AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2016JC012084

Key Points:

- A new approach is proposed for nuisance flood projection using mean sea level
- The approach robustly characterizes the projected uncertainties
- The median nuisance flooding in 2030 will be 400% higher under RCP8.5

Correspondence to:
A. AghaKouchak,
amir.a@uci.edu

Citation:
Vandenberg-Rodes, A., H. R. Moftakhari, A. AghaKouchak, B. Shahbaba, B. F. Sanders, and R. A. Matthew (2016), Projecting nuisance flooding in a warming climate using generalized linear models and Gaussian processes, *J. Geophys. Res. Oceans*, *121*, 8008–8020, doi:10.1002/2016JC012084.

Received 22 JUN 2016
Accepted 12 OCT 2016
Accepted article online 18 OCT 2016
Published online 3 NOV 2016

Projecting nuisance flooding in a warming climate using generalized linear models and Gaussian processes

Alexander Vandenberg-Rodes[1], Hamed R. Moftakhari[2], Amir AghaKouchak[2], Babak Shahbaba[1], Brett F. Sanders[2,3], and Richard A. Matthew[3,4]

[1]Department of Statistics, University of California, Irvine, California, USA, [2]Department of Civil and Environmental Engineering, University of California, Irvine, California, USA, [3]Department of Planning, Policy, and Design, University of California, Irvine, California, USA, [4]Blum Center for Poverty Alleviation, University of California, Irvine, California, USA

Abstract Nuisance flooding corresponds to minor and frequent flood events that have significant socioeconomic and public health impacts on coastal communities. Yearly averaged local mean sea level can be used as proxy to statistically predict the impacts of sea level rise (SLR) on the frequency of nuisance floods (NFs). In this study, we use generalized linear models (GLM) and Gaussian Process (GP) models combined to (i) estimate the frequency of NF associated with the change in mean sea level, and (ii) quantify the associated uncertainties via a novel and statistically robust approach. We calibrate our models to the water level data from 18 tide gauges along the coasts of United States, and after validation, we estimate the frequency of NF associated with the SLR projections in year 2030 (under RCPs 2.6 and 8.5), along with their 90% bands, at each gauge. The historical NF-SLR data are very noisy, and show large changes in variability (heteroscedasticity) with SLR. Prior models in the literature do not properly account for the observed heteroscedasticity, and thus their projected uncertainties are highly suspect. Among the models used in this study, the Negative Binomial Distribution GLM with GP best characterizes the uncertainties associated with NF estimates; on validation data ≈93% of the points fall within the 90% credible limit, showing our approach to be a robust model for uncertainty quantification.

1. Introduction

There is strong evidence for global sea level rise (SLR) over past decades both from in situ tide gauge measurements and satellite altimetry [*Church et al.*, 2013; *Church and White*, 2011; *Church et al.*, 2011; *Church and White*, 2006; *Stocker*, 2014; *Domingues et al.*, 2008; *Calafat and Chambers*, 2013; *Watson et al.*, 2015; *Cazenave et al.*, 2014]. The average rate of SLR has increased from 1.2 ± 0.2 mm/yr in 1901–1990 to 3.0 ± 0.7 mm/yr in 1993–2010 [*Hay et al.*, 2015]. The rate over the past decades is an order of magnitude larger than the last millennia and there is a globally averaged acceleration of 0.022 ± 0.015 mm/yr^2 (between 1952 and 2011) [*Calafat and Chambers*, 2013]. Based on current trajectories of anthropogenic activities and greenhouse gas emissions [*Lyu et al.*, 2014], projections of SLR over the 21st century cannot rule out an increase greater than 1 m [*Milne et al.*, 2009; *Rahmstorf*, 2007; *Nicholls and Cazenave*, 2010; *Kopp et al.*, 2014]. However, there is considerable debate around the significance of the suggested rate of acceleration. Accordingly, it has been suggested that for a reliable detection of SLR acceleration rate, one needs to apply differently characterized methods and to remove interannual to multidecadal variability from the observed signal [*Haigh et al.*, 2014; *Visser et al.*, 2015].

This rise in sea level reduces the freeboard between high tidal datum and flood stage, leading to an increase in the risk that extreme water levels overtop flood defenses. By 2100, it is expected that 0.2–4.6% of the global population will be flooded annually [*Hinkel et al.*, 2014] and the global cost of flooding could reach US$210 billion per year, if no adaptation measures are implemented [*Hinkel et al.*, 2013]. Eight out of the top twenty most vulnerable cities of the world are in the United States [*Hallegatte et al.*, 2013], and over half of the population of the United States resides in coastal regions [*Scavia et al.*, 2002]. Without further flood management measures, New York City alone is projected to experience a US$174 million per year loss due to flooding [*Aerts et al.*, 2014].

Nuisance flooding is a relatively new term that describes nondestructive flooding. Nonetheless, it is capable of causing substantial negative socioeconomic impacts [*Gornitz et al.*, 2001], compromising infrastructure such as surface transportation [*Suarez et al.*, 2005] and sewer systems [*Cherqui et al.*, 2015; *Flood and*

Cahoon, 2011], and posing public health risks [*Ten Veldhuis et al*., 2010]. Previous studies on socioeconomic impacts of future coastal flooding have primarily focused on extreme flood heights and therefore the least frequent flood events [*Muis et al*., 2016; *Reed et al*., 2015; *Woodruff et al*., 2013]. *Sweet and Park* [2014] estimated the projections of annual exceedance above local nuisance levels at U.S. tide gauges by shifting probability estimates of daily maximum water levels over a contemporary 5 year period following probabilistic relative SLR projections of *Kopp et al*. [2014] for representative concentration pathways (RCP) 2.6, 4.5, and 8.5. *Sweet and Park* [2014] concluded that the majority of locations are expected to experience even more NF as "tipping points" for inundation will be surpassed by 2050 under the local median SLR projections. *Moftakhari et al*. [2015] described the relationship between the observed local annually averaged mean sea level and NF through nonlinear regression analysis since 1920 in 12 tide gauges located along the coasts of United States. They use Monte-Carlo analysis to project the 90% confidence interval of NF following near-(2030) and mid-term (2050) future projections of SLR, under RCPs 2.6 and 8.5 [*Kopp et al*., 2014].

Moftakhari et al. [2015] assumed that the relationship between mean sea level (MSL) and nuisance flooding (NF) follows a nonlinear trend of the form:

$$NF = \alpha + \beta \cdot (MSL)^{\theta}, \tag{1}$$

where the parameters (α, β, θ) are estimated by nonlinear least squares.

It is well known that nonlinear least squares can provide inefficient estimates for count-valued data, especially for low-valued counts [*Cameron and Trivedi*, 2013]. Since nuisance flooding is a more recent phenomenon, a significant fraction of the historical NF data consists of low-valued counts.

Furthermore, reported standard errors under nonlinear least squares are inconsistent without post hoc correction [*White*, 1981]. This can be seen through the *residuals*

$$\hat{\epsilon}_i := NF_i - \hat{\alpha} - \hat{\beta} \cdot (MSL_i)^{\hat{\theta}}, \tag{2}$$

where $\hat{\alpha}, \hat{\beta}, \hat{\theta}$ are the estimated parameters in (1). Under standard regression assumptions, the distribution of $\hat{\epsilon}_1, \ldots \hat{\epsilon}_N$ should be independent, mean zero, and with constant variance (*homoscedasticity*). However, the observed residuals in the fitted MSL-NF relationship appear to have variances that strongly depend on the MSL predictor (see section 2.2.4); over the last few decades, both the amount of sea level rise and the year-to-year *variability* or *dispersion* of nuisance flooding about its trend has increased substantially, which can be partially attributed to the multidecadal variations in extreme sea levels relative to changes in mean sea level [*Wahl and Chambers*, 2015]. Thus, even if we assume validity of the specific nonlinear relationship (1), this presence of significant *heteroscedasticity* can result in inaccurate projections for future NF.

