STAT COE-Report-11-2021

A Conceptual Framework for the Establishment of Model Readiness Levels

Dr. Darryl Ahner
Nicholas Jones, CTR
Dr. Melissa Key, CTR
Wayne Adams, CTR
Dr. Sarah Burke, CTR
Corinne Weeks, CTR

October 6, 2021

The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. COE products are available at www.afit.edu/STAT.

STAT Center of Excellence
2950 Hobson Way – Wright-Patterson AFB, OH 45433

Table of Contents

Executive Summary

As engineered defense systems become progressively more complex, the Department of Defense (DoD) plans to further integrate modeling and simulation (M&S) in the design, development, and engineering of new capabilities. With a growing reliance on M&S for test and evaluation, in some cases before physical articles exist to be tested, it is critical that decision makers and developers have a metric to understand the readiness of models.

The processes used to determine how well a model is built and how well it represents its subject are verification and validation, respectively. MIL-STD-3022 and DoDI 5000.61 establish policy and provide instruction on how, when, and by whom verification and validation of M&S are to be performed. However, these instructions do not address how to determine whether a model is ready to support development of a system, especially in the absence of physical test data for the whole system.

The development and use of a mechanism to assign a Model Readiness Level (MRL) is critical for practitioners to quantify the level of trust and/or readiness appropriate for users to place in a model for a specific intended use. MRLs allow decision makers to understand the level of risk they accept when using model results. Additionally, MRLs provide an operational understanding to model developers indicating where their models may need to be improved to better reflect reality and understand the resources needed for those improvements.

A MRL should take into account the fidelity of the model relative to a referent for the specific intended use, the degree of authority a referent is assessed to have in representing reality, and the scope, or set of parameters, behaviors, and constraints in the model which are addressed by the referent. The MRL represents an appropriate level of trust in the model, which is a function of user and decision makers' level of trust in the referent, according to the similarity between the model and the referent in scope and fidelity. Metadata for the model must also be updated to track and clearly communicate the intended use, environment(s), mission thread(s), and range of model inputs and outputs over which the model was validated. Continuous Validation throughout the system development lifecycle enables determining and updating a MRL to ensure that it is a true indicator of the model's readiness for a specific intended use.

Introduction

As engineered defense systems become progressively more complex, the Department of Defense (DoD) plans to further integrate modeling and simulation (M&S) in the design, development, and engineering of new capabilities. As new systems are designed to serve broader Joint missions in challenging situations, often M&S is the safest, most economical, and in some cases, the only means available to test systems under development. Furthermore, the 2018 DoD Digital Engineering Strategy reiterated the importance of models in DoD engineering, and established a goal for Enterprise-level Digital Engineering (DE) across the DoD to accelerate the pace of the acquisition lifecycle and alleviate stove-piping in technology development (Deputy Assistant Secretary of Defense for Systems Engineering (DASD(SE)), 2018). DE is an integrated digital approach to systems engineering that uses authoritative sources of system data and models as a continuum across disciplines to support lifecycle activities from concept through disposal. With a growing reliance on M&S for test and evaluation, in some cases before physical articles exist to be tested, it is critical that decision makers and developers have a metric to understand the readiness of models to represent systems under development.

The processes of verification and validation determine how well a model is built and how well it represents its subject, respectively. MIL-STD-3022 and DoDI 5000.61 establish policy and provide instruction on how, when, and by whom verification and validation of M&S are to be performed. However, these policy and instructional documents do not address how to determine whether a model is ready to support system development, especially in the absence of physical test data for the whole system.

Validation compares the output of the model to some authoritative basis over a specified range of operations and under specified environmental conditions to determine whether or not the model is good enough. Some of the questions that arise when validating a model and determining the quality of the validation are: How should such a comparison be made? From where should that authoritative value be derived? What is “good enough?” Without well-defined answers to these questions, thorough validation of M&S is rare.

Numerous types of models are in use across the DoD. Figure 1 depicts some of the more prevalent ones, and their relations to each other. Although DE leverages all of these models, this work will focus solely on validation of operational analysis M&S, with other types of models to be addressed in future work(s).

