Dear Sir/Madam:

We once more thank the Referees for their comments to the revised version of our submission. They again helped us to improve the manuscript. As usual, we have given careful consideration to all remarks; in particular, we have added a detailed discussion of goodness of fits of different theoretical distribution using the Kolmogorov-Smirnov test.

The relevant analysis was performed by Nadezhda Kudryavtseva. As this part is an essential constituent of the manuscript now, we think that she should be a co-author, and have included her into the authors' list.

Also, we have removed a few typos and checked once more formatting of references.

The comments of Referees are represented using normal font below and our response and a description of corrections using italics. All references to the changes in the manuscript are to the version that contains all corrections and is presented below.

Sincerely

Tarmo Soomere, Katri Pindsoo, Nadezda Kudryavtseva and Maris Eelsalu
07 April 2020

Referee #1

/---/ As some of my comments might change the results, I have to label the possible revisions as major.

Thank you; we understand that.

Main comment #1 (to the original manuscript): The authors have fully cleared up why the results are now different than in the previous submission to Earth System Dynamics. This matter has been fully addressed, and I leave it up to the editor how this should be reflected (if at all) in the final published manuscript.

We shall be happy with any decision of the Editor on this issue.

Main comment #2 (to the original manuscript): My second main concern was the accuracy of the data. I recommended that the model would be run for a shorter time with a "full set-up" to validate the results. The authors chose to use measurements to validate their data, and while they are only available at one point, I think that this is a valid alternative to making additional model runs with a wind forcing from an atmospheric model. I accept this general approach taken by the authors to address my concerns. I also agree with the authors that a bias of 0.05 m is about as good as it gets. The revised manuscript states that the Tallinnamadal wave data are buoy data (although I have been under the impression that it is based on a pressure sensor). Is there a reference for these data?

Thank you for noticing this. We re-checked this. The data indeed come from a pressure sensor (and this is why wave direction is not present). The accuracy of wave measurements is ±0.1 m. The description of the location and parameters of the sensor: http://efficiensea.org/files/mainoutputs/wp4/efficiensea_wp4_27.pdf. There are some attempts to evaluate wave heights using standard navigation buoys but the relevant data are not yet reliable.

/---/ I'm left feeling a bit uneasy about the results of the validation. The authors claim that "The appearance of the relevant empirical probability distributions of the occurrence of different wave heights is similar for both data sets (Fig. 3)". Still, I can't really see those two distributions being similar. The most important distinction is that a fit to measured data over 1 m would be reasonable well represented with a straight line (i.e. exponential fit), while a fit to the modelled data would be concave upwards.

Doing the analysis on the model data would probably lead to a positive quadratic exponential, while a similar analysis on the measured data would result in a quadratic exponential close to zero. I understand that this is wave height and not wave set-up, but the former is a "forcing" for the latter. How can we trust the results of a quadratic exponential in the modelled wave set-up, if the shape of the distribution is not estimated correctly in the incoming wave height data?

We do not want to claim that our hindcast data is perfect. Most probably not. Our point is: Let's think that wave model distorts the statistics so that the replicated wave data show a shape of this distribution concave upwards (like a Weibull *distribution in log-linear coordinates that is actually expected here) instead of measured basically exponential distribution. It is then likely that a similar distortion (too large frequency of "intermediate"-height set-up events) should be evident in the distribution of set-up heights. Still the distribution of set-up heights is in quite many occasions concave downwards. This aspect is discussed on page 25, lines 11–19 below (equivalently, page 17, two last lines and page 18, lines 1–8 in the clean version of the manuscript.*

We understand that this is not a particularly strong argument; however, it still says that concave downwards distributions of set-up heights probably reflect some specific features of the "conversion" of wave properties in to set-up heights.

To shed more light to the situation, we have considerably extended the analysis of distributions of set-up heights towards systematic comparison of their shape with several frequently occurring distributions. We created a new Section 3.3 to accommodate this description.

Main comment #3 (to the original manuscript): This comment was concerning the fitting procedures and how they were presented. First, the authors have expanded on the description of the function they are fitting and what they are doing. This is now perfectly clear.

I also now understand why the authors want to use the gaps as a cut of the fitting: you want to avoid fitting the distribution to the "tail" of the data, which is not that statistically stable. I, however, still disagree that the gaps would have some kind of relevant meaning here. I would suggest that the upper limit of the fit would be a certain multiple of the mean wave set-up, or a certain amount of standard deviations above the mean. This would be a more objective and robust measure for the upper point than the gaps, which depend on the resolution of the binning, and also have a larger statistical variability (especially if this analysis should ever be reproduced using measurement data).

Thank you for these suggestions. We have done this kind of analysis (Fig. 13) but using a somewhat „stronger" criteria for the match of the empirical and theoretical distributions. As discussed elsewhere, the gaps do not carry any specific meaning. For us, their presence signals that uncertainties of probabilities of even larger set-up heights may contain large uncertainties and it is not really justified to rely on these values.

I also still disagree with the use of probability distributions instead of cumulative distribution. The authors said that the cumulative distribution can smooth out differences. Another way to see it is that using the probability distribution exaggerates differences. Reviewer #1 also suggested using cumulative distributions, and while he mentioned extreme value analysis, I want to point out that the fitting of cumulative distributions are in no way restricted to performing extreme value analysis (which I know that you are specifically not doing). Still, since I haven't seen this type of fitting before, I can't say how big of a deal this really is. I would strongly suggest fitting to the cumulative distribution (as this is standard practice), but if you are set on fitting to the probability distribution, then, ultimately, I won't stop you.

Thank you for understanding. We are not totally against the use of cumulative distributions. It is just our understanding that they do not shed light to our problem and may even provide deceptive confidence. To make our point clear, we have added Figure 12 where this difference is clearly demonstrates. To avoid such possibly deceptive signal, we have chosen another way to compare empirical distributions with possible theoretical ones via D-values of Kolmogorov-Smirnov test

I also disagree with the authors claim that the highest values would only be a part of some "extreme value distribution" and therefore outside the basic distribution (page 13, lines 16-17). This is not the case. When calculating e.g. block maxima the points in the resulting extreme value distribution are still members of the original underlying distribution. For example, sufficiently long block maxima of an exponentially distributed variable should follow a Gumbel distribution, but all of the block maxima were still points (but vary far up the tail) in the original exponential distribution. If you want to exclude outliers, don't use the previously mentioned claim to do it, since it is fundamentally incorrect.

Thank you; our formulation was inappropriate indeed. Our wish was to stress that the largest values may belong to some other population of events similarly to the idea of (Suursaar and Sooäär, 2007). The sentence is deleted.

In summary, I do believe that the basic idea adopted by the authors is correct, namely that the last (highest) points are infamously unstable, and excluding them from the fit might be warranted. My disagreement is with the motivation (citing extreme value distributions) and the fitting technique (probability distributions and use of gaps) adopted to implement this idea.

Main comment #4 (to the revised manuscript): This comment is based on information that was not available in the original manuscript. You use two simplifications in calculating the refraction and shoaling. Assumption of shallow water and Snell's law. You mention that the water depth can be between 4 and 27 m. This would mean a (deep water) wavelength of about 260 m and 1700 m meter to satisfy the shallow water condition (wave period up to 33 s!). This is not reasonable, especially in the Gulf of Finland. Calculating the true phase and group speed using iteration from the full dispersion relation is trivial, and I see no reason not to use it.

This must be misunderstanding. We do use the full dispersion relation for tracing the waves from the last wave model grid cell to the breaker line and only apply shallow water approximation at the breaker line. This is explained a few lines after Eq. (9). We added the word "full" to make this clear. We only use shallow water approximation at the breaker line. We are aware that this approximation and the evaluation of set-up height based on wave properties at this line are not perfect as dispersion still works to some extent in the surf zone. However, we think that the uncertainty or error generated via this approximation is still acceptable.

The second problem is Snell's law, since it assumes that the isolines in the bathymetry are straight and parallel. This might often be a good approximation, but looking at Fig. 10, this is a questionable assumption in Tallinn Bay. The proper refraction can be calculated using the full equations (as done in e.g. WAM)

The WAM model with the resolution we use provides wave data at a distance of a few hundred of metres from the breaker line. It is of course possible to evaluate refraction and shoaling using more elaborate methods; however, this would need phase-resolving models and bathymetry at much higher resolution.

As one of your contributions in this study is to investigate the complex geometry and it's effect on the wave set-up, I find it surprising that you have neglected effects that seems to be important in this geographical area.

We agree with this point. In essence, we study the problem by assuming that all coastal segments are (i) favourable for the formation of high set-up and (ii) are approximately homogeneous alongshore. These assumptions are only valid for selected segments of the study area. We added the relevant comment to Discussion (page 25, lines 22–24 below; equivalently, page 18, lines 10–12 in the clean version). However, we are of opinion that our study is still of some value as it establishes the typical shape of the distribution of wave set-up and its possible variation in ideal conditions. Another (quite weak) argument is that a sandy coastal area (favourable for set-up) with a similar complexity simply cannot exist.

Specific comment #1: To really see the effect of the fitting range, please continue the blue line as (for example) a dashed blue line outside the fitting range to illustrate how well it captures the data that were not used for fitting.

It would be nice to explore this item for single points; however, we are of opinion that doing so would defocus the presentation. Therefore, instead, we chose to present the magnitude of this effect in terms of the variation in the Kolmogorov-Smirnov statistic for the entire study area for three examples of the fitting range (Figure 13).

Specific comment #2: On page, lines you write: "In other words, in these locations the leading term a of the quadratic polynomial az^2+ bz+ c is positive at a 95% significance level."

This isn't really true, since you didn't do the analysis for only one point. If you do the analysis for all points then, on average, around 5% of the points should give a "false positive" when using a 95% significance (if the null hypothesis was true). For you 25% >> 5%, so the effect is real, but the confidence in a single point is no longer positive at a 95% confidence level, and treating it as such seems to be some unintentional version of p-hacking. (I understand that there was no malicious intent here even though I used the phrase p-hacking. It was used only to make the point clear. This comment was of a pure technical nature.)

Thank you; this is a good point, and we adjusted our formulations to make clear what we mean: this estimate of significance is valid for single points but not for the set of such points (page 22, lines 19–22 below; equivalently, page 15, lines 6–9).

Specific comment #3: Page 14, lines 16-17 “For large values of z this function behaves similarly to the probability density function of a Gaussian distribution.”

Did you mean for small values of z?

Thank you for this question. It’s actually a stupid typo (corrected now) as the inverse Gaussian goes as an exponential function ~exp(-az) for very large arguments (when z^(-3/2) is small compared to exp) while Gaussian goes as ~exp(-az^2).

Specific comment #4: Figure 8. I find it hard to believe that it wouldn’t be possible to get a better Weibull fit than that presented in the figure, since the exponential distribution is contained in the Weibull family. Have you calculated the empirical cumulative distribution and plotted that in Weibull coordinates? You are claiming that Weibull is insufficient, so we need to make sure that this is right. As some other reviewer commented, it takes a lot to discard a three parameter Weibull.

We performed the relevant analysis (expanding substantially Section 3) in terms of D-values of Kolmogorov-Smirnov test and added a couple of images to make these aspects very clear. Just a small note: we use 2-parameter Weibull distribution as we do not consider extreme values of set-up. And of course, the more parameters a fitting function has, the better fit can be produced.

Specific comment #5: As a suggestion, perhaps amend the “convex upwards” and “concave upwards” with (light tailed) and (heavy tailed) remarks. This might be especially useful the first time they are mentioned in the text, and perhaps in the discussion and conclusions, which some might read without going through the entire text and all the figures.

Thank you for this idea. It has been implemented in a slightly different context as the Gaussian and generally Weibull distributions are light-tailed (~exp(-az^2)) and exponential and inverse Gaussian are heavy-tailed (~exp(-az)). This is now explained on page 21, lines 10–14 (equivalently, at the beginning of page 14 of the clean version). Btw, highlighting this difference partially explains why inverse Gaussian provides a good fit.

So e.g. on page 15, lines 26-27 the text would then read “The appearance of the distribution of modelled wave heights in the offshore (Fig. 3) is convex upwards (thin tailed) in the range of relatively frequent wave heights of 0.5–1.7 m.”

As pointed above, the property of being convex and having a heavy/light tail are not so directly related.

Referee #2

The Authors did a thorough revision of the text which is now much more clear and readable. In particular Section 2 (Methods and data) and 3 (Results) have been greatly improved. Still the main problem of the study is the result, which indicates that in complex morphology the wave set-up can follow several (at least two) different statistical distributions, with no way to predict which one will be occurring at any given coastal point. On the whole the manuscript still suffers from being a bit too qualitative and somewhat lacking in rigour concerning the analysis, see specific comments. The latter would be a serious issue in any situation, but here the results presented are rather hard to accept and, furthermore, no clear explanation is given. So the thesis must be proved `beyond a reasonable doubt'.

We generally agree with this remark and provide more detailed analysis of the match of obtained empirical distributions with commonly used theoretical distributions. However, we do not fully agree with the comment that the results are rather hard to accept. In essence, we show that in some 3/4 of cases the distribution of set-up heights follows (at least qualitatively) an exponential distribution. Thus, we understand that the problematic point is that another, not very often used, distribution pops up in some locations, and we do not have a good explanation for that. Such things are an intrinsic part of research, and we are of opinion that hiding such occasions would be totally wrong. In this sense we hope that even if we do not have a rigorous proof, our communication still has a value.

Specific comments

The only measure of uncertainty considered in the study is the use of the 95% significance level in the evaluation of the coefficients of the quadratic polynomial, which is not the same as to evaluate the goodness of the fit to the theoretical distribution, See Figure 8.

We expanded substantially Section 3 towards filling this gap.

A major weak point is the choice of a particular range of data to be fitted: a fixed interval [0.01,0.4] or a variable interval [.01,α], where α is the first gap in the distribution.

We actually did all the analysis several times for different cut-off schemes, one of these involving all data points. This has been explained in all revisions; now on page 21, 18–27 (equivalently, page 14, lines 7–14 in the clean version). The results were very consistent, so we decided to provide only one version in Figs. 9 and 10. This consistency is now thoroughly illustrated Figure 13.

It especially bothers me that in all cases shown, the frequency does not change with the increase of set-up height above 50 cm.

This is simply because large set-up heights have occurred exactly once during the study period. This is also why we use several variations of cut-off.

Bottom line, in my opinion it appears that both choices are rather arbitrary. The first because the entire domain of setup heights is in the range [0,1] (fig. 5) and the second because the threshold depends on the bin size. If I'm not badly mistaken the gaps would disappear if the frequencies were evaluated using different (i.e. larger) bins. Figure 3e-f show clearly the problem.

This is indeed true, and this is another reason why we have used several cut-off schemes in Fig. 13 to be sure that our results are consistent.

Besides, how many points have been excluded in the region $x > 50$ cm in order to get a straight line in figure 3f? [*7f in the revised manuscript*]

No points have been excluded in calculations of the red dashed line and 13 entries have been excluded in the calculation of the blue solid line.

The procedure of building the empirical distribution is not described in sufficient detail and this should really be fixed. The choice of the bins size in particular should be discussed showing what happens if classes of heights are merged. This is a critical point for the kind of analysis proposed in the manuscript.

In this respect we partially disagree. The procedure of building empirical distribution from time series is a standard counting exercise that, to our understanding, needs not to be described in a research paper.

As the modelled wave heights are evaluated with an accuracy of 1 cm, using even finer resolution for the construction of empirical probability distributions is not justified. As the total number of entries in time series of positive set-up heights is 40 000–70 000, using the resolution of 1 cm ensures that at least a few examples of set-up heights will belong to the relevant "size classes" of set-up heights down to frequencies about 10^{-2}*%. This range is apparently large enough to estimate the main properties of the distributions in question. We added this explanation into the beginning of Section 3.2.*

Even though some details of the appearance of empirical distributions will change owing to a change in the bins' size, no large changes to the properties of approximating functions are expected (until the number of bins decreases below a certain limit). This is because such fitting operations are basically linear or rely on the main statistical parameters (e.g., the method of moments only takes into account the mean and standard deviation of the data set for 2-parameter distributions).

