Prediction of short-term non-linear response using screening combined with multi-fidelity Gaussian Process Regression

van Essen, S.M.; Scholcz, T.P.; Seyffert, Harleigh C.

DOI
10.1115/OMAE2023-100954

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the ASME 2023 42nd International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2023)

Proceedings of Proceedings of the ASME 2023 42nd International Conference on Ocean, Offshore and Arctic Engineering
OMAE2023
June 11-16, 2023, Melbourne, Australia

OMAE2023-100954

PREDICTION OF SHORT-TERM NON-LINEAR RESPONSE USING SCREENING COMBINED WITH MULTI-FIDELITY GAUSSIAN PROCESS REGRESSION

Sanne van Essen
Delft University of Technology (Delft)
MARIN (Wageningen)
The Netherlands
s.m.vanessen@tudelft.nl

Thomas Scholcz
MARIN
t.p.scholcz@marin.nl

Harleigh Seyffert
Delft University of Technology
h.c.seyffert@tudelft.nl

ABSTRACT

Predicting wave impact design loads is crucial for ensuring safety and performance of maritime structures, but it is challenging due to the complexity and rarity of these events. Existing methods are mainly suitable for prediction of weakly non-linear responses, or are very computationally expensive. Highly non-linear responses require a fidelity level that can only be achieved with expensive CFD or experiments, leading to sparsely populated exceedance distributions. A new event-based multi-fidelity method called 'adaptive screening' therefore combines elements of screening, multi-fidelity Gaussian Process Regression and adaptive sampling, to more efficiently predict highly non-linear loads. It is applied at the level of the response peak exceedance probability distributions. A simplified case study using second-order wave data validates the effectiveness of the method in accurately predicting short-term design loads. The new method predicts more accurate MPM results than the conventional method recommended by class societies and the ITTC, while also significantly reducing the required HF simulation time. The new method has a deviation of only 0.3-3.5% from the true 1-hour MPM over all test cases, compared to the conventional method's deviation of 5.2-6.7%. The HF simulation time required to do this is 91 times shorter with the new method (0.033 versus 3 hours per sea state). The new method is not very sensitive to input noise as long as HF samples are selected properly, and the application of the method to the exceedance distributions works.

1 INTRODUCTION AND OBJECTIVES

1.1 Background

Prediction of wave impact loads (green water, slamming, wave-in-deck) is crucial in designing safe and effective maritime structures, as they can cause significant damage, endanger the crew and decrease performance (e.g., [1, 2, 3, 4, 5, 6]). However, obtaining reliable design loads for these impacts is challenging due to their complexity and rarity. High-fidelity (HF) tools are needed to account for the impact complexity, and low-fidelity (LF) tools are necessary to cope with the rare occurrences. A review of multi-fidelity methods to predict wave impact loads is provided by [7], which also highlights the importance of this issue by assembled wave impact accidents. The review includes screening, Gaussian Process Regression (GPR), adaptive sampling (AS), and response-conditioning methods. Alternatively, conventional brute force methods can be applied.

While existing multi-fidelity methods are suitable for weakly non-linear responses, highly non-linear responses require a very high fidelity level that can only be achieved with a few wave events in CFD or experiments. This event-based approach is challenging for the existing methods due to the sparseness of the resulting exceedance distribution. To tackle this issue, we propose combining multi-fidelity event-based screening with the sound statistical basis and uncertainty estimate of GPR. The new method is aimed at providing accurate design value predictions for highly non-linear response, while minimising HF simulation time. Additionally, AS can be used to further improve its efficiency.

1.2 Objectives

Our new method, called 'adaptive screening' is introduced here, and a first validation using weakly non-linear wave data is presented. We aim to answer the following questions:

- Can the method be applied to the response peak exceedance distributions, or is another 'abstraction level' more suitable?

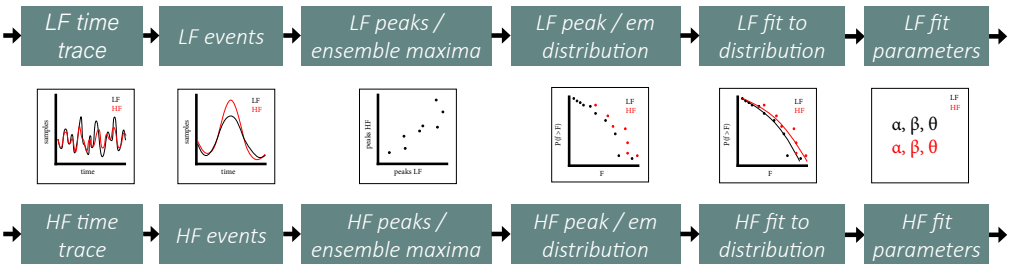


FIGURE 1: *Some of the 'abstraction level' options to learn the relation between LF indicator and HF non-linear response (not exhaustive).*

- Can the new method accurately determine short-term most probable maximum (MPM) values of non-linear responses?
- How does the method perform under conditions of noise or less effective low-fidelity screening indicators?
- How does the accuracy and required simulation time of the method compare to a conventional brute force method?

Design for seakeeping distinguishes long- and short-term responses. Long-term response considers variability over all wave conditions a ship may encounter, whereas short-term response considers variability within one wave condition. In theory, adaptive screening would be able to consider both, but this study focuses on predicting short-term design loads.

1.3 Organisation

Section 2 explains some of the existing methods and Section 3 the required statistical concepts. Section 4 introduces the new adaptive screening method. Section 5 describes the simplified case study used in the present publication, and Section 6 presents the results for this case study. Finally, Section 7 contains conclusions and planned follow-up work.

2 EXISTING METHODS

This section introduces the screening, (multi-fidelity) GPR, and AS methods. The assessment of non-linear ship responses requires some form of multi-fidelity evaluation. These methods establish a link between LF and HF variables at different abstraction levels or stages in the time trace analysis (see Figure 1). However, as explained in the introduction, the existing methods cannot predict design values for highly non-linear responses. They are therefore not widely used by designers for such responses. Section 2.4 describes a conventional wave impact load prediction method used in the industry.

2.1 Screening

Screening methods rely on an LF indicator or surrogate variable that reflects the order statistics of the target HF non-linear variable [8, 9, 10, 11]. By performing Monte-Carlo Simulation (MCS) on the LF indicator, many wave conditions or realizations can be screened quickly for non-linear responses in both short and long-term scenarios. Critical events identified through this screening can then be evaluated using HF methods such as CFD calculations or experiments. Finally, combining the LF statistics with the HF loads enables obtaining HF design loads with a specified probability of exceedance.

An ideal indicator would have identical order statistics as the target non-linear response, where the highest indicator value appears in the same wave event as the highest non-linear response value, and so on. The indicator can take various forms, such as the rise time of relative wave elevation, wave crest front steepness, or a combination of wave steepness, pitch amplitude, and relative wave elevation amplitude. Studies validating wave impact indicators are reviewed in [7]. The screening process includes the following steps:

1. Define long-term wave information for the operational profile (e.g., scatter diagrams).
2. Select critical sea states (LF long-term screening).
3. Generate wave and response realisation time traces.
4. Select critical events (LF short-term screening).
5. Generate input conditions for a HF tool for these events.
6. Perform HF calculations for the events to obtain HF loads.
7. Combine the LF probability from steps 2 and 4 with the HF loads from step 6 in a long-term load distribution.

Here, we focus on short-term screening (steps 3 to 7). Results for a full short-term procedure in one example wave condition are provided by [11]. Screening links LF and HF responses at the level of peak or ensemble maximum values in Figure 1.

2.2 Multi-fidelity Gaussian Process Regression

GPR, also known as Kriging, is a form of supervised learning that assumes a considered process can be described by an infinite-dimensional multivariate Gaussian distribution. This approach has advantages over traditional regression, such as less restriction, more influence from data points (although some assumptions need to be made on the smoothness of the underlying curves), and provision of uncertainty bands for new estimates. It also has benefits over more data-driven techniques: it works well with small datasets and is less 'black-box' since hyperparameters are not entirely abstract. GPR is a non-parametric Bayesian method that infers a probability distribution over all possible values using a prior distribution and updated posterior distribution from observations. GPR can be used in single-fidelity [15] or multi-fidelity [16, 17] forms. More information on single- and multi-fidelity GPR is included in Appendix A.

