A hierarchical Bayesian model of storm surge and total water levels across the Great Lakes shoreline – Lake Ontario

Scott Steinschneider

Department of Biological and Environmental Engineering, Cornell University, 111 Wing Drive, Riley-Robb Hall, Ithaca, NY, 14853.

Corresponding author: Email: ss3378@cornell.edu, Phone: 607-255-2155

Abstract

This study presents a novel approach to determine the distribution of compound flood events composed of storm surge and static water levels along the Great Lakes shoreline. A mixture distribution of the bulk (Student-t) and tail (GPD-negative binomial) components of storm surge is estimated in a hierarchical Bayesian modeling framework. Parameters are modeled across the entire shoreline using Gaussian processes and hourly gauged data, with priors for spatial autocorrelation informed by numerical output from a lake hydrodynamic model. The distribution of total water levels is obtained through Monte Carlo sampling that combines the estimated distribution of surge with stochastic traces of static lake levels that account for water level management, seasonality, and plausible variability in water supplies. The approach can therefore support coastal flood risk assessments in cases when the distribution of static water levels changes are due to altered water level management or climate change. The model is applied in a case study on Lake Ontario. Results suggest that spatial variability in parameter estimates varies significantly by month and mixture distribution component. Evaluations of performance indicate the model is able to capture adequately storm surge behavior at gauges across the lakeshore, even under cross-validation. A frequency analysis of total water levels at two ungauged sites is presented, with specific attention given to the implications of model assumptions on uncertainty in design events. The paper concludes with a discussion of model limitations and avenues for future work.

Keywords: POT, Mixture Model, Great Lakes, Storm Surge, Bayesian

Introduction

Communities living along the Great Lakes shoreline have experienced rapidly rising water levels since 2015. Between 2019 and 2020, water levels on all five of the Great Lakes have hit record highs compared to data over the last 100 years. Damages to homes, businesses, and agricultural fields have been substantial (IJC-LOSLR Board, 2018; Gronewold and Rood, 2019). These record-setting floods have left shoreline communities unsure of how to assess and prepare for flood events in the future, and highlight the need for timely and accurate flood risk information along the Great Lakes shoreline. The goal of this study is to advance statistical methods that can provide this information and to demonstrate these methods in a case study on Lake Ontario.

Coastal flooding on the Great Lakes results from a combination of processes. A major driver of flooding is the static lake level, or still water level, defined as the water level that is experienced in the absence of wind and waves. Static level dynamics on the Great Lakes manifest primary at weekly and longer timescales, and exhibit a strong annual cycle linked to seasonal changes in precipitation, temperature, evaporation, snow accumulation and melt, and runoff (Lenters, 2001). Long-term shifts in water supplies also drive significant static water level fluctuations on inter-annual to decadal timescales (Gronewold and Stow, 2014) which may become more intense under climate change (Gronewold and Rood, 2019). In addition to these natural sources of variability, static levels on all lakes, but especially Lake Superior and Lake Ontario, are influenced by management decisions at control points on the St. Marys River and St. Lawrence River, respectively. Operational policies at these control points have recently been updated (IJC, 2012; IJC, 2014), complicating the use of historical static levels as a source of data to quantify future flood risk.

Wind-driven storm surge and seiches are another primary driver of flood risk along the Great Lakes coastline. These processes manifest at sub-hourly to daily timescales and often are linked to the passing of

extratropical cyclones or other storm systems (Trebitz, 2006). Winds exert horizontal force on the water surface, inducing surface currents that cause water levels to rise downwind and fall upwind. Storm surge dynamics vary considerable across the Great Lakes shoreline, and depend on a variety of local conditions. For instance, position relative to a storm track greatly influences the storm surge experienced at a location (Danard et al., 2004; Mahdi et al., 2019). Long-term storm surge risk is influenced significantly by fetch length, lake depth, and the climatological orientation of storm tracks and associated wind speed and direction (Danard et al., 2003). Locally, physical features at the site that influence local water level change over short durations can also influence storm surge (Irish et al., 2011). Similar factors that influence storm surge also dictate offshore wave heights and their propagation onshore, which can further add to flood risk along the shoreline.

Floods linked to extended periods of high static water levels tend to occur during the late spring and summer, while storm-related flooding linked to surge occurs most often during the spring and autumn when storm activity is greatest (Angel, 1995). The out-of-phase seasonality of static levels and surge produces a negative correlation between the two (i.e., higher surge events at lower static lake levels), but the causal drivers of the two processes are generally considered independent. That is, extended shifts in precipitation, runoff, and evaporation that influence static levels over weeks to years exhibit little, systematic relationship with the occurrence of individual storm systems that drive large surge events on hourly-to-daily timescales (Lee, 1993). However, moderate to high static levels and surge events can overlap, combining to form extreme total water levels with significant implications for flood damage along the coastline (Meadows et al., 1997). This study focuses on characterizing the risk of these compound events by modeling and then combining the distribution of both static levels and surge.

A significant body of literature has explored how to estimate the distribution of total water levels along coastlines, particularly along the ocean coast. One family of approaches, known as indirect methods, separately models processes that compose total water levels (e.g. astronomical tide, storm surge) and then recombine distributions for those processes using convolution (Tawn and Vassie, 1989; Tawn, 1992). These indirect approaches tend to outperform direct methods that fit distributions directly to total water levels (Haigh et al., 2010). An important aspect of indirect methods is the consideration of both the central range (bulk) and the extremes within the distribution of individual processes, such as storm surge. Both parts of the distribution are important, because even moderately high levels of surge within the bulk distribution can combine with high values in other processes (e.g., tide, static levels) to cause total water

levels that induce flooding. Therefore, mixture models have been forwarded as a useful tool to separately model the central range and tails of the distribution, and then to combine those distributions to model the full range of values (Behrens et al., 2004; Mazas et al., 2014). For the extremes, it is now widely recommended to use a peaks-over-threshold (POT) approach based on the Generalized Pareto – Poisson model (Hawkes et al., 2008), while for bulk values several distributions have been considered (Mendes and Lopes, 2004; Behrens et al., 2004; MacDonald et al., 2011).

Another major issue is the estimation of total water level distributions along entire stretches of coastline, even in the absence of gauged data. One common approach is to use numerical predictions of coastal processes from hydrodynamic models that are gridded continuously along the coastline. Extreme value distributions can be fit directly to these modeled data (Flather et al., 1998). However, non-trivial error can result from systematic biases in the numerical model structure, errors in input variables to the model, or spatial averaging over a coarse grid. Therefore, other efforts have been forwarded to use information from the numerical model, but after adjustments made in a calibration process. For instance, the spatial revised joint probability method (SRJPM; Dixon and Tawn, 1995; Dixon and Tawn 1997; Swift, 2003) exploits spatial autocorrelation in tidal and surge components of sea levels from numerical storm surge predictions, but first corrects for errors in the numerical model based on comparisons against gauged data. Alternatively, one could forgo correcting the numerical storm surge predictions, and instead only use these data indirectly to estimate spatial autocorrelation in surge properties, which could then be used in a spatial probability model of surge based strictly on gauged data. This latter approach has not yet been explored in the literature, but provides an alternative to a complex, post-process calibration of numerical model output.

To the author's knowledge, certain aspects of the methods discussed above have not been extended to extreme total water level estimation along the Great Lakes coastline. For instance, as part of the Great Lakes Coastal Flood Study, recently updated guidance from FEMA suggests the following steps (Melby et al., 2012; Nadal-Caraballo et al., 2012; FEMA, 2014): 1) define a composite set of storms that adequately represent the range of surge and wave conditions across a lake; 2) simulate each storm using high-resolution, lake-wide 2-D storm surge and wave models, applied with historical static lake levels that existed at the time of the storm; and 3) conduct POT-based statistical analyses of the water level and wave conditions simulated under the aforementioned storms. The above methodology has several important benefits, including the joint consideration of static level, surge, and wave variability in the frequency

analysis of total water levels. However, it is also notable that the quantification and propagation of storm surge biases in the 2-D hydrodynamic model and forcing meteorological data are not addressed in the methods described in FEMA (2014) although other sources of uncertainty are considered. In addition, the approach depends on historical static water levels observed during the selected set of storms, which may not be applicable if the management of static lake levels changes (as was recently the case for Lake Superior and Lake Ontario; IJC, 2012; IJC, 2014). Further, paleo-reconstructions have found that water supply and static water level variability on the Great Lakes can be significantly larger than the variability observed in the instrumental record (Wilcox et al., 2007; Wiles et al., 2009; Argyilan et al., 2010; Ghile et al., 2014). This suggests that extreme event analyses limited to that instrumental record may underestimate the frequency of extreme total water levels.

In response to these methodological gaps, this study presents a novel approach to estimate the distribution of storm surge and total water levels across the Great Lakes shoreline, with an application to Lake Ontario. There are two primary contributions of this work. First, a hierarchical Bayesian model is developed to estimate the distribution of storm surge at any location along the lakeshore. This model indirectly leverages information on the spatial autocorrelation of coastal processes from a lake hydrodynamic model through the careful construction of prior distributions while avoiding direct use of numerical model output in parameter estimation to minimize the propagation of numerical model bias. Second, this work combines the results of the Bayesian surge model with stochastic traces of static water levels that account for seasonal lake level management and better capture the variability of water supplies compared to the instrumental record. The approach results in a frequency analysis of total water level events, with uncertainty, at any location along the shoreline. Importantly, the choice to model separately storm surge and static levels and then to combine them afterwards to assess compound flood risk, can support coastal flood risk assessments in cases when the static water level distribution changes due to climate change or altered water level management.

Data

The methods developed in this work are applied to the shoreline of Lake Ontario. Three datasets are used throughout this analysis, summarized here and then described in more detail below:

- Hourly, gauged-based storm surge: these data are the primary dataset used to fit the Bayesian model of storm surge.

- Hourly, model-based storm surge: these data are only used indirectly to develop prior distributions of spatial autocorrelation in the Bayesian model.
- Quarter-monthly, model-based static levels: these data are used to quantify static water level variability under the most recent water level management plan for Lake Ontario.

Hourly, gauge-based storm surge

Hourly water levels between January 1, 1980 and December 31, 2019 are collected from eight long-term gauges located in both the United States (Olcott, Rochester, Oswego, and Cape Vincent NY) and Canada (Kingston, Cobourg, Toronto, and Port Weller in the Province of Ontario). These data contain infrequent missing values (less than 0.5% for most gauges), which are gap-filled by linear interpolation. Hourly storm surge is calculated at each gauge by first calculating a smoothed hourly water level series using a Gaussian filter with a window length of seven days. The smoothed series is then subtracted from the original hourly data to develop an hourly series of storm surge. These storm surge data form the primary dataset used to fit the Bayesian model of storm surge at any location around Lake Ontario. The 7-day smoothing window is selected because static water levels (described below) are available on a quarter-monthly (~7.6 day) time step, and so surge defined using a 7-day window ensures that the sum of quarter-monthly static levels and hourly surge events approximate hourly total water levels. However, the results of the analysis do not change significantly if surge is defined using an alternative smoothing window. For instance, time series of surge defined using a 3-day vs. 7-day smoothing window have Pearson correlations between 0.95 and 0.98 across the eight gauges, and parameter estimates for probability models of surge (described below) are extremely similar.

Hourly, model-based storm surge

Hourly surge data is also collected from the Lake Ontario Operational Forecast System (LOOFS) that is managed by NOAA's National Ocean Service. The LOOFS is based on a gridded hydrodynamic model that uses atmospheric observations and weather prediction guidance to produce three-dimensional predictions of water temperature and two-dimensional forecasts of water levels for Lake Ontario (Chu et al., 2011). The LOOFS also predicts deviations from the average lake level, i.e. seiche and storm surge events. The LOOFS provides two sources of data, short-term (1-24 hour) forecasts and nowcasts, the latter of which is based on near real-time meteorological observations and provides a continuous estimate of present conditions across the lake. This study utilizes the nowcast data for hourly water level

deviations from the lake level average (i.e., storm surge). These gridded data are available along the entire coastline at a 5 km resolution (M=178 grid cells in total) from January 1, 2006 to December 31, 2019. The LOOFS hourly surge data is used indirectly in the Bayesian model of storm surge, by forming the basis of prior distributions for spatial autocorrelation used in that model.

Quarter-monthly, model-based static levels

Finally, this work utilizes static water level data that represents average Lake Ontario levels on a quarter-monthly time step. Water levels on Lake Ontario are regulated through operations of the Moses Saunders Dam on the St. Lawrence River. In January 2017, the International Joint Commission (IJC) changed the regulation plan of the lake that had been in place since the early 1960's. The new regulation plan, dubbed "Plan 2014", was designed to restore some of the natural variability of Lake Ontario water levels needed to support coastal wetlands. The change in regulation plan complicates the use of historic static water level observations in the estimation of future design events for the lake. That is, the variability of static water levels under Plan 2014 and the old management plan are significantly different, and therefore it would be inappropriate to conduct a flood analysis for Lake Ontario using observed data that occurred under the old plan. Unfortunately, there is a paucity of gauge-based observations (~4 years; 2017-2020) that can be used to reflect static water level variability under Plan 2014. Therefore, this study uses modeled static water levels under Plan 2014 in place of the historical record. These data are only available at a quarter-monthly time step, because the entire modeling framework for simulating lake levels under Plan 2014 was developed at this time step (IJC, 2014). I note that these quarter-monthly data are not aggregated from hourly gauge-based data; rather, the quarter-monthly time step is the native temporal resolution of the Lake Ontario water level management simulation model.

