UNIVERSITY OF CALIFORNIA,
IRVINE

Preparing for what? Design floods and environmental change

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Civil Engineering

by

Adam Luke

Dissertation Committee:
Professor Brett F. Sanders, Chair
Professor Jasper A. Vrugt
Professor Amir AghaKouchak

2018

DEDICATION

This work is dedicated to friends and family whose support was truly incredible throughout this process. In particular, and in no particular order, I dedicate this work to my good friends Adam H., Hannah P., Olivia H., Ethan M., Jason C., Tanner A., Dylan B. & Dylan R., and all ten of the roommates who I've shared space with over the past five years. Our great times together are what made this experience remarkable.

I also dedicate this work to my entire family who has never been anything but encouraging and supportive. Thank you Mom and Dad for all of the time we spent on the phone during my commutes back and forth to Long Beach. They made the time fly :)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank all of the educators I have learned from over years, and in particular, the professors who influenced and guided my research. Dr. Alsdorf, thank you for introducing me to the research world and pointing me in promising directions. "Go west" was great advice. Dr. Judge, I am still applying the programming skills your group taught me all those years ago during the REU program. That experience gave me the confidence to apply for graduate programs in engineering. Dr. AghaKouchak, thank you for encouraging my research ideas and giving me the opportunities to publish in the engineering manuals. With your help, we might be able to get some of this work into practice :) Dr. Vrugt, thank you for pushing me. I may not always seem appreciative, but I am. Without your class on Bayesian analyses I don't think I would have been interested enough to commit to the phD program. And Dr. Sanders, thank you for being a great advisor for both research directions and career aspirations. I am forever grateful that you brought me in as the lead engineering researcher in the FloodRISE project without a background in engineering or a full commitment to the phD. I'm proud of the work we have accomplished over the past five years and I have a feeling that it will be the foundation of something bigger.

All of this work was funded by the National Science Foundation (grant DMS-1331611) whose support I gratefully acknowledge. I would also like to thank all of the co-authors, co-researchers, and focus group participants in the FloodRISE project who made this a success. The full authorship of the published work this dissertation is attributable to Luke et al. [2017] and Luke et al. [2018]. Thanks everyone!

CURRICULUM VITAE

Adam Luke

EDUCATION

Doctor of Philosophy in Civil Engineering — **2018**
University of California, Irvine — *Irvine, California*

Master of Science in Civil Engineering — **2015**
University of California, Irvine — *Irvine, California*

Bachelor of Science in Environmental Science — **2013**
The Ohio State University — *Columbus, Ohio*

RESEARCH EXPERIENCE

Graduate Research Assistant — **2013-2018**
University of California, Irvine Civil and Envr. Engineering — *Irvine, California*

Undergraduate Research Assistant — **2012**
University of Florida Center for Remote Sensing — *Gainesville, Florida*

Undergraduate Research Assistant — **2011 - 2013**
Ohio State University BYRD Polar Research Center — *Columbus, Ohio*

PRACTICE EXPERIENCE

Staff Engineer — **2014-2018**
Everest International Consultants — *Long Beach, California*

TEACHING EXPERIENCE

Teaching Assistant — **2013-2016**
University of California, Irvine — *Irvine, California*

REFEREED JOURNAL PUBLICATIONS

Going beyond the flood insurance rate map: Insights from flood hazard map co-production — **2018**
Natural Hazards and Earth System Sciences

The Influence of Hazard Maps and Trust of Flood Controls on Coastal Flood Spatial Awareness and Risk Perception — **2017**
Environment and Behavior

Predicting nonstationary flood frequencies: Evidence supports and updated stationarity thesis in the United States — **2017**
Water Resources Research

Integrating resident digital sketch maps with expert knowledge to assess spatial knowledge of flood risk: A case study of participatory mapping in Newport Beach, California — **2016**
Applied Geography

Dominant backscattering mechanisms at L-band during dynamic soil moisture conditions for sandy soils — **2016**
Remote Sensing of Environment

Communicating Flood Risk: Looking Back and Forward at Traditional and Social Media Outlets — **2015**
International Journal of Disaster Risk Reduction

Increased nuisance flooding along the coasts of the United States due to sea-level rise: Past and future — **2015**
Geophysical Research Letters

Hydraulic modeling of the 2011 New Madrid Floodway activation: a case study on floodway activation controls — **2015**
Natural Hazards

REFEREED CONFERENCE PUBLICATIONS

Utilizing complementarity of active/passive microwave observations at L-band for soil moisture studies in sandy soils — **July 2013**
IEEE International Geoscience and Remote Sensing

ABSTRACT OF THE DISSERTATION

Preparing for what? Design floods and environmental change

By

Adam Luke

Doctor of Philosophy in Civil Engineering

University of California, Irvine, 2018

Professor Brett F. Sanders, Chair

The United States experienced record-breaking losses from natural disasters in 2017. Damages from floods were particularly costly, largely because of the high exposure and value of capital at risk. Without targeted mitigation strategies, flood losses are expected to escalate due to urbanization and the projected changes to the frequency of extreme weather events. This dissertation provides and evaluates engineering methods for flood risk mitigation in an era of global change. The body of work contains three research articles: two provide and critique methodology for incorporating changing environmental conditions in the design of flood control infrastructure along rivers and coasts, and the third contributes guidance for improving the clarity and utility of commonly produced flood hazard maps. The results demonstrate that attempting to forecast the impact of changing environmental conditions on extreme floods substantially increases predictive uncertainty relative to traditional methods. Thus, so-called "non-stationary" methods have limited utility for decision making and are difficult to use as the basis for infrastructure design or mitigation planning. Our inability to meaningfully forecast changes in extremes over long time periods has important implications for engineering, but the first step is to acknowledge the unknown. Additional factors of safety, adaptable decision making, and better communication of what we actually do know are the keys to successfully reducing losses.

Chapter 1

Predicting non-stationary flood frequencies

1.1 Chapter Summary

Non-stationary extreme value analysis (NEVA) can improve the statistical representation of observed flood peak distributions compared to stationary (ST) analysis, but management of flood risk relies on predictions of out-of-sample distributions for which NEVA has not been comprehensively evaluated. In this study, we apply split-sample testing to 1,250 annual maximum discharge records in the United States and compare the predictive capabilities of NEVA relative to ST extreme value analysis using a log-Pearson Type III (LPIII) distribution. The parameters of the LPIII distribution in the ST and non-stationary (NS) models are estimated from the first half of each record using Bayesian inference. The second half of each record is reserved to evaluate the predictions under the ST and NS models. The NS model is applied for prediction by (1) extrapolating the trend of the NS model parameters throughout the evaluation period and (2) using the NS model parameter values at the end of

the fitting period to predict with an updated ST model (uST). Our analysis shows that the ST predictions are preferred, overall. NS model parameter extrapolation is rarely preferred. However, if fitting period discharges are influenced by physical changes in the watershed, for example from anthropogenic activity, the uST model is strongly preferred relative to ST and NS predictions. The uST model is therefore recommended for evaluation of current flood risk in watersheds that have undergone physical changes.

1.2 Introduction

Dynamic flood risk is common: both natural and anthropogenic influences can alter flood behavior. For example, the rapid development of watersheds and physical alterations to rivers and coastlines contributed to variable flood risk throughout the 20th century [Villarini and Smith, 2009, Peel and Blöschl, 2011, Vogel et al., 2011]. A warming climate increases the water holding capacity of the atmosphere which directly affects precipitation extremes and flood risk over time [Karl et al., 2009, Trenberth, 2011, Cheng and AghaKouchak, 2014]. Several studies have quantified societal exposure to current and intensified flooding in the 21st century [Hallegatte et al., 2013a, Hirabayashi et al., 2013, Milly et al., 2002, Apel et al., 2006], and in the absence of adaptation strategies project a staggering global loss on the order of 1 trillion US dollars per year by 2050 [Hallegatte et al., 2013a]. Indeed, the threat of escalating exposure requires new approaches for both flood risk management and characterization.

The most commonly applied method to characterize flood risk is known as flood frequency analysis (FFA), which estimates the recurrence rate of rare flooding events based on the annual exceedance probability of flood discharges. In the United States, engineers utilize FFA to determine the 100-year flood discharge, which is a design consideration for hydraulic structures and used to delineate areas subject to mandatory flood insurance. The 100-year

discharge is expected to occur or be exceeded once every 100 years, and has an estimated annual exceedance probability of 1%. In general, there are two commonly applied methods to estimate the exceedance probabilities necessary for riverine FFA [Centre for Ecology and Hydrology, 2008]. In the first approach, a computer model simulates the rainfall-runoff relationship by numerical solution of the governing hydrologic equations. The calibrated model then estimates the flood discharges for different rainfall events with established exceedance probabilities. The second and alternative approach does not rely on a numerical model, but uses only historic observations of flood discharges. This method is based on extreme value analysis (EVA), where a probability distribution is fitted to the record of flood discharges themselves. The fitted distribution is then used to estimate the exceedance probability of specific flood discharges. The latter method is preferred in practice if sufficiently long records of flood discharges are available [Centre for Ecology and Hydrology, 2008], and is therefore the subject of the present study.

Both of these commonly applied methods are built on the assumption of constant flood frequency with respect to time, which implies the statistics of extreme discharges (e.g., mean and/or standard deviation) are time-invariant. This so-called stationary (ST) assumption is quite restrictive, since a warming climate creates the possibility of unprecedented changes to observed flood regimes [Min et al., 2011, Kunkel et al., 2013], and anthropogenic watershed influences can alter the mean and variance of peak flood peaks [Villarini and Smith, 2009, Vogel et al., 2011]. The known limitations of ST FFA, along with the alarming societal impacts of intensified flooding, has given rise to the development and use of non-stationary (NS) approaches. This alternative FFA method does not rely on the ST assumption, but rather recognizes time variant flood frequencies.

Several different approaches have been published in the literature to handle the so-called "NS issue" in FFA [Olsen, 2002, Raff et al., 2009, Jain and Lall, 2001], and new national and state wide policies actually require the consideration of changing flood frequencies for planning,

design, and risk management [European Commission, 2007, Salas et al., 2012]. An excellent review of such methods applied in Europe was reported by Madsen et al. [2013]. The two NS FFA methods described in this review include (1) the use of precipitation projections from future climate scenarios in rainfall-runoff models and (2) the use of a safety margin to adjust the design flood estimates derived from ST extreme value analysis (SEVA). While countries such as Norway, United Kingdom, Belgium, and Germany have adopted the safety margin approach in engineering guidelines [Madsen et al., 2013], the underlying ST assumption is often challenged and relaxed in the scientific literature to allow application of NS extreme value analysis (NEVA) [AghaKouchak et al., 2013, Begueria and Vicente-Serrano, 2006, Begueria et al., 2011, Cheng et al., 2014a, Cooley, 2009, Gilleland and Katz, 2011, Griffis and Stedinger, 2007, Katz, 2010, Lopez and Frances, 2013, Salas and Obeysekera, 2014, Silva et al., 2015, Stedinger and Griffis, 2011a, Tramblay et al., 2014, Cheng et al., 2015, Villarini et al., 2009b, Jakob, 2013, Steinschneider and Lall, 2015]. Indeed, the review by Madsen et al. [2013] concludes on page 33 that moving towards a new NS framework, based on the use of NEVA, is an "important aspiration within the European hydrological science community."

NEVA is an important extension of SEVA which enables the parameters of the extreme value distribution to vary with time. This allows the parameters of the distribution to capture trends in the flood frequencies [Salas and Obeysekera, 2014]. In theory, these trends can be extrapolated to estimate, for example, the distribution of flood discharges in 2050. Numerous studies have shown that NEVA improves the statistical representation of historic hydroclimatic data [Lopez and Frances, 2013, Strupczewski et al., 2001, Villarini et al., 2009b], but the authors consistently caution against using NEVA for out-of-sample prediction. In this study, we apply split-sample testing to 1,250 annual maximum discharge records and compare the predictive capabilities of NEVA relative to SEVA using a log-Pearson Type III (LPIII) distribution. Specifically, we reserve the second half of each data record to evaluate predictions under the calibrated ST and NS models. We use two different approaches to

predict the out-of-sample data with the NS model, which are described along with the ST/NS models in Section 1.3. Section 1.4 includes the United States Geological Survey (USGS) records used in this study. The parameter estimation procedure is presented in Sections 1.5.1-1.5.1, where the parameters of the LPIII distribution in the ST and NS models are inferred from the first half of each record using Markov Chain Monte Carlo (MCMC) simulation with the DREAM$_{(ZS)}$ algorithm. The posterior parameter distributions and model fit metrics are compared in Section 1.5.1, along with a discussion of issues associated with NS model section. Section 1.5.2 outlines our application of Bayesian hypothesis testing which we use to evaluate, compare, contrast, and juxtapose the predictions made by the competing models. The results of this analysis are presented in Sections 2.4-2.5, where we assess (1) which of the three predictions, under the LPIII distribution, describe most accurately the distribution of the out-of-sample flood discharges, and (2) which diagnostics are useful for model selection. The paper concludes in Section 1.8 with recommended applications of NEVA for prediction of flood discharges in the LPIII distribution. Appendix A includes the mathematical definition of the LPIII distribution, and the supporting information includes a MATLAB® program based on the methods outlined in Sections 1.5.1-1.5.1 which can be used for inference of the ST/NS LPIII model parameters.

1.3 Stationary or Non-Stationary?

FFA in the United States is conducted according to the guidelines outlined in hydrologic Bulletin 17B [Interagency Committee on Water Data, 1982], which are built on the ST assumption. NS extensions of the standardized FFA guidelines in the United States have been considered in several studies [Stedinger and Griffis, 2011a, Griffis and Stedinger, 2007]. For instance, Stedinger and Griffis [2011a] use a time-dependent mean to account for NS in the Pearson Type III (PIII) distribution used widely for FFA in the United States and

Australia [Pilgrim, 2001]

$$\log_{10}(Q_t) \sim \text{PIII}(\mu_t, \sigma, \gamma) \tag{1.1}$$

where Q_t signifies the annual maximum discharge, t denotes time, and μ_t, σ, and γ characterize the mean, standard deviation, and skewness of the PIII distribution, respectively. The $\log_{10}$-transformation stabilizes the variance of the annual peak discharges and makes the transformed data more amenable to the PIII distribution. We can assume a simple linear trend for the mean of the PIII distribution

$$\mu_t = \mu_o + \alpha t \tag{1.2}$$

where μ_o signifies the intercept and α is the slope of the (t, μ_t) relationship. The value of μ_o is equivalent to the PIII mean of the first year of the discharge record. If flood peaks are assumed to be stationary then $\alpha = 0$ suffices in Equation 1.2 and the resulting model simplifies to the ST PIII distribution. Joint Bayesian inference of the PIII parameters $\{\sigma, \gamma\}$ and latent variable α from a n-record, $\mathbf{Q} = \{Q_1, \ldots, Q_n\}$, of observed annual peak discharges, would help judge which statistical model, NS or ST, would receive most support from the data. This would require knowledge of the marginal likelihood, or model evidence, which integrates model accuracy, uncertainty, and complexity. Appendix A lists analytic expressions for the probability density and cumulative distribution functions of the PIII distribution. We adopt the notation LPIII to clarify the use of log-transformed discharge data, and refer to Equation 1.1 as ST LPIII model ($\alpha = 0$) or NS LPIII model ($\alpha > 0$). For clarity, we use μ to denote the time invariant mean of $\log_{10}(\mathbf{Q})$ in the ST LPIII model.

The use of a linear time-dependency of the LPIII mean in Equation 1.1 is rather convenient and simplistic, but has several important implications. First, a trend in the logarithmic mean of Q_t equates to an exponential trend in the mean of the flood peaks [Read and Vogel, 2016].

This assumption may seem questionable but is supported by analysis of flood trends in the United States [Vogel et al., 2011] and United Kingdom [Prosdocimi et al., 2014]. Second, despite the use of a constant value for σ, the standard deviation of the arithmetic flood peaks will increase with the logarithmic mean, μ_t, of the LPIII distribution. This intrinsic property is desirable since nonstationarity would likely not only alter the mean of the flood peaks but simultaneously also affect their associated dispersion. Third, Equation 1.1 does not provide guidance on the governing physical factors, climate signals, and/or anthropogenic causes that may explain temporal changes in the frequencies of the annual flood peaks. Fortunately, the parameter estimation methodology presented herein supports the use of much more advanced (t, μ_t) relationships with covariates other than time. Examples include changes in land-use, urbanization, and variations in the North Atlantic Oscillation (NAO) and El Nino Southern Oscillation (ENSO). Yet, these covariates are not always readily available, nor may they correlate sufficiently with flooding events [Archfield et al., 2016]. We, therefore, resort herein to a purely statistical description of annual discharge peaks and use time as a proxy and explanatory variable of nonstationarity. In fact, the LPIII model used herein (with $\alpha > 0$) is analogous to the simplest of NS models presented by Stedinger and Griffis [2011a].

Figure 1.1 demonstrates the additional flexibility attained through NEVA relative to SEVA. Figure 1.1A shows the affect of γ on Equation 1.1, while Figure 1.1B shows the affect of α on Equation 1.1. Notice the traditional ST LPIII model represents different tailing behavior as γ changes, but the distribution of $\log_{10}(\mathbf{Q})$ must remain fixed in time. Indeed, the NS LPIII model includes only one additional parameter, α, but adds the dimension of time to our EVA. The NS model can even behave as a ST distribution, when α is zero. However, it is not clear if predictions of flood peak distributions under a NS model are an improvement relative to ST predictions, despite the additional flexibility. The primary concern is whether the NS signal is clear enough to accurately estimate trend parameters, such as α in this study. Moreover, different NS models will lead to different predictions, and there are multiple approaches for prediction under a single NS model. This raises the question of whether we actually do

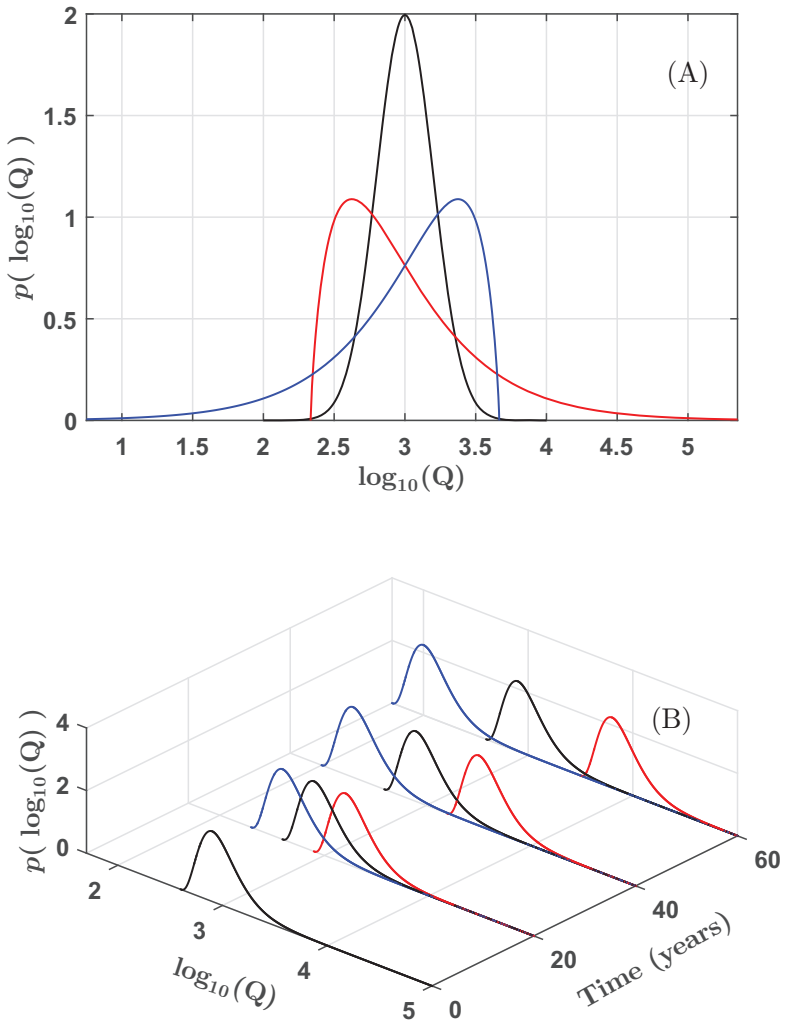

Figure 1.1: (A) ST LPIII distribution with variable shape parameter. The *red*, *black*, and *blue* lines correspond to γ values of 1.5, 0 , and -1.5, respectively. (B) NS LPIII distribution with variable trend parameter. The *red*, *black*, and *blue* lines correspond to α values of 0.015, 0, and -0.015, respectively. In (B), notice that the tail of the *black* distribution becomes closer to the mode of the *red* distribution over time.

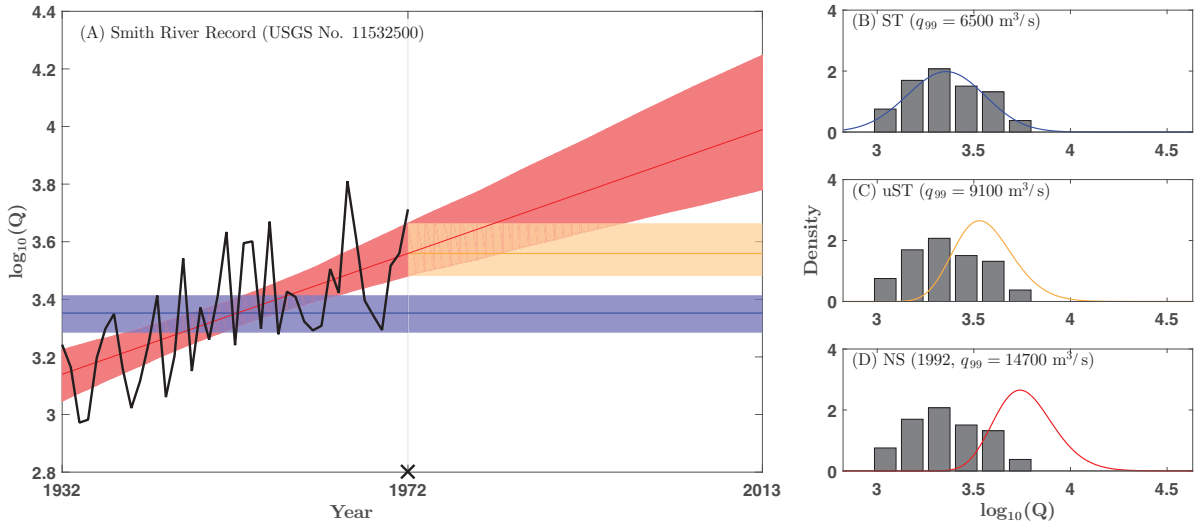

Figure 1.2: (A) Maximum a posteriori (MAP) estimate of the location parameter (mean of $\log_{10}(\mathbf{Q})$) under the ST (*blue* line), uST (*gold* line), and NS (*red* line) models inferred from the Smith River fitting period $\mathbf{Q}$ (*black* line). The colored shading represents the respective 95% credible intervals of the location parameters, and the *black* cross denotes the end of the fitting period. (B - D) Predictions of out-of-sample density under the ST (*blue* line), uST (*gold* line), and NS (*red* line) models derived from the MAP parameter estimates. The *black* histograms represent the empirical density of the fitting period. Notice that (C - D) show predictions under the uST and NS models moving *away* from the observed density, and the 95% credible intervals are wider under the NS and uST models relative to the ST model.

better by including NS in our analysis [Stedinger and Griffis, 2011a].

To illustrate these issues, Figure 1.2A shows the ST and NS LPIII models fit to the first half of the annual maximum discharge record from the Smith River near Crescent City, CA (for details see Section 1.5). Here, we use two different approaches to predict the out-of-sample data with the NS model. Under the first approach, the trend of the NS model parameters is extrapolated throughout the evaluation period, which represents the entirely NS prediction. In the second approach, the NS model parameter values at the end of the fitting period are used to predict with a ST model, or the ST parameters are "updated" by the NS model (hereafter denoted uST). Including the traditional ST prediction leads to three different representations of the out-of-sample density, which are shown relative to the in-sample density in Figure 1.2B-D. During the fitting period, the annual maximum discharge

of the Smith River exhibits a detectable trend at the 0.05 significance level (Mann Kendall trend test [Mann, 1945, Kendall, 1976]), which is reflected by the trend in μ_t. In the NS case, μ_t has substantially increased by the end of the fitting period, indicating that the mean and scale of the annual maximum discharges have changed throughout the record. In the ST case, μ remains constant. The difference in μ between the two approaches is certainly significant from a planning and engineering design perspective. The 99^{th} quantile of the ST LPIII distribution (q_{99}) is used to designate the special flood hazard area and design critical infrastructure such as bridges and levees in the United States [Federal Emergency Management Agency, 2009, Interagency Committee on Water Data, 1982]. Under the NS LPIII model at the end of the fitting period (or uST model), q_{99} is equivalent to 9,100 m^3/s, which is 40% greater than the ST estimate (Figure 1.2B-D). Assuming that flood frequencies are indeed non-stationary, and the distribution of annual maximum discharges continues to change according to the best fit NS model parameters, the value of q_{99} under the NS model is 14,700 m^3/s by the year 1992 and 2.3 times larger than its counterpart of the ST model.

The use of the NS LPIII model for cases when flood frequencies were actually ST can lead to overestimation of flood peaks, and hence over-engineered structures. Even worse, falsely selecting the ST model could lead to under-designed infrastructure and potential disaster. This decision would be difficult in practice, in the absence of guidance on NS model selection and lack of recommendations for FFA in case of a trend in the historic record. Moreover, it is unclear how to use a NS model for prediction. Is it better to assume flood frequencies will continue to change according the trends inferred from the historic record, or should we simply use the most recent NS parameterization for prediction? These issues motivate the split-sample testing used herein, where the predictions of the calibrated ST, uST, and NS models are evaluated using Bayesian hypothesis testing.

1.4 Stream Gage Records

We use a large data-set of annual maximum discharge records to evaluate the predictive abilities of the ST, uST, and NS models. The majority of records used in this study originate from the set of 1861 stream gages which were selected for potential use in the Model Parameter Estimation Project (MOPEX) [Schaake et al., 2006, Slack and Landwehr, 1992, Wallis et al., 1991]. The gages were believed to be unaffected by upstream regulation and watershed development at the time the dataset was compiled (1991/1992), suiting the needs of MOPEX. Since then, many records have been anthropogenically influenced. We rely on USGS data flags to identify influenced records, specifically flags 5 and 6 to identify records affected by upstream regulation and diversion, and flag C to identify records affected by land use changes or channelization. In this study, watersheds that have experienced anthropogenic influences are included in the analysis since the ST assumption is likely violated under these circumstances. An additional 25 watersheds affected by land use changes or channelization were added to the record pool since relatively few gages of this category were included in the original MOPEX dataset. Records with fewer than 60 observations were excluded to ensure the availability of at least 30 observations in the fitting and evaluation period.

We denote the full record of measured annual maximum discharges, $\mathbf{Q}$, and the fitting period data, $\mathbf{X}$. The fitting period data includes the first half of measured peak discharges, $\mathbf{X} = \log_{10}(\{Q_1, ..., Q_n\})$, where Q denotes an observation within the full record, $n = \lceil N/2 \rceil$, and N is the number of observations in the full record. The evaluation data, $\mathbf{X}^*$, includes the remaining observations, or $\mathbf{X}^* = \log_{10}(\{Q_{n+1}, ..., Q_N\})$. Record splitting in this manner resulted in variable fitting period lengths, with 40 observations representing the average data availability for parameter inference. In this study, historic peaks were removed from the record. Also, records with no-flow observations (i.e. zero discharge) were excluded from the analysis to simplify the use of the LPIII distribution. This leads to a data set of 1,250 annual maximum discharge records. The locations of tested records are shown in Figure 1.3A,

while the fitting period characteristics are shown in Figure 1.3B-D. Trend analysis shown in Figure 1.3 was conducted according to the Mann-Kendall test for monotonic trends [Kendall, 1976, Mann, 1945]. Hereafter, a "significant trend" refers to a rejection of the null hypothesis of no trend by the Mann-Kendall trend test at the 0.05 significance level. We acknowledge that there are no good statistical reasons to omit historic peaks and records with zero flow observations from the data-set [Reis Jr. and Stedinger, 2005, Interagency Committee on Water Data, 1982], yet the focus of this study is on assessing predictions made under ST, uST, and NS models fit to a systematic record.

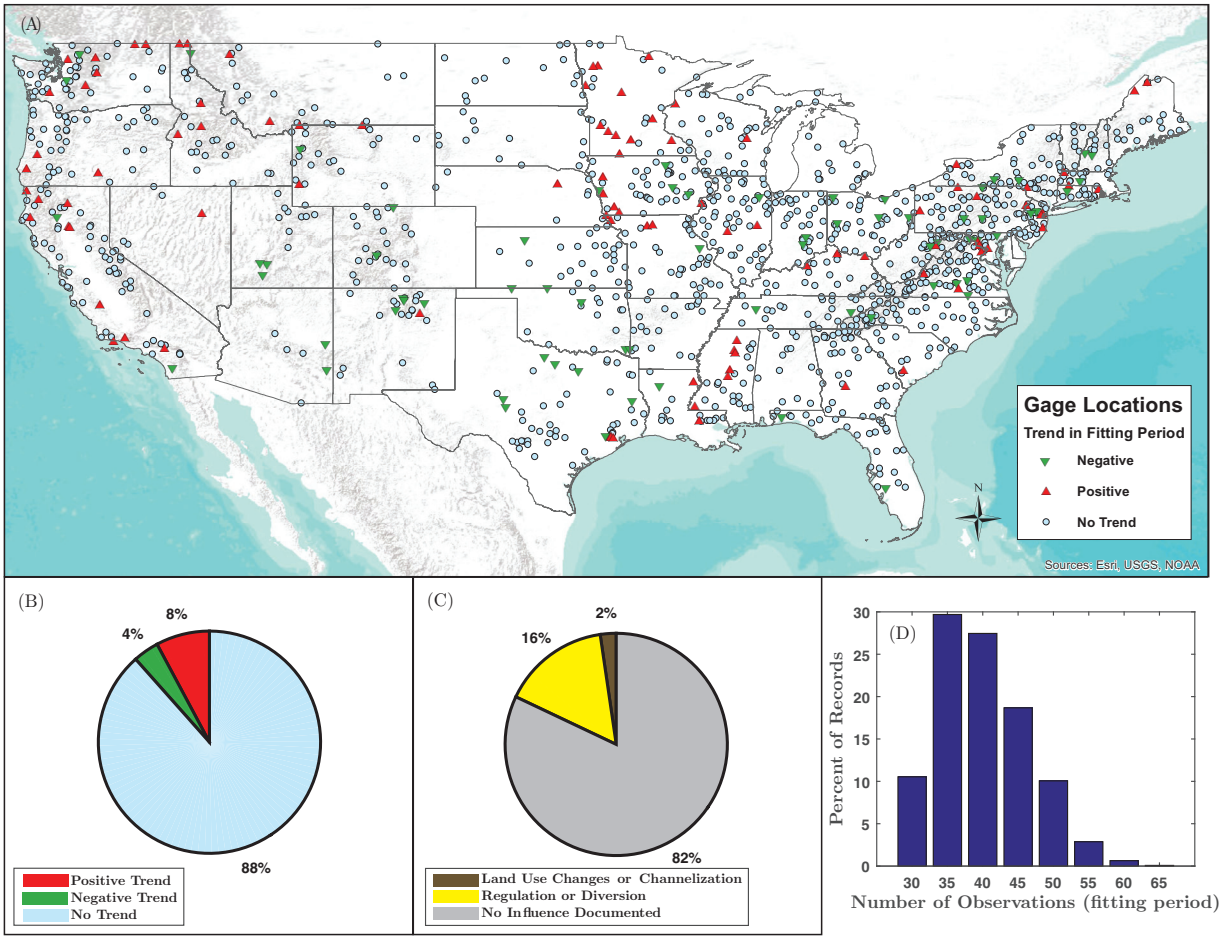

Figure 1.3: (A) Location of annual maximum discharge records and trends detected according to the Mann-Kendall trend test at the 0.05 significance level. (B) Fitting period trend characteristics as a percentage of all records tested. (C) Anthropogenic watershed influences as a percentage of all records tested. Land use changes or channelization refers to USGS record flag "C", while regulation and diversion refers to USGS record flag 5 or 6. (D) Fitting period length as a percentage of all records tested.

1.5 Bayesian Inference

1.5.1 Parameter Estimation

Estimating the parameters of an extreme value distribution is important for reliable frequency analysis. This process is further complicated by the introduction of NS models, where structural-trend parameters must also be inferred from limited data. In this study, we are primarily interested in inferring the ST LPIII parameters, $\boldsymbol{\theta}_{\mathrm{s}} = \{\mu, \sigma, \gamma\}$, and the NS LPIII parameters, $\boldsymbol{\theta}_{\mathrm{n}} = \{\mu_o, \sigma, \gamma, \alpha\}$, where $\mu_t = \mu_o + \alpha t$. While there are certainly other viable parameterizations of the NS LPIII model, we do not consider NS models describing changes in σ and γ because it is unlikely that these trends can be estimated from common record lengths [Yu et al., 2015]. We also do not consider trend models more complex than the simple linear model, due to the uncertainty introduced by additional trend parameters. Uncertainty introduced by NS model complexity has been reported as a major drawback to NEVA, which is the primary argument against widespread application made by Serinaldi and Kilsby [2015a]. Therefore, any meaningful comparison between flood discharge predictions based on ST and NS models must account for parameter uncertainty introduced by the NS model. The issue of parameter uncertainty motivates the Bayesian approach for parameter inference and model comparison. Bayesian parameter estimation results in distributions of parameter values rather than point estimates, which is especially useful for hypothesis testing and uncertainty quantification. Thus, we apply Bayesian parameter estimation via MCMC and the DREAM$_{(\mathrm{ZS})}$ algorithm for robust uncertainty quantification and model comparison.

Under the Bayesian approach for parameter estimation, Bayes' theorem is used to calculate the posterior density of parameter values

$$p(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j) = \frac{p(\boldsymbol{\theta}_j|\mathcal{M}_j)L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)}{p(\mathbf{X}|\mathcal{M}_j)} \tag{1.3}$$

where $p(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$ represents the posterior density of parameter vector, $\boldsymbol{\theta}_j$, given the fitting period data, $\mathbf{X}$, and the model of interest, $\mathcal{M}_j$. The subscript j denotes competing model classes, and in this study $j = \{\text{“s”}, \text{“n”}, \text{“u”}\}$, which stand for the ST, NS, and uST models, respectively. Here, $p(\boldsymbol{\theta}_j|\mathcal{M}_j)$ represents the prior density of $\boldsymbol{\theta}_j$ given $\mathcal{M}_j$, and $L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$ represents the likelihood of parameter values $\boldsymbol{\theta}_j$, given $\mathbf{X}$ and $\mathcal{M}_j$. The denominator, $p(\mathbf{X}|\mathcal{M}_j)$, is known as the marginal likelihood or evidence and acts as a normalization constant so that $p(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$ integrates to one. The evidence is obtained through integration over the parameter space

$$p(\mathbf{X}|\mathcal{M}_j) = \int p(\boldsymbol{\theta}_j|\mathcal{M}_j) L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j) d\boldsymbol{\theta}_j \tag{1.4}$$

where $p(\mathbf{X}|\mathcal{M}_j)$ also quantifies the probability of seeing the data that were actually observed under the competing models. Therefore, the evidence is particularly useful for model selection, because the model which maximizes Equation 1.4 is a better descriptor of the data. Bayesian hypothesis testing and the evidence will be discussed further in Section 1.5.2. However, Bayesian parameter estimation does not necessarily require the value of $p(\mathbf{X}|\mathcal{M}_j)$, since parameter inference can be made from the un-normalized density

$$p(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j) \propto p(\boldsymbol{\theta}_j|\mathcal{M}_j) L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j) \tag{1.5}$$

when MCMC methods are used to sample from the posterior, as applied herein. For now, we are interested in using Equation 1.5 to estimate the un-normalized posterior density of both $\boldsymbol{\theta}_\text{s}$ and $\boldsymbol{\theta}_\text{n}$ given the first half of annual maximum discharge records. Equation 1.5 is not directly applied to infer $\boldsymbol{\theta}_\text{u}$, since the parameters of the uST model are derived from $p(\boldsymbol{\theta}_\text{n}|\mathbf{X}, \mathcal{M}_\text{n})$. In most practical applications, estimating $p(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$ is accomplished by sampling from the posterior after specifying $p(\boldsymbol{\theta}_j|\mathcal{M}_j)$ and $L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$.

Likelihood Function (Fitting Period)

Conceptually, $L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$ represents how plausible a parameter combination is after observing $\mathbf{X}$, which quantifies the information content of the data. Specifically, $L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$ is the probability density of $\mathbf{X}$ according to $\mathcal{M}_j$ with parameterization $\boldsymbol{\theta}_j$. For the parameter inference procedure, the likelihood function is defined as

$$L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j) = \prod_{i=1}^{n} p(X_i|\boldsymbol{\theta}_j, \mathcal{M}_j) \tag{1.6}$$

where n is the number of observations in the fitting period, and X_i represents the i^{th} observation. In this study, $p(X_i|\boldsymbol{\theta}_j, \mathcal{M}_j)$ is obtained by evaluating an analytic expression of the PIII probability density function (Equation A.1 or A.3) at observation $X_i = \log_{10}(Q_i)$ and parameterization $\boldsymbol{\theta}_j$ under $\mathcal{M}_j$. Here, $j = \{\text{“s”}, \text{“n”}\}$, which distinguishes Equation 1.6 applied to determine the likelihood of the ST and NS model parameters, respectively.

Under $\mathcal{M}_\text{s}$, the parameter values within $\boldsymbol{\theta}_\text{s} = \{\mu, \sigma, \gamma\}$ are constant with respect to time, and therefore Equation 1.6 is evaluated at a fixed parameterization of the PIII density function for $i = \{1, ..., n\}$. However, this is not the case under $\mathcal{M}_\text{n}$. When Equation 1.6 is applied to calculate the likelihood of $\boldsymbol{\theta}_\text{n} = \{\mu_o, \sigma, \gamma, \alpha\}$, the parameterization of the PIII density functions changes with $i = \{1, ..., n\}$, according to the linear trend model. Therefore, at each observation X_i, there is a unique parameterization of the PIII density function

$$\boldsymbol{\theta}_\text{n}^i = \{\mu_o + \alpha t_i, \sigma, \gamma\} \tag{1.7}$$

where t_i is the time (in years) between the first and the i^{th} observation. In this study, t_i is considered the covariate, since $\mathcal{M}_\text{n}$ parameter values depend on and change with t_i. For a more general description of covariate modeling see Renard et al. [2013], Coles [2001]. There are several properties of $\mathcal{M}_\text{s}$ and $\mathcal{M}_\text{n}$ with respect to the likelihood function that are worth

noting.

First, $\mathcal{M}_{\mathrm{s}}$ is a special case of $\mathcal{M}_{\mathrm{n}}$ when $\boldsymbol{\theta}_{\mathrm{n}}^{i}$ is constant or $\alpha = 0$. For records without a strong trend, Equation 1.6 will be maximized near $\alpha = 0$ under $\mathcal{M}_{\mathrm{n}}$, and therefore the parameter values of $\boldsymbol{\theta}_{\mathrm{n}}$ will approach $\boldsymbol{\theta}_{\mathrm{s}}$. In this scenario, α becomes a nuisance parameter, and $\mathcal{M}_{\mathrm{n}}$ is over-parameterized. This suggests that if α is inferred using maximum likelihood, there must be a time dependant trend in $\mathbf{X}$ for $\mathcal{M}_{\mathrm{n}}$ to significantly differ from $\mathcal{M}_{\mathrm{s}}$. Secondly, assuming that a trend is observed in $\mathbf{X}$, $\mathcal{M}_{\mathrm{n}}$ has a much greater ability to maximize Equation 1.6 relative to $\mathcal{M}_{\mathrm{s}}$. This is because in the NS case, the location of the distribution changes with the trend in $\mathbf{X}$, enhancing the probability density near the observations relative to the fixed μ under $\mathcal{M}_{\mathrm{s}}$. The NS model's ability to maximize Equation 1.6 relative to the ST model is problematic for comparison of $\mathcal{M}_{\mathrm{n}}$ and $\mathcal{M}_{\mathrm{s}}$ based on likelihood metrics alone, since likelihood based metrics will almost certainly favor $\mathcal{M}_{\mathrm{n}}$ for records with detected trends.

Prior Distributions (Fitting Period)

While $L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$ represents the plausibility of $\boldsymbol{\theta}_j$ after considering the data, $p(\boldsymbol{\theta}_j|\mathcal{M}_j)$ represents knowledge of $\boldsymbol{\theta}_j$ before the data is considered. Often, $p(\boldsymbol{\theta}_j|\mathcal{M}_j)$ is referred to as the prior. The prior represents an analyst's knowledge of $\boldsymbol{\theta}_j$ before any data is collected, which is described mathematically by the joint probability of the parameters within $\boldsymbol{\theta}_j$. The formal inclusion of prior knowledge about $\boldsymbol{\theta}_j$ in the parameter inference process is unique to Bayesian inference. The choice of priors is subjective, but it allows the analyst to use information other than the data for parameter inference.

While several studies have specified highly informative joint-priors for hydrologic data through regional analysis or eliciting expert opinion [Behrens et al., 2004, Perreault et al., 2000, Renard et al., 2006b], we decided to employ relatively uninformative priors so that inferred values of $\boldsymbol{\theta}_{\mathrm{s}}$ are similar to parameterizations produced following standard practice in the

Table 1.1: Model parameters and prior distributions

$\mathcal{M}_n$ Parameters	Scale	Units	Prior	a	b
μ_o	$\log_{10}$	m^3/s	$\mathcal{U}(a,b)$	-10	10
σ	$\log_{10}$	m^3/s	$\mathcal{U}(a,b)$	0	2
γ	$\log_{10}$	–	$\mathcal{N}(\gamma_r, SD_{\gamma_r})$	–	–
α	$\log_{10}$	$m^3/(s \cdot year)$	$\mathcal{U}(a,b)$	-0.15	0.15
$\mathcal{M}_s$ Parameters	Scale	Units	Prior	a	b
μ	$\log_{10}$	m^3/s	$\mathcal{U}(a,b)$	-10	10
σ	$\log_{10}$	m^3/s	$\mathcal{U}(a,b)$	0	2
γ	$\log_{10}$	–	$\mathcal{N}(\gamma_r, SD_{\gamma_r})$	–	–

United States. In the United States, μ and σ are determined entirely through consideration of $\mathbf{X}$, while γ is based on a regional and site specific estimate [Interagency Committee on Water Data, 1982]. An analogous parameterization can be produced in a Bayesian framework by using uninformative priors on μ and σ with an informative prior specified on γ.

For uninformative, yet proper priors, the marginal priors for μ, μ_o, σ and α were specified as a uniform distribution, $\mathcal{U}(a,b)$, where a and b represent the lower and upper bounds of the uniform priors, respectively. Our primary reason for using bounded priors is algorithmic efficiency. Indeed, such prior enhances considerably the convergence speed of the DREAM$_{(ZS)}$ algorithm (Section 1.5.1), a particularly important consideration in the face of the relatively large number of data records used in the present study. Table 1.1 lists the selected bounds on the uniform priors, which were chosen based on physical reasoning. For example, the bounds on μ restrict the location of the distribution to values between 10^{-10} m^3/s and 10^{10} m^3/s, which adequately encompass realistic values of μ. We confirmed that none of the posteriors were truncated by the prior bounds following the fitting procedure.

Selecting a distribution that represents a lack of knowledge is challenging and subjective, and the bounded uniform distribution is certainly not the only viable uninformative prior. For example, Reis Jr. and Stedinger [2005] use Jeffrey's prior as an uninformative prior on the scale parameter, and a normal a distribution with a large variance for an uninformative prior

describing the location parameter. Both approaches will lead to inference of the location and scale parameters based on the likelihood of $\mathbf{X}$, or the information content of the data alone. In contrast, we use an informative prior on γ because information other than $\mathbf{X}$ is typically available and used to estimate shape parameters of hydrologic variables.