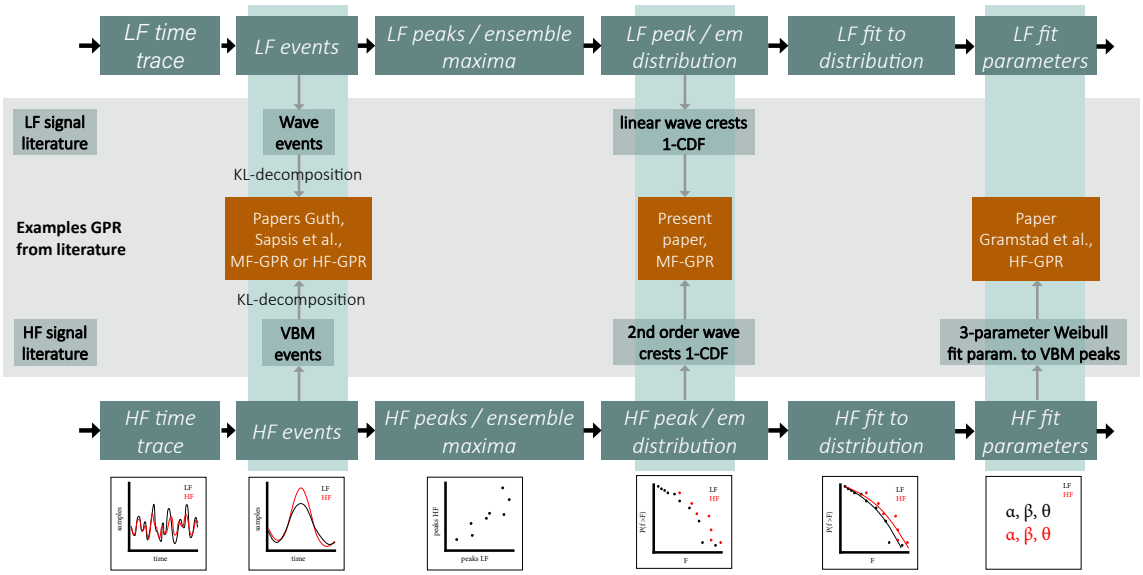


***FIGURE 2**: Abstraction level in Figure 1 for learning the relation between LF indicator and HF non-linear ship response in waves in some of the applications in literature and the new method. References: [12, 13, 14]. KL-decomposition = wave event decomposition in 'Karhunen-Loeve' components (for details see [13]). CDF = cumulative distribution function.*

2.3 Adaptive sampling

An advantage of GPR is that it produces an uncertainty estimate. This can be used to define an acquisition function that finds the optimal location for the next LF or HF sample during an AS method. The acquisition function is designed to achieve a specific target, such as global surrogate optimisation (e.g., [18, 19]) or global surrogate accuracy (e.g., [20]). This process, known as adaptive / sequential sampling or active learning / search, balances exploration and exploitation. Various sampling strategies are available, with popular acquisition functions including Probability of Improvement (PI), Expected Improvement (EI, leading to so-called Efficient Global Optimisation EGO), and Upper Confidence Bound (UCB). When global surrogate accuracy is the target, samples can be added at the location with the largest predictive variance. GPR combined with AS has been applied to time-dependent reliability problems [21]. Applications of multi-fidelity GPR and AS to ship response in waves are limited in literature, but a few were found. For example, [12] used one-step lookahead AS with single-fidelity GPR to learn 3-parameter Weibull fits to vertical bending moment (VBM) maxima for a tanker. [13] used AS with single-fidelity GPR to match wave events to VBM events for a marine vessel, while [14] used multi-fidelity GPR to predict short-term VBM maxima for the same vessel. Figure 2 illustrates the abstraction level at which these studies operated.

2.4 Conventional brute force method

Extreme impact loads are hard to predict, and can have serious consequences for the performance, safety and cost of a marine structure. As a consequence, designers tend to rely on class societies and proven technology to determine such loads. The environmental contour approach (e.g., [22, 23]) is commonly used, in which a few sea states are chosen and modelled in an experiment of HF simulation. Several wave realisations or seeds with a duration of 30 minutes to 3 hours are run. This is recommended by class societies and others for highly non-linear response: e.g., it is advised to run at least 10 seeds [24]; 16 seeds [25, 26] (in the latter reference derived from the recommendations for an exposure duration of 1 hour); 20-30 seeds [27, 28] or even 50-60 seeds [29]. Alternatively, it is recommended to run only 1 seed with a 'sufficient' number of events [26, 30, 31]. 3-parameter Weibull fitting Equation (1) can then be applied to the top 20-40% of the measured non-linear response peaks per seed (see e.g., [28, 32, 33, 34]), from which the MPM value for the target duration can be derived (see Section 3). The uncertainty of this MPM as a function of the number of seeds is studied in [35, 36], showing that this ranges from RMSE error $12\%H_s$ in 1 seed to less than 1% in 20 seeds for the wave crest height MPM, for 1 hour (as is common to use for sailing ships). This conventional approach is 'brute force', but usually fewer seeds are used than would be required for a proper MCS. In practice, more than one sea state, speed, heading and ship configuration needs to be tested, and experimental time is limited. It is therefore common to test only 1 seed per sea state, following the ITTC [30]. Testing 2 or 3 is uncommon, and testing 10 is occasionally done in the offshore industry. Obviously, the conventional method is very (computationally) expensive due to the long HF simulations or experiments.

$$p(x \geq X)_{weibull} = \exp\left(-\left(\frac{x-\theta}{\alpha}\right)^{\beta}\right) \tag{1}$$

3 STATISTICS AND DEFINITIONS

As explained in the introduction, this study aims to predict MPM values for non-linear responses in a given exposure duration. The MPM value can be obtained in different ways, as discussed in e.g., [35, 36]. A design load is related to either a target exceedance probability or a return period; here we use exceedance probability. The exceedance probability distribution (EPD) is defined as one minus the cumulative distribution function (CDF). Different EPD types are distinguished. The EPD of the (ensemble) maximum values of each seed is called DEM, while the EPD of all response peaks in the experimental time traces is called DSR or DNR depending on whether it is based on a single or on multiple realisations. The DEM is defined in Appendix B, but not used here. The DSR and DNR are defined in Equation (2), with $S_C(N)$ all response peak or crest values over N realisations, exceedance probability defined based on the total number of response peaks n_p and probability $1/n_p$ for the maximum value.

$$\begin{aligned} DSR(s) &= p\left(S_C(1) \geq s\right) \\ DNR(s) &= p\left(S_C(N) \geq s\right) \end{aligned} \tag{2}$$

The MPM can be derived from either the DEM or the DNR. The MPM from the DNR $\widehat{S_C}(N)$ is defined in Equation (3), where $n_{p,e}$ is the average number of peaks in the considered exposure duration. It was shown in [35, 36] that these two options lead to similar MPM values for wave crest heights and wave impact force peaks. In our case study, we use Equation (3). For reference, Appendix B shows the MPM values for our case study when derived from the DEM.

$$\begin{aligned} p_{exp} &= 1/n_{p,e} \\ DNR\left(\widehat{S_C}(N)\right) &= p\left(S_C(N) \geq \widehat{S_C}(N)\right) = p_{exp} \\ \widehat{S_C}(N) &= \phi_{P_{exp}}\left[S_C(N)\right] \end{aligned} \tag{3}$$

4 NEW METHOD: ADAPTIVE SCREENING

We introduce a novel approach called adaptive screening, which combines screening discussed in Section 2.1, multi-fidelity GPR in Appendix A.2 and AS in Section 2.3 to predict extreme HF response. Our aim is to address the sparse statistics challenge posed by highly non-linear ship responses and obtain the HF MPM value for a given exposure duration. Generally we only have a few HF data points and many LF data points available. Fitting or regression can reduce the influence of interpolation issues on the sparse HF distribution. The next section presents the steps taken in the new method, after which the advantages of the method compared to existing methods, the utilized assumptions and the abstraction level are discussed.

4.1 Steps

The new method involves the following steps. Steps 1-6 relate mostly to screening, steps 7-8 to GPR and step 9 to AS.

1. Define an LF indicator with a strong statistical relation to the target HF response, as would be done in a screening method. See Section 2.1 for indicator options.
2. Perform LF Monte-Carlo Simulation for a large number N of wave seeds with the same exposure duration. See [35, 36] for an example of the required N for wave crests, green water impact forces and wave-in-deck impact forces.
3. Identify all LF indicator events in the time traces.
4. Select M indicator events from all available seeds, as would be done in a screening method. Diffent sampling strategies can be used, e.g., M events with the highest LF indicator value, or M events around the target exceedance probability of the LF indicator.
5. Find the corresponding HF response for these screened LF events, by running HF CFD or experiments for the selected events.
6. Calculate the DNR of the derived HF data points using (Equation (2)), with an exceedance probability based on the total number of LF indicator data points. In other words, we assume that the order statistics of the HF response are similar to those of the LF indicator and use the full LF event set to define n_p for the selected HF events.
7. Apply 1D single- or multi-fidelity GPR to the LF and HF DNR samples. In order for the GPR to perform well in the tail of the distribution, it is applied to the logarithm of the LF and HF DNR in Equation (2).
8. Estimate the MPM and its uncertainty using Equation (3) applied to the (MF-)GPR prediction for the DNR, with $n_{p,e}$ based on the average number of LF response peaks in the target exposure duration.
9. Compare the MPM uncertainty to a given tolerance. If it is above the tolerance, add a new HF sample using one of the AS strategies discussed in Section 2.3. This new sample has to be selected from the available LF samples. Repeat steps 5 to 9 until the MPM uncertainty is below the tolerance.

4.2 Advantages

Compared to the conventional method in Section 2.4, the new method will require a much shorter HF simulation time in order to reach a similar MPM accuracy. Compared to the existing multi-fidelity methods in Section 2, the new method also has advantages. Firstly, by combining MF-GPR with screening we can generate new HF samples at low exceedance probability values, without performing a very long HF simulation. The LF statistics are used to identify a wave event at any given exceedance probability (within the limits of LF MCS) for which we obtain the HF load. This cannot be done with GPR or regression methods without screening, and makes the new method suitable for highly non-linear responses. Secondly, the new approach uses MF-GPR in step 7, which provides uncertainty estimates for predictions and can take into account features of both LF and HF data points. This can be used as input for AS, which can make the procedure more efficient by reducing the total number of required samples and targeting the sampling at interesting areas of the design space.

Thirdly, by using multi- instead of single-fidelity regression we can reduce the sparse HF statistics problem. We use the LF statistics not only to select the event in screening steps 1-5, but also to improve the regression in step 7.

