Statistical analysis of cavitation erosion impacts in a vibratory apparatus with copulas

IOP Conf. Series: Earth and Environmental Science **240** (2019) 062035 doi:10.1088/1755-1315/240/6/062035

Statistical analysis of cavitation erosion impacts in a vibratory apparatus with copulas

G Taillon[1], K Onishi[1], T Mineshima[1] and K Miyagawa[2]

[1]Department of Applied Mechanics, Waseda University, Shinjuku-ku, Tokyo 169-8555, Japan
[2]Department of Applied Mechanics and Aerospace Engineering, Waseda University, Shinjuku-ku, Tokyo 169-8555, Japan

E-mail: ga.taillon@gmail.com

Abstract. A method of analysis of cavitation peaks (impact events) using copulas is developed. Impact events, otherwise known as peaks, are defined as maximum in the pressure amplitude applied to a material surface. These impact events were measured using a high speed pressure sensor in a cavitation apparatus based on the ASTM G32 standard. A total of 46180 impacts were measured over 100 realizations of 4ms long recording. First, the impact duration and amplitude's joint marginals are modeled as gamma distribution (part of the exponential family), determined by a Kolmogorov-Smirnov test (KS test). Then, copulas enable the study of the dependence structure of the measured impacts characteristics. The measured parameters are shown to not be independent but instead have a complex, asymmetric dependence structure. There are almost no impacts that have a combination of a high amplitude (>12MPa) and low duration (<5μs). The Tawn copula best fitted the data, as determined by a maximum likelihood method. An extension of the KS test to two dimensions demonstrated that the copula is a better fit compared with a joint distribution of independent marginals.

Keywords: Cavitation, Statistical analysis, Copulas, Impact events, Peaks, statistical dependence.

1. Introduction

Cavitation is an interesting phenomenon: vapor forms in a liquid under low pressure, essentially creating voids, or cavities in the fluid. In any liquid flowing through a pipe, an increase in speed results in a static pressure drop. This implies that nozzles or bends that result in higher flow speed can cause cavitation. In the context of the exploitation of turbomachines, these bubbles are mostly cause of concern: they cause instabilities, vibration, noise, and erosion. Rather surprisingly, cavitation bubble implosions cause shockwaves and microjets whose impact pressure amplitude can exceed material's yield strength [1]. For this reason, erosion of turbines by cavitation is a major problem. Typical mass loss can be as high as 200kg after a few years of operation [2]. Unfortunately, it is still quite difficult to predict the erosion rate as a function of the operating condition of a hydraulic machines. As part of efforts to increase the reliability of cavitation erosion models, a statistical analysis method of cavitation impacts is presented here.

Following the work of Franc and Hattori among others [3–6], we will describe a methodology that can be used to analyze the dependence between measured peak parameters, the peak amplitude (impact force or pressure) and duration of cavitation erosion impacts measured using high speed pressure sensors. The impact energy caused by acoustic impacts is directly

proportional to both these parameters [7] and the dynamics of deformation and erosion depend on the strain rate [8]. Certain impact duration and loading conditions observed in cavitation can enable high strain-rate effects, for which models such as the Johnson-Cook dynamic failure are useful [9]. As such, better impact modeling leads to better mass loss prediction.

The current impact measuring method is based on the recording of impact events using a high-speed pressure sensor and an oscilloscope in an ASTM G32 based vibratory apparatus. By measuring the impact amplitude and duration, and fitting this data to random distributions, one can observe if the parameters are independent. If random variables are independent, their joint distributions is simply the multiplication of their univariate marginal distributions. In the case of dependence between these variables, a copula can be used to link these two margins into one multivariate distribution, complete with the dependence structure [10, 11].

We hope that copulas could be part of a package tools used to study cavitation and cavitation erosion from the point of view of statistical analysis. The dependence structure of any number of impacts parameters would be computed using copulas, while the occurrence of impacts in time and space would be randomly generated by stochastic processes. Thus can be created a model that describes the random nature impacts, with parameters fitted empirically to the operating conditions. This model would be a rather simple mathematical object compared with deterministic and time consuming computational fluid dynamics (CFD) currently used to model cavitation. Random impact data could be generated with such a model, then combined with finite element models of erosion to compute mass loss predictions. For now, we present the methodology and results concerning the construction of a copula that describes the dependence structure of cavitation impact parameters, their durations and amplitudes.