Hereby we propose a more robust approach that overcomes the shortcomings of previous studies. We will use the framework of generalized linear models (GLMs) [*McCullagh and Nelder*, 1989] to provide a principled statistical approach for both count data and heteroscedasticity. Straightforward implementations of the GLM framework, however, do not effectively model the apparent heteroscedasticity in the NF data. Our main contribution in this paper is an extension of GLMs, where we model the dispersion of the NF counts nonparametrically using Gaussian processes (GPs). Similar approaches have been proposed for nonparametric regression models [*Goldberg et al*., 1997; *Wang and Neal*, 2012], but have not been proposed in the context of GLMs. We will show that it provides an effective model for the observed variability in nuisance flooding, and gives better projections on held out (validation) data. Importantly, such models are now easily implemented using standard Bayesian software (the Stan package [*Carpenter et al*., 2015]), broadening their appeal beyond a few statisticians and machine learning researchers, and could be an effective model for noisy environmental data in general.

2. Methods and Materials

2.1. Tide Gauge Data and SLR Projections

We use three sets of data here to calibrate our model and project NF in the future. Hourly WL data for tide gauges used in this study (Table 1 and Figure 1) are provided by National Oceanic and Atmospheric Association (NOAA). The yearly averaged WL (available on http://www.psmsl.org/[*Holgate et al*., 2012]) is then subtracted by the long-term mean (e.g., over the whole record) to produce the time series of mean sea level

Table 1. Tide Data Description

	NOAA ID	Latitude	Longitude	Start Year	Completeness (%)
Boston, MA (BMA)	8443970	42.353	−71.053	1921	99
The Battery, NY (BNY)	8518750	40.700	−74.013	1920	99
Sandy Hook, NJ (SNJ)	8531680	40.467	−74.008	1932	98
Atlantic City, NJ (ANJ)	8534720	39.355	−74.418	1920	92
Philadelphia, PA (PPA)	8545240	39.933	−75.142	1920	97
Baltimore, MD (BMD)	8574680	39.267	−76.578	1920	99
Annapolis, MD (AMD)	8575512	38.983	−76.480	1928	95
Washington, DC (WDC)	8594900	38.873	−77.022	1931	98
Sewells Point, VA (SVA)	8638610	36.947	−76.330	1927	99
Wilmington, NC (WNC)	8658120	34.227	−77.953	1935	98
Charleston, SC (CSC)	8665530	32.782	−79.925	1921	99
Fort Pulaski, GA (FGA)	8670870	32.033	−80.902	1935	99
Key West, FL (KFL)	8724580	24.555	−81.807	1920	99
Port Isabel, TX (PTX)	8779770	26.060	−97.215	1944	96
La Jolla, CA (LCA)	9410230	32.867	−117.257	1924	93
San Francisco, CA (SCA)	9414290	37.807	−122.465	1920	99
Seattle, WA (SWA)	9447130	47.602	−122.338	1920	99
Honolulu, HI (HHI)	1612340	21.307	−157.867	1920	99

anomaly (ΔMSL) relative to land, that reflects a combination of WL changes and vertical land motions [*Wöppelmann and Marcos*, 2015; *Nicholls et al.*, 2014]. The cumulative hours that observed WL exceeds a certain threshold, in each meteorological year (MY; May–April), represents the NF at any given tide gauge. This NF threshold (listed in *Sweet and Park* [2014]), determined by local Weather Forecasting Offices of NOAA's National Weather Service, has been historically associated with minor flooding impacts. It is worth mentioning that not necessarily every WL exceedance above this threshold results in noticeable coastal flooding [*Sweet and Park*, 2014]. *Sweet and Park* [2014] provided NF records prior-MY 2012 (May 2012 to April 2013) and *Moftakhari et al.* [2015] extended the records to MY 2013. The probability distribution of local SLR projections under CMIP5 (RCP 2.6 and RCP 8.5 scenarios) are provided by *Kopp et al.* [2014]. We use quantiles 0.05 and 0.95 as representatives of 90% confidence limits of projected SLR for year 2030.

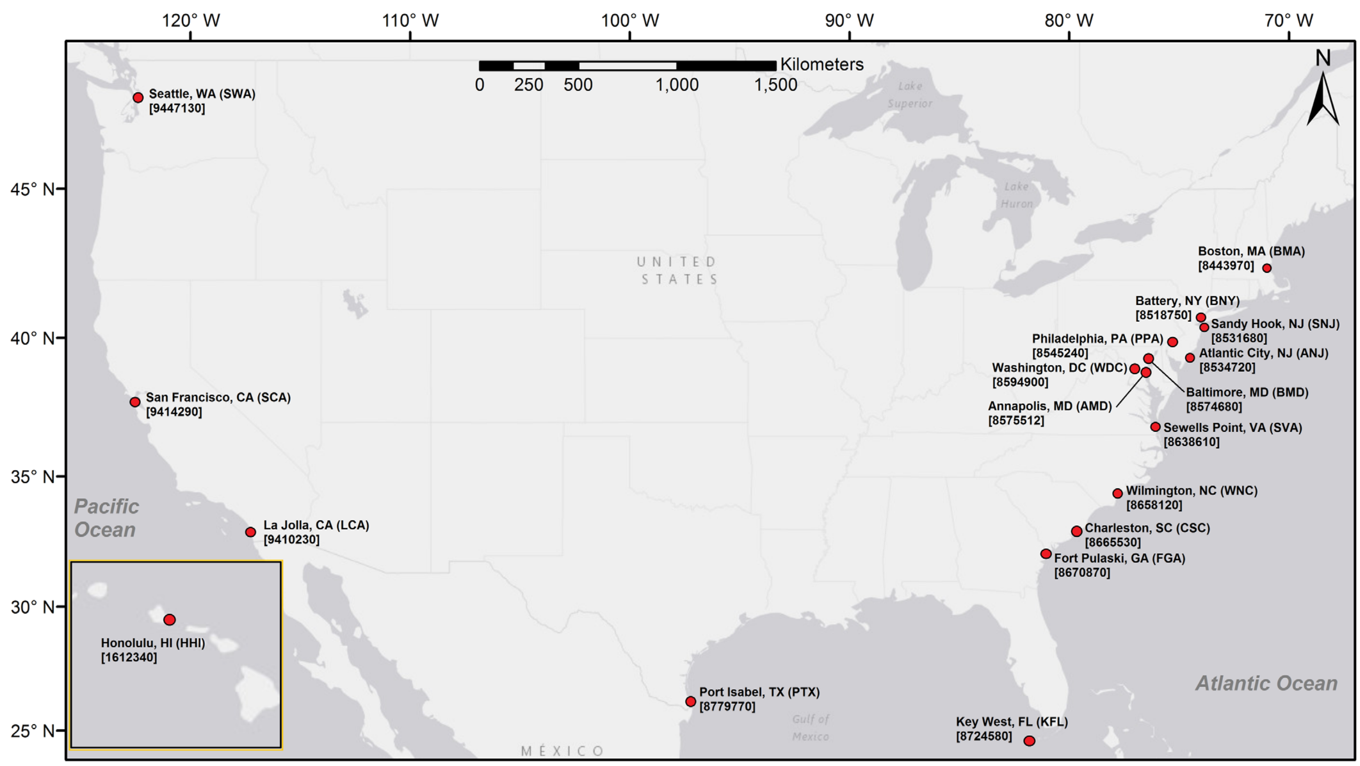

Figure 1. Station locations.

