THE SAFEGUARD VALIDATION DATA-SET AND RECOMMENDATIONS TO IMO TO UPDATE MSC CIRC 1238

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

SUMMARY

Two evacuation model validation data-sets collected as part of the EU FP7 project SAFEGUARD are presented. The data was collected from a RO-PAX ferry operated by ColorLine (RP1) and a cruise ship operated by Royal Caribbean International (CS). The trials were semi-unannounced assembly trials at sea and involved some 1349 and 2500 passengers respectively. The trials took place at an unspecified time however, passengers were aware that on their voyage an assembly exercise would take place. The validation data-sets consist of passenger; response times, starting locations, end locations and arrival times in the assembly stations. The validation data were 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 results from blind simulations using maritimeEXODUS, EVI and ODIGO for these assembly trials are presented and compared with the measured data. Three objective measures are proposed to assess the goodness of fit between the predicted model data and the measured data.

NOMENCLATURE

AS	Assembly Station
CAD	Computer Aided Design
CCTV	Closed Circuit TV
CS	Cruise Ship
DXF	Drawing Exchange Format
E_i	Experimental data for i^{th} passenger
EPC	Euclidean Projection Coefficient
ERD	Euclidean Relative Difference
EU	European Union
FP	Framework Programme
FSEG	Fire Safety Engineering Group
IMO	International Maritime Organization
IR	Infra-Red
m_i	Measured (predicted) data for i^{th} agent
MSC	Maritime Safety Committee
n	Number of data points
RP1	RO-PAX vessel number 1
s	Smoothing factor
SC	Secant Cosine
SGVDS1	Safeguard Validation Data-Set 1
SGVDS2	Safeguard Validation Data-Set 2
t	Data spacing
TAT	Total Assembly Time

1. INTRODUCTION

In 2002 the International Maritime Organization (IMO) introduced guidelines for undertaking full-scale evacuation analysis of large passenger ships using ship evacuation models [1]. These guidelines, known as IMO MSC Circular 1033, were to be used to certify that passenger ship design was appropriate for full-scale evacuation. As part of these guidelines it was identified that appropriate full-scale ship based evacuation validation data was not available to assess the suitability of ship evacuation models. As suitable validation data was not available, a series of test cases were developed which verified the capability of proposed ship evacuation software tools in undertaking simple simulations. However, these verification cases were not based on experimental data. Furthermore, successfully undertaking these verification cases does not imply that the evacuation model is validated or capable of predicting real evacuation performance. In 2007, IMO MSC Circular 1238 (MSC1238) [2], a modified set of protocols for passenger ship evacuation analysis and certification were released however, the issue of validation of passenger ship evacuation models was not addressed. The IMO Fire Protection (FP) Sub-Committee in their modification of MSC Circ. 1033 at the FP51 meeting in February 2007 [3] invited member governments to provide, "...further information on additional scenarios for evacuation analysis and full scale data to be used for validation and calibration purposes of the draft revised interim guideline." The EU Seventh Framework Programme (FP7) project SAFEGUARD aims to address this requirement by providing full-scale data for calibration and validation of ship based evacuation models.

As part of project SAFEGUARD, a series of five semi-unannounced full-scale assemblies were conducted at sea on three different types of passenger vessel. From these trials five passenger response time data-sets and two full-scale validation data-sets were collected. This paper will concentrate on the two Safeguard Validation Data-Sets (SGVDS) which were generated from assembly trials conducted on a large RO-PAX ferry operated by Color

Line (RP1) [4] and a Cruise Ship (CS) operated by Royal Caribbean – SGVDS1 and SGVDS2 respectively.

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 consisted of the ship's Captain sounding the alarm and crew moving the passengers into the designated assembly areas. The trial took place at an unspecified time on the crossing however, passengers were aware that an assembly exercise would take place. The data collected during the assembly trial consisted of passenger; response time data, starting locations, final destination and arrival time at the designated assembly stations. 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 passengers [4]. 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 while it was cruising in the Baltic Sea 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 included 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.

The ship validation data-sets that were generated are unique for a number of reasons. Unlike most evacuation model validation data-sets, SGVDS1 and SGVDS2 incorporate regional information relating to the starting locations of the population in addition to the actual response time distribution for the population. Most evacuation validation data-sets lack these essential details allowing modellers the opportunity to tune their predictions in order to obtain the best fit to the experimental results. Furthermore, the trials were conducted on a real ship, at sea and were semi-unannounced making the results relevant, credible and realistic. Finally, the two data-sets represent the first comprehensive ship evacuation model validation data-sets collected.

In this paper we will present both SGVDS1 and SGVDS2, the blind results from three Advanced Evacuation Modelling tools; maritimeEXODUS [5-11], EVI [12-14] and ODIGO [15,16] of the validation data-sets, an assessment of the level of agreement between model predictions and trial data and a suggested protocol for testing other models.

2. THE SHIP GEOMETRIES

2.1 RP1

This ship contains a mixture of different public spaces spread over three decks (deck 7 to 9) including: business and traveller class seating areas (airline style seating), large retail and restaurant/cafeteria areas, bar areas, indoor and outdoor general seating areas and general circulation spaces (see Figure 1). The cabins located on deck 9 are intended for lorry drivers only and not for regular passengers.

Only the airline seating portion of deck 9 is available to regular passengers. A CAD file was provided (.dxf format) to define the layout of the ship within the evacuation model. The ship has four Assembly Stations (AS), three located on Deck 7 (AS A, B and C) and one located on Deck 8 (AS D). AS B and C are located on the outer decks while AS A and D are internal. AS A is located within the ship's reception/lobby area and AS D is located within the ship's self service restaurant. The outer AS B and C each have two exits/entrances. Internal AS A, located in the ship's lobby area, can be accessed by three routes, while internal AS D, located in the ship's self service restaurant can be accessed by two routes.

The vessel has four main vertical zones and four sets of primary passenger staircases. The aft most staircase (stair 1) is located in the bar on deck 7 and extends up to the bar on deck 8 and is 1.0 m wide. The next staircase (stair 2) is located just outside the bar on deck 7 and extends to deck 9. From deck 7 to deck 9 the staircase consists of two lanes separated by a banister with landings located between deck 7 and deck 8 and deck 8 and deck 9. The width of each stair lane is 1.35 m.

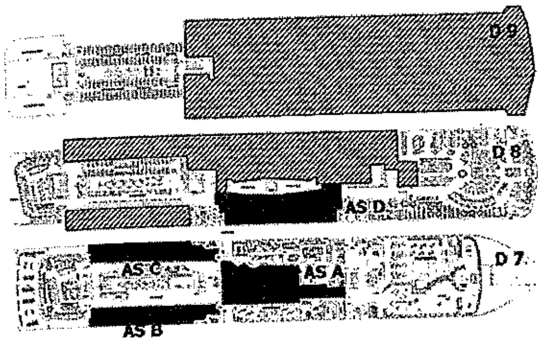


Figure 1. Layout of RP1 showing assembly stations (hatched areas not accessible by passengers)

The next staircase (stair 3) is located near midships between the reception area and the general seating area and also extends between decks 7 and 9. From deck 7 to deck 8 the stair consists of two lanes separated by a banister with a landing located between deck 7 and deck 8. From deck 8 to deck 9, there is only a single stair lane with a landing located between deck 8 and deck 9. As with stair 2, the width of each stair lane is 1.35 m. The final staircase (stair 4) is located outside the duty free shop/buffet restaurant in the forward part of the ship and extends between deck 7 and deck 8. The stair consists of two lanes separated by a banister with a landing located between deck 7 and deck 8. As with stairs 2 and 3, the width of each stair lane is 1.35 m.

2.2 CS

This vessel consists of 13 decks, of which seven decks are accommodation space consisting of passenger cabins. The other five decks consist of general circulation and entertainment spaces such as restaurants, bars, disco, swimming pools, casino, theatre, cinema, spa/health centre, business centre, leisure pursuits (such as gymnasium, climbing wall, crazy golf, cards room) and retail areas (see Figure 2).

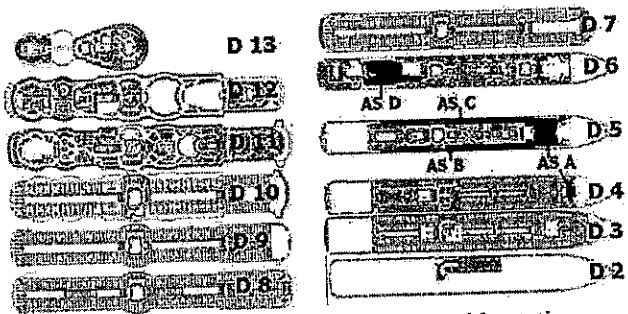


Figure 2. Layout of CS showing assembly stations

The regions not used by passengers have been removed from Figure 2. A CAD file was provided (in .dxf format) to define the layout of the ship within the evacuation model. The ship has 18 AS distributed over two decks, deck 5 and 6, of which 10 are external and eight are internal. The 10 external AS (AS b to AS f and AS r to AS v) are located on deck 5. For the purposes of the validation modelling, these are grouped together and identified as AS B and C, with AS B representing the actual AS v to r and AS C representing the actual AS b to f. AS B has three entrances located near the atrium amidships, in the shopping mall and just outside the theatre at the fore end of the vessel, while AS C has two entrances located outside the theatre in the fore of the vessel and the other located near the atrium amidships.

The eight internal AS are located on deck 5, AS a, and on deck 6, AS g and AS w. These AS are located in the theatre (AS a) and restaurant areas (AS g and AS W). Once again, for the purposes of the validation modelling, these are grouped together and identified as AS A and D, with AS A representing the actual AS a, and AS D representing the actual AS g and AS w. AS A has two entrances located at the entrance to the theatre, while AS D has two entrances located at the atrium (amidships) and from the bar area at the aft of the vessel.

The vessel has seven main vertical zones however only three main vertical passenger staircases were available in the trial. The first staircase is located within the restaurant in the aft section of the vessel, which is spread across deck 4 and deck 5. This stair is curved with a landing. The second staircase is located amidships in the ship's atrium and extends from deck 2 to deck 13 with a varying geometry. The other staircase is located in the forward part of the vessel, next to the theatre, and extends from deck 2 to deck 12. All of the stair runs for this staircase are 1.2 metres wide and 1.9 meters long. There were two double lane runs leading to a landing which measures 5.2m by 1.5m. From the landing there are two more double lane stair runs leading up to the next deck.

