Proceedings of the 2004 Winter Simulation Conference
R .G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

FOUNDATIONS OF VALIDATING REUSABLE BEHAVIORAL MODELS IN ENGINEERING DESIGN PROBLEMS

Richard J. Malak, Jr.
Christiaan J. J. Paredis

Systems Realization Laboratory
G.W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA 30332 U.S.A.

ABSTRACT

We present a conceptual framework for validating reusable behavioral models. The setting for this work is a modern product development environment in which design is performed by teams of specialists that collaborate through model reuse. The various modes of model reuse separate validation-relevant knowledge from the tasks for which it is needed. To enable efficient and effective transfer of this knowledge to the tasks for which it is needed, we propose a framework for validating reusable behavioral models based on formal representations of validation-relevant knowledge. The framework defines the abstract knowledge representation as well as an abstract process for applying this knowledge to validate reusable behavioral models. Although this framework is not a complete solution to the validation problem in design, it forms a foundation for understanding and solving the problem and represents a starting point for future investigation.

1 INTRODUCTION

Engineering design is the process of mapping a set of requirements into a definition of an artifact that will meet those requirements (Pahl and Beitz 1996). It is an iterative process in which designers must evaluate the properties and performance of an alternative—i.e., its *behavior*—relative to the stated requirements. These behavioral evaluations guide decision making when comparing and refining alternatives and thus are central to engineering design.

Designers commonly use computer-based modeling and simulation (M&S) methods to make predictions about artifact behavior. One benefit of computer-based models is their ease of reuse. This trait corresponds well with the iterative and evolutionary nature of engineering design. Furthermore, it enables designers to speed design evaluation and to amortize the cost of model development over many uses. However, that one can easily (re)use a model says nothing about whether one *should* (re)use it. A model of a system is said to be *valid* if one can use it instead of the system to achieve one's objectives (Law and Kelton 2000); the process of establishing validity is known as *model validation*.

The validation needs of engineering design differ from those of other M&S problems. Behavioral model reuse is commonplace in design and advances in technology are making model reuse easier and more widespread. However, model reuse can result in a segmentation of the knowledge required to perform validation across several tasks. Validating a model for a new use is difficult when these tasks are distributed in space and/or time.

We propose a conceptual framework for validating reusable behavioral models. This framework is based upon formal representations of validation-relevant knowledge and defines the abstract process for applying such knowledge. This framework is a departure from prior thinking in several ways. Most significantly:

- We recognize that in addition to validating behavioral models, one must also validate the predictions generated from them and must do so using the same formalisms.
- We identify three complementary validation problems—*validity characterization*, *compatibility assessment* and *adequacy assessment*—that individually provide insight into the properties of a behavioral model or prediction and together solve the validation problem.
- We argue for formal descriptions of the limitations of behavioral models and predictions. These descriptions—called *validity descriptions*—provide a user with assurances about a model's or prediction's accuracy over a specific set of uses and forms an interface between creators and users.

We note that this framework is not a complete solution to the validation problem but serves as a conceptual roadmap to understanding and solving the problem—it represents a starting point for future investigation and dialogue between the design and M&S communities.

The organization of this paper is as follows. Section 2 is a discussion of the challenges of validating reusable behavioral models. Section 3 contains a description of the framework for validating reusable behavioral models. Section 4 contains an example that illustrates the framework on a simple engineering example. Section 5 is a discussion of the implications of the proposed framework and areas for future work.

2 CHALLENGES OF VALIDATING RESUABLE BEHAVIORAL MODELS

2.1 Model Validation

Model validation commonly is defined as the "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" (Schlesinger, Crosbie et al. 1979). Typically, validation is considered to be a process that runs concurrent to model development (Sargent 2001). Many works present process flows for the activities and interactions of subject-matter experts (SMEs) who work throughout a M&S lifecycle to ensure model validity (e.g., (Balci 1998; Law and Kelton 2000; Law and McComas 2001)). The SMEs use various analysis methods, both qualitative and quantitative, to gauge a "degree of credibility" for a model (Balci 1997).

From a procedural standpoint, the challenge of model validation is to bring together various sources of validation-relevant knowledge. This includes knowledge about a model (its "domain of applicability" and "range of accuracy") and an intended application (the corresponding domain and accuracy needs). Much of this knowledge is generated during the model development process, which is why validation typically is integrated with development.

Efficiency is a major concern in engineering design. Pressures to minimize design times while keeping costs low require that one perform M&S—and therefore, model validation—quickly and inexpensively. This is a major motivation for model reuse, which saves the time and expense of model development. However, long validation times can undermine the benefits of model reuse. Intuitively, one might suspect that having validated a model for one use will permit faster validation for a different, but similar use. Although some have speculated that this may be the case (e.g., (Balci, Nance et al. 2002)), the literature provides little guidance on how to accomplish it.

