Independent Verification and Validation of an Industrial Simulation Model

* *Stewart Robinson*

Operational Research and Management Sciences Group, Warwick Business School, University of Warwick, Coventry CV4 7AL, UNITED KINGDOM

Tel: 44-2476-522132, Fax: 44-2476-524539, Email: stewart.robinson@warwick.ac.uk

Roger J. Brooks

Management Science Department, Lancaster University Management School, Lancaster University,

Lancaster LA1 4YX, UNITED KINGDOM

Tel: 44-1524-593866, Fax: 44-1524-844885, Email: roger.brooks@lancaster.ac.uk

* Corresponding author

Author Biographies

Stewart Robinson is a Professor of Operational Research at Warwick Business School in the UK and Associate Dean for Specialist Masters Programmes.

Roger Brooks is a lecturer at Lancaster University Management School in the UK.

ABSTRACT

The independent verification and validation (IV&V) of simulation models is largely restricted to applications in the military and public policy domains. There is little evidence of IV&V for industrial simulation models. This is largely because industrial simulations are normally of a much smaller scale and could not warrant a full IV&V. A procedure for IV&V of industrial simulation models is described that provides a viable alternative where the cost and time of IV&V must be contained. The procedure consists of the following activities: structured walkthrough, review of model assumptions, code examination, review of verification procedures, replications analysis, review of static analysis, review of output reporting, and investigation of results and experimentation. The IV&V of a Sellafield Limited supply chain model is described.

Key Words

Discrete-event simulation, industrial simulation models, verification and validation, independent verification and validation

Independent Verification and Validation of an Industrial Simulation Model

1. INTRODUCTION

Discussions on independent verification and validation (IV&V) centre on the assessment of the large scale models that are typically found in the military and public policy domains. There is, however, a significant use of simulation in industry. In this context there does not appear to be any reference to the idea of IV&V. A key issue is that models in the industrial context are generally smaller than their counterparts for the military and public policy domains. But this does not necessarily mean that the decisions being taken with these models are of an insufficient scale to warrant an independent review of the confidence that should be placed in the results. Many industrial simulation models involve decisions that run into the millions or even billions of dollars.

While IV&V may be necessary, or at least beneficial, for some industrial simulations, standard IV&V assessments are generally too extensive for these models. It is not inconceivable that the cost of IV&V could be greater than the cost of model development. This makes little sense, as it is surely better to use more resources on improving the quality of the model (development) than on assessing its quality (IV&V).

This leads to the question: how can the concept of IV&V be adapted for the industrial context? The purpose of this paper is to answer this question by exploring the idea of IV&V for industrial simulation models. An approach to IV&V is described and applied to a model developed for Sellafield Limited. First there is a review of approaches for IV&V, which centres on the military and public policy domains. The shortcomings of these approaches in relation to industrial simulation models are discussed. The Sellafield Limited model is then described, after which the objectives and methodology for IV&V of the model is explained. A summary of the conclusions and recommendations from the Sellafield Limited IV&V review is provided. The paper concludes with a discussion on IV&V in the industrial context.

2. APPROACHES TO INDEPENDENT VERIFICATION AND VALIDATION (IV&V)

Verification is the process of ensuring that the model design has been transformed into a computer model with sufficient accuracy [1]. Validation is the process of ensuring that the model is sufficiently accurate for the purpose at hand [2]. Although model validity is theoretically a binary decision, a model is either sufficiently accurate or it is not, proving this is another matter [3]. Taking the stand point of falsification in the philosophy of science, it is not possible to prove that a model is valid, only that it is invalid. As a result, the purpose of verification and validation is to increase confidence in a model, by not being able to prove it is invalid, to the point that it will be used for decision-making. The term 'confidence' does not refer to the statistical confidence in the model, but to the overall belief in the model and its results [1]. This confidence is an attribute of the decision-maker [4, 5].

Verification and validation are the responsibility of the model developers. Relying on their judgements about the confidence that can be placed in a model is not always sufficient, particularly if the decisions that are to be based on the model are of a significant scale or critical to an organisation. As a result, it is sometimes deemed necessary to obtain an independent view of a model through IV&V.

The main aim of IV&V is to determine whether a model is suitable for a particular use. Gass [6] defines model assessment (or evaluation) as 'a process by which interested parties (who were not involved in a model's origins, development and implementation) can determine, with some level of confidence, whether or not the model's results can be used in decision-making'. He believes that model assessment is necessary in three circumstances: firstly, when the decision-makers don't have the necessary knowledge and skills to evaluate the model; secondly, when the decision-makers are far removed from the process of developing the model; thirdly, when the model is to be applied to a set of circumstances differing from the original intended use.

There is some confusion over the use of terminology in IV&V. The US Department of Defense uses the term 'accreditation', defining it as 'the official certification that a model, simulation, or federation of models and simulation is acceptable for use for a specific purpose'[7]. Balci [8] points out that this use of terminology is at odds with the International Standards Organization, and suggests the use of the term 'certification' is more appropriate. Tullos-Banks et al [9] identify a clear distinction between IV&V, certification and verification, validation and accreditation (VV&A). By verification and validation they refer to tests and checks during each of the simulation stages, as generally recommended in the literature (e.g. [10]). Accreditation is then the process of reviewing the tests and making a decision as to whether the overall model is acceptable. Overall control for VV&A is by senior personnel in the organisation using the model, who Tullos-Banks et al recommend are separate from those involved in the modelling and V&V, giving some independence to the review. Tullos-Banks et al distinguish VV&A from certification, which they consider to be a more formal process, and from IV&V, which they interpret as meaning verification and validation carried out by a separate independent organisation. In this paper the central concept is the use of independent experts to evaluate a model's suitability for purpose. We shall not be unduly concerned by a discussion on terminology, but instead we shall use IV&V as a generic term to cover the field of independent model assessment.

A variety of assessment procedures have been proposed and applied over the history of IV&V. Most of this work concentrates on the assessment of models for military and public policy applications, although there are some assessment procedures that have application for a wider class of models. The characteristics of these assessment procedures are now discussed. Given the differences in the approaches that are described below, it should be noted that some recent work has been attempting to provide a common framework, or standards, for model assessment [11, 12, 13].

