

COMPARING VALIDATION RESULTS FOR TWO DIME/PMESII MODELS: UNDERSTANDING COVERAGE PROFILES

Dean S. Hartley III

Hartley Consulting
106 Windsong Ln
Oak Ridge, TN 37830, USA

ABSTRACT

Coverage profiles help in visualizing what is modeled and how well it is modeled. Two DIME/PMESII models with initial validation results for their conceptual models are compared. The differences in their coverage profiles are examined and related to the differences in the purposes of the models. These results are used to draw conclusions about general models.

1 INTRODUCTION

Before looking at validation results and comparing them, we need some to ensure a common background. The first subsection introduces various DIME/PMESII concepts and the second subsection reviews verification and validation.

1.1 DIME/PMESII, HSCB, and other Acronyms

The acronyms from the domain that includes DIME/PMESII models can be confusing. The following discussion is taken from the author's website (Hartley 2010).

The challenges facing the world are seen as more complex than those of the Cold War. In addition to conventional wars, there are irregular wars, terrorism (both disorganized and organized, small and large scale), and natural and human-caused disasters. Increasingly, people believe that the world is interconnected and that what happens in distant places can affect them. We have acted first and thought about it later (sometimes).

The actions have gone under several names, with a host of acronyms. They include Peace Operations (PO) of several kinds, Humanitarian Assistance (HA) and Disaster Relief (DR), Counter Insurgency (COIN), Counter Terrorism (CT), and Operations Other Than War (OOTW) or Military OOTW (MOOTW), which include the preceding and other types of operations. Many of these operations are now called Stability and Support Operations (SASO) and Stability, Support, Transition, and Reconstruction operations (SSTRO). Some of these operations are considered to be part of Irregular Warfare (IW).

The technical approaches that have been used to model these operations have generated a new acronym: DIME/PMESII. The acronym PMESII refers to the Political, Military, Economic, Social, Information, and Infrastructure variables that describe the status of a situation (state vector). There have been arguments that other categories should be included in the taxonomy; however, for our purposes, we will use PMESII to refer to all state vector variables, regardless of taxonomy details. The acronym DIME refers to the levers of power that a (nation) state has to influence the PMESII state, Diplomatic, Information, Military, and Economic. As with PMESII, we will use DIME to refer to all such interventions, regardless of taxonomy details.

The performance of operations that required more than kinetic effects (a polite term for destructive effects through physical means) drove the development of DIME/PMESII models. Similarly, the develop-

ment of DIME/PMESII models is driving a need to understand and apply social science theories. Therefore we have a new acronym and term, HSCB Modeling, which stands for Human Social Culture Behavior modeling.

When using the acronym, HSCB, the focus is on the theoretical basis of a model; whereas, DIME/PMESII (or PMESII for short) focuses on the technical details needed to implement a model. When the focus is on the operations being modeled, models may be cited as OOTW, SASO, etc., models. These definitions are not synonyms, as shown in Figure 1 below; however, it has become clear that most of the operations listed above will require DIME/PMESII modeling techniques, supported by a firm HSCB basis.

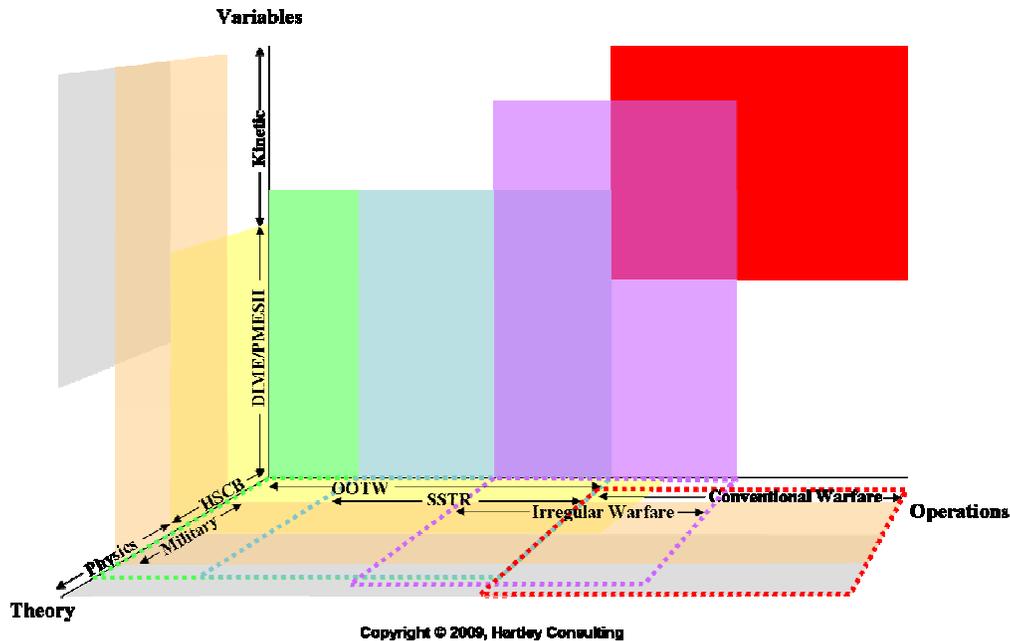


Figure 1: Connections among acronyms

1.2 Thoughts about VV&A

VV&A is about understanding and assessing models. Osman Balci and Robert Sargent have been prolific in writing about Verification and Validation (V&V) for years. Two papers represent their early works (Balci 1986, Sargent 1987). Two later works present excellent overviews of the techniques that can be used in V&V (Balci 2001, Knepell and Arangno 1993). Over time, the U.S. Department of Defense (DoD) came to recognize the importance of V&V and added a third component, Accreditation, yielding the VV&A acronym. DoD also formalized the definitions and required the practice of VV&A (DoD 5000.59, DoDI 5000.61):

Definition 1 Verification is the process of determining that a model or simulation implementation and its associated data accurately represent the developer’s conceptual description and specifications.

