Image Cover Sheet

CLASSIFICATION UNCLASSIFIED	**SYSTEM NUMBER** 144700

TITLE

CALIBRATION/VALIDATION AND SENSITIVITY ISSUES OF SIMULATIONS

AN: 94-06139

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to: JR

CALIBRATION/VALIDATION AND SENSITIVITY ISSUES OF SIMULATIONS

Paul J.F. Bonenfant

Weapon Systems Section, Armaments Division
Defence Research Establishment Valcartier
Courcelette, Québec

1. Introduction.

With the continuing decline of military budgets and the accessibility of increasingly more powerful computer systems, implementing simulation softwares that could help in the decision-making process within the Canadian Forces is becoming more attractive than ever. Whereas before our only source of information came from empirical testing, now we can use computers to gain insight. For example, if the effectiveness of a new missile is to be determined, it is much less expensive to have a computer simulate an attack against a military vehicle than it is to perform the actual attack. With this philosophy in mind, significant effort has been put into implementing simulation models to conduct military studies. Between the battle models and the ballistic flight and penetration simulators lies models that consider the interaction between a single threat and target. In this category, DREV has produced detailed lethality/vulnerability/survivability models for sea (GVAM), air (DRACM) and land (VULCAN) studies.

Since we will be referring to the last of these three often, a brief description is in order. VULCAN is an engineering based discrete-event system simulator that simulates attacks on land combat vehicles by various threats. The geometric target, complete with major components (i.e. main gun, ammunition, personnel etc.), is created with the aid of a CAD-type software. Roughly speaking, when an attack is simulated against a target, a digitized projectile (usually an anti-tank flechette or a shaped charge) is propelled through the target. The components traversed are recorded and the relative damage inflicted on each component is calculated. As well, the effects of behind armour debris (BAD) are also taken into account. Whenever the main penetrator transpierces armour, the model generates a representative spray of fragments propelled from the point of exit of the projectile from the armour. Analogous to the case of the main penetrator, the components which lie in a fragment's trajectory are recorded and their damage assessed. Once all of this information is collected, VULCAN evaluates the overall damage to systems (i.e. Firepower, Mobility, Personnel). The output is then used to assess the survivability/vulnerability of the target.

As with any simulation model, the question which arises most often is: How much credibility can we ascribe to the results obtained? Indeed if the model was built to aid in the decision-making process, it is necessary to address this question very early in the life of the model. Hence a systematic procedure is required in order to ascertain that the model not only behaves as it was intended to, but is also representative of the actual system. The present paper expounds on the various methodologies considered to aid in answering the above question with respect to VULCAN.

2. Definitions.

One usually speaks of *validating* a simulation model. In the literature there is no agreed-upon definition of validation (see [8] sec 3.2 for an example of differing definitions), thus we state the definition which we have adopted: *Validation is a process carried out by comparison of model predictions with independent field observations and experimental measurements. A model cannot be considered validated until sufficient testing has been performed to ensure an acceptable level of predictive accuracy.* This is the formal definition employed by the International Atomic Energy Agency (IAEA) for the purpose of studying their simulation models [8]. By *calibration* we mean the adjusting of the model parameters in order to obtain better predictive accuracy. And by *calibration/validation*, we mean the interactive process of adjusting the model parameters and comparing the resulting simulated results with the system.

Before any consideration be given to validating a model, it should have passed a thorough verification process. By this we mean that the computerized implementation of the *conceptual model* (the mathematical/logical representation of the system to be simulated) has been properly de-bugged and functions as it was intended. The verification and the computer implementation processes have been completed on VULCAN, and so we will not discuss these issues here. Hence when we speak of our simulation model, we are referring to our *computerized* model (i.e. the computer implemented conceptual model). The interested reader can find further information on verification of the computer implementation of conceptual models in [1,2,5,9].

A word of warning: Caution should be exercised when encountering the word verification in the simulation literature. Although at present the usage seems to be mostly as we use it here, in earlier works, such as the landmark paper [3], the word *verification* is used much in the same context as we use *validation* .

3. The Historical Methodologies.

The landmark paper [3], quoted by many articles on model validation, establishes a theoretical framework for validation methodologies. We briefly describe the three basic philosophies contained therein.