2.1 Who Performs the Assessment?

In general a model assessment is performed by an independent third party, hence the use of the term 'independent verification and validation' [14]. Balci [5] recommends that the group of assessors should be made up of people with a knowledge of the real system, modellers, simulation analysts and people with extensive simulation experience. Later he suggests that any organisation that performs a significant amount of simulation work should have a 'simulation quality assurance' group responsible for the total quality management of the simulation work [15]. Such a group may not, of course, be truly independent.

Robinson [16] suggests an alternative approach, in which the clients perform an assessment of the modelling study. This procedure assumes a high level of client involvement and the assessment concentrates on the clients' perceptions of the simulation study. This is quite different to the ideas of independent and expert evaluation as discussed in IV&V, and does not preclude the use of expert evaluation as well.

Sargent [10] discusses the timing of the assessment by a third party. The work can either be performed concurrently with the model building or after the event. Sargent states a clear preference for concurrent assessment, arguing that it is more effective.

2.2 What Criteria are Assessed?

Most of the procedures outline a set of criteria ('factors' or 'indicators') that need to be considered when evaluating a model. The basis of the majority of these criteria is model verification and validation, although other factors such as documentation and training are also considered to be important. Gass and Joel [4] use seven criteria: model definition, model structure, model data, computer model verification, model validation, model usability and model pedigree. Ören [17, 18] also proposes a number of components and criteria for model assessment. More recently, Liu et al [19] suggest five criteria: validity, correctness, reliability, usability and interoperability.

Balci [5] identifies a series of indicators (indirect measures) that can be used in assessing the acceptability of a simulation's results. For every process in a simulation study he identifies the corresponding 'credibility assessment stage'. For each credibility assessment stage there are a set of indicators that can be used for assessment. Beyond the main credibility assessment stages there are also a series of 'other indicators'. Balci suggests that these should be derived from the software quality characteristics, such as accessibility, accuracy and efficiency, devised by Boehm et al [20].

In a later paper, Balci [8] suggests that seven top level indicators should be used for model certification:

- Formulated problem credibility
- Requirements credibility
- Application credibility
- Experimentations credibility
- Project management quality
- Cost
- Risk

These indicators explicitly identify the need to assess not only the product, but also the process (model life-cycle) and project (management issues) in the modelling effort [21]. Balci [22] also identifies the importance of assessing the quality of the model developers.

2.3 How are the Criteria Assessed?

Some of the procedures require subjective scores to be given to each of the criteria. Sargent [23] recommends that the assessor should provide a subjective score between 0 and 10 for each criterion. Balci [5] proposes that a panel of experts should score and weight his set of indicators. He identifies various methods for assessing the indicators, expanding on this in a later discussion [15]. Most of these entail subjective judgements based on evidence, although more objective scores are possible for some indicators. In describing his certification methodology, Balci [8] identifies three types of scoring that might be used: crisp (a single value between 0 and 100; fuzzy (an interval between 0 and 100); nominal (a named score which is assigned a predetermined value).

Others prefer the use of qualitative statements. Sargent [10], for instance, argues against providing quantitative scores because he believes they can be misleading. Fossett et al [24] do not attempt to score their factors, but argue that by collecting and reviewing information on each factor, areas of strength and weakness in relation to the credibility of a simulation can be identified.

Gass and Joel [4] recommend that the decision-maker's requirements for each criterion are taken into account when assessing a model. The independent assessor provides a score for each criterion between 1 (not satisfying) and 5 (fully satisfying). Meanwhile, the decision-makers agree a threshold value for each criterion. If the score for any criterion falls below the threshold value then the confidence in that area must be called into question.

2.4 How is an Overall Score Calculated?

Where subjective scores are assigned various methods are used for weighting and combining the scores to provide an overall score. Balci [5] proposes that an overall score is calculated from the scores and weights assigned by the panel of experts, a higher score indicating that greater confidence can be placed in the results. Balci does point out that a high score does not necessarily guarantee the acceptability of the model and its results, since this is an attribute of the decision-maker, not the model. In a later paper, Balci [22] suggests that it would be beneficial to provide a percentage level of confidence that a model satisfies the acceptability criteria. However, he goes on to recognise that it is not generally possible to derive such numerical measures of confidence and that a qualitative assessment is a more realistic outcome.

Gass [25] proposes that importance weights are assigned to each of the criteria and that, based on a weighted average of the scores, the model is classified as operationally acceptable (can be used with confidence), acceptable (with some deficiencies), or not acceptable (needs major changes). In a later paper Gass [26] employs the Analytic Hierarchy Process [27] to determine weights for each criterion that are based on their importance to the decision-maker. When these weights are combined with a level of satisfaction score for each criterion (intensity), determined by an independent assessor, an overall numerical rating (between 0 and 1) is calculated. Balci [8] also employs the Analytic Hierarchy Process for weighting indicators to calculate a final score. Pohl and Brade [28] suggest the issue is addressed with fuzzy multi-attribute decision theory.

2.5 Software Support for IV&V

Software tools are available that support IV&V. The US Defense Modeling and Simulation Office [29] list and provide some examples of tools that support both verification and validation. Balci et al [30] describes a web-based software system, the Evaluation Environment, which provides an implementation of Balci's certification methodology. Ke et al [31] describe a synthetic environment (HITVICE) for facilitating model assessment.

2.6 Critique of Assessment Procedures in Relation to Industrial Simulation Models

As noted above, the majority of procedures for IV&V have been developed for use with military and public policy simulations. Simulation in industry is quite different. Robinson [32] discusses the facets of simulation studies under different modes of practice. Two specific modes relate to the discussion here: simulation as 'software engineering' (mode 1) and simulation as 'a process of organisational change' (mode 2).

Mode 1 centers on the provision of a product, in this case a simulation model. This involves large models, whose prime motivation is representation of a real system with a view to providing decision support or training. These models are often used over many years. The development of the model involves multiple modellers, developing code in a programming language, whose predominant skill is software development. It may take years (certainly many person years) to develop the model.

Under mode 2 the work centers on the provision of a service, with the prime motivation being to intervene in a problem situation. This involves small scale models that are used for a short period and then normally thrown away. The model is normally developed by a lone modeller, typically using a simulation package, who is skilled in modelling. Model development requires only a matter of weeks or months.

These modes of practice lie on a continuum and so a simulation study may not exactly match a specific mode. Robinson [32] does identify, however, that mode 1 is the most prevalent in military simulation modelling, while mode 2 is more prevalent in industrial simulation modelling.