Another point to discuss in section 2 is the bathymetry, as it is critical factor in the evaluation of the set-up. In particular it should be verified that the bathymetry used to run the model has a resolution greater than or at least comparable with the model grid resolution. If not, the possible effect on the results should be discussed.

To meet suggestions of another referee, we added a more detailed description of the wave model and its bathymetry into Section 2.2. In particular, we explain there that the bathymetry used in the innermost grid was constructed using high-resolution sailing maps by Estonian Maritime Board, with an efficient resolution ~200 m. We think this is enough for wave modelling at a resolution of about 470 m.

Technical corrections

The significant wave height is a statistical measure of the sea state on a wide area and during a long time, compared to a wave height at a fixed point in space and time. In section 2 they are sometimes used improperly (for example in paragraph 20). My suggestion is to define significant wave height as HS or Hm0 and monochromatic wave height as H, but more important is to stick to it in all the manuscript.

We distinguish the height of monochromatic waves (and denote its value at the breaker line by H_b) from significant wave height. The latter is, as recommended, defined in terms of m_0. The output of wave measurements is also given in terms of significant wave height. There is of course certain difference between how this quantity is defined in wave models and measurements; however, its explanation is beyond the scope of our study. The seeming inconsistency may arise at the point where we link the modelled significant wave height with the wave height at the breaker line. We have adjusted the text in a few locations so that this logic is clearer.

A symbol for the set-up height would also help readability of the manuscript.

We introduced the symbol \eta_{max} in Eq. (9) but further on it seemed to us that the use of words "set-up height" is more traditional.

When Weibull distributions are considered is important to mention the value of the shape factor used, because it causes the shape of the curve to change drastically.

Thank you; we indicate now the shape parameter, e.g., for Fig. 11 and 12.

Referee #3

The manuscript discusses wave setup, which in some coastal areas and high wave events can be a significant proportion of the total sea level at coast. The introduction gives a nice overview about the topic.

Thank you for this.

Maybe the recent studies in the Baltic Sea related e.g. to the analyse joint distributions of sea level and surface waves (Leijala et al., 2018. https://doi.org/10.5194/nhess-18-2785-2018) could be added.

Thank you; it is important indeed to remind the reader that the baseline level of wave run-up also includes the local elevation of water level owing to set-up. We added this reference to the relevant remark on page 3, line 3.

My main concern is the reliability of the model results in coastal areas and their suitability for the study in question. Especially when the distributions show some exceptional behaviour at some of the coastal grid points.

We fully agree with this point.

Although the triple nested model setup has been used in earlier studies (Soomere 2005 and Soomere et al. 2013 (missing from the references)) a brief description should be given here as well as it is important in order to interpret the results. I assume that the areas covered area the Baltic Sea, the Gulf of Finland (nest1) and the Estonian coastal area (nest2). What is used as forcing for the Baltic Sea and the Gulf of Finland grids? As they provide the boundary conditions, they play an important role in the accuracy of the results.

We apologise for the missing reference (Soomere et al., 2013) in the original submission; it has been added. We also added a description of the model setup into Section 2.2.

All models were used with Kalbådagrund wind data. Even though wind properties in Baltic proper may be different, waves generated in the sea area to the south of the latitudes of the Gulf of Finland almost never reach Tallinn Bay and thus are irrelevant in the context of our study.

The authors state that the wind reanalysis and NWP systems available are not sufficiently accurate to model waves in coastal areas. This might have been the case ten years ago, but presently there are several sufficiently good quality reanalysis available and they have been successfully used in wave model studies in the Baltic Sea. And even if they are not perfect, they are able to better represent the spatial variability in the wind field than a single point measurement. Although Kalbådagrund is representative station for open sea wind conditions in the Gulf of Finland, the values are most likely too high to be used in coastal wave modelling and might results into too high wave energies.

We agree that the quality of numerically replicated wind information is continuously increasing. To show that our results are adequate, we added a comparison with measured wave data at the northern border of the second nesting area into the end of Section 2.2. Our model forced with Kalbådagrund data lead to a bias of 5 cm and to a reasonably replicated the empirical distribution of occurrence of different wave heights. As the further propagation time of waves from the border of the second nesting until the shore is usually few dozens of minutes, variations in the wind properties in the nearshore apparently play a very minor role in the formation of the wave properties at the breaker line.

It is not also clear to me how the use of ‘precomputed maps of wave properties’ affects the statistics compared to a full reanalysis or hindcast.

This is probably a misunderstanding that has been already resolved in the first revision of the manuscript. We additionally adjusted the relevant parts of the text to make our point clear. In essence, our point is that this simplified scheme of replication of wave time series reasonably replicates also wave statistics. As just discussed, the numerically evaluated wave heights have bias of 5 cm compared to the measured ones and the model reasonably replicates the empirical distribution of occurrence of different wave heights. Thus, we are of opinion that the model, albeit it is greatly simplified, produces a realistic hindcast of time series of wave properties and reasonably replicates wave statistics as shown in Fig. 3. The process of using precomputed maps is quite complicated and is described in detail in (Soomere, 2005).

The Authors refer to Baltic Sea wave model studies, that have shown wave model WAM as a good tool to model Baltic Sea wave fields. But the end results is not only based on the model, it is a combination of model, the selections (physics, numerics) you make within the model, accuracy of bathymetric data, the description of coastline with the given resolution, and the forcing used. Therefore some validation of the model setup used should be performed. If there is no data available from the coastal area in question, then the results from Gulf of Finland grid could be used for validation against the Gulf of Finland wave buoys. That would at least give some estimates on the reliability of the results.

Thank you; we have added into first revision of Section 2.2 quite detailed comparison of hindcast and measured wave properties at the border of the second nested area (Fig. 2, Fig. 3). This is the only location in the vicinity of the study area where reliable wave data is available. It is thus likely that the wave model produces acceptable-quality wave time series in the open sea. Most of the innermost grid area has water depth >20 m and thus waves with typical periods for this area (3–4 s) practically do not “feel” the presence of bottom. As water depth increases rapidly offshore from the nearshore grid cells used for our calculations, WAM model propagates waves through the areas where the waves may experience substantial interaction with seabed only over one, maximum two grid cells. The bathymetry is built using high-resolution marine maps and thus is realistic. The propagation of waves from the closest to land grid cell to the breaker line is resolved within linear theory but its limitations are well known.

Also one could check that the highest waves in the coastal areas are plausible given the accuracy of the offshore value, the bathymetry and the propagation angle of waves.

The analysis of highest waves in the nearshore is provided in (Soomere et al., 2013 and Figure 3 in this source). We added the relevant remark to page 16, lines 25–28 below (equivalently, page 9, lines 7–10 of the clean version). This check is also included into the first revision by means of comparison of empirical probability distributions of simulated and measured wave heights (page 7–8, Fig. 3). This figure shows that during the comparison period three measured and two simulated significant wave heights exceeded 5 m whereas the maximum measured and simulated wave heights differ by 20 cm. This *sort of behaviour of extremes is realistic as significant wave height of 5.2 m has been recorded twice about 50 km from the study area since 2001. As for the further propagation of waves from the border of the innermost grid to the nearshore, we assume that the WAM model adequately does its job.*

Also the fact that the ice conditions are not included might distort the statistics. Have you done calculations by including only the ice-free periods?

We agree that taking into account the presence of ice may change the results to some extent quantitatively. However, as the number of ice-free winters is increasing, we think that it is interesting to present ice-free statistics as a starting point.

I suggest that the quality of the model results is addressed before considering this paper for publication in OS.

As discussed above, we have performed a careful comparison of the model output against available measured wave data at the border of the study area.

Also, I suggest considering to upgrade the model setup to newer WAM versions and using spatially varying wind forcing datasets.

As pointed by other Referees, doing so would basically mean running the entire wave model for many decades and thus is not realistic.

Variability of distributions of wave set-up heights along a shoreline with complicated geometry

Tarmo Soomere[1,2], Katri Pindsoo[1], Nadezhda Kudryavtseva[1], Maris Eelsalu[1]

[1] Laboratory of Wave Engineering, Department of Cybernetics, School of Science, Tallinn University of Technology, Akadeemia tee 21, Tallinn, 12618, Estonia
[2] Estonian Academy of Sciences, Kohtu 6, Tallinn, 10130, Estonia

Correspondence to: Tarmo Soomere (soomere@cs.ioc.ee)

Abstract. The phenomenon of wave set-up may substantially contribute to the formation of devastating coastal flooding in certain coastal areas. We study the appearance and properties of empirical probability density distributions of the occurrence of different set-up heights on an approximately 80 km section of coastline near Tallinn in the Gulf of Finland, the eastern Baltic Sea. The study area is often attacked by high waves propagating from various directions and the typical approach angle of high waves varies considerably along the shore. The distributions in question are approximated by an exponential distribution with a quadratic polynomial as the exponent. Even though different segments of the study area have substantially different wave regimes, the leading term of this polynomial is usually small (between –0.005 and 0.005) and varies insignificantly along the study area. Consequently, the distribution of wave set-up heights substantially deviates from a Rayleigh or Weibull distribution (that usually reflect the distribution of different wave heights). On about ¾ of the occasions it is fairly well approximated by a standard exponential distribution. In about 25% of the coastal segments it qualitatively matches a Wald (inverse Gaussian) distribution. This property signals that very high extreme set-up events may, in some locations, occur substantially more frequently than is expected from the probability of occurrence of severe seas.

1 Introduction

Global sea level rise in most of existing projections of climate change (Cazenave et al., 2014) is often associated with major consequences for the coastal zone (Hallegatte et al., 2013). The resulting economic damages to low-lying coastal areas (Darwin and Tol, 2001) may lead to a significant loss of worldwide welfare by the end of this century (Pycroft et al., 2016). Global sea level rise, however, contributes only a small fraction to the most devastating coastal floods. These events, in addition to being economically extremely damaging (Meyer et al., 2013), may also lead to massive loss of life and destruction of entire coastal communities (Dube et al., 2009).

A devastating flood is usually caused by the interplay of several drivers, with fundamentally different predictability, and physical, dynamic and statistical properties. A reasonable forecast of the joint impact of tides, low atmospheric pressure (inverted barometric effect), wind-driven surge and wave-induced effects requires a cluster of dedicated atmospheric, ocean circulation and wave models. The resulting high water levels may be additionally amplified by specific mechanisms and events such as tide–surge interactions (Batstone et al., 2013; Olbert et al., 2013), meteorologically driven long waves (Pattiarachi and Wijeratne, 2014; Pellikka et al., 2014; Vilibic et al., 2014) or seiches (Vilibic, 2006; Kulikov and Medvedev, 2013). In addition, wave-driven effects at the waterline such as wave set-up and run-up (Stockdon et al., 2006) may greatly contribute to the damaging potential of extreme water levels. These phenomena are driven by momentum carried by waves, and have different time scales and appearance. When a wave crest reaches the shore, the resulting temporary inland movement of the water, with a time scale comparable with the wave period, is termed run-up (see, e.g., Didenkulova, 2009 for an overview and references). In contrast, wave set-up is the increase in the mean water level due to the release of momentum of breaking waves.

Along with contemporary numerical simulations and a direct search for worst-case scenarios (e.g., Averkiev and Klevanny, 2010), the use of the probabilistic approach is another classic way to quantify the properties of extreme water levels and related risks. The relevant pool of literature contains substantial amounts of work on statistical parameters of water level variations (e.g., Serafin and Ruggiero, 2014; Fawcett and Walshaw, 2016), and extreme water levels and their return periods (e.g., Purvis et al., 2008; Haigh et al., 2010; Arns et al., 2013). Similar probabilistic analysis has been extensively applied to the average and extreme wave properties (e.g., Orimolade et al., 2016; Rueda et al., 2016), wave-driven effects at the waterline (Holland and Holman, 1993; Stockdon et al., 2006), and properties of meteotsunamis (Geist et al., 2014, Bechle et al., 2015). On most occasions severe coastal flooding occurs under the joint impact of several drivers. This feature generates the necessity to consider multivariate distributions of their properties. Most often, the simultaneous occurrence of storm surges and large waves is addressed (e.g., Hawkes et al., 2002; Wadey et al., 2015; Rueda et al., 2016b). A few studies also include an analysis of joint distributions of significant wave heights, periods and directions (Masina et al., 2015).

Typical probability distributions of different constituents of extreme water levels may be fundamentally different. The distribution of observed and numerically simulated water levels is usually close to Gaussian (Bortot et al., 2000; Johansson et al., 2001; Mel and Lionello, 2014; Soomere et al., 2015). The total water level in semi-sheltered seas with extensive subtidal or weekly-scale variability may contain two components. In the Baltic Sea, one of these (that reflects the water volume of the entire sea) has a classic quasi-Gaussian distribution whereas the other component (that reflects the local storm surge) has an exponential distribution and apparently mirrors a Poisson process (Soomere et al., 2015) similar to the non-tidal residual in the North Sea (Schmitt et al., 2018). The probabilities of occurrence of different single wave heights are at best approximated either by a Rayleigh (Longuet-Higgins, 1952), Weibull (Forristall, 1978) or Tayfun distribution (Socquet-Juglard et al., 2005). The probability distribution of run-up heights usually follows the relevant distribution for incident wave heights (Denissenko et al., 2011) but can be approximated by a Rayleigh distribution even if the approaching wave field does not represent a Gaussian process (Denissenko et al., 2013). The empirical probabilities of average or significant wave heights in various offshore conditions usually resemble either a Rayleigh or a Weibull distribution (Muraleedharan et al., 2007; Feng et al., 2014) while Pareto-type distributions are more suitable for the analysis of meteotsunami heights (Bechle et al., 2015).

In this paper we focus on the appearance and properties of empirical distributions of wave-driven local water level set-up. This process, called set-up in the following, is a classic phenomenon on open ocean coasts. It may often provide as much as 1/3 of the total water level rise during a storm (Dean and Bender, 2006) and significantly contribute to extreme sea level events (Hoeke et al., 2013; Melet et al., 2016, 2018). The impact of this phenomenon *inter alia* contributes to the overall level of danger in the coastal zone because, for example, the baseline level of wave run-up (Leijala et al., 2018) includes the local elevation of water level owing to set-up. The physics of set-up has been known for half a century (Longuet-Higgins and Stewart, 1964). Adequate parameterizations of this phenomenon have been introduced more than a decade ago (Stockdon et al., 2006) and many models take it into account to a certain extent (SWAN, 2007; Roland et al., 2009; Alari and Kõuts, 2012; Moghimi et al., 2013).

The contribution from wave set-up still provides one of the largest challenges in the modelling of storm surges and flooding (Dukhovskoy and Morey, 2011; Melet et al., 2013). This reflects the intrinsically complicated nature of its formation. First of all, the set-up height strongly depends on the approach angle of waves at the breaker line. This angle is well-defined only if the coastline is almost straight, the nearshore is mostly homogeneous in the alongshore direction and the wave field is close to monochromatic (Larson et al., 2010; Viška and Soomere, 2013; Lopez-Ruiz et al., 2014; 2015). Generally, this angle is a complicated function of shoreline geometry, nearshore bathymetry, wave properties and instantaneous water level. Even if the basic statistical properties of wave fields (usually given in terms of significant wave height, mean or peak period, and mean propagation direction) are perfectly forecast or hindcast in a nearshore location, the evaluation of the further propagation of waves is a major challenge because, for example, refraction properties and the location of the breaking line change with the local water level.

Several studies have focused on the maxima of set-up heights over certain coastal areas (Soomere et al., 2013; O'Grady et al., 2015) or the maximum contribution from set-up to the local water level extremes (Pindsoo and Soomere, 2016). The problem of evaluation of maximum set-up heights has a relatively simple solution on comparatively straight open ocean coasts. The nearshore of such coasts is usually fairly homogeneous in the alongshore direction and the highest waves tend to approach the shore under relatively small angles. These features make it possible to use simplified schemes for the evaluation of the joint impact of refraction and shoaling in the nearshore (e.g., Larson et al., 2010). On many occasions it is acceptable to assume that waves propagate directly onshore (O'Grady et al., 2015) or to reduce the problem to an evaluation of the properties of the highest waves that approach the shore from a relatively narrow range of directions (Soomere et al., 2013). In areas with complicated geometry and especially in coastal segments where high waves may often approach at large angles it is necessary to take into account full refraction and shoaling in the nearshore (Viška and Soomere, 2013; Pindsoo and Soomere, 2015).