There will always be inherent limits on engineers’ abilities to understand a system, but the system aspects which are known and understood can still be captured in a model. Such a model allows the representation of system behavior, however, not all models are equally good representations of the systems they describe, nor are they equally well-suited to all use cases. Thus, it is critical that a mechanism to assign a Model Readiness Level (MRL) be used to quantify the level of trust and/or readiness appropriate for users to have in a model for a specific intended use. This will allow decision makers to understand the level of risk they are accepting when using model results and provide an operational understanding to model developers indicating where their models can be improved to better reflect reality. MRLs may be used to support many different types of decisions such as analysis of alternates, requirement generation, concept of operations development, system development, or operational assessment and acceptance, each of which may require a different level of validation, or MRL, depending on the maturity of the system and the criticality of the decision. To ensure a MRL provides real value to these decision makers and model developers, it is important that any

MRL framework have the qualities of usability, consistency, mathematical rigor, and completeness.

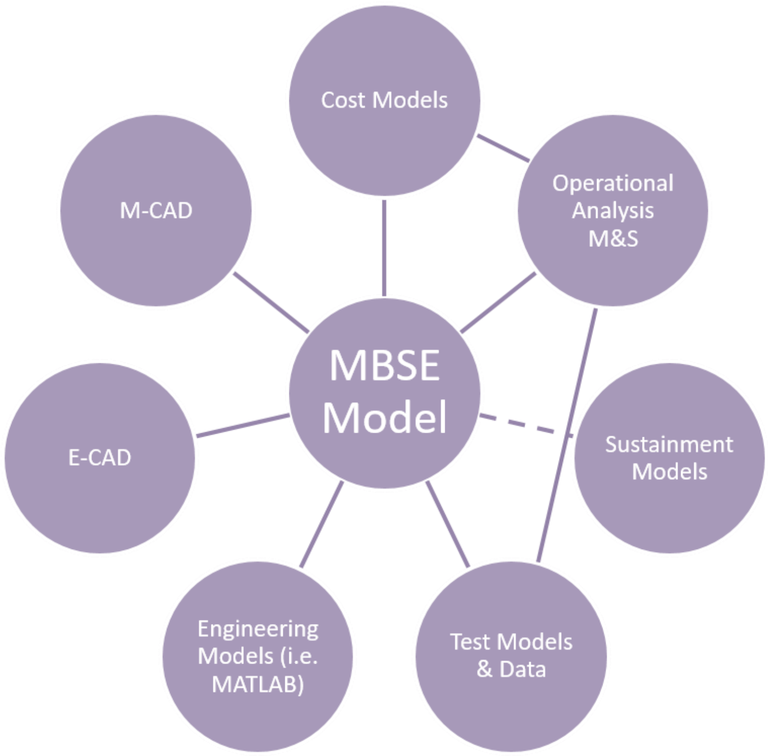

Figure 1: Digital Engineering Models and Their Relationships

Literature Review

There are a number of existing model evaluation frameworks intended to quantify model uncertainty and readiness for use. Notably, Sandia National Laboratories produced two decision architectures for guiding model development; the Predictive Capability Maturity Model (PCMM) and the Modeling & Simulation Technical Readiness Levels (ModSim TRLs) which are derived from the PCMM. Both decision architectures proposed multi-category Technology Readiness Level (TRL)-like rankings to assess model technical maturity in multiple dimensions (Pilch, 2006) (Clay, 2007). A similar matrix for M&S Credibility Assessment is outlined in the NASA Standard of Models and Simulations (National Aeronautics and Space Administration, 2016). While these frameworks are in theory easy to use, they lack technical rigor and have numerous ambiguities that may lead to difficult and inconsistent application in practice. They also produced high-level measures of readiness that may be useful to decision makers in deciding whether or not to use a model in its current state, but are of limited use to model developers seeking to improve a model.

Johns Hopkins University Applied Physics Laboratory (JHU APL) developed a more rigorous approach in the form of the M&S Use Risk Methodology (MURM), a logical framework for determining model readiness in terms of risk level for model use versus risk acceptability for a target use case (Youngblood, Stutzman, Pace, & Pandolfini, 2011). The National Research

Council (NRC) also produced detailed technical analyses of uncertainty quantification methods for model verification and validation, and studied means of quantifying the uncertainties associated with various steps of those processes (National Research Council, 2012). While the MURM and NRC methods are technically rigorous, neither framework can be easily applied to practical situations to produce information that is useful to both decision makers and model developers. In fact, none of the prevailing mechanisms mentioned for assessing the fitness of models for engineering development have provided the appropriate combination of usability, consistency, mathematical rigor, and completeness to facilitate widespread adoption and use across the DoD. This work will explore concepts of MRLs that are grounded in the fundamentals of model validation and which can be built into a rigorous and usable framework to determine model readiness across the DoD.

Key Definitions

To ensure a common understanding of the subject, the following definitions will be used throughout this white paper:

Accuracy - the degree to which a parameter or variable, or a set of parameters or variables, within a model or simulation conforms exactly to reality or to some chosen standard or referent *(Modeling and Simulation Enterprise, 2021)*.