Two key differences in these modes are of importance to the context of IV&V. First, military and public policy simulations are generally much larger in scale than industrial simulations. As a result, the procedures for IV&V reflect the scale of models being developed and used. Because of the cost of model evaluation, Gass [33] suggests that in deciding whether to evaluate a model, it should have involved more than five years of person effort to develop. Albeit that this advice is 30 years old, there is little to suggest that models are getting smaller and that less effort and expense is being put into their development. Indeed, Youngblood et al [34] point to the increasing size of military simulation models, with specific reference to the growing use of distributed simulation that presents its own challenges for IV&V [35]. Industrial simulation models are generally much smaller and require much less time and cost to develop. In a survey of simulation in industrial settings, Cochran et al [36] found that typical simulation projects take between one and three months (although a reasonable proportion, 29%, do take more than 6 months). The IV&V procedures discussed for military and public policy applications would simply not be cost effective for the majority of industrial models. They might even be more expensive to implement than the original cost of model development.

A second difference lies in the assumption that the decision-maker is not involved in the 'model's origins, development and implementation' [6]. Diener et al [37] state that it is almost impossible for all the users of military models to be involved in their development, due to the length of the development cycle. This is not the case for industrial simulation models, where there is an expectation that the clients will be highly involved in the model development process [38, 39]. Because the client is involved in the development of the model, some confidence in the model has already been derived (if the model is deserving of such confidence). The role of IV&V is, therefore, reduced to one of needing to provide additional confidence in the model and provide confidence for other stakeholders that were not so involved in the model's origins, development and implementation. Indeed, it may often be the case that the decision-makers have gained sufficient confidence from their involvement in the simulation study so as not to warrant an IV&V.

It is probably as a result of these two differences that IV&V is not common in industrial settings. Indeed, the authors know of no published examples of IV&V of an industrial simulation model.

3. AN INDUSTRIAL SUPPLY CHAIN MODEL: THE SITE REMEDIATION INTEGRATED (SRI) SIMULATION MODEL

Sellafield Limited is responsible for the safe delivery of contracts at Sellafield and Capenhurst on behalf of the Nuclear Decommissioning Authority. The Sellafield site in Cumbria (north west

England) is one of the world's largest nuclear engineering centres whose key focus is now on decommissioning historic liabilities held in waste repositories. A massive programme of waste retrieval, packaging and storage ('site remediation') is underway with a view to making safe the nuclear liabilities. This entails a supply chain from current waste storage through a series of solid and effluent treatment plants, to final storage. For more information on the process of decommissioning nuclear sites, see Wall and Shaw [40].

Simulation and modelling have been widely used by Sellafield Limited to support the design and planning for site remediation. This includes a hierarchy of models, from high level spreadsheet and database models through to detailed plant simulations. The model that was the subject for IV&V sits in the middle of this hierarchy, and is a discrete event simulation of the complete supply chain. This involves some level of detail for individual plants such as key equipment and resources, process control and equipment failures, as well as transportation between plants. This model is known as the Site Remediation Integrated (SRI) Simulation Model and covers waste repositories, retrieval, treatment plants, transportation and storage. A total of ten individual and separately managed plants are included in the SRI model.

The objectives of the model are as follows:

- To evaluate and compare alternative strategies for site remediation.
- To estimate the likely end date for site remediation.
- To provide a better understanding of the interactions and important factors in the retrievals operation.

As such, the overall aim of the model is to support strategic decision-making concerning the remediation process.

The SRI model was developed by a specialist simulation consulting company, which for reasons of confidentiality is referred to here as ‘Sim Consulting’. Development of the original model started in 2000 and represented about 1 year of person time. This model was subject to IV&V in 2001 and then subject to another review in 2003 following further developments. As such, the model is relatively large scale for an industrial context, albeit small in comparison to many military and public policy simulations.

The SRI model was developed using the Witness simulation software [41]. This provides a visual display of the running model which was helpful for understanding the model and for verification and validation (V&V). All data were held in Excel spreadsheets and results were output to separate spreadsheets for further analysis and reporting. Sim Consulting also developed an Excel based ‘scenario manager’ for setting-up and documenting all experimental scenarios.

In terms of the modes of practice described above, on a continuum from mode 1 to mode 2, the facets of the SRI model place it closer to mode 2. That is, the prime motivation for the model was intervention in a specific problem situation and the model was developed by a lone modeller in a standard simulation package. The scale of the model, however, was somewhat larger than the majority of models developed in mode 2.

4. THE OBJECTIVES OF THE IV&V REVIEW

The IV&V team consisted of the two authors of this paper with input and advice from Professor Michael Pidd of Lancaster University Management School. The agreed purpose of the IV&V was to provide an assessment of the confidence that should be placed in the SRI model. The specific objectives were to:

- Assess the methodology used by Sim Consulting and Sellafield Limited in developing, validating and using the model.
- Assess the quality of the model build and function.
- Identify improvements to the methodology for future model development and use.

Sellafield Limited and Sim Consulting worked together closely in agreeing the model specification (conceptual model) which is documented using process flow diagrams. Consequently Sellafield Limited had a high degree of confidence that the conceptual model represented the key elements of the real system in a suitable way. Agreement was reached that the work would not examine directly the validity of the conceptual model as a representation of the real system, although there was some consideration of the appropriateness of the level of detail for meeting the modelling objectives. Instead the review focused on the fidelity with which the computer model reflected the model design as set out in the conceptual model and on the way that the model had been validated and used.

This paper describes the 2003 review of the model. The process for the 2001 review was very similar. The total time required by the IV&V team for the review was in the region of 40 person days (shared between the team), spread over two to three months. This entailed around 10 face-to-face meetings. There was also a commensurate level of effort required from both Sim Consulting and Sellafield Limited.

The principal client of the IV&V review was the senior manager of the team within Sellafield Limited that commissioned and used the model. This manager had significant involvement in the development of the SRI model. The IV&V report was also aimed at other senior management at Sellafield Limited, not involved in the model development process, but who were responsible for strategic decisions that would be informed by the model results.

5. THE METHODOLOGY FOR IV&V OF THE SRI MODEL

The methodology for IV&V of the SRI model is now described. Due to the time and funding available, a full IV&V as typically described in the military and public policy context was not feasible. Therefore, the proposed methodology aims to provide a relatively rapid and low cost review of a model. It also takes into account the involvement of at least some of the decision-makers in the development and use of the model.

