Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

VALIDATION RISK IN HIERARCHICAL MODELS

John B. Young

The MITRE Corporation
925 Corporate Drive, Suite 301
Stafford, VA 22554, USA

Barry Ezell, PhD

Virginia Modeling Simulation
and Analysis Center
Old Dominion University
1030 University Boulevard
Suffolk, VA 23435, USA

ABSTRACT

Using a model as an input data source for integration into another model carries with it risks to the validity of the model composition. This paper presents research into the inherent risks of model integration. The research decomposes models into sets of semantic concepts allowing for a calculation of structural alignment. Measurable changes in a model's output due to the integration of another model provide an impact assessment. Risks to decisions arise from incompatible assumptions and constructions of models. We present a risk assessment as a tuple containing differences in models' alignment across three axes and in changes in a model's output metrics. Risks to decisions arise from incompatible assumptions and constructions of models.

1 INTRODUCTION

In the analysis of large and complex systems, a number of assumptions are made in model construction to aide in its development and instantiation. A model may use any number of sources for data input, such as expert knowledge, authoritative databases, scientific theory, or other models' outputs. This paper presents research into the conceptual alignments of models and their suitability for use in model integration. We assert that the risk of using a model for a decision purpose is affected by the incorporation of a second model into the decision space and present a means for estimating that risk.

Model integration lacks a formal definition, but is a fundamentally human activity concerned with the management and governance of models and simulations (Dolk and Kottemann 1991). We take integration to be the act of joining two (or more) models together for a larger purpose. Much like model composability, we concern ourselves with the reuse of models and maintaining their validity. Where model interoperability concerns itself with the useful sharing of information during runtime, neither model composability nor integration make runtime concurrence a requirement. Model integration can be conducted for a variety of purposes: concatenation, amplification, parameter discovery, model construction, and model merging (Levis and Abu Jbara 2014). We concern ourselves with parameter discovery wherein a model is used to provide values, assumptions, or constraints onto information required in another model.

In this paper, we will summarize germane concepts of model validity and model integration. We will present a method to decompose models into conceptual elements. We will show a method to use those conceptual elements as a mechanism to evaluate the conceptual alignment of models. We will apply this conceptual alignment into risk literature and develop an ordered set of value to calculate the risk to model validity due to model integration.

2 VALIDITY AND INTEGRATION

2.1 Validation

Validation is a well-understood requirement of successful modeling and simulation projects. There are myriad of definitions of validity in the literature (Sargent, 2013; Bair & Tolk, 2013). Many definitions have varying degrees of the phrase "accurate representation" or "from the perspective of the intended users." The key term is the relationship of a model to its intended users, and by proxy, it's intended use. Models are deliberate abstractions of a real-world system, and there must be some underlying purpose or intent to the model in order to select the components of the real-world that are necessary for a conceptual model to be developed and a simulation system. So, even models that purportedly examine the same phenomenon or systems may have slightly nuanced differences in their instantiations, sometimes inadvertently through developer or user biases, perceptions, and experiences. Models that are known to be different will certainly have differences in their assumptions of what they include, exclude, how they depict the underlying system theories, and how they handle uncertainties and unknowns.

There is a myriad of methods by which to validate models, and interested readers may look to (Balci 1994; Sargent 2013; Law 2007; Petty 2010; Jones 2015; Zeigler et al. 2018) for discussions about model validity and methods to achieve validation. Petty (2010) draws an analogy of model validation being a comparison – comparisons to other knowledge about the system(s) being modeled, empirical tests, or even to other models. (Balci, 1994) presents two principal types of validation error that have been derived from statistical testing and depend on comparison of model results to some other data set. Type I error is model user rejecting a valid model as invalid due to the results of objective tests. This error is sometimes called the model *developer's risk*, as the development would fundamentally be for naught if the model were to be rejected. Type II error is called *the model user's risk* and is a failure to reject an invalid model and accepting it as a valid. Type I error is often times correctable by further refinement or development of the model and the largest consequence of such an error is increased cost in the model development. However, Type II error can be catastrophic as it can lead a model user to make an incorrect decision. An additional form of validation error is sometimes referred to as Type III error, where one has answered the wrong question or formulated the problem incorrectly, an idea first espoused by Mitroff and Featheringham (1974). While the specification of an error type is interesting, it is beyond the scope of this paper, but it is of note that there is already a documented method of the notion that models may be wrong for their purpose and users may not always be able to know.

The formulation of a model from knowledge about the real world and a well-defined research question is conceptual modeling. Ideally, this model development activity will document assumptions, knowledge that is being represented, and its intended purpose. However, even with clear knowledge of the experimental frame, any number of interpretations can be developed into a conceptual model and further developed into an instantiated model (Tolk et al. 2013). Both developers and users of models come with their own unique experiences and biases and apply them to their understanding of the model (Ezell and Crowther 2007). Given the myriad of interpretations and individuals' perceptions, it becomes a significant challenge to validate a model by comparing its results to the results of another model without significant insights into the conceptual models.

2.2 Integration

We take model integration as the application of a model for usage in conjunction with another model without the need for simultaneous runtime or data exchanges. Model integration is by its nature a multi-modeling effort requiring insights from two or more models. Appreciating the contexts, semantic meanings, and conceptual elements are critical for determining the appropriateness of models for integration. Model integration is not arbitrary, the models in question are presumed to offer insights to the same domain of interest, and share some representation of the real-world system under study.

