The need for a verification and validation protocol for evacuation models

Ronchi, Enrico

Published in:
[Host publication title missing]

DOI:
10.13140/2.1.4893.3127

2014

THE NEED FOR A VERIFICATION AND VALIDATION PROTOCOL FOR EVACUATION MODELS

Enrico Ronchi

Department of Fire Safety Engineering, Lund University
P.O. Box 118
Lund, 22100, Sweden
e-mail: enrico.ronchi@brand.lth.se

ABSTRACT

To date, there is no standard Verification and Validation (V&V) protocol for the evaluation of evacuation model predictions. This paper is intended to open a discussion on the main issues associated with the definition of a standard procedure for the V&V of building fire evacuation models. Examples of such issues are discussed, namely 1) the definition of tests able to investigate the capability of evacuation models of representing emergent behaviours, 2) a discussion on the methods employed to study the uncertainty caused by the use of distributions or stochastic variables in evacuation modelling, 3) the definition of the acceptance criteria of a standard V&V protocol.

1. INTRODUCTION

Evacuation modelling is a growing field and research and commercial software are constantly released on the market (Ronchi and Kinsey, 2011). This may be associated with an increasing commercial interest towards their possible applications as well as being a sign of an active research community.

The development of today's most used evacuation modelling tools has been relatively recent, with the majority of the most used tools being developed in the last 20 years (Kuligowski et al., 2010). Modelling techniques have been rapidly evolving towards a higher level of sophistication, i.e. during the passage of time models have adopted an increasing level of resolution in the representation of building geometries (from coarse network models to continuous and hybrid models) (Frantzich et al., 2007).

Although over the years more and more attention has been paid to the "human factor", intended here as the study of the theories on human behaviour in fire, many of today's models are mostly inspired by the movement of fluids rather than humans (Moussaid et al., 2011). This approach makes it difficult to capture and represent all possible behaviours. An accepted standard V&V protocol is therefore needed to understand if current model predictions are a sufficient representation of reality.

In this paper, verification is intended as "*the process of determining that a calculation method implementation accurately represents the developer's conceptual description of the calculation method and the solution to the calculation method*" (International Standards Organization, 2008). Validation is defined as the "*process of determining the degree to which a calculation method is an accurate representation of the real world from the perspective of the intended uses of the calculation method*" (International Standards Organization, 2008).

One of the main issues of evacuation modelling is the lack of a standard V&V protocol. Many research efforts have been made to define appropriate tests and procedures for assessing the capabilities of evacuation models. This research has been conducted by different parties, such as model developers (Galea et al., 1997; Hostikka et al., 2007; Thompson and Marchant, 1995; Thunderhead Engineering, 2014), research groups (Lord et al., 2005; Ronchi et al., 2013a; Wagoum et al., 2012) and International Organizations (International Maritime Organization, 2007; International Standards Organization, 2008). Each contribution represents an important step towards a better assessment of the capabilities of evacuation models. Nevertheless, the lack of a standard V&V protocol creates a relevant problem to the model users: they are in the difficult position of choosing among different models trying to compare features and capabilities that have been tested and documented differently.

From a research perspective, models should be chosen and evaluated in accordance to their predictive capabilities and suitability for the intended use rather than other factors such as cost, marketing strategies adopted by the developers, etc. A survey on evacuation models (Ronchi and Kinsey, 2011) confirmed this point of view, showing that model users (practitioners, researchers, etc.) consider V&V as the most important factor when selecting a model. How can a user evaluate this factor if model V&V is conducted in different ways? A document of the

National Institute of Standards and Technology (Ronchi et al., 2013a) investigated those issues and raised a set of fundamental questions:

1) What are the tests and procedures that should be included in a standard V&V protocol to assess the accuracy of model predictions?
2) Who should decide which V&V tests and procedures should be performed?
3) How should the acceptance criteria in a standard V&V protocol be defined?
4) Who should define the acceptance criteria for assessing the V&V of a model?
5) Who should perform the V&V tests? Should be the model developers, the model users or a third party?

The answers to these questions require a broad debate among all parties involved in the evacuation modelling community. Subsequently, this paper aims at discussing these questions rather than responding to them. A debate on evacuation model V&V is ongoing and several points of views have been discussed in different fora, research publications, social networks, conferences, etc.

It should also be noted that the definition of a standard V&V protocol can be considered as a starting point for an efficient use of evacuation models. In fact, the *user effect* (Ronchi, 2013) - intended as the impact of the user's interpretation of results as well as the user's input calibration - may also play a fundamental role in the calculation, analysis and interpretation of evacuation model results.