2.2 Behavioral Model Reuse in Engineering Design

In cases of model reuse, it is not possible to perform validation and model development concurrently. This is because one cannot know about all possible applications of a model at the time it is developed. Moreover, some validation-relevant knowledge is missing and model validation cannot proceed without it. Conversely, it can be difficult to acquire validation-relevant knowledge about a model at the time that its application is known when the model user is not its creator.

Just like the validation of any other model, the validation of reusable models requires one to bring together various sources of relevant knowledge. However, the challenge with model reuse is that this knowledge can be far removed from where it is needed. To better understand this, it is useful to consider a couple of potential model reuse scenarios.

In engineering design, a prototypical model reuse scenario involves decision-making and design-analysis tasks. Decision making involves identifying preferable design alternatives based upon predictions generated during design analyses. In general, one decision-making task involves predictions from multiple analysis tasks. Design analysis involves the (re)use of behavioral models to generate the desired predictions. In this situation validation-relevant knowledge is distributed among the various tasks: model-specific knowledge is localized in the analysis tasks and application-specific knowledge is localized in the decision making task. The challenge is to integrate these sources of knowledge in a timely and effective manner. This can be difficult in general. Research on design repositories (Szykman, Sriram et al. 1998; Szykman, Sriram et al. 2000) and behavioral knowledge repositories (Mocko, Malak Jr. et al. 2004) promise to make behavioral models and the predictions they yield accessible indefinitely. This means that designers may have to validate the use of models without consulting their creators.

In another scenario, behavioral models can be shared across corporate or organizational boundaries. This mode of reuse could be important in the defense industry where government agencies often perform situational analyses using models of their assets and personnel (e.g., war gaming, logistical analyses, etc.). Rather than modeling an asset (e.g., a tank, ship, building, etc.) themselves, the government agency can require a behavioral model as a deliverable from the contractor who designs and constructs the asset. This would be an efficient reuse of resources, since often the government requests a model to be developed as part of the design process. However, it raises validation issues with respect to applying the model. In this situation, a model is being reused by an entirely different organization than the one who created it. This results in a separation of validation-relevant knowledge on a very large scale.

An efficient approach is required to transfer validation-relevant knowledge to where it is needed. Although systematic documentation may satisfy this requirement in some cases, it is not a good approach in general. A major drawback is that documentation is informal and therefore can be ambiguous. This is particularly problematic in multidisciplinary settings where one designer may not understand the assumptions implicit in the work of another. Incorrect interpretations can lead to mistakes during model validation. Another drawback to documentation is that it does not scale well. As the complexity of models and model applications grow, so does the size and scope of their corresponding documentation. This poses a practical challenge for those who must read and interpret the documents, possibly resulting in long validation times and/or mistakes.

A more suitable approach is to use formal representations to support the transfer and use of validation-relevant knowledge. Formal representations are unambiguous and computer-interpretable. This would admit the use of automation to manage the complexity of large models and model applications. In the next section, we describe a framework for validating reusable behavioral models that is based on formal representations of validation-relevant knowledge. This framework defines the abstract knowledge representation as well as an abstract process for applying this knowledge to validate reusable behavioral models.

3 VALIDATING REUSABLE BEHAVIORAL MODELS

3.1 A Framework for Behavioral Model Validation

We begin this section with a definition: A **validity description** is a formal statement about an *upper bound on the inaccuracy* of a model or prediction over a particular set of conditions, or *context*. The notions of context and inaccuracy correspond to the ideas from Section 2 of "domain applicability" and "range of accuracy," respectively. Thus, a validity description is a representation of the validation-relevant knowledge about a model or a prediction. Specifically, it represents assurances about the properties of a model or prediction under specific conditions. Context and inaccuracy are described in Sections 3.2 and 3.3, respectively.

We extend the notion of validation-relevant knowledge to predictions because they play a key role in determining the validity of models. A prediction from one model can serve as an input to another. As such, predictions define the context for the application of a model. This is discussed in greater depth in Section 3.2.

Given the concept of a validity description, validation can be decomposed into a three-step process:

- **Validity Characterization**. The process of developing a validity description.
- **Compatibility Assessment**. The process of determining whether the context of the intended use of a behavioral model or prediction is consistent with the context of its validity description.
- **Adequacy Assessment**. The process of determining whether the accuracy of a behavioral model or prediction is adequate for the user's objectives.

This three-step process is appropriate for reusable behavioral models because it allows for the efficient creation, transfer and use of validation-relevant knowledge. Validity characterization is performed during model development where knowledge about a model is produced. It results in a formal description of the model that is relevant in subsequent validation steps. Because this description is formal, it can be interpreted unambiguously by both humans and computers. The assessment steps define how this knowledge is used by subsequent M&S tasks.

The following sections discuss the notions of context and inaccuracy in greater detail. The semantics of these concepts are defined. Additionally, well-defined conditions for when two contexts are consistent with one another are presented and the relationship between context and inaccuracy is discussed.