5.1 The Concept of IV&V for an Industrial Simulation Model

Both verification and validation of industrial simulation models is, or at least should be, carried out during the process of model development and use [3]. This expectation that V&V had been performed on the SRI model at the time of the IV&V, significantly affected the approach to the review. Instead of attempting to perform an after the event V&V on the model, the IV&V team focused their efforts on ensuring that Sim Consulting and Sellafield Limited had performed appropriate V&V checks during model development and use. By reviewing model documentation, investigating the V&V that had been carried out and scrutinising reports of the results, conclusions were drawn regarding the confidence that should be placed in the SRI model and the results. Staff from both Sim Consulting and Sellafield Limited were involved in the IV&V, enabling the client to develop confidence (or otherwise) in the model not only from the IV&V report, but also from being involved in the process of IV&V.

5.2 The Process of V&V

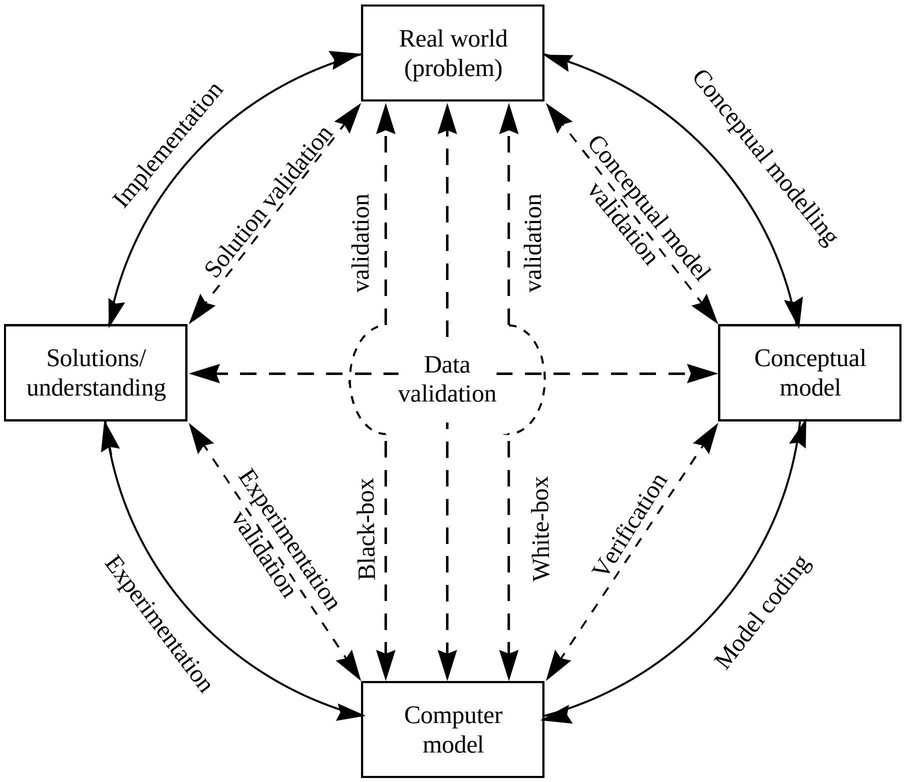
The process of the IV&V review centred on the key V&V activities expected during model development and use. Figure 1 outlines Robinson's [3] life-cycle for model development and use. This is based on the work of Landry et al [42], which is in turn built upon Sargent [43]. It shows four key activities of conceptual modelling, model coding, experimentation and implementation. Along side each of these, and performed in parallel, are four V&V activities. Three other forms of validation are also denoted: white-box validation, black-box validation and data validation.

These V&V activities are defined by Robinson [39] as follows:

- *Conceptual Model Validation*: determining that the content, assumptions and simplifications of the proposed model are sufficiently accurate for the purpose at hand.

- *Data Validation*: determining that the contextual data and the data required for model realisation and validation are sufficiently accurate for the purpose at hand.
- *Verification*: the process of ensuring that the model design (conceptual model) has been transformed into a computer model with sufficient accuracy [1].
- *White-Box Validation*: determining that the constituent parts of the computer model represent the corresponding real world elements with sufficient accuracy for the purpose at hand.
- *Black-Box Validation*: determining that the overall model represents the real world with sufficient accuracy for the purpose at hand.
- *Experimentation Validation*: determining that the experimental procedures adopted are providing results that are sufficiently accurate for the purpose at hand.
- *Solution Validation*: determining that the results obtained from the model of the proposed solution are sufficiently accurate for the purpose at hand.

Figure 1 Life-cycle for model development and use [3]



Although solution validation is similar to black-box validation in concept, it is different in that it compares the results of the proposed solution in the model to the results of the implemented solution in the real world. As a consequence, solution validation can only take place post-implementation of the study's findings.

5.3 The Process of IV&V for an Industrial Simulation Model

The IV&V review was supported by problem domain documentation and model documentation provided by Sellafield Limited and Sim Consulting respectively. A series of face-to-face meetings were held with the two stakeholders. In these meetings the documentation was reviewed in detail and compared to the model, input data and results. The model and associated Excel spreadsheets were available during meetings for inspection and use.

A central requirement of the IV&V activity was the documentation provided by the stakeholders. This documentation included (name in brackets denotes the originator of the documents):

- *Model Requirements*: description of model purpose and requirements. (Sellafield Limited)
- '*Concept Design*': description of the conceptual model, primarily through a series of process flow diagrams. (Sim Consulting)
- *Assumptions and Simplifications*: details of assumptions about the real system and modelling simplifications. (Sim Consulting)
- *Verification*: details of tests carried out, including model changes and further tests performed. (Sim Consulting)
- *Key Changes*: changes to the model since the previous 2001 IV&V review. (Sim Consulting)
- *Scenario Manager*: summary of key experimental scenarios run with the SRI model. (Sim Consulting)
- *Data Tables*: lists of all input data, including documentation on data sources. (Sim Consulting)
- *Static Analysis*: report of comparative analysis of model results to analytical calculations. (Sim Consulting)
- *Replications Analysis*: report of analysis of number of independent replications required with the model. (Sim Consulting)
- *Summary of Results*: results from key experimental scenarios. (Sim Consulting)

- *Sim Consulting Project Life-Cycle*: methodology for model development and use employed by Sim Consulting. (Sim Consulting)
- *Options for Site Remediation*: report of possible options for site remediation supply chain design. (Sellafield Limited)
- *Analysis of Options for Site Remediation*: report of results from model experimentation. (Sellafield Limited)
- *Recommendations for Site Remediation*: management report on recommendations following use of the SRI model. (Sellafield Limited)
- *Meeting notes*: notes of meetings between Sellafield Limited and Sim Consulting during the development of the SRI model. (Sellafield Limited)

All these documents provided useful input to the review, but of greatest utility were the model requirements, concept design, assumptions and simplifications, data tables and static analysis. The first four provided a detailed explanation of the model content, while the static analysis was an important verification and black-box validation test of the model. During the IV&V review some errors and inconsistencies were found in the documents, although none were significant. These were revised and corrected as necessary.

