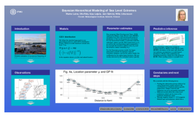
Bayesian Hierarchical Modeling of Sea Level Extremes



Marko Laine, Olle Räty, Ulpu Leijala, Jani Särkkä, Milla Johansson

Finnish Meteorological Institute, Helsinki, Finland

presented at

AGU FALL MEETING
Global Encounters | 12-17 December 2016

INTRODUCTION

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 through the North Sea and the Danish straits. The Finnish coastline of the Baltic Sea is located in the northeastern Baltic Sea (Figure 1). There 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 meteorosensitivity)
2. long-term mean sea level change (global mean sea level, the postglacial land uplift and the Baltic Sea water balance)
3. wind-induced waves

In this study, the focus is on short-lasting 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 modelling and
2. to better quantify uncertainties related to the return level estimates

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

Background

FMI has studied extreme weather, climate and sea level events perennially putting risks in Nuclear Power Plants since 2007 (Jylhä et al., 2014), and presently in the PREDICT project within the SAFIR2022 program.

MODELS

GEV distribution

The basic idea in the method approach to use generalized extreme value (GEV) distribution (Coles, 2001) to model maximum annual sea level Y as

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

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 distribution is strongly controlled by the shape parameter. If $\xi > 0$, y has a bounded upper tail, while in other cases the tail is unbounded and has either exponential ($\xi = 0$) or heavy ($\xi > 0$) decay.

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

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

This equation gives the return level y_p associated with occurrence probability p or, conversely, the return period $1/p$. There is a need to estimate uncertainties related to the return level estimates. The GEV model is fitted to the data to provide point estimates or the posterior predictive distribution $f(y|\theta)$ for maximum annual sea level. As we are particularly interested in reliably quantifying uncertainty in the predicted return levels, Bayesian approach provides a natural way to do this.

Hierarchical model

Individual GEV model fits to the gauge locations were used as the baseline. In addition to modeling Y separately for each tide gauge $Y_i = GEV(\mu_i, \sigma_i, \xi_i)$, 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 (common).
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 rgdal R packages.

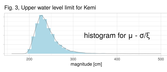
PARAMETER ESTIMATES

Stan program (Stan Development Team, 2020) implements Markov chain Monte Carlo (MCMC) simulations from the multi dimensional posterior distribution of the model parameter uncertainty. For the hierarchical models, it estimates both the individual, station-specific GEV parameters μ_i, σ_i 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 stations correlation distance, and for the spline model, the spline coefficients and the knot locations. We were mainly interested in the individual station specific parameters as well as using the estimated model to predict future return levels.

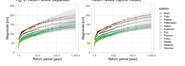
Figure 2. Violin plots showing the GEV parameter estimates and the corresponding uncertainty represents for all statistical models in the sub group Gaussian. The related uncertainty obtained with MCMC.

All the four selected hierarchical models provided quite similar results. This is illustrated in Figure 2, which shows the posterior distributions of the parameters μ, σ and ξ for the Helsinki tide gauge. Because of the relatively strong hierarchical tendency, all the hierarchical models were able to borrow information from neighbouring stations and they provided more informative posterior distributions. Especially the shape parameter ξ seems to be quite constant over the study area, which is also a physically reasonable result. The posterior distributions of ξ are mostly negative, which is in line with the fact that e.g. Weibull type tail behaviour fits the water level maxima. This means that we could use an upper limit of $\mu - \sigma/\xi$ for the estimates (with attached uncertainties) 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 (Vuosaari Tuuli 2011).



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 Bothnia Bay. For with respect to the distance in Kemi are illustrated in Figure 4a-b and 5a-b. The Gaussian process model (Figure 4a-b) is more flexible and can model more complex relationships than the spline or linear 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 the mean parameter, but this approach was not tested here due to the good performance of the Gaussian process model.