Fidelity - the level of consistency between a model and reality, defined in the three dimensions of accuracy, resolution, and repeatability.

Model - a physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process (DoDI 5000.61).

Modeling and Simulation (M&S) - the use of models, including emulators, prototypes, simulators, and stimulators, either statically or over time, to develop data as a basis for making managerial or technical decisions (Modeling and Simulation Enterprise, 2021).

Referent - a codified body of knowledge about a thing being simulated (SISO-REF-002-1999).

Referent Authority - the strength of credibility of a referent’s claim to be a high-fidelity representation of reality.

Repeatability - the similarity of the results obtained from the same model (or referent) over multiple observations under the same input conditions.

Resolution - the degree of granularity with which a parameter or variable can be determined *(Pace, 2015)*.

Scope - the set of model inputs, outputs, assumptions, and limitations representing the mission-relevant system parameters, environmental conditions, constraints, and requirements, and their allowable values.

Simulation - a method for implementing a model over time (DoDI 5000.61).

Specific Intended Use - the set of dimensions, ranges, and assumptions of the model inputs and outputs needed to represent a system’s relevant mission parameters, environmental conditions, constraints, and requirements, combined with the additional constraints imposed by the target modeling environment and the required level of fidelity for the specific stage of program development.

Validation - the process that determines whether a model has sufficient fidelity relative to an appropriate referent(s) for a specific intended use.

Validity - the fidelity of a model over a pre-specified scope relative to an appropriate referent(s).

Model Validation

Model validation is the foundation on which any assessment of model readiness must be made. The process of validation determines whether a model has sufficient fidelity relative to an appropriate referent(s), or sources of validation, for a specific intended use. This process encompasses three major ideas: 1) sufficient fidelity, 2) appropriate referents, and 3) a specific intended use. These are the three Pillars of Model Validation. Fidelity is the level of consistency between the model and reality. Appropriate referents act as surrogates for the unknowable true state of reality. Finally, the specific intended use determines the scope of the model – the set of inputs, the ranges of each input, and the outputs over which the model and referents need to have sufficient fidelity.

Model validation intends to show an appropriate level of agreement between a model and one or more referents that are trusted representations of reality. Figure 2 illustrates this mapping from reality, to referents, and to the new model. The components in each block of Figure 2 inherit authority from the components of the block to the left. Reality, in the left-most block, serves as the ultimate basis for any authoritative understanding of a system. Thus, reality serves as the basis of authority and truth for the referents, and the referents provide authority to the model.

To determine if a model is ready for an intended use, a user will need to validate it against trusted referents that represent the desired use case at the desired fidelity. This will enable users to determine the MRL as a function of the overlap in scope between the model representation and the referent data, the similarity in fidelity between the model and the referent, and the degree of trust that can be extended to the model based on the trustworthiness of the referent.

This white paper proposes a paradigm in which model validation is an ongoing process performed in conjunction with test and evaluation (T&E) activities. This paradigm will be referred to as Continuous Validation. The capabilities and limitations of a model can be better understood over time by comparing a specified model to an increasing set of referents. These referents may include (but are not limited to) data, alternative models, predictions based on subject matter expert (SME) judgement, and equations for known physical phenomena against which we can test the model. The MRL is a metric to express this growth in understanding of model capabilities.