Two sets of modeled static water level data are utilized in the analysis. The first dataset represents static water levels that would have been experienced over the historical record (1900-2000) if the Plan 2014 regulation plan had been in place during the 20th century. These data were generated by forcing the Plan 2014 regulation model with historically observed water supplies over the 20th century (IJC, 2014). These modeled data are appended with static, quarter-monthly water levels observations from January 1 2017 through December 31 2019, which occurred after Plan 2014 had been instituted and contain two record-setting floods (in 2017 and 2019) that are critical in the estimation of future design events. The second dataset of modeled data is based on static water level simulations under Plan 2014 that are forced with stochastically generated (rather than historically observed) water supplies. Specifically, 500 different 100-

year time series of stochastic water supplies were generated using time series models fit to the statistical properties of the historical supplies, and which cover a wider range of plausible variability than the historical supplies (IJC, 2014). The stochastic water level simulations provide a large sample of static levels that account for Plan 2014 and can be combined with storm surge models to estimate the frequency of total water levels on Lake Ontario.

Methods

This paper focuses on two major components of total water levels (Z) along the Great Lakes shoreline - static water levels (W) and storm surge (S):

$$Z(t) = W(t) + S(t) \quad (1)$$

Storm surge is taken to be the combination of wind-driven and pressure-driven (seiche) effects on the free surface of the lake, and includes any measured increase in hourly water levels due to wave setup. Static water levels are taken relative to a datum, while storm surge can be interpreted as positive or negative deviations from that static water level. All quantities above are considered on an hourly time step *t*; quarter-monthly static levels are converted to an hourly time step by repeating each quarter-monthly value by the number of hours in a quarter-month (182).

A primary contribution of this paper is the development of a hierarchical Bayesian model to estimate the distribution of surge at any location along the lakeshore, despite the limited availability of long-term gauges. An exploratory data analysis that provides the motivation for model design is presented below followed by a description of the model formulation. Uniquely, the proposed approach combines gauge-based data with data from a lake hydrodynamic model to help estimate the spatial autocorrelation of model parameters, but bases the magnitude of model parameters only on the gauge-based data, in order to avoid the propagation of numerical model bias. A second contribution of this work is an approach to combine the results of the Bayesian surge model with the distribution of highly seasonal, auto-correlated, and managed static water levels to estimate design events for total water levels (i.e., compound flood events), with uncertainty, at any location along the shoreline. This approach is described further below. Importantly, in this approach I assume independence between the static water level and storm surge distributions, besides any correlation between the two processes associated with seasonality in storm activity and water supplies. That is, I model the distribution of static levels and surge independent of each other, by month, and then combine those distributions across months to estimate the likelihood of key

design events for total water levels. Figure 1 provides an overview of the methods and applications of different datasets.

Exploratory Data Analysis to Support Storm Surge Model Design

Mixture Models of the Distribution of Surge Events

In order to estimate design events for total water levels, it is important to consider the joint impact of static levels and storm surge. During periods of high static levels, small surge events that are not considered extreme (e.g., 0.025-0.050 meters, or ~ 1-2 inches) may still substantively add to flood risk if the static level is already encroaching on the foundations of shoreline structures. This highlights the need to model the entire distribution of storm surge (not just tail events), and to combine that storm surge distribution with the distribution of static levels to estimate the risk associated with total water level events. This situation presents a challenge, because it is not straightforward to model the entire distribution of storm surge accurately. As an example, Figure 2 shows the empirical distribution of hourly storm surge for the Olcott gauge in January. The data generally follow a bell curve shape, but exhibit greater kurtosis than can be modeled with a Gaussian distribution. A Student-t distribution is better able to capture the bulk distribution of the data. However, in the tails of the distribution (inset of Figure 2), the Student-t distribution overestimates the likelihood of some of the largest surge values. While this deviation appears small, it can significantly influence the estimated likelihood of low probability events, which are of greatest interest. For example, if 50 years of January hourly surge at the Olcott gauge are randomly sampled from the normal and Student-t distributions presented in Figure 2, the maximum sampled value across those two probability models differs by over 0.2 meters (not shown). Past work has shown that mixture distributions are effective at correcting discrepancies in model representations of the bulk and tail of a distribution (Behrens et al., 2004; Scarrott and MacDonald, 2012). Figure 2 shows a fitted mixture distribution that combines a Student-t distribution for the central range of surge values with a generalized Pareto distribution for the upper tail. While this approach can impose a slight discontinuity in the density function, it helps adjust the rate of decay in the tail of the distribution and provides the best fit for the data shown. This type of mixture distribution will form the basis of the proposed Bayesian model. A key challenge addressed in this work is the estimation of mixture model parameters at locations without gauged data, discussed next.

Challenges of Estimating Storm Surge Distributions across the Great Lakes Shoreline

As described earlier, the distribution of storm surge varies along the Great Lakes shoreline, depending on the climatological wind speed and direction, lake depth, and other local factors. One strategy to estimate the distribution of surge along the shoreline is to estimate the distribution of surge at each of the available gauges, and then to use the distribution estimated at the nearest gauge for a location of interest. This approach, while straightforward to apply, assumes that the distribution of storm surge exhibits significant spatial autocorrelation around the entire shoreline. This may or may not hold, particularly in regions of significant transition (e.g., moving from the southern to eastern shore of Lake Ontario, where wind fetch can vary significantly over relatively short distances compared to gauge spacing; Mason et al., 2018). An alternative strategy is to estimate surge distributions specific to a location of interest using spatially resolved hourly water level data from the LOOFS nowcast, which integrates meteorological data across the entire lake. However, this approach is sensitive to model error. Figure 3 shows three years of hourly storm surge at the eight long-term water level gauges, as well as LOOFS estimates of storm surge at those same locations. In many cases, the variability in LOOFS storm surge appears to match that of the gauges, but at some locations (e.g., Oswego, Toronto) the LOOFS estimates significantly underestimate the observed variability. Figure 4 provides a spatial representation of this bias across the entire lakeshore, showing the frequency that surge exceeds a fixed threshold of 0.025 m (~1 inch) for both for LOOFS grid cells and the available gauges in four different months throughout the year. Figure 4 highlights how there are downward biases in the LOOFS frequency of surge events around Toronto and Oswego, and upward biases near Cape Vincent and Kingston.

Figures 3 and 4 suggest that the direct use of the LOOFS data in estimating storm surge distributions may lead to biased estimates, depending on the location of interest. This result highlights the importance of using gauged data directly in the estimation process. However, the LOOFS data still can provide value in extreme event estimation along the shoreline. Specifically, the LOOFS data can help estimate the degree of spatial autocorrelation in surge, which is difficult to infer from the sparse set of gauges available along the shoreline. Because the LOOFS integrates meteorological data along the entire shoreline, it provides a unique estimate of how quickly storm surge behavior varies over space, and this information on spatial autocorrelation can be accurate even if the magnitude of surge in the LOOFS is biased. The Bayesian model described next presents an approach to combine spatial information from the LOOFS with gauged observations to support the estimation of mixture distributions for storm surge along the entire shoreline.

Hierarchical Bayesian Model of Storm Surge

The proposed Bayesian model of storm surge is fit using hourly, gauged-based storm surge data, and regionalizes storm surge distributions to locations across the entire Lake Ontario shoreline. The model has two primary components: 1) a Student-t distribution for the central range of surge values and 2) a Peaks-Over-Threshold (POT) model for the upper tail of surge events. The spatial variability of parameters for both model components are captured using Gaussian processes, with priors informed by spatial autocorrelation observed in the LOOFS modeled data. Because climate and storm surge behavior vary seasonally, the model described below is applied separately for each month of the year.

Bulk and Tail Distribution of Storm Surge

Let $S_i(t)$ be the hourly storm surge in hour *t* and location *i*, where locations around the entire lakeshore are discretized based on the same 5 km resolution as the LOOFS modeled data (i.e., *i*=1,…,*M*, with *M*=178). Long-term water level gauges are present at some subset of these locations, $G = \{g | g \in \{1, \dots, M\}\}$, and the remaining locations are considered ungauged. The goal is to estimate storm surge distributions at all *M* locations given only gauged data at the subset of locations *G*.

For the bulk distribution, assume that S_i follows a scaled t distribution:

$$S_i \sim t(\sigma_i, \nu_i) \tag{2}$$

where σ_i and ν_i are the scale and degrees of freedom associated with site *i*.

Let u_i be a threshold for the i^{th} location that separates the bulk distribution from the tail distribution, and define $\{P_{i,1}^S, P_{i,2}^S, \dots, P_{i,N_i}^S\}$ as the sample of N_i surge peak values at site *i* that are greater than u_i. Note that surge values greater than u_i are first declustered by requiring 24 h between the end of one surge event (i.e., sequence of exceedances) and the beginning of another event. Surge peaks are defined as the maximum surge within an event. Then, surge peak exceedances are defined as the difference between the surge peaks and the threshold, $Y_i^S = P_i^S - u_i$.

If the threshold u_i is high enough, the distribution of Y_i^S approaches the generalized Pareto distribution (GPD; Pickands, 1975):

$$Y_i^S \sim GPD(\tau_i, \xi_i) \quad (3)$$

where τ_i, ξ_i are the scale and shape parameters for site *i*, respectively. Note that a modified (threshold-invariant) scale can be defined as $\widetilde{\tau_i} = \tau_i - u_i\xi_i$, which will be considered below.

To ensure the GPD is approximately valid, the threshold u_i is selected using mean residual life plots (Coles, 2001) applied to each site. These plots for each of the eight long-term water level gauges (not shown) suggest that u_i must vary by month to support the GPD approximation, but not by location. Therefore, a single threshold u for all gauges is selected by month, with u=0.075 m for October through May, and u=0.05 m for June through September.

Next, define $\{X_{i,1}^S, X_{i,2}^S, \ldots, X_{i,K}^S\}$ to be the number of surge peaks for site *i* over *K* years of data (1980-2019 in this study). Generally, such count data is assumed to follow a Poisson distribution, but preliminary analyses suggest that the counts of surge peaks are too dispersed (Ver Hoef and Boveng, 2007) to be modeled effectively as Poisson distributed (i.e., the variance exceeds the mean). Therefore, in this study I assume surge peak counts follow a negative binomial distribution:

$$X_i^S \sim NegBin(\lambda_i, \kappa_i) \quad (4)$$

where λ_i is the mean rate of occurrence and κ_i is the dispersion parameter. As κ_i becomes larger, the variance of X_i^S approaches the mean, and the distribution converges to the Poisson distribution.

Spatial Model and Prior Distributions

The parameters for each of the distributions described above need to be estimated at all *M* locations around the lakeshore. Following past work (Stedinger and Lu, 1995; Martins and Stedinger, 2000; Mailhot et al., 2013), I assume that a single regional shape parameter for the GPD is applicable to all sites, such that $\xi_i = \xi$. This regional shape is given a relatively uninformative uniform prior distribution between -0.3 and 0.3.

For all other parameters, Gaussian processes are used to model their spatial variability. Let θ_i be a generic parameter representing any of the parameters at site *i* from the set $\{\sigma_i, \nu_i, \widetilde{\tau_i}, \lambda_i, \kappa_i\}$ and define $\theta_{i,log} = \log \theta_i$. I model the vector of values θ_{log} across the *M* sites as multivariate normal:

$$\theta_{log} \sim MVN\left(\mu_{\theta_{log}}, \Sigma_{\theta_{log}}\right) \tag{5}$$

Here, $\mu_{\theta_{log}}$ is a vector of repeated mean values (i.e., the same mean for all sites), while $\Sigma_{\theta_{log}}$ is a covariance matrix. The covariance $\Sigma_{\theta_{log}}(d)$ between any two locations separated by a distance *d* is estimated via an exponential covariance function:

$$\Sigma_{\theta_{log}}(d) = \phi_1 e^{-\phi_2 d} \tag{6}$$

where ϕ_1 is the variance and ϕ_2 is a scaling parameter. Note that $\mu_{\theta_{log}}, \phi_1$, and ϕ_2 are defined separately for each parameter in the list $\{\sigma_i, \nu_i, \widetilde{\tau_i}, \lambda_i, \kappa_i\}$.

