Direct counting method and its validation

Vladimir Shigunov, *DNV GL SE, Hamburg, Germany,* vladimir.shigunov@dnvgl.com

ABSTRACT

Direct counting of stability failures, required to define failure rate (or probability) in direct stability assessment and operational measures, is possible for a stationary Poisson process. This paper proposes measures to ensure that numerical simulations or model tests satisfy the requirements of such process and develops efficient procedure, illustrated in application examples, to define the 95%-confidence interval of failure rate, which is required for assessment.

Keywords: *Poisson process, Failure rate, Direct stability assessment.*

1. INTRODUCTION

Direct counting, i.e. counting of the number of stability failures per given exposure time, is used in the probabilistic direct stability assessment (DSA) in design situations, as well as (in combination with statistical extrapolation, which itself uses direct counting) in the full probabilistic DSA and in operational measures (OM). Such counting is based on the model of a stationary Poisson process (frequently without knowing this), which raises two questions: first, whether real stability failures can be assumed a Poisson process and if they can, second, how should numerical simulations or model tests be performed to ensure that the resulting process is a Poisson process.

Stationarity seem questionable since ships are sailing in conditions (sea state, ship speed and course) that often change quicker than the time required for roll motion to be considered stationary; in some applications (e.g. transient flooding), also the hydrodynamic model changes too quickly for the stationarity assumption. This apparent contradiction is easy to resolve since design assessment (and operational measures) do not consider a ship operating in a sequence of changing sea states, forward speeds and courses but consider ensemble statistics over a large number of ships, each of which operates unlimited time in stationary conditions.

A Poisson process requires that failures do not occur simultaneously and that failures happen independently, i.e. that the occurrence of one failure does not affect the probability of occurrence of a second failure. The former requirement seems trivial, while the latter is known to be not valid for roll motion since big roll motions tend to appear in groups. The justification is the *clumping heuristic*: although big roll motions tend to appear in groups, occurrence of such groups may be independent, which, in turn, is based on the heuristic that rare events tend to be independent. However, no strict proof is possible.

This paper summarises properties of a Poisson process, illustrates the validity of the Poisson process assumption and proposes direct counting procedures to define the boundaries of the 95%-confidence interval of the failure rate, which is used in the probabilistic DSA and OM.

2. POISSON PROCESS

A *counting process* is a stochastic process $N(t)$, where the integer random variable N counts the total number of some defined events, e.g. stability failures (or more general, exceedances of some reaction level), tasks, phone calls, customers etc. (the usual term in mathematics is *arrivals*) that have occurred in the time interval from 0 up to and including time t. The number of arrivals per time interval $(s,t]$, $N(t)-N(s)$, is called an *increment*. Consider counting processes in which increments are *independent* (i.e. numbers of arrivals in non-overlapping time intervals are independent) and *stationary* (i.e. the number of arrivals depends only on the length of a time interval and not on its location in time).

There are several equivalent definitions of a Poisson process; one, which is convenient for mathematical derivations, defines a Poisson process with a rate $r>0$ as a counting process $N(t)$ which has stationary independent increments and satisfies the following requirements:

$$p\{N(\mathrm{d}t)=1\}=r\,\mathrm{d}t+o(\mathrm{d}t) \qquad (1)$$

$$p\{N(\mathrm{d}t) \geq 2\} = o(\mathrm{d}t) \tag{2}$$

$$p\{N(\mathrm{d}t) = 0\} = 1 - r\,\mathrm{d}t + o(\mathrm{d}t) \tag{3}$$

where the Landau's notation $o(x)$ means that if $f(x)=o(x)$, then $\lim_{x\to 0} f(x)/x = 0$.

Another definition, more useful in applications, defines a Poisson process with a constant rate $r>0$ as a counting process $N(t)$ which has independent increments and where the number of arrivals $N(t)$ in any time interval of length t satisfies the Poisson distribution with the mean rt, i.e.

$$p\{N(t)=k\} = (rt)^k \cdot e^{-rt}/k! \text{ for } k=0, 1, ... \tag{4}$$

(it automatically implies stationary increments).

A Poisson process has several useful properties. The *superposition property* (which is convenient for summing failure rates due to different failure modes) means that a sum of independent Poisson processes $N_1, \ldots, N_k$, i.e. $N_1+\cdots+N_k$, is a Poisson process with the rate $r_1+\cdots+r_k$ (conversely, if the sum of two independent random variables is Poisson distributed, so are each of these two variables). The *random split property* means that if each arrival of a Poisson process $N(t)$ with rate r is randomly tagged as either process $N_1(t)$, with probability p, or $N_2(t)$, with probability $1-p$, then the two resulting processes $N_1(t)$ and $N_2(t)$ are independent Poisson processes with rates rp and $r(1-p)$, respectively. Similarly, the *thinning property* means: if each arrival of a Poisson process with rate r is randomly marked, with probability p, then the marked process is a Poisson process with rate rp. In both the random split and thinning properties, the probability p can depend on time as $p(t)$, then the above properties are valid assuming that $p = (1/t)\int_0^t p(u)\mathrm{d}u$.

The probability density function of the Poisson distribution $f(k)=p\{N(t)=k\}$, eq. (4), expresses the probability of occurrence of k arrivals during a time interval t. A special case of eq. (4) is when $k=0$ (which corresponds to the probability that no failures occur from time 0 to time t):

$$p \equiv p\{N(t)=0\} = e^{-rt} \tag{5}$$

From eq. (5), the probability that at least one failure happens during time t, i.e. that $k>0$, (loosely formulated: "probability of stability failure during time t") is

$$\begin{aligned} p^* &\equiv p\{N(t)>0\} = 1 - p\{N(t)=0\} \\ &= 1 - p = 1 - e^{-rt} \end{aligned} \tag{6}$$

For small rt, p^* from eq. (6) can be linearised as

$$p^* \approx rt \tag{7}$$

The mean of a Poisson process, i.e. the mean number of arrivals per interval t, is

$$E\{N(t)\} = \int_0^\infty t f(t)\mathrm{d}t = rt \tag{8}$$

i.e. r is equal to the expected number of arrivals per time unit. The variance of the Poisson process is equal to the mean, $Var\{N(t)\}=rt$.

