IMO Information paper - The SAFEGUARD Validation Data-Set and Recommendations to IMO to Update MSC/Circ. 1238

E.R. Galea[1], L. Filippidis[1], S. Deere[1], R. Brown[1,2], I. Nicholls[3], Y. Hifi[3], N. Besnard[4]
[1] Fire Safety Engineering Group, University of Greenwich, London, UK
[2] Offshore Safety and Survival Centre, Fisheries and Marine Institute, Memorial University, St. John's, Canada
[3] Safety at Sea Ltd, Glasgow, UK
[4] Principia, Nantes, France

1.0 Introduction

Two evacuation model validation data-sets collected as part of the EU FP7 project SAFEGUARD are presented. In addition, a validation protocol and acceptance criteria are developed based on the collected data. It is proposed that the validation data-sets, suggested validation protocol and the acceptance criteria could be adopted by IMO as part of a validation suite to determine acceptability of maritime evacuation models in a future enhancement to IMO MSC.1/Circ. 1238 [1].

The data-sets are based on two semi-unannounced assemblies at sea for a RO-PAX ferry with 1349 passengers and a Cruise ship with 2292 passengers. The trials took place at an unspecified time, however passengers were aware that on their voyage an assembly exercise would take place. The data-sets consist of passenger: response time data, starting location, end location (assembly station) and arrival time at the designated assembly stations. The response time data was collected using digital video cameras while the other data was collected using a novel data acquisition system consisting of ship-mounted beacons, each emitting unique Infra-Red (IR) signals and IR data logging tags worn by each passenger. The collected data is used to define two unique validation data-sets for ship evacuation models. The data-sets are considered unique for a number of reasons, primarily because unlike most validation data-sets, they contain information defining; occupant response times, starting locations, end locations and final arrival times. Furthermore, the trials were conducted on real ships, at sea and were semi-unannounced making the results relevant, credible and realistic.

A validation protocol and acceptance criteria are proposed based on the collected data. The acceptance criteria are objective and are determined by a metric consisting of three measures, the Euclidean Relative Difference, Euclidean Projection Coefficient and Secant Cosine. Collectively the metric measures the magnitude of the distance between the predicted and experimental data and the similarity of the shapes of the predicted and experimental arrival time curves. As a whole, the metric assesses the goodness of fit between the predicted model data and the measured data. The proposed acceptance criteria take into consideration uncertainties associated with the measured data.

Here we present a summary of the findings and recommendations from project SAFEGUARD relating to the SAFEGUARD Validation Data Sets (SGVDS) proposed for adoption in a modified version of IMO MSC.1/Circ. 1238. A full paper describing this work will be presented at the "SAFEGUARD Passenger Evacuation Seminar" hosted by RINA on 30 November 2012. The full paper will be available shortly after the seminar on the SAFEGUARD website at http://www.safeguardproject.info/downloads/.

2.0 The Trials and the Data-Sets

As part of the SAFEGUARD project, a series of five semi-unannounced full-scale assemblies were conducted at sea on three different types of passenger vessels. From these trials five passenger response time data-sets and two full-scale validation data-sets were collected. This paper concerns the two Safeguard Validation Data-Sets (SGVDS) which were generated from assembly trials conducted on a large RO-PAX ferry operated by ColorLine (RP1) [2] and a Cruise Ship (CS) operated by Royal Caribbean International – SGVDS1 and SGVDS2 respectively.

© 2012: The Royal Institution of Naval Architects

The Color Line vessel can carry approximately 2000 passengers and crew and over 700 vehicles. The route taken by the vessel for collection of SGVDS1 was from Kristiansand in Norway to Hirtshals in Denmark, a trip of 3 hours and 15 minutes. Data from a sailing from Kristiansand to Hirtshals in early September 2009 was collected with 1349 passengers on board. The trial took place at an unspecified time on the crossing, however passengers were aware that on their crossing an assembly exercise would take place. Some 30 digital video cameras were used to collect the response time data. The other validation data was collected using a novel data acquisition system consisting of ship-mounted beacons, each emitting unique Infra-Red (IR) signals and IR data logging tags worn by each passenger [2]. In all, 30 IR beacons were installed and a total of 780 passengers (of the 1349 on board) wore tags.

