Stochastic Impact Model : Poisson Processes and Copulas to Model Cavitation Erosion Impacts

International Journal of Fluid Machinery and Systems
Vol. 12, No. 4, October-December 2019
DOI: http://dx.doi.org/10.5293/IJFMS.2019.12.4.418
ISSN (Online): 1882-9554

Original Paper

Stochastic Impact Model: Poisson Processes and Copulas to Model Cavitation Erosion Impacts

Gabriel Taillon[1] and Kazuyoshi Miyagawa[2]

[1]Department of Applied Mechanics, Waseda University
Shinjuku-ku, Tokyo 169-8555, Japan, taillon.gabriel@asagi.waseda.jp
[2]Department of Applied Mechanics and Aerospace Engineering, Waseda University
Shinjuku-ku, Tokyo 169-8555, Japan, k-miyagawa@waseda.jp

Abstract

A statistical model that represents impacts caused by cavitation bubble collapse on material surfaces is developed, as one step in the development of a novel cavitation erosion model. This model uses stochastic processes, for example the non-homogeneous Poisson process (NHPP), to represent impact events as points randomly distributed in time and space. Then, marks which represent the impact properties are computed using empirical distributions. This model is based on experimental data and empirical observations using a high-speed pressure sensor in a vibratory cavitation apparatus based on the ASTM G32 standard. A total of 41581 impacts were measured over 100 realizations of 4ms long recording. The rate of the non-homogeneous Poisson is shown to be a sinusoidal function, but the process is over-dispersed. The spatial distribution of points is independent from the temporal portion, so modeled independently as a Poisson cluster process. Two marks were measured for each impact event: the impact duration and amplitude. These mark's joint marginals are modeled as gamma distributions, determined by a Kolmogorov-Smirnov (KS) test. The marks are shown to be dependent, prompting the use of copulas to model their joint distribution. The best-fitting copula is a highly asymmetric Tawn copula, in the class of so-called extreme value copulas. Almost no impacts were observed with a combination of a high amplitude (>12MPa) and low duration (<5μs). 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. The combination of the spatio-temporal processes, with the mark's distributions combine to produce the stochastic impact model for cavitation erosion in the vibratory apparatus.

Keywords: Cavitation, statistical analysis, Poisson process, impacts, copulas.

1. Introduction

Cavitation is an interesting phenomenon: vapor forms in a liquid under low pressure, essentially creating voids, or cavities in a 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 cause higher flow speeds can lead to cavitation. In the context of the exploitation of turbomachines, these bubbles are mostly nefarious: they are source of instabilities, vibration, noise, and erosion. Rather surprisingly, cavitation bubble implosions create shockwaves and micro-jets 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 machine. As part of efforts to increase the reliability of cavitation erosion models, a stochastic model for cavitation erosion impacts is proposed here.

Following the work of Franc and Hattori among others [3]–[6], we will describe a model based on the distribution of impacts as points randomly distributed with time and space. The current impact measuring method is based on the recording of impact events using a Müller M60-1L-M3 high-speed pressure sensor and an oscilloscope in an ASTM G32 based vibratory apparatus. The peaks of amplitude in the pressure data can be treated as points randomly distributed in time, essentially as a stochastic process. The cornerstone of stochastic processes, the Poisson process with its useful properties and ease of computation was studied first. Further analysis demonstrated that one of its more complex variants, the non-homogeneous Poisson process, better fit the impact data.

Two additional parameters were measured for each point, the peak amplitude (impact force or pressure) and duration. 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,

Received March 30 2019; accepted for publication December 1 2019: Review conducted by Yoshinobu Tsujimoto. (Paper number O19060S)
Corresponding author: Gabriel Taillon, taillon.gabriel@asagi.waseda.jp

* Part of this paper was presented at the 29th IAHR Symposium on Hydraulic Machinery and Systems, held at Kyoto, Sept. 16-21st, 2018.

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 impact amplitude and duration are treated as so-called "marks" of the Poisson Process, which are auxiliary random variables associated with each point. In the present case, they were assumed independent from the Poisson process and from each other: each were fitted to an univariate distribution. If random variables are independent, their joint distribution is simply the multiplication of their univariate marginal distributions. The marks' joint distribution did not fit with the experimental data which demonstrates dependence. This justifies the use of a copula to link these two margins into one multivariate distribution, complete with the dependence structure [10], [11].

Unless proven otherwise, all parts of the stochastic impact model can be improved upon and refined independently using appropriate data or better analysis method. For example, compared to the previously published rate function, the one presented in this paper is slightly more accurate [12][1]. The combination of the spatial part, temporal part and marks (impact amplitude and duration) makes this an example of the stochastic impact model.

2. Methodology

First, the vibratory cavitation machine in which cavitation impacts were recorded is presented, followed by the peak measurement and statistical analysis methods.

2.1 Vibration apparatus

The ASTM G32 [13] inspired vibratory cavitation erosion apparatus setup is shown in Fig. 1. The tap water used is set to a temperature of 25.0 ± 5.0°C using a recirculating water setup. The piezoelectric transducer's vibration frequency is 19.5kHz for an amplitude of 7μm, the lowest setting to minimize sensor wear. This apparatus creates a high amplitude pressure field in water, which causes cavitation bubbles to nucleate, grow, then collapse. The pressure field's amplitude is easily computable, assuming the emitted wave is plane, see section 2.2 of [14]. These bubbles collapse a certain distance to the material surface, creating shockwaves and microjets that damage 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. In Fig. 1 the latter is shown, sometimes referred to as the indirect setup [1].