3.2 Context

The term *context* refers to the limited domain over which a model or prediction applies. Several researchers within the artificial intelligence community have discussed the formalization of context for knowledge-based systems (e.g., (Guha and Lenat 1992; McCarthy 1993; Akman and Surav 1997); see (Guha and McCarthy 2003) and (Akman and Surav 1996) for surveys). The general approach they take is to state assumptions about the world as propositions in a logic. Falkenhainer and Forbus take such an approach for describing behavioral model components (Falkenhainer and Forbus 1991). The basis for formalizing assumptions comes from the mechanics of mathematical modeling where analysts make simplifications such as assuming a derivative is exactly zero or that a system is completely closed. However, assumptions like these seldom are satisfied exactly. Analysts make these assumptions because the resulting models often are useful as long as the assumptions are "close enough" to reality. For example, an analyst might assume a derivative is exactly zero when developing a mathematical model with the knowledge that the inaccuracy of that model is small as long as the actual value of the derivative is small. The person most qualified to judge whether an assumption is met "closely enough" is the creator of the model. However, this person may not be the user. An assumption-based approach is insufficient for representing context to a user because the user may lack the domain expertise required to make this determination.

A set-based approach is more appropriate for representing context. Conceptually, a context defines a set of "world states" within which one has some assurance of correctness or accuracy. There may be no such assurances outside of this region. In principle, a context specifies allowable values of every variable in the "world." In practice, the concept of near-decomposability states that only a handful of variables affect a system significantly (Simon 1996); all others have so little impact on a model's predictions that they can be assumed unbounded. In the simplest of situations, a context set is a hypercube created by bounds on the problem variables. In more complex cases, a context is a region of space defined by functional relationships among the variables and it may include constraints on variables not present in a model.

Each behavioral prediction contributing to a decision must satisfy its own contextual obligations. These obligations relate to the particular aspect of behavior—or *behavioral attribute*—a decision maker wants to be predicted. Decision makers typically require predictions about different behavioral attributes of a system, and each behavioral attribute can have a different context. For example, a decision maker might require one prediction about structural stress under steady-state conditions and another about the probability of failure under specific dynamic conditions. We refer to the context requirements for a particular behavioral attribute in a particular decision problem as a *behavioral attribute context*. A decision maker performs compatibility assessment for a prediction by comparing its context to that of the corresponding behavioral attribute. A decision maker can use a prediction only if it applies over the entire behavioral attribute context. Otherwise, there will be portions of the behavioral attribute context in which the prediction cannot be trusted. Decision makers take decisions in this circumstance at their own risk; to do so would be like making decisions about a supersonic aircraft based upon subsonic performance predictions.

In general, one can rationally execute a decision if and only if each prediction is of the same or broader context than its corresponding behavioral attribute—that is, they must subsume the behavioral attribute contexts. Figure 1 contains conceptual depictions of two possible decision making scenarios, each with a behavioral attribute context and the context of a corresponding prediction. In Figure 1(a), a rational decision is possible because the context of the prediction information subsumes the behavioral attribute context; they are *context-compatible*. However, a rational decision cannot be made in the situation depicted in Figure 1(b). Here, the prediction and the behavioral attribute are not context-compatible. All is not lost if the context requirements for a decision cannot be met at first. It is often possible to expand the context of a prediction if one is willing to trade a degree of accuracy for it (we discuss accuracy and the context-accuracy relationship in the next subsection).

A simulation experiment is comprised of a model and the inputs and parameters for the model. For a design problem, parameters specialize a behavioral model to a particular design alternative (i.e., they specify physical dimensions or other quantities that remain constant throughout the simulation) and inputs represent external stimuli. Each element of a simulation experiment is associated with a particular context and the context of a prediction made by the simulation is the intersection of these contexts. Figure 2 contains a conceptual depiction of the relationship of a

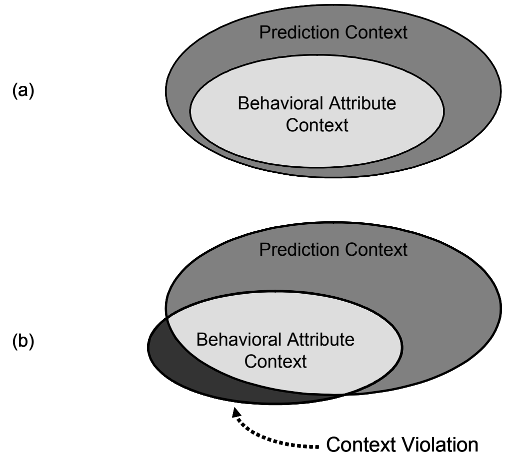

Figure 1: Conceptual Depiction of Contexts that are Compatible (a) and Not Compatible (b)