The above formulation assumes that the model parameters $\{\sigma_i, \nu_i, \widetilde{\tau_i}, \lambda_i, \kappa_i\}$ vary smoothly across space with some degree of spatial autocorrelation that is reflected in the parameter ϕ_2.If $\phi_2 = 0$, then no spatial autocorrelation exists, and the best estimate of any given parameter at some ungauged site *j* is given by the mean estimate $\mu_{\theta_{log}}$ (after appropriate back-transformation). However, as ϕ_2 increases, nearby parameters become correlated; in this case, estimates of $\theta_{j,log}$ at an ungauged site will depend on the values of that parameter at all other sites, including sites with observed data (sites in set *G*). In this way, the model will estimate $\theta_{j,log}$ at the ungauged site based on a balance between the unconditional mean $\mu_{\theta_{log}}$ and the values θ_{log} estimated at gauged sites, with more emphasis given to θ_{log} at nearby gauged sites as ϕ_2 increases.

To inform the estimate of ϕ_2 for each model parameter, I quantify the spatial autocorrelation of parameters based on the LOOFS data. That is, I fit scaled t, GPD, and negative binomial distributions to the hourly surge, peak surge exceedance, and peak surge count data for each of the *M*=178 LOOFS grid cells. All estimation is conducted via maximum likelihood. Peak surge events for the LOOFS are identified using a grid-cell specific threshold equal to the 99.5th percentile of surge values by month at that grid cell. Grid cell specific thresholds are required because in several months, mean residual life plots suggested that no single threshold can support the asymptotic GPD assumption for peak surge exceedances at all grid cells (not shown). GPD scale parameters for the LOOFS are converted to modified

(threshold independent) scale parameters to ensure that their spatial covariance structure can be mapped to the modified scale parameters of the gauged data.

For each parameter estimated using LOOFS data (i.e., one of the parameters from the set $\{\sigma, \nu, \tilde{\tau}, u, \kappa\}^{LOOFS}$, generically denoted θ^{LOOFS}), I randomly sample estimates for $M/2 = 89$ locations without replacement. Using the parameter estimates for those locations, I apply a logarithmic transformation (to derive θ_{log}^{LOOFS}) and estimate $\mu_{\theta_{log}^{LOOFS}}, \phi_1^{LOOFS}$, and ϕ_2^{LOOFS} via maximum likelihood and Equations 5-6. This is repeated 1,000 times, creating an empirical distribution for each of those three parameters. A gamma distribution is then fit to the samples of ϕ_2^{LOOFS}. This gamma distribution is used as the prior distribution for ϕ_2 in the Bayesian model. In this way, the LOOFS data is used to define prior knowledge of the spatial autocorrelation in each model parameter. Note that the prior of ϕ_2 for the mean rate λ of the negative binomial distribution is estimated based on the spatial autocorrelation in the LOOFS site-specific threshold u_i^{LOOFS}, rather than λ_i^{LOOFS}. This is because λ_i^{LOOFS} will vary across the lake based on variations in local storm surge dynamics and the site-specific threshold u_i^{LOOFS}, while for the gauged data with constant threshold, λ_i will only vary across space due to local storm surge dynamics. The threshold u_i^{LOOFS}, which is calculated as a site-specific percentile, will also vary only due to local storm surge dynamics in the LOOFS model, and likely better reflects the spatial autocorrelation that would be seen in λ_i.

All other parameters ($\mu_{\theta_{log}}$ and ϕ_1 for each of $\sigma, \nu, \tilde{\tau}, \lambda,$ and κ) are given relatively uninformative uniform priors, rather than being based on the estimates from the LOOFS data. This approach is taken because the LOOFS-based estimates of $\mu_{\theta_{log}}$ and ϕ_1 exhibit biases compared to at-site model fits to the gauged data, and so constructing priors for these parameters based on the LOOFS data may propagate these biases into final posterior estimates for each parameter. Instead, I isolate my use of the LOOFS data to learn only about the spatial autocorrelation of model parameters, rather than their magnitudes.

Design Event Estimation of Total Water Levels

To estimate the distribution of total water levels at a site, two steps are required: 1) the bulk (Student-t) and tail (GPD/negative binomial) distributions estimated for that site need to be connected into a mixture distribution that captures the full hourly distribution of surge events; and 2) that mixture distribution

needs to be combined with the distribution of static water levels. Other studies have suggested methods to achieve these steps. For instance, in Mazas et al. (2014), the tail distribution of peak surge events was first converted into a distribution of sequential values (based on a model of peak surge event duration) before being combined with the bulk distribution to form a mixture distribution of hourly surge. That mixture distribution was then convoluted with the distribution of hourly static levels (tidal levels in that study) to create the distribution of total water levels. However, in the present work such an approach is made more difficult by three challenges: 1) The static water levels are based on quarter-monthly levels that have been disaggregated to an hourly time step, and therefore exhibit extremely high autocorrelation at the hourly level. This autocorrelation complicates the convolution and inference of total water level design events. 2) Due to significant and out-of-phase seasonality between static levels and surge, bulk and tail distributions are estimated separately by month. Cross-correlation across months, particularly for static levels that are highly auto-correlated throughout the year, needs to be considered when estimating annual return period events for total water levels. 3) The convolution of distributions for static levels and surge is computationally demanding, making it difficult to propagate parametric uncertainty into the final distribution of total water levels, particularly when parametric uncertainty is quantified using Monte Carlo simulations of the posterior distribution for each parameter.

Given the challenges above, this study takes an alternative, sampling based approach that is conceptually straightforward and computationally manageable. The sampling approach is targeted towards the distribution of annual maxima of total water levels at a site of interest (site *i*), and can be achieved via the following steps:

1. A times series of static water levels *W* is sampled at a quarter-monthly time step over a long record (100 years) and disaggregated to an hourly time step. In this work, I sample one of the 500 different 100-year stochastic water level simulations (see Section 2.3). Each of these series reflects a plausible realization of static water levels under Plan 2014.
2. For each month and year of static water level data, a sample is drawn from the posterior distribution of storm surge model parameters, $\{\sigma_i^{samp}, \nu_i^{samp}, \tilde{\tau}_i^{samp}, \xi^{samp}, \lambda_i^{samp}, \kappa_i^{samp}\}$. A site-specific scale parameter τ_i^{samp} is derived from $\tilde{\tau}_i^{samp}$ based on the posterior sample ξ^{samp} and the threshold *u* for that month ($\tau_i^{samp} = \tilde{\tau}_i^{samp} + u\xi^{samp}$).
3. Uniform random deviates $r_1,\ldots,r_H$ between 0 and 1 are sampled for all *H* hours in the current month and year.

4. Each uniform deviate r_h for hour h is compared to the probability of occurrence for a peak surge event, $\frac{\lambda_i^{samp}}{l}$, where l equals the number of hours in the current month. If $r_h > \frac{\lambda_i^{samp}}{l}$, then a sample is drawn from the Student-t distribution with parameters $\sigma_i^{samp}, \nu_i^{samp}$. If $r_h \leq \frac{\lambda_i^{samp}}{l}$, then a sample is drawn from the GPD distribution with parameters $\tau_i^{samp}, \xi^{samp}$.
5. For samples drawn from the Student-t distribution that are greater than the threshold u, these samples are discarded and resampled until all such samples are below u.
6. The hourly static levels and samples of storm surge are added together to form an hourly time series of total water levels for the current month and year. Note that this sampling approach assumes independence between surge and static levels in a particular month.
7. Steps 2-6 are repeated for all months within a year, and the annual maxima of that year is saved. This is done for every year available in the static water level dataset. Note that the distribution of surge varies by month in this sampling procedure, and the quarter-monthly static water level simulations account for seasonality. Therefore, the changing risk of compound flood events is captured correctly across different times of year.
8. Steps 1-7 are repeated many times (e.g., 1000) in order effectively sample from the posterior distribution of surge parameters.

It is worthwhile to note that the sampling of hourly storm surge in steps 3-5 is akin to sampling from a mixture of probability distributions: $\pi(s) = p\pi_1(s) + (1-p)\pi_2(s)$. Here, $\pi(s)$ is the distribution of hourly surge, which is developed as a mixture of the bulk distribution, π_1, and the tail distribution, π_2, with mixing probability p. The mixing probability is determined by the rate parameter of the negative binomial distribution, $p = \frac{\lambda_i^{samp}}{l}$, and a random draw from either the bulk or the tail distribution is determined by comparing the random deviate r_h to that mixing probability. The only caveat is that samples from the bulk distribution above the threshold u are discarded and resampled.

The result of the above methodology in steps 1-8 is an ensemble of annual maxima time series of total water levels. Each member of the ensemble has length n equal to the number of years available in the static water level time series (in this study, 100 years). From this ensemble of annual maxima series, design events associated with a specific return period can be estimated empirically, with uncertainty, by sorting each ensemble member and calculating percentiles of total water levels associated with that return

period (empirically estimated as $T = \frac{n+1}{n+1-j}$, with j equal to the rank of the sorted data). The uncertainty in return levels will reflect uncertainty in both the static levels and the posterior distribution of storm surge parameters at a given site. In this work, I leverage simulations from a previously created stochastic simulation framework of static water levels for Lake Ontario under Plan 2014 to quantify static water level uncertainty. If no such model exists, then a time series model of static levels should be created and used to capture the auto-correlated nature of static levels across months.

The sampling scheme proposed above ignores the fact that peak surge events (those drawn from the GPD) usually occur in clusters. Therefore, the resulting samples do not fully characterize the hourly distribution of total water levels. However, if the only goal is to sample the annual maxima of hourly total water levels, it is not problematic to ignore persistence in peak surge events because I am only interested in the maximum value of the event. If no peak surge events occur within a given month (i.e., $r_h > \frac{\lambda_i^{samp}}{l}$ for all $h \in \{1, \dots, H\}$), then maximum surge values for that month are based on samples from the bulk (i.e., Student-t) distribution of hourly surge.

3.4. Application

The Bayesian model of storm surge is developed in the STAN probabilistic coding language. Posterior distributions are evaluated using the Hamiltonian Monte Carlo sampling method (Duane et al., 1987). Three chains are run for all parameters with overdispersed initial values using 2,500 burn-in simulations and 2,500 iterations afterwards. Convergence is assessed based on chain mixing using the Gelman and Rubin convergence criterion (Gelman and Rubin, 1992).

In the results below, I first show the spatial autocorrelation in model parameters inferred from the LOOFS data, and demonstrate how this information propagates into the posterior distribution of storm surge parameters for different months of the year. I then validate the fitted Bayesian model for storm surge at the eight available long-term gauges along the Lake Ontario shoreline. Finally, I examine return levels and their uncertainty for total water levels at two specific ungauged locations (near Sodus Point, NY and Prince Edward County, CA). In addition to the proposed mixture model, I also consider the distribution of annual maxima total water levels if samples of storm surge are drawn exclusively from the Student-t

distribution estimated for the ungauged locations. This comparison serves to highlight how the mixture model for surge helps constrain the uncertainty around the most extreme total water level design events.

Results and Discussion

Spatial Model and Posterior Distributions

Figure 5 shows Moran scatterplots (Anselin 1996) for each of the six parameters, by month, estimated via maximum likelihood at the M=178 grid cells using LOOFS data. The Moran scatter plot shows the original parameter against a weighted average of that parameter at neighboring locations (i.e., spatially lagged parameter), where the weights are higher for closer locations and are based on inverse Euclidean distance. A positive relationship in the scatter is indicative of positive spatial autocorrelation, while no relationship in the scatter indicates that the parameter has little or no spatial autocorrelation. Recall that because the LOOFS data requires a grid-cell specific threshold, I examine the spatial autocorrelation in this threshold, rather than the mean rate of occurrences of peak exceedances used in the negative binomial model. Figure 5 shows that in all months, there is significant spatial autocorrelation in the LOOFS threshold and the scale of the Student-t distribution. There is also non-trivial spatial autocorrelation in the modified scale of the GPD, the dispersion parameter of the negative binomial distribution, and the degrees of freedom parameter of the Student-t distribution in most months. There is no significant spatial autocorrelation in the GPD shape parameter, justifying the modeling choice above to use a single regional shape for all sites.

The spatial autocorrelation in the LOOFS parameter estimates are summarized in gamma prior distributions for ϕ_2 in the Bayesian model. Figure 6 shows example priors for two parameters (ν, the degrees of freedom parameter for the Student-t distribution; and λ, the mean rate of occurrence parameter for the negative binomial distribution) and two months (January and July). Recall that smaller values of ϕ_2 indicate less spatial autocorrelation. Consistent with Figure 5, the priors in Figure 6 show higher autocorrelation (i.e., higher densities for larger values of ϕ_2) for λ compared to ν. In addition, there is a high degree of certainty that spatial autocorrelation in ν is low, while for λ there tends to be a larger spread in ϕ_2. For both parameters, spatial autocorrelation tends to be higher in the cold season compared to the warm season. However, this is not always the case (see Figure 5).