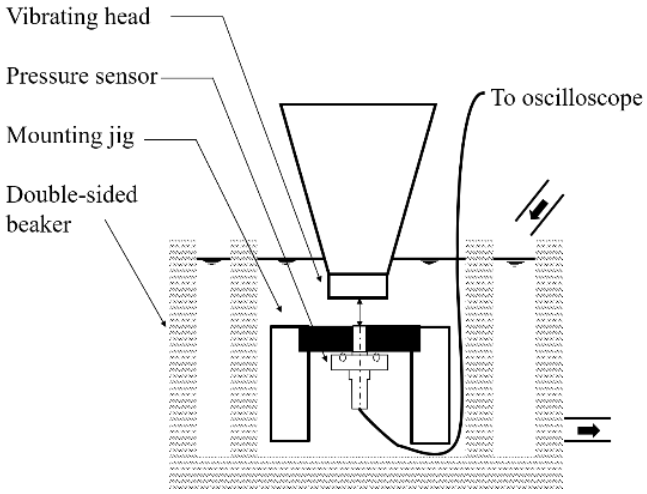


Fig. 1 Vibratory cavitation apparatus setup with pressure sensor and cooling

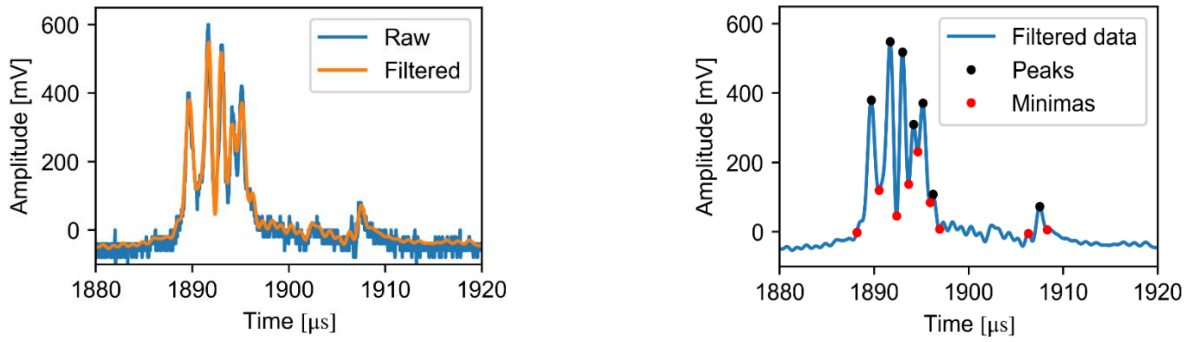
A pressure sensor manufactured by Muller Instruments is used to record the pressure amplitude over time. It has a very fast rise time of 50ns (for 10% to 90% of measured voltage) and has a dynamic range of -3MPa to 40MPa. It was placed at 1.4mm from the vibrating head to record impacts. This sensor uses a 1mm diameter polyvinylidene difluoride (PVDF) piezoelectric sensor to detect the applied pressure. 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. Impacts, or peaks, can have a full-width half maximum value lower than 1μs, so a sampling rate of 10MS/s or faster was necessary to properly resolve the peaks. The pressure amplitude was recorded for a total time of 4ms at a sample rate of 25MS/s for 100 realizations using an oscilloscope.

With the recorded data it is possible to find the peaks, the maximal values of the amplitude above a certain noise threshold. Some bandpass filtering is required, because the data is noisy as shown in Fig. 2. The frequencies ([3kHz, 1MHz]) remove enough noise for the peak finding algorithm, without introducing too much error. The appropriate noise level for the impact measurements is 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.

The impact duration is calculated using an algorithm that finds the interval between the peaks' two closest local minima. If the impact amplitude hit zero closer to the peak than the closest minima, the closest zero was used instead. This method was chosen

[1] Part of the work presented in the paper has been presented and discussed at the 29th IAHR Symposium on Hydraulic Machinery and Systems, Kyoto, 2018.

rather than fitting each peak to a function and finding it's full-width half-maximum for simplicity's sake. Fig. 2(b) shows results of the peak and amplitude finding algorithms. In the next section, the statistical analysis methods of these points using Poisson processes and copulas is presented.



(a) Pass-band filtering frequencies: [3kHz, 1MHz]

(b) Peaks or impact events all have associated amplitudes

Fig. 2 Peaks in pressure amplitude data recorded using the Müller M60-1L-M3

2.2 The stochastic impact model

This model abstracts the complex phenomenon of cavitation bubble collapse into a collection points distributed randomly in time and space, into a stochastic process. Specifically, cavitation impacts are modelled as a marked spatio-temporal Poisson process:

- A Poisson process is the simplest process describing the random distribution of points in any data space,
- Spatio-temporal because the points are distributed in time and space (with time treated as fundamentally different from space [15]),
- Marked refers to properties associated with every point. Each mark has its own distribution.

Any number of marks can be computed for every point, but here two were measured and fitted: the impact amplitude and duration. These mark's distributions were observed to be dependent and as such are modeled using copulas, see section 2.2.2. In the case of independence, a joint distribution simply the multiplication of the independent variables' univariate distributions, written so: $F_{X,Y}(x,y) = F_X(x)F_Y(y)$. The joint distribution of the measured marks, impact duration and amplitude, is observed to diverge significantly from this expected result in section 3.2.

Each part of the stochastic model will be briefly described, starting with an introduction of Poisson processes in the next section. Software packages of the programming languages R and Python were used whenever available, otherwise algorithms were implemented in Python.

This approach contrasts directly with efforts in bubble dynamics to model the behavior of single bubbles. Bubble dynamics describes the growth, collapse, and rebound of a bubble as a function of the initial conditions, with fluid mechanics. Computational fluid dynamics necessitates the calculation of the fluid properties at each timesteps, for each domain cell. As such it is a computationally intensive, microscale and deterministic (excluding turbulence) model of fluid behavior. The stochastic model is derived from empirical measurements in the macroscale time range. It makes uses of randomness to abstract unknown flow parameters into a rather simple mathematical model, much less computationally intensive than bubble dynamics, but suited for random impact generation in long time ranges.

The randomness of these models accounts for local parameters that influence the behavior of cavitation bubbles which are difficult to measure and control. Such parameters include the local temperature, reference pressure, driving pressure, bubble gas contents, initial nuclei size, number and position of closest neighbors, etc. The cavitating apparatus' operation conditions were fixed in this paper, but their influence on the stochastic model can be quantified empirically through change of the model parameters.

For now, a simple dataset with a single set of operating conditions is used to demonstrate the working principle of the model, as well as the analysis methods necessary to construct it. The nature of the relationship between the controllable cavitation conditions and the stochastic impact model parameters will be explored in future studies.

2.2.1 Poisson processes for time and space

The points calculated using a stochastic process represent the physical events, the impacts on a material surface caused by collapsing cavitation bubbles, as measured by a high-speed pressure sensor. The spatial and temporal parts of this point process were assumed independent, and the measurements presented in this paper cannot be used to comment on this assumption. This proves useful because the spatial and temporal parts can be separated into two different point processes, computed as if the spatial process was a mark of the temporal process or vice-versa.

The Poisson process represents independently occurring random events with a fixed rate (or intensity) generally noted λ. For such a process, any number of events in an interval of time t, noted $N(t)$, is distributed as a Poisson random variable with mean λt. This is the expected value of $N(t)$ noted $E[N(t)] \equiv \Lambda = \lambda t$, as well as its variance. The Poisson process has several other key properties, including the exponential distribution of the interarrival times. All the intervals between events, named inter-arrival times, are distributed as $P[\tau_i > t] = e^{-\lambda t}$, with τ_i the inter-arrival time.