2.2. Methodology

2.2.1. Statistical Preliminaries

We let $\mathcal{D}$ denote our data set of N yearly observations of the MSL and NF variables:

$$\mathcal{D} = \{(NF_1, MSL_1), \ldots, (NF_N, MSL_N)\}. \tag{3}$$

Within the statistical paradigm, these data are modeled a priori by a probability density function $p(\cdot)$, which typically depends on additional parameters. In this study, we model each NF observation as depending only on the observed MSL variable for the same year, that is, we assume the conditional independence

$$\begin{aligned} &p(NF_1, \ldots, NF_N \mid MSL_1, \ldots, MSL_N) \\ &\quad = p(NF_1 \mid MSL_1) \cdots p(NF_N \mid MSL_N), \end{aligned} \tag{4}$$

where $p(NF_j \mid MSL_j)$ is the conditional probability density of the NF variable given the mean sea level MSL_j, evaluated at NF_j.

Note that the particular probability density $p(\cdot)$ is inferred from its argument. The reader is here forewarned that this can be confusing in the context of Bayesian inference (section 2.2.3), where parameters are also random variables. The notation makes no distinction between a possible value of random variable and the random variable itself. That is, $p(\alpha)$ denotes the probability density function of the parameter α, evaluated at a particular value α, with the particular density function $p(\cdot)$ (e.g., normal and Poisson) usually specified in the text.

We will use the standard notations $\mathrm{E}[X]$, $\mathrm{Var}[X]$, and $\mathrm{Cov}[X, Y]$ to refer to the mean, variance, and covariance of random variables X and Y.

2.2.2. Generalized Linear Models

In the generalized linear model framework, we specify the observation variable NF to have an *exponential family* distribution. In the context of NF, the observations are count data—the number of hours that the observed water level lies above a certain threshold—so we will focus on the Poisson and Negative Binomial Distributions, which are supported on the nonnegative integers. We then specify the estimated mean μ as a log-linear function of the MSL variable:

$$\log \mu = \alpha + \beta \cdot MSL, \tag{5}$$

where α and β are unknown parameters. The nonlinear regression approach minimizes the sum of squared errors $\sum_{i=1}^{N} (NF_i - \mu_i)^2$, where $\mu_i = \exp(\alpha + \beta \cdot MSL_i)$. To account for the nature of the observations (e.g., Poisson), a common *frequentist* estimate of the parameters α and β is obtained by maximizing the *log likelihood* of the data, which by (4) is

$$\begin{aligned} &\log p(NF_1, \ldots, NF_N \mid \alpha, \beta, MSL_1, \ldots, MSL_N) \\ &\quad = \sum_{i=1}^{N} \log p(NF_i \mid \alpha, \beta, MSL_i). \end{aligned} \tag{6}$$

If $p(NF_j \mid \alpha, \beta, MSL_j)$ are assumed to be Gaussian with constant variance, maximizing (6) is identical to minimizing the sum of squared errors in nonlinear regression.

As discussed in the section 1, a constant variance assumption (homoscedasticity) is not tenable. Now the Poisson distribution has the variance equal to the mean, but this is perhaps still not representative of the data. In fact, the residuals are often in the order of the mean, instead of the square-root of the mean (Figure 2). To better model such overdispersed NF data, we specify the observation variable NF to be a Negative Binomial random variable, parameterized by its mean μ and overdispersion $1/\phi$, such that

$$\mathrm{Var}\, NF = \mu + \frac{\mu^2}{\phi}. \tag{7}$$

In particular, the log likelihood of the data would be

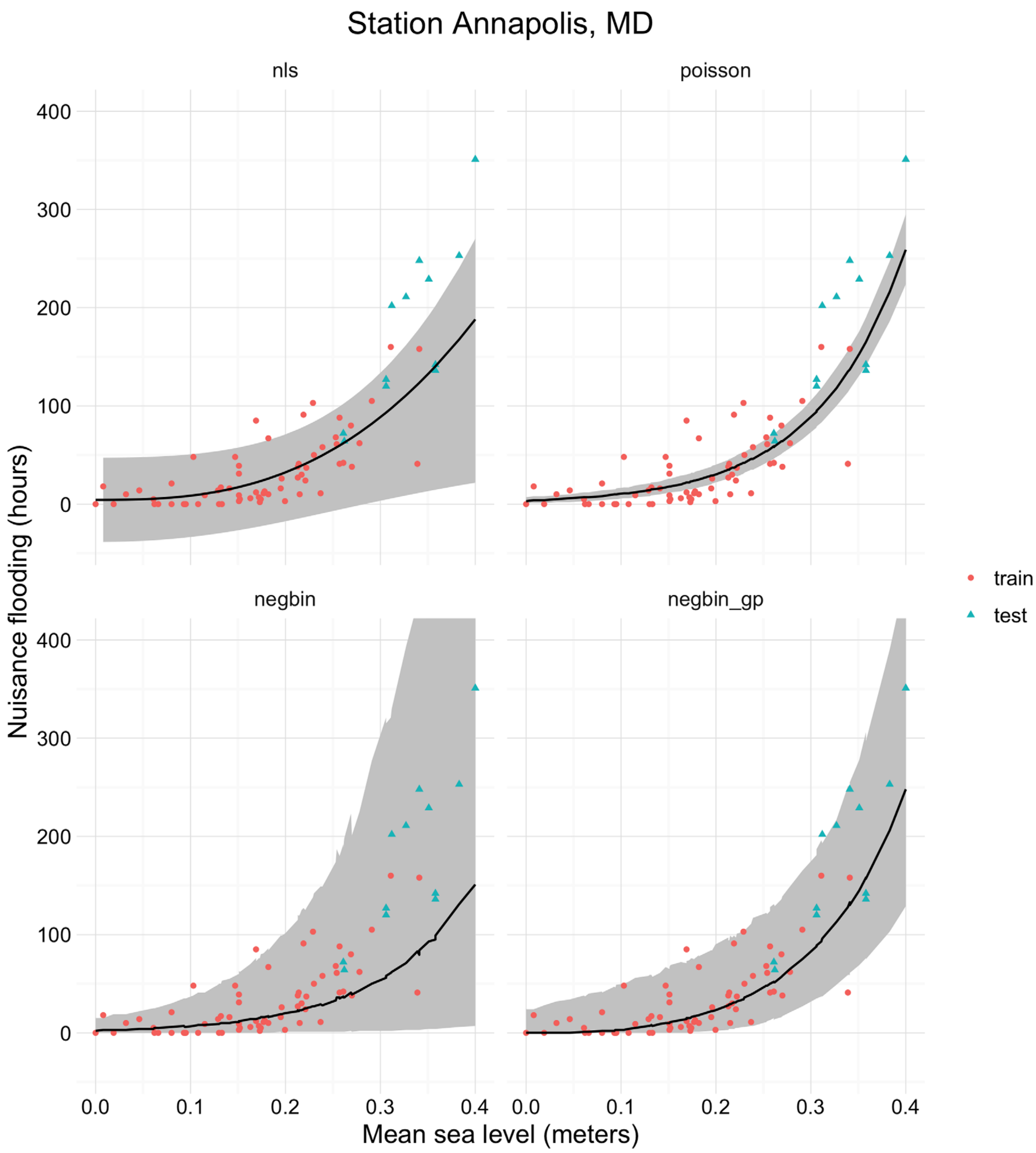

Figure 2. Predicted nuisance flooding given mean sea level for data observed in Annapolis, MD (AMD). Circles: training data (years 1929–2000), blue triangles: validation data (years 2001–2013). Line: predicted NF medians, shading: 90% confidence interval (nls) and credible intervals (Poisson, NegBin, NegBinGP models).

$$\begin{aligned}&\sum_{i=1}^{N}\log p(NF_i \mid \alpha,\beta,\phi,MSL_i)\\ &=\log\prod_{i=1}^{N}\frac{\Gamma(\phi+NF_i)}{(NF_i)!\,\Gamma(\phi)}\left(\frac{\phi}{\phi+\mu_i}\right)^{\phi}\left(\frac{\mu_i}{\phi+\mu_i}\right)^{NF_i}.\end{aligned} \tag{8}$$

Given the observed data, we could maximize the log likelihood function based on (8) to estimate the parameters α, β, and ϕ.