Integration occurs for a variety of reasons; (Levis and Jbara 2014) list five major reasons why models are integrated with one another. The first such reason is *concatenation* wherein models share representations of entities or phenomena and can get instances from one another. This case is far more akin to traditional interoperability issues given the ability to share and use data from two or more sources. The second case is *amplification*. This is a case where a second model offers additional information or representation to the first model. That information can be in the form of greater detail on existing phenomena or in the form of additional considerations that were not explicitly part of the initial model. The third case is *parameter discovery* wherein a second model is used to bound or estimate parameters for inputs into the initial model. This is a case where a model is used as a data source for another model. The fourth case is *model construction* where a model is used to develop a second model. In this case, the initial model may yield insights into the dynamics or cause and effect relationships required in a new model. The last case is *model merging* where one model's algorithmic changes are applied to another model's structure. All five of these cases are extremely useful in understanding the potential space of model integration, and each require care and domain knowledge from model users and analysts.

There are reasons to deliberately separate models and their respective representations of systems under study. Oftentimes analytic studies requiring the models in the first place have different scopes and intentions. The purpose of a model, its applicability to analysis, its efficiency of searching parameters and solutions, and the cost of the model – in terms of both time and money – are all reasons why one may want to divide system representations into separate models (Gallagher et. al 2014). Increasing complexity of the system(s) under study may be an additional reason for the establishment of a separate model (North, 2014). The ability to integrate a finely-detailed model into an aggregated model is a cross-resolution activity (Davis, 1995). However, it not always meaningful to offer additional information into a broader context, and may generate confusion in analyzing results, or may generate artificial dependencies that do not map to a real phenomenon in the real-world system. Users must be cognizant of the contexts and assumptions that drive any and all models that are being used in their studies and decisions.

3 CONCEPTUAL ELEMENTS OF MODELS

To be explicit and clear about the construction of a model and what it deliberately abstracts from the real-world system, and how it makes those representations, we use the Objects-Processes-Relationships method developed by Turnitsa (2012) in his doctoral dissertation. To summarize here, models can be distilled into collections of three sets: Objects which are the artifacts of the system and hold value(s) describing their individual states. Processes are the dynamic elements of models that mark a change in the state of one or more Objects and represent the causality we wish to capture in our models. Relationships are the linkages between two other conceptual elements of the model.

Each of these sets are in turn defined by the concepts which they include. Those concepts are in turn divided into their defining elements. In the case of Objects, they are defined by Attributes which hold values that define the state of the Object. In the case of Processes, they are defined by Characteristics which define the state of the Process. In the case of Relationships, they are defined by Rules that define the applicability and nature of the Relationship. We can define these sets of concepts mathematically which will aide in the calculations of model alignment later.

Let:

M_A indicate Model *A*

M_B indicate Model *B*

Further, let:

$M_{A,O}$ indicates the set of Objects in Model *A*

$M_{B,O}$ indicates the set of Objects in Model *B*

$M_{A,O(n)}$ indicate Object *n* in Model *A*

$M_{B,O(l)}$ indicate Object *l* in Model *B*

$M_{A,O(n),A}$ indicates the set of Attributes in Object *n* in Model *A*

$M_{B,O(n),A}$ indicates the set of Attributes in Object *n* in Model *B*

$M_{A,O(n), A(m)}$ indicate attribute *m* in Object *n* in Model *A*
$M_{B,O(l), A(k)}$ indicate attribute *k* in Object *l* in Model *B*
$M_{A,A}$ indicates the set of all Attributes across all Objects in Model *A*
$M_{B,A}$ indicates the set of all Attributes across all Objects in Model *B*

These statements regarding Objects as sets of information in models have corollary statements about Processes and Relationships. Each of these three sets of major conceptual elements is a dimension of a model's structure that can be evaluated against other models to arrive at model alignment metrics between the models.

4 MISALIGNMENT

We calculate three alignment values between two models, one representing each conceptual set of Objects, Processes, and Relationships. There are seven scenarios in which each of these alignment values can be calculated. On any of the dimensions, there can be a misalignment due to differences in scope, resolution, and structure or combinations of these. We introduce each of these in turn:

Scope refers to the quantity of concepts that are included in each model. It can be thought of as the "breadth" of the model or of individual concepts, and is a count of the concepts that are included in the model either by design or by assumption. As depicted in the Figure 1 below, one model may contain a set of concepts while a second model has a different set of concepts. The two models may have significant overlap or very little overlap. At least one modeling system contains a major concept not found in the other system.

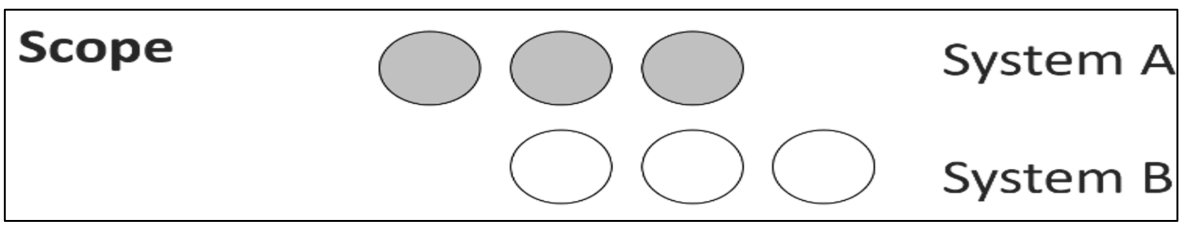

Figure 1: Model concepts misaligned by scope (Tolk, 2012).

