Reply to review #1

We thank Referee #1 for the constructive comments that have helped us to clarify and improve many aspects of our manuscript. In the following, we explain how we plan to address the individual issues raised by the reviewer in the revised manuscript.

1. General Comments

1A We thank the reviewer for this comment. We will add a proper definition of rogue waves and discuss their relevance, as suggested.

1B We thank the reviewer for pointing out this ambiguity in the manuscript. We will draw a clear connection of Rayleigh and Weibull distributions to linear superposition and add an additional reference for second-order models to the text.

1C We appreciate the critical review of the paragraph. We agree that some formulations are misleading and we will re-formulate the passage in agreement with the suggestions of the reviewer.

1D The word “respective” will be removed, as suggested by the reviewer.

1E We thank the reviewer for pointing out additional important conclusions in the literature. We will add these to our introduction.

1F We agree with the reviewer. We will add the suggested references accordingly.

1G We agree with the reviewer. We will add the suggested references accordingly.

1H We understand the objection of the reviewer. In our text, we will replace the expression “nonlinearity” by “nonlinear processes”, where appropriate.

1I We thank the reviewer for the additional reference. We will add it accordingly.

1J We agree with the reviewer on the contradiction in the text. We will rephrase the sentence in question accordingly.

1K We thank the reviewer for several additional important references concerning rogue wave occurrence on varying bathymetry. We will extend the discussion and include the suggested references.

1L We thank the reviewer for this remark. We will replace the unpublished reference by publically available literature.

2. Scientific Improvement

2A We thank the reviewer for this detailed assessment and the comparison with the previous study of Teutsch et al. (2020). The reviewer is completely right that Figures 7 and 9 in Teutsch et al. (2020) do not support the conclusion that rogue wave frequencies are overestimated by the Forristall distribution. However, in these Figures data from different types of instruments (radar and wave buoys) are considered jointly. Figure 2 in Teutsch et al. (2020) clearly shows that data from both instruments show different behaviour with rogue wave frequencies in the radar/buoy data set being higher/lower than that derived from the Forristall distribution. The only exception here were the results from the buoy SEE off Norderney, which showed results comparable to those derived from radar data. This rendered the station SEE outstanding and provided the motivation for this study.

2B We thank the reviewer for his/her discussion of the Ursell number formulation. As suggested, we will rewrite the equation and state in the text that different definitions of the Ursell number exist, which will lead to different threshold values.

2C For the agreement between the Forristall distribution and the measurements, please see our reply to comment 2A. In l. 369 of the discussion, we will add the information that the "nearby stations" were buoy stations as well, to avoid confusion of buoy and radar measurement results from the previous study.
Regarding the second part of the comment, we acknowledge that there is a clear definition of what represents shallow/deep/intermediate water for a wave, while the terms were used here in a broader sense to distinguish sites. We will revise the manuscript to make this clearer and we will use a different terminology. We will also replace second-order theory by Forristall distribution.

2D The mentioned citation refers to non-Gaussianity in decreasing water depth, here in the context of wave run-up. The article is referred to by Sergeeva et al. (2011) to emphasise nonlinear behaviour of waves above a varying bathymetry. We do, however, agree with the reviewer that our formulation concerning the reference is misleading, as the referred article does not concern rogue waves. We will therefore re-formulate the sentence to read "... described e.g. by Huntley et al. (1977) *in the context of wave run-up. It has* gained increased attention in the context of rogue wave occurrence *(e.g. Sergeeva et al.* (2011)*)*".

3. Scientific Issues

3A We thank the reviewer for raising the issue of the shallow water definition. We will present the ranges of kh in our data, as well as the value of the slope, as suggested. The reviewer is right that our article does not solely concern waves in shallow water as defined by $h < L/20$. Since we investigate waves in the context of the KdV equation, we follow the definition of the applicability of the KdV equation as given by Osborne and Petti (1994), p. 1731, and Osborne (1995), p. 2629. We acknowledge that, therefore, our definitions of shallow water ($kh < 1.36$ or $h/L < 0.22$) and deep water ($kh > 1.36$ or $h/L > 0.22$) are different from the definitions of shallow ($h/L < 0.05$), intermediate ($0.05 < h/L < 0.5$) and deep water ($h/L > 0.5$) that are used in the engineering context, and that this may lead to confusion. We will therefore state this difference more clearly in the text, and we will clearly define the terminology for ‘shallow’ that is used in the paper. Furthermore, we will change the title to include “shallow depths” instead of “shallow water”, as suggested by the reviewer.