We remark that the Poisson distribution can be obtained as a limit of the Negative Binomial as $\phi\to\infty$.

2.2.3. Bayesian Inference

Alternatively, we can take the Bayesian modeling approach [*Gelman et al*., 2014], which has at least two advantages in this situation. First, this will allow us to more flexibly model the dispersion parameter ϕ of the above Negative Binomial model using Gaussian Processes, as we will do in the next section. Second, the Bayesian approach easily allows us to jointly account for the uncertainty of model parameters and projections of future MSL.

With a Bayesian approach, we explicitly model the parameters α, β, and ϕ as random variables a priori, and then use Bayes' rule to obtain the posterior distribution of the parameters conditional on the observed data $\mathcal{D}$: (recall (3))

$$p(\alpha,\beta,\phi\,|\,\mathcal{D}) \propto \prod_{i=1}^{N}\left[p(NF_i\,|\,\alpha,\beta,\phi,MSL_i)\right]\cdot p(\alpha,\beta,\phi). \tag{9}$$

For the choice of prior distribution $p(\alpha,\beta,\phi)$, we will assume independence among the parameters, thus $p(\alpha,\beta,\phi)=p(\alpha)p(\beta)p(\phi)$. Note that given the observed data, the parameters are dependent a posteriori. Here we place a lognormal prior on the dispersion parameter ϕ, and normal $(0,100^2)$ priors on the α and β parameters.

Given a future mean sea level projection MSL_{N+1}, the Bayesian framework integrates over the posterior parameter distribution to obtain the posterior distribution of future nuisance flooding:

$$\begin{gathered} p(NF_{N+1}|\mathcal{D},MSL_{N+1})= \\ \int p(NF_{N+1}|\,\alpha,\beta,\phi,MSL_{N+1})p(\alpha,\beta,\phi\,|\,\mathcal{D})\,d\alpha\ \ d\beta\ \ d\phi. \end{gathered} \tag{10}$$

For a distribution $p(MSL_{N+1})$ instead of just a point estimate, the Bayesian framework easily handles this by integrating over this uncertainty as:

$$\begin{aligned} p(NF_{N+1}|\mathcal{D})=\int & p(NF_{N+1}|\,\alpha,\beta,\phi,MSL_{N+1}) \\ & \cdot p(\alpha,\beta,\phi\,|\,\mathcal{D})p(MSL_{N+1})\,d\alpha\ \ d\beta\ \ d\phi\ \ dMSL_{N+1}. \end{aligned} \tag{11}$$

Typically, the above integrals do not have a closed form representation, so we will approximate them using Markov Chain Monte-Carlo simulation.

2.2.4. Gaussian Process Modeling of Heteroscedasticity

Recall that the Negative Binomial model assumes the specific mean-variance relationship (7), with the parameter ϕ controlling the amount of overdispersion compared to the Poisson. However, the apparent mean-variance relationship of most stations is not implied by either the Poisson or the Negative Binomial law. Figure 2 shows the predicted mean and dispersion for the Annapolis, MD, training data, under the nonlinear regression model, the Poisson and Negative Binomial models. Note that the Poisson model provides extremely overconfident projections, while the Negative Binomial model (and nonlinear least squares) is too conservative with larger MSL. The last plot shows our proposed GP heteroscedasticity model—described below—which appears to give a good account of the observed dispersion.

In order to more flexibly model the heteroscedasticity present in the NF data, we propose to replace the dispersion parameter ϕ in the Negative Binomial model with a smoothly varying function $\phi(\cdot)$ of the MSL input. We eschew parametric models for $\phi(\cdot)$, instead choosing a classic Bayesian nonparametric model: the (log-)Gaussian Process (GP) [*Rasmussen and Williams*, 2006; *Gelman et al*., 2014]. We recall that a GP is a distribution on functions $f(\cdot)$, and, as with multivariate Gaussians, a GP is uniquely defined by its mean and covariance. The covariance of a GP is a function $K(\cdot,\cdot)$ of two variables, with $K(x,x')=\mathrm{Cov}[f(x),f(x')]$. Furthermore, the specification is consistent, in the sense that the distribution of any finite set of values $f(x_1)$, $\ldots,f(x_n)$ is multivariate Gaussian with covariance matrix $\{K(x_j,x_k)\}_{j,k=1}^{n}$.

In this work, we specified $\log\phi(\cdot)$ to be a GP, with

$$\phi(x)=\exp\left\{a+bx+\tau\int_0^x B(z)\,dz\right\}. \tag{12}$$

Here $B(\cdot)$ is Brownian motion, x is the MSL input, and a, b, and τ are parameters. The stochastic process defining $\log \phi(\cdot)$ in (12) is often called (one-fold) integrated Brownian motion (IBM) [*Wood and Kohn*, 1998; *Wahba*, 1990], which is a Gaussian process with a linearly varying mean

$$\mathrm{E}[\log \phi(x)] = a + bx, \tag{13}$$

and covariance structure (kernel):

$$\begin{aligned} K_\tau(x, x') &:= \mathrm{Cov}[\log \phi(x), \log \phi(x')] \\ &= \frac{\tau^2}{2}|x - x'| \min(x, x')^2 + \frac{\tau^2}{3} \min(x, x')^3. \end{aligned} \tag{14}$$

IBM has the following advantages compared with other commonly used kernels such as the squared exponential kernel:

1. With an (improper) flat prior on a and b, $p(a, b) \propto 1$, The IBM prior can be interpreted as the Bayesian analog of the cubic smoothing spline [*Kimeldorf and Wahba*, 1970].
2. The squared exponential and many other kernels make the strong assumption of stationarity, which can unduly constrain the posterior of $\phi(x)$ in regions with sparse data, especially when extrapolating past the available data. The IBM prior is nonstationary, and accounts for increasing uncertainty under extrapolation. This is particularly important in the context of modeling the NF-MSL relationship, as the evolution of tidal components [*Jay*, 2009; *Ray*, 2006, 2009; *Flick et al*., 2003; *Müller et al*., 2011] also contributes additional uncertainty which is influenced in part by sea level rise [*Devlin et al*., 2014]. Additional variability due to changes in the frequency of storm surges over time [*Marcos et al*., 2015; *Thompson et al*., 2013] is also absorbed into this IBM component. Of course, the IBM model assumes that such changes have already impacted the historical data and vary smoothly in response to MSL rise.

We placed a broad lognormal prior on the scale parameter τ, along with the flat priors on a and b.

3. Results

We used the data described in section 2.1 to compare the performance of the following four models: nonlinear least squares (nls), Poisson GLM (Poisson), Negative Binomial GLM with (NegBinGP), and Negative Binomial GLM without (NegBin) the GP heteroscedasticity model. We recall that the nls model assumes (1) for the mean, while the latter three models assume equation (8).

3.1. Predictive Performance

For each station, we used the data from all years $t \leq 2000$ as our training (calibration) set, and then obtained posterior medians NF_t^{median} along with 5% and 95% posterior quantiles for the nuisance flooding in years 2001–2013, given the MSL estimate.

3.1.1. Mean Absolute Error

We first used mean absolute error as our metric to compare the performance of the four models:

$$\frac{1}{13} \sum_{t=2001}^{2013} |NF_t^{median} - NF_t|. \tag{15}$$

Table 2 contains the results. With the exception of Philadelphia, PA, the NegBin model is the worst performer by far. The remaining three models show similar validation set error, with relative differences not more than 20%, and often under 10%, except for Philadelphia, PA, San Francisco, CA, and Washington, DC. For Port Isabel, TX, Honolulu, HI, Sewell's Point, VA, and Washington, DC, the nonlinear least squares algorithm failed to converge from the initial parameters after 20 random initializations.

3.1.2. Coverage

We are also interested in the uncertainty quantification of different models proposed in the methodology section. Table 3 shows the coverage of each model, that is, the percentage of validation data across all 18 stations lying between the 5% and 95% posterior predictive intervals.