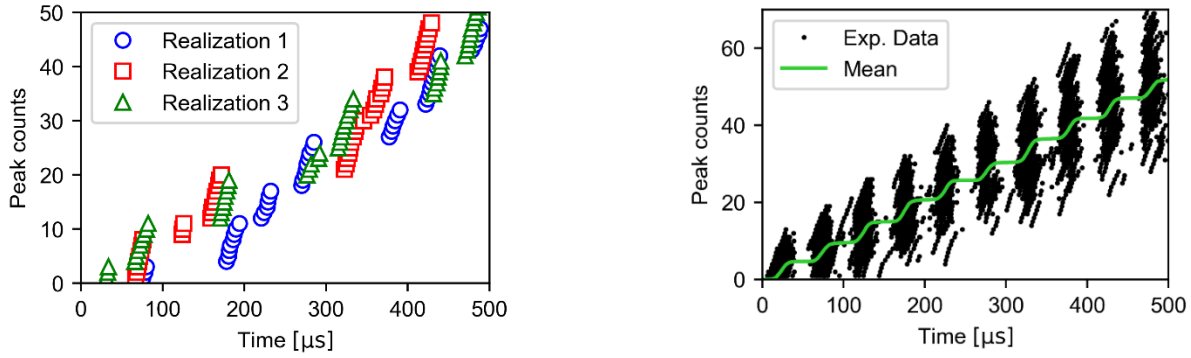
We refer to the statistical literature [16]–[20] for more comprehensive descriptions. The Poisson process' probability density function (PDF), which describes the probability that a number n of events occur in an interval of length t is written [11], [13]:

$$P[N(t) = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \qquad t \geq 0 \tag{1}$$

One convenient aspect of this process is the ease of estimation: the single parameter λ is its rate and can be estimated by dividing the total number of events $N(t)$ by the total measurement time t. The strong law of large numbers implies that this estimation is increasingly accurate with time. Increasing the number of realizations achieves the same effect [14], [16], [19].

The process is also rather simple to simulate, that is, it is easy to generate random points. This is sometimes known as drawing random variates, a specific outcome of a random variable or process. The arrival times of any number of events can be obtained iteratively by injecting uniformly distributed probabilities into the distribution equation of arrival times $P[\tau_i > t] = e^{-\lambda t}$. This can be written as $t_{next} = -\frac{\ln(u)}{\lambda}$ with $u \sim U(0,1)$ a random variate of the uniform distribution U between 0 and 1.

Unfortunately, the experimental impact data could not be modeled by a simple Poisson process, because the rate was observed to vary over time. In Fig. 3(a) the peak counts of 3 different realizations are shown. For a Poisson process, the mean as a function of time is expected to be linear. In Fig. 3(b), the impact peaks of the measured 100 realizations are shown, with its mean oscillating. Other processes exist to model this behavior.



(a) Peak counts for 3 separate realizations

(b) All 100 realizations with mean

Fig. 3 Measured peak counts as a function of time

In the present case, the rate λ itself is a function $\lambda(t)$. These processes are known as non-homogeneous (or inhomogeneous) Poisson processes (NHPP). This process' expectation becomes $\Lambda(t) = \int_0^t \lambda(\tau) d\tau$. With a high number of realizations, the rate function can be estimated using the mean number of events as a function of time. Other estimation procedures exist, such as the one proposed by Leemis [22].

The simulation of a NHPP is similarly a fair bit more complex. A number of algorithms exist, such as can be found in this technical note [23]. The thinning (or acceptance/rejection) algorithm was found most useful to generate events presented here, because the density functions used here are reasonably tractable. Essentially, this algorithm starts by generating points like a Poisson process, with a rate equal to the maximum rate λ_{max} in the interval of interest. Then, points are randomly rejected in proportion to the rate at the point. Points which satisfy the following condition $u \sim U(0,1) \rightarrow u \leq \lambda(t)/\lambda_{max}$ are accepted.

Concerning the spatial part of the stochastic model, pictures of single bubble impact indentations on metal surfaces such as found in [1], [24] were used to build some spatial distributions, made to visually resemble the indentation pictures. The indentations can be clearly understood to be part of a Poisson cluster process: some parents events occur randomly in space, with each parent leading to the creation of a certain number of children events [15]. In this case the physical mechanisms lead to a clear definition for parents and children: the initial bubble collapse leads to the collapse of micro bubbles. These micro bubbles form a toroidal shape before collapse or cluster around a point depending on the collapse parameters. The indentations imply a bubble collapse event, but the high number of collapse events compared to the indentations leads to the conclusion that not all impacts cause indentations.

Simple distributions that visually resemble the impact pictures in [1], [24] were chosen. For parents a gaussian peak is suitable, and for the children some radially symmetric functions are used. The parents and children are generated by thinning with their respective distribution centered on their parent. Collapsing bubbles do not necessarily lead to the collapse of micro bubbles, but the number of micro bubbles is unclear. Here, a uniformly random number of children events between 0 and 5 is chosen. Unfortunately, the high-speed pressure sensor cannot be used to assess the spatial distribution of impacts, nor differentiate between parent and children events. Some generated samples using the thinning algorithm are presented in section 3.1, as a part of the complete stochastic cavitation impact model.

The other part of the stochastic impact model are the marks, two of which were measured here. These marks were observed to be dependent, and as such are modeled using so-called copulas. First, the necessary basics of univariate statistical analysis are introduced.

2.2.2 Univariate statistical analysis for impact properties

The next sections are focused on the statistical analysis of marks: the impact amplitude and duration. They were first assumed to be independent random variables, but some statistical dependence was observed. Comprehensive definitions of dependence can be found in [25], [26]. In this article, it is sufficient to observe that the joint distribution is not equal to the multiplication of the

variable's distribution to reject independence, as is shown in Fig. 11. The marks cannot be modeled using only univariate distributions: a copula is necessary to build the joint distribution, rather than only multiplying the univariate distributions.

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, in the literature [11], [26]. To keep this methodology short, not much detail will be shown beyond citing the methods by name and referring to relevant literature.

The impact data is first fitted to several common probability distributions, then a test is performed to choose the best fitting distribution. The ubiquitous Kolmogorov-Smirnov (KS) test is used in this paper to reject, or not-reject, the hypothesis that univariate data follows the distribution [27]. This test necessitates the construction of empirical cumulative distribution functions (ECDF). This can be written as:

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_i \leq 1} \tag{2}$$

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. Many methods exist to fit a random distribution's parameters, an overview is presented in [27]. Specifically, the *fitdist* package for R and the *scipy* package for Python make use of the maximum likelihood estimation method to fit data to random distributions, among other methods. Using these distributions, it is possible to draw samples from the distribution, also known as random variate generation [25]. The software tools used here provided readymade functions for this purpose. With the fitted marginal distributions, one is ready to use copulas to fit to the multivariate data.

2.2.3 Copulas for bivariate dependence modeling

The present subsection's namesake are functions that model the dependence between random variables. These are functions with the marginal distributions as inputs, and outputs the probability density (sometimes referred to only as density) in the combined data space. If $H(x, y)$ is the joint distribution of two continuous random variables, then the copula C is uniquely defined as:

$$H(x, y) = C\big(F(x), G(y)\big) \tag{3}$$

with F and G the marginal distributions of random variables x and y respectively. The name 'copula' was chosen to emphasize the manner in which it couples univariate margins into one joint distribution [11], [25], [26], [28]. For example, the Gumbel copula has only one parameter $\theta \in [1, \infty)$:

$$C_G(u, v) = \exp\left[-\left((-\log(u))^\theta + (-\log(v))^\theta\right)^{1/\theta}\right] \tag{4}$$

The case of independence can also be modeled as a copula: $C(x, y) = xy$. We also refer to [10], [25], [26] for a descriptions of copulas, their properties, their construction and applications. Many R-based software tools such as the packages *copula* and *VineCopula* proved useful to test and estimate copulas for the present bivariate data.

