BOREAL ENVIRONMENT RESEARCH 26: 79–87
ISSN 1797-2469 (online)

Helsinki 31 May 2021

Southern Baltic sea level extremes: tide gauge data, historic storms and confidence intervals

Magnus Hieronymus[1)*] and Fredrik Hieronymus[2)3)]

[1)] *Swedish Meteorological and Hydrological Institute, Folkborgsvägen 17, 601 76 Norrköping, Sweden (*corresponding author's e-mail: hieronymus.magnus@gmail.com)*
[2)] *Department of Clinical Medicine, Aarhus University, Incuba Skejby, Building 2, Palle Juul-Jensens Blvd 82, DK-8200 Aarhus, Denmark*
[3)] *Institute of Neuroscience and Physiology, University of Gothenburg, P.O. Box 430, 405 30 Gothenburg, Sweden*

Received 18 Dec. 2020, final version received 12 Mar. 2021, accepted 23 Apr. 2021

Hieronymus M. & Hieronymus F. 2021: Southern Baltic sea level extremes: tide gauge data, historic storms and confidence intervals. *Boreal Env. Res.* 26: 79–87.

Knowledge of extreme sea levels is important when planning housing developments and other infrastructure in coastal locations. The natural science basis of such plans are often return level-return period plots derived from tide gauge records that typically stretch from a few decades to a century. Coastal planners, however, often require return levels associated with return periods that are much longer than these tide gauge records in their planning. Moreover, return level estimates are known to be sensitive to outliers and can have significant biases. Here, we quantify different confidence intervals that are applicable to such return level estimates, and discuss their usability for coastal planning in the context of historic data from the Baltic Sea flood in 1872. Two types of commonly used confidence intervals are found to be too narrow to capture a plausible range that includes the 1872 Baltic Sea flood. A parametric bootstrapping method is then introduced, which gives a reasonable range even when this extreme flood is considered.

Introduction

The sea level is currently rising at an accelerating pace, with further acceleration, increased flooding and more severe economic consequences to be expected in the future (Vousdoukas *et al.* 2018, Oppenheimer *et al.* 2019, Seroussi *et al.* 2020). Estimates of susceptibility to flooding are therefore important when planning new infrastructure in the coastal area, protection for existing infrastructure or even retreat. Projections of mean sea level rise (Johansson *et al.* 2014, Hieronymus and Kalén 2020) and knowledge of sea level extremes are key ingredients when mapping such susceptibilities (Leijala *et al.* 2018, Pellikka *et al.* 2018, Gordeeva and Klevannyy 2020). Sea level extremes used for planning purposes are often cast in terms of return levels with corresponding return periods, while the uncertainty in such estimates is most often characterized by a 95% confidence interval. A 100 year return level has a 1/100 probability of being exceeded in any given year, and is a commonly used metric that can be derived from tide gauge records or model experiments.

However, much higher return levels than the 100 year return level are also often requested for planning purposes. One example from our

Editor in charge of this article: Kai Myrberg

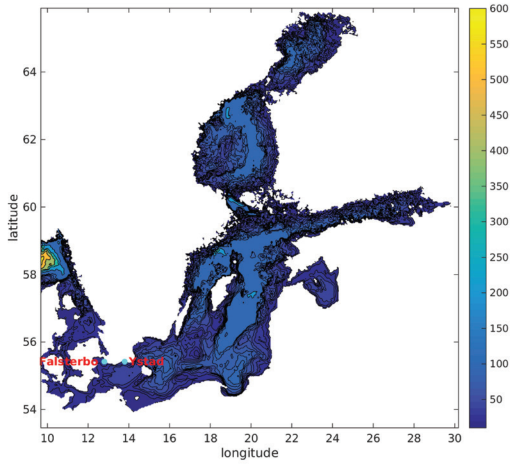

Fig 1. Bathymetric chart of the Baltic Sea showing the location of our tide gauge station Ystad and the Falsterbo peninsula. Bathymetric contours are drawn every 10m up to a depth of 100 m then every 100 m.

study area is the Swedish National Board of Housing, Building and Planning, who in their "Starting points for flood risk assessment" (translation of Swedish original title: Utgångspunkter för bedömning av översvämningsrisk) suggest that developments of new important infrastructure and settlements should be done with a 10 000 year return level in mind (Boverket 2020). Of course, an event with a yearly probability of exceedance of 1/10 000 is extremely unlikely to feature in the observed record from any nearby tide gauge, given that such records are typically around 50 years long and records exceeding 100 years are very rare. Multi-millennial runs with hydrodynamic models that faithfully capture high frequency sea level variability are similarly extreme rarities (Lang and Mikolajweicz 2019). Moreover, it has been demonstrated that return level estimates are sensitive to outliers. That is, a single yearly maximum being pulled from or added to a record can have large consequences for the estimated return levels (Dangendorf *et al.* 2016, Hieronymus and Kalén 2020).