Since γ controls the tail behavior and extrapolation to the unobserved quantiles of the LPIII distribution, it is difficult to determine from a single record. Therefore, estimates of γ are usually based on a site specific and regional approximation. Bulletin 17B approximates γ by weighting the station estimate, γ_{s}, with a regional estimate, γ_{r}. In the Bulletin 17B guidelines, γ_{s} is determined by calculating the sample skew of $\mathbf{X}$, and γ_{r} is typically taken from an isocline map of regional skews. The weights on γ_{s} and γ_{r} are inversely proportional to their respective mean square errors. A Bayesian adaptation of this weighting process can be accomplished for $\mathcal{M}_{\mathrm{n}}$ and $\mathcal{M}_{\mathrm{s}}$ through an informative marginal prior

$$p(\gamma|\mathcal{M}_j) \sim \mathcal{N}(\gamma_{\mathrm{r}}, \mathrm{SD}_{\gamma_{\mathrm{r}}}) \tag{1.8}$$

where $\mathrm{SD}_{\gamma_{\mathrm{r}}}$ is the root mean-square error of the regional skew, γ_{r}, and $\mathcal{N}$ denotes the normal distribution. Values of γ close to the regional estimate are assigned a high prior probability, and thus the prior reflects regional knowledge of γ. In this study, γ_{r} was obtained for each record by extracting values of γ_{r} from a digitized version of the Bulletin 17B Plate 1 generalized skew map. Here, $\mathrm{SD}_{\gamma_{\mathrm{r}}}$ was set equal to 0.55 per Bulletin 17B guidelines. Without considering the regional skews through the informative prior, posterior γ would be fairly different from γ estimated in practice. We also note that using an informative prior for only the shape parameter has been recommended for NS model parameter estimation in the generalized extreme value distribution (GEV) [El Adlouni et al., 2007], and this approach was previously designated the generalized maximum likelihood (GML) method in the ST case [Martins and Stedinger, 2000]. Ouarda and El-Adlouni [2011] apply the GML method for inference of NS GEV parameters in a Bayesian framework, which parallel the methods

applied in this study. The difference herein is that the generalized prior on γ has been adapted to incorporate regional information.

MCMC Simulation with the DREAM$_{(ZS)}$ Algorithm

Now that $p(\boldsymbol{\theta}_j|\mathcal{M}_j)$ and $L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$ have been defined, the posterior distribution of $\boldsymbol{\theta}_\mathrm{s}$ and $\boldsymbol{\theta}_\mathrm{n}$ can be approximated through sampling of the posterior. Sampling methods such as MCMC are required for most applications of Bayes' theorem, since $p(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$ is often analytically intractable. Essentially, MCMC methods provide a sample from the posterior distribution, and parameter inference is made from the generated sample. For a complete description of posterior sampling via Monte Carlo simulation, please see Section 2 of Vrugt [2016].

In this study, we use the DiffeRential Evolution Adaptive Metropolis algorithm (DREAM$_{(ZS)}$) to generate samples of the target distribution [Vrugt et al., 2009, Laloy and Vrugt, 2012a], and assess convergence through the $\hat{R}$-statistic of Gelman and Rubin [1992]. The MCMC procedure utilized three Markov chains, each including 8000 generations of $\boldsymbol{\theta}_j$. Fifty percent of the generations from each chain were discarded as burn-in, and the remaining generations were combined into a single matrix. This procedure results in a posterior sample size of 12,000 generations of $\boldsymbol{\theta}_j$ for each record and model. Parameter inference and model comparison is now based on the posterior samples of $\boldsymbol{\theta}_\mathrm{s}$ and $\boldsymbol{\theta}_\mathrm{n}$ produced by the MCMC algorithm. The parameters of the uST distribution, $\boldsymbol{\theta}_\mathrm{u} = \{\mu_n, \sigma, \gamma\}$, are derived from the posterior sample of $\boldsymbol{\theta}_\mathrm{n}$. Here, μ_n represents the parameter value of μ_t at the end fitting period under $\mathcal{M}_\mathrm{n}$. The distribution of μ_n was obtained by applying Equation 1.2 at each posterior sample of $\boldsymbol{\theta}_\mathrm{n}$, with t equal to the time between the first and the last year of the fitting period. The uST σ and γ parameters were taken from the posterior sample of $\boldsymbol{\theta}_\mathrm{n}$, and not $\boldsymbol{\theta}_\mathrm{s}$.

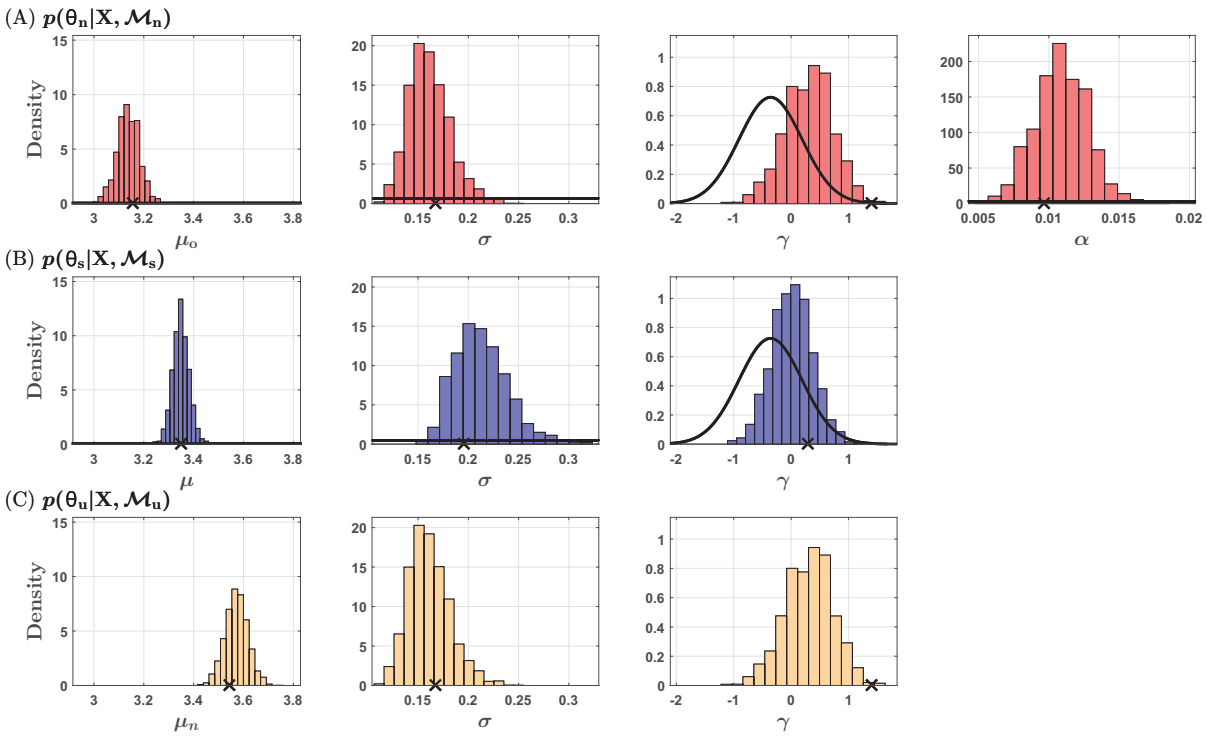

Figure 1.4: Marginal posterior distributions of $\boldsymbol{\theta}_{\mathrm{n}}$ under the NS model (*red* histograms), $\boldsymbol{\theta}_{\mathrm{s}}$ under the ST model (*blue* histograms), and $\boldsymbol{\theta}_{\mathrm{u}}$ under the uST model (*gold* histograms), given the data in the Smith River fitting period. *Black* lines represent the marginal prior distributions, and the *black* crosses show the MLE of $\boldsymbol{\theta}_j$ under the three models. Notice that the MLE of γ under $\mathcal{M}_{\mathrm{n}}$ is very unlikely according to the prior distribution, and posterior $\boldsymbol{\theta}_{\mathrm{n}}$ favors lower values of σ than posterior $\boldsymbol{\theta}_{\mathrm{s}}$.

Posterior Comparisons

We now return our attention to the Smith River data record, previously shown in Figure 1.2. Figure 1.4 presents the marginal posterior distributions of $\boldsymbol{\theta}_{\mathrm{n}}$ (top row), $\boldsymbol{\theta}_{\mathrm{s}}$ (middle row), and $\boldsymbol{\theta}_{\mathrm{u}}$ (bottom row) given the data in the Smith River fitting period. The prior density and maximum likelihood parameter estimates are represented by the black lines and black crosses, respectively. The maximum likelihood estimate (MLE) of $\boldsymbol{\theta}_j$, denoted $\hat{\boldsymbol{\theta}}_j$, is taken as the parameter combination from the posterior sample which maximized Equation 1.6. This Figure highlights several important results. Firstly, Figure 1.4 shows that the informative prior on γ causes the marginal posterior distribution to be centered between the MLE and the mode of the informative prior. Since the MLE of γ is the site specific skew, and the

prior mode is the regional skew, posterior γ represents a combination of the site specific and regional information. This result demonstrates the utility of informative priors for the inference of shape parameters.

Secondly, notice that the posterior under $\mathcal{M}_{\mathrm{n}}$ favors lower values of σ than the posterior under $\mathcal{M}_{\mathrm{s}}$. This difference can be explained by how the two models represent records that exhibit a strong trend. The apparent trend in the data, which is shown in Figure 1.2, is simply variance about the mean of the data in the context of a ST model. Therefore, the observable difference in the magnitude of the annual maximum discharges from the beginning to the end of the fitting period is accounted for through a relatively large value of σ. However, under $\mathcal{M}_{\mathrm{n}}$, the trend causes μ_t to increase throughout the fitting period. Thus, the time-dependant increase in the observed discharges is represented by a changing location parameter, not a large variance. In the context of $\mathcal{M}_{\mathrm{n}}$, σ does not describe the variance of the entire sample, but rather the variance about the changing location parameter. This concept is further illustrated in Figure 1.1B-D, where the ST distribution exhibits a variance large enough to encompass all of the data, but the NS distributions do not. The apparent difference in σ between the $\mathcal{M}_{\mathrm{n}}$ and $\mathcal{M}_{\mathrm{s}}$ has implications for inference of γ.

Thirdly, let us compare the marginal posteriors of γ between $\boldsymbol{\theta}_{\mathrm{s}}$ and $\boldsymbol{\theta}_{\mathrm{n}}$. While the posterior distribution of γ is relatively similar between the two models, the marginal distribution under $\mathcal{M}_{\mathrm{n}}$ includes more density away from the prior mode. This is due to the difference in the value of the MLE between the two models. Under $\mathcal{M}_{\mathrm{s}}$, the MLE of γ is closer to the regional estimate relative to $\mathcal{M}_{\mathrm{n}}$. Indeed, the MLE of γ under $\mathcal{M}_{\mathrm{n}}$ is found in the tail of the prior distribution. In other words, the MLE is very unlikely according to our prior knowledge. As a consequence, the NS posterior is centered further to the right. This suggests that the likelihood function under $\mathcal{M}_{\mathrm{n}}$ favors values of γ that are unrealistic according to past studies, which can be explained by the relatively low value of σ. Previous studies have also shown that σ and γ are inversely related in the LPIII distribution [Srikanthan and McMahon, 1981,

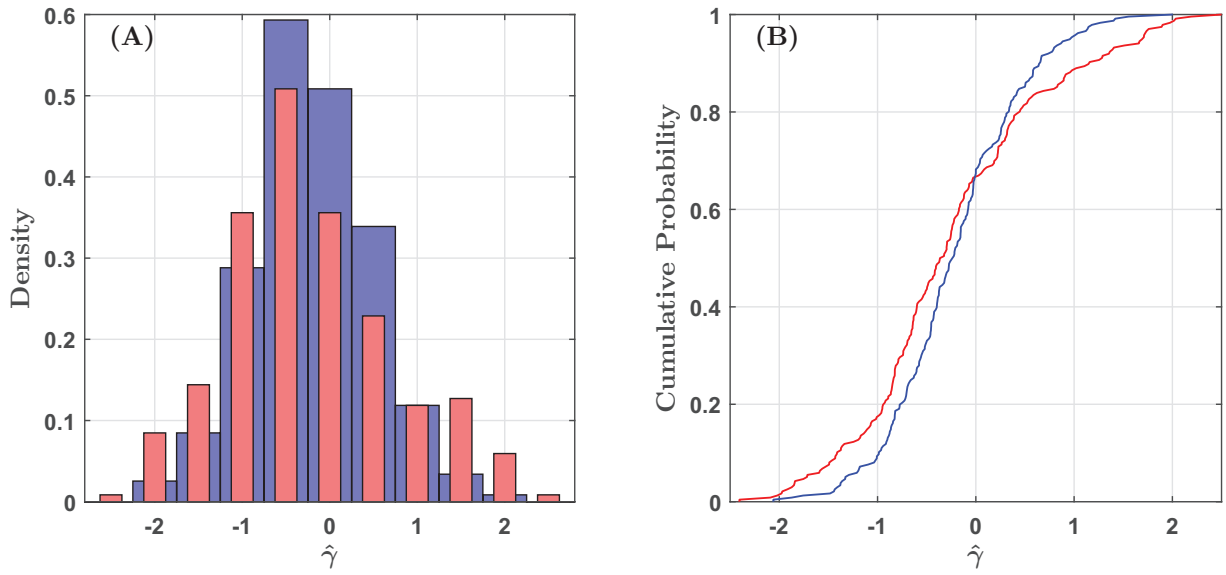

Figure 1.5: (A) Empirical density of $\hat{\gamma}$ under the NS model (*red* histogram) and the ST model (*blue* histogram) based on the estimates from fitting periods where $\hat{\alpha}$ was one standard deviation away from the mean of zero (236 total). The circumflex denotes the MLE of γ and α. (B) Empirical CDF of $\hat{\gamma}$ under the NS model (*red* line) and ST model (*blue* line). The CDF of $\hat{\gamma}$ under the NS model is different from the ST model counterpart, indicating that the likelihood function favors larger values of $\hat{\gamma}$ (i.e. ± 1.4) under the NS model. Indeed, there is a difference between the two CDFs at the 0.05 significance level according to the two sample Kolmogorov-Smirnov test [Massey Jr, 1951].

Nozdryn-Plotnicki and Watt, 1979], so in general low σ estimates will produce large skews. Let us explore this further by comparing the MLE of γ under $\mathcal{M}_{\text{n}}$ and $\mathcal{M}_{\text{s}}$ for other records tested.

Figure 1.5A shows the empirical density of $\hat{\gamma}$ under $\mathcal{M}_{\text{n}}$ (red histogram) and $\mathcal{M}_{\text{s}}$ (blue histogram) for estimates obtained from fitting periods with relatively large trends. Large trends are defined as $\hat{\alpha}$ values one standard deviation away from the mean value of zero, which were found for 236 fitting periods. Values shown in Figure 1.5 were produced by the procedure outlined in Section 1.5.1, except here the prior on γ was completely uninformative, or $p(\gamma|\mathcal{M}_j) \sim \mathcal{U}(-5, 5)$. This ensures that $\hat{\gamma}$ is adequately approximated, since the MCMC algorithm is no longer constrained by the regional estimate. We show how $\hat{\gamma}$ is distributed for records with large trends because the difference in the estimates under the two models

generally increase with increasing values of $\hat{\alpha}$. Figure 1.5B also compares the empirical cumulative distribution function (CDF) of $\hat{\gamma}$ under $\mathcal{M}_{\mathrm{n}}$ (red line) and $\mathcal{M}_{\mathrm{s}}$ (blue line), and reveals that there is a greater probability of large $\hat{\gamma}$ values (i.e. ± 1.4) when estimated under $\mathcal{M}_{\mathrm{n}}$ relative to $\mathcal{M}_{\mathrm{s}}$. Indeed, the difference between the CDFs under the two models is apparent at the 0.05 significance level, according to the two sample Kolmogorov-Smirnov test [Massey Jr, 1951]. While it appears that there is only a slightly higher probability of large $\hat{\gamma}$ values under $\mathcal{M}_{\mathrm{n}}$, extreme quantiles of the LPIII distribution are very sensitive to γ. Previous studies suggest γ should be with within ± 1.4 [Reis Jr. and Stedinger, 2005]. Therefore, estimation of $\boldsymbol{\theta}_{\mathrm{n}}$ should not be based solely on MLE in the LPIII model. Incorporating regional information through an informative prior is an attractive methodology to restrict the inference of γ to realistic values.

Lastly, let us discuss how the posterior samples derived from the MCMC procedure can be used for model comparison and selection. Several common model comparison metrics include the small sample Akaike Information Criterion ($\mathrm{AIC_c}$), the Bayesian Information Criterion (BIC), and the Deviance Information Criterion (DIC) [Burnham and Anderson, 2003, Spiegelhalter et al., 2002]. Under all three metrics, the goodness of fit term depends on $L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$, whereas each metric applies a different penalty for model complexity. The model which minimizes the metric value is preferred given the data in the fitting procedure, and the difference in the metric values between competing models is used for model selection. The differences in $\mathrm{AIC_c}$, BIC, and DIC values between $\mathcal{M}_{\mathrm{s}}$ and $\mathcal{M}_{\mathrm{n}}$ for the Smith River fitting period are 23.7, 22.5, and 22.1, respectively, each supporting $\mathcal{M}_{\mathrm{n}}$. Rules of thumb for interpreting these differences are given by Burnham and Anderson [2003]. Based on the listed differences of each model selection metric, we conclude that there is *essentially no empirical support* for $\mathcal{M}_{\mathrm{s}}$ relative to $\mathcal{M}_{\mathrm{n}}$ given the data in the Smith River fitting period. Similar conclusions are drawn for the other data records.

Figure 1.6 shows the differences in fit metric values between $\mathcal{M}_{\mathrm{s}}$ and $\mathcal{M}_{\mathrm{n}}$ for fitting periods

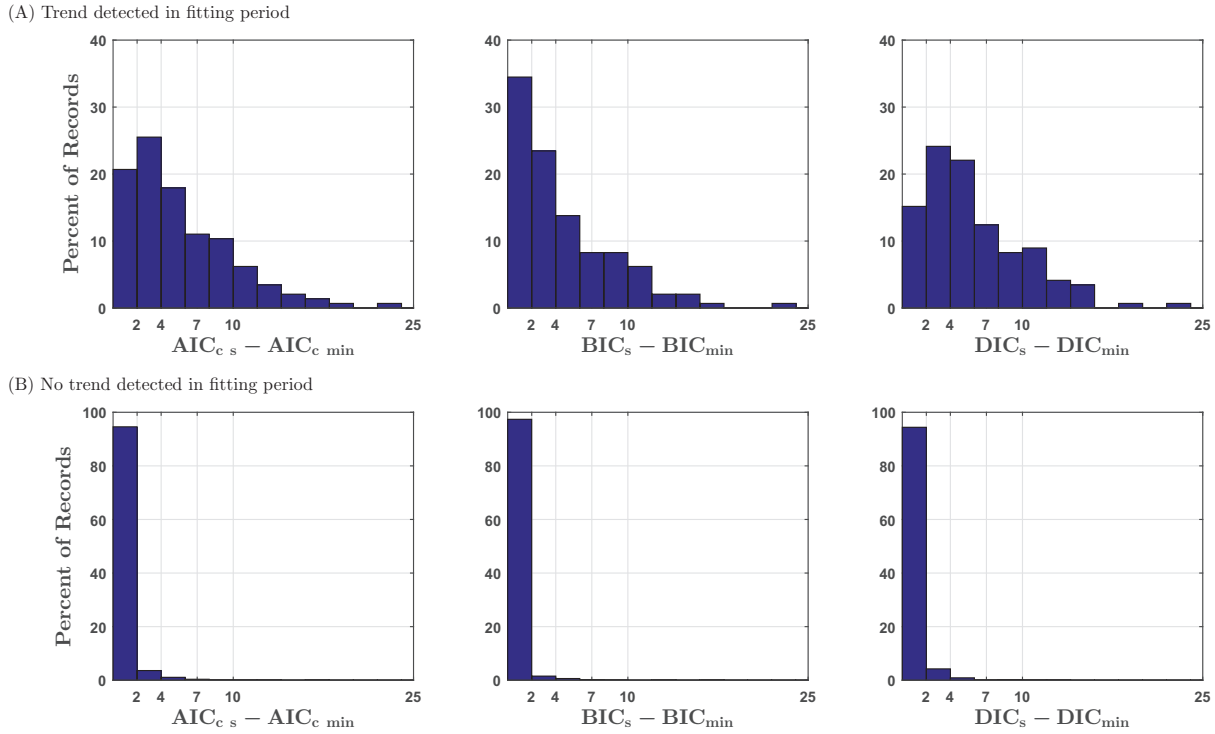

Figure 1.6: Comparison of model fit between the ST and NS models. Row (A) displays the differences in metric values for fitting periods with a detected trend (145 total), and row (B) displays the differences for fitting periods without detected trends (1105 total). Each column represents a different goodness of fit metric. The magnitude of the differences represents the level of empirical support for the ST model. Values from 0-2 in each panel indicate substantial support, values from 4-7 indicate considerably less, and values greater than 10 indicate essentially no empirical support the for ST model [Burnham and Anderson, 2003]. These results demonstrate that the presence of a trend substantially reduces the level of empirical support for the ST model relative to the NS model.

with detected trends (first row) and without detected trends (second row). Based on the differences presented in Figure 1.6, it is clear that the presence of a trend in the fitting period causes AIC_c, BIC, and DIC to significantly favor $\mathcal{M}_n$. Indeed, even for records without a significant trend, fit metrics can still indicate considerably less support for $\mathcal{M}_s$ relative to $\mathcal{M}_n$ because of the ST model's inferior ability to maximize the likelihood function. Model selection based solely on goodness of fit metrics may cause an analyst to sensibly select $\mathcal{M}_n$ for prediction given the presence of a trend in the data. However, goodness of fit metrics are primarily dependant on $L(\boldsymbol{\theta}_j|\mathbf{X}, \mathcal{M}_j)$, even if different penalties are applied for model complexity. Therefore, model selection rests on the ability to reproduce the historic record. Yet, this begs the question of whether the ability to reproduce historic records is relevant for NS model selection. When comparing multiple ST models, this paradigm is useful, since we are not anticipating change in the historic distribution (by definition of stationarity). At least for engineering applications, the value of NS models should be dependant on their predictive ability, not their descriptive ability, as measured by goodness of fit. This issue of NS model selection and validation has been raised by other authors [Renard et al., 2013], and Burnham and Anderson [2003] note that their proposed guidelines for model selection cannot be expected to hold if observations are not independent. In truth, the predictive capabilities of a model can only be properly quantified using data that was not used in the fitting procedure. In this study, the out-of-sample predictions of the competing models are evaluated using Bayesian hypothesis testing.

1.5.2 Bayesian Hypothesis Testing

Bayesian hypothesis testing depends on Bayes factor, which summarizes the evidence given by the data in favor of one statistical model relative to another [Kass and Raftery, 1995]. The "evidence" was introduced in Equation 1.4, but here we are interested in determining the evidence given by the *out-of-sample* data, or $p(\mathbf{X}^*|\mathcal{M}_j)$. The out-of-sample data, $\mathbf{X}^*$,

was previously defined in Section 1.4. In this study, the evidence represents the probability of seeing the out-of-sample data that were actually observed under the competing models. Consequently, the evidence is extremely useful for model selection: models with larger values of $p(\mathbf{X}^*|\mathcal{M}_j)$ are statistically preferred for predicting the out-of-sample data. Model comparison is formally accomplished using the Bayes factor

$$B_{j,k} = 2\log_e\left[\frac{p(\mathbf{X}^*|\mathcal{M}_j)}{p(\mathbf{X}^*|\mathcal{M}_k)}\right] \tag{1.9}$$

where $B_{j,k}$ denotes the Bayes factor for $\mathcal{M}_j$ and against $\mathcal{M}_k$. Here, we present a $\log_e$-scale formulation of the Bayes factor for simpler interpretation. Positive values of $B_{j,k}$ mean the evidence supports $\mathcal{M}_j$, and negative values support $\mathcal{M}_k$. The Bayes factor can be viewed as a predictive score, which measures the relative success of $\mathcal{M}_j$ and $\mathcal{M}_k$ at predicting $\mathbf{X}^*$. Interpretation of this Bayes factor formulation is given by Kass and Raftery [1995]. There are several important properties of Bayes factors and the evidence which make them especially useful for model selection.

First, the evidence automatically penalizes model complexity and parameter uncertainty. As the prior density diffuses through either larger parameter uncertainty or increasing dimensionality of the prior (i.e. increasing complexity), the value of the evidence will generally decrease (Equation 1.4). However, if increasing the dimensionality of the prior also causes the likelihood of the data to increase, the evidence could favor the more complex model. In this light, Bayes factor offers a formal means to measure if increased model complexity is justified by the data. Aside from favoring models with relatively concentrated priors, the evidence will also favor models where the prior parameter knowledge agrees with the data, which means that the likelihood function is maximized at or near the prior mode. This property allows different parameterizations of identical models to be evaluated according to Bayes factor. Overall, the Bayes factor is very sensitive to the choice of the prior, which is generally considered a downside of Bayesian hypothesis testing [Kass and Raftery, 1995].

However, sensitivity to the prior also creates opportunities for flexible applications.

Likelihood Function (Evaluation Period)

To compute Bayes factor, the evidence must be calculated under the competing models. This requires specification of the evaluation period prior and likelihood function for $\mathcal{M}_{\text{s}}$, $\mathcal{M}_{\text{u}}$, and $\mathcal{M}_{\text{n}}$, as well as evaluating the Bayes factor integral (Equation 1.4), which is a very challenging integration. So let us first define the likelihood function, since it is relatively intuitive. To measure the predictive ability of the models, we must evaluate the likelihood function at the out-of-sample data

$$L(\boldsymbol{\theta}_j|\mathbf{X}^*, \mathcal{M}_j) = \prod_{i=1}^{n^*} p(X_i|\boldsymbol{\theta}_j, \mathcal{M}_j) \tag{1.10}$$

where n^* represents the number of observations in the evaluation data. Equation 1.10 is identical to the likelihood function of the fitting period, except here we are evaluating the PIII density function at the out-of-sample data. Also, $j = \{\text{“s”}, \text{“n”}, \text{“u”}\}$, since we are evaluating the predictions of all three models. When Equation 1.10 is applied to calculate the likelihood of $\boldsymbol{\theta}_{\text{n}}$ under $\mathcal{M}_{\text{n}}$, again there is a time dependant parameterization of the PIII density function (Equation 1.7), and t_i represents the time between the first observation of the fitting period and the i^{th} observation of the evaluation period. We note this to emphasize that the trend parameter under $\mathcal{M}_{\text{n}}$ is extrapolated throughout the evaluation period.

Notice that the likelihood function distinguishes $\mathcal{M}_{\text{n}}$ from $\mathcal{M}_{\text{s}}$ and $\mathcal{M}_{\text{u}}$, but does not distinguish $\mathcal{M}_{\text{u}}$ from $\mathcal{M}_{\text{s}}$. Remember, under both $\mathcal{M}_{\text{s}}$ and $\mathcal{M}_{\text{u}}$, we predict future discharges are distributed according to a ST LPIII model. This justifies identical likelihood function evaluations. However, the parameters of the ST LPIII model are different under $\mathcal{M}_{\text{s}}$ and $\mathcal{M}_{\text{u}}$, since $\boldsymbol{\theta}_{\text{u}}$ was derived from the NS model. We can account for the difference between $\mathcal{M}_{\text{s}}$ and $\mathcal{M}_{\text{u}}$ through specification of the prior.

Prior Distributions (Evaluation Period)

In this study, our prior knowledge of the evaluation period model parameters is based on the posterior distributions inferred from the fitting period data. Even though observations were used to create this prior knowledge, we can still consider this a prior belief about the evaluation period parameters, since the out-of-sample data was not used in the fitting procedure. This technique provides informative, joint priors describing the evaluation period model parameters, which allows for different parameterizations of the same model to be tested according to Bayes factor. Of course, we still need to formalize this prior knowledge as a probability distribution.

The evaluation period priors under $\mathcal{M}_{\mathrm{s}}$, $\mathcal{M}_{\mathrm{u}}$ and $\mathcal{M}_{\mathrm{n}}$ were defined by a multivariate Gaussian mixture model (GMM) fit to the posterior samples of $\boldsymbol{\theta}_{\mathrm{s}}$, $\boldsymbol{\theta}_{\mathrm{u}}$, and $\boldsymbol{\theta}_{\mathrm{n}}$, respectively. A GMM was chosen to describe the evaluation period priors since GMMs can represent a large class of distributions and approximate arbitrarily shaped densities. This property is ideal for approximating the joint densities of the posterior samples, since no assumption is required regarding the shape of marginal posteriors. In this study, the hyperparameters of the GMM were inferred through the Expectation Maximization (EM) algorithm. The number of Gaussians in the mixture, K, was determined by iteratively increasing the number of Gaussian components until $\mathrm{BIC}_K - \mathrm{BIC}_{K+1} < 2$. For more on mixture models and the EM algorithm, please see Picard [2007].

An example of the GMM fit to the posterior samples of the Smith River fitting period is shown in Figure 1.7. Figure 1.7 shows marginal prior densities of $\boldsymbol{\theta}_{\mathrm{n}}$ (top row), $\boldsymbol{\theta}_{\mathrm{s}}$ (middle row), and $\boldsymbol{\theta}_{\mathrm{u}}$ (bottom row), although the fitted GMM also defines the joint density of parameter values. Black lines represent the prior density defined by the GMM, and the histograms are the posteriors from the Smith River fitting period. There are several important properties of the evaluation period priors that are highlighted by Figure 1.7. First, notice the difference

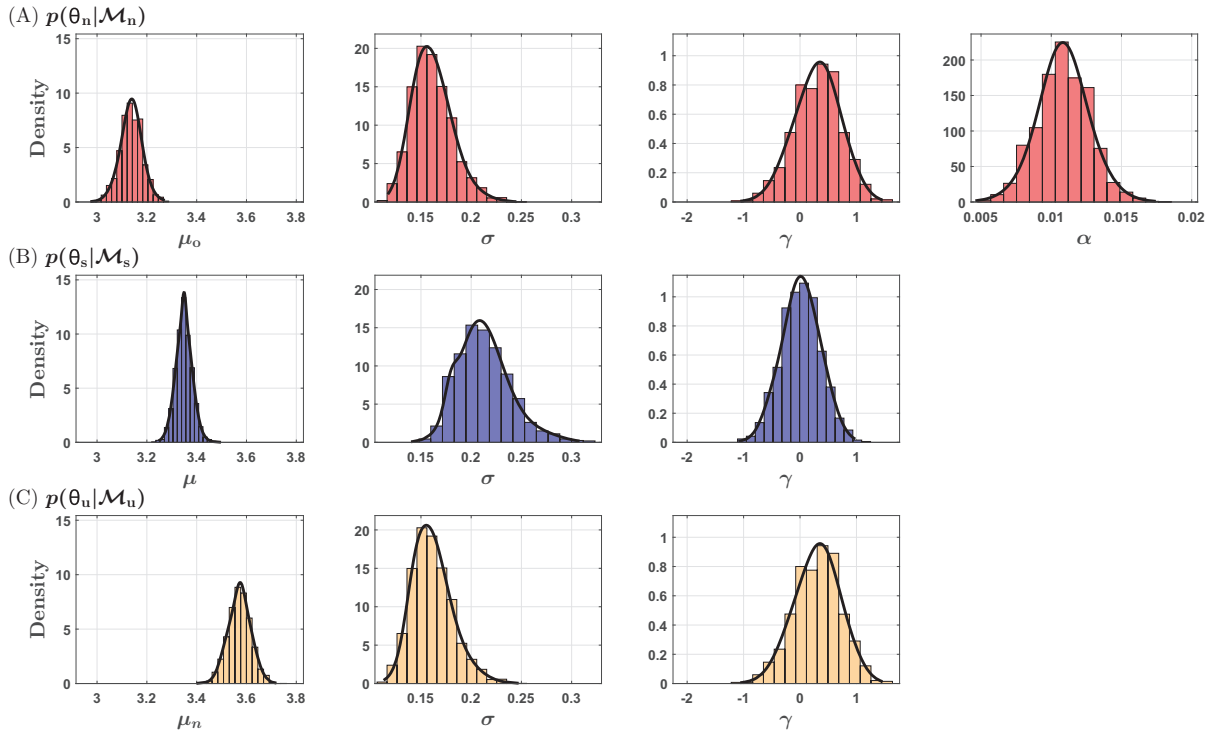

Figure 1.7: Prior parameter density for the evaluation period (*black* lines) under $\mathcal{M}_n$ (row A), $\mathcal{M}_s$ (row B), and $\mathcal{M}_u$ (row C) defined by the GMM fit to the posterior distributions of the Smith River fitting period (*red*, *blue*, and *gold* historgrams). The additional complexity under $\mathcal{M}_n$ is accounted for through its four dimensional prior. Increased parameter uncertainty of μ_n is represented by its lower and more dispersed prior density relative to μ.

in the prior modes under $\mathcal{M}_{\mathrm{s}}$ compared to $\mathcal{M}_{\mathrm{u}}$. Under $\mathcal{M}_{\mathrm{s}}$, the prior favors lower values of μ relative to the μ_n prior density. This is because $\mathcal{M}_{\mathrm{s}}$ assumes no change in the location of the distribution, so the out-of-sample μ should be similar to the stationary μ inferred from the fitting period. Conversely, under $\mathcal{M}_{\mathrm{u}}$, a persistent shift in the distribution has occurred, so predictions are best made with the updated (larger) μ_n. The important concept here is that the predictions under $\mathcal{M}_{\mathrm{s}}$ and $\mathcal{M}_{\mathrm{u}}$ are set apart by their respective priors. In this study, the evidence will generally favor the model where the prior mode is closest to the out-of-sample μ and σ. However, this generalization does not consider how parameter uncertainty affects the model comparisons.

Notice also that the posterior samples of μ_n are significantly more dispersed than μ. Since samples of μ_n depend on the uncertain intercept and trend parameter of the linear model (Equation 1.2), they are understandably more dispersed than μ, which simply represents the mean of the fitting period. Figure 1.7 demonstrates the larger parameter uncertainty under $\mathcal{M}_{\mathrm{u}}$, as shown by the lower and more dispersed prior probability density of μ_n relative to μ. Therefore, according to Bayes factor, $\mathcal{M}_{\mathrm{u}}$ will only be favored relative to $\mathcal{M}_{\mathrm{s}}$ if the increased parameter uncertainty is offset by a substantially more accurate prediction of the out-of-sample μ and σ. Following similar reasoning, $\mathcal{M}_{\mathrm{n}}$ will only be favored relative to $\mathcal{M}_{\mathrm{s}}$ and $\mathcal{M}_{\mathrm{u}}$ if two conditions are met. First, $\mathcal{M}_{\mathrm{n}}$ must substantially improve the the likelihood of $\mathbf{X}^*$ in order to offset the diffusion of prior density caused by the addition of the trend parameter, α. Secondly, the trend inferred from fitting period must be similar to the out-of-sample trend, or else $p(\boldsymbol{\theta}_{\mathrm{n}}|\mathcal{M}_{\mathrm{n}})$ will not agree with $L(\boldsymbol{\theta}_{\mathrm{n}}|\mathbf{X}^*, \mathcal{M}_{\mathrm{n}})$, thus significantly diminishing the evidence. Therefore, Bayes factor will evaluate if predictions under the NS model improve the statistical representation of the out-of-sample data enough to justify the additional complexity and uncertainty introduced by extrapolation of the NS trend parameter. The last step towards implementing the proposed test requires integration of Equation 1.4.

Integration of the Bayes Factor Integral

Analytic evaluation of the Bayes factor integral is only possible for a narrow class of models. In the absence of analytic solutions, the evidence is approximated using a variety of methods which are reviewed by Kass and Raftery [1995]. In this study, Gaussian Mixture Importance Sampling (GAME) is applied for approximation of the evidence [**?**]. Essentially, GMIS is a Monte Carlo integration technique where the efficiency is improved through importance sampling, and the importance distribution is defined by a GMM fit to a posterior sample. The GMIS estimator was calculated for each record and model by (1) generating another posterior sample with $\text{DREAM}_{(\text{ZS})}$ using the evaluation period likelihood function and priors defined in Section 1.5.2, (2) fitting a GMM to the second generated sample using the EM algorithm, and (3) applying the standard importance sampling integration strategy presented by Kass and Raftery [1995] using the fitted GMM as the importance distribution. The strength of the GMIS estimator is that it is free of theoretical assumptions regarding the shape of the posterior, and it is computationally tractable through use of the importance distribution.

1.6 Results: Evaluation of Predictive Ability

Let us begin by revealing the evaluation period of the Smith River Record. Figure 1.8 shows the data in the full Smith River record (Figure 1.8A), as well as predictions derived from the maximum a posteriori (MAP) parameter estimates compared to the out-of-sample density (Figure 1.8B-D). Undeniably, the trend that was detected in the fitting period did not persist in the evaluation period. Thus, predictions made under both $\mathcal{M}_{\text{n}}$ and $\mathcal{M}_{\text{u}}$ drastically over-estimate the out-of-sample μ. Bayes factor reflects these poor predictions by very strongly supporting $\mathcal{M}_{\text{s}}$ relative to both $\mathcal{M}_{\text{n}}$ and $\mathcal{M}_{\text{u}}$ (values given in Figure 1.8C, interpretation based on Kass and Raftery [1995]). Under $\mathcal{M}_{\text{s}}$, the trend that occurred throughout the fitting period was adequately represented through a large variance parameter, reflecting the

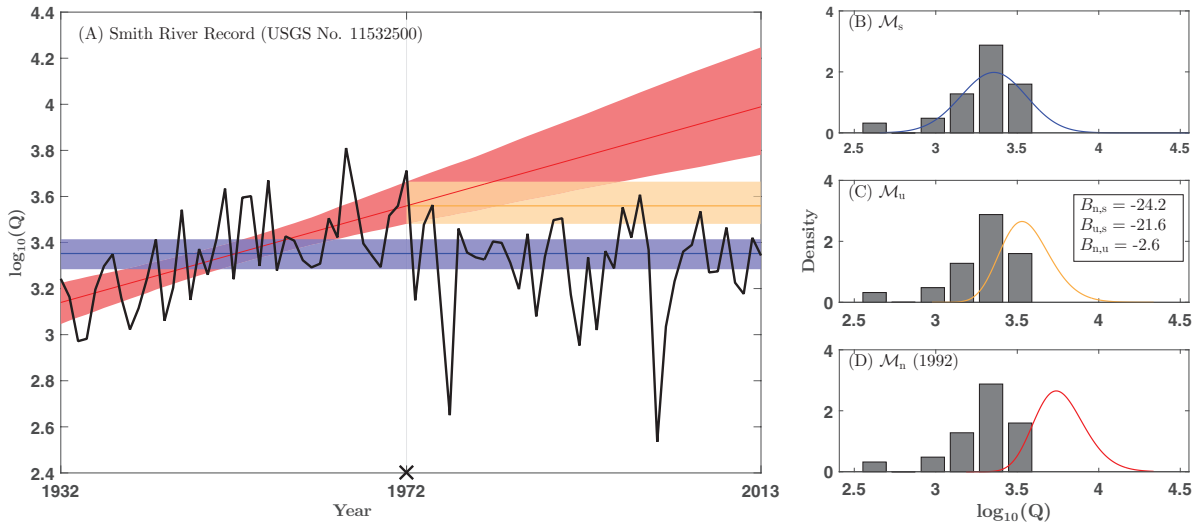

Figure 1.8: (A) The MAP estimate of location parameters under $\mathcal{M}_s$ (*blue* line), $\mathcal{M}_u$ (*gold* line), and $\mathcal{M}_n$ (*red* line) shown over the full record length (*black* line). The colored shading represents the respective 95% credible intervals of the location parameters. (B - D) Predictions of out-of-sample density under $\mathcal{M}_s$ (*blue* line), $\mathcal{M}_u$ (*gold* line), and $\mathcal{M}_n$ (*red* line) derived from the MAP parameter estimate. The *black* histograms show the empirical density in the evaluation period. For the Smith River record, $\mathcal{M}_s$ most accurately predicted the out-of-sample data, which is reflected by $B_{n,s}$ and $B_{u,s}$ shown in (C).

variability of the hydrologic system. The apparent increase in discharge magnitude from 1932 to 1972 could be considered long term persistence and not a significant "trend", or the increase could be considered a significant trend which simply did not persist throughout the prediction period. A proper physical understanding of the relationship between climatology, land use, and flood discharges in the Smith River watershed is needed to elaborate further. Indeed, the detection of a trend alone offers little insight towards persistence, and the significance of detected trends is particularly uncertain for hydroclimatic data because dependance structures are poorly understood [Cohn and Lins, 2005, Lins and Cohn, 2011]. This point is not only supported by previous studies and Smith River watershed, but also by the results of our evaluation for other records tested.