In order to make a fair comparison, for the nls method, we calculated the 90% confidence interval using a second-order Taylor method (R package propagate [*Spiess*, 2014]). As seen in Figure 2, such confidence intervals do not take into account the nonnegativity of the data; indeed the estimated 5% quantile is

Table 2. Mean Absolute Errors for All 18 Stations

	AMD	ANJ	BMA	BMD	BNY	CSC	FGA	HHI	KFL
nls	62.6	13.1	4.9	26.2	13.2	8.9	8.4	NaN	4.9
Poisson	53.2	11.9	3.8	26.2	12.4	8.9	7.9	17.7	4.8
NegBin	96.4	15.5	4.0	37.2	15.4	11.9	8.0	19.6	5.2
NegBinGP	57.7	13.1	4.8	27.9	14.0	9.3	7.9	18.8	5.2

	LCA	PPA	PTX	SCA	SNJ	SVA	SWA	WDC	WNC
nls	5.5	13.0	NaN	5.0	14.7	NaN	3.3	NaN	32.4
Poisson	5.8	9.9	44.4	5.5	12.8	17.8	3.1	28.5	30.6
NegBin	5.8	8.1	77.7	7.6	14.3	18.5	3.3	41.2	38.6
NegBinGP	5.9	10.4	51.6	7.6	13.1	15.9	3.2	34.1	36.2

typically negative. Since the nls has less than 75% coverage, this implies that at least a quarter of the test points lie above the estimated 95% quantile.

Note that the NegBinGP model provides a reasonably accurate uncertainty prediction ($\approx$ 93%), which is closest to the ideal of 90%. The original NegBin model is overly conservative ($\approx$ 98% of points lie in 90% interval).

3.1.3. Leave-One-Out Predictive Density

A more powerful test of out-of-sample predictive accuracy in the Bayesian setting is the Leave-one-out cross-validation estimate of the out of sample log predictive density (lpd):

$$\mathrm{lpd}_{loo}=\sum_{i=1}^{N}\log p(Y_i|Y_{-i}), \tag{16}$$

where

$$p(Y_i|Y_{-i})=\int p(Y_i|\theta)p(\theta|Y_{-i})d\theta, \tag{17}$$

is the posterior probability density of the observation Y_i, given Y_{-i} – the data set with the ith observation removed.

While lpd_{loo} is extremely computationally intensive to compute, approximations such as the Watanabe-Aikake Information Criterion (WAIC) and Pareto-smoothed importance sampling of lpd_{loo} (PSIS-LOO) have been shown to work very well for low-dimensional models such as ours [*Vehtari et al.*, 2016]. Using the PSIS-LOO criterion, the approximate difference between the two Negative Binomial models (with and without GP heteroscedasticity), across all stations, was

$$\mathrm{lpd}_{loo}^{NegbinGP}-\mathrm{lpd}_{loo}^{Negbin}\approx 111.5, \tag{18}$$

with a standard error of 15.6, vastly preferring the GP heteroscedasticity model. Individually, only two stations (La Jolla and San Francisco) preferred the non-GP model, although these differences in log probability (-1.3 and -0.6, respectively) were within their individual standard errors (1.4 and 1.6).

3.2. Projections

Given its good performance in calibration, we used the Negative Binomial with GP heteroscedasticity model to give uncertainty projections for future nuisance flooding at year 2030, using the entire historical data set as training data. In order to jointly model the uncertainties of the SLR projections along with the NF-SLR response, we took a fully Bayesian approach as described in section 2.2.3. In particular, for each station we modeled the underlying Gaussian Process-based dispersion $\phi(\cdot)$ on a set of points $\vec{x}$ containing both the historical MSL values and a fine grid of points that overlay the sea level rise (SLR) projections. We drew 10, 000 Hamiltonian Monte-Carlo samples from the posterior distribution of the parameters $(\phi(\vec{x}), a, b, \tau, \alpha, \beta)$, and for each sample evaluated, the projected nuisance flooding at a sample from the future distribution of MSL x_{2030} for the year 2030. We used linear interpolation of the given SLR quantiles to form this approximate future MSL distribution. Table 4 records the resulting NF posterior prediction quantiles due to projected SLR under the RCP2.6 and RCP8.5 scenarios.

Table 3. Coverage Percentages on Validation Data

nls	Poisson	NegBin	NegBinGP
74.3	46.2	97.8	92.9

Table 4. Posterior Predictive Quantiles for NF in Year 2030 Under the NegBinGP Model

	2.6					8.5						
RCP	5%	17%	50%	83%	95%	5%	17%	50%	83%	95%	Highest Experienced	Average 2000–2014
AMD	193	338	735	1591	2722	190	342	755	1679	3159	351	180
ANJ	40	83	186	401	731	42	84	190	432	749	96	40
BMA	4	9	23	51	100	4	10	25	59	112	21	8
BMD	27	59	135	280	469	33	64	145	307	534	111	44
BNY	6	18	48	108	188	6	17	52	140	288	61	24
CSC	91	148	302	578	961	84	151	318	678	1250	61	33
FGA	55	96	185	366	618	54	87	180	399	751	45	20
HHI	86	210	641	1948	4796	87	218	646	2081	5982	119	31
KFL	2	8	34	138	437	2	8	38	166	559	20	5
LCA	3	17	62	258	846	4	17	60	250	711	23	9
PPA	8	20	52	124	226	9	23	59	129	251	69	20
PTX	303	714	2401	8539	23184	349	815	2602	9113	23837	201	81
SCA	10	26	78	206	442	10	27	78	199	427	64	15
SNJ	37	67	150	318	567	30	67	170	434	843	85	38
SVA	8	22	57	121	206	8	23	60	135	231	95	25
SWA	0	1	6	20	46	0	1	6	18	44	14	3
WDC	41	82	178	346	585	41	85	190	372	618	210	85
WNC	93	251	825	2908	8961	79	227	910	4286	14317	127	57

Figure 3 also graphs the resulting predicted mean and 90% confidence intervals for all 18 stations using the entire historical data, along with the (intermediate) RCP4.5 estimates for MSL in year 2030.

The results suggest that the median of projected NF in 2030 will be on-average $\approx$ 400% higher under RCP 8.5, compared with the recent (2000–2014) experienced NF along the coasts of United States. Among all the studied gauges, Fort Pulaski, GA, and Charleston, SC with $\approx$ 860%, and Seattle, WA, and Washington, DC, with $\approx$ 105% increased NF are expected to have the highest and lowest predicted median of NF in 2030, respectively. We caution (see discussion below) that these NF projections assume a log-linear NF-MSL relationship and are highly suspect when extrapolating far away from the historical data. We feel that this is particularly the case for Atlantic City, NJ, Charleston, SC, Fort Pulaski, GA, Honolulu, HI, Key West, FL, Port Isabel, TX, and Wilmington, NC, for which the 90% interval for MSL at year 2030 does not contain any historical MSL observations.

4. Discussion

We have used the generalized linear model framework to effectively deal with the count nature of the data, along with nonparametric modeling of the heteroscedasticity, quantifying the variability of nuisance flooding as a function of the MSL predictor. Although our model proves to be effective for the available historical data, caution is certainly warrented when extrapolating to future years, where MSL is projected to rise significantly beyond the available data on which we trained the model. Even with our better accounting of the variability of nuisance flooding, we emphasize that future projections depend heavily on the hypothesized form of the NF-MSL relationship, as well as the MSL projections themselves. In *Moftakhari et al*. [2015], this relationship was assumed to follow a power function $NF=\alpha+\beta \cdot MSL^{\theta}$, while in this study, the natural choice for GLM's with count data is the log-linear relationship $NF=\exp(\alpha+\beta \cdot MSL)$. Even for stations like Annapolis, MD, and Charleston, SC, which exhibit a relatively strong relationship between MSL and NF, the historical data appear insufficient to choose between these two functional forms.