3B We apply the definition of the KdV equation as given in Osborne (2010) , p. 9, which defines the linear phase speed as $c_0 = \sqrt{gh}$. We agree that the term ‘shallow-water wave celerity’ is misleading when applying the KdV equation to relative depths larger than *h/L* = 0.05. Nevertheless, the linear phase speed is used in the KdV equation within the range of applicability. We will use the term ‘linear phase speed’ instead.

3C We thank the reviewer for pointing out that the category “normal” does not account for the fact that waves slightly below the threshold $H/H_s = 2.0$ are influenced by nonlinear processes and can become highly dangerous, similarly to waves with $H/H_s > 2.0$. We will change the category “normal” to “non-rogue”, to emphasise that these samples do not include waves according to the definitions $H/H_s \geq 2.0$ or $C/Hs \geq 1.25$.

3D We agree with the reviewer that the number of samples is a vague quantity for the reader and that the total number of waves should be more informative. We will include the precise number of measured waves in Table 1.

3E We thank the reviewer for the valuable suggestion that the results from the crucial papers Osborne et al. (1991) and Bruehl and Oumeraci (2016) should be explained to the reader more thoroughly. For explanation, we will additionally refer to earlier crucial references, the original numerical studies by Zabusky and Kruskal (1965) and Osborne & Bergamasco (1986). Osborne et al. (1991) applied the approach to ocean measurement data and Bruehl and Oumeraci (2016) performed an experimental study. Note that there is a technical difference between our approach and the approach in the three last mentioned works. We use the NLFT for vanishing boundary conditions, while Osborne & Bergamosco (1986), Osborne et al. (1991) and Bruehl and Oumeraci (2016) apply the NLFT for periodic boundary conditions. The relation between the two transforms is somehow similar to the relation between the linear Fourier transform and the linear Fourier series (which is what the FFT computes). They are related, and both have their advantages and disadvantages, but not all results can be directly compared for this reason. We will add paragraphs accordingly, additionally referring to earlier crucial references that explain the behaviour of solitons for the NLFT employed in our paper, like Hammack and Segur (1974) and Ablowitz & Kodama (1982). In our context, the sea surface elevation is described by a discrete spectrum indicating solitons and a continuous spectrum indicating a dispersive wave train. Of these two parts, we only discuss the soliton spectrum further in this article. However, it is known that for vanishing boundary conditions the soliton spectrum completely describes the behaviour of the wave train in the far field. After the complete dissipation of the dispersive waves, only the solitons are left in the far field. When the distance between these solitons is sufficiently large, no interactions occur between them and all solitons are clearly visible with their characteristic shapes. Assuming frictionless propagation, their amplitudes can already be read from the nonlinear spectrum of the initial time series. Therefore, we prefer to add the equation for the surface elevation in the far field, resulting from the solitons, given e.g. by Equation (4) in Prins & Wahls (2019), with reference to Schuur (1984), Eq. 17, Schuur (1986), p. 83, Eq. 33 and Ablowitz & Kodama (1982), Eq. 2.20a.

We further agree to add plots that explain the meaning of the soliton spectrum (Figure 1 in this response). In addition to an exemplary time series (blue line in the first plot, zoomed in to the rogue wave), we will add its linear FFT spectrum (second plot) and, in addition to the soliton spectrum (last plot), the nonlinear continuous spectrum (third plot, which will not be analysed further in this article). Each of the solitons in the soliton spectrum would be a physical soliton if the signal is propagated according to the KdV equation. After sufficiently long propagation, each of these solitons will appear isolated with its characteristic shape. Within the time series, the solitons are close together; they overtake and interact with each other. For visualisation of the role of the solitons in the time series, the first plot shows the soliton train (red line) that is obtained by nonlinear superposition of the solitons (considering their interactions) using the algorithm from Prins & Wahls (2021). Note that inverting large soliton spectra is numerically very difficult (Prins & Wahls, 2021). We therefore had to use a shortened time series for the figure.

