaluating the difference between univariate and bivariate analysis with results demonstrating the importance of considering correlation between variables when performing corrections. All methods are evaluated under a nonstationary climate with extensions developed to BC methodologies for this new context.

A primary difficulty with quantifying future flooding hazards in estuaries is that extremes are controlled by a variety of forcing (e.g. waves, streamflow, El Nino, winds, etc.), all of which may change in the future. This is investigated in Chapter 3 using a multi-component modeling framework which considers the combined evolution of estuarine forcings under climate change. Forcings are simulated with sub-models and then used as inputs for decadal-scale hydrodynamic simulations. This modeling framework is utilized for two climatological periods (historic and future) and for two study sites (Coos and Tillamook Bays). It is found that extreme water levels (e.g., 100 year return interval events) exhibit spatial variability with magnitudes increasing with distance into the estuaries. For the climate model studied, forcing of extreme events relaxes into the future, but this effect is eventually overcome by sea level rise (SLR). Overall results show significant system complexity, highlighting the importance of physically modeling multiple processes relevant to flooding rather than considering a simple SLR adjustment.

The process-based modeling in Chapter 3 has the advantage of resolving many physical processes, but has the disadvantage of requiring significant computational resources. Chapter 4 evaluates an alternative hybrid methodology (process-based and statistical) that seeks to calculate water levels with much greater computational efficiency. Specifically, an emulation-based approach is evaluated which replaces the process-based model with a fast statistical representation. Emulation is validated against a variety of observations and is found to be skillful with reasonably low overall error. A hierarchical validation finds that the largest loss of skill in

emulator construction is associated with the process-based model attempting to reproduce observations. After emulator construction, simulation time is nearly instantaneous, making emulation a promising strategy for a variety of research questions such as probabilistic future flood hazard assessments.

©Copyright by Kai Parker
September 6, 2018
All Rights Reserved

Climate Change Impacts to Estuarine Hydrodynamics: Tools, Methodologies, and Predictions

by
Kai Parker

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented September 6, 2018
Commencement June 2019

Doctor of Philosophy dissertation of Kai Parker presented on September 6, 2018.

APPROVED:

Major Professor, representing Civil Engineering

Head of the School of Civil and Construction Engineering

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Kai Parker, Author

ACKNOWLEDGEMENTS

To Dave Hill, I would like to express sincere appreciation and gratitude. As my advisor your continuous encouragement, support, and guidance was a major source of strength in making this dissertation possible. You taught me a lot about what it means to be an academic and how to swim in these unknown waters.

To my second “unofficial advisor”, Peter Ruggiero, your belief and support in me and my sometimes outlandish ideas were critical in keeping me going and growing as a researcher. Your enthusiasm for science and discovery were fuel to keep me excited about the world waiting to be explored. Thank you for everything you put into this thesis and my development as a researcher.

To my other committee members, Tuba Ozkan-Haller, Sarah Emerson, and Emily Shroyer, I would like to express sincere thanks. Your time, guidance and feedback helped to make this thesis stronger. The other faculty members in coastal and ocean engineering, ocean engineering, and statistics were all foundational in helping me understand my corner of the coastal engineering world. Little bits and pieces of you all are scattered about this thesis.

To my many friends in Corvallis, you were the support structure that holds this structure up. You made life here in Oregon wonderful and I couldn’t have asked for a better group of people to spend the last 5 years with. Writing this thesis was bitter sweet knowing it meant closing the curtain on this stage of life that was spent with you all.

Finally to Isabel, you are the person who made this all possible. You convinced me everything was going to be ok when I was overworked and on the edge of giving up. You listened to me babble on and on about ideas when I was over-excited about research. Playing with you in the mountains and streams of Oregon is what recharged my batteries and my mind. You are the counterbalance to this whole unstable out-of-control apparatus and I would make you a co-author on this whole thesis if I could.

I would also like to express thanks to Oregon Sea Grant for supporting various aspects of my research both through a grant (project number R/CNH-25) and the Robert E. Malouf marine studies scholarship (Project number E/INT-143). Beyond this, Sea Grant provided a scientific family of others committed to using science to directly help coastal communities. Your steadfast

commitment to this goal changed the very way I view science. I was additionally supported by the NOAA Regional Integrated Sciences and Assessments Program (RISA) (NA15OAR4310145) through OSU's CIRC program. Working with CIRC further fueled my commitment to help bridge the gap between hard science and coastal stakeholders. While this research is not featured in this thesis, working with CIRC was invaluable in forming my mentality as a researcher. Finally I would like to thank the ARCS program, especially my donor Ann Emerson. Your kindness, interest in my research, and financial support were critical in making this PhD possible. Thank you.

CONTRIBUTION OF AUTHORS

Dr. David F. Hill contributed ideas and revisions to Chapters 1 through 4.

Dr. Peter Ruggiero contributed ideas and revisions for Chapter 4.

Dr. Gabriel García-Medina contributed wave model results for Chapter 3

Dr. Jordan Beamer contributed hydrological modeling to Chapter 3

Dr. Katy Serafin assisted with formulation of ideas for Chapter 5 as well as data analysis on tide gauge decomposition.

TABLE OF CONTENTS

	<u>Page</u>
1. Chapter 1 – General Introduction	1
1.1 Estuaries in the Pacific Northwest	2
1.2 Climate Change in the PNW	3
1.3 Research Strategies	4
1.3.1 A process-based modeling perspective	4
1.3.2 A statistical (emulator) based perspective	5
1.4 Organization of Dissertation	5
2. Chapter 2 - Evaluation of Bias Correction Methods for Wave Modeling Output	7
Abstract:.....	8
2.1 Introduction.....	8
2.2 Data.....	13
2.3 Methods	15
2.3.1 Bias Correction Methods	15
2.3.2 Validation Methods.....	23
2.4 Results.....	25
2.4.1 Statistical Bias.....	25
2.4.2 Correlation	29
2.4.3 Bivariate PDF/CDF Shape and RMSE	30
2.4.4 Extreme Quantiles.....	32
2.5 Discussion.....	34
2.6 Conclusions.....	37
Acknowledgments	38
3. Chapter 3 - The Effects of Changing Climate on Estuarine Water Levels: A Pacific Northwest Case Study.....	39
Abstract:.....	40
3.1 Introduction.....	40
3.2 Study Sites	43
3.3 Methods	44
3.3.1 Climate Model	45
3.3.2 Wave Model.....	46

TABLE OF CONTENTS (Continued)

	<u>Page</u>
3.3.3 Hydrological Model	47
3.3.4 Monthly Anomaly Model	48
3.3.5 Sea Level Rise.....	50
3.3.6 Local Hydrodynamic/Wave Model.....	50
3.4. Nonstationary Recurrence Intervals.....	52
3.5 Results.....	53
3.5.1 Recurrence Intervals (Tide Gauge Locations)	54
3.5.2 Recurrence Intervals (Spatially Variable).....	55
3.5.3 Changes to RI spatial variability.....	58
3.6. Discussion.....	59
3.6.1 Drivers of Extreme WLs	59
3.6.2 Spatial variability in RIs	61
3.6.3 Changing Extreme Events.....	62
3.6.4 Modeling limitations.....	64
3.7 Conclusions.....	66
Acknowledgments	67
4. Chapter 4 - Emulation as an Approach for Rapid Estuarine Modeling	68
Abstract:.....	69
4.1 Introduction.....	69
4.2 Methods	72
4.2.1 Study Site	73
4.2.2 Observational Data.....	74
4.2.3 Simulator Configuration	76
4.2.4 Dimensional Reduction and Levels of Simplifications	77
4.2.5 Experimental Design.....	83
4.2.6 Emulator Configuration	84
4.3 Results.....	86
4.3.1 Error Introduced by Model Simplifications.....	86
4.3.2 Emulator Validation.....	87
4.3.3 Emulator Performance: Westport, WA Tide Gauge	89

TABLE OF CONTENTS (Continued)

	<u>Page</u>
4.3.4 Emulator Performance: USACE field campaign	91
4.3.5 Example Emulator Application: Extreme WL decomposition	93
4.4 Discussion	95
4.4.1 Effect of simulator simplifications.....	95
4.4.2 Computational Cost Considerations	97
4.4.3 Emulation Beyond WLs.....	99
4.4.4 Extreme WL Event Decomposition.....	101
4.5 Conclusions.....	102
Acknowledgments	104
5. Chapter 5 – General Conclusion	105
5.1 Chapter Conclusions	105
5.2 Conclusions on Methodologies.....	106
5.3 Future Work.....	109
Bibliography	111
Appendix.....	134
Appendix A: Nonstationary Extensions for the Binning and Direct Method	135
The Binning Method (Future).....	135
The Direct Method (Future).....	136

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Figure 1: Site map showing the buoy locations used for bias correction marked as stars and all other buoys (both historic and active) marked with dots. The external tick marks represent the WW3 model 1 degree resolution grid.....	9
Figure 2: Monthly means of significant wave height (H_s) and peak wave period (T_p) at each buoy location. Data shown are observational (black), historic model (red) and future model (blue).....	14
Figure 3: The Quantile Mapping procedure as applied to a single historic model value.	15
Figure 4: Equidistant Quantile Matching as applied to a single future model value.....	18
Figure 5: The Binning Method as applied to the second variable of a bivariate bias correction. Example deciles of bivariate CDF are marked with red dots.	19
Figure 6: The Direct Method as applied to the second variable of a bivariate bias correction.....	21
Figure 7: The Shuffle Method (Schaake Shuffle). Numbers indicate rank in terms of magnitude. Color gradient indicates time with values later in the time series being darker.....	22
Figure 8: Bias Correction results for the first three statistical moments: Mean, Standard Deviation, and Skewness. Raw column shows the agreement of raw model output with observations. All data are presented as relative differences (RD) and all boxplots represent an ensemble of 40 data points. Note that some outliers are not visible due to y-axis scaling.	28
Figure 9: Bias Correction results for correlation coefficients: Kendall's τ , Spearman's ρ , and Pearson's r . Raw column shows the agreement of raw model output with observations. All data are presented as relative differences and all boxplots represent an ensemble.	29
Figure 10: Comparison of bivariate PDFs (left columns) and CDFs (right columns) for the summer seasons of the historic period. Plotted results are an average of all 10 jack-knife bins. All subplots have the same scale and axes as shown on the Raw Model subplot.	31
Figure 11: Root mean square errors for comparison of bivariate PDFs and CDFs.	32
Figure 12: Bias Correction results for extreme quantiles (95th and 99th). Raw column shows the agreement of raw model output with observations. All data are	

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
presented as relative differences (RD) and all boxplots represent an ensemble of 4 data points.	33
Figure 13: Overview map (left) and detail views (inset boxes) of Tillamook and Coos Bays on the Oregon coastline. Both insets are at the same scale (shown on Tillamook Bay inset).	43
Figure 14: Overall general modeling framework. The dark gray triangle labeled AOGCM in the top left corner is the “parent” model. Dark gray rectangles represent sub-models. Light gray ovals represent variables that are passed between modeling components.	45
Figure 15: Station output locations at Tillamook and Coos Bay study sites, panel (a) and panel (b) respectively. Grey arrows show hydrologic inputs into the model domain. Red points and numbers represent transect station locations (see section 3.6.1). Both figures are at the same scale (shown in panel a).	47
Figure 16: MMSLA regression for the Tillamook study site in the format of Thompson et al. [2014]. Fitted contributions to MMSLA from predictor variables τ_{eq} , τ_{ls} , and τ_{xy} are shown as thin black lines (full, dotted and dashed respectively). The bold black line is the observed MMSLA signal while the bold red line is the total fitted MMSLA signal.	49
Figure 17: Effective RI curves for Tillamook Bay. The increase in effective RIs through time corresponds to the rate of SLR. Black dots show the simulated max annual WLs at the tide gauge (in NAVD 88).	53
Figure 18: RIs for the Tillamook and Coos Bay tide gauges, panel (a) and panel (b) respectively. WLs are in NAVD88. Confidence intervals are calculated using a likelihood-based method and are plotted as dotted lines. SLR has not been included for future RI curves.	54
Figure 19: Coos Bay effective RI WLs for the historic (dashed line) and future (solid line) periods. The location of intersection between the historic and future effective RI curves is plotted on the time axis as a dot with color corresponding to return period.....	55
Figure 20: Historic period 100-year flood surface results for Tillamook (a) and Coos (b) Bays. Individual stations are plotted with the color scale indicating the individual 100-year RI WL magnitude. Note that the color scale is different between the two study sites. The calculated flood surface is plotted as a blue transparent region.....	56

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
Figure 21: Comparison of 100-year flood inundation zones for FEMA (blue cross hatch) and this study (black cross hatch). The plotted region is the lower eastern region of Coos bay. The calculated 100-year flood zone is limited to the DEM domain (shown by the topo/bathy DEM) while the FEMA flood zone extent is larger.	57
Figure 22: Changes to 100-year RIs plotted at output stations. Changes are calculated as the future effective RI WL (at year 2050, 0.172 m of SLR) minus the historic RI WL. The dark black line represents the edge of the computational domain and light gray lines are the -5, 0, and 5 meter contours.	58
Figure 23: WL contributions from various forcings during the largest simulated WL event. Stations transition from the tide gauge (station 1) to the far river outlet of the estuary (see Figure 15). Panel (a) and (b) are for the Coos historic and future periods. Panel (c) and (d) are for the Tillamook historic and future periods. All panels have the same y-axis scaling.....	59
Figure 24: Conceptual framework for developing an estuarine hydrodynamic emulator.	72
Figure 25: Grays Harbor, WA study site and locations of observational datasets. Circles and triangles represent USACE deployments with co-located instruments labeled with a single number representing both WL and Hs stations. The main panel shows the bathymetry and topography of the local estuary in the NAVD88 Datum. The inset panel shows the larger geographical context of the estuary with the thin black line delineating the domain of the hydrodynamic model.....	74
Figure 26: Hierarchy of model simplifications between observations (top) and emulator output (bottom). Each simplification is associated with a level (right side of the figure) and some amount of error (defined on the left side of the figure).	78
Figure 27: Panel (a): Comparison of NTR during Storm 1 from a fully dynamic simulation (black line) and simplified stationary simulations (black dots). Panel (b): Example stationary run (at the peak of the storm) showing how the stationary NTR is calculated. The horizontal bold dotted line represents the time of the stationary run. At this time, NTR is calculated by subtracting the value of a tide only run from the value of the stationary run (Bold red line). This NTR value is plotted as a red outlined dot in panel (a).	83
Figure 28: Panel (a) shows three sample functions drawn from the prior distribution. Panel (b) shows 3 sample functions from the posterior distribution after 4 training observations (dark black points). In panel (b), the shaded region represents the 95% Bayesian credible interval. Figure after Rasmussen and Williams [2006].....	85
Figure 29: Error between calculated WLs between full and stationary (Level 1 and Level 3) simulator runs. The histogram includes comparisons for Storms 1-5 (Table 3).	

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
The bold line represents zero error while the dotted line is the mean error of the ensemble. Subplot (a) is computed at the tide gauge location while subplot (b) is at WL station 7 (see Figure 25).....	87
Figure 30: Panel (a): Simulated vs. emulated WLs for the full training dataset. WLs were produced using a 5-fold validation procedure. Panel (b): histogram of the error between simulated and emulated WLs.....	88
Figure 31: Comparison of emulated (red dashed line) and observed (black line) hourly WLs at the Grays Harbor tide gauge for January 2010. Coefficient of determination is for the entire tidal record (2006 - 2016).....	89
Figure 32: Tide Gauge comparison of observed and modeled NTR at various simplification levels. Each subpanel (a through e) is one of the top 5 storms of record (Table 3). Due to windowing for spectral filtering, storm 4 observed NTR is calculated using the subtraction method rather than the Bromirski method. All panels have the same y-axis scaling.....	90
Figure 33: Comparison of observations (USACE deployments, Figure 25), full simulator and emulated WLs. All panels have the same y-axis scaling.	92
Figure 34: Average WL contribution from forcing components during extreme events. Two transects are plotted with subplot (b) showing transect locations with approximate station locations marked as ticks. Subplot (a) is the East-West transect and subplot (c) is the North-South transect.	94
Figure 35: Impact of training dataset size on emulator skill. RMSE is calculated with a 5-fold cross-validation and represented by a standard boxplot. The dotted vertical line represents the theoretical dataset size from Loepky et al., [2009].....	99
Figure 36: Comparison of observations (USACE deployments), full simulator and emulated Hs time series. Symbols are used for the observations for the sake of visual clarity. All panels have the same y-axis scaling.....	100
Figure 37: Binning Method using equidistant quantile mapping.....	136
Figure 38: Direct Method using equidistant quantile mapping.	138

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 1: Coos and Tillamook estuarine characteristics (after Engle et al., [2007]). Wind characteristics are at gauge locations. Wave characteristics are offshore at buoys 46002 and 46005. Streamflow values are from this study's hydrological analysis	43
Table 2: Comparison of forcing for Tillamook Bay under the historic and future climatological periods. “Ave. Overall” is a full time series average while “Ave. Annual Max” is an average of forcing during observed annual WL maximum events.....	63
Table 3: Summary of forcing for the five largest NTR events at Grays Harbor, WA. Forcing values are reported at the occurrence of maximum NTR.	76
Table 4: RMSE values comparing WL model output to observations at various simplification levels. Rows are locations / variables while the two column groupings represent instrument deployments. The length of the time series comparison varies depending on station deployment length.	93
Table 5: RMSE values comparing Hs model output to observations at various simplification levels. Rows represent station locations while the two column groupings represent instrument deployments. Hs station 0/7 was relocated so deployment 1 values represent the Hs station 0 location and deployment 2 values representing the station 7 location.....	101

Chapter 1 – General Introduction

Throughout history, the coast has been a hub for human development, providing transportation, industry, food, aesthetics, and countless other motivations for settling along the shoreline. This has led to the coastal zone being one of the most densely populated on earth [Nicholls and Small, 2002], a trend that is projected to increase into the future [Neumann et al., 2015]. While coastlines provide many diverse benefits, they additionally represent an added risk in the form of coastal hazards such as large waves, storm surges, erosion, and flooding. Society's disproportionate investment in coastlines means that coastal hazards represent a rapidly growing exposure to risk for communities, infrastructure, and economies [Jongman et al., 2012]. Desire to manage this risk has resulted in a demand for coastal hazards research and for tools capable of accurately quantifying and predicting hazards.

The push for advancing coastal hazards science has recently come into a new focus through the lens of climate change. The modern era has forced an abandonment of baseline assumptions of stationarity in our coastal systems [Milly et al., 2008]. For hazards research, this new paradigm has necessitated a broad reframing of the very definition of risk [Read and Vogel, 2015]. How do we understand common concepts like 100-year return interval events when we can no longer rely on the past as a valid analog for the future? How do we incorporate extreme uncertainty in the future climate into our hazard assessments and how we build our communities? These questions and many more are especially pertinent in the context of coastal research where the effects of a changing ocean (via sea level, chemistry, temperature, dynamics and many more) are already visible in observational records [Rintoul et al., 2013].

This dissertation brings these themes together through the goal of trying to better understand the future of coastal hazards in a changing climate. In particular this work is focused on flooding hazards in Pacific Northwest estuaries. The following subsections describe the general context of this research question as well as this dissertation's strategy for approaching the problem.

1.1 Estuaries in the Pacific Northwest

This dissertation is focused on the U.S. Pacific Northwest (PNW), both as a study site and as a testbed for the ideas developed within. The PNW is subject to strong wave forcing with winter significant wave heights averaging 3-4 meters and annual events commonly exceeding 10 meters [Ruggiero et al., 2010]. This is coupled with an equally energetic storm climate highlighted by extratropical cyclones that produce extreme coastal winds and precipitation [Allan and Komar, 2002a; Mass and Dotson, 2010]. Tides are mesotidal and observed water levels exhibit a variety of large-scale oscillations, the most visible of which is the El Nino - Southern Oscillation (ENSO) [Chelton and Davis, 1982]. The PNW has one of the雨iest climates in the United States producing significant freshwater discharges to the coast which can dominate nearshore dynamics [Mazzini et al., 2014]. From a coastal hazards perspective, this results in an environment defined by significant system complexity and a broad range of hazard drivers [Cheng et al., 2015b; Olabarrieta et al., 2011].

As a function of diverse forcing, PNW extreme events are often the result of compound events [Serafin et al., 2014]. This means that extremes are commonly caused by combinations of forcings for which individual components may not necessarily be extreme [Moftakhari et al., 2017; Wahl et al., 2015; Leonard et al., 2014]. Consideration of compound events and changes to their probability of occurrence under climate change has been shown to be of considerable importance to flooding [Wahl et al., 2015]. However, compound events are difficult to model as they require multivariate considerations and are challenging to predict. This precludes commonly used event based approaches which assume that extreme water levels are always associated with a certain type of event (generally hurricanes) [Orton et al., 2016]. Extremes are less clear in the PNW with non-intuitive combinations of forcing leading to flooding.

This research additionally focuses on estuarine systems rather than open coastlines. This is for a variety of reasons, the first of which is that estuaries are of considerable societal importance. Estuaries provide a wide range of ecosystem services [Barbier et al., 2011] as well as often being hubs of coastal infrastructure due to harbors and ports. The second reason is that flooding in PNW estuaries is less well understood than open coasts. Flooding risk along PNW open coastlines has seen significant research in recent decades [Serafin and Ruggiero, 2014;

Barnard et al., 2014; Ruggiero, 2013]. This is less true for estuaries, mainly due to a larger complexity and broader range of forcing controls. This mismatch between societal importance and relative understanding make PNW estuaries a rich subject for research.

1.2 Climate Change in the PNW

Coastal communities in the PNW are actively engaged with climate change research because they are already seeing the effects of a changing environment. For example, research has documented an increase in extratropical storms in recent decades [Graham and Dias, 2001; Favre and Gershunov, 2006]. Similarly, and likely linked, there is evidence of wave climate intensification [Allan and Komar, 2002b; Allan and Komar, 2006; Ruggiero et al., 2010]. Tide gauge analysis shows that sea level is rising at around 1.6 mm/year for the PNW, although this rate is currently being buffered by vertical land motion [National Research Council, 2012]. Climate projections show wetter winters and drier summers for the PNW [Mote and Salanthe, 2010] so stream hydrographs are additionally expected to change. PNW streams are particularly vulnerable to climate change as a large percentage of the snowpack is at or near the freezing point [Watts et al., 2016].

How these various changes will combine to produce future flooding hazards remains a significant unknown. This is particularly true for estuarine systems because of the numerous pathways through which climate change can affect estuarine hydrodynamics. Considerable research has shown that sea level rise (SLR) is expected to intensify flooding events in estuaries [Dawson et al., 2007; Rilo et al., 2013; Smith et al., 2010]. However the specific impact of SLR on flooding has been shown to be complex and non-linear [Atkinson et al., 2013, Ding et al., 2013]. Furthermore, SLR impact has been shown to be coupled with other processes (such as changes to tropical cyclones) further increasing exposure to flood risk [Frey et al., 2010; Mousavi et al., 2011]. Even dynamics that are often considered deterministic, such as tides, have been shown to be variable into the future. Specifically, SLR can modify tidal wave propagation speed, basin shape, resonance and other factors creating spatially complex and sometimes drastic changes to tidally induced water levels [Pelling and Green, 2013].

Overall there is both evidence that the climate in the PNW is changing and that this will impact estuarine flooding hazards. This dissertation hopes to provide tools and conclusion for understanding the future of estuarine flood hazards in the PNW.

1.3 Research Strategies

This thesis investigates climate change impacts to flooding using two approaches. A fully process-based approach is implemented as well as a “hybrid” approach that is targeted at bringing statistical model simulation speeds to process-based models. Exploring the same question from different perspectives allows a holistic view of the scope of the problem as well as a deeper understanding of the built in biases, strengths, and limitations of each methodology. The following two sections describe the research strategies implemented within each dissertation chapter.

1.3.1 A process-based modeling perspective

Chapter 3 of this dissertation approaches climate change impacts to estuarine flooding using a process-based modeling framework. This approach allows an exploration of how multiple changing forcing components will result in an integrated change to estuarine hydrodynamics. Research has shown that considering a holistic view of climate change, through viewing multiple changing physical processes, is vital as the elements of extreme water levels are coupled [Wolf, 2009; Odigie and Warrick, 2018] and nonlinear [Ding et al., 2013]. This is accomplished by using a variety of sub-models that track how each forcing component of extreme water levels (e.g. waves, tides, winds, streamflow, etc.) evolves under a changing climate. The outputs from these sub-models provide inputs to a hydrodynamic model which is run continuously for the periods of interest. Climate change impacts are quantified through decadal scale simulations run for historic and future periods at two study sites.

A process-based modeling perspective requires some quantification of change to the system. This is commonly provided by global climate models (GCMs), although statistical [Wahl et al., 2015] or semi-empirical projections [Rahmstorf, 2007] are also used. GCMs rely on our

understanding of physical mechanisms to predict how changes to the environment will result in changes to the climate [Solomon et al., 2007]. The complexity of global climate physics means that there are a variety of issues in GCM predictions [Parker, 2011; Petersen, 2000], including systematic bias and an inability to reproduce observations [Jun et al., 2008; Maraun et al., 2010]. These issues become problematic when GCM data serve as forcing for other models, transferring the bias and potentially resulting in unphysical results [Christensen et al., 2008]. This problem was encountered when trying to integrate GCM data into Chapter 3's modeling framework and so was further explored. Chapter 2 details a study investigating a variety of methodologies for correcting GCM bias, although through the specific lens of wave modeling.

1.3.2 A statistical (emulator) based perspective

One limitation of process-based models is that they are computationally expensive which can limit the types of questions that can be asked. For example, the extensive computation time required for the approach taken in Chapter 3 forced the usage of a single climate model. This is problematic in that climate models predict a wide range of potential future climates [Solomon et al., 2007]. Being able to encapsulate this uncertainty into future hazard assessments makes probabilistic approaches an attractive option for climate change studies. This said, probabilistic modeling requires multiple model implementations and this is currently not viable for flooding hazard studies of the complexity exhibited in Chapter 3. This sets up a strong motivation to develop some way of bridging computationally intensive physical models and fast simulation times. Chapter 4 explores an approach to this problem through an “emulator” that uses a statistical representation of a physical model to rapidly produce water level predictions.

1.4 Organization of Dissertation

This dissertation consists of three core chapters, each of which is in journal manuscript form covering a distinct topic in regards to this thesis's overall research question.

- Chapter 2: Evaluation of Bias Correction Methods for Wave Modeling Output
 - Published: Ocean Modeling, 2017.

- Chapter 3: Climate Change Impacts to estuarine flooding: A Pacific Northwest Case Study
 - In Prep: Coastal Engineering, 2018
- Chapter 4: An Emulation Based Approach to Probabilistic Estuarine Flood Hazard Assessment
 - In Prep: JGR Oceans, 2018

These chapters are followed by a brief discussion and synthesis in Chapter 5.

Chapter 2 - Evaluation of Bias Correction Methods for Wave Modeling Output

Kai Parker, David Hill

Abstract:

Models that seek to predict environmental variables invariably demonstrate bias when compared to observations. Bias correction (BC) techniques are common in the climate and hydrological modeling communities, but have seen fewer applications to the field of wave modeling. In particular there has been no investigation as to which BC methodology performs best for wave modeling. This paper introduces and compares a subset of BC methods with the goal of clarifying a “best practice” methodology for application of BC in studies of wave-related processes. Specific focus is paid to comparing parametric vs. empirical methods as well as univariate vs. bivariate methods. The techniques are tested on global WAVEWATCH III historic and future period datasets with comparison to buoy observations at multiple locations. Both wave height and period are considered in order to investigate BC effects on inter-variable correlation. Results show that all methods perform uniformly in terms of correcting statistical moments for individual variables with the exception of a copula based method underperforming for wave period. When comparing parametric and empirical methods, no difference is found. Between bivariate and univariate methods, results show that bivariate methods greatly improve inter-variable correlations. Of the bivariate methods tested the copula based method is found to be not as effective at correcting correlation while a “shuffling” method is unable to handle changes in correlation from historic to future periods. In summary, this study demonstrates that BC methods are effective when applied to wave model data and that it is essential to employ methods that consider dependence between variables.

Keywords: Wave Modeling, Bias Correction, Climate Change, Multivariate Statistics

2.1 Introduction

One of the key drivers for the development of wave models has been the need for high resolution data distributed across both large areas and large time windows. These data are integral for navigation, hazard forecasting, recreational purposes, and a broad array of ocean science applications. Due in part to high operational costs, the observational in-situ record is sparse and cannot practically cover all areas of the ocean at all times (Figure 1). Therefore,

models serve to “fill the gaps” and provide a more complete understanding of the wave climate. Models are further essential for any study of future wave conditions where data clearly do not exist. However, data from wave models can and do consistently exhibit bias (defined in this study as a systematic deviation from the corresponding observed “true value”) that results from a variety of factors including inherent simplifications and inadequate model physics (parameterizations, assumptions, etc.), numerical solution schemes, resolution, insufficient or imperfect calibration datasets, and incorrect boundary forcing data.

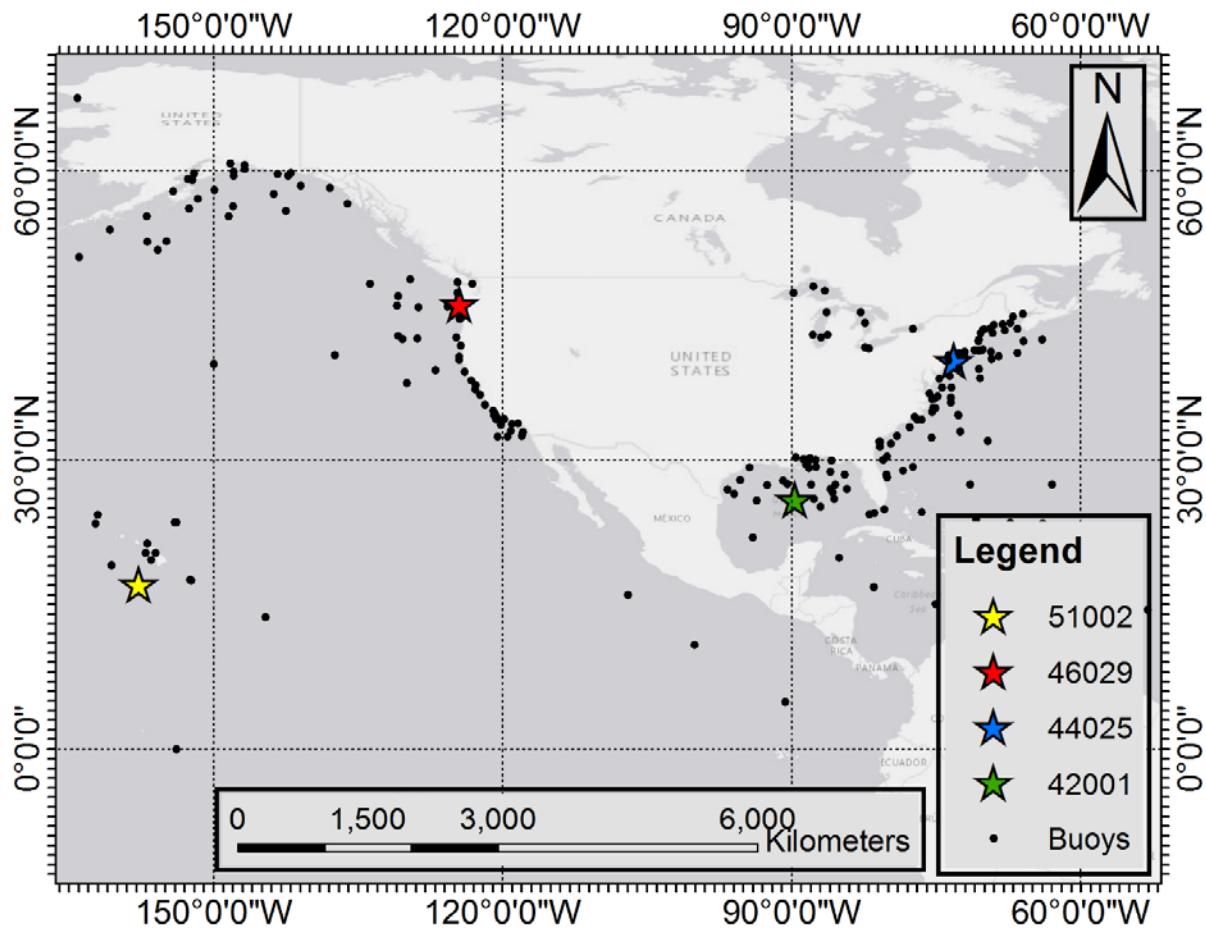


Figure 1: Site map showing the buoy locations used for bias correction marked as stars and all other buoys (both historic and active) marked with dots. The external tick marks represent the WW3 model 1 degree resolution grid.

Model bias is not unique to the field of ocean sciences. In particular, the atmospheric and hydrologic science communities have developed a mature body of literature dealing with the

subject. This is primarily due to a reliance on general circulation model (GCMs), which are highly prone to bias [Mehran et al., 2014; Mueller and Seneviratne, 2014]. Since climate model output is often the input for other models, GCM biases propagate downstream and detrimentally impact other modeling results [Xu, 1999; Christensen et al., 2008]. To resolve this, methods for bringing model output back into alignment with observations have been sought. This demand is the foundational driver for the development of bias correction (BC) procedures.

At the conceptual level, BC methods (as explored in this paper) define a transfer function that transforms model data to a new dataset with fewer statistical biases. How this transfer function is defined ranges from a simple shift in the mean value to increasingly complex techniques that can fully correct statistical distributions. For example, the widely used quantile mapping methods [Panofsky and Brier, 1958; Wood et al., 2004; Déqué, 2007; Piani et al., 2009] attempt to match the CDF (Cumulative Distribution Function) of the model time series to that of a target, typically an observational time series. Recent advances in BC techniques of the type considered in this paper have expanded into the multivariate domain and attempt to incorporate the relationship between variables as well. In cases of dependent variables, well-intentioned univariate BC can lead to incorrect inter-variable correlations and non-physical results [Chen et al., 2011; Thrasher et al., 2012]. This is important in the field of wave modeling as wave parameters (height, direction, period, etc.) are highly correlated [Mathisen and Bitner-Gregersen, 1990; Ferreira and Soares, 2002; Repko et al., 2004; De Waal and van Gelder, 2006; Corbella and Stretch, 2012]. Recent contributions to multivariate BC differ in how they treat inter-variable relationships and include a data binning technique [Piani and Haerter, 2012], a direct bivariate distribution approach based on copulas [Li et al., 2014], and a shuffling technique [Vrac and Friederichs, 2014]. For convenience these methods will be called the Binning Method, the Direct Method, and the Shuffling Method respectively. These three BC techniques will be explained in detail in the methods section of the paper.

From a broad perspective, many modeling practices can be considered a form of “bias correction.” For example, model tuning or data assimilation are both procedural ways of attempting to bring model output into agreement with observations. In the meteorological community, Model Output Statistics (MOS) are routinely used to remove bias in numerical weather prediction, albeit in a format that may be more easily recognized as statistical

downscaling. For brevity, this study does not consider all bias-reducing techniques. Instead, it focuses on statistical bias correction methods that correct statistical distributions of model variables. With this constraint, there are three options when looking at bias correction in a wave modeling problem:

- A) Apply no BC [Leake et al., 2007; Lionello et al., 2008; Grabemann and Weisse 2008; Mori et al., 2010]
- B) Apply BC to the input data [Wang and Swail, 2002; Hemer et al., 2011, Hemer et al., 2012; Durrant, 2013; Wang et al., 2014]
- C) Apply BC to the output wave fields [Caires and Sterl, 2005, Cavaleri and Sclavo, 2006; Andrade et al., 2007; Tomas et al., 2008; Charles et al., 2012].

Consideration should be given as to which of these methodologies is the most applicable to the particular study since each has associated strengths and weaknesses. Option (A) is the ideal case and the most theoretically robust. The gradual improvement of physical models is undeniably the end solution to bias. BC can be thought of a temporary solution to bring currently flawed model predictions into alignment with reality but with associated limitations. Ehret et al. [2012] reviews the broad issues with BC, including the lack of a sound physical basis [Haerter et al., 2011], impossibly restrictive assumptions, a masking of uncertainty, introduction of physical inconsistencies between other model variables, and alteration of the spatial and temporal covariance structure of variable fields [Johnson and Sharma, 2012]. This being said, if model results are sufficiently biased, analysis may be restricted to only being relative (non-absolute). While this is acceptable for a comparison of results (say historic and future simulations), using uncorrected wave model output to force additional models (e.g., coastal sediment transport) will simply propagate the bias.

Option (B) has proven effective at improving modeled wave parameters [Caires et al., 2004; Hemer et al., 2011; Durrant et al., 2013] since many wave modeling errors can be directly traced to input wind fields [Cardone et al., 1996; Rogers and Wittmann, 2002; Durrant et al., 2013]. Additionally, this option has the advantage of correcting model output over the entire model domain. This said, bias correction of wind fields has significant disadvantages including being computationally expensive [Wang and Swail, 2002] and oftentimes practically difficult. The

process is complicated by sparse observational information across the ocean basins, both spatially and temporally, leading to target datasets of limited length, spatial coverage, and accuracy. Furthermore, even with BC of wind fields the wave model output will likely still exhibit bias due to wave modeling errors [Rogers et al., 2005].

Option **(C)** is well positioned to ensure that wave model output will be statistically in agreement with wave climate observations. Despite this, there have been relatively few studies of the application of BC methods to wave model output. To discuss a few examples, Andrade et al. [2007] used a variant of quantile mapping that fits a log-normal distribution to significant wave height and changes the parameters to match probability distribution functions (PDFs). Additionally, Caires and Sterl [2005] used nonparametric regression estimators, Cavaleri and Sclavo [2006] used a parametric correction, and Charles et al. [2012] used a univariate quantile mapping method to independently correct wave height, wave period, and wave direction. It should be noted that we are considering a “local” BC problem in the strict sense that model output is corrected only at the observation location. In general, wave model bias can be considered slowly varying (e.g., figure 7 of Hemer et al. [2012]) and transfer functions derived at one location can be used to inform the correction at other nearby locations. In this sense, option **(C)** is well suited as an intermediate step in a nested modeling approach. Basin-scale wave model output can be extracted at an observation location, corrected, and then used as open boundary forcing for a local-scale domain. If considering a location with rapidly varying bias structure (e.g. complex nearshore coastal configurations) or looking at corrections across larger regions, Tomas et al. [2008] introduces a spatial-temporal correction based on a nonlinear parametrization of Empirical Orthogonal Functions (EOFs).

The main contribution of this paper is to provide a comparative study of BC methods applied to wave model output (option **(C)** as listed above). This study is the first quantitative comparison of univariate and bivariate methods and is the first to apply bivariate methods to wave model output applications. This paper focuses specifically on the “application” of various BC techniques, leaving more complete expositions of the individual methods to the relevant citations. Section 2.2 of this paper describes the geographic location and the relevant data used for this study and section 2.3 introduces the BC techniques, including a brief theoretical and technical overview as well as the methodology for comparison between them. Comparative

results are provided in section 2.4 and a discussion of the results, including limitations, is provided in section 2.5.

2.2 Data

The data utilized for inter-comparison of BC techniques were provided by the Australian Commonwealth Scientific and Industrial Research Organization (CSIRO). CSIRO has produced wind-wave climate projection datasets [Hemer et al., 2015] using WAVEWATCH III [v3.14; henceforth WW3; Tolman, 1991; Tolman, 2009] forced by an ensemble of atmosphere-ocean general circulation models (AOGCMs). These datasets cover runs from many different AOGCMs, multiple emissions scenarios (CMIP3 and CMIP5) and several climatological periods (1980-2005; 2026-2045; 2080-2099). The model implementation is detailed in Hemer et al. [2012; 2013] and the reader is directed to these primary sources for details regarding the model setup and production methodology for the data used in this study.

The CSIRO datasets are an ideal testbed for this intercomparison of BC methods since: (a) the results were found to demonstrate bias [Hemer et al., 2013], (b) the data are available globally, allowing for the investigation of regional variability, and (c) the models are available for several climatological periods, allowing for testing of nonstationary BC. A single AOGCM, MRI-CGCM3 [Yukimoto et al., 2012], was arbitrarily chosen from the CSIRO ensemble. This choice should not be significant as the statistical testing methodology described in this paper is designed to reveal robust differences between methodologies regardless of the specific dataset. Two timeslices were used for this study, CSIRO's historical (1980-2005) and mid 21st century (2026-2050) simulations with the future run using the RCP 4.5 climate scenario.

It should be noted that Hemer et al. [2013] define bias as the difference between the model being forced by AOGCM winds and the model being forced by reanalysis (CFSR; Saha et al., 2010) winds (either raw or bias corrected). In this paper, bias is instead defined as the difference between buoy observations and model output. This choice was made to include the full range of model biases, including those attributable to the wave model itself. Four National Data Buoy Center (NDBC) locations were chosen as test points (Figure 1):

- #42001: Mid Gulf of Mexico (25.888 N, 89.658 W)

- #44025: Long Island (40.251 N, 73.164 W)
- #46029: Columbia River Bar (46.159 N, 124.514 W)
- #51002: South West Hawaii (17.056 N, 157.803 W)

These buoys were chosen to sample a wide variety of wave climates and therefore test the spatial variability of the BC effectiveness. Figure 2 shows the monthly climatology at each of the chosen buoys. Monthly average significant wave height (H_s) and peak wave period (T_p) time series are plotted for the buoy, historic model, and future model. All locations show a distinctive seasonal cycle but with significant spread in terms of magnitude at each location. Of particular interest to this study, each location and variable shows a unique bias structure. This said, wave heights are generally underpredicted while wave period is overpredicted in some cases and underpredicted in others.