The present work discusses the need for a V&V protocol and presents two explanatory examples of issues that should be addressed, namely 1) the definition of a complete list of tests for the verification and validation of the methods adopted by evacuation models to represent emergent behaviours, 2) the assessment of methods able to investigate the uncertainty associated with the modelling techniques used to represent the stochastic nature of human behaviour. This second issue is associated with the representation of human behaviour in evacuation models using distribution laws or stochastic variables.

2. CURRENT STATE OF V&V TESTS: THE IMO TESTS

This section discusses the definition of V&V tests for the analysis of the predictive capabilities of evacuation models. To date, the tests of the International Maritime Organization (IMO tests) provided within the MSC/Circ. 1238 (International Maritime Organization, 2007) are the most used verification tests. Model developers (Gwynne et al., 1998; Hostikka et al., 2007; Thompson and Marchant, 1995; Thunderhead Engineering, 2014) generally use the IMO tests as part of their V&V procedure, often coupling them with additional verification and validation tests.

The IMO tests are often employed for the verification of models designed for any context of use despite they have been specifically designed for maritime applications. Population demographics are based on maritime populations and the egress components and strategies under consideration are based on ship evacuation procedures. For instance, stair evacuation generally occurs upwards in ships rather than the case of downward evacuation in buildings.

It should be noted that the IMO tests do not include validation tests. For this reason, model developers and testers often adopt different experimental data-sets for model validation.

Different research efforts have been made in order to review and expand the tests presented by the IMO to different applications. For example, the list of tests available in the MSC/Circ1238 (International Maritime Organization, 2007) has been reviewed during the RIMEA project (Meyer-König et al., 2007). However, the modifications suggested by the RIMEA group (which are focused on building evacuation rather than ship evacuation) do not include validation tests and they do not include testing of many of the features available in today's evacuation models (Kuligowski et al., 2010).

Modifications on the MSC/Circ1238 (International Maritime Organization, 2007) have been proposed during the SAFEGUARD project (Galea et al., 2012). This project was focused on ship evacuations, thus the applicability of the suggested modifications may not be suitable for context different than maritime (e.g. buildings).

Ronchi et al (2013a) reviewed the tests in the MSC/Circ1238 (International Maritime Organization, 2007) and provided a modified list of tests for building fire evacuation models. This list is based on the analysis of five main elements (Gwynne et al., 2013) in evacuation modelling, namely 1) pre-evacuation time, 2) movement and navigation, 3) exit usage, 4) route availability and selection, and 5) flow conditions/constraints.

The present paper uses some examples of suggested tests and methods provided in Ronchi et al (2013a) to discuss the need for standard V&V protocol for evacuation models.

3. TESTING EMERGENT BEHAVIOURS

Data on human behaviour are scarce in both quantity and quality (Averill, 2011). This issue affects the development of a standardized list of V&V tests. In fact, an immediate consequence is the lack of a "*robust, comprehensive and validated conceptual model of occupant behaviour during building fires*" (Kuligowski, 2011) which can be used as benchmark for the evaluation of the model predictions.

Data-sets potentially suitable for validation are generally chosen in relation to different factors such as their availability to the public, documentation, and the data collection/analysis method employed.

Ronchi et al (2013a) recommended a categorization of tests in relation to the core behavioural elements included in evacuation models as discussed by Gwynne et al (2013). These core elements can be compared against ideal cases or experimental data. Ideal cases are intended as simple evacuation scenarios for which expected results can be obtained by evidence or simple mathematical equations. Ideal cases are currently employed in different V&V procedures (Galea et al., 1997; International Maritime Organization, 2007; Meyer-König et al., 2007).

Ideal cases are particularly useful to perform the study of emergent behaviours. The use of this type of tests may be needed in case of lack of experimental data suitable for a comparison with model predictions. The results of the models are therefore compared with expected behaviours which are in line with the currently accepted theory representing certain behaviours of the occupants during fire emergencies (this is also called in this paper *qualitative verification*).

Two important questions should be addressed in order to define a standard V&V protocol for evacuation models:

1) What are the features that should be included in an evacuation model?
2) Which theories are widely accepted by the evacuation community? Do current evacuation models include them? Do they reflect the current understanding on evacuation behaviour?

The first question is linked to the context of use of the model. The features needed in a model can be indeed significantly different in relation to its possible applications. The second question leads to more general considerations such as the need for methods to evaluate the theoretical background of a model in order to perform its validation.

In the following section, two tests are presented in order to provide examples of issues associated with these aspects. Examples are presented in order to discuss emergent behaviours which can be included in a standard V&V protocol for building evacuation models, but which need to be discussed in the evacuation modelling community. These tests have been previously presented by Ronchi et al (2013a) and they include the description of the geometry, scenarios, expected results, test method and user actions. The tests are mostly based on the IMO tests but the features under consideration are expanded, in line with the model review made by Kuligowski et al (2010).