Figure 1.9 summarizes the Bayes factor results for all records tested (1250 total). The pie charts display the Bayes factor value as a percentage of records tested. The columns compare

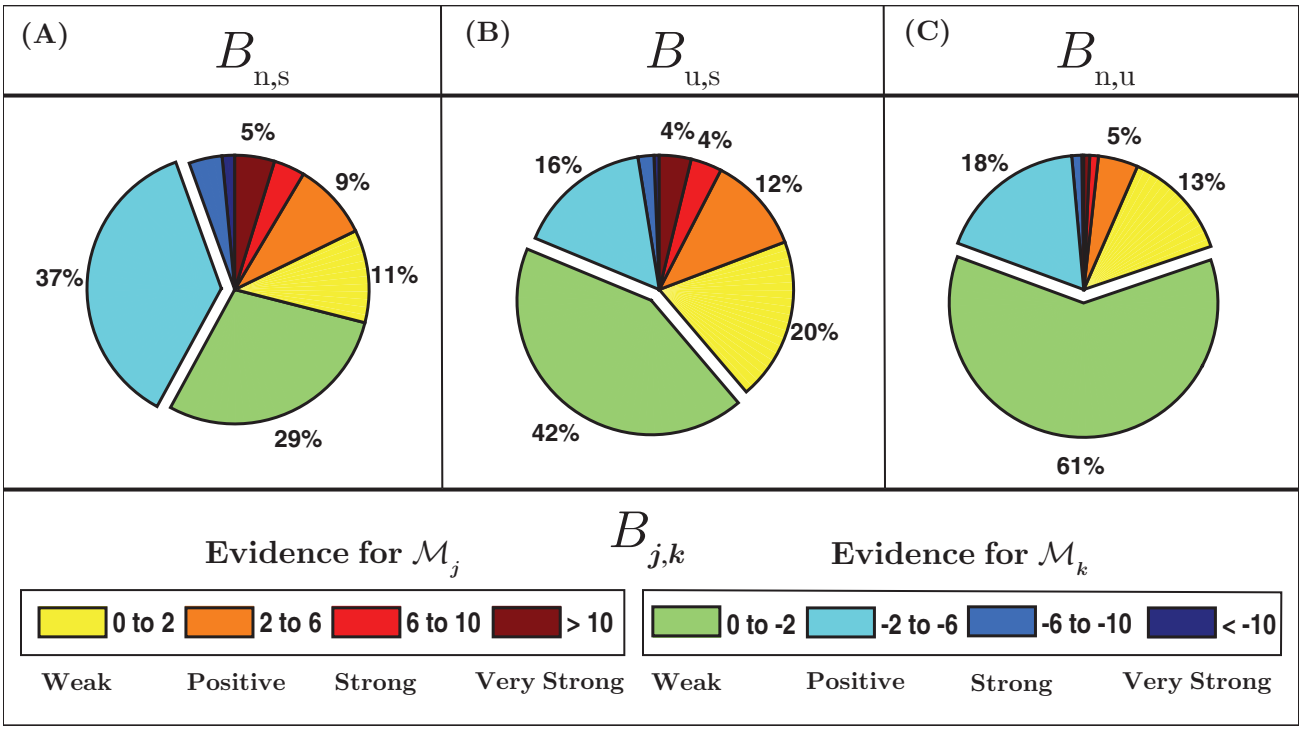

Figure 1.9: Bayes factor value as a percentage of all records tested (1250 Total). (A) Bayes factor for $\mathcal{M}_{\rm n}$ and against $\mathcal{M}_{\rm s}$. (B) Bayes factor for $\mathcal{M}_{\rm u}$ and against $\mathcal{M}_{\rm s}$. (C) Bayes factor for $\mathcal{M}_{\rm n}$ and against $\mathcal{M}_{\rm u}$. *Yellow* to *red* values support $\mathcal{M}_j$, which is $\mathcal{M}_{\rm n}$ in (A and C), and $\mathcal{M}_{\rm u}$ in (B). *Green* to *blue* colors support $\mathcal{M}_k$, which is $\mathcal{M}_{\rm s}$ in (A and B) and $\mathcal{M}_{\rm u}$ in (C). Results in (A) show that the evidence supports predictions under $\mathcal{M}_{\rm s}$ relative to $\mathcal{M}_{\rm n}$ for about 70% of records tested, with 40% of records exhibiting at least positive evidence for $\mathcal{M}_{\rm s}$. Results in (B) show that predictions under $\mathcal{M}_{\rm s}$ are also preferred relative to $\mathcal{M}_{\rm u}$, except here the majority of evidence only weakly supports $\mathcal{M}_{\rm s}$. Interpretation of the Bayes factor values is based on Kass and Raftery [1995].

the evidence between different models, or Bayes factor for $\mathcal{M}_n$ and against $\mathcal{M}_s$ (Figure 1.9A), for $\mathcal{M}_u$ and against $\mathcal{M}_s$ (Figure 1.9B), and for $\mathcal{M}_n$ and against $\mathcal{M}_u$ (Figure 1.9C). For example, blue to green colors in Figure 1.9A-B represent records where the evidence favors $\mathcal{M}_s$ for prediction, whereas yellow to red colors in Figure 1.9A,C favor $\mathcal{M}_n$ for prediction. Interpretation of Bayes factor shown in the Figure 1.9 legend is based on Kass and Raftery [1995]. Several important results are highlighted by examining the Bayes factor results over all records tested. First, Figure 1.9A shows that evidence favors $\mathcal{M}_s$ relative to $\mathcal{M}_n$ for roughly 70% of all records tested, with 40% of records exhibiting at least positive evidence towards $\mathcal{M}_s$. This means that the ST model was a better predictor of the out-of-sample data relative to the NS model extrapolations for about 70% of records tested. Bayes factor between $\mathcal{M}_s$ and $\mathcal{M}_u$ yield similar results (Figure 1.9B), except here the majority of tests offer only weak evidence towards $\mathcal{M}_s$ for prediction. Secondly, notice also that the evidence is not strongly in favor of any competing model when we examine the record pool as a whole, with the majority of the evidence falling within the weak to positive range. This is not surprising, since 88% of the records tested do not exhibit a statistically significant trend in the fitting period. Therefore, predictions under both $\mathcal{M}_n$ and $\mathcal{M}_u$ will be similar to $\mathcal{M}_s$ for most records. When models produce similar representations of data, Bayes factor naturally favors the simplest, so it is not surprising that the ST model is the apparent favorite for prediction over all records tested. However, this result does raise an important point: $\mathcal{M}_s$, or predictions of flood discharges using a ST model, should remain the default technique. The majority of flood records in the United States do not exhibit a statistically detectable trend, and therefore the trend parameter of the NS model is simply a nuisance parameter in most situations. Now, let us examine records where NS and uST model predictions are substantially different from the ST model.

Figure 1.10 summarizes the results for subsets of the record pool with detected trends in the fitting period. The subsets shown in Figure 1.10 include records with a detected trend in the fitting period and DIC significantly favoring the NS model, or $\mathrm{DIC_s} - \mathrm{DIC_n} > 2$

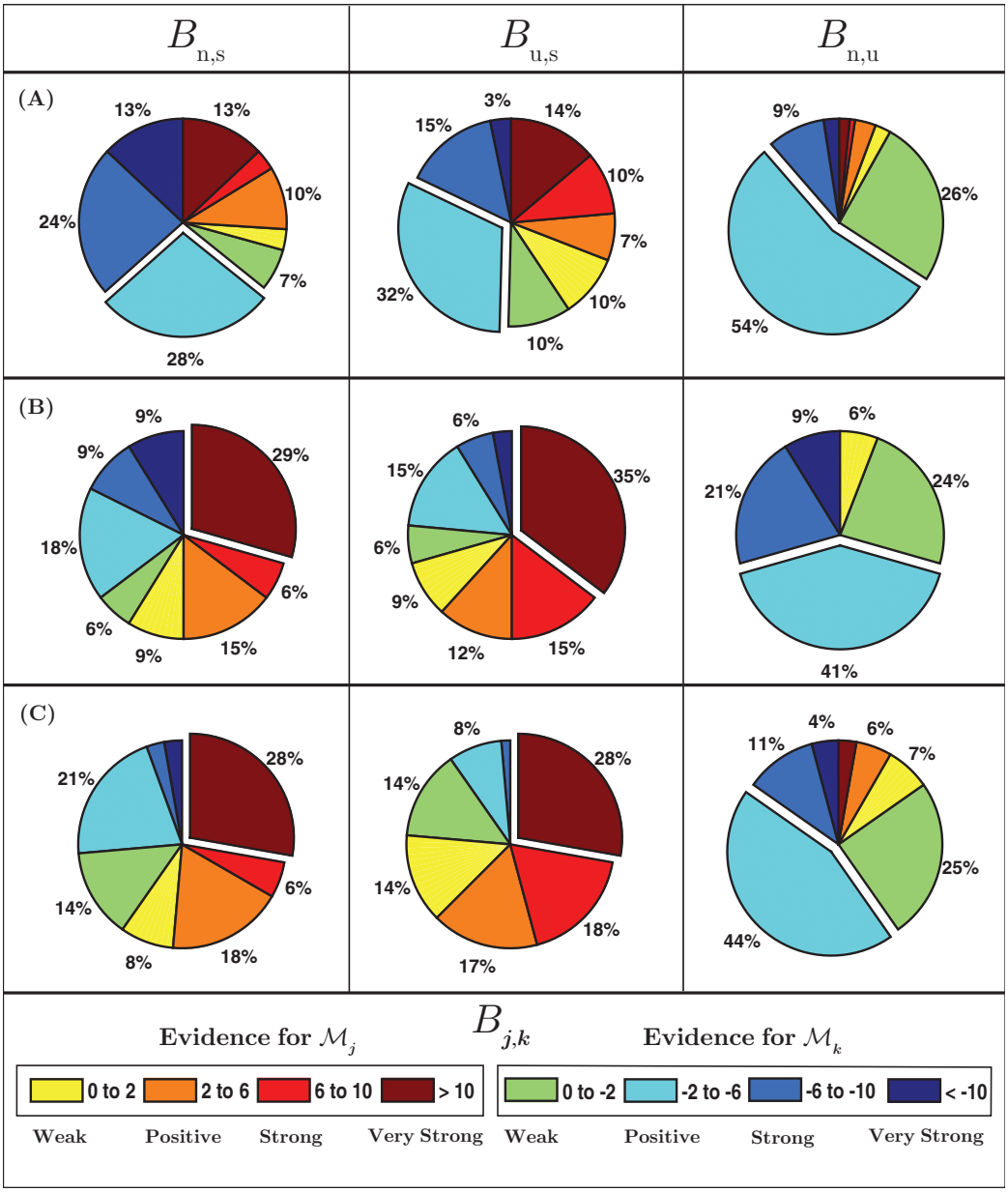

Figure 1.10: Bayes factor value as a percentage of records with detected trends in the fitting period and (A) DIC significantly favoring $\mathcal{M}_{\mathrm{n}}$ (123 total), (B) discharges affected by land use changes, regulation, or diversion (34 total), (C) trend also detected using full record length (72 total). The columns compare the evidence between competing models in the same manner as Figure 1.9. (A) Reveals that the presence of a trend and DIC favoring the NS model does not improve the relative success of $\mathcal{M}_{\mathrm{n}}$ compared to all records tested in Figure 1.9. (B) Shows that $\mathcal{M}_{\mathrm{u}}$ is strongly preferred for prediction relative to $\mathcal{M}_{\mathrm{s}}$ and $\mathcal{M}_{\mathrm{n}}$ if a trend is accompanied by physical watershed changes.

(top row). The middle row shows the results for records with detected trends and discharges affected by land use changes, regulation, or diversion. The bottom row shows records where a trend was detected in the fitting period and full record length. Again, the columns compare the competing models. First, let us discuss records where trends were accompanied by fit metrics favoring the NS model (Figure 1.10A). Here, the statistical metrics determined from the fitting period suggest that the ST assumption is violated and the NS model is a preferred alternative. Despite these metrics, Figure 1.10A shows $\mathcal{M}_{\mathrm{s}}$ is strongly preferred for prediction given records of this category. Bayes factor between $\mathcal{M}_{\mathrm{s}}$ and $\mathcal{M}_{\mathrm{n}}$ shows that the evidence is in favor of $\mathcal{M}_{\mathrm{s}}$ for roughly 75% of these trending records. Moreover, about 40% exhibit at least *strong* evidence in favor of $\mathcal{M}_{\mathrm{s}}$. Remarkably, comparison between Figure 1.9A and Figure 1.10A illustrates that $\mathcal{M}_{\mathrm{n}}$ is less preferred relative to $\mathcal{M}_{\mathrm{s}}$ when a trend is detected and goodness of fit favors the NS model. Similarly, the evidence favors $\mathcal{M}_{\mathrm{s}}$ relative to $\mathcal{M}_{\mathrm{u}}$ for these records, although $\mathcal{M}_{\mathrm{u}}$ did perform better than $\mathcal{M}_{\mathrm{n}}$. This result confirms our earlier assertion that a statistically significant trend offers little insight towards persistence, and therefore little insight towards the predictive ability of the NS model. Furthermore, metrics of model fit measure the probability of reproducing the historic record, which is not related to trend persistence. To illustrate this important point further, Figure 1.11 shows the predictions under the three models compared to the data in the full record, given a trend and fit metrics favoring $\mathcal{M}_{\mathrm{n}}$. Indeed, trends in the shown fitting periods are visually apparent and statistically significant, yet both $\mathcal{M}_{\mathrm{n}}$ and $\mathcal{M}_{\mathrm{u}}$ are poor predictors of the second half of the record. Thus, additional criteria are needed to confidently and reliably select NS models for prediction. Let us now examine records where trends were accompanied by indications of persistent change.

Figure 1.10B summarizes the results for trending fitting periods where anthropogenic influences were known to affect discharges. Records in Figure 1.10B were affected by upstream regulation, diversion, land use changes, or channelization during the fitting period. Again, we rely on USGS data flagging to find influenced records (Section 1.4). Only 34 records

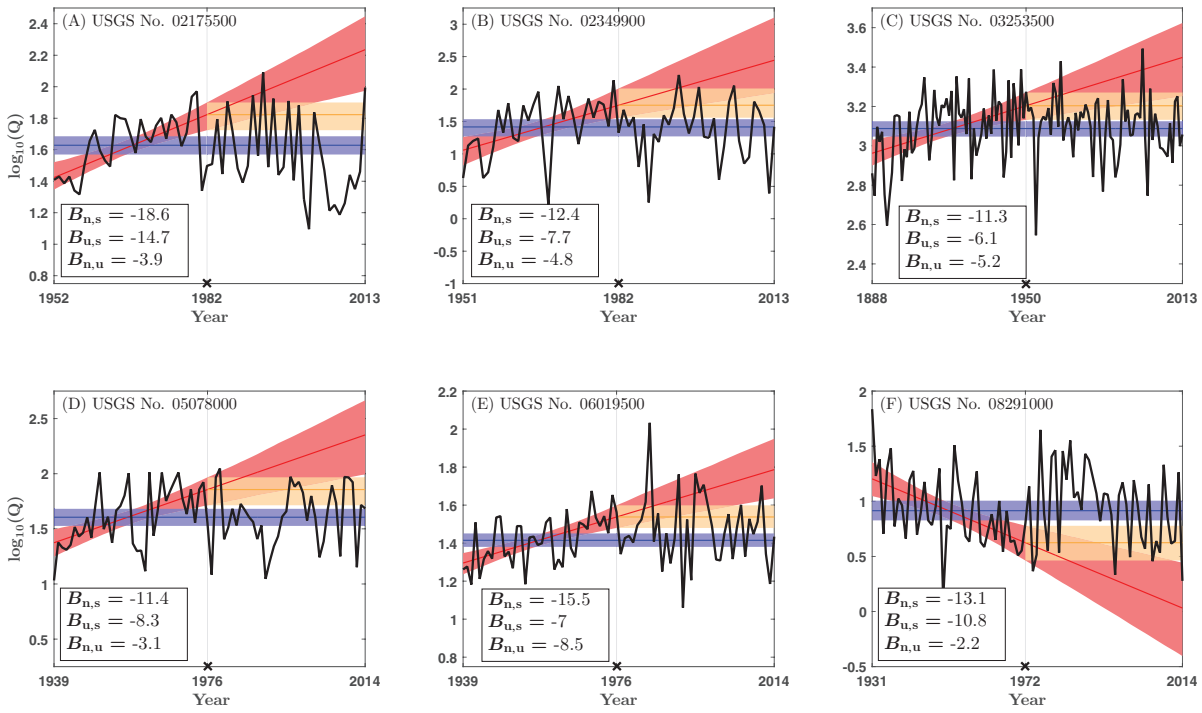

Figure 1.11: The MAP estimate of location parameters under $\mathcal{M}_s$ (*blue* line), $\mathcal{M}_u$ (*gold* line), and $\mathcal{M}_n$ (*red* line) shown over the full data record (*black* line). The colored shading represents the respective 90% credible intervals of the location parameters, and the *black* cross denotes the end of the fitting period. The Bayes factor values are shown in the text boxes. Records (A-F) belong to category (A) of Figure 1.10, i.e. they exhibit a significant trend in the fitting period and DIC favors the NS model.

tested meet this criteria, primarily because the first half of USGS records rarely contain data flags *and* significant trends. For these records, both $\mathcal{M}_{\mathrm{u}}$ and $\mathcal{M}_{\mathrm{n}}$ are preferred relative to $\mathcal{M}_{\mathrm{s}}$. Notably, the evidence is *very strongly* in favor of $\mathcal{M}_{\mathrm{u}}$ relative to $\mathcal{M}_{\mathrm{s}}$ for more than a third of the records in Figure 1.10B, with 50% of records exhibiting at least *strong* evidence for $\mathcal{M}_{\mathrm{u}}$. The relative success of the predictions derived from the NS model can be attributed to the permanent nature of anthropogenic watershed changes. Here, we know the cause of the observed trend will persist without intervention, so it is much more likely for NS and uST predictions to be successful. Figure 1.12 shows successful predictions under $\mathcal{M}_{\mathrm{u}}$ relative to $\mathcal{M}_{\mathrm{s}}$ for selected records in Figure 1.10B. It is very important to note that $\mathcal{M}_{\mathrm{n}}$ was preferred relative to $\mathcal{M}_{\mathrm{u}}$ for only 2 of the 34 records in Figure 1.10B, and still the evidence only weakly supported $\mathcal{M}_{\mathrm{n}}$. Thus, for records with a detected trend in the fitting period likely caused by anthropogenic influences, the uST model is preferred for prediction. This means that extrapolation of inferred NS model parameters is almost never preferred, even when the physical cause of the observed trend is known to be persistent. We emphasize this point through examination of one more sub-set of the record pool.

Figure 1.10C summarizes the results for records where the trend detected in the fitting period was also detected using the full record length. In other words, the monotonic trend detected in the fitting period persists in the evaluation period, although not necessarily at the same rate. In total, 72 records meet this criteria. Again, both $\mathcal{M}_{\mathrm{u}}$ and $\mathcal{M}_{\mathrm{n}}$ are preferred relative to $\mathcal{M}_{\mathrm{s}}$. Figure 1.10C also shows that $\mathcal{M}_{\mathrm{n}}$ only significantly improves predictions relative to $\mathcal{M}_{\mathrm{u}}$ for about 10% of the records with an observed trend in the fitting period and full record length. In context of Bayes factor, this means the statistical representation of the out-of-sample data is *rarely* sufficiently improved to justify the additional complexity and uncertainty associated with the NS model extrapolations. Figure 1.12E-F strongly illustrates this result. For these records, the extrapolations under $\mathcal{M}_{\mathrm{n}}$ match the prediction period data relatively well compared to predictions under $\mathcal{M}_{\mathrm{u}}$, which appear to be biased-low. However, Bayes factor only weakly supports $\mathcal{M}_{\mathrm{n}}$ relative to $\mathcal{M}_{\mathrm{u}}$ in Figure 1.12E, and strongly supports

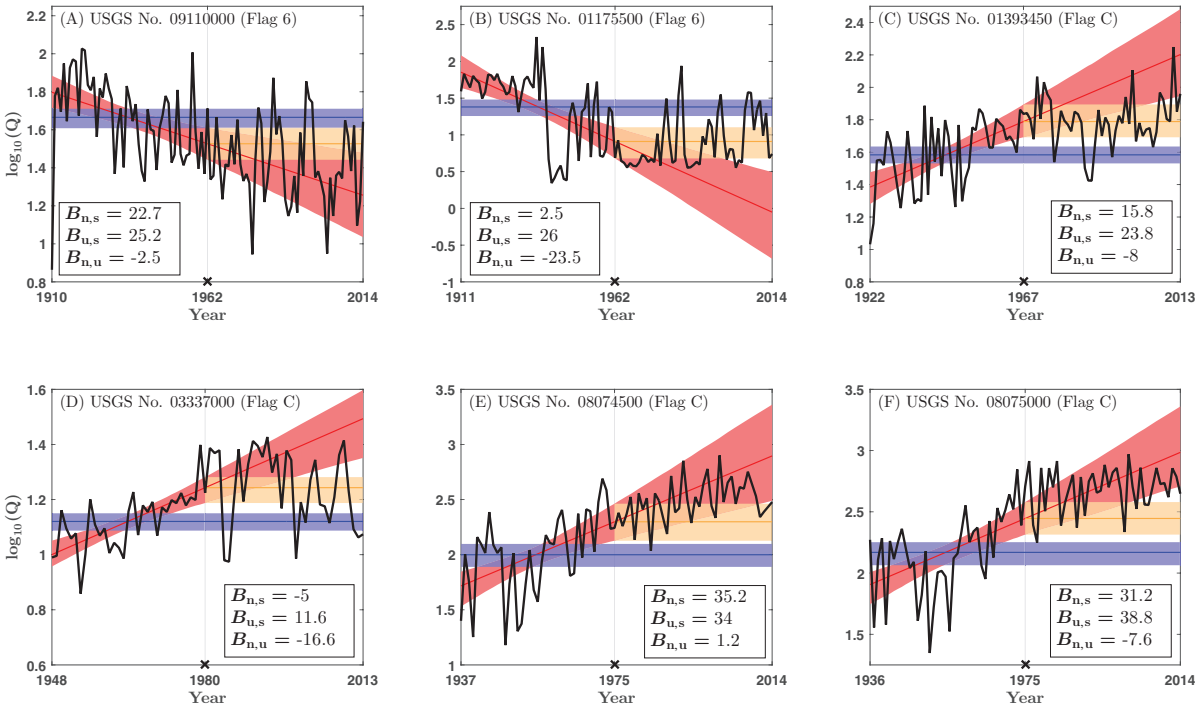

Figure 1.12: The MAP estimate of location parameters under $\mathcal{M}_s$ (*blue* line), $\mathcal{M}_u$ (*gold* line), and $\mathcal{M}_n$ (*red* line) shown over the full data record (*black* line). The colored shading represents the respective 90% credible intervals of the location parameters, and the *black* cross denotes the end of the fitting period. The Bayes factor values are shown in the text boxes. Records (A-F) belong to category (B) of Figure 1.10, i.e. they exhibit a significant trend in the fitting period and discharges are affected by land use changes or channelization (Flag C), upstream regulation, or diversion (Flag 6).

$\mathcal{M}_u$ relative to $\mathcal{M}_n$ in Figure 1.12F. Although this outcome appears odd, the prediction period μ is well within the 90% credible intervals under $\mathcal{M}_u$. Therefore, an out-of-sample trend could be accounted for by making predictions with a conservative value of μ selected from within the uncertainty bounds under $\mathcal{M}_u$, and this approach is likely preferred relative to extrapolations of an estimated trend under $\mathcal{M}_n$. Appropriate quantiles of uST μ that are useful for prediction in the case of an out-of-sample (future) trend could be recommend via simulation in future studies. While often overlooked, the ST assumption appears to have some utility even in the context of NS [Matalas, 2012]. Before stating conclusions, let us discuss the implications and limitations of these results.

1.7 Discussion

Results from this study show that the extrapolation of parameter trends in the tested NS model does not significantly improve prediction, even when the observed trend persists into the future. However, results also show that extrapolation is not necessary to improve predictions relative to the traditional ST approach. The uST model, derived from the most recent in-sample NS model parameters, is preferred for prediction when a detected trend can be attributed to physical watershed changes (see Viglione et al. [2016] and references therein for recent NS attribution methods). Other authors have emphasized the importance of considering NS when the causes are well understood [Lins and Cohn, 2011], and Vogel et al. [2011] stressed that much greater attention should be given to anthropogenically influenced watersheds. Yet, in the United States, there are no federal guidelines for FFA in watersheds where documented physical alterations have changed the local hydrology. Using NS models to update ST distributions is an attractive approach for the statistical treatment of these watersheds. Predictions under the uST distribution only assume that there has been a persistent change in the distribution of flood discharges, and our results show that even if

the distribution continues to change, the uST distribution is likely preferred relative to extrapolations of the NS model parameters conditioned on time. Also, by making predictions within the ST paradigm, the concepts of return period and risk do not have to be extended to a fully NS framework for the design of hydraulic structures and hazard mapping. Until research shows *how* our prediction of flood peak distributions can be significantly improved by the extrapolation of NS model trend parameters, the ST assumption still holds for planning and design [Montanari and Koutsoyiannis, 2014].

For illustrative purposes, Figure 1.13B shows the return periods derived from the quantiles of the uST LPIII distribution using the full Whiteoak Bayou record near Houston, Texas for parameter inference (Figure 1.13A). The uncertainty in the predicted 100-year discharge (q_{99}) is also shown (Figure 1.13C), demonstrating the significant uncertainty in the estimate and the advantage of the Bayesian approach for parameter inference. We provide a MATLAB® program in the supporting information which can be used for inference of the uST distribution parameters from complete stream flow records, but we must emphasize several limitations. First, the Bayesian procedure applied herein measured the relative success of predicting the *overall* out-of-sample distribution, and not the precision of extreme quantiles. While the Bayesian approach is particularly useful for assessing the justified level of model complexity, traditional statistical techniques are more appropriate for testing the accuracy of specific quantiles. Under the traditional or frequentist approach, simulation is used to evaluate alternative quantile estimators in terms of bias and variance with respect to design flood estimates (see Hosking and Wallis [1997], Section 6.4 for a detailed example). Therefore, the results show that the uST model is most appropriate among the tested models when NS behavior is obvious, but the uST model does not necessarily predict q_{99} accurately, for example. Future studies could quantify the error in uST predictions of extreme quantiles through Monte Carlo simulation, similar to those conducted by Yu et al. [2015]. Secondly, although the uST model is *statistically* preferred for prediction relative to NS extrapolations, this result ignores the flood damages associated with under prediction. If we considered the

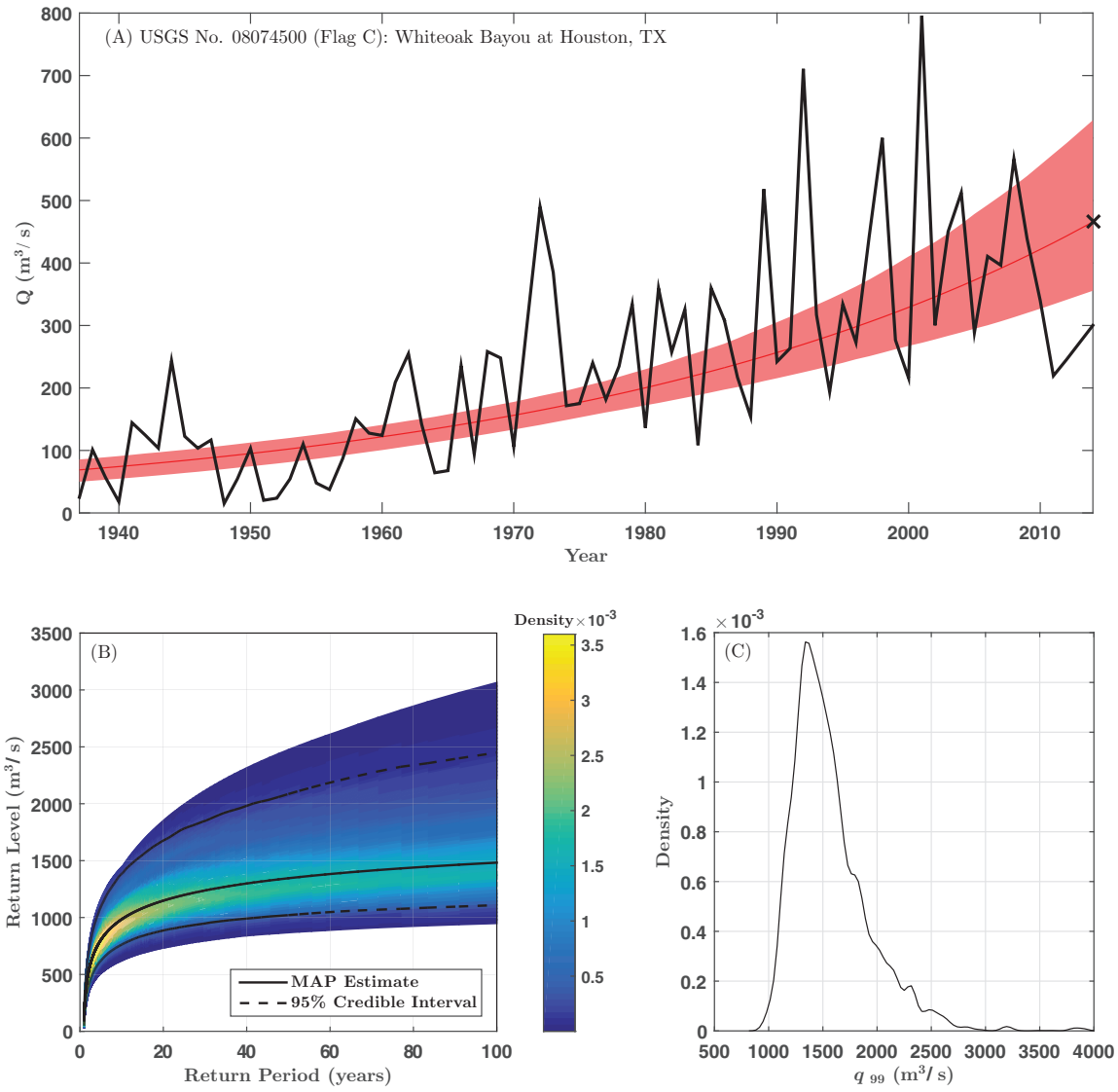

Figure 1.13: (A) The MAP estimate of the NS location parameter and 95% credible interval (*red* line and shading) inferred from the full record length (*black* line), shown outside of the log-space. The *black* cross denotes the location of the uST distribution. (B) Return level vs. return period plot derived from the uST distribution. The solid *black* line represents the MAP estimate of each return level, and the dashed *black* lines represent the 95% credible interval. The density of the return level estimates is shown on the colorbar, which are readily available from the posterior sample of θ_u. (C) Distribution of the q_{99} estimates (or 100 year return level) under the uST model.

losses associated with exceedances of design flood estimates, a more conservative prediction derived from the NS model extrapolations could be warranted, despite the additional complexity and uncertainty. Indeed, if an observed trend continues, it is unlikely for any derived uncertainty intervals of the uST quantiles to be accurate over a long time period (i.e. greater than 40 years). Therefore, we do not recommend basing design or insurance products on the return periods derived from the uST quantiles, but they can be used for evaluation of present flood risk in urbanizing or significantly altered catchments. While our results show that extrapolation of NS model parameters is not a statistically preferred alternative, this conclusion is somewhat limited to the tested NS model.

The primary limitation of this study is that only one NS model was tested for prediction. The tested NS LPIII model has notable advantages and disadvantages. From a practical point of view, modeling non-stationarity in the LPIII model is advantageous because of its widespread use in the United States and Australia for FFA. The LPIII distribution will continue to be utilized in the United States moving forward, since the forthcoming update to Hydrologic Bulletin 17B will not change the recommended distribution for FFA [Lamontagne et al., 2013]. Working with the LPIII distribution allows us to take advantage of the substantial prior knowledge of the distribution parameters. From a modeling point of view, the main advantage of the tested NS LPIII model is that changes in the mean and standard deviation of $\mathbf{Q}$ are captured with one additional trend parameter. This is a desirable property, since hydrologic non-stationarity would likely result in changes to both the location and scale [Stedinger and Griffis, 2011a]. Therefore, the tested NS model is physically realistic and relatively simple. However, the linear trend in the log-mean equates to an exponential trend in the mean of $\mathbf{Q}$. Therefore, extrapolation of the NS model parameters leads to non-linear predictions of the changing mean. This is not desirable for out-of-sample predictions, which is reflected by the relative failure of NS extrapolations. It is possible that NS predictions would improve through extrapolation of a truly linear NS model. We also note that the sample size available for NS model parameter inference was particularly

limited. Excluding half the available data was necessary for predictive analysis, but the annual maxima sampling method is wasteful of data. Sample size can be increased using a peak over threshold approach for extreme value sampling [Renard et al., 2006b, Kyselỳ et al., 2010], or taking advantage of regional trend information [Cunderlik and Burn, 2003, Renard et al., 2006a]. Both techniques would likely result in improved parameter inference and more reliable trend detection. Short term trends and cycles that appear significant could also be avoided if the NS model incorporates additional information on historic floods [Reis Jr. and Stedinger, 2005, Salinas et al., 2016]. Another limitation of the tested model is that changes in the distribution of flood discharges are conditioned upon time.

By extrapolating a NS model for prediction using *time* as a covariate, we assume that the change in the distribution of flood discharges will continue at the same rate we have estimated from a limited sample. Furthermore, we assume a change in the distribution without a physical understanding of what caused the change. Extrapolations are therefore prone to failure since we have no insight towards the persistence of observed trends. Other authors have discussed the problems associated with NS trend extrapolation [Koutsoyiannis and Montanari, 2015], but the dominance of NS models in terms of traditional model selection metrics can be very misleading and actually cloud our judgement. The findings of this study largely support the claim by Merz et al. [2014] "Although statistical approaches have played and will play an important role, they have to be complemented by the search for the causal mechanisms and dominant processes in the atmosphere, catchment and river system that leave their fingerprints on flood characteristics". An understanding of the dominant mechanisms responsible for changes in flood discharge distributions allows for the development of physically based covariates [Hall et al., 2014, Delgado et al., 2014, Steinschneider and Lall, 2015, Machado et al., 2015, Lima et al., 2015]. Under this approach, the historic record is used to develop a relationship between the distribution and a physical process, such as atmospheric circulation patterns. Future changes to the distribution are then based on the inferred relationship and the prevalence of the physical process under general circula-

tion model (GCM) based climate change projections. Here, *the GCM is applied to provide insight towards the trend of important flood generating mechanisms, rather than to predict local and extreme precipitation events.* For example, Šraj et al. [2016] present a simple NS model where changes in the frequency distribution are conditioned on annual precipitation. Design discharges can then be estimated for different projected values of annual precipitation that occur over the project lifespan. Indeed, physically based NS modeling is a promising direction towards meaningful extrapolation of NS model trend parameters, but continued research aimed at identifying important climate related signals in flood records is needed [Archfield et al., 2016]. We note that extrapolation of inferred trends should only be *considered* when there is strong evidence that the observed trend is likely to continue in the future. Examples include frequency analysis of temperature [Cheng et al., 2015], sea level extremes [Mudersbach and Jensen, 2010], or FFA in a watershed where urbanization is expected to continue in the future.

1.8 Conclusions and Recommendations

Results show that the ST model is preferred for out-of-sample prediction, overall. Even for records with a detected trend in the fitting period and goodness of fit metrics significantly favoring the NS model, the ST predictions are preferred relative to the uST and NS predictions. Therefore, the Mann-Kendall trend test applied to peak discharges and goodness of fit metrics alone are not sufficient to warrant the selection of a NS or uST model for prediction. This conclusion has been stressed by other authors [Renard et al., 2013, Serinaldi and Kilsby, 2015a], and our results demonstrate empirical evidence towards this important assertion. For records with a detected trend and known physical watershed alterations (specifically USGS data flags 5/6 or C), the uST distribution is strongly preferred for prediction over the average evaluation period length of 40 years. Extrapolation of the NS

model parameters is rarely preferred for prediction, even for records with a detected trend in the fitting period and full record. All things considered, we recommend the uST LPIII distribution for evaluation of current flood flood risk in watersheds with a detectable trend in $\mathbf{Q}$ that can be attributed to persistent physical changes. At this time, we do not recommend basing design or insurance products on predictions under the uST distribution because (1) the accuracy of extreme quantiles has not been comprehensively evaluated and (2) if the trend continues it is unlikely for any derived confidence bounds to encompass true quantiles over a long time period. We also recommend a fully Bayesian approach for parameter estimation so that (1) the bias of the NS skew estimate can be reduced with an informative prior, and (2) parameter uncertainty is explicitly characterized. The supporting information provides a MATLAB® program based on the methods outlined in Section 1.5.1 (applied to full record length) for inference of the ST, uST, and NS LPIII parameters. The program also includes basic post-processing options. In conclusion, we recommend that future research should focus on the evaluation of predictions derived from promising NS models and the development of physically based covariates. Many stochastic models have been proposed, but we need to learn how to use them for prediction. Undoubtedly, practitioners are limited to standard statistical methods in the absence of accepted alternatives.

Chapter 2

Using sea level rise projections for engineering design and planning

2.1 Chapter Summary

Policy makers are increasingly requiring engineers to consider sea level rise projections in the planning and design of coastal infrastructure. However, using sea level rise projections is complicated by many factors including 1) speculation about the likelihood of future emission scenarios, 2) the uncertainty in sea level rise projections for a given emission scenario, and 3) the challenge of translating projections of mean sea level into traditional design concepts such as return periods and risk. In this paper, we present methodology for addressing these three challenges for coastal infrastructure planning and engineering design. First, we present an approach for using a combination of extreme value analysis and sea level rise projections to estimate 1) the risk of an exceedance event over the design life of a proposed project and 2) the expected waiting time until water levels exceed a specified elevation. We introduce a "speculation" index to quantify the sensitivity of risk and waiting time estimates to weights

assigned to future emission scenarios. The speculation index is used to derive a design water level for critical infrastructure and an objective planning and design horizon for regular decisions. We illustrate the methodology using historic data and sea level rise projections for Los Angeles, California. Results show that waiting time and risk estimates based on sea level rise projections are remarkably uncertain and speculative over long time horizons (i.e., late 21st century and beyond). In Los Angeles, there is a brief time horizon (prior to 2055) where risk and waiting time estimates are insensitive to speculation and have relatively low variance. For longer time horizons, estimates of risk and waiting times can only be considered objective for extremely high water levels. The results imply that risk-based and cost-effective design is only possible over a short time horizon, and that most long-term decisions should be adaptable to incorporate new information.

2.2 Introduction

The widespread perception of changing environmental conditions severely undermines traditional design concepts in the field of Civil Engineering. The average recurrence intervals of extreme conditions (or return periods) are estimated based on the assumption that historic data samples reasonably approximate future conditions. Estimates of risk, or the probability of an exceedance event during the design lifespan of a structure, are also based on historic data. In more formal terms, estimates of fundamental design parameters such as return periods and risk of extreme events rely on the stationary assumption, or the assumption that the statistical properties of the historic data record (i.e. mean, standard deviation) are time-invariant. The stationary assumption is very useful because it allows engineers to make data-driven design decisions based upon the information content of the historic record alone. However, climate change's anticipated impacts on the probabilities of extreme events [Hartmann et al., 2013, Rhein et al., 2013] question the validity of the stationary assumption

and the common practice of relying on historic data for engineering design.

In the field of hydrology, there is a debate regarding the "successor" to the stationary assumption for estimating design flow rates, e.g., 100 year discharge estimates commonly used for the design of bridges and mapping floodplains. The climate modeling community tends to argue for incorporating general circulation model (GCM) projections into hydrologic design and risk assessment [Milly et al., 2008, 2015], applied statisticians often present more complex stochastic models that explicitly characterize changing probabilities [Katz, 2013, Cheng et al., 2014b, Renard et al., 2006b], whereas several researchers have argued and demonstrated that more complex approaches increase predictive uncertainty and therefore the stationary assumption should be preferred despite its limitations [Stedinger and Griffis, 2011b, Luke et al., 2017, Serinaldi and Kilsby, 2015b]. In engineering practice, the most commonly adopted approach to account for the so-called "stationarity issue" in design flows is to multiply estimates derived from the traditional methods by a safety factor to account for the uncertainty in future climate [Madsen et al., 2013]. Turning our attention now to coastal processes, the stationarity issue is different and it is unlikely to be resolved with a safety factor approach.

There have already been documented changes in the magnitude of extreme sea level events [Menéndez and Woodworth, 2010], unlike changes in hydrologic (riverine) extremes that are yet to be observed [Villarini et al., 2009a]. Most tide gages contain trends in extremes due to mean sea level rise, and there is widespread consensus that rates of sea level rise are likely to accelerate throughout the 21st century [Church et al., 2013]. These factors make it difficult to justify infrastructure design and planning based on historic data alone, since historic records do not contain information on the anticipated acceleration of sea level rise. Moreover, projections of mean sea level throughout the 21st century and beyond are widely available and relatively well developed compared to GCM projections of precipitation at regional scales [Kopp et al., 2014, Flato et al., 2013]. From a practical point of view, academic arguments

for or against using sea level rise projections in engineering are becoming irrelevant: many agencies now require projections to be used in design and planning [Engineers Australia, 2013, California Coastal Commission, 2015, USACE, 2013]. So, what's problem? Why can't engineers simply adopt projections of mean sea level for design purposes? There are several key issues which create barriers for making sea level rise projections useful for engineering.

First and foremost, modern projections of mean sea level all result from hypothetical emission scenarios adopted by the Intergovernmental Panel on Climate Change (IPPC)[Van Vuuren et al., 2011a]. Despite recent advances that assign probability distributions to sea level rise projections [Kopp et al., 2014], each sea level rise distribution is conditioned upon an emission scenario. Moreover, the IPCC refrains from assigning relative likelihoods to the emission scenarios [Van Vuuren et al., 2011a], making them very difficult to use for engineering purposes. For example, there is no distinction between the likelihood of representative concentration pathway (RCP) 8.5, which depends on socioeconomic narratives where coal dominates the market share of primary energy supply after 2040 [Ritchie and Dowlatabadi, 2017], compared to RCP 2.6, which is attainable if society aggressively adopts carbon capture technology [Van Vuuren et al., 2011b]. Since sea level rise projections under the different emission scenarios vary considerably after 2050 [Kopp et al., 2014], the choice of RCP scenario becomes important in the design process. Choosing an RCP for design without information on the likelihood of these scenarios changes a relatively objective design process founded in extreme value analysis to a highly *subjective* scenario planning exercise that is dependant upon the selected (or prescribed) emission scenario. This problem clearly complicates design and planning in coastal areas.

Secondly, within a given emission scenario there is a considerably wide range of sea level rise projections. For example, Kopp et al. [2014] project with 95% confidence between -2 cm and 85 cm of relative sea level rise for Crescent City, CA by the year 2100 under RCP scenario 4.5. This implies that there could be negative sea level rise relative to land north of Cape

Mendocino, or a rise of 85 cm (2.8 ft) by the year 2100. Variance in models' predictions of thermal expansion, glacier mass loss, and ice-sheet mass account for the majority of uncertainty in sea level rise projections [Church et al., 2013]. Indeed, the additional uncertainty introduced by model projections is new for engineers: uncertainty associated with traditional estimates of risk and return periods is attributable to limited sample size and the selected frequency distribution. These uncertainties on their own are already significant [Stedinger and Griffis, 2011b].

The last issue addressed in this paper is the challenge of translating mean sea level rise projections into corresponding estimates of the return periods and risk of extreme water levels. The problem is that sea level rise projections do not directly express concepts that have traditionally provided the basis for design decisions. Projected changes in mean sea level are often added to the present day extreme value distribution to estimate future risk, e.g., the current 100 year water level plus projected mean sea level rise is used to assess vulnerability [Mudersbach and Jensen, 2010]. However, this approach does not provide information describing 1) expected waiting time until an exceedance event, which is analogous to a return period and can be useful for evaluating a project's lifespan, or 2) the probability of an exceedance event throughout a design life, which can be used to understand a project's risk. By using non-stationary extreme value analysis, or assuming the statistical properties of the extreme value distribution vary with mean sea level, it is possible to estimate both of these fundamental design metrics given different sea level rise trajectories.

The purpose of this paper is to present methodology for engineering design that addresses the three issues outlined above: subjective likelihood of emission scenarios, highly uncertain sea level rise projections, and translating projections into traditional design concepts. We emphasize uncertainty analysis of waiting time and risk estimates to provide guidance on the utility of these metrics in the context of sea level rise projections. The remainder of the paper is organized as follows. First, we briefly review the concepts of return periods

and risk and show how they can be extended to a non-stationary framework that assumes time-variant probabilities (Section 2.3.1). Then we present a non-stationary (NS) model that can be used to synthesize extreme value analysis with projections of mean sea level (Section 2.3.2). After presenting the NS model, we introduce a Monte Carlo experiment that is used to quantify uncertainty in expected waiting time and risk estimates resulting from uncertain emission scenarios, sea level rise projections, and limited historic sample size (Section 2.3.3-2.3.4). Our strategy for assigning probabilities to the RCPs is explained in Section 2.3.4, along with a brief overview of the emission scenarios. We then introduce the "speculation" index, which measures the sensitivity of expected waiting time and risk estimates to weights assigned to future emission scenarios (Section 2.3.6). In sections 2.4, we present the results of the methodology applied to the Los Angeles gage and demonstrate how the speculation index can be used to derive a critical infrastructure design water level and an objective planning and design horizon. We discuss the implication of these results in Section 2.5 and summarize and conclude the paper in Section 2.6. Appendix B includes details on the Bayesian methods used for parameter inference.

2.3 Methods

2.3.1 Return Periods and Risk of Non-stationary Water Levels

There are several definitions of return periods in the literature [Olsen et al., 1998, Cooley, 2013], and herein we adopt the expected waiting time definition because it is easier to interpret for design life applications. In this paper, the return period $T(z)$ represents the expected waiting time (in years) until water levels exceed elevation z. The term "expected" is used in a statistical context, i.e., the expected waiting time is the average waiting time until an exceedance of z. For simplicity and consistency with standard frequency analysis of tide

gages in the U.S. [e.g., Zervas, 2013b], we apply frequency analysis to annual maxima water levels throughout this paper. By assuming 1) annual maxima water levels are independent events and 2) the probability of an exceedance event remains constant, or the distribution of annual maxima is stationary, calculating $T(z)$ is simple [Gumbel, 1941]

$$T(z) = \frac{1}{1 - F(z)} \tag{2.1}$$

where $F(z)$ is the cumulative distribution function of random variable Z, which defines the annual probability that Z is less than or equal to water level z, or $F(z) = \mathrm{P}(Z \leq z)$. Notice that the expected waiting time until an exceedance of level z is simply the inverse of the annual exceedance probability under the independence and stationary assumption.