Instead of looking at the number of arrivals $N(t)$, a Poisson process can be seen as a sequence of time intervals T_1 (time between $t=0$ and the first arrival), T_2 (time between the first and second arrival) etc., which are also random variables.

Note that the probability that the time until the first arrival exceeds t, i.e. that $p\{T_1>t\}$, is the same as $p\{N(t)=0\}=e^{-rt}$, eq. (5), i.e. T_1 is exponentially distributed (similarly, it can be shown that all time intervals T_i are exponentially distributed random variables with the same rate r).

Therefore, a Poisson process can also be defined as a counting process $N(t)$ in which time intervals between arrivals are independent random variables, exponentially distributed with rate r:

$$p\{T>t\} = e^{-rt} \text{ for } t>0 \text{ and } 0 \text{ otherwise} \tag{9}$$

(note that this definition automatically implies independent and stationary increments).

An important property of the exponential distribution is its *memoryless property*: a memoryless random variable T is a variable for which

$$p\{T>t+\tau\} = p\{T>t\} \cdot p\{T>\tau\}, \tag{10}$$

which also can be written in the following form: $p\{T>t+\tau | T>t\} = p\{T>t\}$, which means that if an arrival has not occurred until time τ, the distribution of the remaining waiting time is the same as the distribution of the original waiting time, which means that the remaining waiting time has no memory of previous waiting time.

Moreover, the exponential distribution is the only continuous distribution with the memoryless property: if the time intervals between arrivals are not exponential, the process will not be a Poisson process since it does not satisfy the requirements of stationary and independent increments. The memoryless property of the exponential distribution is what makes the Poisson process unique (sometimes called a *purely* or *completely random process*).

Correspondingly to a sum of Poisson processes, if $T_1, \ldots, T_k$ are independent exponentially distributed random variables with rates $r_1, \ldots, r_k$, then $\min(T_1, \ldots, T_k)$ is exponentially distributed with rate $r_1+\cdots+r_k$. The index of the variable that achieves the minimum is distributed according to the law $p\{i \,|\, T_i=\min(T_1,\ldots,T_k)\}=r_i/(r_1+\cdots+r_k)$.

From eq. (9), the cumulative density function of time to failure is

$$F(t)=p\{0<T<t\}=1-p\{T>t\}=1-e^{-rt} \tag{11}$$

for $t>0$ and 0 otherwise. In the reliability theory, the *survival function* is frequently used, defined as $R(t)=p\{T>t\}=1-F(t)=e^{-rt}$.

The probability density function of an exponential distribution, i.e. of the time intervals between arrivals in a Poisson process, is

$$f(t)=\mathrm{d}F(t)/\mathrm{d}t=re^{-rt} \text{ for } t>0 \text{ and 0 otherwise} \tag{12}$$

The mean of an exponentially distributed random variable T (i.e. mean time between arrivals) is

$$E\{T\} \equiv \overline{T} = \int_0^\infty tre^{-rt}\mathrm{d}t = 1/r \tag{13}$$

The second moment $E\{T^2\} = \int_0^\infty t^2re^{-rt}\mathrm{d}t = 2/r^2$, and thus the variance and the standard deviation are

$$Var\{T\}=E\{T^2\}-E^2\{T\}=1/r^2 \tag{14}$$

$$\sigma\{T\}=(Var\{T\})^{1/2}=1/r=\overline{T} \tag{15}$$

respectively, i.e. $\sigma\{T\}$ is equal to the mean. Shigunov (2016, 2017a) verified eq. (15) by numerical simulations, Fig. 1.

To define parameters of distributions from a series of numerical simulations or model tests, time intervals T_i to each failure should be defined; for a set of N values T_i, i=1, 2, …, N, the *sample mean time to failure* is

$$\hat{T}=(1/N)\sum_{i=1}^N T_i \tag{16}$$

and the *maximum likelihood estimate of the stability failure rate* is

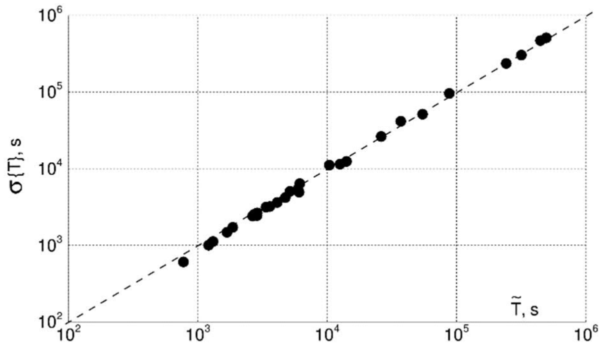

Figure 1: An estimate of the standard deviation of the time to failure vs. estimate of the mean time to failure from 200 simulated failures (from Shigunov, 2017a).

$$\hat{r}=1/\hat{T} \tag{17}$$

The variance of the individual time intervals T_i to stability failure is

$$Var\{T_i\}=\overline{T}^2 \tag{18}$$

thus the standard deviation of the individual times to stability failure is

$$\sigma\{T_i\}=(Var\{T_i\})^{1/2}=\overline{T} \tag{19}$$

To estimate the confidence interval for the rate, Shigunov (2017) proposed an approximation based on the central limit theorem, according to which, for a sufficiently big sample, the sample mean can be assumed normally distributed with the standard deviation $\sigma_T=\sigma\{T_i\}N^{-1/2}$, i.e. $\sigma_T=\overline{T}N^{-1/2}$, eq. (19). Then, a 95%-confidence interval for the failure rate can be defined as

$$\hat{r}(1-1.96N^{-1/2})<r<\hat{r}(1+1.96N^{-1/2}) \tag{20}$$

Fig. 2, based on numerical simulations of several cases of parametric (P1, P2) and synchronous (S1 to S4) resonance, Shigunov (2017), confirms that $\sigma_T/\sigma\{T_i\}$ really decreases as $N^{-1/2}$ with increasing N, and that this happens already for small sample sizes N.