4.3 Assumptions

The screening part of the method (steps 1 to 6) is only useful when the order statistics of the LF indicator and HF response are sufficiently similar. See Section 2.1 for studies that validated this assumption for wave impact indicators. In steps 7-8, the GPR assumes some degree of smoothness of the distributions, and the multi-fidelity GPR version assumes that the shapes of the LF and HF distributions are similar.

4.4 Abstraction level

The abstraction level at which the new approach links the LF and HF responses is depicted in Figure 2. In the present study we choose to operate at the level of the peak exceedance probability distributions (DNR), in 1D with the exceedance probability as only input variable. For sparse non-linear statistics it is sensible move towards the right of the figure and apply some form of regression, as we are not able to run HF simulations for long time traces or for many events. The selected level is closest to the value we are trying to predict, and it aligns with the conventional method in Section 2.4. However, it is not the only option. If we conclude that we miss important physics we may have to go back in abstraction level or add more input variables (multi-variate regression). If the prediction is easy on the other hand we may try to go a step further in abstraction level.

5 SIMPLIFIED CASE STUDY

The present paper applies the new approach from Section 4 to a simplified case study. The last step of the new approach, adaptive sampling, is not yet applied.

5.1 Monte-Carlo simulations

The LF indicator in this case is a linear long-crested Gaussian wave signal, and the HF response is the second-order long-crested wave signal. We perform MCS for both variables, allowing for 'true' HF validation material. We also have the theoretical linear Rayleigh and second-order Forristall [37] 'true' reference exceedance distributions. The linear wave elevation time traces were generated using a long-crested JONSWAP wave spectrum with $H_s = 10$ m, $T_p = 11$ s, $\gamma = 3.3$. The sea state steepness is $H_s/T_p^2 \sim 0.083$, so relatively steep. The water depth was set at 30 m (relatively shallow to generate large second-order contributions). No current is assumed. The second-order wave signals were generated using the Python toolbox Pyseawave of the Co-operative Research Ships (which uses the formulations of [38]). Frequency limits 0-5 rad/s were used for the interaction between first-order components. A small part of the resulting time traces is shown in Figure 3. We used $T_1 = T_p/1.198$ for JONSWAP to obtain the reference Forristall distribution. To ensure convergence of the wave crest height extreme value statistics for 1 hour duration, we require at least 8-22 realisations for a root mean square error (RMSE) convergence criterion of 3% H_s, according to [35,36]. In this case, we use 50 realisations of 1 hour duration for the MCS.

5.2 Noise

The data above perfectly follow linear and second-order wave theory. In a real non-linear response problem, especially the HF data points are expected to be noisy due to numerical issues in CFD, basin effects in experiments, etc. Noise in this definition may also indicate differences in the order statistics between the LF indicator variable and the HF response variable, due to e.g., additional physics modelled in the HF tool or a less good LF indicator. The LF data points would be generated using a lower-order tool (e.g., linear or weakly non-linear potential flow), so these will include less noise. Two 'noise' versions are considered here:

- 'MeasNoise': Gaussian measurement noise with zero mean and 0.6 m standard deviation, added to the HF time series.
- 'PhysNoise': Time series based on another linear wave spectrum, added to the HF time series. This situation could for instance represent HF experiments where there is still a residual wave system in the basin from a previous experiment. A JONSWAP spectrum with $H_s = 1.5$ m, $T_p = 12$ s, $\gamma = 3.3$ was selected, with random phases that are independent from the phases of the main wave system.

This changes the input data such that the LF indicator (peaks linear waves) is a less good indicator for the HF response (peaks second-order waves + noise). Similar plots as presented above for the 'clean' data are provided in Appendix C for the data including noise. These show that noise introduces more scatter in the relation between the LF and HF wave crests, as expected.

5.3 Truth and considered approaches

The target of the case study is to obtain the MPM value of the HF response (second-order wave crest height) for an exposure duration of 1 hour. We want to compare the MPM value from different versions of the new method (Section 4) to the 'true' value and to the value derived from the conventional approach (Section 2.4). For the 'truth' we have two references: the full MCS second-order wave crest DNR, and the Forristall theoretical DNR (which is a sanity check for the MCS). This gives us the following options to derive the 1-hour MPM value for the HF non-linear wave crest heights:

1. Truth based on the DNR from HF MCS.
2. Truth based on the theoretical Forristall distribution.
3. Conventional prediction based on approach in Section 2.4.
4. New prediction based on 'screening + single-fidelity GPR' applied to only the HF training data (Section 4).
5. New prediction based on 'screening + multi-fidelity GPR' applied to both the LF and HF training data (Section 4).
6. New prediction based on 'screening + Weibull-fitting' applied to only the HF training data.

The settings for new methods 4 to 6 are discussed in Section 5.4 and those for conventional method 3 in Section 5.5. For

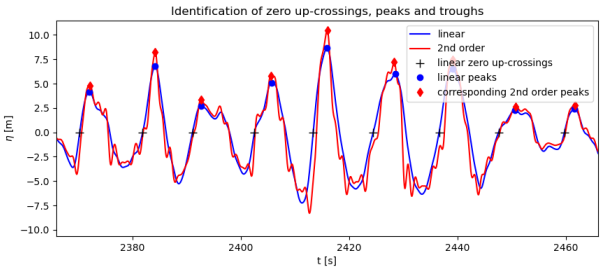


***FIGURE 3**: Matched LF indicator peaks (linear wave crests) and HF response peaks (second-order wave crests) without noise - zoom-in around the highest crest in the first realisation.*

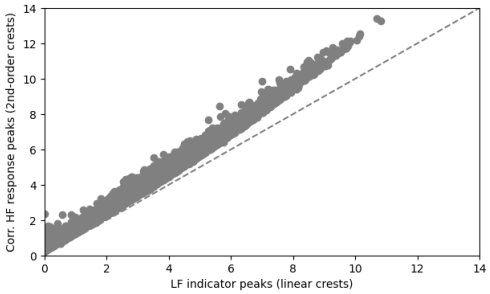


***FIGURE 4**: Scatter plot of matched LF indicator peaks (linear wave crests) and HF response peaks (second-order wave crests) over all 50 1-hour realisations. Time traces without noise.*

method 6, steps 1-6 are identical to the steps for the adaptive screening procedure in Section 4. Next, a 3-parameter Weibull distribution (see Equation (1)) was fitted using least-squares fitting to the HF DNR. The reasons for selecting Weibull and least-squares fitting are explained in [35]. This approach was added in order to evaluate whether GPR adds something with respect to simpler regression. The 1-hour MPM for all methods and the truth can be estimated using Equation (3), with $n_{p,e}$ based on the average number of LF linear wave crests in 1 hour.

5.4 New method choices

For the new approaches in Section 5.3, some choices were made: LF and HF sampling, GPR implementation and settings.

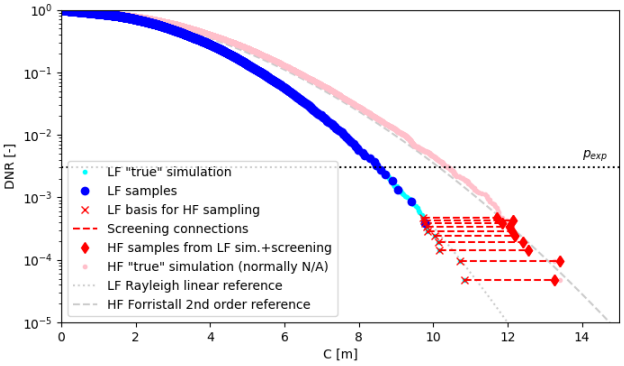
Sampling of LF training data: In order to generate training data, we need a sampling strategy for the LF and HF data in step 4 of the approach. First, we discuss the LF samples. Using all LF wave crests in multiple 1-hour realisations as LF training data leads to memory issues in the generation of the covariance matrices in Equations (14) to (16). The LF training data therefore consist of every 10^{th} wave crest in the DNR based on the LF Monte-Carlo simulation, covering the whole range of exceedance probabilities.

Sampling of HF training data: In order to avoid having to do CFD simulations for each selected event in step 5 of the procedure in Section 4 in our simplified case study, we used the HF MCS results to sample the HF training data. In the case study we can do this because we have a HF MCS available - in a real application we do not, so then we must do CFD or experiments in step 5. The case study therefore assumes that the generated CFD HF loads for the selected events are accurate, and validates the rest of the approach. In order to use the HF MCS results, we match the LF indicator peaks from the LF MCS (linear wave crests) in time to the response peaks from the HF MCS (second-order wave crests) using the following procedure. Firstly, the LF zero up-crossing indicator peaks were identified ('linear peaks' in Figure 3). Next, the corresponding maximum values of the HF response ('corresponding 2nd order peaks'), between each set of LF indicator zero up-crossings were identified. Because we use the maximum HF value between LF zero up-crossings, it does not matter if the LF and HF peak are slightly shifted in time. This can be useful when you use an indicator value that has a phase difference with the HF response. The scatter plot resulting from the peak matching (Figure 4) shows that the second-order wave crests are larger than the linear crests. It also shows that the LF indicator is not perfect; there is some scatter in the data. For the HF training data we used two screening sampling strategies:

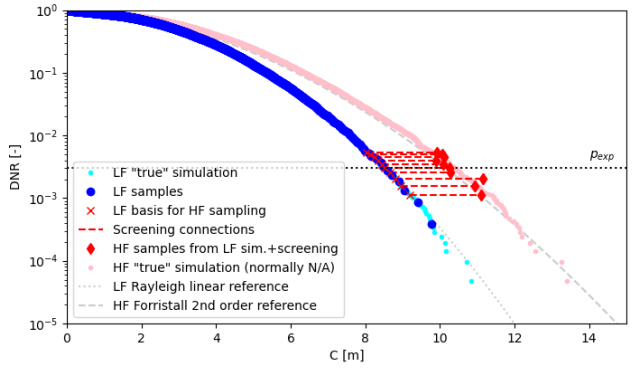
- 'SampMax': The HF peaks corresponding to the 10 maximum LF indicator peaks.
- 'SampPexp': The HF peaks corresponding to 10 LF indicator peaks around p_{exp}. 'Around' is defined by the LF DNR indices [-50,-40,-30,-20,-10,0,10,20,30,40]+i_{pexp} (where i_{pexp} is the closest LF index for p_{exp} in the DNR).