Firstly there is *Rationalism.* The idea here is that a simulation model should be comprised of self-evident truthful statements (or *synthetic a priories*) interconnected by valid logical flows. For a model built by rigorously applying these principles, there is no question with respect to its validity. Unfortunately one cannot find many self-evident truthful statements in complex systems. Indeed, some argue that there exists no such statements [3]. At the opposite end of the spectrum lies the the concept of *Empiricism.* Here no statement is accepted which cannot be empirically verified. Hence a validation would consist of an exhaustive experimentation of every component of a model. One can quickly see how a strict adherence to this philosophy would be much too expensive or even utterly impossible. The third view on validation is that of *Positive Economics*. Here less emphasis is placed on the processes involved, the validity of the underlying assumptions is not even considered. A model is deemed valid if it can accurately predict the behaviour of the dependent variables which are treated by the model. Although it is true that the ultimate goal is to have accurate predictions, accepting underlying assumptions without regard to any scientific scrutiny is rather dubious.

Naylor and Finger [3] argue that not any one of the above schools of thought is sufficient, but a *multi-stage* validation, in which all three of the above are incorporated, is in order. This validation method consists of (1) developing the (conceptual) model on theory, knowledge of the system and intuition; (2) verifying the model assumptions where possible by empirically testing them; and (3) comparing the outputs produced by the (computerized) model with the real system output given similar initial conditions [5]. It should be noted that Naylor and Finger do not address the question of computer implementation.

4. The Task at Hand.

In Figure 1 we display the flowchart for the VULCAN Validation Project. The project can be broken down into four sub-tasks: The first is ascertaining that the geometric model is faithfully represented in the computer, the second considers simulation results without the effects of Behind Armour Debris (BAD), the third investigates the different BAD generating modules and the fourth examines the collected sensitivity information (which may entail more gathering of information).

In order to speak of a valid simulation model, one must first be sure that the model has the proper input. In general *Data Validity* is the determination that the necessary data is both complete and correct [5]. For VULCAN the largest single source of input data is the target vehicle (Indeed it is here that we give the armour thicknesses and basic characteristics of the components of the vehicle i.e. location, size and relative strength) . Thus our Phase I corresponds to the data validity stage of the validation process . DREV has contracted this phase out to local industry and will not be discussed here. The other main source of input data is the threat. For example, in the case of a shaped charge, threat information such as standoff distance, fusing angle and penetration ability versus standoff distance are given as input. This information is either obtained empirically at DREV or is furnished by industry.

In the second and third phases we begin looking at the output produced by the model and comparing it to the real system. Hence that is where the validation, as defined earlier, occurs. Before getting into the specifics of our assignment we first consider some general methodologies for performing the actual validation.

5. The Validation Process.

Of paramount importance in this stage of model development is the participation of the decision makers that will ultimately rely on the outputs of the model for guidance. It is well known that users (in our case military officers) are more inclined to accept a simulator in which they had an active role in implementing. One of the first validity tests that a model undergoes is the so-called *face validity* or expert opinion test. Here someone knowledgeable in the system modelled is asked to peruse the model's output in search of gross discrepancies with the actual system. Once the model is deemed to produce reasonable results, a closer scrutiny of its agreement with the actual system is performed. From sample runs of VULCAN on a T-72 MBT it was found [10] that initial conditions on the personnel within the vehicle were such that we had a crew of supermen! This problem has since been rectified.

The comparison phase (which is the most conclusive) is done within the context of *experimental frames*. An experimental frame defines a limited set of circumstances under which the system (or the model) is to be observed or subjected to verification (sic) [4].