Here, as a more general approach, it is proposed to define the $(1-\alpha)\cdot 100\%$-confidence interval (note that α denotes some small value: e.g. for a 95%-confidence interval, α=0.05) for the rate r of an exponential distribution as

$$\frac{2N}{\hat{r}\chi^2_{1-\alpha/2,2N}}<\frac{1}{r}<\frac{2N}{\hat{r}\chi^2_{\alpha/2,2N}} \tag{21}$$

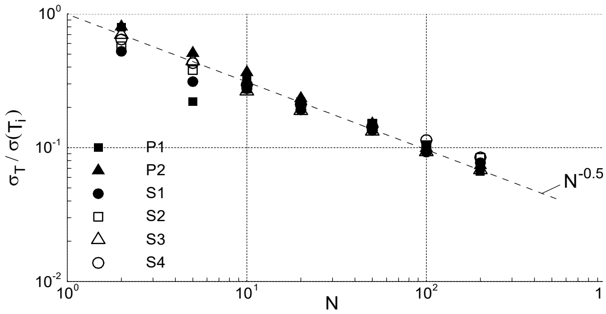

Figure 2: The standard deviation of the estimate of the mean time to failure vs. the number of counted failures from Shigunov (2017)

or, resolving eq. (21) with respect to r,

$$0.5\chi^2_{\alpha/2,2N}/N < r/\hat{r} < 0.5\chi^2_{1-\alpha/2,2N}/N \tag{22}$$

where $\chi^2_{p,f}$ is the quantile function (corresponding to a lower tail area, equal to the cumulative probability p) of the χ^2 distribution with f degrees of freedom. The function $\chi^2_{p,f}$ is available in many software packages: e.g. in MS Excel, it can be calculated as chisq.inv(p;f) or, alternatively, as chisq.inv.rt(1-p;f). Correspondingly, the upper-tail quantile function $\chi^2_{1-p,f}$ can be calculated as chisq.inv.rt(p;f) or chisq.inv(1-p;f), Fig. 3. Eqs. (20) and (22) are compared in section 5.

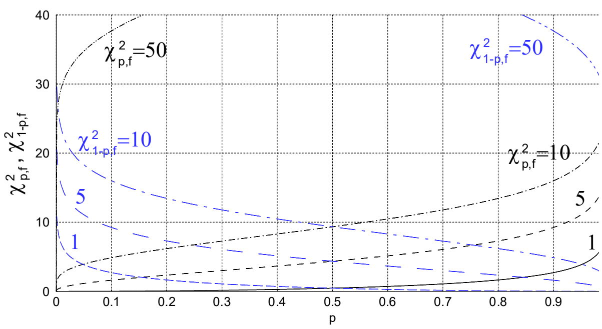

Figure 3: Examples of functions $\chi^2_{p,f}$ and $\chi^2_{1-p,f}$, calculated with MS Excel as chisq.inv(p;f) and chisq.inv.rt(p;f), respectively, for $f = 1$, 5, 10 and 50 degrees of freedom vs. cumulative probability p

3. NON-STATIONARY PROCESSES

Section 2 considered the Poisson process with a constant rate r, which is called a *stationary* (or *homogeneous*) Poisson process. The definition of the Poisson process can be generalised to a *non-stationary* (*non-homogeneous*) process by letting the rate to change in time (which happens due to changing sea state, forward speed and course) as $r(t)$: a non-stationary Poisson process is a counting process $N(t)$ which has independent increments and, for any $dt > 0$, satisfies the following conditions:

$$p\{N(t+dt)-N(t)=1\} = r(t)\,dt + o(dt) \tag{23}$$

$$p\{N(t+dt)-N(t)\geq 2\} = o(dt) \tag{24}$$

$$p\{N(t+dt)-N(t)=0\} = 1 - r(t)\,dt + o(dt) \tag{25}$$

The mean of a non-stationary Poisson process (i.e. mean number of arrivals) per infinitesimally small time interval $(t, t+dt]$, can be calculated as $E\{N(t+dt)-N(t)\} = \sum_{k=0}^{\infty} k \cdot p\{N(t+dt)-N(t)=k\}$ $= r(t)dt + o(dt)$, and, therefore, the mean number of arrivals in a finite time interval from t to $t+\tau$ is

$$\begin{aligned} E\{N(t+\tau)-N(t)\} &= E\left\{\int_t^{t+\tau}[N(u+du)-N(u)]\right\} \\ &= \int_t^{t+\tau} E\{N(u+du)-N(u)\} = \int_t^{t+\tau} r(u)du = \bar{r}\tau \end{aligned} \tag{26}$$

where

$$\bar{r} = (1/\tau)\int_t^{t+\tau} r(u)du \tag{27}$$

which is the average rate in the time interval $(t, t+\tau]$. Comparison of eq. (26) with eq. (8) shows that $N(t+\tau)-N(t)$ is a Poisson-distributed random variable with the mean $\bar{r}\tau$ and the rate $\bar{r}$, i.e. all properties of a Poisson distribution apply by using the mean rate $\bar{r}$. In particular, the distribution of the number of arrivals in a time interval $(t, t+\tau]$, i.e. the probability of k arrivals in this time interval, is

$$p\{N(t+\tau)-N(t)=k\} = (\bar{r}\tau)^k e^{-\bar{r}\tau}/k! \tag{28}$$

Similarly, the probability that no failures occur from time t to time $t+\tau$ is

$$p \equiv p\{N(t+\tau)-N(t)=0\} = e^{-\bar{r}\tau} \tag{29}$$

and the probability that at least one failure happens between time t and time $t+\tau$ ("probability of failure during time τ") is

$$p^* \equiv p\{N(t+\tau)-N(t)>0\} = 1-e^{-\bar{r}\tau} \tag{30}$$

One way of looking at long-term statistics is to consider the ship as operating a "long time" in a changing sequence of stationary situations, each of which is specified by a stationary sea state and constant ship forward speed and course. This allows considering the "long-term" operation as a non-stationary Poisson process with variable failure rate $r(t)$, thus allowing using the usual formulae like eq. (28)-(30), where the mean "long-term" rate $\bar{r}$ is, according to eq. (27),

$$\bar{r} = (1/t_o)\int_0^{t_o} r(u)du \approx (1/t_o)\sum_i r_i t_i = \sum_i r_i p_i \tag{31}$$

where t_i is the duration of each stationary "short-term" situation (specified by a stationary sea state and constant ship forward speed and course), t_o is the total time at sea and $p_i = t_i / t_o$ are the probabilities of each stationary "short-term" situation.