3.1. Example 1: People with movement disabilities

The first example presented in this paper is the analysis of the evacuation of people with movement disabilities. Different studies have shown that people with disabilities represent a significant part of the world population, such as around 12% in the US (Erickson et al., 2010), between 14-17% in the UK (Boyce et al., 1999) and between 10-20 % in Europe (Kecklund et al., 2012).

Evacuation models can be used to study total evacuation strategies, in contrast with the traditional approach in which mobility-impaired occupants were instructed to wait in areas of refuge (Kuligowski et al., 2013). In case of total evacuation, people with disabilities can have a significant impact on the evacuation process (in particular on the total evacuation time), due to different issues such as the inability to access certain egress components or reduced walking speeds (Ronchi and Nilsson, 2013a).

This leads to the need for evacuation modelling tools able to take into consideration people with movement disabilities (Kuligowski et al., 2010). The possible variability of people impairments make it difficult to define a single test able to comprehensively investigate the model capabilities in representing the behaviours of people with mobility impairments. Nevertheless, as a first step, a test has been designed by Ronchi et al (2013a) for the verification of simple emerging behaviours.

The test aims at testing the possibility of simulating an occupant with reduced mobility (e.g. decreased travel speeds and increased space occupied by the occupants) as well as representing the interactions between impaired individuals and the rest of the population and the environment

3.1.1 Test description

This section presents the test on people with movement disabilities by Ronchi et al (2013a).

Geometry

Construct two rooms at different heights, namely room 1 (1 m above the ground level) and room 2 (at ground level), connected by a ramp (or a corridor/stair if the model does not represent ramps). Insert one exit (1 m wide) at the end of room 2 (see Figure 1a-1b for the schematic representation of the rooms).

Scenarios

Scenario 1: Room 1 is populated with a sub-population consisting of 24 occupants in zone 1 (with an unimpeded walking speed of 1.25 m/s and the default body size assumed by the model) and 1 disabled occupant in zone 2 (the occupant is assumed to have an unimpeded walking speed equal to 0.8 m/s on horizontal surfaces and 0.4 on the ramp (see Figure 1a). The disabled occupant is also assumed to occupy an area bigger than half the width of the ramp (>0.75 m) (e.g., a wheelchair user). All occupants have to reach the exit in room 2.

Scenario 2: Re-run the test and populate zone 2 with an occupant having the same characteristics of the other 24 occupants in zone 1 (i.e. no disabled occupants are simulated, see Figure 1b). All occupants have to reach the exit in room 2.

Scenario 1

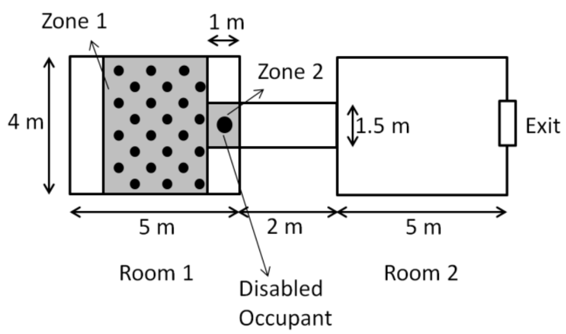

Figure 1a. Schematic geometric layout of the test on people with movement disabilities as presented by Ronchi et al (2013a).

Scenario 2

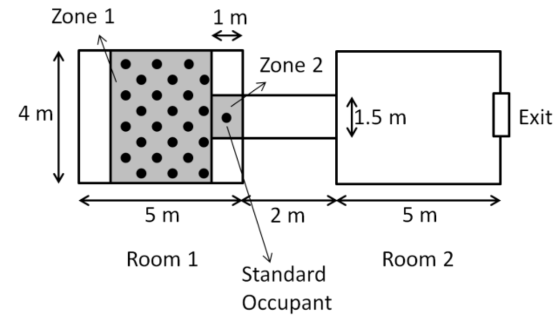

Figure 1b. Schematic geometric layout of the test on people with movement disabilities as presented by Ronchi et al (2013a).

Expected result

The expected result is that occupants in zone 1 in Scenario 1 reach the exit in a time slower than occupants in zone 1 in Scenario 2. If possible, this qualitative verification can be performed using the visualization tool of the model.

Test method

The test is a qualitative verification of emergent behaviours. The tester should qualitatively evaluate if the model is able to simulate disabled populations and their possible impact on the evacuation times.