The VULCAN Validation Project

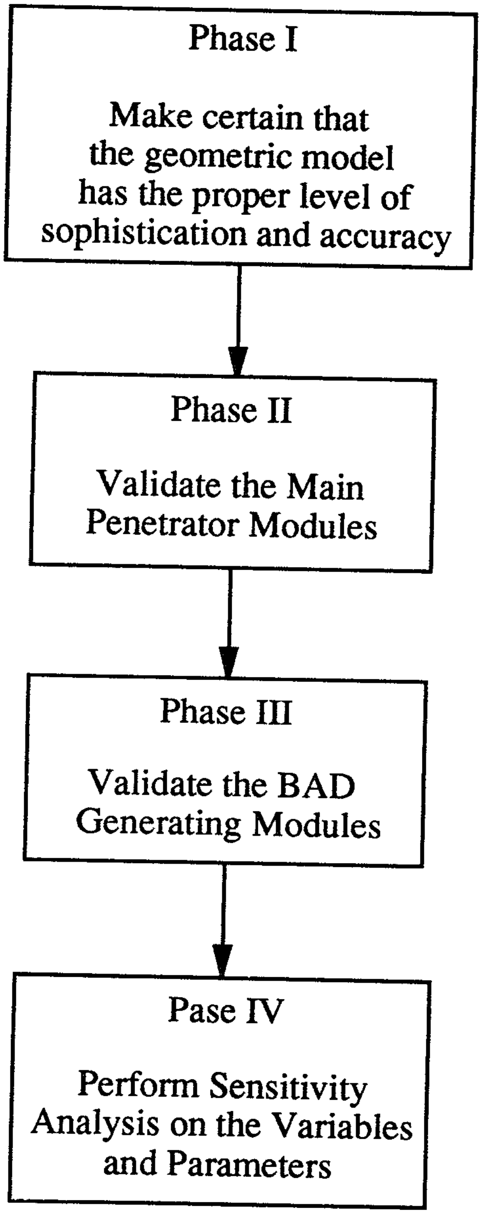

Figure 1

In the literature one says a model is *replicatively valid* if, under the same experimental frames, it produces comparable results as did the system some time in the past. Similarly a model is said to be *predictively valid* if it accurately predicts the outcome of the system under the same experimental frames. Ideally we would have enough observations of the system for every conceivable experimental frame. We would then perform simulation runs and then simply compare the simulated results with the real outcome. Fortunately we do not live in an ideal world and, for any complex system, it is impossible to have all of the possible outcomes for an infinite number of different experimental frames. Further, if we did have all of this information at our disposal, there would be no need for a simulation model. We would need only perform some type of optimal search to find, for example, the best policy decision, or the best missile to use against a T-72 MBT. But we have to make due with a finite number of outcomes from a finite set of experimental frames.

For systems that are already in existence, the most reliable source of system output data is historical data. If the model can be shown to be replicatively valid, then confidence in its predictive ability should be quite high. One comparison method which uses historical data and involves people knowledgeable in the system is the so-called *Turing Test* [7]. In this test certain experimental frames, for which historical data on the system exist, are chosen and simulated. The outputs of both the real system and the model are put in the same document form. Extraneous information is removed so that differentiating between simulated and original documents is not trivial. Given, say, five simulated and five genuine forms one then shuffles these and asks the participant (manager or military officer) to try and distinguish between the real and simulated documents. After having decided on a particular document, he/she is then asked to enumerate the reasons for which the particular decision was made. Studies show that this creates a healthy exchange of information which aids in the calibration/validation of the model. Two such historical experiments as well as a statistical method for interpreting the results are documented in [7].

As our system consists of the interaction of a threat with an armoured vehicle, historical data are the results of live-fire trials. Since there is a limited number of live-fire trials at our disposal, we must also resort to alternate sources of information to perform our validation. One such source can be another (validated) simulation model which performs similar tasks. The American VAMP assessment module will be used at DREV for these purposes. Some international cooperation in this field of study is also anticipated. An action group, under the auspices of The Technical Cooperation Program (TTCP; An agreement between Australia, Canada, New-Zealand, the U.K., and the United States for exchange of information in R&D), will soon look at the sensitivity issues (see below) pertaining to these types of models. This will also create a good forum for the exchange of data (live-fire and/or simulated test results) for validation purposes.

If the simulation model is comprised of several independent modules, then it may be possible to validate each of these independently. This allows for sharper focus on fewer parameters, easier comparisons of the results (real and simulated) and provides an added safeguard verifying for degeneracy (i.e. the quality that independent modules have no influence on one another, see [6]). As for VULCAN the BAD and the main penetrator modules are disjoint in the damage that they inflict to the vehicle components. Thus we have decided to break down the actual validation process, as defined earlier, into two steps: The first step (phase II in Figure 1) consists of calibrating/validating the main penetrator module by using simulation runs without the use of BAD.

The second (phase III in Figure 1) involves calibrating/validating the BAD generating modules. In this case we will use simulation runs with the BAD engaged.