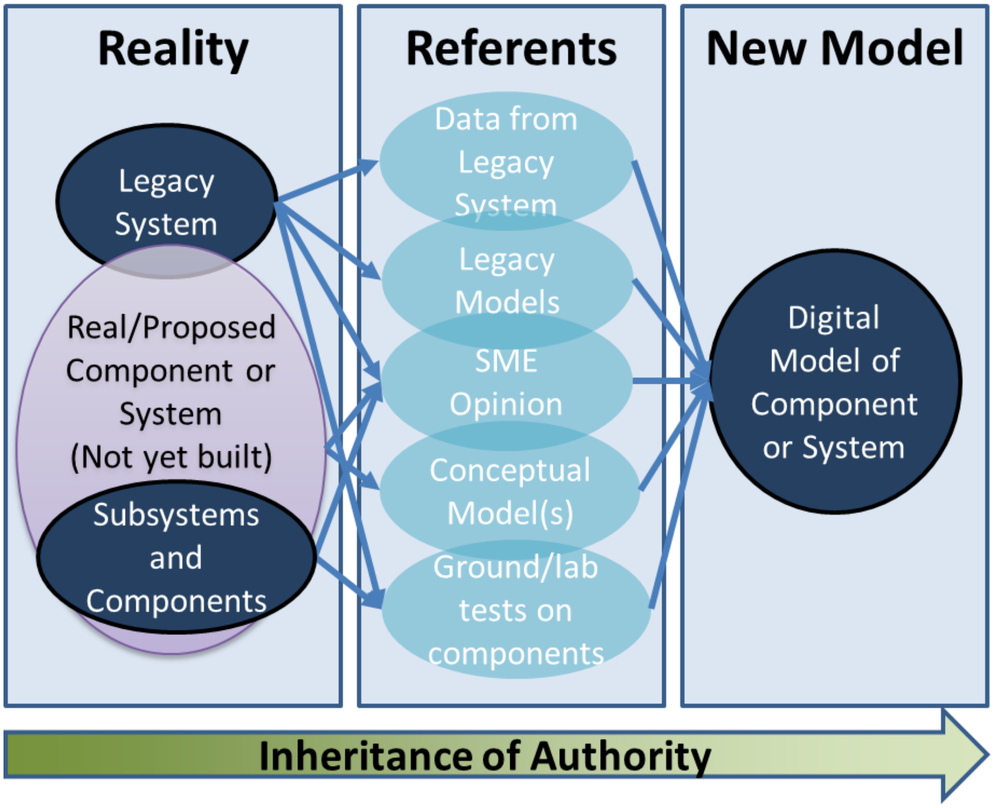

Figure 2: Validating a Model Against a Referent as a Stand-in for "Reality"

Components of Model Readiness

A MRL should take into account each of the Pillars of Model Validation when making a comparison between a model and its intended use: the fidelity of the model relative to a referent for the specific intended use, the degree of authority the referent is believed to have in representing reality, and the scope, or set of parameters, behaviors, and constraints in the model which are addressed by the referent. The MRL represents an appropriate level of trust in the model, which is a function of our level of trust in the referent, according to the similarity between the model and the referent in scope and fidelity. This concept is the core of model readiness, and can be expressed as shown below in Equation 1.

$$MRL = f(Fidelity, Referent\ Authority, Scope)$$

The MRL is designed to provide the means of characterizing the validity (i.e. fidelity using referents as a surrogate for reality) of the model across the input space of interest, and to summarize this information into a single score. For the purpose of this work, each of the Pillars of Validation contributing to model readiness will be addressed as an independent concern. In practice, situations may occur in which there is some overlap between these three aspects of comparison; these situations will be addressed in a future work.

Fidelity

The first element of model readiness is fidelity. The fidelity of a model is its level of consistency with reality, which can be expressed in terms of three dimensions: accuracy, resolution, and repeatability. Accuracy is the degree to which a parameter or variable, or a set of parameters or variables, within a model or simulation conforms exactly to reality or to some chosen standard or referent (Modeling and Simulation Enterprise, 2021). Resolution is the degree of granularity with which a parameter or variable can be determined (Pace, 2015). Repeatability measures the similarity of the results obtained from the same model (or referent) under the same input conditions. These dimensions address the statistical concepts of the center (accuracy) and spread (resolution and repeatability), and can be applied to both the model and the referent. Resolution and repeatability are the central topics of the field of uncertainty quantification (UQ), where they are addressed as epistemic and aleatory uncertainties, respectively. Epistemic uncertainties can, in principle, be reduced by improving experimental or modeling methodologies, while aleatory uncertainties are statistical in nature and can only be quantified. Note that while epistemic uncertainty is in principle reducible, resolution and repeatability are intrinsic properties of the model, while accuracy requires some referent for comparison. These three dimensions can be used to quantify the level of agreement of the model with the referent in statistical terms. The three dimensions of fidelity are illustrated in Figure 3, with their relationships for different levels of fidelity illustrated in the bottom row. Fidelity will be discussed in greater detail in an upcoming STAT COE Best Practice on Modeling and Simulation Validation.