Even though high storm surges are often associated with severe seas, the formation of high set-up depends on many details of the storms and the impacted nearshore. It does not necessarily exhibit its maximum level in the coastal sections that are affected by the highest waves. The maximum storm surge and maximum set-up usually do not occur simultaneously (Pindsoo and Soomere, 2015). On the contrary, in coastal areas with complicated geometry each short segment may have its own 'perfect storm' that creates the all-highest sum of storm surge and set-up (Soomere et al., 2013). These observations call for further analysis of the properties of the set-up phenomenon.

As described above, research into the statistical properties of the main drivers of high local water levels and the reach of swash generated by large waves that attack the shore have revealed that the relevant distributions of the magnitude of these drivers are very different. They may include a Gaussian distribution for the water volume of the Baltic Sea (Soomere et al., 2015), an exponential distribution for storm surges (Schmitt et al., 2018), a quasi-Gaussian distribution for water levels at the shores of the Baltic Sea (Johansson et al., 2011), a Weibull distribution for different significant wave heights (Feng et al., 2014), and a Weibull or Rayleigh distribution for wave run-up heights (Denissenko et al., 2013). The knowledge of the shape and parameters of such distributions is often crucial in various forecasts and management decisions.

In this paper we address the basic features of statistical distributions of set-up heights along an approximately 80 km coastal section in the vicinity of Tallinn Bay in the Gulf of Finland, the Baltic Sea. The shoreline of the study area has a complicated geometry and contains segments with greatly different orientations. The goal is to identify the typical shapes of the distributions of the probability of occurrence of simulated wave set-up heights and to analyse the alongshore variability of these distributions.

The paper is organised as follows. Section 2 introduces the method of evaluation of the maximum set-up height for obliquely approaching waves. It also provides a short overview of the simplified method for rapid reconstruction of long-term wave climate, the forcing data for the underlying wave model and the procedure of evaluation of properties of breaking waves based on the output of the wave model. Section 3 presents an analysis of spatial variations in the appearance of the empirical distribution of wave set-up heights in the study area. The core result is an estimate of the typical shape of empirical probability distributions of different set-up heights along the coast. A part of these distributions substantially deviate from the listed distributions and exhibit an unexpectedly large proportion of high set-up events compared to the classic Gaussian, Rayleigh or Weibull statistics. The Kolmogorov-Smirnov test is applied to estimate the goodness of match of the empirical distribution of set-up heights with common theoretical distributions. Several implications of the results are discussed in Section 4.

2 Methods and data

2.1 Set-up height for obliquely incident waves

The classic concept of wave set-up (Longuet-Higgins and Stewart, 1962) relates the local increase in the water level with the release of the onshore component of radiation stress in the process of wave breaking. Based on this concept, it has been demonstrated that, in ideal conditions, the maximum set-up height $\eta_{\max}$ (with respect to the still water level) created by a train of monochromatic waves with a constant height propagating directly onshore along a planar impermeable beach is (McDougal and Hudspeth, 1983; Hsu et al., 2006)

$$\eta_{\max} = d_b \frac{40\gamma_b^2 - 3\gamma_b^4}{128} \approx \frac{5}{16}\gamma_b H_b, \qquad (1)$$

where $\gamma_b = H_b / d_b$ is the breaking index that is assumed to be constant all over the surf zone, d_b is the still water depth at the breaker line and H_b is the wave height at the breaker line (Fig. 1). Expression (1) is used in many engineering applications (Dean and Dalrymple, 1991) and studies into the properties of set-up of waves that approach the shore at a relatively small angle (see Soomere et al., 2013 and references therein).

If waves approach under a non-negligible angle θ with respect to the shore-normal, the situation is much more complicated. Shi and Kirby (2008) argue that the water level set-down at the breaker line is invariant with respect to the approach angle of waves. The average deviation η of the sea surface from the still water level within the surf zone of an impermeable planar beach is (Hsu et al., 2006; Shi and Kirby, 2008; the power of γ_b in the first term at the right-hand side of their expression being corrected):

$$\eta = \frac{\gamma_b^2 \sin^2\theta_b}{2h_b\left(8 + 3\gamma_b^2 - 2\gamma_b^2 \sin^2\theta_b\right)}\left(d^2 - d_b^2\right) - \frac{3\gamma_b^2 - 2\gamma_b^2 \sin^2\theta_b}{8 + 3\gamma_b^2 - 2\gamma_b^2 \sin^2\theta_b}\left(d - d_b\right) - \frac{\gamma_b^2}{16} d_b . \qquad (2)$$

The last term at the right-hand side of Eq. (2) represents the water level set-down η_b at the breaker line and θ_b is the wave approach direction at breaking. Here $d = d(x)$ represents the water depth counted from the still water level at a particular distance *x* from the shoreline and η is a function of *x*. The maximum wave set-up $\eta_{\max}$ occurs somewhere inland where

$d_{\max}$ is negative and the thickness of the water sheet $d^* = d + \eta_{\max} = 0$, and thus $\eta_{\max} = -d_{\max}$. For this location, Eq. (2) reduces to:

$$\eta_{\max} + d_{\max} = \frac{\gamma_b^2 \sin^2 \theta_b}{2h_b\left(8 + 3\gamma_b^2 - 2\gamma_b^2 \sin^2 \theta_b\right)}\left(d_{\max}^2 - d_b^2\right) - \frac{3\gamma_b^2 - 2\gamma_b^2 \sin^2 \theta_b}{8 + 3\gamma_b^2 - 2\gamma_b^2 \sin^2 \theta_b}\left(d_{\max} - d_b\right) - \frac{\gamma_b^2}{16} d_b + d_{\max} = 0 \,. \qquad (3)$$

For shore-normal waves $\theta_b = 0$ and Eq. (3) reduces to a linear equation:

$$-\frac{3\gamma_b^2}{8 + 3\gamma_b^2}\left(d_{\max} - d_b\right) - \frac{\gamma_b^2}{16} d_b + d_{\max} = 0 \,. \qquad (4)$$

In this case the maximum set-up height $\eta_{\max}$ is defined by Eq. (1).

For obliquely approaching waves Eq. (3) is a quadratic equation with respect to $q = d_{\max}/d_b$:

$$\frac{\gamma_b^2 \sin^2 \theta_b}{2\left(8 + 3\gamma_b^2 - 2\gamma_b^2 \sin^2 \theta_b\right)}\left(q^2 - 1\right) - \frac{3\gamma_b^2 - 2\gamma_b^2 \sin^2 \theta_b}{8 + 3\gamma_b^2 - 2\gamma_b^2 \sin^2 \theta_b}\left(q - 1\right) - \frac{\gamma_b^2}{16} + q = 0 \,. \qquad (5)$$

This equation can be rewritten as

$$8q^2 \gamma_b^2 \sin^2 \theta_b + 128q + \gamma_b^2\left(40 - 3\gamma_b^2\right) - \gamma_b^2 \sin^2 \theta_b\left(40 - 2\gamma_b^2\right) = 0 \,. \qquad (6)$$

Equation (6) has two negative solutions for physically reasonable values of γ_b. The physically relevant solution to Eq. (6) must be bounded and should be almost equal to $q \approx -5\gamma_b^2/16$ for very small approach angles $\theta_b \approx 0$. Therefore, the expression

$$q_1 = \frac{-32 + \sqrt{1024 - 2\gamma_b^4 \sin^2 \theta_b\left[40 - 3\gamma_b^2 - \sin^2 \theta_b\left(40 - 2\gamma_b^2\right)\right]}}{4\gamma_b^2 \sin^2 \theta_b} \qquad (7)$$

provides the desired solution. Equation (7) deviates from expression (30) of Hsu et al. (2006) by reasons discussed by Shi and Kirby (2008). The maximum set-up height for obliquely approaching waves is thus

$$\eta_{\max} = -q_1 d_b = -\frac{q_1 H_b}{\gamma_b} \,. \qquad (8)$$

This quantity is called simply set-up height below.

2.2 Wave time series in the nearshore of the study area

We evaluate the shape and parameters of the empirical probability distribution set-up heights along an approximately 80 km coastal segment of Tallinn Bay and Muuga Bay (Fig. 2). The study area is an example of a wave-dominated micro-tidal region. The shoreline is locally almost straight for scales up to a kilometre or two. Several relatively straight parts along the Suurupi Peninsula (grid points 1–10 in Fig. 2) and the area of Saviranna (grid points 137–143 in Fig. 2) are open to the north. However, at larger scales (from a few kilometres) the coast contains large peninsulas and bays deeply cut into the mainland. The shores of these landforms are open to different directions and have greatly different wave regimes (Soomere, 2005). As the formation of set-up crucially depends on the wave height and direction (or approach angle), this type of coastal landscape makes it possible to analyse the wave set-up distribution for coastal sections with radically different wave climates, and also the associated magnitudes of set-up (Soomere et al., 2013).

The fetch length in the Gulf of Finland is >200 km for westerly and easterly winds but <100 km for all other wind directions. The highest significant wave height (5.2 m) in the Gulf of Finland has been recorded twice in a location just a few tens of km to the north of the study area (Tuomi et al., 2011). The strong winds in this region blow predominantly from the south-west and north-north-west. Easterly storms are less frequent but may generate waves as high as those generated by westerly storms (Soomere et al., 2008). Strong storms with winds from the north-north-west may generate significant wave heights >4 m in the interior of Tallinn Bay (Soomere, 2005). The varying mutual orientation of high winds, propagation direction of waves and single shoreline segments makes it possible to identify potential alongshore variations in the distributions of set-up heights.

We employ time series of wave properties (significant wave height, wave period and propagation direction) reconstructed using the wave model WAM cycle 4 and one-point high-quality wind information from the vicinity of the study area. The wave model is implemented in a triple nested version with the resolution of the innermost grid about 470 m (Soomere, 2005). The regular rectangular grid of the coarse model covered the whole Baltic Sea with a step of 3 × 6′ (3′ along latitudes and 6′ along longitudes, that is, about three nautical miles). We applied 24 evenly spaced directions and 25 frequencies ranging from 0.042 to 0.41 Hz with an increment of 1.1.

Experience with this model in the Baltic Sea and Finnish archipelago indicates that it is important to adequately represent the wave growth in low wind and short fetch conditions (Tuomi et al., 2011; 2012). To meet this requirement, in calculations with the wind speed below 10 m/s we applied an increased frequency range of waves up to 2.08 Hz in order to correctly represent wave growth in low wind conditions after calm situations (Soomere, 2005). Full wave spectrum from the coarse model was used as boundary conditions for the first nesting with a similar regular grid with a step 1 × 2′ (about 1 nautical mile). This grid covered the interior of the Gulf of Finland to the east of 23°18′E. The bathymetry of these two models is based on data from (Seifert et al., 2001) with a resolution 1 × 2′. Full wave spectrum from the model for the Gulf of Finland was used as boundary conditions for the second nesting. The study area (Fig. 2) was covered with a regular grid with a step of 1/4 × 1/2′ between 24°28′ and 25°16′E and to the south of 59°41′N. The bathymetry for the second nesting bathymetry is constructed based on maps issued by the Estonian Maritime Board with typical resolution of about 200 m. The wave height is presented in terms of significant wave height H. This quantity, often denoted as H_S and below called wave height, is

Deleted: the model uses

Field Code Changed

Field Code Changed

defined as $H=\sqrt{m_0}$, where m_0 is the zero-order moment of the one-dimensional wave spectrum, and is estimated with a resolution of 1 cm.

Field Code Changed

Field Code Changed

The nearshore of the study area is divided into 174 coastal segments with a length of about 500 m (Fig. 2). Each segment corresponds to a nearshore wave model grid cell. Ignoring the presence of sea ice may lead to a certain overestimation of the overall wave energy in the region but apparently does not significantly distort the shape of the probability distribution of different wave heights (Fig. 3). It is therefore likely that the shape of distributions of set-up heights and the variation in these distributions along the shoreline are also reconstructed adequately.

We employ a simplified method for rapid reconstruction of long-term wave climate. The computations are speeded up by replacing exact calculations of wave generation, interactions and propagation by an analysis of precomputed maps of wave properties for different wind speed, direction, and duration. This simplification avoids reconstruction of all the wave time series and relies on a favourable feature of the local wave regime, namely that wave fields rapidly become saturated and have relatively short memory in the study area (Soomere, 2005). Consequently, a reasonable reproduction of wave statistics is possible by the assumption that an instant wave field in Tallinn Bay is a function of a short time period of wind dynamics. This assumption justifies splitting the calculations of time series of wave properties into independent sections with duration of 3–12 hours. The details of the model set-up, bathymetry used, and the implementation and validation of the outcome have been repeatedly discussed in the literature (Soomere, 2005; Soomere et al., 2013).

Deleted: the time series of sea state

The described approach makes it possible to circumvent one of the major issues of replication of Baltic Sea wave fields, namely, the frequent inconsistency of different modelled wind data sets (Nikolkina et al., 2014). Similar to, for example, wave-driven sediment transport, wave set-up is intrinsically sensitive with respect to the wave propagation direction. As the nearshore wave directions in areas with complex geometry and bathymetry may be greatly impacted by local features, it is crucial to properly reconstruct the offshore wave directions. This is only possible if the wave model has correct information about wind directions. This is an issue in the Gulf of Finland where atmospheric models often fail to reproduce wind directions (Keevallik and Soomere, 2010). To overcome this issue, we use wind data from an offshore location in the central part of this gulf. The wind recordings at Kalbådagrund (59°59′ N, 25°36′ E, a caisson lighthouse located on an offshore shoal) are known to represent marine wind properties well (Soomere et al., 2008). Even though this site is located at a distance of about 60 km from the study area, it is expected to correctly record wind properties in the offshore that govern the generation of surface waves in the open sea.

Deleted: exceptionally

Wind properties at Kalbådagrund were recorded starting from 1981 once every 3 hours for more than two decades, but since then they have been filed at a higher time resolution. To ensure that the forcing data is homogeneous, we downsampled the newer higher-resolution recordings by selecting the data entries once in 3 hours. The entire simulation interval 1981–2016 contained 103 498 wind measurement instants with a time step of 3 h. In about 9000 cases (less than 10% of the entire set) either wind speed or direction was missing. These time instants were excluded from the further analysis. As some of these instants involved quite strong winds, our analysis may underestimate the highest wave set-up events for some segments of the shore. However, as we are interested in the statistical properties of most frequently occurring set-up heights and in the alongshore variations of these properties, it is likely that omitting these data does not substantially impact the results.

To roughly estimate the adequacy of the described method for speeding up the estimates of the properties of wave climate via rapid reconstruction of time series of approximate wave properties, its outcome is compared with the results of wave measurements made by Marine Systems Institute, Tallinn University of Technology, at Tallinnamadal (Fig. 2, 59°42.723′ N, 24°43.890′ E). The data come from a pressure sensor. The accuracy of estimates of significant wave height is ±0.1 m. The description of the location and parameters of the sensor are presented in http://efficiensea.org/files/mainoutputs/wp4/efficiensea_wp4_27.pdf. The data set is available at https://www.emodnet-physics.eu/map/platinfo/piroosplot.aspx?platformid=8974.

As the measurements are available starting from 2012, the comparison is performed for the time interval of 2012–August 2016. The measured wave properties are compared with the modelled properties at the closest grid point of the sea area represented at 470 m resolution at 59°41′ N, 24°45′ E. The buoy is located about 3 km from the border of this area and the distance between the buoy and the centre of the closest grid cell is 3.34 km. The comparison (Fig. 3) only includes the instants when both measured (green) and modelled (red) wave parameters were available. We use for comparison the significant wave height.

Deleted: results of buoy

Deleted: H (often denoted as H_S and below called wave height) defined as $H = \sqrt{m_0}$, where m_0 is the zero-order moment of the one-dimensional wave spectrum.

