JOURNAL OF COASTAL AND HYDRAULIC STRUCTURES

Review and rebuttal of the paper

Assessment of maintenance efforts and probabilities of failure at German inland waterways to advance the design of bank revetments

J. Sorgatz and J. Kayser

Editor handling the paper: Anton Schleiss

The reviewers remain anonymous.

Review Round 1:

Replies to the reviewers' comments on the manuscript "Assessment of maintenance efforts and probabilities of failure at German inland waterways to advance the design of bank revetments"

First of all, we would like to thank the reviewers for commenting on our manuscript, which greatly assisted in improving the manuscript. We believe that we could clarify and include all issues raised by the reviewers adequately.

In this letter, we reply to the questions which were raised during the review, and explain the changes made to the manuscript. Reviewer comments are in Times New Roman italic, our replies in Times New Roman (black color). Quotation marks indicate excerpts from the manuscript. Changes to the manuscript that are related to comments of the reviewers are given in green color.

In addition, all changes made to the text are highlighted in the manuscript as well. In few cases, we made minor adjustments to the wording in order to make a sentence more concise or easier to follow, e g., "For the presented investigations," (formerly L282) was replaced by "For the stability assessment," (L389). These minor changes are not outlined in this reply to the reviewers' comments, but highlighted in the text.

Reviewer A:

My recommendation is a revision mainly due to a need to clarify and structure the method, some additions for discussion and minor comments for Tables and Figures. However, I must admit that the topic of the paper is slightly out of my field of expertise. I suggest in case of a second round of review to find a more qualified reviewer for the sake of good quality of the review process.

Research gap and objective
The article aims at an objective maintenance strategy specifically for riprap bank revetments of waterways based on a structural reliability analyses which - to my knowledge - has not been investigated with this kind of supporting data yet. However, this assessment is tailored to German inland waterways and associated design practices. Hence, the methods are applicable only to structures with the same boundary conditions.

Methods
The study demonstrates a probabilistic approach supplemented by a deterministic one in order to link maintenance measures to revetment stability. Maintenance effort is represented by the amount of armor stone displacement. The stability criterion, the limit state function, is based on a semi-empirical hydraulic design criteria: the required mean armor stone diameter. The hydraulic load is taken into account by the stern wave height. The measured maintenance effort (placed armor stones) is compared to the probability of failure (probability of armor stone displacement).

Comment 1 (formerly L290ff): While the armor stone placements are measured over several years, the stern wave height is only measured for short time periods of 7 up to 14 days. The wave heights are taken as the representation of the load situation for a larger canal section. Hence, they are extrapolated temporally and spatially. Ship-induced currents were not measured. However, the authors rely on the available statement according to Sorgatz (2021) that a minimum number of 250 wave events are needed for the probabilistic analyses. It is mentioned in the discussion that 3 out of 12 cases do not meet this requirement. At this point, I am missing a discussion if these 3 cases should *still be considered for the conclusion. Moreover, it could be pointed out which cases they are or at least refer to Table 2 where the count can be found. For DEK-n the requirement is just met with nr=253. That is the one where 5 campaigns (Altenrheine, Gleesen, Hesselte, Rodde, Venhaus) are originated in. This could be pointed out together with the conclusion of a need for site-specific data. An even higher number of wave events is recommended for the deterministic analyses.*

The minimum number of observations (Sorgatz, 2021) is a robust estimate which considers a larger variability of the sample. For lower variabilities as considered in the present investigations (cov < 0.5), it is acceptable to work with a small number of samples. With the now included modifications in the text we tried to relate this information to the reader. Moreover, we tried to differentiate more clearly between the prerequisites for the deterministic and the probabilistic calculations and highlighted the need for site-specific data as suggested by Reviewer A.

L397ff:

> "When considering annual probabilities of failure, representative load assumptions are required for a period of one year. In the presented investigations, the annual probability of armor stone displacement is extrapolated spatially and temporarily. This implies that the site-specific probability density functions of the wave heights represent the annual traffic over an entire section. […]
>
> Furthermore, based on probabilistic calculations by Sorgatz (2021), at least 250 measured wave events are required for a reliability-based hydraulic revetment design. In the case of the deterministic analyses, the number of required measurements is higher (Sorgatz, 2021) which is supported by the fact that the deterministic analysis yields $p_{f,d} = 0$ for two case studies with less than 250 observations (DEK(S)-1 and MDK-2). For 9 out of 12 case studies the requirement of more than 250 measurements is fulfilled (see Table 2). For DEK(S)-1, MDK-1 and MDK-2 only 200 observations are available. Despite this fact, we decided to use all data for the analyses since the required number of measurements also depends on the variability of the sample. The results of Sorgatz (2021) are a robust estimate for cov = 3.0. For lower variabilities as considered in the present investigations (cov < 0.5), it seems acceptable to work with a smaller number of samples."

Comment 2-1 (L82): n refers to north while it is a blockage ratio in L63. Suggestion: use N for north.

The notation has been changed to DEK(N) and DEK(S) throughout the entire paper.

Comment 2-2 (formerly Table 3 and L112, now Table 3 and L116ff): Together with the survey of maintenance measures, the yearly costs are displayed. The costs are not addressed again in any other section besides the aim of an economic strategy referring to optimal lifetime costs. In my opinion, the maintenance costs do not contribute significantly to the thesis of this paper and I would recommend to reduce them in Table 3 to the sum or to the scaled sum them with the replaced amount of armor stones or even dismiss them totally if you cannot demonstrate a needed way to incorporate them.

We agree with Reviewer A that the provided data may confuse the reader. We have concluded that the costs are not essential for the reader to understand this article. Nevertheless, in order to make the data available to other researchers for further analyses, the table section with the costs has been moved from the main body to the appendix. The following explanations have been added to the text and moved to the passage that explains the elicitation:

> "The following investigations focus on the maintenance efforts in terms of the annual mass of placed armor stones rather than on the maintenance costs, since the costs may be affected by factors such as the availability of personnel and machinery as well as by locally and over time differing costs for armor stones. However, for the

purpose of completeness and in order to make the data available to other researchers, the elicited maintenance costs are presented in the Appendix."