A more correct treatment is to consider the "long-term" operation as a stationary Poisson process with a constant rate $\bar{r}$, consisting of an infinite number of stationary Poisson processes with constant rates r_i, each of which happens in stationary "short-term" conditions with probability p_i; then, the application of either the splitting property of a stationary Poisson process or, alternatively, sum and tagging properties, leads to the same formula (30).

Thus, the problem reduces to the definition of a constant rate r_i in each stationary situation from numerical simulations or model tests in a way ensuring a stationary Poisson process.

4. PRACTICAL PROBLEMS

In practice, ensemble statistics in stationary conditions is replaced by time history statistics using the ergodicity property. Thus, in principle, it is required to run one, long enough, simulation of roll motion, updating (after each failure) the number N of failures, the total simulation time t_t, the estimate of failure rate

$$\hat{r} = N / t_t \tag{32}$$

and the estimate of the upper boundary r_U of the 95%-confidence interval of failure rate, eq. (22),

$$r_U = 0.5 \hat{r} \chi_{0.05/2,2N} / N \tag{33}$$

Once r_U reduces below the acceptance standard, the simulation can be stopped and the loading condition can be considered acceptable.

However, if the sea state is modelled, following Longuet-Higgins (1952), as a finite sum of harmonic components,

$$\zeta(t) = \sum_{i=1}^{M} a_i \cos(\omega_i t + \varepsilon_i) \tag{34}$$

where $a_i = \{2 S_{\zeta\zeta}(\omega_i) D(\mu_i) \Delta\omega_i \Delta\mu_i\}^{1/2}$ are amplitudes, ω_i frequencies, μ_i directions and ε_i phases of components, collecting sufficient statistics in one sufficiently long run is impossible because of self-repetition effects, which violate the requirement of independence of failures (note that exact repetition is not required: a partial repetition in a relevant band of encounter frequencies is bad enough). A solution is to generate multiple independent *realisations* of the same sea state, randomly varying phases ε_i in the interval $[0,2\pi)$ for each realisation, and to simulate ship motions for a limited time in each such realisation (in model tests, the duration of each realisation is limited also by wave reflection effects). Note that eq. (32) still can be used, so that N and t_t are the total number of failures and total simulation time, respectively, over all realisations.

In the examples in this paper, also frequencies ω_i and directions μ_i of components were randomly varied between realisations (using normal distributions with specified standard deviations). Some authors recommend also using a random variation of amplitudes a_i. To generate random values, pseudo-random number generators are frequently used, which apply a user-specified integer as a seed number; in the examples in this paper, the internal computer timer was used as the seed number. Short-crested seaway with a JONSWAP wave energy spectrum was used (with the peak parameter equal to 3.3 and the $\cos^2$-wave energy spreading), discretised by 19 wave directions with 10^3 harmonic components of equal amplitudes per direction.

Another problem is the transient hydrodynamic effects at the beginning of each simulation, which violate the stationarity requirement. To address this, some time after the start of each simulation (in the examples in this paper, 50 roll periods were used) was excluded from post-processing, i.e. not included in t_t, and failures during this time were not counted in N in eq. (32). Note that whereas it is useful to increase the number and reduce duration of random realisations using parallel simulations, this will also increase wasted simulation time.

Finally, the independence of stability failures in numerical simulations is violated by autocorrelation of large roll motions: large roll amplitudes tend to appear in groups (note that neglecting this effect would lead to an over-estimation of the stability failure rate, i.e. a conservative error). One way to neutralize this effect is to run each simulation up to the first encountered failure; another way is to switch off both the simulation timer t_t and the failure counter N after an encountered failure until the envelope of the autocorrelation of roll motion reduces to a specified level. These two options lead to similar unproductive losses: in the former method, due to transient effects at the start of a new simulation, in the latter, for the decay of the autocorrelation function of roll motion. However, the former method (which was used in the examples in this paper) is simpler and, besides, another benefit is that restarting also takes care of repeatability effects.

After removing portions of time histories of roll motion affected by repeatability effects, initial transients and stability failures, the remaining pieces can be treated as a single stationary Poisson process: removed pieces do not matter due to the memoryless property, durations of the remaining pieces may be arbitrary, and it does not matter whether a failure was encountered in each simulation or not, Fig. 4.hus, eq. (32) can be used for the most likelihood estimate of the failure rate, with N and t_t summed over all remaining pieces of the simulations. Similarly, the sample mean time to failure can be calculated as $\hat{T} = t_t / N$, eq. (16), and all formulae from section 2 can be directly applied.

To investigate repeatability effects, parametric resonance in head waves and synchronous resonance in beam waves were simulated for a systematically varied significant wave height in two types of simulations. In one, denoted 'limited' for brevity, the simulation time was limited to 3 hours (while simulations were stopped after first failure), and in the other ('unlimited'), simulations were run always until first failure. Thus, after each 'limited' simulation, N was increased by 1 in case of failure (t_t was increased by the time to failure) and by 0 if simulation ended in 3 hours without failure (t_t was increased by 3 hours). After each 'unlimited' simulation, N was increased by 1 and t_t was increased by the time to failure. Fig. 5 compares the sample mean time to failure from these simulations, plotted vs. the number of failures, together with the boundaries of the 95% confidence interval of the mean time to failure, eq. (22), for examples with mean time to failure above and below 2 hours.

Fig. 5 shows no clear difference; for more insight, Fig. 6 shows quantile diagrams (QQ diagrams) derived from these simulations. Since the cumulative density function of an exponentially distributed time to failure is $F(t) = 1 - e^{-rt}$ for $t > 0$, $T_i / \bar{T}$ should be equal to $-\ln(1 - F_i)$. The cumulative density function F_i was calculated from the sample data as $i/(N+1)$, where i is the index of a stability failure when stability failures are sorted in ascending order of T_i. Fig. 6 shows that the 'unlimited' simulations over-estimate failure time compared to the exponential distribution for simulations with large time to failure due to self-repetition effects: the same 'uncritical' realisation repeats itself again and again (since repetition is not exact, failure may eventually happen but much later than it should).