Definition 2 Validation is the process of determining the degree to which a model or simulation and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model.

Definition 3 Accreditation is the official certification that a model or simulation and its associated data are acceptable for use for a specific purpose.

Figure 2 illustrates the elements of model construction and V&V. The real world, conceptual model (or models), coded model, and data are common elements in all computer models. Many models, particularly DIME/PMESII models require a fifth element, a proxy for the real world or a representation of the human perception of the real world. This element is required when our understanding of the real world is

imperfect. In the PMESII domain, this need is significant because we know that our HSCB theories are poor approximations, at best, and are wrong in some cases. The author participated in a major report on V&V of DIME/PMESII models (Clemence 2007), which expands on this discussion and provides a large bibliography on the subject of VV&A.

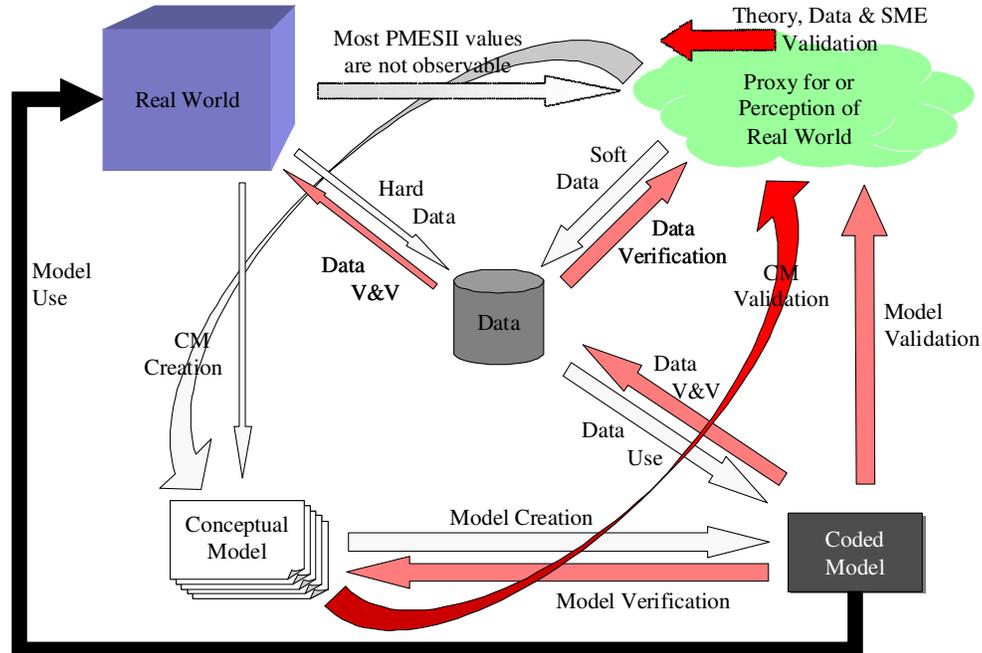


Figure 2: Relationships among model creation and V&V elements

This paper is concerned with the Conceptual Model (CM) validation and our knowledge of the validity of the theories upon which the CM is largely based, for instance: what parts of DIME/PMESII space are modeled (coverage); what theories are the bases for each part; and how good are those theories (profile level). The coverage profile helps in visualizing what is modeled and how well it is modeled. Radar charts (shown later) are useful because they show levels without implying connections among variables.

2 CM VALIDATION PROCESS

The CM validation process is illustrated by the darker red curved arrow in Figure 2 and includes the theory part of the left-pointing arrow in the upper right corner of the figure.

2.1 Two Models

Two models will be used to illustrate how coverage profiles are useful in understanding conceptual model validity: Senturion and the Interim Semi-static Stability Model (ISSM). Senturion is used to determine the outcome of political competition over some issue. Its theoretical basis is Spatial Bargaining, which is a well-tested theory. Senturion is a narrow, focused model that runs rapidly and has moderately difficult data requirements. The ISSM is used to understand and track the situation in some complex emergency (e.g., peace operation or irregular war). Its theoretical basis is subject matter expert opinion. The ISSM is a broad, comprehensive model that runs rapidly and has simple data requirements. The validation results used here are for version 8.05 of Senturion and 4.03 of ISSM Main. The Senturion validation can be characterized as independent validation. Although superficial in nature, it was thorough enough to be illustrative for this paper (Hartley 2009b). The ISSM validation, described by Hartley (2005a, 2005b, 2006a, 2006b), cannot be characterized as independent, as it was performed by the software's developer;

however, the portions used for this paper are straightforward associations with either external group expert opinion (value 2.5 from the chart below) or developer’s expert opinion (value 2.0).