Comment 3 (formerly L130, now L155f): Please clarify the definition of beta and its link to the probability of failure.

Indeed, there are several definitions of the reliability index β and, thus, its link to the probability of failure p_f. In the case of the presented study, the following functional relationship is used:

$$p_f = \Phi(-\beta)$$

where Φ is the standard Gaussian distribution. To make this relation clear to the reader, we added the following lines to the text:

> "They can be expressed via the probability of failure p_f or the reliability index β. A functional relationship between p_f and β is given via the standard Gaussian distribution Φ as:
>
> $$p_f = \Phi(-\beta) \quad (1)$$
>
> Commonly a similarity between coastal structures and revetments is assumed which may allow to transfer target reliabilities proposed for coastal structures to the assessment and design of revetments in inland areas."

Comment 4 (Table 5): Recommendation of a separating vertical line in between ULS and SLS

To increase the readability, we followed the Reviewer A's recommendation and added a separating vertical line in Table 5.

Comment 5 (formerly L164, now L187ff): Here, pf is defined for the probabilistic analysis but you use it in the deterministic case in table 9. Plus, it is redefined in L222 (results!). Please clarify both cases in the methods. Generally, try to clarify the method applied for the deterministic and probabilistic analyses.

We agree with the reviewer's concern that the former Sec 2.2.2 was difficult to follow. We therefore decided to completely revise it. For the purpose of clarification, we introduced a new section which is called "2.2.3 Deterministic and probabilistic stability assessment". This section covers both, the equations for the deterministic and probabilistic calculation method. In addition, it explains how traffic related probabilities of failure are derived. We thus moved the parts which cover the probabilistic approach from section 2.2.2 and the former lines L208 – 213 to this section (see also *Comment 7*), included an explanation regarding the denotation of the two calculation approaches at the start of Section 2.2.3 and introduced sub-headings for the different calculation approaches:

> "In this section, two calculation approaches for the stability assessment of revetments are presented. In the first approach, the hydraulic loads are described via their measured values. Characteristic values of slope geometry and material properties are used. In the second approach, the input variables are described via their probability density functions. In both cases, a "probability of failure" can be determined. For simplicity, we refer to calculation method and probability of failure of the former as "deterministic" and the latter as "probabilistic"."

To clarify the procedure of the deterministic approach, we also added additional explanatory lines at the end of section 2.2.3 (see text below). Furthermore, deterministic and probabilistic probabilities of failure are now differentiated with the subscript "p" for probabilistic and "d" for deterministic.

> "In the case of the deterministic calculations, firstly, a utilization rate η is defined; $\eta > 1$ indicates that the required armor stone diameter exceeds the armor stone diameter in-situ which will result in armor stone displacements. Equation (5) features a set of deterministic variables $\mathbf{X}$ and is defined as the ratio of the maximum of the required armor stone diameter $D_{50,\mathrm{req}}$ and the armor stone size in-situ $D_{50,\mathrm{site}}$.
>
> $$\eta(\mathbf{X}) = \frac{\max D_{50,req}(\mathbf{X})}{D_{50,site}} \quad (5)$$
>
> For the purpose of stability assessment, the maximum of the observed wave heights and the mean values of the armor stone characteristics (see Table 6 and Table 7) are used. A conservative (damage and maintenance are not permitted) and a less strict (moderate damage and maintenance are allowed) design are considered by means of the B_B'* values. Failure is defined as follows:
>
> $$I_d(\mathbf{X}) = \begin{cases} \text{stable } (0), & \text{if } \eta(\mathbf{X}) \leq 1 \\ \text{unstable } (1), & \text{if } \eta(\mathbf{X}) > 1 \end{cases} \quad (6)$$
>
> Only if $\eta > 1$ is observed, the maintenance over one year can be extrapolated from the number of daily exceedances of η. For this purpose, the required armor stone diameter is computed for each of the observed wave events and, subsequently, compared to the armor stone diameter in-situ. The ratio of number of failures and total number of observations n_{total} results in a "deterministic" probability of failure $p_{f,d}$, see eq. (7).
>
> $$p_{f,d} = \frac{1}{n_{total}} \sum_{i=1}^{n_{total}} \mathbf{1}_{\{\eta(\mathbf{X}) > 1\}} \quad (7)$$"

Comment 6 (formerly Figure 3, now Figure 5): the number displayed as the mean does not match with the cumulative density function's mean which should be around 0.25 m. Maybe it is scaled? Please check.

This is indeed a mistake. We modified the illustrations to correct this mistake.

Comment 7 (formerly L208-213, now 192ff) This is part of the method, not the results.

We tried to distinguish more clearly between method and results. Thus, these lines are now included in the new section "Deterministic and probabilistic stability assessment" (see also *Comment 5*).

Comment 8 (formerly L223-224, now L283ff): The presented pf values considered as implausible: The deterministic assessment reflects the observed frequency. Hence, the values are plausible as the observed frequency in this short time period measured, but it is unlikely to represent the actual probability of failure.

We agree with this observation and adapted the paragraph accordingly as follows:

> "Only if $\eta > 1$ is observed, the maintenance over one year can be extrapolated from the number of daily exceedances of η. Thus, only the conservative design is used in the following. For the results of these calculations ($p_{f,d}$, $p_{v,d}$, $p_{a,d}$ in Table 9) it must be noted that $p_{f,d} = 0.00$ for DEK(S)-1 and MDK-2 are unlikely to represent the actual probability of failure. The deterministic analyses rely on the observed frequency of limit state exceedances, which, in turn, is affected by the total number of available observations. Thus, the $p_{f,d} = 0.00$ values are most likely a result of the short observation period."

Comment 9 (formerly Figure 4 and 5, now Figure 6): I'd recommend that the name of the axis correspond with the description

The graphs were completely revised to accommodate for the comments of Reviewer B. Additionally, to increase the comprehensibility of the paper, we adapted the recommendation of Reviewer A.

