

Bayesian Hierarchical Modeling of Sea Level Extremes

Marko Laine^{1,1}, Olle Rätty^{1,1}, Jani Särkkä^{1,1}, Ulpu Leijala^{1,1}, and Milla Johansson^{1,1}

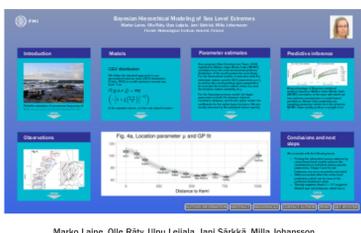
¹Finnish Meteorological Institute

November 30, 2022

Abstract

Reliable estimates of occurrence probabilities of sea level extremes are required in coastal planning (e.g. design floods) and to mitigate risks related to flooding. Probabilities of specific extreme events have been traditionally estimated from the observed extremes independently at each tide gauge location. However, this approach has shortcomings. Firstly, sea level observations often cover a relatively short historical time period and thus contain only a small number of extreme cases (e.g. annual maxima). This causes substantial uncertainties when estimating the distribution parameters. Secondly, exact information on sea level extremes between the tide gauge locations and incorporation of dependencies of adjacent stations is often lacking in the analysis. A partial remedy to these issues is to exploit spatial dependencies exhibited by the sea level extremes. These dependencies emerge from the fact that sea level variations are often driven by the same physical and dynamical factors at the neighboring stations. Bayesian hierarchical modeling offers a way to model these dependencies. The model structure allows to share information on sea level extremes between the neighboring stations and also provides a natural way to represent modeling uncertainties. In this study, we use Bayesian hierarchical modeling to estimate return levels of annual sea level maximum in the Finnish coastal region, located along the north-east part of the Baltic Sea. As annual maxima are studied, we use the generalized extreme value (GEV) distribution as the basis of our model. To tailor the model specifically for the target region, spatial dependencies are modeled using physical covariates which reflect the distinct geometry of the Baltic Sea. We illustrate the added value of the hierarchical model in comparison to the traditional one using the available long-term tide gauge time series in Finland. Careful analysis of the sources of uncertainties is necessary when extrapolating the return level estimates into the future. This work is a part of project PREDICT (Predicting extreme weather and sea level for nuclear power plant safety) that supports nuclear power plant safety in Finland.

Bayesian Hierarchical Modeling of Sea Level Extremes



Marko Laine, Olli Rätty, Ulpu Leijala, Jani Särkkä, Milja Johansson

Finnish Meteorological Institute, Helsinki, Finland



PRESENTED AT



INTRODUCTION



Photo credit: Jarmo Vähäkallio

Reliable estimation of occurrence frequency of high consequence sea level extremes is very important to support coastal planning and safety in all coastal regions of the world. The Baltic Sea is a semi-enclosed shallow sea connected to the Atlantic Ocean via narrow Danish straits. Finnish coastal region – the focus region of this study – is located in the northeastern Baltic Sea (Figure 1). Three factors affect the sea surface level on the Finnish coast:

1. short-term sea level variations (storm surges, internal oscillations in the Baltic Sea, tides, and meteorological),
2. long-term mean sea level change (global mean sea level, the postglacial land uplift and the Baltic Sea water balance), and
3. wind-induced waves.

In this study, the focus is on short-term sea level variations. The main aims of this study are:

1. to provide updated return level estimates for annual sea level maxima along the Finnish coastal region with Bayesian hierarchical modeling and
2. to better quantify uncertainties related to the return level estimates.

According to recent studies, it is forecasted that the flooding risks will increase during the 21st century at all places on the Finnish coast except in the Bothnian Bay (Pelkkä et al., 2018).

Background

FMI has studied extreme weather, climate and sea level events potentially posing risks to Nuclear Power Plants since 2007 (Ojala et al., 2018), and presently in the PREDICT project within the SAFIR2022 program.