The basic properties of wave heights such as the maximum (measured 5.58 m, modelled 5.77 m), mean (measured 0.643 m, modelled 0.697 m) and median (measured 0.40 m, modelled 0.54 m) are represented reasonably. The bias of the model (about 0.05 m) is at the same level as the typical bias for modelled wave properties in the Baltic Sea in the most recent simulations (Björkqvist et al., 2018). As our study basically relies on statistical properties of wave fields (the probability of occurrence of seas with different significant wave height, period, and direction), the analysis below, strictly speaking, does not require an exact reconstruction of the sequence of wave events. In this context, the root mean square difference of the modelled and measured time series wave heights (0.5 m) is reasonable. This value is about twice as large as in (Björkqvist et al., 2018) for the Gulf of Finland (0.20–0.31 m) or northern Baltic Proper (0.26 m) and comparable to the level of this quantity for the Sea of Bothnia (0.31–0.56 m, Björkqvist et al., 2018).

The highest waves in the coastal areas over the entire simulation interval are plausible (Soomere et al., 2013, Figure 3). The maximum wave height in 1981–2012 was 5.2 m in a section that was fully open to one of the directions of the strongest winds from north-north-west. Such wave conditions have been measured twice since 2001 in the interior of the Gulf of Finland (Soomere, 2005; Pettersson et al., 2013).

The appearance of the relevant empirical probability distributions of the occurrence of different wave heights is similar for both data sets (Fig. 3). The location and height of the peaks of these distributions (that represent the properties of most frequently occurring waves) have a reasonable match. The model overestimates to some extent the frequency of waves with heights of 0.8–1.5 m and underestimates the frequency of highest waves in the area (>2 m). The overall appearance of the distribution for modelled wave heights resembles a Weibull distribution. The empirical distribution for measured wave

Deleted: almost perfect

Deleted: whereas this

heights higher than 0.4 m better matches an exponential distribution and exhibits much larger variability in the frequency of very high (>3 m) waves.

2.3 Nearshore refraction and shoaling

The nearshore grid cells selected for the analysis (Fig. 2) are located in water depth ≥4 m in order to avoid problems with reconstruction of wave heights under possible intense wave breaking in these cells in the strongest storms. Some of the cells are located in much deeper water, at a depth of 20–27 m. The nearshore of the study area contains various underwater features and bottom inhomogeneities. This means that shoaling and refraction may considerably impact the wave fields even along the relatively short paths (normally ≤1 km in our model setup) from the model grid cells to the breaker line. The predominant wind directions in strong storms are from the south-west, north-north-west and west (Soomere et al., 2008). Consequently, high waves often approach some of the selected grid cells at large angles with respect to the shore-normal. Therefore, it is not acceptable to assume that the incidence angles are small. As a result, oversimplified approaches to replicate the changes in wave properties in the immediate nearshore (Lopez-Ruiz et al., 2014; 2015) and even advanced approximations of refraction and shoaling (Hansen and Larson, 2010) may fail.

For this reason we calculate the joint impact of shoaling and refraction of approaching waves in the framework of linear wave theory. Following (Soomere et al., 2013), we assume that the numerically evaluated wave field for each time instant is monochromatic. The wave height is characterised by the numerically simulated (significant) wave height H, peak period T_p and mean propagation direction θ (clockwise with respect to the direction to the North). These properties are evaluated at the centre of each selected grid cell. The significant wave height at this location is denoted as H_0. The approach direction θ_0 at this location with respect to the onshore-directed normal to the shoreline is calculated from θ based on an approximation of the relevant (about 500 m long) coastal segment by a straight line that follows the average orientation of the shoreline in this segment. Similarly, it is assumed that the nearshore seabed from the centre of each grid cell to the waterline is a plane with isobaths strictly parallel to this straight line. Finally, we assume that breaking waves are long waves. Then the wave height H_b at the breaking line can be found as the smaller real solution of the following algebraic equation of 6th order (Viška and Soomere, 2013; Soomere et al., 2013):

$$\frac{H_b^5 g}{H_0^4 \gamma_b}\left(1-\frac{gH_b}{\gamma_b}\frac{\sin^2\theta_0}{c_{f0}^2}\right)=c_{g0}^2\left(1-\sin^2\theta_0\right). \qquad (9)$$

Here c_g is the group speed, c_f is the phase speed and the subscripts “0” and “b” indicate the relevant value at the centre of the particular wave model grid cell and at the breaker line, respectively. The phase and group speed at the wave model grid cell are estimated based on the standard expressions of linear theory using the wave number k evaluated from the full dispersion relation for linear monochromatic waves $2\pi/T_p=\sqrt{gk\tanh kd_0}$ with the period equal to the peak period T_p and the water depth d_0 equal to the model depth for the particular grid cell. We assume that waves at the breaker line (where the water depth is d_b) are long waves. The set of assumptions is completed with the common notion that the breaking index is $\gamma_b = H_b / d_b = 0.8$ (Dean and Dalrymple, 1991). Part of the introduced assumptions, such as a plane seabed and dry coast without any vegetation, monochromatic wave fields and a constant value of the breaking index for wind-seas as well as the ignoring of the wave period (or steepness) in the calculations are not fully realistic. The potential impact of these approximations is discussed in Section 4.

The procedure of evaluation of set-up heights is thus as follows. We start from the numerically simulated significant wave height H_0, peak period T_p, and mean propagation direction θ with respect to the direction to the North. Next we calculate the phase (c_{fb}) and group speed (c_{gb}) of such waves at the model grid cell and find the propagation direction with respect to the shore-normal θ_0. Equation (9) is employed subsequently to evaluate the changes to the wave height owing to refraction and shoaling on its way from the model grid point to the breaker line. In essence, it links the model output H_0 (given in terms of significant wave height) with the height H_b of breaking waves. The phase and group speed at the breaker line are estimated from the dispersion relation for long waves: $c_{gb} = c_{fb} = \sqrt{gd_b}$. The wave approach direction θ_b at the breaker line with respect to the shore-normal is calculated from Snell's law $\sin\theta / c_f = const$. Thereafter we employ Eqs. (7) and (8) to find the set-up height for the particular time instant.

Several earlier studies of extreme set-up heights (Soomere et al., 2013; Pindsoo and Soomere, 2015) followed this procedure but took into account only waves that propagated under high angles (not larger than ±15°) with respect to the shore-normal and ignored the correction expressed in Eqs. (7, 8) for waves that approached under a nonzero angle. This approach is denoted S2013 below. It is adequate on the open ocean coasts where waves usually approach the shore under relatively small angles but it may fail in semi-sheltered basins with short fetch.

3 Results

3.1 Maximum set-up heights

The phenomenon of wave set-up is only significant if large waves propagate towards the shore. This is usually the case on open ocean coasts where swells almost always create set-up. The situation may be different in sheltered sea areas with complicated geometry where intense swells may be infrequent and the majority of the wind wave energy may propagate in an offshore direction, this being common in the study area. The wind regime of the study area is a superposition of four wind systems (Soomere et al., 2008). The most frequent wind direction is from south-west (that is, from the mainland to the sea). The proportion of wave fields that propagate onshore is 40–70% along the entire study shoreline (Fig. 4). The statistical properties of set-up heights discussed below thus represent 40 000–70 000 examples of wave fields in each coastal segment.

Field Code Changed

Field Code Changed

The only exception is grid cell 107 (Figs. 2, 4) between Viimsi Peninsula and the island of Aegna that is sheltered for almost all wind directions.

We start from a comparison of maximum set-up heights evaluated using the above-described approach and the method employed in S2013. The two sets of estimates differ insignificantly (by less than 0.1 m) in about 80% of the coastal segments (Fig. 5). The alongshore variations in the maxima of set-up heights evaluated from Eq. (8) are considerably smaller than those estimated using the approach of S2013. The largest examples of set-up heights reach 1 m and the majority of maximum set-up heights for single coastal sections are 0.6–0.8 m in both sets of estimates.

The largest differences between the two sets become evident in segments that are sheltered from predominant storm directions, most notably in deeply cut bays. Estimates based on Eqs. (7, 8) are often remarkably (by up to 50%) higher in these sections than those derived using S2013. This feature signals that the highest waves approach the shore at a relatively large angle in such sections. This property shows the importance of the generation of remarkable set-up heights by obliquely approaching high waves. Therefore, ignoring waves that approach under large angles may substantially underestimate the maximum set-up height in some coastal segments.

In other words, the impact of refraction often overrides the effect of geometric blocking of waves by changing the orientation of the coastline. Refraction thus often redirects wave energy so that even beaches that are seemingly well sheltered geometrically may at times receive remarkable amounts of wave energy (*cf* Caliskan and Valle-Levinson, 2008). The differences in the maxima of set-up heights evaluated using the two approaches for such coastal sections are often 0.2–0.3 m and reach up to 0.5 m. Such a strong impact of refraction is thought to be responsible for a local increase in wave heights in the Baltic Sea (Soomere, 2003) and also in extreme ocean conditions (Babanin et al., 2011). The processes that are not resolved by phase-averaged wave models such as reflection and diffraction may add even more wave energy to seemingly sheltered coastal segments.

On the contrary, S2013 overestimates the maximum set-up height in a few locations at headlands that are fully open to the Gulf of Finland (Fig. 5). A likely reason for such a sporadic overestimation is the sensitivity of the formation of set-up with respect to the approach angle of waves. The magnitude of set-up rapidly decreases with an increase in the approach angle. This decrease is ignored in S2013. As a result, the height of set-up created by waves that approached at angles of 10–15 degrees with respect to shore-normal was overestimated. This feature also demonstrates the importance of the correct evaluation of refraction and shoaling.

Some differences between the results presented in this paper and those described in Soomere et al. (2013) and Pindsoo and Soomere (2015) stem from the different time intervals used in the calculations. Simulations for 1981–2012 indicate that the maximum set-up heights in coastal areas open to the east were mostly from the 1980s (Soomere et al., 2013) even though the maximum wave heights occurred starting from the mid-1990s. This feature may be related to a change in the directional structure of strong winds with easterly storms being relatively weak for about two decades. There is increasing evidence, however, that this process has reversed and strong easterly storms have returned to the area. Evidence of this change is the event with significant wave height 5.2 m that was recorded in the Gulf of Finland during an extreme easterly storm on 29–

30 November 2012 (Pettersson et al., 2013). Pindsoo and Soomere (2015) observed that many new all-time highest set-up events apparently occurred in coastal segments open to the east since 2012. This process has led to the generation of maximum simulated waves at a number of locations on the eastern Viimsi Peninsula near Leppneeme (grid cells 115–117 in Fig. 2) in 2010 (Fig. 6). These aspects will be addressed in more detail elsewhere.

3.2 Frequency of occurrence of set-up heights

As the modelled wave heights are evaluated with an accuracy of 1 cm, using even finer resolution for the construction of empirical probability distributions is not justified. The total number of entries in time series of positive set-up heights is 40 000–70 000. Therefore, using the resolution of 1 cm ensures that at least a few examples of set-up heights will belong to the relevant "size classes" of set-up heights down to frequencies about 10^{-2}%. This range is apparently large enough to estimate the main properties of the distributions in question.

The shape of the empirical distributions of the occurrence of set-up heights varies extensively in the study area (Fig. 7). This shape matches an exponential distribution in the majority (about 75%) of the model coastal segments[1]. Such a distribution is represented by a straight line in the semilogarithmic (log-linear) coordinates used in Figs. 7, 8. It apparently reflects a background Poisson process that also describes storm surges in the study area (Soomere et al., 2015). This type of distribution appears for coastal segments that are open to the common strong wave directions. For example, grid point 23 (Fig. 7a) is open to the northwest and north, grid point 96 (Fig. 7c) is located near the western shore of Viimsi Peninsula and is open to the west and northwest, and grid point 145 (Fig. 7f) is widely open to the directions from the northwest to the northeast.

The distribution is convex upwards at a few locations that are sheltered from most of predominant approach directions of strong waves, including north-north-west. This shape is evident in the most sheltered location of eastern Kopli Bay (grid point 43, Fig. 7b) and to a lesser extent in coastal sections sheltered by the island of Aegna (grid point 106, Fig. 6d). The relevant empirical distributions of set-up heights can be reasonably approximated by a two-parameter Weibull or Gaussian distribution. They both have a convex upwards shape in log-linear coordinates.

Deleted: ~~of which~~

A subset (about ¼) of the presented distributions exhibit a different, clearly concave upwards shape in log-linear coordinates. This feature is evident in coastal sections that are sheltered from a few predominant wave directions. Strong winds blow in this region usually from three directions: west, north-north-west, and east (Soomere et al., 2008). The segments that exhibit a concave upwards shape of the distribution (e.g., grid point 129, Fig. 6e) are mostly sheltered against waves that approach

---

[1] An early discussion version of this paper (available at https://www.earth-syst-dynam-discuss.net/esd-2016-76) contained a bug in the script for the calculation of set-up heights and for the subsequent evaluation of the parameters of their probability density function. This bug led to an erroneous conclusion about the frequency of occurrence of various kinds of distributions of set-up heights in different coastal sections as well as to severe overestimation of the frequency of matches of these distributions with an inverse Gaussian distribution.

from the west or north-north-west but are open to the east. However, segments that are widely open to all strong wave directions (e.g., grid point 1, Fig. 7) may also exhibit a concave upwards shape of the empirical distribution of set-up heights. This concave upwards appearance clearly differs from the shape of the usual distributions of the magnitude of wave phenomena (Fig. 8) such as the classic (Rayleigh) distribution of single wave heights (Longuet-Higgins, 1952), the Tayfun distribution of the heights of largest waves, the Weibull family of distributions for the occurrence of various wave conditions, or the Rayleigh distribution for run-up of (narrow-banded) Gaussian wave fields (Didenkulova et al., 2008). Therefore, none of the above-mentioned distributions can be used for the universal approximation of the probabilities of different set-up heights.

To further explore the shape of the distributions of set-up heights and their possible variations along the shoreline, we assume that these distributions belong to the family of general exponential distributions. The overall appearance of empirical distributions in log-linear coordinates (Fig. 7) suggests that their shape can be, as a first approximation, matched with a quadratic polynomial $az^2 + bz + c$, where z is the set-up height. In other words, the empirical probability density $P(z)$ is approximated by the following function:

$$P(z) = \exp\left(az^2 + bz + c\right) \tag{10}$$

The fitting is performed in log-linear coordinates used in Figs. 7, 8. In the case $a = 0$, distribution (10) reduces to a classic exponential distribution $\sim A\exp(bz)$. The values $a < 0$ correspond to convex upwards distributions that eventually can be approximated by a Weibull or normal distribution whereas $a > 0$ indicates that the distribution is locally concave upwards. A more important difference between distributions with $a = 0$ and $a \neq 0$ is in the nature of their tails. If $a = 0$, the probability of large set-up heights decreases as $\sim A\exp(bz)$ when $z >> 1$ while in the case $a < 0$ this probability decreases much faster, as $\sim A\exp\left(az^2\right)$. It is often said that the former distribution has a heavy tail (and a comparatively large probability of very large set-up heights) whereas the latter distribution has a light tail (and a lower probability of very large set-up heights). The case $a > 0$ is only possible locally for a certain range of set-up heights and serves as an indication that some larger set-up height are more probable that expected from distributions with $a \leq 0$.

Such a fitting procedure is not straightforward for several reasons. Firstly, the number of nonzero points of the distributions in Fig. 7 is highly variable along the study area similar to the variation in the typical magnitude of the set-up. Secondly, the relevant empirical distributions have gaps for some value(s) of the set-up height. A natural reason for this feature is that we are looking at very low probabilities (down to 0.001%, that is, a few occasions) of occurrence of relatively high set-up events in 1981–2016. Thirdly, a few locations have several outliers. There are remarkably high set-up events that do not follow the general appearance of the empirical distribution of set-up heights for the particular location (Fig. 7b, d, f). Such events apparently reflect severe storms in which the wind pattern was favourable for the development of very large waves that approached a certain coastal segment at a small angle. The presence of similar outliers is characteristic, for example, for time

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

series of sea level in Estonian waters (Suursaar and Sooäär, 2007) and is associated with situations when strong storms blow from a specific direction.