2.2 Validity Scores

Table 1 displays the range of scoring values and their meanings. The two columns labeled Component and Ensemble express a key problem with model validity: models with high scoring components may not have high overall scores because the interfaces among the components do not have high validity. DIME/PMESII models have the additional problem of resting on social theories. As shown by the shaded area of the Component column, social theories do not meet the highest standards for validity and most lie at the lower end of the scale. Further, most social theories are isolated (few or no high validity connections to other theories), meaning that ensembles are unlikely to rise above the lower levels of the scale.

Table 1: Scoring validity is meaningful at both the component and ensemble levels.

Value	Component	Ensemble (Multi-model or SubCategory)
5	Expresses fully validated theory, e.g., Newtonian physics with caveats on operations near light speed or in regimes subject to quantum effects	Expresses multiple Level 5 theories with fully engineered interfaces, e.g., fly-out model of ground-to-air rocket, involving, chemical reactions of propellants, ballistics, air flow, electronics, etc.
4	Expresses well researched theory involving considerable data checking and peer review, e.g., economic theory earning Nobel prize	Expresses multiple Level 4 or Level 5 theories with well researched interfaces, e.g., economic model ensemble used by Federal Reserve in setting U.S. interest and discount rates
3	Expresses theory supported by data and published in peer-reviewed literature	Expresses multiple Level 3, 4 or 5 theories with considerable peer-reviewed interfaces, e.g., some U.S. combat models
2	Expresses theory with rational basis, accepted by some experts as plausible (SWAG)	Expresses multiple Level 2 - 5 theories with plausible interfaces (SWAG)
1	Expresses a codified theory (WAG)	Expresses multiple Level 1 - 5 theories with codified interfaces (WAG)
0	Uncodified, mental model of uncertain consistency and completeness	Uncodified processes for connecting models of uncertain consistency and completeness

Validity scoring has been discussed by other authors. For example, Balci et al (2002) describe an automated environment for documenting and scoring the results of what they call Credibility Assessments, which can be conducted under VV&A process guidance. On the other hand, Sargent (2004) does not believe in the use of scoring models for several reasons involving misplaced confidence and hidden subjectivity, preferring developer-alone, developer and user, or IV&V methods, depending on the situation. Coverage profiles are introduced here because the complexity of DIME/PMESII validation requires some aggregation of scores, yet Sargent is correct in distrusting aggregated scores.

2.3 DIME/PMESII Space

For validation purposes, the DIME/PMESII taxonomy serves as a useful organizational tool. However, some care is needed in evaluating the results.

First, it is important to realize that with our current understanding of the underlying reality, any taxonomy has an element of the arbitrary in its definition.

Definition 4 We will label as Variables the elements with the finest granularity.

Definition 5 *The Variables are collected into SubCategories.*

Definition 6 *The SubCategories are collected into Categories.*

The collection choices are derived from observing which Variables are most closely related to other Variables and, thus, there is a rationale; however, in many cases, a good argument could be made for alternate groupings.

Second, it is important to realize that combining the scores upward in the taxonomy has limited value. In an ideal taxonomy for validity, each Variable would be independent of all other Variables, so that the validity of the conceptual model with respect to one Variable would be independent of the validity with respect to any other Variable. For example, the “administration of justice is effective and fair” and the “financial system is solid” are two Variables for which one can easily envision real scenarios with any combination of values. It is hard to imagine how the validity of the submodel for one Variable could depend on the validity of the submodel of the other Variable. Thus, combining the validity values, while mathematically possible, would convey little information about the validity of the model and would obscure what is known. On the other hand, the current taxonomy is known not to be an ideal taxonomy: some Variables overlap or depend on other Variables. In the list below, there are Variables for common crime, drug crime, and organized crime and there is a Variable for [overall] crime. The rationale is that some models may have the detail to split crime into pieces, while others might lump all crime together. To the extent that the SubCategories embody valid and useful concepts (that is, the Variables in each SubCategory do refer to a single concept), it is possible that a combined validity at that level provides some useful information.

Third, the granularity of the current taxonomy is not adequate for all models. This taxonomy is based on Variables that are useful for general DIME/PMESII models that have as their highest level Measures of Merit (MoMs) at the Measure of Political Effectiveness (MoPE) level. A model that includes the details of the psychological processes of individuals might require decomposing the Variables into SubVariables or SubSubVariables. The current taxonomy would not be adequate for distinguishing validity differences among models of this type.

A complete discussion of the contents of the taxonomy is beyond the scope of this paper; however, the following description should provide enough concrete examples to allow the reader to understand the examples. The Categories are DIME [including all types of interventions], Political (Pol), Military (Mil), Economics (Econ), Social (Soc), Information (Info), Infrastructure (Infra) [the components of PMESII], Environment (Env), and Kinetics [not represented by these two models]. The SubCategories divide the Categories; for example, the Political Category is divided into Political-Gov[ernment], Political-Pol[itical], Political-ROL (Rule of Law), Political-Sec[urity], Political-z (Measures of Political Effectiveness), and Political Intraconnections. Each SubCategory encompasses a number of Variables. For example, Political-ROL includes the following variables:

- Administration of justice is effective and fair,
- Common crime is not a problem,
- Corruption in central authority is not rampant,
- Corruption in law enforcement is not rampant,
- Corruption in public office is not part of culture,
- Corruption in social services is not rampant,
- Crime is not a problem,
- Drug crime is not a problem,
- Government police force is effective against crime,
- Human rights are protected,
- Organized crime is not a problem,
- Police are distinct from the military, and
- Prison structure is adequate.