Resolution refers to the level of precision that is incorporated into the model and its concepts to describe each concept. Where one model may have a succinct description for its own purposes, a second model may have a more detailed description of the same concept. The detail used to describe the components may be by explicit design, or may be implicit assumptions in the model. Figure 2 below depicts System A as having 3 major concepts where System B has 4 concepts in place of each concept in A, for a total of 12 concepts. The ratio of concepts in B to concepts in A need not be fixed, nor need be consistent from one concept to another.

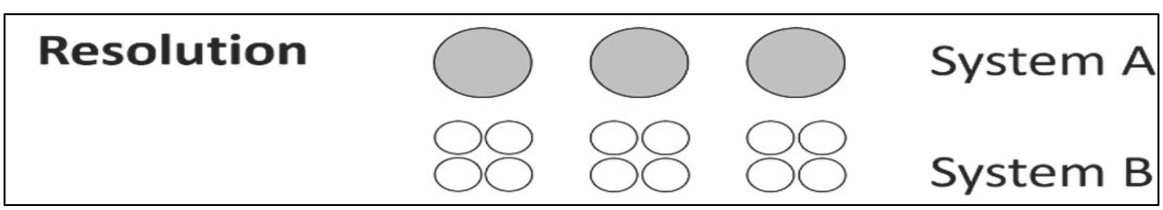

Figure 2: Model concepts misaligned by resolution (Tolk, 2012).

Structure refers to the grouping of one or more concepts in describing a larger concept. These groupings of subcomponents may not mirror one another across multiple models. To complicate matters, sub-concepts may be included in the grouping of another major concept in another model. In the Figure 3 below, System A includes two entities, each with two descriptive components. Likewise, System B has two entities, each with two differing descriptive components, though some of those sub components have been swapped between major conceptual entities.

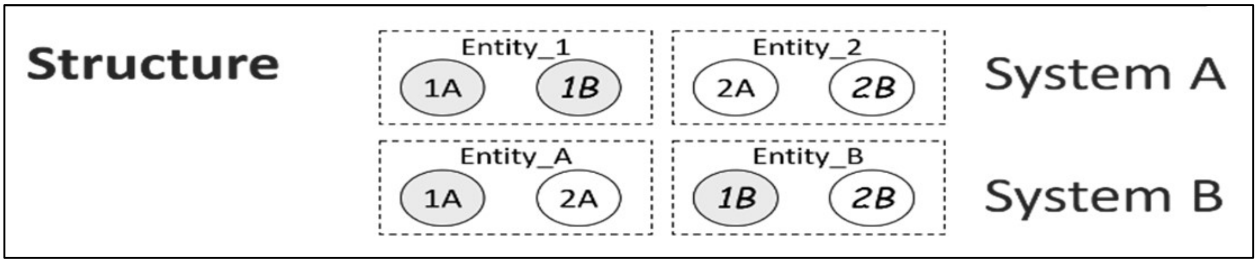

Figure 3: Model concepts misaligned by structure (Tolk, 2012).

More complicated misalignment scenarios arise with the combination of these misalignments. A misalignment of scope and resolution is depicted in the figure below. System A may have any number of concepts describing its scope breadth – Figure 4 shows three as an example. System B may share some non-zero number of major concepts with System A but replaces some number of System A's concepts with higher levels of detail.

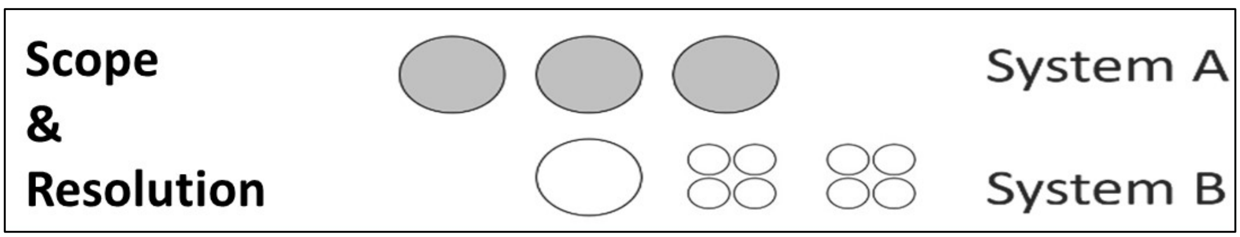

Figure 4: Misaligned scope and resolution, adapted from Tolk (2012).

Figure 5 below depicts models that are conceptually misaligned in both scope and structure. At least one of the two systems contain a major concept not included in the other system for a misalignment of scope. In the example below, System A contains "concept 1" which has no corollary in System B while System B contains "concept 4" which has no mapping in System A. In the major concepts that are shared between the models, there is a mismatch of which sub-components are included in each major concepts' definition. It is possible that one or more sub-concepts may exist in one model with no mapping to the other model, as depicted in sub-component 2C.