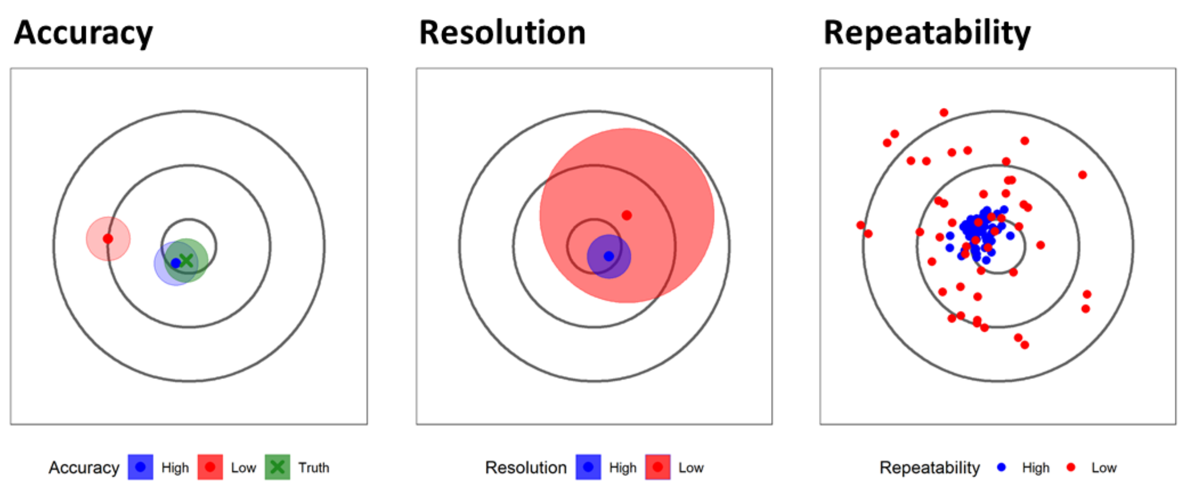

Figure 3: The Three Dimensions of Fidelity

For model readiness, model fidelity relative to a referent will be determined by combining metrics which compare the statistical properties of the model and referent data. The STAT COE is investigating methods for comparison using relative measures of center and variability between the model and referent, and incorporating the concept of statistical signal-to-noise ratio (SNR). An effective method should have strong discriminatory power for differences between the model and referent in each of the dimensions of fidelity and produce a metric which can be used as a de-weighting factor for the similarity of the model and the referent; the nearer the factor is to one, the more fidelity the model has to the referent, with a score of one denoting perfect agreement. If a model agrees perfectly with a referent, then that model's predictions can be trusted for any application where that referent would be trusted as a source of information about a system.

Referent Authority

The second element of model readiness is referent authority. All referents used must be authoritative in order for validation and MRL assessment to serve as means for building trust in a model. Since model validation treats referents as representations of some aspects of reality it is clear that validators are placing some level of trust in them, and they must therefore have some authority. However, not all referents are equally authoritative. A set of recorded performance data for a system and the judgement of a SME may both be drawn from observation of the same real-world event, but one is more objective than the other, and can be considered to be closer to the real world. Both are authoritative in some measure, and as such both can be used to validate the model. It is imperative to determine how authoritative a referent is in order to know how much to trust it, and in turn, how much trust to place in a model that matches it with high fidelity. One possible method is to build a list of possible referents, and rank them by their typical perceived trustworthiness in the T&E community. STAT COE has identified potential data sources that might be used as authoritative referents in Table 1.

Authoritative Referents for Model Performance
Component lab test data
First principles/physics predictions
Hardware in the loop data
Lab-scale system test data
Legacy/similar sub-system models predictions
Legacy/similar sub-system test data
Legacy/similar system model predictions
Legacy/similar system test data
Live system test data
Operational real-world data
Prototype test data
SME concurrence with observed model performance
SME estimates of expected model performance

Table 1: Potential Referents for Model Performance

It should be noted that the list of referents in Table 1 is not complete. The STAT COE has requested the assistance of the DoD T&E community in determining an appropriate list of referents, ranked according to their level of authority, which could be used to determine the degree of trust that is reasonable to place in a referent, or in a model that matches that referent with high fidelity. For a referent of a particular type, the applicable referent authority value is de-weighted according to the model's fidelity to that referent. This determines the authority that can be transferred from the referent to the model; the model can only be trusted to the degree that it agrees with real referents, and only as far as those referents would be trusted.

In many cases, more than one referent will be utilized to assess model readiness. All applicable referent authority values must be combined to determine an overall referent authority score for use in calculating model readiness. In practice, the referents used in calculating the MRL for a particular model will each be systematically assigned an authority weighting, determined based on an ordered ranking of the applicable referent authorities. The STAT COE is developing a method to determine referent authority using weighted average fidelity of referents at each authority level. In cases where all the referent authorities have different authority levels, the ordered ranking is straightforward; the referent authority with the highest authority level is ranked first, followed by the next highest referent authority, and so on. Once the referent authorities have been ranked, the MRL can be calculated to account for the model's fidelity relative to the most authoritative referent available. The advantage of this method is that lowerlevel referents are down-weighted in magnitude relative to the higher-level referents.