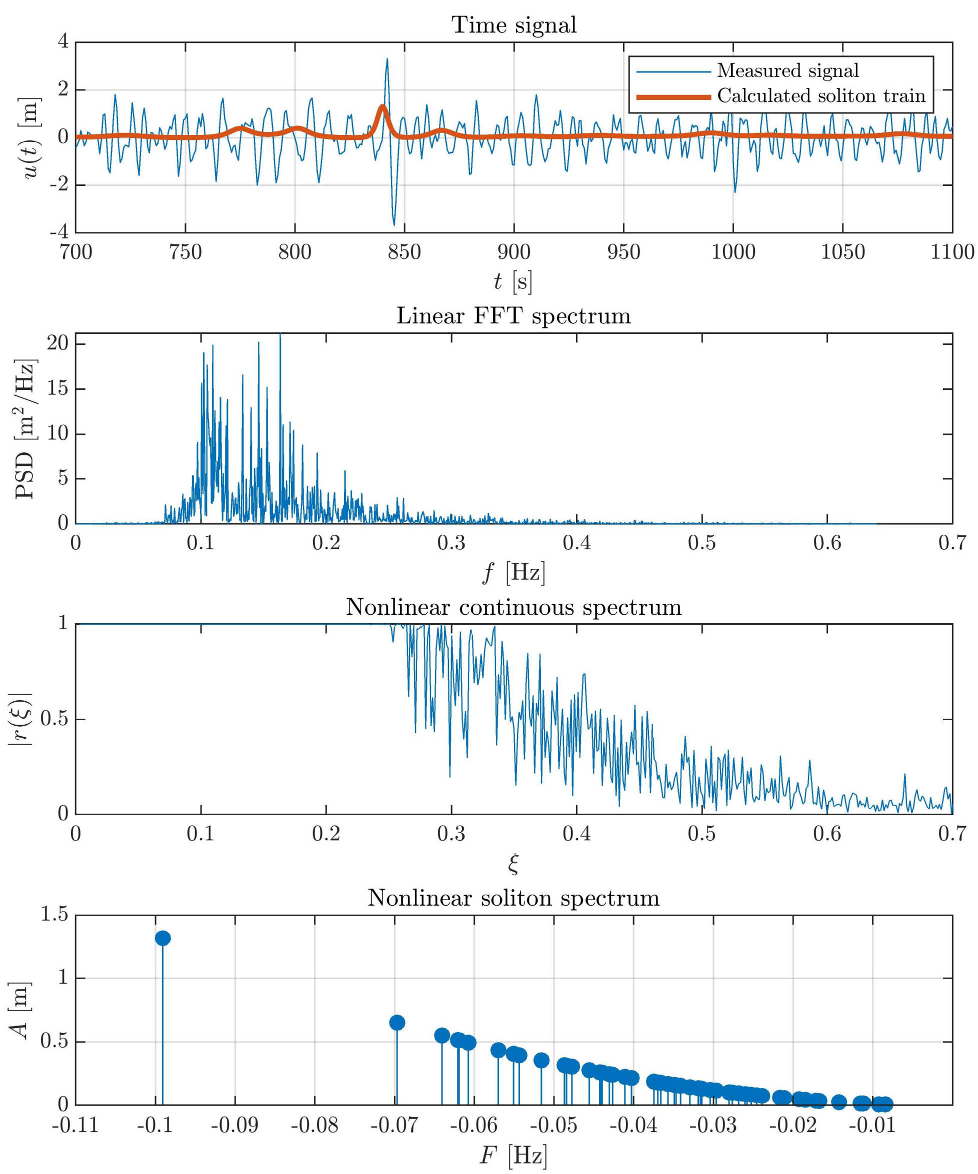

Figure 1: a) time series and soliton train, as calculated from the inverted soliton spectrum. b) linear Fourier spectrum. c) continuous spectrum. d) discrete soliton spectrum.

3F The reviewer is right in that the NLFT is currently employed as a trace method. The reason is that we do not know how the nonlinear spectrum changes during propagation around the buoy. If the KdV equation describes the propagation around the buoy reasonably well, then the nonlinear spectrum would be approximately constant during propagation, and the method could single out certain time series that lead (or led) to extreme rogue waves. This aspect however requires more research that is beyond the scope of the current paper. Therefore, we use the approach as a trace method, which has the additional potential to provide further information in future applications. Our work is nevertheless a necessary step in the direction of recognising potentially dangerous time series. If the method does not work for visible rogue waves, there is little hope for it to work for hidden rogue waves. We demonstrate for the first time that certain distinctive patterns in the nonlinear spectrum of real-world time series indicate extreme rogue waves (at a specific measurement site). Finally, we note that even if the KdV does not describe propagation well, the NLFT could still be a better transform to analyse data in this area than the linear Fourier transform (where the spectrum also only develops in a simple way if the propagation is linear and the depth is constant, but it is nevertheless applied in different contexts). We will discuss these points in the revised manuscript.