Comment 10 (formerly L240-245): This is part of the method, not the results.

To distinguish more clearly between methods and results we moved the first part to the newly introduced section "Deterministic and probabilistic stability assessment" (see also *Comment 5* and *Comment 7*).

L206ff:

> "The probabilistic calculations are performed with the Python package OpenTURNS (Baudin et al., 2015). Due to a highly non-linear limit state function Monte-Carlo simulations were employed to determine the probability of failure."

The second part was deleted as it is already explained in in the newly introduced Section 2.2.3

Comment 11 (formerly L241): I recommend the convergence criteria and strategy to be specified, to have the possibility to evaluate the reliability of the estimated probabilities (especially confidence or accuracy on the probability estimated). A discussion about the convergence and the estimates can be taken under Results or Discussion.

We agree that the convergence criteria and the accuracy of the probability estimate are important to evaluate the precision of the results. To make this information available to the reader, we included the following lines:

L206ff:

> "The probabilistic calculations are performed with the Python package OpenTURNS (Baudin et al., 2015). […] The ratio between the asymptotic standard deviation of the probability estimate and its mean value is used as convergence criteria. It is a relative measure of dispersion and must be lower than 0.05 in order to end a simulation."

Moreover, we included a new column in Table 10 with the standard deviation of probability of failure determined with Monte-Carlo simulations and a brief discussion regarding the uncertainty of the probability estimates in the section "Results":

L292ff:

> "Table 10 presents the results of the probabilistic analyses. Since the probability estimate $p_{f,p}$ was determined by means Monte-Carlo simulations, the standard deviation σ of the probability estimate is given to evaluate the accuracy of the results. Considering the propagation of σ through eq. (12) and eq. (13), $p_{a,p}$ deviates by a maximum of +/-5 % as a consequence of the variability of the initial probability estimate $p_{f,p}$. To put this into numbers using the example of DEK(N)-1, if $p_{f,a}$ = 3.44E-02 with σ = 1.66E-03, then $p_{a,p}$ deviates by +/- 1.80E-02."

A discussion regarding the interpretation of the results under the constraint of the above-mentioned uncertainty is included in the section "Discussion".

L397ff:

> "When considering annual probabilities of failure, representative load assumptions are required for a period of one year. In the presented investigations, the annual probability of armor stone displacement is extrapolated spatially and temporarily. This implies that the site-specific probability density functions of the wave heights represent the annual traffic over an entire section. However, the distributions are statistical approximations of the wave events observed in real life. The Monte Carlo simulations yield an error of +/- 5 % in relation to the determined probabilities of failure. Since for revetments these uncertainties do not translate into failure, but into a change in maintenance, for now, these uncertainties are acceptable. The accuracy of the predictions can be increased by supplementing the database.
>
> Furthermore, based on probabilistic calculations by Sorgatz (2021), at least 250 measured wave events are required for a reliability-based hydraulic revetment design."

Comment 12 (formerly L241, now L206ff): Is there a specific reason for using this algorithm against another (possibly also available on OpenTURNS)? However, I have no experience with OpenTURNS and therefore do not feel confident enough of judging the sampling algorithm further.

Initially, we used the Adaptive Directional Sampling algorithm which is an importance sampling method. It is particularly suitable for the evaluation of a low probability of failure in combination with a small number of stochastic variables. Yet, since the results do not indicate low probabilities of failure we decided to make the calculations more accessible to the reader by using straight-forward Monte-Carlo simulations. Thus, the analyses were re-calculated. The explanation of the employed algorithm in the "Methods" part and Table 10 in the "Results" part were revised. Additionally, to account for this change in the calculation method, the failure criterion was slightly reformulated:

> "The probabilistic calculations are performed with the Python package OpenTURNS (Baudin et al., 2015). Due to a highly non-linear limit state function Monte-Carlo simulations were employed to determine the probability of failure. […] The "probabilistic" probability of failure $p_{f,p}$ is then defined as follows:
>
> $$p_{f,p} = \frac{1}{n_{total}} \sum_{i=1}^{n_{total}} \mathbf{1}_{\{g(\mathbf{X})>1\}} \quad (9)$$
>
> To account for the correlation of the random variables, a joined probability density function is required. It is found via a Gaussian copula C(X) for modeling multivariate dependence during the reliability analyses:"

Comment 13 (formerly Table 11, now Table 12): Separating the parameters, the results from the equations and the resulting differences with a vertical line would make the table clearer. Moreover, "deterministic equation" and "probabilistic equation" are unclear. The displayed equations are linear equations with parameters based on the deterministic / probabilistic approach. Please clarify in the table by adjusting the headings. Especially because those are not directly stated in the text as equations before (only shown in the Figures).

First of all, we must point out that Table 12 was revised as a result of Reviewers' comments. Despite this fact, we tried to follow the above recommendations of Reviewer A and adjusted the annotation and the formatting of Table 12.

"Table 12| Results of parameter study for a comparison of the linear and exponential equations found with the deterministically and the probabilistically determined failure probabilities. Regression equations that were determined with the "deterministic" failure probabilities are denoted by "-DET" and $p_{a,d}$, whereas "-PROB" and the $p_{a,p}$ in the equation indicates that the equation is derived with probabilities of failure obtained from Monte-Carlo simulations."

Case studies			Linear regression			Exponential regression		
Vessel passages per year	p_v	p_a	LIN-DET $A = 7.89\ p_{a,d}$	LIN-PROB $A = 9.31\ p_{a,p}$	Diff	EXP-DET $A = e^{2.62 p_{a,d}}$	EXP-PROB $A = e^{3.15 p_{a,p}}$	Diff
	--	--	kg/m²	kg/m²	%	kg/m²	kg/m²	%

Comment 14 (formerly L297ff and L322ff):

In the end the target reliability is to be compared with the actual structural reliability. To predict maintenance measures, the link between maintenance and revetment stability was investigated. For both methods, a linear relationship has been found, while the probabilistic method should be favored as expected.

