On extending multifidelity uncertainty quantification methods from non-rare to rare problems

Brendan Brown, *University of North Carolina at Chapel Hill*
Vladas Pipiras, *University of North Carolina at Chapel Hill*

ABSTRACT

When modeling a random phenomenon (e.g. ship motions in irregular seas), data are often available from multiple sources, or models, of varying fidelity, those with higher fidelity carrying higher costs. Multifidelity uncertainty quantification (UQ) offers tools that allow using lower-fidelity and lower-cost models to inform decisions being made about high-fidelity models. With a few exceptions though, much of the focus of the multifidelity UQ literature has been on characterizing uncertainty related to averages, in the context of non-rare problems where data are available to estimate these averages directly. In this work, we extend some multifidelity UQ methods to estimation of probabilities of rare events, possibly those that have not been observed in high-fidelity data. The suggested approach is based on bivariate extreme value theory, applied to simultaneously large observations from low-fidelity and high-fidelity models. The ideas are illustrated on simulated data associated with ship motions. It is not assumed that the reader is familiar with multifidelity UQ, with the discussion focusing on the most basic setting and building naturally from the recalled methods for non-rare problems.

Keywords: *Uncertainty Quantification, Multifidelity Estimators, Bivariate Extremes, Sampling Variability, Probability of Rare Event, Nonlinear Random Oscillator, Ship Motions.*

1. INTRODUCTION

When studying random phenomena of interest, it is common to examine data from multiple sources or models. For example, ship motions or loads data can be collected from a model basin or sea trials, or generated from various computer programs, e.g. SimpleCode (Weems and Wunrow, 2013), LAMP (Lin and Yue, 1991). With data at hand, a common goal is to estimate some quantities of interest, for example, mean, single significance amplitude (SSA), etc. In this case, how should different estimates of the same quantities of interest obtained across multiple models be interpreted? If one of the models is less "expensive" to run but less accurate, how can it be used in conjunction with the more expensive and more accurate models in order to estimate better the quantities of interest? What does this say about differences among the models?

These questions have been studied from various angles as part of the Uncertainty Quantification (UQ) literature, in particular, in the direction of the so-called multifidelity (MF) methods. See, for example, a recent survey paper by Peherstorfer *et al.* (2018). As above, at the most basic level, the underlying assumption of these methods is the availability of two sets of data, one associated with the variable X_e and the other with the variable X_s, referring to expensive (true, high-fidelity, etc.) model and simple (low-fidelity, surrogate, etc.) model, respectively. (We shall use the terms and subscripts for "expensive" and "simple" throughout this work, in lieu of perhaps more sophisticated "high-fidelity" and "low-fidelity.") For example, X_e could refer to CFD and X_s to LAMP calculations. The interest is in estimating the mean (or the expected value) $\mathbb{E}(X_e)$ of the expensive response, or some function thereof, having the data from both expensive and simple models. Construction and calibration of simple models also make an important part of MF methods, but these will not be our focus here. That is, we suppose that data on X_e and X_s are given and ask questions about implications of this setting.

Estimation of the mean through available MF methods concerns non-rare behavior of the studied random phenomenon in that there is enough variability in collected data to make an informed decision about behavior of the mean. In this work, we are interested in analogous MF methods but for

rare problems. A working example throughout this paper is that of estimating an exceedance probability $\mathbb{P}(X_e > x)$ for some large threshold x, so that $X_e > x$ is a rare event. The latter event might be so rare that it is not even observed in the data from the expensive model. It should be noted nevertheless that in the latter case, the rare probability could, in principle, still be estimated through the approach of the statistical Extreme Value Theory (EVT); see e.g. Coles (2001). This approach for extreme ship motions, capsizing and other rare phenomena has been studied quite extensively by the second author of this work and collaborators over the past number of years (e.g. Campbell *et al.*, 2016; Belenky *et al.*, 2018, Belenky *et al.*, 2019).