In the case study the HF samples in both strategies are generated using the time trace matching discussed above (visualised by the horizontal lines in the DNR of Figure 5). The first strategy relates most to the traditional way to use screening data in the studies discussed in Section 2.1. However, for derivation of MPM values from the DNR (see Section 3), the focus around p_{exp} in the second option is more logical. Adaptive sampling is not yet applied. The resulting true and training data are summarised in Figure 5. For reference, the theoretical linear Rayleigh and second-order Forristall wave crest distributions are also included in this plot, showing that the MCS data follow them relatively well (as expected). As said, the shown full DNR HF result is available for the present simplified case study, but we only use it for validation and for sampling the HF training data.

GPR implementation and settings: The MF-GPR theory was first implemented and tested based on the papers of [16, 17]. This was great for the understanding of the principles. However, for the final results a more robust open-source Python toolbox was used: Emukit v0.4.9 [39] (which in turn uses GPy v1.10.0 [40]). All (MF-)GPR kernels were selected to be square exponential (Equation (5)). Some other kernels such as exponential quadratic were tried, but gave almost identical results. The (MF-)GPR noise settings were imposed instead of learned. They were all set at 0.1 m standard deviation (for input waves with and without noise). A small sensitivity study showed that some noise is always better than no noise in the GPR; this makes the resulting uncertainty band slightly bigger, but also stabilises the results. Applying GPR



(a) Sampling: SampMax.



(b) Sampling: SampPexp.

***FIGURE 5**: DNR of the true and training data over all 50 1-hour realisations for the two sampling options, without noise. Both sample options with in total 10 HF samples.*

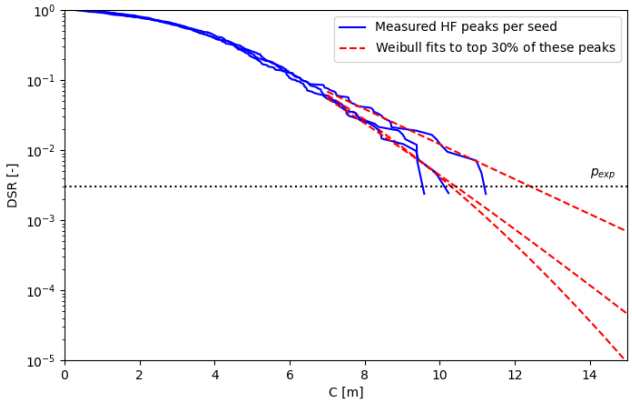


***FIGURE 6**: Peaks and Weibull fits in the conventional procedure for 3 randomly picked seeds, without noise.*

without noise to noisy input data yields unstable results.

5.5 Conventional method choices

In order to show the improvement of the new approach compared to the conventional experimental method discussed in Section 2.4, results for this method are also provided. No experiments were performed, but it is assumed that they would deliver second-order wave peaks for a number of 1 hour seeds. We selected 3 seeds, and applied 3-parameter Weibull fitting to the top 30% of the resulting HF peaks in each seed. For the case without noise, this results in Figure 6. Each time 3 seeds are randomly picked, the result will be slightly different. This was accounted for by a bootstrap analysis over 200 seed picking realisations. For each set of 3 seeds, the MPM crest height value was calculated from the fits using Equations (20) and (21). The mean and U95 uncertainty (1.96 times standard deviation) over the 200 seed picking realisations were calculated, in order to be compared to the new method and the full MCS results.

5.6 Differences between simplified case study and wave impact problem

Because our ultimate focus is on wave impacts, the LF points may relate to an indicator variable and the HF points to a load variable. This differs from the case study definition of LF and HF, where both are the same variable at different fidelity levels. However, if the indicator LF and load HF statistics are sufficiently similar for screening (see Section 2.1), we can still use them in multi-fidelity approaches. Another difference with the simplified study is that the indicators for the wave impact problem are probably less good, which will complicate the regression. Whether the new method can still be applied will be evaluated in a follow-up (see Section 8).

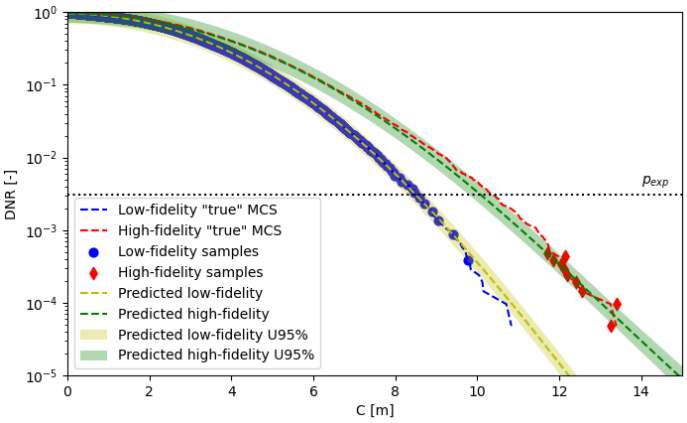
6 RESULTS CASE STUDY

The results for the case study without noise and with noise are discussed in the present section.

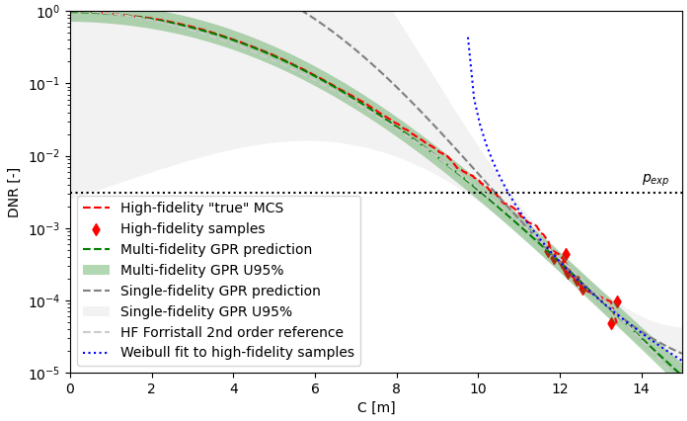
6.1 Results - no noise

In Figure 7 and Figure 8, we present predicted DNR curves without noise for two different sampling options. Each figure includes two plots. The top one shows both the LF and the HF DNR approximation from screening + multi-fidelity GPR (approach 5 from Section 5.3). The bottom one shows the same HF result, the result of screening + single-fidelity GPR (approach 4) and the result of the screening + Weibull-fitting (approach 6). The 1h MPM value for all can be read from the horizontal dotted line (equivalent to Equation (3)), and the results are listed in Table 1. Note that uncertainty is not available for the screening + Weibull approach.

In Figure 7, the MF-GPR result based on the SampMax samples is good, except for a small underestimation of intermediate crest heights. This is also visible in the 1h MPM (Table 1): -2.8% w.r.t. MCS. The single-fidelity GPR performs worse for higher probabilities due to HF training data being centered around maxima, but it still performs well around target probability p_{exp}. The Weibull fit is worse than the single-fidelity GPR for the full domain, but it performs reasonably around p_{exp} and good around the maxima. In Figure 8 for sample option SampPexp around p_{exp}, all three methods show better results. In general, the three methods perform well around the center of gravity of the sampling, but MF-GPR outperforms the single-fidelity GPR and Weibull fitting over

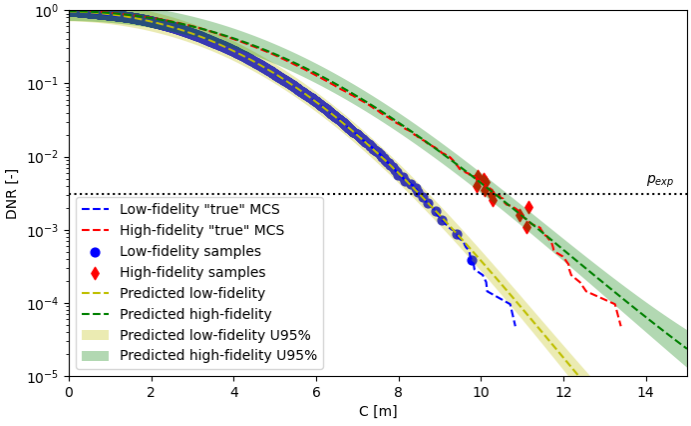


(a) Details screening + multi-fidelity GPR result.