Deleted: . Similar to outliers in sea level time series, these events likely follow an extreme value distribution (Coles, 2004) and thus should be left out of the analysis of the shape of the distribution of set-up heights

To estimate the impact of these aspects on the results, we performed three versions of the fitting procedure. Firstly, we used all data points in the relevant distributions starting from the height of 0.01 m to evaluate the coefficients *a*, *b* and *c*. Secondly, we used for the same purpose only set-up heights from 0.01 m to 0.4 m (Fig. 7). This approach was not applicable in some locations where set-up heights did not reach 0.4 m. Thirdly, we evaluated these coefficients starting from the height of 0.01 m to the first gap in the empirical distribution (the lowest set-up height that did not occur in 1981–2016). Doing so made it possible to check whether the shape of the distribution is governed by the majority of events or if it is dominated by the presence of a few very large set-up heights (Fig. 8).

The particular values of the coefficients *a*, *b* and *c* depend to some extent on the chosen version (Fig. 7). The shape of the approximate distribution is invariant with respect to the particular choice. All distributions also match reasonably well data points corresponding to the largest set-up heights. The differences between the resulting theoretical distributions are mostly insignificant. The relevant estimates are located almost in the middle of the 95% confidence intervals of each other (Fig. 7).

The coefficients *a* at the leading term of the approximating polynomial (Fig. 9, 10) are mostly very small, in the range of (–0.005, 0.005). Their 95% confidence intervals normally include the zero value. This feature indicates that on most occasions the parameter *a* can be set to zero and the distribution of set-up heights can be reasonably approximated with an exponential distribution at a 95% level of statistical significance. On such occasions, the entire process can be adequately approximated by a Poisson process and the parameter *b* characterises the vulnerability of the particular coastal segment with respect to the set-up phenomenon similarly to the analysis of storm-driven high water levels (Soomere et al., 2015).

A few outliers of the parameter *a* in relatively sheltered coastal segments were negative and reached values down to –0.08 (Fig. 9). These values correspond to distributions with convex upwards shape in semilogarithmic coordinates and are thus qualitatively similar to the family of Gaussian or Weibull distributions.

3.3 Fitting the empirical distribution with theoretical distributions

Importantly, in about a quarter of the coastal segments in the study area, the parameter *a* is positive and its 95% confidence intervals do not include the zero value. In other words, in each of such locations, the leading term *a* of the quadratic polynomial $az^2 + bz + c$, estimated from the time series of wave properties for this location, is positive at a 95% significance level. Note that this estimate is valid for single points only and is not applicable for the set of such points. A positive leading term corresponds to the concave-up appearance of the relevant distributions of set-up heights for a certain range of these heights. This means that large set-up events may be systematically much higher and/or occur much more frequently than one could expect from the classic Gaussian, Weibull or Poisson-type statistics. The described features indicate that the empirical distribution of set-up heights can be, at least locally, approximated using an inverse Gaussian (Wald) distribution with a probability density function (Folks and Chhikara, 1978)

Deleted: these

Deleted: This feature

Deleted: very

Deleted: are

$$P=\sqrt{\frac{\lambda}{2\pi z^3}}\exp\left[-\frac{\lambda(z-\mu)^2}{2\mu^2 z}\right]. \qquad (11)$$

For a certain set of parameters λ (the shape parameter) and μ (the mean), a part of the graph of this function has a concave upward shape in semilogarithmic coordinates (Fig. 8). It thus fairly well approximates the empirical distributions of wave set-up at the relevant locations. For large values of z this function behaves similarly to the probability density function of an exponential distribution and thus, in the above-mentioned notion, has a “heavy” tail and signals that large set-up heights are more frequent than for a Gaussian or Weibull process.

Deleted: and

Deleted: Gaussian

All coefficients of the quadratic approximation of the exponent vary insignificantly along the study area. This is remarkable because the shape of the relevant Weibull distribution (and thus the shape parameter of this distribution) for different wave conditions varies considerably in the study area (Soomere, 2005). The variations in the leading coefficient a are uncorrelated with the values of maximum set-up heights in the study area. It is thus likely that the locations where an inverse Gaussian distribution governs the properties of set-up heights appear because of a specific match of the directional structure of winds and the orientation of the coastline. This feature also signals that the basic features of the distribution are only weakly (if at all) connected with the properties of local wave climate. This conjecture is supported by comparatively small variations in the values of other parameters in the polynomial approximation (Fig. 9b,c). The values of c are all positive and mostly in the range of 2.5–4 (Fig. 9). The values of parameter b are, as expected, almost everywhere negative, concentrated around –0.2, typically varying between –0.1 and –0.4. A few locations with positive values of this parameter correspond to large negative values of a.

To shed more light on which theoretical distribution is most suitable for the description of the probabilities of occurrence of wave set-up heights, we fitted the distributions using the *R* packages “fitdistrplus” for fitting any distribution to the data (Delignette-Muller and Dutang, 2005; v. 1.0-14), “actuar” (v. 2.3-3) for fitting with inverse Gaussian distribution, and “goft” (v. 1.3.4) to make an initial guess of the parameters of the inverse Gaussian distribution under *R* version 3.6.1. As the probability density function of an inverse Gaussian distribution is everywhere positive, this fit can only be applied to positive values of wave set-up, that is, for the cases when waves are propagating towards the shore.

As an example of the appearance of the fit by different theoretical distributions, we start from the analysis of the data in the westernmost point of the study area presented in Fig. 8. The underlying data set contains 3724 (~6%) negative values out of 61720 entries. The fit was performed for the set-up values in the range from 0.01 to 0.4 m (Fig. 11). Attempts to fit the data in question with an exponential, Gaussian and Weibull distributions (Fig. 11) demonstrate that these distributions do not replicate the location and height of the maximum and inaccurately follow the empirical distribution for larger set-up values. A variation in the higher and lower cutoff values leads to a certain change in the parameters of the fitted distribution change, but the inverse Gaussian fit remains appropriate for all cut-off values.

The goodness of fit of the cumulative distribution function (CDF) with different theoretical distributions is much better (Fig. 12). The empirical data do not follow a Gaussian distribution, whereas both Weibull and exponential distributions provide an acceptable match to the observed distributions. The inverse Gaussian shows the best agreement among the tested CDFs. As expected, the differences between all fitted distributions are relatively small for larger set-up heights but become more evident for the most frequently occurring set-up heights (Fig. 12).

We applied a Kolmogorov-Smirnov test to clarify which theoretical distribution describes the data in question at best. The lower is the D-value, the smaller is the difference between the distributions. The corresponding D-values from this test (Kolmogorov-Smirnov statistic) are 0.039 for the inverse Gaussian distribution, 0.067 for the Weibull distribution, 0.120 for the exponential distribution, and 0.158 for the Gaussian distribution. Therefore, an inverse Gaussian distribution at best matches the empirical distribution of set-up heights at this location. However, the hypothesis that the data and random points from the fitted inverse Gaussian distribution are from the same distribution was rejected, even at an 80% confidence level. However, the probability that the empirical distribution represents a data set with other tested theoretical distributions was much smaller. Therefore, even though there is no rigorous evidence that the empirical data follow an inverse Gaussian distribution, it provides the closest fit to the data compared with the other three tested distributions.

This proportion of the "remoteness" of all four distributions from the empirical probability distribution of set-up heights persists for the entire study area and different cut-off values (Fig. 13). The cut-off values were tested in the range from 40% to 95% of the maximum set-up height in a particular coastal section. In terms of D-values of the Kolmogorov-Smirnov test, an inverse Gaussian distribution shows a better approximation to the empirical distribution of set-up heights than all other distributions. The Weibull distribution overall provides a slightly worse fit except for several (<7%) coastline locations (Fig. 13). An exponential distribution, even though it seems to follow the empirical distribution in most occasions closely, has a smaller probability to describe set-up heights adequately. This feature apparently reflects the fact that an exponential distributions is a particular (one-parameter) case of the family of (two-parameter) Weibull distributions.

4 Discussion and conclusions

The analysis reveals that numerical estimates of maxima of wave set-up heights are relatively sensitive with respect to how the impact of radiation stress and the transformation of wave properties in the nearshore are evaluated. The magnitude of the related effects substantially depends on the bathymetry. Refraction can easily override the purely geometric effects of shoreline orientation changes and redirect substantial levels of wave energy into seemingly sheltered shore sections. This means that high-resolution information about wind (including wind directions) and bathymetry, together with advanced methods for the evaluation of propagation and impact of radiation stress in the nearshore in operational and hindcast models of coastal flooding, are required.

The core message from the analysis is that the empirical probability distribution of different set-up heights can usually be fairly well approximated by a standard exponential distribution $\exp(-\lambda z)$. When the exponent function of the general exponential distribution is approximated using a quadratic function, the coefficient of its leading term does not differ from zero at a 95% significance level for more than ¾ of the coastal segments of the study area. As the study area contains a variety of sections with different orientations and with radically different wave properties, it is likely that the qualitative shape of the distribution only weakly depends on the properties of the local wave climate.

Another important message is that the basic shape of this distribution function is concave upwards in a log-linear plot for a substantial number of coastal segments. The local shape of the relevant empirical distributions of wave set-up heights can be adequately approximated with a family of inverse Gaussian (Wald) distributions. Similarly to the exponential distribution, the probability density function of an inverse Gaussian distribution decays as $\exp(-\lambda z)$ for $z \gg 1$. Even though the absolute values of the coefficients of the leading term of such a quadratic approximation are relatively small, the goodness of fit with other classic distributions, such as Rayleigh or Weibull distributions (Fig. 8) that decay as $\exp\left(-\lambda z^2\right)$ for $z \gg 1$, is clearly worse. As the coefficient of the linear term of this quadratic approximation is relatively small (Fig. 9b), the use of a Lévy distribution might also be appropriate.

Field Code Changed

Field Code Changed

Deleted: approximations

Field Code Changed

Field Code Changed

Deleted: are evidently inappropriate

This result is intriguing because sensible approximations with inverse Gaussian (Wald) distributions are scarce in descriptions of geophysical phenomena. Perhaps the most well-known example of the use of a Wald distribution is to describe the time a Brownian motion (with positive drift) takes to reach a fixed positive level. Other examples include statistical properties of soil phosphorus (Manunta et al., 2002), long-distance seed dispersal by wind (Katul et al., 2005), and some models of failure (Park and Padgett, 2005).

The appearance of the distribution of modelled wave heights in the offshore (Fig. 3) is convex upwards in the range of relatively frequent wave heights of 0.5–1.7 m. It would thus be natural to expect that this property also becomes evident in setup heights. The distribution of measured wave heights largely follows an exponential distribution for 1.2–3.2 m high waves and is only slightly convex upwards for 0.5–1.5 m high waves. This difference in the distributions for modelled and measured wave heights suggests that the presence of convex upwards distributions of set-up heights in nature may be even more pronounced than is demonstrated in Figs. 7 and 9. This difference also signals that the approximations employed to evaluate wave properties are not responsible for the presence of convex upwards distributions of set-up heights. For example, a natural conjecture from Fig. 3 is that ignoring ice cover and the use of discontinuous wind data have at least partially supported the convex upwards shape of the distribution of wave heights.

Deleted: several

Deleted: described above

Some of the introduced assumptions such as the ideal plane and rigid seabed, the presence of a dry coast without any vegetation, and ignoring the wave period and the particular value of the coastal slope (and thus wave reflection) in the calculations are not fully realistic. In other words, it is assumed that all coastal segments are (i) favourable for the formation of high set-up and (ii) approximately homogeneous alongshore. These assumptions are only valid for selected segments of the study area. They all generally lead to an overestimation of set-up heights (Dean and Bender, 2006). As they impact set-up heights in the same manner, independently of wave properties, it is likely that they mainly stretch the resulting distributions of set-up heights towards larger values but do not modify their basic shapes. It might be expected that the impact of other simplifications such as the assumption of monochromatic wave fields, using a constant value of the breaking index and employing long wave approximation for breaking waves, generally emphasizes the role of approach directions.

Therefore, it is likely that the set of assumptions used makes the established features more noticeable than they would be for real wave fields.

Finally, we note that the presented results do not require any modification of the classic estimates of extreme values of set-up heights and their return periods based on, for example, the block maximum method. Namely, the limiting distributions of independent block maxima follow either a Gumbel, Weibull or Fréchet distribution notwithstanding the distribution of the underlying values (Coles, 2004). This general theorem is obviously also valid for any time series that follows an inverse Gaussian distribution. A subtle implication from the qualitative match of statistics of set-up heights with an inverse Gaussian distribution is that set-up events with heights close to extreme heights may be much more frequent than their estimates based on classic Gaussian or Weibull statistics and also clearly more frequent than similar estimates for Poisson processes. This increase in the probability of large wave set-up events is balanced by a similar decrease in the relative number of events with an average magnitude compared to normally or Weibull distributed events. The described features basically indicate that the frequency and role of close to extreme set-up events (and their contribution to damages and economic losses) may be underestimated based on observations of similar events of average height. In particular, severe set-up events may occur substantially more frequently that could be expected from the probability of the occurrence of severe seas.

Deleted: e

Code availability: From the authors on request (Matlab and *R* scripts).

Data availability: Kalbådagrund wind data are available from the Finnish Meteorological Institute.

Wave measurements: https://www.emodnet-physics.eu/map/platinfo/piroosplot.aspx?platformid=8974

Sample availability: not applicable.

Authors team: Tarmo Soomere, Katri Pindsoo, Nadezhda Kudryavtseva, and Maris Eelsalu,

Author contribution. T. Soomere designed the study, derived the equations and approximations used in the paper, produced Figure 5, compiled the introduction and discussion, checked the consistency of the results, and polished the text. K. Pindsoo developed the scripts, ran the simulations, produced most of graphics, and drafted the body parts of the manuscript. N. Kudryavtseva performed testing of fits of empirical data with different theoretical distributions, wrote the relevant parts of the text and prepared Figures 11–13. M. Eelsalu recalculated the graphics, performed a comparison of the properties of the reconstructed waves with measured waves near the study area, evaluated the distributions of set-up heights for time series of wave properties measured in the northern Baltic Proper, drafted the relevant parts of the manuscript, and produced updated maps.

Competing interests: The authors declare that they have no conflict of interest.

Disclaimer: All authors have approved the submitted version.

Acknowledgements

The research was financed by the institutional support of the Estonian Ministry of Education and Research (IUT33-3), the project “Sebastian Checkpoints – Lot 3 Baltic” of the call MARE/2014/09 and by the Estonian Research Council (ETAg) to two networks: the ERA-NET Rus+ network EXOSYSTEM and FLAG-ERA network FuturICT2.0. The authors are greatly thankful to the Finnish Meteorological Institute for making the Kalbådagrund wind data public, the Marine Systems Institute, Tallinn University of Technology, for providing wave data at Tallinnamadal, and to three anonymous referees for valuable suggestions towards improvement of the manuscript.

References

Alari, V. and Kõuts, T.: Simulating wave–surge interaction in a non-tidal bay during cyclone Gudrun in January 2005, in: Proceedings of the IEEE/OES Baltic 2012 International Symposium “Ocean: Past, Present and Future. Climate Change Research, Ocean Observation & Advanced Technologies for Regional Sustainability,” May 8–11, Klaipėda, Lithuania. IEEE Conference Publications, https://doi.org./10.1109/BALTIC.2012.6249185, 2012.

Arns, A., Wahl, T., Haigh, I. D., Jensen, J., and Pattiaratchi, C.: Estimating extreme water level probabilities: A comparison of the direct methods and recommendations for best practise, Coast. Eng., 81, 51–66, https://doi.org./10.1016/j.coastaleng.2013.07.003, 2013.

Averkiev, A. S. and Klevanny, K. A.: A case study of the impact of cyclonic trajectories on sea-level extremes in the Gulf of Finland, Cont. Shelf Res., 30, 707–714, https://doi.org/10.1016/j.csr.2009.10.010, 2010.

Babanin, A. V., Hsu, T.-W., Roland, A., Ou, S.-H., Doong, D.-J., and Kao, C. C.: Spectral wave modelling of Typhoon Krosa, Nat. Hazards Earth Syst. Sci., 11, 501–511, https://doi.org./10.5194/nhess-11-501-2011, 2011.

Batstone, C., Lawless, M., Tawn, J., Horsburgh, K., Blackman, D., McMillan, A., Worth, D., Laeger, S., and Hunt, T.: A UK best-practice approach for extreme sea-level analysis along complex topographic coastlines, Ocean Eng., 71, 28–39 https://doi.org/10.1016/j.oceaneng.2013.02.003, 2013.