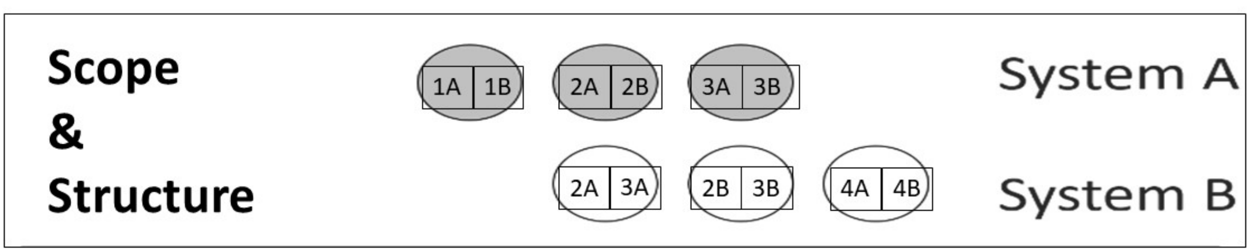

Figure 5: Misaligned scope and structure, adapted from Tolk (2012).

Figure 6 below depicts the next major Risk scenario where two models are misaligned in both resolution and structure. Both modeling systems include the same major concepts, but at least one of the two models

– in this case System B – includes greater detail in one or more of the concepts. Where major concepts are shared in each model, there may be different structures of supporting detail. In the example below, concept 2A moved from describing one major concept in System A to describing another major concept in System B. Likewise, concepts 1B and 3B describe different major concepts between the two models.

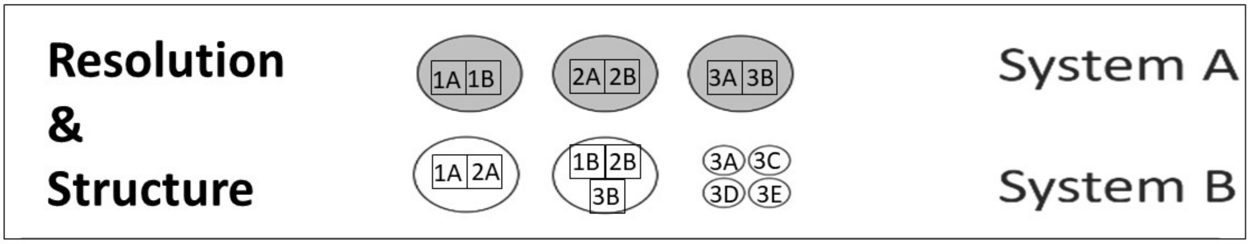

Figure 6: Misaligned resolution and structure, adapted from Tolk (2012).

The most complex of the misalignments is a misalignment across all three major definitions of misalignment – scope, resolution, and structure. Each model may have different major concepts from one another, supporting sub-concepts may be grouped differently in each model to describe different major concepts, and one model may have more detail in place of simplified assumptions in the other model. In practicality, this is the most likely scenario, where models have been developed and applied independently with different assumptions, different levels of detail, and different purposes – perhaps nuancedly different, but different nonetheless. This complex scenario is depicted in Figure 7 below.

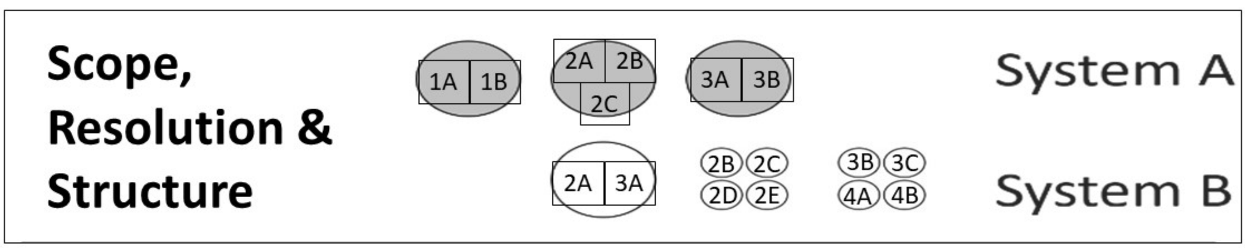

Figure 7: Misaligned scope, resolution, and structure, adapted from Tolk (2012).

5 RISK

Thus far, this paper has reviewed at a somewhat high-level concepts related to model theory and model validation. To apply a risk assessment to the usage of two or more models in a single decision space, an overview of what risk is and how it is assessed is required.

Generally, risk is some function of uncertainty and of damage (Kaplan and Garrick 1981). Oftentimes risk is colloquially seen as the product of uncertainty and damage, but this need not always be the case. The multiplication of uncertainty and damage assumes that the decision maker is risk neutral and doesn't have a particular preference in mind (Hubbard 2009). In reality, the calculation of uncertainty or the calculation of damage might be non-linear and there are particular outcomes that may be significantly worse than others. Kaplan and Garrick stress the need for some sort of a loss or damage as a key component of risk, beyond simple uncertainty. They also espouse risk as a triplet, wherein each potential outcome is enumerated as a scenario, a probability, and a consequence. These scenarios aid in the development and enumeration of outcomes that are undesirable so that they can be addressed and mitigated. Therefore, risk can then be expressed as $R = \langle S,P,C \rangle$, where R = Risk, S = Scenario, P = Probability, and C = Consequence. Then, when risk is assessed, a table is generated wherein each scenario is listed, its likelihood or uncertainty, and the damage that could be expected if this scenario were to come to pass.