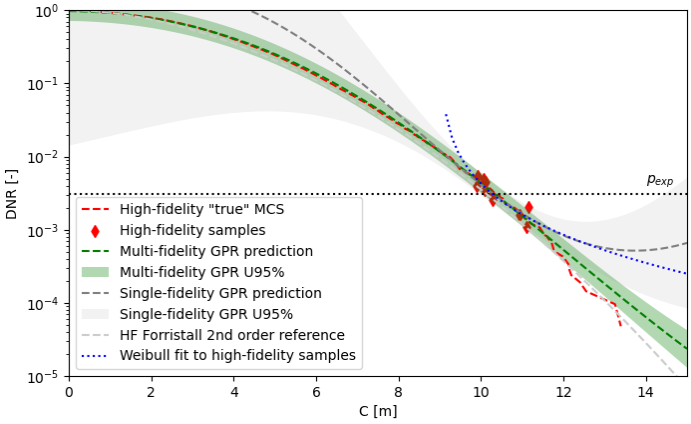


(b) Screening + single- / multi-fidelity GPR / Weibull fit results.

***FIGURE 7**: Results for sample option SampMax, no noise.*



(a) Details screening + multi-fidelity GPR result.



(b) Screening + single- / multi-fidelity GPR / Weibull fit results.

***FIGURE 8**: Results for sample option SampPexp, no noise.*

the full curve. We recommend to focus the sampling around the probability of interest, and adaptive sampling may further improve results (or obtain them with fewer samples). Table 1 also includes the result from the conventional procedure, which performs much worse than the new methods with sampling around p_{exp} (+6.7% MPM deviation versus -0.3 to -0.9%).

If there is no noise in the training data, the size of the MF-GPR uncertainty band mainly depends on the GPR noise settings. The size of the MF-GPR uncertainty band in Figures 7 and 8 therefore does not say much. However, even then, with proper sampling the noise level is lower than that based on the conventional method.

The main advantage of using a multi-fidelity method compared to a conventional brute force approach is the significant reduction of HF simulation or experiment time. Assuming that we need 12 sec CFD simulation time for each selected event (as used in CFD with imposed ship motions from the LF calculation in [11, 41]), we need 120 sec HF simulation time for the 10 selected events in each sea state. We also need an LF MCS, which can be very quick with linear potential flow. In contrast, we need 3 hours HF simulation time per sea state for the conventional approach with 3 seeds and 50 hours for the full HF MCS (see Table 2; note that this table shows the full-scale time to be simulated, not the CPU time). Comparing the two tables shows that the new method provides more accurate MPM results than the conventional method, also with a significant reduction in HF simulation time. The results of the conventional method will be more accurate with more seeds, but this also strongly increases the HF simulation time.

6.2 Results - with noise

Similar results were obtained for the two types of noise defined in Section 5.2. We exclude the detailed MF-GPR plots and only apply SampPexp sampling because the previous section shows that this performs best. Forristall is not included in the plots as it does not account for noise. The required HF simulation time with and without noise is the same for the different methods (see Table 2).

The first noise type, MeasNoise, adds Gaussian measurement noise to the HF signal. Results are shown in Figure 9. Compared to the case without noise, the samples have a larger scatter, indicating that linear wave crests are now a less reliable indicator for second-order wave crests. However, all three new methods are still very close to the MCS result around p_{exp}. MPM values in Table 1

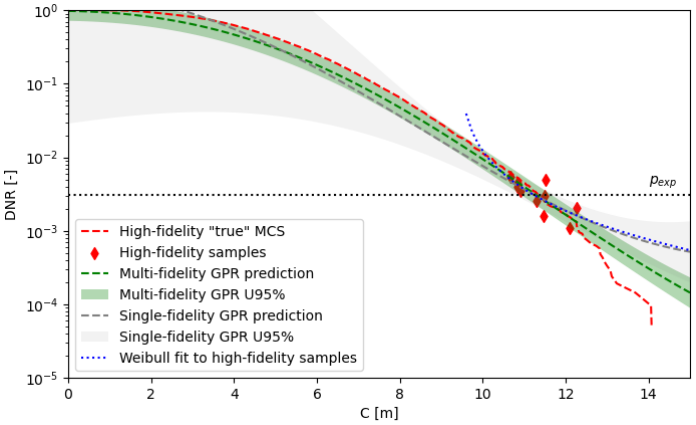


***FIGURE 9**: Results for sample option SampPexp, with MeasNoise. Screening + single- / multi-fidelity GPR / Weibull fit results.*

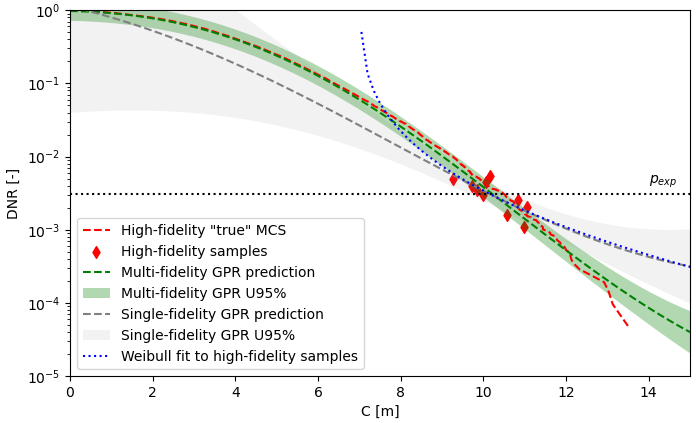


***FIGURE 10**: Results for sample option SampPexp, with PhysNoise. Screening + single- / multi-fidelity GPR / Weibull fit results.*

deviate ranging from -0.3% to -0.8% from the MCS value, showing that these results are much closer to the MCS result than the conventional method (+5.2%), and have lower uncertainty (which can be further reduced by optimizing GPR noise settings).

The second noise type, PhysNoise, adds 'physical noise' such as residual basin waves to the HF signal. Results for this type of noise, sampled using the SampPexp option, are shown in Figure 10. The predictions around p_{exp} are again close to the true MCS result for all three new methods, but slightly less accurate than for the measurement noise case. MPM values in Table 1 deviate ranging from -2.7% to -3.3% from the MCS value for the new methods, which is still better than the +5.8% of the conventional method.

These results were generated with the input GPR noise settings provided in Section 5.4. Results may further improve by optimising the noise hyperparameters instead of prescribing them. Based

***TABLE 1**: HF second-order wave crest height 1-hour MPM, its deviation with respect to the true MCS value and its uncertainty. Result of different approaches. The input noise cases refer to Section 5.2 and the input sample options to Section 5.4.*

Input noise case	**None**	**None**	**Meas**	**Phys**
Sampling	**Max**	**Pexp**	**Pexp**	**Pexp**
1-hour MPM in [m]				
True MCS	10.382	10.382	11.334	10.508
True Forristall	10.145	10.145	-	-
Conventional (3 seeds)	11.075	11.075	11.923	11.118
Screening + MF-GPR	10.088	10.346	11.301	10.228
Screening + GPR	10.428	10.315	11.262	10.174
Screening + Weibull	10.749	10.288	11.241	10.166
Deviation w.r.t. true MCS in [%]				
Conventional (3 seeds)	+6.7	+6.7	+5.2	+5.8
Screening + MF-GPR	-2.8	-0.3	-0.3	-2.7
Screening + GPR	+0.4	-0.6	-0.6	-3.2
Screening + Weibull	+3.5	-0.9	-0.8	-3.3
Uncertainty (U95) of 1-hour MPM [m]				
Conventional (3 seeds)	1.290	1.290	1.255	1.487
Screening + MF-GPR	0.626	0.636	0.640	0.636
Screening + GPR	1.673	0.652	0.651	0.651

***TABLE 2**: Required HF and LF simulation time per sea state for prediction of the 1-hour MPM.*

Method	**HF [h]**	**LF [h]**
True MCS	50	0
Conventional (3 seeds)	3	0
All versions of new method (10 events)	0.033	50

on these results, it can be concluded that some measurement noise in the HF data or a residual extra wave do not matter much for the MPM prediction (the correct HF MPM including noise is predicted). The methods are not very sensitive to noise, as long as the HF samples are focused around the probability of interest.

Finally, we evaluate the application of the new methods at the abstraction level of the DNR. The present results are good. However, with focused sampling around p_{exp}, the HF samples are close together on the probability axis. Especially when there is a lot of noise in the HF data or the indicator is less good, the samples will form a cloud in the DNR plot, which makes regression difficult. Assuming that all the HF samples have exactly probability p_{exp} in the DNR, assembling them in a DEM and then applying the new methods to this DEM could therefore improve the results. This

would shift analysis to another level of abstraction and probably make the regression more stable, especially for more non-linear or noisy problems.

7 CONCLUSIONS

It can be concluded that all three versions of the new method (screening + MF-GPR, screening + GPR, and screening + Weibull-fitting) perform well for the simplified case study, where MF-GPR performs the best overall. Weibull fitting and single-fidelity GPR can introduce larger deviations if samples are not well chosen. Focusing the HF sampling around the exceedance probability of interest is advisable. The new method is rather insensitive to noise as long as HF samples are selected properly. The application to the DNR exceedance distribution works, and application to the DEM may further improve results. Most importantly, the new method produces more accurate MPM results than the conventional method while also significantly reducing HF simulation time. The new method has a deviation of only 0.3-3.5% from the true 1-hour MPM over all test cases, compared to the conventional method's deviation of 5.2-6.7%. The HF simulation time required to do this is 91 times shorter with the new method (0.033 versus 3 hours per sea state).