It should be noted that in principle, BC methodologies can be extended to the n-dimensional case, incorporating more than the two variables considered here. For wave modeling applications this would mainly be of interest in order to include wave direction within the BC framework. This is excluded from the present paper for two reasons. First, BC methods work best with climatological (typically ~ 30 years) periods of record and this duration of wave direction time series is generally unavailable. Second, the comparative bivariate analysis presented here is relatively complex and lengthy. We therefore defer tri-variate methods to a future study in order to maintain clarity and a reasonable length for this study.

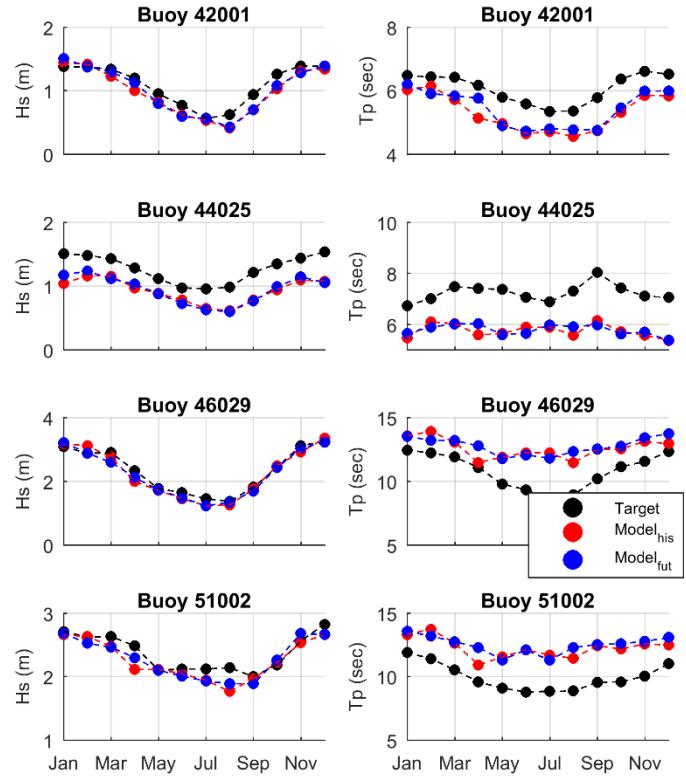


Figure 2: Monthly means of significant wave height (H_s) and peak wave period (T_p) at each buoy location. Data shown are observational (black), historic model (red) and future model (blue).

2.3 Methods

2.3.1 Bias Correction Methods

A subset of available BC methods were chosen to illustrate the differences between univariate and multivariate methods, parametric and empirical methods, and to look at the performance of available multivariate methods. In this section we will indicate model data with (M), target buoy data with (T), and bias corrected model data with (B).

2.3.1.1 Univariate Quantile Mapping

Several studies have found that distribution based methodologies, in particular quantile mapping, are most effective for BC of climate model data [Themeßl et al., 2011; Teutschbein and Seibert, 2013]. Quantile mapping attempts to match the CDF of a modeled time series to that of an observed time series. This is done most commonly using the direct constraint of Déqué [2007] which can be expressed as:

$$B = CDF_T^{-1}[CDF_M(M)]$$

Equation 1

where CDF refers to the cumulative distribution of the time series and the subscripts follow the notation specified above. For example,

CDF_T^{-1} is the inverse of the target dataset's CDF. Equation 1 specifies that each individual model value is replaced with the target value having the same quantile. This will result in a “scaling” of the model time series such that the CDF of the bias-corrected model data exactly matches that of the target.

Figure 3 shows this procedure as applied to a single model variable value

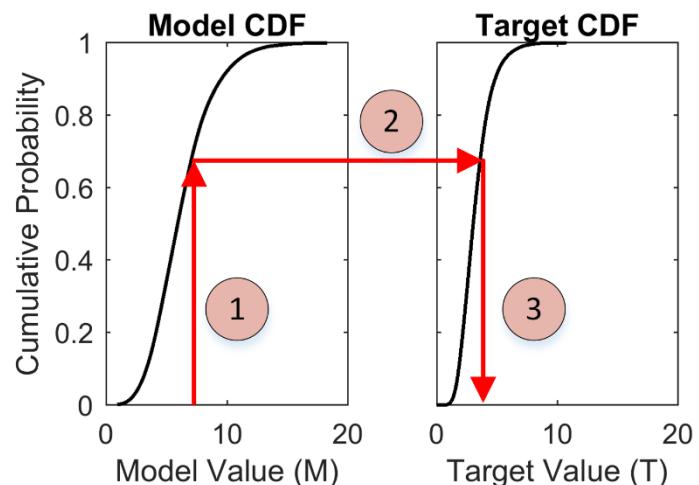


Figure 3: The Quantile Mapping procedure as applied to a single historic model value.

(M). Note that in section 2.3, all variables will be kept generic and only replaced with wave parameters in the results sections. Also note that the captions for the methodological figures are kept brief, with full explanation in the text. In Figure 3, the first step of quantile mapping is to determine the quantile of the model value of interest (1). This quantile is then transferred over to the target dataset (2) and a variable value is found by inverting the CDF of the target dataset (3). This value is the new bias corrected model value. This procedure is repeated for each value in the model dataset.

Parametric and empirical versions of quantile mapping differ in how the CDF and its inverse are calculated. The parametric method involves fitting a theoretical (parametric) distribution to the data. Various “long-term” wave parameter distributions have been identified by the literature with the most common being the Log-normal or Weibull for both Hs and Tp [Bauer and Staabs, 1998; Ferreira and Soares, 2000; Ferreira and Soares, 2002; De Waal and Van Gelder, 2006; Holthuijsen, 2007]. It should be noted that there is no universal best fitting distribution because the statistical properties of waves vary both temporally and spatially, requiring different techniques for each individual case. For this reason multiple commonly used distributions were tried with the best fitting, determined via a goodness of fit test, being ultimately selected. The Rayleigh, Log-normal, Generalized Extreme Value (GEV), Weibull, and Burr distributions were fitted for Hs. For Tp, the same distributions were tried but with the Burr replaced with a Gamma distribution. This fitting and selection process was repeated for each BC operation.

The empirical method determines CDF values directly from the data with no assumption of a distribution. For this study a kernel smoothing function was used to provide smooth estimates of CDF quantiles and for interpolation when inverting the CDF. Calculated values were checked against the direct empirical estimates to choose a proper bandwidth that avoided oversmoothing.

There exists some disagreement in the literature as to which method (empirical or parametric) is preferable. Li et al. [2010] state that use of an empirical distribution will result in frequent interpolation and extrapolation with unsatisfactory results. This should only be an issue for small datasets and can be circumvented by a kernel smoother with proper bandwidth. Lafon

et al. [2013] looked at a comparison of empirical and parametric techniques, with a focus on the effect of using a variable number of quantiles. They cautioned that too many quantiles will result in an “overfitting” to the target dataset, causing a loss of robustness due to an oversensitivity to the calibration period. The climate, if treated as a stochastic process, would be expected to have some random variability around its “true” distribution. A perfect fitting of the model to a training segment of the climate time series being considered (target) could theoretically overfit the correction, causing error between the BC and the true distribution. This issue is subtle and many authors simply fit the data to the empirical CDF at every data point (each model value has its own quantile). Attempting to resolve this uncertainty is one of the goals of this study.

Equation 1 is formulated to bias correct model data when observations are available. To bias correct future model data (where observations are not available), the so called equidistant quantile matching method of Li et al. [2010] can be used. This method is described by Equation (2) and can be seen conceptually in Figure 4.

$$B_f = M_f + CDF_{T,h}^{-1}[CDF_{M,f}(M_f)] - CDF_{M,h}^{-1}[CDF_{M,f}(M_f)]$$

Equation 2

In this equation, the subscripts h and f are used to denote historic and future values. As shown in Figure 4, the first step **(1)** is to find the quantile corresponding to the future model variable value $CDF_{M,f}(M_f)$. The variable value of both the target and historic model are evaluated (using the inverse CDF) at this quantile value and the difference is found **(2)**. This difference is the model predicted change from the historic run to the future run. The difference between these two values is then applied to the model future value to acquire the BC value **(3)**. Of note, this method necessarily makes the assumption that the statistical properties of the AOGCM systematic errors are stationary, a result that is not necessarily demonstrated in the context of climate change and is itself a topic of research [Charles et al., 2012; Deutschbein and Seibert, 2013].

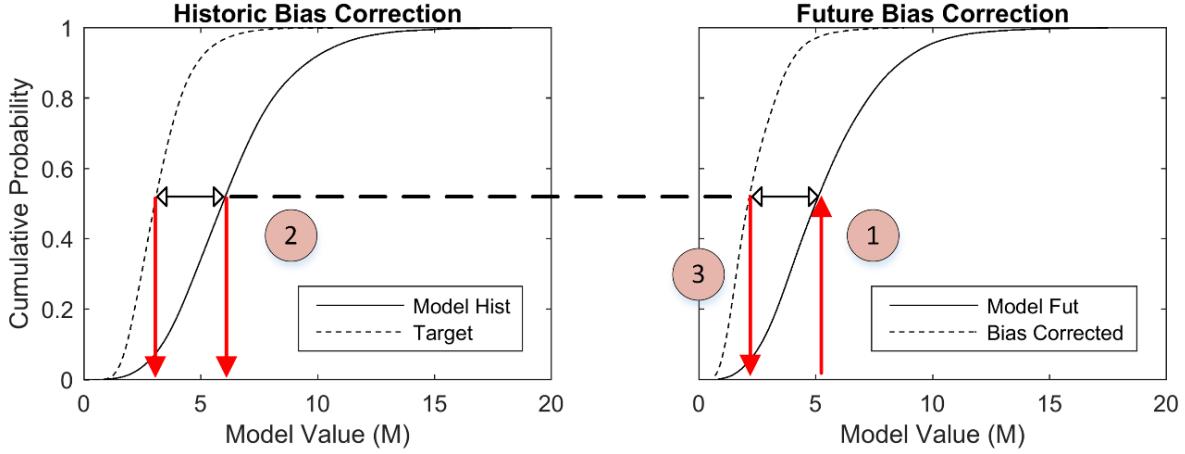


Figure 4: Equidistant Quantile Matching as applied to a single future model value.

2.3.1.2 The Binning Method

The Binning method [Piani and Haerter, 2012] is a multivariate method that was originally developed for jointly correcting surface temperature and precipitation meteorological fields. This technique handles correlation by first correcting one of the variables with a standard 1D BC method and then correcting the other variable conditionally based upon bins of the first variable. This is conceptually illustrated in Figure 5. The first step (**1**), not shown in the figure, is to use standard quantile mapping on one of the variables (marked with an X subscript). The second step (**2**) is to select a bin (this study uses 10 bins, or deciles) of the first variable (X) from both the target and model datasets. Note that the bins are selected in CDF space so a decile is shown in the figure as a single bar of data ($.2 > CDF_X \geq .1$), indicated with red symbols. The second step is to determine CDFs from these bins of data (**3**). Finally a normal quantile mapping is used with these decile dependent CDFs (**4**) so as to bias correct the second variable (Y). Step 3 is repeated for each model value within the decile. Then a new bin is selected and the process is repeated.

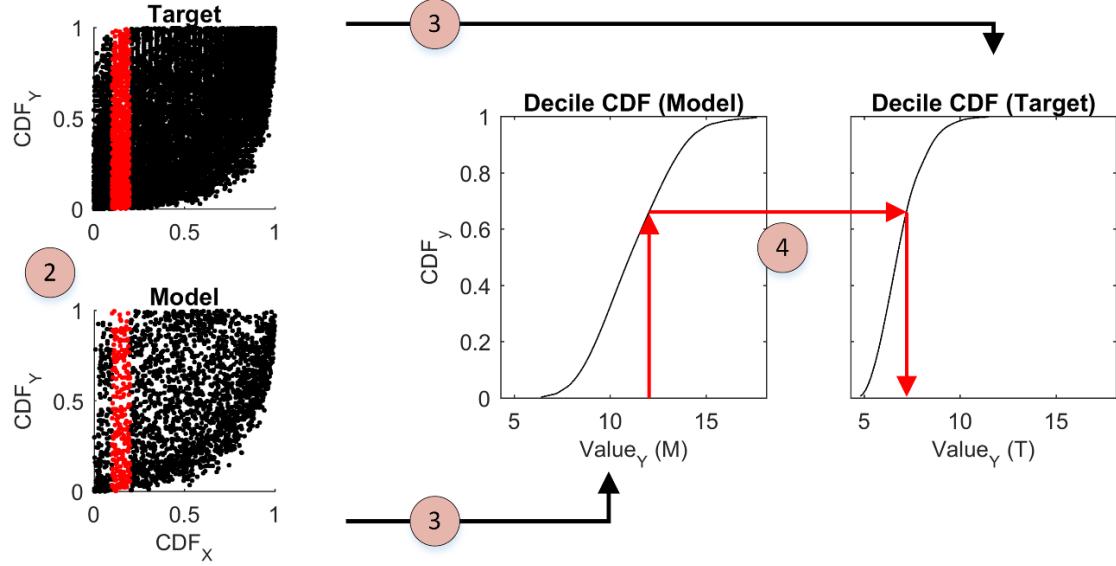


Figure 5: The Binning Method as applied to the second variable of a bivariate bias correction. Example deciles of bivariate CDF are marked with red dots.

The Binning method has a free variable in the number of bins to be used. The number must be large enough to properly distinguish correlation but small enough that sufficient data are within each bin to allow an accurate inversion of the CDF. Piani and Haerter [2012] found that surprisingly few bins (<5) were required to accurately correct the correlation. This study's choice of deciles was arbitrary and it is hypothesized, though untested, that a smaller number of bins would have produced similar results.

Note that the Binning method is a framework for dealing with correlation and is not dependent on the univariate BC technique. Any univariate BC could be substituted for quantile mapping and the methodology would be the same. An interesting question is whether or not the choice of the first variable to be corrected is significant. As these methods correct this first variable in a univariate way, and then correct the second conditionally based on this first variable, it is not unreasonable to assume that the first variable will be better matched to the target statistically. Colloquially, the second variable would “sacrifice” its ability to perfectly match the target statistically in order to match the correlation between the variables. For all results presented in this paper H_s was used as the first variable corrected. However, a single buoy (#46029) was used to determine the sensitivity of results to this decision by comparing results to the case when T_p was used as the first variable.

While univariate quantile mapping has an accepted methodology for “future” runs (Equation 2; Figure 4), the authors know of no similar procedure for the multivariate BC methods. As the Binning method is a framework built around univariate BC, an extension based on the principles of equidistant quantile matching was developed for this paper and is detailed in Appendix A.

2.3.1.3 *The Direct Method*

A second technique for bivariate BC was proposed by Li et al. [2014] and is based on a direct defining of a bivariate distribution between the variables of interest using copulas. This allows conditional correcting of one of the variables based upon the bivariate distribution’s definition of the relationship between the two variables. The mathematics of the Li et al. [2014] method are significantly more in-depth than those of the previous methods and this paper only presents a procedural overview. Figure 6 shows the general process for correcting a single model output Hs, Tp pair.

The first step of the methodology (1) involves a univariate BC of the first variable, similar to that of Binning method (not shown in Figure 6). The next step is to fit a copula to both the model and target datasets (2). A variety of Archimedean and Elliptical copulas (Clayton, Frank, Gumbel, Gaussian, and Student T) were considered and the best fitting was chosen for each dataset. Copulas are a way to model dependence between variables, separate from the marginal distributions. Marginal distributions are defined as the PDFs of the individual variables, without reference to the other correlated variable. By looking at CDFs, the marginal distribution is removed since CDFs are both non-dimensional and approximately uniformly distributed. This allows for a more direct look at correlation between variables as the variables have been reduced to a comparable state. A primer on copula mathematics is provided by De Wall and Van Gilder [2006], and De Michele et al. [2007] among others.

Once a copula has been fit to the data, the next step is to determine the bivariate CDF value of the model copula (3). For normal quantile mapping, BC is based on evaluating the model quantiles with target data and this is similar to what will be done here, just in a bivariate capacity. Following this template, the analysis then moves to the fitted target copula. We are trying to find the comparable quantile to that of the model copula but since copulas are bivariate

there are multiple locations of identical CDF magnitude, i.e., a so called level curve. This complicates analysis as more than just the bivariate CDF value is required for determining a unique location on the CDF surface. We can escape this problem because of the two bias corrected value pairs, one is already known from the univariate BC (step (1)). Using this information it is possible to take a slice of the bivariate CDF at the known BC value (4). Since we know the bivariate CDF value of the model (evaluated at (3)), we can take the inverse of this slice to find the CDF value for the second value (based on the target copula) (5). With this quantile we can then invert the target CDF to get the bias corrected model value (6).

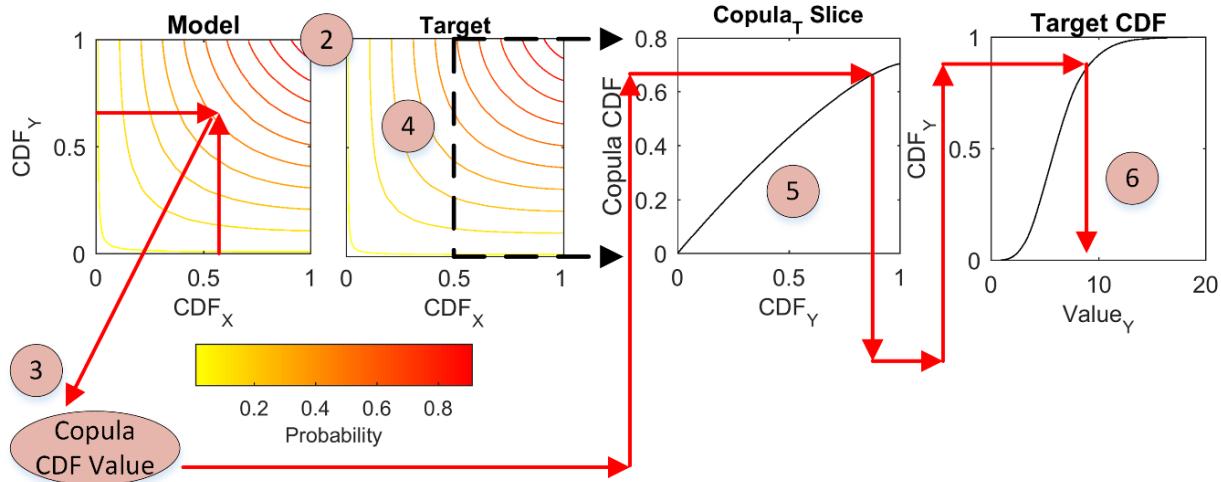


Figure 6: The Direct Method as applied to the second variable of a bivariate bias correction.

Similar to the Binning method, a nonstationary methodology for future runs was developed by the authors and is presented in Appendix A.

2.3.1.4 The Shuffle Method

The “Schaake Shuffle,” or Shuffle method for this paper, is a method of multivariate BC that Clark et al. [2004] developed as a method for correcting both spatial and temporal covariability in forecast models. Vrac and Friderichs [2014] showed that the method could further be used for multivariate BC. The Shuffle method takes advantage of the fact that the CDF

of a time series is not dependent upon the order of the time series. This means it is possible to univariately bias correct the CDF of each variable in a dataset to fix the individual variable's bias and then shuffle each variable's time series to fix the intervariable correlation. Correlation is, simply put, a relationship between two variables and this can be created by properly organizing each variable's time series to exhibit the wanted relationship. The target time series contains a template for the relationship between variables (as displayed by the way each variable moves in relation to the others) so by rearranging the model time series of each variable to mimic that of the target, the method can recreate the wanted correlation. The process is easy to implement and involves first univariately bias correcting each variable to remove the statistical bias. The next step is ranking, in terms of magnitude, both the target and model output (already univariately bias corrected) for each variable. The ranked model output for each variable is then placed into the same order as that of the ranked target. Figure 7 shows this procedure graphically. The numbers within the circles are the rank of the circles in terms of magnitude for each variable time series (represented by the colors brown and blue). For the Target and Model time series the colors become darker from right to left representing the original order of the variables. For the shuffling, we use the order of ranks from the target series to rearrange the model to the BC time series. This results in the model having the fifth ranked variable moved to position one from position two for Var_x and

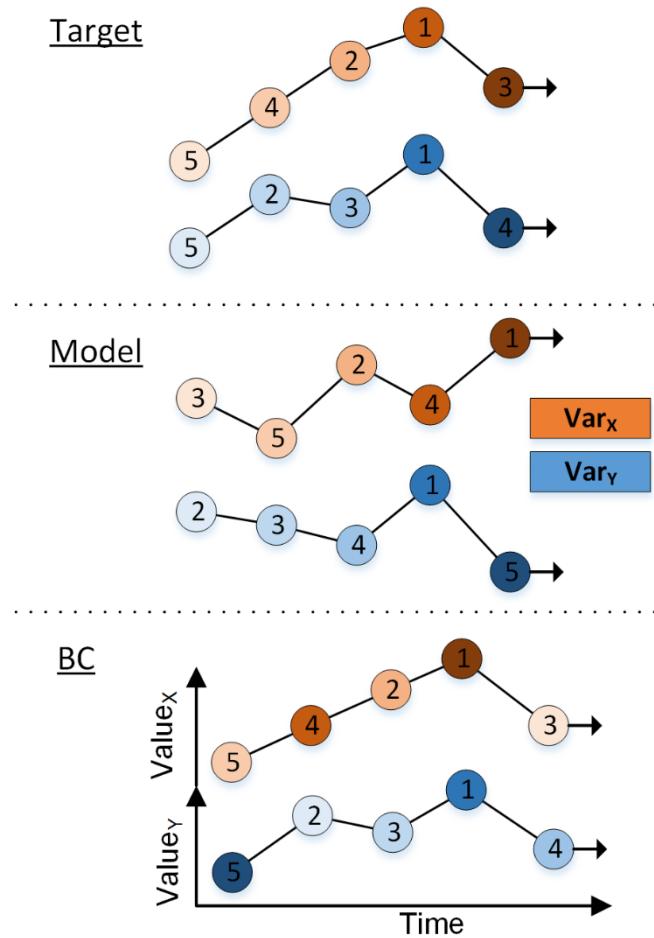


Figure 7: The Shuffle Method (Schaake Shuffle). Numbers indicate rank in terms of magnitude. Color gradient indicates time with values later in the time series being darker.

to position one from position five for Var_y . This shuffling continues for all values in the time series.

The Shuffle method is very powerful in that temporal, spatial, and intervariable relationships can be corrected simultaneously. This is only possible because the relationship in time, space, and location (as preserved by the rank) of the target time series is being “stolen” for the model time series. This differentiates the Shuffle method from other BC techniques as it produces a scaled version of the target time series rather than a scaled version of the model output. This further requires that the target time series and model time series have the same sampling rate and size which can be a significant limitation.

2.3.2 Validation Methods

The primary method of assessing the effectiveness of a BC methodology is to look at how well the method is able to change the statistical properties of the model to that of the target. Because our validation methodology involves an ensemble, these statistics were normalized as the relative difference (RD), defined as:

$$RD = \frac{B - T}{T}$$

Equation 3

in which the variables T and B follow the notation from above but represent some metric of performance (mean, standard deviation, etc.).

A comparison of statistical moments from individual variables ignores the relationship between variables. Since preservation of this relationship is the primary goal of a multivariate BC, we also looked at a comparison of correlation coefficients (a measure of the statistical dependence between two variables). Three common correlation coefficients were examined, (Kendall’s τ , Spearman’s ρ , and Pearson’s r) all of which measure a different form of dependence. Kendall’s τ measures the degree of similarity between ranked time series (how the variables move together). Spearman’s ρ measures how well the relationship between two

variables can be described by a monotonic function. Pearson's r measures the strength and direction of the linear relationship between two variables.

A further metric used as a quantification of BC effectiveness was the shape of the bivariate PDF and CDF. Correlation coefficients would be expected to capture some of the relationships shown by the shape of these distribution functions but not all. Piani and Haerter [2012] noted this when they found that the shape of the bias corrected bivariate CDF was one of the primary means of seeing the difference between a bivariate and univariate correction. Their study noted this qualitatively but did not propose a quantitative metric of this result. We expand upon this by considering the root mean square error (RMSE) between the target's bivariate CDF and the bias corrected model's CDF as a way of quantifying how well correlation is corrected. PDF RMSE was also investigated as an additional perspective into changes of overall distribution shape.

Oftentimes wave modeling studies are more concerned with extreme events than the mean of a wave climate. High return level events are drivers of coastal hazards and can dominate related processes signals (such as geomorphology) [Anderson et al., 2010; Zhang et al., 2002]. This importance brings an interest in the ability of BC to address incorrect model prediction of high quantile events. For this reason the 95th and 99th quantiles were compared between the target and BC time series as a further metric for comparison.

As mentioned earlier, previous studies have brought into question the robustness of BC techniques [Li et al., 2010; Lafon et al., 2013] and in particular empirical methods due to overfitting and the sensitivity of BC to the training period. We decided to address this issue by considering measures of performance using a modified jack-knife (also known as a k-fold) cross-validation technique following the procedure of Bissell and Ferguson [1975] and Lafon et al. [2013]. The cross-validation component of this methodology involved breaking up the total target record into a training segment and a validation segment. The metrics of performance are then assessed by developing the BC from the training segment and evaluating its performance with the validation segment. The jack-knife, or k-fold, component of this methodology means that an ensemble of statistics are calculated using different training periods. Specifically, the total target record is broken up into a number of sub-records. For each iteration, one of these sub-

records is removed as the validation segment while the others are used as the training segment. For each of these iterations the RD is calculated, resulting in an ensemble of normalized “errors” that will quantify the robustness of the BC technique to the calibration period.

Wave parameters often exhibit a strong seasonal signal (See Figure 2) that is important to capture in a BC methodology. For this reason it was decided to additionally break the total record up into seasons. As each of these seasons has a characteristically different distribution shape, this can be considered a further test of robustness in how well the BC method performs on different shaped distributions. Each of these seasons was run through the cross-validation procedure. With 4 seasons, each with a 10 bin jack-knife test, the resulting ensemble for each measure of performance was 40 data points. As the RD is normalized, the data were not regrouped by season but treated as a single sample.

2.4 Results

Although BC was tested on 4 locations, only results from buoy 42001 will be presented graphically in this section. This location was chosen for presentation as being a characteristic case with all other locations exhibiting the same general behavior. It was additionally chosen as the most illustrative with model output being generally the most biased across all metrics. Any cases where individual results vary from the displayed data will be discussed within the text. Furthermore, additional plots for all locations have been provided in the supplementary materials.

2.4.1 Statistical Bias

Figure 8 presents results of the BC procedure in terms of summary statistics for Hs and Tp. The ensemble, as described above, provides important information about the sample spread, leading to the presentation of the data as box plots. All presented box plots have the centerline of the box at the median value, the edges of the box at the 25th and 75th percentiles, and the whiskers extending to the most extreme data points not considered outliers (as defined using the Tukey boxplot convention [Frigge et al., 1989] where the end of the whiskers are +/- 1.5 times the interquartile range from the box edges). Outliers are plotted as black dots. Each methodology has both historic and future results represented by the blue and red colored boxes in each

column, respectively. A heavy line is used to connect the means to provide a visual cue as to the change from historic to future runs.

When examining a condensed information plot such as Figure 8, it is useful to consider what an “ideal” result would be. The y-axis is a RD, meaning that the values are both normalized (unitless) and comparative. A value of 0 on this axis (plotted in the figure as a dotted line) for the historic case would mean a perfect match between the target and the bias corrected data. The second aspect to consider is the width of the boxes (and their corresponding whiskers). This is the sample spread and represents the variance in RD, or how well the BC performed for a variety of test cases (the ensemble). An ideal BC technique would have a width approaching zero, meaning that it is robust to both calibration period (k-fold validation) and distribution shape (seasons).

The first column in each panel of Figure 8 shows the raw model output. The y-axis value in this case indicates the performance of the wave model itself, when compared to the observational data. For example, in looking at RD for the mean values, it is clear that the model under-predicts Hs (negative value for RD) and as well as Tp (also see Figure 2). The spread in the raw model results is due to the jack-knife procedure, where only a portion of the buoy data is considered the target at a time. The displayed sample spread therefore indicates the natural variability of the target dataset over time.

Since there is no observation (target) dataset for the future, the future result metric is calculated using the historic observational data for normalization. The differences between the historic and the future run results therefore is an indicator of changes in the wave climate. An ideal BC technique would preserve the relationship between the historic and future runs. Conceptually this means that despite biases in the simulated wave data, the future period is assumed to have a useful signal if interpreted relative to the historic period rather than that of the observed [Wood et al., 2002]. If the amount of change is the same for the model column and the bias corrected column, then the BC has successfully maintained the relationship between the historic and future runs.

Looking at the Hs column of Figure 8, all BC techniques perform consistently across the metrics. Furthermore all methods perform very well with the RD of metrics being

close to zero with low variance. Noting the change of y-axis scale, it is seen that all techniques perform decreasingly well (both with matching the target and with sample spread) for increasingly higher moments. This is an expected result as higher moments are more variable depending on calibration period. It may be initially counterintuitive that all BC methods perform so uniformly for Hs but recall that all tested multivariate BC methods require one of the variables to be corrected with a univariate BC as a step in the process. For this study this initial variable was chosen to be Hs so all methods are essentially treating this variable the same.

The Tp column of Figure 8 indicates a similar general level of success in correcting the moments of the Tp distribution. This result is contrary to our hypothesis that the choice of primary BC variable was significant and that the second variable would be less accurately corrected. The only outlier to this generally good performance is the Direct method which did not perform as well for higher statistical moments. It should be noted that the Direct method additionally showed poor performance for mean and standard deviation in some of the other buoys tested (buoy #44025 and #46029).

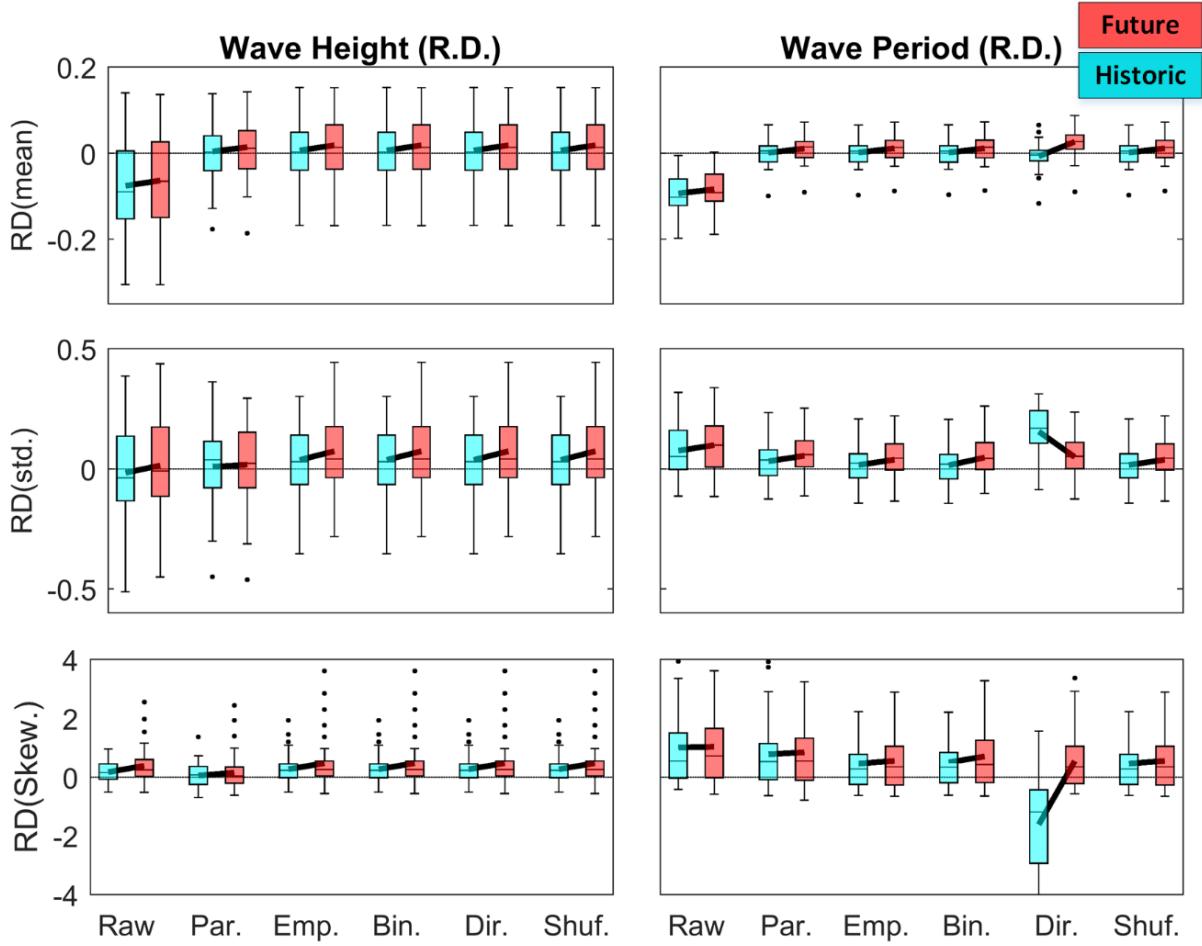


Figure 8: Bias Correction results for the first three statistical moments: Mean, Standard Deviation, and Skewness. Raw column shows the agreement of raw model output with observations. All data are presented as relative differences (RD) and all boxplots represent an ensemble of 40 data points. Note that some outliers are not visible due to y-axis scaling.

A sensitivity analysis was performed to explore the significance of Hs being arbitrarily chosen as the primary variable by rerunning a single buoy (#46029) with TP as the primary variable. This test corroborated the above result with no difference found between the means of boxplots run with Hs or Tp as a primary variable.

2.4.2 Correlation

Figure 9 shows correlation plotted using the same boxplot conventions as Figure 8. It is in correlation that multivariate BC techniques would be expected to outperform univariate methods. Figure 9 confirms this with all univariate procedures having boxplots that have nonzero relative differences, showing that the bias corrected data has the same incorrect correlation as the model itself. This is contrasted with the bivariate methods which all have boxplots shifted towards zero representing an improvement in correlation. Furthermore, this result is shown across all three correlation coefficients hinting at a true correction in correlation, although this will be more accurately be accessed by looking at the full bivariate PDF/CDF. In terms of the three bivariate methods, the Binning and Shuffle methods are found to outperform the Direct method. The Shuffle method displays the undesirable characteristic of an inability to preserve the historic-future relationship of correlation (represented by an incorrect bold line slope between historic and future simulation boxes). This is a result of the Shuffle technique methodology which exactly matches both the historic and future periods to the target time series, not allowing for a correct modeling of change in correlation over time.

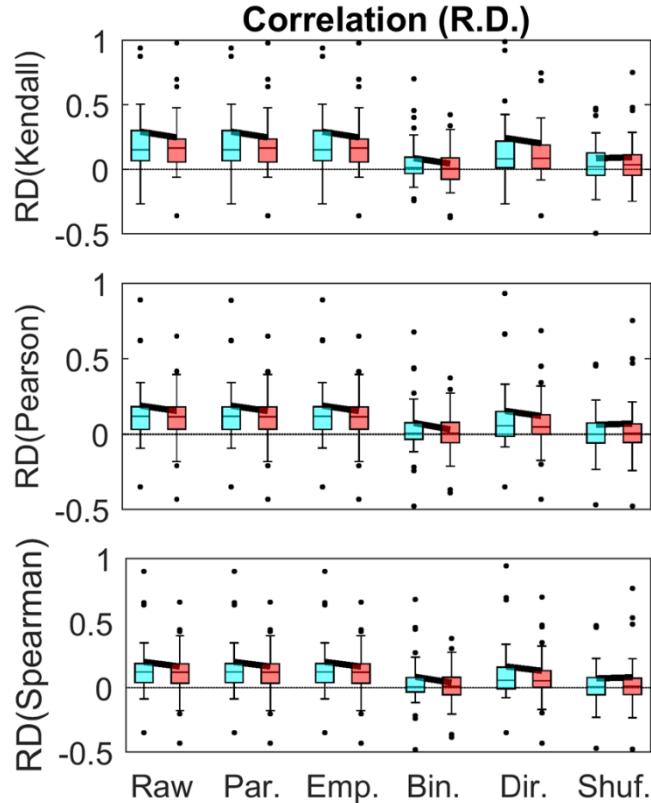


Figure 9: Bias Correction results for correlation coefficients: Kendall's τ , Spearman's ρ , and Pearson's r . Raw column shows the agreement of raw model output with observations. All data are presented as relative differences and all boxplots represent an ensemble.

2.4.3 Bivariate PDF/CDF Shape and RMSE

While bulk statistics can quickly show the properties of a dataset, analysis of the full distribution provides a more visual and intuitive understanding of changes. Figure 10 shows bivariate PDFs and CDFs of all model, target and BC results for the summer seasons of the historic period. The plotted bivariate distributions are an average of the 10 jack-knife runs for the chosen season, each of which has been calculated on a consistent grid for comparison. In terms of the plots types, bivariate PDFs are easier to conceptualize physically (having real units) but are somewhat difficult to compare quantitatively due to noisiness and scale issues. CDF plots are less intuitive but have the advantage of more directly showing the correlation between the variables, a result of removing the marginal distributions.

First considering the bivariate PDFs (displayed as the 2 left columns), Figure 10 shows that both univariate quantile mapping methods (empirical and parametric) still have the same general shape of the raw model. The distribution has been scaled by the BC (a shrinking or expansion along the x or y axes) but the characteristics of the shape have not been significantly altered. This is contrasted with the bivariate BC methods that not only scale the distribution but also transform the shape into closer alignment with the target. In this case, both the Binning and Shuffle method closely match the target while the Direct method appears to be somewhere in the middle, performing better than the univariate methods but worse than the other bivariate methods. This result is in alignment with what was found by comparing statistical moments and correlation coefficients.

The CDF plots tell a similar story although the removing of the marginal distribution allows us to see that the univariate BC methods are entirely unable to fix correlation. The shape of the model's CDF remains unchanged in each of the univariate methods CDF plots. This is contrasted by the bivariate methods, which for the Binning and Shuffle methods, match the target to the point that it is difficult to see the dotted line representing the target. The CDF plots show again the direct method is somewhere in between the univariate and bivariate methods.

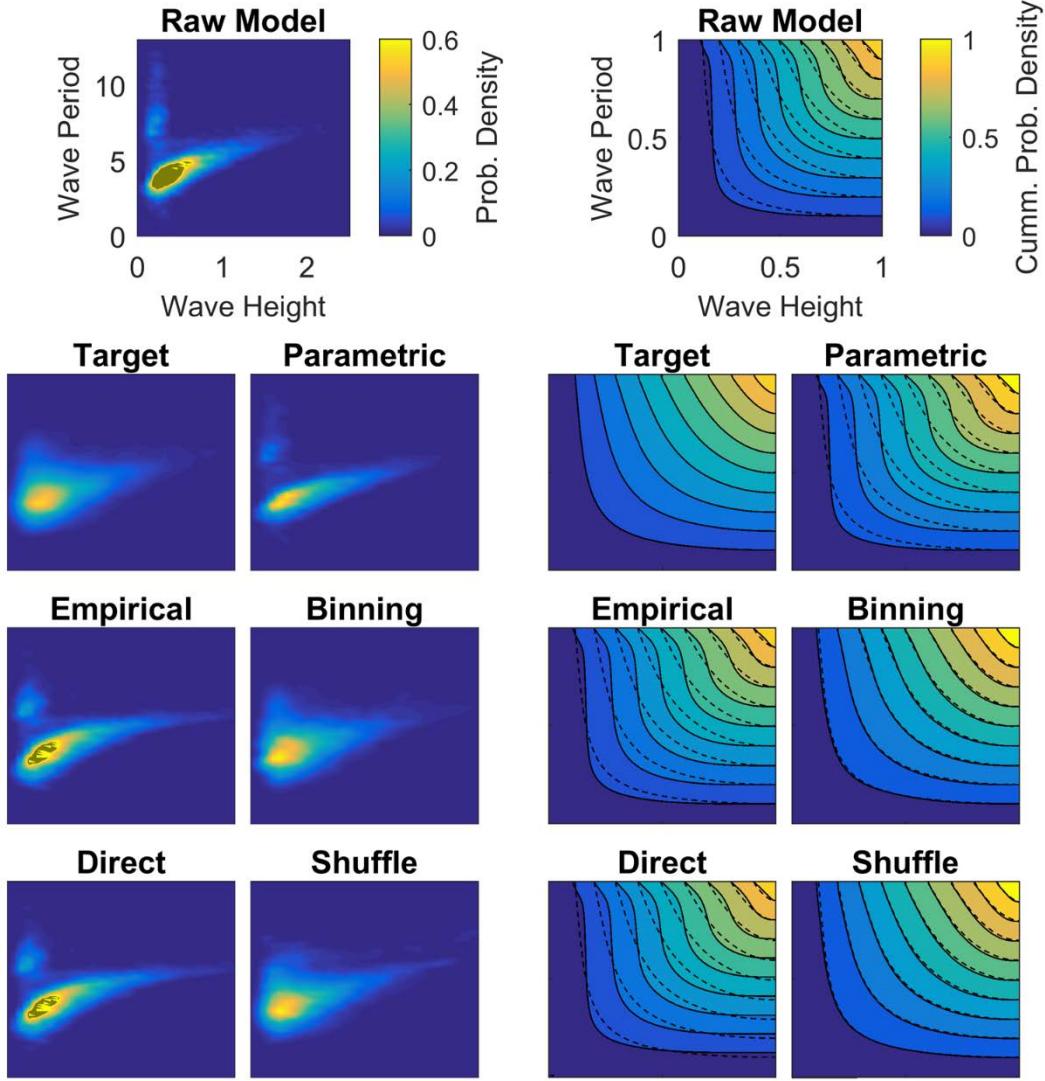


Figure 10: Comparison of bivariate PDFs (left columns) and CDFs (right columns) for the summer seasons of the historic period. Plotted results are an average of all 10 jack-knife bins. All subplots have the same scale and axes as shown on the Raw Model subplot.