In the Bayesian inference, the priors for ϕ_2 are combined with the spatial autocorrelation inferred for parameters across the gauged sites to determine the posteriors of ϕ_2, and consequently, the posteriors for parameters at all sites across the lake. This can be seen in Figure 7, which shows the posterior mean estimates for each of the spatially varying model parameters, $\{\sigma, \nu, \tilde{\tau}, \lambda, \kappa\}$, for January and July. The posterior means at the grid cells associated with the gauged sites are highlighted separately. Figure 7 demonstrates how posterior mean estimates at gauged sites propagate into the estimates at ungauged sites over relatively large distances for parameters with higher spatial autocorrelation. For example, the mean rate of occurrence for exceedances (λ) in January is highly spatially autocorrelated, which manifests in the slow transitions in posterior mean estimates between gauges. Conversely, this same parameter is less spatially autocorrelated in the summer, and consequently there are faster transitions from estimated values at the gauged sites to the global mean at nearby ungauged sites. A similar pattern is seen for other parameters with only moderate spatial autocorrelation (σ, ν). These more abrupt transitions at specific gauges could be caused by sampling variability in certain model parameters, e.g., the degree of freedom parameter (ν) that can be highly sensitive to outlier surge values. Alternatively, the location of a particular gauge could have some unique properties that alter the behavior of hourly water levels at that specific site. For example, the gauges at Toronto and Cobourg in Canada are located in harbors, while the gauges at Cape Vincent and Kingston are located at the inlet to the St. Lawrence River. The hydrodynamics at those locations could uniquely influence surge behavior, and this might explain some of the more abrupt transitions for certain parameters at those sites.

Model Validation and Total Water Level Design Events

Figures 8 and 9 show the goodness of fit of Bayesian model estimates for each of the eight gauges. Figure 8 shows probability plots of the peak surge exceedance data and the generalized Pareto distribution with posterior mean estimates for (τ, ξ); probability plots for all hourly surge values and the Student-t distribution with posterior mean estimates for (σ, ν); and histograms of exceedance counts and fitted negative binomial densities with posterior means for (λ, κ). All results are shown for the month of January. For all sites and all variables, the models with posterior mean parameter estimates provide very good fits to the data. Formal goodness of fit tests are provided in Figure 9 for all months. Here, Kolmogorov-Smirnov (KS) tests are conducted to assess goodness of fit for the GPD and Student-t distributions with posterior mean parameter estimates, while a Chi-Square test is used to assess goodness of fit for the negative binomial distribution. Figure 9 shows the distribution of p-values of these tests across all sites for each month. Results suggest that both the fitted GPD and negative binomial distributions provide adequate fits to the data at the vast majority of site-month combinations. KS tests on

the bulk distribution suggest poor model fits at most sites in all months. However, this result is an artifact of sample size. There are tens of thousands of hourly surge values for each site and month combination, and so even very small deviations between the fitted Student-t distribution and the data will lead to rejections of the null hypothesis under the KS test (Johnson and Wichern, 1992). However, as shown in Figure 8 (and Figure 2), a Student-t distribution does provide a very reasonable fit for the bulk distribution of the hourly surge data.

To confirm out-of-sample performance, the Bayesian model was re-estimated without data from the Oswego gauge. The Bayesian model was then used to regionalize GPD parameters to the Oswego location, and the (posterior mean) estimated model was evaluated against storm surge magnitudes at that location (Figure 10). For comparison, GPD distributions fit at the two nearest gauges on either side of the Oswego gauge (Cape Vincent and Rochester) were applied to the Oswego site. Figure 10 shows probability plots for surge in spring and autumn months (March and October) when storms are more active. The results suggest that the Bayesian model produces reasonable out-of-sample estimates of the storm surge distribution at the Oswego site, especially when compared to the use of the fitted GPD at the nearest gauge (Cape Vincent). The Bayesian model also leads to moderate improvements over the use of the fitted GPD at the Rochester gauge (the second nearest). In other months (not shown), the Bayesian model estimates are either equivalent to or slightly better than the use of the GPD estimated at the Rochester gauge.

To this point, all model evaluation has been conducted using posterior mean parameter estimates. However, uncertainty quantification is a major benefit of Bayesian modeling. In Figure 11, the full posterior distribution of storm surge model parameters is propagated into the estimates of total water level design events, along with uncertainty in static water levels, as described in Section 3.3. Return levels, with 95% uncertainty intervals, are shown for two ungauged locations near Sodus Point, NY (Figure 11a) and Prince Edwards County, ON (Figure 11b) under two cases: 1) using the proposed mixture model of surge, and 2) using storm surge samples drawn exclusively from the Student-t distribution estimated at the ungauged location (i.e., no mixture with a GPD targeted for tail events). These models are compared against historic (1900-2000 + 2017-2019), annual maxima water level data at a quarter-monthly time step (see Section 2.3; recall that hourly historical data that accounts for Plan 2014 operations is not available). Figure 11c shows the difference in median return level estimates between the two sites (Sodus Point – Prince Edward County) for the two different probability models.

Three major insights emerge from Figure 11. First, median estimates from both models and both sites general predict larger return levels than suggested by the observed, quarter-monthly data for small design events (< 5-year return periods), but align with or only slightly exceed the empirical return levels for larger events (Figure 11a,b). This is unsurprising because for more frequent events, it is highly likely that storm surge (which is accounted for by the models but not the quarter-monthly observations) would add significantly to the total water level. However, for less frequent quarter-monthly events (e.g., those greater than 75.5 m, a level that can cause widespread flooding), it is simply less likely that a large hourly surge value would occur concurrently with that high quarter-monthly level.

Second, the mixture model and Student-t only model predict similar median return levels across the full range of return periods at both sites, with the Student-t only model generally exceeding the mixture model by between 2 and 4 cm (Figure 11a,b). However, the models differ significantly in their uncertainty bounds, especially for the largest events. For instance, for Sodus Point (Figure 11a), the Bayesian mixture model estimates the upper bound of the 95% uncertainty interval for the 100-year event at 76.5 m, while the model based entirely on the Student-t distribution estimates this upper bound at 77.5 m (a meter higher). The latter is unrealistically high. For instance, in the 50,000 years of stochastic quarter-monthly data (without storm surge), the maximum level reached is 76.66 m. The maximum hourly storm surge observed between 1980 and 2019 at the two nearest gauges to Sodus Point NY is 0.34 m and 0.40 m, respectively. Therefore, even with storm surge, it is highly unrealistic that the 100-year total water level event would reach 77.5 m. The uncertainty bounds estimated by the model using only the Student-t distribution are too wide because that single distribution cannot accurately model both the spread in the bulk distribution and in the tails. The fitted distribution is overly influenced by the central range of values, leading to overestimated probabilities of very large surge values (see Figure 2). Conversely, the mixture model is able to tailor a separate GPD distribution to the tails, thus constraining probability estimates of the largest values. The result is a more realistic upper bound on the highest return levels for total water level events.

Finally, when comparing the return levels between Sodus Bay, NY and Prince Edward County, CA, both the mixture and Student-t only models agree that compound flood risk is higher at Sodus Bay, NY (Figure 11c). Return level differences range near 0.075 m for smaller return periods and grow to between 0.15-0.20 meters for the highest return periods. The higher flood risk at Sodus Bay is attributable to more

storm surge at that site. Specifically, the Bayesian model estimates that the frequency of surge events over the threshold (i.e., the mean rate of surge occurrence, λ) is substantially larger at Sodus Bay than Prince Edward County, ON between the months of March through June. These months coincide with the time of highest static water levels on Lake Ontario, thus increasing the compound risk of high static water and high surge events at Sodus Point NY compared to Prince Edward County, ON.

Conclusions

This study forwarded two major contributions that add to a developing literature around improved flood frequency approaches with uncertainty for coastal engineering (Gonzalez et al., 2019), tailored for the unique circumstances observed on the Great Lakes. First, this work developed a hierarchical Bayesian model that can estimate storm surge distributions, with uncertainty, across the Great Lakes shoreline. Uniquely, the model uses prior distributions for parameters controlling spatial autocorrelation to leverage credible information from hydrodynamical models of the Great Lakes, without propagating biases embedded in those modeled data. This approach highlights a key strength of Bayesian modeling, where the use of information can be tailored based on its credibility through carefully crafted prior distributions. The Lake Ontario case study highlighted some interesting features of storm surge, including distinct seasonality in both the magnitude and spatial autocorrelation of storm surge model parameters. The model was also able to capture adequately the behavior of surge across all available gauges on the lake, both in the central range and tail of their distributions.

In addition, this work advanced a new approach to estimate design events for total water levels at ungauged sites in the Great Lakes basin, taking into account the highly seasonal and auto-correlated nature of static water levels, static water level management through control structures, and uncertainty in both static water level and storm surge distributions. The proposed approach is directly generalizable to the other Laurentian Great Lakes, which have similar gauging networks, operational forecast systems, regulated static water levels, and stochastic static water level datasets (IJC 2012). Importantly, the approach to model separately and then combine storm surge and static level distributions provides a way to support coastal flood risk assessments in cases when static water level management changes, as recently was the case in both the Upper Great Lakes (IJC 2012) and on Lake Ontario (IJC 2014). Results from this study showed the importance of accounting for storm surge when developing total water level design events, especially for more frequent return periods. In addition, results highlighted the importance of correct distributional assumptions when estimating uncertainty bounds around the most extreme design

events. As shown in Figure 11a,b, the final return level used for design of capital-intensive projects (e.., seawalls, bulkheads, etc.) can be significantly influenced by these modeling choices, especially if a conservative return level estimate is used based on the estimated uncertainty bounds.

While the results of this work are promising, several limitations require additional discussion. First, it is important to recognize that the Bayesian model regionalizes the behavior of surge around the entire lakeshore based on observed surge at a limited number of gauged locations. Any unique factors effecting surge behavior at a particular gauge (e.g., hydrodynamics within a harbor) could negatively impact the regionalization process to other, open-water locations. The Bayesian model was designed to reduce this impact by accounting for the degree of spatial autocorrelation observed across gauges for different seasons, but these impacts can likely never be eliminated. Likewise, further model adjustments would be needed to estimate surge specifically within embayments or other adjacent water bodies separated from Lake Ontario by barrier beach systems.

The total water level design events estimated using the proposed model are based on storm surge and static water level datasets that reflect stationary conditions. The stochastic water level dataset does not account for changes in the Great Lakes water balance that could occur under climate change, including increased evaporative water loss, shifts in seasonality due to changes in snow accumulation and melt, and changes in regional precipitation (Hayhoe et al., 2010; Gronewold et al. 2013; Lofgren and Rouhana, 2016; Byun and Hamlet, 2018). Rather, the stochastic water levels used here for design event estimation only account for natural variability in the system, not long-term change. Future work is needed to develop a similar dataset of stochastic levels that can account for both natural variability and systematic climate change, including uncertainty around those climate changes. However, the methods forwarded in this study are immediately applicable if such future static level scenarios are available, which is a unique benefit of the approach. Similarly, the historic hourly water levels used to estimate storm surge models do not account for potential changes in surge due to climate change, e.g., shifts in wind speeds or storm tracks across different seasons (Desai et al., 2009). Additional work is needed to explore the credibility of these projected changes and their potential effects on total water level design estimates.

Finally, the total water level design events estimated using the current model do not account for wave runup. Wave activity can add significantly to total water levels and flooding along the shoreline, and

current guidelines for total water level frequency analysis correctly emphasize the need to incorporate the impact of waves (FEMA, 2014). However, wave runup is significantly influenced not only by local bathymetric and other nearshore conditions, but also by design specifications of shoreline protection infrastructure. Therefore, wave runup needs to be calculated on a parcel basis taking into consideration data on local conditions and infrastructure, making regionalization of such a model difficult. In addition, any proposed model to account for wave runup in total water level calculations should consider the correlation between wave activity and surge, which both are associated with storm activity and wind events (Paprotny et al., 2018), as well as possible trends in wave activity due to climate variability and change (Grieco and DeGaetano, 2019). These efforts are left for future work.

Acknowledgements

This work was prepared under award NA19OAR4310313 from the National Oceanic & Atmospheric Administration/Department of Commerce to Syracuse University (SU), and also under project R/CHD-15 funded under award NA18OAR4170096 from the National Sea Grant College Program of the U.S. Department of Commerce's National Oceanic and Atmospheric Administration, to the Research Foundation for State University of New York on behalf of New York Sea Grant. The statements, findings, conclusions, views and recommendations are those of the author(s) and do not necessarily reflect the views of any of those organizations.

References

Angel, J.R. (1995). Large-scale storm damage on the U.S. shores of the Great Lakes, Journal of Great Lakes Research, 21 (3), 287-293.

Anselin, L. (1996). The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In *Spatial Analytical Perspectives on Gis in Environmental and Socio-Economic Sciences*, edited by Manfred Fischer, Henk Scholten, and David Unwin, 111–25. London: Taylor; Francis.

Argyilan EP, Forman SL, Thompson TA. Variability of Lake Michigan water level during the past 1000 years reconstructed through optical dating of a coastal strandplain. The Holocene. 2010;20(5):723-731. doi:10.1177/0959683609358913

Behrens, C.N., Lopes, H.F., Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. Stat. Model. 4, 227–244.

Byun, K.; Hamlet, A.F. (2018). Projected changes in future climate over the Midwest and Great Lakes region using downscaled CMIP5 ensembles. Int. J. Climatol., 38, 531–553.