The Royal Caribbean vessel can carry approximately 2500 passengers and 842 crew. The vessel performs several cruise holidays in the Caribbean and the Baltic Sea. Data was collected on the vessel during a cruise at the end of July 2010 with the assembly trial being performed on the first leg of the vessel's journey, between Harwich in the UK and Copenhagen in Denmark. As with the RP1, the trial took place at an unspecified time however passengers were aware that an assembly exercise would take place during the first leg of the trip. The trial was undertaken during the morning on the day after the ship left Harwich and involved some 2292 passengers. The ship's alarm was sounded towards the end of breakfast and passengers, with the help of the crew, moved to their assigned assembly stations. Each passenger was designated an assembly station, which was indicated to them on their key card (that provided access to their cabins). The same type of data as that collected during the RP1 trial was collected. Some 106 video cameras were used to capture the response times of passengers. These were made up of the ship's CCTV system (94 cameras) and specially installed digital video cameras (12 cameras). Given the larger size of this ship, a total of 70 IR beacons were installed and 1950 tags were worn by passengers.

SGVDS1 and SGVDS2 consist of: (a) ship geometry, supplied as a CAD DXF file, (b) passenger starting locations, specified as a region, (c) passenger assembly station, (d) passenger response time distributions and (e) passenger assembly curves, for each assembly station and the combined curve for the vessel. All the detailed information required to setup the evacuation analysis will be provided on an open website [3]. All other data required to run the simulations are specified in the IMO MSC.1/Circ. 1238 guidelines. For modelling purposes, each of the two vessels are assumed to have four assembly stations (AS).

3.0 Validation Metric

It is desirable to have objective measures of the level of agreement between predicted and measured performance rather than subjective assessments based on visual inspection of how well the predicted and measured curves agree. This is particularly important if the validation analysis is to be used by regulatory authorities to determine the suitability of an evacuation modelling tool. Thus it is necessary to quantify the level of agreement between predicted and measured performance. This is achieved by using a validation metric based on quantifiable differences between the predicted and measured curves.

In [4] several metrics are presented which can be used to quantify the level of agreement between predicted and measured values. However, the mathematical formulations presented in [4] have a number of typographical errors [5] and are here presented correctly. Before presenting the formulation of the metrics it is necessary to introduce some terminology. The series of measured experimental data is represented by the n-dimensional vector E = (E1, …., En), where E_i represents the measured assembly time for the i^{th} passenger. Similarly, the series of predicted model data is represented by the vector $m = (m_1, \ldots, m_n)$, where m_i represents the predicted assembly time for the i^{th} agent. The metric used to quantify the level of agreement between predicted and measured values consists of three measures (see equations 1 to 3).

© 2012: The Royal Institution of Naval Architects

$$\frac{\|E-m\|}{\|E\|}=\frac{\sqrt{\sum_{i=1}^{n}(E_i-m_i)^2}}{\sqrt{\sum_{i=1}^{n}E_i^2}} \qquad (1)$$

$$\frac{\langle E,m\rangle}{\|m\|^2}=\frac{\sum_{i=1}^{n}E_i m_i}{\sum_{i=1}^{n}m_i^2} \qquad (2)$$

$$\frac{\langle E,m\rangle}{\|E\|\|m\|}=\frac{\sum_{i=s+1}^{n}\frac{(E_i-E_{i-s})(m_i-m_{i-s})}{s^2(t_i-t_{i-1})}}{\sqrt{\sum_{i=s+1}^{n}\frac{(E_i-E_{i-s})^2}{s^2(t_i-t_{i-1})}\sum_{i=s+1}^{n}\frac{(m_i-m_{i-s})^2}{s^2(t_i-t_{i-1})}}} \qquad (3)$$

The First is the Euclidean Relative Difference (ERD) defined by equation 1. This is used to assess the average difference between the experimental data (E_i) and the model data (m_i). This equation should return a value of 0 if the two curves are identical. The smaller the value for the ERD, the better the overall agreement. An ERD of 0.2 suggests that the average difference between the model and experimental data points, taken over all the data points is 20%.

The second measure is the Euclidean Projection Coefficient (EPC) defined by equation 2. The EPC calculates a factor which when multiplied by each model data point (m_i) reduces the distance between the model (m) and experimental (E) vectors to its minimum. Thus the EPC provides a measure of the best possible level of agreement between the model (m) and experimental (E) curves. An EPC of 1.0 suggests that the difference between the model (m) and experimental (E) vectors are as small as possible. The third measure is the Secant Cosine (SC) defined by equation 3. Unlike the other two measures, it provides a measure of how well the shape of the model data curve matches that of the experimental data curve. It makes use of secants (which approximate to tangents) through both curves. An SC of 1.0 suggests that the shape of the model (m) curve is identical to that of the experimental (E) curve.