User actions

If the model under consideration does not permit the simulation of people with movement disabilities or it does not permit the simulation of agents of different dimensions, the tester is recommended to discuss this limitation in the documentation associated with the V&V of the model.

3.1.2 Discussion on the test

The test presented in the previous section can be used to discuss some of the issues about the simulation of people with disabilities. Three exemplary issues are discussed in this paper.

First, this test is deliberately aimed at evaluating only the simulation of people with specific mobility impairments (e.g. people on a wheelchair), and the inclusion of model features able to reproduce further types of impairments (Kinateder et al., 2014) should be discussed, i.e. it is important to assess what behaviours cannot be predicted with the model.

Second, there is a need to define the underlying theories that can be used as reference to model group interactions (e.g., social influence, cooperative behaviours, etc.) between an occupant with disability and the rest of the population of the building.

Third, there is a need to discuss the implementation of a simulation framework able to represent the accessibility of people with mobility impairments to different egress components.

This example highlights the complexity of the definition of the tests that should be included in a standard V&V protocol for the features for which both limited experimental data-sets are available, as well as there is not an accepted model to represent the variability of the possible behaviours that may take place.

3.2. Example 2: Simulation of affiliation

The simulation of *affiliation* is the second example presented in this paper to discuss the need for a broad discussion on the tests to be included in a standard V&V protocol to evaluate emergent behaviours. The concept of affiliation has been introduced by Sime (1985) and it relates to the likelihood of a person to use familiar routes/exits over unfamiliar ones during evacuation. For instance an occupant may try to evacuate a building through the route he/she used to enter it.

Ronchi et al (2013a) suggest a test aimed at qualitatively verifying the capabilities of evacuation models to simulate the effect of an individual's familiarity with an exit on exit usage. Such a test requires an exit choice sub-model able to represent (implicitly or explicitly) the affiliation with the exits. This example is used here to discuss the fact that a test can be interpreted as a validation of emergent behaviours or analytical verification depending on the type of sub-algorithm available in the model (deterministic or probabilistic).

3.2.1 Test description

This section presents the test on affiliation by Ronchi et al (2013a).

Geometry

Construct a room of size 10 m by 15 m. Two exits (1 m wide) are available on the 15 m walls of the room and they are equally distant from the 10 m long wall at the end of the room (see Figure 2).

Scenarios

Scenario 1: Insert an occupant in the room with a response time equal to 0 s and a constant walking speed equal to 1 m/s as shown in Figure 2 (the black dot represents the occupant which is 1 m away from the 10 m long wall on the bottom of Figure 2). The occupant should always be placed in the same position among different runs and his/her position should be equidistant to both exits. The occupant is assumed to be unfamiliar with the exits. Run the test several times until you get a stable percentage of exit usage for both exits i.e., exit usage does not vary more than 1% with an additional run. Annotate the exit usage for the two exits

Scenario 2: Insert an occupant in the central area at the beginning of the corridor with an instant response time and a constant walking speed equal to 1 m/s as shown in Figure 2. This occupant is affiliated with Exit 2. The same occupant is not affiliated with Exit 1 (e.g. Exit 2 is the favoured exit chosen by the occupant if all the other conditions affecting choice are the same for all exits). Run the test several times until you get a stable percentage of exit usage for both exits i.e., exit usage does not vary more than 1% with an additional run. Annotate the exit usage for both exits.

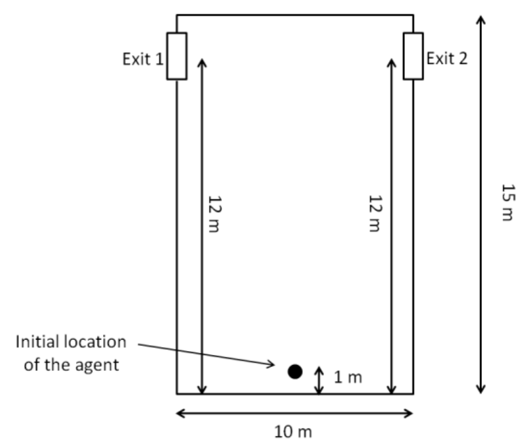

Figure 2. Schematic geometric layout of the test on affiliation as presented by Ronchi et al (2013a).

Expected result

The expected result is that the usage of exit 2 in scenario 2 is higher than the exit 2 usage in scenario 1.

Test method

The evaluation method of this test is a quantitative verification of model results in terms of exit usage.

User actions

The model tester should document if the model includes a dedicated algorithm for the simulation of affiliation and if the exit choice sub-model is based on deterministic assumptions (i.e. user defined percentage of exit usage) or if it includes a predictive sub-algorithm.