Chu, P.Y., Kelley, J.G.W., Mott, G. V., Zhang, A., and G.A. Lang. (2011). Development, Implementation, and Skill Assessment of the NOAA/NOS Great Lakes Operational Forecast System. *Ocean Dynamics* 61: 1305–1316. https://doi.org/10.1007/s10236-011-0424-5

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer, London.

Danard, M.B., A., Munro, A., and Murty, T.S. (2003). Storm surge hazard in Canada. Natural Hazards, 28, 407–431. https://doi.org/10.1023/A:1022990310410

Danard, M.B., Dube, S.K., Gonnert, G., Munroe, A., Murty, T.S., Chittibabu, P., Rao, A.D., Sinha, P.C. (2004). Storm surges from extra-tropical cyclones. Natural Hazards, 32, 177–190.

Desai, A., Austin, J., Bennington, V. et al. (2009). Stronger winds over a large lake in response to weakening air-to-lake temperature gradient. Nature Geosci 2, 855–858. https://doi.org/10.1038/ngeo693

Dixon, M.J., Tawn, J.A. (1995). Extreme sea-levels at the UK A-class sites: optimal site-by-site analyses and spatial analyses for the east coast. Lancaster University and the Proudman Oceanographic Laboratory, Tech. Rep., August 1995.

Dixon, M.J., Tawn, J.A. (1997). Spatial analyses for the UK coast. Lancaster University and the Proudman Oceanographic Laboratory, Tech. Rep., June 1997.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth D. (1987). Hybrid Monte Carlo, Phys. Lett. B, 195(2), 216–222.

Flather, R.A., Smith, J.A., Richards, J.D., Bell, C., Blackman, D.L. (1998). Direct estimates of extreme storm surge elevations from a 40-year numerical model simulation and from observations. The Global Atmosphere and Ocean Systems 6, 165–176. WASA-Special issue. Waves and Storms in the North Atlantic Editor: Hans Von Storch.

FEMA. (2014). Great Lakes Coastal Guidelines, Appendix D.3 Update. Chicago, IL: FEMA Region V.

Gelman, A., and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences, Stat. Sci., 7(4), 457–511.

Ghile, Y., Moody, P. & Brown, C. (2014). Paleo-reconstructed net basin supply scenarios and their effect on lake levels in the upper great lakes. Climatic Change 127, 305–319. https://doi.org/10.1007/s10584-014-1251-8

Gonzalez, V.M., Nadal-Caraballo, N.C., Melby, J.A., and Cialone, M.A. (2019). Quantification of uncertainty in probabilistic storm surge models: literature review. ERDC/CHL SR-19-1. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Grieco, M.B., DeGaetano, A.T. (2019). A climatology of extreme wave height events impacting eastern Lake Ontario shorelines. Theor. Appl. Climatol., 136, 543–552. https://doi.org/10.1007/s00704-018-2502-9

Gronewold, A.D., Fortin, V., Lofgren, B. et al. (2013). Coasts, water levels, and climate change: A Great Lakes perspective. Climatic Change 120, 697–711. https://doi.org/10.1007/s10584-013-0840-2

Gronewold, A.D., and Rood, R.B. (2019), Recent water level changes across Earth's largest lake system and implications for future variability, Journal of Great Lakes Research, 45 (1), 1-3.

Haigh, I., Nicholls, R., Wells, N. (2010). A comparison of the main methods for estimating probabilities of extreme still water levels. Coast. Eng. 57, 838–849.

Hawkes, P., Gonzalez-Marco, D., Sánchez-Arcilla, A., Prinos, P. (2008). Best practice for the estimation of extremes: a review. J Hydraul Res, 46 (2), 324–332.

Hayhoe, K., VanDorn, J., Croley, T., Schlegal, N., Wuebbles, D. (2010). Regional climate change projections for Chicago and the US Great Lakes. J Great Lakes Res 36:7–21

Johnson, R. A., & Wichern, D. W. (1992). Applied multivariate statistical analysis (3rd ed.). New Jersey: Prentice-Hall.

International Joint Commission (IJC). (2012). Lake Superior Regulation: Addressing uncertainty in upper Great Lakes water levels, Final Rep., 236 pp., 28 March.

International Joint Commission (IJC). (2014). Lake Ontario St. Lawrence River Plan 2014: Protecting against extreme water levels, restoring wetlands, and preparing for climate change. http://www.ijc.org/files/tinymce/uploaded/LOSLR/IJC_LOSR_EN_Web.pdf

International Joint Commission (IJC) Lake Ontario - St. Lawrence River (LOSLR) Board. 2018. Observed Conditions & Regulated Outflows in 2017. https://ijc.org/sites/default/files/2018-08/ILOSLRB_FloodReport2017.pdf

Irish, J. L., D. T. Resio, and D. Divoky. (2011). Statistical Properties of Hurricane Surge along a Coast. Journal of Geophysical Research: Oceans, 116 (C10), C10007.

Lee, D.H. (1993), Great Lakes water level statistical techniques, NOAA Technical Memorandum ERL GLERL-78.

Lenters, J. D. (2001), Long-term trends in the seasonal cycle of Great Lakes water levels, J. Great Lakes Res., 27(3), 342–353.

Lofgren, B. M., and Rouhana, J. (2016). Physically Plausible Methods for Projecting Changes in Great Lakes Water Levels under Climate Change Scenarios. *J. Hydrometeor.*, **17**, 2209–2223, https://doi.org/10.1175/JHM-D-15-0220.1.

MacDonald, A., Scarrott, C.J., Lee, D., Darlow, B., Reale, M., Russell, G., 2011. A flexible extreme value mixture model. Comput. Stat. Data Anal. 55, 2137–2157. https://doi.org/10.1016/j.csda.2011.01.005

Mahdi, T., Jain, G., Patel, S., Sidhu, A.K. (2019). A review of cyclone track shifts over the Great Lakes of North America: implications for storm surges. Nat Hazards 98, 119–135. https://doi.org/10.1007/s11069-018-3429-2

Mailhot A., Lachance-Cloutier S., Talbot G., Favre A-C. (2013). Regional estimates of intense rainfall based on the peak-over-threshold (POT) approach. J Hydrol, 476, 188–199.

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

Mason, L. A., Riseng, C. M., Layman, A. J., and Jensen, R. (2018). Effective fetch and relative exposure index maps for the Laurentian Great Lakes. Scientific Data 5, 180295. https://doi.org/10.1038/sdata.2018.295.

Mazas, F., Kergadallan, X., Garat, P., and Hamm, L. (2014), Applying POT methods to the Revised Joint Probability Method for determining extreme sea levels, Coastal Engineering, 91, 140-150.

Meadows, G.A., Meadows, L.A., Wood, W.L., Hubertz, J.M., and Perlin, M. (1997). The relationship between Great Lakes water levels, wave energies, and shoreline damage, Bulletin of the American Meteorological Society, 78 (4), 675-683.

Melby, J. A., Nadal-Caraballo, N.C., Pagan-Albelo, Y., and Ebersole, B.A. (2012). Wave Height and Water Level Variability for Lakes Michigan and St. Clair. ERDC/CHL TR-12-23. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Mendes, B., Lopes, H.F., 2004. Data driven estimates for mixtures. Comput. Stat. Data Anal.47, 583–598.

Nadal-Caraballo, N. C., Melby, J. A., and Ebersole, B. A. (2012). Statistical Analysis and Storm Sampling Approach for Lakes Michigan and St. Clair. ERDC/CHL TR-12- 19. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Paprotny, D., Vousdoukas, M. I., Morales-Nápoles, O., Jonkman, S. N., & Feyen, L. (2018). Compound flood potential in Europe. Hydrology and Earth System Sciences Discussions, 1–34.

Pickands, J. (1975). Statistical inference using extreme order statistics. Ann. Stat. 3 (1), 119–131.

Scarrott, C., and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification, Revstat, 10, 33–60. https://www.ine.pt/revstat/pdf/rs120102.pdf

Swift, R. (2003). Extreme water levels — an interregional comparison. ABPmer Internal Report Number R1058.

Stedinger, J.R., Lu, L.-. (1995). Appraisal of regional and index flood quantile estimators. Stochastic Hydrol Hydraul 9, 49–75. https://doi.org/10.1007/BF01581758

Tawn, J. (1992). Estimating probabilities of extreme sea-levels. Appl. Stat. 41 (1), 77–93.

Tawn, J.A., Vassie, J.M. (1989). Extreme sea levels: the joint probabilities method revisited and revised. Proc Instn Civ Engrs, Part 2, 87, pp. 429–442.

Ver Hoef, J.M. and Boveng, P.L. (2007), Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology, 88, 2766-2772. doi:10.1890/07-0043.1

Wilcox, D.A., Thompson, T.A., Booth, R.K., Nicholas, J.R. (2007). Lake-Level Variability and Water Availability in the Great Lakes, U.S. Geological Survey. Circular 1311 25 pp.

Wiles, G. C., Krawiec, A. C., and D'Arrigo, R. D. (2009). A 265-year reconstruction of Lake Erie water levels based on North Pacific tree rings, Geophys. Res. Lett., 36, L05705, doi:10.1029/2009GL037164.

Figure Captions

Figure 1. Flowchart of the methodology and application of different datasets.

Figure 2. Empirical distribution of January, hourly storm surge at the Olcott gauge, along with fitted normal, Student-t, and mixture distributions. The orange box delimits the area of the distribution highlighted in the inset.

Figure 3. Gauge-based and LOOFS-based hourly storm surge at eight locations on Lake Ontario.

Figure 4. Frequency (expressed as a percentage) that hourly storm surge exceeds a fixed threshold of 0.025 m (~1 inch) at each LOOFS grid cell, as well as at each of the eight long-term gauges.

Figure 5. Moran I plots for LOOFS parameters, including GDP modified scale, GDP shape, negative binomial dispersion, threshold value, Student-t scale, and Student-t degrees of freedom. Different colors show values for different months (cold months in blue, warmer months in red).

Figure 6. Gamma prior distributions for ϕ_2 applied to ν, the degrees of freedom parameter for the Student-t distribution, and λ, the mean rate of occurrence parameter for the negative binomial distribution. Priors are shown separately for the months of January and July.

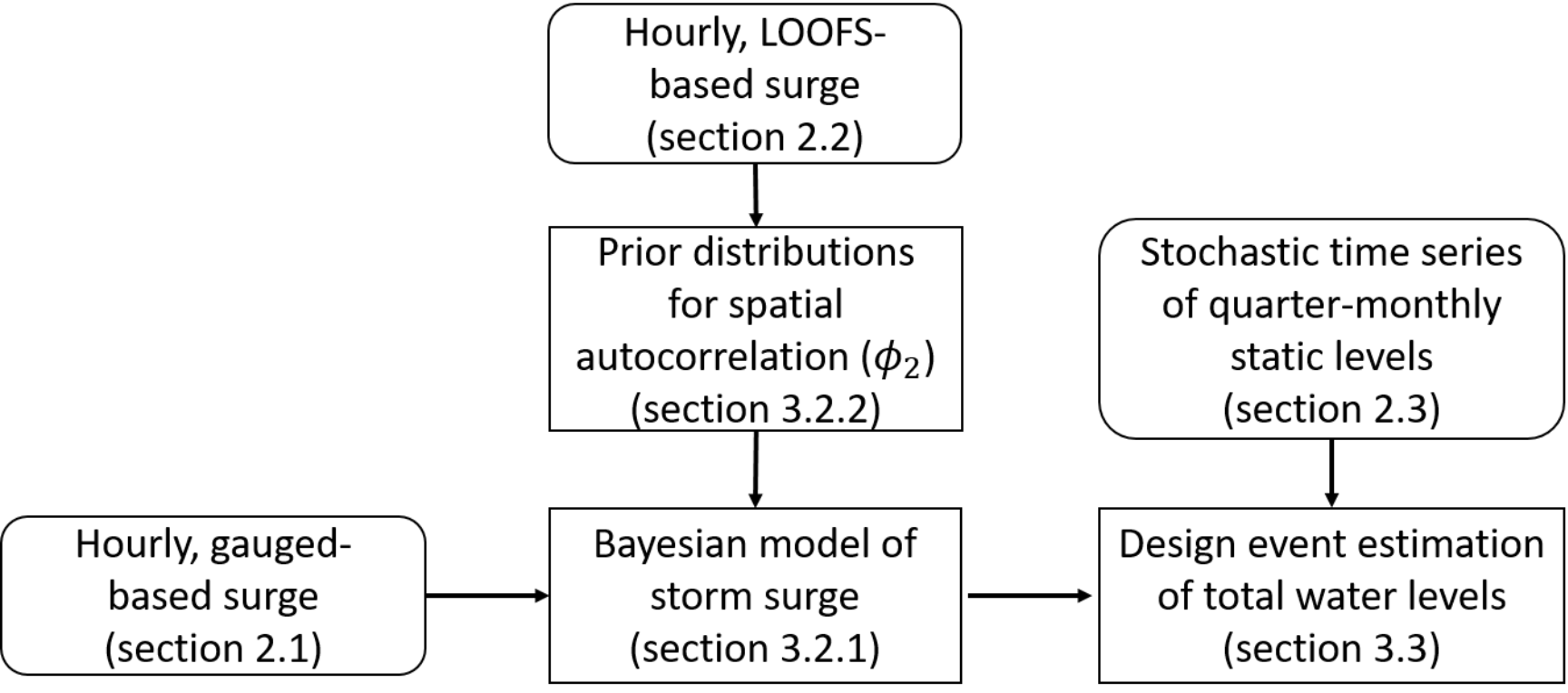
Figure 7. Bayesian posterior mean estimates of parameters that vary across the lake for the months of January and July.