In the context of estimating a rare exceedance probability $\mathbb{P}(X_e > x)$, we are thus interested in whether and how the data for the variable X_s from the simple model might be useful. For example, since the simple model is thought to be inexpensive to run, the events $X_s > x$ could, in principle, be observed in really long records of the model. Then, could one use the direct estimate of the probability $\mathbb{P}(X_s > x)$ for that of $\mathbb{P}(X_e > x)$? These are the kind of questions that will be discussed in this work, within an introduced mathematically justified framework.

We are not aware of other works pursuing this exact line of investigation. The closest are perhaps MF methods for failure probabilities as in e.g. Peherstorfer *et al.* (2017). These failure probabilities though are still estimated directly, perhaps in conjunction with importance sampling, whereas in this work, we do so indirectly through EVT. For this reason, the reader should also expect our approach to appear more complex, especially to those unfamiliar with EVT.

The MF methods discussed in this work will be illustrated on synthetic data generated from a non-linear random oscillator mimicking ship rolling. It should be noted that the synthetic data framework is for illustration purposes only; there is nothing more expensive or simpler about either model in the synthesis.

The rest of this work is organized as follows. Section 2 sets some notation and introduces the more probabilistic notions used throughout this work. In Sections 3 and 4, we discuss and illustrate the most basic available MF estimator when making inference about the expected value (mean) in non-rare problems. Sections 5 and 6 extend this MF approach to estimating probabilities of rare events. Basic bivariate EVT is recalled and employed in developing the approach in Section 5. Section 7 concludes.

2. BASIC SETTING AND NOTATION

At the most basic level, we assume the following setting. We observe two signals: $X_e(t)$, $t \in [0, T_e]$, from an expensive, true, high- (or maybe engineering-level) fidelity model, and $X_s(t)$, $t \in [0, T_s]$, from simple, surrogate, low-fidelity model. Again, the terms "expensive" and "simple" will be used exclusively below. The observation window sizes T_e and T_s are such that $T_e \ll T_s$, reflecting the idea that the simple model can be run for a much longer period of time at low cost, though the exact costs will be mostly ignored here. More importantly, we assume that the expensive and simple models are run under the same "conditions" in that the error process

$$\epsilon(t) = X_e(t) - X_s(t), \; t \in [0, T_e], \tag{1}$$

is meaningful over the smaller observation window $[0, T_e]$.

Furthermore, the following notation will be used: $\mu(Z) = \mathbb{E}(Z)$, $\sigma^2(Z) = Var(Z)$ will stand for a theoretical mean and variance, respectively, of a variable Z or a stationary process $Z(t)$; $\hat{\mu}(Z)$, $\hat{\sigma}^2(Z)$ will denote statistical estimators of the latter quantities from data; $\bar{Z}_T$ will refer to the sample average of $Z(t)$ over time interval $[0, T]$. The hats used for other quantities will also refer to estimators. For example, $\widehat{\mathbb{P}}$ will refer to a probability estimate.

3. METHODS FOR NON-RARE PROBLEMS

In the setting described in Section 2, suppose that one is interested in estimating the mean $\mu(X_e)$ of X_e. A multifidelity (MF) estimator of the mean is defined as

$$\hat{\mu}_{mf}(X_e) = \bar{X}_{s,T_s} + \bar{\epsilon}_{T_e} \tag{2}$$

or, equivalently, as

$$\begin{aligned} \hat{\mu}_{mf}(X_e) &= \bar{X}_{s,T_s} + \left(\bar{X}_{e,T_e} - \bar{X}_{s,T_e}\right) \\ &= \bar{X}_{e,T_e} + \left(\bar{X}_{s,T_s} - \bar{X}_{s,T_e}\right), \end{aligned} \tag{3}$$

where the last expression is a simple rearrangement of the previous one. For comparison, let also

$$\hat{\mu}_0(X_e) = \bar{X}_{e,T_e} \tag{4}$$

be the baseline estimator of the mean that uses only the expensive data.