The results (Figure 4 and Table 4) show a spatial pattern in the rate of increase in NF. Looking at the regions identified by *Wahl and Chambers* [2016], we can see that the gauges that fall in the same region show somehow the same expected rate of increase. The gauges in North Atlantic and Mid-Atlantic are expected to experience $\approx$ 220$\pm$100% higher NF in 2030, while Charleston, SC, and Fort Pulaski, GA, that are located in southern Atlantic are expected to experience $\approx$ 860$\pm$10% more NF by then. The decreasing pattern in NF can be detected from south to north along the West Coast of United States. La Jolla, CA, located in the south of South Pacific region is expected to experience $\approx$780% more NF by 2030, while San Francisco, CA, and Seattle, WA, located in the north of the South Pacific and North Pacific regions are expected to experience $\approx$430% and $\approx$100% increased NF, respectively. The components like climate-driven dynamic

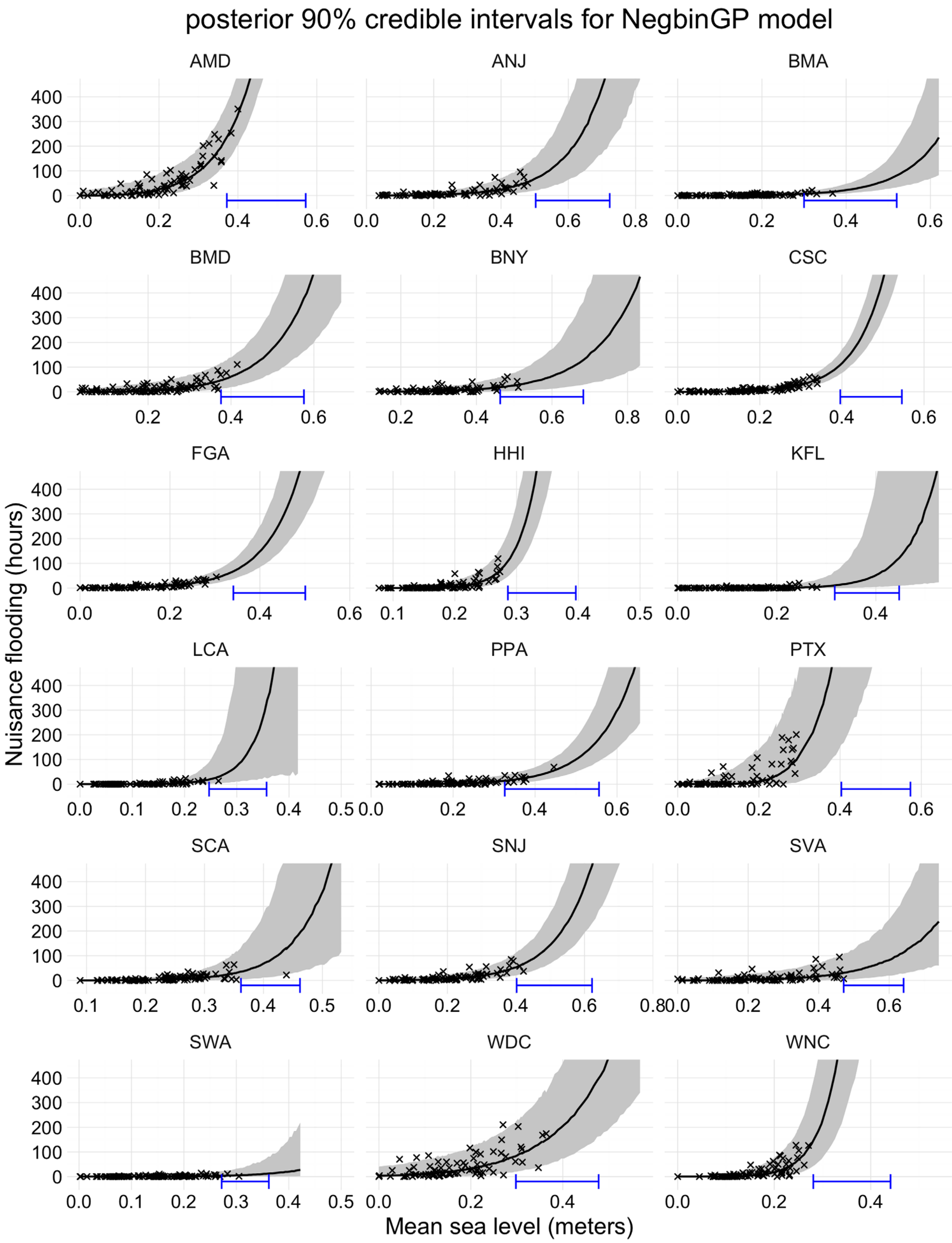

Figure 3. Posterior predictive median and shaded 90% intervals for all 18 stations with the NegBinGP model, overlaid with Kopp et al.'s 90% confidence interval for MSL in year 2030 (blue errorbars).

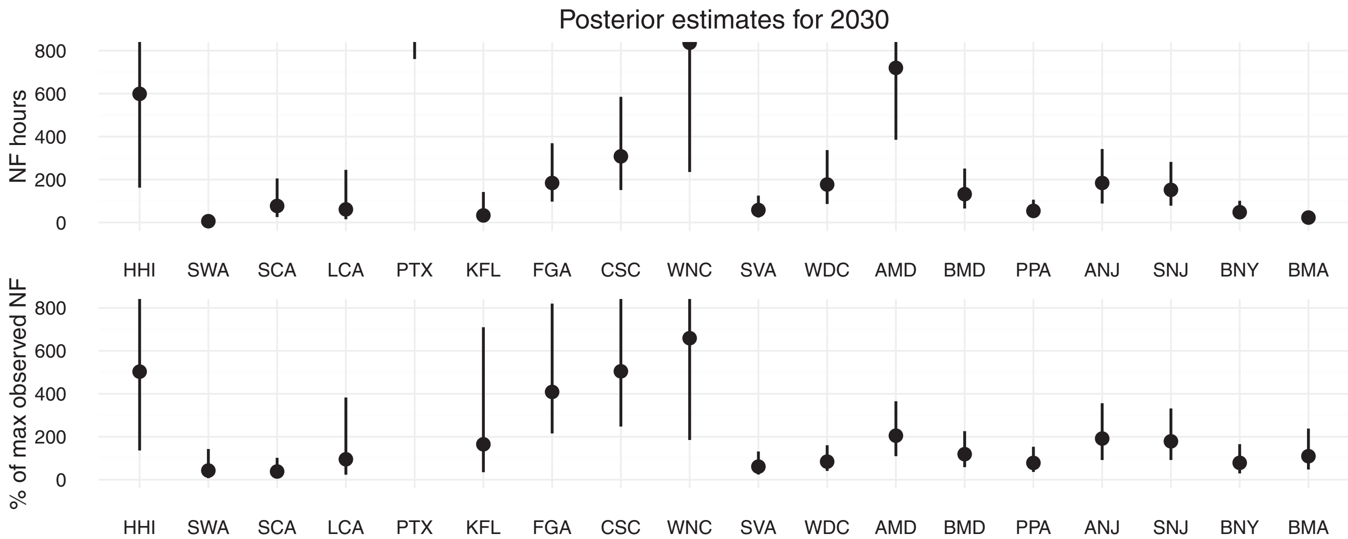

Figure 4. Medians and 67% credible intervals for nuisance flooding in 2030. (top) Hours of NF, and (bottom) percentage of the highest experienced NF through 2014.

processes (e.g., coastal/oceanic circulation) [*Sallenger et al.*, 2012; *Bromirski et al.*, 2011] and vertical land motions (e.g., tectonics activity) [*Wahl et al.*, 2013; *Burgette et al.*, 2009; *Williams*, 2013] might be contributing factors that produce these spatial patterns. The proposed nuisance flooding prediction model is not sensitive to human responses to coastal flooding (e.g., coastal flooding adaptation projects and improved flood defenses). However, human adaptation will reduce the magnitude of impacts of SLR [*Nicholls and Tol*, 2006; *Nicholls et al.*, 2011; *Hinkel et al.*, 2010; *Oppenheimer*, 2013; *Nicholls et al.*, 2014], which will affect NF projections, especially over long time scales (i.e., multidecadal and century scale). Future studies may evaluate the impact of these coastal infrastructural adaptations to the frequency of nuisance events.