The impacts amplitude and duration marks were better fitted by the Tawn copula (or asymmetric logistic copula) [28]. This copula's definition is based around so-called Pickands dependence functions. Equation (5) presents the density in the probability space using a Pickands function P:

$$C(u, v) = (u, v)^{P(\omega)}, \text{ with } \omega = \frac{\ln(u)}{\ln(uv)} \tag{5}$$

The Tawn copula's Pickands function is written as:

$$P(t) = (1-t)(1-\psi_2) + t(1-\psi_1) + \left[\left(\psi_1(1-t)\right)^\theta + (\psi_2 t)^\theta\right]^{1/\theta} \tag{6}$$

with $t \in [0,1], 0 \leq \psi_1, \psi_2 \leq 1$ and $\theta \in [1, \infty)$. The Tawn copula is an extension of the 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, the type 1 and 2 Tawn copulas refer to $\psi_1 = 1$ or $\psi_2 = 1$ respectively, and this package also includes rotated version of these copulas.

3. Results and discussion

The dataset presented here is the same one used in the [29] conference proceedings, which presents part of the methodology, some results in the construction of the stochastic impact model.

3.1 Spatio-Temporal Poisson process

First, the spatial distribution of the impacts is constructed. In Fig. 4 is shown an example of the indentations caused by the repeated impacts on the surface of the material. On the center are the indentations caused by the initial bubble, and around are the impacts caused by children microbubbles. The distributions presented in Fig. 5 show the artificial distributions based on these pictures [24]. This part of the stochastic impact model is treated as an additional mark of the process, independent of the temporal process or other marks.

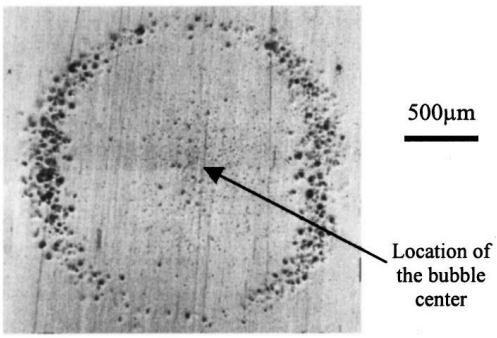
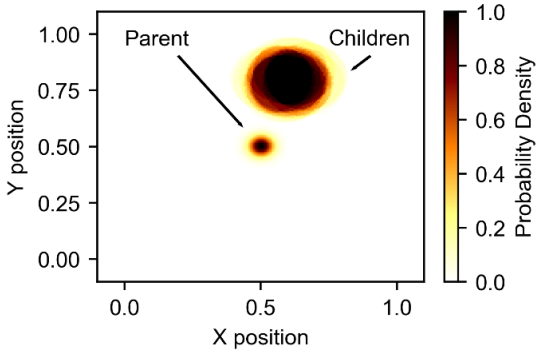
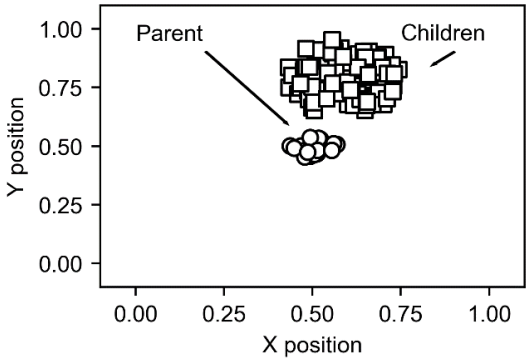


Fig. 4 Image of impacts on aluminium caused by repeated single bubble collapse.

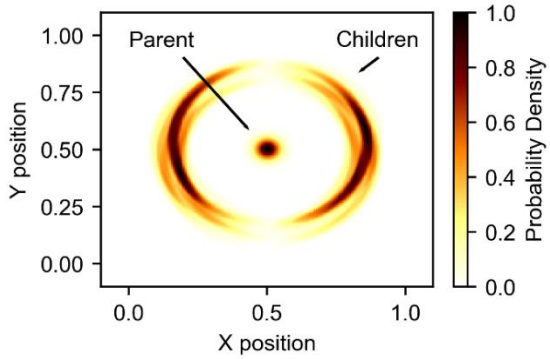
A total of 20 parent events were simulated, with a random number between 0 and 5 of child events generated using a distribution centered around the parent point. The parent always is a normal distribution centered on [0.5, 0.5] with [0.001, 0.001] variance. The parent and children distributions are presented in Fig. 5. All points were generated by thinning: first points are generated uniformly in the space, then accepted randomly in proportion to the probability distribution shown in Fig. 5 column 1. In this case, each children distribution is centered around the parent point's position, as observable in the variability of the children distributions.



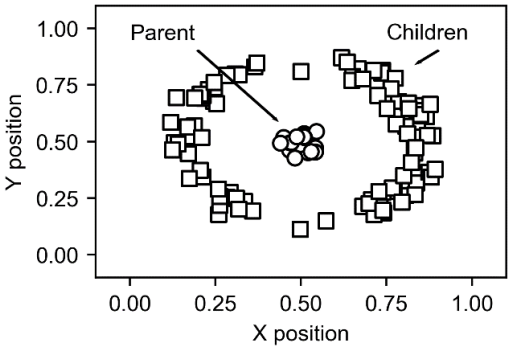
(a.1) Parent distribution: Centered gaussian peak.
Children distribution: Constant circle of radius $r = 0.15$ at a distance of $r = [0.1, 0.3]$ centered on parent point



(a.2) Parent and children points generated by thinning.
Parent events: 20
Children events: random between 0 and 5 for each parent



(b.1) Parent distribution: Centered gaussian peak.
Children distribution: Radial parabola with maximum at $r = 0.35$ that falls to 0 at the limits $r = [0.325, 0.375]$, multiplied by the cosine of the polar angle



(b.2) Parent and children points generated by thinning.
Parent events: 20
Children events: random between 0 and 5 for each parent

Fig. 5 Left (column 1): Spatial distribution of impacts for parents and children.
Right (column 2): Associated parent and children random variates generated by thinning.
The distributions are based on pictures of cavitation erosion indentations in [24].

This demonstrates the working principle of the thinning algorithm to generate points of a Poisson cluster process, for the spatial portion of the stochastic impact model. Row b in Fig. 5 best represents the indentations in Fig. 4. While the distributions here are arbitrary and constructed to visually resemble pictures, empirical distributions can be estimated using kernel density estimation for example. For now, let us move on to an example of temporal process for the stochastic model using impact data.