3. THE SIMULATION SOFTWARE

3.1 maritimeEXODUS

The ship evacuation model maritimeEXODUS [5-11] produced by the Fire Safety Engineering Group (FSEG) of the University of Greenwich was used to perform the evacuation simulations presented in this paper. The software has been described in detail in many publications [5-11] and so only a brief description of the software will be presented here. EXODUS is a suite of software to simulate the evacuation and circulation of large numbers of people within a variety of complex enclosures. maritimeEXODUS is the ship version of the software. The software takes into consideration people-people, people-fire and people-structure interactions. It comprises five core interacting sub-models: the Passenger, Movement, Behaviour, Toxicity and Hazard sub-models. The software describing these sub-models is rule-based, the progressive motion and behaviour of each individual being determined by a set of heuristics or rules. Many of the rules are stochastic in nature and thus if a simulation is repeated without any change in its input parameters, a slightly different set of results will be generated. It is therefore necessary to run the software a number of times as part of any analysis. The submodels operate on a region of space defined by the geometry of the enclosure. The geometry can be specified automatically using a DXF file produced by a CAD package or manually using the interactive tools provided. In addition to the representation of the structure itself, the abandonment system can also be explicitly represented within the model, enabling components of the abandonment system to be modelled individually.

The software has a number of unique features such as the ability to incorporate the effects of fire products (e.g. heat, smoke, toxic and irritant gases) on crew and passengers and the ability to include the impact of heel and trim on passenger and crew performance. The software also has the capability to represent the performance of both naval personnel and civilians in the operation of watertight doors, vertical ladders, hatches and 60 degree stairs. Another feature of the software is the ability to assign passengers and crew a list of tasks to perform. This feature can be used when simulating emergency or normal operating conditions. The version of the software used for this analysis was V5.0 beta.

3.2 EVI

EVI [12-14], developed by Safety at Sea Ltd, is an evacuation tool which utilises a continuous space modelling approach. The simulation can be performed in real-time with direct and interactive feed back from the 3D virtual reality environment. EVI consists of two models. There is macroscopic modelling for high-level planning such as how to get from one location to another and there is microscopic modelling which occurs at agent level allowing agents to avoid boundaries and each other according to a set of pre-defined rules (containment, collision avoidance, lane formation and conflict resolution). In combination, this is termed mesoscopic modelling.

A geometric model of the layout is developed in a pre-builder called EVE. This is developed from existing representations of the vessel such as CAD Models and general arrangements. Distributions of agent locations and the evacuation schemes are added to this model (muster stations, primary routes etc). Semantics which relate to additional information agents receive from the environment such as signage can also be added. The environmental modelling is provided by the user in the form of a database. A graph topology is formed from shape definition linked with doors. Routes are then formed through these spaces. These routes, when reviewed in the real world are the path plans that the agents follow to get from the initial location to mustering.

As agents move, they must avoid the walls of the environment in a process called containment. As other agents are introduced collision avoidance must be considered to prevent people running into each other. As the number of agents increase, lane formation is modelled to enable flows of people to pass. Counterflow and arch resolution are introduced which ensure that blockages do not occur in corridors and doors respectively. Individual agents are able to be programmed with objectives allowing crew procedures and specific passengers movement to be defined. The effect of smoke, heel and trim can be introduced to the environment and the agents respond accordingly.

3.3 ODIGO

ODIGO (ancient Greek standing for "I guide"), developed by Principa, is a tool to simulate crowd movement onboard ships [15,16]. It is an integrated tool including a pre-processor, a simulation engine, and a post-processor. The model comprehends areas representing public spaces created on decks and related staircases. The simulation engine uses a multiagent method of a cognitive/reactive hybrid type. The simulation uses an exact geometry, i.e. agents may move anywhere in areas provided that they respect margin distances between themselves and walls. The agent definition (features and starting position) is made using random allocation. The agents act upon objectives (got to AS, to cabin, move to craft, etc.) and they may chain several objectives together. The main application area of ODIGO is evacuation simulation. Nevertheless, it can be used to simulate crowd flows in other situations:

- Embarkation or disembarkation,
- End of a show,
- Queues in restaurants.

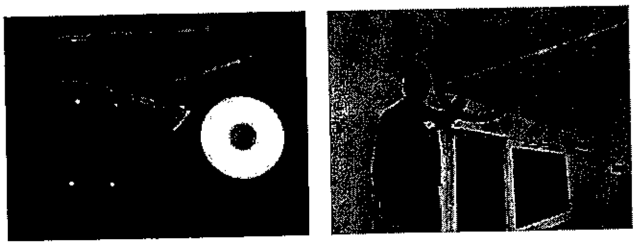
The pre-processor allows the geometry of public spaces of the ship to be created from DXF files describing general arrangement of the decks. When simulating human behaviour, many random items must be integrated to take into account the stochastic aspect of the problem. The results that are directly available for post-processing are:

- Time history of flows: crossing passages, reaching targets.
- 3D visualization of the simulation
- Display of the routes
- Densities of population in the areas

4. THE INFRA-RED TRACKING SYSTEM

The system used to track and time the movement of passengers from their starting locations to their assigned

AS was an Infra-Red (IR) system based on the TagMobile system developed by the RFID Centre Ltd. The RFID Centre worked with FSEG to modify this system to make it more appropriate for use in evacuation research applications [4]. The system deployed consisted of a number of IR Beacons strategically located throughout the vessel, and IR data logging tags worn by each passenger (see Figure 3a). Each beacon generates a unique IR light field. As a tagged individual passes through the IR field, IR light sensors in the tag detect the IR light and log its ID and the time at which it was detected in the tag's internal memory. Following the trial, all the tags must be retrieved in order to determine the occupant's starting and end locations and the arrival time at the assembly station. The IR beacons are strategically placed at the main locations where passengers congregate and at the entrances to each of the AS (see Figure 3b). In this way, the initial location of each tagged passenger can be determined, which AS they go to and at what time they enter the AS. Testing of the IR tracking system demonstrated that the system was able to identify the number of passengers passing a point, even in very large crowds and record the time at which they passed the measuring point [4].



(a) (b)

Figure 3: IR beacon and tag (a) and installing IR beacon at the entrance to an external AS (b)

To test the accuracy of the arrival times derived from the IR system, video cameras were installed at the entrance to several of the AS on the CS. This enabled a comparison of the arrival time derived from the IR system with the arrival times manually determined from the video camera record. In addition, this analysis allowed for a comparison of the total number of passengers passing through the entrance to the AS as counted by the IR system with the actual number that could be seen in the video record. The comparison was carried-out for two locations, both on the ship's starboard side of Deck 5 – one forward and one near midships. The forward location (at Beacon location 73, Camera UOG12) was a doorway with a vestibule leading to AS B. The location near midships (at Beacon 50, Camera UOG10) was a doorway that opened directly into the same external AS. These two locations were selected as they represented examples of locations in which the beacons were expected to perform well i.e. location 50 and those which would pose a challenge for the beacons i.e. location 73.

When analysing the video for both locations, the time at which a passenger passed across the door line was taken as their entry time. Because a comparison was being made to the IR data, times were recorded only for passengers that could be clearly seen wearing or holding an IR tag. In addition, because of the way the IR tag data was analysed, the entry times were recorded only for passengers who entered the assembly station and remained there.

Results of the comparisons at Beacon location 73 are provided in Figure 4. It is clearly seen that the IR data collection system matched quite closely with the data manually derived from the video record.

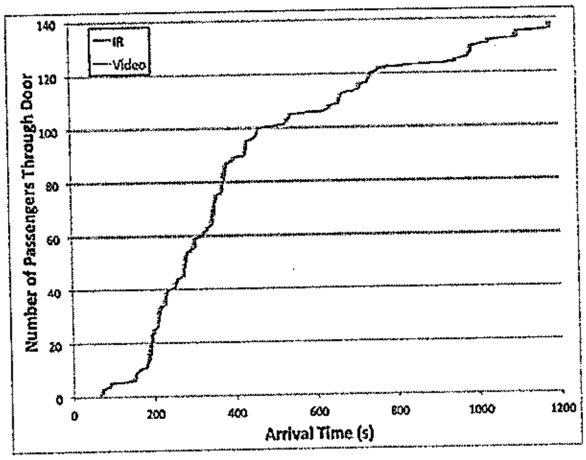


Figure 4. Comparison of passenger arrival times at Beacon 73 (Camera UOG12)

The IR system correctly counted the number of passengers through the door (138) and timing results consistently lagged the camera results by 5.0 s on average with a standard deviation of 1.11 s (maximum difference was 10 s and minimum difference was 2 s). It is noted that the IR system accurately counted the number of passengers even in the high density situation encountered at this location. These results suggest that the IR system provides an accurate measure of the arrival times for passengers when compared against a synchronised video system, despite a small lag between the actual arrival time and what the IR data collection system actually measures. In addition, the IR system accurately counts the number of people that arrive at the measuring location, even in high density situations.

5. THE INITIAL POPULATION DISTRIBUTION

The initial distribution of the population was determined through the use of the IR tracking system. Using the IR tag information, the initial location of each tagged passenger was determined using data related to the first

IR field that the tagged passengers passed through. Using this information the initial location of each tagged passenger can be confined to a region of space on a deck defined by the IR beacons. Typical regions correspond to the physical compartments on the ship, so a region may be a restaurant, or bar area, or communal seating area.

The starting deck of the 764 tagged passengers on RP1 is shown in Table 1. As shown in Table 1, 413 tagged passengers were on deck 8 at the start of the evacuation. The starting region within deck 8 for each of the 413 tagged passengers is also known. For example, 39 tagged passengers were initially in the aft bar on deck 8, while 4 tagged passengers were located in the airline seating area on deck 8. Similarly, the starting deck for the 1779 tagged passengers included in the assembly analysis on the CS is shown in Table 2. Also presented in Tables 1 and 2 are the final AS that passengers starting on each deck ended up in.