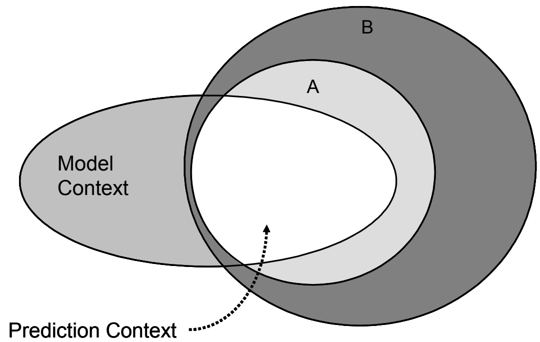

Figure 2: Conceptual Depiction of Prediction Context as it Relates to the Contexts of a Model and Two Model Inputs, A and B

prediction context to the contexts in a simulation experiment. Mathematically, one can state this relationship as

$$C_P = C_M \cap \left(\bigcap_{j=1\dots n} C_j \right), \tag{1}$$

where C_P is the context of the prediction, C_M is the model context and C_j is the context of the j[th] input or parameter to the model. This means that the context of a prediction is never more general than the least general context from which it is formed.

One can assess the compatibility of a model on two levels. First, one can ask whether a model is appropriate for a given simulation experiment. To answer this question, one compares the context of a model to those of the parameters and inputs of an experiment. One can say that the model is compatible with the other elements of the experiment if the intersection of these contexts—i.e., the context of the resulting prediction—is not the empty set.

More commonly in design, one performs a simulation experiment to predict a specific behavioral attribute for use in a decision. In this case, one assesses the compatibility of a model relative to whether the resulting prediction is context-compatible with the behavioral attribute. Given inputs and parameters for a model and a desired behavioral attribute, compatibility assessment shows that the a behavioral model is compatible if the prediction yielded by the simulation experiment is context-compatible with the behavioral attribute. This combines the concepts illustrated in Figures 1 and 2. Combining the relationship in Equation (1) with the notion of context compatibility, we can say that a simulation experiment is context-compatible with a behavioral attribute if and only if

$$C_{BA} \subseteq \left(C_M \cap \left(\bigcap_{j=1\ldots n} C_j \right) \right), \qquad (2)$$

where C_{BA} is the behavioral attribute context. With respect to validating the use of a behavioral model, the condition $C_{BA} \subseteq C_M$ is necessary for Equation (2) to hold. If a behavioral model is context compatible with a desired behavioral attribute, one can proceed to assess its adequacy.

3.3 Inaccuracy

Inaccuracy refers to the total uncertainty in a prediction or model. There are many ways in which one can characterize uncertainty. For the discussion here, we distinguish between two types: *aleatory uncertainty* and *epistemic uncertainty*. Aleatory uncertainty is a potential deviation from reality in a prediction or model due to natural random behavior (Parry 1996) and is also known as variability, stochastic uncertainty, objective uncertainty (Ferson and Ginzburg 1996) and irreducible uncertainty. Examples of phenomena that involve or exhibit aleatory uncertainty include machining error, annealing, errors in communications systems, many measurement errors and radioactive decay.

Epistemic uncertainty is a potential deviation from reality in a prediction or model due to a lack of knowledge or information (Parry 1996) and sometimes is called imprecision (Antonsson and Otto 1995), reducible uncertainty or subjective uncertainty (Ferson and Ginzburg 1996). Epistemic uncertainty often results from ignorance or modeling decisions, such as selecting one model over another or choosing to make particular approximations and simplifications. Oberkapmf and coauthors distinguish between epistemic uncertainty and error, which they describe as resulting from deliberate simplifications or inadvertent mistakes in modeling (Oberkampf, DeLand et al. 2002). However, error is a type of knowledge deficiency and is therefore better viewed as a subclass of epistemic uncertainty.

One can represent aleatory uncertainty using classical probability theory, but it is not generally correct to represent epistemic uncertainty in this way. This can be because one does not have enough information under epistemic uncertainty to describe the relative likelihoods of events or because a probabilistic interpretation is altogether invalid. The latter case can result from modeling assumptions. For example, an analyst might simplify a model by ignoring an energy loss (e.g., friction, thermal losses, etc.). This can result in a systematic deviation from reality where the precise deviation is uncertain. It would be wrong to represent this uncertainty with classical probability theory because the deviation from reality is systemic, not stochastic. Formal approaches for representing and making decisions under epistemic or combined epistemic-aleatory uncertainty are still a topic of research. Researchers have explored several alternatives to classical probability theory, including possibility theory (Dubois 1988), fuzzy set theory (Zadeh 1965), Dempster-Shafer theory (Yager, Kacprzyk et al. 1994), probability-bounds analysis (Ferson 2000) and interval analysis (Ferson and Ginzburg 1996). How one performs adequacy assessment depends on the chosen inaccuracy representation.

Designers contend with both aleatory and epistemic uncertainty. Aleatory uncertainty is particularly important when considering the impacts of manufacturing and environmental variations (e.g., random deviations in part sizes or loading conditions). Because they are approximations of reality, all models have epistemic uncertainty and, by virtue of being computed from a model, all predictions have epistemic uncertainty as well. Another source of epistemic uncertainty is the incompleteness of a design specification (i.e., there is a lack of knowledge about what the final design will be). This source of uncertainty manifests itself in behavioral models since the model cannot "know" more about a design than is present in its specification. Also note that there can be epistemic uncertainty about an aleatory uncertainty. For example, one may not know the precise mean of a probability distribution.