3.2.2 Discussion on the test

The test on affiliation can be used as an example to discuss the use of deterministic or probabilistic algorithms within evacuation models and the subsequent methods to be used for their V&V.

For example, the sub-algorithm embedded in a model to represent exit choice may be based on a user-defined assignment of probabilities of exit usage or on an implicit probability of exit usage based on the characteristics of the agents and their interactions (e.g., affiliation, social influence) or external conditions (e.g. visibility of the doors, line of sights).

While the external conditions could be represented with a sufficient level of accuracy, to date, there is not a single accepted theory for the representation of the characteristics of individuals which affects exit choice and their interactions with other people.

Today, the results of models may depend on the use of a certain theory or a data-set to represent exit usage. But which and how many behavioural theories should be implemented for the simulation of exit choice? How should this be made? What experimental data-sets (or data from real case studies) can be used as a benchmark for model predictions?

The test presented above can be interpreted as a test of emergent behaviours if the model already includes a sub-algorithm for the representation of affiliation (assuming that affiliation is an accepted theory to represent exit choice) or simple analytical verification if the user assigned manually a deterministic probability of exit usage. This dual interpretation is reflected in the need for a standard V&V protocol which also addresses the documentation to be produced by the model developers. In fact, it should be clearly stated what is actually "predicted" by the model and what instead the model "can or cannot predict" if supported by experimental data or theory for the calibration of the model input.

4. THE ANALYSIS OF BEHAVIOURAL UNCERTAINTY

This section discusses an issue which has not been addressed in current V&V procedures, i.e., the study of *behavioural uncertainty* (Ronchi et al., 2013b). Uncertainties in the context of fire safety engineering and modelling are generally categorized into three different components, namely *model input uncertainty*, *measurement uncertainty*, and *intrinsic uncertainty* (Hamins and McGrattan, 2007).

Model input uncertainty depends on the parameters used as input of the model which are obtained from experimental measurements. *Measurement uncertainty* is linked with the experimental measurement itself. *Intrinsic uncertainty* depends on the physical and mathematical assumptions and methods used for model formulation.

Modelling evacuation behaviour presents an additional component of uncertainty which depends on the methods adopted to simulate the possible variability of human behaviour, namely *behavioural uncertainty*. Behavioural uncertainty is linked with the use of stochastic variables (e.g. the use of distribution laws) employed in evacuation modelling to represent human behaviour. The concept is based on the fact that a single experiment or model run is not representative of the possible range of occupant behaviours that may take place.

A group of people evacuating a building can be used as an example to explain the different types of uncertainties. The uncertainty which depends on the measurements of the walking speeds is the *measurement uncertainty*. The distribution employed to approximate the occupant speeds in the model input is the *model input uncertainty*. The uncertainty which is linked with the calculation method to represent the movement of the occupants is the *intrinsic uncertainty*. The fact that occupants, under identical (or very similar) conditions, may have behaved differently (e.g. walking faster or slower) is the *behavioural uncertainty*.

The use of stochastic variables is associated with the inability to confidently represent within a modelling framework all cues and factors affecting human behaviour. This is reflected in different methods of agent representation (e.g., the use of probabilistic variables or distribution laws). This choice of the model developers can be interpreted in two different ways. The first interpretation is that the "human element" introduces factors that are not entirely predictable. A second interpretation may instead rely on the fact that given the limited knowledge on human behaviour, currently there is not enough information to predict human response with any degree of certainty using a deterministic approach.

The use of random sampling methods to generate input distributions lead to multiple occupant-evacuation time curves for the same scenario using the same model inputs. In addition, random variables can be implicitly included in a model algorithm, thus not permitting the user to control/access to them. For this reason, a V&V protocol should include a method to study the variability of results associated with the random variables embedded in the models, i.e. the impact of behavioural uncertainty on model results. This is needed in order to capture the possible impact of the variability of occupant behaviour on results.

To date, the levels of sophistication employed by model users/testers to address this issue vary significantly. They may range from a qualitative evaluation of the number of runs to be simulated to a quantitative treatment of the problem.

The purpose of the evacuation simulations and the subsequent results of interest play also an important role on this matter. In the case of applications in the context of Fire Safety Engineering for Performance Based Design, evacuation models are often employed to calculate the RSET (Required Safe Egress Time) or total evacuation time, which corresponds to the time needed by the last agent in the model to leave the building. This time is crucial since it is then compared to the Available Safe Egress Time in order to assess the level of safety of a building. Nevertheless, in some cases, the sole information about the RSET could not represent the only scope of the application of a model and there could be a need for a more detailed analysis of the model evacuation results (Galea et al., 2012). This may be the case of the evaluation of evacuation procedures and strategies. For instance, the result of interest can be the whole occupant-evacuation time curve, i.e. a complete analysis of the time needed by each occupant to evacuate the building. This issue should be addressed in model V&V procedures, in which the comparison between an experimental occupant-evacuation time curve and the curve produced by an evacuation model is investigated.