PREDICTIVE INFERENCE

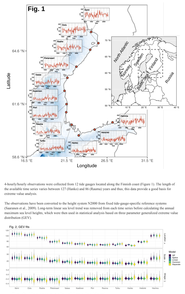


A big advantage of Bayesian statistical analysis based on Markov chain Monte Carlo (MCMC) simulations is the ease with which we can estimate uncertainties in model based predictions. Monte Carlo predictions by sampling parameter values from the posterior MCMC chain readily produce a sample from the corresponding posterior predictive distribution. This sample can be used to form the 95% credible intervals for the return level estimates as is done in Figure 6. Figure 6 is also a realistic description of the maximum annual sea levels. In the present case it does not to provide a robust analysis of the available observations, as even the estimated return levels with very small probabilities should be well heard, at least in the current climate, and they can be used as a part of risk analysis in planning coastal buildings etc.

Figure 6 displays predictive analysis done with station specific separate models as well as with hierarchical models, which quote shows the advantage from pooling the information.

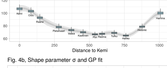
More specific examples are given in Figures 7 and 8, which shows the 50-year and 1000-year return level estimates simultaneously for all used models. The most notable feature in these figures is the larger predictive uncertainty for separate models for most tide gauges.

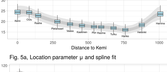
OBSERVATIONS

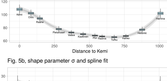


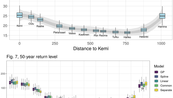
A hourly-hourly observations were collected from 13 tide gauges located along the Finnish coast (Figure 1). The length of the available time series varies between 127 (Hanko) and 90 (Kemi) years and thus, this data provide a good basis for extreme value analysis.

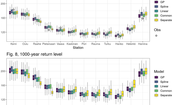
The observations have been converted to the height system N2000 from local tide gauge-specific reference system (Vuosaari Tuuli 2011). The annual maxima were then extracted from the data, and they were used to fit the annual maximum sea level heights, which were then used in statistical analysis based on their parameter generalized extreme value distribution (GEV).











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. Shape ξ and the tail behaviour are more accurately estimated.
- Different models affect the return level predictions which can be seen in the predictive distribution plots.
- Sharply negative shape ($\xi < -0.2$) 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 in physics-based analysis of the Baltic Sea.

Acknowledgements

The poster presents preliminary results of the project PREDICT (producing extreme weather and sea level for nuclear power plant safety) of the SAFIR2022 program. The SAFIR2022 is the 3rd Safety Research Program of Finland (2019-2022). The 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 Olle Räty in Meteorological Research Unit, Ulpu Leijala, Jani Särkkä, and Milla Johansson in Marine Research Unit. You can contact the authors by email using firstname.lastname@fmi.fi.

ABSTRACT

Reliable estimation of occurrence probabilities of sea level extremes are important in coastal planning (e.g. design floods) and to support risk analysis (e.g. flooding). Probabilities of specific extreme events have been traditionally estimated from the observed maxima 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 differences between the tide gauge locations and interpretation of dipole- theories of adjacent stations is often lacking in the analysis.

A partial remedy in these issues is to exploit spatial dependencies exhibited by the sea level extremes. These dependencies origin from meteorological forcing, which is similar for adjacent locations, and from the mutual interaction between 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. We model annual sea level maxima Y with the GEV distribution, and the model parameters are tied together with common value (GEV) distribution at 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 benefits of the hierarchical modeling approach by comparing the results to models where the tide gauge locations are modeled as fully independent. 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.

REFERENCES

Introduction photo: Jarmo Vehkalahti

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

Jylhä, K., Kauppinen, H., Fortelius, C., Gregow, H., Hyttinen, O., Johansson, M., Kauppinen, A., Kaunisto, K., Koistinen, K., Kuitunen, E., Leijala, U., Mäkelä, A., Pellikka, H., Saku, S., Sandberg, J., Sarlin, M., Tapio, A., Venäläinen, A., and Vita, E.: Recent meteorological and marine studies to support nuclear power plant safety in Finland, Energy, 145, 1102–1119, https://doi.org/https://doi.org/10.1016/j.energy.2018.01.053, 2018.

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

Stenman, F., Jokimäkeläinen, P., Poutiainen, P., Tabari, H., Mikkonen, J., and Pusenius, M.: The new Finnish height reference N2000, in: Geodetic Reference Frames, pp. 295–302, Springer, 2009.

Stan Development Team: Stan Modeling Language User's Guide and Reference Manual, version 2.21. 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/