The purpose of a validity description is to provide a user with assurances about the inaccuracy of a behavioral model or prediction over a well-defined set of situations. In general, a user has no way of knowing if the inaccuracy of a model or prediction is actually larger than what is reported. Because of this, creators must ensure that their characterizations of inaccuracy do not understate the actual inaccuracy. This suggests that a conservative approach to inaccuracy characterization is best. Overly conservative characterizations are undesirable because they artificially limit the usefulness of an item. However, it is better to err on the side of conservativeness than to take a chance that understating the inaccuracy will not matter. If one models inaccuracy with a set-based approach (e.g., (Ben-Haim 2001)), then the objective of inaccuracy characterization is to find the least upper bound, or *supremum*, of the inaccuracy set. Any upper bound on the set is acceptable, though.

For most models and predictions, its inaccuracy depends upon the context in question. For instance, a linear deflection model for a beam may be very accurate when the displacement is less than some upper bound, but inaccurate otherwise. In general, inaccuracy never decreases—and likely increases—as the context expands. This results in an important tradeoff for analysts as they develop behavioral models: too narrow of a context can yield a very accurate model that is seldom useful (i.e., it frequently fails compatibility assessment), while too broad of a context can result in a model too inaccurate to be useful (i.e., it frequently fails adequacy assessment).

4 EXAMPLE PROBLEM

In the following example, we demonstrate the semantics of validity descriptions and how one can propagate a validity description through a model. We use a particular formulation of Newton's second law of motion as an example. Using this example, we are also able to highlight the relationship between context and inaccuracy. Readers should note that the approach that we adopt for developing the validity characterization in this example is only one of many possibilities. The purpose of this example is to demonstrate the semantics of the concepts described in Section 3 rather than to promote a particular method.

This example is based on a common formulation of Newton's second law of motion,

$$\mathbf{F}(t) = m\mathbf{a}(t), \tag{3}$$

where $\mathbf{F}(t)$ is the net force vector on a particle as a function of time, m is the particle mass and $\mathbf{a}(t)$ is the particle acceleration vector as a function of time. This is the model of interest for this example. Our first objective is to characterize it—that is, to define a context and find the inaccuracy over that context.

Equation (3) is a simplification of the more general relationship

$$\mathbf{F}(t) = \frac{d}{dt}(m(t)\mathbf{v}(t)), \tag{4}$$

where $m(t)$ is the time-varying particle mass and $\mathbf{v}(t)$ is the time-varying velocity vector. Expanding the derivative, we have $\mathbf{F}(t) = m(t)\dot{\mathbf{v}}(t) + \dot{m}(t)\mathbf{v}(t)$, where $\dot{m}$ and $\dot{\mathbf{v}}$ are the time-derivatives of mass and velocity, respectively. The difference between the two model formulations is the term $\dot{m}(t)\mathbf{v}(t)$. Thus, the inaccuracy in the model in Equation (3) will depend on the velocity and the time-derivative of the mass. Assume the context of interest is defined by bounds on these variables, with all other variables assumed to be free. Moreover, the context is the set of world states for which the following inequalities hold:

$$|\dot{m}(t)| \le \Delta_{\dot{m}}, \quad \|\mathbf{v}(t)\|_2 \le \Delta_{\mathbf{v}}.$$

With a context is defined, it now is possible to characterize inaccuracy. For this example, we adopt a set-based approach to representing inaccuracy with an additive inaccuracy model, $\mathbf{F}(t) = m\mathbf{a}(t) + \mathbf{e}$, where $\mathbf{e}$ is a vector inaccuracy term. One can develop more complex inaccuracy models; see (Ben-Haim 2001) for examples. In this case, a bound on the magnitude of the inaccuracy term over the given context defines the inaccuracy of the behavioral model. The ideal bound is the supremum. Assuming that the model in Equation (4) is perfectly accurate, one can compute the inaccuracy term as the difference between the two models. Thus, we have $\mathbf{e} = \dot{m}(t)\mathbf{v}(t)$ which has a magnitude of $\varepsilon = \|\mathbf{e}\|_2 = \dot{m}\|\mathbf{v}\|_2$. The supremum is maximum of this magnitude over the context set, C, or

$$\begin{aligned} \varepsilon_{\text{sup}} &= \max_{C}(\|\mathbf{e}\|_2 = \dot{m}\|\mathbf{v}\|_2) \\ \varepsilon_{\text{sup}} &= \Delta_{\dot{m}}\Delta_v \end{aligned}.$$

Combining the model in Equation (3) with the context and inaccuracy, one gets

$$\left\{\mathbf{F}(t), m, \dot{\mathbf{v}}(t), \mathbf{e} : \mathbf{F}(t) = m\dot{\mathbf{v}}(t) + \mathbf{e}, \|\mathbf{e}\|_2 \le \Delta_{\dot{m}}\Delta_v\right\},$$

assuming that $|\dot{m}(t)| \le \Delta_{\dot{m}}$ and $\|\mathbf{v}(t)\|_2 \le \Delta_{\mathbf{v}}$ due to context restrictions. This is a set of functions (due to the inaccuracy) over a set of conditions (the context) and represents the limitations of one's knowledge about the model. This characterization is unambiguous and, with it, a user can perform compatibility and adequacy assessment without consulting its creator.