MODELS

GEV distribution

We follow the standard approach to use generalized extreme value (GEV) distribution (Coles, 2001) to model maximum annual sea-level Y as

$$G(\mu, \sigma, \xi) = \exp\left(-\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)^{-\xi}\right]^{-1/\xi}\right)$$

In the equation above, μ is the real-valued location parameter, $\sigma > 0$ is the scale parameter and ξ denotes the shape parameter. The tail behavior of GEV distributions is strongly controlled by the shape parameter: If $\xi < 0$, there is a bounded upper tail, while in other cases the tail is unbounded and has either exponential ($\xi = 0$) or polynomial ($\xi > 0$) decay.

The quantile distribution for the annual maximum sea level is obtained by inverting the GEV equation:

$$y_p = \begin{cases} \mu - \frac{\sigma}{\xi} \left[1 - \{-\log[1-p]\}^{-\xi}\right], & \text{for } \xi \neq 0, \\ \mu - \sigma \log[1 - \log[1-p]], & \text{for } \xi = 0. \end{cases}$$

This equation gives the return level y_p associated with occurrence probability p or conversely the return period $1/p$. There are several ways to estimate the associated distribution parameters μ, σ, ξ . Here, we take Bayesian approach to make inferences on the posterior predictive distribution (PPD) for maximum annual sea-level. As we are particularly interested in reliably quantifying uncertainties in the predicted return levels, Bayesian approach provides a natural way to do this.

Individual GEV model fits to the gauge locations were used as the baseline. In addition to modeling Y separately for each tide gauge $Y_i \sim \text{GEV}(\mu_i, \sigma_i, \xi_i)$, we experimented with four different hierarchical model formulations, where the station-specific parameter values are tied together with hyper distributions. The models used are:

1. The individual values for all three GEV parameters are coming from a joint Gaussian hyper distribution, whose parameters are estimated from data.
2. Individual location and scale parameters μ and σ depend linearly on the distance to the Bothnian Bay (linear).
3. Individual location and scale parameters are modeled with B-splines using distance as the covariate (spline).
4. Individual location and scale are modeled using Gaussian processes with squared exponential covariance function depending on the distance (GP).

The distribution parameters were estimated with Stan program and the rstan and rstanif R packages.

PARAMETER ESTIMATES

Stan program (Stan Development Team, 2020) implements Markov chain Monte Carlo (MCMC) simulation from the multi-dimensional posterior distribution of the model parameters' uncertainties. For the hierarchical models, it estimates both the individual, station-specific GEV parameters μ, σ, ξ , as well as the corresponding hyper-parameters, for example the between station mean of μ and the between station variability of μ .

For the Gaussian process model, the hyper parameters include the between station correlation distance, and for the spline model the coefficients for the spline basis functions. We are mostly interested in the individual station-specific parameters as well as using the estimated model to predict future return levels.



Figure 2. Violin plots showing the GEV parameter estimates and the corresponding uncertainty separately for all modeled models in the 40-year baseline. The colored uncertainty obtained using MCMC.

All the four selected hierarchical models provided quite similar results. This is illustrated in Figure 2, which shows the model estimates of parameter values for each tide gauge. Overall, it can be seen that compared to the individual estimates without hierarchy, all the hierarchical models were able to borrow information from neighboring stations and thus, provided more informative parameter distributions. Especially the shape parameter ξ seems to be quite constant over the study area, and it makes sense to combine this information. Interestingly, ξ seems to estimate mostly around -0.2 to 0 , which suggests a Weibull type tail behavior for the sea level maxima. This means that one could use an upper tail of $\sim \exp(-x^2)$ as an estimate (with attached uncertainty) of the maximum attainable sea level height at any given location. An example distribution for the upper limit is given for the Kemi tide gauge in Figure 3, as this gauge has one of the highest observed maximum annual sea-level heights in Finland (see Figure 1).