We make several observations that might be useful to readers unfamiliar with MF estimators. Note that $\hat{\mu}_{mf}(X_e)$ is unbiased for $\mu(X_e)$ even if $\mu(X_e) \neq \mu(X_s)$. This follows from Eq. (3) since

$$\mathbb{E}\hat{\mu}_{mf}(X_e) = \mathbb{E}\bar{X}_{s,T_s} + \left(\mathbb{E}\bar{X}_{e,T_e} - \mathbb{E}\bar{X}_{s,T_e}\right) = \mu(X_s) + \left(\mu(X_e) - \mu(X_s)\right) = \mu(X_e). \quad (5)$$

Another key observation and the crux of MF methods is that $\hat{\mu}_{mf}(X_e)$ can potentially do better in estimating the mean than the baseline estimator $\hat{\mu}_0(X_e)$, in the sense that

$$Var\left(\hat{\mu}_{mf}(X_e)\right) < Var\left(\hat{\mu}_0(X_e)\right). \quad (6)$$

Indeed, suppose for simplicity that the two terms in Eq. (2) are independent so that the variance of their sum is the sum of the variances. The variance of the sample average $\bar{Z}_T$ of a stationary process $Z(t)$ behaves for large T as

$$Var(\bar{Z}_T) \approx \frac{\Pi(Z)}{T}, \quad (7)$$

where the so-called long-run variance $\Pi(Z) = \int_{-\infty}^{\infty} \gamma_Z(u)du$ accounts for temporal dependence in the process $Z(t)$ having the auto-covariance function $\gamma_Z(u) = Cov(Z(t+u), Z(t))$ at lag u. Then, Eq. (5) is equivalent (for large T_e and T_s) to

$$\frac{\Pi(X_s)}{T_s} + \frac{\Pi(\epsilon)}{T_e} < \frac{\Pi(X_e)}{T_e}, \quad (8)$$

which provides a verifiable condition for the MF estimator to outperform the baseline. As seen from Eq. (8), this will happen if

$$T_e \ll T_s \quad \text{and} \quad \Pi(\epsilon) < \Pi(X_e). \quad (9)$$

The first relation of Eq. (9) is natural in the scenario of low costs for the simple model. The second relation in Eq. (9) states effectively that the error between the signals of the simple and expensive models has to be small compared to the original expensive signal. This is also intuitive as the simple model should be useful only if it approximates the expensive model well. We should also note that the key consequence of Eqs. (6) and (8) is that a normal confidence interval for the mean $\mu(X_e)$ would be smaller when using $\hat{\mu}_{mf}(X_e)$ as its length is determined by $Var(\hat{\mu}_{mf}(X_e))$.

In practice, the above discussion also suggests how to proceed in estimating the mean with the simple and expensive model data. First, estimate the long-run variances $\Pi(\epsilon)$ and $\Pi(X_e)$. Estimation of these quantities is discussed in detail in e.g. Pipiras *et al.* (2018). Second, compare the resulting estimates $\widehat{\Pi}(\epsilon)$ and $\widehat{\Pi}(X_e)$. If $\widehat{\Pi}(\epsilon)$ is smaller than $\widehat{\Pi}(X_e)$, then the MF estimator should be preferred to for sufficiently large T_s. Again, this would translate into smaller confidence intervals for $\mu(X_e)$.

It should also be stressed that though the case of the mean seems simplistic, it is at the core of estimation of many quantities. For example, the variance $Var(X_e) = \mu(X_e{}^2) - (\mu(X_e))^2$ is expressed through the means of a process and its square and can be dealt with similarly.

4. ILLUSTRATION FOR NON-RARE PROBLEMS

To illustrate the procedure of Section 3, we use synthetic data from a non-linear random oscillator model describing qualitatively ship rolling. More specifically, suppose the dynamics of a stationary process $X(t)$ is governed by the equation

$$\ddot{X}(t) + 2\delta\dot{X}(t) + r\big(X(t)\big) = Z(t), \quad (10)$$

where $\delta > 0$ is a damping parameter, $r(x)$ is a restoring force and $Z(t)$ is a zero-mean random excitation. The excitation $Z(t)$ is commonly assumed to be a Gaussian stationary process, with the spectral density suggested by e.g. the Bretschneider spectrum for wave elevations. We further assume a piecewise linear restoring force $r(x)$, given by

$$r(x) = \begin{cases} \omega_0^2 x, & \text{if } |x| \le x_m, \\ -k\omega_0^2(x - x_m) + \omega_0^2 x_m, & \text{if } |x| > x_m, \end{cases} \quad (11)$$

where ω_0 is a natural frequency of the system, x_m is referred to as a knuckle point (separating the linear and nonlinear regimes) and $k > 0$ enters into the negative slope of the nonlinear part. The restoring force has a softening shape for $|x| > x_m$, typical in modeling ship motions.