This means that simulations that are too long lead to deviation from the Poisson process in relevant sea states, i.e. the notion of failure rate and the formulae from section 2 are not applicable (note that using these formulae as if the process were Poisson would lead to an under-estimation of the failure rate, i.e. a non-conservative error). Thus, the maximum duration of simulations should be limited: when at least 10^3 frequencies per wave direction are used, simulations up to 3 hours seem acceptable. More general recommendations require further studies; until that, quantile plots can be used to verify the absence of repetition effects.

To check whether the considered measures are sufficient to assume the Poisson process, the χ^2 goodness-of-fit test was applied to several cases of parametric and synchronous resonance in head and beam, respectively, waves at systematically varied significant wave height, using 'limited' simulations of 3-hour duration (or until failure if it happened).

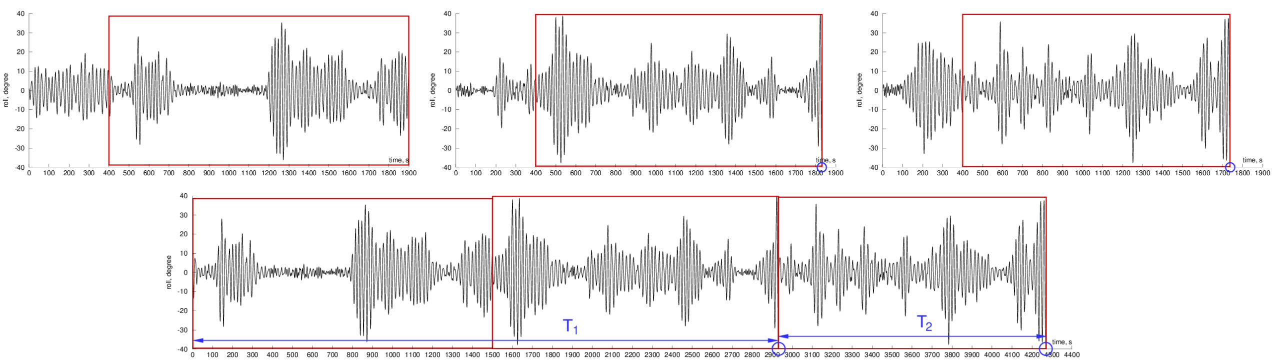

Figure 4: Roll motions in multiple realisations of a sea state (top) and the resulting Poisson process (bottom)

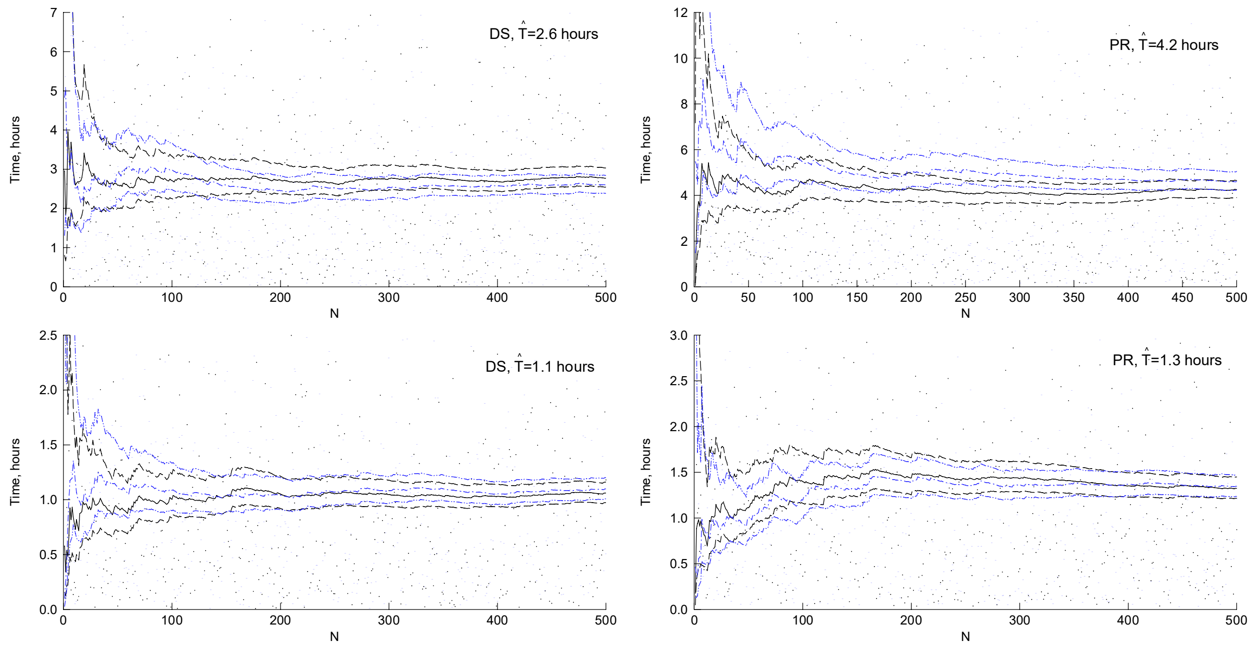

Figure 5: Time until failure (dots) and sample mean time to failure (lines) from 'limited' (solid black lines) and 'unlimited' (dash-dot lines) simulations together with bounds of the 95%-confidence interval of mean time to failure vs. the number of counted failures for synchronous (left) and parametric (right) resonance cases