The IV&V activities performed in the review consisted of the following: structured walkthrough, review of model assumptions, code examination, review of verification procedures, replications analysis, review of static analysis, output reporting, results and experimentation, and review of previous recommendations. Each is now briefly discussed.

Structured Walkthrough

A detailed walkthrough with the model developer of the conceptual model design and data tables, and how these relate to the computer model. This entailed walking through each part of the conceptual model and matching it with the corresponding fields in the data tables and components in the computer model. A detailed understanding of both the conceptual model and how it had been implemented in the software was gained through this activity. As a result of the walkthrough omissions from the documented conceptual model (concept design) and minor errors in the data tables were identified. Suggestions were also made to improve the presentation of the data tables and to provide additional outputs. The detailed discussions revealed a number of assumptions and simplifications to the model that had not been documented; the documentation was revised accordingly. The structured walkthrough was carried out at Sim Consulting's offices and took a total of four days. A member of staff from Sellafield Limited, who was a user of the model, observed the walkthrough in order to gain a better understanding of the model.

Review Model Assumptions

Following the structured walkthrough Sim Consulting and Sellafield Limited were asked to compile a complete list of assumptions (including the additional ones identified in the structured walkthrough) and classify them by confidence level (high, medium, low) and likely level of impact on the results (high, medium, low). The list of assumptions was then reviewed with the IV&V team, focusing on the most important assumptions identified, particularly the high impact, low confidence assumptions. The IV&V team needed to understand how these assumptions might impact on the results of, and therefore the confidence in, the model. It was recommended that sensitivity analysis should be carried out on the most critical assumptions. The assessment of the assumptions was carried out in a meeting between Sim Consulting and Sellafield Limited, without input from the IV&V team, who then reviewed the outcome of the assessment process with the two other parties.

Code Examination:

A review of the simulation model code. This did not entail a detailed line-by-line code check since this was not deemed necessary to meet the IV&V review objectives and would not have been possible within the time-scales available. Instead, one member of the IV&V team spent a day at Sim Consulting's offices looking through the code, understanding its structure and sample testing parts of the code. This was possible because the model had been developed in a simulation package with which the reviewer was very familiar. The model developer was available during this day to answer questions on points that needed clarification. The code check identified one minor error, some elements of redundant code, a lack of commenting in sections of the code and some revisions to the concept design. Since only one minor code error was found in the day, it was concluded that there was no need for further code checking.

Review of Verification Procedures

A review of the documentation provided by Sim Consulting for their verification of changes to the model code. This entailed the IV&V team reading through the documentation and then discussing it with Sim Consulting in order to understand the range of verification activities performed and the approach to rectifying coding errors that were identified during verification.

Replications Analysis

Reviewing the approach taken by Sim Consulting to ensuring that the experimentation had properly taken account of the stochastic nature of the model. For the SRI model this specifically involved Sim Consulting performing multiple replications with their own random number generator and the one provided with the Witness simulation software. The results were compared to satisfy the IV&V team that the Sim Consulting generator (which had been previously tested) was working satisfactorily.

Review of Static Analysis

Static analysis provides a means of comparing the model results to analytical, but approximate, results. An analysis performed by Sim Consulting was reviewed by a member of the IV&V team. The analysis entailed a comparison of the results from individual plants in the model, running unconstrained, with an analysis of the throughput of the bottleneck operation in each plant. This showed a small variance of less than 2% between the model and static analysis for most plants, with the maximum difference being around 6%. These differences were seen as acceptable given the stochastic nature of the simulation model, which cannot be fully accounted for in the static analysis.

Output Reporting

The nature of the output information provided by the model was reviewed. Sim Consulting demonstrated the output reports to the IV&V team and the code for generating those reports was discussed. The review identified some requirements to improve the outputs, especially with respect to providing additional time-series for some key outputs.

Results and Experimentation

The analysis of results and procedures followed for experimentation were reviewed, for instance, the management of scenarios. This review was carried out with Sim Consulting and the user of the model from Sellafield Limited. A particular focus of this review was whether it was possible to explain why different scenarios led to different outcomes. This is valuable for providing some confidence in the results and for more fully understanding the behaviour of the system.

Review Previous Recommendations

A review of the actions taken on the recommendations from the previous IV&V review in 2001. In a discussion between the IV&V team and Sim Consulting it was found that the majority of recommendations had been implemented in developing the latest version of the SRI model.

Time constraints had meant, for instance, that a sensitivity analysis on breakdown data had not been carried out at the time of the review.

Where required, Sim Consulting were asked to make model changes and re-run the model to determine the effect. The results of these re-runs were reported back to the IV&V team. In particularly, the effect of model errors that were identified during the review was addressed using this approach. The errors identified were found to have little effect on the model results.

Table 1 shows how each of these activities relates to the V&V activities shown in figure 1. All the V&V activities are covered by the IV&V review, with the exception of solution validation. This was not possible, since it cannot be assessed until the site remediation supply chain is in operation. Indeed, the accuracy of the estimate of end date for site remediation cannot be assessed until the operation of the supply chain is complete.

Table 1 IV&V Procedures and Forms of V&V

IV&V activity	**V&V activity**
Structured walkthrough	Conceptual model validation Data validation Verification White-box validation
Review model assumptions	Conceptual model validation
Code examination	Verification
Review verification procedures	Verification
Replications analysis	Experimentation validation Verification
Review of static analysis	Verification Black-box validation
Output reporting	Experimentation validation
Results and experimentation	Experimentation validation
Previous recommendations	Various – different for each recommendation

It should be noted that the specific IV&V activities can be adapted to the requirements of the review. In the 2001 IV&V, four other activities were included in the review. Because Sim Consulting were using their own random number generator, the random number streams being used in the model were tested. The IV&V team also tested the consequent random sampling functions that had been developed by Sim Consulting. A third activity involved a review of reports on some sensitivity analysis that had been carried out with the SRI model. Finally, a review of intended future developments to the model was carried out.