Figure 1 presents time plots of two realizations of the random oscillator model in Eq. (10), labeled expensive and simple. For the expensive signal X_e, the values $w_0 = 0.6$, $\delta = 0.15w_0$, $k = 1$, $x_m = 30\pi/180$ are taken. The same values were used for the simple signal X_s, except that $k = 0.3$ and the variance for the excitation is smaller. We emphasize again that these expensive and simple signals are called so for illustration purposes only; there is nothing more expensive or simpler, or high- or low-fidelity about either of the signals. The signals were generated for $T_e = 3600$ seconds (1 hour) and $T_s =$

36000 seconds (10 hours). Figure 2 depicts the time plot of the error process $\epsilon(t)$ between the two signals for the first 360 seconds. Note that the vertical scale in Figure 2 is much smaller compared to that in Figure 1, suggesting that the simple model might be a good approximation for the expensive model.

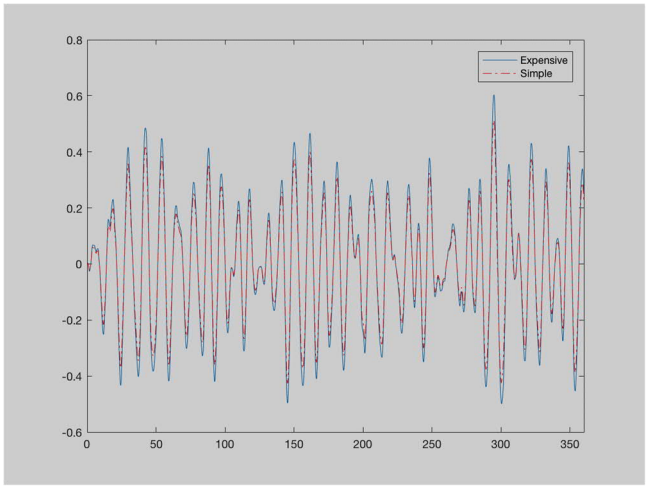

Figure 1: Two realizations of random oscillator model.

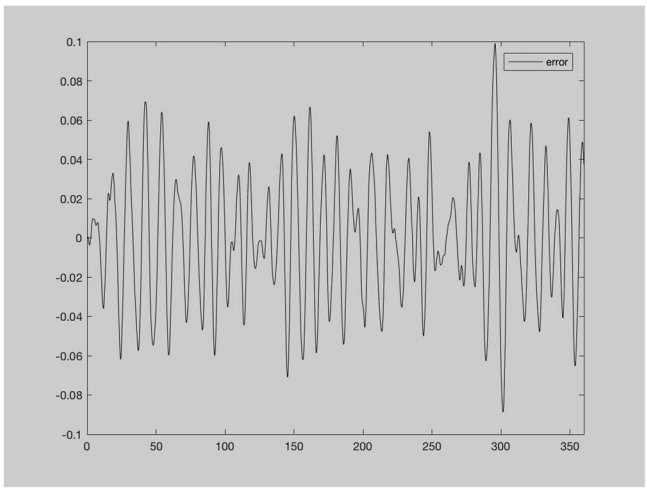

Figure 2: The error process for two realizations in Figure 1.

For the two signals, the long-run variances estimated through triangular kernel and "decorrelation time" bandwidth (see Pipiras *et al.*, 2018) were $\widehat{\Pi}(\epsilon) = 0.0001$ and $\widehat{\Pi}(X_e) = 0.0051$. Clearly, $\widehat{\Pi}(\epsilon)$ is smaller than $\widehat{\Pi}(X_e)$ by an order of magnitude. In this case, the MF estimator is preferred for T_s larger than T_e. The confidence interval for the mean resulting from the MF estimator is depicted in Figure 3 (the right vertical segment) in comparison to the confidence interval if the baseline estimator is used (the left vertical segment). The two mean estimates are indicated as circles on the two confidence intervals.