Return level estimates derived from short tide gauge records could thus potentially often be biased. Most often we would expect them to be low biased, since none of the most extreme events are likely to have occurred during the measurement period. However, high biases are also possible if very extreme conditions indeed do feature in the record. Historic storms can sometimes offer clues, insofar as there is anecdotal evidence of sea levels reached during flooding events that predate measurement stations. One prominent example is the Baltic Sea flood in 1872, which ravaged the southern Baltic Sea, claiming the lives of 271 people while leaving 15 000 homeless. This storm temporarily raised sea levels by over three meters in parts of Germany and by more than two meters in parts of Sweden (Feuchter *et al.* 2013, Jensen and Müller-Navarra 2008, Fredriksson *et al.* 2016, Johansson and Nerheim 2020).

Fredriksson *et al.* (2016) estimated multi-millennial return periods for the 1872 Baltic Sea flood based on estimates of sea levels reached on the Falsterbo Peninsula and current tide gauge data. The location of Falsterbo peninsula is marked on the map (Fig. 1). The map also shows the location of Ystad, the tide gauge that we will use in our investigation, which also experienced record sea levels during the same event. The Ystad tide gauge is, in fact, of particular interest because a sea level 1.96 m above the mean was recorded by the lighthouse keeper in Ystad lighthouse during the 1872 storm (Johansson and Nerheim 2020). The sea level reached during the 1872 Baltic Sea flood was undoubtedly extreme. However, it having a multi-millennial return period appears at odds with other historical evidence that suggest similar events may have taken place during the last millennium (Fredriksson *et al.* 2016, Johansson and Nerheim 2020). Moreover, it is not only the estimated return period that is extremely long for this event. The 95% confidence interval also appears much too narrow to accommodate such extremes as the 1872 Baltic Sea flood with a reasonable return period. This brings us to the primary topic of this brief communication; namely whether the most commonly used confidence intervals for return level estimates adequately capture the uncertainty in those estimates or whether some other metric may be preferable for planners needing information about what extremes might be possible.

Material and methods

In what follows, we describe how our return level curves are estimated from tide gauge data.

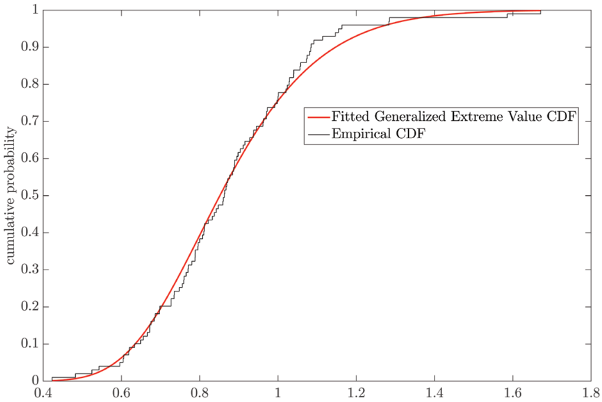

Fig 2. Comparison of the fitted GEV cumulative probability function and the empirical one for the Ystad station. Both the fitted and empirical distributions reflect the full 99 year data set.

Our approach is typical of both industry and scientific practise. However, we should note that it is certainly not the only possible approach to derive extreme value statistics. and that methodological differences in how return levels are estimated, such as whether block maxima or peak over threshold is used, can lead to considerable differences in the results (Wahl *et al.* 2017). Considerable differences may also occur owing to choices of fitting method. The two most common methods are the maximum likelihood method and the probability weighted moments method, but other more novel methods also exist (Makkonen 2008, Makkonen and Tikanmäki 2019).