Over the course of the 100 realizations, all of which are 4ms long, a total of 41581 impact events were recorded. The band pass filtering at [3kHz, 1MHz] with example results of the peak and minima finding algorithm are shown in Fig. 2. The filtered data taken from the high-pressure sensor is shown in Fig. 6, with some periodic behavior observable. Peaks seem to bunch up every ~50μs or so. The vibratory apparatus oscillates at a frequency of 19.5kHz, which produces a pressure wave with a period of 51.2μs.

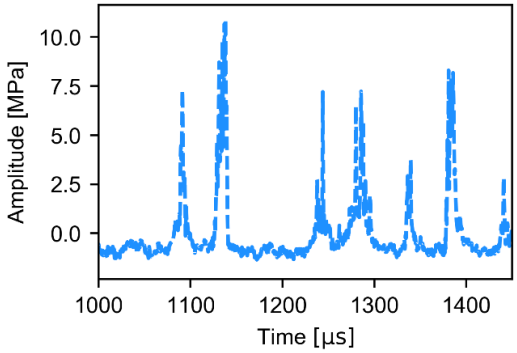
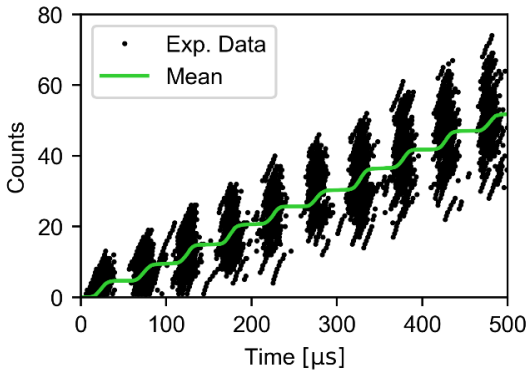


Fig. 6 Periodic behavior of the pressure amplitude. Impacts seem to occur every ~50μs, consistent with the vibratory apparatus' 19.5kHz frequency.

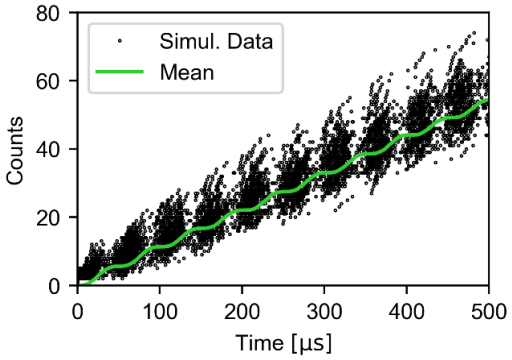
All impacts events from the 100 realizations are shown in Fig. 7(a), with the associated mean oscillating, contrary to the expected linear λt of a Poisson process. This implies that the process is at least non-homogeneous. Further analysis proved that a slightly different function to previously published in [29] is 0.51% more accurate:

$$\lambda(t) = \max(0, A(\sin(2\pi\nu t + \phi))\,) = [A(\sin(2\pi\nu t + \phi))]^{+} \tag{7}$$

with A the amplitude, ν the frequency and the phase ϕ. This function effectively outputs only the positive portion of the sinus wave, and 0 otherwise as necessarily $\lambda(t) \geq 0$. The fitted parameters obtained are $\nu = 19.79\,\text{kHz}$ and $\phi \approx \frac{3\pi}{2}$, $A = 3.266 \times 10^5 \text{s}^{-1}$ which is the average impact rate multiplied by π. In the previously presented function [29], the impact rate over time was equal to A because it was this function's average. The area under the curve of eq. (7) is lesser by a factor of π.



(a) Measured impacts and mean.



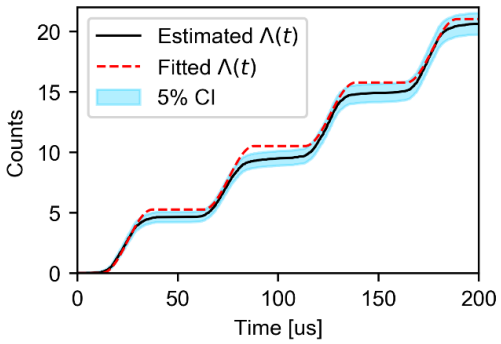
(b) Simulated impacts generated by thinning using a sinusoidal rate function. 100 of realizations were simulated.

Fig. 7 Comparing measured impacts for all 100 realizations with simulations using a sinusoidal rate function.

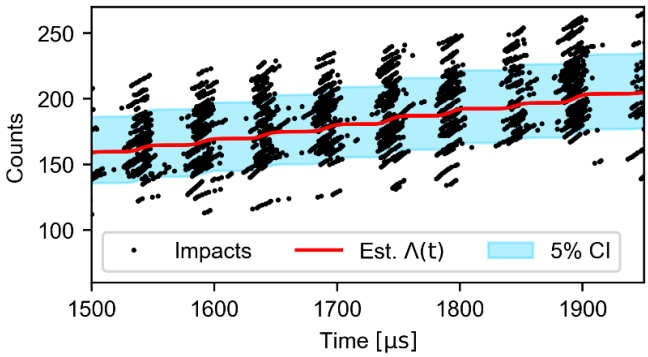
The rate function's frequency is similar to the vibratory apparatus's frequency of 19.5kHz. Pressure waves caused by circular pistons can be modeled as plane waves with frequency equal to that of the piston, in nearfield conditions [30]. The rate function's frequency is expected to be close to the vibration apparatus' if some bubbles undergo cycles of growth and collapse over a pressure wave period. According to [31], this depends on the collapse time of bubbles: if it is longer than a half-period (25μs) then the bubble lifecycle of the bubble can stretch over 2-3 pressure periods.

The cumulative rate function $\Lambda(t)$ is plotted in Fig. 8, which show two confidence intervals (CI) at a significance level of 5%, for the rate function in (a) and for the data points in (b). In Fig. 8, the rate function is outside these CI 8.47% of the time, which is notably high, but 0.51% lower than the previously reported function [29]. The mean of the impact data is a little smoother than the modeled rate function. This implies that some impacts occur during the low-pressure phase and that a minority of bubbles have a collapse time slightly longer than half-period multiples. Also, the experimental data is outside the CI bounds in 31.6% cases, which indicates unexpectedly high variance in the data points.

Contrary to what is expected for a NHPP, the experimental mean and the variance are not equal. The simulated data in Fig. 9(b) shows the expected behavior, while the variance in Fig. 9(a) increases parabolically. This behavior is expected for Poisson process whose rate is not deterministic [20]. For example, the Mixed Poisson processes' rate is randomly distributed, while for the Cox process the rate itself is a stochastic process.



