Ocean Engineering 146 (2017) 234–245

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

The use of confidence intervals to determine convergence of the total evacuation time for stochastic evacuation models

Angus Grandison *, Steven Deere, Peter Lawrence, Edwin Richard Galea

Fire Safety Engineering Group (FSEG), University of Greenwich, Old Royal Naval College, Park Row, Greenwich, London SE10 9LS, UK

ARTICLE INFO

Keywords:
95th percentile
Confidence intervals
Evacuation modelling
Convergence
Maritime safety
IMO

ABSTRACT

International guidelines (IMO MSC.Circ 1533) specify that evacuation models used to certify evacuation performance of passenger ships must demonstrate that the calculated representative evacuation time, the *sample* 95th percentile time τ^S, is lower than a prescribed Pass/Fail Criterion Time (PFCT). In this paper a Confidence Interval Convergence Test (CICT) method is presented that minimises the computational burden required to demonstrate that a model design has passed/failed by calculating a CI for the population 95th percentile time, τ^P, rather than simply relying on τ^S determined from an arbitrary sample of 500 simulations as specified in the current guidelines. The CICT has comparable pass/fail accuracy to using 500 simulations whilst significantly reducing the number of simulations required when the PFCT is far from the τ^P. In addition, the proposed method has superior accuracy to the convergent method described in the IMO guidelines. Furthermore, the methodology described in the guidelines fails to identify situations where there may be uncertainty in the pass/fail status due to proximity of τ^P to PFCT. The CICT identifies these situations and provides a means of resolving the uncertainty. The CICT can be applied to any stochastic evacuation model to determine parameter convergence.

1. Introduction

Many evacuation simulation models (Gwynne et al., 1999; Kuligowski et al., 2010) employ a stochastic approach for the representation of behaviour and movement (Gwynne et al., 2001, 2003; Ha et al., 2012; Korhonen et al., 2008; Meyer-König et al., 2005; Park et al., 2004; Pradillon, 2003; Thompson and Marchant, 1995; Vassalos et al., 2002) as they attempt to reflect the probabilistic nature of human behaviour (Averill, 2011). This is consistent with real behaviour since if any evacuation experiment is repeated using the same population and same starting conditions it is likely that the evacuation will progress differently and result in a different total evacuation time (TET). However, two key questions that arise when using a stochastic evacuation model concerns how many simulations are required to obtain a given level of confidence that the predicted results provide a true indication of the expected outcome for the scenario and what should be considered the representative value of predicted parameters such as TET for a given scenario. Given a distribution of predicted TETs there are a number of possible candidate values for the representative TET such as the longest TET, the mean TET, the median TET, or the 95th percentile TET. To a certain extent, the predicted parameter used to represent the distribution of possible results is dependent on the purpose for undertaking the analysis. If it is part of a risk analysis, it may be appropriate to take a reasonable worse case and so the 95th percentile TET may be appropriate, if the analysis is more concerned with typical performance, then the mean TET may be appropriate.

While there has been some interest in these issues (Meacham et al., 2004; Ronchi et al., 2014) for building applications, there are currently no internationally agreed guidelines on how to address this issue for building applications. However, the International Maritime Organization (IMO) in their guidelines for evacuation analysis (IMO, 2016) specifies that when assessing the evacuation capability of a passenger ship using an advanced egress model, a minimum of 500 simulations must be performed and that the representative TET is the 95th percentile TET, τ, from those simulations. The possible use of the 95th percentile (of a sample of simulations) TET has also been suggested for the building (Meacham et al., 2004) and aviation (Galea, 2006; Galea et al., 2010) industries. The IMO (2016) guidelines further stipulate the minimum number (four) and nature of scenarios that must be investigated for each new ship design. This includes the nature of the population (of agents) distribution (age, gender and number of disabled occupants) and the range and distribution of key parameters such as occupant response times and walking speeds. In addition the guidelines stipulate that each scenario must be repeated with the key parameters varied between the given

---

* Corresponding author.
E-mail address: a.j.grandison@gre.ac.uk (A. Grandison).

https://doi.org/10.1016/j.oceaneng.2017.09.047
Received 25 April 2017; Received in revised form 29 August 2017; Accepted 24 September 2017

Nomenclature

CI	Confidence Interval
CI(x%)	CI with an x% confidence level
CICT	Confidence Interval Convergence Test
IMO	International Maritime Organization
PFCT	Pass/Fail Criterion Time
TET	Total Evacuation Time (s)
n	sample size of simulations
R_{CI}	range of the CI (s)
T_b	TET of simulation b (s)
τ	95th Percentile TET (s)
τ^P	*population* 95th Percentile TET (s)
$\tau^{S(=n)}$	*sample* (of size n) 95th Percentile TET (s)

ranges for each repeat simulation. Thus, in addition to the natural variation in evacuation output that can be expected due to the stochastic nature of behaviour (even if none of the input parameters are altered), varying the key parameters between each of the repeat simulations will result in even greater variation in the predicted output. As stochastic evacuation models generally use pseudo-random numbers then there will be a finite number of possible different simulations that can be produced, but the number of unique simulations could be very large, 2^{19937}–1 ($>10^{6001}$) if a Mersenne-Twister Random Number Generator (RNG) (Matsumoto and Nishimura, 1998) or 2^{249} ($>10^{74}$) for a R250 (Kirkpatrick and Stoll, 1981) RNG is used. From a practical point of view it is generally only possible to take a relatively small sample (<10,000) of all possible simulations and so the population of simulations that these are drawn from is effectively infinite.