For estimating risk, we use the interpretation presented by Salas and Obeysekera [2013]. Herein, risk $R(z,t)$ is defined as the probability of an exceedance of design level z before or during the last year of the design lifespan t. Under this definition of risk, $R(z,t)$ is the probability of overtopping before the end of the design life. By applying the same independence and stationary assumption, calculating $R(z,t)$ is also straightforward [Salas and Obeysekera, 2013]:

$$R(z,t) = 1 - F(z)^t \tag{2.2}$$

Notice that under the stationary assumption, both $T(z)$ and $R(z,t)$ are defined by a constant distribution function $F(z)$. The limitation of this assumption in the context of sea level rise is that $F(z)$ will change as mean sea level increases, and therefore $T(z)$ and $R(z,t)$ are not meaningful over long time horizons where we anticipate significant changes to $F(z)$.

We now drop the stationary assumption to allow $F(z)$ to vary with time, and in this paper $F(z)$ varies as a function of sea level rise trajectory. The expected waiting time for water levels to exceed level z when we assume the probability of exceedance changes with respect

to time is given by Olsen et al. [1998], Cooley [2013], Salas and Obeysekera [2013]:

$$T(z) = 1 + \sum_{y=1}^{\infty} \prod_{i=1}^{y} F_i(z) \tag{2.3}$$

where $F_i(z)$ is the time-varying distribution function of z, and the variables y and i are both indices of years. It is important to note that the expected waiting time until an exceedance of level z is infinite for cases when $F_i(z)$ approaches unity, or the exceedance probability of z goes to zero. In the context of extreme ocean levels, this situation occurs when projections of mean sea level predict a decrease in sea levels relative to land. On the other hand, when projections predict and increase in sea level relative to land, $F_i(z)$ goes to zero as the exceedance probability of z approaches unity. In this case, the infinite sum in Equation 2.3 can be stopped in the year $y = y_{\max}$, or the year when $F_i(z)$ equals zero. When $F_i(z)$ approaches zero, $T(z)$ converges to a finite value that represents the average waiting time until exceeding z. Approximations of Equation 2.3 are also given by Cooley [2013]. The risk of water levels exceeding level z before the end of design life t is easier to interpret under non-stationary conditions and is also given by Salas and Obeysekera [2013]:

$$R(z,t) = 1 - \prod_{i=1}^{t} F_i(z) \tag{2.4}$$

Notice that Equation 2.4 is a fairly straightforward extension of Equation 2.1 to allow $F(z)$ to vary with time.

The important consequence of calculating $T(z)$ and $R(z,t)$ as a function of non-sationary (time-varying) probabilities is that $T(z)$ and $R(z,t)$ are now relative to a specific starting year. For example, the expected waiting time until an exceedance event starting from a reference year of 2000 will be different from the expected waiting time from 2050. This issue can be addressed by adding $T(z)$ to the reference year, or the year corresponding to $t = 0$, so that $T(z)$ is converted to a specific year when an exceedance event is expected.

The authors consider this consequence an advantage; non-stationary $T(z)$ can be used to estimate when sea level rise is expected to cause water levels to exceed proposed or existing infrastructure. Likewise, non-stationary $R(z,t)$ can be used to estimate the probability sea levels will exceed the elevation of existing infrastructure by year t. To determine $R(z,t)$ and $T(z)$ under non-stationary conditions, the time varying distribution function $F_i(z)$ must be defined. In this paper, we use non-stationary extreme value analysis (NEVA) to define $F_i(z)$.

2.3.2 Modeling Water Levels with Non-stationary Extreme Value Analysis

NEVA is an important extension of stationary extreme value analysis which allows the distribution of extremes to change. The change in the distribution is modeled by expressing one or more of the distribution's parameters as a function of time or another external covariate [Coles, 2001, AghaKouchak et al., 2013]. In this study, we use a non-stationary, generalized extreme value (NS-GEV) distribution to define the time dependant distribution function $F_i(z)$:

$$Z \sim \text{GEV}(\mu_i, \sigma, \xi) \tag{2.5}$$

where the random variable Z is the annual maximum ocean level, and σ and ξ denote the scale and shape parameter of the GEV distribution, respectively. The time-variant location parameter μ_i is formulated as a function of changes in mean sea level

$$\mu_i = \Delta\text{MSL}_i + \mu_o \tag{2.6}$$

where ΔMSL_i is the change in mean sea level from a year 2000 baseline to year i, and μ_o is a constant off-set between the location of the GEV distribution and mean sea level. Notice

that μ_i is the location of the GEV distribution relative to mean sea level in the year 2000, and μ_o is the location parameter during the year 2000. Using changes in mean sea level to shift the distribution of annual maxima water levels allows us to couple extreme value analysis with sea level rise projections [Obeysekera and Park, 2012, Hunter, 2010].

According to Equations 2.5 and 2.6, we assume that ΔMSL_i changes the location of the extreme ocean level distribution but does not cause the scale or tail behavior of the distribution to change. Physically, this means that we assume changes in mean sea level increases extreme ocean levels on average, but the magnitude of storm surges (which governs the scale and shape parameters) will not change. This assumption is supported by tide gage records, which demonstrate that changes in extreme ocean levels are primarily attributed to changes in mean sea level, and that relatively few tide gages worldwide exhibit trends in storm surges. It is important to note that a different model should be used if changes in surge magnitude and frequency are detected, which has been reported for several tide gages along the gulf coast of the US [Menéndez and Woodworth, 2010].

To illustrate the concepts presented in Sections 2.3.1 - 2.3.2, Figure 2.1 shows $T(z)$ estimates for extreme water levels in Southern California (CA). Figure 2.1A shows the NS-GEV model (Equations 2.5 - 2.6) fitted to the record of annual maxima ocean levels from the Los Angeles, CA tide gage. The parameter estimation procedure is included in Appendix B. Figure 2.1B shows the projected change in mean sea level defined by the 25^{th} percentile of the RCP 4.5 projections [Kopp et al., 2014], the corresponding change in the GEV distribution according to the fitted model, and the water levels expected to be exceeded by 2050, 2100, and 2150. The blue line represents ΔMSL_i, which is measurable prior to 2018. We use a centered 19-year moving average to define mean sea level for consistency with the Kopp et al. [2014] projections, which project changes in mean sea level relative to a 19-year window centered on the year 2000. After 2018, ΔMSL_i is defined by projections. The red line is the location of the fitted NS-GEV model μ_i, and the red shading shows the 1^{st} - 99^{th} quantiles of the

distribution. Notice that the difference between red and blue lines is μ_o.

The black lines are the water levels we would expect to be exceeded by 2050, 2100, and 2150 for the displayed sea level rise projection and parameterization of the NS-GEV model. Here, return level z was determined by setting $T(z)$ equal to the desired return periods (33, 83, and 133 years) and solving for z using Equation 2.3. The fitted NS-GEV model is used to define $F_i(z)$. Notice that $T(z)$ is relative to a starting year and can be used to determine specific years when exceedances of different elevations are expected. Notice also that the probability of exceeding level z increases after the expected exceedance year. This is shown graphically in Figure 2.1B by the larger probability density (red shading) above return levels as sea levels increase. Since exceedance probabilities increase after the expected year of the first exceedance, $T(z)$ is appropriate for estimating the useful design life of a structure with elevation z (e.g. a sea wall or finished floor elevation). In other words, the protective value of a structure with elevation z decreases significantly after $T(z)$, because sea level rise is expected to cause exceedance probabilities to increase indefinitely. Of course, this is based on the assumption that projected changes in mean sea level materialize.

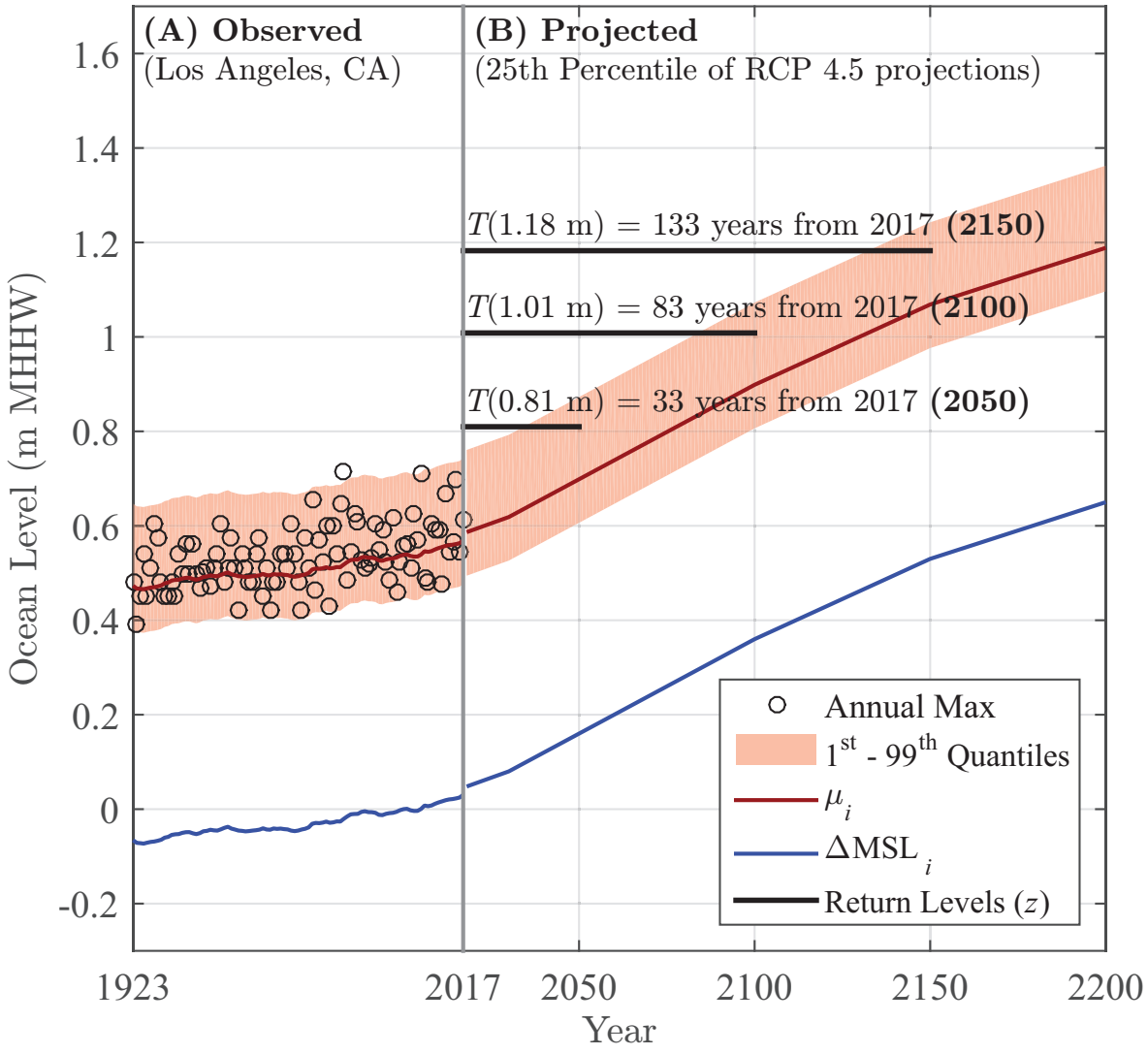

Figure 2.1: A) Observed annual maximum ocean levels, ΔMSL_i, and fitted NS-GEV model. B) Projected ΔMSL_i, changes to fitted NS-GEV distribution, and return levels. The black lines are the water levels we would expect to be exceeded by 2050, 2100, and 2150 for the given sea level rise projection and NS-GEV model fit.

2.3.3 Uncertainty Analysis

Figure 2.1 also demonstrates that in the context of sea level rise, $T(z)$ is a now a function of the probabilities defined by the NS model and projections of sea level rise. These inputs to the $T(z)$ and $R(z,t)$ calculations are inherently uncertain. To estimate confidence intervals for $T(z)$ or $R(z,t)$ based on the uncertain inputs, we can use Monte Carlo experimentation. Monte Carlo experiments are a broad class of algorithms that rely on random sampling to obtain results, and in summary, Monte Carlo based algorithms usually follow a similar pattern:

1. A domain of possible inputs is defined
2. Inputs are randomly generated from a probability distribution over the domain of inputs
3. A computation is performed using the randomly generated inputs
4. Repeat steps 2-3
5. Aggregate results

Accordingly, using Monte Carlo experiments to estimate uncertainty in $T(z)$ and $R(z,t)$ requires us to define probability distributions over the sources of uncertainty. Table 2.1 summarizes the known sources of uncertainty associated with $F_i(z)$ and projected values of its input variable ΔMSL_i. Formal classifications of the sources of uncertainty are provided in the parenthesis according to Kennedy and O'Hagan [2001], except here we use the term "input variability" to classify uncertainties in model forcing variables. The colored sources in Table 2.1 are the sources of uncertainty that we consider in the Monte Carlo experiment. The Table 2.1 footnotes briefly describe our approach for defining distributions over these sources of uncertainty for Monte Carlo sampling, and the following sections describe our strategies in more detail.

Table 2.1: Sources of uncertainty in $T(z)$ and $R(z,t)$ when determined using sea level rise projections. Formal classifications of uncertainty sources are provided in parenthesis.

Extreme Value Probabilities, $F_i(z)$
Observed ΔMSL_i (Model Input Variability)
Unknown Extreme Value Distribution (Model Inadequacy)
Limited Sample Size for Parameter Inference[a] (Parameter Uncertainty)
Imperfect Measurements of Extreme Ocean Levels (Observational uncertainty)
Projected Sea Level Change, ΔMSL_i
Climatic Uncertainties
Future Emissions[b] (Model Input Variability)
Ice Sheet Contributions[c] (Model Inadequacy and Parameter Uncertainty)
Glaciers and Ice Cap Contributions [d] (Model Inadequacy and Parameter Uncertainty)
Thermal Expansion Contributions [e] (Model Inadequacy and Parameter Uncertainty)
Non-Climatic Uncertainties
Water from Land Contributions [f] (Model Inadequacy and Parameter Uncertainty)
Vertical Land Movement [g] (Model Inadequacy and Parameter Uncertainty)

[a] Distribution defined by Bayesian parameter estimation (Appendix B)
[b] Distribution defined by emission scenario weights (Section 2.3.4)
[c - g] Joint distribution defined by Kopp et al. [2014]

The most straightforward source of uncertainty to define probabilistically is the parametric uncertainty of the NS model, which results from limited sample size for parameter inference. The parametric uncertainty is specifically the uncertainty in the estimates of μ_o, σ, and ξ, which are denoted collectively as $\boldsymbol{\theta}$. To define a joint distribution of $\boldsymbol{\theta}$, we use Bayesian inference to estimate the parameters of the NS model. Bayesian inference results in a full distribution of parameter values conditioned on the data, or the probability of $\boldsymbol{\theta}$ given the annual maximum observations from the historic record. In this study, we implement Bayesian parameter estimation using Markov Chain Monte Carlo simulation with the $\text{DREAM}_{(\text{ZS})}$ algorithm developed by Vrugt et al. [2009], Laloy and Vrugt [2012b]. For the sake of brevity, we will not review Bayesian methods in this paper, however we refer the interested reader to Cheng et al. [2014b], Luke et al. [2017], and Vrugt [2016] for applied examples and software for implementation. In the context of this paper, Bayesian parameter estimation provides us with a distribution of parameter values to sample from for the Monte Carlo experiment. In Appendix B, we proved the posterior distribution of parameter values and more details

about the parameter estimation procedure.

Notice that we do not consider the uncertainty in observed ΔMSL_i, ocean level measurements, or extreme value distribution selection. Uncertainties in observed ΔMSL_i and ocean level measurements both result from observational errors in water level measurements, which are on the order of ± 0.005 m for average water levels and ± 0.02 m for individual measurements [NOAA, 2013]. Thus, errors in computed ΔMSL_i and observed water levels are relatively small compared to the uncertainties in projected sea level rise. Regarding extreme value model uncertainty, Arns et al. [2013] showed that probabilities defined by different extreme value distributions are minor for long record lengths. It is therefore unlikely that extreme value distribution selection will contribute significantly to the overall uncertainty in light of the relatively long tidal records available for parameter inference. However, it would be interesting to assess the sensitivity of the estimates to stochastic model selection in a future study. Indeed, the many sources of uncertainty associated with projections of ΔMSL_i are likely to dominate in the $T(z)$ and $R(z,t)$ calculations.

Kopp et al. [2014] provide a probability distribution of ΔMSL_i for a global network of tide gages. The probability distribution reflects the uncertain contribution of glaciers, ice sheets, ice caps, thermal expansion, and water from land to cumulative sea level rise (Table 2.1). The distribution also reflects uncertainties in vertical land movement at the local tide gages, so the local distributions of ΔMSL_i describe uncertainty in sea level rise projections relative to land. To construct the distribution over these sources of uncertainty, Kopp et al. [2014] randomly sampled from distributions of sea level rise contributions resulting from each individual component presented in Table 2.1. Expert elicitation was used to determine the tail of the contributions from the Antarctic ice sheet [Bamber and Aspinall, 2013]. The aggregate of expert elicitation and the random samples provide the full distribution of ΔMSL_i at each tide gage. The Kopp et al. [2014] distributions primarily reflect the contributions of model and parametric uncertainty to the variation in sea level rise projections. The distributions

do not reflect uncertainty in GCM model forcing, or the unknown emission concentrations throughout the 21st century. The Kopp et al. [2014] probabilistic projections of sea level rise are conditioned upon different emission scenarios.

2.3.4 Reviewing and Weighting the RCP Scenarios

Since the RCPs were developed primarily to evaluate climate change impacts resulting from different policies, there are no probabilities or relative likelihoods explicitly attached to the RCP scenarios [Van Vuuren et al., 2011a]. This makes them difficult to use for engineering purposes, since the cost of design will vary with the RCP scenario and there is no guidance from the IPCC regarding their relative likelihoods. However, it is possible to at least subjectively estimate the probabilities of RCP scenarios by evaluating the likelihood of their underlying socioeconomic conditions. In this study, we assign probabilities to the RCP scenarios based on different interpretations of literature that assess the plausibility of the conditions leading to the respective RCPs. Weighting RCPs based on literature interpretation is of course subjective, so we test the sensitivity of the Monte Carlo experiments to the probabilities assigned to the RCP scenarios. The sensitivity of the Monte Carlo experiment to RCP weights expresses the subjectivity of the resulting $T(z)$ and $R(z,t)$ estimates, which we quantify in Section 2.3.6.

In this study, we limit the domain of emission scenarios to RCP 2.6, 4.5, and 8.5 because RCP 4.5 and 6.0 lead to identical sea level rise projections [Kopp et al., 2014]. The domain of emission scenarios is summarized as follows: RCP 8.5 is characterized by increasing greenhouse gas emissions over time and represents scenarios from the literature where no climate policy is enacted [Riahi et al., 2007], RCP 4.5 represents greenhouse gas stabilization scenarios where radiative forcing is balanced after 2100 [Clarke et al., 2007], and RCP 2.6 represents scenarios that lead to very low greenhouse gas concentrations by implementing

carbon capture technology and aggressive mitigation strategies [Van Vuuren et al., 2007]. The RCPs are not tied to a specific scenario that leads to the respective levels of radiative forcing. However, the baseline emission scenarios are characterized by relatively strong assumptions regarding fossil fuel usage and availability throughout the 21st century.

Baseline emission scenarios (i.e. no climate policy scenarios) generally assume compounding growth in energy demand, reduction in the availability of oil and gas, and consequently reliance on coal to meet the growing energy demand [Ritchie and Dowlatabadi, 2017]. Thus, baseline emission scenarios such as RCP 8.5 assume high levels of future coal combustion. The RCP 8.5 scenario in particular assumes an extraordinarily unlikely amount of coal combustion. RCP 8.5 indicates a "return to coal" future where coal is poised to dominate all energy forms by the end of the century [Ritchie and Dowlatabadi, 2017]. The "return to coal" future is based on the assumption that as demand for energy increases, coal reserves will expand indefinitely to meet the growing energy demand. This assumption has not materialized thus far throughout the 21st century. Ritchie and Dowlatabadi [2017] showed that coal reserves have declined by roughly 15% since 2000 despite rising market prices. Moreover, the total carbon emissions necessary to reach or exceed RCP 8.5 would require combusting 80% to 270% more coal than is presently available according to modern reserve estimates [Ritchie and Dowlatabadi, 2018]. The combination of declining coal reserves and the significant expansion that would be necessary to meet RCP 8.5 makes this scenario appear very unlikely. Capellán-Pérez et al. [2016] puts the probability of exceeding RCP 8.5 at 12% based on fossil fuel resource availability, although one could argue that exceeding RCP 8.5 is economically impossible based upon Ritchie and Dowlatabadi [2017]. Thus, weighting the RCP scenarios using trends in coal reserves and current reserve estimates would assign relatively low likelihood to RCP 8.5 compared to 4.5 and 2.6. Depending on one's perspective, an opposing weighting scheme could be argued for by analyzing intended policies resulting from the Paris climate agreement.

The Paris Agreement aims to keep global warming below 2° Celsius and pursue efforts to limit warming to 1.5° Celsius. Mitigation scenarios that limit warming to 2° Celsius are represented by RCP 2.6 [Van Vuuren et al., 2011b], so the likelihood of accomplishing the Paris Agreement is analogous to the likelihood of RCP 2.6. Rogelj et al. [2016] analyzed national plans that describe participating nations' intentions for addressing the climate change challenge after 2020. The submitted plans, known as Intended Nationally Determined Contributions (INDCs), imply a median warming of 2.6° - 3.1° Celsius by 2100 if implemented successfully [Rogelj et al., 2016]. Substantial over-delivery on INDCs is required to maintain a chance of keeping warming below 2° Celsius and thus realizing the RCP 2.6 scenario. Van Vuuren et al. [2011b] reports that to reach RCP 2.6, global emissions need to peak around 2020 which is not realistic even with successful implementation of INDCs. Thus, using INDCs to estimate the likelihood of RCP scenarios would assign relatively low weights to RCP 2.6 compared to 4.5 and 8.5. The INDCs are also "intended", so there is no guarantee of delivery implying that baseline (no-policy) RCPs are still plausible even considering the submitted intentions.

Indeed, the speculative nature of determining the likelihood of future emission scenarios was a major factor in the IPCC's decision to not provide guidance on weighting future warming scenarios [Schneider, 2001]. In the "probability vacuum" that followed, the most common assumption is to assign equal probabilities to the future warming scenarios. In this study, we acknowledge the subjectivity in weighting the RCPs by repeating the Monte Carlo experiment with relative likelihoods based on 1) coal reserve trends and current estimates, 2) analyses of INDCs, and 3) no preference guidance. Table 2.2 shows the weights assigned to the RCP scenarios for the different bases used to assign probabilities. The weights based on recent estimates of coal reserves assume that more recent data on coal availability and usage rules out RCP 8.5, and we will refer to this assumption as the "optimistic" outlook. Weights based on INDCs assume the national plans to limit emissions will not keep warming below 2° Celsius, so the Paris Agreement fails and RCP 2.6 is not possible. We denote weights

Table 2.2: Probabilities assigned to RCP Scenarios based on different literature. The columns are the different RCP Scenarios, the rows are the weighting bases and pseudonyms, and the values in the table are the probabilities assigned to each scenario for the respective weighting schemes.

Weighting Name, Subscript	Basis	RCP 2.6	RCP 4.5	RCP 8.5
Optimistic, "o"	Limited Coal Reserves	0.5	0.5	0.0
Pessimistic, "p"	Paris Agreement Fails	0	0.5	0.5
Neutral, "n"	No Official Guidance	0.33	0.33	0.33

based on the INDC analyses as the "pessimistic" perspective. Lastly, weights based on the no preference guidance assign equal probabilities to all scenarios, and we call this weighting scheme "neutral". We note that there are many more approaches for weighting the RCP scenarios that could be justified by the literature other than those presented in Table 2.2. By assigning probabilities of zero to the extreme scenarios of RCP 2.6 and 8.5 in the pessimistic and optimistic outlooks, respectively, we provide distributions over the RCP domains that approximate a limit to the many possible interpretations of the literature. We could not find strong justifications for ruling out the middle scenario, RCP 4.5. With probabilities defined for emission scenarios, sea level trajectory, and NS-GEV model parameters, we can proceed with the Monte Carlo experiment to produce confidence intervals in the $T(z)$ and $R(z,t)$ estimates.

2.3.5 Summary of Monte Carlo Experiments

The Monte Carlo experiment used to estimate confidence intervals in $T(z)$ and $R(z,t)$ is summarized as follows:

1. Randomly select:
   (a) RCP scenario with probabilities given in Table 2.2
   (b) ΔMSL_i from distribution provided by Kopp et al. [2014] conditioned upon the RCP scenario draw

(c) $\boldsymbol{\theta}$ from the posterior distribution provided by Bayesian parameter estimation (Appendix B)

2. Define $F_i(z)$ using randomly selected ΔMSL_i and $\boldsymbol{\theta}$

3. Calculate $T(z)$ and $R(z,t)$ using Equation 2.3 and 2.4, respectively

4. Repeat Steps 1-3 for $n = 10,000$ trials for each z and t value tested

The Monte Carlo experiment is repeated for the three RCP weighting schemes described in Table 2.2, resulting in three different samples of $T(z)$ and $R(z,t)$ values for each of z and t values tested. For clarity, we denote the random samples of $T(z)$ values resulting from the Monte Carlo experiments as $\mathbf{t}_{\text{w}}^{z} = [T_1(z), T_2(z), ...T_{n=10,000}(z)]$, where the subscript denotes the random sample corresponding to the weighting scheme $\mathbf{w}$ and the superscript expresses the z value. Here, $\mathbf{w} = \{\text{“o”}, \text{“p”}, \text{“n”}\}$, which refers to the different RCP weights in Table 2.2. The random sample of $R(z,t)$ values are represented by $\mathbf{r}_{\text{w}}^{z,t} = [R_1(z,t), R_2(z,t), ...R_{n=10,000}(z,t)]$, where the subscript again denotes weighting scheme and the superscript expresses different z,t combinations.

2.3.6 Measuring Sensitivity to Scenario Weights (Speculation)

The distributions of the $\mathbf{t}_{\text{w}}^{z}$ and $\mathbf{r}_{\text{w}}^{z,t}$ samples for a constant weighting scheme $\mathbf{w}$ characterize the uncertainty of these design metrics as a function of future emissions, sea level rise projections, and limited sample size for NS model parameter inference. However, the sample distributions and the corresponding confidence intervals of these metrics will change with the RCP weights. The similarity between $\mathbf{t}_{\text{w}}^{z}$ and $\mathbf{r}_{\text{w}}^{z,t}$ samples for changing values of $\mathbf{w}$ describes how sensitive the estimates are to the weights assigned to the emission scenarios. Physically, the similarity expresses the sensitivity of the estimates to future emission scenarios. Conceptually, similarity between the $\mathbf{t}_{\text{w}}^{z}$ and $\mathbf{r}_{\text{w}}^{z,t}$ samples for different emission scenario weights

expresses the sensitivity of the estimates to speculation about the likelihood of RCPs.

In this study, we use a re-scaled extension of the Bhattacharyya coefficient to measure the similarity of $\mathbf{t}_{\mathrm{w}}^{z}$ and $\mathbf{r}_{\mathrm{w}}^{z,t}$ samples for different values of $\mathbf{w}$. The d-dimensional Bhattacharyya coefficient measures the similarity between d statistical samples by integration of the samples' overlap [Kang and Wildes, 2015]:

$$B(\mathbf{f}_1, ..., \mathbf{f}_{\mathrm{d}}) = \sum_{k=1}^{p} \Big[\prod_{j=1}^{d} f_j(x_k) \Big]^{1/d} \tag{2.7}$$

where $B(\mathbf{f}_1, ..., \mathbf{f}_{\mathrm{d}})$ is the Bhattacharyya coefficient between samples $\mathbf{f}_1, ..., \mathbf{f}_d$, k is an index of discrete partitions of the samples, and p is the number of partitions. Here, $f_j(x_k)$ is the probability that a member x is within partition k given the sample j. The probabilities are defined by counting the number of members in the sample j that fall within the k^{th} partition and dividing by the total number of members in the j^{th} sample (n). The value of $B(\mathbf{f}_1, ..., \mathbf{f}_{\mathrm{d}})$ is zero if any of the samples do not overlap due to multiplication by zero in each partition, whereas $B(\mathbf{f}_1, ..., \mathbf{f}_{\mathrm{d}})$ is unity for identical samples [Kang and Wildes, 2015].

In the context of this paper, $d = 3$ and we use the 3-dimensional Bhattacharyya coefficient to measure the similarity between the samples of $\mathbf{t}_{\mathrm{w}}^{z}$ and $\mathbf{r}_{\mathrm{w}}^{z,t}$ given the pessimistic, optimistic, and neutral weighting schemes provided in Table 2.2. Specifically, we compute:

$$B_T(z) \quad = B(\mathbf{t}_{\mathrm{o}}^{z}, \mathbf{t}_{\mathrm{p}}^{z}, \mathbf{t}_{\mathrm{n}}^{z}), \qquad \text{for } z = \{z_1, ..., z_{\max}\}; \tag{2.8}$$

$$B_R(z,t) = B(\mathbf{r}_{\mathrm{o}}^{z,t}, \mathbf{r}_{\mathrm{p}}^{z,t}, \mathbf{r}_{\mathrm{n}}^{z,t}), \text{ for } z = \{z_1, ..., z_{\max}\};\ t = \{t_1, ..., t_{\max}\} \tag{2.9}$$

where $B_R(z,t)$ represents the sensitivity of risk samples to RCP weights as a function z and t, and $B_T(z)$ represents the sensitivity of the waiting time samples to RCP weights as

a function of z. Values of B are straightforward to interpret for cases when $B = 0$ and $B = 1$. If $B = 1$, then there is complete overlap between the samples and thus the risk and waiting time values are not sensitive to weights on future emission scenarios for given z or t. In these cases, we can make inferences from the samples objectively. On the other hand, if $B = 0$, then one or more of the samples do not overlap. Completely non-overlapping samples indicates that values are dependant upon emission scenario weights, and thus inferences from the samples are sensitive to one's perspective. When values of B are between zero and unity, they are more difficult to interpret. In these cases, it is more useful to compare the values relative to other locations in the z, t space rather than interpret the meaning of partially overlapping samples. To compare the sensitivity of $R(z,t)$ and $T(z)$ estimates to weights given to future emission scenarios throughout the z, t space, we re-scale the $B_T(z)$ and $B_R(z)$ values:

$$S_T(z) = 1 - \frac{B_T(z) - \min(\{B_T(z_1), ..., B_T(z_{\max})\})}{\max(\{B_T(z_1), ..., B_T(z_{\max})\}) - \min(\{B_T(z_1), ..., B_T(z_{\max})\})} \tag{2.10}$$

$$S_R(z,t) = 1 - \frac{B_R(z,t) - \min(\{B_R(z_1,t_1), ..., B_R(z_{\max},t_{\max})\})}{\max(\{B_R(z_1,t_1), ..., B_R(z_{\max},t_{\max})\}) - \min(\{B_R(z_1,t_1), ..., B_R(z_{\max},t_{\max})\})} \tag{2.11}$$

where $S_T(z)$ is the sensitivity of $T(z)$ estimates to emission scenario weights relative to the other tested values of z, and $S_R(z,t)$ is the sensitivity of $R(z,t)$ estimates to scenario weights relative to the other z, t combinations. Notice that in Equations 2.10 and 2.11, re-scaling results in values of unity for the $T(z)$ and $R(z,t)$ samples most sensitive to the RCP weights, while values of zero are assigned to the least sensitive samples. The measure of relative sensitivity to scenario weights provided by $S_T(z)$ and $S_R(z,t)$ highlights estimates in the z, t space that are most useful for decision making. We refer to this metric as the speculation index.

2.4 Results

2.4.1 Risk Estimates

Let us begin by presenting the risk estimates obtained from the Monte Carlo experiment using water level data from the Los Angeles tide gage (NOAA tide gage no. 9410660) and the Kopp et al. [2014] probabilistic projections of ΔMSL_i for Los Angeles. In Figure 2.2, the box plots illustrate the percentiles of the $\mathbf{r}_{\text{n}}^{z,t}$ samples, or the uncertainty in $R(z,t)$ based on the neutral weights on RCP scenarios. Recall that $R(z,t)$ is an exceedance probability in this study (Equation 2.4), and therefore ranges from 0 - 1. When the box plots are drawn as a single line, this indicates that all Monte Carlo trials resulted in the same $R(z,t)$ value. The $R(z,t)$ values are most uncertain for box plots where the middle 33% is stretched between the lower and upper bounds, indicating that the risk values are approximately uniformly distributed between 0 - 1. In Figure 2.2, t increases from the top panel to the bottom panel, with different values of z shown on the x-axis.

There are two properties of how the $R(z,t)$ distributions change with time that are worth noting. First, as t increases, the range of "at-risk" elevations expands, and the z value of "risk-averse" elevations increases. In this paper, we define "at-risk" elevations as elevations where $R(z,t)$ is distributed between 0 and 1, and "risk-averse" elevations as elevations where 97.5% of the $R(z,t)$ estimates are less than 0.001. For example, Figure 2.2A shows that by 2030, elevations less than 0.5 m MHHW are projected to be exceeded, $R(z,t)$ estimates range from 0.1 to 0.95 for elevations between 0.5 and 1 m MHHW, and $R(z,t)$ estimates are less than 0.001 for elevations 1.25 m and greater at 97.5% confidence. Thus, the range of at-risk elevations is between 0.5 m and 1 m MHHW, while elevations 1.25 m MHHW and greater are defined as risk-averse.

Contrast the $R(z,t)$ distributions by 2030 shown in 2.2A with the $R(z,t)$ distributions by

2200 shown in 2.2D. By 2200, at-risk elevations are between 1 and 4.5 m MHHW, and the risk-averse elevations are now 4.75 m MHHW or greater. Comparison between 2.2A and 2.2D demonstrates that risk-averse elevations for long time horizons are significantly higher than risk-averse elevations in the short-term. This indicates that risk-averse planning and design for long time horizons is significantly more expensive than risk-averse planning in the short-term. Notice also the uncertainty in $R(z,t)$ increases with t for values of z we define as "at-risk".

The second important property of $R(z,t)$ distributions is their increasing variance with t for elevations we define as at-risk. By 2030 and 2050, $R(z,t)$ distributions for at-risk elevations are relatively constrained and easy to interpret. For example, Figure 2.2A shows that for $z = 0.75$ m MHHW and $t = 13$ years from 2017, $R(z,t)$ is about 0.5. For some planning and design decisions, this level of risk may be tolerable and the $R(z,t)$ estimate is very useful. By 2200, the $R(z,t)$ distributions for at-risk elevations become very difficult to use. Figure 2.2D shows that by 2200, elevations between 1 and 4.25 m MHHW have an exceedance probability between 0 and 1. In other words, we do not know the probability of water levels exceeding these elevations by 2200 given the neutral weighting scheme and current sea level rise projections. Using $T(z)$ estimates for long planning horizons is also difficult when $T(z)$ estimates are based on sea level rise projections.

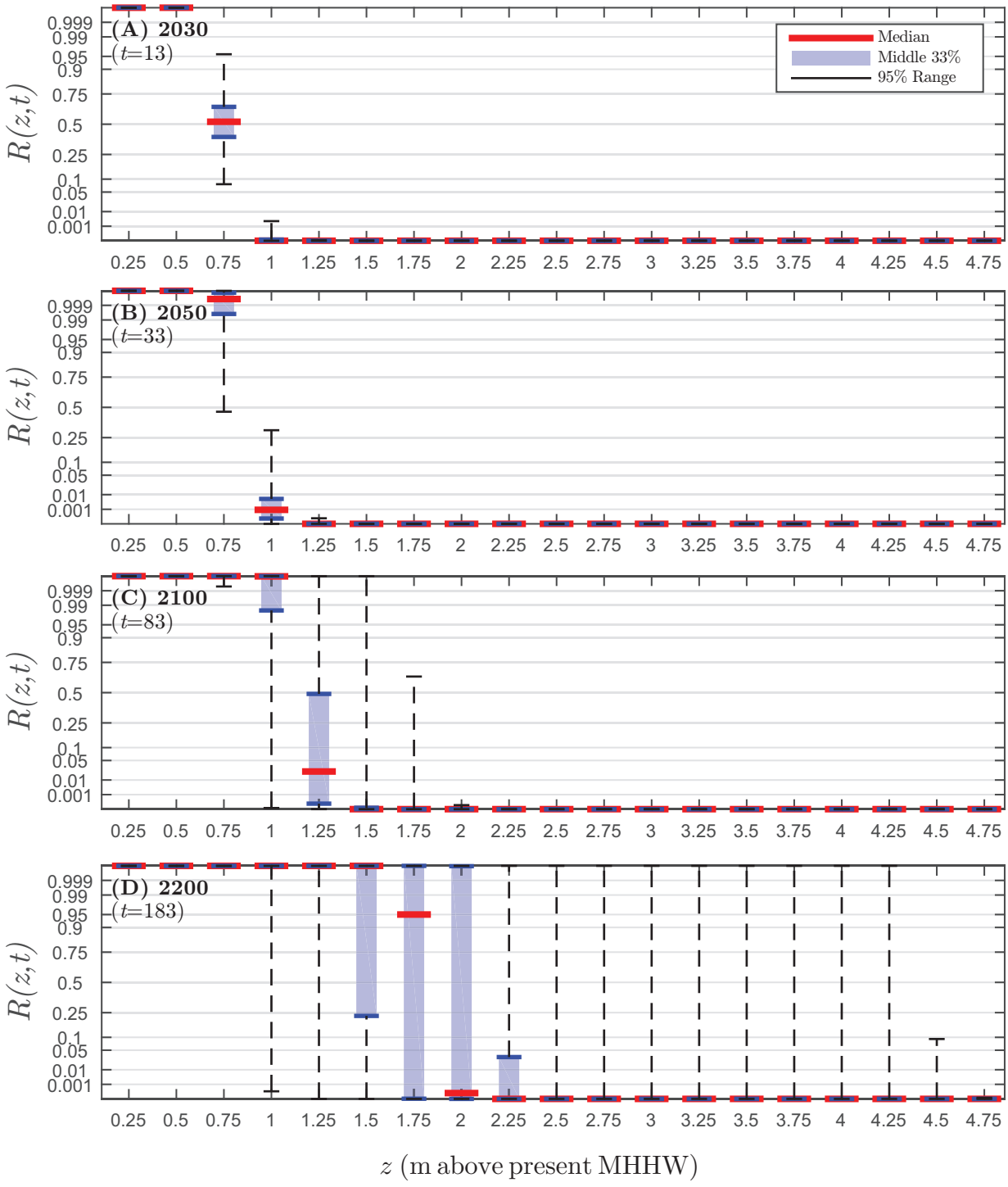

Figure 2.2: Risk of water levels exceeding z by 2030 (A), 2050 (B), 2100 (C), and 2200 (D). The box plots illustrate the distribution of $R(z,t)$ estimates obtained from the neutral weighting scheme. The dashed line encompasses 95% of the estimates (2.5th - 97.5th percentiles), the blue lines contain the middle third (33rd - 67th percentiles), and the red line marks the median estimate (50th percentile). Risk estimates are most certain when the box plots are drawn as a single line, and least certain when the middle third is stretched between 0 and 1.

2.4.2 Waiting Time Estimates

Figure 2.3 shows the expected waiting time estimates from the year 2000. The box plots in the left panel illustrate the percentiles of the $\mathbf{t}_{\mathrm{n}}^{z}$ samples, and the right panel shows the percent of Monte Carlo trials where the randomly selected $\Delta\mathrm{MSL}_i$ and $\boldsymbol{\theta}$ combination did not cause z to be exceeded by 2200. For the Monte Carlo trials where z had zero probability of exceedance by 2200, we classified the $T(z)$ value as greater than 200 years from the year 2000. Since $\Delta\mathrm{MSL}_i$ projections end in the year 2200, it is not possible to calculate the expected year of exceedance for these trials without making strong assumptions about the trajectory of $\Delta\mathrm{MSL}_i$ after 2200. However, it is important to retain trials where z is not exceeded by 2200, because these trials provide information describing the probability that the expected waiting time is greater than 200 years. Retaining $T(z)$ values in the sample categorized as "greater than 200" creates an ordinal data-set, i.e., the values have a natural order but we do not know the distance between all values. Hence, we restrict our analysis of the $\mathbf{t}_{\mathrm{w}}^{z}$ sample to positional measures such as the median and different percentiles rather than the mean or standard deviation. We conducted the Monte Carlo experiment for increasing values of z until 97.5% of trials resulted in $T(z)$ classification greater than 200 years. Thus, the maximum z value on the y-axis is very unlikely to be exceeded by 2200.

The most important characteristic of the $T(z)$ distributions is the increasing uncertainty in $T(z)$ values a function of z. As z increases, the variance in $T(z)$ estimates less than 200 years increases significantly, while a larger percentage of estimates become classified as greater than 200 years. For example, we expect an exceedance of 0.75 m MHHW at the Los Angeles gage between 2020 and 2050 with 95% confidence, where the likely exceedance year is close to 2030. The expected waiting time for an exceedance of just 1 m above MHHW becomes very uncertain, with a 95% confidence range between 2050 and after 2200. The uncertainty in the $T(z)$ estimates is primarily attributable to the large variance in the sea level rise projections. Under RCP 2.6 and 4.5, mean sea levels stabilize and even decrease

after 2100 near the low range of the projections, whereas under RCP 8.5, sea levels are very un-likely to stabilize and appear to increase indefinitely (please see Figure 3 of Kopp et al. [2014]). The possibility of either stabilizing sea levels or indefinite rise causes the expected waiting times of higher z values to be especially uncertain.

While the box plots are drawn horizontally, the uncertainty bounds can also be interpreted vertically to estimate the level z we would expect to be exceeded in year $T(z) + 2000$. For example, Figure 2.3 shows that the water level we expect to be exceeded in Los Angeles by 2100 is between 1 and 1.75 m MHHW with 95% confidence, where the likely value is around 1.25 m. The water level we would expect to be exceeded by 2200 is even more uncertain, with a 95% confidence range between 1 and 4.25 m MHHW. The increasing uncertainty in the estimates for high values of z and $T(z)$ makes them very difficult to use for long planning horizons, even when considering only the neutral weighting scheme shown in Figure 2.3. If we consider the other plausible weighting schemes on future climates, risk and weighting time estimates become even more difficult to use for long-term planning.