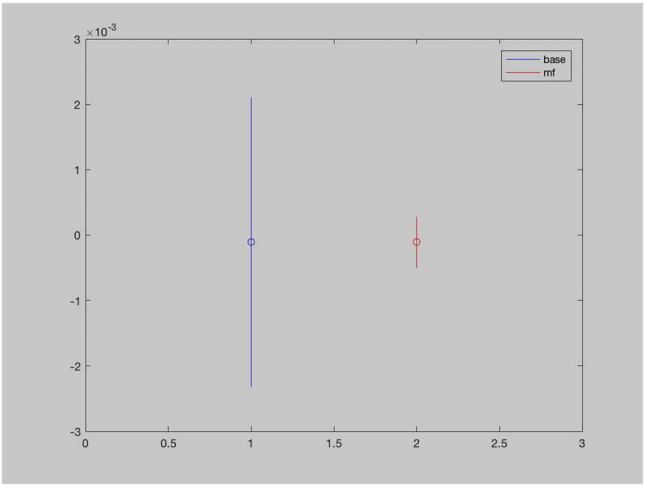

Figure 3: Confidence intervals for the mean based on the baseline and the MF estimator.

Figure 3 shows a clear benefit of the MF estimator and the simple signal in this case. Again, what makes this possible is a relatively small variance of the error process for the two signals and the fact that $T_s > T_e$. This should not be taken for granted in a given situation and might require proper calibration of the simple and expensive models.

5. METHODS FOR RARE PROBLEMS

We would like to extend the methods described in Sections 3 and 4 to estimation of an exceedance probability $\mathbb{P}(X_e > x)$ for large target x. For oscillating signals related to ship dynamics, X_e in the exceedance probability typically represents suitable peaks of the signal, perhaps even peaks of an envelope (e.g. Campbell *et al.*, 2016). To simplify the discussion and for technical reasons, we shall further assume that X_e represents block maxima of the peaks. If needed, block maxima exceeding a critical value could be translated to peak exceedances per unit time.

Figure 4 illustrates the notions of peaks and block maxima on the same synthetic data as in Section 3, where 10 hours of data are used with both expensive and simple models. The figure depicts a scatterplot of peaks from the expensive and simple signals, and in a darker shade, the respective block maxima are marked for 39 blocks of size 30. In this setting, for example, one might be interested to estimate $\mathbb{P}(X_e > 1.5)$, with no occurrences of the event $X_e > 1.5$ in the data as can be seen from Figure 4. Would having potentially larger simple model data for X_s help in this case, and through what method?

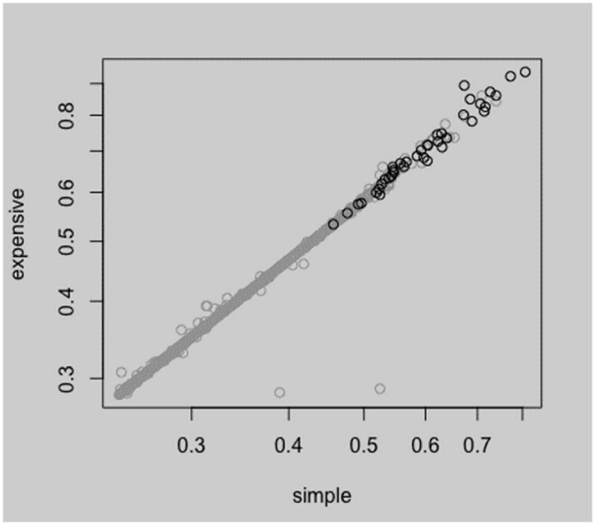

Figure 4: Scatterplot of peaks and block maxima for the simple and expensive models.