Our analysis is based on data from the Ystad tide gauge (Fig. 1). The tide gauge data is detrended to remove mean sea level rise and vertical land motion. The time series exists for the years 1886–1987, with hourly data and has no missing values. The largest sea level measured during this period is 1.67 m above the mean. Return value estimates are derived by fitting a Generalized Extreme Value (GEV) distribution to a time series of yearly sea level maxima. The analysis is thus making use of the Fisher-Tippett-Gnedenko theorem, which states that block maximum of samples can only converge in distribution to the GEV distribution. A block length of one year is used, but instead of a calendar year we use a year starting in July and ending in June. The Baltic Sea has strong, low frequency sea level variability (Hünicke *et al.* 2015, Johansson and Kahma 2016) that affects the extremes, but extremes nearly never occur during summer months (Männikus *et al.* 2020). The approach is thus tailored to give rise to independent blocks. In total we have 99 such blocks.

Algorithmically on a user level, our procedure of estimating return levels is simple. The first step is to fit a GEV distribution, which has three parameters: shape (ξ), scale (σ) and location (μ), to the time series of yearly sea level maxima from the tide gauge. The second step is to calculate return levels from the inverse of the GEV cumulative distribution function using the three parameters derived in step one. We used MATLAB (ver. 2015a) for this analysis, making use of the maximum likelihood based *gevfit* function in step one and *gevinv* in step two. A comparison of the fitted GEV cumulative distribution function to the empirical one indicates a reasonable fit (Fig. 2).

Confidence intervals for the return levels can be produced in several different ways. One commonly used technique is the delta method (Coles 2001), which uses a truncated Taylor series expansion of the return level and a likelihood based estimate of the uncertainty in the GEV parameters. The delta method is elegant and computationally cheap, while a drawback of the method is that confidence intervals are symmetric around the central estimate. Another commonly used, and often preferable, confidence interval is the profile likelihood based one (Caires 2011, Mathworks 2020). This method, unlike the delta method, does not rely on linearization and its interval is not symmetric around the central estimate. Here we will compare delta method and profile likelihood based confidence intervals to intervals produced using parametric bootstrapping methods similar to those used in Kyselý (2008) that more directly emulates the way return levels are estimated in practice.

The first bootstrapping method is used to model the spread in return level estimates that owes only to the finite length of the observational record. To do so, we assume here that the GEV parameters estimated from the station data by the gevfit function are identical to those of the real underlying yearly maximum sea level distribution at the site. This estimate therefore underestimates the uncertainty in the derived return levels, because the real parameters of the GEV distribution are never known a priori in real applications of this nature. The likelihood that a sample, like the observational record, or indeed any finite sequence of random numbers drawn from some distribution will be fitted to a different distribution decreases as the length of the sample increases. This effect is quantified by drawing $m \times n$ random numbers from the, here assumed real, GEV distribution, where m = 30 000 is the ensemble size and n is the number of years of data. These numbers are drawn using MATLAB's *gevrnd* function with the shape, scale and location parameter derived from the tide gauge data with the *gevfit* function. A return level curve is calculated for each ensemble member using the methodology described above and 95% confidence intervals are then calculated as the 0.025 and 0.975 quantiles of this ensemble.

Our second bootstrapping method is based on the idea that there might be tide gauges that produce very extreme and very rare outliers, whose real return level curves are significantly higher than the one we inferred from observations, but which also create samples that are comparable to our observed historical data. We thus attempt to find the most extreme tide gauge whose 95% confidence interval contains our best estimate of the return level at the site.

In practise, we think of (ξ, σ, μ) triplets as plausible tide gauges, and we ask the question which of these gauges give rise to return level estimates whose 95% confidence interval contains our estimate. Formally, we seek the set S, defined as:

$$S = [(\xi, \sigma, \mu) : \mathrm{RL}_{\mathrm{obs}} \in 95\%\ \mathrm{CI}_{\mathrm{RL}}], \quad (1)$$

where $\mathrm{RL}_{\mathrm{obs}}$ is the observationally-based return level estimate and 95% $\mathrm{CI}_{\mathrm{RL}}$ is the confidence interval on the return level of the (ξ, σ, μ) triplet estimated using the first bootstrapping method. To bring down the computational cost, we restrict ourselves to check that $\mathrm{RL}_{\mathrm{obs}} \in 95\%\ \mathrm{CI}_{\mathrm{RL}}$ for the 10 000 year return level. Even so, it is, of course, impossible to test all (ξ, σ, μ)triplets. Moreover, it seems unlikely that triplets that are extremely far from our best estimates of ξ, σ and μ would be relevant. We therefore impose the additional constraint that we are only interested in the triplets belonging to the set Z, defined as:

$$Z = [(\xi, \sigma, \mu) : \xi \in 95\%\mathrm{CI}_{\xi}, \sigma \in 95\%\mathrm{CI}_{\sigma}, \mu \in 95\%\mathrm{CI}_{\mu}]. \quad (2)$$