Bechle, A. J., Kristovich, D. A. R., and Wu, C. H.: Meteotsunami occurrences and causes in Lake Michigan, J. Geophys. Res-Oceans, 120, 8422–8438, https://doi.org./10.1002/2015JC011317, 2015.

Björkqvist, J.-V., Lukas, I., Alari, V., van Vledder, G. Ph., Hulst, S., Pettersson, H., Behrens, A., and Männik, A.: Comparing a 41-year model hindcast with decades of wave measurements from the Baltic Sea, Ocean Eng., 152, 57–71, https://doi.org./10.1016/j.oceaneng.2018.01.048, 2018.

Bortot, P., Coles, S., and Tawn, J.: The multivariate Gaussian tail model: an application to oceanographic data, J. Roy. Stat. Soc. A-Appl. Stat. 49, 31–49, https://doi.org./10.1111/1467-9876.00177, 2000.

Deleted: :

Deleted:

Deleted: doi:

Deleted: doi:

Deleted: doi:

Deleted: .

Deleted: doi:

Deleted: .

Deleted: doi:

Caliskan, H. and Valle-Levinson, A.: Wind-wave transformations in an elongated bay Cont. Shelf Res., 28, 1702–1710, doi: 10.1016/j.csr.2008.03.009, 2008.

Cazenave, A., Dieng B., Meyssignac, B., von Schuckmann, K., Decharme, B., and Berthier, E.: The rate of sea-level rise, Nat. Clim. Change, 4, 358–361, https://doi.org/ 10.1038/nclimate2159, 2014.

Coles, S.: An introduction to statistical modeling of extreme values. Springer, 3rd printing, Springer, London, 208 pp., https://doi.org/10.1007/978-1-4471-3675-0, 2004.

Darwin, R. F. and Tol, R. S. J.: Estimates of the economic effects of sea level rise, Environ. Resource Economics, 19, 113–129, https://doi.org/ 0.1023/A:1011136417375, 2001.

Dean R.G. and Dalrymple R.A: Water Wave Mechanics for Engineers and Scientists. World Scientific, https://doi.org/10.1142/1232, 1991.

Dean, R. G. and Bender, C. J.: Static wave set-up with emphasis on damping effects by vegetation and bottom friction, Coast. Eng., 53, 149–165, https://doi.org/10.1016/j.coastaleng.2005.10.005, 2006.

Delignette-Muller, M. L., and Dutang, C.: fitdistrplus: An R package for fitting distributions, J. Statistical Software, 64(4), 1–34, https://doi.org/10.18637/jss.v064.i04, 2015.

Denissenko, P., Didenkulova, I., Pelinovsky, E., and Pearson, J.: Influence of the nonlinearity on statistical characteristics of long wave runup, Nonlin. Proc. Geophys., 18(6), 967–975, https://doi.org/ 10.5194/npg-18-967-2011, 2011.

Denissenko, P., Didenkulova, I., Rodin, A., Listak, M., and Pelinovsky, E.: Experimental statistics of long wave runup on a plane beach, J. Coast. Res., Special Issue 65, 195–200, https://doi.org/ 10.2112/SI65-034.1, 2013.

Didenkulova, I.: New trends in the analytical theory of long sea wave runup. In: Applied Wave Mathematics (Quak, E. and Soomere, T., eds.), Springer, 265–296, https://doi.org/ 10.1007/978-3-642-00585-5_14, 2009.

Didenkulova, I., Pelinovsky, E., and Sergeeva, A.: Statistical characteristics of long waves nearshore Coast. Eng. 58, 94–102, https://doi.org/ 10.1016/j.coastaleng.2010.08.005, 2008.

Dube, S. K., Jain, I., Rao, A. D., and Murty, T. S.: Storm surge modelling for the Bay of Bengal and Arabian Sea, Nat. Hazards, 51, 3–27, https://doi.org/ 10.1007/s11069-009-9397-9, 2009.

Dukhovskoy, D.S. and Morey, S. L.: Simulation of the Hurricane Dennis storm surge and considerations for vertical resolution, Nat. Hazards, 58, 511–540, https://doi.org/ 10.1007/s11069-010-9684-5, 2011.

Fawcett, L. and Walshaw, D.: Sea-surge and wind speed extremes: optimal estimation strategies for planners and engineers, Stoch. Environ. Res. Risk Assess., 30, 463–480, https://doi.org/ 10.1007/s00477-015-1132-3, 2016.

Feng, X., Tsimplis, M. N., Quartly, G. D., and Yelland, M. J.: Wave height analysis from 10 years of observations in the Norwegian Sea. Cont. Shelf. Res. 72, 47–56, https://doi.org/ 10.1016/j.csr.2013.10.013, 2013.

Forristall, G. Z.: Statistical distribution of wave heights in a storm, J. Geophys. Res-Oceans, 83(NC5), 2353–2358, 1978.

Deleted: ,

Deleted: .

Deleted: doi:

Deleted: ,

Deleted: doi:

Deleted: ,

Deleted: ,

Deleted: doi:

Deleted: doi:

Deleted: doi:

Deleted: .

Deleted: doi:

Deleted: doi:

Deleted: ,

Deleted: doi:

Deleted: ,

Deleted: :

Deleted: DOI

Deleted: -

Deleted: doi:

Folks, J. L. and Chhikara, R. S.: The inverse Gaussian distribution and its statistical application–A review, J. R. Stat. Soc. Ser. B-Methodol., 40, 263–289, https://doi.org/ 10.2307/2984691, 1978.

Deleted: .

Deleted: doi:

Geist, E. L., ten Brink, U. S., and Gove, M.: A framework for the probabilistic analysis of meteotsunamis, Nat. Hazards, 74, 123–142, https://doi.org/ 10.1007/s11069-014-1294-1, 2014.

Deleted: doi:

Hallegatte, S., Green, C., Nicholls, R. J., and Corfee-Morlot, J.: Future flood losses in major coastal cities, Nat. Clim. Change, 3, 802–806, https://doi.org/10.1038/nclimate1979, 2013.

Hawkes, P. J., Gouldby, B. R., Tawn, J. A., and Owen, M. W.: The joint probability of waves and water levels in coastal engineering design, J. Hydraulic Res., 40, 241–251, https://doi.org/10.1080/00221680209499940, 2002.

Haigh, I. D., Nicholls, R., and Wells, N.: A comparison of the main methods for estimating probabilities of extreme still water levels, Coast. Eng., 57, 838–849, https://doi.org/10.1016/j.coastaleng.2010.04.002, 2010.

Hoeke, R. K., McInnes, K. L., Kruger, J. C., McNaught, R. J., Hunter, J. R., and Smithers, S. G.: Widespread inundation of Pacific islands triggered by distant-source wind-waves, Glob. Planet Change, 108, 128–138, https://doi.org/10.1016/j.gloplacha.2013.06.006, 2013.

Holland, K. T. and Holman, R. A.: The statistical distribution of swash maxima on natural beaches, J. Geophys. Res.–Oceans, 98(C6), 10271–10278, https://doi.org/ 10.1029/93JC00035, 1993.

Deleted: doi:

Hsu, T.-W., Hsu, J. R.-C., Weng, W.-K., Wang, S.-K., and Ou, S.-H.: Wave setup and setdown generated by obliquely incident waves, Coast. Eng., 53, 865–877, https://doi.org/10.1016/j.coastaleng.2006.05.002, 2006.

Johansson, M., Boman, H., Kahma, K., and Launiainen, J.: Trends in sea level variability in the Baltic Sea, Boreal Environ. Res. 6, 159–179, 2001.

Katul, G. G., Porporato, A., Nathan, R., Siqueira, M., Soons, M. B., Poggi, D., Horn, H. S., and Levin, S. A.: Mechanistic analytical models for long-distance seed dispersal by wind, American Naturalist, 166, 368–381, https://doi.org/ 10.1086/432589, 2005.

Deleted: doi:

Keevallik, S. and Soomere, T.: Towards quantifying variations in wind parameters across the Gulf of Finland, Estonian J. Earth Sci., 59, 288–297, https://doi.org/10.3176/earth.2010.4.05, 2010.

Deleted: ,

Deleted:

Kulikov, E. A. and Medvedev, I. P.: Variability of the Baltic Sea level and floods in the Gulf of Finland, Oceanology, 53, 145–151, https://doi.org/ 10.1134/S0001437013020094, 2013.

Deleted: ,

Deleted: doi:

Larson, M., Hoan, L. X., and Hanson, H.: Direct formula to compute wave height and angle at incipient breaking, J. Waterw. Port C-ASCE, 136, 119–122, doi: 10.1061/ASCEWW.1943-5460.0000030, 2010.

Deleted: .

Leijala, U., Björkqvist, J./V., Johansson, M. M., Pellikka, H., Laakso, L., Kahma, K. K.: Combining probability distributions of sea level variations and wave run-up to evaluate coastal flooding risks, Nat. Hazards Earth Syst. Sci., 18, 2785–2799, https://doi.org/10.5194/nhess-18-2785-2018, 2018.

Longuet-Higgins, M. S.: On the statistical distribution of the heights of sea waves, J. Mar. Res., 11, 1245–1266, 1952.

Longuet-Higgins, M. S. and Stewart, R. W.: Radiation stress and mass transport in gravity waves with application to ”surf-beats”, J. Fluid Mech., 8, 565–583, https://doi.org/10.1017/S0022112062000877, 1962.

Longuet-Higgins, M. S. and Stewart, R. W.: Radiation stresses in water waves: a physical discussion with applications, Deep-Sea Res., 11, 529–562, https://doi.org/10.1016/0011-7471(64)90001-4, 1964.

Lopez-Ruiz, A., Ortega-Sanchez, M., Baquerizo, A., and Losada, M. A.: A note on alongshore sediment transport on weakly curvilinear coasts and its implications, Coast. Eng., 88, 143–153, https://doi.org/ 10.1016/j.coastaleng.2014.03.001, 2014.

Lopez-Ruiz, A., Solari, S., Ortega-Sanchez, M., and Losada, M.: A simple approximation for wave refraction – Application to the assessment of the nearshore wave directionality, Ocean Modell., 96, 324–333, https://doi.org/ 10.1016/j.ocemod.2015.09.007, 2015.

Manunta, P., Feng, Y., Goddard, T., Anderson, A. M., and Cannon, K.: Analysis of soil test phosphorus to assess the risk of P transport in a watershed, Comm. Soil Sci. Plant Analysis, 33, 3481–3492, https://doi.org/ 10.1081/CSS-12001452, 2002.

McDougal, W. G., and Hudspeth, R. T.: Wave setup/setdown and longshore current on non-planar beaches, Coast. Eng., 7, 103–117, https://doi.org/10.1016/0378-3839(83)90007-8, 1983.

Masina, M., Lamberti, A., and Archetti, R.: Coastal flooding: A copula based approach for estimating the joint probability of water levels and waves, Coast. Eng., 97, 37–52, https://doi.org/ 10.1016/j.coastaleng.2014.12.010, 2015.

Mel, R. and Lionello, P.: Verification of an ensemble prediction system for storm surge forecast in the Adriatic Sea, Ocean Dyn., 64, 1803–2814. https://doi.org/ 10.1007/s10236-014-0782-x, 2014.

Melet, A., Almar, R., and Meyssignac, B.: What dominates sea level at the coast: a case study for the Gulf of Guinea, Ocean Dyn., 66, 623–636, https://doi.org/ 10.1007/s10236-016-0942-2, 2016.

Melet, A., Meyssignac, B., Almar, R., and Le Cozannet, G.: 2018. Under-estimated wave contribution to coastal sea-level rise, Nature Clim. Change, 8, 234–239, https://doi.org/ 10.1038/s41558-018-0088-y, 2018.

Meyer, V., Becker, N., Markantonis, V., Schwarze, R., van den Bergh, J. C. J. M., Bouwer, L. M., Bubeck, P., Ciavola,P., Genovese, E., Green, C., Hallegatte, S., Kreibich, H., Lequeux, Q., Logar, I., Papyrakis, E., Pfurtscheller, C., Poussin, J., Przyluski, V., Thieken, A. H., and Viavattene, C.: Assessing the costs of natural hazards – state of the art and knowledge gaps, Nat. Hazards Earth Syst. Sci., 13, 1351–1373, https://doi.org/ 10.5194/nhess-13-1351-2013, 2013.

Moghimi, S., Klingbeil, K., Gräwe, U., and Burchard, H.: A direct comparison of a depth-dependent Radiation stress formulation and a Vortex force formulation within a three-dimensional coastal ocean model, Ocean Modell., 70, 132–144, https://doi.org/ 10.1016/j.ocemod.2012.10.002, 2013.

Deleted: ,

Deleted: ,

Deleted: doi:

Deleted: doi:

Deleted: .

Deleted: doi:

Deleted:

Deleted: doi:

Deleted: ,

Deleted: doi:

Deleted: doi:

Deleted: .

Deleted: doi:

Deleted: ;

Deleted: Review article

Deleted: doi:

Deleted: http://dx.doi.org/

Muraleedharan, G., Rao, A. D., Kurup, P. G., Nair, N. U., and Sinha, M.: Modified Weibull distribution for maximum and significant wave height simulation and prediction, Coast. Eng. 54, 630–638, https://doi.org/10.1016/j.coastaleng.2007.05.001, 2007.

Deleted: doi:

Nikolkina, I., Soomere, T., and Räämet, A.: Multidecadal ensemble hindcast of wave fields in the Baltic Sea. In: The 6th IEEE/OES Baltic Symposium Measuring and Modeling of Multi-Scale Interactions in the Marine Environment, May 26–29, Tallinn Estonia. IEEE Conference Publications, 9 pp., https://doi.org/10.1109/BALTIC.2014.6887854, 2014.

Deleted: doi:

O'Grady, J. G., McInnes, K. L., and Hoeke, R. K.: Forecasting maximum wave setup hazards around Australia, Coasts & Ports Conference 2015, 15–18 September 2015, Pullman Hotel, Auckland, 636–641, 2015.

Deleted: .

Olbert, A. I., Nash, S., Cunnane, C., and Hartnett, M.: Tide-surge interactions and their effects on total sea levels in Irish coastal waters, Ocean Dyn., 63, 599–614, https://doi.org/10.1007/s10236-013-0618-0, 2013.

Orimolade, A. P., Haver, S., and Gudmestad, O. T.: Estimation of extreme significant wave heights and the associated uncertainties: A case study using NORA10 hindcast data for the Barents Sea, Marine Structures 49. 1e17, https://doi.org/10.1016/j.marstruc.2016.05.004, 2016.

Deleted: http://dx.doi.org/

Park, C. and Padgett, W. J.: Accelerated degradation models for failure based on geometric Brownian motion and gamma processes, Lifetime Data Anal., 11, 511–527, https://doi.org/10.1007/s10985-005-5237-8, 2005.

Deleted: doi:

Pattiaratchi, C. and Wijeratne, E. M. S.: Observations of meteorological tsunamis along the south-west Australian coast, Nat. Hazards, 74, 281–303, https://doi.org/10.1007/s11069-014-1263-8, 2014.

Deleted: ,

Pellikka, H., Rauhala, J., Kahma, K.K., Stipa, T., Boman, H., and Kangas, A.: Recent observations of meteotsunamis on the Finnish coast, Nat. Hazards, 74, 197–215, https://doi.org/10.1007/s11069-014-1150-3, 2014.

Deleted: doi:

Pettersson, H., Lindow, H., and Brüning, T.: Wave climate in the Baltic Sea 2012, HELCOM Baltic Sea Environment Fact Sheets 2012. http://helcom.fi/baltic-sea-trends/environment-fact-sheets/hydrography/wave-climate-in-the-baltic-sea/, 2013.

Deleted: .

Pindsoo, K. and Soomere, T.: Contribution of wave set-up into the total water level in the Tallinn area, Proc. Estonian Acad. Sci., 64(3S), 338–348, https://doi.org/10.3176/proc.2015.3S.03, 2015.