The cases of conceptual misalignment serve as scenarios, identifying the type of misalignment. The tuple from (Kaplan and Garrick 1981), R= ⟨S,P,C⟩ will be extended to include five different probabilities. The first three values are the alignment values for each dimension of model alignment and stand in place of the otherwise straightforward probability. The consequence dimension of risk is extended to include two metrics: changes in model MOEs and model MOPs.

Thus, the new extended tuple for risk is R= ⟨S, D(Obj), D(Proc), D(Rel), D(MOE), D(MOP)⟩, where D(Obj) is the difference in the alignment value for the models' objects; D(Proc) is the difference in the alignment value for the models' processes; D(Rel) is the difference in the alignment value for the models' relationships; D(MOE) is the change in a model's effectiveness measures when incorporating a second model; and D(MOP) is the change in a model's performance measures when incorporating a second model.

5.1 Calculating Alignment Values

The calculation of alignment depends on the scenario of misalignment. We take a value of 1.0 to mean that models are perfectly aligned conceptually and 0.0 to mean that models have no overlapping semantic meaning. Using set theory, where each dimension of Objects, Process and Relationships are expressed as sets, and their constituent Attributes, Processes, and Rules are likewise expressed as sets.

As an example of alignment calculation, consider two models that are misaligned in scope and resolution. A Venn diagram of their shared concepts is depicted in Figure 8. In this simple example, each model contains a unique concept. But of the two objects that are similar, one offers more detail than the other model in one of the concepts. In this case, we count the number of attributes, characteristics, or rules (depending if this was a representation of objects, process, or relationships, respectively) that are within the shared space of these two models. This sum is divided by the total number of components in the first object plus the ratio of conceptual components in both models divided by the components in the first model. Mathematically, this is expressed as:

$$\frac{(M_{A,O(n),A}) \cup (M_{B,O\,(n),A})}{(M_{A,O(n),A}) + \dfrac{(M_{A,O(n),A}) \cap (M_{B,O\,(n),A})}{(M_{A,O(n),A})}}$$

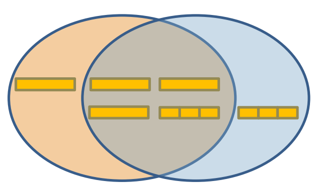

Figure 8: Two models with misaligned scope and resolution.

In this simple example, this formula would yield (1) / (2 + ((2+4)/2)) = 1 / 5 = 0.2 as an alignment value. Such set theoretic calculations are conducted for each possible permutation of shared conceptual elements.

In this illustrative example, it was assumed that the single concept with similar resolution in each model was semantically compatible with the other model. However, it is possible or even likely that a conceptual component will not be semantically the same, and its value will not strictly be 1.0. Wartik et al. (2001) present a method for assessing the alignment between seemingly similar concepts. Their method is a value table with discrete levels of semantic meaning. An analyst would assess the alignment between two concepts across models and assign it a value score from 0.0 to 1.0 based upon the alignment of the concepts. Table 1 below is adapted from their report to illustrate how these assessments are made.

In the simple illustrative example above, it becomes possible to see the impacts of an imperfect alignment in a conceptual element. If we change the attribute of the one shared concept from a 1.0 to a 0.5

due to a medium degree of alignment, the formula now yields: (0.5) / (2 + ((0.5+3.5)/2)) = 0.5 / 4 = 0.125. We now see that the addition of this assessment unsurprisingly reduced the degree of alignment between these two concepts from 0.2 to 0.125.

Assessments can be made for the entire structural space of the models, yielding a value for each of the three dimensions of alignment.

Table 1: Alignment assessment from Wartik et al., 2001.

Value	Standard Phrase	Definition
0%	No Alignment	This value is assigned in either of the following circumstances: There is no overlap between the models. One model contains an instance of an element that has no analog in the other. Lack of information in one model prevents alignment analysis.
25%	Low Degree of Alignment	There is some overlap, but it seems coincidental. Overlap might have been achieved by using some attributes in ways that its designers did not originally intend.
50%	Medium Degree of Alignment	There is a moderate amount of overlap, but still a significant disconnect between the models.
75%	High Degree of Alignment	Perfect alignment can probably be achieved by small changes to one model or the other.
100%	Perfect Alignment	There is an exact, unambiguous mapping between the models.

5.2 Calculating Impacts

The most basic definitions of risk indicate that there must be a consequence associated with a chance or probabilistic event. Having derived a method to assess models' alignments as a proxy for those chances, we turn now to defining the consequences of models integration. To find such consequences, we assume that a model has been run in isolation on its own merits to produce a set of metrics, both measures of effectiveness and measures of performance. After such a model run has been conducted, the model is rerun with the integration of a second model that will have some effect – great or small – on the output metrics of the model.