Obvious limitations of the analyses have been discussed: Maintenance measures result from subjective assessments before. Available data. Ship-induced flows are not taken into account for the load.

However, the objective was to find a methodology to incorporate into a sustainable maintenance strategy for waterways. The linear relationship predicts the required maintenance as a function of the annual probability of failure. But as it is pointed out, based on the assumptions, is enough only for "first investigations" of the correlation. Hence, the recommended target reliability seems like a rough estimate still and it does not seem obvious to choose ß=1.3 or make a decision with confidence. I suggest to describe the decision process and the consequences. The target reliabilities ranges proposed by JCCS (2001) could be included in Figure 5 and consequences could be discussed.

We agree that the decision process should be made more transparent. Firstly, we believe we made the choice of the regression equations including the uncertainties inherent to the analyses more transparent by reorganizing the section "Results". In the subsection "Regression analyses with deterministic and probabilistic analyses" we now present the results of the regression analyses of both, the deterministic and the probabilistic approach in more detail. For instance, the p-values and R^2 values are now included for different trend functions and parameter combinations. The target reliabilities proposed by JCSS (2001) are added to Figure 6 (formerly Figure 4 and Figure 5). For the choice of a first estimate of the target reliability, we included the following explanations:

L413ff:

"Revetments are simple structures. However, since they cover large areas at canals or rivers, the relative cost of safety measures is comparatively high. If larger armor stones have to be installed, the costs rise quickly as a result of the corresponding increase in required armor layer thickness. The risk associated with damage, on the other hand, is low under the constraint that inspections are conducted on a regular basis. If one applies the risk classification according to JCSS (2001), see Table 7, the structures should at least satisfy a reliability index $\beta >$ 1.3 ($p_a < 0.1$).

The annual probabilities of armor stone displacements obtained with the deterministic approach (see Table 9) exceed this target reliability. On the one hand, this can be a result of generally low reliability standards at inland waterways which, in (conservative) deterministic analyses, are included by the B_B'* values and partial factors equal 1. On the other hand, the results emphasize that the deterministic approach may not cope well with the limited number of observations.

In contrast, the probabilistic approach seems to be characterized by a greater robustness regarding the underestimation of actions. The probability functions account for the possibility that values occur, which exceed in nature observed loads. In the case that only maintenance efforts are of relevance for design considerations, an annual target reliability of $p_a = 0.1$ ($\beta = 1.3$) may be derived from the presented probabilistic analyses. Figure 6 shows low maintenance measures for $p_a < 0.1$. Only for $p_a > 0.5$ maintenance measures are observed on a larger scale. It is emphasized that the proposed target reliability is a first estimate which is affected by the uncertainties discussed in the previous paragraphs. Certainly, further data and analyses are required to specify a generally valid target reliability. This is particularly important as the analyses of DEK (North) show that maintenance costs can be quite high in the case of heavy traffic and/or large hydraulic loads."

In the section "Conclusions" we tried to emphasize that the proposed target reliability is a first estimate which still requires further research.

L445ff:

"Especially when the number of observations is small, the probabilistic approach should be favored over the deterministic approach. […] If only maintenance efforts are of relevance for design considerations, the results indicate that $\beta = 1.3$ ($p_f \approx 10^{-1}$) may be a suitable first estimate for a target reliability resulting in moderate maintenance. However, once more it is emphasized that further data and analyses are required to specify a generally valid target reliability."

Thus, we also re-phrased the findings in the Abstract slightly in order to clarify what results can be derived from our analyses:

"It was found that a positive correlation between revetment stability and maintenance must be assumed. A comparison between the deterministic and the probabilistic stability assessment shows that particularly for small sample sizes and low failure probabilities, the probabilistic approach should be favored over the deterministic approach. In the case that only maintenance is of relevance for design considerations, the results of the probabilistic analyses indicate that $\beta = 1.3$ ($p_f \approx 10^{-1}$) may be a suitable first estimate for an annual target reliability."

Comment 15: The main issue is the clarification of the methods. Also, since, as mentioned above, parts of the method are described in the results instead.

As outlined in the previous comments in more detail, we addressed this comment as follows:

- Re-organization of section "Methods" and section "Results"
    - Introduction of new subsection "2.2. Deterministic and probabilistic stability assessment" in the section "Methods"
    - Introduction of subsections "Probabilities of failure from deterministic and probabilistic analyses" and "Regression analyses with deterministic and probabilistic analyses" in the section "Results"
- Additional information on applied algorithm and convergence criteria
- New subscripts and additional explanations to distinguish between deterministic and probabilistic calculations and results more clearly

Reviewer B:

Comment 1: If possible, the authors could add a map of Germany where the reader could see the location of the canals that were studied. Better, if the 12 studied locations are included, that would be even better for the reader to understand the work.

We agree with Reviewer B that a map of the observation location is beneficial to understanding the paper. Thus, a new image (Figure 1) was added.

Comment 2 (formerly L67, now L65ff): Please mention the variation of the canals discharge, water level and blockage ratio throughout the year. Are they constant or suffer seasonal fluctuations? Please discuss that the damage on the revetment will be different if there aren't significant variations in the water level (so the ship waves will always impact at the same location and cause a concentrated damage that will lead to a sooner damage to the filter layer), compared to the case where a water level variation will cause the wave to impact at different heights and spread the damage over the slope height, making it less likely for the damage to reach the filter layer. This is related to the sentence in Line 86 where you indeed mention that damage occurs at the water level.