Deleted: ,

Deleted: doi:

Purvis, M. J., Bates, P. D., and Hayes, C. M.: A probabilistic methodology to estimate future coastal flood risk due to sea level rise, Coast. Eng. 55, 1062–1073, https://doi.org/10.1016/j.coastaleng.2008.04.008, 2008.

Deleted: doi:

Pycroft, J., Abrell, J., and Ciscar, J. C.: The global impacts of extreme sea-level rise: A comprehensive economic assessment, Environ. Resource Economics, 64, 225–253, https://doi.org/10.1007/s10640-014-9866-9, 2016.

Deleted: doi:

Roland, A., Cucco, A., Ferrarin, C., Hsu, T.-W., Liau, J.-M., Ou, S.-H., Umgiesser, G., and Zanke, U.: On the development and verification of a 2-D coupled wave-current model on unstructured meshes, J. Mar. Syst., 78, S244–S254, https://doi.org/10.1016/j.jmarsys.2009.01.026, 2009.

Rueda, A., Camus, P., Mendez, F. J., Tomas, A., and Luceno, A.: An extreme value model for maximum wave heights based on weather types, J. Geophys. Res-Oceans, 121, 1262–1273, https://doi.org/ 10.1002/2015JC01095, 2016a.

Deleted: doi:

Rueda, A., Camus, P., Tomas, A., Vitousek, S., and Mendez, F. J.: A multivariate extreme wave and storm surge climate emulator based on weather patterns, Ocean Modell., 104, 242–251, https://doi.org/ 10.1016/j.ocemod.2016.06.008, 2016b.

Deleted: .

Deleted: doi:

Sallenger Jr, A. H., List, J. H., Gelfenbaum, G., Stumpf, R. P., and Hansen, M.: Large wave at Daytona Beach, Florida, explained as a squall-line surge, J. Coast. Res., 11, 1383–1388, https://www.jstor.org/stable/4298439, 1995.

Schmitt, F. G., Crapoulet, A., Hequette, A., and Huang, Y.: Nonlinear dynamics of the sea level time series in the eastern English Channel, Nat. Hazards, 91(1), 267–285, https://doi.org/ 10.1007/s11069-017-3125-7, 2018.

Deleted: doi:

Seifert, T., Tauber, F., and Kayser, B.: A high resolution spherical grid topography of the Baltic Sea, 2nd edition, Baltic Sea Science Congress, Stockholm 25–29 November 2001, Poster 147, http://www.io-warnemuende.de/topography-of-the-baltic-sea.html, 2001.

Serafin, K. A., and Ruggiero, P.: Simulating extreme total water levels using a time-dependent, extreme value approach, J. Geophys. Res-Oceans, 119, 6305–6329, https://doi.org/ 10.1002/2014JC010093, 2014.

Deleted: doi:

Socquet-Juglard, H., Dysthe, K., Trulsen, K., Krogstad, H. E., and Liu, J. D.: Probability distributions of surface gravity waves during spectral changes, J. Fluid Mech., 542, 195–216, https://doi.org/ 10.1017/S0022112005006312, 2005.

Deleted: doi:

Soomere, T.: Anisotropy of wind and wave regimes in the Baltic Proper, J. Sea Res., 49, 305–316, https://doi.org/10.1016/S1385-1101(03)00034-0, 2003.

Soomere, T.: Wind wave statistics in Tallinn Bay, Boreal Environ. Res., 10, 103–118, 2005.

Soomere, T., Myrberg, K., Leppäranta, M., and Nekrasov, A.: The progress in knowledge of physical oceanography of the Gulf of Finland: a review for 1997–2007, Oceanologia, 50(3), 287–362, 2008.

Soomere, T., Pindsoo, K., Bishop, S. R., Käärd, A, and Valdmann, A.: Mapping wave set-up near a complex geometric urban coastline, Nat. Hazards Earth Syst. Sci., 13, 3049–3061, https://doi.org/ 10.5194/nhess-13-3049-2013, 2013.

Deleted: .

Deleted: doi:

Soomere, T., Eelsalu, M., Kurkin, A., and Rybin, A.: Separation of the Baltic Sea water level into daily and multi-weekly components, Cont. Shelf Res., 103, 23–32, https://doi.org/ 10.1016/j.csr.2015.04.018, 2015.

Deleted: doi:

Stockdon, H. F., Holman, R. A., Howd, P. A., and Sallenger, A. H.: Empirical parameterization of setup, swash, and runup, Coast. Eng., 53, 573–588, https://doi.org/10.1016/j.coastaleng.2005.12.005, 2006.

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

Deleted: .

Deleted: doi:

[SWAN]: SWAN Technical Documentation. Delft University of Technology. The Netherlands, 2007.

Tuomi, L., Kahma, K. K., and Pettersson, H.: Wave hindcast statistics in the seasonally ice-covered Baltic Sea, Boreal Environ. Res., 16, 451–472, 2011.

Tuomi, L., Kahma, K. K., and Fortelius, C.: Modelling fetch-limited wave growth from an irregular shoreline, J. Mar. Syst., 105, 96–105, https://doi.org/10.1016/j.jmarsys.2012.06.004, 2012.

Vilibic, I.: The role of the fundamental seiche in the Adriatic coastal floods, Cont. Shelf Res., 26, 206–216, https://doi.org./10.1016/j.csr.2005.11.001, 2006.

Vilibic, I., Monserrat, S., and Rabinovich, A. B.: Meteorological tsunamis on the US East Coast and in other regions of the World Ocean, Nat. Hazards, 74, 1–9, https://doi.org./10.1007/s11069-014-1350-x, 2014.

Viška, M., and Soomere, T.: Simulated and observed reversals of wave-driven alongshore sediment transport at the eastern Baltic Sea coast, Baltica, 26 (2), 145–156, https://doi.org./10.5200/baltica.2013.26.15, 2013.

Wadey, M. P., Brown, J. M., Haigh, I. D., Dolphin, T., and Wisse, P.: Assessment and comparison of extreme sea levels and waves during the 2013/14 storm season in two UK coastal regions, Nat. Hazards Earth Syst. Sci., 15, 2209–2225, https://doi.org./10.5194/nhess-15-2209-2015, 2015.

Deleted:

Deleted: doi:

Deleted: doi:

Deleted: doi:

Deleted: doi:

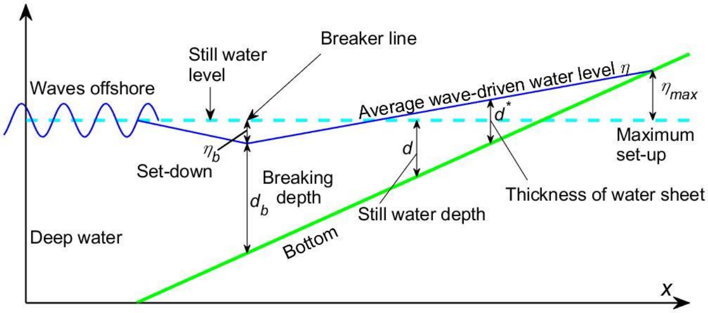


Figure 1: Scheme of the cross-section of a coastal area with wave set-up. The sign of η is positive if the average wave-driven water level exceeds the still water level and negative in the opposite case. The sign of d is positive in the area covered by still water and negative in the otherwise dry section of the coast. The quantity d^* is non-negative.

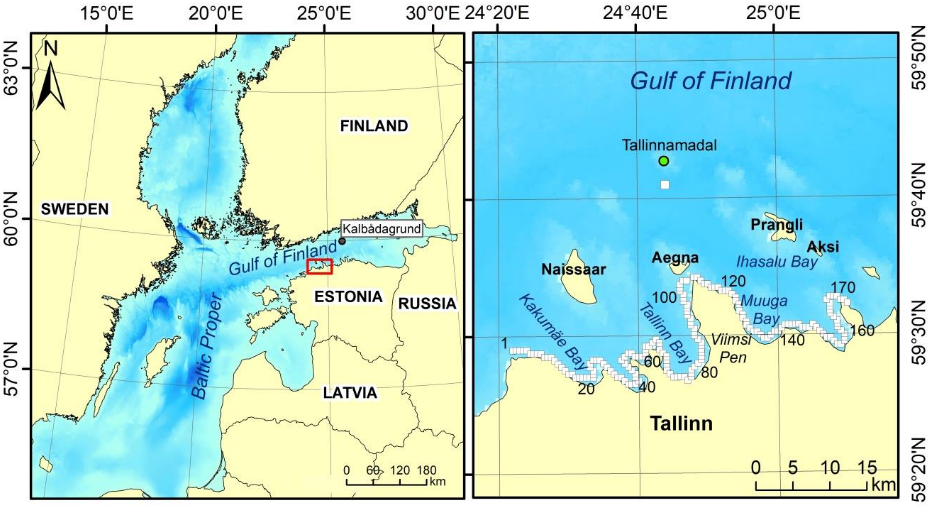


Figure 2: Study area (red box in the left panel) in the vicinity of Tallinn Bay. Small circles along the shoreline in the right panel indicate the nearshore grid cells of the wave model WAM with a resolution of about 470 m. The grid cells are numbered consecutively from the west to the east. The green circle shows the location of the wave buoy at Tallinnamadal and the white square to the south of it the closest grid point of the wave model used for comparison of modelled and measured wave data.

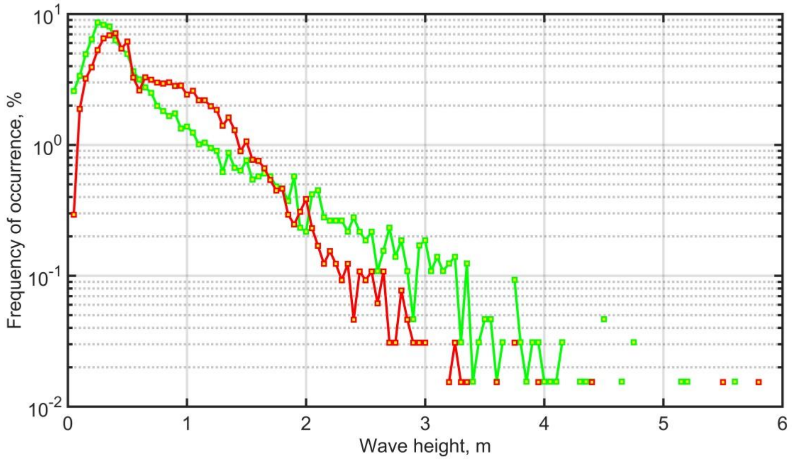


Figure 3: Empirical probability distributions of measured wave heights at Tallinnamadal (green) and modelled wave heights in the closest grid cell (red) with a resolution of 1 cm in 2012–2016. The missing lines between the data points indicate gaps in the sets of frequencies.

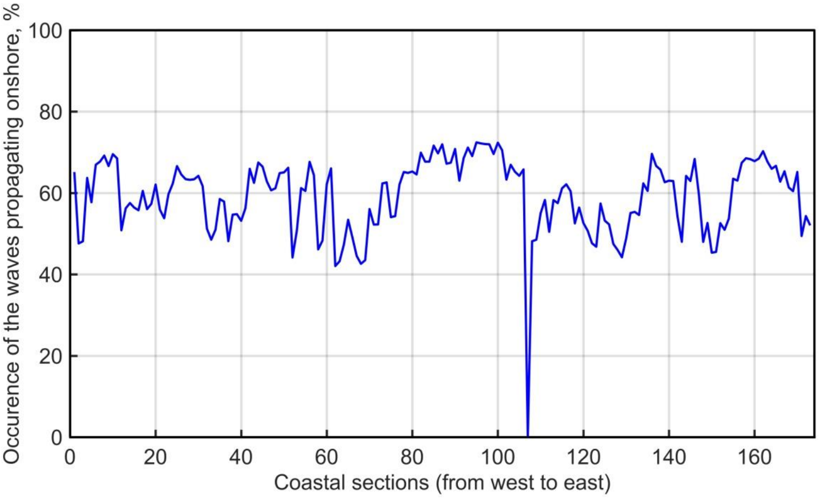


Figure 4: The percentage of occurrence of waves that propagate onshore and produce elevated wave set-up events in the study area.

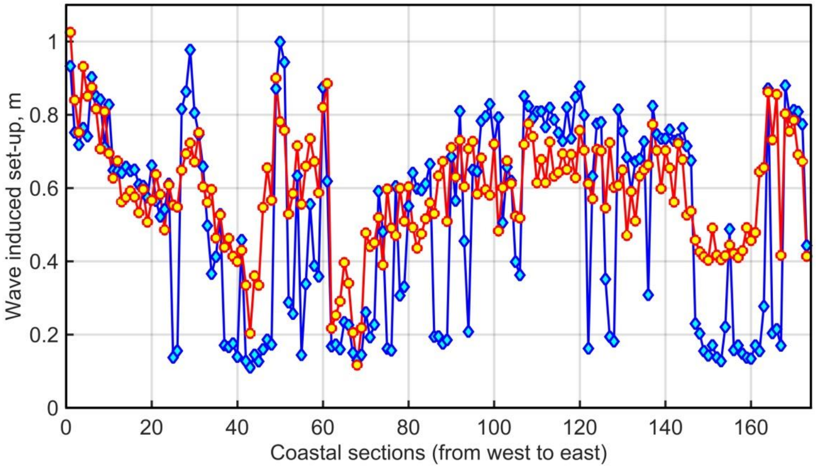


Figure 5: Maximum set-up heights evaluated using all onshore-propagating wave fields and Eqs. (1–8) (red circles) and similar heights evaluated using only those waves that approach the shore at an angle less than ±15° from shore-normal (blue diamonds).

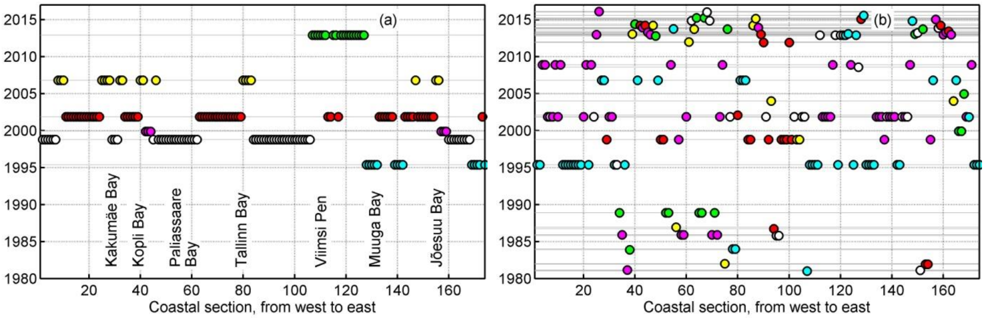
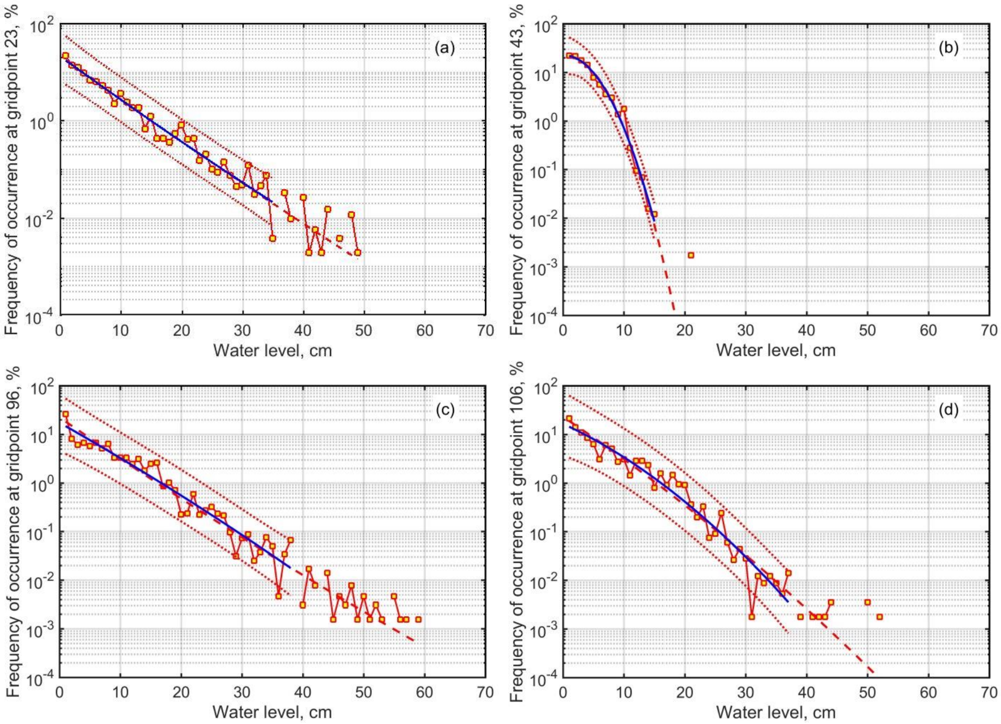


Figure 6: (a) Six storms that caused the highest waves in different coastal sections of the study area in January 1981–May 2016. Notice the cluster of green circles along the eastern coast of Viimsi Peninsula in an autumn storm of 2013. (b) 58 storms that caused the highest wave set-up in these sections in January 1981–May 2016. The set-up heights are evaluated similarly to the procedure in (Pindsoo and Soomere, 2015) using only waves approaching at an angle ±15° with respect to shore-normal. Colours vary cyclically in time and correspond to different storms in single years.