3G

- As suggested, we will support the explanation of the results by describing the insights from Bruehl and Oumeraci (2016) and the relationship between the soliton spectrum and the far field behaviour under KdV, in a meaningful connection with our reply to 3E.
- The use of the term “determined” was supposed to imply that the soliton was determined by the use of NLFT. For clarity, we will replace the terms “determined” and “specific” with “individual”.
- The reviewer criticises that scaling down a rogue wave to 80% is not linked to any physical explanation. We would like to reply that an established method for the treatment of NLFT spectra does not exist. In contrast to the linear case, where the impact of a window on the spectrum can be expressed analytically in a way that is easy to interpret, no such result is known for the nonlinear case. Windowing of the time series and calculating separate nonlinear spectra is common, but does not have any theoretical grounding. In contrast to windowing, which is a general purpose technique that impacts large parts of the time series, our method is local and aims specifically at rogue waves. By scaling only the rogue wave, the changes in the time series are as small as possible. The hope is thus that the danger of evoking additional, unrelated changes in the nonlinear spectrum is minimised by this approach.
  We intend to localise the influence of a change in rogue wave height in the soliton spectrum to establish a connection between a measured rogue wave and individual solitons. The underlying idea of the method is that if local changes of the rogue wave lead to local changes in the spectrum, the changing soliton components are associated with the rogue wave. Since a rogue wave is a particular wave event, it is reasonable to explore changes in the spectrum when only this wave is changed. Furthermore, also the (hidden) solitons in the data are localised components and changes to the particular rogue-wave event are expected to have effects to the soliton spectrum only when a soliton is located sufficiently close to the modified region. The changes in the soliton spectrum only affect a few solitons, whereas all other solitons remain constant. Since only a few solitons are modified, we can conclude that these solitons are located in the modified rogue-wave region within the time series. Regarding the request to remove the rogue wave from the time series, we prefer reducing its height as opposed to cutting it out, which would

introduce an artificial gap to the time series. The method shows that gradually reducing the height of the rogue waves leads to the gradual reduction of individual solitons. A change of the rogue wave will not have an impact on all soliton components in the spectrum. By this straightforward approach, solitons that are directly linked to the rogue wave are easily identified.

- Solitons linked to the rogue wave are not always the largest solitons in the spectrum. As also pointed out in the manuscript, the soliton alone is not sufficient to explain the rogue wave. Only by interaction with components from the continuous spectrum, the rogue wave is formed. In order for this to happen, the soliton and the other components must interact constructively. When the interactions are not constructive, it is very well possible that a larger soliton leads to a smaller hump in the time series. Hence, the dispersive waves and nonlinear interactions have a strong impact, and the largest soliton is not necessarily associated with the largest wave in the time series. For a visual illustration, we would like to refer to Figure 1 (a) in Osborne et al. (1991), in which the largest soliton is also not associated with the highest wave in the time series. In contrast, in our example given above, the largest soliton is located close to the position of the rogue wave.
- Against this background, changing Figure 8, in which rogue wave heights are compared with the associated solitons, would not make sense. Comparing with the highest soliton in the spectrum would include some solitons that are linked to wave groups without rogue waves. While we do not plan to change this in the paper, we nevertheless followed the suggestion of the reviewer and calculated A_{max} with respect to H_{max}. We present the updated Figure 8 in comparison with the original Figure 8 below (Figure 2-Figure 4). Here, grey dots show cases, in which the highest attributed soliton is identical with the maximum soliton in the discrete spectrum ($A_s^1 = A_{max}$). For rogue samples, this is true in most cases (extreme: 87%, double: 85%, crest 78%, height 71%). For non-rogue samples, this is true in 42% of the cases. The figures show that the results are in a comparable range when rogue wave (or maximum wave) heights are related to the maximum instead of the highest attributed soliton.