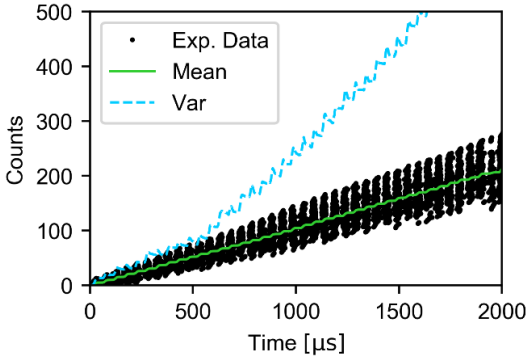
(a) The sinusoidal rate function is outside the 5% confidence intervals 8.47% of the time.



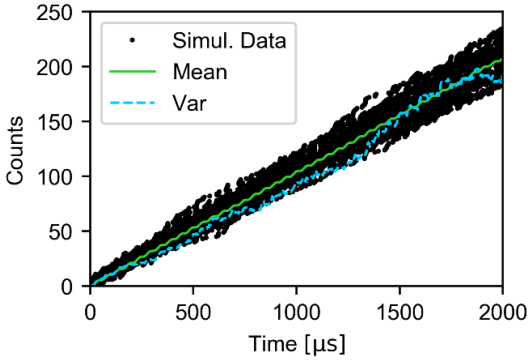
(b) 31.6% of impacts are outside the 5% confidence intervals.

Fig. 8 Confidence intervals to analyze the validity of the temporal non-homogeneous Poisson process' rate function.

Another possibility is the influence of varying external parameters, such as the water temperature or the amount of dissolved gas. The temperature was observed to increase slightly, even with the current temperature control setup. As such, the present dataset cannot discriminate between these possibilities. Future impact measurements will include accurate temperature and dissolved oxygen monitoring, as well as the observation of the influence of the vibration apparatus' parameters on the impact model.



(a) Measured impacts: mean and parabolic variance

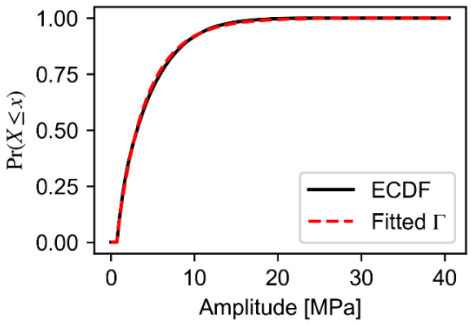


(b) Mean and variance converge for simulated data

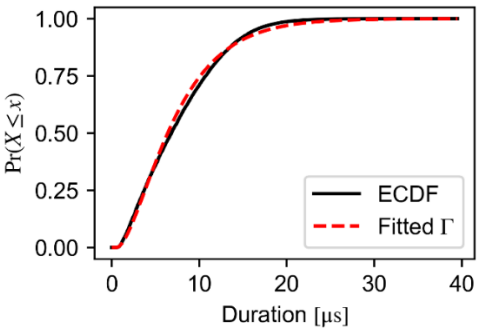
Fig. 9 Over dispersed experimental Poisson process compared with expected behavior.

3.2 Modeling dependent marks using copulas

For every impact event, two marks were measured: the pressure amplitude and the impact duration. An example of the peaks and minima around it used to measure the duration are presented in Fig. 2. A total number of 41581 peaks were detected in the 100 realizations.



(a) Impact amplitude fitted to a Γ distribution $\boldsymbol{k} = 0{,}894$ $\phi = 172mV^{-1}$, $loc = 30.0mV$



(b) Impact duration fitted to a Γ distribution $\boldsymbol{k} = 1.85$ $\phi = 3.83\mu s^{-1}$, $loc = 0.537\mu s$

Fig. 10 Maximum likelihood fitting of marginal distributions using experimental cumulative density functions (ECDF)

The KS test was used, which demonstrated that the gamma distribution is appropriate for both the amplitude and duration distribution at a significance level of 5%. Other distributions are also not rejected, 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 Fig. 10.

The gamma distribution has two parameters, either a shape parameter k and scale parameter φ, or a shape $\alpha = k$ and rate parameter $\beta = 1/\varphi$. The first parametrization was used here, with the additional location parameter loc that describes the shift

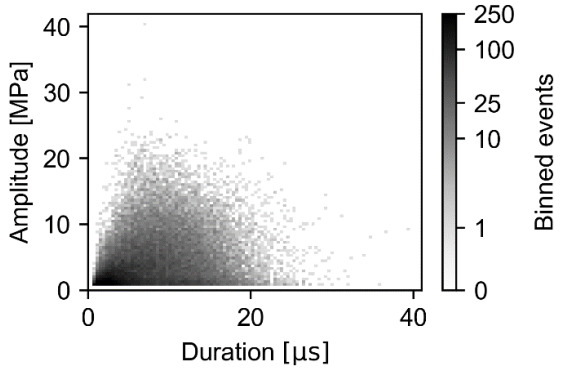
of the random variable compared to 0. With the marginal distributions determined, the bivariate distribution with dependence structure using copulas can be computed.

As a first step to check the data, all the marks were binned in a 2D histogram, as illustrated in Fig. 11. Then, points generated from the joint distribution computed by multiplying the two marginal distributions (valid if independent) were generated to visualize the difference.

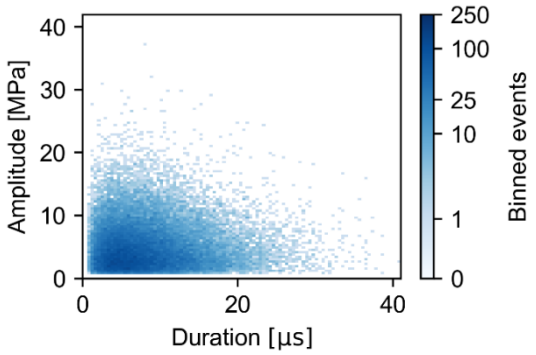
As one can observe, the experimental points are skewed to the high duration region. There are almost no impacts that have a combination of low duration (lower than 5μs) and high amplitude (higher then 500mV or about 12MPa) in the experimental data. This indicates that the random variables are not independent. Also, the highest number of impacts seem to have a combination of low amplitude and short duration, with the number of impacts in higher amplitude and longer duration region decreasing exponentially. It seems that impacts that have higher amplitude also have longer durations in general. Impacts of low amplitude, by contrast have a much wider spread of duration, but still tend to have low duration.

The copula that bests fit the data is the type 2 Tawn copula (with $\psi_2 = 1$) rotated 180°, also known as the survival type 2 Tawn copula, one so-called extreme-value copulas. After the selection process, this copula was fitted to the data, then the copula density and multivariate distribution were computed in Fig. 12. As is visible, the copula is highly asymmetric, skewed to the low 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. With this copula it is possible to generate random samples.