None of these were applicable for the 2003 IV&V. The random number streams and random sampling functions were the same as in 2001. No specific sensitivity analysis had been carried out with the model at the time of the review. The model was near completion and so no significant future developments were planned following the 2003 review.

5.4 Reporting the IV&V Findings

An IV&V report summarised the findings of the IV&V team. The sections of the report covered the aims and objectives of the review, the approach taken, findings, conclusions, recommendations and further work. The section on findings summarised the findings from each activity in the IV&V review. Appendices provided detailed reports on each IV&V activity.

The conclusions section drew together the findings and discussed them in relation to the V&V activities identified in figure 1. An overall conclusion was given in relation to the objectives of the SRI model and the extent to which confidence could be placed in the model for meeting these objectives.

Following the approach of Fossett et al [24], no effort was made to provide a score for the credibility of the model. The review informed the clients about the strengths and weaknesses of the model by providing qualitative statements, in some cases backed by quantitative evidence, for or against the validity of the model. Assigning a numerical value would not have assisted this process, particularly in the absence of comparative scores for other models that might provide a context.

Recommendations were split into recommendations for the SRI model and recommendations for future modelling practice by Sim Consulting when working with Sellafield Limited. Recommendations were identified on the basis of the detailed findings from the IV&V review.

The report was delivered and then discussed at a meeting with Sim Consulting and Sellafield Limited. Both organisations were given opportunity to provide a written response to the report, particularly identifying any actions that were to be taken to implement the recommendations. A final meeting was held with the three parties to discuss the response.

6. CONCLUSIONS AND RECOMMENDATIONS OF THE IV&V REVIEW OF THE SRI MODEL

In order to provide some sense of the outcome of the IV&V process, a brief summary of the conclusions and recommendations of the SRI model review is provided here.

6.1 Conclusions of the IV&V Review

The overall conclusion of the IV&V team was as follows:

> In our opinion, based on the documentation and information provided and the work carried out, Sim Consulting and Sellafield Limited have followed due process in the development of the SRI model, and the model build and function are of good quality. Some aspects of good modelling practice, particularly documentation, were not kept up-to-date due to the rapid pace of model development. However, the documentation has been completed for the IV&V review. Although some errors have been found in the documentation and the model, these are relatively minor with respect to the scale of the model, and none of the errors have affected the model output to a significant degree. Sim Consulting have corrected all errors identified during the IV&V review.

Following this general conclusion, there were specific points concerning each of the three modelling objectives. These supported the use of simulation as the best approach to evaluate and analyse strategies for site remediation at Sellafield due to the complexities of the system, and concluded that a high level of confidence could be placed in the implementation of the conceptual model and that appropriate experimental practice had been adopted. The dependence of the model results on input data and model assumptions was highlighted with inevitable uncertainties in data on future systems' operations. Sensitivity analysis was therefore recommended in order to understand the range of possible outcomes.

The use of the model as a means for promoting team work between different managers within the site remediation supply chain was also identified as a very important strength. Prior to the existence of the model, plant managers communicated little with one another, causing a lack of coordination across the supply chain. The model helped the managers to understand the interactions between the various parts of the supply chain.

6.2 Recommendations of the IV&V Review

A number of recommendations concerning the SRI model were made as a result of the IV&V review. For instance, some additional output reports were suggested. It was recommended that redundant code, which was a legacy from previous versions of the model, should be removed. Further use of sensitivity analysis was suggested. It was also recommended that Sellafield Limited consider developing a higher level, first cut model that could be adapted more quickly and run much faster. This would enable rapid, but less accurate, experimentation with a wide range of scenarios, in order to identify preferred candidates. These could then be looked at in detail with the SRI model. Subsequently, Sellafield Limited investigated the use of both system dynamics and spreadsheet tools, in the end opting to develop a spreadsheet model.

Further to this, recommendations for future modelling practice were also given. These included recommendations to extend the documentation of the models developed by Sim Consulting, to improve procedures for agreeing model changes and to perform cross validation with the more detailed plant models.

7. DISCUSSION: IV&V IN THE INDUSTRIAL CONTEXT

The description above outlines the review process carried out for Sellafield Limited. Our attention now turns to a discussion of the approach taken, with a view to highlighting the key facets of the IV&V approach.

The underlying philosophy of the IV&V review was that an appropriate approach to model development, V&V and experimentation will lead to a model and results of sufficient quality. In taking this view, the need for after the event V&V and detailed code checking was largely negated. This was important since involvement in such activities would have extended the IV&V beyond feasible limits in terms of time and cost.

The assessment of the SRI model was performed by people with extensive experience of simulation (the IV&V team), people knowledgeable about the real system (Sellafield Limited) and modellers and simulation analysts (Sim Consulting). As such, the review followed the advice of Balci [5] in terms of make-up of the group for model assessment.

Unlike the case with many military and public policy models, Sellafield Limited had been highly involved in the development of the model. This shifted the focus of the IV&V from one of convincing the clients of the quality and worth of a model that they knew little about, to one of enhancing (or otherwise) their confidence in a model about which they already had a reasonable level of knowledge. What they lacked was a detailed knowledge of simulation making it impossible for them to draw conclusions about the quality of the model with any confidence. Involvement in the IV&V process enabled them to develop a greater knowledge of simulation and the SRI model; another important facet of the IV&V process.

Such involvement in model development and the IV&V review was not the case for all decision-makers at Sellafield Limited. Some had little or no involvement. For these decision-makers, the

IV&V report along with discussions with those staff involved in the model development could be used as a basis for their decisions about what confidence to place in the model.

The activities carried out during the review centred on determining the extent to which the different aspects of V&V had been carried out. Six forms of V&V were considered; that is, all forms shown in figure 1 minus solution validation. Critical to this process was the existence of documentation about the context of the problem, the model and its use. Sim Consulting were very thorough in this respect and also provided additional documentation where required. This could not be said of most simulation studies in an industrial context. There are no agreed standards for simulation model documentation and simulation software providers remain relatively quiet on the subject. There is, of course, a more limited need for documentation in most cases. Industrial simulation models are relatively small, they are normally the product of a lone modeller and they are often thrown away soon after experimentation is complete.