We agree that the above aspect is important for the interpretation of the results. German inland canals are commonly regulated by weirs and locks. Thus, the water level is close to constant throughout the year. Moreover, there is hardly any discharge. Damage only occurs in a limited area of the slope. To convey this information to the reader, we added the following lines:

> "In fairways confined laterally and in depth, the hydraulic loads are significantly governed by the vessels' velocity in relation to the blockage ratio *n* that describes the ratio between the cross-sectional area of the waterway A_W and the cross-sectional area of the submerged part of a vessel A_S ($n = A_W / A_S$). The smaller the blockage ratio, the slower the vessels can pass. For modern waterways in Germany, the blockage-ratio of $n \geq 5.3$ is to be considered as reasonable compromise between economic navigation and ease of navigation (BMVBS, 2011). For modern waterways in Germany, the blockage-ratio of $n \geq 5.3$ is to be considered as reasonable compromise between economic navigation and ease of navigation (BMVBS, 2011). Water level and blockage ratio of the investigated canals are nearly constant throughout the year. To ensure safe travels, the water level is regulated by weirs and locks in order to not fall below or exceed the lower or upper operating water level. As a result, the water level varies by a maximum of 0.65 m throughout the year. In contrast to waterways with significant water level fluctuations such as rivers and estuaries, where damage can spread over the entire slope, the ship waves will impact a smaller area and cause concentrated damage. The exposure of the filter layer is therefore more likely."

Comment 3 (formerly Figure 1): The author mentions in the caption "Impression of the revetment constructions…". Nevertheless, those figures (besides figure d) only show a general image of the canals and do not show the details of the revetments. So, if the authors want to give a better description of the revetments, I would suggest including closer/detailed images.

We agree with Reviewer B that the provided images are not suitable to show the details of the revetments. We thus decided to delete the images.

Comment 4 (formerly L86, now L85ff): I would ask the authors to add a few more details about the instrumentation used (i.e. models, ranges, precise locations, etc.). Also, please discuss the choice of having measurements with 2 Hz, as wave measurements usually are done with higher frequencies. In summary, just discuss in some more detail (but relatively briefly) the instrumentation and the choices made during the measurement and analysis of the data.

Certainly, the measurement set-up and subsequent data analyses are a central feature of the investigations described in the paper. We therefore provided the reader with two additional images and supplementary explanations on this aspect:

> "The waves were recorded by means of absolute pressure probes (Driesen + Kern GmbH, type P-LOG125-B and P-LOG520-A, see Figure 2, left) which were positioned at two (approx. 0.5 m and 1.0 m below water level) or three different depths (approx. 0.5 m, 1.0 m and 1.5 m below water level) of the sloped embankment. The pressure probes have the following characteristics:
>
> - high-precision piezoresistive sensors (0 to 400 kPa),
> - measurement of air pressure (800 to 1200 kPa) and temperature in water and air (0.2 to +80.0°C) and
> - measuring accuracy: 0.1 % of the measuring range.
>
> As a result of the limited data storage capacity of the loggers 15 - 20 years ago, a recording frequency of 2 Hz was chosen. Fortunately, preliminary investigations at that time with a recording frequency of 4 Hz showed that a recording frequency of 2 Hz is sufficient to provide accurate maximum wave height readings. From the measured water level fluctuations, waves are determined for each vessel passage. This process is illustrated in Figure 2, right for H_{stern}. For each wave event the points marked in the graph are picked manually by an experienced consultant. The difference between P4 and P3 yields H_{stern}."

Comment 5 (formerly L92-98, now L116ff): The authors should work on improving the clarity of the data shown in this table, making it clearer to the reader. For example, to reduce/clarify data from Table 3, and maybe present it in other graphs or plots. From what I see there are more than 30 columns, when the key ones are only the first couple of columns and the very last one. If you don't discuss the results in the columns in the middle, just remove them from the table, or put it in annexes. Also, I didn't see a lot of discussion regarding costs, so all this data here on costs confuses the understanding of the important data you want to show (i.e. the average weight of material per year used in the maintenance of each section). In summary, select what is really important and make it easier/clear for the reader to follow your nice data/study.

We agree with Reviewer B that the provided data may confuse the reader. We have concluded that the costs are not essential for the reader to understand this article. Nevertheless, in order to make the data available to other researchers for further analysis, the table section with the costs has been moved from the main body to the appendix. The following explanations have been added in the text:

> "The following investigations focus on the maintenance efforts in terms of the annual mass of placed armor stones rather than on the maintenance costs, since the costs may be affected by factors such as the availability of personnel and machinery as well as by locally and over time differing costs for armor stones. However, in order to make the data available to other researchers, the elicited maintenance costs are presented in the Appendix."

Comment 6: On measurement/maintenance locations, it took me some time to understand: So, you make field measurements in 7 locations (Table 2). Then you have maintenance data for 13 sections (Table 3). But in Table

3 you say 2 sections are not used because there are no field observations (grey marked), which bring us to 11 sections. The question here is: if you don't use it, why is it included in Table 3? Then (in Lines 114-116 you mention that you use the data from one section in another section, right? Lastly, you studied 12 sections as shown in Tables 6 and 7 (i.e. 7 general locations, plus 5 on the DEK (north). In summary, please make sure in text/table/graph exactly which sections you study, in which sections you have all data is available, and for which sections you had to assume or extrapolate data. This will be important later on when the final conclusions are made. This way we could see if the conclusions/discrepancies found can be related to the characteristics of the available data.

Indeed, it may be difficult to link the different data sources due to the quantity of available data and the number of data sources. To accommodate the reviewer's concerns, we made a number of changes:

- We included Figure 3 which visualizes the maintenance measures over time. This allows the reader to see the location and extent of the maintenance measures at one glance.
- We moved the paragraph which outlines which elicited data is used and which not in the case study section to clearly highlight the connection between case studies and data (*formerly L114ff, now L245ff*) and added additional explanatory lines:

> "Due to the unusual construction, the elicited maintenance measures of DEK (South), *Los 14* will not be considered for further analyses. Instead, the elicited maintenance measures of DEK (South), *Los 15* will be used for the stability assessment of DEK (South). This section has a comparable geometry and traffic situation. Furthermore, the elicited maintenance measures for MDK, *Forchheim* are not considered as the canal geometry differs significantly compared to that of MDK, *Strullendorf*, which means that the wave measurements are not transferable. DEK (North), *unterh. Gleesen* and *Bevergen* are not included in subsequent analyses, as these sections are rather short canal sections, one at the confluence of the Mittelland Canal and one at the confluence of the river Ems. In these sections, the traffic behavior is not comparable with the traffic behavior in the other canal sections."