5. Conclusions

In this study, we used a combination of generalized linear models (GLMs) and Gaussian Process (GP) models to estimate nuisance flooding associated with local mean sea level. We have shown the superiority of the proposed model (Negative Binomial GLM with a Gaussian Process overdispersion model) in comparison to three other models (i.e., nonlinear least squares, Poisson, and Negative Binomial models without GP overdispersion). The comparison shows that the proposed model improves upon previous methodology, accounts well for the nonstationarity in the variance of NF with respect to MSL, and better quantifies the uncertainty of projections via a statistically robust approach. The results suggest an $\approx$ 400% rise in NF by 2030 under RCP 8.5 compared with the recent experienced average, with a considerable spatial pattern that causes the rise to be different between gauges located in different regions along the coasts of United States.

Acknowledgments
This study is supported by the National Science Foundation award DMS 1331611. Hourly water level data used in this study are available from the National Oceanic and Atmospheric Administration (NOAA; http://tidesandcurrents.noaa.gov/). The local sea level projections are provided by *Kopp et al.* [2014]. The data and all code used for the analysis are available at https://bitbucket.org/vandenbe/sea_level_final/src.

References

Aerts, J., W. W. Botzen, K. Emanuel, N. Lin, H. de Moel, and E. O. Michel-Kerjan (2014), Evaluating flood resilience strategies for coastal megacities, *Science*, *344*(6183), 473–475.

Bromirski, P. D., A. J. Miller, R. E. Flick, and G. Auad (2011), Dynamical suppression of sea level rise along the pacific coast of north America: Indications for imminent acceleration, *J. Geophys. Res.*, *116*, C07005, doi:10.1029/2010JC006759.

Burgette, R. J., R. J. Weldon, and D. A. Schmidt (2009), Interseismic uplift rates for western Oregon and along-strike variation in locking on the Cascadia subduction zone, *J. Geophys. Res.*, *114*, B01408, doi:10.1029/2008JB005679.

Calafat, F., and D. Chambers (2013), Quantifying recent acceleration in sea level unrelated to internal climate variability, *Geophys. Res. Lett.*, *40*, 3661–3666, doi:10.1002/grl.50731.

Cameron, A. C., and P. K. Trivedi (2013), *Regression Analysis of Count Data*, vol. 53, Cambridge Univ. Press, Cambridge, U. K.

Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell (2015), Stan: A probabilistic programming language, *J. Stat. Software.*

Cazenave, A., H.-B. Dieng, B. Meyssignac, K. von Schuckmann, B. Decharme, and E. Berthier (2014), The rate of sea-level rise, *Nat. Clim. Change*, *4*(5), 358–361.

Cherqui, F., A. Belmeziti, D. Granger, A. Sourdril, and P. Le Gauffre (2015), Assessing urban potential flooding risk and identifying effective risk-reduction measures, *Sci. Total Environ.*, *514*, 418–425.

Church, J. A., and N. J. White (2006), A 20th century acceleration in global sea-level rise, *Geophys. Res. Lett.*, *33*, L01602, doi:10.1029/2005GL024826.

Church, J. A., and N. J. White (2011), Sea-level rise from the late 19th to the early 21st century, *Surv. Geophys.*, *32*(4-5), 585–602.

Church, J. A., N. J. White, L. F. Konikow, C. M. Domingues, J. G. Cogley, E. Rignot, J. M. Gregory, M. R. van den Broeke, A. J. Monaghan, and I. Velicogna (2011), Revisiting the earth's sea-level and energy budgets from 1961 to 2008, *Geophys. Res. Lett.*, *38*, L18601, doi:10.1029/2011GL048794.

Church, J. A., et al. (2013), Climate change 2013: The physical science basis; sea level change, in *IPCC Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chap.13, pp. 1137–1216, Cambridge Univ. Press, Cambridge, U. K., and N. Y.

Devlin, A. T., D. A. Jay, S. A. Talke, and E. Zaron (2014), Can tidal perturbations associated with sea level variations in the western pacific ocean be used to understand future effects of tidal evolution?, *Ocean Dyn.*, *64*(8), 1093–1120.

Domingues, C. M., J. A. Church, N. J. White, P. J. Gleckler, S. E. Wijffels, P. M. Barker, and J. R. Dunn (2008), Improved estimates of upper-ocean warming and multi-decadal sea-level rise, *Nature*, *453*(7198), 1090–1093.

Flick, R. E., J. F. Murray, and L. C. Ewing (2003), Trends in united states tidal datum statistics and tide range, *J. Waterway Port Coastal Ocean Eng.*, *129*(4), 155–164.

Flood, J. F., and L. B. Cahoon (2011), Risks to coastal wastewater collection systems from sea-level rise and climate change, *J. Coastal Res.*, *27*(4), 652–660.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2014), *Bayesian Data Analysis*, vol. 2, Taylor and Francis, Chapman and Hall/CRC, Boca Raton, Fla.

Goldberg, P. W., C. K. Williams, and C. M. Bishop (1997), Regression with input-dependent noise: A Gaussian process treatment, *Adv. Neural Inf. Process. Syst.*, *10*, 493–499.

Gornitz, V., S. Couch, and E. K. Hartig (2001), Impacts of sea level rise in the New York city metropolitan area, *Global Planet. Change*, *32*(1), 61–88.

Haigh, I. D., T. Wahl, E. J. Rohling, R. M. Price, C. B. Pattiaratchi, F. M. Calafat, and S. Dangendorf (2014), Timescales for detecting a significant acceleration in sea level rise, *Nat. Commun.*, *5*, 3635.

Hallegatte, S., C. Green, R. J. Nicholls, and J. Corfee-Morlot (2013), Future flood losses in major coastal cities, *Nat. Clim. Change*, *3*(9), 802–806.

Hay, C. C., E. Morrow, R. E. Kopp, and J. X. Mitrovica (2015), Probabilistic reanalysis of twentieth-century sea-level rise, *Nature*, *517*(7535), 481–484.

Hinkel, J., R. J. Nicholls, A. T. Vafeidis, R. S. Tol, and T. Avagianou (2010), Assessing risk of and adaptation to sea-level rise in the European union: An application of diva, *Mitigation Adaptation Strategies Global Change*, *15*(7), 703–719.

Hinkel, J., D. P. van Vuuren, R. J. Nicholls, and R. J. Klein (2013), The effects of adaptation and mitigation on coastal flood impacts during the 21st century. An application of the diva and image models, *Clim. Change*, *117*(4), 783–794.

Hinkel, J., D. Lincke, A. T. Vafeidis, M. Perrette, R. J. Nicholls, R. S. Tol, B. Marzeion, X. Fettweis, C. Ionescu, and A. Levermann (2014), Coastal flood damage and adaptation costs under 21st century sea-level rise, *Proc. Natl. Acad. Sci.*, *111*(9), 3292–3297.

Holgate, S. J., A. Matthews, P. L. Woodworth, L. J. Rickards, M. E. Tamisiea, E. Bradshaw, P. R. Foden, K. M. Gordon, S. Jevrejeva, and J. Pugh (2012), New data systems and products at the permanent service for mean sea level, *J. Coastal Res.*, *29*(3), 493–504.

Jay, D. A. (2009), Evolution of tidal amplitudes in the eastern Pacific Ocean, *Geophys. Res. Lett.*, *36*, L04603, doi:10.1029/2008GL036185.