2. Methodology

First, some time will be spent on presenting the vibratory cavitation machine in which cavitation impacts were recorded. Then, the peak measurement and statistical analysis methods based on copulas will be explained.

2.1. Vibration apparatus

The vibratory cavitation erosion apparatus setup is shown in figure 1. The tap water used was set to a temperature of $25 \pm 5°C$ using a recirculating water setup. The piezoelectric transducer's vibration frequency was 19.5kHz for an amplitude of 7µm. Inspired by the ASTM G32 standard [12], it does not conform to the standard so erosion results cannot be compared directly.

This apparatus creates a high amplitude pressure field in water, which causes cavitation bubbles to nucleate, grow, then collapse. These bubbles collapse a certain distance to the material surface, creating shockwaves and microjets that cause pits and damage to the material surface. This setup can be used in two different ways: by using the vibrating head as a sample, or by putting a sample a certain distance away from a more resistant head, such as Ti.

A pressure sensor manufactured by Muller Instruments was used to record the pressure amplitude over time. It has a very fast rise time, 50ns for 10% to 90%, has a dynamic range of -3MPa to 40MPa, and was placed at a distance of 1.4mm from the vibrating head. This sensor uses a 1mm diameter polyvinylidene difluoride (PVDF) piezoelectric sensor to detect the applied pressure. The actual mechanisms of bubble collapse cannot be observed with the sensor, only the impact pressure they cause on the material surface. The sensor outputs a voltage that can be converted to pressure with the factor 23.3kPa/mV, unique to the sensor's combination of sensitivity, cable capacitance and other parameters.

Knowing that impacts, or peaks, can have a full-width half maximum value lower than 1µs, a sampling rate of 10MS/s or faster was found to be necessary to properly resolve the peaks. The pressure amplitude was recorded for a total record time of 4ms at a sample rate of 25MS/s for

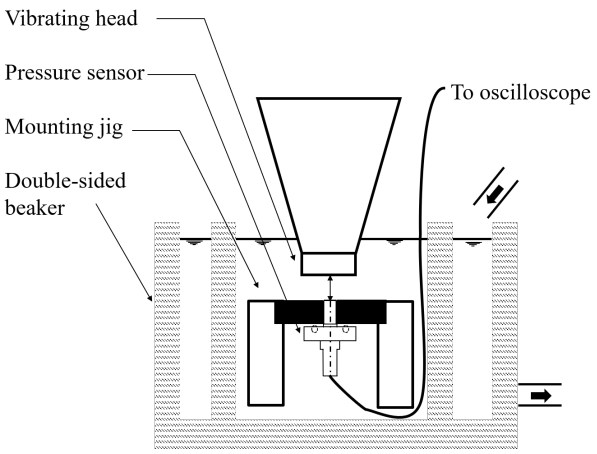


Figure 1. Vibratory cavitation apparatus setup with pressure sensor and cooling.

one hundred repetitions. A setup with an oscilloscope proved sufficient for these needs. With the recorded data, and after bandpass filtering ([3kHz, 1MHz]), it is possible to find the peaks, the maximal values of the amplitude above a certain noise threshold. The appropriate noise level for the impact measurements was found to be 20mV, converted to impacts of about 0.5MPa or lower being ignored.

These repeated recordings are known as realizations in statistical analysis and stochastic processes literature, and will be referred as such for the remainder of this text. The impact duration was computed by finding two points around each peak, the closest local minimas. If the impact amplitude hit zero closer to the peak than the closest minima, the closest zero was used instead. This method was chosen rather than fitting each peak to a function and finding it's full-width half-maximum for simplicity's sake. In the next section, the statistical analysis methods of these points using copulas is presented.

2.2. Univariate statistical analysis

Sections 2.2 and 2.3 show how to model the dependence between random variables. We will describe detailed step that are critical to building a copula: testing, selecting, fitting... starting with univariate data. To keep this methodology short, we will not show much detail beyond citing the methods by name and referring to relevant literature. For example, we refer to [13] and [14] for a discussion on the concept of dependence in statistics. The construction of a copula that models the dependence between random variables starts by fitting univariate data otherwise known as the marginal distributions, or margins.