Since performing the SRI model review, the IV&V team have suggested an enhancement to the approach for documenting the verification tests performed on a model. This consists of a report with the following headings: identification number, category (area of the model), test, how done, pass/fail/accepted (where there is a good reason for a difference), detail, whether fixed (if test failed), how fixed, model version tested.

The findings of the review were given in a report. There was no attempt to provide a quantitative score for the confidence that should be placed in the model as this would have relied largely on assigning scores to a set of qualitative assessments. Summarising such assessments in a single quantitative score would be misleading and could be open to misinterpretation. Instead, the reader of the report was left to form an opinion about the confidence he/she would place in the model based on the evidence presented. As such, the report helped the reader to form a judgement about the 'acceptability' of the model; acceptability being an attribute of the decision-maker [5].

No consideration has been given to the idea of software support for this IV&V review process. If, however, the process were repeated many times, it should be possible to identify specific activities that could be aided with the use of relevant software. Sim Consulting are now using off-the-shelf software for aiding the documentation of their models.

The SRI model is larger than the typical industrial simulation models identified by Cochran et al [36]. It represented something more than one year of development, and indeed nearly two years by the time of the 2003 review. The size and cost of the model, however, was not the key driver for requiring an independent review. Instead, the criticality and cost of the decisions being made with the SRI model drove the need for IV&V.

The approach used for the SRI model of building on the V&V already carried out during the modelling process provides a way in which IV&V can be used effectively for other industrial or commercial models (e.g. large automotive manufacturers or airports). V&V itself consists of testing the model and critically evaluating the modelling processes with the aim of building up sufficient confidence to use the model in the decision making process (and rejecting or amending the model if there are critical test failures). IV&V can be seen simply as an additional test which can increase the confidence in the model further. As such it can be tailored for the particular circumstances such as the scale of the model, the existing confidence in the model and the importance of the decisions being taken. For example, for a small model with significant client involvement a one-day high-level review in which the modeller presents the modelling process that was followed may be a good way of providing some additional assurance without incurring much extra cost. This approach is, of course, dependent on the modeller performing appropriate V&V during model development and use. That said, if appropriate V&V activities had not been carried out by the modeller, then this would lead to an unfavourable result from the IV&V and the subsequent rejection of the model.

The IV&V process also provides significant additional benefits in helping to improve the general simulation procedures being followed within an organisation, in generating ideas for model use and development, and in adding to the knowledge of the client about simulation. However, it is necessary for the clients to balance the cost of model development and use with the cost of IV&V. It would make little sense to skew the investment too heavily in favour of the latter, while sacrificing the quality of the former.

8. CONCLUSION

The independent verification and validation of an industrial simulation model is described. The approach is to determine whether due process has been followed in the development and use of a model, without the need for an after the event V&V of the model. In doing so, the scale and cost of IV&V has been significantly reduced. This is necessary to balance the cost of IV&V with model development in the industrial context.

The approach leads to a series of qualitative statements about the model and the process of model development. The aim is to provide the decision-makers with evidence to support their confidence in the model, beyond that obtained from being involved with the development of the model. The IV&V approach also provides suggestions on how to improve model development and use in the future.

The approach described here has proved very beneficial to Sellafield Limited. It has helped inform their judgements about what confidence to place in the model. It has also helped to improve the modelling practice of Sellafield Limited and Sim Consulting. Similar IV&V reviews have since been carried out on other models developed for Sellafield Limited.

9. ACKNOWLEDGEMENT

The authors acknowledge the support and involvement of Sellafield Limited in performing the IV&V review work described in this paper.

10. REFERENCES

1. Davis, P.K. 1992. Generalizing Concepts of Verification, Validation and Accreditation (VV&A) for Military Simulation. R-4249-ACQ, October 1992, RAND, Santa Monica, CA.
2. Carson, J.S. 1986. Convincing users of model's validity is challenging aspect of modeler's job. *Industrial Engineering,* 18 (6): 74-85.

3. Robinson, S. 1999. Simulation Verification, Validation and Confidence: A Tutorial. *Transactions of the Society for Computer Simulation International*, 16 (2): 63-69.

4. Gass, S.I. and Joel, L.S. 1981. Concepts of model confidence. *Computers and Operations Research*, 8 (4): 341-346.
5. Balci, O. 1985. Guidelines for successful simulation studies. Technical Report TR-85-2, Department of Computer Science, Virginia Tech, Blacksburg, VA.
6. Gass, S.I. 1983. Decision-aiding models: validation, assessment, and related issues for policy analysis. *Operations Research*, 31 (4): 603-631.
7. DoD 1996. DoD modeling and simulation verification, validation, and accreditation. Department of Defense Instruction 5000.61, Apr.
8. Balci, O. 2001. A methodology for certification of modeling and simulation applications. *ACM Transaction on Modeling and Computer Simulation*, 11 (4): 352-377.
9. Tullos-Banks, H.L., Parker, C.T. and Collins, K.W. 2005. Verification, validation and accreditation of federations. *Proceedings of the 2005 Spring Simulation Interoperability Workshop*, www.sisostds.org (accessed June 2009).
10. Sargent, R.G. 2007. Verification and validation of simulation models. *Proceedings of the 2007 Winter Simulation Conference* (Henderson, S.G., Biller, G., Hsieh, M.-H., Shortle, J., Tew, J.D., Barton, R.R., eds.), IEEE, Piscataway, NJ: 124-137.
11. Jacquart, R., Bouc, P. and Brade, D. 2004. A common validation, verification and accreditation framework for simulatons: project JP 11.20, REVVA. *Proceedings of the 2004 Spring Simulation Interoperability Workshop*, www.sisostds.org (accessed June 2009).

12. Schwartzenburg, F., Oates, W., Park, J., Johnson, D., Stutzman, M., Bailey, M. and Youngblood, S. 2007. Verification, validation, and accreditation (VV&A). One voice – unified, common and cross-cutting. *Proceedings of the 2007 Summer Computer Simulation Conference* (G.A. Wainer, ed.), Society for Modeling and Simulation International, San Diego, CA: 429-436.

13. Stutzman, M. and Park, J. 2004. Navy modeling and simulation standards for verification, validation, and accreditation. *Proceedings of the 2004 Fall Simulation Interoperability Workshop*, www.sisostds.org (accessed June 2009).