Changes in both MOEs and MOPs could range from minor to significant. The introduction of a new conceptual components from an additional model may augment, change, or contradict the metrics of a single model on its own. Cases where metrics change significantly or new metrics contradict previous metrics are the scenarios of highest consequence to the overarching purposes of the models and resulting decisions. Developing a hierarchy of preferences for MOEs and MOPs as consequences is relatively straightforward. The principle of maximum information entropy will be applied here to determine weightings for MOEs and MOPs in the Consequence component of risk. When the only piece of information is a general preference order of categories, we will equally divide the consequence space from zero to one and take the centroid value of each subspace. To develop a hierarchy table of MOPs and MOEs, we need only list a preference order of categories. Characteristics of these categories are the significance of changes in MOP values upon the integration of an input model – minor, moderate, or significant, the introduction of new attributes as part of the MOPs, and if new attributes exist whether they contradict the original model's MOPs or not yields 18 categories to measure consequences of model integration (Table 2). Similarly, a value hierarchy for MOEs can be constructed with the same categories and definitions of categories (Table 3).

These tables yield lookup values when a domain expert has seen that changes in model outputs are minor or significant, whether new information has been added to the outputs, and if so, does that new information conflict with existing information.

Table 2. Value hierarchy for MOPs.

Preference Order	Change in MOPs' Values	New Attributes in the MOPs	Conflicting Attributes	Upper Bound of Level	Centroid Weighting
1	Minor	No	NA	0.111	0.056
2	Minor	Yes	No	0.222	0.167
3	Minor	Yes	Yes	0.333	0.278
4	Moderate	No	NA	0.444	0.389
5	Moderate	Yes	No	0.556	0.500
6	Moderate	Yes	Yes	0.667	0.611
7	Significant	No	NA	0.778	0.722
8	Significant	Yes	No	0.889	0.833
9	Significant	Yes	Yes	1.000	0.944

Table 3. Value hierarchy for MOEs.

Preference Order	Change in MOEs' Values	New Attributes in the MOEs	Conflicting Attributes	Upper Bound of Level	Centroid Weighting
1	Minor	No	NA	0.111	0.056
2	Minor	Yes	No	0.222	0.167
3	Minor	Yes	Yes	0.333	0.278
4	Moderate	No	NA	0.444	0.389
5	Moderate	Yes	No	0.556	0.500
6	Moderate	Yes	Yes	0.667	0.611
7	Significant	No	NA	0.778	0.722
8	Significant	Yes	No	0.889	0.833
9	Significant	Yes	Yes	1.000	0.944

5.3 Overall Risk Calculation

From the previous subsections on the dimensions of Risk, there are a number of model misalignments that can generate risk in model integration. The risk tuple $R = \langle S, D(Obj), D(Proc), D(Rel), D(MOE), D(MOP)\rangle$ can help identify the risk profile of permutations among the alignment of conceptual elements and the potential changes in metrics from the model. In plain words, the risk tuple reads that integration risk is a function of the alignment scenario, the differences of each of three conceptual dimensions of the model, and the impact the integration of the models has on the outputs of the modeling process.

The calculation of the misalignment between two models follows the general form of:

P(misalignment) = p(misalignment of Objects) ∪ p(misalignment of Processes) ∪ p(misalignment of Relationships)

The total misalignment of the two models is $D(ModelMisalignment) = D(Obj) + D(Proc) + D(Rel) - (D(Obj) \times D(Proc)) - (D(Obj) \times D(Rel)) - (D(Proc) \times D(Rel)) + (D(Obj) \times D(Proc) \times D(Rel))$, where $D(Obj) = 1 - A(Obj)$, representing the misalignment between models' Objects; $D(Proc) = 1 - A(Proc)$, representing the misalignment between models' Processes; and $D(Rel) = 1 - A(Rel)$, representing the misalignment between models' Relationships.

Changes in both MOEs and MOPs are likewise combined using probability statements. In a simple case of two values, the numbers can simply be averaged. In more complex situations with multiple MOEs or MOPs, the combination of MOEs and MOPs follow the general form combing metrics: P(metrics) = p(change in MOE_1) ∪ p(change in MOE_2) ∪ … ∪ p(change in MOEn) ∪ p(change in MOP_1) ∪ p(change in MOP_2) ∪ … ∪ p(change in MOE_n).

With methods to calculate probabilities of misalignment and consequences available, overall risk can be calculated. With the changes of model metrics – both MOEs and MOPs – a result of the inclusion of an additional feeder model, then the Risk due to Model Integration is defined as Model Results will adversely affect the decision because of Model Integration and that Model Results are worsened because of Model Integration and that Model Results adversely affect the decision. This is mathematically defined as:

$$\text{Integration Risk} = \text{p(misalignment) x p(metrics) x } [1 - \text{p(misalignment)} + \text{p(misalignment) x p(metrics)}]$$

Values for the misalignment, the metrics, and therefore the overall risk will range between 0 and 1. Higher values of misalignment indicate that the models have relatively poor alignment in their conceptual components. Higher values in the metrics mean that there are significant changes to the model's outputs. Unsurprisingly, there is higher risk to the decision from model integration when alignment is poor and when metrics change significantly. Likewise, there is lower risk when the models are well-aligned and the changes to metrics are small. However, the value of this analysis is identifying risk values for moderate changes in either alignments or in metrics.

Figure 9 presents the risk surface response to changing combinations of models' structural alignments and changes to model outputs. The X axis represents changing values of the aggregate of misalignments, scaled from 0 to 1 where 0 represents a perfect alignment between the two models and 1 means complete misalignment. The Y axis represents changes to models' outputs in both MOEs and MOPs, ranging from 0 to 1 where 0 means no change and 1 means significant change. The Z axis represents the calculated integration risk where 0 represents no risk and 1 represents a high risk to the quality of the decision and model credibility. The surface area of this curve is larger in regions of lower risk, and smaller in regions of higher risk. This indicates that the risk of model integration may in fact be skewed towards smaller risks.