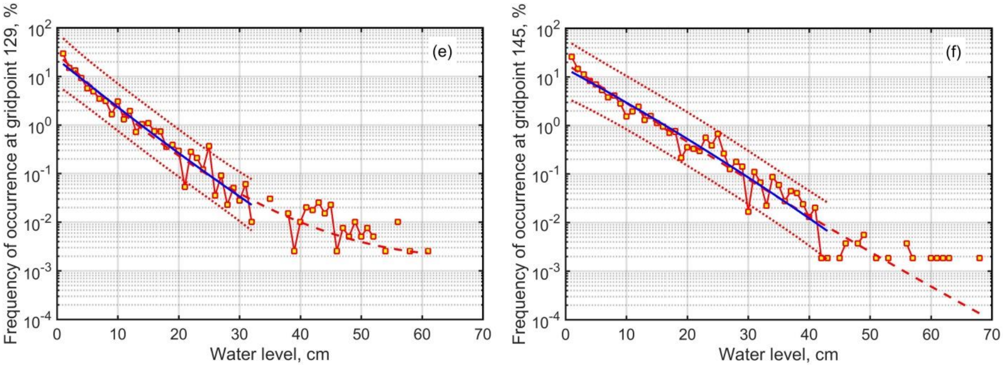


Figure 7: Simulated distributions of various set-up heights (red squares) at various locations in the Tallinn Bay area open to different directions. Blue line: interpolation with a quadratic function from the set-up height of 0.01 m to the first gap in the empirical distribution; red dotted lines: its 95% confidence intervals; red dashed line: similar interpolation using all data points. The interpolating lines evaluated using only the data points from 0.01 m to 0.4 m are fully masked by blue lines. The resolution of all distributions is 1 cm.

Deleted: 3

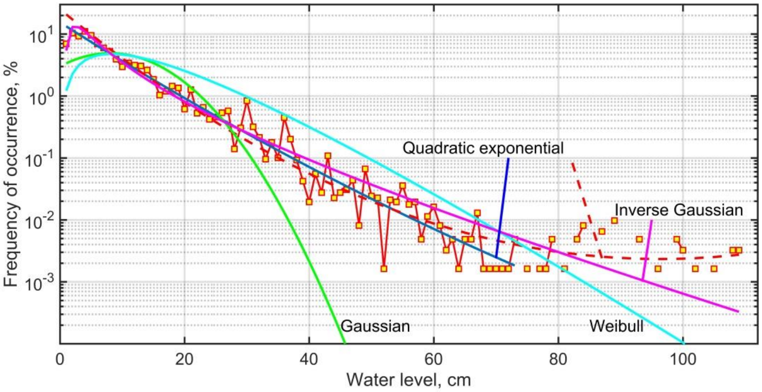


Figure 8: Simulated distributions of various set-up heights (red squares) with a resolution of 1 cm in the westernmost coastal segment of the study area (grid point 1) and the relevant Gaussian, Weibull and inverse Gaussian (Wald) distributions evaluated using the method of moments. Blue line: interpolation of the empirical distribution in semilogarithmic coordinates with a quadratic function (equivalently, the formal local exponential distribution with a general quadratic exponent) from the set-up height of 0.01 m to the first gap in the empirical distribution; red dashed line: similar interpolation using all data points.

Deleted: a

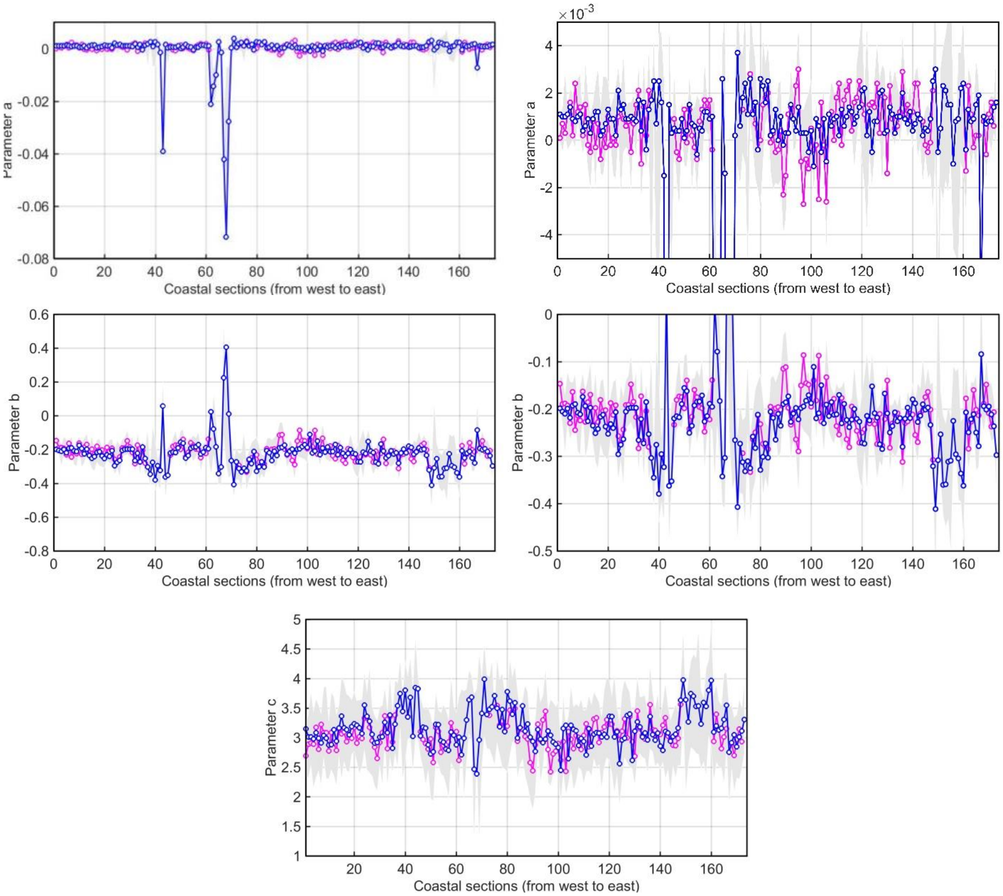


Figure 9: Alongshore variation of the coefficients *a*, *b*, *c* of the quadratic approximation $az^2 + bz + c$ of the exponent of empirical distributions of set-up heights in the Tallinn Bay area. For the parameters *a* and *b*, more detailed alongshore variation is presented in graphs with vertically stretched scales. Blue line: the respective parameter calculated for the range of set-up heights from 0.01 m to the first gap in the empirical distribution; the grey area marks the 95% confidence interval of this value, the pink line describes the values of the relevant parameter for the range of set-up heights from 0.01 m to 0.04 m.

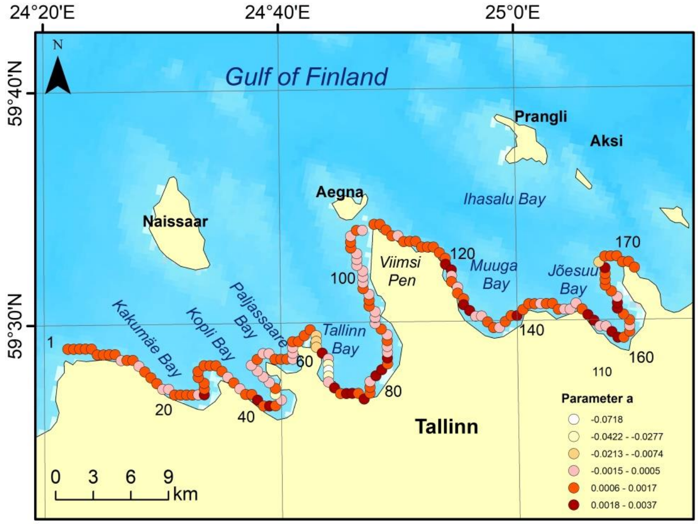
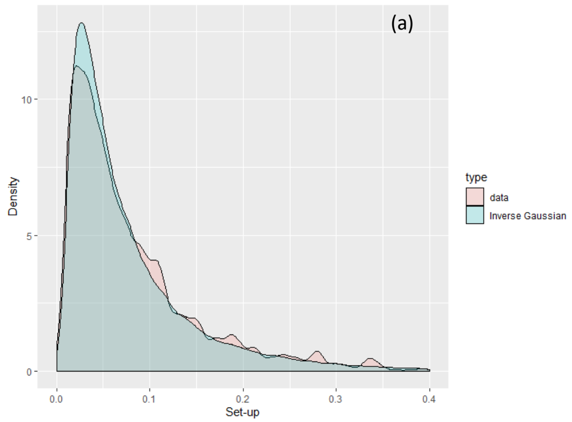


Figure 10: Alongshore variation in the coefficient of the leading term (colour code) in the approximation of the exponent of empirical distribution of set-up heights at single locations. The grid cells are numbered consecutively from the west to the east.



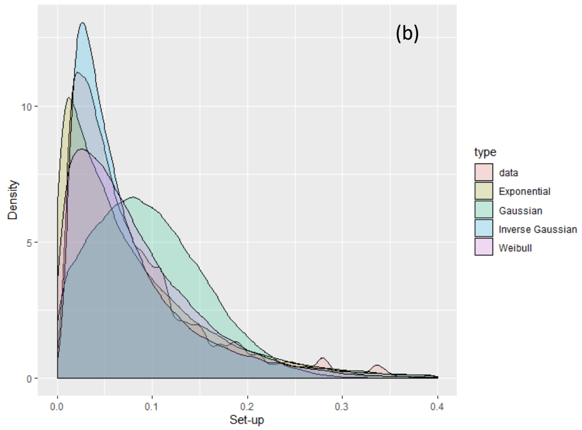
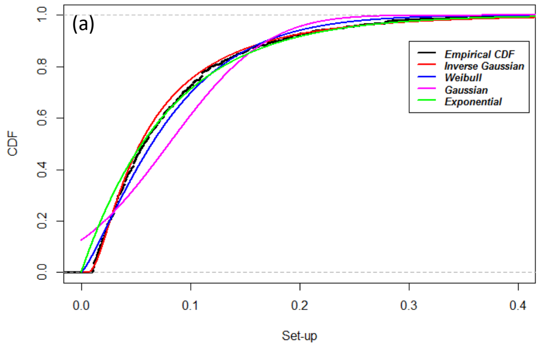


Figure 11. (a) The best fit of the data set presented in Fig. 8 with an inverse Gaussian distribution with the mean $\mu = 0.0804 \pm 0.0004$ and shape $\lambda = 0.0793 \pm 0.0005$ and the best fit of the same data set with an exponential (rate 12.44±0.06), Gaussian (mean 0.0804±0.0003, std 0.0700± 0.0002) and Weibull (shape parameter 1.247±0.004, scale parameter 0.0867±0.0003) distribution.

Field Code Changed

Field Code Changed



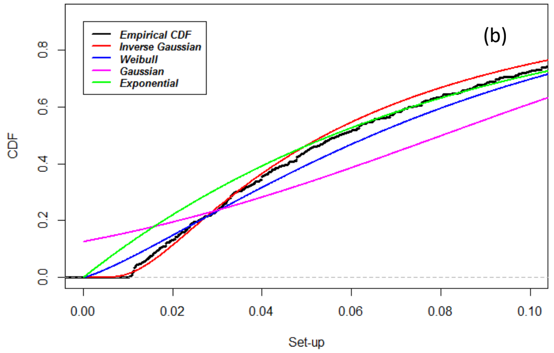
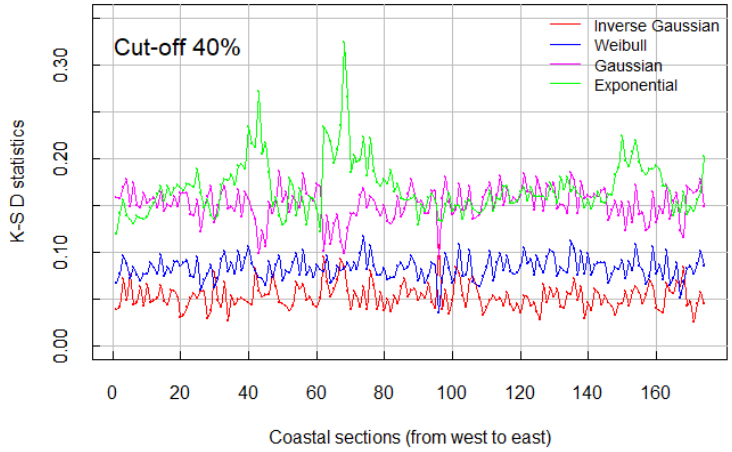


Figure 12: (a) Match of the cumulative distribution function of the empirical distribution of occurrence of different set-up heights for the data presented in Fig. 8 with inverse Gaussian, Weibull, Gaussian and exponential distributions. (b) The same match for set-up heights <0.11 m.



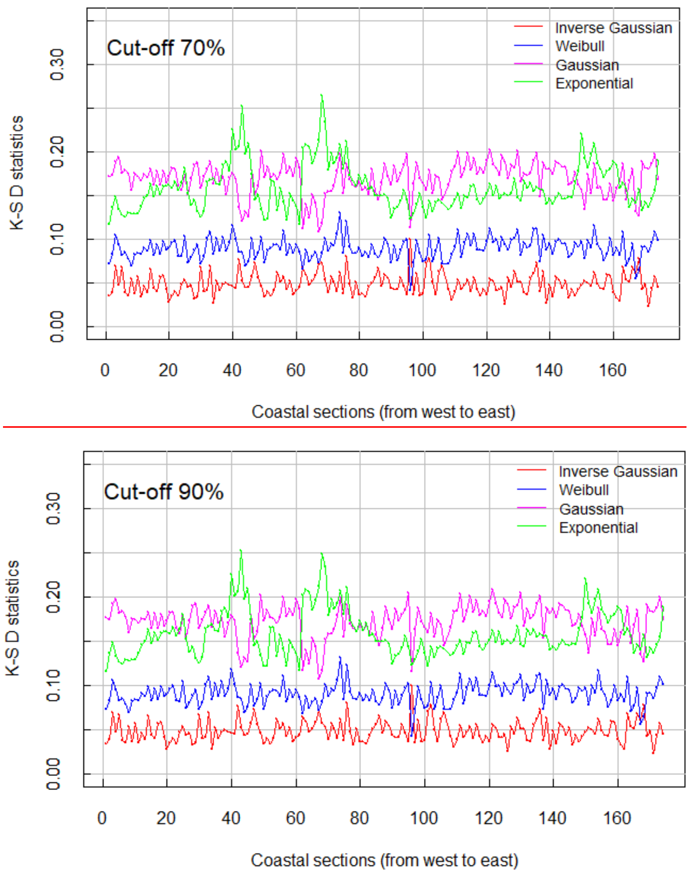


Figure 13: Alongshore variation in the D-value of Kolmogorov-Smirnov test for the goodness of fit of the empirical probability distribution of set-up heights with an inverse Gaussian, Weibull, exponential, and Gaussian distribution for different cut-off values (40%, 70% or 90% of the maximum set-up height in a particular coastal section).