These plots qualitatively show how the BC process is working but by calculating the RMSE we can attempt to quantitatively compare the methods. Figure 11 shows this comparison once again as a boxplot with the same format as detailed above. Only historic data are shown since a comparison of the future run to the target would not hold any useful information. The

PDF RMSE shows that all methods improve the PDF but that the Binning and Shuffle methods perform the best (RMSE closest to zero and the smallest spread). This is because these methods are correcting both the marginal and the shape of the distribution best. Of interest this metric shows the direct method performing quite poorly, at about the same level as the parametric and empirical methods. Upon further investigation this poor performance is season specific and is likely due to the fitted copula not properly describing the correlation structure of the data sufficiently. This said, it should be noted that a direct comparison of bivariate PDFs is a noisy estimate of BC success since values at specific grid cells are very sensitive to scaling.

The CDF RMSE shows the same relationship with the Binning and Shuffle methods outperforming all other methods. Since the marginal distributions are removed, the parametric and empirical methods perform the same as the model. In this case the Direct method is shown to perform worse than the univariate methods and even the raw model. This is evidence that, although the Direct method seems to match correlation coefficients fairly well, it is not properly correcting the full structure of dependence and potentially even making the structure worst in certain cases.

2.4.4 Extreme Quantiles

Figure 12 shows the results of an upper quantile (95th and 99th percentile) comparison in the same boxplot format as used for Figure 8 and Figure 9. An important difference is that no jack-knife cross-validation was used to generate the displayed ensemble, only seasonal bins resulting in a total of 4 data points for each method. Extremes are highly dependent on record length and a comparison of longer BC segment to a smaller validation segment would not be a

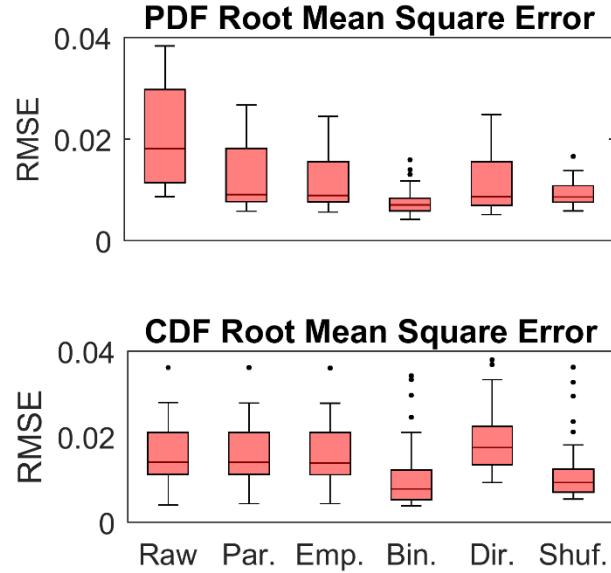


Figure 11: Root mean square errors for comparison of bivariate PDFs and CDFs.

valid comparison for this metric. This forced methodology change means that displayed performance is not directly comparable to that seen in Figure 8 and Figure 9 as the entire record length is used both for bias correction and validation. It would be expected that uncertainty and robustness in this comparison is significantly higher so results should be viewed with caution.

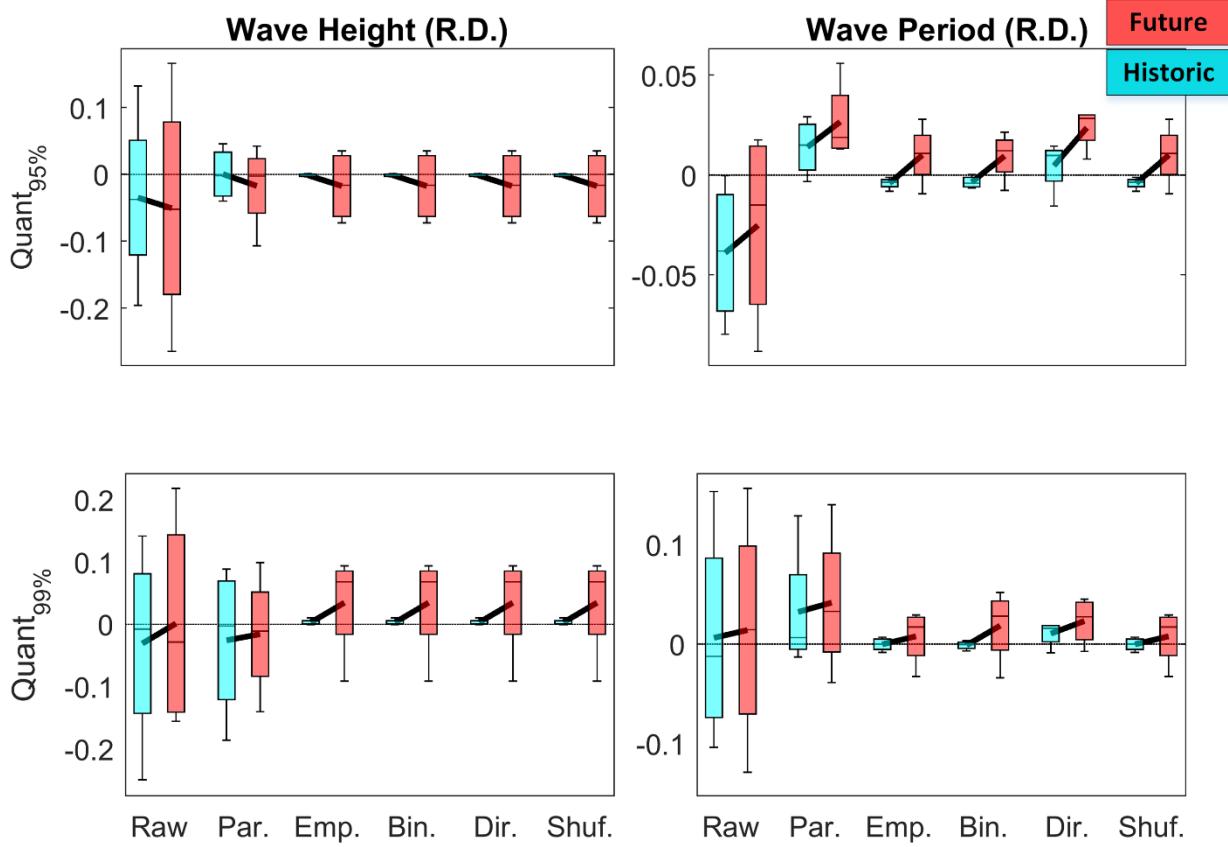


Figure 12: Bias Correction results for extreme quantiles (95th and 99th). Raw column shows the agreement of raw model output with observations. All data are presented as relative differences (RD) and all boxplots represent an ensemble of 4 data points.

Proceeding with this caveat, Figure 12 shows an improvement of the 95th and 99th percentile for both wave height and wave period. The only outlier to this general improvement is the Parametric method which shows little to no improving for the 99th percentile of wave height and a decrease in accuracy from the model for both the 95th and 99th percentile of wave period.

2.5 Discussion

This study is intended as an investigation of bias correction applied to wave modeling applications and as a comparative guide to which methods are most applicable in this context. In particular, consideration was given to the relative performances of (1) empirical vs. parametric quantile mapping and (2) univariate vs. bivariate bias correction techniques. The final goal of this paper was, through a rigorous comparison of methods, to identify the best performing method. This section will explore these research questions through statistical testing under the standard framework of the null hypothesis being no difference in the parameter being considered. The reported p-value represents evidence to reject this null hypothesis. For succinctness, this paper will withhold full expositions of the used tests and readers are encouraged to reference a statistics resource as needed.

The first comparison was between parametric and empirical quantile mapping. The literature is somewhat inconclusive in this regard with some studies specifying a loss of robustness due to empirical methods and others a loss of accuracy due to parametric assumptions. This study found little evidence of a gain of accuracy from using the Empirical method nor a gain in robustness from using the Parametric method. This was statistically verified with a paired t-test for statistical moments and correlation parameters. For wave height, the test showed an inability to reject the null hypothesis of no difference ($p\text{-values} > 0.03$) for both mean and standard deviation while wave period was unable to reject the null for mean but did show a statistical difference for standard deviation ($p\text{-value} < 2 \times 10^{-15}$). That said, the relative difference was very small (95% confidence interval of .0125 to .0187) representing a statistically significant but practically non-relevant difference between methods. This result was further corroborated by the RMSE comparison of bivariate CDF and PDF distributions which showed no noticeable difference between the two methods. Finally, both wave height and period visually showed a less effective correction of extreme events but this result was not found to be statistically significant ($p\text{-values} > .06$). Overall this result of no difference in any metric could either be because the distributions at the buoy locations chosen were well enough fit that the error introduced was negligible or that quantile mapping is insensitive to a well fitted distribution. A goodness of fit test failed at the 95% significance level for the vast majority of cases but this may be because the data record was very long providing strict error bars on the statistical test.

The next comparison was between univariate and multivariate BC techniques. Using a standard one-way analysis of variance (ANOVA) test, it was found that, in terms of statistical moments, there was no evidence of a difference between univariate and bivariate techniques (1 way ANOVA, p-value > .5) for wave height or wave period. This result was somewhat surprising as it was hypothesized that all bivariate methods would perform differently for the second variable due to the additional constraint of matching correlation. When looking at correlation it was found that there was evidence that bivariate techniques improved incorrect model correlations in comparison to univariate methods (1-way ANOVA, p-value $\approx .07$). Of note this larger p-value is a result of the model correlation being only moderately incorrect, making the BC correction smaller and therefore less significant statistically. Evidence of the bivariate methods effectiveness was further solidified qualitatively by looking at bivariate PDF shapes and quantitatively by looking at the RMSE of bivariate CDFs. No difference in high percentile corrections were found between univariate and bivariate methods.

As the preservation of the dynamic link between wave parameters (in this case Hs and Tp) is critical for realistic modeling, this result strongly supports the use of bivariate techniques over univariate. While this study looks only at bivariate BC, other correlated parameters (e.g. wave direction) could be included in the BC process. The extension of some copulas (for the Direct method) to multiple dimensions is straightforward and it is hypothesized that the Binning method could also be generalized to an N-dimensional case. As the Shuffle method involves a direct reordering according to the target time series, it is already a multivariate BC technique that would only require keeping track of another variable.

Since bivariate BC was concluded to be superior to univariate, the determination of a “best practice” methodology was limited to a comparison of the three tested multivariate techniques. In terms of statistical moments, strong evidence was found that the Direct method performed less well at correcting the standard deviation of the second variable than the other two bivariate techniques (Multiple Comparison Test, Tukey-Kramer adjustment, p-value $< 1 \times 10^{-9}$) as well as higher moments. This result was even more distinct in other buoys (#44025 and #46029), showing a significant incorrect matching of all statistical moments for the second variable. While visually the Direct method appear to not correct correlation as well as other bivariate methods, this result was not found to be statistically significant (Multiple Comparison Test, Tukey-Kramer

adjustment, p-value > .5). This result was shared by the correction of extreme quantiles with the Direct method appearing to be less effective but not to the point of being statistically significant (Multiple Comparison Test, Tukey-Kramer adjustment, p-value > .3). A comparison of CDF RMSEs showed the Direct method only slightly outperforming the univariate methods and significantly underperforming in relation to the Shuffle and Binning methods. This result hints at the importance of not just looking at correlation coefficients (which only measure a small subspace of the many ways variables can be related) to measure dependence. Piani and Haerter [2012] found a similar result and concluded that a “full appreciation” of the benefits of a bivariate methodology could not be explained by correlation coefficients alone. We hypothesize that the comparatively poor performance of the Direct method may be a result of the fitted copula not being able to adequately model the correlation between the two variables. It is possible that consideration of a larger number of candidate copulas for fitting could alleviate this problem. The other two methods, the Binning and Shuffle techniques, performed equally well at correcting statistical moments and correlations. This said, the Shuffle technique has the additional requirement of an exact match of sampling rate and record length between the target, model historic and model future time series, a somewhat restrictive condition.

This study also considered the performance of the tested BC methods by examining how well they preserved the nonstationary change from historic to future model runs. It was found that all methods performed well for the first variable (H_s) but that for the second (T_p) the Direct method struggled at reproducing changes for higher moments. In terms of correlation all methods preserved changes well except for the Shuffle technique which is unable to account for changes in correlation. The Shuffle technique also suffers from the philosophical issue that both the historic and future runs have a temporally identical time series (only scaling of variable magnitudes differ).

With current scientific focus on quantifying uncertainty, many wave modeling efforts have evolved into ensemble based frameworks requiring many simulations. This approach brings additional scrutiny to the computational burden of each step in the modeling process. A full computational analysis of the methodologies explored in this paper is beyond the scope of this study but a broad comparative investigation was explored. Looking at the time to correct one season of 30 years of data, Parametric corrections are very fast (order of 10's of seconds).

Empirical methods take longer (order of 10's of minutes) but still not to the point of being significant in relation to the computational time of model runs. Of Empirical methods, the Direct method is slowest and the Binning method is fastest but with the difference being in the order of minutes. These results are presented cautiously as the magnitude of this comparison is undoubtedly variable dependent on the implementation and system the code is being run on. Overall the computational time of BC was found to be minimal and not a driver of the choice between methodologies or the use of BC in general.

2.6 Conclusions

In summary, wave forcing is a key component in the accurate modeling of ocean and coastal systems. Oftentimes a direct wave model application will result in biased output due to a variety of factors ranging from errors within the wave model itself to biases within the input meteorological fields (especially with AOGCM derived products). Biases in wave model output can affect the entire system being considered making BC oftentimes requisite for bringing results back into agreement with reality. This said, BC should be used only after noting the various limitations and assumptions ranging from a masking of uncertainty to the potential introduction of physical inconsistencies between model output fields. While BC of model input data is widely accepted, this study investigates the alternative approach of a direct BC to the wave model output itself. In this context, a subset of BC techniques are compared using a statistically robust cross-validation technique in order to investigate which methodologies are most effective.

Key findings of this study were that: (i) the BC of wave model output is an effective tool for bringing wave model output into alignment with observations; (ii) no statistical difference was found between univariate empirical and parametric methods; (iii) bivariate methods were shown to be superior to univariate methods due to improvement in matching correlation; and, (iv) the empirical binning method of Piani and Haerter [2012] (extended in this study to apply to future projection model runs) proved to be the best overall approach across the various metrics considered.

Acknowledgments

The first author conducted this study with support from Oregon Sea Grant (Grant No. NA150AR4170224) as well as the Malouf Scholarship. Further financial support was provided by the ARCS foundation Portland. The data for this study was provided by Mark Hemer at the Commonwealth Scientific and Industrial Research Organization (CSIRO). The utilized data can be found on the CSIRO server at: <http://doi.org/10.4225/08/55C991CC>.

Chapter 3 - The Effects of Changing Climate on Estuarine Water Levels: A Pacific Northwest Case Study

Kai Parker, David Hill, Gabriel García-Medina, Jordan Beamer

Abstract:

Climate change impacts to extreme water levels at two U.S. Pacific Northwest estuaries is investigated using a multi-component process-based modeling framework. The integrated impact of climate change on estuarine forcing is considered using a series of sub-models that track changes to oceanic, atmospheric, and hydrologic controls on hydrodynamics. This modeling framework is run at decadal scales for historic and future periods with changes to flooding quantified across the two study sites. It is found that there is spatial variability in extreme water levels at both study sites with all recurrence interval events increasing with further distance into the estuary. This spatial variability is found to increase for the 100-year event moving into the future. It is found that sea level rise is buffered by a decrease in forcing. Small recurrence interval events are less buffered and therefore more impacted by sea level rise than higher return interval events. Finally, results show that annual extremes at the study sites are defined by compound events with a variety of forcing contributing to high water levels.

3.1 Introduction

Estuaries are important intersections of human and natural systems, serving as both some of the most resource-rich ecosystems on Earth and some of the most densely populated. Asked to meet many, at times conflicting, needs, estuaries require careful management. Unfortunately, coastal planning is limited by an insufficient understanding of how estuaries will respond to future conditions. In particular, extreme water levels (WLS) are of first-order importance with flooding putting both lives and physical infrastructure at risk.

In the U.S. Pacific Northwest (PNW), considerable progress has been made towards an understanding of flooding risk at open beaches as controlled by a combination of forcing drivers, including waves, tides, winds, and others [Barnard et al., 2014; Ruggiero, 2013; Serafin and Ruggiero, 2014]. Flooding risk in PNW estuaries is less well understood, primarily due to the greater complexity of the estuarine environment. Estuarine hydrodynamic climate has the additional driver of streamflow and is controlled by a much more complicated topographical context (e.g. embayment and complex bathymetry). This makes estuaries difficult to simplify as

they exhibit nonlinear water column response [Ding et al., 2013] and are difficult to uncouple [Wolf, 2009].

Attempting to constrain future risk adds an additional challenge through a mismatch in time scales. Climate change signals are significant at scales of decades while individual extreme events occur at timescales of hours to days. This requires both high model temporal resolution and long model simulations. Long time series are additionally necessary for extreme value analysis and constraining event recurrence intervals (RIs). Balancing these needs with computational cost has remained a major obstacle and led to a variety modeling strategies. One approach takes advantage of the fact that extreme value analysis only requires information of the maxima events. Lin et al. [2012] use a unique approach based on multiple model complexities; an efficient model to determine which events will cause flooding and a more complex model to accurately quantify the dynamics of those events. As a second example, Orton et al. [2016] simply model the coastal responses to a large set of hurricane events. Both of these studies focus on the eastern U.S. coastline where tropical cyclones are the principal driver of flooding events. Locations less dominated by tropical cyclones have a more diverse and balanced set of contributions to flooding. In these locations, extreme events can occur due to combinations of forcings which are not themselves individually extreme, a phenomenon discussed in the literature as “compound events” [Leonard et al., 2014; Moftakhari et al., 2017; Wahl et al., 2015]. This makes it difficult to know *a priori* which events will result in maximum WLs.

Recent advances in computing power and parallel processing have opened up an alternative possibility of running continuous time series hydrodynamic models at climate change scales (decades to centuries). This allows examination of extremes without *a priori* knowledge of which events will cause extremes. Additionally, a continuous time series analysis has other desirable properties. For example, an event based approach limits analysis to only large RI events, eliminating information on higher probability, lower magnitude, events. This is undesirable since so-called “nuisance flooding” can, over time, lead to a higher aggregate cost than extreme events, especially when considering sea level rise (SLR) [Maftakhari et al., 2015; 2017]. A continuous time series approach also allows a continuous consideration of the integrated and coupled effect of the various changing controls on estuarine hydrodynamics. Many studies have focused on individual components of climate change (e.g., just SLR) but few

have addressed their combined effects on estuarine flooding. This is problematic since there can be interactions among processes. As an example, SLR has been shown to non-linearly modulate storm surge [Smith et al., 2010].

Studies that have attempted to holistically model estuarine flooding include Cloern et al. [2011] who studied century-scale change in San Francisco Bay. Their study used a hybrid approach, coupling process-based and statistical sub-models to evolve water column properties over time. Barnard et al.'s [2014] Coastal Storm Modeling System (CoSMoS) also used an interlinked model framework but with less focus on estuary WLs (although CoSMoS 2.0 improves upon this). Cheng et al. [2015a] did a preliminary study of a single PNW estuary using a fully process-based model framework. The current study builds upon this effort by including additional physical processes, conducting a comparative study, and applying the results to the production of flood mapping products.

The objective of the present study is to further develop this research direction through applying a comprehensive process-based modeling framework to the problem of estuarine flooding under current and future conditions. A process-based approach allows direct modeling of climate induced changes to all drivers (streamflow, wave forcing, etc.) of estuarine WLs. This study hypothesizes that considering integrated forcing on estuaries results in significant spatial variability in extreme WLs. This hypothesis is in contrast to the common procedure of a static "bath-tub" approximation (i.e., the assumption of a horizontal water surface), which has been shown to result in significant errors [Gallien et al., 2014]. Results from this study will quantify this error as well as provide information on how it may be evolving as a result of climate change. Additionally, flood surface information will be combined with a high resolution digital elevation model (DEM) to determine the extent of flooding and how it may be changing over time.

This paper is organized with the initial section providing a description of the two study sites (section 3.2). The overall modeling framework is then introduced (section 3.3) with description of the individual sub-model components. This is followed by information on nonstationary RIs (section 3.4). The paper then moves on to the results (section 3.5) and closes with a discussion of findings (section 3.6).

3.2 Study Sites

This study focuses on two PNW estuaries, Coos and Tillamook Bays (Figure 13). These estuaries were selected for two reasons. First, each has an active watershed / estuary organization (Coos Watershed Association / Tillamook Estuaries Partnership) which allowed for data sharing and project collaboration. Secondly, the estuaries have similar forcing profiles (Table 1) which maintains some degree of comparability between the two locations. However, the estuaries do have very different physical configurations which allows some exploration as to the importance of local bay configuration.

Table 1: Coos and Tillamook estuarine characteristics (after Engle et al., [2007]). Wind characteristics are at gauge locations. Wave characteristics are offshore at buoys 46002 and 46005. Streamflow values are from this study's hydrological analysis

	Coos	Till.	
Estuary Area	43.8	37.9	km^2
Estuary Drainage Area	1520	1430	km^2
Estuary Volume	0.207	0.071	$\text{m}^3 \times 10^9$
Ave. Daily Flow	61	85	m^3/s
Tidal Prism	0.066	0.061	$\text{m}^3 \times 10^9$
Mean Tidal Range	1.7	1.9	m
Ave Summer Salinity	25.8	24.5	psu
Ave. Depth	3.5	1.4	m
Ave. Wave Height	2.7	2.8	m
Ave. Wind Speed	2.4	4.6	m/s
Ave. Wind Direction	164	190	deg.

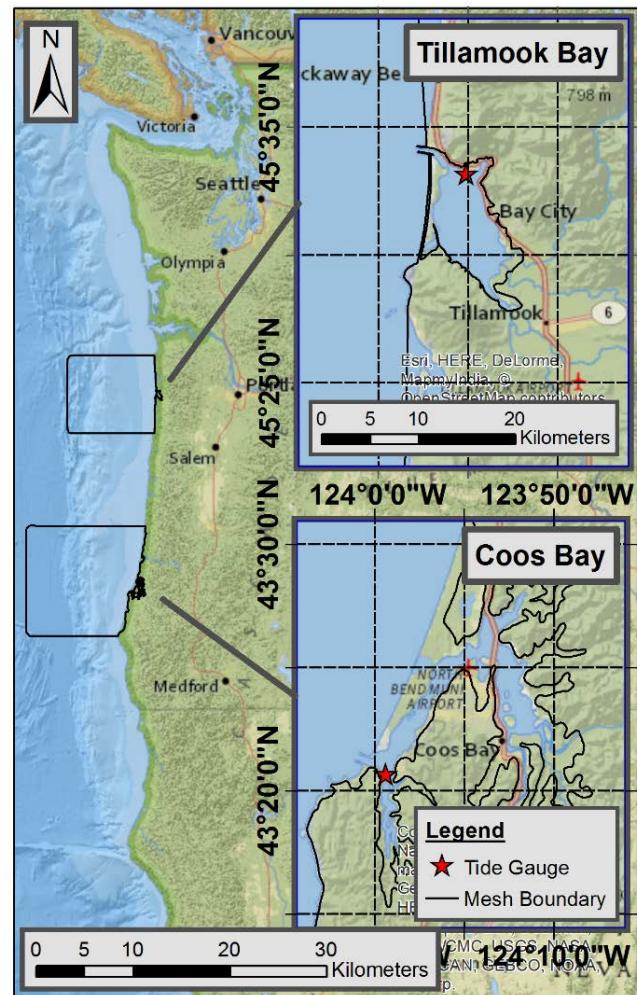


Figure 13: Overview map (left) and detail views (inset boxes) of Tillamook and Coos Bays on the Oregon coastline. Both insets are at the same scale (shown on Tillamook Bay inset).

In terms of physical layout, Coos has a unique hook shape while Tillamook has a more classical bay form with an enclosing sand spit defining the western edge (Figure 13). Both estuaries have a jettied inlet and U.S. Army Corps of Engineers maintained channel. Tillamook's entrance is maintained at 5.5 meters deep, 60 meters wide while Coos has a significantly larger deep-draft channel at 18 meters deep and 210 meters wide. Coos has a larger average depth and surface area giving it a larger estuary volume than Tillamook by around 130 million cubic meters. Coos is also skinnier with a deep channel along the majority of its length. Tillamook, on the other hand, has a channel only near the entrance with the bulk of its area being defined by broad shallow tidal flats.

In terms of forcing, the estuaries are quite similar although with Tillamook experiencing slightly higher environmental forcing. Tillamook's tidal range is approximately 20 cm larger than Coos and mean wave, wind and streamflow forcing are all modestly more intensive.

3.3 Methods

This study utilizes a suite of models and data sources to determine the hydrodynamic response of the study sites to climatic forcing. The overall workflow is that an Atmosphere-Ocean Global Climate Model (AOGCM) serves as the “parent” model providing forcing to a suite of “child” models. These in turn provide the forcing to the hydrodynamic model of the estuaries themselves. This modeling framework is conceptually illustrated in Figure 14. In terms of final-stage output, this modeling chain produces continuous time series of WLs (and other variables) at chosen output points. Provided that a sufficiently long-term simulation is carried out, this spatially explicit information allows for the development of flooding inundation maps at a variety of RIs. The following sections describe each component of the model framework identified in Figure 14.

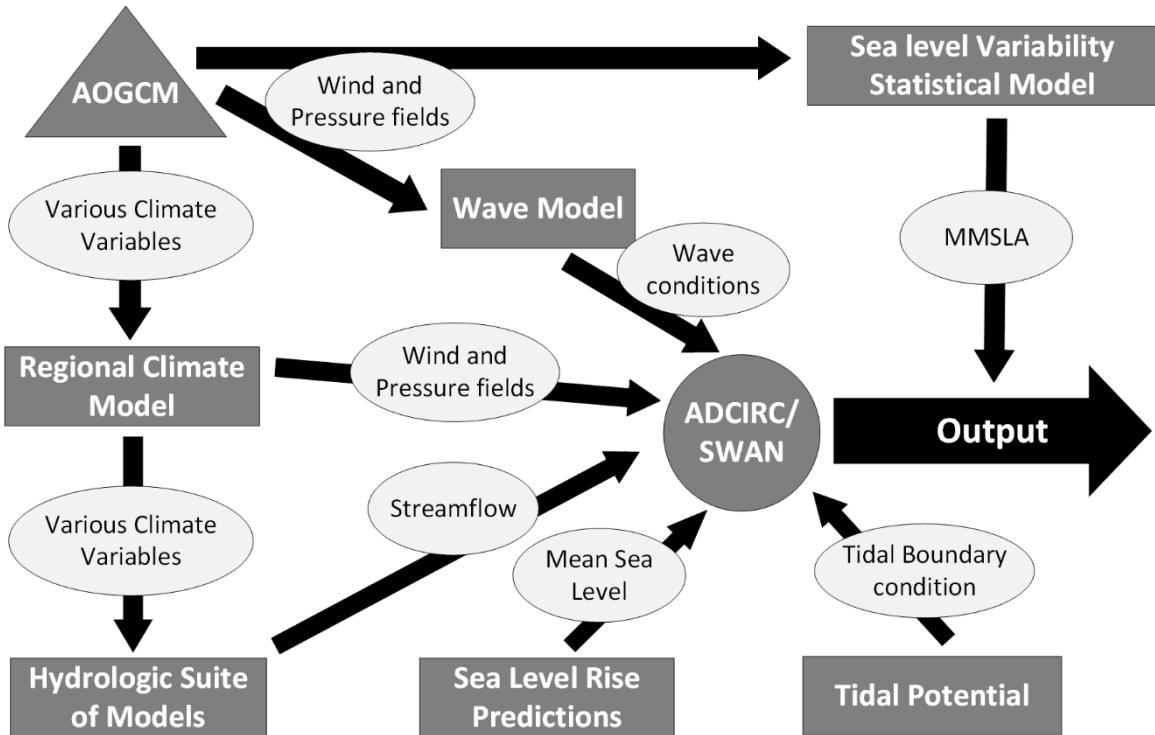


Figure 14: Overall general modeling framework. The dark gray triangle labeled AOGCM in the top left corner is the “parent” model. Dark gray rectangles represent sub-models. Light gray ovals represent variables that are passed between modeling components.

As is common in climate change impact studies, this study uses paired simulations with hindcast and forecast boundary conditions. Two simulations were carried out; one for the period 1979-1999 and the other for the period 2041-2070. The hindcast period is used rather than direct observations to control for biases in the AOGCM and modeling framework.

3.3.1 Climate Model

This study uses data from the North American Regional Climate Change Assessment Program (NARCCAP) [Mearns et al., 2009, 2013]. NARCCAP provides an ensemble of AOGCMs paired with Regional Climate Models (RCMs) focused on the North American continent. The project’s future runs are forced by the Special Report on Emissions Scenarios (SRES) A2 emissions scenario which represents one of the higher emission, anthropologically controlled climate scenarios for the fourth IPCC assessments report [Nakicenovic et al., 2000].

This scenario was chosen (by NARCAAP) as a conservative but plausible climate trajectory that is in-line with current emissions / population patterns. There are other downscaled climate products available (e.g., MACA; Abatzoglou, [2013]) that are based on more current CMIP5 scenarios; however, NARCCAP was the only climate product, at the start of this project, that provided the necessary offshore coverage. Most other products are masked so that data are only available on land surfaces.

From the model pairings available in NARCCAP, the Community Climate Science Model/Canadian Regional Climate Model (CCSM/CRCM) combination was chosen as providing the best agreement with local in-situ meteorological data. The AOGCM component of this pairing was used for datasets requiring larger spatial coverage while the more finely resolved (both temporally and spatially) RCM was used for nearshore or terrestrial sub-models. When using RCM data, it was found that atmospheric parameters were biased so a univariate statistical bias correction (quantile mapping, Déqué, [2007]) was performed. The “target” dataset used for the bias correction was the North American Regional Reanalysis (NARR) [Mesinger et al., 2006] dataset interpolated to the location of the CRCM grid nodes. AOGCM data was also found to be biased but was not bias corrected as there is no target dataset that spans the full global climate model extent. Instead this bias correction was handled within the relevant sub-model.

3.3.2 Wave Model

A basin-scale Wavewatch III v3.14 (WW3) [Tolman, 2009] simulation was performed in order to characterize wave climate and provide offshore wave boundary conditions for the two hydrodynamic model domains. WW3 has seen significant success in the PNW for reproducing wave conditions [Garcia-Medina et al., 2013] and is well suited to application at large scales [Hanson et al., 2009]. The model was run with two nested grids based on the National Centers for Environmental Prediction (NCEP)’s operational Global and North East Pacific models with a resolution of (1° by 1.25°) and (0.25° by 0.25°) respectively.

Wind forcing for the WW3 model was provided by the CCSM global model. Wave model predictions were found to exhibit a significant bias in comparison to observed wave parameters in the PNW study areas. This bias is likely a result of the low-resolution wind fields [Holthuijsen et al., 1996] or the AOGCM’s inability to reproduce marine winds [Hemer et al.,

2011]. This transfer of bias from wind fields to wave model output has been similarly reproduced in other studies [Feng et al., 2006, Hemer et al., 2011]. Sensitivity analysis showed that running the hydrodynamic model with over-predicted wave heights produced unrealistic flooding values and overwhelmed the influence of other signals contributing to extreme WLs. We therefore utilized a bivariate statistical bias correction technique that corrects both the marginal distribution of wave height and period as well as maintains the correlation structure [Parker and Hill, 2017].

3.3.3 Hydrological Model

Streamflow inputs were developed using a series of weather, snowmelt, and hydrological routing models. Specifically, the MicroMet / SnowModel / HydroFlow suite of spatially distributed models were used [Liston & Elder, 2006a; Liston and Elder, 2006b; Liston and Mernild, 2012]. Readers are directed to the source citations for full details of the models. In summary, this suite distributes meteorological forcing variables, uses energy-balance methods to calculate surface energy exchanges, and then uses a simple linear reservoir-routing procedure to route water across the landscape to the coastline. Cheng et al. [2015a] successfully applied this modeling suite to the PNW and Beamer et al. [2017] to the Gulf of Alaska under climate change scenarios.

The models were calibrated and validated using the NARR dataset as forcing.

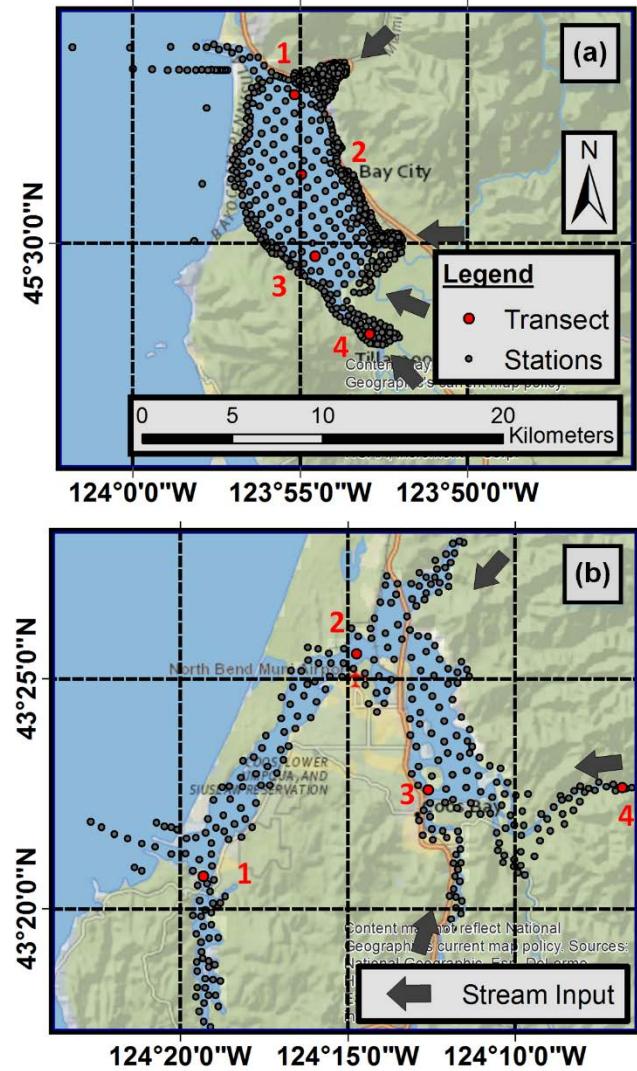


Figure 15: Station output locations at Tillamook and Coos Bay study sites, panel (a) and panel (b) respectively. Grey arrows show hydrologic inputs into the model domain. Red points and numbers represent transect station locations (see section 3.6.1). Both figures are at the same scale (shown in panel a).

Both study sites have multiple USGS stream gauges for comparison and Nash-Sutcliffe efficiencies (NSE; Nash and Sutcliffe, [1970]) as well as coefficients of determination (r^2) were found to be high in all cases. This indicates that the hydrological model adequately captures the hydrology of the Coos and Tillamook watersheds and provides confidence moving on to simulations using CRCM input. For consistency with other aspects of the model framework, the CRCM variables were all bias corrected using quantile mapping with the NARR dataset as the target.

The gridded hydrologic model produces runoff values at every model cell along the coastline. However, most cells produce very low flows so only significant points were selected for inclusion in the hydrodynamic model. For Tillamook, four rivers were included (the Kilchis, Wilson, Miami, and Trask) which were found to capture 95% of the basin streamflow. For Coos, seven points were chosen for inclusion representing 90% of the basin streamflow. The streamflow from these points was aggregated into three inputs into the hydrodynamic model at the location of the Coos River, Palouse Slough and Noble Creek (Figure 15).

3.3.4 Monthly Anomaly Model

An important component of measured WLs along the west coast of the U.S. comes from Monthly Mean Sea-Level Anomalies (MMSLA). These low frequency variations in WLs are caused by a wide variety of forcings ranging from local wind stress to large scale climate patterns (such as El Nino) [Chelton and Davis, 1982]. A comprehensive understanding of MMSLA remains complex, as they integrate a large number of simultaneous processes operating over a wide range of spatial and temporal scales. However, there have been many attempts to model the leading order terms found in MMSLA signals. This study follows the work of Thompson et al. [2014] who used statistical regression and a subset of wind stress metrics to successfully reproduce the bulk of MMSLA variability across the west coast. The utilized regression is given by:

$$\eta = \eta_0 + a\tau_{eq} + b\tau_{ls} + c\tau_{xy} + \epsilon$$

Equation 4

where η is monthly mean sea level, η_0 is the regression y-intercept term, τ_{eq} is the equatorial wind stress, τ_{ls} is the local alongshore wind stress, τ_{xy} is the wind stress curl, ϵ is the residual error, and (a, b, c) are regression coefficients. The reader is directed to the original source publication for additional information regarding the specifics of the wind stress coefficients as well as the scientific basis.

Figure 16 shows the result of this regression for the Tillamook study site (Coos is not shown). While this formulation does not capture all variability in MMSLA ($R^2 \approx .6$), it does qualitatively capture the MMSLA signal and is based entirely on variables that are readily available from the NARCCAP dataset. MMSLA WLs are added to the hydrodynamic model time series as a post-processing step (see Figure 14). Not including MMSLA within the hydrodynamic model does exclude some potential non-linear interactions between MMSLA, WLs and other forcings.

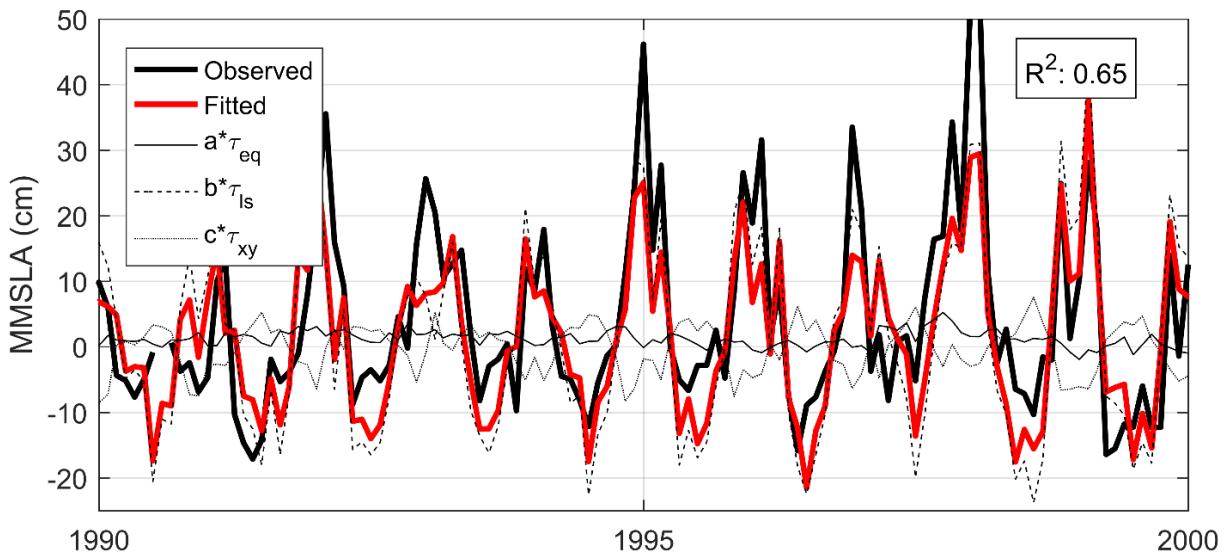


Figure 16: MMSLA regression for the Tillamook study site in the format of Thompson et al. [2014]. Fitted contributions to MMSLA from predictor variables τ_{eq} , τ_{ls} , and τ_{xy} are shown as thin black lines (full, dotted and dashed respectively). The bold black line is the observed MMSLA signal while the bold red line is the total fitted MMSLA signal.