Before addressing these questions, it is instructive to discuss what the baseline estimator for $\mathbb{P}(X_e > x)$ is, without the availability of X_s from the simple model. It is known from the statistical Extreme Value Theory (EVT) that the distribution of the block maxima follows approximately that of a generalized extreme value (GEV) distribution as

$$\mathbb{P}(X_e \le z_1) = e^{-y_1}, \qquad (12)$$

where

$$y_1 = y_1(z_1) = \left(1 + \xi_1 \frac{z_1 - \mu_1}{\sigma_1}\right)_+^{-1/\xi_1} \qquad (13)$$

with location, scale and shape parameters μ_1, σ_1 and ξ_1, respectively, and the subscript + indicating the positive part of the function. After fitting these parameters to the data, the GEV distribution in Eq. (12) would be used to "extrapolate" into the tail $X_e > x$. A confidence interval for $\mathbb{P}(X_e > x)$ could also be provided.

Suppose now that the block maxima X_s are available for the simple model as well. To see how they could be used together with X_e, we need an analogue of Eq. (5). At the population (theoretical) level, consider

$$\mathbb{P}(X_e > x) = \mathbb{P}(X_s > y) + (\mathbb{P}(X_e > x) - \mathbb{P}(X_s > y))$$

and, after rewriting the difference in the parentheses,

$$\mathbb{P}(X_e > x) = \mathbb{P}(X_s > y) + \mathbb{P}(\epsilon), \qquad (14)$$

where

$$\mathbb{P}(\epsilon) = \mathbb{P}(X_e > x, X_s \le y) - \mathbb{P}(X_e \le x, X_s > y). \qquad (15)$$

We view Eqs. (14) and (15) as analogues of Eq. (5). That is, $\mathbb{P}(X_e > x)$ for the expensive model is being replaced by $\mathbb{P}(X_s > y)$ for the simple model, with the error probability $\mathbb{P}(\epsilon)$. The error probability in Eq. (15) is expressed in terms of the joint behavior of X_e and X_s, and could be expected small if the simple model is a good approximation to the expensive model at the extremes. The value y could but does not have to be equal to x; in fact, in analogy to Eq. (5) where $\mu(X_e)$ and $\mu(X_s)$ can be different, having different y and x can be critical.

Turning to estimation, the probability $\mathbb{P}(X_s > y)$ in Eq. (14) could, in principle, be estimated directly if needed, by taking a large enough T_s. The probabilities in Eq. (15), however, need to be estimated from the data on X_e and X_s gathered under the same conditions over the smaller observation window of size T_e. This is where bivariate GEV distributions come in, as those modeling the joint behavior of X_e and X_s. As in Eqs. (12) and (13), let

$$\mathbb{P}(X_s \le z_2) = e^{-y_2}, \qquad (16)$$

where

$$y_2 = y_2(z_2) = \left(1 + \xi_2 \frac{z_2 - \mu_2}{\sigma_2}\right)_+^{-1/\xi_2} \qquad (17)$$

with a similar set of parameters. The cross-dependence between the two variables X_e and X_s of a bivariate GEV is described through a dependence function A, for example, as in

$$\mathbb{P}(X_e \le z_1, X_s \le z_2) = e^{-(y_1+y_2)A\left(\frac{y_1}{y_1+y_2}\right)}, \qquad (18)$$

where y_1 and y_2 are given in Eqs. (13) and (17). The function $A(t)$ is defined for $t \in [0,1]$, is convex and satisfies $\max(t, 1-t) \le A(t) \le 1, A(0) = A(1) = 1$. (See Figure 5 for a plot of such functions.) The case of $A(t) = 1$ for all $t \in [0,1]$ corresponds to independence of X_e and X_s, since in this case $\mathbb{P}(X_e \le z_1, X_s \le z_2) = e^{-(y_1+y_2)}$ is the product of the marginals in Eqs. (12) and (16), and that of $A(0.5) = 0.5$ to their complete dependence. There are parametric models for $A(t)$ that can be fitted in practice.