Given a behavioral model with a validity characterization, our objective is to compute a prediction and its corresponding validity description. Assume we know that some force, $\mathbf{F}_{\max}$, is the maximum net force that will occur on a system of interest and that the system has an uncertain mass that is represented as an interval, $m_s \pm \rho m_s$, $0 < \rho < 1$. Also, assume that these are context-consistent with the model. The desired behavioral attribute is the instantaneous acceleration, $\dot{\mathbf{v}}$. Reformulating the model in terms of $\dot{\mathbf{v}}$ and substituting in the force and mass terms yields

$$\left\{\dot{\mathbf{v}}, \mathbf{e}, \varepsilon_m : \dot{\mathbf{v}} = \tfrac{\mathbf{F}_{\max} - \mathbf{e}}{m_s + \varepsilon_m}, \|\mathbf{e}\|_2 \le \Delta_{\dot{m}}\Delta_v, |e_m| \le \rho m_s\right\},$$

assuming that $|\dot{m}(t)| \le \Delta_{\dot{m}}$ and $\|\mathbf{v}(t)\|_2 \le \Delta_{\mathbf{v}}$. Let,

$$\begin{aligned} \mathbf{F}_{\max} &= \begin{bmatrix} 10 & 0 & 0 \end{bmatrix}^T \quad N \\ m_s &= 100 \quad kg \\ \rho &= 0.1 \\ \Delta_{\dot{m}} &= 10^{-9} \quad {}^{kg}\!/\!_{s} \\ \Delta_v &= 10^3 \quad {}^{m}\!/\!_{s} \end{aligned}$$

This reduces to the scalar case of

$$\left\{ \dot{v}_x, \varepsilon, \varepsilon_m : \dot{v}_x = \tfrac{10-\varepsilon}{100+\varepsilon_m}, |\varepsilon| \le 10^{-6}, |\varepsilon_m| \le 10 \right\}.$$

One can condense this representation into a single inaccuracy parameter:

$$\left\{ \dot{v}_x : \dot{v}_x = 0.101 + \varepsilon_{\dot{v}_x}, |\varepsilon_{\dot{v}_x}| \le 0.011 \right\}.$$

The context for this prediction is the intersection of the contexts of the inputs and model. In this case, they are all identical, so the prediction has the same context.

There are three particularly interesting features about this example. First, the context for the model is defined by terms not actually in the model, namely $\dot{m}$ and $\mathbf{v}$. This happens when an analyst makes simplifications during modeling. When terms are eliminated from a model because they are assumed "insignificant," they must be bounded in its context. Specifying bounds on these terms in a context defines the semantics of the assumption (i.e., what it means to be "insignificant"). This explicit definition is necessary for information exchange and knowledge reuse to proceed effectively.

The second feature illustrated in this example is the relationship between context and inaccuracy. In this case, expanding the context (i.e., raising the bound on either or both context terms) results in an increased inaccuracy. In general, an expansion of a context cannot result in a decrease in inaccuracy, and often will result in an increase. For the single-parameter case illustrated above, this means that

$$C_1 \subset C_2 \rightarrow \varepsilon_1 \le \varepsilon_2$$

where C_1 and C_2 are contexts and ε_1 and ε_2 are the corresponding inaccuracy parameters.

The third interesting feature of this example is that it fails to capture all of the inaccuracy in the model. This is because it ignores the inaccuracy in Equation (4) and therefore the derived inaccuracy is not an upper bound on absolute inaccuracy, but on the relative inaccuracy between the models in Equations (3) and (4). To arrive at an inaccuracy bound that is more faithful to reality, one must either characterize a bound using empirical means or relative to the best available model (i.e., one that is considered "ground truth" by experts in that domain). For the present relationship, the best generally accepted model is Einstein's theory of special relativity, which implies that (Ohanian 1995)

$$\mathbf{F}(t) = \frac{d}{dt}\left(\frac{m(t)\mathbf{v}(t)}{\sqrt{1 - {v^2(t)}/{c^2}}} \right),$$

where $v(t) = \|\mathbf{v}(t)\|_2$ is the particle speed and c is the speed of light. This model is nearly identical to that of Equation (4) as long as the particle speed does not become a significant fraction of the speed of light. However, the inaccuracy can grow to be significant in other contexts.

