STUDY ON SIMULATION CREDIBILITY METRICS

Fei Liu
Ming Yang
Zicai Wang

Control and Simulation Center
Harbin Institute of Technology
Harbin, Heilongjiang 150001, P.R.C.

ABSTRACT

Currently, there appears to be an over-preoccupation with building simulation validity in simulation credibility evaluation. However, today's simulation systems become more complex and larger, the only validity metric can't represent simulation credibility, and there is a need for other credibility metrics. Therefore, we should rethink the basic problem in the simulation community: what are the metrics of simulation credibility. In this paper, credibility metrics are deeply investigated and presented, measurement methods for credibility metrics are discussed, a new approach to synthesis of credibility metrics is presented, and a credibility metrics driven VV&A process is discussed.

1 INTRODUCTION

Model or simulation credibility is the user or decision maker's confidence in the model or simulation, and verification, validation, and accreditation (VV&A) is the necessary way to build up model or simulation credibility. Any model or simulation must be credible before used. However, what make up simulation credibility? Which metrics can be used to represent and quantify simulation credibility?

Currently, there appears to be an over-preoccupation with building simulation validity in simulation credibility evaluation. However, both the size and complexity of modeling and simulation (M&S) application are growing; the domains of application continue to expand; the larger M&S projects are requiring increased levels of involvement of people with diverse capabilities and background; the expected lifetime of M&S application is continually increasing (Arthur et al. 1999). All these changes introduce more uncertainties and complexities into simulation credibility evaluation, which make credibility evaluation more difficult and make the users more difficult to demonstrate whether the simulation is credible. In this case, we cannot simply compare the simulation outputs with the real out-

puts to assess the simulation validity that cannot represent simulation credibility yet, and there is a need for other credibility metrics.

Therefore, we should rethink the basic problem in the simulation community: what are the metrics of simulation credibility. Furthermore, to solve the problem is the all-important problem for current credibility evaluation.

The remainder of this paper is organized as follows: Section 2 presents the credibility metrics, discusses measurement methods for credibility metrics, and presents the approach to synthesis of credibility metrics; Section 3 presents the credibility metrics driven VV&A process; and Section 4 presents the summary.

2 CREDIBILITY METRICS

This session will present the definition of credibility metrics, discuss measurement methods for credibility metrics, and present the approach to synthesis of credibility metrics.

2.1 Definition of Credibility Metrics

The credibility of a model or simulation is an expression of the degree to which one is convinced that a particular model or simulation are suitable for an intended purpose. The degree of confidence in the model or simulation is only based on some metrics that reflect the property of the model or simulation. Based on our experiences and existed references, we think that at least five metrics contribute to the evaluation of simulation credibility, which are validity, correctness, reliability, usability and interoperability. Figure 1 gives the dependencies between these metrics and VV&A. Then these metrics will be in detail discussed.

2.1.1 Validity

Validity is defined as the quality of being inferred, deduced or calculated correctly enough to suit a specific application (Gross 1999). The validation process establishes the valid-

ity of the model or simulation, and provides a crucial piece of evidence to support model or simulation credibility for a particular application. Although the validity of a model or simulation supports but cannot guarantee its credibility in all cases, validity is still the most important metric that represents and quantifies simulation credibility (DMSO 2000).
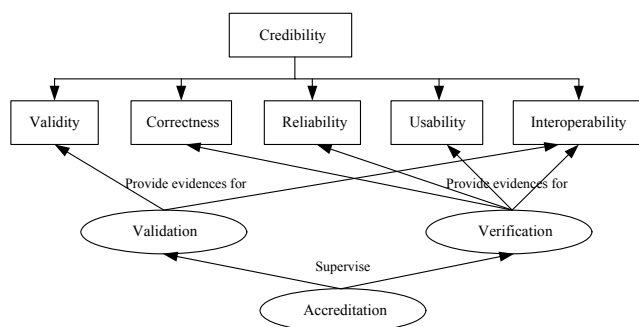

