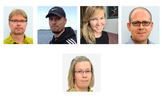
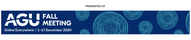
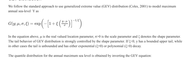
Bayesian Hierarchical Modeling of Sea Level Extremes

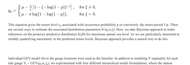
Marko Laine, Olle Räty, Ulpu Leijala, Jani Särkkä, Milla Johansson

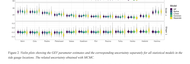
Finnish Meteorological Institute, Helsinki, Finland

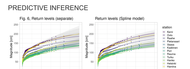


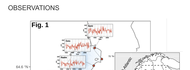


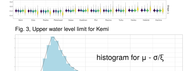


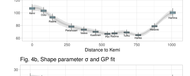


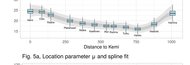


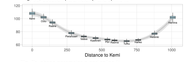


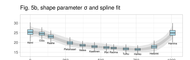


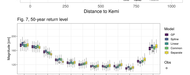


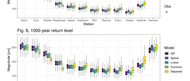












presented at

AGU FALL MEETING
Virtual, Everywhere | 1–17 December 2020

INTRODUCTION

Photo credit: Jenni Attalmaa

Reliable estimation of occurrence frequency of high-consequence sea level extremes is very important in region-wised 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). There become effect the sea surface level on the Finnish coast:

1. short-term sea level variations (storm surges, internal oscillations in the Baltic Sea, tides, and instrumentation);
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-lasting sea level variations. The main aims of this study are

- to provide updated storm level estimates for annual sea level maximum along the Finnish coastal region with Bayesian hierarchical modeling and
- to further 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 potentially posing risks to Nuclear Power Plants since 2007 (Jylhävä et al., 2010), and presently in the PROTECT 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

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

In the equation above, μ is the real valued location parameter, $\sigma > 0$ is the scale parameter and ξ denotes the shape parameter so that $1 + \xi(y - \mu)/\sigma > 0$. The shape parameter governs the tail behavior so that if $\xi < 0$ then y has 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

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

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 $\theta = (\mu, \xi, \sigma)$. Here, we take Bayesian approach to make inferences on the posterior predictive distribution $\tilde{y}|y$ 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.

Individual GEV models fit to the gauge locations were used as the baseline. In addition to modeling Y separately for each site, we fitted a hierarchical GEV model to the maximum annual sea level data from all gauging locations, 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 (Oulu)
3. Individual location and scale parameters are modelled 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 aphsim 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 parameter uncertainty. For the hierarchical models, it estimates both the individual station specific parameter values as well as the hyper values, 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 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.

Fig. 2. HPD fits

Figure 2. Violin plots showing the GEV parameter estimates and the corresponding uncertainty separately for all statistical models in the sub gauge locations. 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 model estimates of parameter values for each sub gauge. Overall, it can be said that compared to the individual estimates (Individual), the hierarchical models tend to shrink the estimates towards the overall mean (Common), and thus provide more information-pressive 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 nicely around -0.2 (±0.05), which suggests that the distribution of annual sea level maxima is bounded from above. The spline model (Spline) and the GP model (GP) estimate (with attached uncertainty) of the maximum attainable sea level height at any given location. For 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 7).

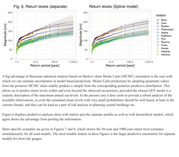
Fig. 3. Upper water level limit for Kemi

Moreover, both the Gaussian process and the spline model can predict model parameters for between station locations, too. This is illustrated in Figure 4, which shows the location parameter μ and GP fit (Figure 4a), and the shape parameter σ and GP fit (Figure 4b). 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. Increasing the number of knots would improve the mean parameter, but this approach was not tested here due to the good performance of the Gaussian process model.

PREDICTIVE INFERENCE

Fig. 6. Return levels (separate)

Fig. 6. Return levels (Spline model)

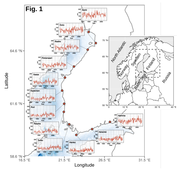


A big advantage of Bayesian statistical analysis based on Markov chain Monte Carlo (MCMC) simulation is the ease with which we can estimate uncertainties in model based predictands. Monte Carlo predictions by sampling parameters' values from the posterior MCMC chain readily produce predictive distributions for any combination of parameters. Figure 6 shows the posterior predictive return levels with 95% credible intervals for the corresponding return periods for the eleven GEV models fit to the realistic description of the maximum annual sea levels. In the present case it does seem to provide a robust analysis of the available observations, as even the estimated return levels with very small probabilities should be well bound, 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 return levels with station specific separate models as well as with hierarchical models, which again shows the advantage from pooling the information.

Above specific examples are given in Figures 3 and 4, 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 linearly bicubic observations were collected from 12 tide gauges located along the Finnish coast (Figure 1). The length of the observation records varies between 127 (Hanko) and 40 (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 (Saaranen et al., 2009). Long-term linear 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).

Fig. 3. HPD fits

Fig. 3. Upper water level limit for Kemi

Fig. 4a, Location parameter μ and GP fit

Fig. 4b, Shape parameter σ and GP fit

Fig. 5a, Location parameter μ and spline fit

Fig. 5b, shape parameter σ and spline fit

Fig. 7. 50-year return level

Fig. 8. 1000 year return level

CONCLUSIONS AND NEXT STEPS

We conclude with the following points:

- 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.
- Mostly negative shape ($\xi < 0$) suggests Weibull type tail behaviour, which has a fixed 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 analyses of the Baltic Sea.

Acknowledgements

This poster presents preliminary results of the project PROTECT (producing extreme weather and sea level for nuclear power plant safety) that is funded by the Nuss Nuclear Waste Management Fund through SAFIR2022, the Finnish Nuclear Power Plant Safety Research Programme 2019-2022 and from the Finnish Meteorological Institute (FMI). PROTECT aims to develop and implement comprehensive and modern methods needed for assessing probabilities of occurrences 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 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 traditionally estimated from the observed annual maximum sea levels using the generalized extreme value distribution (GEV). However, the limited length of observations often cover a relatively short historical time period and thus contain only a small number of extreme issues (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 incorporation of dispers- dences of adjacent stations is also lacking in the analyses.

A partial remedy to these issues is to exploit spatial dependencies exhibited by the sea level extremes. These dependencies can be modeled using hierarchical statistical models, which allow for sharing information between adjacent tide gauging 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 compare different models, where the GEV parameters are allowed to vary over the study area and the extreme value GEV distribution at the basis of sea level. To infer 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 using observations from 12 tide gauges located along the Finnish coast and in Gulf of Finland. Careful analysis of the structure of uncertainties is necessary when extrapolating the return level estimates into the future.

This work is a part of project PROTECT (Producing 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., Tuomenvirta, H., Ruosteenoja, K., Drebs, A., Hyvönen, V., Jokinen, P., Kaas, E., Karhila, V., Kaurola, J., Koskela, J., Niemelä, S., Peltola, H., Rissanen, P., Saku, S., Savijärvi, H., Seitola, T., Tuomi, J., Venäläinen, A., and Vita, A.: Recent meteorological and marine studies to support nuclear power plant safety in Finland, Energy, 35, 1502–1516, https://doi.org/https://doi.org/10.1016/j.energy.2010.09.023, 2010.

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

Saaranen, V., Lehmuskoski, P., Ruokokoski, P., Takalo, M., Mäkinen, J., and Poutanen, M.: The new Finnish height reference N2000, in: Geodetic Reference Frames, pp. 297–302, Springer, 2009.

Stan Development Team: Stan Modeling Language User's 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/