While relativistic effects are insignificant on most engineering problems, it is still important to consider them when characterizing the model in Equation (3). This is because characterization is the only means by which the creator of a model can control its use. Failure to describe the limitations of a model adequately can result in improper use no matter how unlikely that use may seem to at the time it is created.

Ultimately, it is the responsibility of the modeler to characterize the inaccuracy conservatively without being so conservative that it becomes useless. For the current example, a modeler could restrict the context to include only velocities for which the inaccuracy due to relativistic effects is small or could develop an inaccuracy model that accounts for these effects. Complex problems will require numerical approaches for approximating the inaccuracy bound as well as approaches that admit empirical data. Also, any practical method must allow modeling experts to utilize their judgment. The identification of such methods is an open research issue.

While behavioral models are developed by domain experts, predictions are generated by performing simulation experiments. Given a simulation experiment and a behavioral model, one must ensure that the resulting prediction and its validity description are computed properly. Existing simulation tools are not capable of propagating validity descriptions directly. Research into propagation methods is an important issue to be tackled.

5 DISCUSSION

Given the conceptual nature of this paper, there remain many questions about how to implement our vision. Among these, there are three major issues: establishing a basis for trust among collaborating designers, formalisms and methods for validity description and computational issues.

A basis of trust is necessary in order for designers to accept the conclusions of others. A validity description provides guarantees about a model or prediction, but a user must take these guarantees on faith. The literature identifies two ways of establishing such a baseline of trust: accreditation and certification. Under the ISO definitions for these terms, people and organizations can be accredited, while products, processes and services can be certified (Balci 2001). Accreditation helps to identify companies and individuals that meet minimum standards on some task, such as modeling in a particular domain. Certification increases a user's confidence that a particular result—a validity description, for instance—is as specified. Alternately, certification can apply to the methods used to develop particular results. From a validation perspective, accreditation and certification can provide the basis for trust among design specialists. For instance, a user could accept a validity description on faith given that it has been certified or that its creator is accredited. The definition of certification and accreditation procedures associated with the three validation problems identified in this paper is an issue for future development.

Further research also is needed on formalisms for representing and methods for creating validity descriptions. While we present some preliminary ideas in this paper, they are insufficient for practical implementation. For context, the pressing issues are how to choose an appropriate context and how to decide whether a context captures all of the relevant variables. For accuracy, designers need representations and computational methods that can incorporate both aleatory and epistemic uncertainty. Some researchers have begun to look at this issue (e.g., (Ferson and Ginzburg 1996; Ben-Haim 2001)). These representations and methods must also be compatible with a decision theory in order to be useful.

Computational issues must also be addressed. One question is how to compute a validity description for a prediction based upon those of a model and its parameters and inputs. We believe that current methods (deterministic or based on Monte-Carlo simulation) are ineffective for the evaluation of models that include epistemic uncertainty and that fundamentally different computation methods are required. Another question is how to determine whether a validity description and a use are context-compatible. On some problems, the number of variables in a context may grow large and, in principle, the context set may become non-convex. Thus, general solutions for these problems may be computationally intensive. It will be a challenge to balance the desire for fast computation with the desire for high fidelity representation.

6 SUMMARY

Collaborations among specialists create special challenges to validation in engineering design. This paper contains a description of a conceptual framework for understanding these challenges. The framework takes a novel approach by decomposing validation into three complementary processes—validity characterization, compatibility assessment and adequacy assessment—based on a formal representation of validation-relevant knowledge. This allows such knowledge to be acquired, transferred and used efficiently and enables effective validation of reusable behavioral models.

Although the framework presented here is not a complete solution to the problem of validation in engineering design, it does serve as is a conceptual roadmap to understanding and solving the problem. Future research issues include the development of formal representations for accuracy, methods for developing inaccuracy and context representations and methods for propagating validity descriptions through simulations.

ACKNOWLEDGMENTS

The authors thank Tarun Rathnam, Steven Rekuc, Jason Aughenbaugh and the other members of the Systems Realization Laboratory for their contributions to this work. This work is supported by the G.W.W. School of Mechanical Engineering at the Georgia Institute of Technology.

REFERENCES

Akman, V. and M. Surav 1996. Steps toward Formalizing Context. *AI Magazine*, 17(3): 55-72.

Akman, V. and M. Surav 1997. The Use of Situation Theory in Context Modeling. *Computational Intelligence*, 13(3): 427-438.

Antonsson, E. K. and K. N. Otto 1995. Imprecision in Engineering Design. *ASME Journal of Mechanical Design*, 117: 25-32.

Balci, O. 1997. Principles of Simulation Model Validation, Verification and Testing. *Transactions of the Society for Computer Simulation International*, 14(1): 3-12.

Balci, O. 1998. Validation, Verification and Accreditation. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 41-48. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Balci, O. 2001. A Methodology for Certification of Modeling and Simulation Applications. *ACM Transactions on Modeling and Computer Simulation*, 11(4): 352-377.