Figure 1: Dependencies Between Credibility Metrics and VV&A

Validity is the basic problem of the study on credibility. Previous credibility study basically focused on validity, and this can be seen from Figure 2 (Sargent 2000).
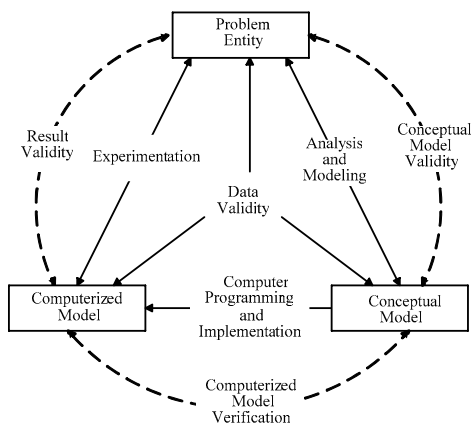

Figure 2: Simplified Version of the Modeling Process

From Figure 2, validity can be divided into three aspects: conceptual model validity, data validity and result validity. Conceptual model validity is determining that (1) the theories and assumptions underlying the conceptual model are correct, and (2) the model representation of the problem entity and the model's structure, logic, and mathematical and causal relationships are "reasonable" for the intended purpose of the model. Data validity is determining that data is appropriate, accurate, and sufficient and if any data transformations are made, they are correctly performed. Result validity is concerned with determining that the model's output behavior has the accuracy required for the model's intended purpose over the domain of its intended applicability (Sargent 2000).

Figure 2 also applies to complex simulation systems, but the latter increases the difficulty to quantify validity. Therefore, new approaches to measurement of validity of complex simulation systems are in demand. The mature of fidelity theory improves quantification of validity and a new way to measure validity based on fidelity in Figure 3 is hopeful (Gross 1999).
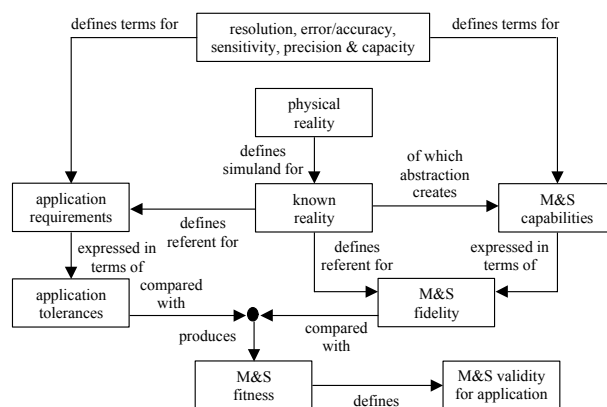

Figure 3: Fidelity Conceptual Framework

In summary, validity is the most important metric that contributes to simulation credibility evaluation. Although the fidelity theory provides a new way to measure the validity, this is still an important problem for further study.

2.1.2 Correctness

Correctness is dealt with by verification. When using a model or simulation, one expects them to be internally consistent, correctly described in all their different representation forms, and completely consistent with each other. All these problems are concerned with correctness. The term "correctness" is also used in related fields, e.g., in software verification. However, software development is only one aspect of M&S, and M&S includes many other things, e.g., life simulation and virtual simulation. Thus, in the context of verification of models and simulation, the concept of correctness is used in a wider sense than in software verification (Brade 2003).

Usually, model or simulation verification provides the following evidences for correctness:

- Simulation requirement verification results,
- Simulation design verification results,
- Simulation implementation verification results.

From the nature of verification and validation, correctness and validity are the basic elements of credibility. However, they are not sufficient for credibility evaluation in the context of complex simulation systems.

2.1.3 Reliability

Advances in computing technologies are enabling the simulation community to assume more ambitious requirements in order to increase the overall accuracy and realism of simulation exercises. A complex simulation system often consists of many subsystems or components, which are interconnected in a local area network or in a wide area network. Such a complex simulation environment mainly includes network, simulation host platforms, operating system, run-time support software and simulation application. On the one hand, this complexity causes such effects as communication delays, noise or sudden cut-offs; on the other hand, such a complex system hides many errors and faults. All these disadvantages momentously affect the simulation results, and consequently affect the use's confidence in simulation results. Credible simulation outputs are based on a reliable simulation system. Based on these considerations, we think that reliability is one of important metrics that represent and quantify simulation credibility. For complex simulation systems, reliability should be taken into account in two levels: system level reliability and subsystem level reliability, and system level reliability is based on subsystem level reliability. Then, we will deeply discuss how to measure reliability of complex simulation systems.