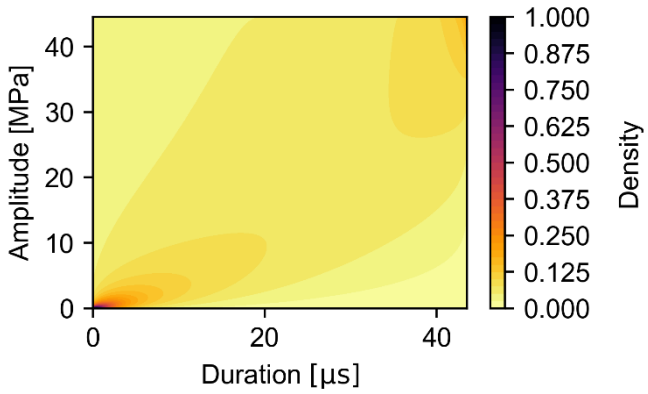
(a) Experimental marks in the data space



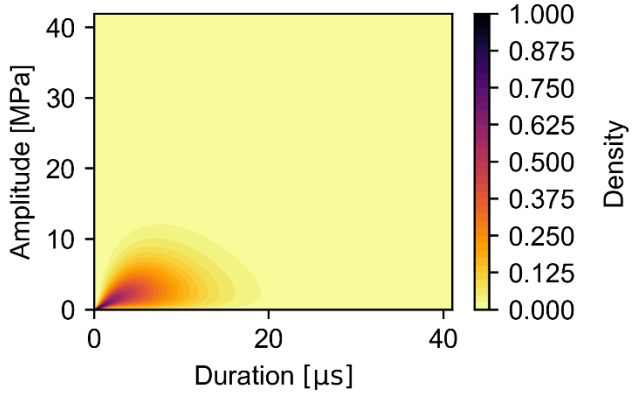
(b) Marks simulated using the independent joint distribution

Fig. 11 Impact amplitude and duration samples marks as 2D histograms. The total number of events are 41581. The bins are approximately 0.4MPa and 0.4μs wide, for a total of 100×100 bins shown

To compare with experimental data, an equal number of samples to the experimental peaks have been generated in Fig. 13. The appearance of the random points more closely matches the experimental data shown in Fig. 11(b). To test for goodness-of-fit, an extension of the Kolmogorov-Smirnov test to two dimensions (2DKS) [33]–[35] was implemented with Python.



(a) Copula density $C(x, y)$



(b) Bivariate distribution $C\big(\Gamma(x), \Gamma(y)\big)$

Fig. 12 Fitted type 2 Tawn Copula $\boldsymbol{\theta} = 1.69$, $\boldsymbol{\psi_1} = 0.370$, $\boldsymbol{\psi_2} = 1$

This test compares two datasets of bi-dimensional data, and once again outputs a statistic which indicates whether 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. This decrease indicates that the

points generated with the copula is harder to distinguish from the experimental data: impact properties are better modeled with dependence.

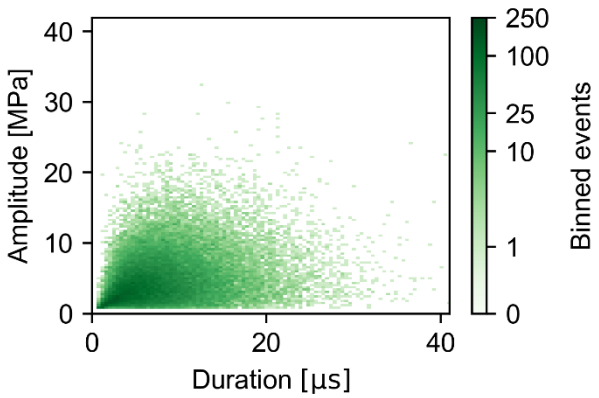


Fig. 13 Two-dimensional histogram of 41581 samples drawn from the fitted survival type 2 Tawn Copula. 2DKS test statistic with experimental data → 7.65%

This completes the construction of a stochastic impact model. The temporal and spatial distributions of impacts were modeled respectively as a non-homogeneous Poisson process and a cluster process. The temporal non-homogeneous Poisson process was over-dispersed, illustrating the need for further experiments with accurate temperature and dissolved oxygen monitoring. Points were simulated using artificial spatial distributions as a proof-of-concept to have a complete picture of the stochastic impact model. Statistical tools such as copulas were used to increase the modelling accuracy of the marks.

The different parts of this model, the marks, the spatial or temporal processes, can all be improved and refined independently, unless dependence is observed. For now, a single set of operating conditions were used to derive this model, but simple empirical relationships can be established with further experiments. For example, the amplitude of the rate function of the temporal Poisson process may well increase linearly with the vibration amplitude. This will be explored in further studies.

4. Conclusion

Cavitation pressure was measured over time using a high-speed pressure sensor. The peaks in detected pressure were treated as points randomly distributed in time and space, to be described using Poisson processes. For each point, the impact amplitude and duration were measured and treated as marks of the Poisson process. This mathematical construction amounts to the stochastic impact model.

- A total of 41581 impacts were recorded over 100 realizations of a 4ms long measurement, for an average impact rate of 1.040×10^5 impacts per second with a high-speed PVDF pressure sensor.
- Analysis of the distribution of impacts in time revealed that the process used to model it must at least be a non-homogeneous Poisson process. The rate function varies over time, and was found to best modeled by:
$$\lambda(t) = \max(0, A(\sin(2\pi\nu t + \phi))\,) = [A(\sin(2\pi\nu t + \phi))]^{+}$$
with parameters: $A = 3.266 \times 10^5 s^{-1}$ which is the average impact rate multiplied by π, $\nu = 19.79$kHz and $\phi \approx \frac{3\pi}{2}$. The impact points were over-dispersed, demonstrating that either the rate of the process itself is random, or that external parameters such as water temperature affected the measurements.
- The spatial distributions were model using pictures of impacts, with gaussian and radially symmetric distributions. Assuming independence, any arbitrary probability density distribution can be combined with the other parts of the stochastic impact model.
- Both marks were appropriately modeled as gamma distributions at a significance level of 5%. These were used as marginals for a copula, used to model the observed dependence. The asymmetric type 2 survival Tawn copula was found most appropriate to model the dependence between the impacts' amplitudes and durations. 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%. It decreased to 7.65% for data simulated using the copula, indicating that the data generated by the copula better fits the experimental data.

Poisson processes and copulas were proven to be useful tools to model cavitation erosion impacts. The stochastic impact model is a rather simple mathematical object compared time-consuming computational fluid dynamics (CFD) currently used to model cavitation. The parameters of this model are fitted empirically to impact data acquired using a high-speed pressure sensor. The acquisition of impact data in any hydraulic machine, pumps, turbines etc. where cavitation impacts occur can be analyzed using the stochastic impact model. Impacts in many circumstances can be modeled using the stochastic impact model.

The stochastic impact model is introduced as an empirical complement to bubble dynamics. This statistics-based approach provides a wide array of tools for analysis and validation. The stochastic model built can be used compute a very high number of impacts with limited computational expense, suited for large time and space scales. The impact data can be combined with finite element models of erosion to compute mass loss predictions. The lack of a theoretical framework and the necessity for many

empirical observations, unique to each cavitation apparatus are disadvantages of this method. In the future, the authors will endeavor to combine the stochastic impact model with material models to predict cavitation erosion

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”.