3.3.5 Sea Level Rise

SLR was included in the modeling framework as a change to mean sea level (MSL) within the hydrodynamic model. Projections were taken from the National Resource Council (NRC) report [NRC, 2012] which developed local estimates for SLR along the Pacific coast. These estimates include contributions from steric/dynamic ocean modifications, glaciers and ice caps, sea-level fingerprint effects, and vertical land motion (e.g., isostatic adjustments). In calculating local SLR estimates, the NRC used a combination of IPCC 4th assessment projections (mid-range scenario) and an extrapolation methodology for the cryosphere components. This produced values larger than either the IPCC 4th or 5th assessments but still below some estimates for mean 2100 global SLR [Vermeer and Rahmstorf, 2009].

SLR data were taken from the nearest reported location to Coos and Tillamook Bay, Newport Oregon, which is situated approximately between the two study sites. The NRC report provides projection values (and ranges) for 2030, 2050, and 2100. A cubic spline was fitted to these values to allow a smooth interpolation to intermediate years. Multi-decade model runs of the hydrodynamic model were broken into smaller 3-month periods for model stability reasons and MSL was updated accordingly for each of these simulation blocks. This allowed for changes in MSL in a step-wise but nearly continuous fashion.

3.3.6 Local Hydrodynamic/Wave Model

The coupled ADCIRC-SWAN (ADCSWAN) model [Dietrich et al., 2011] was used for this study. ADCSWAN is a highly configurable in terms of implemented physics and readers are directed to the source publication and model manuals for a full description of options and parameters. ADCIRC [Luettich and Westerink, 1992] solves the hydrodynamic portion of this pairing through the shallow-water equations. It uses an unstructured horizontal grid which allows for finer spatial resolution in regions of complex bathymetry. Bathymetry for the model grid was developed through blending Oregon Department of Geology and Mineral Industries (DOGAMI) LiDAR [DOGAMI, 2009] and a variety of National Oceanic and Atmospheric Administration DEMs [NOAA, 2018]. Wetting and drying were enabled due to the significant intertidal areas

present in both bays. Non-linear bottom friction was used with a spatially variable friction factor set based on general land use classes [Dietrich et al., 2011; Homer et al., 2011].

ADCIRC was run in the 2D depth-integrated mode. ADCIRC has 3D barotropic/baroclinic modes which would capture additional physics not included with this study's methodology. However, this comes at a substantial additional computational cost that is prohibitive for the climatological timescale of this study. Additionally, our primary focus is on WLs (not full water column properties), and research has shown that storm surge and tidal signals can be accurately resolved by 2D barotropic models. Specifically, Resio and Westerink [2008] point out that 3D effects are readily absorbed by model calibration coefficients. Additionally, a sensitivity study by Weaver and Luettich [2010] found that differences in predicted WLs between 3D and 2D models were on the order of 5% over most of the domain. These modest differences suggest that a 2D model can be an effective and efficient choice for studies of this type.

SWAN is a third-generation spectral model that solves the spectral action balance equation to compute the spectral evolution of wind waves. The unstructured format of SWAN [Zijlema, 2010] was utilized to allow tight coupling (on the same grid) with ADCIRC. SWAN was run in a nonstationary mode with offshore forcing provided by a temporally varying JONSWAP spectrum fitted to bulk wave parameters.

ADCSWAN was run with atmospheric forcing provided as gridded horizontal wind components and surface pressure; wave forcing using SWAN's nonstationary TPAR parametric spectrum file; hydrologic input as a normal flux into the domain, and tidal forcing at the ocean boundaries. Tidal forcing was defined as the eight locally dominant constituents (K1, O1, P1, Q1, M2, S2, N2, and K2) with location dependent amplitude and phase defined by the ENPAC tidal database [Mark et al., 2004] and the simulation time dependent nodal factor and equilibrium argument defined by the T_tide harmonic analysis package [Pawlowicz et al., 2002]. The ADCSWAN model was first validated at each location using a tidal simulation compared against NOAA tide gauge predictions. Using a one month simulation, both the Tillamook and Coos grids were found to have R^2 values greater than 0.97. The Tillamook model was additionally validated

against the largest storm of record (the Great Coastal Gale of 2007; Crout et al., [2008]) with good agreements to extreme WLs [Cheng et al., 2015b].

3.4. Nonstationary Recurrence Intervals

This study primarily considers extremes which will be quantified in terms of RI events, since engineering design and community planning often rely on this concept. The traditional definition of an RI is built on an assumption of stationarity, or time-invariance. This assumption makes the definition of a RI simultaneously the inverse of the probability that an event of a given magnitude will be exceeded in a given year and the expected recurrence period of that event. This definition breaks down under nonstationary conditions, which can be experienced due to climate change and SLR. Reconciling a nonstationary environment with traditional design methods based on stationary assumptions is an ongoing challenge. Proposed alternatives include effective design value [Katz et al., 2002], expected waiting time [Olsen et al., 1998; Sala and Obeysekera, 2014], expected number of events [Parey et al., 2007, 2010], design exceedance probability, and design life level [Rootzen, 2013]. Each of these definitions represents a unique projection of the stationary case for nonstationary conditions. Problematically, the chosen metric can result in significantly different calculated RIs while most users simply interpret the result as comparable to the stationary case. This highlights the importance of rigorously defining utilized nonstationary RI formulation as well as considering if the utilized metric fits design conceptions.

This study uses the effective design value [Katz et al., 2002] interpretation of nonstationary RIs. This defines a temporally varying RI (termed an effective RI, or design value, by Katz et al. [2002]) which holds the probability of occurrence for an event constant through time. This preserves an intuitive definition of RIs as well as how nonstationarity impacts extremes. While this definition does add an additional dimension of time to RIs, this does not significantly impact usability as the user can simply choose the RI value from a standard RI curve plot and then add the amount of nonstationarity (for this study, SLR) corresponding to the year of interest. Figure 17 shows an alternative presentation with calculated effective RIs for Tillamook Bay shown as a family of curves.

The effective RI definition was primarily chosen for its intuitive interpretation but also because it pairs well with the format of the data from this study. WL data are output from the model in reference to MSL. This results in a WL time series that does not show any discontinuity or trend from changing sea level, a signal that would only be visible if viewing WLs relative to a non-tidal datum. This stationarity, as a function of datum, makes it possible to separate the calculation of RIs and the nonstationarity of the time series. This avoids the complications of fitting a nonstationary GEV (Generalized Extreme Value distribution) and the corresponding loss of degrees of freedom from estimating the nonstationary trend from the data. Furthermore, most nonstationary GEV analyses are forced to use linear trends due to limited degrees of freedom. This approach allows a more complicated trend that follows experienced SLR. This can be seen in Figure 17 by the quadratic nature of calculated effective RIs.

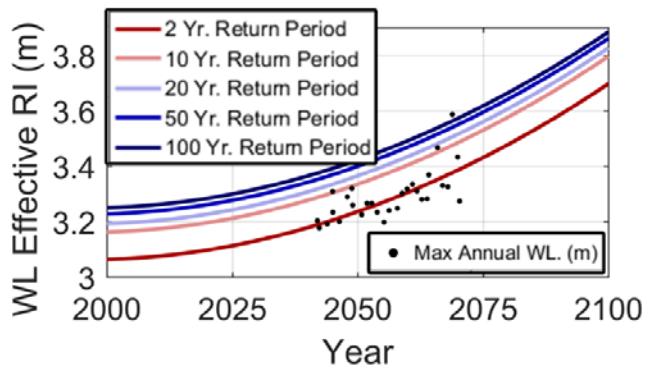


Figure 17: Effective RI curves for Tillamook Bay. The increase in effective RIs through time corresponds to the rate of SLR. Black dots show the simulated max annual WLs at the tide gauge (in NAVD 88).

3.5 Results

Model output was saved at a subset of model nodes (stations) in order to keep output files manageable in size. Output stations were spaced evenly across the bay in order to capture spatial variability (Figure 15). While ADCSWAN has the ability to write out numerous variables (including wave heights, periods, etc.) the focus of this study is on WLs so discussion here will be limited to that variable. The model was run in 3-month long segments with a 2 week overlap to avoid discontinuities in dynamic processes. Output data from these segments were then recombined into a single continuous time series.

3.5.1 Recurrence Intervals (Tide Gauge Locations)

RIs were calculated using a GEV distribution fitted to annual block maxima events. Figure 18 shows this analysis for observed, modeled historic and modeled future time series at the Coos and Tillamook Bay tide gauge location. The calculated historic period RI curve for Tillamook (Figure 18a) exhibits good agreement with observations with a maximum offset of around 4 cm. The agreement is less favorable for Coos Bay (Figure 18b), with a maximum offset of around 9 cm. For both locations, the largest difference between observed and modeled RIs is for medium (approximately 10-year) RI events. This is because both modeled RI curves exhibit a different curvature than the observed.

Figure 18 additionally includes future RI values plotted with SLR not included. Using the effective design value interpretation, this can be thought of as future RIs but with the chosen design year being 2000. This is not a realistic design condition (as future RIs are calculated over the years 2041-2070) but shows future RIs as a function of only changing forcing (no SLR). Both plots show the future RI curve as exhibiting smaller WLs than the historic RI curve, hinting at a reduction in extreme WL forcing into the future.

Figure 18 presents these results in a classic RI curve format but they can additionally be viewed as effective RI (as described in section 3.4). Effective RIs were developed by calculating

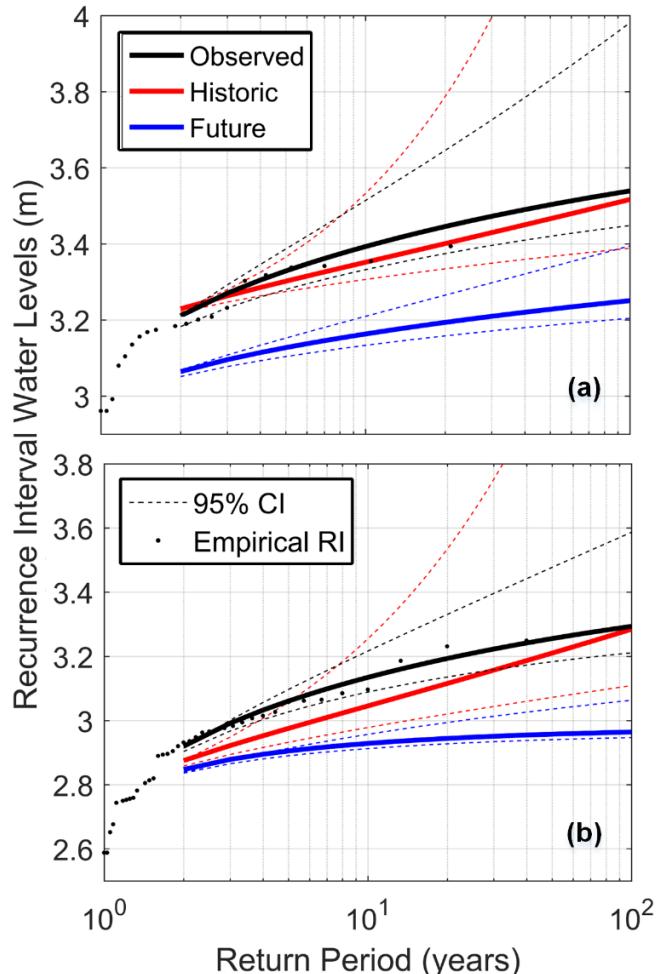


Figure 18: RIs for the Tillamook and Coos Bay tide gauges, panel (a) and panel (b) respectively. WLs are in NAVD88. Confidence intervals are calculated using a likelihood-based method and are plotted as dotted lines. SLR has not been included for future RI curves.

stationary RIs from the WL time series (relative to MSL) and then adding SLR (Figure 19). Figure 19 contains both the future and historic effective RIs on the same plot to visualize how RIs change upon entering a nonstationary climate regime. Based off model assumptions, historic effective RIs are flat lines through time (as they include no nonstationary component). As a start, Figure 19 shows the same RI behavior as above with historic RIs being higher than future RIs for the current period (year 2000).

Moving forward in time, this plot demonstrates that SLR eventually overtakes this effect to result in a higher future flood. The point of overtake (where nonstationary RIs begin to predict a larger flood) is plotted in Figure 19 as colored dots on the x-axis. This intersection is significantly earlier for lower RIs (within 20 years for a 2-year event) than for higher RI events (over 60 years for a 100-year event). This result is similarly shown for Tillamook Bay, although with the spread in overtake time being smaller.

3.5.2 Recurrence Intervals (Spatially Variable)

Section 3.5.1 discusses RIs at a single location (the tide gauge), but a key feature of this study's methodology is the ability to explore spatial variability in RIs across the study sites. Figure 20 demonstrates this by plotting the 100-year RI WL calculated at each output station. Analysis was limited to stations that were wet over 75% of the record length. This was implemented to limit uncertainty in GEV analysis due to insufficient record lengths at only periodically wet stations.

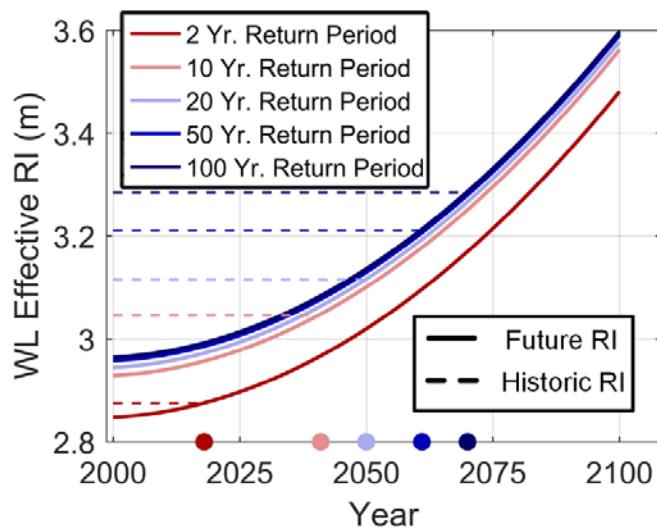


Figure 19: Coos Bay effective RI WLs for the historic (dashed line) and future (solid line) periods. The location of intersection between the historic and future effective RI curves is plotted on the time axis as a dot with color corresponding to return period.

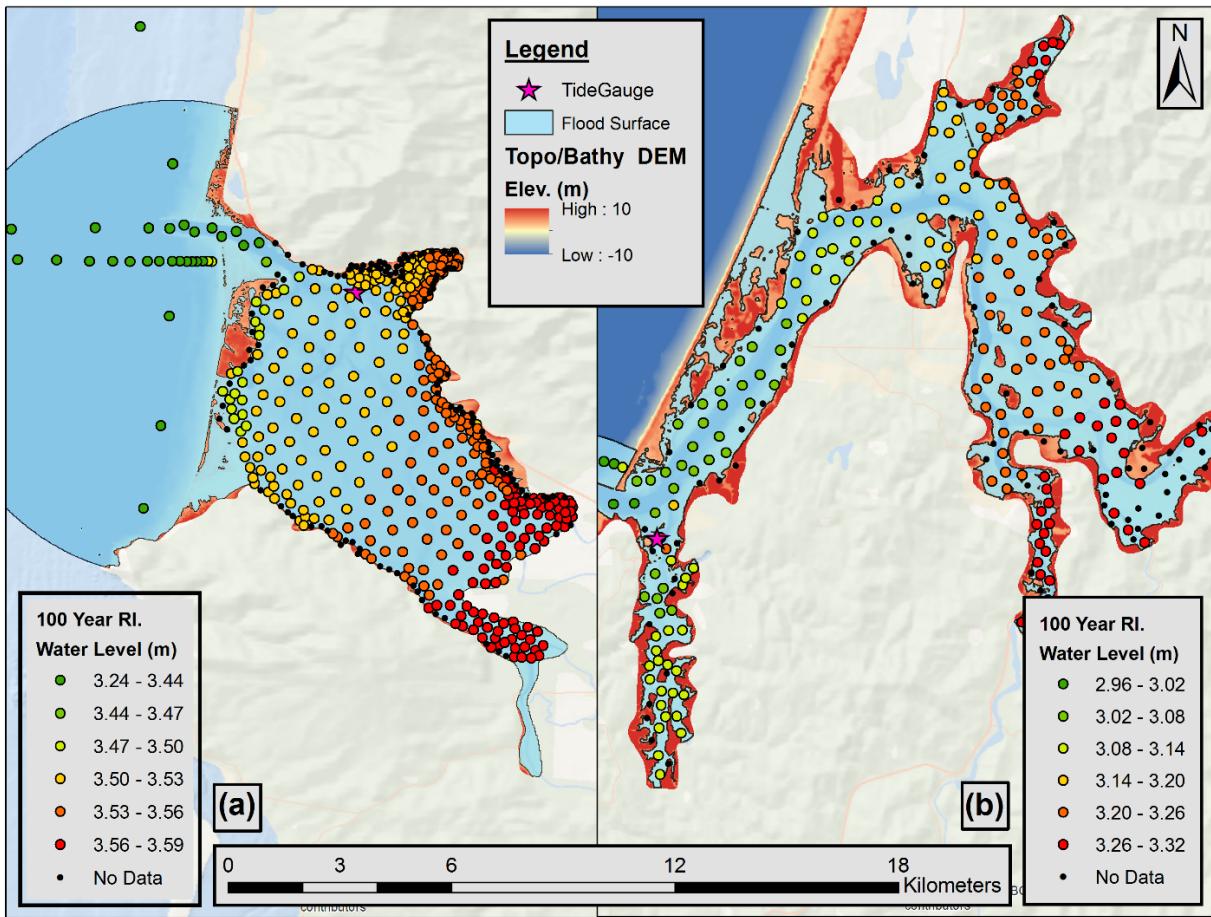


Figure 20: Historic period 100-year flood surface results for Tillamook (a) and Coos (b) Bays. Individual stations are plotted with the color scale indicating the individual 100-year RI WL magnitude. Note that the color scale is different between the two study sites. The calculated flood surface is plotted as a blue transparent region.

It is apparent in Figure 20 that there is a significant gradient in extreme WLs across both study site estuaries. For Tillamook, WLs differ by approximately 25 cm and for Coos by around 35 cm. The gradient in WLs is orientated such that the minimum WL is located near the estuary entrance with extreme WL height increasing with further distance into the estuary. Of particular interest, this trend is the same for both study sites suggesting that this pattern may be more generally applicable. While Figure 20 only plots the 100-year event, this gradient is maintained across various RI periods.

The spatially variable WLs produced from this analysis provide the information necessary for building flood maps. This was accomplished through fitting a smooth surface to

the scattered stations using spatial spline interpolation [ESRI, 2016]. The 100-year RI WL surface was then intersected with the estuary DEM with all locations below the extreme WL surface defined as “flooded”. This methodology makes some assumptions, such as extrapolation at the edge of the surface where station output was not available. The calculated flood surface is shown in Figure 20 as a light blue surface.

It is useful to compare these results to other flooding products. Figure 21 shows the calculated present-day 100-year flood surface overlaid with the Federal Emergency Management Agency (FEMA) 100-year flood plain for Coos Bay [FEMA, 2014]. It can be seen that the produced flood inundation zones are in general agreement with only small local differences. Note that zones in the upper tributaries that are shown as not inundated by the present method simply reflect the boundary of the model grid. Flood zones in the FEMA study were generated using a combined probability analysis for flood forcing as well as selected process-based modeling. Flood elevations were calculated using Hydraulic Engineering Center (HEC) models and the Dynamic Estuary Computer Model [Federal Water Quality Administration, 1975]. The relative agreement in 100-year flood zones between this study’s methodology and FEMA’s very different approach gives confidence in the overall flood surface moving forward with consideration of future flooding extent. Note that FEMA flood zones do not consider climate change, a primary objective of this study.

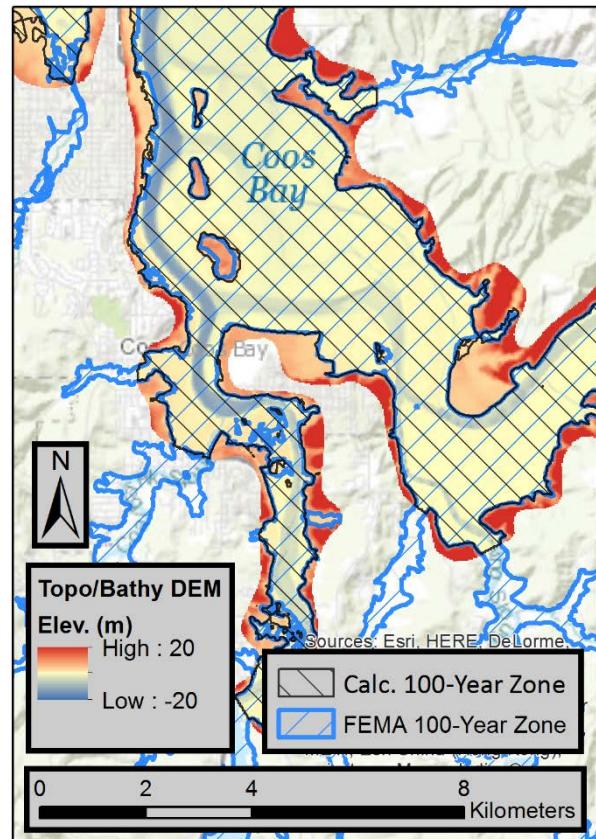


Figure 21: Comparison of 100-year flood inundation zones for FEMA (blue cross hatch) and this study (black cross hatch). The plotted region is the lower eastern region of Coos bay. The calculated 100-year flood zone is limited to the DEM domain (shown by the topo/bathy DEM) while the FEMA flood zone extent is larger.

3.5.3 Changes to RI spatial variability

In order to investigate how RIs may change into the future, an additional RI analysis was performed for the future simulation period. With an identical configuration (both for the process-based and statistical models) to the historic period simulation, the only free variable is the AOGCM forcing under climate change. Therefore, a comparison shows how extreme events can be expected to change (in a relative manner) over time.

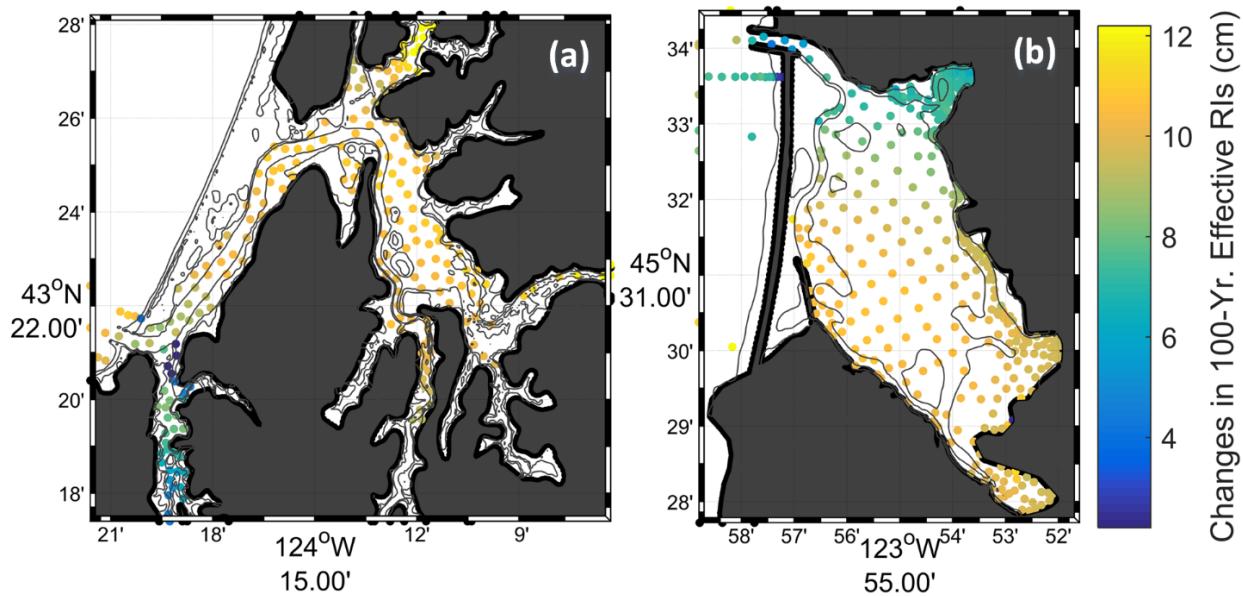


Figure 22: Changes to 100-year RIs plotted at output stations. Changes are calculated as the future effective RI WL (at year 2050, 0.172 m of SLR) minus the historic RI WL. The dark black line represents the edge of the computational domain and light gray lines are the -5, 0, and 5 meter contours.

Figure 22 shows the difference between future effective 100-year RI WLs (calculated at year 2050) and the historic 100-year RI WLs. If the RI difference plot (Figure 22) were to show no spatial variability, then the same pattern seen in Figure 20 would be replicated in the future (only the magnitude would change). This would be an important result signifying that the spatial pattern of current extreme flooding events will remain unchanged into the future with only an estuary-wide SLR adjustment being required to update hazard assessments. Instead, Figure 22 shows a spatial pattern that is qualitatively similar to that seen in the historic RI plots. Specifically, results show the change in WLs increases with further distance into the estuary, a

result that signals an increase in the spatial gradient of RIs as the study sites move into the future. Looking at other RI event periods (not shown), results show a gradual decrease in changes to spatial variability from around 10 cm for the 100-year RI event to approximately no change for the 2-year RI event.

3.6. Discussion

3.6.1 Drivers of Extreme WLs

Figure 20 provides a starting point for understanding the physical processes driving extreme WLs. In particular, both study sites exhibit a similar pattern with spatial correlation between locations of hydrological input and locations of higher extreme WLs. Comparison to salinity measurements by Lee and Brown [2009] show that salinity and extreme WLs are approximately inversely related. This was further investigated by re-simulating the largest annual maximum events at the two study sites for the historic and future period. A series of 17 day simulations bracketing the largest events were performed with various forcing components turned off to determine the relative contributions to WLs (Figure 23). The WL contribution from each forcing was calculated by subtracting a simulation with the forcing of

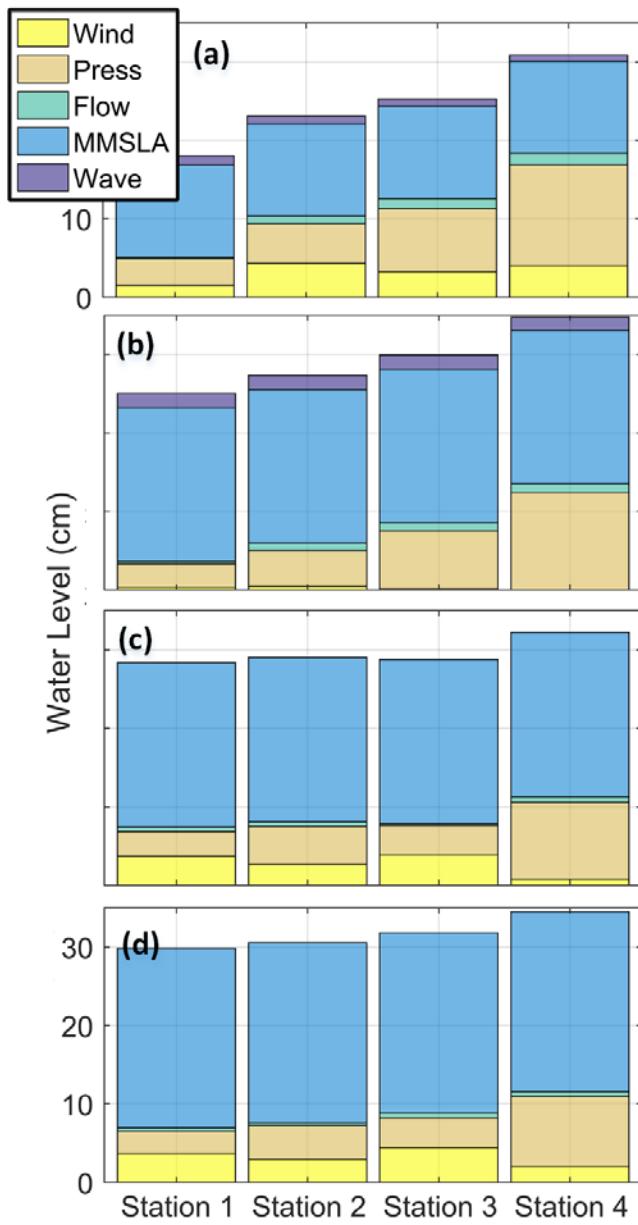


Figure 23: WL contributions from various forcings during the largest simulated WL event. Stations transition from the tide gauge (station 1) to the far river outlet of the estuary (see Figure 15). Panel (a) and (b) are for the Coos historic and future periods. Panel (c) and (d) are for the Tillamook historic and future periods. All panels have the same y-axis scaling

interest turned off from a simulation with full forcing. Components are plotted at the time of maximum WL occurrence.

The results show that, for the simulated events, pressure and MMSLA are the largest components of non-tidal residual. Streamflow and wave forcing were found to be an order of magnitude less important than pressure and MMSLA. This is especially true for the two events at the Tillamook study site (subplot c and d) where streamflow and wave forcing were found to be of negligible importance to extreme event WLs. The wind contribution was found to be variable across stations and events. This is expected as wind setup is very dependent on the estuary geometry and the wind direction of the specific event.

However, this analysis is only for the maximum annual event and others events likely have different compositions in terms of forcing contributions. Further investigation shows that, for both study sites and both climatological periods, extreme event magnitude is not correlated with any individual forcing other than MMSLA and tides (p -value > 0.05 when calculating correlation between WL magnitude and forcing magnitude at annual maximum events). This is reinforced by the fact that annual maximum events are found to often occur during below average non-tidal forcing conditions (e.g. below average wind or waves). This is even true for MMSLA with a few annual events occurring during negative MMSLA forcing. This supports the conclusion of variable WL composition, in terms of forcing contributions, and the predominance of compound events in the PNW. A full analysis of WL contributions would therefore require simulating many events to understand this variability.

Tides are additionally a key driver of extreme events in the Pacific Northwest. Tides were not plotted in Figure 23 for scale reasons but were found to be the largest fraction of WLs (an average of 185 cm for Till and 145 cm for Coos). Tides are the dominant source of WL variability in the Pacific Northwest [Allan and Komar, 2002a] so it follows that extreme WLs would most often occur during a high tide. Therefore, the concurrent timing of tides and non-tidal forcing becomes a major control on WL magnitude. A comparison of the forcing time series and WLs shows that the occurrence of a high tide is generally not concurrent with the maximum of other forcing components. This represents a mechanism explaining the importance of compound events in the PNW, in which extreme water levels are not associated with extreme

forcing. This is also a potential reason for some forcing components showing a negligible influence to extreme water levels (e.g. waves in Figure 23). While waves have been shown to be important drivers of NTR in PNW estuaries [Cheng et al., 2015b; Olabarrieta et al., 2011], tidal modulation means extreme WLs do not necessarily occur during maximum wave energy events. Finally, this provides evidence that the common methodology of simply adding the largest storm surge to a high tide may result in significant overestimations of event magnitude.

3.6.2 Spatial variability in RIs

It is common practice to calculate WLs at a convenient location (such as a tide gauge) and then apply this value across the entire study domain. While larger estuaries (e.g., Delaware Bay) may have multiple tide gauges, smaller estuaries typical of the PNW tend to have one or none. A spatially constant assumption represents a major simplification as even tides can produce significant spatial WL variability in semi-enclosed basins [Holleman and Stacey, 2014]. Additionally, spatial variability is particularly important for estuaries as they are often regions of low-gradient topography where a modest change in water elevation can correspond to a large change in inundated area. Results (Figure 20) show variability in WLs for both locations in excess of 25 cm. Furthermore, the minimum of this gradient is at the estuary mouth where, in the PNW, the tide gauge is generally located. This means that estimating flooding from a tide gauge will result in under-predictions for flooding with errors increasing with upstream distance into the estuary. This result strongly supports the importance of considering spatial variability in WLs within flood hazard assessments.

Between the two study sites, Coos is found to have a larger WL gradients from the mouth to the interior bay by about 10 cm. This is contrary to the expected result that Tillamook, with its proportionally larger forcing, would have a larger gradient. Streamflow was expected to have significant impact on WLs but the results for Tillamook found a minimal contribution. Figure 23 shows that, while the streamflow component does increase moving shoreward for Coos bay, the majority of the WL differential for both sites is driven by pressure. An additional component of the WL gradient is from tidal forcing which produces around a 10 cm differential between the estuary mouth and the inner bay for both locations.

Results show that spatial variability is predicted to change into the future. However, this result is primarily shown for high RI events with lower RI events not showing any change in spatial variability for the future scenario. Lower RI period events are better constrained statistically than higher RI period events so it is possible that some proportion of the predicted spatial variability is a function of GEV analysis on a temporally limited record. A modeled record longer than the period used here (20 years for historic, 30 years for future) could help to illuminate if this conclusion is a physical result.

3.6.3 *Changing Extreme Events*

With analysis showing RIs evolving through time, a natural next question is how climate change is modifying the estuarine system. To help illustrate this, Table 2 shows a comparison of forcing statistics under the historic and future climatological periods for Tillamook. Results show a decrease in most forcing variables, both as an overall average and for average forcing during extreme events. This result is similarly seen for Coos Bay (not shown). This suggest that the decrease in RI shown in Figure 18 is caused by a broad scale reduction of forcing on the estuary for the GCM scenario considered. Unfortunately, with all modeled forcing shown to be reduced for the future period, it is difficult to conclusively differentiate which drivers are controlling changing RIs. Furthermore, as GEV analysis is based on a parametric fit of multiple annual maximum events, it is difficult to characterize the cause of changing RIs without considering the aggregate behavior of all the events to which the GEV is fitted (rather than a single event as shown in Figure 23).

Table 2: Comparison of forcing for Tillamook Bay under the historic and future climatological periods. “Ave. Overall” is a full time series average while “Ave. Annual Max” is an average of forcing during observed annual WL maximum events.

	<u>Historic Period</u>						
	WL	Hs	Flow	Wind Mag.	Wind Dir.	Press.	MMSLA
	(m)	(m)	(m ³ /s)	(m/s)	(deg.)	(Pa.)	(cm)
Ave. Overall	0.0	2.8	85	4.5	160	98,400	0
Std. Overall	0.8	1.4	82	2.6	109	660	11
Ave. Annual Max	2.0	4.5	172	7.5	103	97,800	17
	<u>Future Period</u>						
	WL	Hs	Flow	Wind Mag.	Wind Dir.	Press.	MMSLA
	(m)	(m)	(m ³ /s)	(m/s)	(deg.)	(Pa.)	(cm)
Ave. Overall	0.0	2.6	85	4.5	159	98,400	0
Std. Overall	0.8	1.4	89	2.6	107	630	11
Ave. Annual Max	2.0	3.8	111	6.2	164	98,300	15

A further exploration of changing RI was shown through effective RIs (Figure 19). This result found that the reduction in forcing is eventually overcome by SLR, although with the timing being controlled by the size of the RI event. The idea of an “overtake point” is a simplification based on the assumptions within the model, specifically that of stationarity/nonstationarity (see section 3.6.4.3) and projecting the future period RI’s backward in time (e.g. RIs calculated for the future period applied to 2000). Another way of viewing this result is that SLR represents a constant nonstationarity across all RIs. By year the 2050, both 2-year and 100-year RI events increase by 17 cm due to SLR. On the other hand, the change in RI magnitude from forcing is variable across return periods. From forcing, the change in 2-year RI is only 3 cm while for the 100 year RI it is much larger at 32 cm. This means that low RI events are less buffered by a change in forcing than high RI events. The conclusion is the same from this interpretation in that small RI events will be more impacted by SLR than high RI events.

3.6.4 Modeling limitations

3.6.4.1 Excluded Processes

In this study bathymetry/topography was held constant through time. Bathymetry is a first order control on flooding and so an ideal future projection would include morphological evolution of the estuary. This said, morphological projections at climate change scales are extremely uncertain. The combination of high uncertainty and high dependence would leave resulting flood predictions dominated by an uncertain morphology projection with all other signals obscured. This study therefore does not consider morphological evolution in order to specifically highlight how changing forcing impacts extreme WLs.

Both estuaries have significant anthropological modifications ranging from coastal infrastructure to dredged channels. Coos in particular has an engineered coastline along the majority of its southern boundary. A key factor in future extreme events is the interaction between human intervention and the estuarine system. For example, estuary WL characteristics under tidal forcing show high sensitivity to anthropological changes [Gallien et al., 2011; Wang et al., 2017] and modifications to land use have been shown, in certain cases, to be of the same order of importance to WLs as SLR [Bilskie et al., 2014]. Dredging, for the same reason as morphological evolution, can cause drastic changes to estuarine hydrodynamics. However, similarly to morphological predictions, anthropological controls are highly uncertain. This is often handled in climate modeling through multiple scenarios but this was not possible for computational reasons.

3.6.4.2 Climate Model Variability

The results shown in Figure 18 provide a unique validation tool via a comparison of modeled and observed extremes. However, it is pertinent that the modeled results and observed results are not based on the exact same forcing time series, but rather two different iterations of the historic climate. There is inherent variability in the climate so there is correspondingly variability in RIs depending on the random climate iteration experienced. Since both the Coos and Tillamook RI curves have a straighter shape, this could be a result of the specific climate that

was used for the simulation. This can be generalized to other conclusions made in this paper in that there is an unquantified uncertainty as a function of the specific climate model used.

The common solution to this problem is the usage of ensembles rather than a single AOGCM [Murphrey et al., 2004]. Unfortunately running ensembles is computationally expensive given the level of complexity included in this study. Conceptually this study make the compromise of including more physical complexity at the cost of uncertainty quantification. Results from this project should therefore not be viewed as a probabilistic or a “most-likely” result from climate change. Instead, they should be thought of as a single possibility of what could happen.

Not including multiple model iterations is also problematic in constraining the current climate’s RIs. Results in section 3.6.1 show the importance of compound events and forcing timing (especially tides) at these study sites. While forcing was bias corrected using a methodology that has been shown to perform well for extreme quantiles [Parker and Hill, 2017], timing of forcing occurring on a high tide or during a high MMSLA is additionally critical. This study examines only one possible combination of forcing timings that may or may not be representative of the overall extreme behavior of the system. This could once again be addressed through usage of ensembles or multiple iterations of the current climate.

3.6.4.3 Assumptions of Stationarity

Another important assumption in this study is that of stationarity when SLR is removed. The reality is somewhat more complicated as climate change will result in a forcing driven nonstationarity in addition to that seen from SLR. For our case, each time series segment is statistically stationary (Augmented Dickey–Fuller test, p-value < 0.001) but the overall time series (from 2000 to 2070) must be nonstationary. This is because the two segments (historic and future) show distinctly different calculated RI curves (Figure 18) so WL behavior, as controlled by forcing, must be changing. Simply put, this study analyzes two segments that are not long enough to reveal the longer term nonstationary behavior of the overall time series as controlled by changing forcing. This is not problematic for this study, which compares two snapshots, but a comprehensive analysis would need to resolve the overall nonstationarity from forcing as well as

from SLR. This is a factor in the calculation of the overlap timing from effective RIs (Figure 19) since the overall nonstationarity from forcing is not included in the analysis.

3.7 Conclusions

This paper introduced a process-based modeling framework for analyzing climate change impacts to various controls on estuarine flooding. In particular this study focused on extremes and changes to RI events at two PNW estuaries. This study described the difficulty of using RIs in the context of nonstationarity and showed how “effective RIs” can be an intuitive way of understanding changing flood hazards. Effective RIs showed that predicted changes to forcing results in a decrease in extreme event magnitude moving into the future. This buffers the full effect of SLR on extreme events. This buffering effect was shown to be smaller for low RI events than high RI events suggesting that increasing extremes will be felt first for low RI events.

This study used multiple study sites and climatological periods to explore drivers of extreme events. It was found that extreme events for both locations were not controlled by a single forcing but rather by compound events. Tides were shown to be a key control on extremes and that the concurrence of strong forcing and high tides is required for extreme WLs. It is unlikely that maximum annual forcing coincides with a high tide suggesting that the common procedure of adding an uncoupled high tide and high storm surge will result in a significant over-prediction of flooding magnitude.

An additional outcome of this study was the demonstration that extreme WLs are spatially variable in an estuary. The results showed that WLs varied by more than 25 cm across each estuary domain. Relying only on predictions at the tide gauge and the assumption of a horizontal water surface will therefore mischaracterize flood risk. Since this study found that WL gradients for high RI events increased in the future, errors associated with ‘bathtub’ approximations of flooding surfaces will similarly increase. Overall this study highlighted the importance of limiting conclusions drawn from point analysis to regions spatially near observations or to rigorously define uncertainty from not sampling the full spatial variability of flood surfaces.

Acknowledgments

The utilized NARCCAP climate data are open access and available through the Earth System Grid climate data gateway. The NARR climate dataset is available through NOAA's Earth System Research Laboratory website. Tide gauge records are available through the National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS) website. River discharge is available from the USGS through the National Water Information System. Wave buoy information was obtained through NOAA's National Data Buoy Center website.