8 FURTHER WORK

Future work on this project is foreseen in several areas. Firstly, we plan to include adaptive sampling, which is expected to result in more accurate results with fewer HF samples. Secondly, we plan an application to long duration wave impact data, to validate the new method for predicting wave impact design loads. There will be more scatter in the HF data compared to the present case study, and the LF indicator and HF response will be different variables. It needs to be evaluated whether the assumptions in Section 4.3 are still sufficiently valid in that case. Thirdly, we will evaluate the application to the DEM instead of the DNR. It is expected that this makes the regression more stable for highly non-linear problems, but it needs to be evaluated whether this is necessary for the wave impact problem. Fourthly, we will investigate suitable (MF-)GPR settings (noise parameters, kernels, optimisation algorithm for the hyperparameters, etc.) in more detail. The noise parameters can be learned instead of imposed. The mean of the GPR can be non-constant and based on prior knowledge, such as a fitted extreme value distribution. It could be interesting to compare this approach with the multi-fidelity GPR. Also, the present MF-GPR model is linear in y_l, which puts a lot of modelling burden on $\delta(x)$. In future work, it may therefore be considered to change Equation (13) to e.g., $\mathbf{y_h}(\mathbf{x}) = f(\mathbf{x}, \mathbf{y_l}(\mathbf{x}))$. This may improve results, especially for cases with a large scale factor between the LF and HF data (such as indicator and wave impact values in a wave impact study). Finally, when it would be concluded that the method works for wave impact problems, it may also work for other non-linear problems inside or outside the maritime field. Generalisation of the method for more dimensions and other applications could be evaluated.

ACKNOWLEDGEMENTS

This publication is part of the project "Multi-fidelity Probabilistic Design Framework for Complex Marine Structures" (project number TWM.BL.019.007) of the research programme "Topsector Water & Maritime: the Blue route" which is (partly) financed by the Dutch Research Council (NWO).

A SINGLE- AND MULTI-FIDELITY GPR DETAILS

A.1 Single-fidelity GPR

GPR or Kriging [15, 42] is a method for estimating new dependent variable values $\mathbf{y}^*$ (and their uncertainty) at new input variable values $\mathbf{x}^*$, using a data set of n existing observations $\mathbf{y_{obs}} = (y_1, y_2, \ldots, y_n)$ at $\mathbf{x_{obs}} = (x_1, x_2, \ldots, x_n)$. The underlying process $f(x)$ is assumed to have an infinite-dimensional multivariate Gaussian distribution with a mean function, $\mu(x)$, and covariance function or kernel, $k(x,x')$. Prior knowledge can be incorporated using these functions, and the mean is often set to the mean of the training data. This can already be done when defining $\mathbf{x_{obs}}$ and $\mathbf{y_{obs}}$, such that the mean in the subsequent formulations can be set to zero. Gaussian noise $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$ can be included by adding it to the right-hand side of Equation (4). GPR roughly has the following steps:

$$f(x) \sim \mathcal{G}\left(\mu(x), k(x,x')\right) \tag{4}$$

1. Select a kernel, assume *a priori* hyperparameters.
2. Optimise hyperparameters based on available data.
3. Use the optimised GPR (with *a posteriori* hyperparameters) to estimate the process values $\mathbf{y}^*$ at $\mathbf{x}^*$.

The kernel is a covariance function that describes the relation between x-points. We use the squared exponential kernel (see Equation (5)), which includes length parameter l and maximum allowable covariance parameter σ. The kernel can also include a noise term with an additional hyperparameter σ_n. If two points x and x' are far apart, their covariance is zero, and the length parameter determines how quickly this interaction decreases. Kernels have sets of hyperparameters that are optimised in the GPR process; for the squared exponential, this is $\theta = (l)$ or $\theta = (l, \sigma_n)$ with noise. σ is usually set to the standard deviation of the data set instead of being optimised.

$$k(x,x') = \sigma^2 \exp\left[\frac{-(x-x')^2}{2l^2}\right] \tag{5}$$

The GPR needs the value of the kernel for interaction of each point with all the other points. This results in the interaction matrix in Equation (6).

$$\mathbf{K}(\mathbf{x},\mathbf{x}') = \begin{bmatrix} k(x_1,x_1)\, k(x_1,x_2) & \dots & k(x_1,x_n) \\ k(x_2,x_1)\, k(x_2,x_2) & \dots & k(x_2,x_n) \\ \vdots & \vdots & \vdots \\ k(x_n,x_1)\, k(x_n,x_2) & \dots & k(x_n,x_n) \end{bmatrix} \tag{6}$$

To predict new values based on existing observations, we need a total interaction matrix that accounts for interactions between existing points, existing points and new points, and new points with themselves. This results in covariance matrix $\mathbf{C}$ in Equation (7).

$$\mathbf{C} = \begin{bmatrix} \mathbf{K}(\mathbf{x},\mathbf{x}') & \mathbf{K}(\mathbf{x}^*,\mathbf{x}') \\ \mathbf{K}(\mathbf{x},\mathbf{x}^{*\prime}) & \mathbf{K}(\mathbf{x}^*,\mathbf{x}^{*\prime}) \end{bmatrix} = \begin{bmatrix} \mathbf{K_e} & \mathbf{K}^{*\mathbf{T}} \\ \mathbf{K}^* & \mathbf{K}^{**} \end{bmatrix} \tag{7}$$

When the mean is subtracted from the observations in $\mathbf{x_{obs}}$ and $\mathbf{y_{obs}}$, this reduces the GPR process to Equation (8).

$$\begin{bmatrix} \mathbf{y_{obs}} \\ \mathbf{y}^* \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}) \tag{8}$$

Now Equation (9) provides the conditional probability of new values based on existing observations (for a detailed derivation see [42]). This equation yields the best estimate for new values $\overline{\mathbf{y}^*}$ and their variance $\mathrm{var}(\mathbf{y}^*)$. The 95% confidence interval can be calculated from this variance using $\mathbf{U95} = 1.96 \cdot \mathrm{var}(\mathbf{y}^*)$.

$$\begin{aligned} \mathbf{y}^*|\mathbf{y_{obs}} &\sim \mathcal{N}\left(\mathbf{K}^*{\mathbf{K_e}}^{-1}\mathbf{y_{obs}}, \mathbf{K}^{**} - \mathbf{K}^*{\mathbf{K_e}}^{-1}\mathbf{K}^{*\mathbf{T}}\right) \\ \overline{\mathbf{y}^*} &= \mathbf{K}^*{\mathbf{K_e}}^{-1}\mathbf{y_{obs}} \\ \mathrm{var}(\mathbf{y}^*) &= \mathbf{K}^{**} - \mathbf{K}^*{\mathbf{K_e}}^{-1}\mathbf{K}^{*\mathbf{T}} \end{aligned} \tag{9}$$

To find these new values, we need the values of the kernel hyperparameters. In step one of the GPR, we assume arbitrary *a priori* values, and then maximise the conditional probability $p(\theta|\mathbf{x_{obs}}, \mathbf{y_{obs}})$ to optimise them. According to Bayes' theorem, this is equivalent to maximising the log-likelihood given in Equation (10), where n is the number of existing observations (see e.g., [42]). We can find the maximum of this formulation using a multi-variate optimisation algorithm to solve for the hyperparameter set θ. Assuming a uniform prior distribution $\mathscr{U}(\theta_{min}, \theta_{max})$ gives the best parameters θ.

$$\begin{aligned} &\log(p(\mathbf{y_{obs}}|\mathbf{x_{obs}}, \theta)) = \\ &-\frac{1}{2}{\mathbf{y_{obs}}}^T\mathbf{C}^{-1}\mathbf{y_{obs}} - \frac{1}{2}\log|\mathbf{C}| - \frac{n}{2}\log(2\pi) \end{aligned} \tag{10}$$

Once this *a posteriori* hyperparameter set θ is found, Equation (9) can be used to obtain the new values and their uncertainty. As we subtracted the mean of the data set at the start of the procedure, it has to be added again at the last step.

A.2 Multi-fidelity GPR

In MF-GPR or co-kriging [16] we have existing LF and HF observations $(\mathbf{x_{l,obs}}, \mathbf{y_{l,obs}})$ and $(\mathbf{x_{h,obs}}, \mathbf{y_{h,obs}})$, respectively. These are organised in Equation (11). We also have high-fidelity new points $\mathbf{x}^*$ and the corresponding new values to predict $\mathbf{y}^*$.

$$\mathbf{x_{m,obs}} = \begin{bmatrix} \mathbf{x_{l,obs}} \\ \mathbf{x_{h,obs}} \end{bmatrix} \quad \mathbf{y_{m,obs}} = \begin{bmatrix} \mathbf{y_{l,obs}} \\ \mathbf{y_{h,obs}} \end{bmatrix} \tag{11}$$

We use two levels of fidelity, as in [17]. Certain assumptions must hold in order to efficiently apply MF-GPR: the different fidelity levels must be correlated in some way (necessary for any multi-fidelity model), both the LF and HF data sets should have some degree of smoothness and prior beliefs about each level of the code can be represented using a Gaussian process. Satisfying the second assumption may be hard for some types of non-linear ship responses; wave impact loads for instance can be very noisy and may have different load regimes. However, the extent to which MF-GPR can be applied can be assessed.