The image depicts break points along the surface of the curve at 20% intervals of integration risk. The green region is the lowest risk portion of the curve and accounts for 63.38% of the surface area. The blue region is the region where integration risk ranges from 20% to 40% and accounts for 21.02% of the surface area. The yellow region is the portion of the curve where risk is between 40% and 60%, representing 9.86% of the surface area. The light red is the region where risk is between 60% and 80%, representing 4.45% of the surface area. The upper most, dark red, region depicts the portion of the curve where integration risk exceeds 80%, and represents 1.29% of the total surface area.

6 CONCLUSIONS

We have developed this risk assessment methodology to aide modeling and simulation analysts and users in their metamodeling activities. Models that are ostensibly in the same domain are not always as semantically compatible as they may seem on the surface, and scrutinizing the structure and impacts of model integration is an important activity in understanding and explaining the behavior of models. By combining the ideas of conceptual and structural alignments with risk values, using established approaches and formulae, we are providing a repeatable, unbiased, and well document method to support the decision makers when model compositions are considered. The method is domain agnostic and supported by established practices for validation and verification, as presented in this conferences several times. As such, this method helps analysts in determining the appropriateness of using models to inform, integrate, or tune other models' parameters. Conceptual and structural differences can create sensitivities in models to concepts that are not included in the conceptual models of others and may not meet expectations.

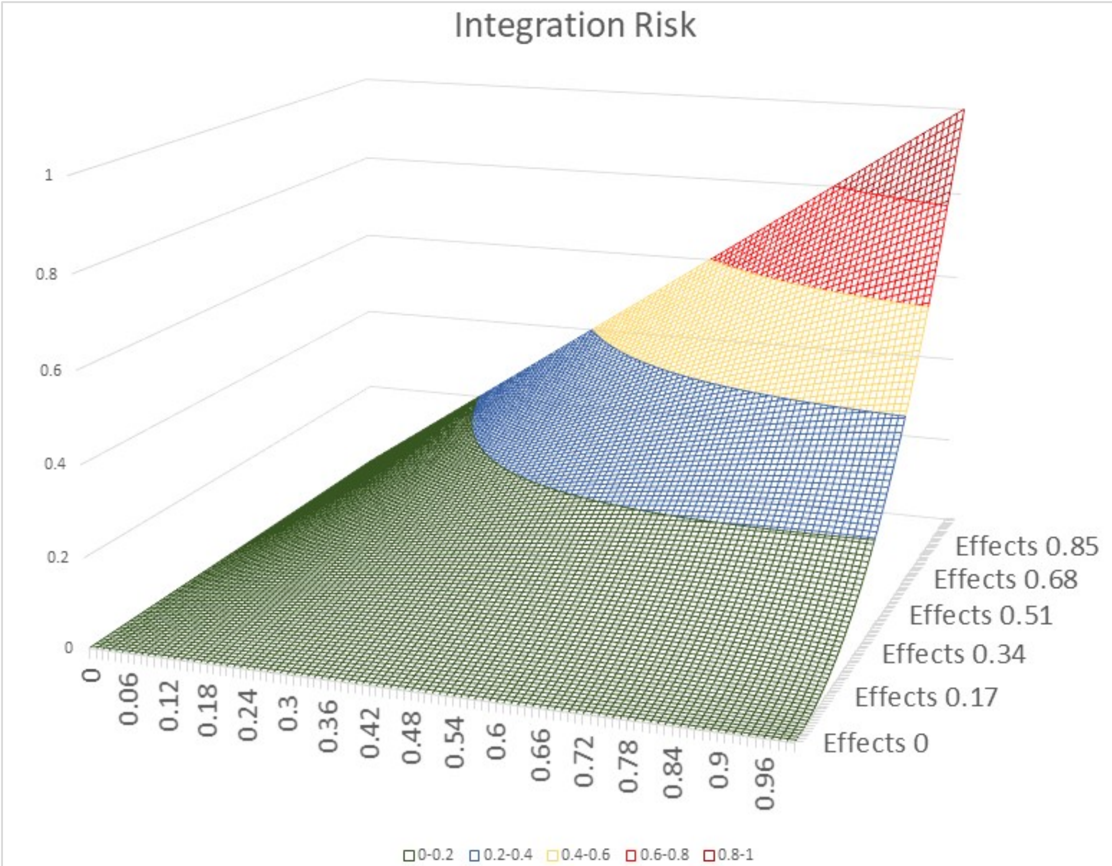

Figure 9: Model integration risk plot.

This theory is applicable to a variety of domains, particularly those where multiple models are used in concert with one another. A notable example is the military and defense establishment which uses a variety of modeling and simulation applications for analytic, wargame, and experimentation purposes. Using models to establish parameters with one another or to inform the same decision naively overlooks the contexts and structures of the models and may not account for covariance and interdependencies of modeled phenomena. This method allows simulation practioners to quantify the risks to their models' validity. Future research should be applied to this domain to examine different weighting schemas to model outputs or different methodologies to measure models' alignments.