Scope

The last element of model readiness is the scope of a model. Scope captures the inputs, outputs, and assumptions of a model in terms of explicit, defined value ranges and data types. The mission-relevant values of the same parameters in the referent define the referent's scope. The model and the referent can be directly compared where their parameters are the same and are defined over the same values. This comparison, performed by determining the fidelity of the model with respect to the referent, in each area of scope, allows authority to be transferred from the referent to the model. Likewise, a model's fitness for a specific intended use can be assessed where the model parameters are the same as the required system parameters to meet the intended use case, and are defined over the same values. The fidelity of the model with respect to the requirements for the specific intended use establishes the authority of the model in each area of scope as a representation of the system in that use case. It is important to note that the model scope, referent scope, and scope of intended use are all three defined separately – each with its own inputs, outputs, and assumptions. The model can be trusted in regions where all three of these scopes overlap, and where the model has sufficient fidelity relative to both the authoritative referent and the intended use.

First, the stakeholders of the system must divide the input space into regions of interest. The regions of interest should cover the set of inputs relevant to the specific intended use of the model. A comparison of model and referent input and output dimensions determines whether the model covers the same behaviors and phenomena of interest as the referent. If there are known discontinuities in the behavior of the model (for example, subsonic versus supersonic speed), these discontinuities create natural bounds for these regions. In many cases, the boundary itself might also be treated as a region of interest (e.g. transonic speed). Other boundaries are imposed by the intended use of the model: some regions of model operation may be of greater interest to the model users than others, and the scope should be divided to focus on these regions. Determining appropriate bounds for these regions will likely require SMEs, especially when the boundary is dependent on more than one input.

Once determined, these dimensions and their applicable value ranges should be documented, both to ensure that the scope of the MRL derived from them is well understood, and to facilitate repeatability of MRL determination and future use of the scope information. This documentation may be done in a number of ways, including via metadata associated with the model, referent data set, or system design, or in a Model Based Systems Engineering (MBSE) model of the system or modeled system, or through some combination of these mechanisms. In addition to documenting the relevant scope parameters of a system, a MBSE model also supports verifying that the scope parameters of subsystem models are inter-compatible. It also helps assure that the implications of system-level requirements and design constraints for subsystem scopes can be taken into account in model selection and MRL determination. The cross-verification of model scope parameters in metadata with those recorded in MBSE system models should be facilitated by the use of common data formats and standard units for metadata and MBSE.

Next, one must determine the set of validation points within each region. Validation points are the input settings at which data is available from the model and from one or more referents. Especially for complex models, determining the coverage of the validation points can be difficult. These validation points will be used to assess the model fidelity relative to the referent within the region, and the fidelity will determine the appropriate degree of authority that can be transferred from the referent(s) to the model within that region. Not all data in a referent may lie within a single (or any) region of the scope of the model or the intended use.

If the model inputs differ from the measured factors impacting the referent response, then the model assumptions must be considered. If a model assumption applies to an input dimension that is not fully represented in the model, and is both valid for the intended use and representative of the referent data, the referent may still be used within the scope where the assumption holds. While the ranges of the validation points within the input dimensions may also differ from the ranges represented by the model, the referent can still support the model in the portion of the input space where the ranges overlap with each other and with the intended use. These types of restrictions are likely to be common in practice, and multiple referents may frequently be required to completely cover the scope of interest.

It should be noted that different regions of interest may require separate validation plans to determine if the model is valid within the region, especially when different test facilities are required for different regions, or discontinuities make it difficult to combine information collected across regions. If the assessment of fidelity shows that the model has sufficient fidelity in the region to be useful, then the MRL will indicate that the model is ready for use in that region. So long as the risks associated with not having every possible input are quantified, understood, and acceptable, a conclusion can be drawn about whether or not the model is ready. However, the number of dimensions in the scopes of both the model and the intended use are critical concerns for ensuring the MRL can be calculated. As the volume of the model's scope increases so does the required validation effort. To keep validation work on budget in terms of time and resources it may be necessary to limit the scope of the model to be as small as possible while retaining the necessary decision making framework. Reducing the scope but still being able to make rigorous defensible decisions can be accomplished by application of the STAT process, shown in Figure 4. The application of the STAT Process is covered in the STAT COE Test Planning guide (Adams, et al., 2020).