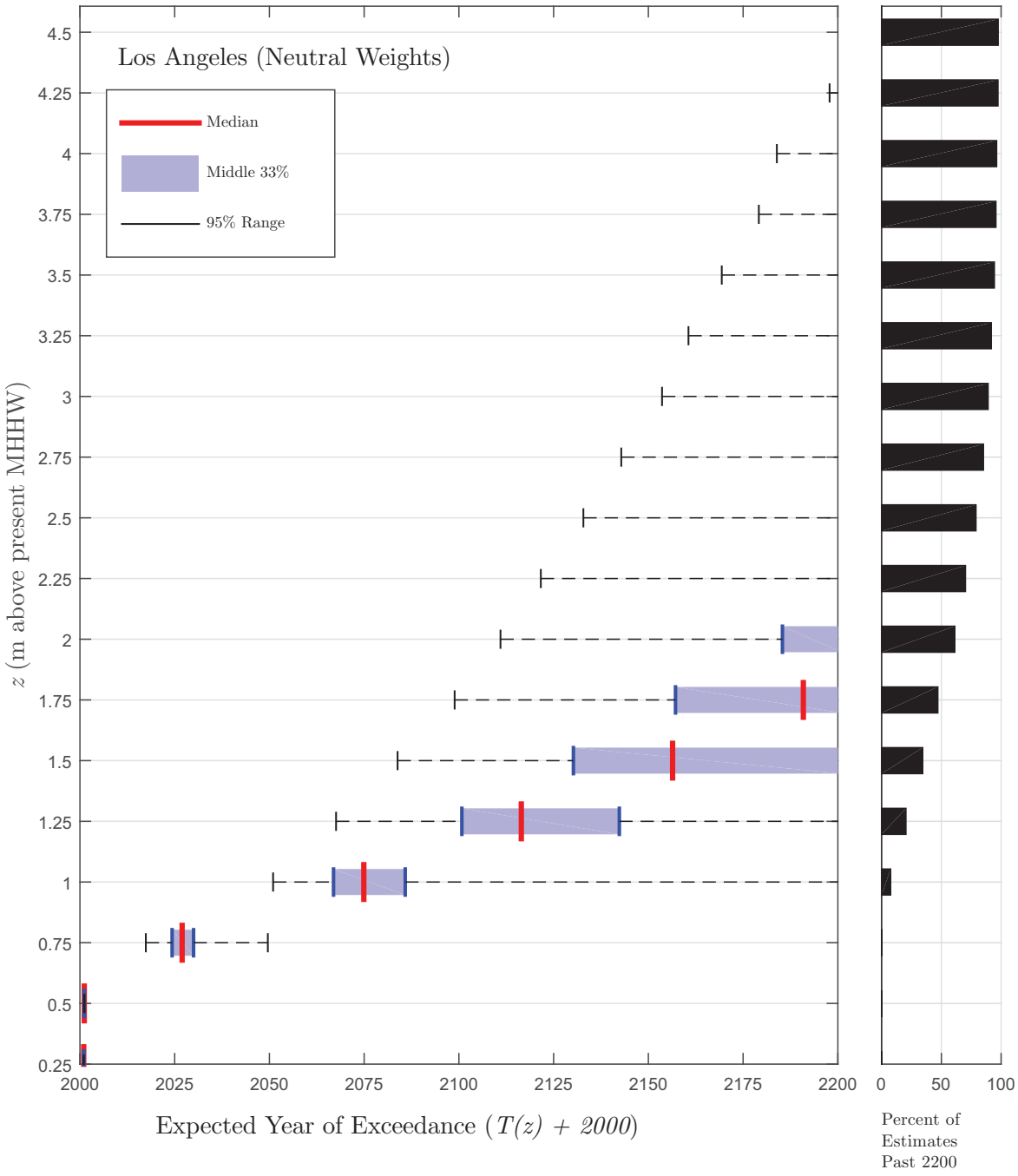

Figure 2.3: Waiting time from the year 2000 water level are expected to exceed z. The box plots describe the distribution of $T(z)$ estimates obtained from the neutral weighting scheme. The dashed line encompasses 95% of the estimates (2.5th - 97.5th percentiles), the blue lines contain the middle third (33rd - 67th percentiles), and the red line marks the median estimate (50th percentile). The black bars on the right panel display the percent of Monte Carlo trials where the $T(z)$ value was classified as greater than 200 years.

2.4.3 Sensitivity to Scenario Weights

In Figure 2.4, we plot the distribution of the $R(z,t)$ samples for each RCP weighting scheme. The weighting schemes are distinguished by color as displayed in the Figure 2.4A legend, and the width of the box plots correspond to the distribution percentiles. The widest (horizontal) line of the box plot is the 50th percentile, the next thickest line spans the middle 33% of the estimates, and the narrowest line is the 95% range.

Notice that the $R(z,t)$ samples overlap almost completely for $t = 13$ and $t = 33$ years from 2017 shown in Figure 2.4A and B, respectively. The sample overlap between the weighting schemes demonstrates that inference of $R(z,t)$ estimates for these t values is not sensitive to weights given to future climate scenarios. Here, inference of $R(z,t)$ would be the same regardless of one's perspective about future emissions, and hence the value of the speculation index will be relatively low. Conversely, for $t = 83$ years from 2017, the distributions are sensitive to weights given to future RCP scenarios. For example, the median $R(z,t)$ value for $z = 1.25$ m MHHW and $t = 83$ years from 2017 is about 0.2 for the pessimistic weighting scheme, whereas the median $R(z,t)$ value is less than 0.001 for the same z and t values according to the optimistic weights. Hence, the median $R(z,t)$ estimate differs by two orders of magnitude between the optimistic and pessimistic perspectives and is therefore highly speculative.

To highlight areas in the z and t domain that are least sensitive to scenario weights, Figure 2.5 plots the speculation index ($S_R(z,t)$, Equation 2.11) for different combinations of z and t. The color bar represents how sensitive the $R(z,t)$ distributions are to scenario weights, with white-yellow colors indicating insensitivity and red-black colors denoting sensitivity. Figure 2.5 demonstrates that the $R(z,t)$ estimates are relatively insensitive to scenario weights for $t + 2017 < 2055$ and $z < 0.75$ m MHHW. These areas of the $R(z,t)$ domain are insensitive to scenario weights because sea level rise projections under the RCP scenarios overlap in

Figure 2.4: Risk of exceeding water level z by 2030 (A), 2050 (B), 2100 (C), and 2200 (D) according to the different perspectives on emission scenarios provided in Table 2.2. The width of the box plots corresponds to the distribution percentiles, where the widest (horizontal) line of the box plot is the 50^{th} percentile, the next thickest line spans the middle 33% of the estimates (33rd - 67th percentile), and the narrowest line is the 95% range (2.5th - 97.5th percentile). Differences between the distributions indicate that inference of $R(z,t)$ values is sensitive to perspective about future emission scenarios.

the near term (i.e. low t values), and relatively low elevations are projected to be exceeded regardless of emission scenario (i.e. $z < 0.75$ m MHHW).

In Figure 2.5, we plot a line that marks the length of a relatively objective planning horizon. The objective planning horizon represents the length of time where the risk of exceeding all levels of z is relatively objective. Specifically, we define the objective planning horizon as the largest value of t where $S_R(z,t)$ is less than average for all values of z. When computing the average value of $S_R(z,t)$, we only use $S_R(z,t)$ values greater than 0 so that the average value is not sensitive to expanding the z, t domain to higher or lower z values. Notice that within the objective planning horizon, we have relatively high confidence in $R(z,t)$ values for all z, and estimates are not sensitive to scenario weights. Notice also that $R(z,t)$ estimates are insensitive to scenario weights for high values of z for increasing t.

$R(z,t)$ estimates for relatively high values of z are also insensitive to scenario weights because they are not at-risk of exceedance regardless of weights on future emission scenarios. Elevations that are not at-risk of exceedance according to all weighting schemes are useful for the design of critical infrastructure because we can objectively state that these elevations have a very low probability of exceedance. We define the critical infrastructure water level plotted in Figure 2.5 as the "risk-averse" elevation according to all weighting schemes, or the z value where $R(z,t)$ is less than 0.001 at 97.5% confidence for each weighting scheme. The critical infrastructure water level can be viewed in Figure 2.4 as the elevation where the 95% range is less than 0.001 for all boxplots. Notice that the critical infrastructure water level, which can be thought of as the objectively risk-averse elevation, increases rapidly with t. The pattern of insensitivity to scenario weights for relatively low and high z values is also apparent in the $T(z)$ estimates.

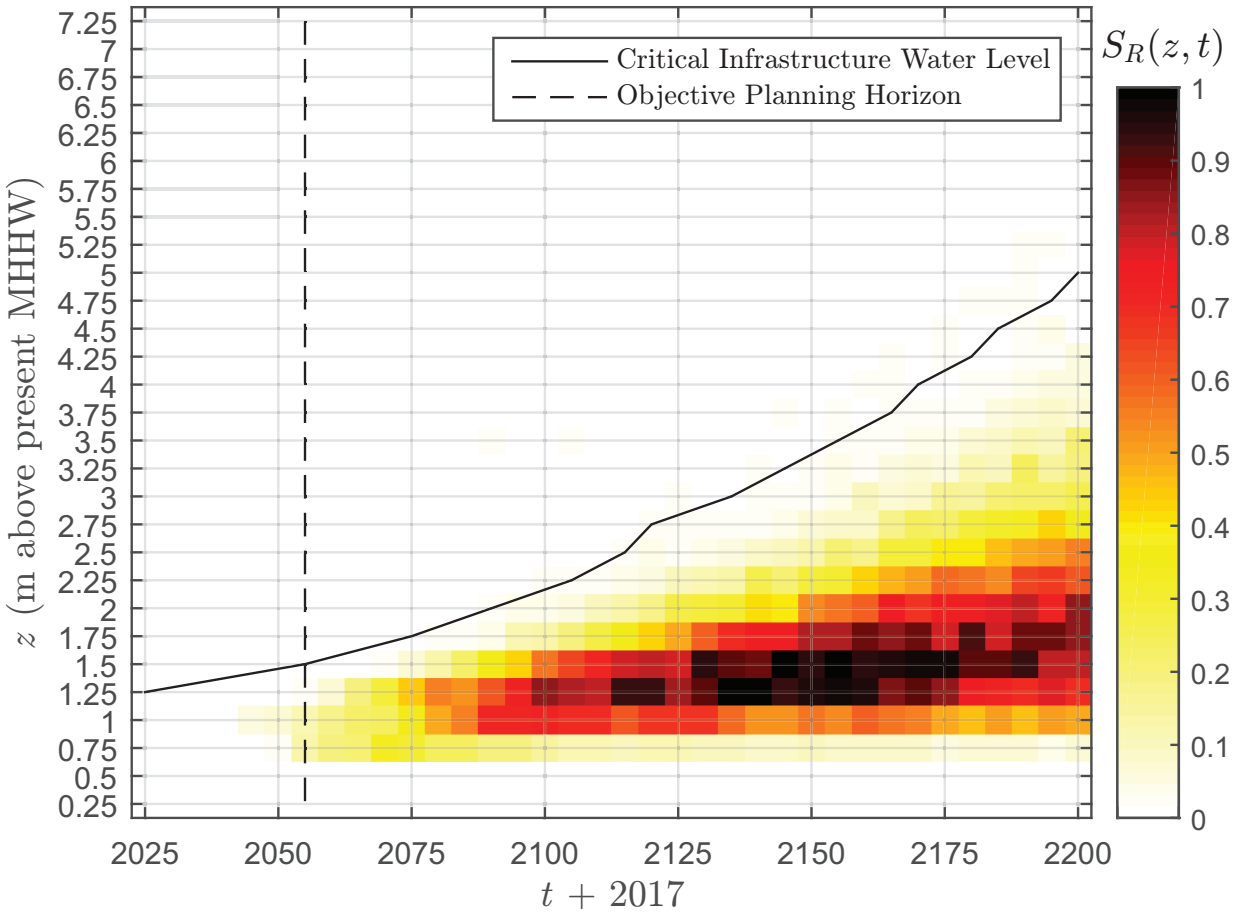

Figure 2.5: The risk speculation index, $S_R(z,t)$, for different z, t combinations. Red/black colors indicate that $R(z,t)$ estimates are relatively sensitive to RCP scenario weights, and white/yellow colors indicate insensitivity for the given z and t combination. The critical infrastructure water level is the elevation that has an $R(z,t)$ value of less than 0.001 for all weighting schemes at 97.5% confidence. We can consider this water level objectively risk-averse. The objective planning horizon marks the largest t value where the speculation index is less than average for all z values. Within the objective planning horizon, we have high confidence in $R(z,t)$ estimates for all values of z.

Figure 2.6 shows the $T(z)$ distribution for all weighting schemes tested. For each z value shown on the y axis, the central panel includes the box plots of the $\mathbf{t}_{\mathrm{o}}^{z}$, $\mathbf{t}_{\mathrm{p}}^{z}$, and $\mathbf{t}_{\mathrm{n}}^{z}$ samples, which are designated by the different colors shown in the legend. The width of the box-plot corresponds to distribution percentiles, with the thickest (vertical) line denoting the median, the second thickest spanning the 33% range, and the narrowest lines covering the 95% range of the estimates. The right panel shows the percent of Monte Carlo trials where z was not exceeded by 2200 according to the different weighting schemes.

Notice that the samples over-lap completely for low values of z, but begin to diverge as z increases. The samples eventually converge again for the highest values of z on the y axis. For relatively high values of z, i.e., z values greater than 4.75 m MHHW at the LA gage, the vast majority of Monte Carlo trials predict that these elevations are not expected to be exceeded until after 2200. This is demonstrated in the right panel, which shows greater than 97.5% of estimates past 2200 regardless of the weighting scheme used. While the year of exceedance remains unknown for high values of z, the Monte Carlo experiment provides very valuable information for these elevations: An exceedance of 4.75 m MHHW is very unlikely prior to the year 2200 regardless of weight given to emission scenario.

Conversely, the Monte Carlo experiments provide ambiguous information about $T(z)$ estimates for z values near 1.5 m MHHW. Figure 2.6 shows that the median waiting time estimate varies by 50 years between the pessimistic and optimistic emission scenario weights. Hence, inference of $T(z)$ values near $z = 1.5$ m MHHW is sensitive to weight given to future emission scenario and is therefore highly speculative. The presence of sample overlap for low z values, divergence for medium z values, and eventual convergence for the highest elevations is captured by $S_T(z)$ shown in the left panel. The speculation index indicates that the Monte Carlo experiment is most sensitive to scenario weights for z values near 1.5 m MLLW, and least sensitive for values below 0.75 m and above 4.75 m.

It is useful to interpret the $S_T(z)$ and $S_R(z,t)$ indices in the context of their physical meaning

and from a planning and design perspective. In the physical context, low values indicate that the $T(z)$ and $R(z,t)$ estimates are relatively insensitive to the difference in emission concentrations between the RCP scenarios. Emission scenarios do not diverge significantly in the short-term [Van Vuuren et al., 2011a], which explains why $T(z)$ and $R(z,t)$ samples overlap for low z and t values. For long time horizons, emission concentrations are of course very different between the RCPs. However, $T(z)$ and $R(z,t)$ estimates are also insensitive to emission concentrations for very high z values, since very high z values have no probability of exceedance regardless of emission concentrations.

From a planning and design perspective, the $S_T(z)$ and $S_R(z,t)$ indices are also significant. Low values indicate that inference of risk and return period estimates is independent of how much weight is given to the different emission scenarios. In these cases, selection of appropriate design water levels is relatively objective. On the other hand, high $S_T(z)$ and $S_R(z,t)$ values indicate that inference of waiting times and risk is dependant on the weights given to future emission scenarios. Dependance on emission scenario weights complicates design and planning by introducing speculation in the inference of appropriate design water levels. From this perspective, the $S_T(z)$ and $S_R(z,t)$ indices highlight areas in the z, t space where risk and waiting times estimates are the most straightforward to use for planning and design. Inference that is sensitive to speculation should be avoided in order to support objective and risk-based planning.

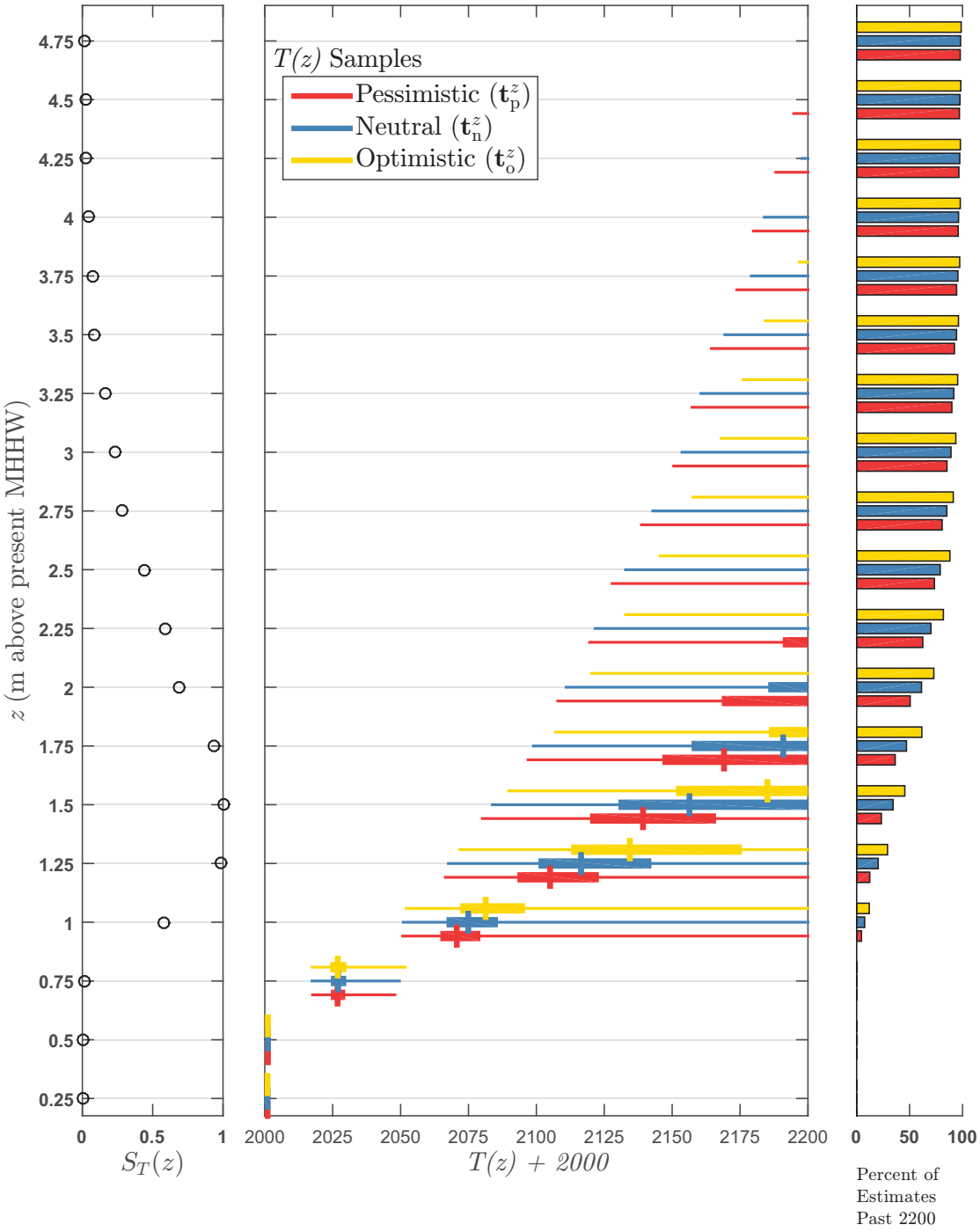

Figure 2.6: The expected waiting time speculation index, $S_T(z)$, for increasing values of z (left panel). The central panel shows the distribution of $T(z)$ estimates for all weighting schemes, which are distinguished by color shown in the legend. The width of the box-plot corresponds to distribution percentiles, with the thickest (vertical) line denoting the median, the second thickest spanning the 33% range, and the narrowest lines covering the 95% range of the estimates. The right panel shows the percent of $T(z)$ estimates we classified as greater than 200 years for each Monte Carlo experiment.

2.5 Discussion

The results of this study highlight the problems with long-term decision making in the context of sea level rise. There are essentially two options for selecting a suitable water level for long-term planning or infrastructure design: we can make costly decisions and drastic plans based relatively high, yet objectively risk-averse design water levels (Figure 2.5), or we can base long-term decisions on lower design water levels where our estimates of risk and waiting times are highly uncertain (Figures 2.3 and 2.2) and sensitive to speculation about the likelihood of future emissions scenarios (Figures 2.4 and 2.6). Naturally, planners and engineers are inclined to choose the former option. However, assuming that the cost of design and planning increases with design water level, long-term risk-aversion for all decisions would be extremely costly.

In the short-term (i.e. prior to 2055 at the Los Angeles gage), we have a relatively good understanding of risk and waiting time estimates. Moreover, risk-averse elevations are relatively low in the short-term compared to the long-term, and inference of design water levels is not sensitive to perspective about future emission scenarios. Clear understanding of risk and expected waiting times allows engineers and planners to 1) make risk based decisions when selecting design water levels and 2) estimate how long a protective measure will be useful. Knowledge of the useful design life allows planners and engineers to understand *when* protective measures need to be re-evaluated and possibly enhanced. Thus, in order to support risk based decision making and utilization of familiar engineering design concepts, we recommend designing protective infrastructure to have short (yet predictable) lifespans with the option of elevating or enhancing the protective measure in the long-term.

The design paradigm of protecting in the near term with the option of enhancement in the long-term supports the need for *adaptive* infrastructure design. By adaptive, we mean designs that allow for low-cost additions in the future to protect against higher sea levels.

An example of an adaptive design can be found in the City of Newport Beach. The City is preparing to add a 0.22 m (9 inch) cap to existing sea walls around Balboa Island (personal communication with City Engineer). There will be periodic 1 m (3 ft) gaps in the addition in order to preserve ease of beach access. The gaps are designed to accommodate removable 0.22 m (9 inch) high barriers to provide flood protection during high tide events. The project is a part of the City's plan to raise protective sea walls incrementally through their repair and maintenance program. The addition brings the top of the sea wall to a design elevation between 1 and 1.1 m MHHW across the length of the sea wall. Our analysis shows that this addition buys the City a considerable amount of time.

Figure 2.2 shows that elevations between 1 and 1.1 m MHHW are very unlikely to be exceeded by 2030 (Figure 2.2A), and they probably won't be exceeded by 2050 (Figure 2.2B). Here, we interpret the risk of water levels exceeding 1 - 1.1 m MHHW prior to 2050, so this interpretation is insensitive to perspective about future emission scenarios (Figure 2.5). The expected waiting time until an exceedance of 1 - 1.1 m MHHW is slightly speculative (Figure 2.6), but each perspective about future emissions puts the median value of the exceedance year near 2075. Notice that the expected waiting time estimates indicate that is it possible, yet unlikely, that 1 m MHHW will not be exceeded until after 2200.

The practical advantage an adaptive design approach is that the structure will be useful regardless of how fast sea levels rise throughout the 21st century. If sea level rise tracks towards the low end of projections, then the projective infrastructure will be useful longer than expected. On the other hand, if sea level rise accelerates quicker than expected, then at least it is cost-effective to make additions to the existing structure. Non-adaptive (i.e. permanent) planning and infrastructure decisions that are based on long-term risk-aversion will *only* be useful under the worst-case RCP scenarios and high end of the sea level rise projections. If all decisions are based on long-term risk-aversion, there is a substantial possibility of over designing for sea level rise and generating unnecessary resistance to resiliency

efforts by affected stakeholders. Hence, the critical design water level presented in Figure 2.5 should be reserved for situations where adaptive design and decision making is not possible or extremely costly.

A significant limitation of this study is that the introduced methodology was only applied to the Los Angeles tide gage. Extreme water levels of the California coast are relatively unique in the sense that the surge to tide ratio is very low [Menéndez and Woodworth, 2010], and therefore sea level rise will have a relatively large impact on extreme water levels compared to regions with larger storm surges. However, the limited application scope presents opportunities for further research. What does the critical design water level look like for areas with large surge to tide ratios? Do the areas of the t, z space that are insensitive to climate weights vary with surge to tide ratio? How does incorporation of storm surge projections affect the sensitivity of risk and waiting time estimates to climate scenario weights? Indeed, all of these questions are worth further investigation.

A secondary limitation of this study is the use of only one extreme value model for the Monte Carlo experiment. The shape parameter of the GEV distribution is negative for extreme water level distributions of the California coast [Zervas, 2013b], which causes the distribution of annual maxima to have an upper bound. The upper bound on the GEV distribution certainly produces clearer risk and waiting time estimates, but also raises the question of how sensitive the risk and waiting time estimates are to the selected extreme value model. Indeed, extreme value *model* uncertainty was not explored in this paper and this topic is also worth future investigation.

2.6 Conclusion

In this paper, we present methodology for engineering design that addresses three problems with using sea level rise projections to derive design water levels: subjective likelihood of emission scenarios, highly uncertain sea level rise projections, and translating projections into traditional design metrics. The methodology uses a non-stationary, generalized extreme value model for pairing sea level rise projections with historic data to determine 1) the expected waiting time until water levels exceed a specified elevation, and 2) the risk of water levels exceeding a design elevation by the end of the project's design life. We provide a Monte Carlo procedure to estimate the confidence intervals in the risk and waiting time estimates, and we introduce a "speculation" index to quantify the sensitivity of the Monte Carlo procedure to perspective about the likelihood of future emission scenarios. The speculation index highlights risk and waiting time estimates that are relatively objective as a function of design water level and planning time frame. We use the speculation index to derive a critical infrastructure design water level and an objective planning window for common decisions.

The methodology is illustrated using the historic data and sea level rise projections for Los Angeles, California. Results show that estimates of risk and expected waiting times are remarkably uncertain and speculative in the long-term, i.e., late 21st century and beyond in Los Angeles. Risk and waiting time estimates can only be considered objective over long time periods for extremely high water levels. However, in the short-term (i.e. prior to 2055 for Los Angeles), risk and waiting time estimates are relatively stable and objective for all water levels. The results imply that risk-based and cost-effective decisions are only possible over relatively short design lifespans. The great challenge of planning and designing for highly uncertain and speculative conditions naturally leads to high risk-aversion. However, extreme risk-aversion exposes us to over designing and planning for sea level rise. Hence, we recommend that long-term risk-averse planning for sea level rise be reserved for only the most sensitive infrastructure and non-adaptable decisions. The majority of planning and

design should still be based on what is predictable and not speculative guesses.

Chapter 3

Going beyond the flood insurance rate map

3.1 Chapter Summary

Flood hazard mapping in the United States (US) is deeply tied to the National Flood Insurance Program (NFIP). Consequently, publicly available flood maps provide essential information for insurance purposes, but do not necessarily provide relevant information for non-insurance aspects of flood risk management (FRM) such as public education and emergency planning. Recent calls for flood hazard maps that support a wider variety of FRM tasks highlight the need to deepen our understanding about the factors that make flood maps useful and understandable for local end-users. In this study, social scientists and engineers explore opportunities for improving the utility and relevance of flood hazard maps through the co-production of maps responsive to end-users' FRM needs. Specifically, two-dimensional flood modeling produced a set of baseline hazard maps for stakeholders of the Tijuana River Valley, US, and Los Laureles Canyon in Tijuana, Mexico. Focus groups with natural resource

managers, city planners, emergency managers, academia, non-profit, and community leaders refined the baseline hazard maps by triggering additional modeling scenarios and map revisions. Several important end-user preferences emerged, such as 1) legends that frame flood intensity both qualitatively and quantitatively, and 2) flood scenario descriptions that report flood magnitude in terms of rainfall, streamflow, and its relation to an historic event. Regarding desired hazard map content, end-users' requests revealed general consistency with mapping needs reported in European studies and guidelines published in Australia. However, requested map content that is not commonly produced included: 1) standing water depths following the flood, 2) the erosive potential of flowing water, and 3) *pluvial* flood hazards, or flooding caused directly by rainfall. We conclude that the relevance and utility of commonly produced flood hazard maps can be most improved by illustrating pluvial flood hazards and by using concrete reference points to describe flooding scenarios rather than exceedance probabilities or frequencies.

3.2 Introduction

Management of flooding is a major societal challenge that is only expected to worsen due to several trends including population growth and urbanization [Sundermann et al., 2014], sea level rise [Hallegatte et al., 2013b], intensification of precipitation extremes [Lenderink and Van Meijgaard, 2008, Coumou and Rahmstorf, 2012], and the compounding effects of sea level rise and terrestrial flooding [Moftakhari et al., 2017]. Insured losses from natural disasters have increased globally [Munich Re, 2005], and while the causes of growing losses are complex and debatable, the increasing exposure and value of capital at risk has undoubtedly played a major role [Bouwer, 2011]. Exposure to flooding is particularly acute in the United States (US), where a combination of subsidized flood insurance and homeowner tax incentive has actually encouraged risky development in floodplains and coastal zones [Bagstad et al.,

2007]. Losses from floods and hurricanes in the (US) have *tripled* over the past fifty years [Gall et al., 2011], and the National Flood Insurance Program (NFIP) is operating at a deficit of about $1 billion annually with a debt of over $20 billion owed to the US treasury before considering insured losses from the 2017 hurricane season [Pasterick, 1998, Brown, 2016]. In fact, properties insured by the NFIP represent the second largest liability of the US federal government after the Social Security program [Gall et al., 2011].

The American Society of Civil Engineers (ASCE) has called for a national strategy to address the escalation of flood losses and threats to public safety, but reports that the US public and policy makers have been unwilling to take action despite major hurricanes such as Katrina and Sandy [Traver, 2014]. The ASCE directive aligns with a global paradigm shift in management philosophy away from *flood control* and towards *flood risk management*. Flood risk management (FRM) refers to a portfolio of approaches for reducing risk that is not limited to controlling flood waters with engineered structures, but also includes effective land use planning, emergency response, and personal preparedness. Importantly, FRM accepts that absolute protection is not possible. Comprehensive FRM reduces the reliance on engineered flood defenses, which is of paramount importance in the US due to the marginal condition of levees and lack of federal resources available for maintenance and necessary upgrades [Traver, 2014]. Studies have shown that robust FRM does indeed lead to significant reductions in fatalities and monetary losses [Kreibich et al., 2017, 2005], however Traver [2014] and Merz et al. [2007] both report that effectively implementing FRM relies on stakeholders who understand their exposure and also have access to tools that are useful for managing personal, household, and community risks.

Flood hazard maps are the most commonly used tool for flood risk communication and management. In the European Union (EU), member countries are under a mandate to develop national flood hazard maps, flood risk maps, and FRM plans based upon the mapped information [Council of European Union, 2007]. General guidelines for meeting end-user needs

have been developed based on participatory processes [Meyer et al., 2012, Hagemeier-Klose and Wagner, 2009, Martini and Loat, 2007], which reflect the varying needs of different end-users for different types of information, as well as the need for context-sensitive information. For example, Meyer et al. [2012] present distinctions between the mapping needs for strategic planning personnel, emergency management personnel and the public, and show that geographical factors (e.g., mountains, polders) influence the need for velocity data.

In the US, flood mapping is tied to the NFIP and the resulting Flood Insurance Rate Maps (FIRMs) delineate the spatial extent of inundation with a 1% and 0.2% annual exceedance probability (AEP). As a vehicle designed to administer an insurance program, the FIRM provides essential information for insurance purposes. Properties with federally backed mortgages located within the 1% AEP floodplain are required to purchase flood insurance, while the flood elevations associated with the FIRM are used for insurance underwriting. However, the binary "in or out" floodplain designation by the FIRMs' *thin grey lines* have been criticized for presenting flood risk as definitive and therefore discouraging important flood hazard discourse [Soden et al., 2017]. Burby [2001] also suggests that the effectiveness of the NFIP is limited because FIRMs lack information necessary to integrate flood hazard considerations into local planning. The Federal Emergency Management Agency (FEMA) has recently expanded its mapping efforts through the Risk MAP program [FEMA, 2014], which produces "non-regulatory" flood hazard data such as depth, velocity, and exceedance probability grids in addition to the standard FIRM [FEMA, 2016]. However, the availability of non-regulatory flood hazard data is limited, and the system is not configured to align mapping products with context-sensitive needs for decision-making. In fact, there is a need for US-centric studies and guidelines for producing maps that are useful for a variety of FRM tasks, i.e., making flood hazard data useable for local end-users across vast hydrologic and social conditions. Flood mapping technology has evolved rapidly over the past decade with modern two-dimensional (2D) hydraulic flood modeling software, computing systems, and increasingly available high quality data to produce point-wise flood hazard information

including flood depths, velocities, flood forces and shear stresses [Sanders, 2017]. Hence, a key issue is making this advanced and complex modeling output useable in FRM.

Producing useable scientific information for non-expert end-users is quite challenging, however, as demonstrated by the climate science community [Dilling and Lemos, 2011]. Useable scientific information (or flood hazard maps in this study) must bridge the gap between what the producers of scientific knowledge deem useful and what is actually helpful in practice. If scientists produce information absent of end-user input, then the produced knowledge is not always applicable to solving problems. Conversely, if end-users set the agenda, end-users may require information that is not possible to produce or weak scientifically. The co-production of scientific knowledge attempts to avoid these issues through an iterative process involving domain experts and end-users. Studies of co-production suggest information must be "end-to-end" useful - applicable to the needs of many users, adaptable for disaster planning as well as mitigation, and able to incorporate local knowledge of threats and hazards [Agrawala et al., 2001, Feldman and Ingram, 2009]. Dilling and Lemos [2011] reported that nearly every example of the successful use of climate knowledge resulted from iteration between the producers and users of scientific information, while the process of iteration itself can also uncover new uses that have not been previously identified. Herein, we recognize the lessons learned from the climate science community by engaging in the co-production of flood hazard maps.

Within the study of engaged scholarship there is a broad spectrum of knowledge "co-production", depending on the degree of iterativity and duration of scholarly engagement. In this paper, we report a single iteration of a broader co-production effort known as Flood-RISE, which aims to promote resilience to flooding in Southern California through continued, meaningful efforts of engagement. The goal of the present study is to both deepen understanding about the factors that make flood hazard maps useable at the local scale and to expand the applications of flood hazard mapping via lessons learned from the co-production

of flood hazard maps. In order to meet this objective, an interdisciplinary team of engineers and social scientists developed a set of baseline flood hazard maps for the stakeholders of the Tijuana River Valley in California, US and Los Laureles in Baja California, Mexico (MX). Following baseline hazard mapping by engineers at the two sites, focus groups were held with a diverse group of end-users comprised of 52 local professionals and community members. The focus groups were designed to understand 1) how to improve the clarity and utility of the baseline hazard maps and 2) how to re-configure the hydraulic models to produce relevant data and useful maps for the communities. In addition to many map revisions, several original flood hazard maps were produced as a direct result of this iteration of knowledge co-production.

The paper continues with a description of the two study sites involved in the co-production of flood hazard maps (Section 3.3). In Section 3.4, we present the baseline (pre-focus group) flood hazard maps produced for each site. Detailed descriptions of the methods used to produce each hazard map are included in Appendix C. Section 3.5 outlines the implementation and design of our end-user focus groups. We present the results of the end-user focus groups in Section 3.1-3.6.2, and the new hazard maps that resulted from the focus groups in Section 3.6.3. This paper concludes with a discussion of these Sections in the context of previous studies and current flood mapping practice in Europe, Australia, and the US.

3.3 Study Site Descriptions

The Tijuana River Valley (TRV) and Los Laureles (LL) communities were selected for the co-production exercise due to their on-going FRM challenges. The TRV is adjacent to the US/MX border and forms the Southwestern most corner of the contiguous US (Figure 3.1A-B). The Tijuana River watershed encompasses 4530 km^2 (1750 mi^2) and is characterized by four major water-supply reservoirs which control about 75% of storm water flow. Due to

its location at the mouth of the Tijuana River, the TRV is prone to flooding from Tijuana River discharges, storm tides, ocean swell, and intermittent flows from LL and Smugglers Gulch (Figure 3.1A, Appendix C.1). The concrete-lined Tijuana River Channel was designed to convey spillway discharges from upstream reservoirs safely through downtown Tijuana. Sediment accumulation has reduced the capacity of the Tijuana River channel near the US/MX border and requires dredging to maintain its design conveyance. The TRV has rural housing, an equestrian (business) presence, and government facilities, but land use in the TRV is primarily preserved for natural habitats or agricultural uses [TRNERR, 2010]. Land use in the TRV contrasts the Mexican side of the border where population density is high due to the presence of colonias and formal settlements.

LL is a canyon community of Tijuana located south of the US/MX border in the LL catchment (Figure 3.1C). The catchment is relatively small compared to the Tijuana River Watershed, and flooding is caused by locally intense precipitation. Flows from LL enter the TRV from the south after passing through a small culvert at the watershed outlet. The culvert is fed by a network of concrete lined storm-water channels, which is continually expanding to prevent stream bed erosion and protect adjacent housing developments. The network of concrete lined channels and expanding urban development has increased the discharges that must be conveyed by the culvert. In both study sites, flooding is a known and recurring issue. Ponding has been observed behind the LL culvert during rainstorms, and a culvert blockage lead to severe flooding and subsequent evacuations in LL and southern Imperial Beach (personal communication, March, 2016). In the TRV, flooding occurs when the upstream reservoirs reach capacity and are forced to open their spillways. Spillway discharges have occurred seven times from 1940 to the present [IBWC, 2006]. Indeed, hazard maps which only delineate at risk areas have limited use for locals; stakeholders already know the TRV and LL are flood prone. Hazard maps that depict more than the traditional 100 year flood extent are, however, scarce.

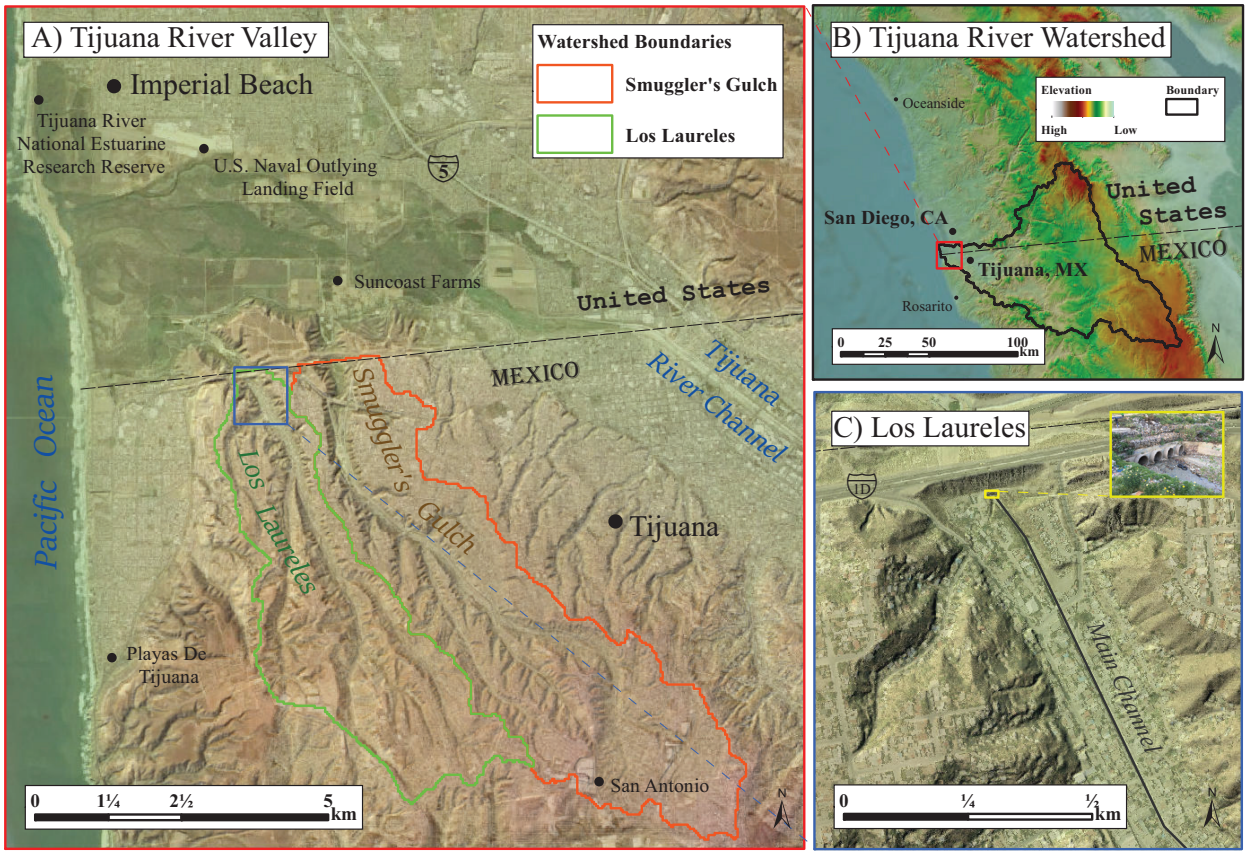

Figure 3.1: A) Tijuana River Valley and relevant features. The Valley is bounded by the City of Imperial Beach to the North, the City of San Diego to the East, and the City of Tijuana to the South. B) Tijuana River Watershed and broader geographical context. Notice the international aspect of the Watershed; about one third of the Watershed area is within the US, and the rest is within Mexico. C) Los Laureles community and Los Laureles main channel. The culvert is also shown, which conveys storm water discharges from the channelized section of the Los Laureles stream network.

3.4 Baseline Flood Hazard Maps

Baseline flood hazard maps were produced to demonstrate to end-users the range of flood hazard data that can be produced with modern methods and to stimulate discussion about hazard map improvement and desired content. At this point, it is helpful to define "flood hazard" as the physical and probabilistic characteristics of floods. The baseline flood hazard maps therefore depict the intensity of spatially varying properties of floods, such as the depth of flooding for a specified probability. We explicitly define flood hazard to distinguish between

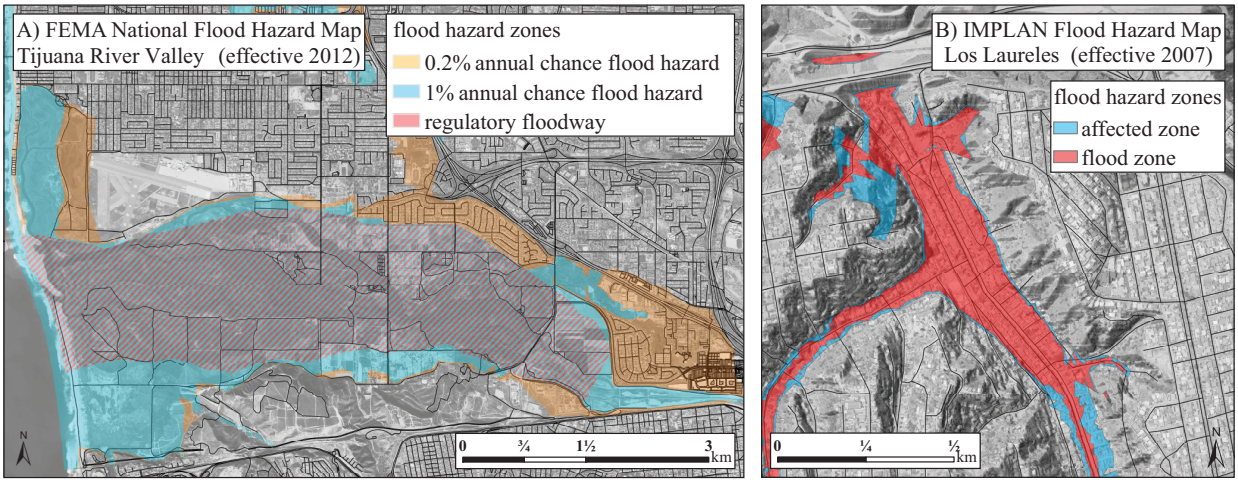

Figure 3.2: A) Federal Emergency Management Agency Flood Insurance Rate Map for the Tijuana River Valley. The *orange* area represents the flooding extent associated with the 0.2% annual exceedance probability (AEP) flood, while the *blue* area represents the flooding extent of the 1% AEP flood. The *red* hashed area delineates the regulatory floodway, where development must not increase the designated base flood elevation by more than 0.3 m (1 ft). B) IMPLAN flood hazard map for the Los Laureles subbasin. The *red* and *blue* zones represent inundation extent for different exceedance probability flooding events. However, details regarding mapping methodology and precise hazard zone explanations were not available.

maps of *hazard* and *vulernability*. Merz et al. [2007] defines vulnerability as the combination of loss susceptibility and damage potential. Thus, flood vulnerability maps show what could be affected by floods and by how much what is affected could be damaged. We restrict our analysis to flood hazard maps because producing vulnerability maps generally requires data that are unavailable in the absence of extensive surveys. In this study, end-users analyzed maps depicting six different flood hazards: depth, force, exceedance probabilities, dominant causes, durations, and extents. The proceeding section presents the hazard maps analyzed by the end-users most applicable to other sites, i.e. we do not present all of the baseline hazard maps herein. The hazard maps we produced were analyzed together with the publicly available hazard maps.