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Resources were provided by the XSEDE STAMPEDE cluster at the Texas Advanced Computing Center through a series of allocations. We thank the XSEDE program for providing computational resources that made this project possible.

This paper was funded in part by Oregon Sea Grant under award (grant) number NA14OAR4170064 (CFDA No. 11.417) (project number R/CNH-25) from the National Oceanic and Atmospheric Administration's National Sea Grant College Program, U.S. Department of Commerce, and by appropriations made by the Oregon State Legislature. Additional funding was from the Oregon Sea Grant Robert E. Malouf marine studies scholarship (Project number E/INT-143). The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of these funders.

Chapter 4 - Emulation as an Approach for Rapid Estuarine Modeling

Kai Parker, Peter Ruggiero, Katherine A. Serafin, David Hill

Abstract:

Many estuarine hydrodynamic modeling questions are currently limited by prohibitively long simulation times. This problem is addressed here through the development of an emulator, or surrogate model, which replaces the simulator (in this case the coupled ADCIRC-SWAN) with a statistical representation that is able to rapidly predict estuarine variables such as water levels. Emulation of estuarine water levels is explored in Grays Harbor, Washington USA using Gaussian process regression. After a variety of model simplifications, the effectiveness of the methodology is validated at various simplification levels to determine where error is being sourced. Emulated water levels are found to be skillful when compared to over a decade of tide gage observations (root mean square error <15 cm). The largest loss of skill in the method originates with ADCIRC-SWAN's inability to perfectly reproduce observations, even when the majority of relevant physics are included. Subsequent simplifications to the simulator (input reduction techniques) and the emulator itself are found to introduce a trivial amount of error (increase in RMSE of 1 cm). As an example application, the emulator is employed to decompose the relative forcing contributions to extreme water levels across the estuary. Results show that all forcing contribute to extremes but with streamflow having the least influence and the wave contribution varying significantly across the estuary. The approach is shown to be effective in reproducing other variables in addition to water levels such as non-tidal residuals and significant wave height suggesting that emulation may play a key role in improving our ability to probabilistically assess flood risk in complex hydrodynamic environments.

4.1 Introduction

Modeling estuarine hydrodynamics remains both a challenge and a goal for the scientific community. Estuaries and bays are often densely populated with significant economic and cultural investment [Pendleton, 2010]. They are also subject to a unique flood hazard environment, with high water levels (WLs) driven by numerous contributing processes including both offshore and local waves, storm surge, and river inflows, among others. Over the past several decades, research efforts have led to improved computational models and increased physical understanding of estuarine flood dynamics [Bode and Hardy, 1997; Ganju et al., 2014; Kantha and Clayson, 2000]. However, increasing hydrodynamic model predictive skill has often

been at the cost of simplicity. Complex physics has gradually replaced empiricism, leaving models more accurate and yet often more difficult to implement. This has led to computational time, rather than a physical understanding of the problem, being a primary control on the ability to answer questions about estuarine flooding.

Steadily increasing computer processing power and code parallelization has pushed the boundary for what can be explored with complex computer codes. Even with these advances, many questions still cannot be comprehensively addressed due to computational limitations. One example is the recent focus by the scientific community on uncertainty in model results [Green et al., 2011; Mastrandrea et al., 2010]. In the field of flood hazards, a major thrust area has been probabilistic assessments, which brings the benefits of uncertainty quantification, utility as a stakeholder-centered decision making tool, better handling of extreme events, and more skillful flooding estimates [Croke and Pappenberger, 2009; Dale et al., 2014, Di Baldassarre et al., 2010]. However, the combination of multiple model iterations required for probabilistic modeling and large per-run computational costs has provided a barrier for moving forward.

Often the solution is a compromise, such as simplifying or eliminating various forcing components [Lin et al., 2010; Purvis et al., 2008]; using smaller ensemble sizes [Davis et al., 2010]; or simplifying model physics [Dawson et al., 2007; Moel et al., 2012]. A promising recent development has been to implement variable model complexity, with a fast coarse model being run to determine relevant or extreme events and then a more highly-resolved, accurate model being used to simulate the extremes [Lin et al., 2010, 2012; Orton et al., 2016]. This technique has been demonstrated to be successful for hurricane-induced flooding but is problematic for other regions. For example, environments not dominated by tropical cyclones often are defined by compound events where combinations of non-extreme forcings can combine to create extremes [Leonard et al., 2014; Moftakhi et al., 2017; Wahl et al., 2015]. In addition, event based techniques can still be considered computationally limited as the full parameter space cannot usually be explored. There remains a need for a modeling technique that can bridge the gap between complex, time-intensive models and fast simulation times.

This paper investigates emulation as a technique for the efficient prediction of estuarine hydrodynamic variables in Grays Harbor, Washington USA. The foundational idea of emulation

(also referred to as surrogate modeling, response surface modeling, meta-modeling among others) is the replacement of a slower processes-based model (a simulator) with a fast statistical model (an emulator) [O'Hagan, 2006; Razavi et al., 2012]. In the standard modeling paradigm, the map between simulator inputs and outputs is based on the laws of physics as implemented within a process-based model [Castelletti et al., 2012]. In emulation, this map is approximated using a statistical model. Following an upfront computational expense to create a training dataset, the emulator is nearly instantaneous. Thus, emulation represents a tradeoff between short simulation times and errors from the approximation. This tradeoff suggests that emulation may be ideal for probabilistic flood modeling along with other potential applications such as model uncertainty, model optimization, sensitivity analysis, real time prediction, and extreme event analysis [Kennedy et al., 2006; Levy and Steinberg, 2010; Oakley, 1999].

The general concept of emulation originated in the 1980s through the idea of computer experiments [Sacks et al., 1989]. Since then, emulation ideas have spread widely, resulting in a rich literature of applications, emulator formulations, and theories from numerous fields. Razavi et al. [2012] reviews emulation just in the field of water resources, with over 30 studies revealing a wide range of applications and emulator formulations. In terms of coastal problems, Malde et al. [2016] and Rueda et al. [2016] successfully implemented emulators for wave prediction problems using SWAN [Booij et al., 1997] as a simulator. Timmermans [2015] used emulation to explore how tuning parameters effect uncertainty in results from the Wave Watch III [Tolman, 2009] wave model. Liu and Guillas [2016] investigated the effect of uncertainty in bathymetry on Tsunami height predictions using a novel merging of Gaussian process emulation with dimensional reduction techniques.

In the context of flooding, emulation has been applied to river channel flooding [Apel et al., 2008] and coastal dyke systems [Moel et al., 2012], although from the relatively simplistic perspective of lookup tables. Surge response functions [SRF; Resio et al., 2009; Song et al., 2012] can be considered a specific case of emulation through regression of dimensionless cyclone scaling terms. However, SRFs are limited in application to tropical cyclones, and have been shown to perform poorly in complex environments [Taylor et al., 2015]. As an alternative to SRFs, Kim et al. [2014] used an artificial neural network to emulate coupled ADCIRC-STWAVE calculated surge from tropical cyclones. This approach was added to by Bass and

Bedient [2017] who used a similar strategy but with the addition of a coupled hydrologic model and kriging as the emulator formulation.

Overall, multiple studies have demonstrated the potential of emulation in a coastal hazards setting. Surge from tropical cyclones has, in particular, seen a variety of successful implementations of emulators. This study builds on these previous studies but explores a location in the Pacific Northwest (PNW) outside of tropical cyclone forcing. This results in a unique challenge in terms of handling diverse forcings and a larger input parameter space. For example, the similarly targeted studies (in terms of looking at WLs) by Kim et al. [2014] and Bass and Bedient [2017] were able to reduce input dimensionality through considering only cyclones and using cyclone characteristics as input dimensionality. This study considers a general application of emulating the coupled ADCIRC-SWAN ([ADCSWAN; Dietrich et al., 2011] simulator in which any combination of forcings can be used to calculate WLs. This paper is intended as a rigorous investigation into the applicability of emulation in this new context. Therefore, this paper is focused primarily on describing the methodology and validation although an application to decomposing extreme water levels is also presented.

4.2 Methods

A conceptual overview of the process used for constructing an emulator, in the context of this study, is provided in Figure 24. The following sections will describe the study site (section 4.2.1) as well as the observations that will be used as forcing and for evaluating emulator performance (section 4.2.2). The specifics of each step in developing and testing the emulator, as well as a discussion of the various options for emulator development are then presented.

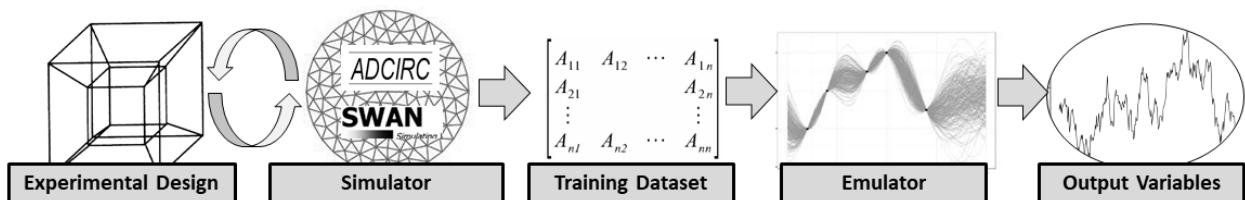


Figure 24: Conceptual framework for developing an estuarine hydrodynamic emulator.

4.2.1 Study Site

Grays Harbor, WA (Figure 25) is an excellent candidate for testing emulation as it exhibits many of the complexities that make estuarine modeling difficult. Grays Harbor is predominantly shallow, dominated by depths averaging less than 5 meters, but also contains a maintained (United States Army Corps of Engineers; USACE) deep-water navigation channel giving it significant depth variability (Figure 25). The bay exhibits spatial variability in WLs [Cialone et al., 2001] as a result of its size [approximately 235.3 km², Engle et al., 2007], shape, and gradients in forcing. Grays Harbor is located in the U.S. Pacific Northwest (PNW, Figure 25) and is subject to an energetic storm and wave climate. The Global Ocean Wave 2 (GOW2) reanalysis [Perez, et al., 2017] node nearest to the study site (see Figure 25) shows a mean offshore significant wave height (H_s) of 2.5 meters with events exceeding 7.5 meters annually. Extreme storm events are generally associated with extratropical cyclones that can produce strong winds, pressure differentials, and precipitation [Allan and Komar, 2002a; Mass and Dotson, 2010]. These events are often associated with significant non-tidal residuals (NTR) [Allan and Komar, 2002a, 2006; Serafin et al., 2018] although of a smaller magnitude than locations impacted by tropical cyclones or with broader continental shelves [Zhang et al., 1999]. Grays Harbor has significant hydrological input from the Chehalis, Humptulips, Hoquiam, Elk and Johns Rivers which collectively drain a watershed of over 7,000 km² for an average monthly river flow of over 22 million m³/month [Engle et al., 2007].

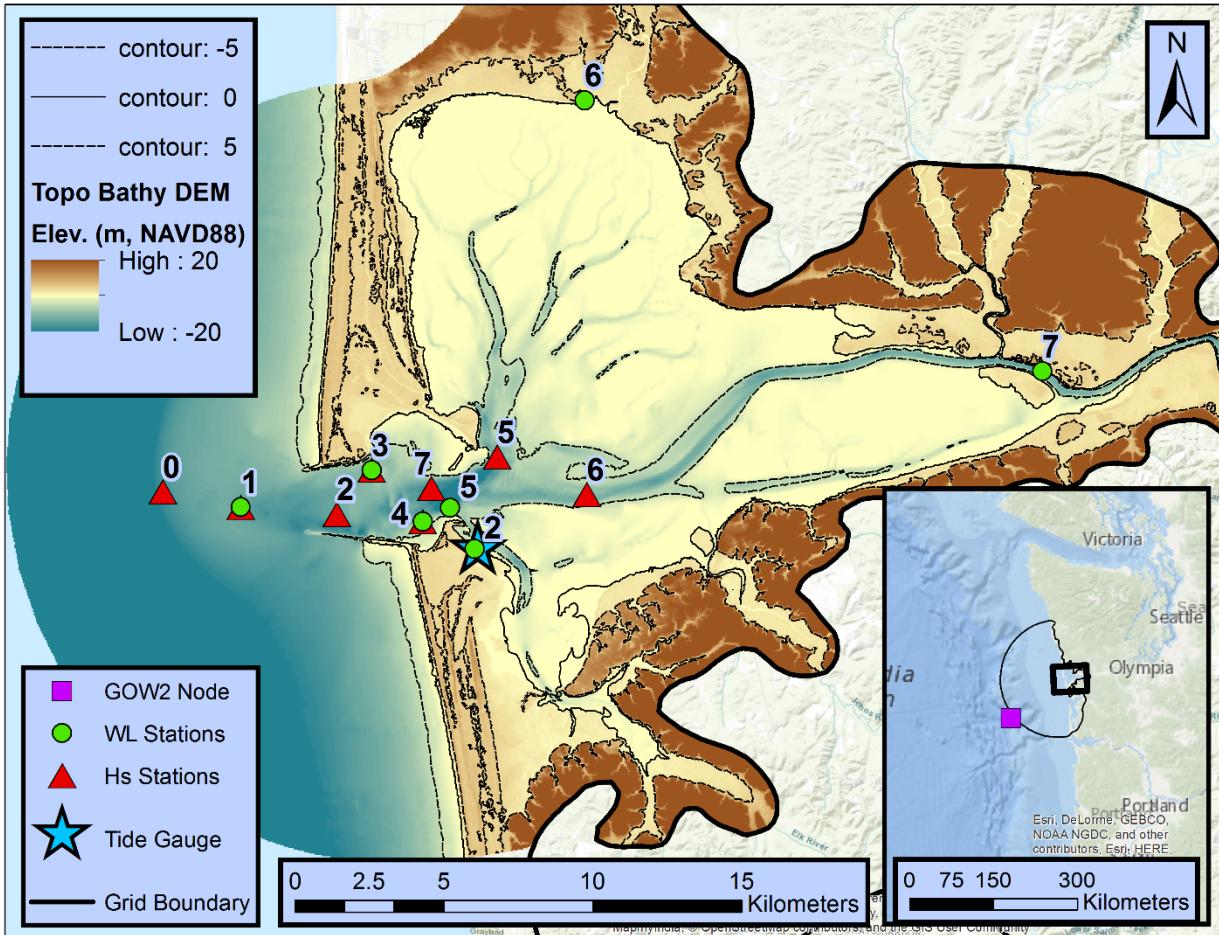


Figure 25: Grays Harbor, WA study site and locations of observational datasets. Circles and triangles represent USACE deployments with co-located instruments labeled with a single number representing both WL and Hs stations. The main panel shows the bathymetry and topography of the local estuary in the NAVD88 Datum. The inset panel shows the larger geographical context of the estuary with the thin black line delineating the domain of the hydrodynamic model.

4.2.2 Observational Data

This study utilizes a variety of observational datasets ranging from instrument deployments to reanalysis products. Forcing and model development datasets are explained in the following section (4.2.2.1) while section 4.2.2.2 details observations specifically used for model validation.

4.2.2.1 Forcing and Model Development Datasets

Wave forcing for the model was obtained from the GOW2 reanalysis of Perez et al. [2017] with output selected from a node located at (Lat: 47°, Lon: -125°; Figure 25). Atmospheric forcing was provided by the North American Regional Reanalysis (NARR) [Mesinger et al., 2006]. NARR provides a wide range of gridded atmospheric variables from which the 3-hourly 10m wind fields and 3-hourly surface pressure fields were utilized. Streamflow was obtained from USGS river gauges with total estuary inflow constructed as the sum of three gauged rivers, the Chehalis, Satsop, and Wynoochee (USGS stations 12031000, 12035000, and 12037400 respectively).

The bathymetry data for the simulator grid were developed by blending two National Oceanic and Atmospheric Administration (NOAA) digital elevation models (DEMs): the Astoria, OR tsunami DEM (1/3 arc second) and the coastal relief model (3 arc seconds) [Love et al., 2012; NOAA National Centers for Environmental Information, 2003]. Bay topography was sourced from Oregon Department of Geology and Mineral Industries (DOGAMI) LiDAR scans [DOGAMI, 2010].

4.2.2.2 Validation Datasets

In addition to forcing, a series of observational datasets were used to validate simulated and emulated variables within the study site. The first dataset is the Westport, WA tide gauge (NOAA station ID # 9441102) which provides continuous hourly WL data from 2006 until the present. Water level observations were decomposed into constituent components (e.g., deterministic tide, monthly mean water level anomalies, storm surge etc.) using the approach described in Serafin and Ruggiero [2014]. The five largest NTR events on record were extracted to test simulator and emulator skill. A brief summary of these storm events is provided in Table 3.

Table 3: Summary of forcing for the five largest NTR events at Grays Harbor, WA. Forcing values are reported at the occurrence of maximum NTR.

	Storm 1	Storm 2	Storm 3	Storm 4	Storm 5
Date	12/03/07	12/15/06	01/12/14	12/11/15	12/12/14
Non-Tidal Residual (m)	0.93	0.73	0.66	1.11	0.58
Significant Wave Height (m)	11.1	8.0	8.5	11.1	6.2
Peak Wave Period (sec)	15.4	12.7	14.3	18.2	12.2
Wave Direction (deg.)	195	243	259	248	229
Surface Pressure (HPa.)	989	977	989	984	986
Wind Speed (m/s)	19.2	17.5	14.9	12.3	9.1
Wind Direction (deg.)	201	217	238	198	218
Streamflow (m ³ /s)	2020	127	859	1510	671

Water level observations at the tide gage were supplemented by a field campaign carried out the U.S. Army Corps of Engineering (USACE) from September-December, 1999 [Figure 25, Cialone et al., 2001; 2002]. This dataset includes seven locations near the inlet with bottom mounted tripods measuring wave characteristics, WLs, tidal currents, and suspended sediment. Additionally, five surface stations were distributed throughout the bay measuring WLs, conductivity and temperature.

The USACE field campaign was broken up into two deployments (with a small maintenance/data collection break between the two deployments). Instruments were replaced in approximately the same location except for Hs station 0 which was moved to location 7 for the second deployment [Cialone et al., 2002]. Figure 25 illustrates the spatial distribution of the various observation stations which have been renamed in this paper for clarity.

4.2.3 Simulator Configuration

This study utilizes the coupled Advanced Circulation [ADCIRC; Luettich and Westerink, 1992] and unstructured Simulating Waves Nearshore [SWAN, Zijlema, 2010] simulator [ADCSWAN; Dietrich et al., 2011]. The coupled ADCSWAN simulator has seen extensive

validation and success in predicting WLs and NTR at various estuaries around the world [Bhaskaran et al., 2013; Dietrich et al., 2012; Krien et al., 2015]. Recently, ADCSWAN has been successfully implemented in the PNW with good agreement between simulator output and observations of WLs, NTR, and currents [Cheng et al., 2015b; Cialone et al., 2002]. ADCSWAN was chosen for this study due to its balance of physical complexity for the variables of interest and computational efficiency. ADCSWAN is implemented in the 2D depth-integrated barotropic mode which has been shown to perform with acceptable error for computing WLs and depth integrated currents in embayed regions [Resio and Westerink, 2008; Weaver and Luettich, 2010]. ADCIRC is run in the fully 2-way coupled implementation with SWAN, which is important for resolving the interacting physics of wave fields and nearshore hydrodynamics [Cialone et. al., 2002; Dietrich et al., 2010, 2011; Funakoshi et al., 2008].

ADCSWAN is run on an unstructured mesh which, in this study, extends beyond the continental shelf (approximately 115 kilometers offshore; Figure 25). Unstructured meshes provide flexibility in simulator resolution with the model grid having element sizes ranging from around 7,000 meters offshore to under 20 meters within the inner Grays Harbor channel.

4.2.4 Dimensional Reduction and Levels of Simplifications

Emulator construction requires sampling the full input parameter space. This constraint dictates the size of the training dataset to be proportional to the number of dimensions included as inputs. In general, process-based hydrodynamic simulators are based on many inputs making some form of dimensional reduction necessary. An important part of emulator construction is

finding a balance between minimizing the number of inputs and maintaining sufficient complexity to acceptably resolve variables of interest.

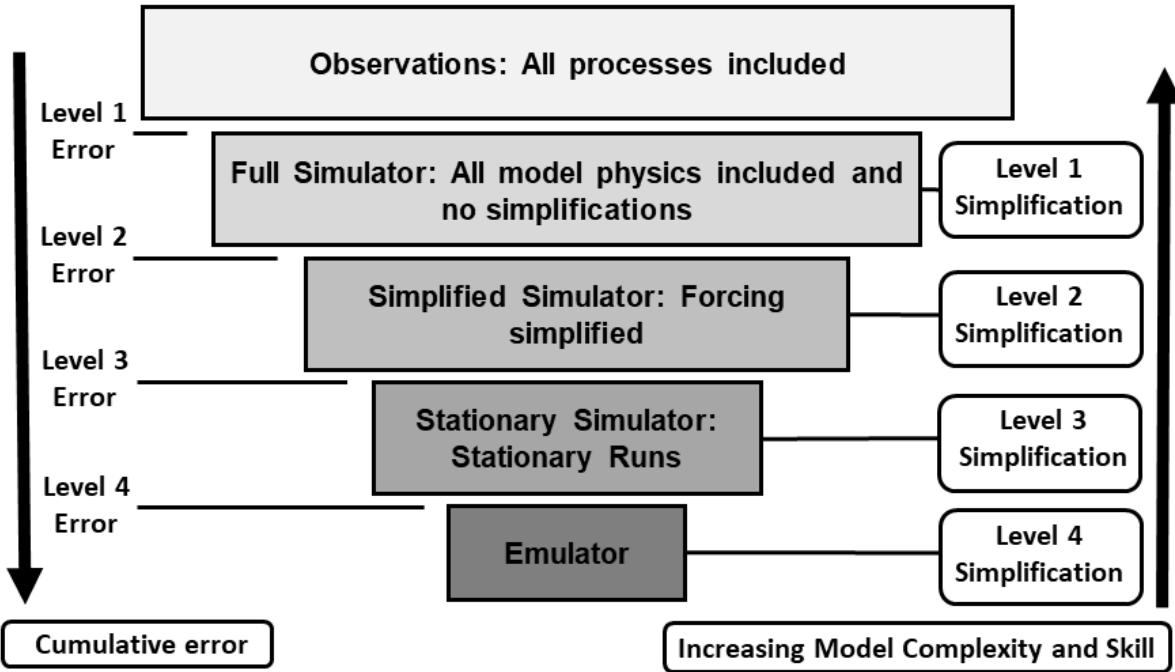


Figure 26: Hierarchy of model simplifications between observations (top) and emulator output (bottom). Each simplification is associated with a level (right side of the figure) and some amount of error (defined on the left side of the figure).

Figure 26 provides a conceptual model of each simplification that went into transforming the full process based simulator (ADCSWAN) into an emulator for this study. Each of these simplifications theoretically introduces some amount of error into the output, defined in Figure 26 as a “level” of error. This paper also refers to simplification levels which correspond to the simplification that induced the error (right side of Figure 26). Finally, this paper will discuss discrete error (error from a single simplification) and cumulative error which refers to the sum of all preceding discrete errors. For example, a comparison of model output from the level 3 simplification (stationary simulator) to observations (no simplification) quantifies the cumulative level 3 error. The following methods sections (4.2.4.1 – 4.2.4.3) explain each simplification in this hierarchy while corresponding error is later quantified in the Results section (section 4.3).

4.2.4.1 Simulator Simplifications

The first level of simplification is simply that of using a process-based simulator. Simulators are unable to exactly reproduce observations for a variety of reasons ranging from incorrect or unresolved physics (e.g., assumptions, parameterizations, etc.) to numerical approximations (truncation errors, etc.). The error induced by this simplification is primarily a function of the chosen model, the implemented physics, and model tuning. Therefore the specific configuration and choice of ADCSWAN (section 4.2.3) is the primary control on the strength of this simplification.

This study considers emulation of a specific implementation of the ADCSWAN model and therefore the model grid (bathymetry, resolution, etc.) is held constant. Additionally, simulator configuration parameters are not included in the emulation framework. ADCSWAN contains a large number of input switches, tuning parameters, forcing options, numerical configurations, and other choices [Westerink et al., 1992]. This study holds all general model configuration parameters constant leaving the various forcing components of water level variability as the sole input dimensionality of the emulator.

4.2.4.2 Forcing Simplifications

Even with the simplification of holding the model configuration fixed, the input dimensionality remains high, due to the numerous physical forcing mechanisms. Here the simulator is simplified to 16 forcing dimensions in order to investigate how well emulation can work at a more basic level. This approach is desirable as less input dimensionality corresponds to a smaller training dataset and therefore a more efficient emulator construction.

Wave simplification

It is well known that offshore wave energy can impact water levels with bays such as Grays Harbor [Cheng et al., 2015b, Olabarrieta et al., 2011]. This forcing is implemented in the simulator using a JONSWAP spectrum fitted to a peak wave period (T_p), H_s , mean wave direction, and directional spread parameters. While research has shown the importance of forcing with full directional spectra for reproducing wave observations [Montoya et al. 2013; Rogers et

al., 2007], most studies accounting for wave influence on WLs use simpler bulk parameter based formulations. Therefore a fitted JONSWAP spectrum is used for both the full (level 1) and simplified (level 2) simulator comparisons. Based on previous research in the PNW [Cheng et al., 2015a], directional spread is held constant at 20 degrees, and wave forcing is applied uniformly along the full simulator open boundary. With these simplifications, wave forcing is included in the emulator as three dimensions: Hs, Tp, and mean wave direction.

Atmospheric simplification

Atmospheric forcing represents a unique challenge for emulation due to spatially variability. Gridded inputs represent a high degree of dimensionality, with every node potentially representing an input dimension. For this reason, a sensitivity study was undertaken to see if spatially constant atmospheric forcing could be used as an approximation of the full forcing fields. WL output from simulator runs with full gridded forcing were compared to runs with spatially constant forcing. Results indicated that the error introduced in predicted WLs by the spatially constant assumption was acceptable in comparison to the corresponding reduction in dimensionality. This error is quantified in the results section (along with other simulator simplifications) as level 2 error. Adopting the spatially constant assumption, atmospheric forcing is reduced in the emulator framework to three dimensions: wind speed, wind direction, and sea level atmospheric pressure.

Tidal Simplification

Tidal forcing in hydrodynamic models is generally represented through harmonic constituents. NOAA tidal predictions at gauge locations (which are based on harmonic analysis) typically contain around 40 individual constituents, each of which is specified by an amplitude and phase. From an emulation perspective, including all constituents would be computationally prohibitive so some form of simplification is desirable. Many studies using ADCIRC are forced with eight or fewer constituents, mainly because global databases of tidal constituents (e.g., TPXO [Dushaw et al., 1997], or LeProvost [Le Provost et al., 1994]) are typically limited to that number. Despite this, these simulations are found to agree well with both harmonic analysis derived and observed tidal elevations [Blain et al., 1998; Blain et al., 2001; Westerink et al.,

1992]. ADCIRC simulates tidal forcing as a boundary elevation time series [Luettich et al., 1992] determined by a spatially variable, temporally constant phase and amplitude and a temporally variable, spatially constant equilibrium argument and nodal factor. Amplitudes and phases are determined by the simulator boundary location, and are therefore not an emulator input dimension when considering a fixed study site. The nodal factor represents adjustments of the amplitude/phase of each constituent that results from the nodal tide cycle. The equilibrium argument (deterministic based on date and time) controls the timing of the harmonic.

While tides are deterministic, they are included within the emulator as forcing for a variety of reasons which will be described in section 4.2.3.3. In approaching simplifications, a sensitivity test was performed to determine the tidal dimensionality required for accurately reproducing maximum WLs during storm events (Table 3). It was found that removing the nodal factor did not significantly change simulated WLs. After this simplification, results showed that eight harmonics (without nodal factors) were sufficient for accurately producing WLs. This allows tides to be included in the emulator as eight input dimensions: 8 harmonic equilibrium arguments, each ranging from 0 to 360 degrees.

Streamflow Simplification

Streamflow is represented in ADCIRC as a flux of water into the domain (specified as a normal flow per unit width of boundary). This allows the simulation of large rivers that have significant cross-channel velocity profiles and for calibration where data on these cross-channel profiles are available. For this study, we instead specify a laterally constant velocity profile across the Chehalis River boundary. This simplification is common [McKay and Blain, 2010; Bunya et al., 2010], especially if the boundary is far enough away from the area of interest that a natural flow profile can develop. This allows streamflow to be represented as a single input dimension (the total volumetric flow rate) for each river inlet.

Grays Harbor has multiple river inlets but the majority of the input flow is concentrated at the Chehalis River which captures around 80% of the watershed area. For simplicity, only the Chehalis input is included in the simulator as a single input dimension. All other stream inputs are assumed to be minimal with only local influences on WLs.

Base Water Level Simplifications

A final (16th) input dimension is considered within the emulator framework as a “Base WL” parameter. This is included to account for large scale changes to estuary sea level, as is experienced through monthly mean sea level anomalies (MMSLA), sea level rise (SLR), seasonality, etc. [Serafin et al., 2018]. These are defined in the simulator as a static change to mean sea level and are therefore included in the emulator as a single input dimension. Shorter time scale variations in WLs are computed by the simulator as controlled by forcing.

4.2.3.3 Simulator Stationary Simplification

ADCSWAN and other process-based hydrodynamic simulators are dynamic in that both inputs and outputs are functions of time and the instantaneous simulator state is determined, in part, by the previous state. Seeking simplicity, this study makes the assumption that the dynamic system can be approximated using a series of stationary simulations. Precedents for such an assumption exist in less complex systems, including spectral evolution in wave modeling (SWAN) approximated using a series of steady-state simulations [Rogers et al., 2007; Rusu et al., 2008].

Simplifying tidal forcing with stationary simulations is difficult since there is no tidal equilibrium in WLs. One approach is to consider tides as a series of horizontal water surfaces of different elevations (corresponding to tidal phases). This would reduce tidal forcing dimensionality to a single value (tidal WL), but at the cost of losing spatial variability. Testing showed that, for the Grays Harbor study site, tidal wave evolution and propagation across the estuary results in significant spatial variability in tidally forced WLs. A second approach is to decouple NTR and tidal WLs and add the two as a linear summation. However, testing showed that this simplification results in significant error. Therefore, a hybrid solution was developed in which all non-tidal forcing is stationary but tides are computed dynamically with model output recorded only at the specific moment of interest. This approach is appropriate since tides are deterministic and, for a specific set of equilibrium arguments, the previous state of tide induced WLs will always be the same. This approach allows tidal forcing to be simplified but retains the spatial variability in tidal WLs and the nonlinear interactions with other processes.

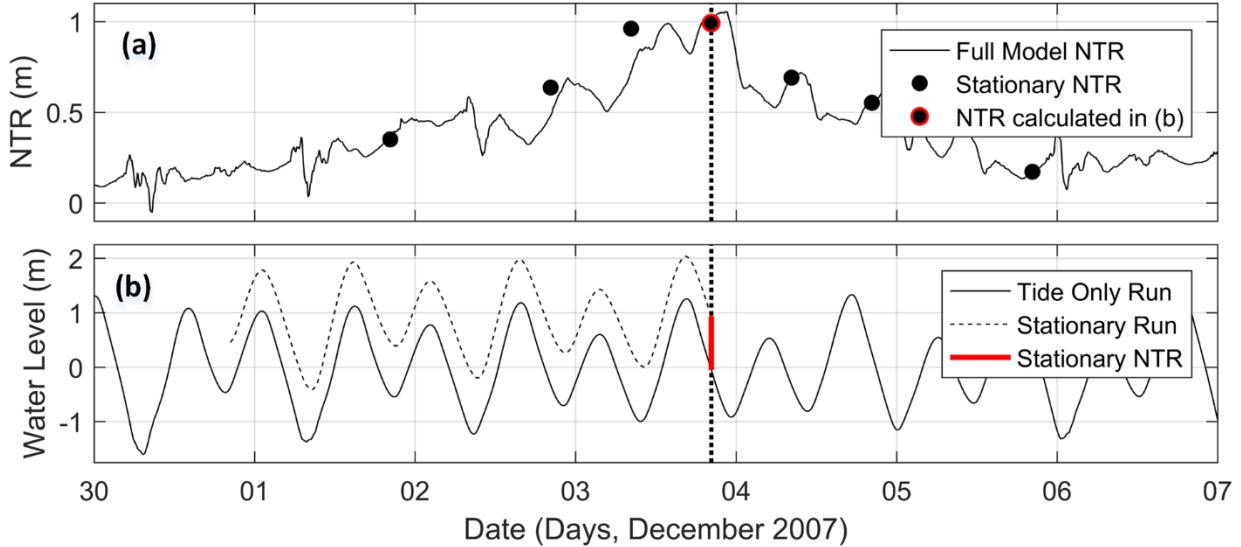


Figure 27: Panel (a): Comparison of NTR during Storm 1 from a fully dynamic simulation (black line) and simplified stationary simulations (black dots). Panel (b): Example stationary run (at the peak of the storm) showing how the stationary NTR is calculated. The horizontal bold dotted line represents the time of the stationary run. At this time, NTR is calculated by subtracting the value of a tide only run from the value of the stationary run (Bold red line). This NTR value is plotted as a red outlined dot in panel (a).

Figure 27 illustrates how stationary runs compare to the full dynamic simulator. Figure 27a compares NTR from the fully forced ADCSWAN (simplification level 1; black line) and 7 stationary ADCSWAN runs (simplification level 3; black dots) during Storm 1. NTRs are computed for both cases by subtracting a ‘tides only’ simulation from the fully forced model. Figure 27b demonstrates how the stationary NTR is computed for the peak of Storm 1. This NTR value is plotted in Figure 27a as a red outlined dot. The agreement between the fully dynamic run and the stationary runs is overall good, with an RMSE error for Storm 1 of 11 cm.

4.2.5 Experimental Design

Following the framework outlined in Figure 24, the first step in building an emulator is the selection of design points (experimental design) to create the training dataset. This study implements a design from the commonly utilized Latin Hypercube sampling (LHS) family of schemes first explored by McKay et al. [1979]. LHS is one of the older and more popular experimental designs and has been found to perform well for complex simulators [Jones and Johnson, 2009]. The specific experimental design for this study was created using a “maximin”

LHS design [Johnson, 1990; Morris and Mitchell, 1992] from the LHS package in R [Carnell, 2017].

Parameters required for a LHS design are the number of dimensions to be included, the range of each dimension, and the number of design points. As detailed in section 4.2.4, this study used an input parameter dimensionality of 16, including wind speed and direction, sea surface pressure, Hs, Tp, and wave direction, river flow, base WL, and eight tidal equilibrium arguments. LHS considers only the maximum and minimum values of each dimension with design points spaced approximately uniformly across dimensions. Ranges were chosen for each parameter in an attempt to span all plausible forcing scenarios. This was determined by looking at 100 year return level events as calculated from the observational records. The size of the training dataset is typically controlled by the cost of running the simulator but Loepky et al. [2009] provide the general guidance that the training dataset should be approximately 10 times the number of dimensions of the input space. Given the 16 input dimensions of this study, this suggests a theoretical training dataset size of 160 runs. To explore the relationship between training dataset size and emulator skill and to validate the emulator's overall effectiveness, this study developed a larger training dataset consisting of 480 ADCSWAN runs.

4.2.6 Emulator Configuration

A variety of formulations have previously been used in an emulation context, including support vector machines, artificial neural nets, and many others [Jin et al., 2001; Gano et al., 2006; Razavi et al., 2012]. This study uses Gaussian process regression (GPR, also referred to as Kriging), a Bayesian statistical nonlinear regression model well suited to this particular application as it scales well to high-dimensional input and intrinsically considers model uncertainty [Levy and Steinberg, 2010; O'Hagan, 2006]. Furthermore, GPR is a general and flexible framework that can be optimized for a variety of modeling problems [Rasmussen and Williams, 2006]. For example, many other common emulator formulations, such as neural nets [Rasmussen and Williams, 2006] and radial basis functions [Anjuo and Lewis, 2011], can be shown to be a form of GPR under certain conditions.

The foundational definition of a Gaussian process is that of an infinite collection of variables for which any finite subset is described by a multivariate Gaussian distribution. Every

point in the input space can be modeled as a random variable (due to uncertainty about the functional response to uncertain input) and the Gaussian process governs how these variables are related. A common way of thinking about GPR is as a distribution over functions [Rasmussen and Williams, 2006]. This is mathematically tractable as a GPR can be completely defined by a mean and covariance function (due to being modeled as a multivariate Gaussian distribution). From a Bayesian perspective, this means a GPR is specified as a prior mean and covariance function. The data then updates this prior with information about the true form of the function to develop the posterior.

This process is conceptualized for a one-dimensional case in Figure 28. The effect of the Bayesian conditioning on the emulator can be seen as an “anchoring” of the posterior sample functions (and uncertainty) at locations of observations. This limits the possible functions to those that go through these observed points. Uncertainty is quantified by the possible functions that pass through these training points.

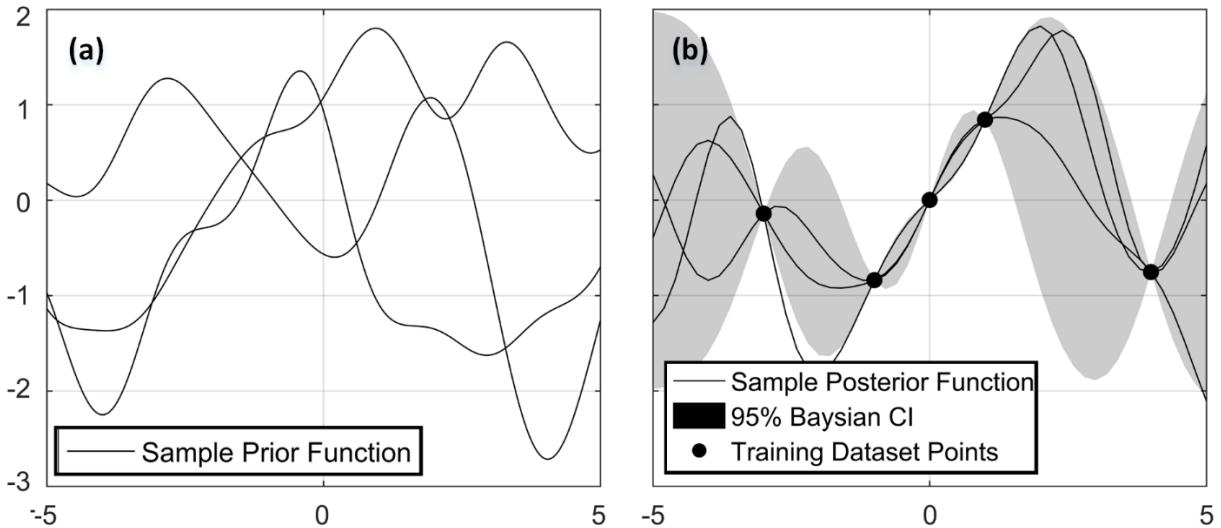


Figure 28: Panel (a) shows three sample functions drawn from the prior distribution. Panel (b) shows 3 sample functions from the posterior distribution after 4 training observations (dark black points). In panel (b), the shaded region represents the 95% Bayesian credible interval. Figure after Rasmussen and Williams [2006].

The first component of a Gaussian process is the mean function. A common choice is to set the prior mean to zero and this is demonstrated in Figure 28 (as shown by the approximate mean of the sample prior functions being zero). As an alternative, this study follows the

methodology of Timmermans, [2015] who used a simple linear regression to obtain information about the mean function's form. Residual analysis of our data showed a cubic relationship for the tidal equilibrium argument terms, a somewhat expected result due to the cyclic nature of tides. Based on this result, the prior mean function was defined with a cubic term for all tidal equilibrium argument inputs and a linear term for all other inputs. A comparison of emulators with and without the modified mean function showed a significant gain in skill by using the modified mean function.

The covariance function of a Gaussian process is the second necessary component for defining the emulator. The covariance function can be thought of as describing the relationship between points in the process. Practically this describes the smoothness of the resulting GPR. A comparison of 4 commonly used covariance functions, the Gaussian, squared exponential, Matern ($\nu = 1.5$) and Matern ($\nu = 2.5$) [Rasmussen and Williams, 2006], found that the Matern ($\nu = 2.5$) was the best performing and this was utilized for all results found in the following section. The training of the emulator was performed using the Managing Uncertainty in Complex Models (MUCM) package in R [Malde et al., 2016].

4.3 Results

4.3.1 Error Introduced by Model Simplifications

With the construction of the emulator being hierarchical (Figure 26), it is important to assess the skill of the emulator at multiple simplification levels to determine where errors are being introduced. This was investigated for simulator simplifications by looking at 5 of the largest storms on record (detailed in section 4.2.2). Storm events were chosen for this analysis as it is expected that the strong forcing gradients and rapidly changing dynamics of these events would provide the most robust test of simulator simplifications. For each storm, WLs were calculated using the full simulator (dynamic, non-simplified) and the stationary simulator (all simplifications except for emulation) to quantify the sum of level 2 and level 3 errors. This comparison was performed at six or seven temporally random points spread out across the duration of each of the storms. Figure 29 shows the difference between the simulators (level 1

minus level 3) at two locations: the tide gauge (Figure 29a) and WL station 7 (Figure 29b). This difference is denoted here as the “error” resulting from simplifying the simulator.