Table 1: Starting deck location and final assembly station for each passenger in the RP1 trial

Deck	7	8	9	Total
AS A	101	54	2	157
AS B	102	32	45	179
AS C	64	32	6	102
AS D	7	295	24	326
Total	274	413	77	764

Table 2: Starting deck location and final assembly station for each passenger in the CS trial

	2	3	4	5	6	7	
AS A	2	26	59	26	13	33	
AS B	1	56	101	45	7	36	
AS C	12	41	71	31	11	35	
AS D	4	10	52	25	21	35	
Total	19	133	283	127	52	139	
	8	9	10	11	12	13	Total
AS A	31	25	12	153	22	0	402
AS B	38	42	40	178	25	6	575
AS C	27	29	24	133	21	2	437
AS D	30	57	28	81	22	0	365
	126	153	104	545	90	8	1779

6. THE TRIAL RESULTS

The main results for the assembly trials conducted on the RP1 and the CS are presented and discussed. These concern the measured assembly time for each of the tagged passengers and response time distribution associated with each starting region.

6.1 THE RESULTS FOR THE RP1 ASSEMBLY TRIAL – SGVDS1

6.1.1 Final Locations of Tagged Passengers at the end of the Assembly Trial

Of the 1349 passengers on board RP1, 780 wore tags and so were tracked during the trial. However, only 764 of these are included in the assembly analysis (see section 6.1.2). Presented in Table 1 are the locations of the tagged passengers on completion of the assembly trial. For example, 326 tagged passengers ended up in AS D, of which seven came from deck 7, 295 came from deck 8 and 24 came from deck 9. Furthermore, while not shown in Table 1, the starting region for each passenger is also known. For example, 147 tagged passengers entered AS D from the neighbouring forward restaurant on deck 8, 7 tagged passengers from the neighbouring general seating area to the aft of AS D entered AS D while 139 tagged passengers who were already located in the AS at the sounding of the alarm remained in AS D.

6.1.2 Passenger Response Time Distribution

The passenger response time distribution was determined from data collected from the 30 digital video cameras located throughout the vessel [4]. Shown in Figure 5 is a video camera used to record response phase behaviours of passengers located in the airline seating area on deck 8. In total the response times for 470 passengers were determined. The response times were determined on a regional basis and associated with the type of space that was occupied. Thus a response time distribution was determined for passengers located in the airline seating areas, bars, restaurants, shops and general seating areas (see Figure 5). In addition, an overall response time distribution was also determined making a total of six response time distributions. Each response time data-set was fitted with a log normal curve, producing for each response time distribution the minimum and maximum observed response times, the natural log of the mean response time and the natural log of the standard deviation.

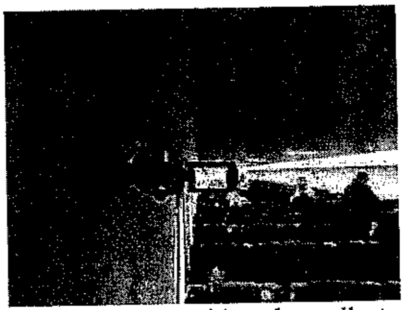


Figure 5. Video camera positioned to collect passenger response time data

Using this information the log normal response time distribution for each region and the overall response time distribution can be constructed. For example, for the response time distribution for occupants in the bar region (see Figure 6), the minimum and maximum response times are 0 and 402 s, while the log of the mean response time is 3.432 and the log of the standard deviation is 0.924 (data for day 2 trial).

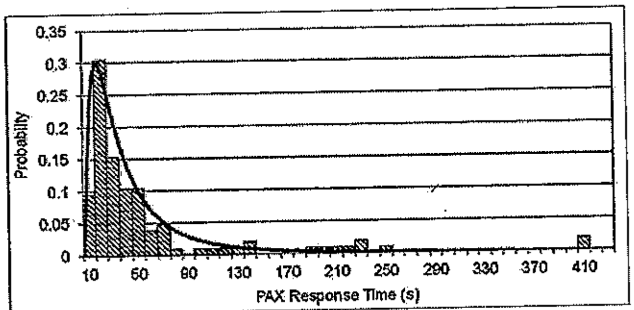


Figure 6. Bar area log normal response time distribution for RP1

The Captain officially ended the assembly exercise 10 minutes after its start. The IR tracking system recorded the time that each tagged passenger entered an AS providing a good indication of the overall assembly time. The IR data suggests that the last tagged passenger arrived in AS A after 585 s (9 min 45 s). In addition, the IR tracking system enables the generation of the arrival curve for each AS and hence the overall arrival curve (see Figure 8). As such the SGVDS1 provides a means of determining not only how well an evacuation model can predict the overall assembly time (i.e. when the final passenger enters the AS), but more importantly, how well the evacuation model can predict the overall assembly process (i.e. the time at which each passenger enters the AS).

In principle, this data-set is ideal for validation purposes, as the starting locations and response times of the population is known. This means that it should be possible to remove most of the uncertainty associated with the input parameters associated with response time and starting location. However, there are several complications associated with SGVDS1.

Firstly, of the 1349 passengers on board, 780 wore the IR tags and participated in the assembly trial. Of these, 16 appeared in the AS after the trial ended and so are not included in the analysis, giving a total of 764 tagged passengers who were actually included in the assembly exercise data-set. The majority of the 569 passengers who did not take the tags indicated that they did not want to participate in the assembly exercise – which was not compulsory for ethical and legal reasons. A small number indicated that they did not want to wear the tag. However, of the 569 passengers who did not take tags, a significant number did eventually decide to participate in the assembly exercise. This was determined by a combination of analysis of video footage, passenger questionnaire responses and team members who were in the assembly stations collecting the IR tags from the participants. By participating in the trial, the presence of the untagged individuals in the evacuation routes will have had an impact on the overall evacuation, especially in the highly congested areas. However, their assembly times will not have been recorded in the overall assembly data.

Secondly, the exact starting location of the tagged participants was not known, but the region where they were located was known. Spatial regions were between 24m and 48m long; thus not knowing the precise starting location of an individual may increase/decrease their arrival time by 25-50 seconds.

Thirdly, the response time is not associated with a unique individual but to a region. Thus the precise response time of each unique individual is not known, but the response time distribution associated with a starting region is known. All of these factors must be taken into consideration when determining how well the evacuation model predicts the assembly exercise.

6.2 THE RESULTS FOR THE CS ASSEMBLY TRIAL – SGVDS2

6.2.1 Final Locations of Tagged Passengers at the end of the Assembly Trial

Of the 2500 passengers on board the CS, 1950 wore tags and so were tracked during the trial and 1779 of these are included in the assembly analysis (see section 6.2.3). Presented in Table 2 are the locations of the tagged passengers on completion of the assembly trial. For example, 402 tagged passengers ended up in AS A, of which two came from deck 2, 26 came from deck 3, 59 came from deck 4, etc. As with the RP1 data, the starting region for each passenger is also known.

6.2.2 Passenger Response Time Distribution

The passenger response time distribution was determined from data collected from the 106 digital video cameras located throughout the vessel. The response times for

1228 passengers were determined producing an overall response time distribution which is presented in Figure 7. The response time data-set was fitted with a log normal curve, with the following key parameters; the minimum and maximum response times are 0 s and 1379 s, while the log of the mean response time is 5.012 and the log of the standard deviation is 0.89. Since response times were not collected for all the passengers in all the various regions of the ship (due to its size) the overall response time distribution is used for SGVDS2.

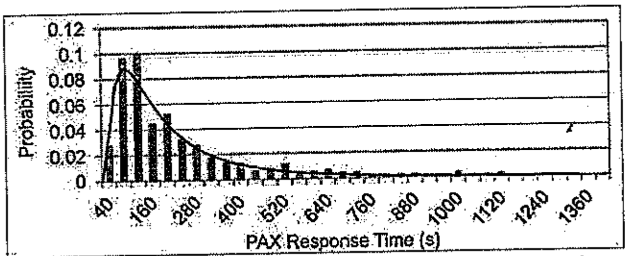


Figure 7. Overall log normal response time distribution for the CS assembly trial

6.2.3 Assembly Times

The Captain officially ended the assembly exercise 29 minutes after its start. The IR data suggests that the last tagged passenger arrived in AS A after 1637 s (27 min 17 s). The arrival curves for each AS and the overall arrival curve, generated using the IR data, is presented in Figure 9. In principle, this data-set is ideal for validation purposes, as the starting locations and response times of the population is known. This means that it should be possible to remove most of the uncertainty associated with input parameters associated with response time and starting location. However, as with the SGVDS1, there are several complications associated with the validation data-set which introduces some degree of uncertainty in the trial results.

First, of the 2292 passengers on board, 1950 wore the IR tags and participated in the assembly trial. Of these, 171 tagged participants were excluded from the data-set for various reasons e.g. a number of participants arrived at the AS after the trial was declared over, several participants had response times considerably longer than that measured using the video camera data, another participant possibly took a circuitous route to the AS, such as going up stairs for several decks when they should have been going down, or delayed their egress for some other reason etc. The 342 passengers that did not have tags were: (1) children under the age of 12 who were not permitted to take part in the validation study, (2) passengers who did not take part in the trial and (3) a number of passengers who decided not to wear the IR tag or forgot to wear the IR tag while participating in the trial. The number in the latter category is believed to be small (through analysis of video footage from the entrance to the AS) and estimated to be less than 10% of the number participating who wore tags. Unlike in the case of the RP1 trial, the impact of these passengers on the overall results is expected to be small and is ignored.

Secondly, the exact starting location of the tagged participants was not known, but the region where they were located was known. Spatial regions were between 50m and 95m long; thus not knowing the precise starting location of an individual may increase/decrease their arrival time by 50-95 seconds. Thirdly, the response time distribution is not associated with a unique individual but represents the overall response time distribution for the entire vessel. Thus, unlike in the RP trial, the zonal response time distributions on the CS are not known to sufficient resolution to be meaningful. The impact that this will have on an evacuation analysis is difficult to estimate as each time the simulation is run, a different random allocation of response times is made for all agents. Thus an agent may be allocated a very long response time in one simulation and in the next simulation may be allocated a very short response time. The error associated with the random allocation of the global response time may be minimised if the average predicted assembly time distribution is considered. However, MSC 1238 requires that the 95th percentile case is used to represent the vessel assembly performance. All of these factors must be taken into consideration when determining how well the evacuation model predicts the assembly exercise.