6. Sensitivity Analysis.

The final phase of our project is the sensitivity analysis/parameter variability test. Much sensitivity information will have been collected during the first phases of validation. The interplay between calibration and validation in those earlier phases will produce, for many variables and parameters, a converging sequence of parameter values. Given one such parameter, any given member of the aforementioned sequence (which would necessarily converge to its validated value) will have its effect on the output recorded, and thus the sensitivity information can be readily extracted. However for some variables additional investigation may be required. Usually this is done by changing the input and internal parameters of a model one at a time in varying increments to determine the influence this has on the output of the model. At the beginning one may start with extreme values for the parameters and then converge onto the pre-assigned (validated) value. Of particular interest, for all of the variables and parameters, is whether or not any thresholds exist. These will serve as guide in determining the level of measurement precision best suited for each case. If the output turns out to be sensitive to certain parameter changes, then more attention should be given to the measurement of that particular parameter. Conversely if the output seems immune to certain levels of change in some other input variable, we know that measurements of this variable need not be so stringent. There is also the give-and-take relationship that develops between cost considerations and the desire of accurate predictions. This issue can best be resolved with the use of well-documented sensitivity information.

7. Configuration Management.

Once a model is deemed to be valid it should be certified as such and then, and only then should the model be made available for its intended use (lest confidence in the model is precluded by erroneous results obtained from a premature model). Moreover, any computer model which is used on an ongoing basis will evolve through time to reflect changing environmental conditions (i.e. for us new armored vehicles and/or weapons to evaluate). Some safeguards must be in place to assure the validity of the model from one generation to the next. There should be a complete set of verification and validation tests to be used whenever any changes are made to the model to go from one generation to the next (of course not all of these tests need be necessary given a particular change). Also care should be taken to preserve at least one past generation of the certified model as a benchmark for other generations [6]. Schlesinger *et al.* argue further that "only certified versions of the model preserved in inviolate form should be used in actual simulation studies." This is especially true in large organizations like the federal government. There it is not uncommon for one agency to develop a model, another to implement the model and yet another to hold the software management of the model (and we have not considered any users yet!). It is clear that some body must take on the responsibility of maintaining, through strict configuration management, the integrity of the model. This body would also be responsible for advertising the existence of the model and helping interested parties in using the model. As well as providing user manuals and information (through documentation, seminars, tutorials etc.), clear guidelines outlining the model's capabilities and, more importantly, it's limitations should be made available to potential users.

It should be clearly spelled out that a model which is used outside of it's scope of applicability, should not be expected to give valid results.

8. Concluding Remarks.

It is evident that the above methodology entails many hours of preparation, simulation runs and analysis. Therefore it should be clear that a treatment as rigorous as this may not always be in order. Consideration should be given as to whether the impact of decisions taken with the aid of the model will warrant the investment. The potential for uses of VULCAN in procurement alone render such an extensive study germane.

In this paper we have briefly discussed some issues pertaining to simulation model calibration/validation and sensitivity analysis. We have outlined a framework for a thorough validation procedure for the simulation model VULCAN. The general methods described herein were collected from the referenced literature and adapted for our purposes. Again we wish to stress the importance of strict configuration management of a simulation model. It is worth noting in closing that a current area of active research is a search for a comprehensive framework to assess the adequacy of any simulation study [8].

REFERENCES

1. Gass, S. I., Evaluation of Complex Models. *Computers and Operations Research*, **4**, (1979), pp. 27-35.

2. Law, A.M., Validation of Simulation Models, I: An Overview and Survey of Real-World Practice. Technical Report 78-15 (1979), Department of Industrial Engineering, University of Wisconsin, Madison.

3. Naylor, T.H. & Finger J.M., Verification of Computer Simulation Models. *Management Science*, **14**, No. 2, (1967), pp. B92-101.

4. Ören, T.I., Concepts and Criteria to Assess Acceptability of Simulation Studies: A Frame of Reference. *CACM,* **24**, No. 4, (1981), pp. 180-189.

5. Sargent, R.G., Verification and Validation of Simulation Models. *Progress in Modelling and Simulation*, ed. by F.E. Cellier, Academic Press, London (1981), pp. 159-169.

6. Schlesinger, S. *et al.*, Developing Standard Procedures for Simulation Validation and Verification. Proceedings of 1974 SCSC, (1974), pp. 927-933.

7. Schruben, L.W., Establishing The Credibility of Simulations. *Simulation*, **34**, (1980), pp 101-105.

8. Sheng, G. *et al.*, Model Validation: A Systematic Approach. In *Reliability Engineering and System Safety Journal.*, (In Press-1993).

9. Winter, E.M. *et al.*, Verification of Engineering Simulations with Minimal Data. Proceedings of 1976 Summer Computer Simulation Conference, (1976), pp. 718-723.

10. Wubbolts, H. Capt., Private Communication. Defence Research Establishment Valcartier, (Dec. 1992).