- We introduced one new line in Table 6 (Maintenance location) to show the connection between elicited maintenance measures and field observations and highlighted the extrapolated data with the help of a footnote.

"Table 6| Summary of revetment constructions in the different cross-sections where field measurements are available, subsequently referred to as general case studies.

	DEK(S)-1	DEK(N)-1	MDK-1	MDK-2	WDK-1	WDK-2	WDK-3
Slope inclination	1:3.0	1:2.1	1:3.0	1:3.0	1:3.0	1:3.0	1:3.0
$D_{50,site}$ in mm	175	170	230	230	170	170	190
Maintenance location	Los 15 [a)]	Rodde	Strullendorf	Strullendorf	I	I	II

[a)] Extrapolated data, the measurement location DEK(S)-1 was located at km 81.6, whereas the elicited maintenance measures refer to km 84.0-89.0."

- Additional explanations were supplied for Table 7 to show which data is used for the DEK (North) case studies:

L232ff and Table 7:

> "In the case of DEK (North), detailed data on maintenance measures for each canal section is available (see Table 3). The wave heights measured at DEK (North) km 111.200 (Rodde) are transferred to the other sections allowing for a more site-specific analysis for DEK (North) with respect to geometry and revetment characteristics (see Table 7)."
>
> "Table 7| Summary of revetment constructions in the different cross-sections where field measurements are available, subsequently referred to as DEK (North) case studies. The wave heights originate from DEK (North), km 111.200 (Rodde)."

Comment 7 (formerly L150-155, now 170ff): On the design of rock armoured slopes under wave action, I would also suggest to reference the work from Van der Meer. Also, looking at Section 2.2, I would like to ask the authors to quickly review the equations, nomenclature, indexes, units and the list of symbols to make sure that there is no important typo in the manuscript.

To address the reviewer's concerns, we reviewed the equations, nomenclature, indexes, units and the list of symbols to make carefully.

We acknowledge that there are a number of other authors who contributed to the design equations for loose riprap revetment. However, since the boundary conditions of neither the van der Meer nor the Hudson formula cover waves travelling parallel to the shore, the German design standard "Principles for the Design of Bank and Bottom Protection for Inland Waterway" (GBB) only relies on the "basic features" of these formulas. In order to clarify this issue, we decided to explain the selection of the design equations in more detail to the reader:

> "In 1959, Hudson (1959) proposed a semi-empirical equation to determine the required armor stone diameter against waves based on a large number of flume tests. The Hudson equation offers the advantage of simplicity and is valid for a wide range of armor units and conditions. In 1987, van der Meer (1987) proposed two design equations, one for plunging and one for surging waves which are nowadays also frequently used for the design of loose riprap structures. Neither the van der Meer or the Hudson formula or any related extensions cover waves travelling parallel to the shore, thus, the design equation against armor stone displacement that is formulated in the German design standard "Principles for the Design of Bank and Bottom Protection for Inland Waterway" (GBB) only relies on the "basic features" of the van der Meer nor the Hudson formula. The wave height at which damage occurs is proportional to the relative density, a characteristic grain size of the loose riprap, and a function of the slope inclination. By means of field tests (BAW, 2009) a derivative of the Hudson equation with a proportionality constant (B_B' or B_B^*) was calibrated that can be used for the design of loose armor stone revetments subjected to ship-induced waves in fairways confined laterally and in depth."

Comment 8 (Section 3): My main concern in this review is when reading Section 3. I have seen very interesting conclusions and results, but I will focus on what gives me some doubts. This is mainly found in (formerly) Figures 4 and 5. Let's then look at (formerly) Figure 4, where it is presented that a relation exists between stability (x-axis) and maintenance (y-axis). Looking at the left figure, the "general cases" (in red) show a *decreasing trend for 6 out of the 7 cases, plus an "outlier" corresponding to the DEK (north) section. Furthermore, when looking only at the "DEK (north) sections", a very unclear trend is found, and according to my understanding, I do not see a relation between stability and maintenance. When combining all data, a trend can be found. But when looking in more detail I do not see such a clear trend. And this is also observed in the right figure, and also in (formerly) Figure 5. So, in my opinion, all these results should be re-evaluated and discussed. To this end, previous comments in this review such as the analysis on which locations data were complete, and where data was assumed could help to clarify the conclusions. Maybe one possibility is to separate the analysis of the general cases to the DEK (north) sections. This way the authors could go into more detail on each section and clarify better the results obtained.*

Firstly, as outlined in the responses to other comments we tried to clarify which data was measured (completely) and which data was extrapolated (see also our response to Comment 6 of Reviewer B).

Despite this fact, we still believe that the results of the general case studies and the site specific DEK (North) case studies can be considered together for the regression analyses. The division into case studies is not a methodological distinction, but rather a didactical distinction to present the results to the reader.

Failure probabilities are determined using measured wave events and information on the canal geometry and the revetment construction. These are related to the collected data on maintenance measures. The data on wave heights are extrapolated from the measurement in a cross-sectional profile for the entire canal section. The only difference between the general case studies and the DEK (North) case studies is that in the former the extrapolation of the wave events is carried out for a maximum of approximately 10 km. In the latter, the wave events are extrapolated for a distance of 30 km. The assumption that measured wave events are transferable to other canal sections is a common procedure in revetment design. It is justifiable under the constraint that geometry and fairway in the canal do not change significantly. This is the case for the considered sections of DEK (North).