The t in equation 3 is a measure of the spacing of the data. For passenger assembly data presented in this paper, the spacing of the data is 1 i.e. there is a data point for each passenger/agent that enters an AS. Thus the difference in t consecutive values in equation 3 is 1. The s in equation 3 is a factor that represents the period of noise in the data, or variations in the experimental data resulting from microscopic behaviour not possible to reproduce in the model. Selecting a value of s which is greater than the period of the noise in the data provides a means to smooth out the effect of the noise. However, care must be taken in selecting the value of s. If s is too large the natural variation in the data may be lost, while if s is too small, the variation in the data created by noise may dominate the analysis. Selecting an appropriate value of s is dependent on the number of data points in the data-set, given by n. Thus it is desirable to keep the ratio s/n as low as possible.

For data-sets in which an experimental and model data point are available for each person, if the ERD = 0.0, then it would not be necessary to consider other measures as the two data-sets would be identical. In all other cases it is necessary to consider the three measures together in order to get a good indication of how well the two data-sets match each other. As the model data curve can cross the experimental data curve one or multiple times, EPC can return a value close to 1.0 while there is a difference between the two curves. Similarly, the SC can return a value of 1.0 even though the model and experimental data curves are off set by a constant value. In general, for

 © 2012: The Royal Institution of Naval Architects

the model and experimental curves to be considered a perfect match, it is necessary to have all three measures at their optimal values i.e. ERD = 0.0, EPC = 1.0 and SC =1.0.

4.0 Validation Protocol
Each of the validation cases are to be run 50 times using the evacuation software tool as specified in IMO MSC.1/Circ. 1238. The comparison between model and reality is based on selecting a representative case from the batch of 50 repeat simulations. The question is how to select the representative case from the 50 simulations? One way of selecting the representative case is to select the curve which produces the smallest difference in the total assembly time (TAT) between the predicted curve and the experimental data. However, the TAT measures the level of agreement between the simulation and the experiment at only one point, the time for the last person to assemble. A better measure of the overall level of agreement is derived from the ERD (equation 1) which measures the difference between the two data-sets at each point. Thus, the particular model prediction that the validation analysis will be based on is the simulation producing the best ERD for the overall assembly curve.

Thus the suggested validation protocol is as follows:

- Perform 50 simulations of the validation scenario.
- Rank each simulation according to the ERD (see equation 1) determined for the total assembly.
- Select the simulation producing the smallest ERD which will be the basis of the validation comparison.
- For the selected simulation case go through the two phase assessment process which consists of the following:
  - **Phase 1**: For the predicted total assembly curve, determine ERD, EPC, SC (see equations 1, 2 and 3) and % TAT. Determine if all four predicted parameters satisfy the acceptance criteria. If so go to Phase 2. If not, the software has failed the assessment.
  - **Phase 2**: For the predicted assembly curve for each of the four assembly stations (AS), determine ERD, EPC and SC. Determine which of the 12 predicted parameters (three for each assembly station) satisfy the acceptance criteria. At least 9 out of 12 criteria must be met for SGVDS1 and 10 out of 12 criteria must be met for SGVDS2 to satisfy the criteria and it is not acceptable to have two or more failed criteria in any one assembly station.
- The process must be repeated for SGVDS1 and SGVDS2.

The acceptance criteria for each of the validation data-sets are:

Acceptance Criteria for SGVDS1	**Acceptance Criteria for SGVDS2**
ERD ≤ 0.45	ERD ≤ 0.25
0.6 ≤ EPC ≤ 1.4	0.8 ≤ EPC ≤ 1.2
SC ≥ 0.6 with S/n = 0.05	SC ≥ 0.8 with S/n = 0.03
% TAT < 45% (Phase 1 only)	% TAT < 15% (Phase 1 only)

The acceptance criteria for the two SGVDS are different due to the uncertainties associated with each data-set. As there are more uncertainties associated with SGVDS1, this validation data-set has the weaker of the two sets of acceptance criteria.