Here, we use the ubiquitous Kolmogorov-Smirnov (KS) test to reject, or not-reject, the hypothesis that univariate data follows a certain distribution [15]. This test necessitates the construction of empirical cumulative distribution functions (ECDF), by way of counting the number of data points with equal or lesser value, and dividing by the total number of measured points. This can be mathematically written as:

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_i \leq t} \tag{1}$$

with $\hat{F}_n(t)$ the ECDF, $\mathbf{1}_A$ the indicator of event A which is equal to 1 if the condition A is met. After finding at least one distribution that fits each data set, one can turn to fitting, or parameter estimation. Many methods exist to fit a random's distribution parameters, we refer to [15] for

an overview. In the present context, software packages in the R and Python languages' libraries were used: the *fitdist* package [16] and the *scipy* package [17]. These packages contain functions that make use, among other methods, of the maximum likelihood estimation method [15] to fit data to random distributions. This way, one obtains fitted marginal distributions using experimental data, ready to build a copula to fit to the multivariate data.

2.3. Copulas for bivariate dependence modeling

The present subsection's namesake are functions that model the dependence between random variables. They take as inputs the marginal distributions, the distributions of the random variables estimated as if independent, and outputs the density in the combined data space.

If $H(x, y)$ is the joint distribution of two continuous random variables, then we can uniquely define the copula C as:

$$H(x, y) = C(F(x), G(y)) \tag{2}$$

with F and G the marginal distributions of random variables x and y respectively. The name copulas was chosen to emphasize the manner in which it couples univariate margins into one joint distribution [11, 13, 14]. For example, the Gumbel copula has only one parameter $\theta \in [1, \infty)$:

$$C_G(u, v) = \exp\left[-\left((-\log(u))^\theta + (-\log(v))^\theta\right)^{1/\theta}\right] \tag{3}$$

The case of independence can also be modeled as a copula: $C(x, y) = xy$. We also refer to [10,13,14] for a descriptions of copulas, their properties, their construction and applications. We found that many R-based software tools such as the packages *copula* [18–21] and *VineCopula* [22] proved useful to test and estimate copulas using bivariate data. In particular, the Tawn copula (or asymmetric logistic copula) [23, 24] of the *VineCopula* Package was found to fit the present data well. This copula's definition is based around so-called Pickands dependence functions. Equation 4 presents the way one can calculate the density in the probability space using a Pickands function A:

$$C(u, v) = (u, v)^{A(\omega)}, \text{with} \quad \omega = \frac{\ln(u)}{\ln(uv)} \tag{4}$$

The Tawn copula's pickand function is written as:

$$A(t) = (1 - \psi_2)(1 - t) + (1 - \psi_1)t + \left[(\psi_1(1 - t))^\theta + (\psi_2 t)^\theta\right]^{1/\theta} \tag{5}$$

with $t \in [0, 1]$, $0 \leq \psi_1, \psi_2 \leq 1$ and $\theta \in [1, \infty)$. The Tawn copula is in actuality a Gumbel copula with two additional asymmetry parameters: ψ_1 and ψ_2.

If these two parameters are equal to unity the Gumbel copula is obtained. In the *VineCopula* package [22], the type 1 and 2 Tawn copulas refer to $\psi_1 = 1$ or $\psi_2 = 1$, and this package also includes rotated version of these copulas.

3. Results and Discussion

3.1. Fitting the marginals

The raw and filtered amplitude data is presented in 2, the found peaks and duration points are presented in figure 3. Any frequencies outside of the [3kHz, 1MHz] interval were filtered before applying the peak finding and duration measuring algorithms. As one can observe, this filtering introduces some amount of error on the maximal value of the peaks, but it is necessary to remove the high frequency noise visible on the raw data. For every impact event, two marks were measured: the pressure amplitude and the impact duration. An example of the peaks and minimas around it used to measure the duration are presented in figure 3. A total number of

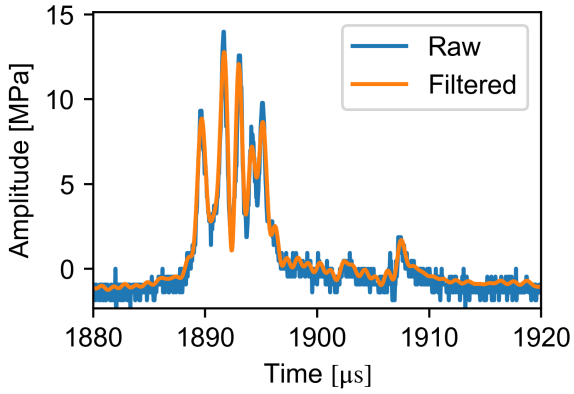


Figure 2. Pressure amplitude as a function of time.