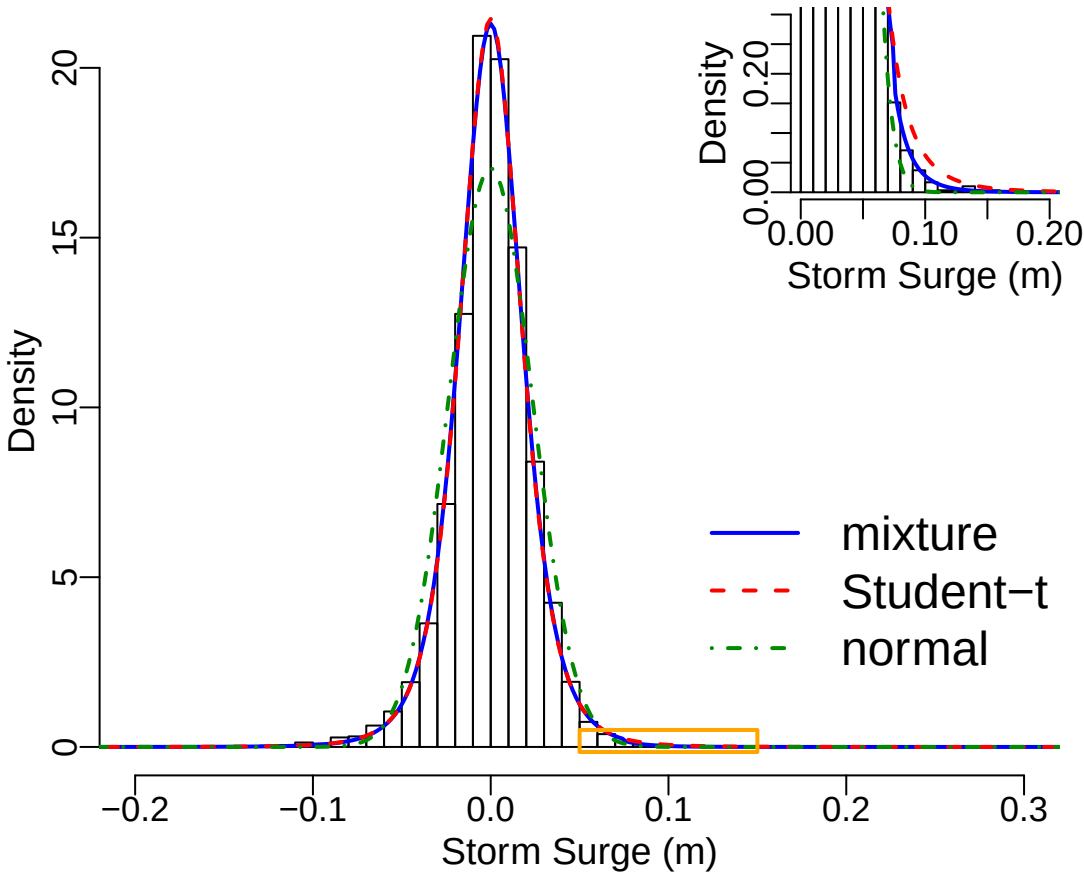
Figure 8. Goodness of fit for Bayesian posterior mean estimates of generalized Pareto, Student-t, and negative binomial models at the eight gauged locations in January. For the GPD and Student-t distributions, goodness of fit is shown with probability plots against the surge peak exceedances and hourly surge values, respectively. For the negative binomial, goodness of fit is shown by overlaying the fitted density on the histogram of the number of surge peak events across years.

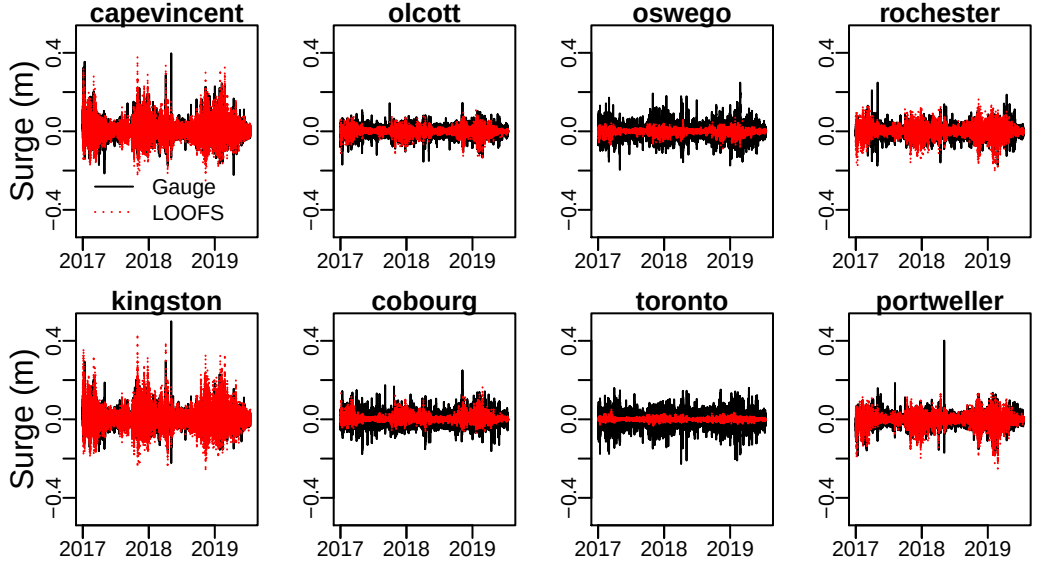
Figure 9. Distribution of p-values for Kolmogorov-Smirnov tests - applied to Student-t (bulk) and GPD (POT) fits, and Chi Square tests - applied to the negative binomial (POT) fit. The distribution of p-values is shown for Bayesian posterior mean estimates across all eight gauged sites for each month.

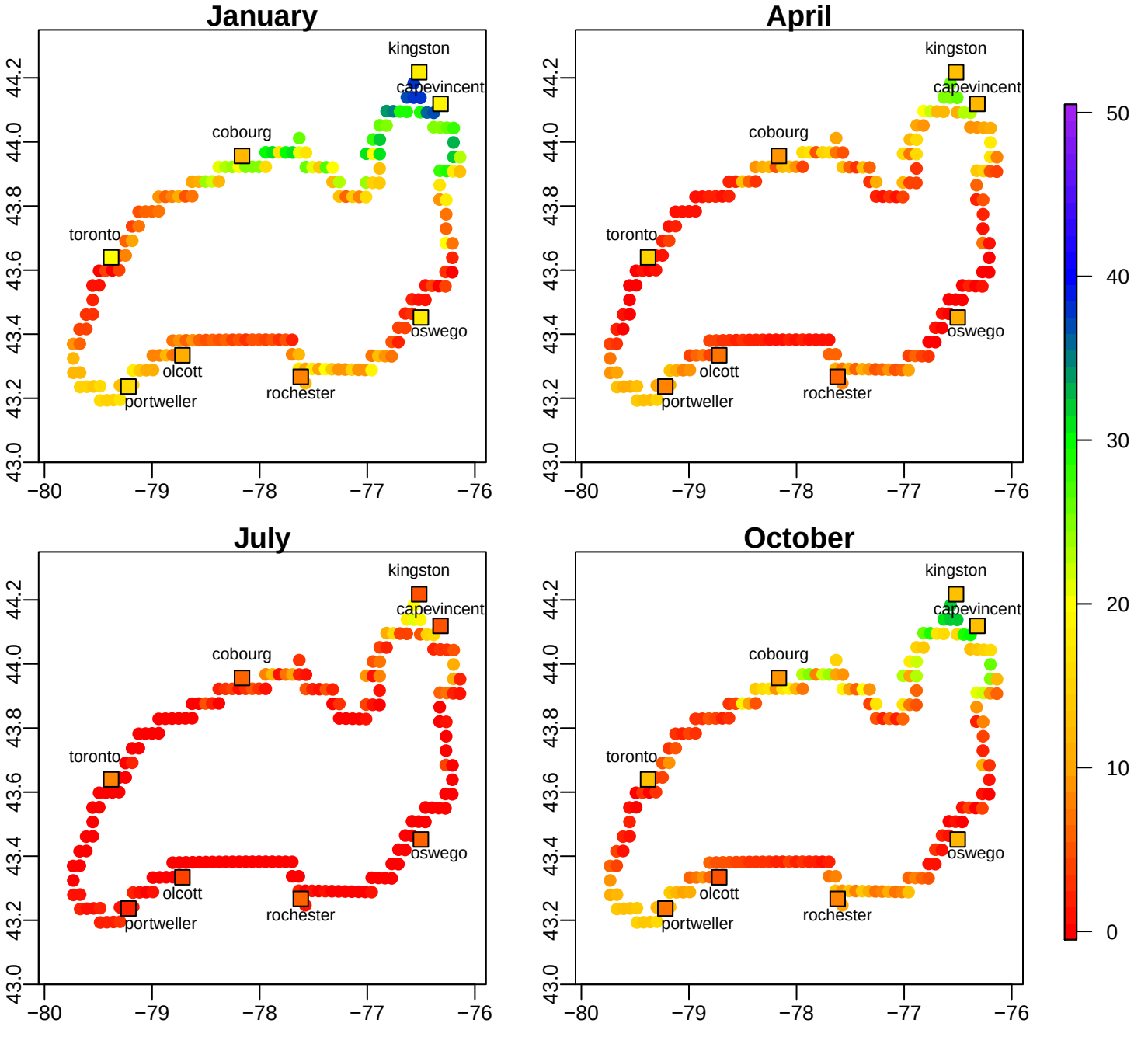
Figure 10. Goodness of fit for out-of-sample Bayesian posterior mean estimates of the generalized Pareto model at Oswego NY. Goodness of fit is shown with probability plots against the surge peak exceedances, and is compared against the goodness of fit based on (maximum likelihood) estimated parameters at the two nearest gauges (Cape Vincent and Rochester).

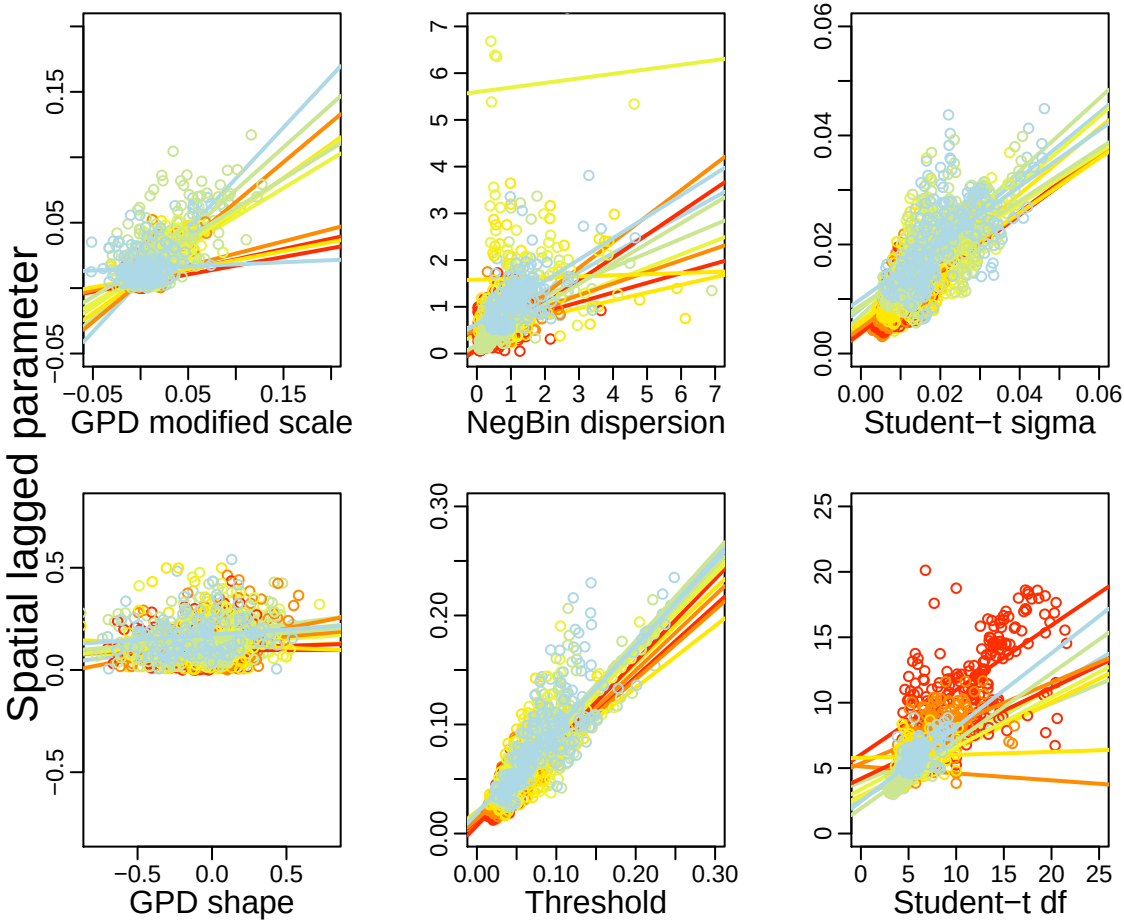
Figure 11. Return level plot of total water levels for a) Sodus Point, NY and b) Prince Edward County, ON Canada. Median estimates (solid) and 95% uncertainty intervals (dashed) are shown for models that account for storm surge using a mixture of Student-t and GPD distributions, and using only a Student-t distribution. c) The difference between median return levels at Sodus Point, NY and Prince Edward County, ON Canada for both the mixture and Student-t only models.