Figure 3.2 shows the publicly available hazard maps, which will serve as the baseline flood *extent* hazard maps in this study. The FEMA FIRM is shown in Figure 3.2A, while Tijuana's

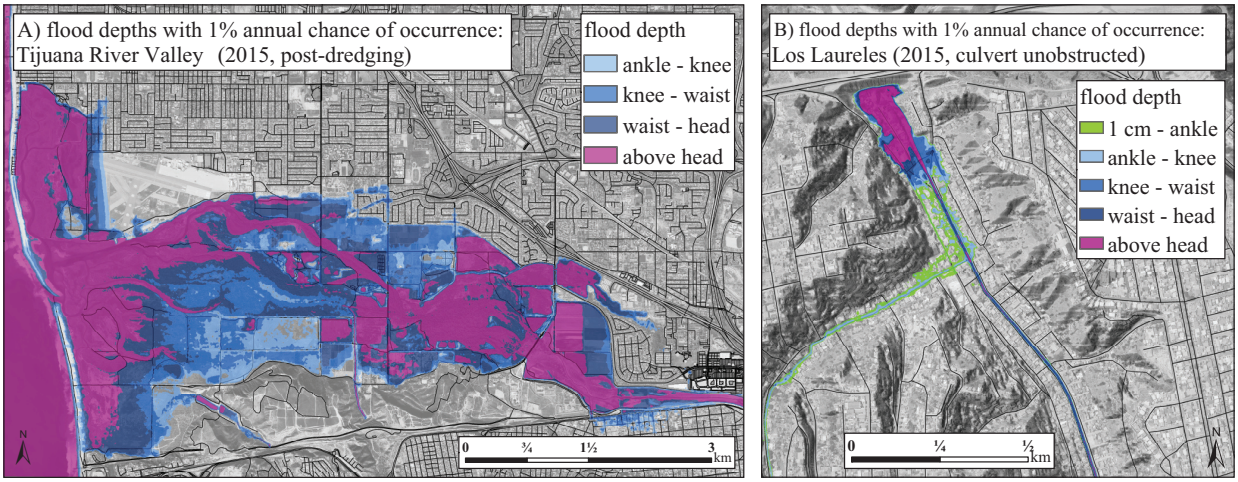

Figure 3.3: A) 1% annual exceedance probability (AEP) flood depths in the Tijuana River Valley. Mapped flood depths result from either storm tides, canyon flows, or Tijuana River discharges (Appendix C.1). The elevation data reflects 2015 topography, after dredging of the Tijuana River channel near the US/MX border. B) 1% AEP flood depths in Los Laureles. Mapped flood depths are caused by streamflows from local precipitation, assuming that the culvert was unobstructed during the duration of the flood. The depth ranges are $0.11 - 0.45$ m for ankle to knee, $0.45 - 1.0$ m for knee to waist, $1.0 - 1.69$ m for waist to head, and greater than 1.69 m for above head depths.

Instituto Municipal de Planeacion (IMPLAN) Flood Hazard map for LL is shown in Figure 3.2B. Both hazard maps depict the inundation extent associated with different AEPs, while the FEMA FIRM also depicts a regulatory floodway. Development in the regulatory floodway is prohibited unless the proposed structure will not increase the base flood elevation by more than 0.3 m (1 ft). Notice also that the flood hazard is described by its associated annual probability, which is consistent with official FEMA mapping legends and terminology. The IMPLAN Flood Hazard map is similar; areas at risk of flooding are shown, but flooding extent is the only characteristic depicted. Flooding extent is also the most commonly mapped hazard in Europe, but many European countries also provide maps of flood depth [Moel et al., 2009, Nones, 2017].

Figure 3.3A shows the analyzed flood *depth* map for the TRV, while the depth map for Los Laureles is shown in Figure 3.3B. Both maps characterize depths using a body scale, and depict depths associated with 1% AEP events. The body scale is based on the average

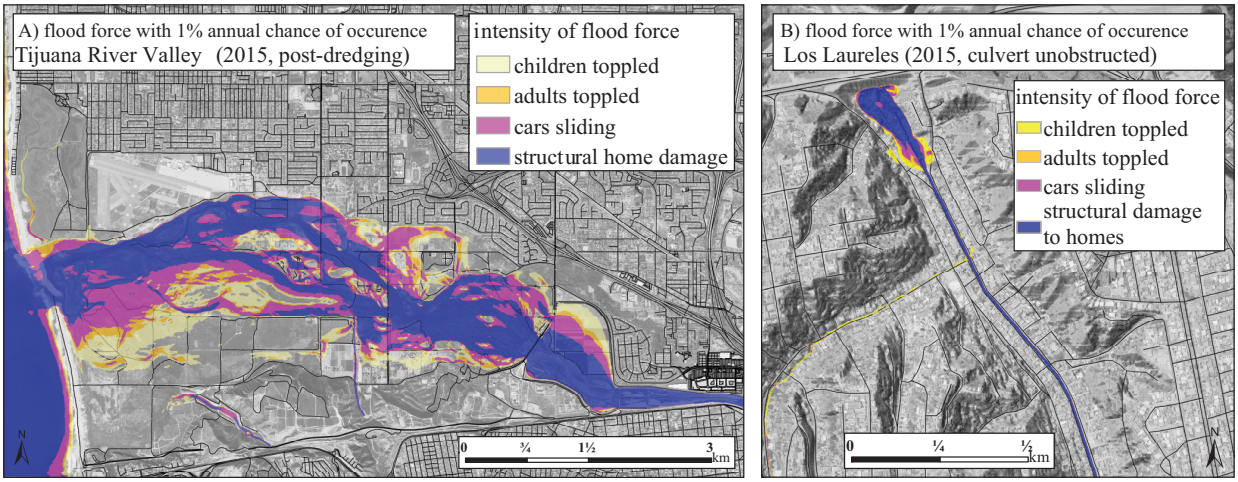

Figure 3.4: A) Flood force with 1% annual exceedance probability (AEP) in the Tijuana River Valley. B) Flood force with 1% AEP in Los Laureles. We use vh as a *proxy* for the flood force, since contours are specifically different thresholds of vh: children toppled at 0.4 m^2/s, adults toppled at 0.65 m^2/s [Xia et al., 2014], cars sliding at 0.8 m^2/s [Xia et al., 2011], and structural damage beginning at 1.5 m^2/s [Kreibich et al., 2009].

person height reported by Fryar et al. [2012], with the body part thresholds defined using the 7.5 heads rule from the field of artistic anatomy [Richer, 1986]. We use a body scale to contour flood depths with the intention of producing flood hazard data that is more relatable to end-users. The clear advantage of providing maps depicting the depth of flooding is that potentially unsafe areas during extreme events within the floodplain are easily identified. Figure 3.3 also demonstrates our (pre-focus group) approach for communicating the mapped hazard. Specifically, we elected to describe flood hazards by their annual probabilities and contour intensity of the hazard with a qualitative scale. Data shown on the map other than the flood hazard was limited to allow end-users to focus requests and revisions on the hazard data itself. All of the baseline hazard maps were presented and described in a similar manner.

Figure 3.4 shows the analyzed flood hazard maps which incorporate flow velocity information, where the flow velocity, v, multiplied by flood depth, h, is contoured using a scale reflecting the strength or "force" of the flood waters. We use vh as a proxy for the force of the flood waters because thresholds for toppling people and moving cars have been reported in terms

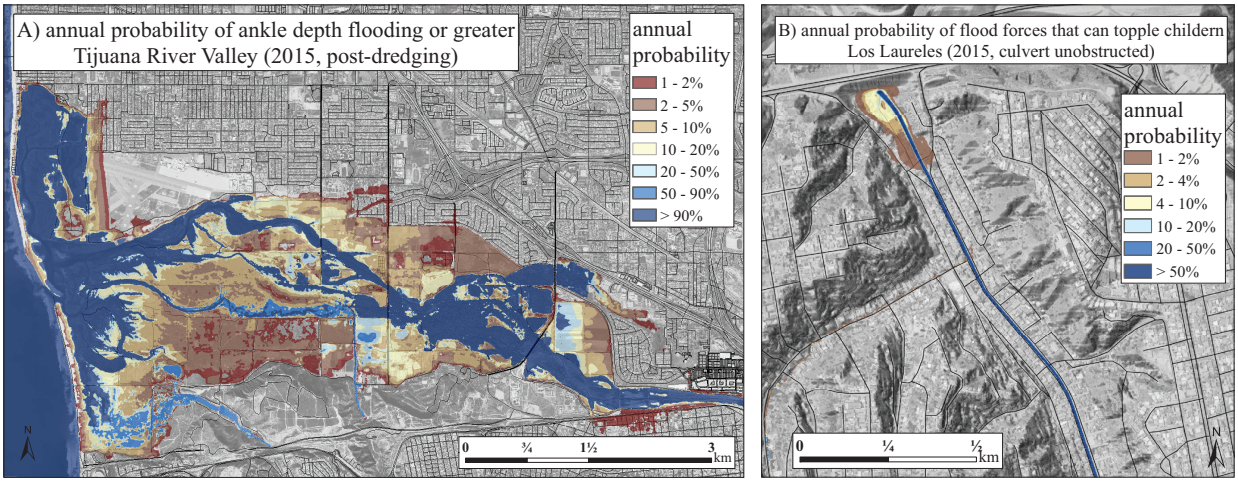

Figure 3.5: A) Annual exceedance probability (AEP) of ankle depth flooding in the Tijuana River Valley. Contours represent the annual probability that either storm tides, canyon flows, or Tijuana River discharges can cause flooding that exceeds ankle depth. B) AEP of flood forces that can topple children in Los Laureles. Here, contours represent the annual probability that streamflow will be intense enough to topple children. The ankle depth threshold was defined as 0.11 m, and the children toppled threshold was defined as 0.4 m^2/s.

of vh, or discharge per unit width [Xia et al., 2014, 2011]. A vh criterion was also found suitable for predicting structural damage to homes [Kreibich et al., 2009, Gallegos et al., 2012], so vh can be used to describe a range of hazardous conditions. Although the "force" map is more precisely described as a flood discharge per unit width map, we use the word force for simpler communication. By relating the intensity of vh to the vulnerability of people, cars, and homes, the map could be used for strategic land use planning or emergency response. We note that Australian flood studies produce information describing the severity of v and h, and recommend that hazard zones are defined by vh thresholds [AEMI, 2013].

The baseline hazard maps were associated with the 1% AEP event for consistency with the usual presentation of flood hazard information, although each hazard map presented thus far could be produced for an event more or less likely than 1% AEP. Hazard maps can also display multiple AEP events on a single map. This is accomplished by contouring the AEP of a particular hazard threshold, rather than contouring the intensity of a flood hazard with a specified AEP. Figure 3.5 presents the analyzed maps which contour AEP

rather than intensity of depth or force. The AEP of ankle depth flooding in the TRV is shown in Figure 3.5A, while Figure 3.5B displays the AEP of flood forces strong enough to topple children in LL. AEP maps of either h or vh thresholds are attainable by simulating multiple floods and combining the results on a single map. The maps shown in Figure 3.5 can be produced using a wide variety of hazard thresholds not limited to ankle depth flooding or children toppled. Here, mapping exceedance probabilities of a specific hazard threshold quantifies and communicates the likelihood of a particularly negative consequence of flooding. However, the AEP map is difficult to produce (Appendix C.3) and presents highly technical information. Thus, it may present challenges in communicating with end-users.

The maps presented in Figure 3.2-3.5 can be produced for a wide variety of locations and are not limited to site specific hydrologic conditions. We do not present the site specific baseline hazard maps produced for the end-users in this paper; however, the site specific hazard maps illustrating dominant drivers and duration of inundation can be viewed online following the links in the *data availability* section. The hazard maps presented in Figure 3.2-3.5 exemplify the range of flood hazard data presented to the end-users and our original approach for communicating the mapped hazard. We note that Figure 3.2-3.5 are not identical to the maps analyzed by the end-users since they were adapted for presentation format herein. Specifically, legend placement and fontsize is slightly different in Figure 3.2-3.5 from the maps analyzed by the end-users, while Figure 3.2-3.5 also do not include street and landmark labels that were included in the baseline maps. However, the legend content and text, map titles, hazard data, base maps, and color schemes are identical between maps analyzed by end-users and Figure 3.2-3.5. The cartographic design of the baseline hazard maps generally followed recommendations for designing flood risk maps provided by Fuchs et al. [2009].

3.5 Stakeholder Focus Groups

Four focus groups were held at each study site, which were distinguished by job-function related to FRM. The TRV focus groups included 22 total participants, with six representing public works professionals and city planners, eight representing emergency managers, five representing natural resource managers, and three categorized as non-governmental organizations or community. The LL focus groups included 33 participants total, with nine individuals representing an academic affiliation, six representing government, eleven representing emergency managers, and seven representing non-governmental organizations or other community members. Participants of each focus group were recruited through personal communication and referral sampling. It was not possible to strengthen reliability of the focus groups by replicating cohorts because a limited number of individuals are involved with FRM at the study sites. Thus, by limiting our focus groups to individuals who play a role in local FRM, we strengthened the study cohort by including the most relevant participants based on job function. Representation was not identical at the two sites because 1) governmental and organizational structures vary between the two countries and 2) participants were chosen based upon their importance to decisions related to FRM in the respective sites.

The TRV and LL focus groups were each two and a half hours in length, during which a facilitator elicited information regarding the focus group's perception of the baseline flood hazard maps. A large format hardcopy of each map was distributed to all participants along with a glossary that described baseline map terms. After individually examining the map, the facilitator would ask the engineer who produced the map to explain the hazard map and establish a common understanding. Participants were able to ask the engineer questions, while the facilitator would ask questions to test understanding. Data was generated throughout the focus groups by the facilitator adhering to a script with questions specific to each baseline hazard map. For example, the facilitator would ask participants about their

opinions of the indicators used in the map legends, the utility of the mapped hazard data, and information they would like to see depicted but was not included. The complete facilitation plan is available in the supplemental materials. This process repeated for each of the baseline hazard maps.

Following the presentation and discussion of the baseline hazard maps, survey data was collected. The surveys were designed to elicit information that would be useful for re-configuring the hydraulic models and producing flood hazard data relevant to the focus group participants. Results from the final "exit" survey addressed several important aspects of producing relevant flood hazard data: end-user job function, planning time frames of interest, flood return periods of interest, flood drivers of interest, relevant FRM strategies, and environmental conditions of interest. Each question in the survey was followed by a range of options, and the focus group participants were prompted to select a single option in response to each question. The full exit survey is also provided in the supplementary information.

Data was collected throughout the focus groups by recording conversations and reinforced by note-taking. Furthermore, transcripts were prepared of all conversations based on audio recordings. Transcripts were translated into English then analyzed using open coding to identify general themes and concepts followed by axial coding (categorization) and identification of patterns and relationships among the concepts [Saldaña, 2015, Feldman, 1995]. Transcripts of the focus groups were analyzed independently using codes including "Requested Map Revisions" and "Requested Map Scenarios" specifically to determine end-user expressed (1) requested improvements to baseline maps and (2) desirable flood hazard maps. "Requested Map Revisions" included requests and inquiries related to the map legend, units, and contextual information. Information that was useful for re-configuring the hydraulic models (new scenarios) and producing new hazard maps were categorized as "Requested Map Scenarios." Flood mapping scenarios that were specifically mentioned or requested by

participants were verified by independent analysis. Data obtained from the exit surveys were also categorized as "Requested Map Scenarios."

3.6 Results

3.6.1 Requested Map Revisions

Table 3.1 summarizes the requested map revisions of the end-users from both sites. The requested revisions generally fell into three categories: requests for the hydrologic/meteorologic conditions of the mapped hazard, further clarification of the map legends, or requests for additional geospatial data to be shown on the map. The requests presented in Table 3.1 were specifically mentioned by participants and confirmed by discussions recorded in the transcripts.

There were several important requested map revisions that were common to the end-users of both sites. First, end-users were particularly interested in the amount of rainfall or streamflow that caused the flood hazard on the map. The amount and duration of precipitation leading to the mapped hazard was often requested for both sites, while TRV end-users re-

Table 3.1: Summary of the end-users' requested map revisions. These specific requests were identified through transcript content analysis.

	Tijuana River Valley Stakeholders	**Los Laureles Stakeholders**
requested hydrologic/meteorologic information:	- amount and duration of precipitation - volumetric flow rate	- amount and duration of precipitation - flooding mechanism (cause) - conditions related to previous flood
legend requests/comments:	- quantify legend thresholds - scientific units - erosion scale - children toppled threshold not appropriate	- quantitative information - scientific units - relate flooding to erosion - relate velocity to infrastructure damage - non-technical language
additional geospatial data requests:	- access roads - river channel - levees and dikes - landmarks - park trails - sediments and vegetation	- access roads - channels - population/demographics - sediment basins - topography - aerial photo with year - locations of dwellings and shelters

quested the flow rate of the Tijuana River associated with the hazard maps. This illustrates the desire to relate the mapped hazard to information that is available in real time or other publicly available information. LL end-users also asked how the mapped scenario related to previous flooding events and noted that the precise frequency of the storm is not necessary for many users. Notice that the baseline hazard maps were described by the probability of the mapped hazard only (Figure 3.3 - 3.5), and not by the conditions leading to the mapped flood hazard such as the amount of rainfall.

Second, participants of both sites were interested in the quantitative or scientific units of the hazard legends. In general, the participants found the qualitative legends helpful and informative, but clarification was needed to explain the basis for the qualitative thresholds. For example, the "children toppled" hazard criteria (Figure 3.4) was received with mixed reactions. Several participants were confused by the distinction between hazardous conditions for children versus adults, which required additional explanation, and others viewed this criteria as alarmist. A participant representing public works professionals said: "I think [children toppled] is an alarming metric. I don't know if it would be useful for my city work... maybe a different way of identifying intensity would be more useful." Another participant representing natural resource managers noted: "I would be interested in knowing the velocity or what the force impacts are. For what I do, I can't write in an EIS [Environmental Impact Statement] that a child would be toppled in this area, so that wouldn't work for me." End-users representing natural resource managers were generally more interested in relating the velocity information to the erosive potential of the flowing water, and several participants recommended an erosion scale based on velocity and land cover data. On the other hand, end-users representing emergency managers did find the metrics in the force map (Figure 3.4) useful.

Emergency managers noted that the flood intensity descriptors of the force map would help them decide where to allocate resources during a flood based on the depicted severity, while

the "cars sliding" metric would help determine which roads would be inaccessible during an extreme event. Emergency managers also said the force map would help them prepare for hazardous debris flow during a flood: "Let me tell you why this [force map] is really important... what I'm really concerned about down in the Valley is...anything that's a hazard in the river that's actually moving. If I see a barn, and I know that its structural home damage in this area, I'll be a little more on alert because... hey those barns and stuff, they're gone. And so now I know there's more things in the river that not only can create a hazard to me but that can also adjust the flow of the river as well." Indeed, participating emergency responders were primarily interested in the relative severity of the flood hazard within the floodplain.

Lastly, end-users requested a wide variety of additional geospatial data to provide additional context, i.e, information beyond the illustrated flood hazard. The common geospatial data requests between the two sites included depictions of access roads, the river channels, and important flood control infrastructure. Of course, the relevant geospatial data will depend on the end-user and site specific characteristics, but many end-users requested the locations of access roads and river channels.

3.6.2 Requested Map Scenarios

Transcripts were examined to extract information that was useful for re-configuring the hydraulic models for new scenarios. This information includes environmental conditions of interest, relevant magnitudes and types of flooding events, and requested flood hazard data (model output). Re-configuring and re-running the models involves changes to model inputs such as a topographic data and boundary conditions. Table 3.2 shows the data we collected from the transcripts that were relevant for producing the new flood hazard maps. In addition to the focus group transcripts, data from the exit surveys were used to configure relevant

Table 3.2: Summary of information used to re-run the hydraulic models and produce new hazard maps (requested map scenarios). These specific requests were identified through transcript content analysis.

	Tijuana River Valley Stakeholders	Los Laureles Stakeholders
relevant flood magnitudes and drivers	- historic floods - more frequent than 100 year/ 1% AEP - storm water runoff - nuisance flooding/ non-extreme events - wave overtopping (Imperial Beach)	- historic floods - more frequent than 100 year/ 1% AEP - pluvial flooding
environmental conditions of interest	- blockages/obstructions - worst case flooding/infrastructure fails - different channel capacities - different channel locations - early season/late season - sea level rise	- blockages/obstructions - worst case scenario - future channelization - future land use
requested flood hazard data/model output	- output related to erosion potential - output that depicts standing water - duration of flooding event - velocity of flood waters - real-time data - suite of maps related to different rainfall depths	- output that predicts erosion potential - output based on forecasted rain - duration of flooding - maximum velocity of flood

flood modeling scenarios. Figure 3.6 shows the results of the exit survey as a percentage of the participants' responses. We note that these results are not statistically significant nor representative of all end-users of flood hazard data, but they were useful for complementing the qualitative information obtained from the transcripts. Comparison between the exit survey results (Figure 3.6) and the modeling requests documented in the transcripts (Table 3.2) reveal several important stakeholder preferences and are considered highly relevant to the specific end-users in this study.

Regarding the magnitude of the mapped flooding event, end-users were interested in events more frequent, or smaller in magnitude, than the 100 year (or 1% AEP) flood. TRV and LL end-users specifically asked to see hazards associated with more frequent events (Table 3.2), and when asked "What return period is most useful to be mapped?", more than half of end-users from both the TRV and LL focus groups selected return periods of 20 years or less. The 100 year flood was still considered relevant, however end-users generally considered the hazards of frequent floods more useful for day-to-day decision making. One participant noted: "We often focus a lot on the 1% [flood] because of FEMA and things like that, but in terms of what people are experiencing right now, they're noodling around with like 5 or 10 year events which is what's causing them problems... hitting those frequencies that people

are more likely to see, that would be a valuable communication tool."

Another important finding that was particularly evident from the LL focus groups was interest in the hazards of pluvial flooding, or flooding caused by storm water runoff. Indeed, twice as many LL end-users selected "pooled rainfall" compared to "excessive streamflow" as the most useful flood driver to be mapped. This interest was documented in the transcripts for both sites, i.e., TRV end-users noted that storm water runoff can cause flooding in neighborhoods just north of the River Valley. One of the TRV end-users described how nuisance flooding associated with storm water is noticeably absent from publicly available maps: "We have areas in the city that experience nuisance flooding, and even though it doesn't show up on the FIRM maps as an area of special flood hazard, ...you almost have to find out from property owners in there, knock on their doors, [and ask] do you ever get floods here? [owners respond] 'Oh yeah all the time.'." LL end-users were also concerned with flooding caused by storm water runoff, but they were more interested in flood hazards caused by extreme rainfall events along steep, canyon terraces that are not adjacent to the storm water channels. Mapping the hazards associated with direct rainfall is considerably different from traditional floodplain mapping, whereby the floodplain is delineated by routing discharges that exceed channel capacities and spread across the floodplain (e.g. Appendices C.2-C.2.4). Hence, it appears that participating end-users astutely discerned a limitation of this traditional flood modeling approach.

When asked "What conditions do you most want to see?", responses were relatively uniform among the choices offered. The exit survey revealed a slight preference for current conditions in the TRV and obstructions in river channels for LL end-users. The transcripts further support interest in scenarios involving blockages. At both sites, end-users were concerned about trash and debris obstructing flow beneath bridges and through the culvert in LL (Figure 3.1). Accordingly, end-users requested to see the impact of obstructions in river channels (Table 3.2), and also "worst case" flooding scenarios involving some kind of infrastructure

failure. We also note that maps showing the possible impacts of climate change on flood characteristics were not especially relevant to the participating end-users. Indeed, LL participants selected "changes in rainfall patterns" least frequently when asked "what conditions do you most want to see?", while sea level rise scenarios were slightly less preferred relative to current conditions in the TRV. Again, these results should not be viewed as representative of all end-users of flood hazard data, but it is important to note that these end-users did not view climate change impacts on flood hazards as more relevant than scenarios involving infrastructure failures or channel blockages.

By far, the most requested flood hazard data, or hydraulic model output, was for data that described the erosive potential of flowing water. Both TRV and LL participants requested model output that can describe erosion potential or susceptibility (Table 3.2), and after presenting the flood force map (Figure 3.4), end-users asked for the velocity of the flood waters to be related to erosion (Table 3.1). It is also worth noting that before being shown the flood force map, end-users at both sites requested information describing the velocity of the floodwaters. Aside from maps depicting velocities or erosion potential, end-users also requested inundation maps based on real time information or rain forecasts. One of the participants suggested a collection of maps related to different rainfall depths as a substitute for real time mapping. Lastly, community members of the TRV requested hazard maps depicting areas susceptible to standing water. Areas susceptible to standing water are a concern among community members due to increased likelihood of mosquito activity and pollutant exposure. The following section describes both the new hazard maps and improved presentation of the hazard data that was supported by the end-user focus groups.

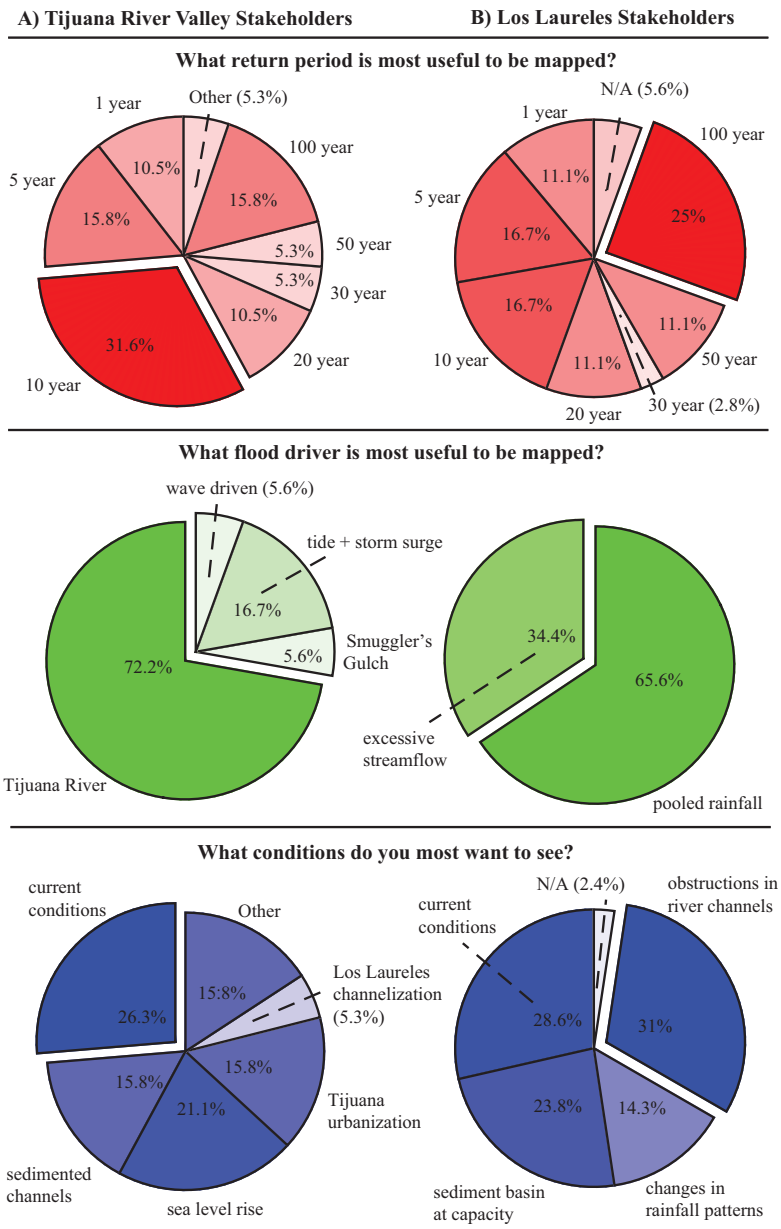

Figure 3.6: Results of the exit survey designed to provide insight on relevant flood frequencies, drivers, and environmental conditions. Column (A) includes TRV responses (22 participants), and column (B) includes LL responses (33 participants). The most distinct responses include the reported utility of maps illustrating more frequent events (TRV) and pooled rainfall (LL). The lack of a clear preference for relevant environmental conditions is worth noting. Here, "environmental conditions" are defined as the physical conditions of study area or the state of climate related variables during the simulated flood. Maps depicting possible impacts of climate change (e.g. sea level rise and changes in rainfall patterns) were not the most desired conditions for end-users of either site.

3.6.3 Co-produced Flood Hazard Maps

In this section, we present the flood hazard maps that were produced after re-configuring the hydraulic models based on the results of the focus groups. The maps presented herein, Figures 3.7 - 3.9, do not include all of the maps we produced post-focus group, but rather the hazard maps that both support and expand upon previous studies offering guidance for improving flood hazard maps. All hazard maps produced in this study in response to end-user feedback can be viewed online by following the links in the *data availability* section. Herein, we also present the mapped hazard data according to the end-users' requested revisions to exemplify effective means of communicating the mapped hazard according to these end-users.

First, we focus on end-user interest in "pooled rainfall" or storm water runoff. Figure 3.7 shows the pre and post-focus group flood depth maps for LL. The pre-focus group flood depth map (Figure 3.7A) was produced by injecting flow hydrographs from a hydrologic model into the channel network of the hydraulic model, so the mapped flooded area corresponds to so-called *fluvial flooding*, or streamflows that exceed the capacity of the channel network (Appendix C.2.4). In the post-focus group flood map, on the other hand, runoff (precipitation minus infiltration) was computed for every numerical grid of the hydraulic model and routed as overland flow. Hence, the mapped flooded area corresponds to so-called *pluvial flooding*, or the combined effects of storm-water runoff and excess streamflow. Comparison between Figures 3.7A and 3.7B reveals that the flood hazard can be significantly underestimated if only fluvial flooding is considered, showing the importance of pluvial flooding (direct storm-water runoff) in this system.

The baseline hazard maps based on fluvial flooding did not align with stakeholders' experience, whereas the post-focus group flood hazard maps that account for pluvial flooding clearly show accumulation of floodwater along roadways and locations adjacent to the main channel. Notice that the pluvial hazard map illustrates ankle to waist deep flooding adjacent

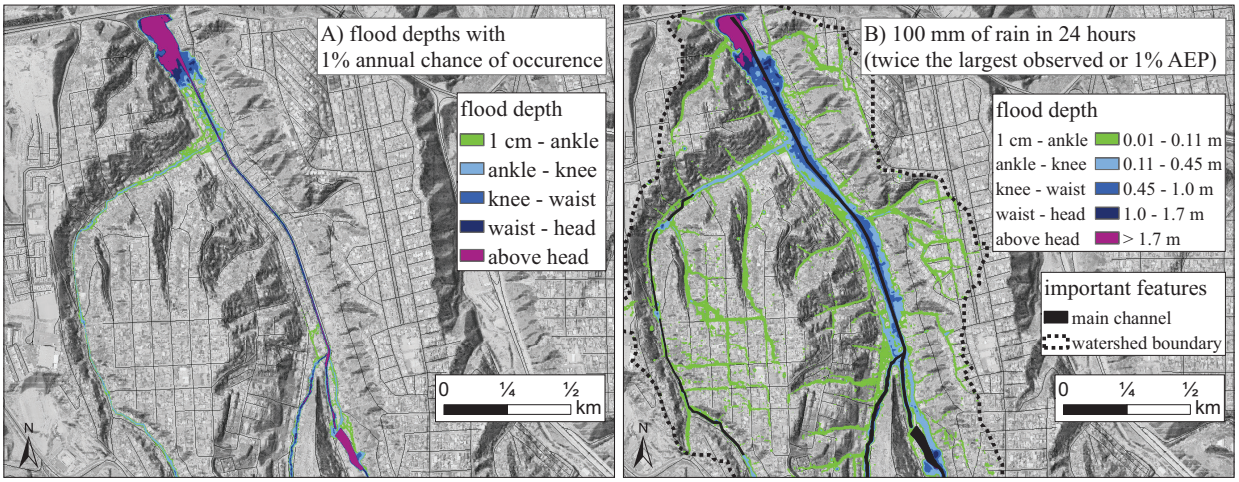

Figure 3.7: A) Pre-focus group Los Laureles flood depth map based on traditional coupling of hydrologic and hydraulic models. B) Post-focus group Los Laureles flood depth map based on routing storm-water runoff using the hydraulic model grid. Comparison between A) and B) shows that traditional hydrologic-hydraulic model coupling can misrepresent flood hazards in catchments where storm-water runoff is severe. We also present B) following the stakeholder requests. Notice that A) is described by the AEP of the flood depths, whereas B) is described by the amount of rainfall that caused the mapped hazard. The post-focus group map also includes a qualatative/quantative legend, the location of the main channel, and the watershed boundary.

to the main channel, whereas no flooding is predicted adjacent to the main channel in the baseline map (Figure 3.7). Hence, this site calls for hydrologic modeling approaches that resolve overland flow caused by rainfall. Pluvial flood hazard studies are still fairly uncommon [Apel et al., 2016], yet there are several examples in the literature and methodological advancements are growing [Nuswantoro et al., 2016, Blanc et al., 2012, Guerreiro et al., 2017, Simões et al., 2015]. Maps produced by such methods have significant potential to support FRM. For example, from an emergency management perspective, pluvial flood maps can highlight streets and low-lying areas in the floodplain that are likely to become hazardous during a severe rain storm, and from a planning and design perspective, pluvial maps can identify areas in the floodplain with relatively poor drainage.

Notice also that the pluvial hazard map presents the data differently from the baseline map. The pre-focus group map describes the mapped hazard by the exceedance probability of

the flood depths, whereas the post-focus group map describes the mapped hazard as the flood depths resulting from the corresponding rainfall depth and duration. In the latter case, the exceedance probability is included but not emphasized. This may seem like a minor difference, but end-users suggested that a description based on rainfall depth is much more useful. Rainfall depths are commonly forecast and reported publicly, so the rainfall totals are generally more relevant to end-users than frequency of occurrence or probability of exceedance. End-users' requests for the amount of rainfall and streamflow associated with the hazard map point to the need to link mapped flooding scenarios to familiar reference points.

Linking flooding scenarios to familiar reference points is difficult, however. Most flooding scenarios are defined by probabilities which are intangible to many end-users, while the significance of rainfall totals or volumetric flow rates can be unknown to users not familiar with typical site conditions. A reference point requested by end-users which has the potential to be accessible to a wide audience was the magnitude of the mapped flooding event relative to historic measurements. Notice that the pluvial flood map (Figure 3.7B) also describes the rainfall event as "twice the largest observed", which means the 1% AEP 24hr rainfall depth is roughly twice the largest recorded at the nearest rainfall gage. Describing the flooding scenario as "twice the largest observed" adds important context concerning the magnitude of the mapped event by giving the impression of a rare scenario yet within the realm of physical plausibility.

In addition to the flood hazards caused by storm-water runoff, LL and TRV stakeholders were also concerned about the erosive potential of flowing water (Table 3.1 and 3.2). Erosion potential was particularly relevant for the TRV and LL, since highly erosive soils lead to continual dredging of flood control channels and maintenance of engineered sediment basins. Accordingly, we produced maps that illustrate the erosive potential of various floods. Figure 3.8 shows an example of an erosion potential map we produced for the TRV stakeholders.

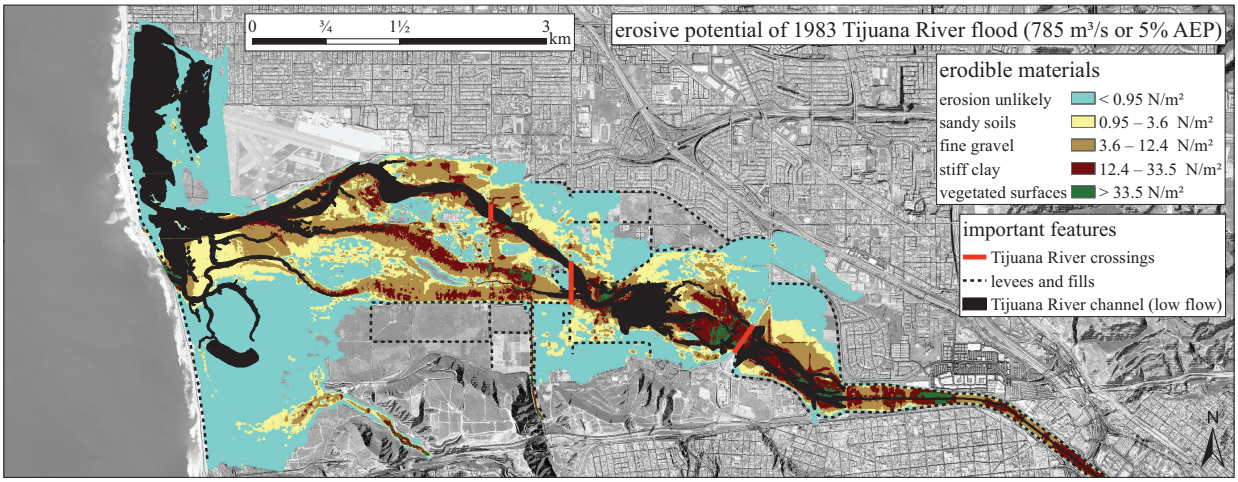

Figure 3.8: Erosion potential map of the 1983 Tijuana River flood produced for the TRV stakeholders. The contoured flood hazard variable is the maximum depth-averaged shear stress predicted by the hydraulic model during the simulated flood event. Legend thresholds were estimated from Fischenich [2001]. The map also includes the locations of levees and fills, Tijuana River crossings, and the Tijuana river channel.

The erosion potential map contours the maximum depth-averaged shear stress predicted by the hydraulic model during the course of the simulated flood. We contour shear stress to illustrate erosion potential because shear stress is commonly used to model soil erosion and design stream stabilization or restoration projects [Knapen et al., 2007, Fischenich, 2001]. The qualitative scale in Figure 3.8 that describes materials susceptible to erosion during the simulated flood is based on the permissible shear stresses of different stream restoration materials reported by Fischenich [2001]. Notice that the units of the erosion thresholds are also included in the hazard map legend. Although it is unlikely that even technical end-users can interpret the magnitude of different shear stress values off hand (i.e. Figure 3.8), the units provide the basis for the qualitative scale and improve credibility. The shear stress values by themselves are fairly abstract, so the quantitative and qualitative scales compliment each other quite nicely.

Natural resource managers, less often targeted as end users of flood hazard data, were an audience that sought maps of erosion potential in this study. Most flood hazard maps are tailored for public communication, strategic planning, emergency response, or insurance

purposes [Meyer et al., 2012]. Yet for many locations, managing land use and controlling erosion is essential to FRM, while erosion caused by flood waters can be the most damaging aspect of a major flood [Luke et al., 2015]. Maps contouring shear stress and illustrating erosion potential can be very useful for developing local land use strategies and suppressing erosion. For example, natural resource managers could restrict sensitive land uses in areas of the floodplain that are likely to experience relatively erosive flows. By overlaying the predicted shear stresses with land use/land cover data, the erosion potential map could also be used to highlight areas in need of armoring or erosion resistant vegetation.

Notice also that the example erosion potential map displays the shear stresses associated with a hindcast of the 1983 Tijuana River flood, i.e., a historic flood. Stakeholders of both sites were interested in hazard maps of historic flooding (Table 3.2), which we also find particularly valuable due to their relevance to end-users and ease of communication. Maps of historic floods are relevant precisely because they actually occurred; the map does not need to communicate an abstract exceedance probability or frequency. In the case of historic flood maps, the year of the flood or possibly the name of the storm can be used to describe the flooding scenario. The 1983 Tijuana River flood was associated with a particularly strong El Nino year, which several stakeholders recall. The 1983 flood also represents a relatively frequent flood (about 20 year return period) which was deemed useful to be shown on a map by end-users of both sites (Figure 3.6, Table 3.2).

Lastly, we present one additional flood hazard map produced in response to community members' concerns about stagnant water in the TRV. Standing water creates breeding sites for mosquitos and a corresponding proliferation of mosquito-borne diseases such as West Nile Virus, so it is an important public health consideration after a major flood [Kouadio et al., 2012]. Figure 3.9 shows the depth of flood waters remaining in the TRV immediately following recession of a flood with a peak flow rate of 2300 m^3/s, assuming no evaporation or infiltration. Therefore, the map should be interpreted as the standing water depth

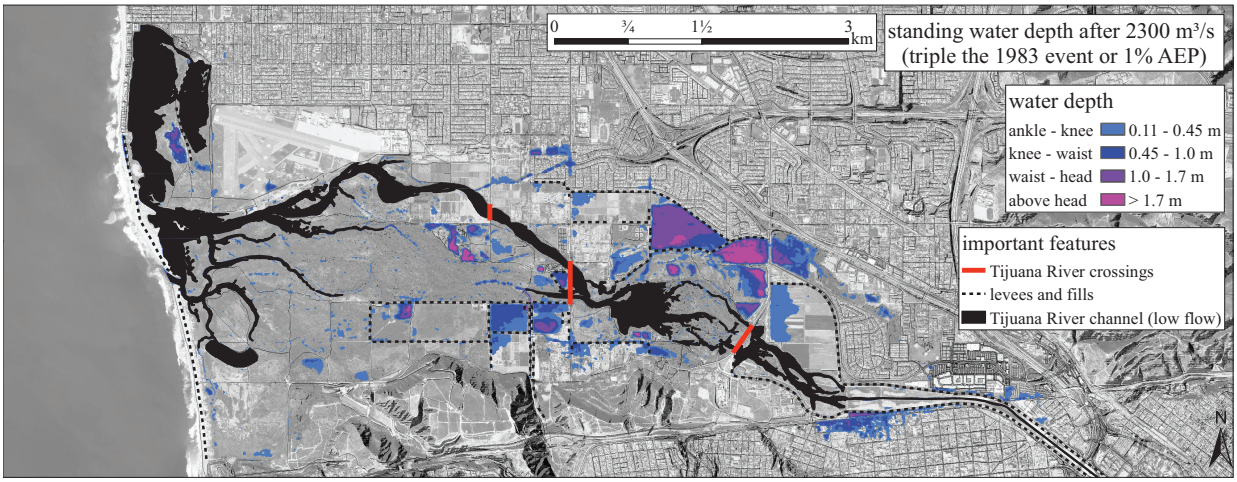

Figure 3.9: Flood hazard map illustrating pooled water that did not freely drain following a peak Tijuana River flow of 2300 m^3/s, which has an estimated 1% AEP or 100 year frequency. The flood depths shown on the map are the hydraulic model solution after the hydrograph receded to baseflow conditions. The areas that do not freely drain are likely to experience increased mosquito activity and exposure to pollutants following a major flooding event.

immediately following the flood.

The standing water map is unique relative to other hazard maps because it illustrates conditions immediately after rather than during the event, and therefore represents a mechanism of supporting of *recovery* efforts - a key component of the disaster management cycle [Khan et al., 2008]. The standing water map could help authorities with mosquito abatement, deployment of pumps, and strategic positioning of emergency shelters. Additionally, a standing water map can support floodplain planning by enabling enhanced assessments of public health risks for proposed floodplain development projects.

3.7 Discussion and Recommendations

We acknowledge that the requested map scenarios and revisions of the TRV and LL end-users do not represent all end-users of flood hazard data. However, for the sake of improving the practice of flood hazard mapping, let us compare the end-users' requests herein to previous

studies and flood mapping guidelines in Europe, Australia, and the US. The majority of end-user preferences from this study align with recommendations for improving flood hazard maps from European studies by Meyer et al. [2012] and Hagemeier-Klose and Wagner [2009]. Concerning the presentation or communication of the mapped content, Hagemeier-Klose and Wagner [2009] recommended that flood hazard maps should be linked to real time information such as river stage, while Bell and Tobin [2007] questioned the effectiveness of communicating floods based solely on probabilities. Both studies point to the need to link the mapped flooding scenario to concrete reference points other than frequency or exceedance probability. The need for tangible reference points was supported in this study by end-users' requests for the rainfall totals, streamflow, and relative magnitude of the mapped scenario. In flood mapping practice, however, communication of the mapped hazard is still very much tied to probabilities.

The European handbook [Martini and Loat, 2007] states the title of the map should include the hazard parameter and probability, while the FEMA FIRM describes hazard zones by AEPs only (i.e. Figure 3.2). Yet there is increasing evidence that these descriptions are ineffective. Relatively little guidance is provided regarding the map legends as well, which is another important aspect of communicating the mapped hazard. Legends are either completely described by numerical values (i.e. depth of flooding in meters or feet) or a qualitative flood severity zone described by terms such as "low" to "severe" (FEMA, 2014). We recommend that future mapping guidance documents provide advice for different ways to communicate the mapped hazard scenario and more complete legend descriptors. Alternatives supported by this study include 1) providing qualitative and quantitative scales, and 2) describing flooding scenarios by the flood magnitude (in corresponding scientific units), the magnitude relative to an historic event, and finally the probability of the flood. The magnitude and probability of the flood provides relevant information for technical end-users, while the magnitude related to an historic event is a tangible reference point for lay-persons. It is our opinion that least emphasis should be given to the probability when describing mapped

flooding scenarios. Not only are concrete references preferred for describing flood risk [Bell and Tobin, 2007], but flood probabilities and corresponding frequencies are highly uncertain [e.g. Appendices C.1.1 - C.1.3, Di Baldassarre et al., 2010, Kjeldsen et al., 2014, Merwade et al., 2008, Merz and Blöschl, 2008, Stedinger and Griffis, 2008]. Indeed, all hazard variables illustrated in flood maps are inherently uncertain, however it is remarkable that perhaps the most uncertain and complex characteristic of floods is also the primary descriptor.