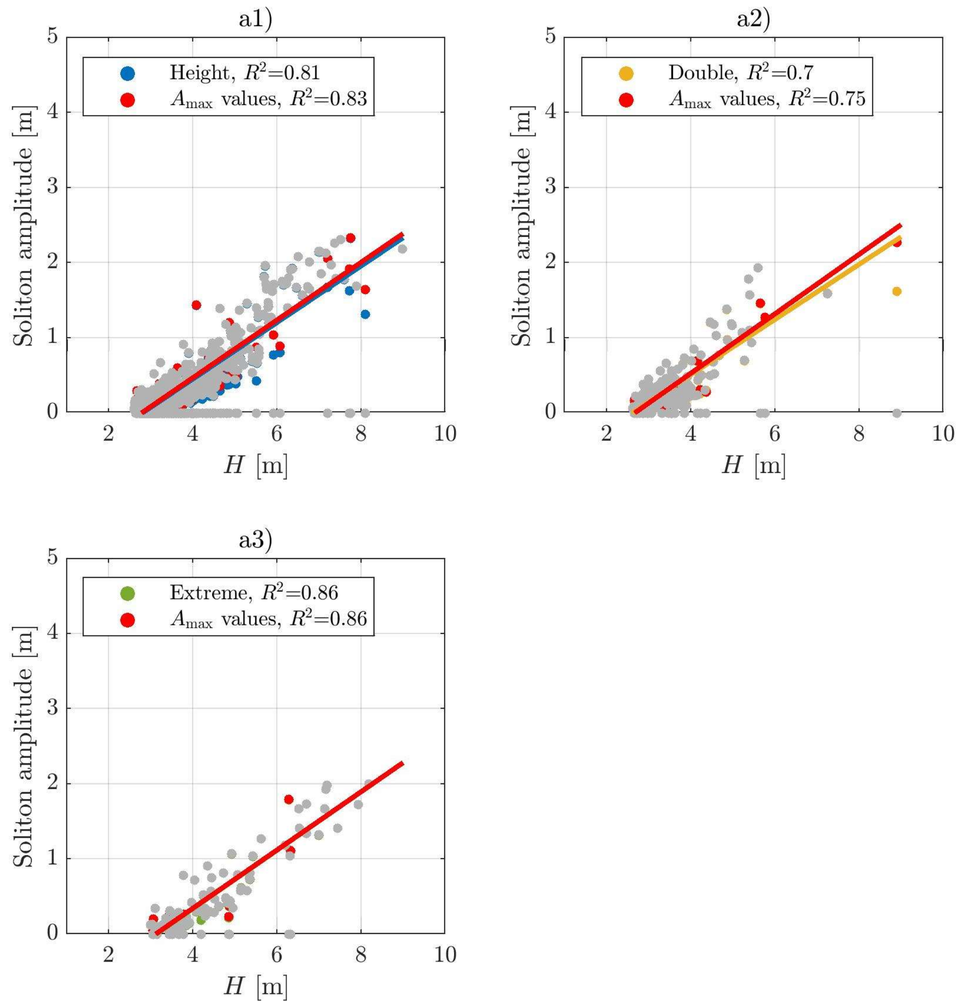

Figure 2: Update of Figure 8a in the preprint. Blue, yellow and green dots, in the legends referred to as "Height", "Double" and "Extreme", represent values from the original figures. When using maximum soliton amplitudes instead of the amplitudes of the attributed solitons, these values change to values represented by red markers. Grey markers show values of attributed soliton amplitude that are identical with the maximum soliton amplitude.