Kimeldorf, G. S., and G. Wahba (1970), A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, *Ann. Math. Stat.*, *41*(2), 495–502.

Kopp, R. E., R. M. Horton, C. M. Little, J. X. Mitrovica, M. Oppenheimer, D. Rasmussen, B. H. Strauss, and C. Tebaldi (2014), Probabilistic 21st and 22nd century sea-level projections at a global network of tide-gauge sites, *Earth's Future*, *2*(8), 383–406.

Lyu, K., X. Zhang, J. A. Church, A. B. Slangen, and J. Hu (2014), Time of emergence for regional sea-level change, *Nat. Clim. Change*, *4*(11), 1006–1010.

Marcos, M., F. M. Calafat, Á. Berihuete, and S. Dangendorf (2015), Long-term variations in global sea level extremes, *J. Geophys. Res. Oceans*, *120*, 8115–8134, doi:10.1002/2015JC011173.

McCullagh, P., and J. A. Nelder (1989), *Generalized Linear Models*, vol. 37, CRC Press, Boca Raton, Fla.

Milne, G. A., W. R. Gehrels, C. W. Hughes, and M. E. Tamisiea (2009), Identifying the causes of sea-level change, *Nat. Geosci.*, *2*(7), 471–478.

Moftakhari, H. R., A. AghaKouchak, B. F. Sanders, D. L. Feldman, W. Sweet, R. A. Matthew, and A. Luke (2015), Increased nuisance flooding along the coasts of the united states due to sea level rise: Past and future, *Geophys. Res. Lett.*, *42*, 9846–9852, doi:10.1002/2015GL066072.

Muis, S., M. Verlaan, H. C. Winsemius, J. C. Aerts, and P. J. Ward (2016), A global reanalysis of storm surges and extreme sea levels, *Nat. Commun.*, *7*, 1196.

Müller, M., B. Arbic, and J. Mitrovica (2011), Secular trends in ocean tides: Observations and model results, *J. Geophys. Res.*, *116*, C05013, doi:10.1029/2010JC006387.

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

Nicholls, R. J., and R. S. Tol (2006), Impacts and responses to sea-level rise: A global analysis of the sres scenarios over the twenty-first century, *Philos. Trans. R. Soc. A*, *364*(1841), 1073–1095.

Nicholls, R. J., N. Marinova, J. A. Lowe, S. Brown, P. Vellinga, D. De Gusmao, J. Hinkel, and R. S. Tol (2011), Sea-level rise and its possible impacts given a 'beyond 4 c world' in the twenty-first century, *Philos. Trans. R. Soc. A*, *369*(1934), 161–181.

Nicholls, R. J., S. E. Hanson, J. A. Lowe, R. A. Warrick, X. Lu, and A. J. Long (2014), Sea-level scenarios for evaluating coastal impacts, *WIREs Clim. Change*, *5*(1), 129–150.

Oppenheimer, M. (2013), Climate change impacts: Accounting for the human response, *Clim. Change*, *117*(3), 439–449.

Rahmstorf, S. (2007), A semi-empirical approach to projecting future sea-level rise, *Science*, *315*(5810), 368–370.

Rasmussen, C., and C. Williams (2006), *Gaussian Processes for Machine Learning*, vol. 1, pp. 40–43, MIT Press, Cambridge, Mass.

Ray, R. D. (2006), Secular changes of the m2 tide in the Gulf of Maine, *Cont. Shelf Res.*, *26*(3), 422–427.

Ray, R. D. (2009), Secular changes in the solar semidiurnal tide of the western north Atlantic Ocean, *Geophys. Res. Lett.*, *36*, L19601, doi:10.1029/2009GL040217.

Reed, A. J., M. E. Mann, K. A. Emanuel, N. Lin, B. P. Horton, A. C. Kemp, and J. P. Donnelly (2015), Increased threat of tropical cyclones and coastal flooding to New York city during the anthropogenic era, *Proc. Natl. Acad. Sci. U. S. A.*, *112*(41), 12,610–12,615.

Sallenger, A. H., Jr., K. S. Doran, and P. A. Howd (2012), Hotspot of accelerated sea-level rise on the Atlantic coast of north America, *Nat. Clim. Change*, *2*(12), 884–888.

Scavia, D., et al. (2002), Climate change impacts on us coastal and marine ecosystems, *Estuaries*, *25*(2), 149–164.

Spiess, A.-N. (2014), *Propagate: Propagation of Uncertainty*, R package version 1.0-4. [Available at https://CRAN.R-project.org/package=propagate.]

Stocker, T. F. (2014), *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press, Cambridge, U. K.

Suarez, P., W. Anderson, V. Mahal, and T. Lakshmanan (2005), Impacts of flooding and climate change on urban transportation: A systemwide performance assessment of the Boston metro area, *Transp. Res., Part D*, *10*(3), 231–244.

Sweet, W. V., and J. Park (2014), From the extreme to the mean: Acceleration and tipping points of coastal inundation from sea level rise, *Earth's Future*, *2*(12), 579–600.

Ten Veldhuis, J., F. Clemens, G. Sterk, and B. Berends (2010), Microbial risks associated with exposure to pathogens in contaminated urban flood water, *Water Res.*, *44*(9), 2910–2918.

Thompson, P. R., G. T. Mitchum, C. Vonesch, and J. Li (2013), Variability of winter storminess in the eastern united states during the twentieth century from tide gauges, *J. Clim.*, *26*(23), 9713–9726.

Vehtari, A., A. Gelman, and J. Gabry (2016), Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC, arXiv preprint, arXiv:1507.04544.

Visser, H., S. Dangendorf, and A. C. Petersen (2015), A review of trend models applied to sea level data with reference to the acceleration-deceleration debate, *J. Geophys. Res. Oceans*, *120*, 3873–3895, doi:10.1002/2015JC010716.

Wahba, G. (1990), *Spline Models for Observational Data*, vol. 59, SIAM, Philadelphia, Pa.

Wahl, T., and D. P. Chambers (2015), Evidence for multidecadal variability in us extreme sea level records, *J. Geophys. Res. Oceans*, *120*, 1527–1544, doi:10.1002/2014JC010443.

Wahl, T., and D. P. Chambers (2016), Climate controls multidecadal variability in us extreme sea level records, *J. Geophys. Res. Oceans*, *121*, 1274–1290, doi:10.1002/2015JC01105.

Wahl, T., I. Haigh, P. L. Woodworth, F. Albrecht, D. Dillingh, J. Jensen, R. J. Nicholls, R. Weisse, and G. Wöppelmann (2013), Observed mean sea level changes around the north sea coastline from 1800 to present, *Earth Sci. Rev.*, *124*, 51–67.

Wang, C. and R. M. Neal (2012), Gaussian process regression with heteroscedastic or non-Gaussian residuals, arXiv preprint, arXiv: 1212.6246.

Watson, C. S., N. J. White, J. A. Church, M. A. King, R. J. Burgette, and B. Legresy (2015), Unabated global mean sea-level rise over the satellite altimeter era, *Nat. Clim. Change*, *5*(6), 565–568.

White, H. (1981), Consequences and detection of misspecified nonlinear regression models, *J. Am. Stat. Assoc.*, *76*(374), 419–433.

Williams, S. J. (2013), Sea-level rise implications for coastal regions, *J. Coastal Res.*, *63*, 184–196.

Wood, S., and R. Kohn (1998), A Bayesian approach to robust binary nonparametric regression, *J. Am. Stat. Assoc.*, *93*(441), 203–213.

Woodruff, J. D., J. L. Irish, and S. J. Camargo (2013), Coastal flooding by tropical cyclones and sea-level rise, *Nature*, *504*(7478), 44–52.

Wöppelmann, G., and M. Marcos (2015), Vertical land motion as a key to understanding sea level change and variability, *Rev. Geophys.*, *54*, 64–92, doi:10.1002/2015RG000502.