In addition, there are intervention (or DIME) Variables that serve the dual purposes of describing the fact that a type of intervention is taking place and (if present within the model in question) reflecting the changes to the situation created by that intervention. The DIME Variables that fall into the Political-ROL SubCategory (and are also part of the DIME-Political SubCategory) are the following:

- (Re)building & monitoring new police force,
- Assisting in establishing humane penal systems,
- Assisting in establishing/reforming legitimate legal system,
- Conducting constabulary operations,
- Conducting war crimes investigations, tribunals, etc.,
- Monitoring and reporting on corruption by govt officials,
- Monitoring human rights practices,
- Providing advisors to police & criminal justice organizations & supporting, establishment of operations, and
- Training police forces.

There are about 300 Variables, of which not quite half are DIME, or intervention, variables. With a few exceptions, these Variables are at a similar level of granularity and represent a fairly good coverage at that level of granularity of all of the PMESII indicators and DIME interventions for international intervention operations. The indicator Variables are derived from Hayes and Sands (1997), with a few additions, and the intervention Variables are derived from various lists of actual operations over the last 20 years (Hartley 2006b and Hartley 2009a).

2.4 Aggregation and Display

Validation values are defined at the Variable level. For the conceptual model, there are two components: coverage and value.

The value depends on the validity value of the theory (Table 1 component column) and the comparison of the granularity of the module containing the Variable and the granularity of the model as a whole. For example, the value is lowered for a Variable in a module that has only national level granularity where the module is contained in a model that should measure district level changes.

Coverage has two components: absolute and relative. Individual variables that are not covered are given a 0 value, as they are not connected to any theories. However, if the variable is not relevant to the use of a model, it is not displayed or averaged in SubCategories and Categories.

Validation values for the coded model depend on the CM values, quality of implementation, match of implementation to model use, match of direction of change (reality to model results), and match of size of change, all based on suitability to the model's purpose. These values are out of the scope of this paper.

3 COMPARISONS

The comparisons for Senturion and the ISSM are described at the most aggregated (Category), the SubCategory, and the most detailed (Variable) levels. The data and charts in this section were derived from the author's DIME/PMESII VV&A Tool (Hartley 2009a). Tables are provided to show the coverage counts so that the reader doesn't have to read the details of the radar charts. The radar charts show the aggregated values at the Category and SubCategory levels and the detailed values at the Variable level. These comparisons illustrate the need for the detailed views and substantiate Sargent's comments about aggregated scores.

3.1 Category Level Comparison

At the Category level, the uniformly high results for Senturion dominate the results. The two extra Categories that are covered by the ISSM do not appear to outweigh Senturion’s high values.

Table 2: Coverage comparison at the Category level

Categories	DIME	Pol	Mil	Econ	Soc	Info	Infra	Env	Total	Comment
Senturion	1	1	1	1	1	1	0	0	6	Relatively high, uniform values
ISSM	1	1	1	1	1	1	1	1	8	Low, non-uniform values

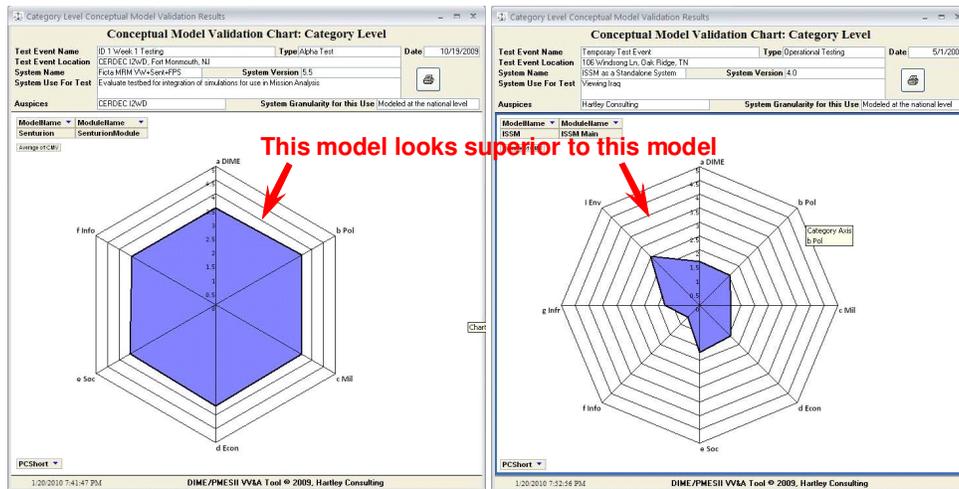


Figure 3: Category level radar chart comparison

3.2 SubCategory Level Comparison

At the SubCategory level, the picture begins to change. The ISSM covers almost three times as many SubCategories as Senturion, although Senturion continues to have higher values for those SubCategories that it does cover. Note that four figures are required to show all of the SubCategories.