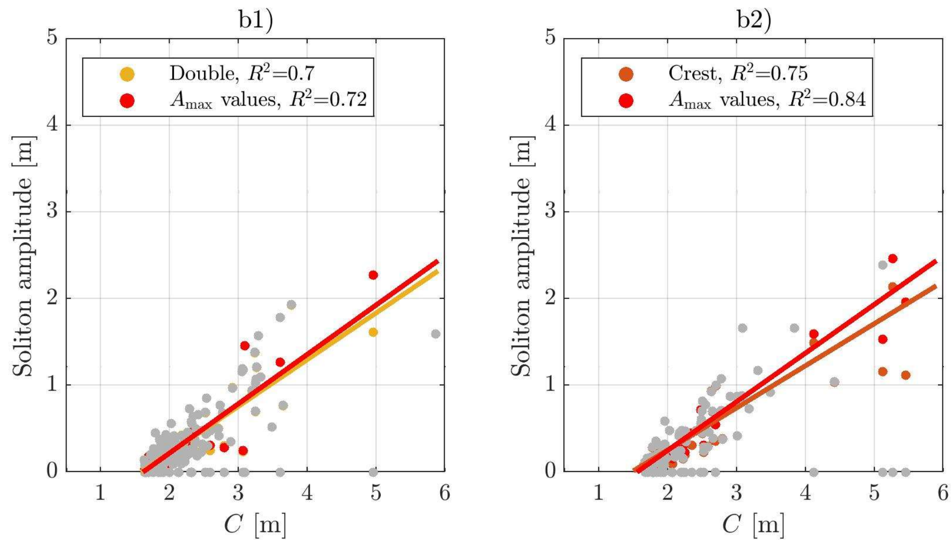

Figure 3: Update of Figure 8b in the preprint. Yellow and orange dots, in the legends referred to as "Double" and "Crest", represent values from the original figures. When using maximum soliton amplitudes instead of the amplitudes of the attributed solitons, these values change to values represented by red markers. Grey markers show values of attributed soliton amplitude that are identical with the maximum soliton amplitude.

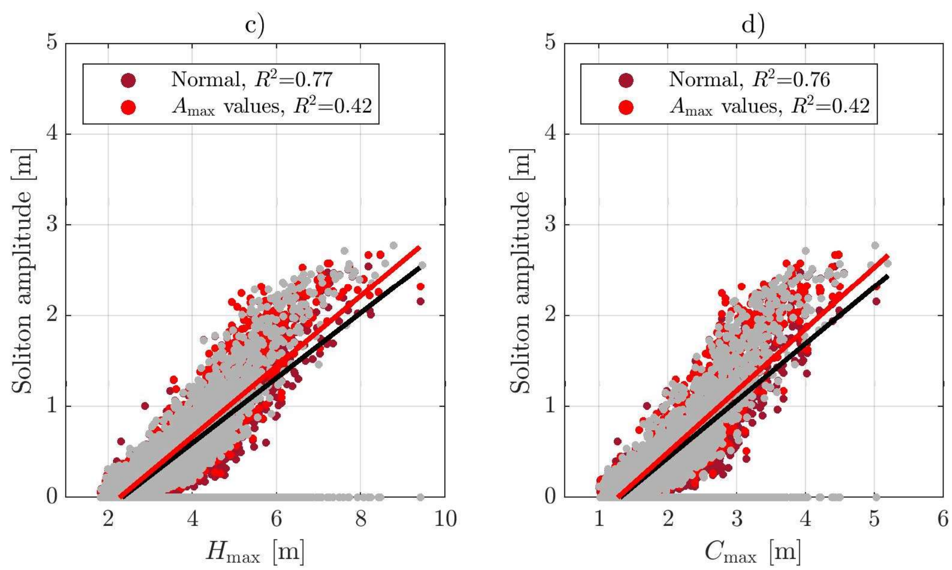

Figure 4: Update of Figures 8c and 8d in the preprint. Dark red dots, in the legends referred to as "Normal", represent values from the original figures. When using maximum soliton amplitudes instead of the amplitudes of the attributed solitons, these values change to values represented by red markers. Grey markers show values of attributed soliton amplitude that are identical with the maximum soliton amplitude.

- In our study, soliton amplitudes were always smaller than rogue wave crests, which is in agreement with Figure 1a in Osborne et al. (1991).
- We understand that the formulation "To remove the influence of the underlying sea state" in line 261 is misleading. We agree with the reviewer that the influence of the sea state is not only characterised by the significant wave height. Our intention with normalising by H_s is to create dimensionless values, to be able to compare different samples. Since rogue waves are defined on the basis of the significant wave height, we find this parameter suitable for the normalization. The influence of the sea state in terms of the parameters suggested by the reviewer (steepness, Ursell number, *kh*, bandwidth) affect the wave components in the continuous part of the nonlinear spectrum, which we do not discuss further in this article. The soliton spectrum is not affected. Through the continuous spectrum, the sea state parameters possibly influence the wave distribution. We have tested the method that was suggested by the reviewer. As an example, Figure 5 shows the exceedance probability of H/H_s in all samples of a defined Ursell number range. It is seen that the distribution behaves differently in the different ranges. However, this cannot be stated for certain, as the results rely on few data, due to the binning into ranges. The few rogue waves in the samples are distributed randomly, which leads to uncertain results. Together with the consideration that the sea state may indeed affect the continuous part of the spectrum, and thus the probability distribution may change with the sea state parameters, we have come to the conclusion that it is not possible to deduce the influence of solitons from these exceedance probability plots. Therefore, we have decided not to include the plots in the revised article.