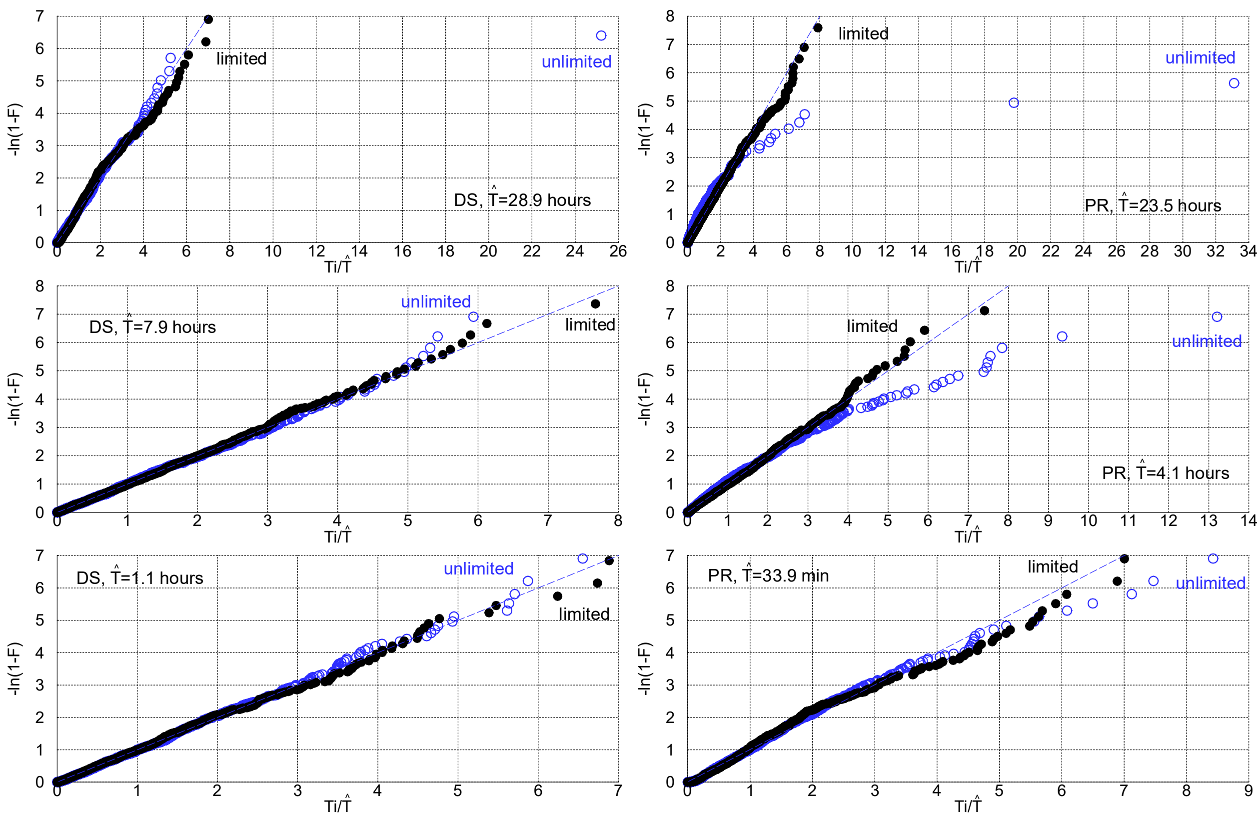

Figure 6: Quantile diagrams from 'limited' and 'unlimited' simulations (black full and empty blue circles, respectively) for synchronous (left) and parametric (right) resonance cases

Random realisations of the same sea state were repeated until about 10^3 failures were encountered in each sea state. Observed times to failure were compared with the exponential distribution (which used the most likelihood estimate $\hat{r}=1/\hat{T}$ for the failure rate). The full range $t \geq 0$ of time to failure was sub-divided into $k \geq 5$ intervals of equal probability $\Delta F = \Delta(1-e^{-rt}) = 1/k$; the number of intervals was systematically increased up to a maximum $k = N/5$.

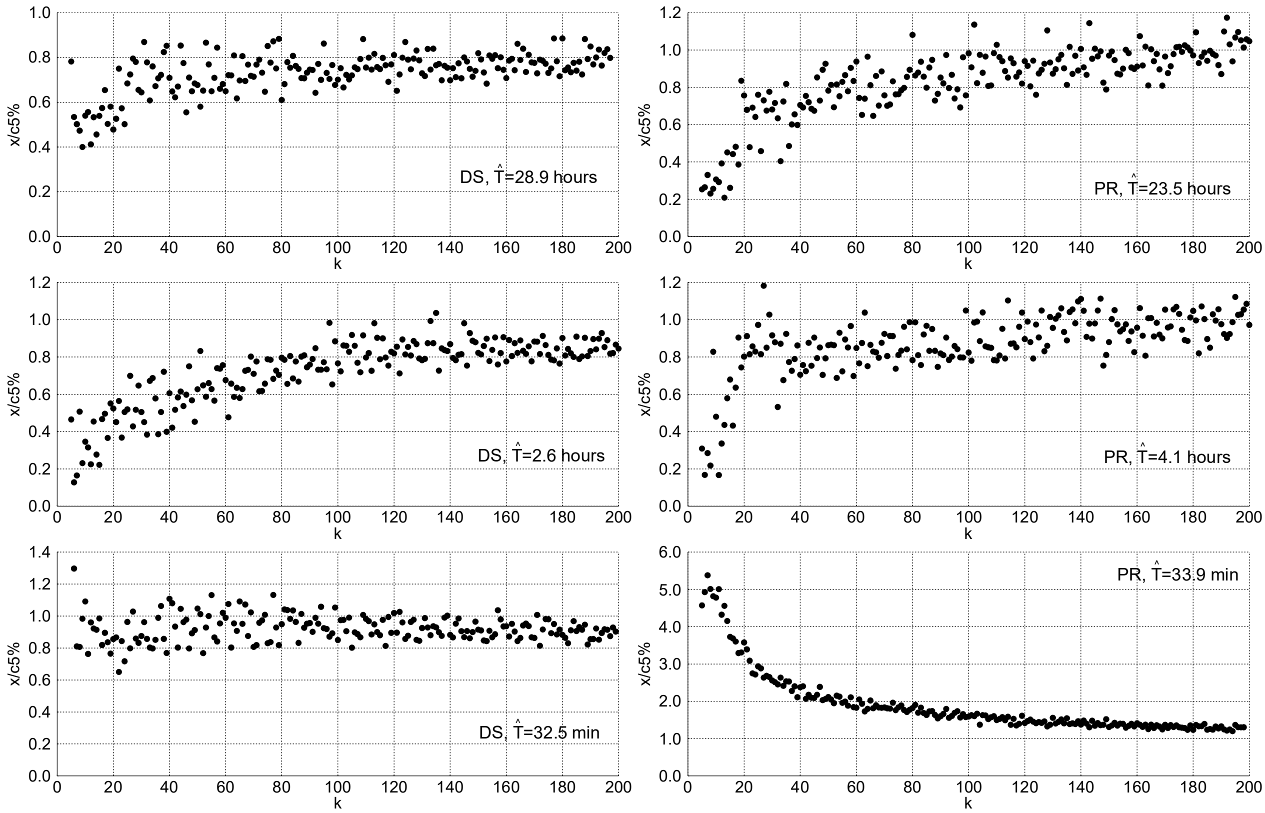

Figure 7: The ratio of χ^2-test statistic to critical value $\chi^2_{1-\alpha,k-2}$ for significance level α=5% vs. number of intervals k of time to failure for synchronous (left) and parametric (right) resonance for varied mean time to failure