Prior to the recently updated IMO guidelines, IMO (2007) specified that the 95th percentile value from a 50 simulation trial sample, $\tau^{S=50}$, was sufficient to represent the predicted evacuation time for the vessel design. When undertaking an evacuation analysis, the representative TET is compared to the relevant **Pass Fail Criterion Time** (PFCT) and the design is deemed to have passed if the τ is less than the PFCT. However, the variability of τ between samples was not examined and there is no requirement for error bars to be specified for the representative value. Thus, in the previous IMO guidelines, $\tau^{S=50}$ was assumed to be a good estimation of the 95th percentile value of the entire population of predicted TETs for the given scenario, τ^P. However, there is considerable variation in $\tau^{S=50}$ (see Fig. 1) and using $\tau^{S=50}$ to represent τ^P can lead to an increasing number of false positives (type I error where a poor design is deemed to have passed) and false negatives (type II error where a satisfactory design is deemed to have failed) as τ^P for the vessel design and scenarios gets closer to the pass fail criterion time, PFCT. It is noted that the actual τ^P is generally impractical to determine as it would require running a very large number of simulations to ensure that all possible permutations of model input parameters and all the natural inherent model variability was accounted for. It is further noted that the τ^P for a particular model cannot be assumed to exactly represent reality due to assumptions used to specify the artificial benchmark scenarios, the simplifications within the model and a lack of data defining the performance of the population in general and particularly in emergency situations. The IMO (2016) guidelines add a 25% safety factor to account for these uncertainties.

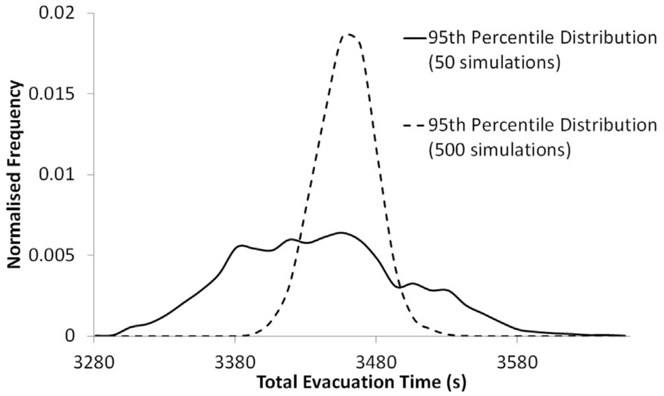

Fig. 1. Example variability of τ^S across one million experiments for 50 and 500 simulation sized samples.

A false positive occurs when τ^S is less than the PFCT but τ^P is greater than the PFCT. Similarly, a false negative occurs when τ^S is greater than the PFCT but τ^P is less than the PFCT. In the most recent version of the guidelines, IMO (2016) attempted to address this problem by increasing the sample size to a minimum of 500 simulations. In this approach it is assumed that $\tau^{S=500}$ is likely to be a more precise estimate of τ^P thereby reducing, but not eliminating, the probability of false positives or false negatives. This can be seen in Fig. 1, where the difference between the maximum and minimum values of $\tau^{S=50}$ for a sample involving 1 million simulation experiments is 540s, compared to the more precise extent of $\tau^{S=500}$ which is 191s.

However, while the IMO (2016) Guidelines state that the 500 simulations are considered a minimum, they provide no advice as to what circumstances may require additional simulations to be considered. The inevitable effect of this omission is that most engineers will treat the stated minimum as effectively the required number of simulations. They may be motivated to undertake more simulations in the event that the design failed. Furthermore, performing 500 simulations is potentially a considerable computational burden when evaluating the design of a large passenger ship with many thousands of passengers and so engineers are unlikely to voluntarily perform more simulation unless required to. This is considered a serious omission as no proof is required to demonstrate that the sample $\tau^{S=500}$ provides a good representation of the population τ^P.

Given that performing 500 simulations may be more than required in some cases and acknowledging the computational burden of undertaking the task, IMO provided the option of performing fewer than the specified 500 simulations if it could be demonstrated that the sample 95th percentile time had converged, as stated in the IMO (2016) guidelines, "*The minimum of 500 different simulations can be reduced when a convergence is determined by an appropriate method* …". Within the guidelines a suggested convergent method that increases the precision that is required for τ^S as the PFCT gets closer to τ^S is presented however, the efficiency of this approach is not discussed.

While the IMO Guidelines provide a means for demonstrating that fewer than 500 simulations may be required, it does not provide a means for demonstrating that 500 simulations may be insufficient. Thus a motivation for this paper is to provide a methodology that has comparable pass/fail accuracy as using 500 simulations whilst minimising the computational burden required when a design clearly passes or fails the PFCT by a significant margin and which indicates that more than 500 simulations may be required to make a decision on the suitability of the design.

Ronchi et al. (2014) have proposed convergence criteria for stochastic evacuation models based on five measures. The first two measures are based on comparing the difference between the mean and standard deviation of TETs for *j* simulations against the mean and standard deviation obtained for *j*-1 simulations against a specified tolerance. The other three measures are based on functional analysis (Peacock et al., 1999) which compare properties of the average overall egress curve (i.e. the number of exited agents vs time) for *j* simulations against properties of the average overall egress curve for *j*-1 simulations (note that the metrics specified in Peacock et al. (1999) are incorrectly specified and are corrected in Galea et al. (2013)). In their work the representative TET is the mean value of all the TETs generated together with the standard deviation of the TETs and is therefore, in its current form, unsuitable for examining the