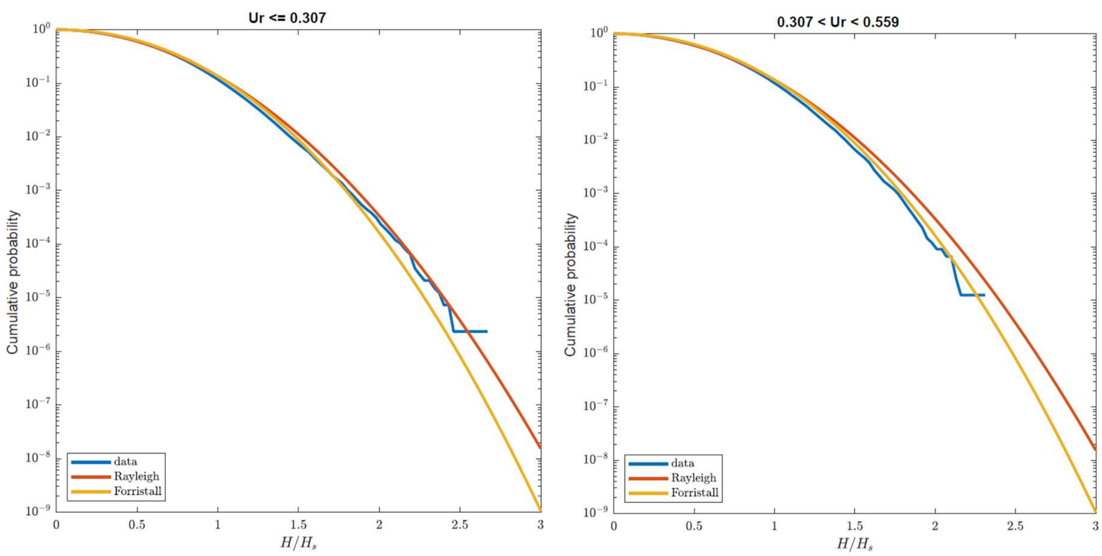

Figure 5: Cumulative exceedance probability of H/H_s in all samples that belong to the category of $0.2 < A_1^S/H_s \leq 0.3$, and two different Ursell number ranges.

3H We would like to point out that we do not intend to explain a rogue wave by one soliton alone (as can be seen by comparison of the free-surface elevation ad and the soliton train in Figure 1 in this document). Our hypothesis is that solitons *contribute* to the formation of rogue waves, and we have shown that there is/are always one or several solitons involved when a rogue wave is present in the sample. However, the surface elevation is described not only by solitons, but also by dispersive waves and by the interaction of wave components. This statement is supported by Figure 10 in the preprint, which we therefore would like to keep in the article. Bruehl et al. (2016) have shown that soliton-like waves can form waves that seem to have linear shapes.

3I The soliton spectrum alone cannot reveal the formation of rogue waves in general, but solitons may be directly attributed to rogue waves and as such are involved in their presence. We would like to emphasise that it is actually the first time that this has been verified by the nonlinear Fourier analysis of real-world data. Furthermore, as discussed later in the manuscript in relation to Figure 15, certain configurations of the soliton spectrum (A_2/A_1) actually do indicate the presence of rogue waves with high probability. The attribution is shown by the method discussed in comment 3G, which we would like to retain in the paper. Furthermore, the size of a soliton is not sufficient to explain the height and the shape of a rogue wave all by itself (see for example Figure 1). As suggested by the reviewer, we will transfer these issues to the discussion of the article to present the line of argument in a straight order.