7. MODELLING PROCEDURES

The simulations for each modelling tool were performed by the respective model developers. The simulations were performed blind i.e. without sight of the full set of experimental results. The bulk of the parameters used in the simulation are compliant with those specified in MSC1238 [2] with the exception of the response time distribution and the initial location of the passengers; these are determined from the trial data. For the RP1 simulations, the regional response time data is used (see Section 6.1.2) and the initial starting locations of the passengers as defined in Section 5 are used. For the CS simulations, the global response time data is used (see Section 6.2.2) and the initial starting locations of the passengers as defined in Section 5 are used. It is noted that as the population demographics used in the validation analysis are derived from MSC1238 and not the actual vessels, they may not necessarily reflect the actual population demographics of the passengers involved in the trials. This may introduce some error in the overall numerical predictions of the assembly process.

Furthermore, given the starting zone that an agent is assigned to, the AS that they should use is known. This information is also imposed on the simulations presented here. Thus the agent will go to the correct AS as defined by the trial. As already noted, in the RP1 trials a significant number of passengers (569 or 42% of those on board) did not wear IR tags, however, some of these untagged passengers actually participated in the trial and so had an effect on the movement of those passengers wearing the IR tags during the assembly exercise. It is not known how many of the 569 passengers participated in the trial but we cannot ignore the fact that a large number of passengers, who were not wearing IR tags participated in the assembly exercise and so had an impact on the overall result. In an attempt to take this into account, it is assumed that 250 of these passengers, approximately half, did actually participate in the assembly exercise. These passengers are included in the evacuation simulation as moving passengers, but are not included in the analysis of the AS arrival curves and the total assembly times. These 250 agents are distributed throughout the vessel according to the population distribution of the known 764 passengers.

As is required by MSC1238 [2] a total of 50 repeat simulations are produced, where the starting locations of the passengers within the various starting regions are randomised. In the regulatory analysis, the 95th percentile case is selected to represent the prediction of the assembly process, with the Total Assembly Time (TAT) derived from the 95th percentile time representing the overall assembly time for the vessel. The regulations assume that evacuation models will under-predict the likely total assembly time by 25% and so require that an additional 25% safety factor is added to the predicted total assembly time. The purpose of the validation exercise is to determine how well the evacuation software predicts the overall assembly process, not simply the TAT. It is possible that a poor software tool may incorrectly predict the overall assembly process but randomly produce a reasonable prediction of the TAT.

While the TAT may be the only number that the regulatory authority is concerned with, confidence in the reliability of the TAT prediction is based on how well the software predicts the overall assembly process. Thus, the validation exercise must evaluate how well the software reproduces the overall assembly process (arrival times for each passenger) and not simply the TAT. Furthermore, just as there is a spread in the results for the numerical simulations, there would also be a spread in the experimental results if the experiment were repeated, even if all the passengers started from the same locations with the same response times as it is unlikely the passengers would do the exact same thing twice. While we have a range of numerical results for the assembly, we only have one experimental result and it is impossible to determine if the experimental result is representative of the average result for the experiment or if it is an outlier and how wide the range in experimental results is likely to be. Thus, the best numerical result will be compared with the experimental result to determine how well the software predicts the trial assembly.

From a simple visual observation of the predicted assembly curves it is difficult to identify which of the 50 curves produces the best level of agreement with the experimental results. As the regulatory authorities are primarily concerned with the prediction of the TAT, the numerical prediction producing the TAT with the smallest error is arbitrarily selected to represent the best prediction. For comparison purposes, the numerical prediction producing the TAT with the largest error is also considered. In addition, a more objective method for identifying the numerical prediction which produces the best level of agreement with the experimental data is identified.

8. COMPARING MODEL PREDICTIONS WITH TRIAL RESULTS

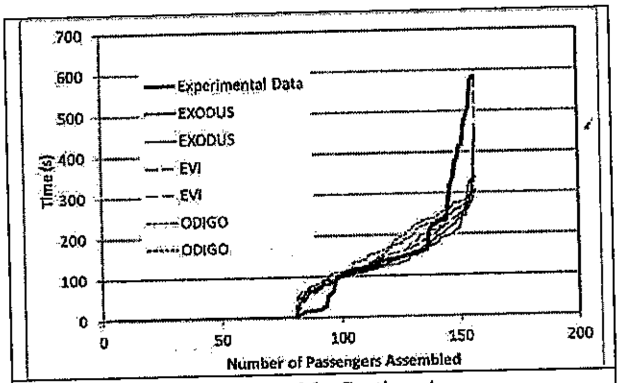
8.1 THE RP1 MODEL PREDICTIONS – SGVDS1

The numerical predictions producing the TAT with the smallest and greatest error for the RP1 are presented in Figure 8 along with the experimental data. Presented are the measured and predicted arrival curves for each AS (Figure 8a to Figure 8d) and the overall arrival curve (Figure 8e). The arrival curves presented in Figure 8 appear shifted to the right by varying degrees. This is due to a number of tagged passengers who were already in the AS at the start of the assembly process.

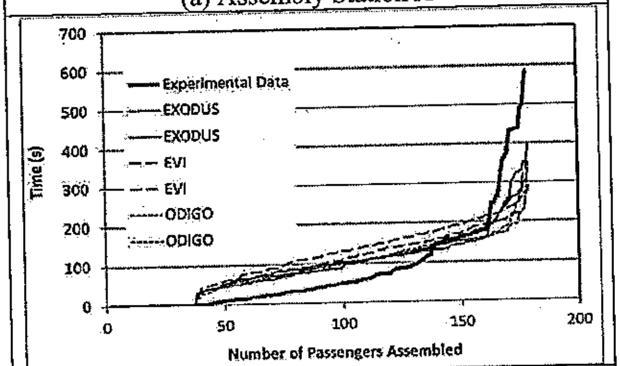
As can be seen from Figure 8, the numerical simulations for all three software tools under-predict the TAT for each AS and the overall assembly process, with the exception of EVI and AS A. For maritimeEXODUS, the simulation producing the best/worst TAT under-predict (negative values) the TAT for each AS, A, B, C and D by; -22%/-42%, -33%/-43%, -27%/-29% and -28%/-16% respectively and the overall TAT is under-predicted by -22%/-34% (see Table 3). Thus at best, maritimeEXODUS under-predicts the TAT by 22%, while at worst it under-predicts the TAT by 34%. The error in predicting the TAT for each assembly station varies from -16% to -43%.

For EVI, the simulation producing the best/worst TAT over-predict (positive values) or under-predict (negative values) the TAT for each AS by; 3%/-47%, -29%/-51%, -43%/-54% and -23%/-28% respectively and the overall TAT is under- or over-predicted by 3%/-47% (see Table 4). Thus at best, EVI over-predicts the TAT by 3%, while at worst it under-predicts the TAT by 47%. The error in predicting the TAT for each assembly station varies from 3% to -54%.

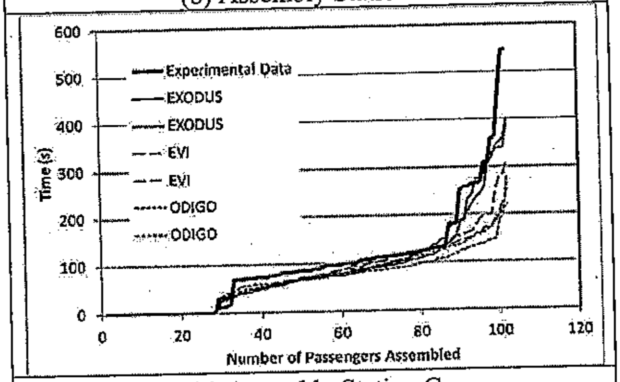
For ODIGO, the simulation producing the best/worst TAT under-predict (negative values) the TAT for each AS by; -22%/-47%, -39%/-52%, -60%/-49% and -23%/-30% respectively and the overall TAT is under-predicted by -22%/-47% (see Table 5). Thus at best, ODIGO under-predicts the TAT by 22%, while at worst it under-predicts the TAT by 47%. The error in predicting the TAT for each assembly station varies from -22% to -60%.



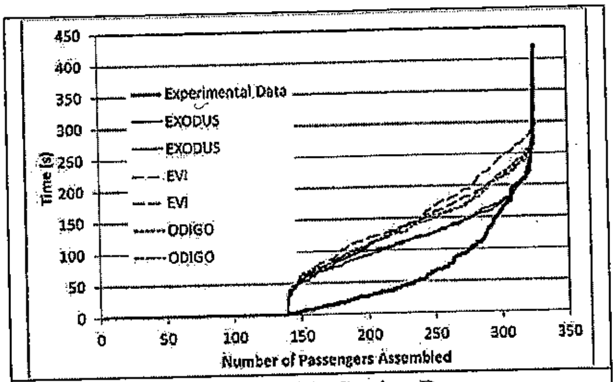
(a) Assembly Station A



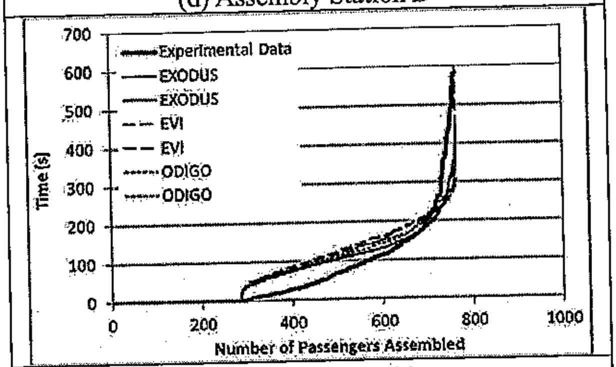
(b) Assembly Station B



(c) Assembly Station C



(d) Assembly Station D



(e) Overall Assembly

Figure 8. Comparison of model predictions (best and worst TAT) with experimental data for RP1 for maritimeEXODUS, EVI and ODIGO

However, as can be seen from Figure 8, the number of passengers in the later stage of the assembly, resulting in the under-predictions in the TATs, is very small and is possibly caused by a few stragglers exhibiting behaviours not represented within evacuation models, for example stopping on the way to the AS to talk or make some observation. Furthermore, as described earlier, there are several other uncertainties introduced in the experimental data which may contribute to these differences.