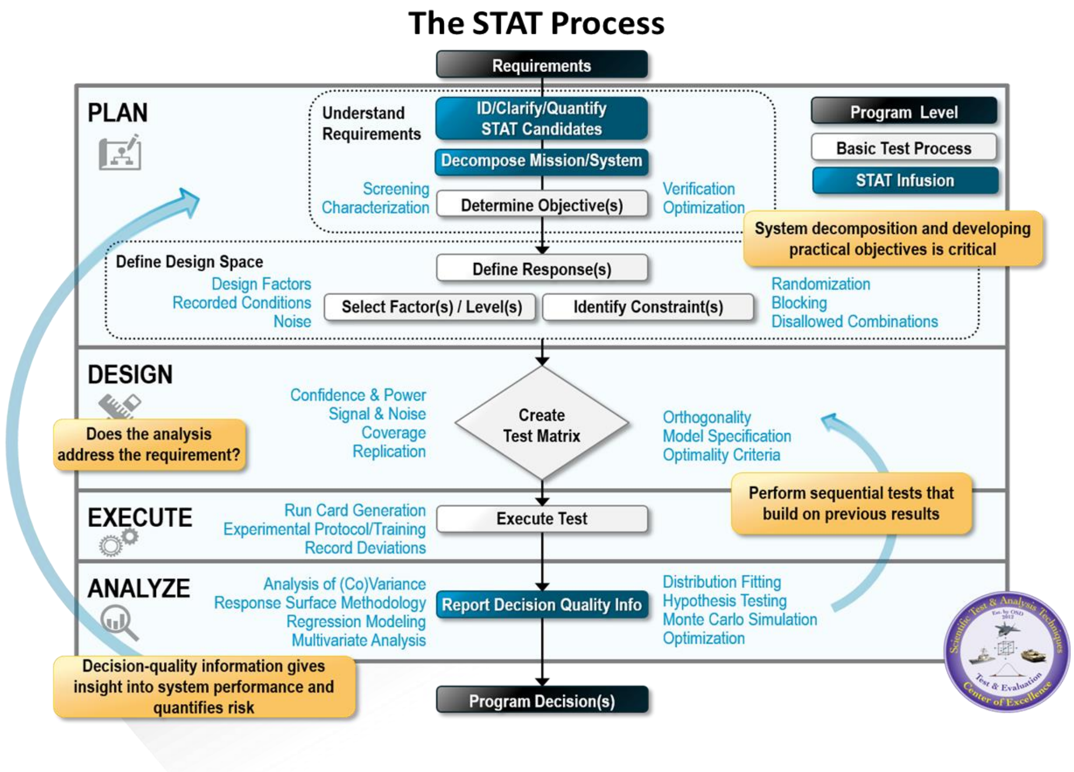

FIGURE 4: THE STAT PROCESS

Continuous Validation

MRLs are intended as a framework for determining the rigor and range of validity of a model which can then be extended to determine the readiness (suitability for intended use and validity in the desired bounds) of a valid model to support a defined new use case. A key concern for validation in M&S is that it must be repeated throughout the system development lifecycle. This concern leads to the concept of Continuous Validation, where the set of appropriate referents is updated over time to improve the assessment of the fidelity and expand the scope over which the model has been validated. The objective of Continuous Validation is continuous improvement, and expansion of the types and completeness of the referent data along with updated analysis. In a M&S environment, especially early in system development when referents based on test data are scarce, SMEs can often provide an initial assessment of what the output of a new model should look like. Legacy data and models can be used as a referent when the scope between the referent and model are sufficiently similar. As lab-scale, ground, and flight test results successively become available, the validation assessment is updated to incorporate this information. As the model is developed and refined over its lifetime, it should be repeatedly re-validated against the best available data. The MRL should be re-assessed at each validation and updated using new referents as they become available. At each re-assessment, metadata for the model must also be updated to track and clearly communicate the intended use, environment(s), mission thread(s) and range of model inputs as well as outputs over which the model was validated. The Modeling and Simulation (M&S) Community of Interest (COI) Discovery Metadata Specification (MSC-DMS), published by the Modeling & Simulation Enterprise (MSE), provides a standard for recording and documenting model metadata on a wide variety of model attributes, including model validation (Modeling and Simulation Coordination Office (MSCO), 2012). This will allow users to determine the appropriate degree of trust to place in a model at a given stage of development and for a given use, how the trustworthiness of the model has evolved over time, and what the bounds of that trust should be, based on a scaling of the authority of the referents used in the validation of different parts of the model space. Ideally, within-program validation will culminate with operational data as the system is used in field. Once a MRL has been established, it can support prudent model reuse in an environment that supports model composability. Composability is defined by MSE as “the capability to select and assemble reusable modeling and simulation components in various combinations into simulation systems to meet user requirements” (Modeling and Simulation Enterprise, 2021). The DoD future composable model architecture is outlined in the Defense Modeling and Simulation Reference Architecture (DMSRA) standard (Defense Modeling and Simulation Coordination Office, 2020). If different programs choose to use the same composable models, their own lab, ground, and flight tests will provide additional referents for Continuous Validation to further improve the overall understanding of the model and provide information to those interested in applying the model to a new intended use. Continuous Validation is a key process that enables determining and updating a MRL to ensure that it is a true indicator of the model’s readiness for a specific intended use.