41682 peaks were detected in the 100 realizations of the 4ms long measurement, for an average impact rate of 1.042×10^5 impacts per second.

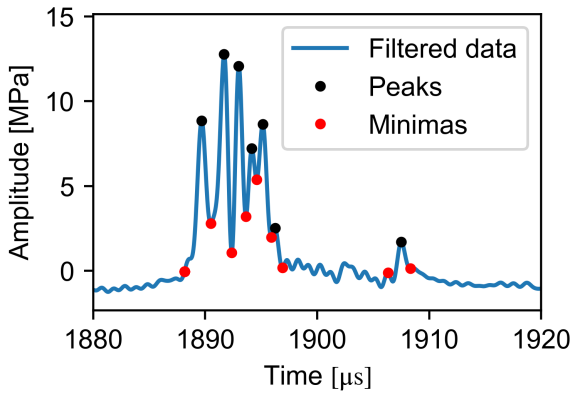


Figure 3. Peaks and surrounding mimimas.

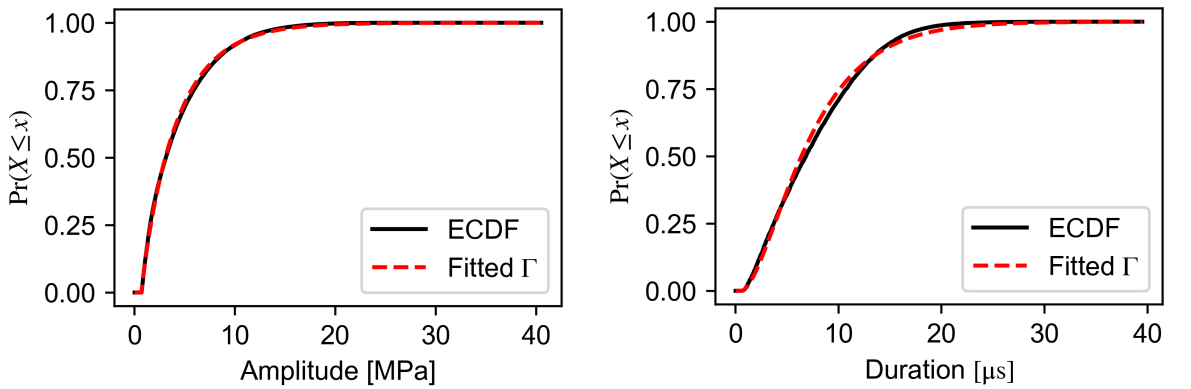
The KS test was then used to determine that the gamma distribution is appropriate to model both the amplitude and duration distribution at a significance level of 5%. By appropriate we mean that these distributions were not rejected by the KS test.

Now, other distributions are also appropriate for this data, such as the general exponential distribution for the amplitude data, but the gamma distribution was appropriate for both datasets. The experimentally determined ECDF as well as the fitted gamma distributions are presented in figure 4. The gamma distribution has two parameters, which can be described in two ways:

- the shape parameter k and scale parameter ϕ
- the shape $\alpha = k$ and rate parameter $\beta = 1/\phi$

The first parametrization was used here, with the additional location parameter *loc* that describes the shift of the random variable compared to 0.

Using these parameters it is possible to draw samples from the distribution, also known as random variate generation. The software tools used here provided ready made functions for this



(a) Amplitude fitted to a Γ distribution: $k = 0.894, \phi = 172\text{mV}^{-1}, loc = 30.0\text{mV}$.