Thus, we probe only the cuboid in ξ-σ-μ space, whose boundaries are defined by the 95% confidence intervals of the individual parameters. Here we use the parameters confidence intervals derived from the observational estimate in our basic case. In all other cases, where we look at the impact of changing n, we use the average of the confidence intervals for the GEV parameters derived when fitting the 30 000 ensemble members used in the first bootstrapping method for the same purpose. Note, that this cuboid has a larger volume in ξ-σ-μ space than the region used to compute the profile likelihood based confidence interval (Mathworks 2020), and thus admit more extreme return levels.

Formally, we define our new interval as the infimum and supremum of:

$$B = [\mathrm{RL}(\xi, \sigma, \mu):(\xi, \sigma, \mu) \in S \cap Z]. \quad (3)$$

Some salient points about B is that; B is not a classical confidence interval associated with some given probability, although it is related to these constructs since return levels in B are related to triplets whose 95% confidence intervals contain our return level estimate. This difference is of no concern to us since our goal is to explore measures that suitably highlights the uncertainties in return levels derived from observational time series. The infimum and supremum of B must come from triplets on the boundary of either S, Z or on the boundaries of both. This fol-

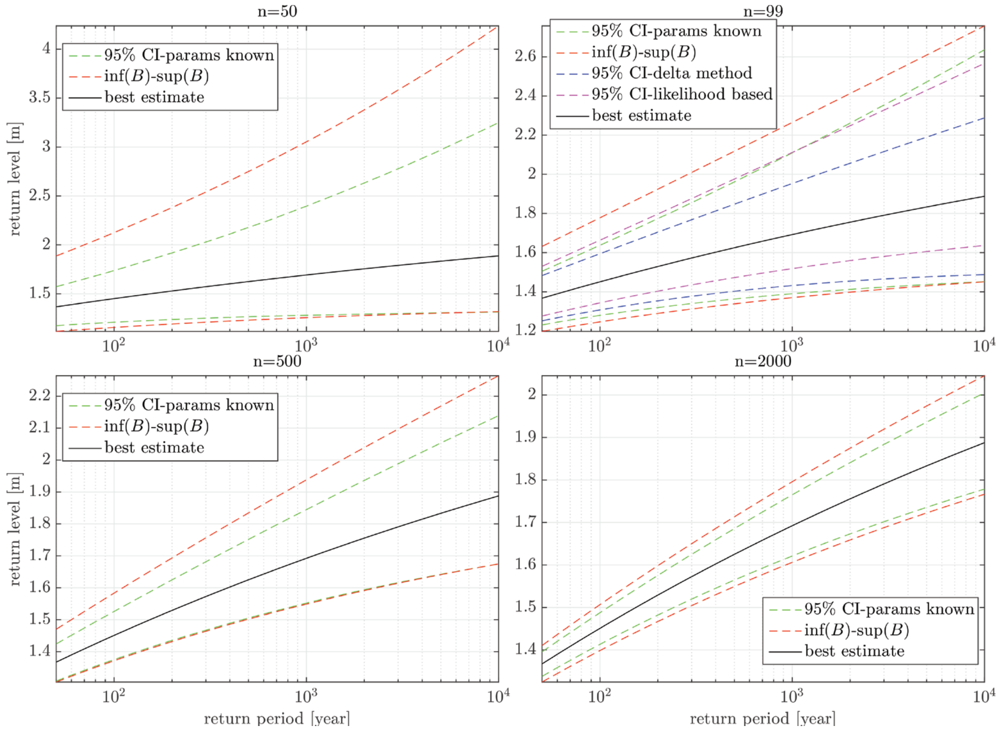

Fig 3. Return level as a function of return period for different record lengths and with several examples of confidence intervals for the Ystad tide gauge. The abbreviation "CI-params known" reflects the 95% confidence intervals from the first bootstrapping method that assumes that the estimated GEV parameter values are their true values. inf(*B*)-sup(*B*) shows the infimum and supremum of the set *B*. 95% CI-delta method and 95% CI-likelihood based represent the 95% confidence intervals of the delta method and the profile likelihood method respectively.