5.0 Example Applications
SGVDS1 and SGVDS2, the validation protocol and the acceptance criteria were tested through blind applications of three established evacuation modelling software tools; maritime EXODUS, EVI and ODIGO. An example of the analysis performed using the three software tools for SGVDS2 is presented in Table 1. The assembly curves produced by each software tool for AS D and the overall assembly process are presented in Figure 1.

Table 1. Metric values for maritimeEXODUS, EVI and ODIGO predictions of SGVDS2

© 2012: The Royal Institution of Naval Architects

		SC					n	ERD	EPC	% diff TAT
	s/n	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**				
maritimeEXODUS	**Overall**	**0.9**	**1.0**	**1.0**	**1.0**	**1.0**	**1743**	**0.08**	**1.1**	**-2.2**
	AS A	**0.8**	**0.9**	**0.9**	**0.9**	**0.9**	**397**	**0.13**	**1.1**	**-18.0**
	AS B	**0.8**	**0.9**	**0.9**	**0.9**	**1.0**	**561**	**0.10**	**1.0**	**-5.7**
	AS C	**0.8**	**0.8**	**0.9**	**0.9**	**0.9**	**434**	**0.10**	**1.1**	**9.5**
	AS D	**0.8**	**0.9**	**0.9**	**0.9**	**0.9**	**351**	**0.15**	**1.0**	**8.7**
EVI	**Overall**	**0.9**	**0.9**	**1.0**	**1.0**	**0.9**	**1743**	**0.17**	**0.9**	**-1.1**
	AS A	**0.5**	**0.6**	**0.7**	**0.7**	**0.7**	**397**	**0.14**	**0.9**	**-1.1**
	AS B	**0.8**	**0.9**	**0.9**	**0.9**	**0.9**	**561**	**0.21**	**0.8**	**-8.0**
	AS C	**0.7**	**0.7**	**0.8**	**0.8**	**0.9**	**434**	**0.25**	**0.8**	**23.5**
	AS D	**0.8**	**0.9**	**0.9**	**0.9**	**0.9**	**351**	**0.24**	**0.9**	**-12.3**
ODIGO	**Overall**	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**1743**	**0.14**	**0.9**	**-0.8**
	AS A	**0.7**	**0.9**	**0.9**	**0.9**	**1.0**	**397**	**0.12**	**1.1**	**-0.8**
	AS B	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**561**	**0.21**	**0.8**	**-10.1**
	AS C	**0.6**	**0.7**	**0.8**	**0.9**	**0.9**	**434**	**0.22**	**0.8**	**24.9**
	AS D	**0.4**	**0.7**	**0.8**	**0.8**	**0.8**	**351**	**0.21**	**0.9**	**-7.6**

Applying the suggested validation protocol to the maritimeEXODUS data presented in **Table 1,** we note that in the first step the model predictions satisfy all four criteria and hence the second step of the validation protocol is considered. In the second step each AS satisfies all the criteria. As the model predictions have satisfied all four criteria in step 1 and 12 of the 12 criteria in step 2, the model is considered to have satisfied the acceptance criteria.

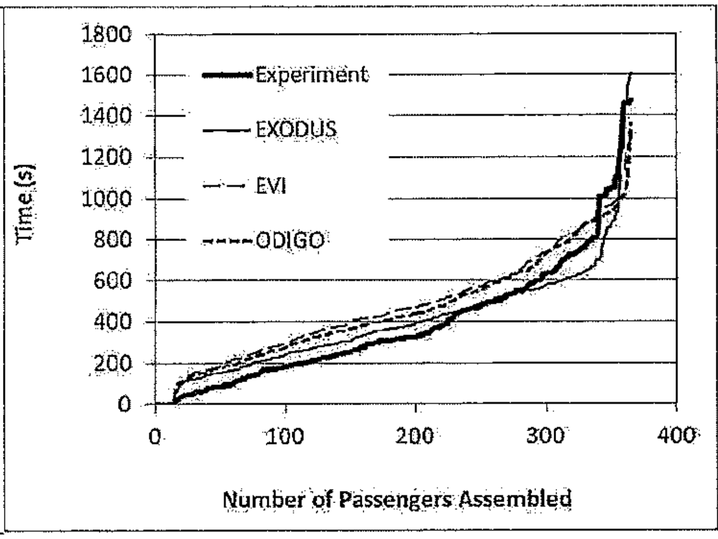
Applying the suggested validation protocol to the EVI data presented in **Table 1,** we note that in the first step the model predictions satisfy all four criteria and hence the second step of the validation protocol is considered. In the second step, AS A fails to meet criteria (iii) (see highlight in Table 1) but all other criteria are satisfied. As the model predictions have satisfied all four criteria in step 1, and 11 of the 12 criteria in step 2, the model is considered to have satisfied the acceptance criteria.