The convergence of the repeated simulations towards an "average behaviour" (expressed as an average RSET or an average occupant-evacuation time curve) can be used to assess the impact of behavioural uncertainty on model results. Four different methods for the assessment of the number of repeated runs of an evacuation scenario are discussed in this paper:

1) The use of an arbitrary number of runs for a V&V study (International Maritime Organization, 2007)
2) The assessment of the number of runs for a V&V study based on simple acceptance criteria, i.e. accepted errors (Ronchi and Nilsson, 2013b)
3) The assessment of the number of runs for a V&V study based on functional analysis operators (Ronchi et al., 2013b)
4) The assessment of number of runs for a V&V study based on functional analysis operators and inferential statistics (Lovreglio et al., 2014).

The use of different methods is reflected in the assessment of behavioural uncertainty during the process of V&V of evacuation models.

The simplest method consists of the simulation of an arbitrary fixed number of runs and the use of the average total evacuation time and its standard deviation to perform the comparison (sometimes including confidence intervals (Kling et al., 2012)). This method is widely used in today's evacuation model applications and validation tests (International Maritime Organization, 2007). It presents the immediate advantage to require a low modeller effort and it is recommended within IMO Guidelines (International Maritime Organization, 2007). The main limitation of this approach is that results may be affected by the specific impact of the randomness of the modelling assumptions adopted (models with a higher "degree of randomness" may be very sensitive to the chosen number of runs). In addition, this method does not provide a quantitative investigation of behavioural uncertainty.

An alternative method relies on the estimation of the number of simulations to be performed based on the error of two consecutive averaged total evacuation times (or based on the study of average standard deviations) (Ronchi and Nilsson, 2013b). The assessment of the number of runs to be used for the comparison with benchmark data is made using pre-defined accepted errors. In order to extend this approach to the study of the entire occupant-evacuation time curve, this method can be easily expanded to different percentages of evacuees (e.g. studying when 25%, 50%, 75%, 98%, etc. of the population has left the building). This method represents a higher level of sophistication if compared with the simulation of an arbitrary number of runs, but its efficiency is strongly dependent on the choice of the accepted error.

The third method currently available to study behavioural uncertainty is a method developed by Ronchi et al (2013b) based on the use of functional analysis operators. The method is based on a set of criteria developed in order to study the convergence of the full simulated occupant-evacuation time curves. The curves are studied using vector operators (the Euclidean Relative Distance, the Euclidean

Projection Coefficient and the Secant Cosine) and the method permits estimating an average curve with pre-defined accepted errors (which may depend on the scope of the analysis). This approach relies on the law of large numbers (the central limit theorem), i.e., it is based on the concept that the impact of behavioural uncertainty on average behaviour tends to decrease with an increasing number of runs. Also in this case, the choice of the accepted error is critical for the assessment of the number of runs to be used for the study of the average occupant-evacuation time curve. The main advantage of this method is that it allows studying the entire occupant evacuation curve rather than the total evacuation time only or a limited sub-set of evacuee percentages.

The last method has been presented by Lovreglio et al (2014) and it is an extension of the method developed by Ronchi et al (2013b). It relies on a combination of the use of functional analysis operators and inferential statistics. In this case, the choice of the number of runs for a V&V study is made performing a comparison between experimental/theoretical and simulated behavioural uncertainties.

The methods described above present different advantages and limitations. From a model validation perspective, different approaches can be used to perform comparisons with the benchmark data, namely 1) the sole study of the total evacuation times, 2) the use of the best/worst model estimation for the occupant-evacuation time curve, or 3) the average occupant-evacuation time curve produced by the model. In all cases, the assessment of the impact of behavioural uncertainty should be included. In this context, a standard V&V protocol is needed in order to assess the appropriate method for the definition of the number of runs for the comparison in relation to the type of V&V study conducted.

The impact of behavioural uncertainty on evacuation results may be scenario-dependent, but its assessment is crucial for a complete understanding of the evacuation process. Nevertheless, multiple data-sets of a single evacuation scenario are rarely available in the literature, thus making it difficult to assess behavioural uncertainty experimentally and perform accurate validation studies which fully take this issue into consideration.