We use the autoregressive model of [16]. This uses the assumption in Equation (12), which is a kind of Markov property that enables decomposition of the multi-fidelity GPR problem into independent single-fidelity GPR problems. This speeds up the computation. The assumption can be interpreted as follows: given the nearest point $y_l(x)$, we can learn no more about $y_h(x)$ from any other point $y_l(x')$ for $x' \neq x$.

$$k\left(y_h(x), y_l(x')|y_l(x)\right) = 0 \tag{12}$$

The formulation commonly used for MF-GPR, including in the present paper, is expressed in Equation (13). See Section 8 for a comment on alternatives. Here, the multi-fidelity problem is split into two GPRs: one for the LF data and one for the difference function $\delta(x)$ between the LF and HF data. The parameter ρ is used to establish the correlation between the two datasets. The present process has the following steps:

$$\mathbf{y_h}(\mathbf{x}) = \rho\mathbf{y_l}(\mathbf{x}) + \delta(\mathbf{x}) \tag{13}$$

1. Select a kernel, assume its *a priori* hyperparameters θ_l for 'ordinary' single-fidelity GPR of LF function.
2. Optimise these LF hyperparameters using available LF data.
3. Use optimised single-fidelity LF GPR (based on *a posteriori* hyperparameters θ_l) to estimate the LF values at $\mathbf{x}^*$.

4. Select a kernel, assume its *a priori* hyperparameters θ_δ for multi-fidelity GPR of the difference function $\delta(x)$.
5. Optimise these multi-fidelity hyperparameters using available HF data and evaluated LF GPR values at $\mathbf{x}^*$.
6. Use optimised multi-fidelity GPR (based on *a posteriori* hyperparameters θ_l and θ_δ) to estimate the HF values over $\mathbf{x}^*$.

Steps 1 to 3 are identical to the single-fidelity procedure described in Appendix A.1 applied to the LF data. In steps 4-5, similar hyperparameters are optimised for the difference function, plus ρ from Equation (13). In steps 4 to 6, the multi-fidelity GPR formulations differ from the single-fidelity formulations. The multi-fidelity covariance matrix of the existing observations $\mathbf{D_e}$ is formulated in Equation (14), which uses Equation (6). Similarly, the multi-fidelity covariance matrix of the existing observations with new points $\mathbf{D}^*$ is given in Equation (15) and the multi-fidelity covariance matrix for the new points $\mathbf{D}^{**}$ in Equation (16).

$$\begin{aligned}
\mathbf{D_{e,11}} &= \mathbf{K}(\mathbf{x_l}, \mathbf{x_l}'|\theta_\mathbf{l}) \\
\mathbf{D_{e,21}} &= \rho\mathbf{K}(\mathbf{x_h}, \mathbf{x_l}'|\theta_\mathbf{l}) \\
\mathbf{D_{e,22}} &= \rho^2\mathbf{K}(\mathbf{x_h}, \mathbf{x_h}'|\theta_\mathbf{l}) + \mathbf{K}(\mathbf{x_h}, \mathbf{x_h}'|\theta_\delta) \\
\mathbf{D_e} &= \begin{bmatrix} \mathbf{D_{e,11}} & \mathbf{D_{e,21}}^T \\ \mathbf{D_{e,21}} & \mathbf{D_{e,22}} \end{bmatrix}
\end{aligned} \tag{14}$$

$$\begin{aligned}
\mathbf{D^*_{11}} &= \rho\mathbf{K}(\mathbf{x}^*, \mathbf{x_l}'|\theta_\mathbf{l}) \\
\mathbf{D^*_{12}} &= \rho^2\mathbf{K}(\mathbf{x}^*, \mathbf{x_h}'|\theta_\mathbf{l}) + \mathbf{K}(\mathbf{x}^*, \mathbf{x_h}'|\theta_\delta) \\
\mathbf{D}^* &= \begin{bmatrix} \mathbf{D^*_{11}} & \mathbf{D^*_{12}} \end{bmatrix}
\end{aligned} \tag{15}$$

$$\mathbf{D}^{**} = \rho^2\mathbf{K}(\mathbf{x}^*, \mathbf{x}^{*\prime}|\theta_\mathbf{l}) + \mathbf{K}(\mathbf{x}^*, \mathbf{x}^{*\prime}|\theta_\delta) \tag{16}$$

Now we can assemble the multi-fidelity form of the full covariance matrix (Equation (17)) and of the log-likelihood function (Equation (18)), where n_m is the number of elements in $\mathbf{x_{m,obs}}$.

$$\mathbf{D} = \begin{bmatrix} \mathbf{D_e} & \mathbf{D}^{*\mathbf{T}} \\ \mathbf{D}^* & \mathbf{D}^{**} \end{bmatrix} \tag{17}$$

$$\begin{aligned}
&\log(p(\mathbf{y_{m,obs}}|\mathbf{x_{m,obs}}, \theta_l, \theta_\delta)) = \\
&-\frac{1}{2}\mathbf{y_{m,obs}}^T\mathbf{D}^{-1}\mathbf{y_{m,obs}} - \frac{1}{2}\log|\mathbf{D}| - \frac{n_m}{2}\log(2\pi)
\end{aligned} \tag{18}$$

Finally, the new predicted values and their variance can be calculated using Equation (19). Here, the constant mean of the data set is handled in the same way as in the single-fidelity GPR: subtracted from the observations in step 1, and added again after the last step.

$$\begin{aligned}
\overline{\mathbf{y}^*} &= \mathbf{D}^*{\mathbf{D_e}}^{-1}\mathbf{y_{m,obs}} \\
\mathrm{var}(\mathbf{y}^*) &= \mathbf{D}^{**} - \mathbf{D}^*{\mathbf{D_e}}^{-1}\mathbf{D}^{*\mathbf{T}}
\end{aligned} \tag{19}$$

B TRUE MPM DERIVED FROM DEM

The DEM is defined in Equation (20), with $S_E(N)$ the ensemble maximum response values of N realisations and probability $1/N$ for the maximum value. The MPM from the DEM $\widehat{S_E}(N)$ is defined in Equation (21). The estimated value of 0.632 comes from the exceedance probability of the MPM in a DEM distribution for a Gaussian signal as discussed in, e.g., [43].

$$DEM(s) = p\left(S_E(N) \geq s\right) \tag{20}$$

$$\begin{aligned}
DEM\left(\widehat{S_E}(N)\right) &= p\left(S_E(N) \geq \widehat{S_E}(N)\right) = 0.632 \\
\widehat{S_E}(N) &= \phi_{0.632}\left[S_E(N)\right]
\end{aligned} \tag{21}$$

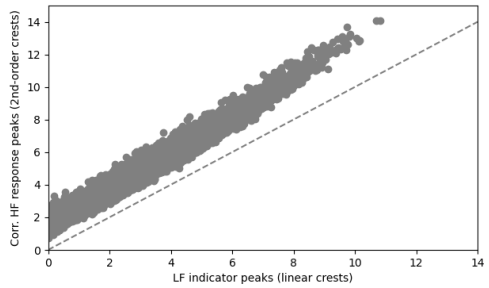
For the HF second-order wave time series in the case study of Section 5 without noise, the 1-hour MPM estimated from the DNR is 10.382 m (see Table 1). The MPM estimated from the DEM using the formulations above for the same case and the same 50 realisations is 10.530 m. The difference between these two values is of the same order as the differences between the MPM from the different approaches in Section 6. This difference is probably caused by wave grouping; individual second-order wave crests are not fully independent. The ensemble maxima in the DEM are fully de-coupled (one 'event' per realisation), whereas the crests in the DNR are not (many 'events' per realisation). It is expected that the estimates are closer for more rare HF phenomena such as wave impact loads (as events are more independent, even when all events are considered).

C INPUT PLOTS WITH NOISE

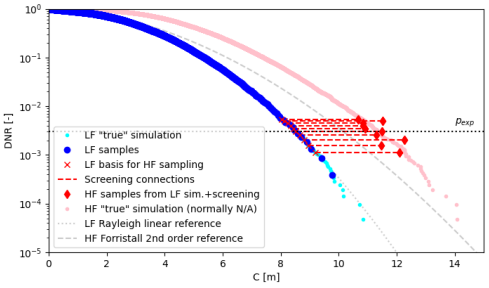
This appendix contains similar plots of the scatter of matched LF versus HF peaks as Figure 4 and true and training DNR data as Figure 5, but here for the two input versions with noise. Figure 11 shows these plots for the option MeasNoise and Figure 12 for the option PhysNoise; both for sampling option SampPexp.

REFERENCES

[1] M. Halsne, O. N., G. Ersdal, M. Langøy, T. Andersen, L. G. Bjørheim, Semisubmersible in service experiences on the Norwegian Continental Shelf, in: 41th OMAE Conf., ASME, Hamburg, Germany, 2022. doi:10.1115/OMAE2022-81289.

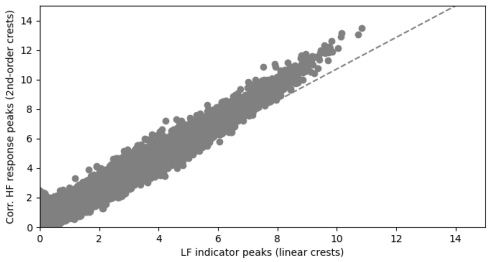


(a) Scatter plot of matched peaks.

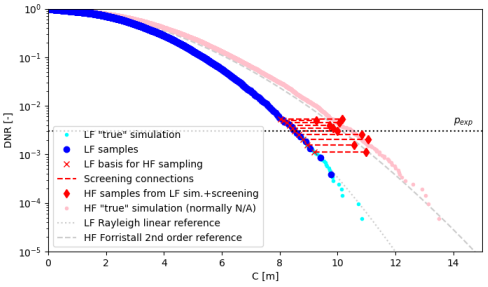


(b) DNR of the true and training data.