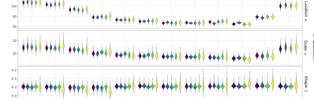


Figure 3. Upper water level limit for Kemi

Moreover, both the Gaussian process and the spline model can predict model parameters for between station locations, too, as they depend on the distance from Bothnian Bay. This is with respect to the distance to Kemi (see Figure 4) in Figures 4 and 5a. As an additional note, Figure 5a shows that the spline model does not capture the mean parameter as well as the Gaussian process model. This is due to the fact that the number of knots was fixed to a small number (5) in order to avoid overfitting. Penalized B-splines could probably better capture these mean parameters, but this approach was not tested here due to the good performance of the Gaussian process model.

PREDICTIVE INFERENCE



Figure 4. Displays predictive analyses done with station specific separate models as well as with hierarchical models, which again shows the advantage from pooling the information.

More specific examples are given in Figures 7 and 8, which shows the 50-year and 1000-year return level estimates simultaneously for all used models. The most notable feature in these figures is the large predictive uncertainty for separate models for most tide gauges.

OBSERVATIONS

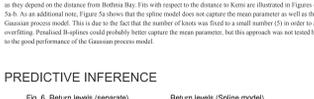


Figure 1. 4-hourly observations were collected from 12 tide gauges located along the Finnish coast (Figure 1). The length of the available time series varies between 127 (Kemi) years and thus, this data provide a good basis for extreme value analysis.

The observations have been converted to the height system N2000 from fixed tide-gauge-specific reference systems (Chambers et al., 2009). Long-term mean sea level trend was removed from each time series before calculating the annual maximum sea level heights, which were then used in statistical analysis based on three parameter generalized extreme value distribution (GEV).

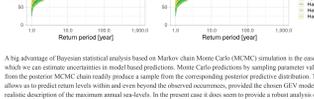


Figure 2. Violin plots showing the GEV parameter estimates and the corresponding uncertainty separately for all modeled models in the 40-year baseline. The colored uncertainty obtained using MCMC.

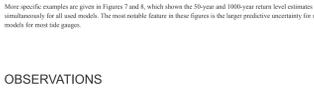


Figure 3. Upper water level limit for Kemi

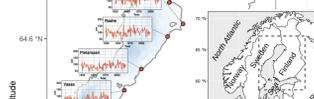


Figure 4a. Location parameter μ and GP fit

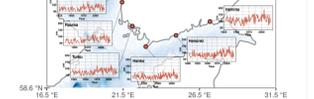


Figure 4b. Shape parameter σ and GP fit

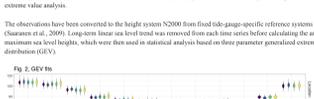


Figure 5a. Location parameter μ and spline fit

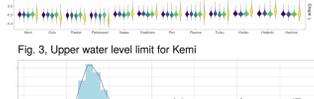


Figure 5b. shape parameter σ and spline fit



Figure 7. 50-year return level



Figure 8. 1000-year return level

CONCLUSIONS AND NEXT STEPS

We conclude with the following items:

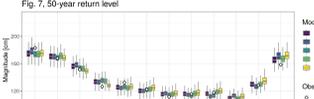
- Pooling the information across stations by using hierarchical models reduces the uncertainties in individual station-specific parameters. Spline and GP tail behaviors were more accurately captured.
- Different models affect the return level predictions which can be seen in the predictive distribution plots.
- Sharply negative slope $\xi \sim -0.2$ suggests Weibull type tail behaviour, which has a hard upper limit at $\mu - \sigma\xi$.
- The Gaussian process or spline hierarchy allows estimating water levels at non-gauged locations, too.

Next steps in our project involves:

- Study the temporal stationarity of the parameters.
- Study the effect of North Atlantic Oscillation (NAO).
- Take into account the changing climate.
- Compare maximum sea level heights to physics based analysis of the Baltic Sea.