Usually, reliability is defined as the ability of a product to perform a required function under stated conditions for a stated period of time, which is measured by reliability model. Reliability model refers to diagram and mathematics model that are used to predict or estimate system reliability. By reliability model, one can decompose system reliability into reliability of simple components, and then quantitatively predict, allocate, estimate, and evaluate reliability. Bruzzone presented a reliability example concerned with complex HLA simulation systems (Bruzzone et al. 2002), which can give us some suggestions on how to measure reliability of complex simulation systems.

2.1.4 Usability

Our definition of simulation usability is not framed in terms of "ease of simulation use", but rather in terms of "reduced probability of simulation misuse". This operational definition stems from the twin observations that simulations are credible only within a well-defined usage context, and only when they are properly used within that context. Any simulation attribute that reduces the probability of simulation misuse enhances its credibility within a given context (Muessig et al. 2000). Thus we consider usability as one of important metrics to represent and quantify credibility.

By usability, then, we refer to that collection of simulation user support features that facilitate credible use of the simulation and reduce the probability that it will be employed inappropriately. Examples of such features are: training in proper simulation use and interpretation of outputs; accurate and comprehensive simulation documentation; on-call technical support for simulation users; simulation user groups that meet on a regular basis; the existence and implementation of a sound configuration management process for the simulation, both during and after development; the availability of trained simulation operators and analysts who can run the simulation and interpret its outputs correctly; and any other support feature that can help simulation users ensure credible use of the simulation.

Note that simulation usability is a necessary, but not sufficient, condition for simulation credibility. No simulation user support feature, no matter how well designed to minimize the probability of simulation misuse, will militate against improper use of the simulation (Muessig et al. 2000).

2.1.5 Interoperability

As described above, complex simulation systems often connect lots of different kinds of subsystems. On the one hand, this expands the application range and purpose of simulation; on the other hand, this makes interoperability between simulation systems more difficult. Therefore, DMSO bring forward "fostering of the interoperability, reuse, and affordability of M&S" (DODD 1994). Now interoperability has become an important but difficult problem in simulation development and validation. How the interoperability is realized heavily affects how the users' needs are met, therefore, interoperability has become one of important metrics that affect credibility.

At present, there is not a commonly accepted definition of interoperability. DMSO gives below definition "the ability of a model or simulation to provide services to and accept services from other models and simulations, and to use the services so exchanged to enable them to operate effectively together" (DMSO 1997). Like many other definitions of interoperability, this definition still doesn't address the basic issues that simulation focuses on. Interoperability is much more than data exchange. Interoperability has to include all the elements that make it possible for the interconnected models and simulations to operate effectively together, such as the context of the problem being addressed, the syntax and semantics of the data being provided and the level of detail and fidelity consistent across the entire simulation system (Clark et al. 2001).

In order to measure interoperability, the elements of interoperability need to be understood. Therefore, below interoperability framework is developed, shown in Figure 4, which decomposes interoperability into three levels: physical, technical, and functional.

The physical interoperability components refer to real, physical objects that have specific geo-spatial coordinates in real-time, including networks, nodes, circuits, equipment,

and platforms which are used to implement physical transportation of simulation data. Interoperability of this level mainly includes hardware (e.g., network cards) compatibility, network protocols compatibility, and operating systems compatibility.
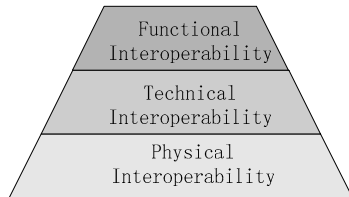

Figure 4: Interoperability Framework

The technical interoperability components include rules, services, interfaces, and standards, which are used to realize information and data management, and date exchange specifications definition. The interoperability of this level addresses Coordinated services (e.g., time management services, synchronization point services, save and restore services, and data distribution services), and standard (e.g., SOM and FOM) compliance.

The functional interoperability components include entities and environments that are used to realize the simulation purpose. This level in fact addresses the problems of substantial interoperability which includes interoperability among entities, and interoperability among entities and environments. Interoperability among entities includes entity level of representation, entity attribution, entity behaviors, temporal resolution, and spatial resolution. Interoperability among entities and environments include the effects of environment on entities, and the effects of entities on environment (Dahmann et al. 1999).

2.2 Measurement of Credibility Metrics

Many methods exist for measuring credibility metrics, and in the followings, only two kinds of these methods are discussed.