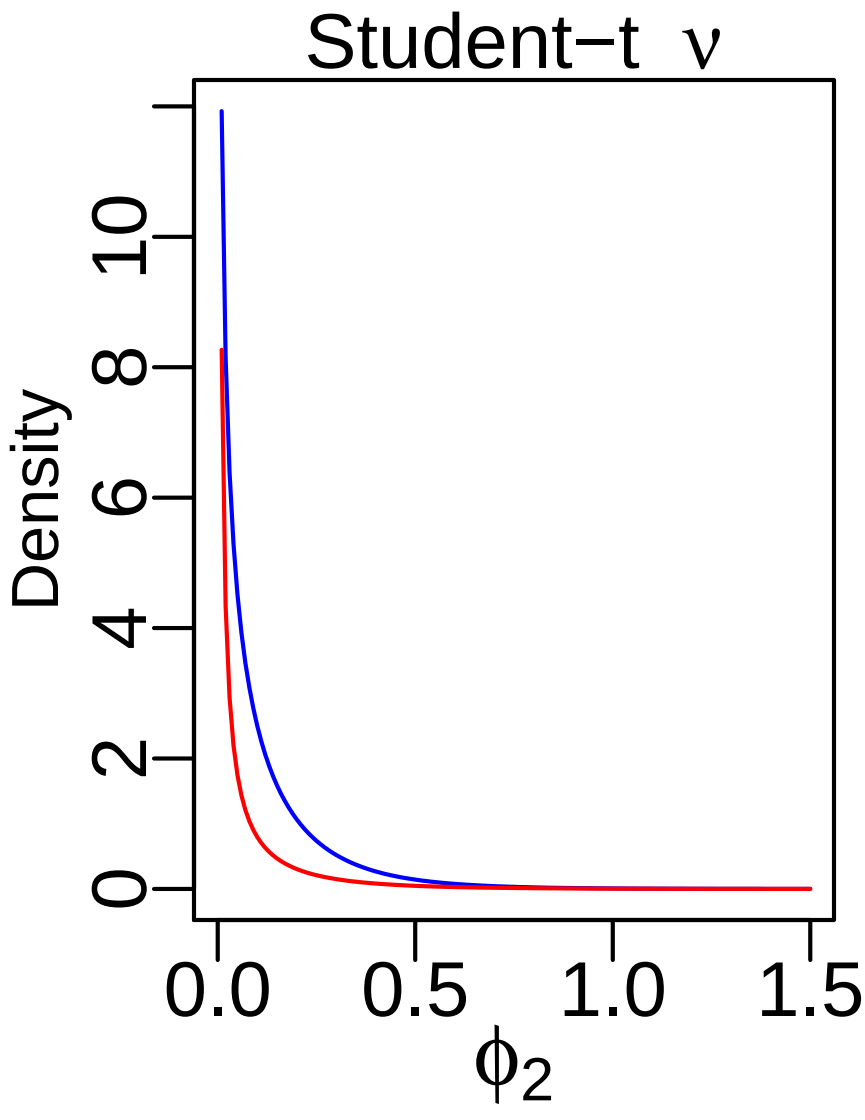

Hourly, LOOFS-based surge (section 2.2)
Prior distributions for spatial autocorrelation (ϕ_2) (section 3.2.2)
Stochastic time series of quarter-monthly static levels (section 2.3)
Hourly, gauged-based surge (section 2.1)
Bayesian model of storm surge (section 3.2.1)
Design event estimation of total water levels (section 3.3)

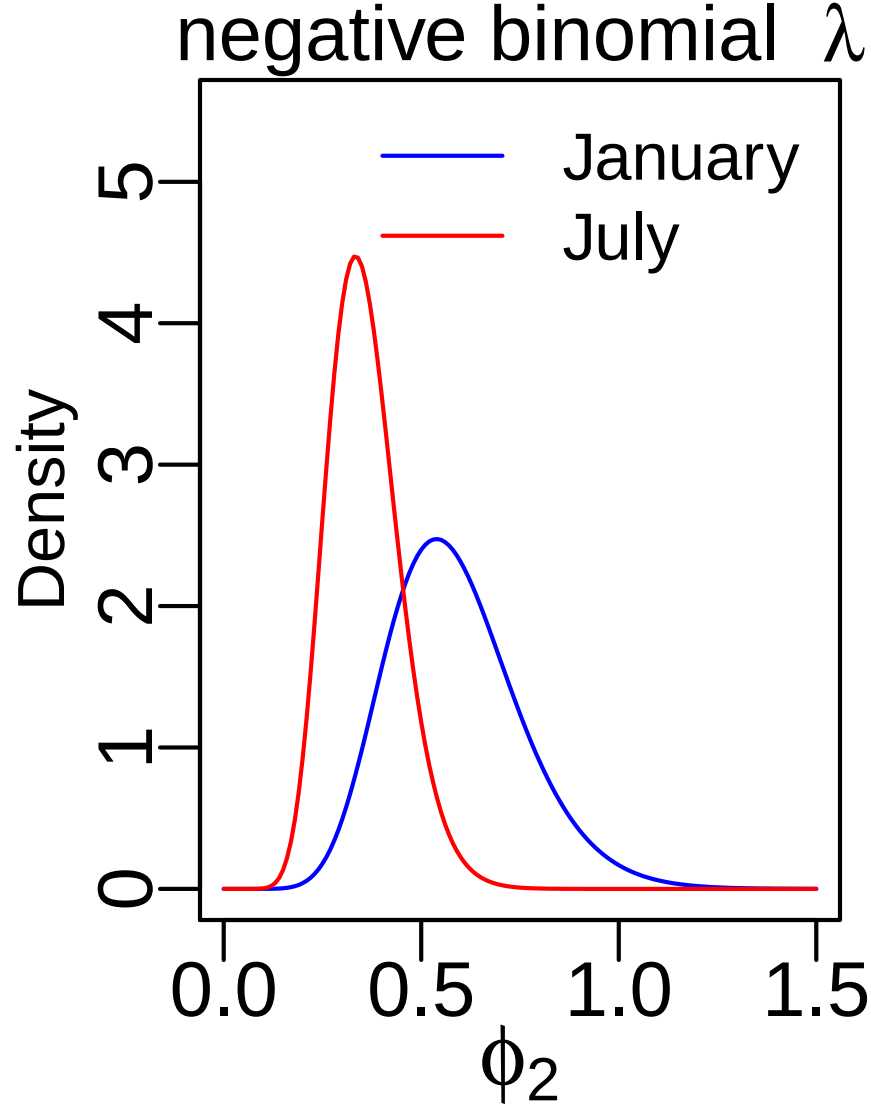

Density
0
5
10
15
20
−0.2
−0.1
0.0
0.1
0.2
0.3
Storm Surge (m)
mixture
Student−t
normal
Density
0.00
0.20
0.00
0.10
0.20
Storm Surge (m)

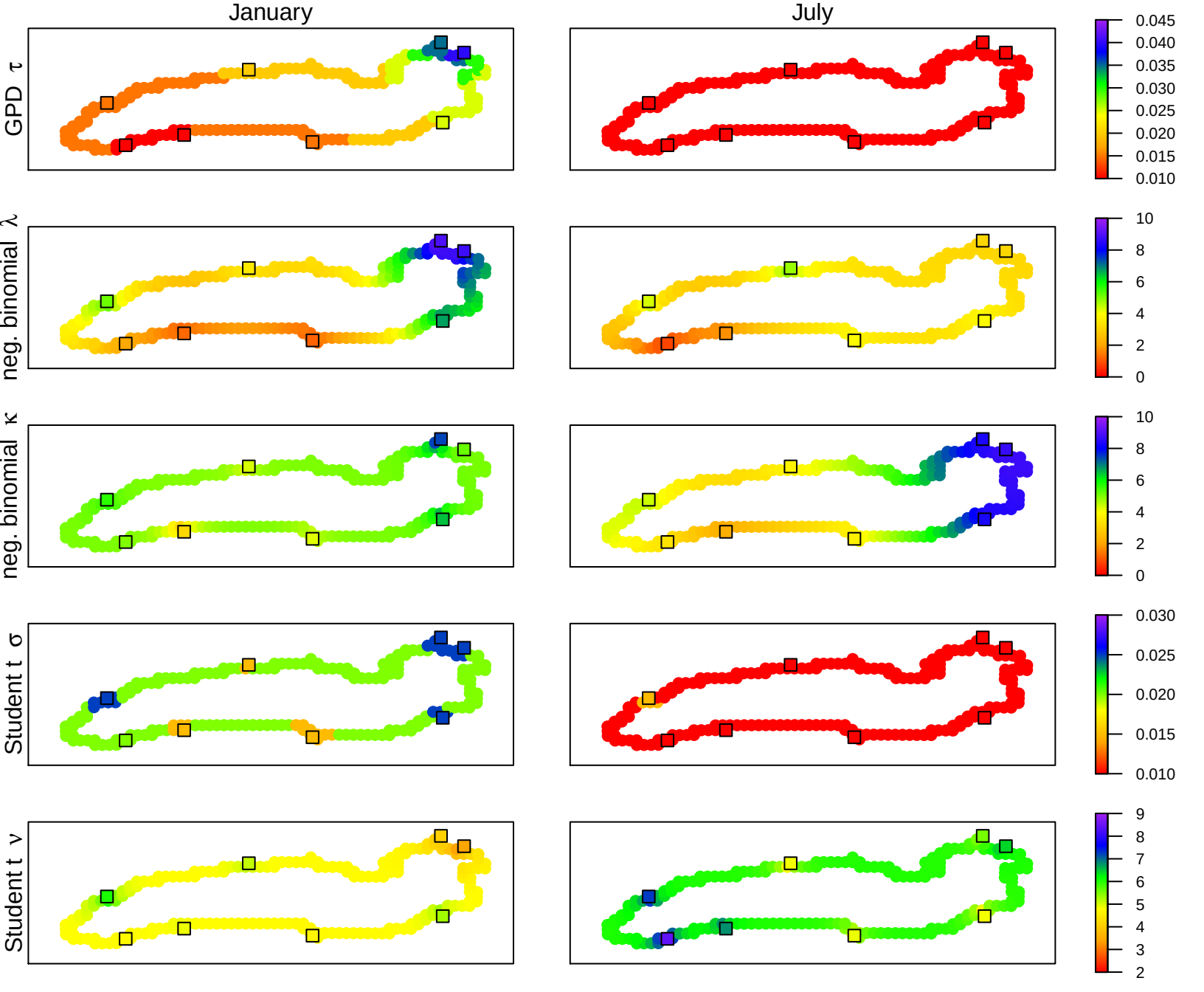

capevincent
olcott
oswego
rochester
kingston
cobourg
toronto
portweller
Surge (m)
Surge (m)
Gauge
LOOFS
0.4
0.0
−0.4
2017
2018
2019