Balci, O., R. E. Nance, J. D. Arthur and W. F. Ormsby 2002. Expanding Our Horizons in Verification, Validation and Accreditation Research and Practice. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yucesan, C.-H. Chen, J.L. Snowdon and J.M. Charnes, 653-663. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Ben-Haim, Y. 2001. *Information Gap Decision Theory*. London: Academic Press.

Dubois, D. 1988. *Possibility Theory*. New York: Plenum Press.

Falkenhainer, B. and K. D. Forbus 1991. Compositional Modeling: Finding the Right Model for the Job. *Artificial Intelligence*, 51: 95-143.

Ferson, S. 2000. Probability Bounds Analysis Solves the Problem of Incomplete Specification in Probabilistic Risk and Safety Assessments. In *Proceedings of the Ninth Conference on: Risk-Based Decision Making in Water Resources IX*, 173-188. Santa Barbara, California: American Society of Civil Engineers.

Ferson, S. and L. Ginzburg 1996. Different Methods Are Needed to Propagate Ignorance and Variability. *Reliability Engineering and System Safety*, 54(2-3): 133-144.

Guha, R. V. and D. B. Lenat 1992. Language, Representation and Contexts. *Journal of Information Processing*, 15(3): 340-349.

Guha, R. V. and J. McCarthy 2003. Varieties of Contexts. In *Modeling and Using Context, 4th International and Interdisciplinary Conference, CONTEXT 2003*, 164-177. Stanford, California: Springer.

Law, A. M. and W. D. Kelton 2000. *Simulation Modeling and Analysis*. Boston: McGraw-Hill.

Law, A. M. and M. G. McComas 2001. How to Build Valid and Credible Simulation Models. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B.A. Peters, J.A. Smith, D.J. Medeiros and M.W. Rohrer, 22-29. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

McCarthy, J. 1993. Notes on Formalizing Context. In *IJCAI-93: Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 555-560, Chambery, France: Morgan Kaufmann Publishers.

Mocko, G., R. J. Malak Jr., C. J. J. Paredis and R. Peak 2004. A Knowledge Repository for Behavioral Models in Engineering Design. In *24th ASME Computers and Information in Engineering Conference*, Salt Lake City, Utah: American Society of Mechanical Engineers.

Oberkampf, W. L., S. M. DeLand, B. M. Rutherford, K. V. Diegert and K. F. Alvin 2002. Error and Uncertainty in Modeling and Simulation. *Reliability Engineering and System Safety*, 75(3): 333-357.

Ohanian, H. C. 1995. *Modern Physics*. Englewood Cliffs, New Jersey: Prentice Hall.

Pahl, G. and W. Beitz 1996. *Engineering Design: A Systematic Approach*. 2nd ed. London: Springer-Verlag.

Parry, G. W. 1996. The Characterization of Uncertainty in Probabilistic Risk Assessment of Complex Systems. *Reliability Engineering and System Safety*, 54(2-3): 119-126.

Sargent, R. G. 2001. Some Approaches and Paradigms for Verifying and Validating Simulation Models. *In Proceedings of the 2001 Winter Simulation Conference*, ed. B.A. Peters, J.A. Smith, D.J. Medeiros and M.W. Rohrer, 106-114. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Schlesinger, S., R. E. Crosbie, G. S. Innis, C. S. Lalwani, J. Loch, R. J. Sylvester, R. O. Wright, N. Kheir and D. Bartos 1979. Terminology for Model Credibility. *Simulation*, 32(3): 103-4.

Simon, H. A. 1996. *The Sciences of the Artificial*. 3rd ed. Cambridge, Massachusetts: MIT Press.

Szykman, S., R. Sriram, C. Bochenek and J. Racz 1998. The Nist Design Repository Project. In *Advances in Soft Computing - Engineering Design and Manufacturing,* ed. R. Roy, T. Furuhashi and P. K. Chawdhry, 5-19. London: Springer-Verlag.

Szykman, S., R. Sriram, C. Bochenek, J. Racz and J. Senfaute 2000. Design Repositories: Engineering Design's New Knowledge Base. *IEEE Intelligent Systems*, 15(3): 48-55

Yager, R. R., J. Kacprzyk and M. Fedrizzi, Eds. 1994. *Advances in the Dempster-Shafer Theory of Evidence*. New York: John Wiley and Sons.

Zadeh, L. A. 1965. Fuzzy sets. *Information and Control* 8: 338-353.

AUTHOR BIOGRAPHIES

RICH MALAK is a graduate student in the G.W. Woodruff School of Mechanical Engineering at Georgia Tech. His research is focused on the representation, application and reuse of information and knowledge in the context of engineering design. His email address is <gtg269k@prism.gatech.edu>.

CHRIS PAREDIS is an Assistant Professor in the G.W. Woodruff School of Mechanical Engineering at Georgia Tech. In his research, he combines aspects of information technology, simulation, and systems theory to support the design of mechatronic systems. The goal of his current research is to develop an integrated IT framework for simulation-based design. His email address is <chris.paredis@me.gatech.edu>.