After a bivariate model is fitted to X_e and X_s, one could obtain an estimate $\widehat{\mathbb{P}}(\epsilon)$ of the error probability, and also the estimate $\widehat{\mathbb{P}}(X_s > y)$ of the error probability (in the same way as the baseline estimator $\widehat{\mathbb{P}}(X_e > x)$), leading to the MF estimator

$$\widehat{\mathbb{P}}(X_e > x) = \widehat{\mathbb{P}}(X_s > y) + \widehat{\mathbb{P}}(\epsilon). \qquad (19)$$

A confidence interval can be constructed to go with $\widehat{\mathbb{P}}(\epsilon)$. For large enough T_s, the variability of $\widehat{\mathbb{P}}(X_s > y)$ can be thought negligible in comparison. If the variability expressed through a confidence

interval on $\widehat{\mathbb{P}}(\epsilon)$ is smaller than that of the baseline estimate $\widehat{\mathbb{P}}(X_e > x)$, then the MF estimate in Eq. (19) should be preferred. In practice, we suggest choosing y as the point corresponding to x through the regression line of X_e on X_s.

6. ILLUSTRATION FOR RARE PROBLEMS

We illustrate the ideas of Section 5 on the same synthetic data used in Section 4 and also in Figure 4. For this example, the estimated marginal parameters (and their standard errors in parentheses) are: $\hat{\mu}_1 = 0.5475\ (0.0083)$, $\hat{\sigma}_1 = 0.0471\ (0.0060)$, $\hat{\xi}_1 = 0.3876\ (0.1118)$ and $\hat{\mu}_2 = 0.6431\ (0.0102)$, $\hat{\sigma}_2 = 0.0577\ (0.0075)$, $\hat{\xi}_2 = 0.3963\ (0.1128)$. Figure 5 presents estimation of the function $A(t)$ entering Eq. (18) and modeling dependence through four parametric models. (For reference, the function $\max(t, 1-t)$ is also plotted in Figure 5.) Since $A(0.5)$ are close to 0.5 (see the discussion following Eq. (18)), the resulting plot suggests that the bivariate block maxima of X_e and X_s are quite strongly correlated. This is also consistent with the scatterplot of the block maxima (in a darker shade) in Figure 4.

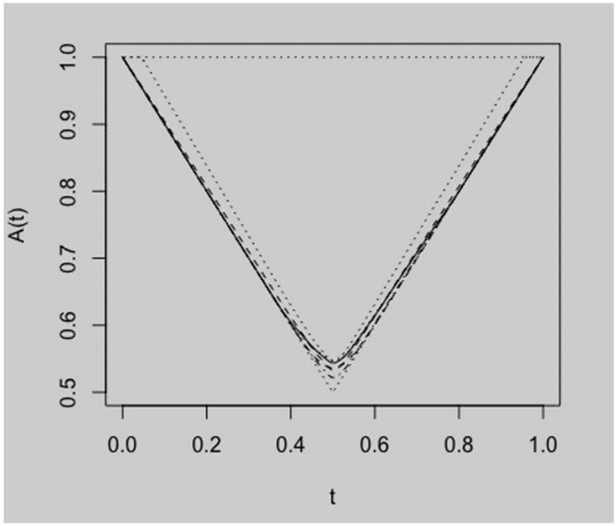

Figure 5: Estimation of $A(t)$ through four parametric models.

Figure 6 depicts the resulting baseline and MF probability estimates and their variability in vertical segments for the target $x = 1.5$. In producing the plot, we treated the fitted bivariate GEV model as the truth, with the horizontal line and the middle circle in the first vertical segment as the baseline estimator representing the true GEV probability $\mathbb{P}(X_e > 1.5)$. Variability is measured by generating data from the bivariate GEV model, re-estimating the probability $\mathbb{P}(X_e > 1.5)$, either through the baseline or the MF estimator, and taking the 0.025th and 0.975th quantiles of the obtained estimates as the endpoints of the vertical segments.