The error computed via this test was found to have a max of 25 cm with a RMSE of 6 cm at the gauge location. The maximum RMSE for all stations (across the estuary domain) was found to be 9 cm. The simplified simulator was found to be only lightly biased with a mean approximately 2 cm lower than the full simulator (represented as a positive mean error in Figure 29). Sensitivity testing for non-storm conditions (not shown) confirmed that these storm results are likely conservative and that the simplified simulator generally performs better for non-storm conditions.

4.3.2 Emulator Validation

The ability of the emulator to replicate the fully simplified simulator (level 4 error) was quantified using a k-fold cross-validation procedure [Arlot and Celisse, 2010; Kohavi, 1995]. K-fold cross-validation breaks the total training dataset into k segments and cycles through every possible combination of withholding one segment for validation and training the emulator with the remaining segments. This results in an ensemble of skill metrics for which the mean is less biased and more robust to training period than a standard validation methodology [Arlot and Celisse, 2010]. Figure 30 shows the results of the cross-validation with 5 segments comparing emulated WLs and simplified simulator WLs.

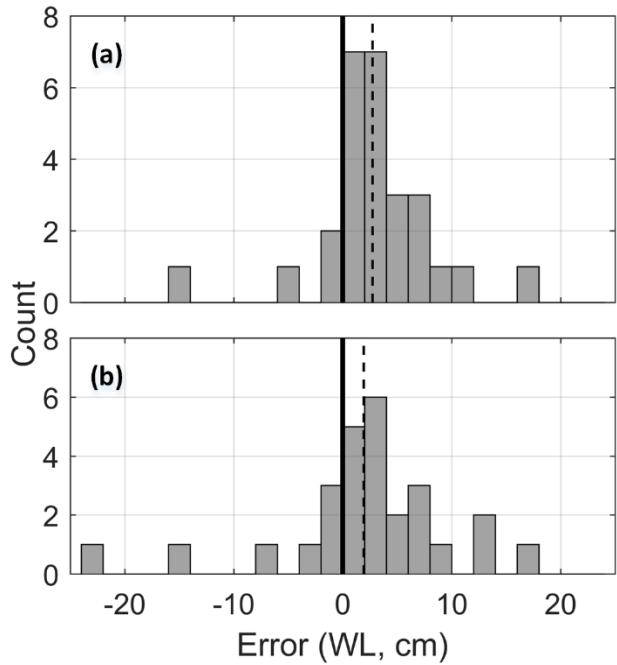


Figure 29: Error between calculated WLs between full and stationary (Level 1 and Level 3) simulator runs. The histogram includes comparisons for Storms 1-5 (Table 3). The bold line represents zero error while the dotted line is the mean error of the ensemble. Subplot (a) is computed at the tide gauge location while subplot (b) is at WL station 7 (see Figure 25).

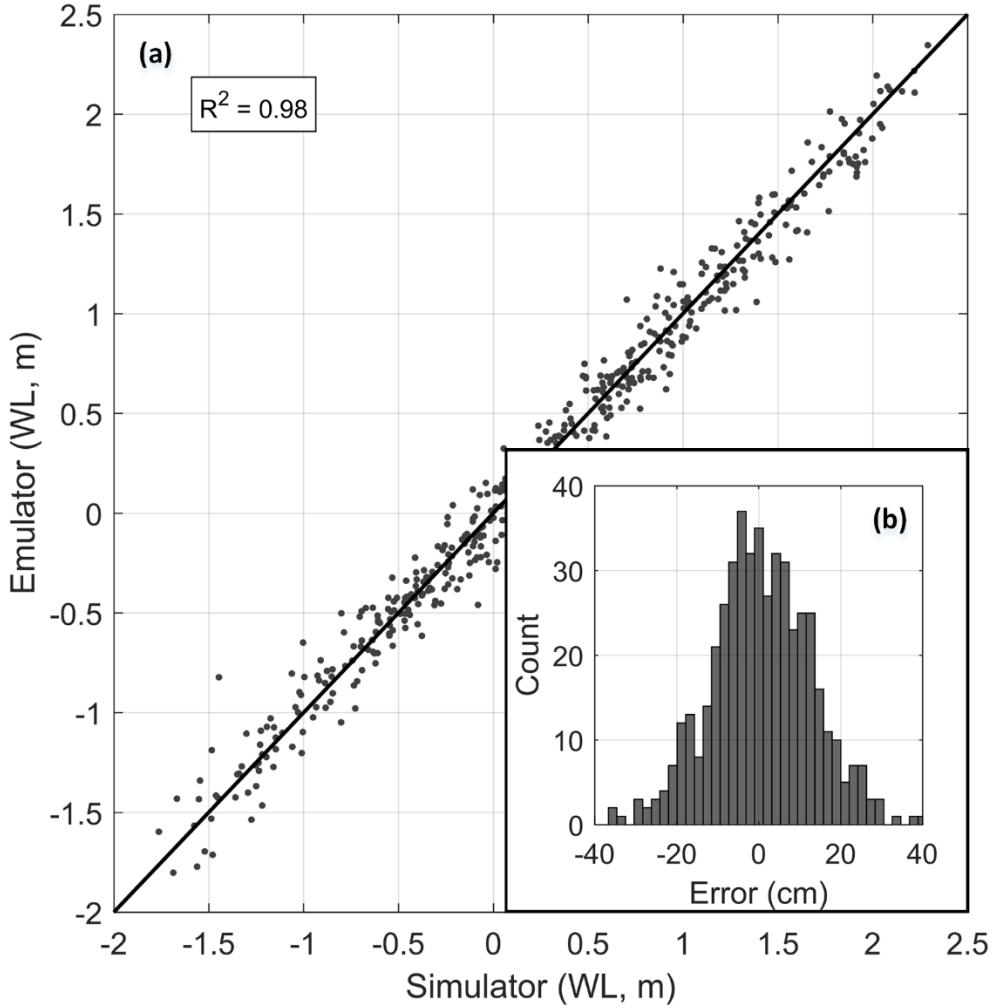


Figure 30: Panel (a): Simulated vs. emulated WLs for the full training dataset. WLs were produced using a 5-fold validation procedure. Panel (b): histogram of the error between simulated and emulated WLs.

Overall the emulator was found to perform well at this comparison level with a high level of skill. The emulator shows no bias (mean of the residuals less than 1 cm), and relatively even variance in residuals across WL magnitude. This said, the width of the histogram in Figure 30b suggests that this step introduces more error than simulator simplifications. The RMSE was found to be around 13 cm which is significantly larger than the calculated simulator simplification error (Figure 29, sum of level 2 and level 3 error) of approximately 6 cm.

4.3.3 Emulator Performance: Westport, WA Tide Gauge

The next stage of quantifying the skill of the emulator is to compare emulated WLs to observations at the tide gauge. This test provides a measure of the cumulative level 4 error, or the total integrated error from predicting observed WLs using emulation. This analysis was performed by using the emulator to predict hourly WLs at the location of the tide gauge for the entire period of record (2006-2016). Comparison between tide gage observations and hourly emulated WLs for a randomly chosen month long segment are shown in Figure 31.

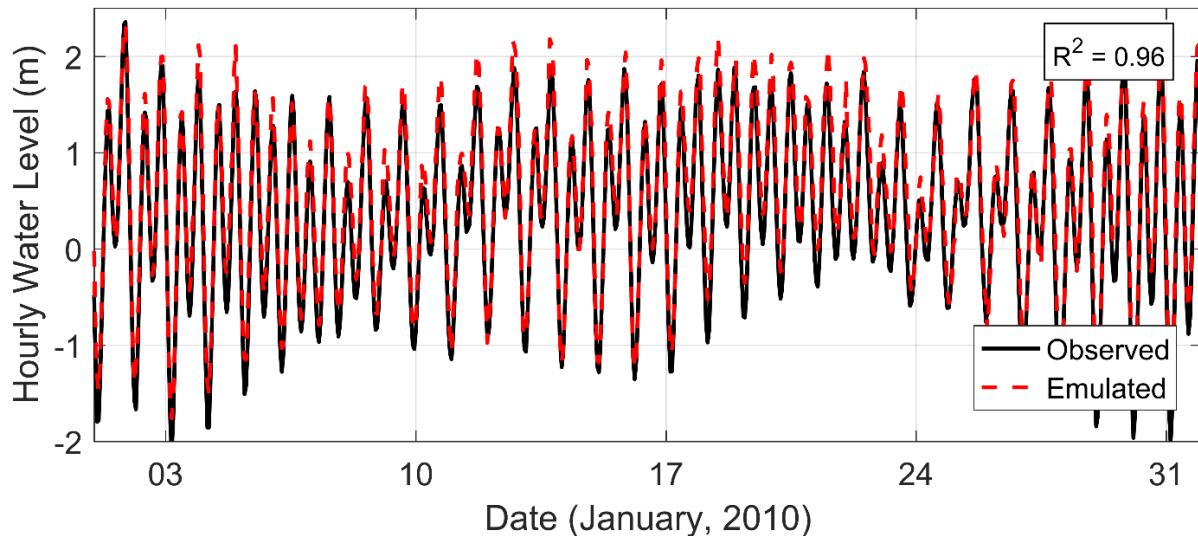


Figure 31: Comparison of emulated (red dashed line) and observed (black line) hourly WLs at the Grays Harbor tide gauge for January 2010. Coefficient of determination is for the entire tidal record (2006 - 2016).

Overall, hourly emulated WLs (for the over 10-year record) compare favorably to the tide gauge with an R^2 value of greater than 0.96, RMSE of approximately 15 cm, and a bias of less than 1 cm.

As with tide gauge records, WL output from an emulator can be considered as the sum between two components: tides and NTRs. In the PNW, tides are the dominant source of WL

variability [Allan and Komar, 2002b] and so the skill of the emulator in predicting WLs is primarily controlled by its ability to reproduce the deterministic tides.

However, coastal hazard research often considers NTR individually as the driver of extreme WLs on top of regular tidal cycles. Therefore, it is additionally important to test the effectiveness of the emulator at reproducing NTR signals. This additionally provides a more robust test of emulator performance as NTR is not explicitly modeled by the emulator as an output.

NTRs are often calculated at tide gauges by subtracting the predicted tide (determined through harmonic analysis) from the measured WL. This procedure can be problematic since NTR can affect tidal phase, resulting in a false NTR signal from out of phase tidal signals [Haigh et al., 2014; Pugh, 1996; Serafin and Ruggiero; 2014]. Therefore, for this study NTRs were calculated from tide gauge data using the procedure of Bromirski et al. [2003]. This method uses spectral filtering to remove energy from tidal bands for a cleaner estimate of NTRs.

For spectral analysis, a windowing function is used making data at the edge of the record unusable. In addition the Bromirski et al.,

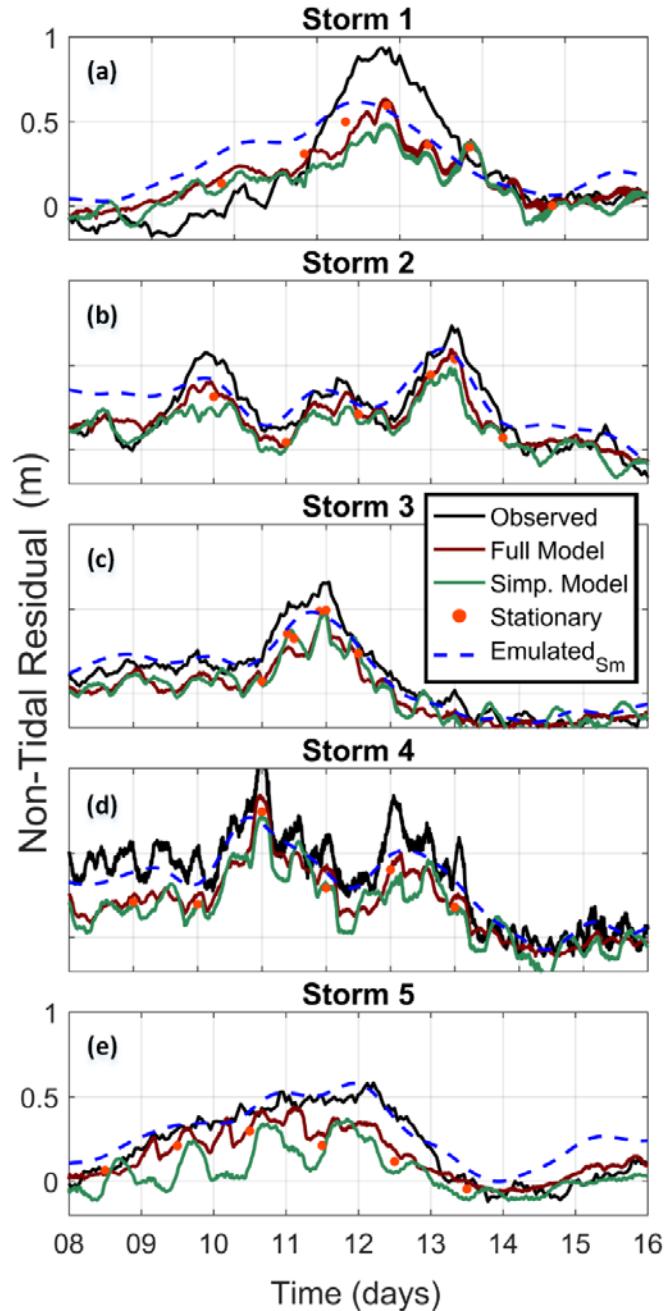


Figure 32: Tide Gauge comparison of observed and modeled NTR at various simplification levels. Each subplot (a through e) is one of the top 5 storms of record (Table 3). Due to windowing for spectral filtering, storm 4 observed NTR is calculated using the subtraction method rather than the Bromirski method. All panels have the same y-axis scaling.

[2003] methodology requires relative large blocks of time (on the order of a year) for properly transforming data into the frequency domain. As storm simulations are on the scale of weeks, NTRs from the ADCSWAN simulations (at all simplification levels) and emulator simulations are calculated as a full forcing run minus a tidal only run. Emulated NTRs are subsequently smoothed with a loess filter to reduce noise associated with tidal phase mismatches between the tide-only and full forcing emulated time series.

Comparison of observed and hindcast NTRs for Storms 1-5 (Figure 32) show a good overall performance of the emulator. To contextualize the comparison, the full, simplified, and stationary simulator calculated NTR time series are all shown. A quantitative comparison of error between observed and modeled NTRs show that all simplification levels (from full ADCSWAN to emulator) have an RMSE of 14 cm plus or minus 1 cm. The similar error across all simplification levels suggest that the largest source of error for NTR is in ADCSWAN itself (level 1). For example, in Figure 32a it is clear that the full ADCSWAN simulation (red line) is unable to reproduce the peak NTR signal in Storm 1. Comparison of bias shows that all simplification levels are biased low by about 15 cm plus or minus 3 cm except for the emulated NTR which has a bias of only 6 cm.

4.3.4 Emulator Performance: USACE field campaign

The tide gauge provides a rich dataset for validating the emulator due to its record length, but is spatially limited to a single comparison point within the study area. One key strength of using an emulator is the ability to provide WL information across study sites – as an emulator can be constructed at any grid node within the ADCSWAN model. Figure 33 summarizes the spatial performance of emulation through a comparison of emulated and observed WL time series for a 1999 field campaign led by the U.S. Army Corps of Engineers in Grays Harbor [Figure 2, Cialone et al., 2001; 2002].

Figure 33 shows good performance between observed and modeled WLs across most locations. The main exception to this is WL station 6 which has comparatively poor agreement between the observed and emulated WLs. This lack of skill is equally shown by the full ADCSWAN simulator and is therefore not a result of the emulation procedure. Table 4 gives RMSE values for a comparison between observations and modeled output at various levels of simplification (level 1 simulations and level 4 emulations are shown in Figure 33). The levels described in the column headers are cumulative error, or a comparison of the model at that simplification level (Figure 26) to observations. No level 3 skill estimate was developed due to the computational constraints of simulating sufficient stationary point runs to accurately quantify model skill.

Table 4 indicates that the largest drop in skill is at the level 1 simulator simplification, the full ADCSWAN model's inability to perfectly reproduce observations. Level 2 simplifications are found to only nominally impact modeled WLs with a small (1 cm) increase in RMSE for some stations. Level 4 simplifications additionally produce very little loss of skill, with some comparisons actually showing a reduction in RMSE from this simplification.

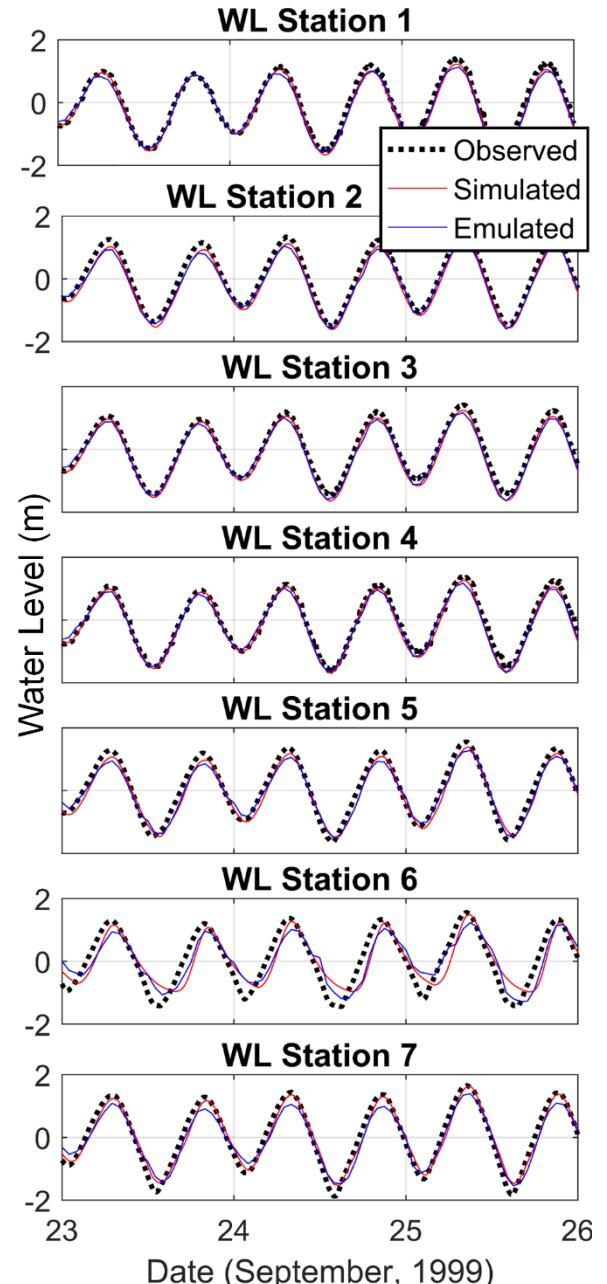


Figure 33: Comparison of observations (USACE deployments, Figure 25), full simulator and emulated WLs. All panels have the same y-axis scaling.

Table 4: RMSE values comparing WL model output to observations at various simplification levels. Rows are locations / variables while the two column groupings represent instrument deployments. The length of the time series comparison varies depending on station deployment length.

Station	Deployment 1: RMSE (cm)			Deployment 2: RMSE (cm)		
	Level 1	Level 2	Level 4	Level 1	Level 2	Level 4
WL 1	24	24	25	21	21	22
WL 2	19	20	22	16	17	19
WL 3	26	26	27	37	37	29
WL 4	19	19	21	22	23	20
WL 5	21	22	22	19	19	20
WL 6	44	44	41	36	36	39
WL 7	21	22	28	23	21	27

4.3.5 Example Emulator Application: Extreme WL decomposition

Outside of validation, it is illustrative to explore an example application of the developed emulator. For this purpose, a decomposition of the relative forcing contributions responsible for extreme WLs within Grays Harbor was performed. For a longer record of extreme WLs, this application uses the time series developed by Serafin et al. [2018] in which the neighboring Toke Point, WA tide gage was merged with the Grays Harbor tide gage to develop a virtually complete 31 year WL time series (1984-2015). To determine the relative contribution of different forcings on extremes, seven versions of the 31 year time series were emulated under different forcing scenarios. As a baseline “full forcing” case, the time series was emulated with the observed forcing at Grays Harbor (comparable to the results in Figure 34). 6 additional forcing scenarios were emulated with one forcing excluded (e.g., waves, wind, pressure, MMSLA, streamflow, and only tidal) to isolate the contribution of each particular forcing on WLs. A particular forcing contribution was calculated as the emulator output WL with full forcing minus the emulator output WL with all forcing except the component of interest. The exception to this is tidal (which cannot be turned off due to how it is included in the emulator) which was calculated simply as emulator output with no other forcing but tidal. WL contributions were calculated at the time of annual maximum WL events as determined by the full forcing time

series. This results in forcing contribution components for 31 annual maxima events. The average of these 31 events was taken with components plotted along East-West and North-South transects crossing the estuary (Figure 34).

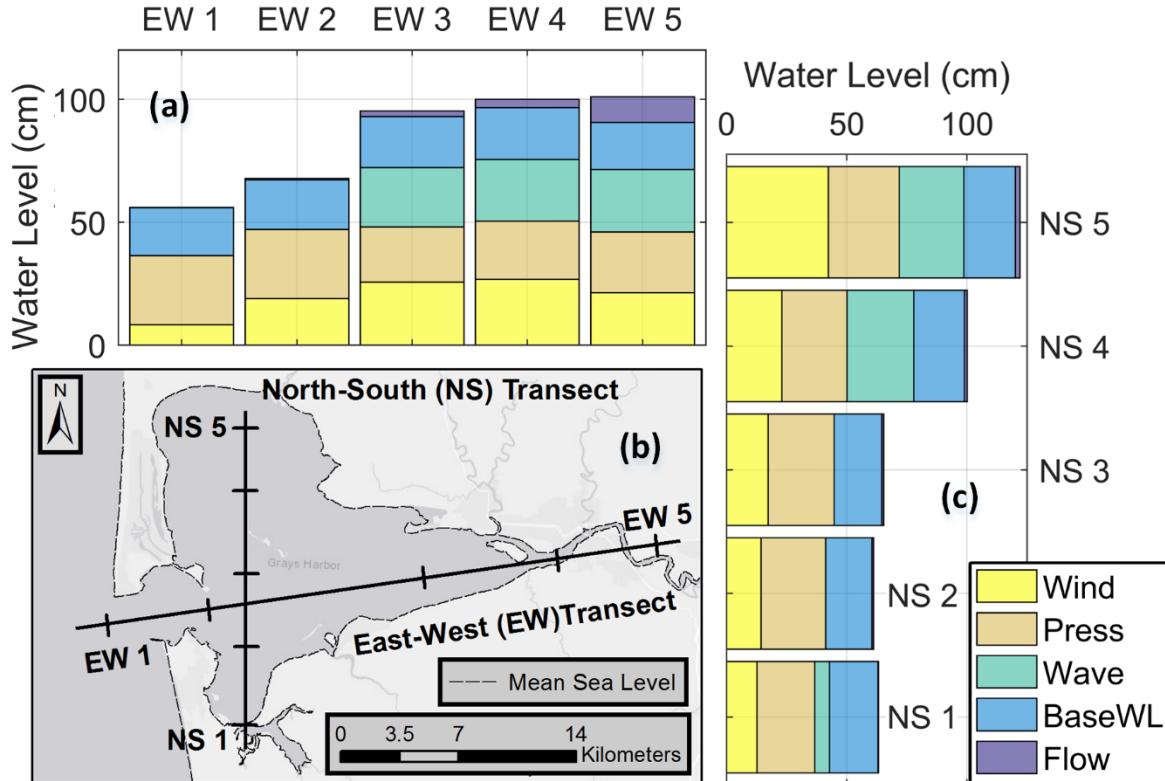


Figure 34: Average WL contribution from forcing components during extreme events. Two transects are plotted with subplot (b) showing transect locations with approximate station locations marked as ticks. Subplot (a) is the East-West transect and subplot (c) is the North-South transect.

The diverse mix of contributions for each bar in Figure 34 shows that extremes events are compound in nature and controlled by multiple forcings. The exception to this is streamflow which is negligible at all locations except the easternmost station (EW5, Figure 11). While some contributions are approximately constant across the domain (e.g. pressure and base WL), others are spatially variable leading to both an East-West and North-South gradient in water levels. This is partly from wave influence which is found to have a significant contribution but only at stations in the northern and eastern reaches of the bay. The influence of wind increases to the

north, most likely due to the mean wind direction emanating from the south during storm events. The influence of pressure anomalies on extreme WLs is found to be uniform but this result is likely from the spatial simplification of sea level pressure fields.

Not shown in Figure 34 is the contribution from tidal forcing. This is primarily for scale reasons as the tidal component is an order of magnitude larger than that from other forcing (average of 140 cm). Tides additionally show a gradient across the estuary domain although of the opposite pattern as that shown in Figure 34. The tidal component of WLs decreases by about 30 cm moving from the center of the estuary moving North or East. As WLs are the sum of these two components (tides and forcing driven NTR), the calculated gradient in total WLs is less than that observed in Figure 34 (under 20 cm across the two transects).

4.4 Discussion

4.4.1 Effect of simulator simplifications

The hierarchical validation used in this study provides a unique quantification of the error budget as sourced from multiple simplification levels. A comparison of model performance at various simplification levels (Table 4) found that the primary source of lost skill is from the fully dynamic simulator rather than from simulator simplifications or from emulation. Averaging across stations and deployment periods, all simplifications and emulation only increased RMSE by 1 cm relative to the level 1 error.

This result is of particular interest when compared to the quantification of discrete error for only emulation (see the histogram in Figure 30) which shows that emulation introduces comparatively significant error into WL estimates. Acknowledging the larger sample size of this analysis and more rigorous k-fold validation, it is still relevant that the RMSE of discrete level 4 error is double that of the combined level 2 and level 3 errors. Furthermore, a cumulative level 4 comparison at the tide gauge (Figure 31) found a RMSE of 15 cm while Level 4 itself (Figure 30) had a RMSE of 13 cm. This result suggests that the error from the emulator construction is not necessarily additive nor linear. In other words, the cumulative error is not the sum of the discrete errors. For this case, the dominance of level 1 error is found to mask that of the other

levels. This may not be true if level 1 error was significantly reduced at which point other simplifications may become relevant to the error budget.

In terms of quantifying model skill, the RMSE for comparisons to the 1999 USACE field data is overall larger than the RMSE in comparison to the tide gauge. A close examination of the USACE WL time series shows significant high frequency noise that is likely the cause of the overall larger RMSE values. The source of this noise is unclear but is most likely measurement noise or local disturbances (e.g. ship wakes). It is expected that the tide gauge, located within a stilling well with controlled calibration and validation procedures, would present a more accurate representation of the modeling error.

This study suggests that the most effective action to improve emulated WL predictions is to reduce level 1 error. One option would be optimized tuning, a process which can be accomplished by including tuning parameters within the emulator framework [Hall et al., 2009; Kennedy et al., 2006]. ADCSWAN could also be replaced with a different simulator or physics implementation that could better resolve processes at the location of interest. For example, ADCSWAN could have been run in 3D baroclinic mode. This would come at the cost of drastically increasing computation time and requiring additional input dimensionality in the form of density, temperature, and salinity fields. An additional source of error for hydrodynamic simulators in general is bathymetry. Bathymetry is a first order control on WLs and unresolved or incorrect topographic features are an important fraction of the error budget. This is likely the source of errors for WL station 6. Figure 35 shows WLs at this station having an asymmetric tidal signal indicative of shallow water while the observations have less asymmetry. This suggests that the water depth at the time of deployment was greater than the depth in the bathymetric data set used to generate the ADCSWAN grid. Therefore investment in more accurate or more recent bathymetry is a viable step for decreasing level 1 error.

Level 2 error could likely be reduced by making less aggressive simplifications of forcing inputs. It is conceptually straight forward to include other input dimensionality such as spatially variable atmospheric forcing or full spectral wave forcing. A promising strategy is the inclusion of field variables through decomposing the field into principal components [Higdon et al., 2008; Liu and Guilliams, 2016].

There are additionally a range of options for avoiding the stationarity assumption made in this study, which would eliminate level 3 error. The incorporation of temporal variability in emulators is reviewed by Reichert et al. [2011] who suggests the following strategies:

- 1) Apply a standard emulation methodology but with time as an additional degree of dimensionality. [Conti et al., 2005]
- 2) Describe the time series using basis functions and then apply emulation to the basis function coefficients. [Bayarri et al., 2007; Higdon et al., 2008]
- 3) Emulate the difference from one time point to the next. [Bhattacharya, 2007; Conti et al., 2009]
- 4) Use a Gaussian stochastic process as a Bayesian prior. [Liu and West, 2009]
- 5) Develop a hybrid dynamic/emulated model, or a “Mechanistic dynamic emulator.” [Albert C., 2012; Reichert et al., 2011]

In the context of this study, strategy 1 is conceptually the simplest but it is not clear a priori how far into the past the system’s memory extends and each included time step exponentially increases the dimensionality of the input space. Strategy 2 is complicated by identifying basis functions that adequately capture the various contributing signals. For example, a Fourier transformation is a natural solution except that storm surge is non-periodic and an important contributor to estuarine WLs. Strategies 3-5 all have potential advantages, but bring additional complexity to an already complex system so were not explored further.

For reducing level 4 error, GPR is a flexible framework and there are likely gains to be made through a more exhaustive approach for emulator specification. In particular, handling of the periodic nature of tides within the covariance function [Roberts et al., 2013] is a promising direction for reducing error.

4.4.2 Computational Cost Considerations

Emulation is an approach to dramatically reduce simulation times and is particularly valuable in situations where the simulator must be run for very long periods or repeatedly for many iterations (i.e., probabilistic risk assessment). While emulation requires an upfront cost, through the running of multiple simulations to construct a training dataset (see section 4.2.5), after this initial computational cost is paid the emulator is comparatively instantaneous. As the

nature of the trade-off is computation time, it is useful to review the costs of building the training dataset.

The first control on computational cost for the training set is the number of input dimensions. The simplifications implemented in this study managed to reduce the input space to 16 dimensions. Each design point took approximately 66 core hours to run so a full experimental design of 160 points would require over 10.5 thousand core hours (although with parallelization the actual time was much less). This study developed a larger experimental design (over 400 points) but this was primarily for validation rather than emulator skill (see discussion below). Full ADCSWAN required approximately 18 core hours per day of simulation time. This means that the comparison in Figure 34 (simulating 31 years of tide gauge data across 7 scenarios) would have required around 80 thousand core hours to complete. Once the emulators had been constructed, this analysis was accomplished in just hours. In this specific case, emulation becomes an efficient option if approximately one and a half years of simulation are required. This limit is highly situationally dependent and is controlled by computer, simulator, emulator, etc. and is intended only as an order of magnitude reference. Furthermore, emulation is primarily targeted at probabilistic methodologies rather than hindcasting for which multiple iterations of time series quickly sum to very large total simulation times.

The above analysis is based on a LHS design and the Loepky et al. [2009] guideline that a training dataset should be around 10 times the number of input dimensions. However, LHS is one of many possible experimental designs [Levy and Steinberg, 2010]. Significant research has focused on optimizing experimental designs beyond LHS and it is possible that a more complex design could reduce the size of the training dataset. For example, LHS does not consider the probability that a particular combination of input parameters may occur. Therefore, some design points are likely poorly utilized exploring space that is physically impossible or highly improbable (for example high wave height, low wave period parameter space).

Finally, the above analysis did not consider the effect of training dataset size on skill. This relationship was tested by quantifying emulator performance at a variety of training dataset sizes. For this analysis, the total body of simulations (480) was partitioned into smaller dataset sizes ranging from 50 to 450 simulations for testing. For each smaller dataset, a k-fold validation

with 5 segments was performed (Figure 35) to quantify emulator skill at this smaller training dataset size. This analysis is identical to that described in section 4.3.2 but with an artificially decreased training dataset size.

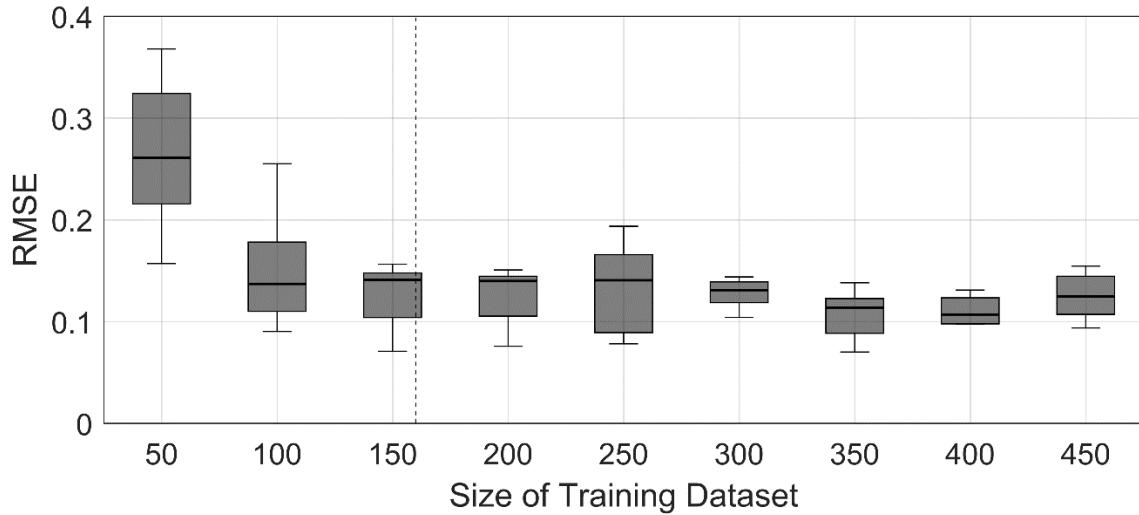


Figure 35: Impact of training dataset size on emulator skill. RMSE is calculated with a 5-fold cross-validation and represented by a standard boxplot. The dotted vertical line represents the theoretical dataset size from Loepky et al. [2009].

Our results are in good agreement with the guidance of Loepky et al. [2009] that ten times the number of input dimensions is sufficient for building a skillful emulator (Figure 35). Beyond this limit, only very small gains in skill are realized, suggesting that it is not efficient to over build the training dataset.

4.4.3 Emulation Beyond WLs

Although this study has primarily focused on WLs, emulation can easily be extended to other variables in a coastal hazards framework. Following a procedure similar to that of WLs, Hs was emulated at the observational Hs stations from the 1999 USACE field campaign (Figure 25) [Cialone et al., 2001; 2002] for validation. The Hs emulator was developed identically to that of WLs except that Hs emulation was found to not need cubic terms for the prior mean function. A comparison to observations (Figure 36) shows that the emulator performs well for Hs with the peak (around 26 September, 1999) being well reproduced by the emulator at stations 0, 1, 2, 3, and 4 (Figure 36). Performance is comparatively poor at stations 5 and 6, which are further

within the estuary and less dominated by offshore waves. This result is quantified in Table 5 which shows poor skill for interior Hs stations. It should be noted, however, that the Hs signals at these two stations have low variance and are barely above the noise floor.

Results show that, similarly to WLs, the largest loss of skill is at level 1. Simplified simulator and emulator results are found to closely track the full simulator. This is most evident at the bay interior stations where the full simulator and emulator are both found to over-predict Hs. Table 5 reveals that level 2, 3 and 4 simplifications produce little loss of skill for calculated Hs (average increase in RMSE of 2 cm). An exception to this is Hs station 5 which shows a significant increase in RMSE at the emulation level. The cause of poor emulator performance at this one location is unclear but it is likely due to a poor emulator specification.

Overall these results suggest that emulation could be integrated into many parts of an estuary modeling system. However, a key assumption of emulation with GPR is smoothness in response characteristics, suggesting that GPR is sub-optimal for “jumpy” variables. Not shown are similar results for Tp which can exhibit discontinuities within estuaries as Tp switches from one wave spectrum component to another. The emulator is

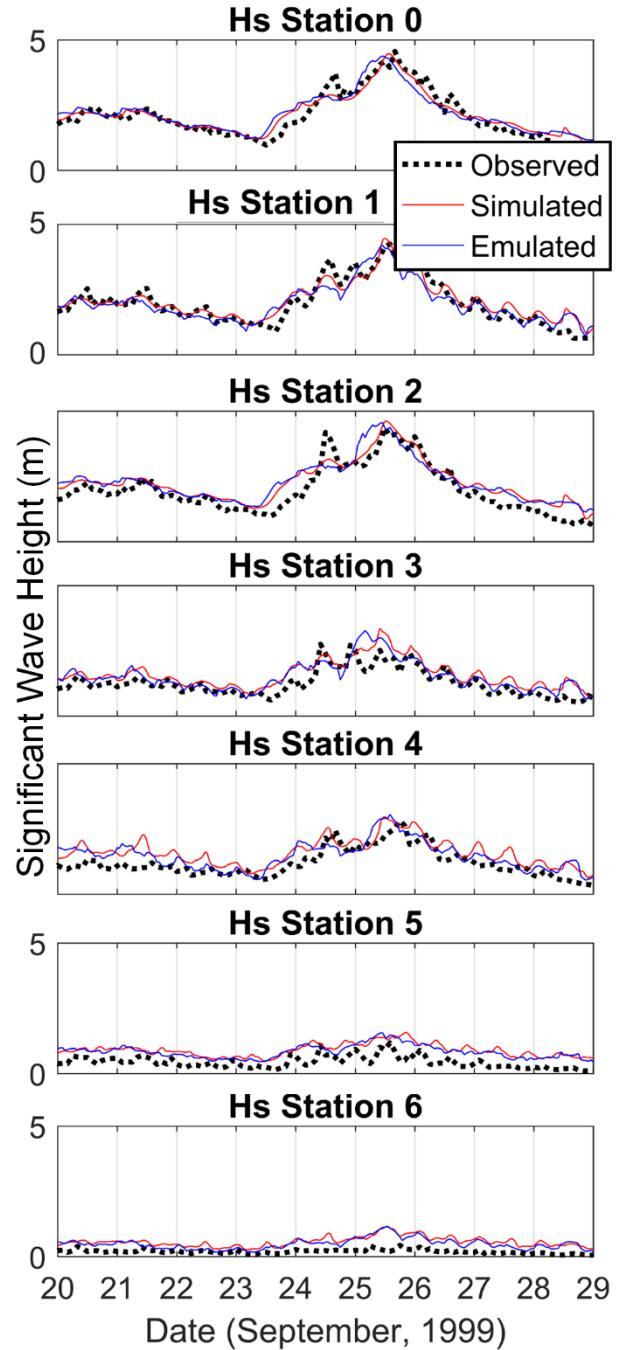


Figure 36: Comparison of observations (USACE deployments), full simulator and emulated Hs time series. Symbols are used for the observations for the sake of visual clarity. All panels have the same y-axis scaling

qualitatively able to capture Tp characteristics but cannot resolve these instantaneous jumps. For this reason, it is important to carefully consider the form of the output variable being emulated and its relation to the input parameters.

Table 5: RMSE values comparing Hs model output to observations at various simplification levels. Rows represent station locations while the two column groupings represent instrument deployments. Hs station 0/7 was relocated so deployment 1 values represent the Hs station 0 location and deployment 2 values representing the station 7 location.

Station	Deployment 1: RMSE (cm)			Deployment 2: RMSE (cm)		
	Level 1	Level 2	Level 4	Level 1	Level 2	Level 4
Hs 0/7	32	32	37	73	73	75
Hs 1	38	39	37	61	61	69
Hs 2	45	46	51	48	49	61
Hs 3	45	45	35	52	52	50
Hs 4	49	50	40	148	148	42
Hs 5	47	47	104	136	136	209
Hs 6	35	35	28	40	40	38

4.4.4 Extreme WL Event Decomposition

The decomposition of annual maximum WLs throughout the bay into its constituent components supports the findings in Chapter 3 that extremes in PNW estuaries are controlled by compound events. This conclusion is further supported by the variance in emulated extreme WL contributions (not shown) which demonstrate that the composition of each annual maximum event varies widely across the 31 yearly events. The mean contribution of each forcing is found to be significant providing evidence that all included forcing processes are important for properly quantifying extreme water levels. The only exception to this is streamflow which is found to be of nominal importance for all regions other than very near the streamflow boundary. This result is likely specific to the Grays Harbor estuary and would be different for a more hydrologically dominated estuary system [Chen et al., 2014, Lavery and Donovan, 2005; Svensson and Jones, 2004].

While most components (other than streamflow) contribute to extreme WLs across the estuary, wave contributions are only found at locations away from the center of the bay. For comparison, Olabarrieta et al. [2011] investigated the influence of wave setup in a PNW estuary (Willapa Bay, WA) using a 3-D fully coupled ocean-wave model. They found that wave induced setup is relatively consistent across the estuary domain with only a small decay of magnitude with distance into the estuary. Willapa Bay has a significantly shallower mouth than Grays Harbor, so wave breaking (and therefore radiation stress induced wave setup), is likely occurring immediately upon entering the estuary. Grays Harbor has a deeper channel allowing waves to penetrate further into the estuary. However, Figure 36 does show some wave attenuation at both station 3 and 4 so it is somewhat unexpected to see no wave contribution at these locations.