The uncertainty in the exact starting location of the passengers can introduce an error of 25s to 50s in the prediction of the assembly times. This uncertainty alone introduces a possible error of some 9% in the overall TAT and an error of some 11% in the prediction of the TAT for the individual AS. In addition, in analysis of the performance of the IR system, it was noted that the IR measured assembly times can lag the actual assembly times by up to 5s. This would introduce a 1% error in the estimation of the measured assembly times. Finally, the error associated with the assembly of the non-tagged passengers is difficult to estimate. While the analysis has attempted to take this into account by introducing half the untagged passengers into the simulation, it is not clear if this is sufficient as the actual number, starting location and AS used by the non-tagged passengers is not known. Taking this uncertainty into consideration, the errors in the predicted assembly times appear reasonable.

As noted earlier, simply predicting the time for the overall assembly within an acceptable tolerance is not sufficient to determine whether or not the simulations provide an accurate representation of the assembly process. It may be possible for a reasonable prediction of the overall assembly time to be generated while the evacuation dynamics is misrepresented by the evacuation simulation. To determine if the evacuation simulation is a good representation of the evacuation dynamics it is necessary to look at how well the predicted arrival curves match the measured arrival curves.

From Figure 8 it is noted that for each software tool, the assembly curves for the best and worst TAT are very similar. This suggests that the primary difference between the simulations producing the best and worst TAT is the value for the TAT. This difference is likely to be driven by random variables such as the response time or starting position of the last agent to assemble. By sight, the predicted and measured assembly curves produced by the three models for the overall assembly appear to be in very good agreement with each other and in reasonable agreement with the experimental data (see Figure 8e). With the exception of AS D (see Figure 8d), the predicted arrival curves produced by each software tool for each of the AS (Figure 8a to Figure 8d) also appear to be in reasonable agreement with the measured curves. All three models appear to produce similar predictions for each AS, with maritimeEXODUS and EVI producing slighter better agreement with the experimental data for AS A, maritimeEXODUS and ODIGO producing slightly better agreement for AS B, maritimeEXODUS and EVI producing slightly better agreement for AS C and maritimeEXODUS producing the better agreement for AS D.

These observations suggest that the evacuation models are reasonably predicting the overall assembly process.

8.2 THE CS MODEL PREDICTIONS – SGVDS2

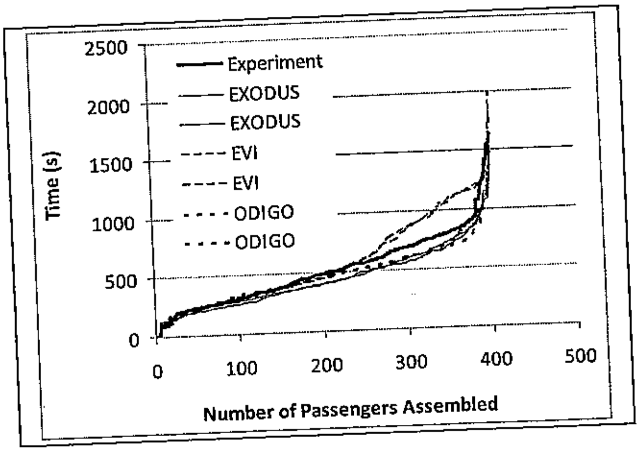
The numerical predictions producing the TAT with the smallest and greatest error are presented in Figure 9 along with the experimental data for the CS. The measured and predicted arrival curves are presented for each AS (Figure 9a to Figure 9d) and the overall arrival curve (Figure 9e). As can be seen from Figure 9, the numerical simulations under-predict the TAT for the overall assembly process and either under- (negative values) or over-predict (positive values) the assembly time for each AS.

For the maritimeEXODUS software, the simulation producing the best/worst TAT under- or over-predicts the TAT for each AS, A, B, C, D by; -0.1%/-16%, 2%/-8%, 12%/-0.3% and 4%/-5% respectively and the overall TAT is under-predicted by -0.1%/-14% (see Table 6). Thus at best, maritimeEXODUS under-predicts the TAT by 0.1%, while at worst it under-predicts the TAT by 14%. The error in predicting the TAT for each assembly station varies from -16% (under-prediction) to 12% (over-prediction).

For the EVI software, the simulation producing the best/worst TAT under- or over-predicts the TAT for each AS by; 0.4%/17%, -11%/-11%, 22%/27% and -5%/-8% respectively and the overall TAT is over-predicted by 0.4%/17% (see Table 7). Thus at best, EVI over-predicts the TAT by 0.4%, while at worst it over-predicts the TAT by 17%. The error in predicting the TAT for each assembly station varies from -11% (under-prediction) to 27% (over-prediction).

For the ODIGO software, the simulation producing the best/worst TAT under- or over-predicts the TAT for each AS by; -9%/23%, 7%/-12%, 20%/29% and 0.5%/-21% respectively and the overall TAT is over-predicted by 0%/23% (see Table 8). Thus at best, ODIGO under-predicts the TAT by 0%, while at worst it over-predicts the TAT by 23%. The error in predicting the TAT for each assembly station varies from -21% (under-prediction) to 29% (over-prediction).

From Figure 9 it is noted that for each software tool, the assembly curves for the best and worst TAT are very similar. This suggests that the primary difference between the simulations producing the best and worst TAT is the value for the TAT. This difference is likely to be driven by random variables such as the response time or starting position of the last agent to assemble. As can be seen by comparing the three model predictions for the overall assembly (Figure 9e), all three models appear to produce quite good predictions for the overall assembly process. With the exception of AS A, the ODIGO and EVI predictions for each AS appear to be quite close to each other while the predictions for maritimeEXODUS appear to agree most closely with the experimental results. As can be seen by comparing the model predictions for the CS assembly trial (Figure 9) with the predictions from the RP1 assembly trial (Figure 8), the CS predictions are significantly closer to the experimental data.



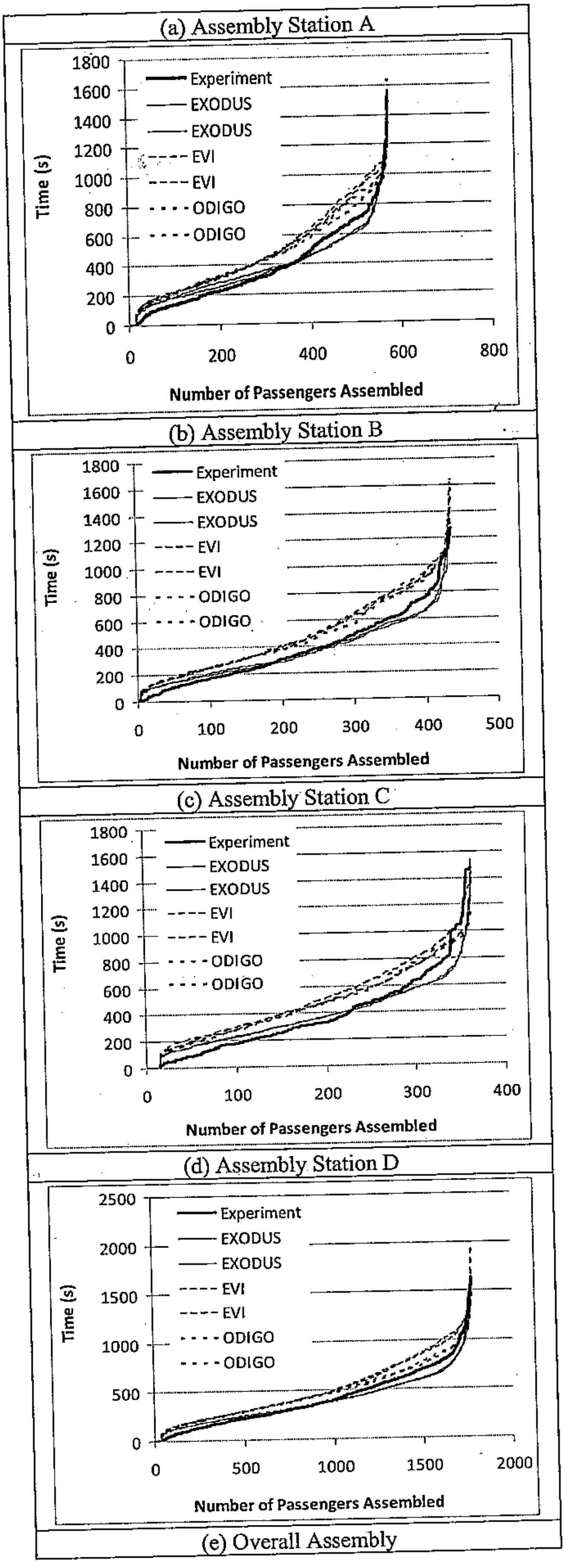


Figure 9. Comparison of model predictions (best and worst TAT) with experimental data for CS for maritimeEXODUS, EVI and ODIGO

As described earlier, there are several uncertainties introduced into the experimental data which should be considered when assessing the level of agreement between model predictions and experimental data. The uncertainty in the exact starting location of the passengers can introduce an error of 50s to 95s in the prediction of the assembly times. This uncertainty alone can introduce a possible error of some 6% in the overall TAT and an error of some 8% in the prediction of the TAT for each individual AS. The error associated with using the global response time distribution rather than the actual response time for an agent is difficult to estimate but may be appreciable. Finally, the error associated with the untagged passengers is expected to be small, and the 5s measurement error in the arrival times associated with using the IR system is considered insignificant for this trial (less than 0.4% of the TAT). Taking these uncertainties into consideration, the errors in the predicted assembly times appear very reasonable.

Furthermore, as noted for the RP1 trials, there does not appear to be a significant difference between the predicted assembly curves for the best and worst TAT. By sight, the predicted and measured assembly curves for the overall assembly appear to be in very good agreement (see Figure 9e). The predicted arrival curves for each of the AS (Figure 9a to Figure 9d) also appear to be in very good agreement with the measured curves. This suggests that the evacuation model is doing a good job of predicting the overall assembly process. Furthermore, the level of agreement with the CS data-set appears to be significantly better than that of the RP1 data-set.

8.3 VALIDATION METRIC

While all three evacuation simulation software appear to be producing reasonable predictions of the assembly process it is desirable to have objective measures of the level of agreement between predicted and measured performance rather than subjective assessments. 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.