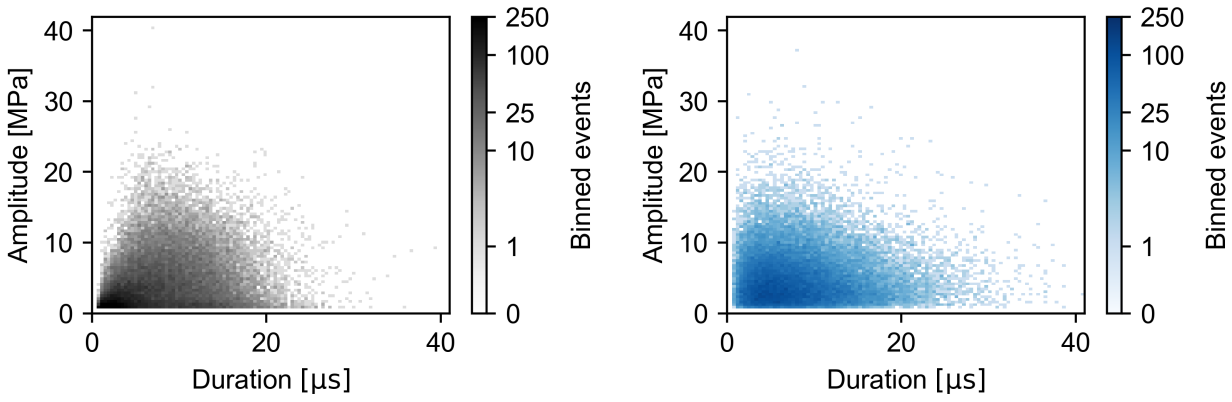
(b) Duration fitted to a Γ distribution: $k = 1.85, \phi = 3.83\mu\text{s}^{-1}, loc = 0.537\mu\text{s}$.

Figure 4. Maximum likelihood fitting of marginal distributions using experimental cumulative density functions (ECDF).

purpose, we refer to literature on generation of random variates [25] for more information on that subject. With the marginal distributions determined, one is ready to compute the dependence structure using copulas.

3.2. Fitting the copula

As a first step to check the data, all the marks were binned in a 2D histogram, as illustrated in figure 5. Then, points generated from the joint distribution computed by multiplying the two marginal distributions (valid if independent) were generated and also binned, to visualize the difference.



(a) Experimental samples.

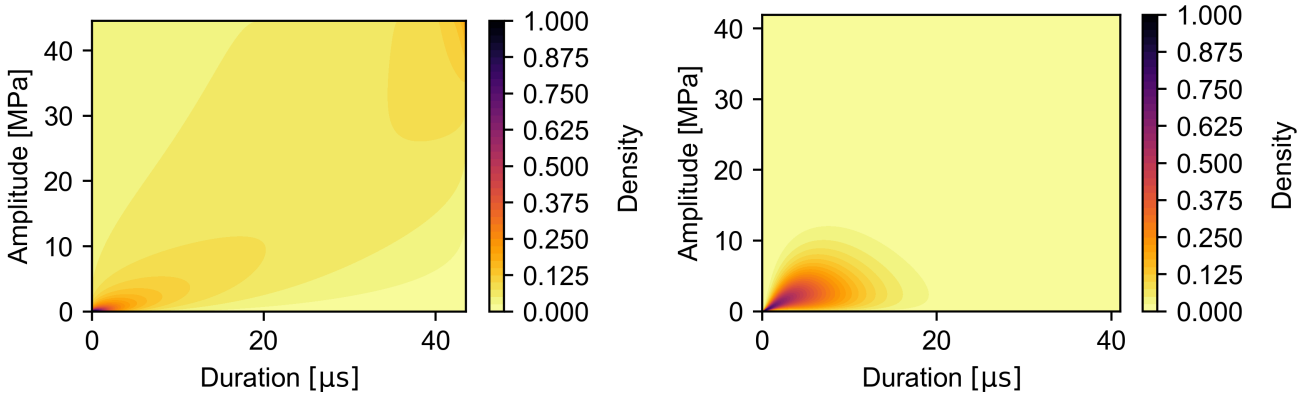
(b) Samples drawn from the joint distribution, assuming fitted marginal distributions are independent.
2DKS test statistic with experimental data → 10.92%.

Figure 5. Impact amplitude and duration samples as 2D histograms. The total number of events is 41682. The bins are approximately 0.4MPa and 0.4 μs wide, for a total of 100×100 square bins shown.

As one can observe, the experimental points are skewed to the high duration region. There are almost no impacts that have a combination of low duration (lower then 5μs) and high amplitude (higher then 500mV or about 12MPa) in the experimental data. This is the first indication that the random variables might not be independent. Also, the highest number of impacts seems to have a combination of low amplitude and short duration, with the number of impacts in higher amplitude and longer duration region decreasing exponentially.

To wit, the copula most likely to represent the data is the type 2 Tawn copula (with $\psi_2 = 1$) rotated 180 degrees, also known as the survival type 2 Tawn copula (one of so-called extreme-value copulas). After the selection process, this copula was fitted to the data, the copula density as well as the multivariate distribution density constructed using copulas were computed, and are shown in figure 6. In this figure, the copula density is computed by injecting the data space in the copula as $C(x, y)$, and the multivariate distribution is found by injecting the marginals, like so $C(F(x), G(y))$.