The number O_i of the observed times to failure within each interval i was counted, and the expected number E_i was calculated, according to the assumed distribution, as N/k; then, the *test statistic* was calculated as

$$x = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i \tag{34}$$

The *critical value* of the test statistic, at the *significance level* α=0.05, was defined as $c_{5\%} = \chi^2_{1-\alpha,f}$, i.e. the value of the χ^2 distribution at the cumulative probability 1 - α=0.95 with the number of degrees of freedom $f=k-p-1$, where $p=1$ is the number of parameters of the assumed distribution estimated from the sample. Fig. 7 shows ratio $x/c_{5\%}$ vs. the number of intervals k: when $x/c_{5\%}<1$, the *null hypothesis* that the data follow the assumed distribution cannot be rejected at the significance level 5%. Fig. 8 shows ratio $x/c_{5\%}$ at k=200 as a function of the sample mean time to failure. For synchronous resonance, the Poisson process model is acceptable (at the 5% significance level) in all studied cases. On the other hand, for parametric resonance, the results disagree with the Poisson process assumption: marginally at $\hat{T} \approx 2$ hours and greater and increasingly for $\hat{T}$ decreasing below 2 hours. Note, however, that the χ^2 test is considered as very strict when the amount of data is large.

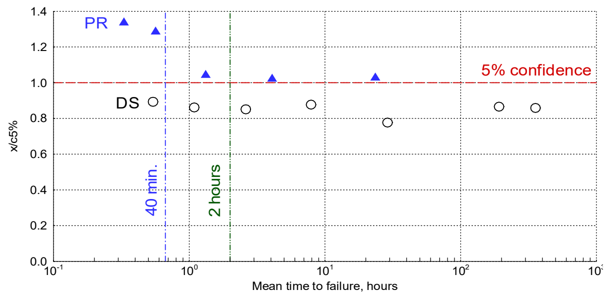

Figure 8: The ratio $x/c_{5\%}$ at k=200 for synchronous (○) and parametric (▲) resonance vs. sample mean time to failure

5. APPLICATION

In the *probabilistic assessment in design situations*, Annex 1 of IMO (2019), section 5.3.3, the *acceptance* requirement is that in all design sea states the upper boundary of the 95%-confidence interval of the failure rate should not exceed a *standard* λ (note that since the remaining 5% outside of the confidence interval include both tails, this means that failure does not happen with the probability 97.5%); the standard is equal to one

Table 1. Significant wave heights for design sea states with probability density 10^{-5} (m·s)$^{-1}$ for unrestricted service, according to Annex 1 of IMO (2019), section 5.3.3

T_z, s	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5	12.5	13.5	14.5	15.5	16.5
h_s, m	2.8	5.5	8.2	10.6	12.5	13.8	14.6	15.1	15.1	14.8	14.1	12.9	10.9

failure in 2 hours in sea states with probability density 10^{-5} (m·s)$^{-1}$. For parametric roll assessment, design situations are defined by 0° and 180° mean wave directions, zero forward speed, all wave periods and significant wave heights vs. zero-upcrossing period per Table 1.

This means that for the acceptance of a loading condition, the upper boundary of the 95%-confidence interval of the failure rate should be estimated and compared with the standard λ in all design situations. To save simulation (or model testing) time, it is useful to estimate also the lower boundary of the 95%-confidence interval of the failure rate and stop further simulations or tests (and consider the loading condition as *unacceptable*) once this estimate exceeds the standard λ in at least one design situation.

The upper r_U and lower r_L boundaries of the 95%-confidence interval of the failure rate can be estimated from eq. (22) as

$$r_U = 0.5\chi^2_{1-\alpha/2,2N}\hat{r}/N \qquad (36)$$

$$r_L = 0.5\chi^2_{\alpha/2,2N}\hat{r}/N \qquad (37)$$

where N is the number of counted stability failures, $\alpha = 1-95/100 = 0.05$ and $\hat{r} = 1/\hat{T}$.

For *acceptance*, it should be required that r_U is less than λ, which leads to the following condition:

$$\hat{T} > \hat{T}_A \equiv \beta_1(N)/\lambda \qquad (38)$$

where $1/\lambda = 2$ hours and $\beta_1(N) = 0.5\chi^2_{1-\alpha/2,2N}/N$; β_1 is shown as a function of N in Fig. 9; $\chi^2_{1-\alpha/2,2N}$ was calculated with MS Excel as chisq.inv.rt(α/2;2*N).

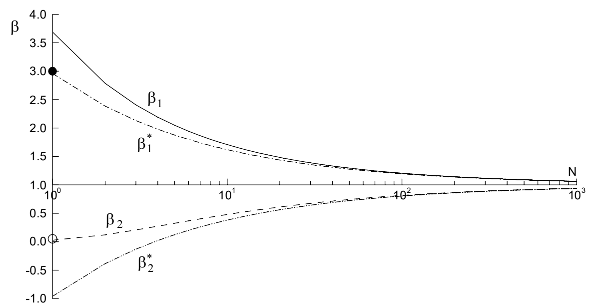

Figure 9: χ^2 vs. normal approximation for r_U and r_L: β_1, β_2, β_1^* and β_2^* together with approximations for N=1 from Shigunov (2016), 3/λ (●) and 0.05/λ (○)

For *not acceptance*, it can be required that r_L is greater than λ, which leads to the condition

$$\hat{T} < \hat{T}_F \equiv \beta_2(N)/\lambda \qquad (39)$$

where $\beta_2(N) = 0.5\chi^2_{\alpha/2,2N}/N$, Fig. 9; $\chi^2_{\alpha/2,2N}$ was calculated with MS Excel as chisq.inv(α/2;2*N).