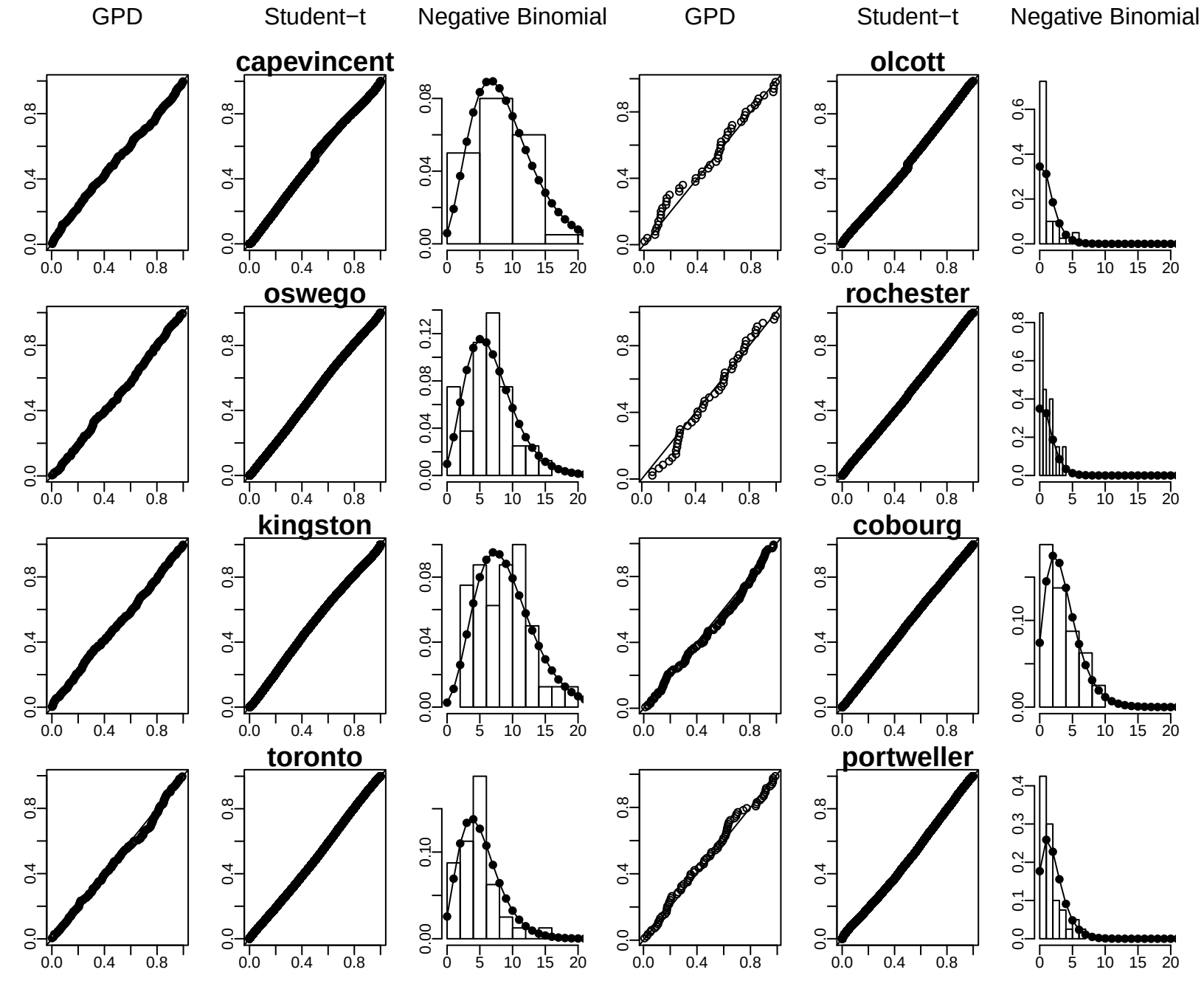

January
April
July
October
kingston
capevincent
cobourg
toronto
oswego
olcott
portweller
rochester
44.2
44.0
43.8
43.6
43.4
43.2
43.0
−80
−79
−78
−77
−76
50
40
30
20
10
0

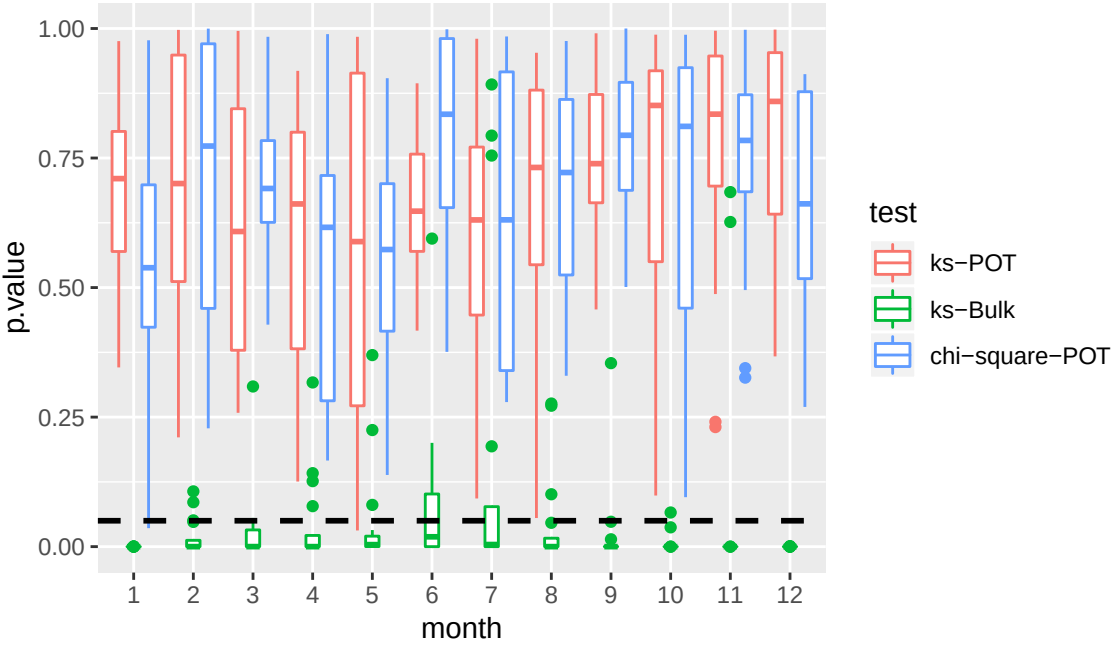

Spatial lagged parameter
GPD modified scale
NegBin dispersion
Student−t sigma
GPD shape
Threshold
Student−t df

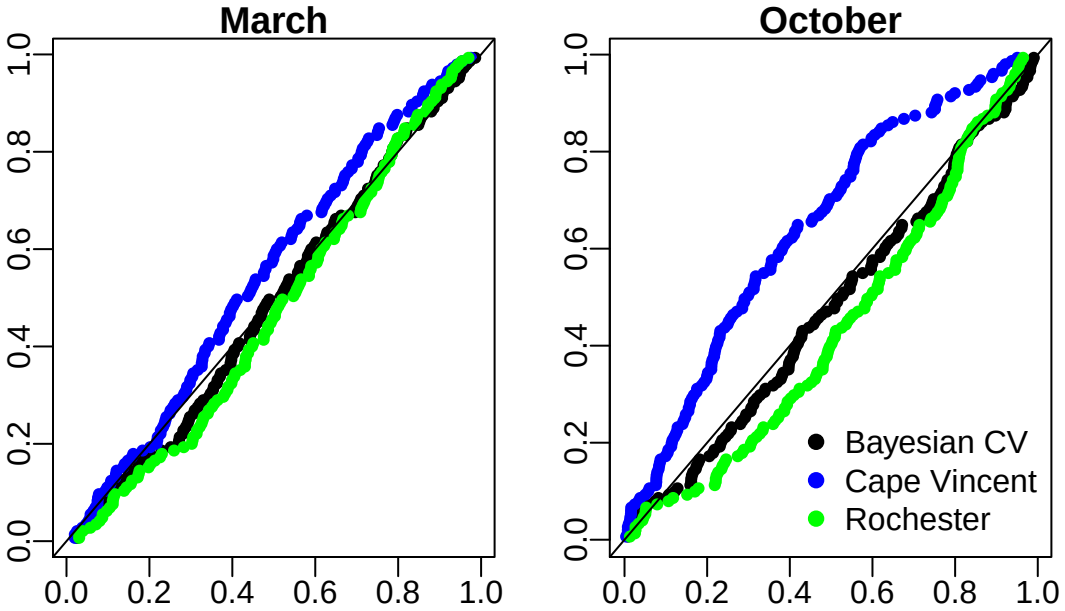

Student−t ν
Density
0
2
4
6
8
10
0.0
0.5
1.0
1.5
ϕ2

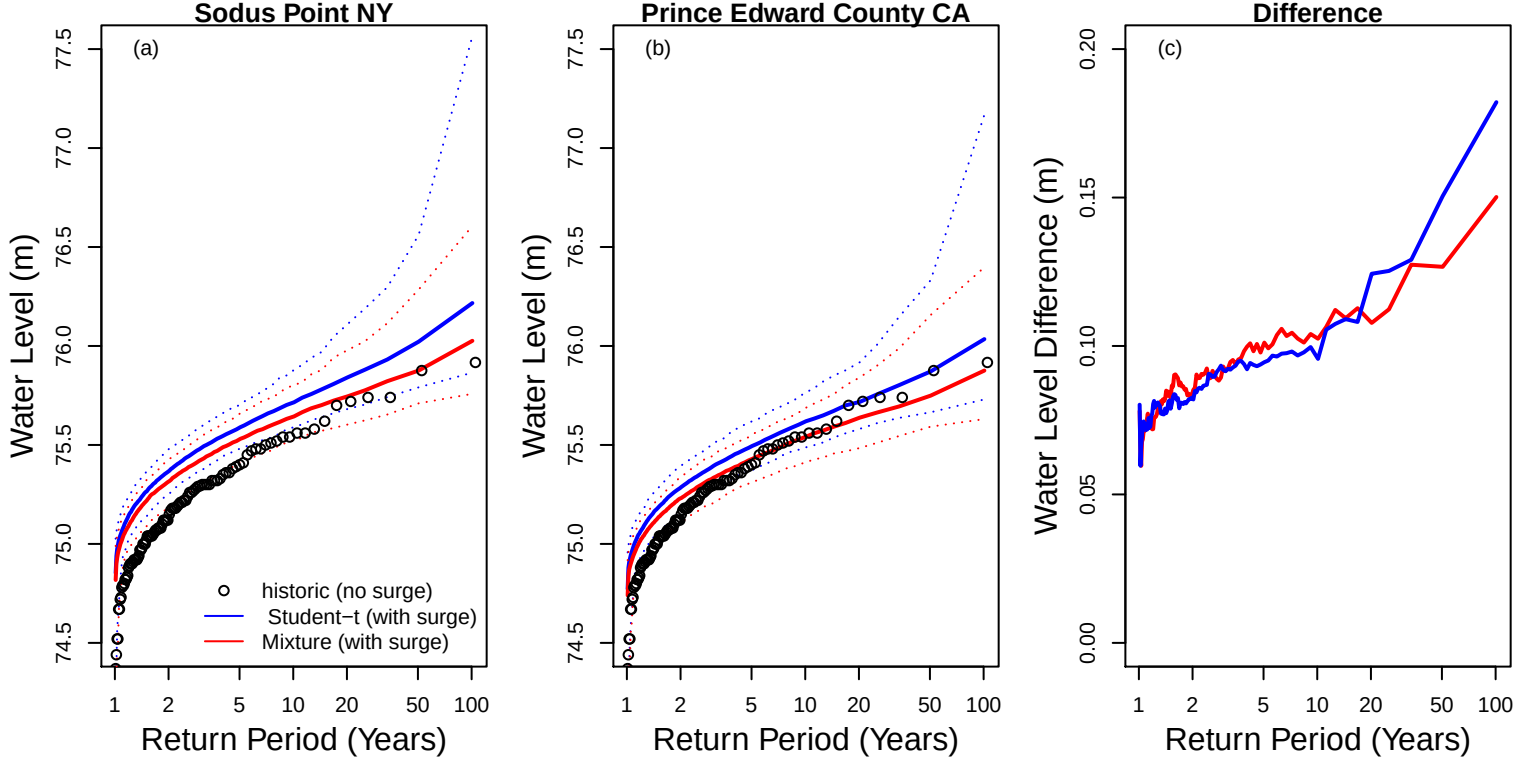

negative binomial λ
January
July
Density
0
1
2
3
4
5
0.0
0.5
1.0
1.5
ϕ2


January
July
GPD τ
neg. binomial λ
neg. binomial κ
Student t σ
Student t ν
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
10
8
6
4
2
0
10
8
6
4
2
0
0.030
0.025
0.020
0.015
0.010
9
8
7
6
5
4
3
2


GPD
Student−t
Negative Binomial
GPD
Student−t
Negative Binomial
capevincent
olcott
oswego
rochester
kingston
cobourg
toronto
portweller


1.00
0.75
0.50
0.25
0.00
p.value
1
2
3
4
5
6
7
8
9
10
11
12
month
test
ks−POT
ks−Bulk
chi−square−POT


March
October
0.0
0.2
0.4
0.6
0.8
1.0
Bayesian CV
Cape Vincent
Rochester


Sodus Point NY
(a)
Water Level (m)
74.5
75.0
75.5
76.0
76.5
77.0
77.5
historic (no surge)
Student−t (with surge)
Mixture (with surge)
1
2
5
10
20
50
100
Return Period (Years)
Prince Edward County CA
(b)
Water Level (m)
74.5
75.0
75.5
76.0
76.5
77.0
77.5
1
2
5
10
20
50
100
Return Period (Years)
Difference
(c)
Water Level Difference (m)
0.00
0.05
0.10
0.15
0.20
1
2
5
10
20
50
100
Return Period (Years)