Table 3: Coverage comparison at the SubCategory level

SubCategories	DIME	Pol	Mil	Econ	Soc	Info	Infra	Env	Total	Comment
Senturion	1	4	4	2	3	2	0	0	16	Relatively high, uniform values
ISSM	6	5	5	8	7	5	9	1	46	Low, non-uniform values

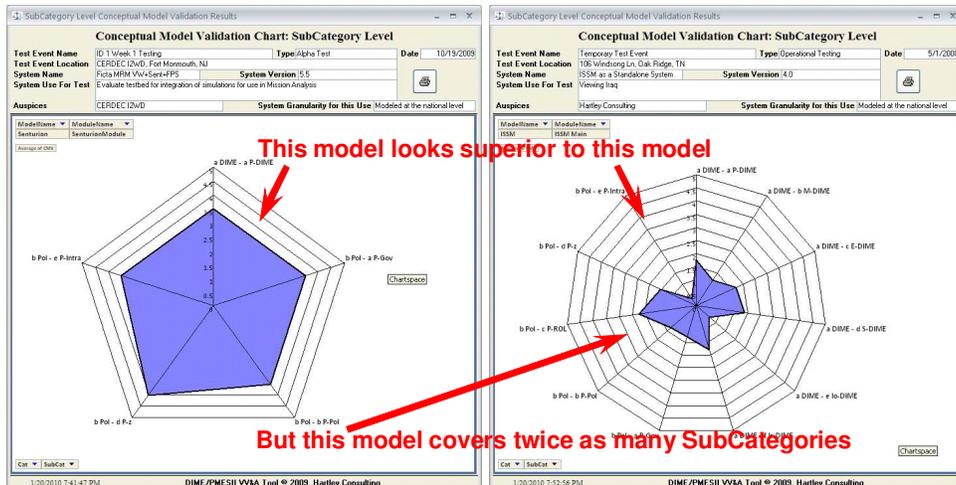


Figure 4: DIME and Pol SubCategories radar chart comparison

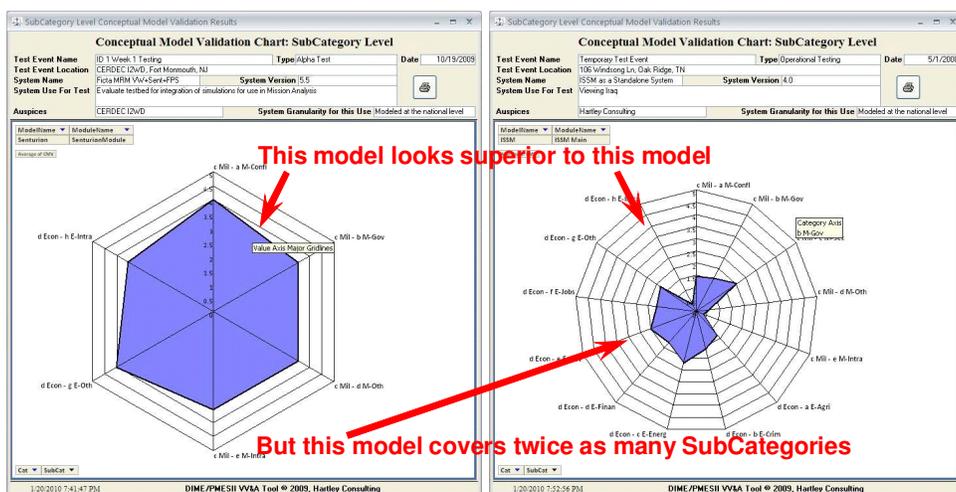


Figure 5: Mil and Econ SubCategories radar chart comparison

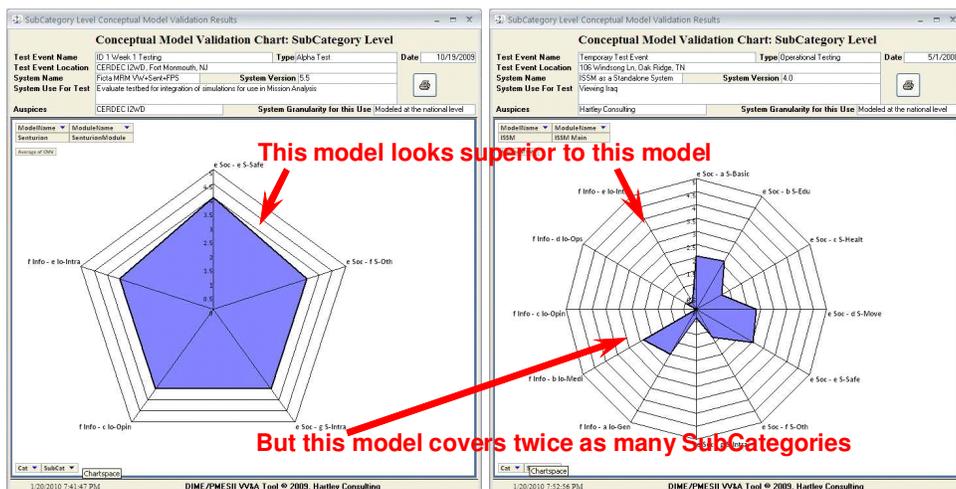


Figure 6: Soc and Info SubCategories radar chart comparison

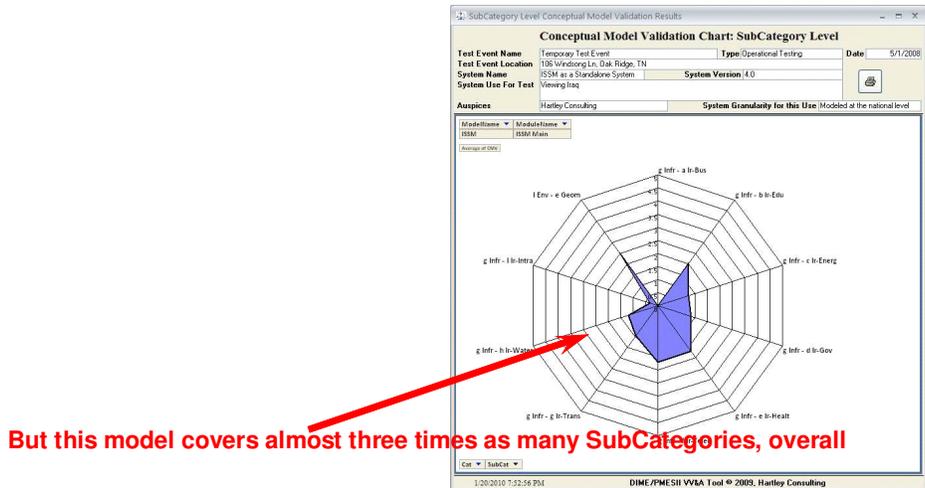


Figure 7: Infra and Env SubCategories radar chart comparison

3.3 Variable Level Comparison

At the Variable level, the differences between Senturion and the ISSM are most clearly shown. Senturion has very narrow coverage, although at high levels, while the ISSM has very broad, but lower level coverage. Only three charts are required to show the Senturion Variables versus 13 charts for the ISSM.

Table 3: Coverage comparison at the Variable level

Variables	DIME	Pol	Mil	Econ	Soc	Info	Infra	Env	Total	Comment
Senturion	1	12	7	6	7	10	0	0	43	Relatively high, uniform values
ISSM	122	60	46	49	45	21	59	2	404	Low, non-uniform values