lows directly from the fact that if the supremum of B was not a triplet on the boundary of either S or Z, then there must be a triplet in $S \cap Z$ whose return level is higher than the supremum, given that the partial derivatives of the return level with respect to the GEV parameters are non-zero. Instinctively, we would be more comfortable with an upper bound found on the boundary of S than Z, although we cannot prove that this is preferable. It does appear to us a stronger constraint that no other triplet whose 95% CI_{RL} contains RL_{obs} gives a larger return level, than that no other triplet within our region of interest does the same.

Estimating the interval defined by the infimum and supremum of B is computationally costly, even when using the simplification described above. In our implementation we resolve the cuboid in ξ-σ-μ space, with 50 points in each coordinate direction and we use an ensemble of 400 members when calculating the return level confidence intervals for each triplet. Making good initial guesses of what the infimum and supremum of B might be, and only probing triplets that give more extreme values than those guesses has proven to be an efficient way of saving computational time.

Results

The return level curve and confidence intervals are derived from the Ystad tide gauge record for some different lengths of the observational record (Fig. 3). The maximum likelihood-based estimates of the GEV parameters are $\xi = -0.0910$, $\sigma = 0.1764$ and $\mu = 0.7880$. The lighthouse keeper in Ystad lighthouse recorded a sea level 1.96 m above the mean during the 1872 Baltic Sea flood (Johansson and Nerheim 2020). Based on the return level curve derived here; the return period of the sea level reached during the

1872 flood in Ystad is 27 000 years. This, is even longer than the return period estimated for the same event in Fredriksson *et al.* (2016). The difference between our estimates stems both from methodological differences and differences in the estimates used for the sea level reached during the 1872 event.

The basic case in the figure is that with $n = 99$, as it uses the whole observational time series. The other cases are illustrations of how the confidence intervals converge with n. Already from the basic case it is evident that the different confidence intervals give rather different results, especially for the upper bound. In particular, it is clear that the symmetric assumption used for the delta method underestimates the upper bound. Moreover, both the commonly used delta method and the profile likelihood based confidence intervals appear to be too narrow, as they are considerably narrower than the first bootstrapping method's confidence interval, which is derived using the unrealistic assumption that the real GEV parameters are known.

The lighthouse keeper's estimate of a sea level 1.96 m above the mean only enters the delta method confidence interval at a return period of 1056 years. Moreover, other eyewitness accounts have suggested that a sea level 2.4 m above the mean was reached at the Falsterbo peninsula (see Fig. 1) about 60 km west of Ystad (Fredriksson *et al.* 2016). The same authors also found that on average the yearly maximum sea level at the Falsterbo peninsula is 7.7 cm higher than that in Ystad, and that the two stations are highly correlated. This average difference is so much smaller than the difference between the two estimates for the 1872 Baltic Sea flood that our estimate appears to be conservative. Thus, the delta method, the profile likelihood based method and the first bootstrapping method all appear to produce confidence intervals that are too narrow to accommodate this event with a reasonable probability.

The second bootstrapping method, in contrast to the first and the two common methods, gives estimates that are plausible even considering extreme events such as the 1872 Baltic Sea flood. The bootstrapping techniques are also useful for determining how fast the uncertainty goes down as the length of the instrumental record increases. A typical Swedish tide gauge record with hourly resolution is about 50 years long, while the longest are more than 100 years long. The convergence of the confidence intervals toward the real return level is fastest for small n. However, convergence is slow enough that high return levels will be extremely uncertain for the foreseeable future (Fig. 3). The interval given for the 10 000 year return level with the second bootstrapping method with $n = 50$ for example, is more than four times larger than the IPCC's projected mean sea level rise at Ystad between today and 2100, under their highest emission scenario RCP8.5 (Oppenheimer *et al.* 2019, Hieronymus and Kalén 2020).