As is visible, the copula is highly asymmetric, skewed to the high duration values, with lower probability density in the high amplitude-low duration region, as expected. The fitted copula parameters are: $\theta = 1.69, \psi_1 = 0.370, \psi_2 = 1$. The density is significantly higher closer to a point just right of the origin, and decreases rapidly as the duration and amplitude increase.



(a) Copula density. **(b)** Multivariate distribution density.

Figure 6. Fitted type 2 Tawn Copula $\theta = 1.69, \psi_1 = 0.370, \psi_2 = 1$.

With this copula it is possible to generate a number of random samples. To compare with experimental data, an equal number of samples to the experimental peaks have been generated in figure 7. The appearance of the random points more closely matches the experimental data shown in figure 5 (b). To test for goodness-of-fit, an extension of the Kolmogorov-Smirnov test to two dimensions (2DKS) [26–28] was implemented with Python (available at `https://github.com/Gabinou/2DKS`) and applied.

This test compares two datasets of bi-dimensional data, and once again outputs a statistic which indicates whether or not the hypothesis that the two datasets were derived from the same distribution should be rejected. The 2DKS statistic for the samples generated using the independent marginals is 10.92%, and decreases to 7.65% when generated using the copulas. A decrease in the 2DKS statistic signifies that it is more difficult to reject the hypothesis that the samples were derived from the same distribution. It seems more correct to assume that the cavitation impact parameters are not independent.

Knowing this distribution, and the dependence structure, random variates that fit with experiments can be generated, which can be of use for deformation and erosion modeling. Such

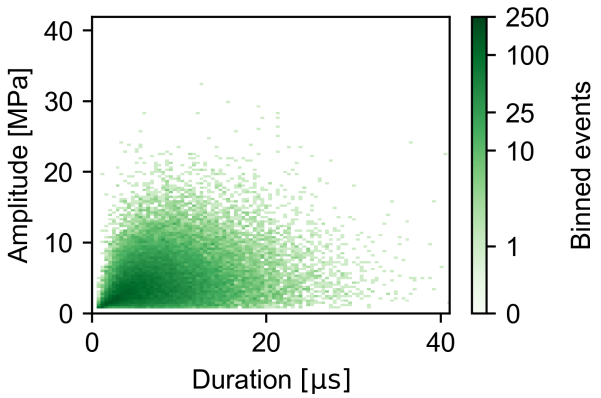


Figure 7. Two dimensional histogram of 41682 samples drawn from the fitted survival type 2 Tawn Copula. 2DKS test statistic with experimental data $\rightarrow$ 7.65%.

a model can be part of more generalized fluid dynamics or finite element computations aimed at predicting the mass loss of cavitating turbomachines. In subsequent studies, the variation of the fitted parameters as a function of operation of the vibratory apparatus will be performed.

4. Conclusion

Cavitation pressure was measured over time using a high-speed pressure sensor. Statistical analysis was performed to determine the appropriate distributions with which to model the cavitation impact characteristics: the impact amplitude and duration. A total of 41682 were recorded over 100 realization of a 4ms long measure, for an average impact rate of 1.042×10^5 impacts per second. At a significance level of 5% with the KS test, the gamma distribution among others was found appropriate to model these random variables.

Copulas were then used to model the dependence structure between the measured impact amplitude and impact duration. It was found that an asymmetric type 2 survival Tawn copula to be the most likely to fit the experimental multivariate data. By fitting the data to this copula is obtained a multivariate density that enables the generation of random samples that fit the experimental data better than assuming independence. This copula has a lower density in the high amplitude (>12MPa) and low duration (<5μs) region. The two dimensional KS statistic for the samples generated using the independent marginals is 10.92%, and decreased to 7.65% indicating that the data generated by the copulas better fits with the experimental distribution.

5. Acknowledgments

The authors would like to thank the Waseda Research Institute for Science and Engineering (WISE) for providing support to the presented research, in context of the project: “High performance and high reliability research for hydraulic turbomachinery systems”.