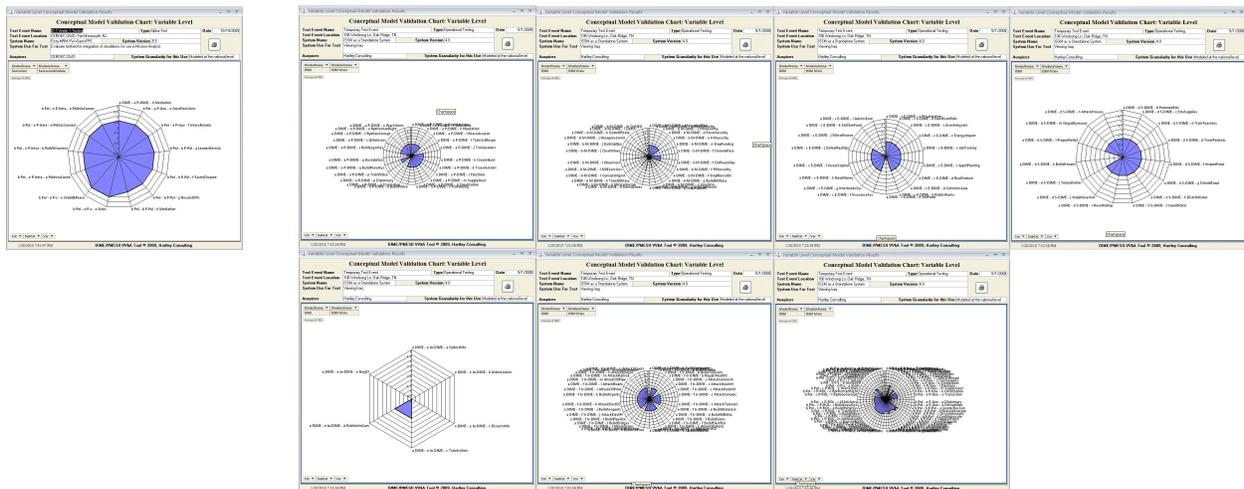


Figure 8: DIME and Pol Variables radar chart comparison

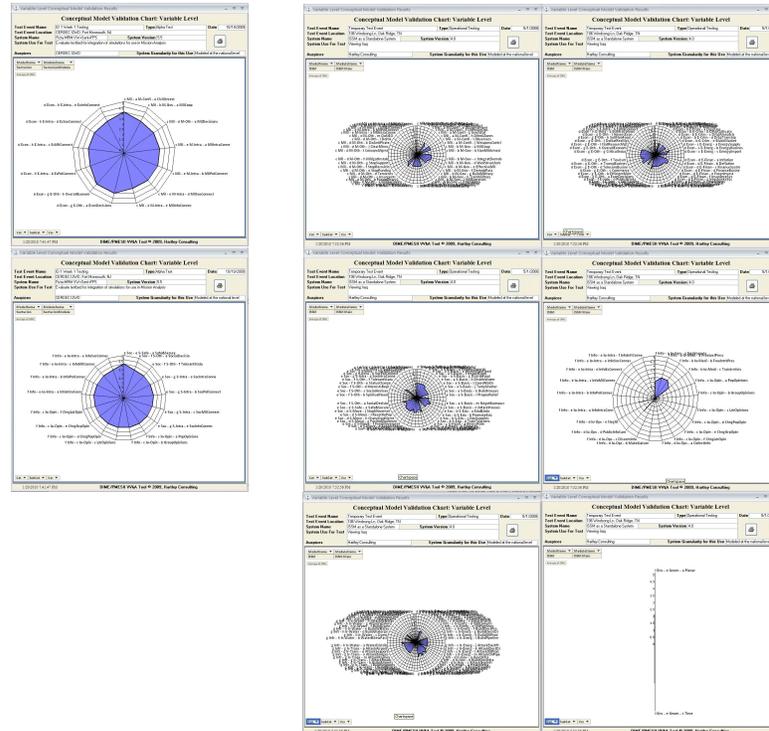


Figure 9: Mil & Econ, Soc & Info, Infra & Env Variables radar chart comparison

3.4 Implications

Figure 10 provides a recap of the Senturion versus ISSM comparison at the SubCategory level; a complete recap at the Variable level would be too messy. At the SubCategory level, the ISSM provides three times the coverage and, at the Variable level, nine times the coverage as does Senturion; however, Senturion has much higher levels of validity where it provides coverage. However, these are data concerning the validity of the conceptual models. For acceptance (Accreditation), what is lacking is an answer to the fitness for purpose question. For this paper we will ignore questions of validity of the coded model.

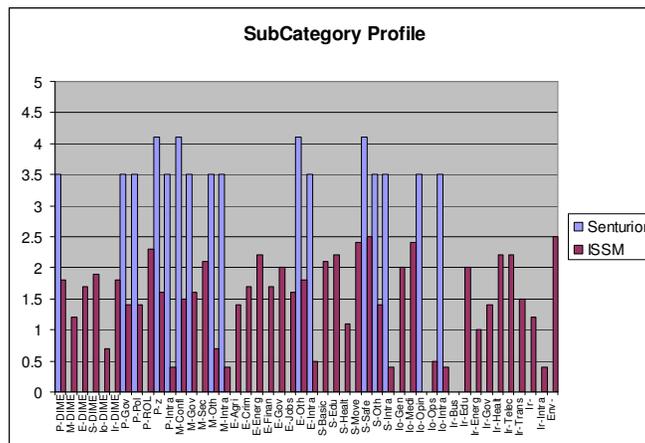


Figure 10: Comparison of Senturion and ISSM profiles at the SubCategory level

As stated earlier, Senturion is used to determine the outcome of political competition over some issue and the ISSM is used to understand and track the situation in some complex emergency. For its purpose, Senturion does not need broad coverage and the ISSM has been demonstrated not to need more than moderate validity for its purpose, but does need very broad coverage. Although the information presented here is not adequate for making the judgment, the reader may be interested in knowing that each model is satisfactory for its purpose.