We find all of the triplets giving rise to the infimum and supremum of B, except the supremum in the $n = 99$ case, to belong to the boundary of S. Three of the seven triplets that belong to the boundary of S, however, also belong to the boundary of Z.

Discussion

Given that return levels are commonly used in coastal spatial planning and may dictate for example if housing development is allowed in certain locations, it is of great importance to have a good understanding of the uncertainty in such estimates. Here we conclude that different 95% confidence intervals give considerably different ranges, especially on their important upper bounds. Moreover, we find the commonly used delta method and profile likelihood based method give too narrow confidence intervals to capture plausible ranges for the real return levels in the Southern Baltic Sea when compared to historic floods. Note that, the confidence intervals discussed are really statements about the confidence in the estimation process, rather than statements about the confidence that the true return level value should fall within the specific intervals. Therefore, it is in a sense wrong to call them too narrow, as it implies judging them on metrics they were not designed to measure. However, they are too narrow in the sense that they do not appear to accurately capture a plausible range for the true return level, which is ultimately what is needed for coastal infrastructure planning and which is also how planners typically interpret confidence intervals.

It should be noted that the above conclusion relies on eyewitness accounts of sea surface heights reached during past storms. Such accounts typically include also effects of wind waves. Comparing them to still water levels measured by tide gauges as done here may thus exaggerate the inferred historical still water level (Wahl *et al.* 2012, Leijala *et al.* 2018). Other uncertainties owing to for example human errors and timing of old observations could also affect our conclusions. Moreover, so can uncertainties owing to methodological differences in how return values are estimated (Wahl *et al.* 2017, Makkonen and Tikanmäki 2019). Non-stationarity in sea level extremes is another potential explanation for the mismatch between derived return levels and historic observations. Linear trends in the location parameters have, for example, been found in the Baltic Sea in model experiments covering the years 1979–2012 (Kudryavtseva *et al.* 2018). Nevertheless, in the case of the 1872 Baltic Sea flood, we don't believe that either of these potential sources of error fully explain the mismatch between our derived return levels and the historic observations.

We believe a more plausible explanation for the presumed shortcoming of the GEV analysis presented here is tied to the very prominent low frequency sea level variability in the Baltic Sea. Certain atmospheric conditions can raise the sea level in the Baltic Sea by more than half a meter for several weeks (Hünicke *et al.* 2015, Johansson and Kahma 2016). Storms that arrive during such episodes can create extreme outliers in sea level that fit poorly with derived return level curves (Suursaar and Sooäär 2007). Such extremes are perhaps best viewed as coming from a different population than other yearly maxima, and a GEV analysis with a longer block length might be needed to ensure that the different blocks are in fact drawn from the same distribution.

Our and others presumption that return level estimates are often biased low (Suursaar and Sooääar 2007, Fredriksson *et al.* 2016, Gordeeva and Klevannyy 2020), suggests that the probability of flooding may be underestimated especially for long return periods. On the other hand, the confidence intervals from the second bootstrapping method for the 10 000 year return level, which is sometimes favoured for planning (Boverket 2020), are so wide that the information is likely useless for most planning purposes. Thus, it appears that there is a considerable difference between what planners want and what one can reasonably expect to extract from data. Given the slow convergence of the return level estimates with new data, it seems that notable improvements in these uncertainty estimates can presently only be achieved through modelling experiments where longer time periods than the observed history can be modelled (see e.g., Särkkä *et al.* (2017)). A computationally cost efficient way of achieving this could be to use machine learning-based regression models (Hieronymus *et al.* 2019). Another plausible path is to rely more on physical insights or scaling experiments such as those of Hieronymus *et al.* (2018), where the effect of increased wind speeds on return levels are estimated — than on statistical inferences. However, while much can still be learnt about the physics of flooding, it is often hard to translate such physical insights into frequency metrics such as return periods. Moreover, hydrodynamical models contain biases and do not resolve all processes that occur in nature, corrections are therefore often needed before accurate return levels can be inferred (Björkqvist *et al.* 2020).