***FIGURE 11**: Input figures for noise option MeasNoise, sampling option SampPexp.*



(a) Scatter plot of matched peaks.



(b) DNR of the true and training data.

***FIGURE 12**: Input figures for noise option PhysNoise, sampling option SampPexp.*

[2] G. K. Kapsenberg, On the slamming of ships, PhD thesis, Delft University of Technology, Delft, The Netherlands (2018). doi:10.4233/uuid:14eac2bb-63ee-47e4-8218-1ba3830a97b4.

[3] I. Viste-Ollestad, T. L. Andersen, N. Oma, S. Zachariassen, Investigation report Petroleumstilsynet - Investigation of an accident with fatal consequences on COSLInnovator, 30 December 2015, Tech. rep. (2016).

[4] R. Dallinga, G. Gaillarde, Hatch cover loads experienced by M.V. Derbyshire during typhoon 'Orchid', in: Glasgow Marine Fair And Int. Workshop On Safety Of Bulk Carriers, Glasgow, Scotland, UK, 2001.

[5] G. Ersdal, A. Kvitrud, Green water on Norwegian production ships, in: 10th ISOPE Conf., Int. Soc. of Offshore and Polar Eng. (ISOPE), Seattle, USA, 2000.

[6] Panama Maritime Authority, Dutch Safety Board, Bundesstelle fur Seeunfalluntersuchung, Loss of containers overboard from MSC Zoe on 1-2 January 2019, Tech. rep. (2020).

[7] S. M. van Essen, H. C. Seyffert, Finding dangerous waves – Review of methods to obtain wave impact design loads for marine structures, J. Offshore Mechanics and Arctic Eng. (OMAE-22-1110) (2023). doi:10.1115/1.4056888.

[8] T. H. J. Bunnik, C. T. Stansberg, C. Pákozdi, S. Fouques, L. Somers, Useful indicators for screening of sea states for wave impacts on fixed and floating platforms, in: 37th OMAE Conf., ASME, Madrid, Spain, 2018. doi:10.1115/OMAE2018-78544.

[9] T. H. J. Bunnik, J. Scharnke, E. J. de Ridder, Efficient indicators for screening of random waves for wave impacts on a jacket platform and a fixed offshore wind turbine, in: 38th OMAE Conf., ASME, Glasgow, UK, 2019. doi:10.1115/OMAE2019-95481.

[10] C. T. Stansberg, Wave Front Steepness and Influence on Horizontal Deck Impact Loads, J. Mar. Sci. Eng. 8, 314 (2020). doi:10.3390/jmse8050314.

[11] S. M. van Essen, C. Monroy, Z. Shen, J. A. Helder, D.-H. Kim, S. Seng, Z. Ge, Screening wave conditions for the occurrence of green water events on sailing ships, Ocean Eng. 234 (109218) (2021). doi:10.1016/j.oceaneng.2021.109218.

[12] O. Gramstad, C. Agrell, E. Bitner-Gregersen, B. Guo, E. Ruth, E. Vanem, Sequential sampling method using Gaussian process regression for estimating extreme structural response, Mar. Struct. 72 (2020). doi:10.1016/j.marstruc.2020.102780.

[13] S. Guth, T. Sapsis, Wave episode based Gaussian process regression for extreme event statistics in ship dynamics: Between the Scylla of Karhunen–Loeve convergence and the Charybdis of transient features, Ocean Eng. 266 (112633) (2022). doi:10.1016/j.oceaneng.2022.112633.

[14] S. Guth, B. Champenois, T. P. Sapsis, Application of Gaussian process multi-fidelity optimal sampling to ship structural modeling, in: 34th Symp. Naval Hydrodynamics, Washington DC, USA, 2022.

[15] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, MIT Press, Massachusetts Institute of Technology, 2006.

[16] M. C. Kennedy, A. O'Hagan, Predicting the output from a

complex computer code when fast approximations are available, Biometrika 87 (1) (2000) 1–13.

[17] I. J. Forrester, A. Sobester, A. J. Keane, Multi-fidelity optimization via surrogate modelling, Proc. R. Soc. A 463 (2007) 3251–3269. doi:10.1098/rspa.2007.1900.

[18] F. Grassi, G. Manganini, M. Garraffa, L. Mainini, Resource aware multifidelity active learning for efficient optimization, in: AIAA SciTech Forum, 2021. doi:10.2514/6.2021-0894.

[19] D. Huang, T. T. Allen, W. I. Notz, R. A. Miller, Sequential kriging optimization using multiple-fidelity evaluations, Structural and Multidisciplinary Optimization 32 (5) (2006) 369–382. doi:10.1007/s00158-005-0587-0.

[20] J. H. S. de Baar, Z. Leylek, A. J. Neely, Efficient uncertainty quantification for an axial compressor, using adaptive multi-fidelity kriging, in: 23rd International Symposium on Air Breathing Engines (ISABE 2017): Economy, Efficiency and Environment, Vol. 2, 2017.

[21] Z. Hu, S. Mahadevan, A single-loop kriging surrogate modeling for time-dependent reliability analysis, J. Mechanical Design 138 (061406) (2021). doi:10.1115/1.4033428.

[22] S. R. Winterstein, T. C. Ude, C. A. Cornell, P. Bjerager, S. Haver, Environmental parameters for extreme response: inverse FORM with omission factors, in: Int. Conf. Struct. Saf. Reliab., Innsbruck, Austria, 1993.

[23] S. Haver, S. R. Winterstein, Environmental contour lines: a method for estimating long term extremes by a short term analysis, in: SNAME Annu. Meet., Society of Naval Architects and Marine Engineers, Houston, USA, 2008. doi:10.5957/SMC-2008-067.

[24] ABS, Guidance notes on air gap and wave impact analysis for semisubmersibles, American Bureau of Shipping, Spring, USA, 2020.

[25] DNV-GL, Offshore Technical Guidance, Horizontal wave impact loads for column stabilised units, DNVGL-OTG-14, Det Norske Veritas, Oslo, Norway, 2019.

[26] BV, Rule Note NR583: Whipping and springing assessment, Bureau Veritas, Paris, France, 2015.

[27] A. Naess, O. Gaidai, P. Teigen, Extreme response prediction for nonlinear floating offshore structures by monte carlo simulation, Applied Ocean Research 29 (4) (2007) 221–230. doi:10.1016/j.apor.2007.12.001.

[28] DNV, Det Norske Veritas, Class Guideline DNVGL-CG-0130: Wave loads, Det Norske Veritas, Oslo, Norway, 2018.

[29] G. Lian, Slamming loads on large volume structures from breaking waves, Phd thesis, University of Stavanger, Stavanger, Norway (2018).

[30] ITTC, Recommended Practice 7.5-02-07-02.3: Experiments on rarely occurring events, International Towing Tank Conference, 2017.

[31] NORDFORSK, Assessment of ship performance in a seaway: results of a Nordic co-operative project on seakeeping performance of ships, Nordic Co-operative Organisation For Applied Research c/o Technoconsult, Copenhagen, Denmark, 1987.

[32] ITTC, Recommended Practice 7.5-02-07-02.6: Global loads seakeeping procedure, International Towing Tank Conference, 2017.

[33] S. Haver, Metocean modelling and prediction of extremes, no. May, Haver & havet, University in Stavanger, NTNU, Stavanger, Norway, 2017.

[34] F. Mauro, L. Braidotti, U. la Monaca, R. Nabergoj, Extreme loads determination on complex slender structures, Int. Shipbuilding Progress 66 (2019) 57–76. doi:10.3233/ISP-180256.

[35] S. M. van Essen, J. Scharnke, H. C. Seyffert, Required test durations for converged short-term wave and impact extreme value statistics - Part 1: ferry dataset, Marine Structures (103410) (2023). doi:10.1016/j.marstruc.2023.103410.

[36] J. Scharnke, S. M. van Essen, H. C. Seyffert, Required test durations for converged short-term wave and impact extreme value statistics - Part 2: deck box dataset, Marine Structures (103411) (2023). doi:10.1016/j.marstruc.2023.103411.

[37] G. Z. Forristall, Wave crest distributions: observations and second-order theory, Phys. Oceanogr. 30 (8) (2000) 1931–1943. doi:10.1175/1520-0485(2000)030<1931:WCDOAS>2.0.CO;2.

[38] J. N. Sharma, R. G. Dean, Development and evaluation of a procedure for simulating a random directional second order sea surface and associated wave forces, Ocean Eng. Rep. 20 (1979).

[39] A. Paleyes, M. Pullin, M. Mahsereci, N. Lawrence, J. González, Emulation of physical processes with emukit, in: 2nd Workshop on Machine Learning and the Physical Sciences, NeurIPS, Vancouver, Canada, 2019. doi:10.48550/arXiv.2110.13293.

[40] GPy, GPy: A gaussian process framework in python, http://github.com/SheffieldML/GPy (since 2012).

[41] H. J. A. Bandringa, H., S. M. van Essen, On the validity of CFD for simulating extreme green water loads on ocean-going vessels, in: 39th Int. Conf. Ocean Offshore Arct. Eng. (OMAE), ASME, Virtual, Online, 2020. doi:10.1115/OMAE2020-18290.

[42] Ebden, M., Gaussian Processes for Regression: a quick introduction, https://arxiv.org/abs/1505.02965v2 (2008).

[43] M. Ochi, Applied Probability and Stochastic Processes in Engineering and Physical Sciences, John Wiley & Sons, Singapore, 1990.