2.2.1 Statistical Techniques

The statistical techniques require the system being modeled to be completely observable, i.e., that all data required for model validation can be collected from the system. The model is validated by using the statistical techniques to compare the model output data with the corresponding system output data after the model is run with the same input data as the real system. The classical and widely used statistical techniques are confidence interval and hypothesis testing. Statistical techniques provide objective measurement for credibility metrics, so they should be made full use of. However, it is frequently not possible in practice to use either of these techniques because (a) the statistical assumptions required cannot be satisfied or only with great

difficulty (assumptions usually necessary are data independence and normality) and/or (b) there is insufficient quantity of system data available that causes the statistical results not to be "meaningful" (Sargent 2000).

2.2.2 Expert Assessments

Although statistical techniques are the most effective methods, expert assessments are the most widely used methods. Classical expert assessment methods contain inspection, walkthrough, Turing test, and face validation, which are used throughout the life cycle of simulation development. This kind of validation technique is widely used both because its capabilities can provide insight about model and simulation capabilities and because adequate real world (experimental) data are often not available to allow robust quantitative validation.

To make expert assessments more effective, Pace in detail discusses the ways that experts are selected, managed, and used (Pace and Sheehan 2002). Besides, the fuzzy set theory can be used in acquiring expert assessment opinions, and rough set theory can be used in simplifying expert assessment opinions and credibility metrics.

2.3 Synthesis of Credibility Metrics

After measurement of credibility metrics, lots of certain and uncertain assessment results are produced, and then we need synthesize these assessment results in order to obtain the ultimate assessment values for each credibility metrics. As these assessments results are the mixtures of uncertain and certain information, we find that the Dempster Shafer (D-S) theory (Shafer 1976) is an effective method to synthesize these results.

The D-S theory is based on the specification of a triple $(S, £, m)$, where (i) S is a frame of discernment that contains everything that could occur in the particular universe under consideration, (ii) $£$ is a countable collection of subsets of S, and (iii) m is a function defined on subsets of S such that $m(E) > 0$ if E $£$, $m(E)=0$ if $E \subset S$ and $S \notin £$, and $\Sigma_{E \ £} m(E)=1$. For a subset E of S, $m(E)$ characterizes the amount of likelihood that can be assigned to E but to no proper subset of E.

In the terminology of the D-S theory, (i) S is the sample space or universal set, (ii) $£$ is the set of focal elements for S and m, and (iii) $m(E)$ is the basic probability assignment (BPA) associated with a subset E of S.

The belief, Bel(E), and plausibility, Pl(E), for a subset E of S are defined by Bel(E)= $\Sigma_{U \ E} m(U)$ and Pl(E)= $\Sigma_{U \cap E \neq \Phi} m(U)$.

In concept, Bel(E) is a measure of the total amount of belief in E and not the amount committed precisely to E by the evidence corresponding to the belief function m. The quantity Pl(E) expresses the plausibility of E, i.e., the

maximum extent to which the current evidence could allow one to believe *E*.

The information contained in Belief concerning a given subset *E* may be conveniently expressed by the Belief interval [Bel(*E*), Pl(*E*)].

By means of the Bel(*E*) and Pl(*E*), uncertain information can be represented in Figure 5. Therefore, the Belief interval is also an effective way to represent uncertain expert opinions in credibility evaluation.
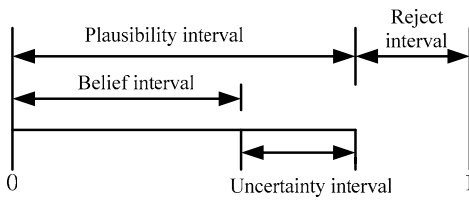

Figure 5: Representation of Uncertain Information

Let Bel_1, Bel_2 and m_1, m_2 denote two belief functions and their corresponding BPAs, respectively. And the focal elements are $B_1,\ldots, B_I$, and $C_1,\ldots, C_J$, respectively. The D-S combination rule defines a new BPA, denoted $m_1 \oplus m_2$, which represents the combined effect of m_1 and m_2. The corresponding belief function, denoted $Bel_1 \oplus Bel_2$, may then be computed from $m_1 \oplus m_2$ by definition of a belief function. And $m_1 \oplus m_2$ is expressed as

$$m_1 \oplus m_2(A) = \begin{cases} 0 & A = \Phi \\ \dfrac{1}{K} \sum_{B_i \cap C_j = A} m_1(B_i) \times m_2(C_j) & A \neq \Phi \end{cases} \quad (1)$$

where $i=1,\ldots, I$; $j=1,\ldots, J$; and $K = 1 - \sum_{B_i \cap C_j = \Phi} m_1(B_i) \times m_2(C_j)$.