In [17] 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 [17] have a number of typographical errors [18] 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 = (E_1, ... E_n), 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, ... 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).

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

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

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

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 in magnitude. 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 the 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 the assembly data presented in Figure 8 and Figure 9, 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 (as shown in Figure 8 and Figure 9) the 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 offset by a constant value. In general, for 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.

8.3.1 Validation Metric Applied to Numerical Predictions of SGVDS1

If the validation metric is applied to the data shown in Figure 8 it produces the values presented in Table 3 to Table 5 for maritimeEXODUS, EVI and ODIGO respectively. First consider the data relating to the overall assembly curve produced by each software tool. With an *s/n* of 0.05, the values for the SC for the best and worst TAT are quite large (> 0.7) suggesting that the shape of the overall assembly curves resembles that of the experimental data. This is consistent with the conclusion drawn from a visual inspection of Figure 8e. Note that an *s/n* of 0.05 represents 5% of the data-set and implies s = 24 for this data-set. Thus for the 480 point data-set, the gradients used in the evaluation of equation 3 are spread over 24 data points, which is not unreasonable. Furthermore, the ERD produced by each software tool for the overall assembly is reasonably low (< 0.42) and the EPC is close to 1.0 (between 0.9 and 1.1) suggesting that the overall predicted assembly curve for each software tool is reasonably close to the measured curve, again consistent with a visual inspection of Figure 8e.

It is also noted that maritimeEXODUS marginally produces the closest level of agreement with the overall assembly curve (ERD < 0.33, EPC = 1.1 and SC = 1.0), closely followed by ODIGO (ERD < 0.40, EPC = 1.0 and SC = 0.8), followed by EVI (ERD < 0.42, EPC between 0.9 and 1.0 and SC between 0.7 and 0.8) again consistent with a visual inspection of Figure 8e. However, it is noted that the metric suggests that the difference between the three models is small, consistent with a visual inspection of Figure 8e.

Next consider each AS. Here we find that for an s/n of 0.05, the SC values produced by each software tool for each AS are close to 1.0 for both the best and worst TAT cases (> 0.6 for all cases except AS A for EVI and ODIGO), suggesting that the shapes of the predicted curves are in reasonable agreement with the measured curves, again supporting the conclusions of the visual inspection. For the smallest of the AS data-sets (AS C), an s/n of 0.05 represents an s value of 4, while for the largest of the AS data-sets (AS D), this represents an s value of 9. For both the best and worst TAT cases, with the exception of AS D, the ERD values are reasonably low (< 0.5), and the EPC values are reasonably close to 1.0 (between 0.6 and 1.4) with the exception of that for AS A. These values suggest that, given the uncertainties in the data-set, and with the exception of AS D, the predicted values are reasonably close to the measured values, which again is consistent with a visual inspection of Figure 8.

Table 3. Metric values for maritimeEXODUS prediction of SGVDS1

	SC					ERD	EPC	% diff
s/n	0.01	0.03	0.05	0.07	0.09			TAT
BEST TAT								
Overall	0.8	0.9	1.0	1.0	1.0	0.31	1.1	-21.8
AS A	0.3	0.5	0.7	0.7	0.8	0.35	1.3	-21.8
AS B	0.3	0.8	0.9	1.0	1.0	0.34	1.1	-32.8
AS C	0.3	0.6	0.8	0.9	0.9	0.22	1.2	-26.9
AS D	0.7	0.9	0.9	0.9	0.9	0.54	0.7	-28.4
WORST TAT								
Overall	0.8	1.0	1.0	1.0	1.0	0.33	1.1	-33.7
AS A	0.5	0.7	0.8	0.8	0.9	0.37	1.4	-42.3
AS B	0.3	0.9	0.9	0.9	0.9	0.42	1.1	-42.8
AS C	0.2	0.5	0.7	0.7	0.8	0.23	1.2	-29.3
AS D	0.8	0.9	0.9	0.9	0.9	0.52	0.7	-15.8
BEST ERD								
Overall	0.8	0.9	1.0	1.0	1.0	0.29	1.1	-27.5
AS A	0.4	0.6	0.8	0.8	0.9	0.36	1.4	-29.4
AS B	0.7	0.8	0.8	0.9	0.9	0.38	1.2	-27.3
AS C	0.5	0.6	0.7	0.8	0.9	0.21	1.2	-24.8
AS D	0.8	0.8	0.8	0.9	0.9	0.52	0.7	-23.1

Table 4. Metric values for EVI prediction of SGVDS1

	SC					ERD	EPC	% diff
s/n	0.01	0.03	0.05	0.07	0.09			TAT
BEST TAT								
Overall	0.4	0.7	0.7	0.8	0.8	0.42	0.9	2.6
AS A	0.1	0.2	0.5	0.5	0.6	0.36	1.2	2.6
AS B	0.3	0.7	0.7	0.7	0.8	0.49	0.9	-29.3
AS C	0.4	0.7	0.9	0.9	0.9	0.35	1.4	-43.0
AS D	0.8	0.9	0.9	0.9	0.9	0.72	0.6	-23.4
WORST TAT								
Overall	0.7	0.7	0.8	0.8	0.8	0.41	1.0	-47.4
AS A	0.3	0.4	0.6	0.6	0.7	0.37	1.3	-47.4
AS B	0.5	0.7	0.6	0.7	0.7	0.50	1.0	-50.5
AS C	0.7	0.8	0.9	0.9	0.9	0.44	1.5	-54.1
AS D	0.5	0.6	0.7	0.7	0.8	0.83	0.6	-28.5
BEST ERD								
Overall	0.6	0.8	0.8	0.8	0.9	0.39	0.9	-35.0
AS A	0.4	0.6	0.7	0.7	0.7	0.35	1.2	-45.3
AS B	0.2	0.8	0.8	0.9	0.9	0.43	1.0	-34.7
AS C	0.3	0.6	0.8	0.8	0.8	0.38	1.4	-30.6
AS D	0.5	0.7	0.7	0.8	0.9	0.80	0.6	-22.9

The validation analysis should attempt to judge how closely the predictions produced by the software agree with the experimental results. This analysis is based on a representative case from the batch of 50 repeat simulations. Thus far, the representative case has been based on the best TAT i.e. the predicted curve which produces the best TAT is compared with the experimental data to determine how well the software can reproduce the experimental results. 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 which measures the difference between the two data-sets at each point. Thus, in addition to producing a metric analysis based on the best and worst TAT, the metric analysis is repeated for the simulation results producing the best ERD for the overall assembly curve. It is noted that if this curve was also plotted in Figure 8 it would also be tightly clustered together with the other curves. The conclusions drawn from the metric analysis for the best ERD are similar to the analysis for the best TAT. It is also noted that the best ERD produces a larger percentage error in the TAT than that produced for the best TAT.

Table 5. Metric values for ODIGO prediction of SGVDS1

	SC					ERD	EPC	% diff
s/n	0.01	0.03	0.05	0.07	0.09			TAT
BEST TAT								
Overall	0.6	0.8	0.8	0.9	0.9	0.38	1.0	-22.5
AS A	0.2	0.3	0.5	0.6	0.6	0.35	1.1	-22.5
AS B	0.4	0.7	0.7	0.7	0.8	0.49	1.2	-39.4
AS C	0.4	0.7	0.9	0.9	0.9	0.46	1.6	-59.7
AS D	0.7	0.9	0.9	0.9	0.9	0.72	0.6	-23.2
WORST TAT								
Overall	0.7	0.8	0.8	0.8	0.8	0.40	1.0	-47.5
AS A	0.4	0.5	0.6	0.6	0.6	0.38	1.2	-47.5
AS B	0.3	0.7	0.8	0.9	0.9	0.47	1.2	-52.0
AS C	0.4	0.7	0.8	0.9	0.9	0.49	1.7	-48.5
AS D	0.8	0.8	0.8	0.8	0.9	0.70	0.6	-29.8
BEST ERD								
Overall	0.7	0.8	0.8	0.8	0.9	0.38	1.0	-42.9
AS A	0.3	0.5	0.6	0.6	0.7	0.38	1.1	-46.3
AS B	0.2	0.7	0.7	0.8	0.8	0.51	1.2	-50.4
AS C	0.3	0.6	0.8	0.8	0.9	0.44	1.6	-45.9
AS D	0.8	0.8	0.9	0.9	0.9	0.69	0.6	-21.1

Based on this analysis, a set of acceptance criteria can be defined for SGVDS1 that takes into consideration the uncertainties in the experimental data and that confirms that the predictions produced by the three software tools presented in Figure 8 are arguably a reasonable match for the experimental data based on a visual inspection of the data. A general two-step validation protocol is suggested based in part on the philosophy of MSC 1238, which currently only focuses on the overall assembly time.

In the first step of the validation protocol, the acceptance criteria are applied to the model predictions of the overall assembly. To be deemed acceptable, the model predictions must satisfy all elements of the acceptance criteria. If successful, the second step of the validation protocol is considered. In the second step, the acceptance criteria are applied to each of the four AS with a minimum of nine passes out of a possible 12 being deemed to be acceptable. Furthermore, no more than one failure can occur in any one AS. The validation protocol and acceptance criteria are applied to the model predictions which produce the best ERD. If the protocol is applied in this manner and the software meets the criteria, it demonstrates that the software is capable of producing an acceptable level of agreement with the experimental data for the entire assembly process. The suggested acceptance criteria are as follows:

(i) ERD ≤ 0.45
(ii) $0.6 \leq$ EPC ≤ 1.4
(iii) SC ≥ 0.6 with $s/n = 0.05$
(iv) Predicted TAT for the overall assembly to be within 45% of the measured value. This criterion is only applied to step 1 of the acceptance process.

Applying the suggested validation protocol to the maritimeEXODUS data presented in Table 3, 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 D fails to meet criteria (i) 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 EVI data presented in Table 4, 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 D fails to meet criteria (i) 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 5, 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 B and D fail to meet criteria (i) and AS C fails to meet criteria (ii) but all other criteria are satisfied. As the model predictions have satisfied all four criteria in step 1, and 9 of the 12 criteria in step 2, the model is considered to have satisfied the acceptance criteria.