References

[1] Kim K-H, Chahine G, Franc J-P and Karimi A (ed) 2014 *Advanced Experimental and Numerical Techniques for Cavitation Erosion Prediction*, (Fluid Mechanics and Its Applications vol 106) (Dordrecht: Springer)
[2] Singh R, Tiwari S K, and Mishra S K 2012 Cavitation Erosion in Hydraulic Turbine Components and Mitigation by Coatings: Current Status and Future Needs. Journal of Materials Engineering and Performance, 21(7), 1539–1551.
[3] Hattori S, Hirose T and Sugiyama K 2010 Prediction method for cavitation erosion based on measurement of bubble collapse impact loads. *Wear* **269** 507–514

[4] Franc J P, Riondet M, Karimi A and Chahine G L 2011 Impact load measurements in an erosive cavitating flow. *J. Fluids Eng.* **133** 121301-8
[5] Momma T and Lichtarowicz A 1995 A study of pressures and erosion produced by collapsing cavitation *Wear* **186** 425-436
[6] Singh S, Choi J-K and Chahine G L 2013 Characterization of cavitation fields from measured pressure signals of cavitating jets and ultrasonic horns. *J. Fluids Eng.* **135** 9 091302-091302-11.
[7] Soyama H and Kumano H 2002 The fundamental threshold level—a new parameter for predicting cavitation erosion resistance *J. Ttesting and Eval.* **30** 5 421-431
[8] Callister W D 2007 *Materials science and engineering: An introduction* Seventh Edition (John Wiley & Sons, Inc.)
[9] Johnson G R and Cook W H 1983 *A constitutive model and data for metals subjected to large strains, high strain rates and high temperatures 7th Int. Symp. on Ballistics* (The Hague) pp 541–547
[10] Balakrishnan N and Lai C-D 2009 *Continuous Bivariate Distributions* (Dordrecht: Springer Science+Business Media, LLC)
[11] Joe H 2015 *Dependence Modeling with Copulas* (CRC Press)
[12] ASTM G32-16, Standard Test Method for Cavitation Erosion Using Vibratory Apparatus, ASTM International, West Conshohocken, PA, 2016, www.astm.org
[13] Joe H 1997 *Multivariate models and dependence concepts* (Monographs on Statistics and Applied Probability vol 73) (Dordrecht: Springer)
[14] Nelsen R B 2006 *An Introduction to Copulas* (Springer Series in Statistics vol 42) (New York: Springer Science+Business Media, Inc.)
[15] Lehmann E L, and Romano J P 2005 *Testing Statistical Hypotheses* (New York: Springer Science+Business Media, LLC)
[16] Delignette-Muller M L and Dutang C 2015 fitdistrplus: An R Package for Fitting Distributions. *J. Stat. Soft.* **64** 4 1-34
[17] Jones E, Oliphant E, Peterson P, et al. SciPy: Open Source Scientific Tools for Python, 2001-, http://www.scipy.org/ [Online; accessed 2018-04-25]
[18] Hofert M, Kojadinovic I, Maechler M and Yan J 2017 copula: Multivariate Dependence with Copulas R package version 0.999-18
[19] Yan J 2007 Enjoy the Joy of Copulas: With a Package copula *J. Stats Soft.* **21** 4 1-21
[20] Kojadinovic I and Yan J 2010 Modeling Multivariate Distributions with Continuous Margins Using the copula R Package. *J. Stats Soft.*, **34** 9 1-20
[21] Hofert M, Maechler M 2011 Nested Archimedean Copulas Meet R: The nacopula Package. *J. Stats Soft.* **39** 9 1-20
[22] Schepsmeier U, Stoeber J, Brechmann E C, Graeler B, Nagler T and Erhardt T 2018 VineCopula: Statistical Inference of Vine Copulas R package version 2.1.4.
[23] Bernard C and Czado C 2015 Conditional quantiles and tail dependence *J. Multivar. Anal.* **138** 104-126.
[24] Eschenburg P 2013 *Properties of extreme-value copulas* Technische Universitat Munchen.
[25] Law A M, Kelton W D 2007 *Simulation modeling and analysis* (New York: McGraw-Hill)
[26] Peacock J A 1983 Two-dimensional goodness-of-fit testing in astronomy *Mon. Not. Roy. Astron. Soc.* **202** 3 615-627
[27] Fasano G and Franceschini A 1987 A multidimensional version of the Kolmogorov–Smirnov test. *Mon. Not. R. Astron. Soc.* **225** 1 155-170
[28] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1996 *Numerical recipes in C (Vol. 2)* (Cambridge: Cambridge university press)