For our objective in this paper, the comparison shows that two models can be very different and that considerable detail is required to understand and evaluate their differences. The mythical model profiled in Figure 11 further illustrates the point. This mythical model has been created and marketed as using four very highly thought of theories and as being useful to forecast the results of complex emergencies.

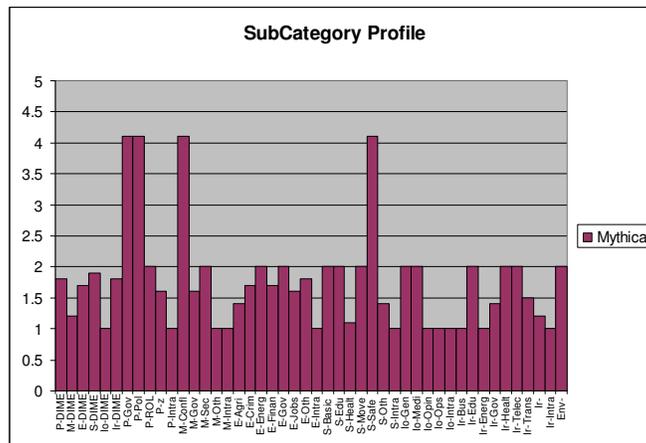


Figure 11: SubCategory profile for a mythical model

The profile in Figure 11 shows that the mythical model has the broad coverage required for its ambitious purpose and that the four areas covered by the highly thought of theories are represented with high scores. However, as is typical of social theories, each theory has rather narrow coverage. The remainder of the coverage is provided by SME conjectures (validity ~2) or connecting code (validity ~1). This profile shows that it is unlikely that the strength of four highly thought of theories is sufficient to provide a high level of validity to this mythical model. On the other hand, the profile might be helpful in determining some purposes for which this model might be very useful.