To increase the comprehensibility, we decided not to label the two subgroups in Figure 6. Additionally, we included the following explanation in our paper:

L236ff:

> "Methodologically, the two case study groups are analyzed in the same way. Failure probabilities are determined using measured wave events and information on the canal geometry and the revetment construction. These are related to the collected data on maintenance measures (see Section 3). The wave heights are spatially extrapolated from the measurement in one cross-sectional profile for an entire canal section. The only difference between the general case studies and the DEK (North) case studies is that in the former the extrapolation is carried out for a maximum of approximately 10 km. In the latter, the wave events are extrapolated for a distance of 30 km. The assumption that measured wave events are transferable to other canal sections is a common procedure in revetment design. It is justifiable under the constraint that geometry and fairway in the canal do not change significantly as it is the case for the considered sections of the DEK (North)."

Furthermore, we reorganized the subsections in the section "Results". In the first subsection "Probabilities of failure from deterministic and probabilistic analyses" we provided the probabilities of failure determined with the deterministic and the probabilistic approach. In a second subsection "Regression analyses with deterministic and probabilistic analyses" we now present the results of the regression analyses of both the deterministic and the probabilistic approach. We believe that this will facilitate the understanding and comparison of the different calculation approaches for the reader.

Additionally, in that way, we could include an investigation of the effects the above outlined assumptions have on the determined regression equations. We investigated two different data compositions and three types of regression equations. To rely the results to the reader, we included a new Table (Table 11) which provides information on p-values, coefficients of determination (R^2) and regression coefficients for the different analyses. The former Figure 4 and Figure 5 were replaced by Figure 6 which now features four images to show the different regression functions and data used throughout this study. We hope that this will make the analysis process, the results and the subsequent conclusions more comprehensible for the reader

L306ff:

"Various regression analyses (linear, exponential and logarithmic) were conducted (see Table 11) to examine the correlation between maintenance measures and probabilities of failure to identify a suitable mathematical model. Each equation is supposed to predicted the required maintenance measures A in kg/m² as a function of the different probabilities of failure. To account for different assumptions during the data assessment, two different data combinations are investigated. Studies that are abbreviated with "-all" consider all available data points (general case studies + DEK (North)). To evaluate the effect of the extrapolation of the wave heights, regression analyses with a reduced data set were conducted ("-general"). For this purpose, only the general case studies are considered.

Table 11| Summary of p-values, coefficients of determination (R^2) and the regression parameters. The following abbreviations and equations are used: LIN – A [kg/m²] = $a \cdot p_{f,a} + b$, LN – A [kg/m²] = $a \cdot \ln(p_{f,a}) + b$, EXP – A [kg/m²] = $e^{a \cdot p_{f,a} + b}$. The grey shaded cells indicate that the investigated relation is not significant (based on a significance level α = 0.05).

Model	Data				
	η	$p_{f,d}$	$p_{a,d}$	$p_{f,p}$	$p_{a,p}$
LIN-all	p = 0.0209, R^2 = 0.4050, a = 9.42, b= –9.80 [a]	p = 0.0010, R^2 = 0.4892, a = 68.76, b = 0.25 [a]	p = 0.0090, R^2 = 0.5036, a = 7.89, b = –0.74 [a]	p = 0.0080, R^2 = 0.5118, a = 126.48, b = 0.41 [a]	p = 0.0047, R^2 = 0.5644, a = 9.31, b = 0.02 [a]
LIN-general	p = 0.1732, R^2 = 0.2023, a = 5.36 [a], b= –5.36 [a]	p = 0.01629, R^2 = 0.6599, a = 103.52, b = –0.28	p = 0.0710, R^2 = 0.4134, a = 5.66 [a], b= –0.34 [a]	p = 0.0003, R^2 = 0.9271, a = 157.05, b = 0.11 [a]	p = 0.0008, R^2 = 0.8925, a = 9.04, b = –0.01 [a]
LN-all [b]	p = 0.0263, R^2 = 0.3769, a = 12.14, b= –0.42 [a]	p = 0.0293, R^2 = 0.4466, a = 2.67, b = 12.96	p = 0.0439, R^2 = 0.3856, a = 3.92, b = 6.59	p = 0.0159, R^2 = 0.4377, a = 1.27, b = 9.91	p = 0.0180, R^2 = 0.4233, a = 1.49, b = 6.45
LN-general [b]	p = 0.2203, R^2 = 0.1381, a = 5.97 [a], b = 0.07 [a]	p = 0.1493, R^2 = 0.4050, a = 2.09 [a], b= 10.27 [a]	p = 0.2305, R^2 = 0.2382, a = 2.33 [a], b = 4.14 [a]	p = 0.0945, R^2 = 0.3507, a = 0.74 [a], b= –0.29 [a]	p = 0.1080, R^2 = 0.3198, a = 0.82 [a], b = 3.78 [a]
EXP-all	p = 0.0014, R^2 = 0.8132, a = 3.19, b = –3.56	p = 0.0085 [c], R^2 = 0.6628, a = 21.41, b= –0.22 [a]	p = 0.0006 [c], R^2 = 0.8605, a = 2.62, b= –0.54 [a]	p = 0.0048 [c], R^2 = 0.7190, a = 41.90, b= –0.29 [a]	p= 0.0009 [c], R^2 = 0.8388, a = 3.15, b= –0.46 [a]
EXP-general	p = 0.0131, R^2 = 0.9609, a = 3.97, b = –4.41	p = 0.0087 [c], R^2 = 0.9740, a = 47.87, b = –0.55	p = 0.0123 [c], R^2 = 0.9633, a = 3.19, b = –0.59	p = 0.0101 [c], R^2 = 0.9700, a = 64.77, b = –0.50	p = 0.0103 [c], R^2 = 0.9693, a = 3.75, b = –0.52

[a] Parameter is not significant (for a significance level α = 0.05).

[b] Analyses do not include MDK-2 and DEK(S)-1, since $p_{f,d} = 0.00$ or $p_{a,d}$ cannot be considered in a logarithmic regression.

[c] Analyses do not include WDK-1, WDK-2 and WDK-3, since zero maintenance cannot be considered in an exponential regression.