Single data-sets are often the only available reference for a study of an individual scenario. Model testers should ideally use a range of evacuation scenarios under similar condition and evaluate if the model prediction produce results which are in line with the range of real-world outcomes. Given the limited availability of data, model testers often rely on a single real-world observation without having a deep understanding on the possible impact of behavioural uncertainty on that scenario, i.e. if that curve is representative or not of the average behavioural performance. It should also be noted that the study of the occupant-evacuation time curve could also include the analysis of the tails of the distribution rather than the analysis of the average values only.

5. THE DEFINITION OF ACCEPTANCE CRITERIA

One of the critical aspects associated with the definition of a standard V&V protocol for evacuation models is the definition of acceptance criteria. This requires a broad debate in the evacuation modelling community since there are many issues to be solved such as:

1) the definition of acceptance criteria can be dependent on the intended use of the models.
2) the definition of acceptance criteria is difficult given the lack of knowledge on certain aspects of human behaviour and the subsequent uncertainty associated with human factors data-sets.
3) Several parties are involved in the evacuation modelling community (model developers, users, international organizations, regulators etc.) and there is a need to discuss on the appropriate party/ies which should define acceptance criteria.

The literature presents possible acceptance criteria which have been so far proposed by model developers (Galea et al., 2012; Meyer-König et al., 2007). Nevertheless, there is a need for a large debate on whether those criteria should be only minimum (they should be met to release a model on the market) or if they should aim at defining “certified model”. This issue is also reflected in the legal responsibility associated with the model outcome (should it be on the model developers, the model users or both?).

The discussion on this matter is ongoing and given the different positions assumed by different parties, the development of an accepted standard V&V protocol for evacuation models appears evident.

ACKNOWLEDGEMENTS

The content of this paper is mostly based on the fruitful discussions which took place at the National Institute of Standards and Technology during the writing of the NIST Technical Note 1822 (Ronchi et al., 2013a). The author wishes to acknowledge Erica Kuligowski, Richard Peacock, Paul Reneke and

Daniel Nilsson who are the co-authors of the NIST Technical Note 1822, for their useful insights on the issues associated with Verification and Validation of evacuation models. The author thanks Jason Averill, Thomas Cleary, Rita Fahy, Anthony Hamins, Steve Gwynne, Blaza Toman and Craig Weinschenk for providing their valuable comments on tests and procedures presented in the Technical Note 1822. The author also would like to thank Max Kinateder for reviewing the manuscript before publication.

REFERENCES

Averill, J.D., 2011. Five Grand Challenges in Pedestrian and Evacuation Dynamics, in: Peacock, R.D., Kuligowski, E.D., Averill, J.D. (Eds.), Pedestrian and Evacuation Dynamics. Springer US, Boston, MA, pp. 1–11.

Boyce, K.E., Shields, T.J., Silcock, G.W., 1999. Toward the Characterization of Building Occupancies for Fire Safety Engineering: Prevalence, Type, and Mobility of Disabled People. Fire Technol. 35, 35–50.

Erickson, W., Lee, C., von Schrader, S., 2010. Disability statistics from the 2008 American community survey (ACS). Cornell University Rehabilitation Research and Training Center on Disability Demographics and Statistics (StatsRRTC), Ithaca, NY.

Frantzich, H., Nilsson, D., Eriksson, O., 2007. Utvärdering och validering av utrymningsprogram [Evaluation and validation of evacuation programs] (No. 3143). Department of Fire Safety Engineering and Systems Safety, Lund University, Lund, Sweden.

Galea, E.R., Deere, S., Brown, R., Nicholls, I., Hifi, Y., Bresnard, N., 2012. The SAFEGUARD validation data-set and recommendations to IMO to update MSC Circ 1238. Presented at the SAFEGUARD Passenger Evacuation Seminar, The Royal Institution of Naval Architects, London, UK, pp. 41–60.

Galea, E.R., University of Greenwich, Centre for Numerical Modelling and Process Analysis, 1997. Validation of evacuation models. CMS Press.

Gwynne, S., University of Greenwich, Centre for Numerical Modelling and Process Analysis, 1998. Validation of the buildingEXODUS evacuation model. CMS Press, London.

Gwynne, S.M.V., Kuligowski, E., Spearpoint, M., Ronchi, E., 2013. Bounding defaults in egress models. Fire Mater. doi:10.1002/fam.2212

Hamins, A., McGrattan, K., 2007. Verification and Validation of Selected Fire Models for Nuclear Power Plant Applications (No. NUREG-1824). National Institute of Standards and Technology, Gaithersburg, MD (US).

Hostikka, S., Korhonen, T., Paloposki, T., Rinne, T., Matikainen, K., Heliövaara, S., 2007. Development and validation of FDS+Evac for evacuation simulations.

International Maritime Organization, 2007. Guidelines for Evacuation Analyses for New and Existing Passenger Ships, MSC/Circ.1238.