Regardless of how one's estimate is derived, it is important that it comes with an with an appropriate quantification of uncertainty. We argue that the second bootstrapping method could be an improvement over the other methods because: 1) it gives more plausible estimates than the other methods as evidenced by historical data; and 2) the approach may be able to find possible tide gauges that create infrequent and very large outliers, but that also create samples that are reasonably compatible with our sampled history. However, as mentioned earlier, both the potential non-stationarity of the extremes and the possibility of extreme outliers belonging to another population could suggest that a proper determination of the GEV parameters can only be done with a much larger block length than one year, or perhaps using non-stationary methods. All of these intervals should thus be thought of as plausible ranges rather than as exact quantifications of uncertainty. This is perhaps the most important piece of information that must be conveyed between the scientists and engineers that produce estimates and data and the decision makers and stakeholders who use them.

Acknowledgements: We thank three anonymous reviewers and Tarmo Soomere for their insightful comments on the manuscript.

Data availability statement: The data presented is freely available through SMHI open data.

References

Björkqvist J.-V., Rikka S., Alari V., Männik A., Tuomi L. & Pettersson H. 2020: Wave height return periods from combined measurement–model data: a Baltic Sea case study. *Natural Hazards and Earth System Sciences*, 20 (12), 3593–3609, doi: 10.5194/nhess-20-3593-2020.

Boverket 2020: Utgångspunkter för bedömning av översvåmningsrisk. Boverket, available at https://www.boverket.se/sv/PBL-kunskapsbanken/planering/detaljplan/lansstyrelsens-tillsyn/tillsynsvagledning_naturolyckor/tillsynsvagledning-oversvamning/stod-till-lansstyrelsen-vid-riskbedomning/utgangspunkter/.

Caires S. 2011: Extreme value analysis: still water level. Jcomm technical report no. 58, World Meteorological Organization, Netherlands.

Coles S. 2001: *An introduction to statistical modeling of extreme values*. 1st ed., Springer, Berlin

Dangendorf S., Arns A., Pinto J.G., Ludwig P. & Jensen J. 2016: The exceptional influence of storm 'Xaver' on design water levels in the German Bight. *Environmental Research Letters*, 11 (5), 054 001, doi: 10.1088/1748-9326/11/5/054001.

Feuchter D., Jörg C., Rosenhagen G., Auchmann R., Martius O. & Brönnimann S. 2013: *The 1872 Baltic Sea storm surge*, In: Brönnimann S. and Martius O. (Eds.) Weather extremes during the past 140 years. Geographica Bernensia G89, pp. 91–98, doi: 10.4480/GB2013.G89.10.

Fredriksson C., Tajvidi N., Hanson H. & Larson M. 2016: Statistical analysis of extreme sea water levels at the Falsterbo Peninsula, South Sweden. *Journal of Water Management and Research*, 72, 129–142.

Gordeeva S.M. & Klevannyy K.A., 2020: Estimation of the maximum and minimum surge levels at the Hanhikivi peninsula, Gulf of Bothnia. *Boreal Env. Res*., 25, 51–63.

Hieronymus M., Dieterich C., Andersson H. & Hordoir R. 2018: The effects of mean sea level rise and strengthened winds on extreme sea levels in the Baltic Sea. *Theoretical and Applied Mechanics Letters*., 8, 366–371.

Hieronymus M., Hieronymus J. & Hieronymus F. 2019: On the application of machine learning techniques to regression problems in sea level studies. *Journal of Atmospheric and Oceanic Technology*, 36 (9), 1889–1902, doi: 10.1175/JTECH-D-19-0033.1.

Hieronymus M. & Kalén O. 2020: Sea-level rise projections for Sweden based on the new IPCC special report: The ocean and cryosphere in a changing climate. *Ambio*, doi: 10.1007/s13280-019-01313-8.

Hünicke B., Zorita E., Soomere T., Madsen K.S., Johansson M. & Suursaar Ü. 2015: *Recent Change — Sea Level and Wind Waves*, 155–185. Springer International Publishing, Cham, doi: 10.1007/978-3-319-16006-1 9.

Jensen J. & Müller-Navarra S.H., 2008: Storm surges on the German coast. *Die Küste*, 92–124.

Johansson L. & Nerheim S. 2020: Extremvattenstånd i Ystad. MSB report, SMHI, MSB, 651 81 Karlstad.

Johansson M. & Kahma K.K. 2016: On the statistical relationship between the geostrophic wind and sea level variations in the Baltic Sea. *Boreal Env. Res*., 21, 25–43.

Johansson M.M., Pellikka H., Kahma K.K. & Ruosteenoja K. 2014: Global sea level rise scenarios adapted to the Finnish coast. *Journal of Marine Systems*, 129, 35–46.