Conclusion

As the DoD engineers defense systems of increasing complexity and comes to rely on M&S to understand and develop those systems, it is imperative that the models developed are well understood and trustworthy. DoDI 5000.61 recognizes model validation as the primary means of establishing this trust, and mandates the validation of M&S “throughout their lifecycles” by DoD Components (US Department of Defense, 2018). As a solution, the STAT COE proposes a MRL system that is founded in Continuous Validation. A MRL ensures the consideration of the scope of validity of a model, its fidelity relative to the available referents, and the authority of those referents in order to quantify a model’s readiness to support its specific intended use (US Department of Defense, 2012) (Modeling and Simulation Coordination Office (MSCO), 2012). A MRL metric would further support Continuous Validation of models throughout their lifecycles to facilitate a better understanding of model readiness, assure compliance with DoD policy, and provide decision makers a tool to understand the risk associated with using a model in support of decisions. Furthermore, the proposed MRL will enable developers to better understand model maturity and identify areas for model improvement. The MRL concept presented in this paper, supported and refined through the use of Continuous Validation, will provide a means for DoD Components to quantify and track the readiness of models to support original and evolving use cases via STAT-based measures and associated metadata. Through this quantitative understanding of model capabilities, MRLs will enable DoD Program Managers and M&S developers to employ M&S confidently in the engineering of complex defense systems.

References

Adams, W., Burns, D., Burke, S., Divis, E., Harmon, M., Jones, K., . . . Truett, L. (2020). *Guide to Developing an Effective STAT Test Strategy V 8.0.* STAT COE, Department of Defense. Air Force Institute of Technology.

Clay, R. L. (2007). *Modeling and Simulation Technology Readiness Levels.* Sandia National Laboratories.

Defense Modeling and Simulation Coordination Office. (2020). *Defense Modeling and Simulation Reference Architecture.* Defense Modeling and Simulation Coordination Office.

Deputy Assistant Secretary of Defense for Systems Engineering (DASD(SE)). (2018). *DoD Digital Engineering Strategy.* United States of America, Department of Defense, Office of the Secretary of Defense.

Government Accountability Office. (2020). *Technology Readiness Assessment Guide: Best Practices for Evaluating the Readiness of Technology for Use in Acquisition Programs and Projects.* Government Accountability Office.

Modeling & Simulation Coordination Office (MSCO). (2011). *Verification, Validation, and Accreditation Recommended Practices Guide.* Retrieved from https://vva.msco.mil/

Modeling and Simulation Coordination Office (MSCO). (2012). Modeling and Simulation (M&S) Community of Interest (COI) Discovery Metadata Specification (MSC-DMS).

Modeling and Simulation Enterprise. (2021, August). *Department of Defense Modeling and Simulation Glossary*. Retrieved from Department of Defense Modeling and Simulation Enterprise: https://www.msco.mil/MSReferences/Glossary

National Aeronautics and Space Administration. (2016). *Standard for Models and Simulations.* NASA.

National Research Council. (2012). *Assessing the reliability of complex models: Mathematical and statistical foundations of verification, validation, and uncertainty quantification.* The National Academies Press. Retrieved from https://doi.org/10.17226/13395

Pace, D. K. (2015). Fidelity, Resolution, Accuracy, and Uncertainty. *Modeling and Simulation in the Systems Engineering Life Cycle, Simulation Foundations, Methods and Applications*, 369. Retrieved from https://doi.org10.1007/978-1-4471-5634-5_10

Pilch, M. (2006). *Predictive Capabiity Maturity Model.* Sandia National Laboratories.

US Department of Defense. (2012). *MIL-STD 3022 Department of Defense Standard Practice Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations.*

US Department of Defense. (2018). Department of Defense Instruction 5000.61.

Youngblood, S. M., Stutzman, M., Pace, D. K., & Pandolfini, P. P. (2011). *Risk Based Methodology for Verification, Validation, and Accreditation (VV&A) M&S Use Risk Methodology (MURM).* The Johns Hopkins University Applied Physics Labratory.