In Shigunov (2016), conditions of *acceptance*

$$\hat{T} > \beta_1^*(N)/\lambda \qquad (40)$$

and *not acceptance*

$$\hat{T} < \beta_2^*(N)/\lambda \qquad (41)$$

respectively, were derived using the central limit theorem (i.e. normal distribution approximation for the sample mean). For the estimate of the 95%-confidence interval for the failure rate, eq. (20), this leads to $\beta_1^*(N) = 1+1.96N^{-1/2}$, $\beta_2^*(N) = 1-1.96N^{-1/2}$; Fig. 9 shows $\beta_1^*(N)$ and $\beta_2^*(N)$ for comparison.

If, after the start of the first simulation, the first failure has not occurred yet, it is useful to know when the achieved simulation time is already sufficient for acceptance. Applying eq. (38), in which it is conservatively assumed that $N=1$ and, correspondingly, $\hat{T} = t$, yields that simulations can be stopped (with the acceptance decision) when the simulation time without failure satisfies the following condition:

$$t \geq t_A \equiv \beta_1(1)/\lambda \qquad (42)$$

i.e. when $t > 3.7/\lambda$ (Shigunov, 2016, proposed a similar approximation $t > 3/\lambda$).

It is useful to extend this idea on the second and further failures: by rearranging, in eq. (38), the definition (16) as $\hat{T} = \hat{T}_{N-1}\cdot(1-1/N)+T_N/N$, where $\hat{T}_{N-1}$ is the sample mean time to failure for N-1 previous failures and assuming, conservatively, $T_N = t$, we obtain that simulation for N-th failure can be stopped with the *acceptance* decision when the simulation time without failure achieves

$$t \geq t_A \equiv \beta_1(N)\cdot N/\lambda - \hat{T}_{N-1}\cdot(N-1) \qquad (43)$$

Examples of the assessment using eq. (38,39) for a 1700 TEU container ship in loading conditions with GM=1.7 m, 1.8 m, …, 2.2 m are shown in Fig. 10 (for each loading condition, only one result is shown, corresponding to the smallest $\hat{T}$ over all T_z). Table 2 shows the number of failures required until a decision can be made; note that a large number of failures was required only in one case (GM=1.9 m).

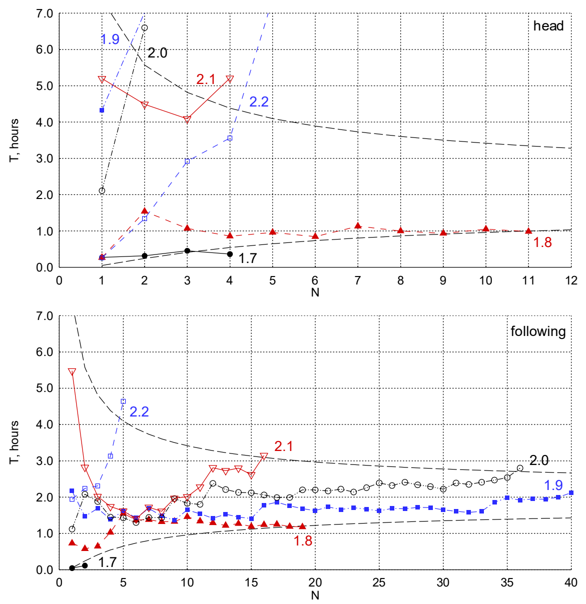

Figure 10: Sample mean time to failure vs. the number of simulations together with the *acceptance* $\beta_1(N)/\lambda$ and *not acceptance* $\beta_2(N)/\lambda$ boundaries in head (top) and following (bottom) waves; values along lines indicate *GM*

Table 2. The number of failures counted until the acceptance (+) or not acceptance (-) for parametric roll assessment in design situations

GM, m	Head waves	Following waves
1.7	-4	-1
1.8	-11	-19
1.9	+2	+175
2.0	+2	+36
2.1	+4	+16
2.2	+5	+5

6. CONCLUSIONS

Counting the number of stability failures, required to define the failure rate and the failure probability, makes sense for a stationary Poisson process. To ensure that stability failures in numerical simulations or model tests satisfy the assumptions of a Poisson process, the procedure should avoid self-repetition effects, transient hydrodynamic effects and the auto-correlation of large roll motions. To avoid self-repetition effects, the duration of simulations should be limited to about 3 hours if about 10^3 frequencies per wave direction are used for discretisation of the wave energy spectrum. During the initial transients at the beginning of simulations, the counter of stability failures and the simulation timer should be switched off. An effective way to avoid the auto-correlation effects is to stop a simulation after the first failure. These measures are sufficient to assume that the process is Poisson for the assessment of synchronous resonance in relevant conditions, whereas for parametric resonance, the process still significantly deviates from the Poisson process when the mean time to failure is less than about 2 hours.

Efficient procedures are proposed to define the 95%-confidence interval of the failure rate, required for the assessment; application examples show that small number of simulations is sufficient in most cases; in few cases (near the acceptance threshold), large number of simulations may be required.

REFERENCES

IMO (2019) Report of the Experts' Group on Intact Stability, Document SDC 6/WP.6, International Maritime Organization, London.

Longuet-Higgins, M.S., 1952, “On the statistical distribution of the heights of sea waves”, J. of Marine Research.

Shigunov, V., 2016, “Probabilistic direct stability assessment”, Proc. 15th Int. Ship Stability Workshop, 13-15 June, Stockholm, Sweden.

Shigunov, V., 2017, “Direct stability assessment: pragmatic solutions”, *Ship Technology Research (Schiffstechnik)* 64(3), pp. 144-162.

Shigunov, V., 2017a, “Possible simplifications of direct stability assessment, Proc. 16th Int. Ship Stability Workshop, 5-7 June, Belgrade, Serbia.

* * *