4 CONCLUSION

Coverage profiles help in visualizing what is modeled and how well it is modeled. The initial validation results of the conceptual models for our two DIME/PMESII models illustrate the point. The differences in their coverage profiles are dramatic. Even without examining the details of the coverage, which would be required for a true understanding of what is modeled and how well, it is clear that they embody different approaches to modeling DIME/PMESII factors. These differences are related to the differences in the purposes of the models. Through the use of a mythical model, we have also shown how coverage profiles can aid in discovering spurious claims of high validity.

REFERENCES

Balci, O. 1986. Credibility Assessment of Simulation Results. In *Proceedings of the 1986 Winter Simulation Conference*. Ed. J Wilson J Henriksen S Roberts. 38-44. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

- Balci, O. 2001. Verification, validation and testing of models, *Encyclopedia of Operations Research and Management Science*. Ed. Saul I. Gass and Carl M. Harris. Kluwer Academic Publishers, Boston, MA.
- Balci, O, R.J. Adams, D.S. Myers, and R.E. Nance. 2002. A Collaborative Evaluation Environment for Credibility Assessment of Modeling and Simulation Applications. In *Proceedings of the 2002 Winter Simulation Conference*. Ed. E Yucesan, C-H Chen, J L Snowdon, and J M Charnes. Washington, DC, 214-220. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Clemence, R. D., Jr., D. S. Hartley III, J. G. McEver, D. F. Noble, A. A. Sciarretta, D. T. Signori, S. Starr, and L. K. Whitney. 2007. *Verification, Validation, and Accreditation (VV&A)*, Evidence Based Research, Inc. Vienna, VA.
- DoD 5000.59 DoD Modeling and Simulation (M&S) Management, Available via <http://www.dtic.mil/whs/directives/corres/pdf/500059p.pdf> [accessed February 25, 2010].
- DoDI 5000.61 "DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)," Available via <http://www.dtic.mil/whs/directives/corres/pdf/500061p.pdf> [accessed February 25, 2010].
- Hartley, D.S., III. 2005a. *MOOTW FAST Prototype Toolbox: FY04 Validation Strategy & Plan*, DRC, Vienna, VA.
- Hartley, D.S., III. 2005b. *MOOTW FAST Prototype Toolbox: FY05 Validation Strategy & Plan*, DRC, Vienna, VA.
- Hartley, D.S., III. 2006a. *Operations Other Than War (OOTW) Flexible Asymmetric Simulation Technologies (FAST) Prototype Toolbox: ISSM v4.00 Analysts' Guide*, Dynamics Research Corporation, Orlando, FL.
- Hartley, D.S., III. 2006b. *Operations Other Than War (OOTW) Flexible Asymmetric Simulation Technologies (FAST) Prototype Toolbox: ISSM v4.00 Users' Guide*, Dynamics Research Corporation, Orlando, FL.
- Hartley, D.S., III. 2009a. DIME/PMESII VV&A Tool, Available via <http://home.comcast.net/~dshartley3/VVATool/VVA.htm> [accessed February 25, 2010].
- Hartley, D.S., III. 2009b. *F²WD HSCB Testbed: 2009 V&V Strategy & Plan*, Hartley Consulting, Oak Ridge, TN.
- Hartley, D.S., III. 2010. DIME/PMESII Community of Interest, Available via <http://home.comcast.net/~dshartley3/DIMEPMESIIGroup/DPGroup.htm> [accessed February 25, 2010].
- Hayes, B.C. and J.I. Sands. 1997. *Doing Windows: Non-Traditional Military Responses to Complex Emergencies*, CCRP, Washington, DC.
- Knepell, P.L. and D.C. Arangno. 1993. *Simulation Validation: A Confidence Assessment Methodology*. IEEE Computer Society Press, Los Alamitos, CA.
- Sargent, R.G. 1987. Overview of Verification and Validation of Simulation Models, An. In *Proceedings of the 1987 Winter Simulation Conference*. Ed. A Thesen; H Grant; W D Kelton. Atlanta, GA, 33-39. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Sargent, R.G. 2004. Validation and Verification of Simulation Models. In *Proceedings of the 2004 Winter Simulation Conference*. Ed. R G Ingalls, M D Rossetti, J S Smith, and BA Peters. Washington, DC. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHY

DEAN S. HARTLEY III is the Principal of Hartley Consulting. Hartley received his A.B. from Wofford College, majoring in mathematics and foreign languages, and his Ph.D. in piecewise linear topology from the University of Georgia. Hartley is a past Vice President of the Institute for Operations Research and Management Science (INFORMS), a past Director of the Military Operations Research Society

Hartley

(MORS), past President of the Military Applications Society (MAS), and a member of the INFORMS Simulation Society (ISIM). Hartley is a Senior Fellow with the George Mason University School of Public Policy, a consulting resource for the Naval Postgraduate School (NPS), Modeling, Virtual Environments & Simulation (MOVES) Institute, and a Research Fellow with the University of Alabama in Huntsville, Center for the Management of Science and Technology (CMOST). Hartley has published *Predicting Combat Effects*, contributed chapters to four other books, and written more than 150 articles and technical documents. His expertise includes modeling of combat and operations other than war (OOTW) and verification, validation, and accreditation (VV&A) of models. His website is <http://dshartley3.home.comcast.net>.