Nomenclature

$\lambda(t)$	Rate (function) of a Poisson process	$C(a,b)$	A copula.
A	Fitted rate function amplitude	$U(a,b)$	Uniform distribution with parameters a,b
ν	Fitted rate function frequency	u	Random variate for uniform distribution.
ϕ	Fitted rate function phase	P	Pickands function for the Tawn copula.
$N(t)$	Total number of events in interval $[0,t]$	ψ_1, ψ_2	Tawn copula asymmetry parameters
$P[\dots]$	Function that outputs a probability	$\hat{F}_n$	Experimental cumulative distribution function
n	Integer variable.	$\mathbf{1}_A$	Indicator of event A, equals 1 if event A occurs.

References

[1] K.-H. Kim, G. Chahine, J.-P. Franc, and A. Karimi, Advanced Experimental and Numerical Techniques for Cavitation Erosion Prediction, vol. 106. Dordrecht: Springer Science+Business Media, 2014.
[2] R. Singh, S. K. Tiwari, and S. K. Mishra, “Cavitation Erosion in Hydraulic Turbine Components and Mitigation by Coatings: Current Status and Future Needs,” J. Mater. Eng. Perform., vol. 21, no. 7, pp. 1539–1551, Oct. 2011.
[3] S. Hattori, T. Hirose, and K. Sugiyama, “Prediction method for cavitation erosion based on measurement of bubble collapse impact loads,” J. Phys. Conf. Ser., vol. 147, p. 012011, 2009.
[4] J.-P. Franc, M. Riondet, A. Karimi, and G. L. Chahine, “Impact Load Measurements in an Erosive Cavitating Flow,” J. Fluids Eng., vol. 133, no. 12, 2011.
[5] T. Momma and A. Lichtarowicz, “A study of pressures and erosion produced by collapsing cavitation,” Wear, vol. 186187, 1995.
[6] G. L. Chahine, “Characterization of Cavitation Fields From Measured Pressure Signals of Cavitating Jets and Ultrasonic Horns,” vol. 135, no. September 2013, 2018.
[7] H. Soyama and H. Kumano, “The Fundamental Threshold Level — a New Parameter for Predicting Cavitation Erosion Resistance,” J. Test. Eval., pp. 421–431, 2014.
[8] W. D. Callister and D. G. Rethwisch, Materials science and engineering: an introduction, Seventh., vol. 94. John Wiley & Sons, Inc., 2007.
[9] G. R. Johnson and W. H. Cook, “A constitutive model and data for metals subjected to large strains, high strain rates and high temperatures,” in 7th International Symposium on Ballistics, 1983, pp. 541–547.
[10] N. Balakrishnan and C.-D. Lai, Continuous Bivariate Distributions, Second. Springer Science+Business Media, LLC, 2009.
[11] H. Joe, Dependence Modeling with Copulas. CRC Press, 2015.
[12] G. Taillon, K. Onishi, T. Mineshima, and K. Miyagawa, “Statistical analysis of cavitation erosion impacts in a vibratory apparatus with copulas,” in 29th Symposium on Hydraulic Machinery and Systems (IAHR2018), 2018.
[13] ASTM, “G32-Standard Test Method for Cavitation Erosion Using Vibratory Apparatus,” vol. i, no. Reapproved 2010. pp. 1–22, 2012.
[14] J.-K. Choi and G. L. Chahine, “Experimental and Numerical Study of Cavitation Erosion Resistance of a Polyurea Coating Layer,” Fourth Int. Symp. Mar. Propulsors, no. June, 2015.
[15] P. Diggle, Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third., vol. 53, no. 9. CRC Press, 2013.
[16] D. L. Snyder and M. I. Miller, Random Point Processes in Time and Space, Second. Springer-Verlag, 1991.
[17] M. West and J. Harrison, An Introduction to the Theory of Point Processes. Springer Science+Business Media, LLC, 1998.
[18] A. Baddeley, E. Rubak, and R. Turner, Spatial point patterns: methodology and applications with R. CRC Press, 2016.
[19] S. I. Resnick, Adventures in Stochastic Process, 3rd ed. New York: Springer Science+Business Media, LLC, 2014.
[20] J. Grandell, Mixed Poisson Processes, First. Dordrecht: Springer Science+Business Media. B.V., 1997.
[21] M. Jacobsen, Point Process Theory and Applications, vol. 36. Boston: Birkhauser, 2006.
[22] L. M. Leemis, “Nonparametric Estimation of the Cumulative Intensity Function for a Nonhomogeneous Poisson Process,” Manage. Sci., vol. 8, no. 2, pp. 149–168, 2017.
[23] R. Pasupathy, “Generating Nonhomogeneous Poisson Processes,” 2011.
[24] J. C. Isselin, A. P. Alloncle, and M. Autric, “On laser induced single bubble near a solid boundary: Contribution to the understanding of erosion phenomena,” J. Appl. Phys., vol. 84, no. 10, pp. 5766–5771, 1998.
[25] H. Joe, Multivariate models and dependence concepts, First., no. 1960. Dordrecht: Springer, 1997.
[26] R. B. Nelsen, An Introduction to Copulas, Springer S., vol. 42, no. 3. Springer, 2006.
[27] E. L. Lehmann and J. P. Romano, Testing Statistical Hypotheses, Third. New York: Springer Science+Business Media, LLC, 2005.
[28] P. Eschenburg, “Properties of extreme-value copulas,” Technische Universitat Munchen, 2013.
[29] G. Taillon, K. Onishi, S. Saito, and K. Miyagawa, “Stochastic processes to model impact events in a vibratory cavitation erosion

apparatus," in 10th International Symposium on Cavitation (CAV2018), 2018, pp. 953–957.
[30] D. T. Blackstock, "Fundamentals of Physical Acoustics." John Wiley & Sons, 2001.
[31] I. Hansson and K. A. Mørch, "The dynamics of cavity clusters in ultrasonic (vibratory) cavitation erosion," J. Appl. Phys., vol. 51, no. 9, pp. 4651–4658, 1980.
[32] J.-P. Franc and J.-M. Michel, Fundamentals of Cavitation, First Edit. Dordrecht: Kluwer, 2005.
[33] J. A. Peacock, "Two-dimensional goodness-of-fit testing in astronomy," Mon. Not. R. Astron. Soc., vol. 202, pp. 615–627, 1983.
[34] G. Fasano and A. Fransceschini, "A multidimensional version of the Kolmogorov-Smirnov test," Mon. Not. R. Astron. Soc., vol. 225, pp. 155–170, 1987.
[35] W. Press et al., Numerical Recipes in C: The Art of Scientific Computing, vol. 29, no. 4. Cambridge: Cambridge University Press, 1987.