Uncertainties associated with flood mapping products are rarely quantified let alone communicated, and in this study, we did not address the important issue of communicating uncertainty in flood maps to end-users. In one of the few studies that has explicitly addressed communicating uncertainty in the FEMA FIRMs' floodplain boundaries, Soden et al. [2017] showed that providing end-users with contrasting information (i.e. the 1% AEP flood extent versus an observed flooding extent) led to important flood hazard discourse and curiosity regarding flood mapping methodology. While it may seem counterintuitive to purposefully expose the limitations of floodplain delineation, such innovative communication strategies force end-users to confront the deterministic standards that our institutions require for regulatory purposes. Explicit confrontation with the limits of science promotes contemplation and is certainly worth further investigation in the context of flood hazard mapping and communication.

Regarding desirable content of flood hazard maps, stakeholder preferences from this study also align with previous work. Meyer et al. [2012] concluded that maps presenting flood hazards at different probabilities is required, velocity information should be provided when available, and the location of flood defenses and access routes should be integrated within the hazard map. All of these recommended contents were requested by the TRV and LL end-users. Hagemeier-Klose and Wagner [2009] also noted the importance of mapping more frequent events than the 1% AEP flood, which was strongly supported by the results of the survey (Figure 3.6). Thus, this study demonstrates consistency between the desired

map content of end-users studied in the US, Mexico, and Europe. Flood mapping guidelines in Europe [Martini and Loat, 2007] and Australia [AEMI, 2013] generally recommend producing this desired content. Guidelines either recommend or require the production of hazard maps associated with different probabilities and even infrastructure failure scenarios. Mapped data includes flooding extent, depths, velocities, and the depth-velocity product, while some studies even provide shear stresses [Martini and Loat, 2007]. Specific guidance is also provided for producing maps that support the FRM activities of distinct European and Australian end-user groups.

Meanwhile in the US, flood mapping guidelines are fairly extensive and standardized for producing the FEMA FIRM only [FEMA, 2016]. There is a lack of guidance and direction available for producing flood hazard maps that support non-insurance aspects of FRM, such as those specifically requested by end-users in this study. The required "non-regulatory" data products of recent FEMA Risk MAP studies [FEMA, 2014] have significant potential to support an expanded portfolio of actionable flood hazard maps in the US. For example, the required velocity grids in Risk MAP studies can be post-processed to produce erosion potential maps, while the required "flood severity grid" contains the depth-velocity data necessary for products designed to support emergency response. We recommend that future FEMA guidelines provide specific directives for producing non-regulatory flood hazard maps tailored to specific FRM objectives including land-use planning, emergency management, and public awareness. The content of flood hazard maps should also be expanded to include pluvial flood hazards when appropriate.

Neither European, Australian, nor US flood mapping guidelines explicitly require or recommend maps characterizing pluvial flood hazard, which were of keen interest to both LL and TRV end-users. However, many EU member states have included pluvial flood hazard assessments in response to the Floods Directive [Nixon et al., 2015]. The Australian technical guidelines for engineers allude to direct rainfall models that can be used to produce pluvial

hazard maps [McCowan, 2016], but because these techniques are relatively new, guidance documents do not require the production of maps depicting the pluvial hazard or intense storm-water runoff. Since techniques for estimating pluvial hazards continue to advance, formal guidelines and requirements for mapping the pluvial hazard zone should be developed, especially in the US. As demonstrated by end-user requests in this study - and the largely pluvial nature of the flooding disaster caused by Hurricane Harvey in the US - pluvial flooding can dominate in urban areas and needs to be considered in future mapping efforts.

3.8 Conclusions and Future Directions

Two-dimensional (2D) flood hazard models developed for the Tijuana River Valley (TRV) and Los Laureles (LL) on both sides of the US-Mexico border supported the co-development of flood hazard maps responsive to end-user management needs. 2D modeling by engineers produced a set of baseline maps that were further refined through end-user focus groups that triggered additional modeling scenarios and map revisions.

This study revealed general consistency between the mapping needs of studied end-users in the US and Mexico with those reported in European studies and guidelines published in Australia. For example, mapping requests included scenarios with different probabilities and even infrastructure failure scenarios, and end-users also requested maps of hazard variables beyond traditional flood extent, such as velocities and standing water. This study also revealed several important flood hazard mapping requests relevant to other sites:

- Flood intensity scales (e.g., depth, force or shear stress) that frame the mapped information both quantitatively and qualitatively. The quantitative scale meets end-user needs for a technical reference point, while the qualitative scale meets end-user needs to easily interpret the mapped information.

- Flood scenario descriptions that report both the magnitude of the flood in terms of rainfall or streamflow amounts and also the flood magnitude relative to an historic event. Use of concrete scenario descriptions increases the utility and relevance of mapped information across different end-users of flood hazard maps.

- Flood hazard maps that depict the erosion potential of flood waters. Erosion potential maps support end-user needs for managing sediment.

- Flood hazard maps that depict standing water following the flood. Standing water maps support recovery planning and public health concerns.

- Flood hazard maps that depict storm-water runoff or pluvial flood hazards. Baseline flood hazard maps depicted *fluvial* flooding hazards only, and after end-user focus groups revealed a deficiency in usefulness, the need for a pluvial flood hazard modeling approach was recognized and implemented. Characterizing pluvial flood hazards is extremely important for urbanized sites with poor drainage.

Of course, the stakeholder preferences herein must be viewed cautiously, since focus group participants do not represent all end-users of flood hazard data. The primary limitation of this study is the limited number of focus group participants (55 total) and narrow geographic scope. Co-production efforts via focus groups acknowledge that community-level knowledge (and mapping preference) varies from locality-to-locality, underscoring how flood hazard knowledge should not be a "one-directional" process but an iterative learning approach that breaks down information gaps between experts and lay users in specific places - thus improving risk communication at the local level. They also produce actionable mapping information useful for reducing flood risks [Spiekermann et al., 2015, Moel et al., 2009]. Indeed, restricted sample size and geographic scope is a common caveat of flood communication and mapping preference studies (e.g Hagemeier-Klose and Wagner [2009], Meyer et al. [2012]).

In future studies, sample size limitations may be overcome by taking advantage of online information systems to present flood hazard data. Online formats offer the opportunity for causal experiments - do different hazard variables make a user more (or less) likely to seek vulnerability reduction measures? How do different presentations of uncertainty in mapped data influence end-users' desire to seek further information? These questions could be answered with so called "A/B" testing, where subjects are presented different web pages and their interactions on the web site are recorded. Our current knowledge of flood mapping preferences and hazard perceptions is based upon empirical studies with relatively small samples [Kellens et al., 2013]. What can "big data" tell us about how end-users respond to, and interact with, flood hazard data?

While online information systems offer avenues for new research, they also provide a medium for presenting an expanded portfolio of hazard maps. Relative to the EU and Australia, flood mapping practice in the US has the greatest opportunity for expansion. Funding for flood mapping in the US remains limited [Traver, 2014], however it is relatively inexpensive to produce additional mapping products from models that are already used to produce Flood Insurance Rate Maps. Furthermore, the availability of free 2D hydraulic modeling software (HEC-RAS 5.0) and increasing abundance of metric resolution topographic data provides practitioners with the means to produce flood hazard data that was previously cost-prohibitive [Sanders, 2017]. While flood mapping methods and data continue to improve - additional criteria must also be addressed to provide decision-makers and citizens with actionable information. To be actionable, map information must help decision-makers: 1) understand vulnerability of properties from flooding; and, 2) select actions that mitigate or reduce this vulnerability [Demeritt and Nobert, 2014, McNutt, 2016, Feldman et al., 2008]. By fully utilizing flood modeling technologies and developing innovative communication strategies, flood hazard maps can more effectively support first responders, natural resource managers, and local residents with the information necessary to manage and respond to flood hazards.

3.9 Data Availability Statement

The University of California's guidelines for maintaining the privacy and confidentiality of human subjects state that data obtained from human subjects should only be accessible on a "need to know" and "minimum necessary" standard. Thus, the transcripts of focus groups conducted in this study are not publicly available. If an interested researcher wishes to review transcripts, please contact the corresponding author with 1) data requests and 2) reasoning for requesting the data.

All of the hazard maps (and several not presented herein) are available to view on an interactive system found here:

FloodRISE (2017). Tijuana River Valley Flood Hazards. University of California, Irvine. `https://bit.ly/floodrise_TRV`

FloodRISE (2017). Los Laureles Flood Hazards. University of California, Irvine. `https://bit.ly/floodrise_GC`

Data that was used to create the hazard maps includes elevation, streamflow, ocean water level, and precipitation data. The elevation data is held by the County of San Diego and could be made available via requests to the corresponding author. The streamflow, water level, and precipitation data is available here:

International Boundary and Water Commission (1960 - 2006). Flow of the Colorado River and other Western Boundary Streams and Related Data. Department of State, USA. `https://ibwc.gov/Water_Data/water_bulletins.html`

National Oceanic and Atmospheric Administration (1924 - 2008). Observed Water Levels at 9410230, La Jolla CA. Department of Commerce, USA. `https://www.tidesandcurrents.noaa.gov/waterlevels.html?id=9410230`

Department of Public Works, Flood Control Section (2003). San Diego County Hydrology Manual. County of San Diego. http://www.sandiegocounty.gov/content/dam/sdc/dpw/FLOOD_CONTROL/floodcontroldocuments/hydro-hydrologymanual.pdf

Bibliography

Amir AghaKouchak, David Easterling, Kuolin Hsu, Siegfried Schubert, and Soroosh Sorooshian, editors. *Extremes in a Changing Climate*. Springer Science and Business Media, Dordrecht, 1st edition, 2013.

Shardul Agrawala, Kenneth Broad, and David H Guston. Integrating climate forecasts and societal decision making: challenges to an emergent boundary organization. *Science, Technology, & Human Values*, 26(4):454–477, 2001.

Heiko Apel, Annegret H Thieken, Bruno Merz, and Günter Blöschl. A probabilistic modelling system for assessing flood risks. *Natural hazards*, 38(1-2):79–100, 2006.

Heiko Apel, Oriol Martínez Trepat, Nguyen Nghia Hung, Bruno Merz, Nguyen Viet Dung, et al. Combined fluvial and pluvial urban flood hazard analysis: concept development and application to can tho city, mekong delta, vietnam. *Natural Hazards & Earth System Sciences*, 16(4), 2016.

SA Archfield, RM Hirsch, A Viglione, and G Blöschl. Fragmented patterns of flood change across the united states. *Geophysical Research Letters*, 2016.

A Arns, T Wahl, ID Haigh, J Jensen, and C Pattiaratchi. Estimating extreme water level probabilities: A comparison of the direct methods and recommendations for best practise. *Coastal Engineering*, 81:51–66, 2013.

Australian Emergency Management Institute (AEMI). Managing the floodplain: a guide to best practice in flood risk management in australia. Technical report, Attorney-General's Department, 2013.

Kenneth J Bagstad, Kevin Stapleton, and John R D'Agostino. Taxes, subsidies, and insurance as drivers of united states coastal development. *Ecological Economics*, 63(2):285–298, 2007.

Jonathan L Bamber and WP Aspinall. An expert judgement assessment of future sea level rise from the ice sheets. *Nature Climate Change*, 3(4):424, 2013.

Santiago Begueria and Sergio M Vicente-Serrano. Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. *J. Clim. Appl. Meteorol.*, 45, 2006. doi: 10.1175/JAM2324.1.

Santiago Begueria, Marta Angulo-Martinez, Sergio M. Vicente-Serrano, J. Ignacio Lopez-Moreno, and Ahmed El-Kenawy. Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold anlysis: a case study in northeast Spain from 1930 to 2006. *Int. J. Climatol.*, 31, 2011. doi: 10.1002/joc.2218.

Cibele N Behrens, Hedibert F Lopes, and Dani Gamerman. Bayesian analysis of extreme events with threshold estimation. *Stat. Model. Int. J.*, 4(3):227–244, 2004.

Heather M Bell and Graham A Tobin. Efficient and effective? the 100-year flood in the communication and perception of flood risk. *Environmental Hazards*, 7(4):302–311, 2007.

J Blanc, JW Hall, N Roche, RJ Dawson, Y Cesses, A Burton, and CG Kilsby. Enhanced efficiency of pluvial flood risk estimation in urban areas using spatial–temporal rainfall simulations. *Journal of Flood Risk Management*, 5(2):143–152, 2012.

Laurens M Bouwer. Have disaster losses increased due to anthropogenic climate change? *Bulletin of the American Meteorological Society*, 92(1):39–46, 2011.

Jared T Brown. Introduction to femas national flood insurance program (nfip). *Washington, DC: Congressional Research Service, August*, 16, 2016.

Raymond J Burby. Flood insurance and floodplain management: the us experience. *Global Environmental Change Part B: Environmental Hazards*, 3(3):111–122, 2001.

Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

California Coastal Commission. California coastal commission sea level rise policy guidance: Interpretive guidelines for addressing sea level rise in local coastal programs and coeatal development permits. Technical report, State of California Natural Resources Agency, 2015.

Iñigo Capellán-Pérez, Inaki Arto, Josué M Polanco-Martínez, Mikel González-Eguino, and Marc B Neumann. Likelihood of climate change pathways under uncertainty on fossil fuel resource availability. *Energy & Environmental Science*, 9(8):2482–2496, 2016.

Centre for Ecology and Hydrology. Flood estimation handbook. Institute of Hydrology, Wallingford, Oxfordshire, 2008.

Linyin Cheng and Amir AghaKouchak. Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific reports*, 4, 2014.

Linyin Cheng, Amir AghaKouchak, and Eric Gilleland. Non-stationary extreme value analysis in a changin climate. *Climatic Change*, 2014a. doi: 10.1007/s10584-014-1254-5.

Linyin Cheng, Amir AghaKouchak, Eric Gilleland, and Richard W Katz. Non-stationary extreme value analysis in a changing climate. *Climatic change*, 127(2):353–369, 2014b.

Linyin Cheng, Thomas J Phillips, and Amir AghaKouchak. Non-stationary return levels of cmip5 multi-model temperature extremes. *Climate Dynamics*, 44(11-12):2947–2963, 2015.

John A Church, Peter U Clark, Anny Cazenave, Jonathan M Gregory, Svetlana Jevrejeva, Anders Levermann, Mark A Merrifield, Glenn A Milne, R Steven Nerem, Patrick D Nunn, et al. Sea level change. In *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. 2013.

Leon Clarke, James Edmonds, Henry Jacoby, Hugh Pitcher, John Reilly, and Richard Richels. Scenarios of greenhouse gas emissions and atmospheric concentrations. *US Department of Energy Publications*, page 6, 2007.

W. J. Cody and K. E. Hillstrom. Chebyshev approximations for the natural logarithm of the *gamma* function. *Math. Comp.*, 21, 1967. doi: 10.1090/S0025-5718-67-99635-4.

Timothy A Cohn and Harry F Lins. Nature's style: Naturally trendy. *Geophysical Research Letters*, 32(23), 2005.

Stuart Coles. *An introduction to the statistical modeling of extreme values*. Springer-Verlag London, London, UK, 1 edition, 2001.

Daniel Cooley. Extreme value analysis and the study of climate change. *Climatic Change*, 2009. doi: 10.1007/s10584-009-9627-x.

Daniel Cooley. Return periods and return levels under climate change. In *Extremes in a changing climate*, pages 97–114. Springer, 2013.

Dim Coumou and Stefan Rahmstorf. A decade of weather extremes. *Nature climate change*, 2(7):491–496, 2012.

Council of European Union. Council directive 2007/60/ec on the assessment and management of flood risks. *OJ L 288*, 2007.

Juraj M Cunderlik and Donald H Burn. Non-stationary pooled flood frequency analysis. *Journal of Hydrology*, 276(1):210–223, 2003.

JM Delgado, B Merz, and H Apel. Projecting flood hazard under climate change: an alternative approach to model chains. *Natural Hazards and Earth System Sciences*, 14(6): 1579–1589, 2014.

David Demeritt and Sebastien Nobert. Models of best practice in flood risk communication and management. *Environmental Hazards*, 13(4):313–328, 2014.

Giuliano Di Baldassarre, Guy Schumann, Paul D Bates, Jim E Freer, and Keith J Beven. Flood-plain mapping: a critical discussion of deterministic and probabilistic approaches. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques*, 55(3):364–376, 2010.

Lisa Dilling and Maria Carmen Lemos. Creating usable science: Opportunities and constraints for climate knowledge use and their implications for science policy. *Global environmental change*, 21(2):680–689, 2011.

S El Adlouni, TBMJ Ouarda, X Zhang, R Roy, and B Bobée. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43(3), 2007.

Engineers Australia. Guidelines for responding to the effects of climate change in coastal and ocean engineering. *The National Committee on Coastal and Ocean Engineering*, 2013.

European Commission. Directive 2007/60/ec on the assessment and management of flood risks. Technical report, Brussels, 2007.

Federal Emergency Management Agency. Guidelines and specifications for flood hazard mapping partners, appendix c: Guidance for riverine flooding analyses and mapping. Technical report, Washington D.C., 2009.

Federal Emergency Management Agency (FEMA). Guidance for flood risk analysis and mapping: Flood depth and analysis grids. Technical Report Guidance Document 4, United States Department of Homeland Security, 2014.

Federal Emergency Management Agency (FEMA). Guidance for flood risk analysis and mapping: Flood insurance rate map (firm) index. Technical Report Guidance Document 7, United States Department of Homeland Security, 2016.

D Feldman and H Ingram. Climate forecast, water management, and knowledge networks: making science useful to decision makers. *Water, Climate and Society*, 1(October 2009): 19–21, 2009.

DL Feldman, KL Jacobs, G Garfin, A Georgakakos, B Morehouse, P Restrepo, R Webb, B Yarnal, D Basketfield, HC Hartmann, et al. Making decision-support information useful, useable and responsive to decision-maker needs. *Decision-support experiments and evaluations using seasonal-to-interannual forecasts and observational data: A focus on water resources*, pages 101–140, 2008.

Martha S Feldman. *Strategies for interpreting qualitative data*, volume 33. Sage, 1995.

Craig Fischenich. Stability thresholds for stream restoration materials. Technical report, Engineering Research and Development Center, Vicksburg, MS Environmental Lab, 2001.

Gregory Flato, Jochem Marotzke, Babatunde Abiodun, Pascale Braconnot, Sin Chan Chou, William J Collins, Peter Cox, Fatima Driouech, Seita Emori, Veronika Eyring, et al. Evaluation of climate models. 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. *Climate Change 2013*, 5:741–866, 2013.

Cheryl D Fryar, Qiuping Gu, and Cynthia L Ogden. Anthropometric reference data for children and adults: United states, 2007-2010. *Vital and health statistics. Series 11, Data from the national health survey*, (252):1–48, 2012.

Sven Fuchs, Karl Spachinger, Wolfgang Dorner, Juliette Rochman, and Kamal Serrhini. Evaluating cartographic design in flood risk mapping. *Environmental Hazards*, 8(1):52–70, 2009.

Melanie Gall, Kevin A Borden, Christopher T Emrich, and Susan L Cutter. The unsustainable trend of natural hazard losses in the united states. *Sustainability*, 3(11):2157–2181, 2011.

Humberto A Gallegos, Jochen E Schubert, and Brett F Sanders. Structural damage prediction in a high-velocity urban dam-break flood: field-scale assessment of predictive skill. *Journal of Engineering Mechanics*, 38(10):1249–1262, 2012.

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Stat. Sci.*, pages 457–472, 1992.

Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre-and post-processing facilities. *International journal for numerical methods in engineering*, 79(11):1309–1331, 2009.

Eric Gilleland and Richard W. Katz. New software to analyze how extremes change over time. *Eos*, 92, 2011. doi: 10.1029/2011EO020001.

Veronica W. Griffis and Jery R. Stedinger. Incorporating climate change and variability into bulletin 17b lp3 model. *World Environmental and Water Resources Congress 2007*, 2007. doi: 10.1061/40927(243)69.

Selma B Guerreiro, Vassilis Glenis, Richard J Dawson, and Chris Kilsby. Pluvial flooding in european citiesa continental approach to urban flood modelling. *Water*, 9(4):296, 2017.

Emil Julius Gumbel. The return period of flood flows. *The annals of mathematical statistics*, 12(2):163–190, 1941.

Maria Hagemeier-Klose and Klaus Wagner. Evaluation of flood hazard maps in print and web mapping services as information tools in flood risk communication. *Natural Hazards & Earth System Sciences*, 9(2), 2009.

J Hall, Berit Arheimer, M Borga, Rudolf Brázdil, Pierluigi Claps, A Kiss, TR Kjeldsen, J Kriauciuniene, ZW Kundzewicz, Michel Lang, et al. Understanding flood regime changes in europe: A state of the art assessment. *Hydrology and Earth System Sciences*, 18(7): 2735–2772, 2014.

Stephane Hallegatte, Colin Green, Robert J. Nicholls, and Jan Corfee-Morlot. Future flood losses in major coastal cities. *Nat. Clim. Change*, 3, 2013a. doi: 10.1038/nclimate1979.

Stephane Hallegatte, Colin Green, Robert J Nicholls, and Jan Corfee-Morlot. Future flood losses in major coastal cities. *Nature climate change*, 3(9):802–806, 2013b.

Dennis L Hartmann, Albert MG Klein Tank, Matilde Rusticucci, Lisa V Alexander, Stefan Brönnimann, Yassine Abdul Rahman Charabi, Frank J Dentener, Edward J Dlugokencky, David R Easterling, Alexey Kaplan, et al. Observations: atmosphere and surface. In *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2013.

Yukiko Hirabayashi, Roobavannan Mahendran, Sujan Koirala, Lisako Konoshima, Dai Yamazaki, Satoshi Watanabe, Hyungjun Kim, and Shinjiro Kanae. Global flood risk under climate change. *Nat. Clim. Change*, 3, 2013. doi: 10.1038/nclimate1911.

J. R. M. Hosking and J. R. Wallis. *Regional Frequency Analysis*. Cambridge University Press, Cambridge, UK, 1st edition, 1997.

John Hunter. Estimating sea-level extremes under conditions of uncertain sea-level rise. *Climatic Change*, 99(3-4):331–350, 2010.

Hydrology Subcommittee. Guidelines for determining flood flow frequency. *Bulletin B*, 17, 1982.

Interagency Committee on Water Data. Guidelines for determining flood flow frequency: Bulletin 17B. Technical Report 28, Hydrol. Subcomm., Washington D.C., 1982.

International Boundary and Water Commission (IBWC). Flow of the colorado river and other western boundary streams and related data. Technical report, Department of State, United States of America, 2006.

Shaleen Jain and Upmanu Lall. Floods in a changing climate: Does the past represent the future? *Water Resources Research*, 37(12):3193–3205, 2001.

J. Jakob. Nonstationarity in extremes and engineering design. In *Extremes in a Changing Climate*. Springer, 2013. doi: 10.1007/978-94-007-4479-0 13.

Soo Min Kang and Richard P Wildes. The n-distribution bhattacharyya coefficient. Technical report, EECS-2015-02, York Univ, 2015.

Thomas R Karl, J. M. Melillo, and T. C. Peterson. *Global climate change impacts in the United States*. Cambridge University Press, 2009.

Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

Richard W. Katz. Statistics of extremes in climate change. *Climatic Change*, 100, 2010. doi: 10.1007/s10584-010-9834-5.

Richard W Katz. Statistical methods for nonstationary extremes. In *Extremes in a changing climate*, pages 15–37. Springer, 2013.

Wim Kellens, Teun Terpstra, and Philippe De Maeyer. Perception and communication of flood risks: a systematic review of empirical research. *Risk analysis*, 33(1):24–49, 2013.

M. Kendall. *Rank correlation methods*. Griffin, London, 4th edition, 1976.

Marc C Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

Himayatullah Khan, Laura Giurca Vasilescu, Asmatullah Khan, et al. Disaster management cycle-a theoretical approach. *Journal of Management and Marketing*, 6(1):43–50, 2008.

Byunghyun Kim, Brett F Sanders, Jochen E Schubert, and James S Famiglietti. Mesh type tradeoffs in 2d hydrodynamic modeling of flooding with a godunov-based flow solver. *Advances in Water Resources*, 68:42–61, 2014.

W. Kirby. Computer oriented Wilson-Hilferty transformation that preserves the first 3 moments and lower bound of the Pearson type 3 distribution. *Water Resour. Res.*, 10, 1972. doi: 10.1029/WR008i005p01251.

Thomas R Kjeldsen, Rob Lamb, and Sarka D Blazkova. Uncertainty in flood frequency analysis. *Applied Uncertainty Analysis for Flood Risk Management*, pages 153–197, 2014.

Anke Knapen, Jean Poesen, Gerard Govers, Gwendolyn Gyssels, and Jeroen Nachtergaele. Resistance of soils to concentrated flow erosion: A review. *Earth-Science Reviews*, 80(1): 75–109, 2007.

Robert E Kopp, Radley M Horton, Christopher M Little, Jerry X Mitrovica, Michael Oppenheimer, DJ Rasmussen, Benjamin H Strauss, and Claudia Tebaldi. Probabilistic 21st and 22nd century sea-level projections at a global network of tide-gauge sites. *Earth's Future*, 2(8):383–406, 2014.

Isidore K Kouadio, Syed Aljunid, Taro Kamigaki, Karen Hammad, and Hitoshi Oshitani. Infectious diseases following natural disasters: prevention and control measures. *Expert review of anti-infective therapy*, 10(1):95–104, 2012.

Demetris Koutsoyiannis and Alberto Montanari. Negligent killing of scientific concepts: the stationarity case. *Hydrological Sciences Journal*, 2015.

H Kreibich, K Piroth, I Seifert, H Maiwald, U Kunert, J Schwarz, B Merz, and AH Thieken. Is flow velocity a significant parameter in flood damage modelling? *Natural Hazards and Earth System Sciences*, 9(5):1679–1692, 2009.

H Kreibich, Giuliano Di Baldassarre, S Vorogushyn, JCJH Aerts, H Apel, GT Aronica, K Arnbjerg-Nielsen, LM Bouwer, P Bubeck, T Caloiero, et al. Adaptation to flood risk–results of international paired flood event studies. *Earth's Future*, 2017.

Heidi Kreibich, Annegret H Thieken, Th Petrow, Meike Müller, and Bruno Merz. Flood loss reduction of private households due to building precautionary measures–lessons learned from the elbe flood in august 2002. *Natural Hazards and Earth System Science*, 5(1): 117–126, 2005.

Kenneth E Kunkel, Thomas R Karl, David R Easterling, Kelly Redmond, John Young, Xungang Yin, and Paula Hennon. Probable maximum precipitation and climate change. *Geophysical Research Letters*, 40(7):1402–1408, 2013.

Jan Kyselỳ, Jan Picek, and Romana Beranová. Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change*, 72(1):55–68, 2010.

Eric Laloy and Jasper A Vrugt. High-dimensional posterior exploration of hydrologic models using multiple-try dream (zs) and high-performance computing. *Water Resour. Res.*, 48 (1), 2012a.

Eric Laloy and Jasper A Vrugt. High-dimensional posterior exploration of hydrologic models using multiple-try dream (zs) and high-performance computing. *Water Resources Research*, 48(1), 2012b.

Jonathan R Lamontagne, Jery R Stedinger, Timothy A Cohn, and Nancy A Barth. Robust national flood frequency guidelines: What is an outlier? In *Proc. World Environmental and Water Resources Congress, ASCE*, 2013.

Geert Lenderink and Erik Van Meijgaard. Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geoscience*, 1(8):511, 2008.

Carlos HR Lima, Upmanu Lall, Tara J Troy, and Naresh Devineni. A climate informed model for nonstationary flood risk prediction: Application to negro river at manaus, amazonia. *Journal of Hydrology*, 522:594–602, 2015.

Harry F Lins and Timothy A Cohn. Stationarity: wanted dead or alive? 1, 2011.

J. Lopez and F. Frances. Non-stationary flood frequency analysis in continental Spanish rivers, using clmate and reservoir indices as external covariates. *Hydrol. Earth Syst. Sci.*, 17, 2013. doi: 10.5194/hess-17-3189-2013.

Adam Luke, Brad Kaplan, Jeff Neal, Jeremiah Lant, Brett Sanders, Paul Bates, and Doug Alsdorf. Hydraulic modeling of the 2011 new madrid floodway activation: a case study on floodway activation controls. *Natural Hazards*, 77(3):1863–1887, 2015.

Adam Luke, Jasper A Vrugt, Amir AghaKouchak, Richard Matthew, and Brett F Sanders. Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the united states. *Water Resources Research*, 53(7):5469–5494, 2017.

Adam Luke, Brett F Sanders, Kristen A Goodrich, David L Feldman, Danielle Boudreau, Ana Eguiarte, Kimberly Serrano, Abigail Reyes, Jochen E Schubert, Amir AghaKouchak, et al. Going beyond the flood insurance rate map: insights from flood hazard map co-production. *Natural Hazards and Earth System Sciences*, 18(4):1097, 2018.

María José Machado, BA Botero, J López, F Francés, A Díez-Herrero, and Gerardo Benito. Flood frequency analysis of historical flood data under stationary and non-stationary modelling. 2015.

H. Madsen, D. Lawrence, M. Lang, M. Martinkova, and T. R. Kjeldsen. A review of applied methods in europe for flood-frequency analysis in a changing environment. Technical report, COST Action ES0901: Flood frequency estimation methods and environmental change, 2013.

H. Mann. Nonparametric tests against trend. *Econometrica*, 13, 1945. doi: 10.2307/1907187.

Frederique Martini and Roberto Loat. Handbook on good practices for flood mapping in europe. 2007.

Eduardo S Martins and Jery R Stedinger. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3): 737–744, 2000.

Frank J Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.

Nicholas C Matalas. Comment on the announced death of stationarity. *Journal of Water Resources Planning and Management*, 138(4):311–312, 2012.

Andrew McCowan. *Flood Hydraulics: Numerical Models, Chapter 4 of Book 6 in Australian Rainfall and Runoff: A Guide to Flood Estimation*. Commonwealth of Australia, 2016.

Marcia McNutt. Hazards without disasters. *Science*, 353(6296):201–201, 2016.

Melisa Menéndez and Philip L Woodworth. Changes in extreme high water levels based on a quasi-global tide-gauge data set. *Journal of Geophysical Research: Oceans*, 115(C10), 2010.

Venkatesh Merwade, Francisco Olivera, Mazdak Arabi, and Scott Edleman. Uncertainty in flood inundation mapping: current issues and future directions. *Journal of Hydrologic Engineering*, 13(7):608–620, 2008.

B Merz, JCJH Aerts, Karsten Arnbjerg-Nielsen, M Baldi, A Becker, Adeline Bichet, G Blöschl, LM Bouwer, Achim Brauer, F Cioffi, et al. Floods and climate: emerging perspectives for flood risk assessment and management. *Natural Hazards and Earth System Sciences*, 14(7):1921–1942, 2014.

Bruno Merz, AH Thieken, and Martin Gocht. Flood risk mapping at the local scale: concepts and challenges. *Flood risk management in Europe*, pages 231–251, 2007.

Ralf Merz and Günter Blöschl. Flood frequency hydrology: 1. temporal, spatial, and causal expansion of information. *Water Resources Research*, 44(8), 2008.

Volker Meyer, Christian Kuhlicke, J Luther, Sebastian Fuchs, S Priest, W Dorner, K Serrhini, Joanna Pardoe, Simon McCarthy, J Seidel, et al. Recommendations for the user-specific enhancement of flood maps. *Natural Hazards and Earth System Sciences*, 12(5):1701, 2012.

P. C. D. Milly, R. T. Wetherald, K.A. Dunne, and T. L. Delworth. Increasing risk of great floods in a changing climate. *Nature*, 415, 2002. doi: 10.1038/415514a.

Paul CD Milly, Julio Betancourt, Malin Falkenmark, Robert M Hirsch, Zbigniew W Kundzewicz, Dennis P Lettenmaier, and Ronald J Stouffer. Stationarity is dead: Whither water management? *Science*, 319(5863):573–574, 2008.

Paul CD Milly, Julio Betancourt, Malin Falkenmark, Robert M Hirsch, Zbigniew W Kundzewicz, Dennis P Lettenmaier, Ronald J Stouffer, Michael D Dettinger, and Valentina Krysanova. On critiques of stationarity is dead: whither water management?. *Water Resources Research*, 51(9):7785–7789, 2015.

Seung-Ki Min, Xuebin Zhang, Francis W Zwiers, and Gabriele C Hegerl. Human contribution to more-intense precipitation extremes. *Nature*, 470(7334):378–381, 2011.

H de Moel, J van Alphen, and JCJH Aerts. Flood maps in europe–methods, availability and use. *Natural Hazards and Earth System Sciences*, 9(2):289–301, 2009.

Hamed R Moftakhari, Gianfausto Salvadori, Amir AghaKouchak, Brett F Sanders, and Richard A Matthew. Compounding effects of sea level rise and fluvial flooding. *Proceedings of the National Academy of Sciences*, 114(37):9785–9790, 2017.

Alberto Montanari and Demetris Koutsoyiannis. Modeling and mitigating natural hazards: Stationarity is immortal! *Water Resources Research*, 50(12):9748–9756, 2014.

Ch Mudersbach and J Jensen. Nonstationary extreme value analysis of annual maximum water levels for designing coastal structures on the german north sea coastline. *Journal of Flood Risk Management*, 3(1):52–62, 2010.

Munich Re. Topics geo annual review: natural catastrophes 2005. *Munich Re, Munich*, 15, 2005.

Natural Resources Conservation Service (NRCS). National engineering handbook section 4. Technical report, United States Department of Agriculture, 1985.

Steve Nixon, Jennifer Horn, Edith Hödl-Kreuzbauer, Arjan ter Harmsel, Dominique Van Erdeghem, and Thomas Dworak. European overview assessment of member states' reports on preliminary flood risk assessment and indentification of areas of potentially significant flood risk. Technical report, European Union, 2015.

NOAA. Environmental measurements systems sensor specifications and measurement algorithms. Technical report, Center for Operational Oceanographic Products and Services, 2013.

Michael Nones. Flood hazard maps in the european context. *Water International*, 42(3): 324–332, 2017.

MJ Nozdryn-Plotnicki and WE Watt. Assessment of fitting techniques for the log pearson type 3 distribution using monte carlo simulation. *Water Resources Research*, 15(3):714–718, 1979.

Riyandi Nuswantoro, F Diermanse, and Frank Molkenthin. Probabilistic flood hazard maps for jakarta derived from a stochastic rain-storm generator. *Journal of Flood Risk Management*, 9(2):105–124, 2016.

Jayantha Obeysekera and Joseph Park. Scenario-based projection of extreme sea levels. *Journal of Coastal Research*, 29(1):1–7, 2012.

J. Rolf Olsen. Climate change and floodplain management in the United States. *Climatic Change*, 76, 2002. doi: 10.1007/s10584-005-9020-3.

J Rolf Olsen, James H Lambert, and Yacov Y Haimes. Risk of extreme events under nonstationary conditions. *Risk Analysis*, 18(4):497–510, 1998.

TBMJ Ouarda and S El-Adlouni. Bayesian nonstationary frequency analysis of hydrological variables1, 2011.

Edward T Pasterick. The national flood insurance program. *Paying the price: The status and role of insurance against natural disasters in the United States*, pages 125–154, 1998.

Murray C Peel and Günter Blöschl. Hydrological modelling in a changing world. *Progress in Physical Geography*, 35(2):249–261, 2011.

L Perreault, J Bernierand, B Bobée, and E Parent. Bayesian change-point analysis in hydrometeorological time series. part 2. comparison of change-point models and forecasting. *J. Hydrol.*, 235(3):242–263, 2000.

Franck Picard. An introduction to mixture models. *Statistics for Systems Biology, Research Report*, (7), 2007.

D Pilgrim. *Australian rainfall and runoff: a guide to flood estimation*, volume 1. 2001.

Victor M Ponce and Richard H Hawkins. Runoff curve number: Has it reached maturity? *Journal of hydrologic engineering*, 1(1):11–19, 1996.

I. Prosdocimi, T. R. Kjeldsen, and C. Svensson. Non-stationarity in annual and season series of peak flow and precipitation in the uk. *Nat. Hazards Earth Syst. Sci.*, 14, 2014. doi: 10.5194/nhess-14-1125-2014.

D. A. Raff, T. Pruitt, and L. D. Brekke. A framework for assessing flood frequency based on climate projection information. *Hydrol. Earth Syst. Sci.*, 13, 2009. doi: 10.1007/s10584-005-9020-3.

Laura K Read and Richard M Vogel. Hazard function analysis for flood planning under nonstationarity. *Water Resources Research*, 2016.

Dirceu S. Reis Jr. and Jery R. Stedinger. Bayesian mcmc flood frequency analysis with histroical information. *J. Hydrol.*, 313, 2005. doi: 10.1016/j.jhydrol.2005.02.028.

B Renard, V Garreta, and M Lang. An application of bayesian analysis and markov chain monte carlo methods to the estimation of a regional trend in annual maxima. *Water resources research*, 42(12), 2006a.

Benjamin Renard, Michel Lang, and Philippe Bois. Statistical analysis of extreme events in a non-stationary context via a bayesian framework: case study with peak-over-threshold data. *Stochastic environmental research and risk assessment*, 21(2):97–112, 2006b.

Benjamin Renard, Xun Sun, and Michel Lang. Bayesian methods for non-stationary extreme value analysis. In *Extremes in a Changing Climate*, pages 39–95. Springer, 2013.

Met al Rhein, Stephen R Rintoul, Shigeru Aoki, Edmo Campos, Don Chambers, Richard A Feely, Sergey Gulev, GC Johnson, SA Josey, A Kostianoy, et al. Observations: ocean. In *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. 2013.

Keywan Riahi, Arnulf Grübler, and Nebojsa Nakicenovic. Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technological Forecasting and Social Change*, 74(7):887–935, 2007.

Paul Marie Louis Pierre Richer. *Artistic anatomy*. Watson-Guptill Publications, 1986.

Justin Ritchie and Hadi Dowlatabadi. The 1000 gtc coal question: Are cases of vastly expanded future coal combustion still plausible? *Energy Economics*, 65:16–31, 2017.

Justin Ritchie and Hadi Dowlatabadi. Defining climate change scenario characteristics with a phase space of cumulative primary energy and carbon intensity. *Environmental Research Letters*, 13(2):024012, 2018.

Joeri Rogelj, Michel Den Elzen, Niklas Höhne, Taryn Fransen, Hanna Fekete, Harald Winkler, Roberto Schaeffer, Fu Sha, Keywan Riahi, and Malte Meinshausen. Paris agreement climate proposals need a boost to keep warming well below 2 c. *Nature*, 534(7609):631, 2016.

Jose D Salas and Jayantha Obeysekera. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering*, 19(3): 554–568, 2013.

Jose D. Salas and Jayantha Obeysekera. Revisiting the concepts of return perios and risk for nonstationary hydrologic extreme events. *J. Hydrol. Eng.*, 19, 2014. doi: 10.1061/(ASCE) HE.1943-5584.0000820.

Jose D Salas, Balaji Rajagopalan, Laurel Saito, and Casey Brown. Climate nonstationarity and water resources management. *Journal of Water Resources Planning and Management*, 138(5):385–388, 2012.

Johnny Saldaña. *The coding manual for qualitative researchers*. Sage, 2015.

José Luis Salinas, Andrea Kiss, Alberto Viglione, Reinhard Viertl, and Günter Blöschl. A fuzzy bayesian approach to flood frequency estimation with imprecise historical information. *Water Resources Research*, 52(9):6730–6750, 2016.

Gianfausto Salvadori and Carlo De Michele. Multivariate extreme value methods. In *Extremes in a changing climate*, pages 115–162. Springer, 2013.

Brett F Sanders. Hydrodynamic modeling of urban flood flows and disaster risk reduction. In *Oxford Research Encyclopedia of Natural Hazard Science*, page DOI: 10.1093/acrefore/9780199389407.013.127. Oxford University Press, 2017.

Brett F Sanders, Jochen E Schubert, and Russell L Detwiler. Parbrezo: A parallel, unstructured grid, godunov-type, shallow-water code for high-resolution flood inundation modeling at the regional scale. *Advances in Water Resources*, 33(12):1456–1467, 2010.

John Schaake, Shuzheng Cong, and Qingyun Duan. U.s. mopex data set. Technical report, Office of Hydrologic Development, NOAA National Weather Service, Silver Spring, MD, 2006.

Stephen H Schneider. What is' dangerous' climate change? *Nature*, 411(6833):17, 2001.

Francesco Serinaldi and Chris G. Kilsby. Stationarity is undead: uncertainty dominates the distribution of extremes. *Adv. Water Resour.*, 77, 2015a. doi: 10.1016/j.advwatres.2014.12.013.

Francesco Serinaldi and Chris G Kilsby. Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, 77:17–36, 2015b.

Jonathan Richard Shewchuk. Triangle: Engineering a 2d quality mesh generator and delaunay triangulator. In *Applied computational geometry towards geometric engineering*, pages 203–222. Springer, 1996.

M Sholders. San diego county hydrology manual. Technical report, County of San Diego, Department of Public Works, Flood Control Section, 2003.

Artur Tiago Silva, Mauro Naghettini, and Maria Manuela Portela. On some aspects of peaks-over-threshold modeling of floods under nonstationarity using climate covariates. *Stoch. Environ. Res. Risk. Assess.*, 2015. doi: 10.1007/s00477-015-1072-y.

Nuno Eduardo Simões, Susana Ochoa-Rodríguez, Li-Pen Wang, Rui Daniel Pina, Alfeu Sá Marques, Christian Onof, and João P Leitão. Stochastic urban pluvial flood hazard maps based upon a spatial-temporal rainfall generator. *Water*, 7(7):3396–3406, 2015.

J.R. Slack and J.M. Landwehr. Hydro-climatic data network (HCDN): A U.S. Geological Survey streamflow data set for the United States for the study of climate variations. Technical Report USGS Open-file Report 92-129, United States Geological Survey, Reston, VA, 1992.

Robert Soden, Leah Sprain, and Leysia Palen. Thin grey lines: Confrontations with risk on colorado's front range. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2042–2053. ACM, 2017.

David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde. Bayesian measures of model complexity and fit. *J. Roy. Stat. Soc. B*, 64, 2002. doi: 10.1111/1467-9868.00353.

Raphael Spiekermann, Stefan Kienberger, John Norton, Fernando Briones, and Juergen Weichselgartner. The disaster-knowledge matrix–reframing and evaluating the knowledge challenges in disaster risk reduction. *International Journal of Disaster Risk Reduction*, 13: 96–108, 2015.

Mojca Šraj, Alberto Viglione, Juraj Parajka, and Günter Blöschl. The influence of non-stationarity in extreme hydrological events on flood frequency estimation. *J. Hydrol. Hydromech., doi*, 10, 2016.

R Srikanthan and TA McMahon. Log pearson iii distributioneffect of dependence, distribution parameters and sample size on peak annual flood estimates. *Journal of Hydrology*, 52 (1-2):149–159, 1981.

Jery R Stedinger and Veronica W Griffis. Flood frequency analysis in the united states: Time to update, 2008.

Jery R. Stedinger and Veronica W. Griffis. Getting from here to where? flood frequency analysis and climate. *J. Am. Water Resour. Assoc.*, 47, 2011a. doi: 10.1111/j.1752-1688.2011.00545.x.

Jery R Stedinger and Veronica W Griffis. Getting from here to where? flood frequency analysis and climate. *JAWRA Journal of the American Water Resources Association*, 47 (3):506–513, 2011b.

Scott Steinschneider and Upmanu Lall. A hierarchical bayesian regional model for non-stationary precipitation extremes in northern california conditioned on tropical moisture exports. *Water Resources Research*, 51(3):1472–1492, 2015.