For the hypothesis tests, a significance level $\alpha = 0.05$ is assumed. In the context of Table 11, this means that the null-hypothesis (model does not explain the data) is rejected if $p < 0.05$. In simple terms, if $p < 0.05$, we can assume that the investigated relation is significant. Comparing the p and R^2 values of functions that indicate an increasing relation between maintenance and failure probability (linear, exponential) with the logarithmic function, then, the majority of case studies indicates that an increase in maintenance is associated with a larger failure probability. Furthermore, for the majority of the linear and exponential regressions, an increase of R^2 is observed for the reduced data (-general) compared to the regression analyses with all data. The best fit can be achieved for the exponential function with the reduced data, but also with all data. Yet, for these regressions, it must be noted that the maintenance measures at WDK were omitted, since the exponential regression cannot handle regression values ≤ 0. As a consequence, EXP-general features only four data pairs.

In view of all regression analyses, it can be noted that the regression analyses with the probabilistically determined failure probabilities show a better agreement with the elicited maintenance measures than the deterministically calculated values. The deterministic approach shows an improvement in the correlation between revetment stability and maintenance (smaller p-value, larger R^2) when using the annual frequency of limit state exceedances ($p_{a,d}$) rather than η. For both the deterministically and the probabilistically calculated failure probabilities, it can be noted that the regression analyses which use p_a as regressor show a better fit than when p_f is used."

Comment 9 (formerly Table 10): Looking at this table I would suggest looking for fitting equations that do not give negative values of material, and considering the differences in percentage. Thus, besides that the differences in material are larger for larger pv values, the percentage decreases (according to the results shown, the differences would be 100% for the smaller pv values, but around 20% for the larger pv values).

We re-assessed the correlations (see also the previous comment) and examined the significance of the regression parameters. Based on the p-values of the intercept values, which were greater 0.05 and, thus, not significant, we decided to use the linear regression function with the intercept $b = 0$. Additionally, we added an exponential regression function to the analyses which shows a good fit to the data, but has the disadvantage that zero maintenance cannot be considered. For all regression equations confidence intervals were calculated to provide information on the uncertainty of the fits. As a result of the new linear regression equations with $b = 0$ and the new exponential equations, the results in Table 12 (formerly Table 10) were also revised. Furthermore, to discuss the decreasing percentage as suggested by Reviewer B, we replaced the column with the absolute differences by a column with percentual differences in Table 12 (formerly Table 10). Section 3.3 was revised as follows:

"To compare the regression equations determined with the "deterministic" and "probabilistic" failure probabilities in more detail, the linear and exponential equations that are based on the complete data are employed in a small parameter study. For this purpose, the required maintenance measures predicted by the different regression equations are compared for a total of 15 different traffic volume and p_v combinations.

The required maintenance that is predicted by the linear equation based on the deterministic analyses is smaller than the required maintenance predicted by the linear equation based on the probabilistic analyses (see Table 12). The differences between the modes LIN-DET and LIN-PROB range from 0.00 kg/m² for small p_v values to 1.40 kg/m² for large p_v values. As a result of the same intercept $b = 0$ in the LIN-DET and LIN-PROB model, the percentual difference between the two approaches stays constant with an increasing number of vessels or an increasing p_v.

The required maintenance that is predicted by the exponential equation based on the deterministic analyses is also smaller than the required maintenance predicted by the exponential equation based on the probabilistic analyses. The differences between the modes LIN-EXP and LIN-PROB range from 0.00 kg/m² for small p_v values to 9.60 kg/m² for large p_v values. In contrast to the linear models, the percentual difference between the two approaches increases with an increasing p_v and an increasing number of vessels. These results confirm a drawback of the deterministic calculation approach which is sensitive to small sample sizes and small probabilities of failure. The p_f may be underestimated since $\eta > 1$ is not found for all observations."

Table 12| Results of parameter study for a comparison of the linear and exponential equations found with the deterministically and the probabilistically determined failure probabilities. Regression equations that were determined with the "deterministic" failure probabilities are denoted by "-DET" and $p_{a,d}$, whereas "-PROB" and the $p_{a,p}$ in the equation indicates that the equation is derived with probabilities of failure obtained from Monte-Carlo simulations.

"Case studies			Linear regression			Exponential regression		
Vessel passages per year	p_v	p_a	LIN-DET $A = 7.89\ p_{a,d}$	LIN-PROB $A = 9.31\ p_{a,p}$	Diff	EXP-DET $A = e^{2.62 p_{a,d}}$	EXP-PROB $A = e^{3.15 p_{a,p}}$	Diff
	--	--	kg/m²	kg/m²	%	kg/m²	kg/m²	%
1000	1.00E-06	1.00E-03	0.01	0.01	15.3	1.00	1.00	0.1
5000	1.00E-06	4.99E-03	0.04	0.05	15.3	1.01	1.02	0.3
10000	1.00E-06	9.95E-03	0.08	0.09	15.3	1.03	1.03	0.5
15000	1.00E-06	1.49E-02	0.12	0.14	15.3	1.04	1.05	0.8
20000	1.00E-06	1.98E-02	0.16	0.18	15.3	1.05	1.06	1.0
1000	1.00E-04	9.52E-02	0.75	0.89	15.3	1.28	1.35	4.9
5000	1.00E-04	3.93E-01	3.10	3.66	15.3	2.80	3.45	18.8
10000	1.00E-04	6.32E-01	4.99	5.89	15.3	5.24	7.32	28.5
15000	1.00E-04	7.77E-01	6.13	7.23	15.3	7.66	11.56	33.8
20000	1.00E-04	8.65E-01	6.82	8.05	15.3	9.64	15.24	36.8
1000	1.00E-02	1.00E+00	7.89	9.31	15.3	13.73	23.33	41.1
5000	1.00E-02	1.00E+00	7.89	9.31	15.3	13.74	23.34	41.1
10000	1.00E-02	1.00E+00	7.89	9.31	15.3	13.74	23.34	41.1
15000	1.00E-02	1.00E+00	7.89	9.31	15.3	13.74	23.34	41.1
20000	1.00E-02	1.00E+00	7.89	9.31	15.3	13.74	23.34	41.1

"