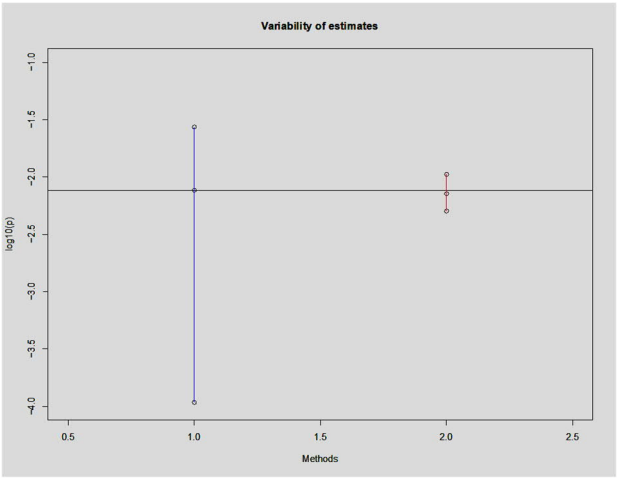

Figure 6: The baseline (left) and MF (right) probability estimates with confidence intervals.

Since the variability of the MF estimator is smaller than that of the baseline, the MF estimator is preferred. It should also be stressed that this is very much a result of strong extremal dependence in the simple and expensive models. Were the dependence not as strong (as expressed through the function $A(t)$ and which can be checked easily), the effect seen in Figure 6 would not be present.

7. CONCLUSIONS

In this work, we showed how a basic MF estimator for low-fidelity and high-fidelity models for non-rare problems could be adapted to estimate probabilities of rare events, especially those that are not observed in high-fidelity data. At a technical level, our approach was rooted in bivariate EVT, that allows modeling simultaneously extremes from the low-fidelity and high-fidelity models. The ideas were illustrated on synthetic data mimicking ship roll motion.

Several directions related to this work could be pursued in the future. First, the methodology should be applied to more realistic models of ship dynamics. Our first attempt in this direction was to compare roll extremes from SimpleCode and LAMP, but their dependence was not strong enough to warrant the use of MF methods. This could partly be a result of the lack of calibration between the two models, which is a topic of its own interest. Second, an even more mathematical treatment of the issues presented in Sections 5 and 6 should also be undertaken, for example, with the introduction of costs, a more

careful construction of confidence intervals, and the use of bivariate peaks-over-threshold methods instead of block maxima, etc.

ACKNOWLEDGMENT

This work has been funded by the Office of Naval Research grant N00014-19-1-2092 under Dr. Woei-Min Lin. The authors also thank Drs. Vadim Belenky and Kenneth Weems at NSWC Carderock Division for their comments on this work.

REFERENCES

Campbell, B., Belenky, V. and V. Pipiras (2016), "Application of the envelope peaks over threshold (EPOT) method for probabilistic assessment of dynamic stability," *Ocean Engineering*, Vol. 120, pp. 298–304.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Belenky, V., Weems, K., Pipiras, V., Glotzer, D. and T. Sapsis (2018), "Tail structure of roll and metric of capsizing in irregular waves," *Proceedings of the 32th Symposium of Naval Hydrodynamics*, Hamburg, Germany.

Belenky, V., Glotzer, D. Pipiras, V. and T. Sapsis (2019), "Distribution tail structure and extreme value analysis of constrained piecewise linear oscillators," *Probabilistic Engineering Mechanics,* To appear.

Lin, W. M. and D. K. P. Yue (1991), "Numerical solutions for large amplitude ship motions in the time-domain," *Proceedings of the 18th Symposium on Naval Hydrodynamics*, Ann Arbor.

Peherstorfer, B., Kramer, B. and K. Willcox (2017), "Combining multiple surrogate models to accelerate failure probability estimation with expensive high-fidelity models," *Journal of Computational Physics*, Vol. 341, pp. 61–75.

Peherstorfer, B., Willcox, K. and M. Gunzburger (2018), "Survey of multifidelity methods in uncertainty propagation, inference, and optimization," *SIAM Review*, Vol. 60, pp. 550–591.

Pipiras, V., Glotzer, D., Belenky, V., Levine, M. and K. Weems (2018), "On confidence intervals of mean and variance estimates of ship motions," *Proceedings of the 13th International Conference on the Stability of Ships and Ocean Vehicles*, Kobe, Japan.

Weems, K. and D. Wundrow (2013), "Hybrid models for fast time-domain simulation of stability failures in irregular waves with volume-based calculations for Froude-Krylov and hydrostatic forces," *Proceedings of the 13th International Ship Stability Workshop*, Brest.