W.G. Strupczewski, V. P. Singh, and W. Feluch. Non-stationary approach to at-site flood frequency modelling i. Maximum likelihood estimation. *J. Hydrol.*, 248, 2001. doi: 10.1016/S0022-1694(01)00397-3.

L Sundermann, O Schelske, and P Hausman. Mind the risk - a global ranking of cities under threat from natural disasters. Technical report, Swiss Reinsurance Company, 2014.

Tijuana River National Estuarine Research Reserve (TRNERR). Comprehensive management plan. Technical report, National Oceanic and Atmospheirc Administration, United States Department of Commerce, 2010.

Y. Tramblay, L. Neppel, J. Carreau, and K. Najib. Non-stationary frequency analysis of heavy rainfall events in southern france. *Hydrolog. Sci. J.*, 14, 2014. doi: 10.5194/nhess-14-1125-2014.

Robert Traver. Flood risk management: call for a national strategy. American Society of Civil Engineers, 2014.

Kevin E Trenberth. Changes in precipitation with climate change. *Climate Research*, 47 (1-2):123–138, 2011.

United States Army Corps of Engineers (USACE). Hydrologic modeling system hec-hms technical reference manual. Technical report, United States Department of Defense, 2000.

USACE. Incorporating sea level change in civil works programs, 2013.

Detlef P Van Vuuren, Michel GJ Den Elzen, Paul L Lucas, Bas Eickhout, Bart J Strengers, Bas Van Ruijven, Steven Wonink, and Roy van Houdt. Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs. *Climatic Change*, 81(2):119–159, 2007.

Detlef P Van Vuuren, Jae Edmonds, Mikiko Kainuma, Keywan Riahi, Allison Thomson, Kathy Hibbard, George C Hurtt, Tom Kram, Volker Krey, Jean-Francois Lamarque, et al. The representative concentration pathways: an overview. *Climatic change*, 109(1-2):5, 2011a.

Detlef P Van Vuuren, Elke Stehfest, Michel GJ den Elzen, Tom Kram, Jasper van Vliet, Sebastiaan Deetman, Morna Isaac, Kees Klein Goldewijk, Andries Hof, Angelica Mendoza Beltran, et al. Rcp2. 6: exploring the possibility to keep global mean temperature increase below 2 c. *Climatic Change*, 109(1-2):95, 2011b.

Alberto Viglione, Bruno Merz, Nguyen Viet Dung, Juraj Parajka, Thomas Nester, and Günter Blöschl. Attribution of regional flood changes based on scaling fingerprints. *Water Resources Research*, 52(7):5322–5340, 2016.

Gabriele Villarini and James A. Smith. Flood peak distributions for the eastern United States. *Water Resour. Res.*, 46, 2009. doi: 10.1029/2009WR008395.

Gabriele Villarini, Francesco Serinaldi, James A Smith, and Witold F Krajewski. On the stationarity of annual flood peaks in the continental united states during the 20th century. *Water Resources Research*, 45(8), 2009a.

Gabriele Villarini, James A. Smith, Francesco Serinaldi, Jerad Bales, Paul D. Bates, and Witold F. Krajewski. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Adv. Water Resour.*, 32, 2009b. doi: 10.1016/j.advwatres.2009.05.003.

Richard M Vogel, Chad Yaindl, and Meghan Walter. Nonstationarity: Flood magnification and recurrence reduction factors in the united states1, 2011.

Jasper A Vrugt. Markov chain monte carlo simulation using the dream software package: Theory, concepts, and matlab implementation. *Environmental Modelling & Software*, 75: 273–316, 2016.

Jasper A Vrugt, CJF Ter Braak, CGH Diks, Bruce A Robinson, James M Hyma, and Dave Higdon. Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *Int. J. Nonlin. Sci. Num.*, 10(3):273–290, 2009.

J.R. Wallis, D.P. Lettenmaier, and E. F. Wood. A daily hydroclimatological data set for the continental United States. *Water Resour. Res.*, 27, 1991. doi: 10.1029/91WR00977.

Junqiang Xia, Fang Yenn Teo, Binliang Lin, and Roger A Falconer. Formula of incipient velocity for flooded vehicles. *Natural Hazards*, 58(1):1–14, 2011.

Junqiang Xia, Roger A Falconer, Yejiang Wang, and Xuanwei Xiao. New criterion for the stability of a human body in floodwaters. *Journal of Hydraulic Research*, 52(1):93–104, 2014.

Xin Yu, Timothy A Cohn, Jery R Stedinger, Karen Karvazy, and Veronica Webster. Flood frequency analysis in the context of climate change. In *World environmental and water resources congress*, pages 2376–2385. American Society of Civil Engineers Reston, 2015.

C Zervas. Noaa technical report nos co-ops 067: extreme water levels of the united states 1893–2010. Technical report, Technical report, National Oceanic and Atmospheric Administration/National Ocean Service/Center for Operational Oceanographic Products and Services, Silver Spring, MD, 2013a.

Chris Zervas. Extreme water levels of the united states 1893-2010. Technical report, National Oceanic and Atmospheric Administration, 2013b.

Appendix A

The Log-Pearson Type III Distribution

A record of block maxima discharges, $\{Q_1, ..., Q_n\}$, follows the LPIII distribution if random variable $\mathbf{X} = \log_{10}(\mathbf{Q})$ is distributed according to the PIII distribution. The PIII distribution parameters are defined by μ, σ, and γ, which represent the mean, standard deviation, and skewness of $\mathbf{X}$, respectively. The probability density function, $f(x)$, and cumulative distribution function, $F(x)$, are defined as

$$f(x) = \frac{(x-\xi)^{\epsilon-1} e^{-(x-\xi)/\beta}}{\beta^{\epsilon}\Gamma(\epsilon)} \tag{A.1}$$

$$F(x) = G\left(\epsilon, \frac{x-\xi}{\beta}\right)/\Gamma(\epsilon) \tag{A.2}$$

where $\epsilon = 4/\gamma^2$, $\beta = \frac{1}{2}\sigma\gamma$, and $\xi = \mu - 2\sigma/\gamma$ [Hosking and Wallis, 1997]. If $\gamma > 0$, then ξ is a lower bound ($\xi \leq x < \infty$). If $\gamma < 0$, then ξ is an upper bound ($-\infty \leq x < \xi$) and

$$f(x) = \frac{(\xi - x)^{\epsilon-1} e^{-(\xi-x)/\beta}}{\beta^{\epsilon}\Gamma(\epsilon)} \tag{A.3}$$

$$F(x) = 1 - G\left(\epsilon, \frac{\xi - x}{\beta}\right)/\Gamma(\epsilon) \tag{A.4}$$

In the special case where $\gamma = 0$, the distribution is normal and the range of x is $-\infty < x < \infty$. Here, $\Gamma(\cdot)$ and $G(\cdot)$ represent the gamma and incomplete gamma functions, respectively. An analytical form of the inverse PIII distribution, $x(F)$, does not exist, thus it is often approximated to calculate discharges associated with specific quantiles, p, or return periods, T

$$\log_{10}(q_p) = \mu + \sigma K_p(\gamma) \tag{A.5}$$

$$K_p(\gamma) = \frac{2}{\gamma}\left(1 + \frac{\gamma n_p}{6} - \frac{\gamma^2}{36}\right)^3 - \frac{2}{\gamma} \tag{A.6}$$

where $\log_{10}(q_p)$ represents the p^{th} quantile of the LPIII distribution, and n_p is the p^{th} quantile of the standard normal distribution. $K_p(\gamma)$ is the frequency factor, or the p^{th} quantile of the PIII distribution with mean 0, standard deviation of 1, and shape γ. The frequency factor is approximated by the Wilson and Hilferty transformation in Equation A.6, which is accurate for $-2 < \gamma < 2$ and $0.01 \leq p \leq 0.99$ or $1.01 \leq T \leq 100$ [Kirby, 1972, Reis Jr. and Stedinger, 2005].

Because of the numerical difficulties in evaluating $\Gamma(\cdot)$ at large values of ϵ (γ values near 0), evaluation of the analytic PIII is often avoided. In this study, evaluation of the analytic PIII distribution near $\gamma = 0$ is accomplished using Chebyshev approximations for $\log_e[\Gamma(\cdot)]$, which do not require evaluation of $\Gamma(\cdot)$ [Cody and Hillstrom, 1967]. Conveniently, $\log_e[\Gamma(\cdot)]$ can be computed using the MATLAB® function gammaln(), and evaluating the analytic

PIII pdf is achieved in the log-space.

Appendix B

Bayesian Parameter Estimation of the NS-GEV Model

We use the Bayesian approach for parameter estimation in this study. The Bayesian approach for parameter estimation is advantageous for several reasons. First, Bayesian parameter estimation produces the joint distribution of parameter values rather than point estimates. The joint distribution is very useful for uncertainty analysis, since we can sample from the joint distribution to characterize the parametric uncertainty of the statistical model. Secondly, Bayesian parameter estimation allows us to stabilize parameter estimates by incorporating information other than the data in the estimation procedure. In the case of NEVA, authors have shown that maximum likelihood estimation can lead to biased estimates of the shape or tail parameter [El Adlouni et al., 2007], and Mudersbach and Jensen [2010] called for more accurate estimation procedures in the context of NEVA applied to extreme ocean levels. The Bayesian approach allows us to improve parameter estimation via the informative prior. The following section briefly outlines our application of Bayesian methods to estimate the parameters of the NS-GEV model (Equations 2.5, 2.6).

In this study, we are interested in inferring the NS-GEV parameters, $\boldsymbol{\theta} = \{\mu_o, \sigma, \xi\}$. The time varying location of the NS-GEV model μ_i is formulated as a function of the parameter μ_o and the covariate ΔMSL_i, where $\mu_i = \mu_o + \Delta\text{MSL}_i$. The covariate ΔMSL_i, or the change in mean sea level from the year 2000, is calculated from the historic record during the fitting period and obtained from Kopp et al. [2014] when the model is used for prediction. For the fitting procedure, we use a centered 19-year moving average of observed (hourly) water levels to define mean sea level. We then take the difference between mean sea level in year i with mean sea level in year 2000 to define a vector of ΔMSL_i values. With the covariate vector defined, we apply Bayes' theorem to obtain the posterior distribution of parameter values given the data $\boldsymbol{y}$ and vector of covariate values $\boldsymbol{x}$

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) \propto p(\boldsymbol{\theta}) L(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) \tag{B.1}$$

where $\boldsymbol{y}$ is the vector of annual maximum water levels, $p(\boldsymbol{\theta})$ is the prior probability density of the parameter values, and $L(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x})$ is the likelihood of the parameter values given the data and covariate vector. Notice that we do not include the marginal likelihood or evidence term (the denominator of Bayes' theorem) in Equation B.1, since parameter inference can be made from the un-normalized density when Markov Chain Monte Carlo (MCMC) methods are used to sample from the posterior. Producing a posterior sample of $\boldsymbol{\theta}$ with MCMC methods requires specification the prior density and likelihood function.

The likelihood function quantifies the probability density of $\boldsymbol{y}$ according to the statistical model with parameterization $\boldsymbol{\theta}$ and vector of covariate values $\boldsymbol{x}$. In this study, the likelihood function is defined as

$$L(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) = \prod_{i=1}^{N} p(y_i|\boldsymbol{\theta}, x_i) \tag{B.2}$$

where i is an index of years in the historic record, N is the number of observations in the

annual maximum record, y_i denotes the i^{th} observation in the data vector $\boldsymbol{y}$, and x_i denotes the i^{th} value in the covariate vector $\boldsymbol{x}$. The term $p(y_i|\boldsymbol{\theta}, x_i)$ is the probability density of observation y_i, which is given by the GEV probability density function evaluated at y_i with parametrization $\boldsymbol{\theta}$ and covariate value x_i. Recall that $\mu_i = \mu_o + x_i$, which causes the parameters of the GEV distribution change with x_i values. Thus, Equation B.2 is evaluated at time-varying parameters. The GEV probability density function is given by Coles [2001].

The prior represents an analyst's knowledge of $\boldsymbol{\theta}$ before any data is collected, which is described mathematically by the joint probability density of the parameters within $\boldsymbol{\theta}$. The prior is useful for estimating the parameters of an extreme value distribution, since the prior allows us to incorporate information other than the limited historic data record in the inference of distribution parameters. In this study, we use uninformative priors on μ_o and σ with an informative marginal prior specified on ξ

$$p(\xi) = \mathcal{N}(\xi_{\text{st}}, \sigma_{\text{st}}) \tag{B.3}$$

where $\mathcal{N}(\xi_{\text{st}}, \sigma_{\text{st}})$ denotes the normal distribution with mean ξ_{st} and standard deviation σ_{st}. In this study, ξ_{st} is the maximum likelihood estimate of ξ at the Los Angeles gage reported by Zervas [2013b], which was obtained by fitting a stationary GEV distribution to the historic data. Here, σ_{st} is the standard deviation of the normal distribution with mean ξ_{st} and 95% confidence interval reported by Zervas [2013b].

We use uninformative priors on μ_o and σ which means that inference of these parameters is based on the data alone. The marginal priors for μ_o and σ and were specified as a uniform distribution, $\mathcal{U}(a, b)$, where a and b represent the lower and upper bounds of the uniform priors, respectively. The uniform distributions assign equal prior density to all values within the bounds, and thus the prior does not provide information about the value of these parameters if the bounds are large enough to encompass realistic values. The joint prior

Table B.1: NS-GEV parameters and marginal prior distributions. The multivariate prior density is simply the product of the marginal distributions evaluated at a proposed parameter combination.

Parameters	Units	Prior	lower bound	upper bound	mean	std.
μ_o	m	$\mathcal{U}(a,b)$	-10	10	–	–
σ	m	$\mathcal{U}(a,b)$	0	2	–	–
ξ	–	$\mathcal{N}(\xi_{\text{st}}, \sigma_{\text{st}})$	–	–	-0.17	0.06

density of $\boldsymbol{\theta}$ is simply the product of the marginal distributions, or $p(\boldsymbol{\theta}) = p(\xi)p(\mu_o)p(\sigma)$. Table B.1 provides the NS-GEV parameters and prior distributions. We note that using an informative prior on just the shape parameter has be coined "generalized maximum likelihood estimation" (GLE) in the hydrologic literature. El Adlouni et al. [2007] showed that GLE is preferable to maximum likelihood estimation in the context of NS-GEV parameter inference, and Martins and Stedinger [2000] reached the same conclusion in the context of stationary GEV parameter inference.

With the prior and likelihood functions defined, we then use MCMC methods to produce a posterior sample of parameter values. Sampling methods such as MCMC provide a sample of $\boldsymbol{\theta}$ where the density of parameter values is proportional to the product of the likelihood function and prior density. For a complete description of posterior sampling via Monte Carlo simulation, please see Section 2 of Vrugt [2016]. In this study, we use the DiffeRential Evolution Adaptive Metropolis algorithm (DREAM$_{(\text{ZS})}$) to generate samples of the target distribution [Vrugt et al., 2009, Laloy and Vrugt, 2012a], and assess convergence through the $\hat{R}$-statistic of Gelman and Rubin [1992]. The MCMC procedure utilized three Markov chains, each including 8000 generations of $\boldsymbol{\theta}$. Fifty percent of the generations from each chain were discarded as burn-in, and the remaining generations were combined into a single matrix. This procedure results in a posterior sample size of 12,000 generations of $\boldsymbol{\theta}$. Parameter inference and random sampling of NS-GEV parameters is now based on the posterior samples of $\boldsymbol{\theta}$ produced by the MCMC algorithm.

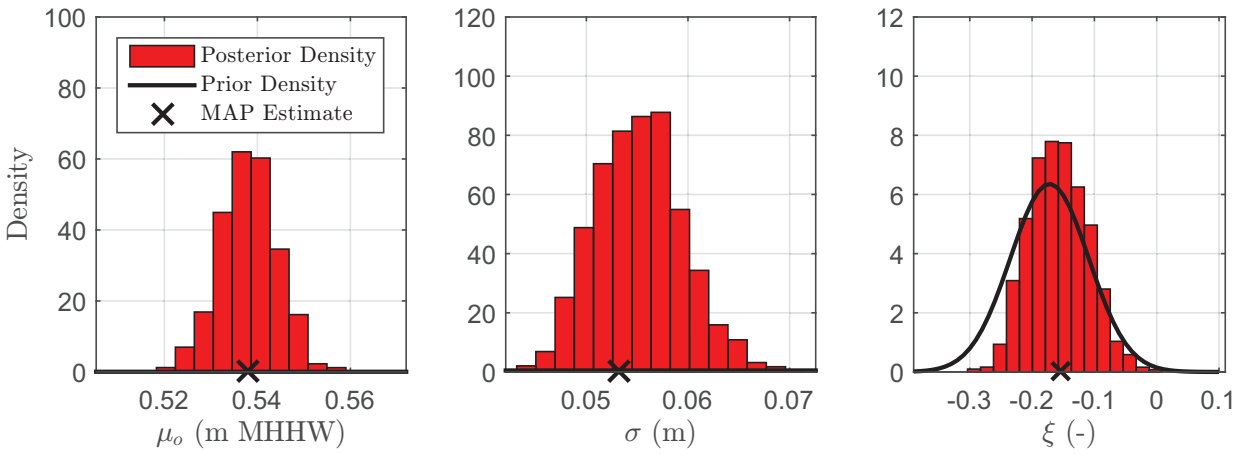

Figure B.1: Posterior sample of NS-GEV parameter values. The black lines show the prior density, and the red histograms illustrate the posterior density. The crosses denote the maximum a posteriori parameter (MAP) estimate, which we use to illustrate the NS-GEV model in Figure 2.1. We sample parameter values from the red histograms in the Monte Carlo experiments.

Figure B.1 shows the posterior density of parameter values obtained from the fitting procedure (red histograms) compared to the prior density (black lines). We note that the posterior density is calculated empirically from the posterior sample. The black crosses denote the maximum a posteriori (MAP) estimate of $\boldsymbol{\theta}$, which is the parameter combination that maximizes Equation B.1. The MAP estimate is the parameter combination used to illustrate the NS GEV model in Figure 2.1. Notice that the shape parameter ξ is negative [Zervas, 2013b], which indicates that the extreme value distribution has an upper bound. Future work could repeat the Monte Carlo experiment with an un-bounded or heavy-tailed distribution to study the sensitivity of $R(z,t)$ and $T(z)$ estimates to distribution tails. In this study, the posterior sample is used to characterize the parametric uncertainty of the NS-GEV model in the Monte Carlo experiments (i.e Step 1c from Section 2.3.5).

To randomly sample from the posterior distribution, there is no need to fit a distribution to the posterior sample of $\boldsymbol{\theta}$ generations. Instead, for each trial in the Monte Carlo Experiment, we generate a random integer value from the uniform distribution on the set $\{1, 2, 3....11998, 11999, 12000\}$. Recall that 12,000 is the number of $\boldsymbol{\theta}$ generations in the

posterior sample. The random integer is used as a index to select a θ generation from the posterior sample, which parameterizes the NS-GEV model in the proceeding steps of the Monte Carlo Experiment. The frequency that different θ values appear in the posterior sample expresses the posterior probability of θ values.

Appendix C

Flood Hazard Mapping Methodology

The flood hazard maps presented in this study resulted from three distinct tasks: flood frequency analysis (FFA), hydrologic and hydraulic modeling, and post-processing of model output. Generally speaking, FFA estimates the recurrence interval of rare flooding events, while hydrologic and hydraulic modeling predicts the hazards associated with simulated floods (depths, velocities, extents, etc.). In this study, post-processing methods are used to combine the results of multiple simulations onto a single map. The following sections outline our FFA, hydraulic modeling, and post-processing methods so that the interested modeler can produce the hazard maps presented herein.

C.1 Flood Frequency Analysis

FFA is complicated in the coastal zone due to the multiple causes or "drivers" of flooding. In this study, we mapped flooding caused by extreme ocean levels, streamflow from the Tijuana (TJ) River, and precipitation over Los Laureles and Smuggler's gulch watersheds (Figure 3.1). The presence of multiple flood drivers often warrants a multivariate approach

Table C.1: Pearson's correlation coefficient matrix between the three drivers of flooding considered in this study. The numeric values in the table describe the correlation between the variables in the row and column headings. Correlation coefficients were determined using de-trended tide gage data at La Jolla (NOAA station 9410230), precipitation measurements from the San Diego International Airport (NOAA network ID GHCND:USW00023188), and TJ River streamflow measurements at the US/MX border recorded by the International Boundary and Water Commission.

	Ocean Level (daily mean)	**TJ Stream flow (daily mean)**	**Precipitation (24 hr sum)**
Ocean Level (daily mean)	1	0.12	0.21
TJ Stream flow (daily mean)	0.12	1	0.17
Precipitation (24hr sum)	0.21	0.17	1

for FFA [Salvadori and De Michele, 2013]. Under this approach, multivariate extreme value analysis (EVA) is used to estimate the probability of scenarios where multiple extremes occur simultaneously. However, we did not conduct multivariate EVA in this study because of the low correlation between flood drivers and the lack of emergent flood hazards caused by the joint occurrence of extremes.

Table C.1 presents the Pearson's correlation coefficient matrix between the flood drivers considered herein. The relatively low correlation is somewhat surprising but understandable. Extended periods of above average rainfall in the upper TJ River Watershed cause large streamflow events, whereas relatively short-lived coastal storm systems can elevate ocean water levels and lead to intense precipitation. The low correlation between flood drivers demonstrates that the simultaneous occurrence of extreme events would be especially rare. Perhaps more importantly, hydraulic model sensitivity analysis revealed that predicted flood depths, extents, and velocities are insensitive to the joint-occurrence of extremes in this system. For example, flood depths predicted by the hydraulic model are not sensitive to the downstream ocean level during large TJ River floods. The lack of "sufficient" correlation between drivers and the hydraulic model's insensitivity to the joint occurrence of extremes allows us to consider the flood drivers independently and use univariate EVA for frequency analysis.

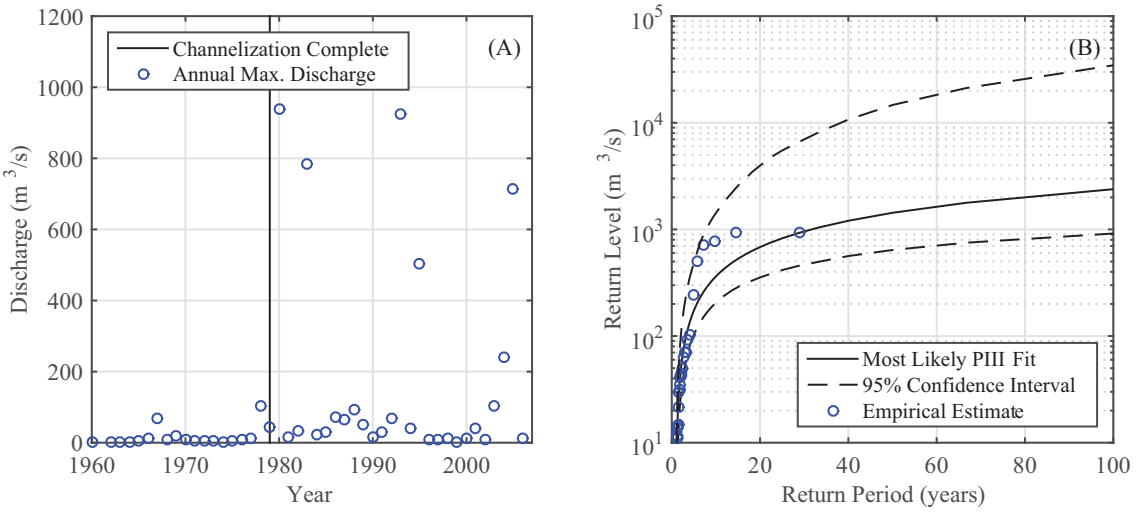

Figure C.1: A) Annual maximum discharge record of the TJ River. Only flood peaks post-channelization were used for PIII parameter inference. B) Flood frequency estimates derived from the PIII distribution fitted to the log-transformed discharge data from 1979 - 2006. Notice the large variance in flood frequency estimates.

C.1.1 Tijuana River Flood Frequency Analysis

FFA of TJ River flows was based on a Pearson Type III (PIII) distribution fitted to the historic record of log-transformed annual maximum discharges. This approach is consistent with the recommended FFA methodology in the US [Hydrology Subcommittee, 1982]. The data record originated from TJ River flow measurements at the US/MX border reported by the International Boundary and Water Commission. To infer the parameters of the PIII distribution, we used the Bayesian parameter estimation technique described by Luke et al. [2017] where an informative prior was used to incorporate regional information about the skewness of the PIII distribution. For the TJ River, parameter estimation was complicated by signs of nonstationarity in the historic record, or time variant statistical properties of the annual maximum discharge data.

Figure C.1A shows the full data record at the US/MX border. At the time of this study, data was not available after 2006. The *black* line in Figure C.1A denotes the year when the TJ River channelization was completed, which appeared to alter the mean and standard

deviation of the flood peaks. Indeed, the pre-channelization distribution is different from the post-channelization distribution at the 0.05 significance level according to the two-sample Kolmogorov-Smirnov test [Massey Jr, 1951]. Due to the apparent change in the distribution of flood peaks following channelization, we did not use data prior to 1979 for estimation of the PIII parameters. The choice to omit data prior to channelization creates a relatively small sample size for parameter estimation and leads to large variance in the estimated return periods (Figure C.1B). Assuming stationarity following the channelization, the return periods in Figure C.1B are simply the inverse of the annual exceedance probabilities associated with the return levels on the y axis. If the pre-channelization flood peaks are included in the frequency analysis, we risk bias in the parameter estimates and resulting return periods.

Notice also that the empirical frequency curve shown in Figure C.1B appears to change shape near the 5 year return period level. We attribute this to the considerable influence of upstream reservoirs on large TJ River flows. Spillway discharges occurred during four of the annual maximum events from 1979 - 2006, but did not affect the majority of the relatively small, run-off driven annual maximum events. The various flood generating mechanisms and the change in the shape of the empirical frequency curve both indicate that the distribution of flood peaks is not the same for small and large annual maximum events. This causes a poor fit of the PIII distribution to the data in the modern period of record and creates even more variance in the frequency estimates. It is unlikely that the variance can be significantly reduced without expanding the sample size through watershed modeling and simulation of peak flows, which was outside the scope of this study. It is very important to note that exceedance probabilities and corresponding frequency estimates based on the historic TJ River discharges alone are unavoidably uncertain.

Table C.2: Most likely estimates of exceedance probabilities associated with extreme ocean levels, TJ Stream flow, and precipitation. These values were used as boundary conditions for the hydrologic and hydraulic modeling.

Annual Exceedance Probability (2015)	Ocean Level (m, NAVD88)	TJ Stream flow (m^3/s)	Precipitation (24 hr sum, mm)
0.01	2.42	2333	101.6
0.02	2.40	1420	88.9
0.05	2.38	688	81.3
0.10	2.36	369	63.5
0.20	2.34	178	50.8

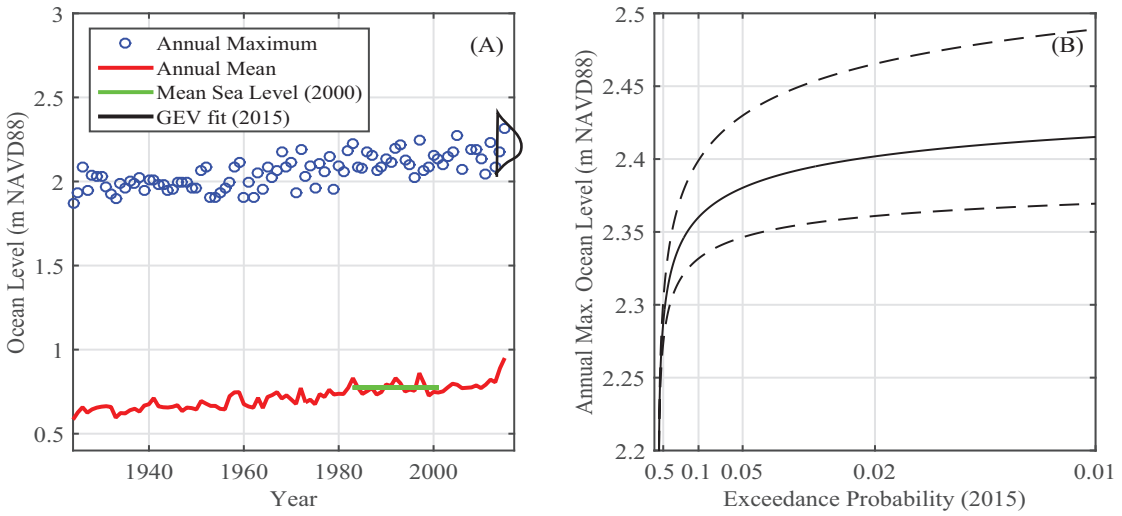

Figure C.2: A) Annual maximum ocean levels compared to annual mean ocean levels recorded at the La Jolla tide gage in CA, USA. The *black* line shows the fitted nonstationary GEV distribution during the year 2015. B) Exceedance probabilities of extreme ocean levels derived from the 2015 GEV distribution. We do not show return periods on the x axis because exceedance probabilities are expected to change as mean sea level increases.

C.1.2 Extreme Ocean Level Frequency Analysis

Extreme ocean levels near the TRV also showed signs of nonstationarity in the historic data record. Figure C.2A shows the annual maximum compared to the annual mean ocean levels recorded at the La Jolla tide gage in CA, US. There is a statistically significant trend in both the annual maximum and mean data at the 0.05 significance level, according to the Man-Kendall trend test for monotonic trends [Mann, 1945, Kendall, 1976]. The persistent trend in ocean levels is not surprising, however it does complicate EVA. In this study, we explicitly modeled the change in extreme ocean levels using a nonstationary, generalized

extreme value (GEV) distribution

$$X \sim \text{GEV}(\mu_t, \sigma, \xi) \tag{C.1}$$

where the random variable X is the annual maximum ocean level, and σ and ξ denote the scale and shape parameter of the GEV distribution, respectively. The time-variant location parameter, μ_t, is formulated as a function of changes in mean sea level

$$\mu_t = \Delta\text{MSL}_o + \mu_o \tag{C.2}$$

where ΔMSL_o is the change in annual mean sea level relative to the mean during the 1983 - 2001 tidal epoch, and μ_o is a constant off-set between the location of the GEV distribution and mean sea level. This model was proposed by Obeysekera and Park [2012] to provide a method for synthesizing extreme value statistics with sea level rise scenarios. We estimated the parameters of the GEV model using Bayesian parameter inference, again with an informative prior on the shape parameter, ξ. The prior on ξ was specified as a normal distribution centered at the La Jolla gage estimate of ξ reported by Zervas [2013a]. Following parameter estimation, exceedance probabilities of extreme ocean levels are estimated as a function of change in mean sea level.

Figure C.2B shows extreme water levels versus exceedance probabilities obtained from the fitted GEV model in the year 2015. Notice that along the x axis, we no longer use return periods to describe the frequency of extreme ocean levels. The common definition of a return period relies on the assumption that exceedance probabilities are time-invariant, which is very unlikely due to anticipated changes in future mean sea level. For our hazard mapping purposes, we used the nonstationary model to estimate the exceedance probabilities of extreme ocean levels associated with present day mean (2015) sea level. The fitted model could also be used to estimate exceedance probabilities associated with future sea levels by using

sea level rise projections to define ΔMSL_o.

C.1.3 Precipitation Frequency Analysis

Precipitation frequency estimates over the Los Laureles and Smuggler's Gulch catchments were obtained from isopluvial maps reported by Sholders [2003]. The isopluvial maps provide 6 hour and 24 hour rainfall depths associated with different return periods. Rainfall depth and frequency estimates were taken from the isopluvial lines nearest to Smuggler's Gulch and Los Laureles catchments. To summarize and conclude the results of our frequency analysis, Table C.2 includes the magnitude (return level) and exceedance probabilities for the three drivers of flooding considered herein. The values in Table C.2 were used as model forcing for the hydrologic and hydraulic modeling.

C.2 Hydrologic and Hydraulic Modeling

In this study, hydrologic modeling was conducted to transform the precipitation totals over the Los Laureles and Smuggler's Gulch catchments (hereafter "the catchments") into flood hydrographs for input to the hydraulic models. Two hydraulic models were developed in this study: one covering the spatial extent of the Los Laureles Catchment, and the other including the Tijuana River Valley.

C.2.1 Hydrologic Modeling

The hydrologic models for the catchments were developed using (1) the Soil Conservation Service (SCS) curve number method to characterize precipitation losses from interception and infiltration [Ponce and Hawkins, 1996], (2) the SCS unit hydrograph method to trans-

form excess precipitation into a hydrograph [NRCS, 1985], and (3) a 24-hour nested storm hyetograph (based on the totals in Table C.2) to define the rainfall distribution within the 24-hour simulation [Sholders, 2003]. The channel flow within the catchments was routed between sub-basins using the kinematic wave model described by [USACE, 2000]. Watershed areas, channel geometries, and basin slopes were estimated from a digital elevation model (DEM) with a 0.76 m (2.5 ft) horizontal resolution, which originated from a 2014 liDar Survey conducted by the County of San Diego. Curve numbers were defined based on land use data from the University of Arizona Remote Sensing Center and literature values from USACE, 2000. Unfortunately, flow measurements within or at the catchment outlets were not available at the time of the study, so hydrologic model calibration was not possible.

C.2.2 Los Laureles Hydraulic Model

Flows in Los Laureles were routed using BreZo [Sanders et al., 2010, Kim et al., 2014], which solves the shallow water equations using a 2D finite volume scheme optimized for applications involving natural topography. BreZo operates on an unstructured grid of triangular or quadrilateral cells, which allows for variable mesh resolution and geometries throughout the modeling domain. The Los Laureles modeling domain covers the entire area of the Los Laureles Watershed (Figure 3.1), with an average cell area of 13.4 m^2. The Los Laureles mesh was generated using Gmsh [Geuzaine and Remacle, 2009] to create a structured, quadrilateral grid along channels and a mixed-mesh of triangular and quadrilateral cells in the floodplain. The structured, quadrilateral portion of the mesh was aligned with trapezoidal channels and small gutters along streets within Los Laureles. We used GPS measurements of channel bank and bottom elevations to define the elevation of mesh nodes aligned with channels. Mesh node elevations within the floodplain were based on the DEM from the 2014 liDAR Survey. Resistance was characterized using spatially-varying Manning's n values, where a value of 0.015 $s/m^{(1/3)}$ was used for concrete surfaces, and 0.035 $s/m^{(1/3)}$ was used for natural areas

of the floodplain. Again, no flow or stage measurements existed within Los Laureles at the time of the study, so the hydraulic model is un-calibrated.

C.2.3 Tijuana River Valley Hydraulic Model

Tijuana River Valley (TRV) flows were also routed using BreZo [Sanders et al., 2010, Kim et al., 2014]. The TRV mesh was generated using Triangle [Shewchuk, 1996], resulting in a triangular mesh of variable resolution throughout the modeling domain. The mesh domain is bounded by the Pacific ocean to the West, Imperial Beach to the North, and the elevated terrain near the US/MX border to the South (Figure 3.1). Mesh edges were aligned with the TJ River channel banks and small levee systems found within the TRV. The resolution of the mesh is highly variable; cells overlapping small channels in the Estuary were assigned an area of 36 m^2, whereas relatively homogeneous regions in the floodplain were assigned a cell area of 100 m^2. Mesh node elevations within the floodplain were also based on the DEM from the 2014 liDAR Survey. Flow resistance was characterized using spatially varying Manning's n values, where the Manning's n value was determined based on land use data. Values ranged from 0.011 $s/m^{(1/3)}$ for the TJ River channel, to 0.1 $s/m^{(1/3)}$ for densely vegetated, riparian areas.

The TRV hydraulic model was validated using observations of water surface elevations in the Estuary and the TJ River at the US/MX border. Observations of Estuary water levels and TJ River stage were obtained from the National Estuarine Research Reserve System and the International Boundary and Water Commission, respectively. In the TJ River channel, comparison of modeled stage to observed stage yielded a root mean square error of 0.25 m for TJ river flow rates of 0 - 1040 m^3/s. The error in modeled water surface elevations is most likely due differences in the sediment level in the TJ channel between the observed and modeled events. To validate the downstream region of the TRV model, we simulated

a 2-week tidal cycle at the ocean boundary and compared modeled water surface elevations to those observed during the same 2-week period. Over the 2 week period, the root mean square error between observed and modeled water surface elevations was 0.07 m. The error in modeled water surface elevations in the Estuary is within the error of the topographic data. Thus, the TRV hydraulic model accurately reproduces water surface elevations for both riverine and tidal forcing.

C.2.4 Hydraulic Model Forcing

Developing hydraulic modeling scenarios appropriate for hazard mapping requires careful consideration. For the TRV hydraulic model, the presence of multiple flood drivers complicates the development of scenarios that represent the exceedance probabilities in Table C.2. This is not an issue for the Los Laureles hydraulic model, since only one driver of flooding was considered in Los Laureles. From a hazard mapping perspective, the challenge in Los Laureles is coupling the hydrologic model with the hydraulic model. Our approaches for addressing both of these issues are outlined in this section.

In this study, the Los Laureles hydraulic model was coupled with the hydrologic model differently before and after end-user focus groups. Prior to the end-user focus groups, the hydrographs generated by the hydrologic model described in Section C.2 were input to the Los Laureles hydraulic model as point-sources of discharge at the sub-watershed outlets within the hydrologic modeling domain. Under this approach, all effective runoff reaches the storm water channels without explicit flow routing in the out-of-bank areas. As discussed in Section 3.6.3, this is not ideal for hazard mapping because only areas susceptible to channel overtopping appear on the hazard map. After end-user focus groups, we added the effective precipitation directly to the 2D modeling grid. Since the 2D modeling grid covers the entire Los Laureles catchment area, there was no need to add fluvial discharges to the modeling

domain using boundary conditions or point-sources. We used the SCS curve number method to estimate the effective precipitation from the rainfall hyetographs, where each 2D model cell was assigned a curve number based on land-use. The effective rainfall hyetographs were added to the 2D grid as spatially distributed source of discharge. Flow was routed for partially wet cells using kinematic wave theory with the friction slope approximated using Manning's equation, whereas flow was routed using the 2D shallow water equations for fully-wetted cells. Explicit routing of overland flow results in the hazard maps similar to those shown in Figure 3.7, which we consider more useful based upon the results of our end-user focus groups.

To address the issue of multiple flood drivers in the TRV, we simulated the extreme conditions of each driver separately with the TRV hydraulic model. It is important to note that if multiple extremes are modeled simultaneously, for example a scenario where an abnormal ocean level coincides with an extreme TJ River flood, the resulting flood hazard would not be associated with the exceedance probabilities of the individual events. Thus, during the TJ River flood simulations, the downstream (ocean) boundary conditions were defined as mean-tidal cycles, and the flows from the catchments were set to zero. To define the the TJ River flow hydrographs, we scaled the hydrograph associated with the 1980 flood to the peak discharges in Table C.2. These scaled hydrographs served as boundary conditions at the upstream boundary of the modeling domain. The resulting TJ river flood hazards predicted by the hydraulic model are associated with the exceedance probabilities defined in Table C.2. The extreme ocean level simulations were developed using the same reasoning. An average, 12-hour tidal cycle was scaled to the extreme ocean levels in Table C.2 to define the ocean boundary of the model. During the extreme ocean level simulations, flows from the catchments and the TJ River were set to zero. Lastly, for the extreme precipitation scenarios, the hydrographs predicted by the hydrologic model associated with the rainfall events in Table C.2 were input as point sources to the hydraulic model at the catchment outlets (Figure 3.1). During these simulations, the ocean boundary conditions were defined as

mean-tidal cycles, while the flows from the TJ River were set to zero. This approach results in an ensemble of hydraulic model output that is a function of 1) exceedance probability and 2) flood drivers. We combined the results of these simulations into hazard maps using probability rules and post-processing techniques.

C.3 Post-Processing Methods

For each hydraulic model simulation, we saved the cell-centered maximum flood depths, unit discharges, depth averaged shear stresses, and durations of depth greater than 0.11 m. These "hazard variables", denoted collectively as H, were processed following simulation to produce the various hazard maps. To create a continuous raster surface from the discrete H values of the hydraulic model cell-centers, we used an inverse distance weighted interpolation scheme. The continuous raster surfaces are the mapped hazard data shown in this study. The Los Laureles hazard maps of H required no further post-processing, since only one driver was considered. However, for the TRV hazard maps, we contoured the maximum value between the three different drivers of flooding

$$\mathbb{H}_i = \max(\{H_i^A,\ H_i^B,\ H_i^C\}) \tag{C.3}$$

where $\mathbb{H}_i$ is the mapped hazard value at raster surface location i, and the superscripts A, B, and C denote H values resulting from extreme ocean level, TJ river flow, and extreme precipitation simulations, respectively. H_i^A, H_i^B, and H_i^C are associated with the same exceedance probabilities when combined in this manner. The TRV hazard maps therefore depict flood hazards with specific exceedance probabilities resulting from either driver of flooding considered, depending on location within the TRV.

The maps contouring the exceedance probabilities of specific flood hazard thresholds required

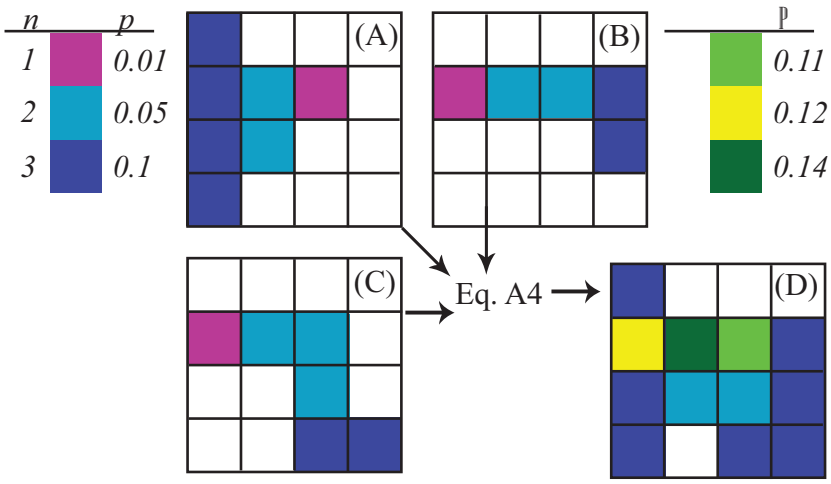

Figure C.3: Illustration of methodology for mapping the exceedance probability of a hazard threshold from multiple flood drivers. A) Raster of P_i^A values for generic flood driver, A. B) Raster of P_i^B values for generic flood driver, B. C) Raster of P_i^C values for generic flood driver, C. D) Exceedance probability raster of $\mathbb{P}_i$ values resulting from flood driver A, B, or C.

additional processing. First, raster surfaces of exceedance probability P_i were created for each flood driver considered. Given a set of hydraulic model output corresponding to n exceedance probabilities, $p_1, p_2, \ldots p_n$, the exceedance probability at raster location i is given by the largest value of p for which the hazard level H_i exceeds a prescribed threshold. To account for the three drivers of flooding in TRV, three probability rasters surfaces were computed: P_i^A, P_i^B and P_i^C, which denote the probability of exceeding the hazard threshold from extreme ocean levels, TJ river floods, and extreme precipitation, respectively. Next, based on the assumption of independence between drivers, the mapped probability is given by

$$\mathbb{P}_i = P_i^A + P_i^B + P_i^C - P_i^A \cdot P_i^B - P_i^A \cdot P_i^C - P_i^B \cdot P_i^C - P_i^A \cdot P_i^B \cdot P_i^C \tag{C.4}$$

where $\mathbb{P}_i$ is the exceedance probability of the hazard threshold at location i resulting from all drivers of flooding considered. In Los Laureles, $\mathbb{P}_i = P_i^C$ since only flooding caused by extreme precipitation was simulated. Notice that Equation C.4 results from probability

addition rules of three independent events, and could be expanded or contracted depending on the number of flood drivers considered. Figure C.3 illustrates this mapping methodology.