The D-S theory can effectively deal with certain and uncertain information, so it is an effective way to combine all kinds of opinions in credibility evaluation. In order to apply the D-S theory to credibility evaluation, we begin at a definition of the frame of discernment *S*. For example, we define *S* as (V_1, V_2, V_3, V_4)=(higher, high, low, lower) for the validity metric and its sub-metrics. Then we decompose the credibility metrics into their sub-metrics, which is shown in Figure 6. Finally, we use the D-S theory to synthesize the expert opinions and then synthesize the sub-metrics. Figure 7 presents the approach to applying the D-S theory to credibility evaluation.
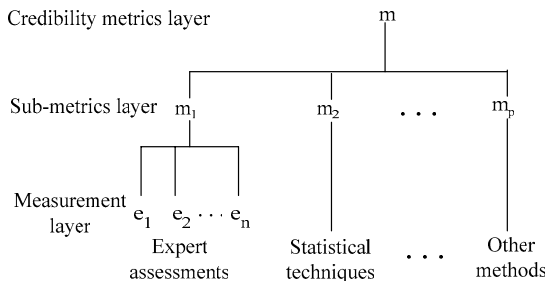

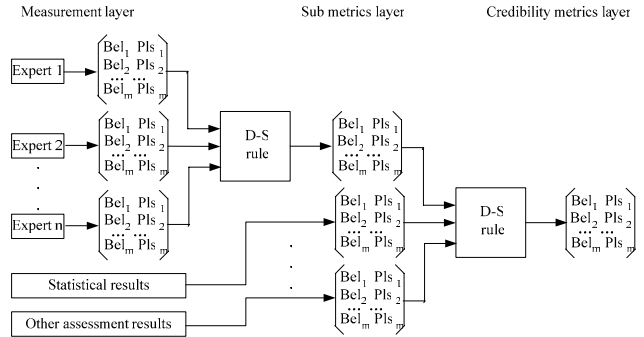Figure 6: Credibility Evaluation Model

Figure 7: Approach to Applying the D-S Theory to Credibility Evaluation

3 CREDIBILITY METRICS DRIVEN VV&A PROCESS

We strongly believe the above five credibility metrics can effectively represent and quantify simulation credibility. As the establishment of confidence in the simulation is the purpose of VV&A, we present below credibility metrics driven VV&A process, shown in Figure 8, under the guidance of which, we can more effectively perform simulation VV&A and credibility evaluation.
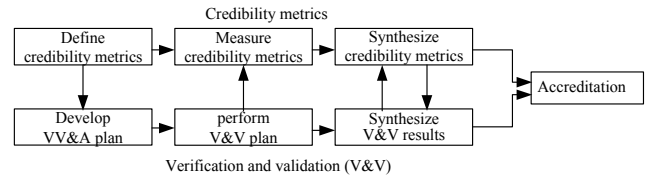

Figure 8: Credibility Metrics Driven VV&A Process

3.1 Develop VV&A Plan

In the beginning, by means of the simulation purpose and requirements, the VV&A agent, experts, and users together define required level of credibility, and further define the levels of credibility metrics. Then the credibility metrics are decomposed into measurable sub-metrics. This process can be done through the dendritic analysis structure (Youngblood and Senko 2002). For example, the interoperability metric is decomposed into functional, technical, and physical aspects. Then the functional interoperability is decomposed into interoperability between entities and entities, interoperability between environments and environments, and interoperability between entities and environments. Besides, this decomposition can be further performed.

According to the definition and decomposition of credibility metrics, the VV&A plan is developed.

3.2 Perform V&V Plan

In this phase, the VV&A agent performs the V&V plan, measure the sub-metrics of credibility, and collect all of valuable assessment results as the sources of the values of sub-metrics of credibility.

3.3 Synthesize V&V Results

In this phase, by means of the collected assessment results, the VV&A agent synthesize the credibility metrics and V&V results, form the final documents, and then present these documents to the accreditation agent for a accreditation decision.

As the credibility metrics driven VV&A process is very complex, such tools as metrics definition tool, data collection tool and statistics tool have to be used to assist in this process. At present, we have developed the CSCW integrated tool platform to assist in this process (Zhang et al. 2002).