8.3.2 Validation Metric Applied to Numerical Predictions of SGVDS2

If the metric is applied to the data shown in Figure 9 it produces the values presented in Table 6 to Table 8. First consider the data relating to the overall assembly curve for all three cases i.e. best ERD and best/worst TAT, for all three software tools. The values for the SC suggest that the shape of the overall assembly curve closely resembles that of the experimental data (SC > 0.8), even with s/n as low as 0.01. This is consistent with the conclusion drawn from a visual inspection of Figure 9e. Note that an s/n of 0.01 represents 1% of the data-set and implies $s = 17$ for this data-set. Thus for the 1743 point data-set, the gradients used in the evaluation of equation 3 are spread over 17 data points, which is considered reasonable. Furthermore, the ERD for the overall assembly is very low (< 0.22) and the EPC is close to 1.0 (between 0.8 and 1.1) suggesting that the predicted overall assembly curves are all very close to the measured curve, again consistent with a visual inspection of Figure 9e.

It is also noted that maritimeEXODUS produces the closest level of agreement with the overall assembly curve (ERD < 0.12, EPC = 1.1 and SC > 0.9), followed by ODIGO (ERD < 0.15, EPC = 0.9 and SC > 0.8), followed by EVI (ERD < 0.22, EPC = 0.8 and SC = 0.9) again consistent with a visual inspection of Figure 9e. In addition, the overall TAT is within 2.2% of the measured value for the best ERD/TAT and within 23% for the worst TAT. This suggests that all three software tools provide a very good level of agreement with the overall assembly data.

Table 6. Metric values for maritimeEXODUS prediction of SGVDS2

	SC					ERD	EPC	% diff TAT
s/n	0.01	0.02	0.03	0.04	0.05			
BEST TAT								
Overall	0.9	1.0	1.0	1.0	1.0	0.11	1.1	-0.1
AS A	0.6	0.9	0.9	0.9	0.9	0.14	1.1	-0.1
AS B	0.9	0.9	1.0	1.0	1.0	0.11	1.0	2.2
AS C	0.7	0.8	0.8	0.9	0.9	0.12	1.1	11.8
AS D	0.5	0.8	0.9	0.9	0.9	0.18	1.1	4.1
WORST TAT								
Overall	1.0	1.0	1.0	1.0	1.0	0.12	1.1	-14.4
AS A	0.7	0.9	0.9	0.9	0.9	0.16	1.2	-16.2
AS B	0.9	1.0	1.0	1.0	1.0	0.11	1.0	-8.3
AS C	0.8	0.9	0.9	1.0	1.0	0.11	1.1	-0.3
AS D	0.6	0.8	0.9	0.9	0.9	0.18	1.1	-5.0
BEST ERD								
Overall	0.9	1.0	1.0	1.0	1.0	0.08	1.1	-2.2
AS A	0.8	0.9	0.9	0.9	0.9	0.13	1.1	-18.0
AS B	0.8	0.9	0.9	0.9	1.0	0.10	1.0	-5.7
AS C	0.8	0.8	0.9	0.9	0.9	0.10	1.1	9.5
AS D	0.8	0.9	0.9	0.9	0.9	0.15	1.0	8.7

Next consider the shape of the predicted AS arrival curves produced by each software tool. For each of the three cases, the predicted assembly curves for each AS show very good agreement with the experimental data. For an *s/n* of 0.03, the SC values for all the AS are close to 1.0, with the poorest level of agreement being SC = 0.7. This was produced by EVI for AS A in the best ERD case. This is consistent with a visual inspection of Figure 9a. These results suggest that the shapes of the predicted assembly curves produced by all three software tools are in good agreement with the measured curves, again supporting the conclusions of the visual inspection. This *s/n* value, representing 3% of the data-set, is larger than that for the overall assembly curve, but is still considered small. For the smallest of the AS data-sets (AS D), this represents an *s* value of 11, while for the largest of the AS data-sets (AS B), this represents an *s* value of 17. These observations are consistent with a visual inspection of Figure 9 which suggests that, with the exception of the EVI prediction of AS A, the shapes of all the predicted AS arrival curves are in good agreement with the shape of the measured curves.

Table 7. Metric values for EVI prediction of SGVDS2

	SC					ERD	EPC	% diff TAT
s/n	0.01	0.02	0.03	0.04	0.05			
BEST TAT								
Overall	0.9	0.9	0.9	0.9	0.9	0.22	0.8	0.4
AS A	0.7	0.8	0.8	0.8	0.8	0.19	0.9	0.4
AS B	0.8	0.9	0.9	0.9	0.9	0.27	0.8	-10.9
AS C	0.7	0.7	0.8	0.9	0.9	0.24	0.8	22.5
AS D	0.8	0.9	0.9	0.9	0.9	0.28	0.8	-5.3
WORST TAT								
Overall	0.9	0.9	0.9	0.9	0.9	0.20	0.9	17.2
AS A	0.6	0.7	0.8	0.8	0.8	0.20	0.9	17.2
AS B	0.8	0.9	0.9	0.9	0.9	0.24	0.8	-11.3
AS C	0.7	0.7	0.8	0.9	0.9	0.24	0.8	27.2
AS D	0.4	0.7	0.8	0.8	0.8	0.24	0.9	-8.1
BEST ERD								
Overall	0.9	0.9	1.0	1.0	0.9	0.17	0.9	-1.1
AS A	0.5	0.6	0.7	0.7	0.7	0.14	0.9	-1.1
AS B	0.8	0.9	0.9	0.9	0.9	0.21	0.8	-8.0
AS C	0.7	0.7	0.8	0.8	0.9	0.25	0.8	23.5
AS D	0.8	0.9	0.9	0.9	0.9	0.24	0.9	-12.3

Next consider the magnitude of the difference between the predicted and measured AS arrival curves for each software tool. The ERD values for each AS are quite low (< 0.28) and the EPC is close to 1.0 (between 0.8 and 1.2) suggesting that the predicted assembly curves for each AS are all very close to the measured curve, again consistent with a visual inspection of Figure 9.

It is also noted that maritimeEXODUS produces the closest level of agreement with the AS assembly curves (ERD < 0.18, EPC between 1.0 and 1.2 and SC > 0.8), with ODIGO (ERD < 0.27, EPC between 0.8 and 1.2 and SC > 0.8) and EVI (ERD < 0.28, EPC between 0.8 and 0.9 and SC > 0.7) producing similar levels of agreement, again consistent with a visual inspection of Figure 9.

These values suggest that the predicted values of all the AS are quite close to the measured values, with the worst TAT case producing the poorest results. These observations are again consistent with a visual inspection of Figure 9.

Table 8. Metric values for ODIGO prediction of SGVDS2

	SC					ERD	EPC	% diff TAT
s/n	0.01	0.02	0.03	0.04	0.05			
BEST TAT								
Overall	0.9	0.9	0.9	0.9	1.0	0.15	0.9	0.0
AS A	0.7	0.9	0.9	0.9	1.0	0.14	1.1	-8.7
AS B	0.9	0.9	0.9	0.9	1.0	0.20	0.9	7.1
AS C	0.7	0.7	0.8	0.8	0.9	0.27	0.8	20.2
AS D	0.4	0.7	0.8	0.8	0.8	0.27	0.9	0.5

WORST TAT								
Overall	0.8	0.9	0.9	0.9	0.9	0.15	0.9	22.5
AS A	0.5	0.7	0.8	0.9	0.9	0.12	1.2	22.5
AS B	0.9	0.9	0.9	0.9	0.9	0.22	0.8	-11.7
AS C	0.7	0.7	0.8	0.8	0.9	0.21	0.8	28.5
AS D	0.5	0.7	0.8	0.8	0.8	0.25	0.9	-20.5
BEST ERD								
Overall	0.9	0.9	0.9	0.9	0.9	0.14	0.9	-0.8
AS A	0.7	0.9	0.9	0.9	1.0	0.12	1.1	-0.8
AS B	0.9	0.9	0.9	0.9	0.9	0.21	0.8	-10.1
AS C	0.6	0.7	0.8	0.9	0.9	0.22	0.8	24.9
AS D	0.4	0.7	0.8	0.8	0.8	0.21	0.9	-7.6

Based on this analysis, a set of acceptance criteria can be defined for SGVDS2 that takes into consideration the uncertainties in the experimental data and that confirms that the predictions produced by all three software tools presented in Figure 9 are arguably good matches for the experimental data based on a visual inspection. The acceptance criteria for SGVDS2 are stricter than for SGVDS1 due to the lower expected uncertainty in the measured values. Furthermore, the validation protocol for SGVDS2 is altered slightly in that only two failures in the AS comparison is considered permissible. As with SGVDS1, the validation protocol and acceptance criteria are applied to the model predictions which produce the best ERD. The suggested acceptance criteria are as follows:

(i) ERD $\leq$ 0.25
(ii) 0.8 $\leq$ EPC $\leq$ 1.2
(iii) SC $\geq$ 0.8 with s/n = 0.03
(iv) Predicted TAT for the overall assembly to be within 15% of the measured value. This criterion is only applied to step 1 of the acceptance process.

Applying the suggested validation protocol to the maritimeEXODUS data presented in Table 6, 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 7, 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) 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 8, 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.

8.3.3 Discussion

8.3.3 (a) Sensitivity to Number of Simulations

The results presented in this paper are for blind predictions of the evacuation performance of the two vessels. By necessity, when used by other researchers, the comparisons will not be blind as the results will have been published. However, this is not considered to detract from the value of the validation data-sets. Indeed, as the geometry, starting locations of the population, population response times and population end points are specified as part of the validation data-set, and all other model parameters are specified by MSC1238, there is little opportunity to tune the evacuation model to produce ideal results. However, due to the nature of the data in the validation data-set, it is possible for users to continually run batches of 50 simulations until an appropriate best ERD case is produced i.e. one that satisfies the criteria. This is due to not knowing the exact starting location of each agent and because the precise response time for each agent is not known, thus each simulation randomly produces a different allocation of response times and precise starting locations, some of which may be more favourable than others. To explore this possibility two additional batches of 50 simulations were produced for SGVDS2 using maritimeEXODUS and the results from the metric analysis are presented in Table 9 and 10.