Further investigation shows that the choice of investigating maximum WLs is significant to this conclusion regarding wave influence. Analyzing forcing during the investigated sample of 31 annual maxima events shows that annual maxima WLs do not represent the annual maxima in terms of forcing. This is because tides are the largest component of extreme WLs and therefore a dominant control on the occurrence of an annual maxima event. Re-calculating WL components based on the annual maxima NTR (rather than WL) reveals an entirely different WL component balance. WL components based on annual maxima NTR events show waves contributing to WLs at all station transects. This suggests that missing wave contribution for interior stations in Figure 34 may be because maxima WLs are not concurrent with large wave heights. The small to medium waves during these events are able to penetrate into the estuary without breaking and therefore do not contribute to WLs. This is not the case during larger wave events.

4.5 Conclusions

This paper has presented an application of emulation, or surrogate modeling, to the problem of rapidly simulating hydrodynamic variables within the Grays Harbor, WA estuary. This methodology is targeted towards a variety of computationally constrained problems including probabilistic modeling, uncertainty quantification, model optimization, and non-parametric extreme event analysis. To facilitate moving forward with these goals, this study has focused on validating and quantifying the error induced by emulation. Additionally, a variety of

simplifications to the simulator have been suggested for reducing input dimensionality, and therefore the size of the training dataset.

The results from this study suggest that the utilized emulator is skillful for calculating a variety of model output variables (WL, NTR, and Hs). Comparisons to observations show that averaging across a variety of observation points and time periods, predicted WL RMSE is less than 24 cm. The bulk of this error is from comparisons to noisy instrumental deployments and a decadal scale comparison to tide gauge data showed the emulator having a smaller RMSE of 15 cm. Emulator performance is evaluated at multiple observation points across the estuary domain providing confidence that emulation is skillful across spatial extents. Decomposing the error from different emulator construction simplification levels shows that the largest source of unexplained variance in emulator hindcasts is from ADCSWAN. Of particular interest, strong simulator simplifications (including that of stationarity) are a relatively low contributor to loss in emulator performance (average increase in RMSE of 1 cm). Therefore, future efforts to improve emulator performance should focus on improving the full simulator before reducing simplifications or optimizing the emulator.

Emulation is found to be very efficient after the construction of the training dataset. Using a LHS experimental design, analysis shows that the training dataset size guidance of 10 times the number of input dimensions [Loeppky et al., 2009] is optimal.

Finally, the emulator was applied to investigating the relative contributions of different forcing variables to extreme WLs at the study site. Results show the importance of compound events to flooding in the PNW with a diverse mix of forcing contributing to annual extreme WLs. Streamflow was found to be, on average, a relatively minor contributor to extremes. Wave contributions were found to be significant but only at locations near the edges of the bay, suggesting strong wave penetration during extreme events. Further investigation showed that this effect is because annual maxima WL events are not necessarily concurrent with large wave events. While this was explored as an example application of emulation to estuary hydrodynamics questions, results hint at the potential of emulation to a broad range of applications.

Overall, this paper introduces and validates a powerful new technique for analyzing estuarine hydrodynamics and flood hazards. Ideally this study will provide a foundation for connecting emulation to a variety of computationally constrained problems in estuarine hydrodynamics.

Acknowledgments

Tide gauge records are available through the National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS) website. River discharge is available from the USGS through the National Water Information System. The NARR climate dataset is available through NOAA's Earth System Research Laboratory website. Bathymetric and topographic data were obtained from NOAA's Bathymetric Data viewer (DEMs) and DOGAMI's LiDAR download portal. We thank Melisa Menendez and Jorge Perez at the Environmental Hydraulics Institute of the Universidad de Cantabria (IH Cantabria) for providing the Global Ocean Wave 2 (GOW2) data. We also thank Mary Cialone and Dave Michalson of the U.S. Army Corps of Engineers for providing the Grays Harbor field observations from 1999. This work was funded by the NOAA Regional Integrated Sciences and Assessments Program (RISA) (NA15OAR4310145) and a contracted grant with the Quileute tribe.

Chapter 5 – General Conclusion

This dissertation has explored challenges, approaches, and conclusions in regards to climate change impacts to flooding hazards in the Pacific Northwest (PNW). Using multiple study sites and modeling perspectives, this dissertation has pushed the current state of understanding for these systems as well as developed tools and methodologies for further progress. The following sections provide a brief synthesis of the preceding chapters.

5.1 Chapter Conclusions

Chapter 2 of this dissertation focused on the problem of bias in global climate models (GCMs). When using GCM output as forcing, bias can propagate leaving hazard assessments out of alignment with reality. Therefore, some form of bias correction is often necessary. This chapter specifically addressed biased wave model data but the developed procedures and conclusions are more broadly applicable and were used for a variety of GCM output variables in other chapters. It was found that statistical bias correction is an effective tool for realigning biased model output data with observations. A robust testing of a variety of different methodologies highlighted the importance of inter-variable correlation structure and that this structure can be distorted without proper consideration. For this reason, bivariate methods that explicitly correct correlation structure were shown to be superior to univariate methods. This chapter additionally explored BC performance under nonstationary conditions through evaluating or extending tested methodologies functionality for a climate change context.

Chapter 3 built upon the tools developed in Chapter 2 to consider climate change impacts to two PNW estuary study sites. A multi-component modeling framework was developed to consider the combined impact of climate change on multiple estuary forcing components. Using decadal scale simulations, water level (WL) time series were analyzed to quantify changes to return interval (RI) event magnitude. Key findings were the importance of spatial variability in flooding characteristics with both study sites exhibiting increasing extreme WL magnitude with further distance into the estuary (a gradient of around 25 cm). As controlled by climate model forcing, this gradient is predicted to increase for 100-year RI events. An investigation of contributions to WLs found that extremes in the Pacific Northwest are defined by compound

events. Tides are the largest component of extreme WLs and an important modulator of extreme event occurrence. Finally this study explored the meaning of RIs in a nonstationary context. Using effective RIs it was found that the full effect of SLR is predicted to be offset by a decrease in forcing. Smaller RI events are buffered less by this effect than large RI events suggesting that low RI flooding events will be more impacted by SLR.

Chapter 4 proposed an alternative methodology to that employed in Chapter 3. An “emulation” based approach was investigated which allows for a fast statistical approximation of complex process-based models. This methodology was found to predict WLs skillfully across a range of observational datasets. A hierarchical validation showed that the largest loss of skill in emulator construction was from the process-based model approximating observations. The emulator was additionally found to predict other estuarine variables (significant wave height and non-tidal residual) skillfully. Finally, an example application of the emulator explored forcing contributions to extreme WLs. All forcings were found to contribute to extremes with streamflow being the smallest in magnitude and wave contributions being confined to the edges of the estuary.

5.2 Conclusions on Methodologies

This thesis provides a unique perspective on climate change impacts to estuarine flooding through approaching the problem with multiple methodologies. Research into climate change impacts to hazards can be broadly categorized into two distinct approaches: (1) observational-based and (2) process-based modeling. While both Chapter 3 and 4 propose distinct perspectives, they are both foundationally process-based modeling approaches. Emulation is best defined as a “hybrid” methodology, being a statistical model of a process-based model. However, emulator predictions are still reliant on process-based model accuracy rather than a function of observed environmental variables (other than through validation). A fully distinct approach would follow an observational-based methodology (most commonly through statistical analysis).

Observational studies have seen great success in investigating flooding [Kirshen et al., 2008, Serafin and Ruggiero 2014, Wahl et al., 2015] and the benefits of using in-situ data are many, including the avoidance of approximations and simplifications inherent to numerical modeling.

This dissertation's reliance on process-based methodologies represents both a strength and a weakness.

One limitation from adopting a process-based modeling approach is visible in the problem of bias in GCM data. Specifically, in the case of GCMs, process-based models are currently not able to reproduce observations due to a variety of insufficiencies ranging from course model resolution to incorrect physics. The BC methodologies presented in Chapter 3 are essentially a post-processing fix to realign incorrect model output with observations. As is mentioned in Chapter 2, it is a fix that comes with a wide range of caveots, theoretical issues, and likely incorrect assumptions [Ehret et al., 2012]. It is expected that, with increasing computation power and better GCM physics, the problem of bias will be removed and BC will no longer be necessary. This conclusion can be extended to process-based models in general: one of their main weaknesses is an inability to reproduce observations. This is for a variety of reasons ranging from incorrect physics, assumptions, numerical approximations, resolution, and many other sources of error. Constantly improving scientific understanding, models, and computing power continue to reduce these sources of model error making process-based modeling more viable through time.

A strength of process-based modeling, and the primary reason why this dissertation does not explore an observation-based statistical model, is the ability to extend information beyond observational sampling points. Most studies using observational information for coastal flood hazard analysis make use of tide gauge records and a large proportion of estuaries, particularly in the PNW, are ungauged or poorly sampled. Early engagement with communities and stakeholders placed a strong emphasis on producing spatial hazard information, in particular flood maps. Flooding maps require either process-based modeling or strong assumptions in regards to extending spatially limited observations. For example, point water level estimates are often extrapolated using a static “bath-tub” approximation (i.e., the assumption of a horizontal water surface). Previous research [Gallien et al., 2014] as well as this dissertation (Chapter 3) has shown that this can result in significant errors due to spatial variability in extreme WLs.

An additional reason why process-based models were chosen for this dissertation is the focus on climate change and predictions into the future. An ideal process-based model would be

as accurate for the future period as it is for the current period. With a perfect understanding of future forcing, a validated process-based model which captures all of the relevant physics should produce a correct forecast for the future. This is, of course, impossible as neither process-based models, nor our understanding of future forcing are perfect. Nevertheless this is a major attraction for using a process-based approach for understanding climate change. Statistical observation-based models can be used for future predictions but only very carefully. Some form of trend extrapolation or other model of change is required. More problematically, it must be assumed that the statistical relationships on which the model are built will not change into the future (or will change in a way specified by the model) [Wilby et al., 2004]. However, these issues can be minimized and the current imperfection of GCM models bring to question how well GCMs are able to predict the future as well.

While not exploring a fully observational perspective, this thesis does develop a contrast between process-based models and emulation. The main differentiation between these two methodologies is that of simulation time. Long simulation time is the second major weakness of process-based modeling for which emulation is a direct response. Reducing simulation times is beneficial in multiple ways, both as allowing analysis which requires long or multiple simulations (e.g. probabilistic assessment, optimization, etc.) as well as allowing more complex simulations. Simulation time and simulator accuracy are, in general, linked as the major sources of error in simulators are often sourced from simplifications to reduce simulation time, for example coarse spatial or temporal resolution, reduced physics, etc. Therefore a method that can reduce simulation times potentially alleviates both main weaknesses of a process-based methodology: both by allowing long/multiple simulations and a more accurate simulator.

However, it is important to remember that emulation is an approximation of a process-based model. This is especially true if, as detailed in Chapter 4, a reduced form of the simulator is used. Results from Chapter 4 showed nominal increases in error from emulation and simulator simplifications but this result is not general. It is expected that, if a more accurate simulator were used, the error from emulation and simulator simplifications would become more predominant. An example of this can be seen in the importance of pressure in producing a cross-estuary gradient in WLs in Chapter 3 results. This effect is lost in Chapter 4 due to simplified pressure fields. Therefore, while emulation can be extremely effective, dynamic models will always be

preferable if computational time is not a factor. With drastic increases to processing power, the benefits of emulation will be reduced.

5.3 Future Work

Future work will be aimed at extending conclusions drawn at specific study sites to the broader PNW and beyond. Of particular interest, these studies show important results in regards to WL contributions across estuary domains. However, the 3 main estuaries investigated in this dissertation are all characteristically similar in terms of size and forcing profiles. Any attempts to generalize conclusions would require a larger distribution of estuary types. This is especially true in regards to streamflow which was found to be a negligible factor in extreme WLs. Preliminary work at a smaller estuary (La Push, Washington) showed a much larger proportional contribution from streamflow. It is expected that there is a gradient in the importance of streamflow ranging from minimal (as seen in Grays, Tillamook, and Coos) to dominant in very small river inlets. Other forcings may also change across this gradient, for example wind which, for small estuaries, would not have a large fetch to build cross-estuary WL differentials. Exploring different size and types of estuaries could provide important conclusions in regards to what processes are important across the full range of estuary systems.

Even without exploring multiple estuary types, this dissertation lays a strong foundation for understanding the processes that effect flooding within estuaries in the PNW. Both Chapter 3 and 4 show the importance of compound events with tides modulating when extreme events occur. The long datasets of extremes developed in this dissertation could provide rich information for further exploring the joint probability of occurrence between the various contributors to extreme WLs. It is clear, from examining annual maximum WLs, that adding maximum NTR to maximum tidal WLs results in significant over-predictions of extreme events. It is also clear that just looking at tidal WLs (the dominant component of extreme WLs) results in a significant under-prediction. Developing a better understanding of where flooding estimates fall between these two extremes could have important implications for flooding hazard assessment. For example, it could result in the ability to predict what conditions will result in an annual flooding event in the PNW. Future work will re-analyze these datasets with a focus on characterizing the nature of PNW compound events.

This dissertation discusses at length the importance of uncertainty in future hazard predictions. This is argued as a major limitation to the modeling framework in Chapter 3 and a primary motivating objective for emulation (Chapter 4). Future research will build on the tools developed in this dissertation towards realizing a fully probabilistic flood hazard projection. In terms of a future forecast, there are multiple sources of uncertainty [Uusitalo et al., 2015]. The first is uncertainty as a function of model error. While not explicitly detailed in this dissertation, GPR was partly chosen due to a natural handling of uncertainty in model predictions. The GPR posterior distribution gives detailed probabilistic information in regards to uncertainty in the GPR's ability to approximate the simulator. This information can be supplemented with information on uncertainty from the simulator trying to reproduce observations [Higdon et al., 2004; Kennedy et al., 2001]. Additional sources of uncertainty come from climate variability which can be assessed through ensembles [Kay, 2015; Murphy et al., 2007] or statistical simulation models [Serafin et al., 2018]. Finally there is uncertainty in future projections a function of the future climate, especially as controlled by uncertain anthropological inputs [Nakicenovic et al., 2000]. This is generally handled through GCM ensembles under various climate scenarios. Future work will attempt to consolidate these multiple uncertainties into an integrated probabilistic future hazard assessment.

Bibliography

- Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1), 121–131. <https://doi.org/10.1002/joc.3413>
- Albert, C. (2012). A mechanistic dynamic emulator. *Nonlinear Analysis: Real World Applications*, 13(6), 2747–2754. <https://doi.org/10.1016/j.nonrwa.2012.04.003>
- Allan, J. C., & Komar, P. D. (2002a). Extreme Storms on the Pacific Northwest Coast during the 1997-98 El Niño and 1998-99 La Niña. *Journal of Coastal Research*. Coastal Education & Research Foundation, Inc. <https://doi.org/10.2307/4299063>
- Allan, J. C., & Komar, P. D. (2002b). Wave Climate Change and Coastal Erosion in the US Pacific Northwest. In *Ocean Wave Measurement and Analysis* (pp. 680–689). Reston, VA: American Society of Civil Engineers. [https://doi.org/10.1061/40604\(273\)70](https://doi.org/10.1061/40604(273)70)
- Allan, J. C., & Komar, P. D. (2006). Climate Controls on US West Coast Erosion Processes. *Journal of Coastal Research*, 223, 511–529. <https://doi.org/10.2112/03-0108.1>
- Anderson, T. R., Frazer, L. N., & Fletcher, C. H. (2010). Transient and persistent shoreline change from a storm. *Geophysical Research Letters*, 37(8), n/a–n/a. <https://doi.org/10.1029/2009GL042252>
- Andrade, C., Pires, H. O., Taborda, R., & Freitas, M. D. C. (2007). Projecting future changes in wave climate and coastal response in Portugal by the end of the 21st century. *Journal of Coastal Research*, SI 50(50), 263–257.
- Anjyo, K., & Lewis, J. P. (2011). RBF interpolation and Gaussian process regression through an RKHS formulation. *Journal of Math-for-Industry*, 3, 63–71. Retrieved from <http://gcoe-mi.jp/english/temp/publish/81a23bab2710961e1da72d05d4ca3e39.pdf>
- Apel, H., Merz, B., & Thielen, A. H. (2008). Quantification of uncertainties in flood risk assessments. *International Journal of River Basin Management*, 6(2), 149–162. <https://doi.org/10.1080/15715124.2008.9635344>
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4(0), 40–79. <https://doi.org/10.1214/09-SS054>
- Atkinson, J., McKee Smith, J., & Bender, C. (2013). Sea-Level Rise Effects on Storm Surge and Nearshore Waves on the Texas Coast: Influence of Landscape and Storm Characteristics. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 139(2), 98–117. [https://doi.org/10.1061/\(ASCE\)WW.1943-5460.0000187](https://doi.org/10.1061/(ASCE)WW.1943-5460.0000187)

- Di Baldassarre, G., Schumann, G., Bates, P. D., Freer, J. E., & Beven, K. J. (2010). Floodplain mapping: a critical discussion of deterministic and probabilistic approaches. *Hydrological Sciences Journal*, 55(3), 364–376. <https://doi.org/10.1080/02626661003683389>
- Barbier, E. B., Hacker, S. D., Kennedy, C., Koch, E. W., Stier, A. C., & Silliman, B. R. (2011). The value of estuarine and coastal ecosystem services. *Ecological Monographs*, 81(2), 169–193. <https://doi.org/10.1890/10-1510.1>
- Barnard, P. L., van Ormondt, M., Erikson, L. H., Eshleman, J., Hapke, C., Ruggiero, P., et al. (2014). Development of the Coastal Storm Modeling System (CoSMoS) for predicting the impact of storms on high-energy, active-margin coasts. *Natural Hazards*, 74(2), 1095–1125. <https://doi.org/10.1007/s11069-014-1236-y>
- Bass, B., & Bedient, P. (2018). Surrogate modeling of joint flood risk across coastal watersheds. *Journal of Hydrology*, 558, 159–173. <https://doi.org/10.1016/j.jhydrol.2018.01.014>
- Bauer, E., & Staabs, C. (1998). Statistical properties of global significant wave heights and their use for validation. *Journal of Geophysical Research: Oceans*, 103(C1), 1153–1166. <https://doi.org/10.1029/97JC02568>
- Bayarri, M. J., Berger, J. O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., et al. (2007). Computer Model Validation with Functional Output. *The Annals of Statistics*, 35(5), 1874–1906. <https://doi.org/10.1214/009053607000000163>
- Beamer, J. P., Hill, D. F., McGrath, D., Arendt, A., & Kienholz, C. (2017). Hydrologic impacts of changes in climate and glacier extent in the Gulf of Alaska watershed. *Water Resources Research*, 53(9), 7502–7520. <https://doi.org/10.1002/2016WR020033>
- Bhaskaran, P. K., Nayak, S., Bonthu, S. R., Murty, P. L. N., & Sen, D. (2013). Performance and validation of a coupled parallel ADCIRC–SWAN model for THANE cyclone in the Bay of Bengal. *Environmental Fluid Mechanics*, 13(6), 601–623. <https://doi.org/10.1007/s10652-013-9284-5>
- Bhattacharya, S. (2007). A simulation approach to Bayesian emulation of complex dynamic computer models. *Bayesian Analysis*, 2(4), 783–815. <https://doi.org/10.1214/07-BA232>
- Bilskie, M. V., Hagen, S. C., Medeiros, S. C., & Passeri, D. L. (2014). Dynamics of sea level rise and coastal flooding on a changing landscape. *Geophysical Research Letters*, 41(3), 927–934. <https://doi.org/10.1002/2013GL058759>
- Bissell, F. A., & Ferguson, R. A. (1975). The Jackknife--Toy, Tool or Two-Edged Weapon? *Journal of the Royal Statistical Society, Vol. 24*(No. 2), 79–100.
- Blain, C. A., & Rogers, W. E. (1998). *Coastal Tide Prediction Using the ADCIRC-2DDI Hydrodynamic Finite Element Model: Model Validation and Sensitivity Analyses in the*

Southern North Sea/English Channel. Retrieved from
<http://www.dtic.mil/docs/citations/ADA358752>

- Bode, L., & Hardy, T. A. (1997). Progress and Recent Developments in Storm Surge Modeling. *Journal of Hydraulic Engineering*, 123(4), 315–331. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1997\)123:4\(315\)](https://doi.org/10.1061/(ASCE)0733-9429(1997)123:4(315))
- Booij, N., Holthuijsen, L. H., & Ris, R. C. (1997). The “SWAN” Wave Model for Shallow Water. In *Coastal Engineering 1996* (pp. 668–676). New York, NY: American Society of Civil Engineers. <https://doi.org/10.1061/9780784402429.053>
- Bromirski, P. D., Flick, R. E., & Cayan, D. R. (2003). Storminess Variability along the California Coast: 1858–2000. *Journal of Climate*, 16(6), 982–993. [https://doi.org/10.1175/1520-0442\(2003\)016<0982:SVATCC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0982:SVATCC>2.0.CO;2)
- Bunya, S., Dietrich, J. C., Westerink, J. J., Ebersole, B. A., Smith, J. M., Atkinson, J. H., et al. (2010). A High-Resolution Coupled Riverine Flow, Tide, Wind, Wind Wave, and Storm Surge Model for Southern Louisiana and Mississippi. Part I: Model Development and Validation. *Monthly Weather Review*, 138(2), 345–377. <https://doi.org/10.1175/2009MWR2906.1>
- Caires, S., & Sterl, A. (2005). A New Nonparametric Method to Correct Model Data: Application to Significant Wave Height from the ERA-40 Re-Analysis. *Journal of Atmospheric and Oceanic Technology*, 22(4), 443–459. <https://doi.org/10.1175/JTECH1707.1>
- Caires, S., Sterl, A., Bidlot, J.-R., Graham, N., & Swail, V. (2004). Intercomparison of Different Wind–Wave Reanalyses. *Journal of Climate*, 17(10), 1893–1913. [https://doi.org/10.1175/1520-0442\(2004\)017<1893:IODWR>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<1893:IODWR>2.0.CO;2)
- Cardone, V. J., Jensen, R. E., Resio, D. T., Swail, V. R., Cox, A. T., Cardone, V. J., et al. (1996). Evaluation of Contemporary Ocean Wave Models in Rare Extreme Events: The “Halloween Storm” of October 1991 and the “Storm of the Century” of March 1993. *Journal of Atmospheric and Oceanic Technology*, 13(1), 198–230. [https://doi.org/10.1175/1520-0426\(1996\)013<0198:EOCOWM>2.0.CO;2](https://doi.org/10.1175/1520-0426(1996)013<0198:EOCOWM>2.0.CO;2)
- Carnell, R. (2017). LHS: Latin Hypercube Samples. R package. Retrieved from <http://lhs.r-forge.r-project.org>
- Castelletti, A., Galelli, S., Restelli, M., & Soncini-Sessa, R. (2012). Data-driven dynamic emulation modelling for the optimal management of environmental systems. *Environmental Modelling & Software*, 34, 30–43. <https://doi.org/10.1016/j.envsoft.2011.09.003>

- Cavaleri, L., & Sclavo, M. (2006). The calibration of wind and wave model data in the Mediterranean Sea. *Coastal Engineering*, 53(7), 613–627.
<https://doi.org/10.1016/j.coastaleng.2005.12.006>
- Charles, E., Idier, D., Delecluse, P., Déqué, M., & Le Cozannet, G. (2012). Climate change impact on waves in the Bay of Biscay, France. *Ocean Dynamics*, 62(6), 831–848.
<https://doi.org/10.1007/s10236-012-0534-8>
- Chelton, D. B., & Davis, R. E. (1982). Monthly Mean Sea-Level Variability Along the West Coast of North America. *Journal of Physical Oceanography*, 12(8), 757–784.
[https://doi.org/10.1175/1520-0485\(1982\)012<0757:MMSLVA>2.0.CO;2](https://doi.org/10.1175/1520-0485(1982)012<0757:MMSLVA>2.0.CO;2)
- Chen, C., Haerter, J. O., Hagemann, S., & Piani, C. (2011). On the contribution of statistical bias correction to the uncertainty in the projected hydrological cycle. *Geophysical Research Letters*, 38(20), n/a–n/a. <https://doi.org/10.1029/2011GL049318>
- Chen, W.-B., Liu, W.-C., Chen, W.-B., & Liu, W.-C. (2014). Modeling Flood Inundation Induced by River Flow and Storm Surges over a River Basin. *Water*, 6(10), 3182–3199.
<https://doi.org/10.3390/w6103182>
- Cheng, T. K., Hill, D. F., Beamer, J., & García-Medina, G. (2015a). Climate change impacts on wave and surge processes in a Pacific Northwest (USA) estuary. *Journal of Geophysical Research: Oceans*, 120(1), 182–200. <https://doi.org/10.1002/2014JC010268>
- Cheng, T. K., Hill, D. F., & Read, W. (2015b). The Contributions to Storm Tides in Pacific Northwest Estuaries: Tillamook Bay, Oregon, and the December 2007 Storm. *Journal of Coastal Research*, 313, 723–734. <https://doi.org/10.2112/JCOASTRES-D-14-00120.1>
- Christensen, J. H., Boberg, F., Christensen, O. B., & Lucas-Picher, P. (2008). On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters*, 35(20), L20709. <https://doi.org/10.1029/2008GL035694>
- Cialone, M. A., & Kraus, N. C. (2001). Engineering Study of Inlet Entrance Hydrodynamics: Grays Harbor, Washington, USA. In *Coastal Dynamics '01* (pp. 413–422). Reston, VA: American Society of Civil Engineers. [https://doi.org/10.1061/40566\(260\)42](https://doi.org/10.1061/40566(260)42)
- Cialone, M. A., Militello, A., Brown, M. E., & Kraus, N. C. (2002). Coupling of Wave and Circulation Numerical Models at Grays Harbor Entrance, Washington, USA. In *Proceedings 28th Coastal Engineering Conference* (pp. 1279–1291). World Scientific Publishing Company. https://doi.org/10.1142/9789812791306_0108
- Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., & Wilby, R. (2004). The Schaake Shuffle: A Method for Reconstructing Space–Time Variability in Forecasted Precipitation and Temperature Fields. *Journal of Hydrometeorology*, 5(1), 243–262.
[https://doi.org/10.1175/1525-7541\(2004\)005<0243:TSSAMF>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2)

- Cloern, J. E., Knowles, N., Brown, L. R., Cayan, D., Dettinger, M. D., Morgan, T. L., et al. (2011). Projected Evolution of California's San Francisco Bay-Delta-River System in a Century of Climate Change. *PLoS ONE*, 6(9), e24465. <https://doi.org/10.1371/journal.pone.0024465>
- Croke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*, 375(3-4), 613–626. <https://doi.org/10.1016/J.JHYDROL.2009.06.005>
- Conti, S., Anderson, C. W., Kennedy, M. C., & O 'Hagan, A. (2005). A Bayesian Analysis of Complex Dynamic Computer Models. *Sensitivity Analysis of Model Output*. Retrieved from https://www.researchgate.net/profile/Marc_Kennedy/publication/228738723_A_Bayesian_analysis_of_complex_dynamic_computer_models/links/09e41507ed0b37c273000000/A-Bayesian-analysis-of-complex-dynamic-computer-models.pdf
- Conti, S., Gosling, J. P., Oakley, J. E., & O 'hagan, A. A. (2009). Gaussian process emulation of dynamic computer codes. *Biometrika*, 96(3), 663–676. <https://doi.org/10.1093/biomet/asp028>
- Corbella, S., & Stretch, D. D. (2012). Predicting coastal erosion trends using non-stationary statistics and process-based models. *Coastal Engineering*, 70, 40–49. <https://doi.org/10.1016/j.coastaleng.2012.06.004>
- Crout, R. L., Sears, I. T., & Locke, L. K. (2008). The Great Coastal Gale of 2007 from Coastal Storms Program Buoy 46089. In *OCEANS* (pp. 1–7). IEEE. <https://doi.org/10.1109/OCEANS.2008.5152026>
- Dale, M., Wicks, J., Mylne, K., Pappenberger, F., Laeger, S., & Taylor, S. (2014). Probabilistic flood forecasting and decision-making: an innovative risk-based approach. *Natural Hazards*, 70(1), 159–172. <https://doi.org/10.1007/s11069-012-0483-z>
- Davis, J. R., Paramygin, V. a., Forrest, D., & Sheng, Y. P. (2010). Toward the Probabilistic Simulation of Storm Surge and Inundation in a Limited-Resource Environment. *Monthly Weather Review*, 138(7), 2953–2974. <https://doi.org/10.1175/2010MWR3136.1>
- Dawson, R. J., Hall, J. W., Bates, P. D., & Nicholls, R. J. (2005). Quantified Analysis of the Probability of Flooding in the Thames Estuary under Imaginable Worst-case Sea Level Rise Scenarios Quantified Analysis of the Probability of Flooding in the Thames Estuary under Imaginable Worst-case Sea Level Rise Scenarios. *International Journal of Water Resources Development*, 21(4), 577–591. <https://doi.org/10.1080/07900620500258380>
- Deque, M. (2007). Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change*, 57(1-2), 16–26. <https://doi.org/10.1016/j.gloplacha.2006.11.030>

- Dietrich, J. C., Bunya, S., Westerink, J. J., Ebersole, B. A., Smith, J. M., Atkinson, J. H., et al. (2010). A High-Resolution Coupled Riverine Flow, Tide, Wind, Wind Wave, and Storm Surge Model for Southern Louisiana and Mississippi. Part II: Synoptic Description and Analysis of Hurricanes Katrina and Rita. *Monthly Weather Review*, 138(2), 378–404. <https://doi.org/10.1175/2009MWR2907.1>
- Dietrich, J. C., Westerink, J. J., Kennedy, A. B., Smith, J. M., Jensen, R. E., Zijlema, M., et al. (2011). Hurricane Gustav (2008) Waves and Storm Surge: Hindcast, Synoptic Analysis, and Validation in Southern Louisiana. *Monthly Weather Review*, 139(8), 2488–2522. <https://doi.org/10.1175/2011MWR3611.1>
- Dietrich, J. C., Zijlema, M., Westerink, J. J., Holthuijsen, L. H., Dawson, C., Luettich, R. A., et al. (2011). Modeling hurricane waves and storm surge using integrally-coupled, scalable computations. *Coastal Engineering*, 58(1), 45–65. <https://doi.org/10.1016/J.COASTALENG.2010.08.001>
- Dietrich, J. C., Tanaka, S., Westerink, J. J., Dawson, C. N., Luettich, R. A., Zijlema, M., et al. (2012). Performance of the Unstructured-Mesh, SWAN+ADCIRC Model in Computing Hurricane Waves and Surge. *Journal of Scientific Computing*, 52(2), 468–497. <https://doi.org/10.1007/s10915-011-9555-6>
- Ding, Y., Nath Kuiry, S., Elgohry, M., Jia, Y., Altinakar, M. S., & Yeh, K.-C. (2013). Impact assessment of sea-level rise and hazardous storms on coasts and estuaries using integrated processes model. *Ocean Engineering*, 71, 74–95. <https://doi.org/10.1016/J.OCEANENG.2013.01.015>
- DOGAMI. (2009). Lidar Remote Sensing Data Collection: Coast of Oregon. Portland, OR.
- DOGAMI. (2010). Lidar Remote Sensing Data Collection: Southwest Washington. Portland, OR: Watershed Sciences.
- Durrant, T. H., Greenslade, D. J. M., & Simmonds, I. (2013). The effect of statistical wind corrections on global wave forecasts. *Ocean Modelling*, 70, 116–131. <https://doi.org/10.1016/j.ocemod.2012.10.006>
- Dushaw, B. D., Egbert, G. D., Worcester, P. F., Cornuelle, B. D., Howe, B. M., & Metzger, K. (1997). A TOPEX/POSEIDON global tidal model (TPXO.2) and barotropic tidal currents determined from long-range acoustic transmissions. *Progress in Oceanography*, 40(1-4), 337–367. [https://doi.org/10.1016/S0079-6611\(98\)00008-1](https://doi.org/10.1016/S0079-6611(98)00008-1)
- Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., & Liebert, J. (2012). HESS Opinions “Should we apply bias correction to global and regional climate model data?” *Hydrology and Earth System Sciences*, 16(9), 3391–3404. <https://doi.org/10.5194/hess-16-3391-2012>

- Engle, V. D., Kurtz, J. C., Smith, L. M., Chancy, C., & Bourgeois, P. (2007). A Classification of U.S. Estuaries Based on Physical and Hydrologic Attributes. *Environmental Monitoring and Assessment*, 129(1-3), 397–412. <https://doi.org/10.1007/s10661-006-9372-9>
- ESRI. (2016). ArcGIS: Spline with Barriers. Spatial Analyst license.
- Favre, A., & Gershunov, A. (2006). Extra-tropical cyclonic/anticyclonic activity in North-Eastern Pacific and air temperature extremes in Western North America. *Climate Dynamics*, 26(6), 617–629. <https://doi.org/10.1007/s00382-005-0101-9>
- FEMA. (2014). Flood Insurance Study: Coos County and Incorporated Areas.
- Feng, H., Vandemark, D., Quilfen, Y., Chapron, B., & Beckley, B. (2006). Assessment of wind-forcing impact on a global wind-wave model using the TOPEX altimeter. *Ocean Engineering*, 33(11-12), 1431–1461. <https://doi.org/10.1016/J.OCEANENG.2005.10.015>
- Ferreira, J. A., & Guedes Soares, C. (2000). Modelling distributions of significant wave height. *Coastal Engineering*, 40(4), 361–374. [https://doi.org/10.1016/S0378-3839\(00\)00018-1](https://doi.org/10.1016/S0378-3839(00)00018-1)
- Ferreira, J. A., & Guedes Soares, C. (2002). Modelling bivariate distributions of significant wave height and mean wave period. *Applied Ocean Research*, 24(1), 31–45. [https://doi.org/10.1016/S0141-1187\(02\)00006-8](https://doi.org/10.1016/S0141-1187(02)00006-8)
- Frey, A. E., Olivera, F., Irish, J. L., Dunkin, L. M., Kaihatu, J. M., Ferreira, C. M., & Edge, B. L. (2010). Potential Impact of Climate Change on Hurricane Flooding Inundation, Population Affected and Property Damages in Corpus Christi. *JAWRA Journal of the American Water Resources Association*, 46(5), 1049–1059. <https://doi.org/10.1111/j.1752-1688.2010.00475.x>
- Frigge, M., Hoaglin, D. C., & Iglewicz, B. (1989). Some Implementations of the Boxplot. *The American Statistician*, 43(1), 50–54. <https://doi.org/10.1080/00031305.1989.10475612>
- Funakoshi, Y., Hagen, S. C., & Bacopoulos, P. (2008). Coupling of Hydrodynamic and Wave Models: Case Study for Hurricane Floyd (1999) Hindcast. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 134(6), 321–335. [https://doi.org/10.1061/\(ASCE\)0733-950X\(2008\)134:6\(321\)](https://doi.org/10.1061/(ASCE)0733-950X(2008)134:6(321))
- Gallien, T. W., Schubert, J. E., & Sanders, B. F. (2011). Predicting tidal flooding of urbanized embayments: A modeling framework and data requirements. *Coastal Engineering*, 58(6), 567–577. <https://doi.org/10.1016/J.COASTALENG.2011.01.011>
- Gallien, T. W., Sanders, B. F., & Flick, R. E. (2014). Urban coastal flood prediction: Integrating wave overtopping, flood defenses and drainage. *Coastal Engineering*, 91, 18–28. <https://doi.org/10.1016/J.COASTALENG.2014.04.007>

- Ganju, N. K., Brush, M. J., Rashleigh, B., Aretxabaleta, A. L., Del Barrio, P., Gear, J. S., et al. (2015). Progress and Challenges in Coupled Hydrodynamic-Ecological Estuarine Modeling. *Estuaries and Coasts*. <https://doi.org/10.1007/s12237-015-0011-y>
- Gano, S., Kim, H., & Brown, D. (2006). Comparison of Three Surrogate Modeling Techniques: Datascape, Kriging, and Second Order Regression. In *11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*. Reston, Virginia: American Institute of Aeronautics and Astronautics. <https://doi.org/10.2514/6.2006-7048>
- García-Medina, G., Özkan-Haller, H. T., Ruggiero, P., Oskamp, J., García-Medina, G., Özkan-Haller, H. T., et al. (2013). An Inner-Shelf Wave Forecasting System for the U.S. Pacific Northwest. *Weather and Forecasting*, 28(3), 681–703. <https://doi.org/10.1175/WAF-D-12-00055.1>
- Grabemann, I., & Weisse, R. (2008). Climate change impact on extreme wave conditions in the North Sea: an ensemble study. *Ocean Dynamics*, 58(3-4), 199–212. <https://doi.org/10.1007/s10236-008-0141-x>
- Graham, N. E., Diaz, H. F., Graham, N. E., & Diaz, H. F. (2001). Evidence for Intensification of North Pacific Winter Cyclones since 1948. *Bulletin of the American Meteorological Society*, 82(9), 1869–1893. [https://doi.org/10.1175/1520-0477\(2001\)082<1869:EFIONP>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<1869:EFIONP>2.3.CO;2)
- Green, C., Viavattene, C., & Thompson, P. (2011). *Guidance for assessing flood losses CONHAZ Report*. Retrieved from https://www.mdx.ac.uk/__data/assets/pdf_file/0031/49189/floodsWP_FINALREPORTsept11.pdf
- Haerter, J. O., Hagemann, S., Moseley, C., & Piani, C. (2011). Climate model bias correction and the role of timescales. *Hydrology and Earth System Sciences*, 15(3), 1065–1079. <https://doi.org/10.5194/hess-15-1065-2011>
- Haigh, I. D., Wijeratne, E. M. S., MacPherson, L. R., Pattiaratchi, C. B., Mason, M. S., Crompton, R. P., & George, S. (2014). Estimating present day extreme water level exceedance probabilities around the coastline of Australia: tides, extra-tropical storm surges and mean sea level. *Climate Dynamics*, 42(1-2), 121–138. <https://doi.org/10.1007/s00382-012-1652-1>
- Hall, J. W., Manning, L. J., & Hankin, R. K. S. (2011). Bayesian calibration of a flood inundation model using spatial data. *Water Resources Research*, 47(5), W05529. <https://doi.org/10.1029/2009WR008541>
- Hanson, J. L. (2009). Pacific Hindcast Performance of Three Numerical Wave Models. *Journal of Atmospheric and Oceanic Technology*, 26(8), 1614–1633. <https://doi.org/10.1175/2009JTECHO650.1>

- Hemer, M. A., McInnes, K. L., & Ranasinghe, R. (2011). Climate and variability bias adjustment of climate model-derived winds for a southeast Australian dynamical wave model. *Ocean Dynamics*, 62(1), 87–104. <https://doi.org/10.1007/s10236-011-0486-4>
- Hemer, M. A., Katzfey, J., & Trenham, C. E. (2013). Global dynamical projections of surface ocean wave climate for a future high greenhouse gas emission scenario. *Ocean Modelling*, 70, 221–245. <https://doi.org/10.1016/j.ocemod.2012.09.008>
- Hemer, M. A., Fan, Y., Mori, N., Semedo, A., & Wang, X. L. (2013). Projected changes in wave climate from a multi-model ensemble. *Nature Climate Change*, 3(5), 471–476. <https://doi.org/10.1038/nclimate1791>
- Higdon, D., Gattiker, J., Williams, B., & Rightley, M. (2008). Computer Model Calibration Using High-Dimensional Output. *Journal of the American Statistical Association*, 103(482), 570–583. <https://doi.org/10.1198/016214507000000888>
- Holleman, R. C., & Stacey, M. T. (2014). Coupling of Sea Level Rise, Tidal Amplification, and Inundation. *Journal of Physical Oceanography*, 44(5), 1439–1455. <https://doi.org/10.1175/JPO-D-13-0214.1>
- Holthuijsen, L. H., Booij, N., & Bertotti, L. (1996). The Propagation of Wind Errors Through Ocean Wave Hindcasts. *Journal of Offshore Mechanics and Arctic Engineering*, 118(3), 184. <https://doi.org/10.1115/1.2828832>
- Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., et al. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*, 81(5), 345–354. Retrieved from <https://www.mrlc.gov/nlcd2011.php>
- Jakob Themeßl, M., Gobiet, A., & Leuprecht, A. (2011). Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, 31(10), 1530–1544. <https://doi.org/10.1002/joc.2168>
- Johnson, F., & Sharma, A. (2012). A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. *Water Resources Research*, 48(1), W01504. <https://doi.org/10.1029/2011WR010464>
- Johnson, M. E., Moore, L. M., & Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2), 131–148. [https://doi.org/10.1016/0378-3758\(90\)90122-B](https://doi.org/10.1016/0378-3758(90)90122-B)
- Jones, B., & Johnson, R. T. (2009). Design and analysis for the Gaussian process model. *Quality and Reliability Engineering International*, 25(5), 515–524. <https://doi.org/10.1002/qre.1044>