The credibility metrics driven VV&A Process has been applied to some large complex simulation system, and it has been proved that this process has below advantages:

- This process effectively implements the definition, measurement, and synthesis of simulation credibility metrics.
- This process makes the VV&A agent focus on the most interested things, thus reducing cost, but increasing efficiency.

4 CONCLUSIONS

In this paper, we discuss a basic problem in the simulation community: what are the metrics of simulation credibility. Solving this problem is essential for current credibility evaluation. The approach to definition, measurement, and synthesis of credibility metrics and the credibility metrics driven VV&A process presented in this paper have been applied to some large complex simulation system and proved effective.

REFERENCES

Arthur, J. D., R. G. Sargent, J. B. Dbney, A. M. Law, and J. D. Morrison. 1999. Verification and validation: what impact should project size and complexity have on attendant V&V activities and supporting infrastructure. In *Proceedings of the 1999 Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 148-155. Phoenix, AZ: IEEE.

Zhang, B., M. Yang, and B. H. Li. 2002. Integration of testing and evaluation for complex simulation system, *Computer Integrated Manufacturing Systems* 8(3).

Brade, D. 2003. A generalized process for the verification and validation of models and simulation results. Phd dissertation, Universität Der Bundeswehr, München.

Bruzzone, A. G., R. Mosca, A. Orsoni, and R. Revetria. 2002. Simulation-based VV&A methodology for HLA federations: an example from the aerospace industry, In *Proceedings of the 35th Annual Simulation Symposium*, 71-76. San Diego, CA.

Clark, D. L., S. K. Numrich, R. J. Howard, and G. Purser. 2001. Meaningful interoperability and the synthetic natural environment. In *Proceedings of the 2001 European Simulation Interoperability Workshop*. Harrow, Middlesex, UK.

Defense Modeling and Simulation Office (DMSO). 2000. Department of Defense Verification, Validation, and Accreditation Recommended Practices Guide. Department of Defense, Alexandria, VA.

Defense Modeling and Simulation Office (DMSO). 1997. DoD Modeling and Simulation Glossary. Department of Defense, Alexandria, VA.

Dahmann, J., M. Salisbury, C. Turrell, P. Barry, and P. Blemberg. 1999. HLA and beyond: interoperability challenges. In *Proceedings of the 1999 Fall Simulation Interoperability Workshop*. Orlando, FL.

DoDD 5000.59. 1994. DoD Modeling and Simulation (M&S) Management. Department of Defense, Alexandria, VA.

Gross, D. C. 1999. Report from the fidelity definition and metrics implementation study group (FDM-ISG). In *Proceedings of the 1999 Spring Simulation Interoperability Workshop*. Orlando, FL.

Muessig, P. R., D. R. Laack, and J. J. Wrobleski. 2000. An integrated approach to evaluating simulation credibility, In *Proceedings of the 2000 Summer Computer Simulation Conference*. Vancouver, British Columbia.

Pace, D. K., and J. Sheehan. 2002. Subject matter expert (SME) peer review in M&S V&V. *Foundations for V&V in the 21st Century Workshop*. Laurel, ML.

Sargent, R. G. 2000. Verification, validation, and accreditation of simulation models. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 50-59. Orlando, FL: IEEE.

Shafer, G. 1976. *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.

Youngblood, S., and B. Senko. 2002. Acceptability criteria: how to define measures and criteria for accrediting simulations. In *Proceedings of the 2002 Fall Simulation Interoperability Workshop*. Orlando, FL.

AUTHOR BIOGRAPHIES

FEI LIU is now a Ph.D candidate majoring in simulation VV&A and credibility evaluation. His research interests include VV&A, and complex simulation modeling method. His email address is <liuf_2005@yahoo.com.cn>.

MING YANG is a Professor at Harbin Institute of Technology. He received his education at Harbin Institute of Technology. His research interests include system simulation, system control and VV&A. His email address is <liuf_2005@yahoo.com.cn>.

ZICAI WANG is a Professor at Harbin Institute of Technology and a member of the Chinese Academy of Engineering. He received his education at Harbin Institute of Technology. Prof. Wang has served his profession for more than forty years. His research interests include system simulation and fuzzy and neural network theory. His email address is <liuf_2005@yahoo.com.cn>.