From the results presented in Table 9 and 10, the results for the SC for batch 3 are marginally better than for batch 1 (see Table 6) while the results for batch 2 are marginally worse than batch 1. All the SC values for batch 3 satisfy the acceptance criteria, while in batch 2, the SC for AS A fails the criteria. The ERD values for batch 3 and 2 are marginally worse than for batch 1, with all the ERD values satisfying the acceptance criteria. The EPC values for batch 2 are marginally better than those for batch 1, while batch 3 are similar to those for batch 1. The largest variation in parameters between the three batches of results occurs for the time for the last agent to assemble overall and in each AS i.e. the TAT. In batch 1, the overall TAT is under-predicted by 2.2%, while in batch 2 it is under-predicted by 7.9% and in batch 3 it is under-predicted by 7.2%. The greatest difference in the AS TAT occurs for AS D, where batch 1 over-predicts by 8.7% while batch 2 under-predicts the

TAT by 1.2% and batch 3 under-predicts by 2.3% - a maximum difference of some 11%. However, as this criteria is only applied to the overall assembly results, all three cases are considered acceptable. Nevertheless, the large variation in the TAT for the AS demonstrates that the TAT is not a reliable measure, especially for validation purposes. Due to the random allocation of precise starting location and response times, it is possible that an agent is assigned a starting location which results in the furthest possible travel distance and the longest possible response time creating an abnormally long TAT. Furthermore, should that agent be associated with the AS that takes longest to assemble it could severly impact the overall TAT. This is why the percentage difference in the TAT criteria should not be applied to individual ASs, and if it is used at all, it should only be applied to the overall TAT.

Based on the metric values, while there are some differences in the precise values for the three components of the metric, the same conclusion with respect to acceptability would be made. Arguably, the results for batch 3 are marginally the best, while the results for batch 2 are marginally the worst. However, these results cannot be generalised to other software tools and so there is some room for users to optimise their results.

Table 9. Metric values for maritimeEXODUS prediction of SGVDS2 – batch 2 best ERD

	SC					ERD	EPC	% diff
s/n	0.01	0.02	0.03	0.04	0.05			TAT
Overall	0.8	0.8	0.9	1.0	1.0	0.09	1.0	-7.9
AS A	0.4	0.4	0.5	0.5	0.6	0.16	1.0	-6.6
AS B	0.9	0.9	0.9	0.9	1.0	0.10	1.0	-12.0
AS C	0.7	0.9	0.9	1.0	1.0	0.11	1.0	16.9
AS D	0.7	0.8	0.9	0.9	1.0	0.13	1.0	-1.2

Table 10. Metric values for maritimeEXODUS prediction of SGVDS2 – batch 3 best ERD

	SC					ERD	EPC	% diff
s/n	0.01	0.02	0.03	0.04	0.05			TAT
Overall	1.0	1.0	1.0	1.0	1.0	0.09	1.1	-7.2
AS A	0.8	0.9	0.9	0.9	0.9	0.14	1.1	-10.7
AS B	0.9	1.0	1.0	1.0	1.0	0.09	1.0	-1.8
AS C	0.8	0.8	0.9	1.0	1.0	0.10	1.0	19.3
AS D	0.8	0.9	0.9	0.9	0.9	0.15	1.1	-2.3

8.3.3 (b) Proposed Validation Data-Sets and Validation Protocols for Incorporation into IMO Evacuation Analysis Guidelines

Two validation data-sets are proposed for inclusion into the IMO evacuation analysis guidelines. It is proposed that before an evacuation simulation tool is considered for use in ship evacuation certification analysis it must demonstrate that it can satisfy the requirements of the proposed validation protocol.

As part of the validation protocol, all information required to setup the evacuation analysis will be provided on a website [19]. This includes: CAD layout of vessel, starting location of passengers, end location of passengers, passenger response time distribution and the assembly curves for each assembly station and the overall assembly process. All other parameters required to perform the simulations will be extracted from IMO MSC Circ 1238.

Based on the analysis presented in this paper, 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 phases:
    - 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, 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 0.6 ≤ EPC ≤ 1.4 SC ≥ 0.6 with S/n = 0.05 % TAT < 45% (Phase 1 only)	ERD ≤ 0.25 0.8 ≤ EPC ≤ 1.2 SC ≥ 0.8 with S/n = 0.03 % TAT < 15% (Phase 1 only)

9. CONCLUSIONS

Data from two semi-unannounced assembly trials at sea for a RO-PAX passenger ferry and a cruise ship have been collected consisting of passenger: response time data, starting locations and arrival time at the designated assembly stations. The response time data was collected using digital video cameras while the start and end locations and the arrival time for the passengers was collected using a novel Infra-Red (IR) data acquisition system consisting of ship-mounted IR beacons 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 have been 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. The proposed acceptance criteria take into consideration uncertainties associated with the measured data in each of the data-sets.

In blind applications of the validation protocol to three commonly used ship evacuation software tools (maritimeEXODUS, EVI and ODIGO), each software tool was found to satisfy the acceptance criteria for each data-set, suggesting that it is capable of predicting the outcome of the assembly process for these two vessels to the specified level of accuracy as defined by the acceptance criteria.

It is proposed that the suggested validation protocol and the acceptance criteria could be used by IMO as part of a validation suite to determine acceptability of maritime evacuation models in a future enhancement to MSC1238. In this way we hope 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.

10. ACKNOWLEDGEMENT

The SAFEGUARD project (contract 218493) was funded under the European Union Framework 7 Transport initiative. The authors acknowledge the co-operation of their project partners.

11. REFERENCES

1. IMO, 'Interim Guidelines for Evacuation Analyses for New and Existing Passenger Ships', *IMO MSC/Circ 1033*, 6 June 2002.
2. IMO, 'Guidelines for Evacuation Analysis for New and Existing Passenger Ships', *IMO MSC/Circ 1238*, 30 Oct 2007.
3. IMO, 'FP 51/WP.3', *Fire Protection Sub-Committee, 51st session, Work Package 3*, 8 Feb 2007.
4. 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).
5. DEERE, S., GALEA, E.R., LAWRENCE, P., FILIPPIDIS, L. and GWYNNE, S., 'The impact of the passenger response time distribution on ship evacuation performance', *International J of Maritime Eng, Vol 148, Part A1, 35-44*, (2006).
6. GALEA, E.R., LAWRENCE, P., GWYNNE, S., SHARP, G., HURST, N., WANG, Z., and EWER, J., 'Integrated fire and evacuation in maritime environments', *Proc of the 2nd Int Maritime Safety Conference on Design for Safety, Sakai Japan, Publisher Ship and Ocean Foundation, pp 161-170*, 27-30 Oct 2004.
7. BOXALL, P., GWYNNE, S., FILIPPIDIS, L., GALEA, E.R., and COONEY. D., 'Advanced Evacuation Simulation Software and its use in Warships', *Proc of the Human Factors in Ship Design, Safety and Operation, London UK, Publisher The Royal Institute of Naval Architects, pp 49-56*, 23-24 Feb 2005.
8. CALDEIRA-SARAIVA, F., GYNGELL, J., WHEELER, R., GALEA, E.R., CARRAN, A., SKJONG, R., VANEM, E., JOHANSSON, K., RUTHERFORD, B., and SIMOES, A.J., 'Simulation of ship evacuation and passenger circulation', *Proc 2nd Int Maritime Safety Conference on Design for Safety, Sakai Japan, Publisher Ship and Ocean Foundation, pp 197-205*, 27-30 Oct 2004.
9. DEERE, S.J., GALEA, E.R., and LAWRENCE, P., 'A Systematic Methodology to Assess the Impact of Human Factors in Ship Design', *Applied Mathematical Modelling, Applied Mathematical Modelling, 33, 867-883, <http://dx.doi.org/10.1016/j.apm.2007.12.014>*, 2009.

10. ANDREWS, D.J., PAWLING, R., CASAROSA, L., GALEA, E.R., DEERE, S., and LAWRENCE, P., 'Integrating Personnel Movement Simulation into Preliminary Ship Design', *International Journal of Maritime Engineering, Volume 150 Part A1 pp 19-34, ISSN 1479-8751, http://www.rina.org.uk/ijme0801.html*, 2008.
11. GALEA, E.R., DEERE, S., SHARP, G., FILIPPIDIS, L., LAWRENCE, P., and GWYNNE, S., 'Recommendations on the nature of the passenger response time distribution to be used in the MSC 1033 assembly time analysis based on data derived from sea trials.' *International J of Maritime Eng, Vol 149, Part A1, 15-29*, (2007).
12. VASSALOS, D., KIM, H., CHRISTIANSEN, G., MAJUMDER, J., 'A Mesoscopic Model for Passenger Evacuation in a Virtual Ship-Sea Environment and Performance-Based Evaluation', *Pedestrian and Evacuation Dynamics – April 4-6, 2001 – Duisburg. pp369-391. ISBN: 3-540-42690-6*, (2001).
13. VASSALOS, D., GUARIN, L., VASSALOS, G.C., BOLE, M., KIM, H.S., and MAJUMDER, J., 'Advanced Evacuation Analysis – Testing the Ground on Ships.' *International Conference in Pedestrian and Evacuation Dynamics (PED) – 2003 – Greenwich. pp147-158, ISBN: 1-904521-08-8*, (2003).
14. VASSALOS, D., GUARIN, L., BOLE, M., MAJUMDER, J., VASSALOS, G.C., and KIM, H.S., 'Effectiveness of passenger evacuation performance for design, operation & training using first-principles simulation tools.' Proceedings of, 1st International Escape, Evacuation and Recovery Conference, March 2004. Inmarsat, London, (2004).
15. PRADILLON, J.Y., 'ODIGO - Modelling and Simulating Crowd Movement onboard Ships', *3rd International Conference on Computer and IT Applications in the Maritime Industries, COMPIT, Siguenza, Spain, pp278-289*, 2004.
16. PRADILLON, J.Y., 'ODIGO: A Crowd Movement Simulation Tool for Passenger Vessels', *2nd International Conference on Computer and IT Applications in the Maritime Industries, COMPIT, Hamburg, Germany, pp 108-117*, 2003.
17. 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).
18. PEACOCK, R.D., *Private Communication with E.R.Galea*, 23 November 2011.
19. http://fseg.gre.ac.uk/validation/ship_evacuation

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