- Jongman, B., Ward, P. J., & Aerts, J. C. J. H. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, 22(4), 823–835. <https://doi.org/10.1016/J.GLOENVCHA.2012.07.004>
- Jun, M., Knutti, R., & Nychka, D. W. (2008). Spatial Analysis to Quantify Numerical Model Bias and Dependence. *Journal of the American Statistical Association*, 103, 934–947. <https://doi.org/10.1198/016214507000001265>
- Kantha, L. H., & Clayson, C. A. (2000). *Numerical models of oceans and oceanic processes*. (Vol. 66). Elsevier.
- Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8-12), 1287–1304. [https://doi.org/10.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8)
- Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society*, 96(8), 1333–1349. <https://doi.org/10.1175/BAMS-D-13-00255.1>
- Kennedy, M. C., Anderson, C. W., Conti, S., & O'Hagan, A. (2006). Case studies in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety*, 91(10-11), 1301–1309. <https://doi.org/10.1016/J.RESS.2005.11.028>
- Kim, S. W., Melby, J. a., Nadal-Caraballo, N. C., & Ratcliff, J. (2015). A time-dependent surrogate model for storm surge prediction based on an artificial neural network using high-fidelity synthetic hurricane modeling. *Natural Hazards*, 76(1), 565–585. <https://doi.org/10.1007/s11069-014-1508-6>
- Kirshen, P., Watson, C., Douglas, E., Gontz, A., Lee, J., & Tian, Y. (2008). Coastal flooding in the Northeastern United States due to climate change. *Mitigation and Adaptation Strategies for Global Change*, 13(5-6), 437–451. <https://doi.org/10.1007/s11027-007-9130-5>
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *International Joint Conference on Arti*

- cial Intelligence*. Retrieved from <http://robotics.stanford.edu/~ronnyk>
- Krien, Y., Dudon, B., Roger, J., & Zahibo, N. (2015). Probabilistic hurricane-induced storm surge hazard assessment in Guadeloupe, Lesser Antilles. *Natural Hazards and Earth System Sciences*, 15(8), 1711–1720. <https://doi.org/10.5194/nhess-15-1711-2015>
- Lafon, T., Dadson, S., Buys, G., & Prudhomme, C. (2013). Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *International Journal of Climatology*, 33(6), 1367–1381. <https://doi.org/10.1002/joc.3518>
- Lavery, S., & Donovan, B. (2005). Flood risk management in the Thames Estuary looking ahead 100 years. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 363(1831), 1455–74. <https://doi.org/10.1098/rsta.2005.1579>
- Leake, J., Wolf, J., Lowe, J., Stansby, P., Jacoub, G., Nichols, R., et al. (2007). Predicted Wave Climate for the UK: Towards an Integrated Model of Coastal Impacts of Climate Change (ASCE). *Estuary and Coastal Modeling*, 393–406. [https://doi.org/10.1061/40990\(324\)24](https://doi.org/10.1061/40990(324)24)
- Lee, H., & Brown, C. A. (2009). *Classification of Regional Patterns of Environmental Drivers and Benthic Habitats in Pacific Northwest Estuaries*. Western Ecology Division. <https://doi.org/EPA/600/R-09/140>.
- Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., et al. (2014). A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change*, 5(1), 113–128. <https://doi.org/10.1002/wcc.252>
- Levy, S., & Steinberg, D. M. (2010). Computer experiments: a review. *Advances in Statistical Analysis*, 94(4), 311–324. <https://doi.org/10.1007/s10182-010-0147-9>
- Li, C., Sinha, E., Horton, D. E., Diffenbaugh, N. S., & Michalak, A. M. (2014). Joint bias correction of temperature and precipitation in climate model simulations. *Journal of Geophysical Research: Atmospheres*, 119(23), 13,153–13,162. <https://doi.org/10.1002/2014JD022514>
- Li, H., Sheffield, J., & Wood, E. F. (2010). Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research*, 115(D10), D10101. <https://doi.org/10.1029/2009JD012882>
- Lin, N., Emanuel, K. A., Smith, J. A., & Vanmarcke, E. (2010). Risk assessment of hurricane storm surge for New York City. *Journal of Geophysical Research*, 115(D18), D18121. <https://doi.org/10.1029/2009JD013630>

- Lin, N., Emanuel, K. A., Oppenheimer, M., & Vanmarcke, E. (2012). Physically-based Assessment of Hurricane Surge Threat under Climate Change. *Nature Climate Change*, 2, 462–467. Retrieved from <https://dspace.mit.edu/handle/1721.1/75773>
- Lionello, P., Cogo, S., Galati, M. B., & Sanna, A. (2008). The Mediterranean surface wave climate inferred from future scenario simulations. *Global and Planetary Change*, 63(2-3), 152–162. <https://doi.org/10.1016/j.gloplacha.2008.03.004>
- Liston, G. E., & Elder, K. (2006a). A Distributed Snow-Evolution Modeling System (SnowModel). *Journal of Hydrometeorology*, 7(6), 1259–1276. <https://doi.org/10.1175/JHM548.1>
- Liston, G. E., & Elder, K. (2006b). A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet). *Journal of Hydrometeorology*, 7(2), 217–234. <https://doi.org/10.1175/JHM486.1>
- Liston, G. E., & Mernild, S. H. (2012). Greenland Freshwater Runoff. Part I: A Runoff Routing Model for Glaciated and Nonglaciated Landscapes (HydroFlow). *Journal of Climate*, 25(17), 5997–6014. <https://doi.org/10.1175/JCLI-D-11-00591.1>
- Liu, F., & West, M. (2009). A dynamic modelling strategy for Bayesian computer model emulation. *Bayesian Analysis*, 4(2), 393–411. <https://doi.org/10.1214/09-BA415>
- Liu, X., & Guillas, S. (2016). Dimension reduction for Gaussian process emulation: an application to the influence of bathymetry on tsunami heights, 1–26. Retrieved from <http://arxiv.org/abs/1603.07888>
- Loepky, J. L., Sacks, J., & Welch, W. J. (2009). Choosing the Sample Size of a Computer Experiment: A Practical Guide. *Technometrics*, 51(4), 366–376. <https://doi.org/10.1198/TECH.2009.08040>
- Love, M. R., Friday, D. Z., Grothe, P. R., Carignan, K. S., Eakins, B. W., & Taylor, L. A. (2012). *Digital Elevation Model of Astoria, Oregon: Procedures, Data Sources and Analysis*.
- Luettich, R A, J., Westerink, J. J., & Scheffner, N. W. (1992). ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts, and Estuaries. Report 1. Theory and Methodology of ADCIRC-2DDI and ADCIRC-3DL. Retrieved from <http://www.dtic.mil/docs/citations/ADA261608>
- Malde, S., Wyncoll, D., Oakley, J., Tozer, N., & Gouldby, B. (2016). Applying emulators for improved flood risk analysis. *E3S Web of Conferences*, 7, 04002. <https://doi.org/10.1051/e3sconf/20160704002>
- Malde, S., Oakley, J., & Wyncott, D. (2016). MUCM: Gaussian Process Emulator. R Package.

- Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., et al. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48(3), RG3003. <https://doi.org/10.1029/2009RG000314>
- Mark, D. J., Spargo, E. A., Westerink, J. J., & Luettich, R. A. , J. (2004). ENPAC 2003: A Tidal Constituent Database for Eastern North Pacific Ocean. Retrieved from <http://www.dtic.mil/docs/citations/ADA429079>
- Mass, C., Dotson, B., Mass, C., & Dotson, B. (2010). Major Extratropical Cyclones of the Northwest United States: Historical Review, Climatology, and Synoptic Environment. *Monthly Weather Review*, 138(7), 2499–2527. <https://doi.org/10.1175/2010MWR3213.1>
- Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., et al. (2010). Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties. Retrieved from <https://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf>
- Mathisen, J., & Bitner-Gregersen, E. (1990). Joint distributions for significant wave height and wave zero-up-crossing period. *Applied Ocean Research*, 12(2), 93–103. [https://doi.org/10.1016/S0141-1187\(05\)80033-1](https://doi.org/10.1016/S0141-1187(05)80033-1)
- Mazzini, P. L. F., Barth, J. A., Shearman, R. K., Erofeev, A., Mazzini, P. L. F., Barth, J. A., et al. (2014). Buoyancy-Driven Coastal Currents off Oregon during Fall and Winter. *Journal of Physical Oceanography*, 44(11), 2854–2876. <https://doi.org/10.1175/JPO-D-14-0012.1>
- McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2), 239–245. <https://doi.org/10.1080/00401706.1979.10489755>
- Mckay, P., & Blain, C. A. (2010). *Toward Developing a Hydrodynamic Flow and Inundation Model of the Lower Pearl River*. Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a513190.pdf>
- Mearns, L. O., Gutowski, W., Jones, R., Leung, R., McGinnis, S., Nunes, A., & Qian, Y. (2009). A Regional Climate Change Assessment Program for North America. *Eos, Transactions American Geophysical Union*, 90(36), 311–311. <https://doi.org/10.1029/2009EO360002>
- Mearns, L. O., Sain, S., Leung, L. R., Bukovsky, M. S., McGinnis, S., Biner, S., et al. (2013). Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP). *Climatic Change*, 120(4), 965–975. <https://doi.org/10.1007/s10584-013-0831-3>

- Mehrhan, A., AghaKouchak, A., & Phillips, T. J. (2014). Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations. *Journal of Geophysical Research: Atmospheres*, 119(4), 1695–1707.
<https://doi.org/10.1002/2013JD021152>
- Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., et al. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, 87(3), 343–360. <https://doi.org/10.1175/BAMS-87-3-343>
- De Michele, C., Salvadori, G., Passoni, G., & Vezzoli, R. (2007). A multivariate model of sea storms using copulas. *Coastal Engineering*, 54(10), 734–751.
<https://doi.org/10.1016/j.coastaleng.2007.05.007>
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Climate change. Stationarity is dead: whither water management? *Science (New York, N.Y.)*, 319(5863), 573–4.
<https://doi.org/10.1126/science.1151915>
- Mínguez, R., Espejo, A., Tomás, A., Méndez, F. J., Losada, I. J., Mínguez, R., et al. (2011). Directional Calibration of Wave Reanalysis Databases Using Instrumental Data. *Journal of Atmospheric and Oceanic Technology*, 28(11), 1466–1485. <https://doi.org/10.1175/JTECH-D-11-00008.1>
- Moel, H., Asselman, N., & Aerts, J. (2012). Uncertainty and sensitivity analysis of coastal flood damage estimates in the west of the Netherlands. *Nat. Hazards Earth Syst. Sci.*, 12, 1045–1058. <https://doi.org/10.5194/nhess-12-1045-2012>
- Moftakhari, H. R., AghaKouchak, A., Sanders, B. F., Feldman, D. L., Sweet, W., Matthew, R. A., & Luke, A. (2015). Increased nuisance flooding along the coasts of the United States due to sea level rise: Past and future. *Geophysical Research Letters*, 42(22), 9846–9852.
<https://doi.org/10.1002/2015GL066072>
- Moftakhari, H. R., AghaKouchak, A., Sanders, B. F., & Matthew, R. A. (2017). Cumulative hazard: The case of nuisance flooding. *Earth's Future*, 5(2), 214–223.
<https://doi.org/10.1002/2016EF000494>
- Montoya, R. D., Osorio Arias, A., Ortiz Royero, J. C., & Ocampo-Torres, F. J. (2013). A wave parameters and directional spectrum analysis for extreme winds. *Ocean Engineering*, 67, 100–118. <https://doi.org/10.1016/J.OCEANENG.2013.04.016>
- Mori, N., Yasuda, T., Mase, H., Tom, T., & Oku, Y. (2010). Projection of Extreme Wave Climate Change under Global Warming. *Hydrological Research Letters*, 4, 15–19.
<https://doi.org/10.3178/hrl.4.15>

- Morris, M. D., & Mitchel, T. J. (1995). Exploratory Designs for Computational Experiments. *Journal of Statistical Planning and Inference*, 43(3), 381–402.
[https://doi.org/10.1016/0378-3758\(94\)00035-T](https://doi.org/10.1016/0378-3758(94)00035-T)
- Mote, P. W., & Salathé, E. P. (2010). Future climate in the Pacific Northwest. *Climatic Change*, 102(1-2), 29–50. <https://doi.org/10.1007/s10584-010-9848-z>
- Mousavi, M. E., Irish, J. L., Frey, A. E., Olivera, F., & Edge, B. L. (2011). Global warming and hurricanes: the potential impact of hurricane intensification and sea level rise on coastal flooding. *Climatic Change*, 104(3-4), 575–597. <https://doi.org/10.1007/s10584-009-9790-0>
- Mueller, B., & Seneviratne, S. I. (2014). Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophysical Research Letters*, 41(1), 128–134.
<https://doi.org/10.1002/2013GL058055>
- Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., & Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430(7001), 768–772.
<https://doi.org/10.1038/nature02771>
- Murphy, J. M., Booth, B. B. B., Collins, M., Harris, G. R., Sexton, D. M. H., & Webb, M. J. (2007). A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 365(1857), 1993–2028.
<https://doi.org/10.1098/rsta.2007.2077>
- Nakicenovic, N., Alcamo, J., Grubler, A., Riahi, K., A., R. R., Rogner, H. H., & Victor, N. (2000). *Special Report on Emissions Scenarios (SRES), A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Geneva, Switzerland: Cambridge University Press. Retrieved from <http://pure.iiasa.ac.at/6101/2/sres-en.pdf>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290.
[https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- National Research Council. (2012). *Sea-Level Rise for the Coasts of California, Oregon, and Washington*. Washington, D.C.: National Academies Press. <https://doi.org/10.17226/13389>
- Neumann, B., Vafeidis, A. T., Zimmermann, J., & Nicholls, R. J. (2015). Future Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding - A Global Assessment. *PLOS ONE*, 10(3), e0118571. <https://doi.org/10.1371/journal.pone.0118571>
- Nicholls, R. J., & Small, C. (2002). Improved estimates of coastal population and exposure to hazards released. *Eos, Transactions American Geophysical Union*, 83(28), 301.
<https://doi.org/10.1029/2002EO000216>

- NOAA. (2018). U.S. Digital Elevation Models. Retrieved from <https://www.ngdc.noaa.gov/mgg/coastal/coastal.html>
- NOAA National Centers for Environmental Information. (2003). Coastal Relief Model. <http://www.ngdc.noaa.gov/mgg/coastal/crm.html>.
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91(10-11), 1290–1300. <https://doi.org/10.1016/j.ress.2005.11.025>
- Oakley, J. (1999). Bayesian uncertainty analysis for complex computer codes. *Thesis*. Retrieved from <http://jeremy-oakley.staff.shef.ac.uk/jeothesis.pdf>
- Odigie, K. O., & Warrick, J. A. (2018). Coherence Between Coastal and River Flooding along the California Coast. *Journal of Coastal Research*, 342, 308–317. <https://doi.org/10.2112/JCOASTRES-D-16-00226.1>
- Olabarrieta, M., Warner, J. C., & Kumar, N. (2011). Wave-current interaction in Willapa Bay. *Journal of Geophysical Research*, 116(C12), C12014. <https://doi.org/10.1029/2011JC007387>
- Olsen, J. R., Lambert, J. H., & Haimes, Y. Y. (1998). Risk of Extreme Events Under Nonstationary Conditions. *Risk Analysis*, 18(4), 497–510. <https://doi.org/10.1111/j.1539-6924.1998.tb00364.x>
- Orton, P. M., Hall, T. M., Talke, S. A., Blumberg, A. F., Georgas, N., & Vinogradov, S. (2016). A validated tropical-extratropical flood hazard assessment for New York Harbor. *Journal of Geophysical Research: Oceans*, 121(12), 8904–8929. <https://doi.org/10.1002/2016JC011679>
- Panofsky, H. A., Brier, G. W., & Best, W. H. (1958). Some application of statistics to meteorology. Retrieved from <http://agris.fao.org/agris-search/search.do?recordID=US201300587097>
- Parey, S., Malek, F., Laurent, C., & Dacunha-Castelle, D. (2007). Trends and climate evolution: Statistical approach for very high temperatures in France. *Climatic Change*, 81(3-4), 331–352. <https://doi.org/10.1007/s10584-006-9116-4>
- Parey, S., Hoang, T. T. H., & Dacunha-Castelle, D. (2010). Different ways to compute temperature return levels in the climate change context. *Environmetrics*, 21(7-8), 698–718. <https://doi.org/10.1002/env.1060>
- Parker, W. S. (2011). When Climate Models Agree: The Significance of Robust Model Predictions. *Philosophy of Science*, 78(4), 579–600. <https://doi.org/10.1086/661566>

- Pawlowicz, R., Beardsley, B., & Lentz, S. (2002). Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. *Computers & Geosciences*, 28(8), 929–937. [https://doi.org/10.1016/S0098-3004\(02\)00013-4](https://doi.org/10.1016/S0098-3004(02)00013-4)
- Pelling, H. E., Mattias Green, J. A., & Ward, S. L. (2013). Modelling tides and sea-level rise: To flood or not to flood. *Ocean Modelling*, 63, 21–29. <https://doi.org/10.1016/J.OCEMOD.2012.12.004>
- Pendleton, L. H. (2010). *The economic and market value of coasts and estuaries: what's at stake?* Arlington: Restore America's Estuaries. Retrieved from <https://www.cabdirect.org/cabdirect/abstract/20103283407>
- Perez, J., Menendez, M., & Losada, I. J. (2017). GOW2: A global wave hindcast for coastal applications. *Coastal Engineering*, 124, 1–11. <https://doi.org/10.1016/J.COASTALENG.2017.03.005>
- Petersen, A. C., & Petersen, A. C. (2000). Philosophy of Climate Science. *Bulletin of the American Meteorological Society*, 81(2), 265–271. [https://doi.org/10.1175/1520-0477\(2000\)081<0265:POCS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<0265:POCS>2.3.CO;2)
- Piani, C., & Haerter, J. O. (2012). Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophysical Research Letters*, 39(20), n/a–n/a. <https://doi.org/10.1029/2012GL053839>
- Piani, C., Haerter, J. O., & Coppola, E. (2009). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, 99(1-2), 187–192. <https://doi.org/10.1007/s00704-009-0134-9>
- Le Provost, C., Genco, M. L., Lyard, F., Vincent, P., & Caneil, P. (1994). Spectroscopy of the world ocean tides from a finite element hydrodynamic model. *Journal of Geophysical Research*, 99(C12), 24777. <https://doi.org/10.1029/94JC01381>
- Pugh, D. T. (1996). *Tides, Surges and Mean Sea-Level*. John Wiley & Sons. Retrieved from <http://nora.nerc.ac.uk/id/eprint/119157/1/sea-level.pdf>
- Purvis, M. J., Bates, P. D., & Hayes, C. M. (2008). A probabilistic methodology to estimate future coastal flood risk due to sea level rise. *Coastal Engineering*, 55(12), 1062–1073. <https://doi.org/10.1016/j.coastaleng.2008.04.008>
- Rahmstorf, S. (2007). A semi-empirical approach to projecting future sea-level rise. *Science (New York, N.Y.)*, 315(5810), 368–70. <https://doi.org/10.1126/science.1135456>
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge: MIT press.

- Razavi, S., Tolson, B. a., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, 48(7). <https://doi.org/10.1029/2011WR011527>
- Read, L. K., & Vogel, R. M. (2015). Reliability, return periods, and risk under nonstationarity. *Water Resources Research*, 51(8), 6381–6398. <https://doi.org/10.1002/2015WR017089>
- Reichert, P., White, G., Bayarri, M. J., & Pitman, E. B. (2011). Mechanism-based emulation of dynamic simulation models: Concept and application in hydrology. *Computational Statistics & Data Analysis*, 55(4), 1638–1655. <https://doi.org/10.1016/j.csda.2010.10.011>
- Repko, A., Van Gelder, P. H. A. J. M., Voortman, H. G., & Vrijling, J. K. (2004). Bivariate description of offshore wave conditions with physics-based extreme value statistics. *Applied Ocean Research*, 26(3-4), 162–170. <https://doi.org/10.1016/j.apor.2004.10.001>
- Resio, D. T., & Westerink, J. J. (2008). Modeling the Physics of Storm Surges. *Physics Today*, 61(9). <https://doi.org/10.1063/1.2982120>
- Rilo, A., Freire, P., Guerreiro, M., Fortunato, A. B., & Taborda, R. (2013). Estuarine margins vulnerability to floods for different sea level rise and human occupation scenarios. *Journal of Coastal Research*, 65, 820–825. <https://doi.org/10.2112/SI65-139.1>
- Rintoul, S., Aoki, S., Campos, E., Chambers, D., Feely, R., Gulev, S., et al. (2013). Observations: Ocean. 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. Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press. Retrieved from https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_Chapter03_FINAL.pdf
- Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., & Aigrain, S. (2013). Gaussian processes for time-series modelling. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 371(1984), 20110550. <https://doi.org/10.1098/rsta.2011.0550>
- Rogers, W. E., & Wivfmann, P. A. (2002). *Quantifying the Role of Wind Field Accuracy in the U.S. Navy's Global Ocean Wave Nowcast/Forecast System*. <https://doi.org/NRL/MR/732-02-8290>
- Rogers, W. E., Wittmann, P. A., Wang, D. W. C., Clancy, R. M., Hsu, Y. L., Rogers, W. E., et al. (2005). Evaluations of Global Wave Prediction at the Fleet Numerical Meteorology and Oceanography Center*. *Weather and Forecasting*, 20(5), 745–760. <https://doi.org/10.1175/WAF882.1>
- Rogers, W. E., Kaihatu, J. M., Hsu, L., Jensen, R. E., Dykes, J. D., & Holland, K. T. (2007). Forecasting and hindcasting waves with the SWAN model in the Southern California Bight. *Coastal Engineering*, 54(1), 1–15. <https://doi.org/10.1016/J.COASTALENG.2006.06.011>

- Rootzén, H., & Katz, R. W. (2013). Design Life Level: Quantifying risk in a changing climate. *Water Resources Research*, 49(9), 5964–5972. <https://doi.org/10.1002/wrcr.20425>
- Rueda, A., Gouldby, B., Mendez, F. J., Tomas, A., Losada, I. J., Lara, J. L., & Diaz-Simal, P. (2016). The use of wave propagation and reduced complexity inundation models and metamodels for coastal flood risk assessment. *Journal of Flood Risk Management*, 9(4), 390–401. <https://doi.org/10.1111/jfr3.12204>
- Ruggiero, P. (2013). Is the Intensifying Wave Climate of the U.S. Pacific Northwest Increasing Flooding and Erosion Risk Faster Than Sea-Level Rise? *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 139(2), 88–97. [https://doi.org/10.1061/\(ASCE\)WW.1943-5460.0000172](https://doi.org/10.1061/(ASCE)WW.1943-5460.0000172)
- Ruggiero, P., Komar, P. D., & Allan, J. C. (2010). Increasing wave heights and extreme value projections: The wave climate of the U.S. Pacific Northwest. *Coastal Engineering*, 57(5), 539–552. <https://doi.org/10.1016/j.coastaleng.2009.12.005>
- Rusu, L., & Pilar, P. (2008). Hindcast of the wave conditions along the west Iberian coast. *Coastal Engineering*, 55(11), 906–919. <https://doi.org/10.1016/J.COASTALENG.2008.02.029>
- Sacks, J., Schiller, S. B., & Welch, W. J. (1989). Designs for Computer Experiments. *Technometrics*, 31(1), 41–47. <https://doi.org/10.1080/00401706.1989.10488474>
- Salas, J. D., & Obeysekera, J. (2014). Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events. *Journal of Hydrologic Engineering*, 19(3), 554–568. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000820](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000820)
- Serafin, K. A., & Ruggiero, P. (2014). Simulating extreme total water levels using a time-dependent, extreme value approach. *Journal of Geophysical Research: Oceans*, 119(9), 6305–6329. <https://doi.org/10.1002/2014JC010093>
- Serafin, K. A., Ruggiero, P., Barnard, P. L., & Stockdon, H. F. (2018). The Influence of Shelf Bathymetry and Beach Topography on Extreme Total Water Levels. *Coastal Engineering*, (In Revision)
- Smith, J. M., Cialone, M. A., Wamsley, T. V., & McAlpin, T. O. (2010). Potential impact of sea level rise on coastal surges in southeast Louisiana. *Ocean Engineering*, 37(1), 37–47. <https://doi.org/10.1016/J.OCEANENG.2009.07.008>
- Solomon, S. (2007). The Physical Science Basis : Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. *Intergovernmental Panel on Climate Change (IPCC), Climate Change 2007*, 996. Retrieved from <https://ci.nii.ac.jp/naid/10029678161/>

- Song, Y. K., Irish, J. L., & Udoh, I. E. (2012). Regional attributes of hurricane surge response functions for hazard assessment. *Natural Hazards*, 64(2), 1475–1490. <https://doi.org/10.1007/s11069-012-0309-z>
- Svensson, C., & Jones, D. A. (2004). Dependence between sea surge, river flow and precipitation in south and west Britain. *Hydrology and Earth System Sciences*, 8(5), 973–992. <https://doi.org/10.5194/hess-8-973-2004>
- Taylor, N. R., Irish, J. L., Udoh, I. E., Bilskie, M. V., & Hagen, S. C. (2015). Development and uncertainty quantification of hurricane surge response functions for hazard assessment in coastal bays. *Natural Hazards*, 77(2), 1103–1123. <https://doi.org/10.1007/s11069-015-1646-5>
- Teutschbein, C., & Seibert, J. (2013). Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? *Hydrology and Earth System Sciences*, 17(12), 5061–5077. <https://doi.org/10.5194/hess-17-5061-2013>
- Thompson, P. R., Merrifield, M. A., Wells, J. R., & Chang, C. M. (2014). Wind-Driven Coastal Sea Level Variability in the Northeast Pacific. *Journal of Climate*, 27(12), 4733–4751. <https://doi.org/10.1175/JCLI-D-13-00225.1>
- Thrasher, B., Maurer, E. P., McKellar, C., & Duffy, P. B. (2012). Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences*, 16(9), 3309–3314. <https://doi.org/10.5194/hess-16-3309-2012>
- Timmermans, B. (2015). *Uncertainty in Numerical Wind-wave Models*. University of Southampton.
- Tolman, H. L. (2009). *User manual and system documentation of WAVEWATCH III TM version 3.14*.
- Tolman, H. L., & Tolman, H. L. (1991). A Third-Generation Model for Wind Waves on Slowly Varying, Unsteady, and Inhomogeneous Depths and Currents. *Journal of Physical Oceanography*, 21(6), 782–797. [https://doi.org/10.1175/1520-0485\(1991\)021<0782:ATGMFW>2.0.CO;2](https://doi.org/10.1175/1520-0485(1991)021<0782:ATGMFW>2.0.CO;2)
- Tomas, A., Mendez, F. J., & Losada, I. J. (2008). A method for spatial calibration of wave hindcast data bases. *Continental Shelf Research*, 28(3), 391–398. <https://doi.org/10.1016/j.csr.2007.09.009>
- U.S. Department of the Interior, F. W. Q. A. (1975). Documentation report: FWQA dynamic estuary model. U.S. Govt. Print. Off. Retrieved from <http://agris.fao.org/agris-search/search.do?recordID=US201300468288>

- Uusitalo, L., Lehtinen, A., Helle, I., & Myrberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling & Software*, 63, 24–31. <https://doi.org/10.1016/J.ENVSOFT.2014.09.017>
- Vermeer, M., & Rahmstorf, S. (2009). Global sea level linked to global temperature. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), 21527–32. <https://doi.org/10.1073/pnas.0907765106>
- Vrac, M., & Friederichs, P. (2015). Multivariate—Intervariable, Spatial, and Temporal—Bias Correction. *Journal of Climate*, 28(1), 218–237. <https://doi.org/10.1175/JCLI-D-14-00059.1>
- De Waal, D. J., & van Gelder, P. H. A. J. M. (2006). Modelling of extreme wave heights and periods through copulas. *Extremes*, 8(4), 345–356. <https://doi.org/10.1007/s10687-006-0006-y>
- Wahl, T., Jain, S., Bender, J., Meyers, S. D., & Luther, M. E. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Climate Change*, 5(12), 1093–1097. <https://doi.org/10.1038/nclimate2736>
- Wang, R.-Q., Herdman, L. M., Erikson, L., Barnard, P., Hummel, M., & Stacey, M. T. (2017). Interactions of Estuarine Shoreline Infrastructure With Multiscale Sea Level Variability. *Journal of Geophysical Research: Oceans*. <https://doi.org/10.1002/2017JC012730>
- Wang, X. L., & Swail, V. R. (2002). Trends of Atlantic Wave Extremes as Simulated in a 40-Yr Wave Hindcast Using Kinematically Reanalyzed Wind Fields. *Journal of Climate*, 15(9), 1020–1035. [https://doi.org/10.1175/1520-0442\(2002\)015<1020:TOAWEA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1020:TOAWEA>2.0.CO;2)
- Wang, X. L., Swail, V. R., & Cox, A. (2010). Dynamical versus statistical downscaling methods for ocean wave heights. *International Journal of Climatology*, n/a–n/a. <https://doi.org/10.1002/joc.1899>
- Wang, X. L., Feng, Y., & Swail, V. R. (2014). Changes in global ocean wave heights as projected using multimodel CMIP5 simulations. *Geophysical Research Letters*, 41(3), 1026–1034. <https://doi.org/10.1002/2013GL058650>
- Watts, A., Grant, G., & Safeeq, M. (2016). *Flows of the future—How will climate change affect streamflows in the Pacific Northwest?* Science Findings 187. Portland, OR. Retrieved from <https://www.fs.usda.gov/treesearch/pubs/52180>
- Weaver, R. J., & Luettich, Jr., R. A. (2010). 2D vs. 3D Storm Surge Sensitivity in ADCIRC: Case Study of Hurricane Isabel. In *Estuarine and Coastal Modeling* (2009) (pp. 762–779). Reston, VA: American Society of Civil Engineers. [https://doi.org/10.1061/41121\(388\)44](https://doi.org/10.1061/41121(388)44)

- Westerink, J. J., Luettich, R. A., Baptists, A. M., Scheffner, N. W., & Farrar, P. (1992). Tide and Storm Surge Predictions Using Finite Element Model. *Journal of Hydraulic Engineering*, 118(10), 1373–1390. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1992\)118:10\(1373\)](https://doi.org/10.1061/(ASCE)0733-9429(1992)118:10(1373))
- Wilby, R., Charles, S., Zorita, E., Timbal, B., Whetton, P., Mearns, L., et al. (2004). Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods Document history Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods. Retrieved from <http://www.narccap.ucar.edu/doc/tgica-guidance-2004.pdf>
- Wolf, J. (2009). Coastal flooding: impacts of coupled wave–surge–tide models. *Natural Hazards*, 49(2), 241–260. <https://doi.org/10.1007/s11069-008-9316-5>
- Wood, A. W. (2002). Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research*, 107(D20), 4429. <https://doi.org/10.1029/2001JD000659>
- Wood, A. W., Leung, L. R., Sridhar, V., & Lettenmaier, D. P. (2004). Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs. *Climatic Change*, 62(1-3), 189–216. <https://doi.org/10.1023/B:CLIM.0000013685.99609.9e>
- Xu, C. -y. (1999). From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Progress in Physical Geography*, 23(2), 229–249. <https://doi.org/10.1177/030913339902300204>
- Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., et al. (2012). A New Global Climate Model of the Meteorological Research Institute: MRI-CGCM3; Model Description and Basic Performance. *Journal of the Meteorological Society of Japan*, 90A, 23–64. <https://doi.org/10.2151/jmsj.2012-A02>
- Zhang, K., Douglas, B. C., & Leatherman, S. P. (1999). Twentieth-Century Storm Activity along the U.S. East Coast. *O U R N A L O F C L I M A T E*, 13. Retrieved from <https://journals.ametsoc.org/doi/pdf/10.1175/1520-0442%282000%29013%3C1748%3ATCSAAT%3E2.0.CO%3B2>
- Zhang, K., Douglas, B., & Leatherman, S. (2002). Do Storms Cause Long Term Beach Erosion along the U.S. East Barrier Coast? *The Journal of Geology*, 110(4), 493–502. <https://doi.org/10.1086/340633>
- Zijlema, M. (2010). Computation of wind-wave spectra in coastal waters with SWAN on unstructured grids. *Coastal Engineering*, 57(3), 267–277. <https://doi.org/10.1016/J.COASTALENG.2009.10.011>

Appendix

Appendix A: Nonstationary Extensions for the Binning and Direct Method

The Binning Method (Future)

The Binning method as extended to the future case is proposed as procedurally a hybrid of the equidistant quantile mapping method (Figure 4) and the standard Binning method (Figure 5). The procedure is illustrated graphically in Figure 37. The first step is to bias correct the first variable (X) using a standard equidistant quantile mapping. This step is not shown in Figure 37 but would be identical to the process outlined in Figure 4. The second step (**2**) is to select a bin of the bias corrected first variable (X) from the target, model historic, and model future datasets. Figure 37 shows the selected decile bin (CDF_X values ranging from 0.2 – 0.3) in red. The third step is to determine CDFs from these bins of data (**3**). With these three bin dependent CDFs it is possible to perform a standard equidistant quantile mapping. This is done by first finding the quantile value corresponding to the model future value for variable (Y) (**4**). The target value and model historic value are found for this quantile and the difference between these two values is quantified (**5**). This difference is then applied to the model historic value so as to find the bias corrected model future value (**6**).

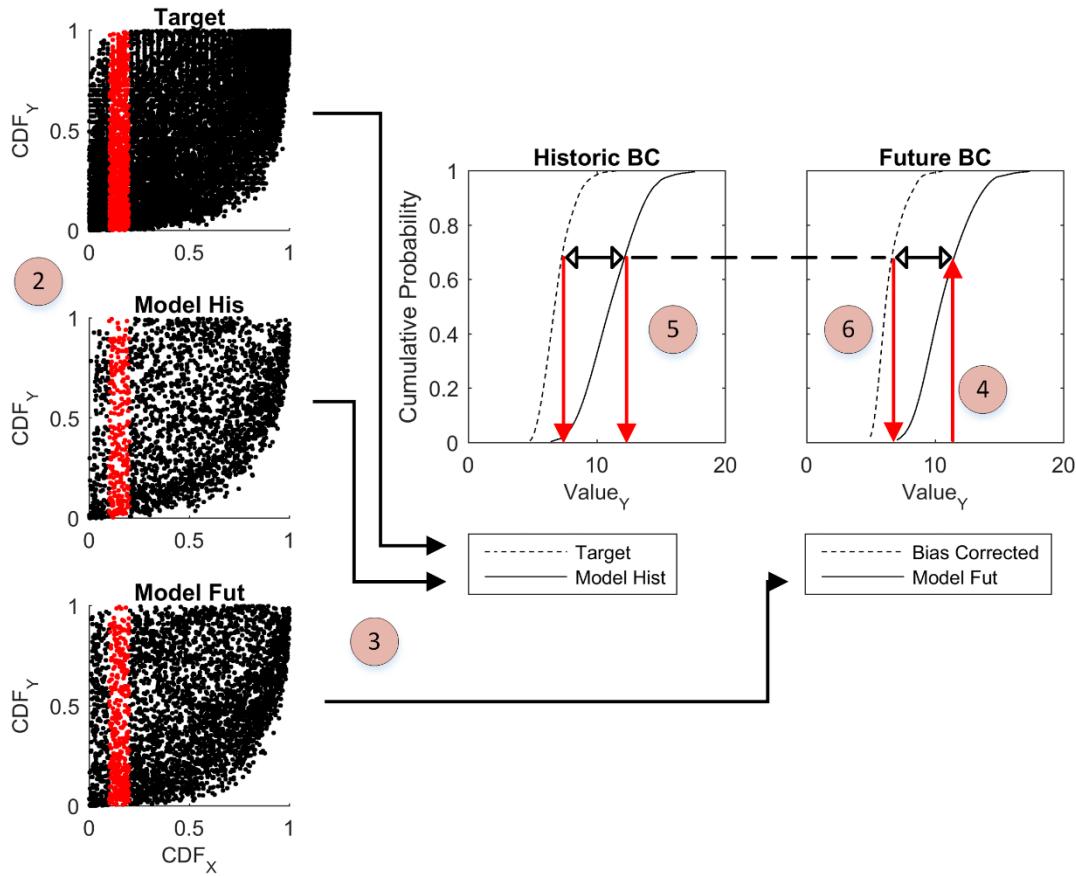


Figure 37: Binning Method using equidistant quantile mapping

The Direct Method (Future)

The future extension for the Direct method is detailed in Figure 38. The overarching goal of the methodology is to find the components necessary for an equidistant quantile mapping: the quantile of the model historic and target datasets (associated by correlation) so that we can apply the difference to the model future quantile of interest. These two quantiles are found using two standard Direct methods applied to their respective datasets. The first step is to bias correct the first variable (X) using a standard equidistant quantile mapping. This step is not shown in Figure 38. The second step (2) is to fit a copula to all three of the model future, model historic, and target datasets. This step is performed only once while all

following steps are repeated for each value in the model future dataset. The next step is to find the bivariate CDF value for the model future copula corresponding the CDF value for model future variable X and Y (**3**). This is done by finding the quantiles corresponding to the model future value pair and evaluating the copula at this point. At this point the two Direct methods diverge with the first branch using the target copula and the second branch using the model historic copula. This divergence corresponds to the goal of finding the quantile of model historic and target datasets. With these two copulas, the next step is to take a slice of the bivariate CDF at the known bias corrected value (variable X). As with the normal direct method (see Figure 6), the value of CDF_X is already known since it is the bias corrected variable X (step 1). If we slice the copula at this value (**4**) and use our knowledge of the bivariate copula CDF value of the model (evaluated in step **(3)**), we can find CDF_Y (**5**). At this point we found the quantiles of the model historic and target datasets associated with the model future Copula value. With these values, a normal equidistant quantile mapping method is performed by finding the ValueY corresponding to these quantiles and quantifying the difference between these values (**6**). Finally, we find the quantile of the model future value (**7**) and apply this difference to find the bias corrected value (**8**).

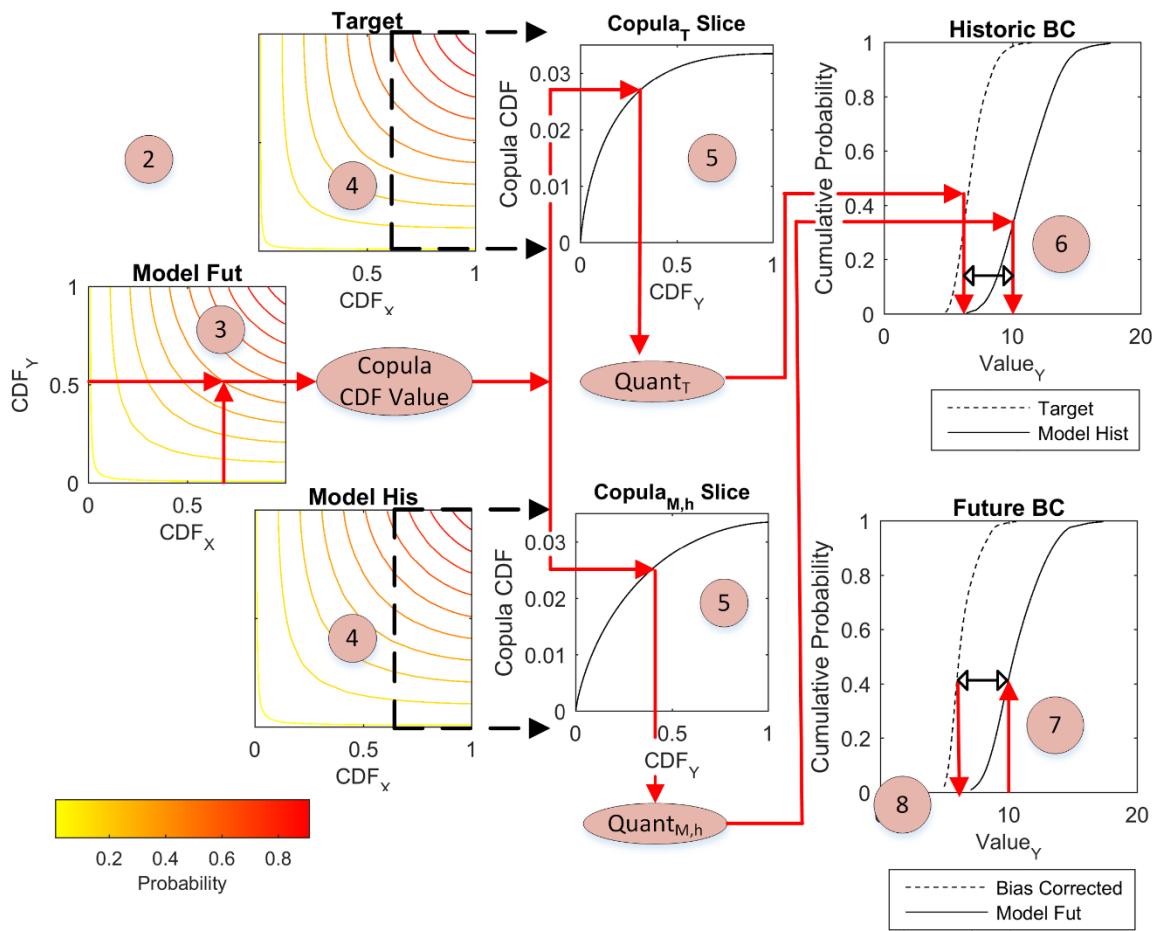


Figure 38: Direct Method using equidistant quantile mapping.