14. Arthur, J.D. and Nance, R.E. 1996. Independent verification and validation: a missing link in simulation methodology? *Proceedings of the 1996 Winter Simulation Conference* (Charnes, J.M., Morrice, D.J., Brunner, D.T. and Swain, J.J., eds.). IEEE, Piscataway, NJ: 230-236.

15. Balci, O. 1998. Verification, validation and testing. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice* (Banks, J. ed.), Wiley, New York: 335-393.

16. Robinson, S. 1998. Measuring service quality in the process of delivering a simulation study: the customer's perspective. *International Transactions in Operational Research*, 5 (5): 357-374.

17. Ören, T.I. 1981. Concepts and criteria to assess acceptability of simulation studies: a frame of reference. *Communications of the ACM*, 28 (2): 190-201.

18. Ören, T.I. 1984. Quality assurance in modelling and simulation: a taxonomy. *Simulation and Model-based Methodologies: An Integrative Approach* (Ören, T.I., Zeigler, B.P. and Elzas, M.S., eds.). Springer-Verlag, Heidelberg, Germany: 477-517.

19. Liu, F., Yang, M. and Zicai, W. 2005. Study on simulation credibility metrics. *Proceeding of the 2005 Winter Simulation Conference* (M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.). IEEE, Piscataway, NJ: 2554-2560.

20. Boehm, B.W., Brown, J.R., Kaspar, H., Lipow, M., Macleod, G.J. and Merrit, M.J. 1978. *Characteristics of Software Quality*. North-Holland, Amsterdam.

21. Balci, O. Nance, R.E., Arthur, J.D. and Ormsby, W.F. 2002. Expanding our Horizons in Verification, Validation, and Accreditation Research and Practice. *Proceedings of the 2002 Winter Simulation Conference* (*Conference* (E. Yücesan, C.-H. Chen, J.L. Snowden and J.M. Charnes, eds.). IEEE, Piscataway, NJ: 653-663.

22. Balci, O. 2004. Quality assessment, verification, and validation of modeling and simulation applications. *Proceeding of the 2004 Winter Simulation Conference* (R,G, Ingalls, M.D. Rossetti, J.S. Smith and B.A Peters, eds.). IEEE, Piscataway, NJ: 122-129.

23. Sargent, R.G. 1981. An assessment procedure and a set of criteria for use in the evaluation of computerized models and computer-based modelling tools. RADC-TR-80-409, Rome Air Development Center, Air Force Systems Command, Griffiths Air Force Base, NY.

24. Fossett, C.A., Harrison, D., Weintrob, H. and Gass, S.I. 1991. An assessment procedure for simulation models: a case study. *Operations Research*, 39 (5): 710-723.

25. Gass, S.I. 1977. A procedure for the evaluation of complex models. *Proceedings First International Conference in Mathematical Models*, University of Missouri.

26. Gass, S.I. 1993. Model accreditation: a rationale and process for determining a numerical rating. *European Journal of Operational Research*, 66: 250-258.

27. Saaty, T.L. 1980. *The Analytic Hierarchy Process*. McGraw-Hill, New York.

28. Pohl, S. and Brade, D. 2005. Using fuzzy multi attribute decision theory in the accreditation of models, simulation , and federations. *Proceedings of the 2005 Spring Simulation Interoperability Workshop,* www.sisostds.org (accessed June 2009).

29. DMSO 2000. V&V tools. vva.dmso.mil/Ref_Docs/VVTools/vvtools.htm (created November 2000, accessed September 2006)

30. Balci, O, Adams. R.J., Myers, D.S. and Nance, R.E. 2002. A collaborative evaluation environment for credibility assessment of modeling and simulation applications. *Proceedings of the 2002 Winter Simulation Conference* (*Conference* (E. Yücesan, C.-H. Chen, J.L. Snowden and J.M. Charnes, eds.). IEEE, Piscataway, NJ: 214-220.

31. Ke, F., Ming, Y. and Zicai, W. 2005.The HITVICE VV&A environment**.** *Proceeding of the 2005 Winter Simulation Conference* (M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.). IEEE, Piscataway, NJ: 1220-1227.

32. Robinson, S. 2002. Modes of simulation practice: approaches to business and military simulation. *Simulation Practice and Theory*, 10: 513-523.

33. Gass, S.I. 1977. Evaluation of complex models. *Computers and Operations Research,* 4: 27-35.

34. Youngblood, S.M., Pace, D.K., Eirich, P.L, Gregg, D.M. and Coolahan, J.E. 2000. Simulation verification, validation, and accreditation. *Johns Hopkins APL Technical Digest*, 21 (3): 359-367.

35. Page, E.H., Canova, B.S. and Tufarolo, J.A. 1997. A case study of verification, validation, and accreditation for advanced distributed simulation. *ACM Transactions on Modeling and Computer Simulation*, 7 (3): 393-424.

36. Cochran, J.K., Mackulak, G.T. and Savory, P.A. 1995. Simulation project characteristics in industrial settings. *Interfaces*, 25 (4): 104-113.

37. Diener, D.A., Hicks, H.R. and Long, L.L. 1992. Comparison of models: ex post facto validation/acceptance? *Proceedings of the 1992 Winter Simulation Conference* (Swain, J.J., Goldsman, D., Crain, R.C. and Wilson, J.R., eds.). IEEE, Piscataway, NJ: 1095-1103.

38. Robinson, S. and Pidd, M. 1998. Provider and customer expectations of successful simulation projects. *Journal of the Operational Research Society*, 49 (3): 200-209.

39. Robinson, S. 2004. *Simulation: The Practice of Model Development and Use*. Wiley, Chichester, UK.

40. Wall, S. and Shaw, I. 2002. Retrieval of intermediate-level waste at Trawsfynydd nuclear power station. *Nuclear Energy: Journal of the British Nuclear Energy Society*, 41 (6): 391-396.

41. Lanner (2006). Witness 2006. www.lanner.com (accessed June 2009)

42. Landry, M., Malouin, J.L. and Oral, M. 1983. Model validation in operations research. *European Journal of Operational Research*, 14 (3): 207-220.

43. Sargent, R.G. 1982. Verification and validation of simulation models. *Progress in Modelling and Simulation* (Cellier, F.E., ed.). Academic Press, London: 159-169.