Acknowledgements

This poster presents preliminary results of the project PREDICT (predicting extreme weather and sea level for nuclear power plant safety) that is funded by the State Nuclear Waste Management Fund through SAFIR2022, the Finnish Nuclear Power Plant Safety Research Programme 2019-2022, and from the Finnish Meteorological Institute (FMI). PREDICT aims to develop and maintain research expertise and methods needed for assessing probabilities of occurrence of safety-relevant single and compound extreme events.



AUTHOR INFORMATION

All the authors work at Finnish Meteorological Institute, Helsinki, Finland. Marko Laine and Olli Rätty in Meteorological Research Unit, Ulpu Leijala, Jani Särkkä, and Milja Johansson in Marine Research Unit. You can contact the authors by email using firstname.lastname@fmi.fi.

ABSTRACT

Reliable estimates of occurrence probabilities of sea level extremes are required in coastal planning (e.g. design floods) and to mitigate risks related to flooding. Probabilities of specific extreme events have been traditionally estimated from the observed extremes independently at each tide gauge location. However, this approach has shortcomings. First, sea level observations offer over a relatively short historical time period and thus contain only a small number of extreme events (e.g. annual maxima). This causes substantial uncertainties when estimating the distribution parameters. Secondly, even information on sea level extremes between the tide gauge locations and incorporation of dependencies of adjacent stations is often lacking in the analysis.

A partial remedy to these issues is to exploit spatial dependencies exhibited by the sea level extremes. These dependencies emerge from the fact that sea level variations are often driven by the same physical and dynamical factors at the neighboring stations. Bayesian hierarchical modeling offers a way to model these dependencies. The modeler encodes a priori their information on sea level extremes between the neighboring stations and of annual sea level maxima in a natural way to represent modeling uncertainties.

In this study, we use Bayesian hierarchical modeling to estimate return levels of annual sea level maximum in the Finnish coastal region, located along the north-east part of the Baltic Sea. As annual maxima are modeled, we use the generalized extreme value (GEV) distribution as the basis of our model. To tailor the model specifically for the target region, spatial dependencies are modeled using generalized extreme values which reflect the distinct geometry of the Baltic Sea. We illustrate the added value of the hierarchical model in comparison to the traditional one using the available long-term tide gauge time series in Finland. Careful analysis of the sources of uncertainties is necessary when incorporating the return level estimates into the future.

This work is a part of project PREDICT (Predicting extreme weather and sea level for nuclear power plant safety) that supports nuclear power plant safety in Finland.

REFERENCES

Coles, S., Bawa, J., Tereza, L., and DeGaazi, P. An introduction to statistical modeling of extreme values, vol. 208, Springer, 2001.

Jylhä, K., Kärräinen, M., Junttila, C., Gergov, H., Hakola, J., Hyytiäinen, O., Johansson, M., Korppi, A., Korpien, A., Koronen, R., Koronen, E., Leppä, I., Mäkelä, A., Pelkkä, H., Saha, J., Sanberg, J., Savolainen, M., Vuola, A., Vuorimies, A., and Vign, E. Recent technological and marine studies to support nuclear power plant safety in Finland. Energy, 165, 108-111, <https://doi.org/10.1016/j.energy.2018.09.033>, 2018.

Pelkkä, H., Leijala, U., Johansson, M., Laitinen, K., Kahra, K. K., 2018. Future probabilities of coastal floods in Finland. Continental Shelf Research, 157, 32-42. DOI: 10.1016/j.csr.2018.02.006.

Saaremaa, V., Lehmiäinen, P., Pöyhönen, P., Takala, M., Mikkelä, L., and Penttinen, M. The new Finnish height reference N2000. In: Geomatics Reference Frames, pp. 265-280, Springer, 2009.

Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, version 2.25. 2020. <https://mc-stan.org>

Stan Development Team. RStan: the R interface to Stan. R package version 2.21.2, 2020. <https://mc-stan.org>