Disclaimer

The author's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions, or viewpoints expressed by the author. This paper has been approved for Public Release; Distribution Unlimited. Public Release Case Number 19-1175.

REFERENCES

Bair, Lisa, and A. Tolk. 2013. "Towards a Unified Theory of Validation". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl. 1,245–1,256. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Balci, O. 1994. "Validation, Verification, and Testing Techniques Throughout the Life Cycle of a Simulation Study." *Annals of Operations Research* 53: 121–173.
Davis, P. K. 1995. "An Introduction to Variable-Resolution Modeling". In *Warfare Modeling*, edited by J. Bracken, M. Kress, and R. E. Rosenthal, 5–36. Danvers, MA: John Wiley & Sons.
Dolk, D. R. and J. E. Kottemann. 1993. "Model Integration and a Theory of Models". *Decision Support Systems* 9: 51–63.
Ezell, B., and K. Crowther. 2007. "Philosophical Issues and their Implications for the Systems Architect." *Foundation of Science* 12: 269–276.
Gallagher, M. A., D. J. Caswell, B. Hanlon, and J. M. Hill. 2014. "Rethinking the Hierarchy of Analytic Models and Simulations for Conflicts". *Military Operations Research* 19(4): 15–24.
Jones, M. C. 2015. "Composability". In *Modeling and Simulation Support for System of Systems Engineering Applications*, edited by L. B. Rainey and A. Tolk. John Wiley & Sons.
Kaplan, S., and B. J. Garrick. 1981. "On the Quantitative Definition of Risk". *Risk Analysis* 1(1): 11–27.
Law, A. M. 2007. *Simulation Modeling & Analysis.* Boston: McGraw Hill, Inc.
Levis, A. H. and Abu Jbara, A. 2014. "Multi-modeling, Meta-modeling and Workflow Languages". In *Theory and Application of Multi-formalism Modeling*, edited by M. Gribaudo and M. Iacono, 56–80. Hershey, PA: IGI Global.
Mitroff, I. I. and T. R. Featheringham. 1974. "On Systemic Problem Solving and the Error of the Third Kind". *Systems Research and Behavioral Science* 19(6): 383–393.
North, M. J. 2014. "A Time and Space Complexity Analysis of Model Integration". In *Proceedings of the 2014 Winter Simulation Conference*. 1,644-1,651. edited by A. Tolk, S.Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
Petty, M. D. 2010. "Verification, Validation, and Accreditation". In *Modeling and Simulation Fundamentals: Theoretical Underpinnings and Practical Domains*, by C. M. Banks and J. A. Sokolowski, 325–372. John Wiley & Sons
Sargent, R. G. 2013. "Verification and Validation of Simulation Models". *Journal of Simulation* 7(1): 12–24.
Tolk, A., S. Y. Diallo, J. J. Padilla, and H. Herencia-Zapana. 2013. "Reference Modelling in Support of M&S—Foundations and Applications". *Journal of Simulation* 7(2): 69–82.
Turnitsa, C. D. 2012. *Exploring the Components of Dynamic Modeling Techniques.* PhD Dissertation, Modeling and Simulation, Old Dominion University, Suffolk, VA.
Wartik, S., B. Haugh, F. Loaiza, and M. Hieb. 2001. *Alignment of Army Integrated Core Data Model*. Alexandria, VA: Institute for Defense Analysis.
Zeigler, B. P., A. Muzy, and E. Kofman. 2018. *Theory of Modeling and Simulation: Discrete Event & Iterative System Computational Foundations*. 3rd edition. Amsterdam: Academic Press.
Tolk, A. 2012. *Engineering Principles of Combat Modeling and Distributed Simulation*. Hoboken, NJ: John Wiley & Sons.

AUTHOR BIOGRAPHIES

JOHN B. YOUNG is a Lead Modeling and Simulation Engineer with the MITRE Corporation in the Strategic Analysis department of the Modeling, Simulation, Experimentation, and Analysis (MSEA) Technical Division. He is a PhD candidate at Old Dominion University in the Modeling, Simulation, and Visualization Engineering (MSVE) department. He holds a masters of engineering in Modeling, Simulation and Visualization Engineering from Old Dominion University and a Bachelors of Science in Industrial and Systems Engineering from Virginia Tech. He has worked in military operations research for the past 16 years both as a government civilian and as a researcher. His research interests include improvement of decision quality, combat modeling and simulation, and campaign simulation. He holds a professional engineering (PE) license and is a Certified Modeling and Simulation Professional (CMSP). He is a voting director for the Military Operations Research Society (MORS) where he served as the Modeling and Simulation Community of Practice Chair and is the program director for the 88th symposium in 2020. His e-mail address is jbyoung@mitre.org.

BARRY C. EZELL, PhD is the Deputy Executive Director at the Virginia Modeling, Analysis, and Simulation Center (VMASC) at Old Dominion University. He holds a PhD in Engineering Management from Old Dominion University. He holds a Master of Science degree in Systems Engineering from the University of Virginia. He holds a Bachelors of Science in Mechanical Engineering Technology from the University of Southern Mississippi. He is a retired US Army Lieutenant Colonel. His research interests include Quantitative Risk Analysis, Decision Analysis, Terrorism Risk Modeling, Systems Analysis, Strategic Portfolio Analysis. His email address is bezell@odu.edu.