International Standards Organization, 2008. Fire Safety Engineering – Assessment, verification and validation of calculation methods. ISO 16730.

Kecklund, L., Andrée, K., Bengston, S., Willander, S., Siré, E., 2012. How Do People with Disabilities Consider Fire Safety and Evacuation Possibilities in Historical Buildings?—A Swedish Case Study. Fire Technol. 48, 27–41. doi:10.1007/s10694-010-0199-0

Kinateder, M., Omori, H., Kuligowski, E.D., 2014. The Use of Elevators for Evacuation in Fire Emergencies in International Buildings (No. Technical Note 1825). National Institute of Standards and Technology, Gaithersburg, MD (US).

Kling, T., Ryynänen, J., Hakkarainen, T., Mikkola, E., Paajanen, A., Hostikka, S., 2012. Numerical tool for simulation of the passenger's evacuation for the train scenarios, TRANSFEU - Development of numerical simulation tools for fire performance, evacuation of people and decision tool for the train design. VTT Technical Research Center of Finland.

Kuligowski, E., 2011. Predicting Human Behavior During Fires. Fire Technol. 49, 101–120. doi:10.1007/s10694-011-0245-6

Kuligowski, E., Peacock, R., Wiess, E., Hoskins, B., 2013. Stair evacuation of older adults and people with mobility impairments. Fire Saf. J. doi:10.1016/j.firesaf.2013.09.027

Kuligowski, E.D., Peacock, R.D., Hoskins, B.L., 2010. A Review of Building Evacuation Models, 2nd Edition, NIST Technical Note 1680.

Lord, J., Meacham, B., Moore, A., Fahy, R., Proulx, G., 2005. Guide for evaluating the predictive capabilities of computer egress models NIST GCR 06-886.

Lovreglio, R., Ronchi, E., Borri, D., 2014. The validation of evacuation simulation models

through the analysis of behavioural uncertainty. Reliab. Eng. Syst. Saf. 131, 166–174. doi:10.1016/j.ress.2014.07.007
Meyer-König, T., Waldau, N., Klüpfel, H., 2007. The RiMEA Project — Development of a new Regulation, in: Waldau, N., Gattermann, P., Knoflacher, H., Schreckenberg, M. (Eds.), Pedestrian and Evacuation Dynamics 2005. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 309–313.
Moussaid, M., Helbing, D., Theraulaz, G., 2011. How simple rules determine pedestrian behavior and crowd disasters. Proc. Natl. Acad. Sci. 108, 6884–6888. doi:10.1073/pnas.1016507108
Ronchi, E., 2013. Testing the predictive capabilities of evacuation models for tunnel fire safety analysis. Saf. Sci. 59, 141–153. doi:10.1016/j.ssci.2013.05.008
Ronchi, E., Kinsey, M., 2011. Evacuation models of the future: Insights from an online survey on user's experiences and needs. Presented at the Advanced Research Workshop Evacuation and Human Behaviour in Emergency Situations EVAC11, Capote, J. et al, Santander, Spain, pp. 145–155.
Ronchi, E., Kuligowski, E.D., Reneke, P.A., Peacock, R.D., Nilsson, D., 2013a. The process of Verification and Validation of Building Fire Evacuation models. Technical Note 1822.
Ronchi, E., Nilsson, D., 2013a. Fire evacuation in high-rise buildings: a review of human behaviour and modelling research. Fire Sci. Rev. 2, 7. doi:10.1186/2193-0414-2-7
Ronchi, E., Nilsson, D., 2013b. Modelling total evacuation strategies for high-rise buildings. Build. Simul. doi:10.1007/s12273-013-0132-9
Ronchi, E., Reneke, P.A., Peacock, R.D., 2013b. A Method for the Analysis of Behavioural Uncertainty in Evacuation Modelling. Fire Technol. doi:10.1007/s10694-013-0352-7
Sime, J.D., 1985. Movement toward the Familiar: Person and Place Affiliation in a Fire Entrapment Setting. Environ. Behav. 17, 697–724. doi:10.1177/0013916585176003
Thompson, P.A., Marchant, E.W., 1995. Testing and application of the computer model "SIMULEX". Fire Saf. J. 24, 149–166. doi:10.1016/0379-7112(95)00020-T
Thunderhead Engineering, 2014. Verification and Validation - Pathfinder 2013.1.
Wagoum, A.U.K., Chraibi, M., Mehlich, J., Seyfried, A., Schadschneider, A., 2012. Efficient and validated simulation of crowds for an evacuation assistant: Simulation of crowds for an evacuation assistant. Comput. Animat. Virtual Worlds 23, 3–15. doi:10.1002/cav.1420