Applying the suggested validation protocol to the ODIGO data presented in **Table 1,** we note that in the first step the model predictions satisfy all four criteria and hence the second step of the validation protocol is considered. In the second step all the ASs satisfy all the criteria. As the model predictions have satisfied all four criteria in step 1 and 12 of the 12 criteria in step 2, the model is considered to have satisfied the acceptance criteria.

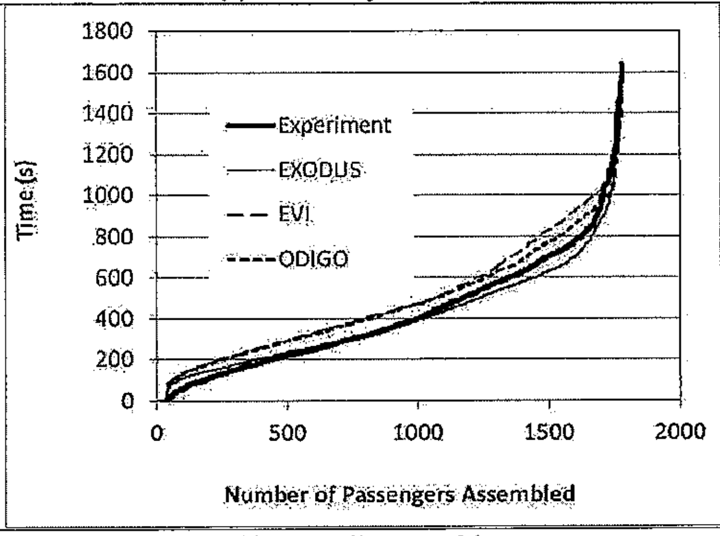
In blind applications of the validation protocol using the ship evacuation simulation software maritimeEXODUS, EVI and ODIGO, each software tool was found to satisfy the acceptance criteria for each of SGVDS1 and SGVDS2, suggesting that they are capable of predicting the outcome of the assembly process for the vessels to a given level of accuracy.

It is proposed that the suggested validation protocol and the acceptance criteria should be adopted by the IMO as part of a validation suite to determine acceptability of maritime evacuation models in a future enhancement to IMO MSC.1/Circ. 1238. In this way it is hoped to improve the reliability of the assessment of ship evacuation capabilities based on computer simulation and hence the safety of all those who travel and work on passenger ships.

 © 2012: The Royal Institution of Naval Architects



(a) Assembly Station D



(b) Overall Assembly

Figure 1. Comparison of model predictions (maritimeEXODUS, EVI and ODIGO - best ERD) with experimental data for SGVDS2 for AS D and the overall assembly

6.0 Acknowledgement.
The SAFEGUARD project (contract 218493) is funded under the European Union Framework 7 Transport initiative.

7.0 References.

1. "Guidelines for Evacuation Analysis for New and Existing Passenger Ships", IMO MSC/Circ 1238, 30 Oct 2007.
2. Galea, E.R., Brown, R.C., Filippidis. L., and Deere. S.: Collection of Evacuation Data for Large Passenger Vessels at Sea, Pedestrian and Evacuation Dynamics 2010. 5th International Conference. Proceedings. March 8-10, 2010, Springer, New York, NY, Peacock, R.D., Kuligowski, E.D., and Averill, J.D., Editor(s), 163-172, (2011).
3. http://fseg.gre.ac.uk/validation/ship_evacuation
4. Peacock, R.D., Reneke, P.A., Davis, W.D., Jones, W.W.: Quantifying Fire Model Evaluation Using Functional Analysis, Fire Safety Journal, 22, 167-184, (1999).
5. Peacock, R.D.: Private Communication with E.R.Galea, 23 November 2011.

SAFEGUARD

Passenger Evacuation Seminar

30 November 2012
London, UK

PAPERS

SEVENTH FRAMEWORK PROGRAMME

THE ROYAL INSTITUTION OF NAVAL ARCHITECTS
10 UPPER BELGRAVE STREET, LONDON, SW1X 8BQ Telephone: +44 (0)20 7235 4622