Kudryavtseva N., Pindsoo K. & Soomere T. 2018: Nonstationary modeling of trends in extreme water level changes along the Baltic Sea coast. *Journal of Coastal Research*, 85, 586–590, doi: 10.2112/SI85-118.1.

Kyselý J. 2008: A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. *Journal of Applied Meteorology and Climatology*, 47 (12), 3236–3251, doi: 10.1175/2008JAMC1763.1.

Lang A. & Mikolajweicz U. 2019: The long-term variability of extreme sea levels in the German Bight. *Ocean Sciences*, 15, 651668, doi: 10.5194/os-15-651-2019.

Leijala U., Björkqvist J.-V., Johansson M.M., Pellikka H., Laakso L. & Kahma K.K. 2018: Combining probability distributions of sea level variations and wave run-up to evaluate coastal flooding risks. *Natural Hazards and Earth System Sciences*, 18 (10), 2785–2799, doi: 10.5194/nhess-18-2785-2018.

Makkonen L. 2008: Problems in the extreme value analysis. *Structural Safety*, 30 (5), 405– 419, doi: 10.1016/j.strusafe.2006.12.001.

Makkonen L. & Tikanmäki M. 2019: An improved method of extreme value analysis. *Journal of Hydrology X*, 2, 100 012, doi: 10.1016/j.hydroa.2018.100012.

Männikus R., Soomere T. & Viŝka M. 2020: Variations in the mean, seasonal and extreme water level on the latvian coast, the eastern Baltic Sea, during 1961–2018. *Estuarine, Coastal and Shelf Science*, 245, 106 827, doi: 10.1016/j.ecss.2020.106827.

Mathworks 2020: Modelling data with the generalized extreme value distribution. Availiable at https://se.mathworks.com/help/stats/modelling-data-with-the-generalized-extreme-value-distribution.html;jsessionid=17f81cdc87516b88ba9364595cd6.

Oppenheimer M., Glavovic B.C., Hinkel J., van de Wal R., Magnan A.K., Abd-Elgawad A., Cai R., CifuentesJara M., DeConto R.M., Ghosh T., Hay J., Isla F., Marzeion B., Meyssignac B. & Sebesvari Z. 2019: Sea Level Rise and Implications for Low Lying Islands, Coasts and Communities. In: Pörtner H.-O., Roberts D.C., Masson-Delmotte V., Zhai P., Tignor M., Poloczanska E., Mintenbeck K., Alegría A., Nicolai M., Okem A., Petzold J., Rama B. & Weyer N.M. (eds.), *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. [In press]

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.

Särkkä J., Kahma K.K., Kämäräinen M., Johansson M. & Saku S. 2017: Simulated extreme sea levels at Helsinki. *Boreal Env. Res.*, 22, 299–315.

Seroussi H., Nowicki S., Payne A.J., Goelzer H., Lipscomb W.H., Abe-Ouchi A., Agosta C., Albrecht T., Asay-Davis X., Barthel A. & Calov R. 2020: ISMIP6 Antarctica: a multi-model ensemble of the Antarctic ice sheet evolution over the 21st century. *The Cryosphere Discussions*, 2020, 1–54, doi: 10.5194/tc-2019-324.

Suursaar Ü. & Sooäär J. 2007: Decadal variations in mean and extreme sea level values along the Estonian coast of the Baltic Sea. *Tellus A: Dynamic Meteorology and Oceanography*, 59 (2), 249–260, doi: 10.1111/j.1600-0870.2006.00220.x.

Vousdoukas M.I., Mentaschi L., Voukouvalas E., Bianchi A., Dottori F. & Feyen L. 2018: Climatic and socioeconomic controls of future coastal flood risk in Europe. *Nature Clim. Change*, 8, 776–780.

Wahl T., Haigh I.D., Nicholls R.J., Arns A., Dangendorf S., Hinkel J. & Slangen A.B. 2017: Understanding extreme sea levels for broad-scale coastal impact and adaptation analysis. *Nature Comm.*, 8.

Wahl T., Mudersbach C. & Jensen J. 2012: Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: a multivariate statistical approach based on Copula functions. *Natural Hazards and Earth System Sciences*, 12 (2), 495–510, doi: 10.5194/nhess-12-495-2012.