3J The soliton gap was chosen after studying the soliton spectra of many rogue samples. It is not arbitrary, and as far as we know there are no existing alternatives from the literature that would have fit our context. Please also note that many existing signal processing tools heavily rely on the linearity of the transform, and applying them with a nonlinear transform in general is meaningful only in the quasi-linear regime. We do not expect the variance of the soliton spectrum to be a better tool because it would involve solitons that are not associated with the rogue wave. We agree that the conclusion should be that outstanding solitons are not good indicators of rogue waves or large waves near the rogue wave threshold of H/H_s = 2.0, but only for extreme rogue waves. We will avoid the term “predictor” in the text, as the soliton spectrum becomes available only after the recording of the time series and the occurrence of the rogue wave (see also reply to comment 3F).

3K The remark is correct, we will adjust the text accordingly, referring to the Forristall distribution (see replies on comments 1B and 2C).

3L We confirm that we have not investigated the interaction with oscillatory waves in the context of rogue wave formation and that we are therefore not entitled to make a statement on the exact nature of the interactions. We agree that we should mention this in the conclusions, so as not to raise wrong expectations with the reader. From Figure 1, it is seen that the soliton train alone does not account for the full height of the rogue wave. This only leaves the continuous spectrum for the explanation of the missing height.

References

Ablowitz, M. J. & Kodama, Y., 1982. Note on Asymptotic Solutions of the Korteweg-de Vries Equation with Solitons. *Studies in Applied Mathematics,* 66(2), pp. 159-170.

Brühl, M. & Oumeraci, H., 2016. Analysis of long-period cosine-wave dispersion in very shallow water using nonlinear Fourier transform based on KdV equation. *Applied Ocean Research,* Volume 61, pp. 81-91.

Hammack, J. L. & Segur, H., 1974. The Korteweg-de Vries equation and water waves. Part 2. Comparison with experiments. *Journal of Fluid Mechanics,* 65(2), pp. 289-314.

Huntley, D. A., Guza, R. T. & Bowen, A. J., 1977. A universal form for shoreline run-up spectra?. *Journal of Geophysical Research,* Volume 82, pp. 2577-2581.

Osborne, A. R., 1995. The inverse scattering transform: Tools for the nonlinear Fourier analysis and filtering of ocean surface waves. *Chaos Solitons Fractals,* 5(12), pp. 2623-2637.

Osborne, A. R., 2010. *Nonlinear ocean waves and the inverse scattering transform.* Amsterdam: Elsevier.

Osborne, A. R. & Bergamasco, L., 1986. The solitons of Zabusky and Kruskal revisited: Perspective in terms of the periodic spectral transform. *Physica D: Nonlinear Phenomena,* 18(1-3), pp. 26-46.

Osborne, A. R. & Petti, M., 1994. Laboratory-generated, shallow-water surface waves: Analysis using the periodic, inverse scattering transform. *Phys. Fluids,* 6(5), pp. 1727-1744.

Osborne, A. R., Segre, E., Boffetta, G. & Cavaleri, L., 1991. Soliton basis states in shallow-water ocean surface waves. *Physical Review Letters,* Volume 67, pp. 592-595.

Prins, P. J. & Wahls, S., 2019. Soliton Phase Shift Calculation for the Korteweg–De Vries Equation. *IEEE Access,* Volume 7, pp. 122914-122930.

Prins, P. J. & Wahls, S., 2021. An accurate O(N2) floating point algorithm for the Crum transform of the KdV equation. *Communications in Nonlinear Science and Numerical Simulation,* Volume 102, p. 105782.

Schuur, P., 1984. Multisoliton phase shifts in the case of a nonzero reflection coefficient. *Phys. Lett. A,* 102(9), pp. 387-392.

Schuur, P. C., 1986. *Asymptotic Analysis of Soliton Problems: An Inverse Scattering Approach.* 1232 ed. (Lecture Notes in Mathematics): A. Dold and B. Eckmann, Eds. New York, NY, USA: Springer-Verlag.

Sergeeva, A., Pelinovsky, E. & Talipova, T., 2011. Nonlinear random wave field in shallow water: variable Korteweg–de Vries framework. *Natural Hazards and Earth System Sciences,* 11(2), p. 323–330.

Teutsch, I., Weisse, R., Moeller, J. & Krueger, O., 2020. A statistical analysis of rogue waves in the southern North Sea.. *Natural Hazards and Earth System Sciences,* 20(10), p. 2665–2680.

Zabusky, N. J. & Kruskal, M. D., 1965. Interaction of "Solitons" in a Collisionless Plasma and the Recurrence of Initial States. *Physical Review Letters,* Volume 15, pp. 240-243.