SCIENTIFIC TEST
& ANALYSIS
TECHNIQUES
CENTER OF EXCELLENCE

# Constructing a Metric for Fidelity in Model Validation

June 2022

Corinne Weeks, Contractor
Nicholas Jones, Contractor
Dr. Melissa Key, Contractor

The STAT COE provides independent STAT consultation to designated acquisition programs and special projects to improve Test & Evaluation (T&E) rigor, effectiveness, & efficiency.

About this Publication:
This work was conducted by the Scientific Test & Analysis Techniques Center of Excellence under contract FA8075-18-D-0002 Task FA8075-21-F-0074.

For more information:
Visit, www.AFIT.edu/STAT
Email, CoE@afit.edu
Call, 937-255-3636 x4736


Scientific Test & Analysis Techniques Center of Excellence
2950 Hobson Way
Wright-Patterson Air Force Base, Ohio



Modernizing *the* Culture *of* Test *&* Evaluation

## Table of Contents

## Executive Summary

As the Department of Defense (DoD) shifts toward a Digital Engineering (DE) approach using Modeling and Simulation (M&S) for system development, models must be considered trustworthy to minimize any risk introduced by the use of models in place of physical articles. Validation aims to establish trust in models but is usually a subjective process resulting in a binary indicator of validity which grants the model validity for its entire lifetime without reassessment. The STAT COE fidelity metric provides a straightforward, quantitative method for identifying the extent to which a model matches available data, a key component of model validation. Fidelity, appropriate referents, and the specific intended use factor into a Model Readiness Level (MRL), which will enable assessment of a model's readiness to be used and trusted in the DE paradigm, with full comprehension of the model's capabilities and risk.

*Keywords: Digital Engineering, Modeling & Simulation, Validation, Fidelity*

## Background

Modeling and Simulation (M&S) are becoming increasingly critical to the development of new capabilities in the Department of Defense (DoD). To further leverage M&S capabilities, the DoD 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 highly dependent on the use of M&S, so the trustworthiness of models used is of critical importance to the sound design of complex systems.

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 (DoDI 5000.61), and therefore establishes whether a model can be trusted to represent the associated physical system. In practice, thorough validation is rare. Often, the outcome of validation is a binary decision designating whether or not a model has been validated, a simplification that fails to address the degree to which a model can be considered trustworthy. In addition, this classification is typically the product of subjectively chosen validation methods combined to establish the validity of a model. Current practice fails to continuously reassess model validity as more data becomes available or as the particular use case of the model changes. There is no universal, objective validation method which can be quickly, repeatedly applied to the wide variety of M&S in the DoD. This deficiency creates an obstacle in the shift toward the DE approach.

The STAT COE is currently working to develop Model Readiness Levels (MRLs), a rigorous validation framework that aims to resolve the shortcomings of current validation practices by providing an objective understanding of the trustworthiness of M&S in DE (Ahner et al. 2021). The STAT COE Whitepaper “A Conceptual Framework for the Establishment of Model Readiness Levels” redefines validation in terms of three pillars that can quantify the trustworthiness of a model: sufficient fidelity, appropriate referents, and a specific intended use (Ahner et al. 2021). The Whitepaper defines fidelity as the level of consistency between the model and reality, appropriate referents as trusted representations of reality, and the specific intended use as the scope of the model – the set of inputs and outputs over which the model and referents need to have sufficient fidelity. Referents may include sources such as data, alternative models, predictions based on subject matter expert (SME) judgement, and equations for known physical phenomena against which developers can test the model (Ahner et al. 2021).

To develop a complete MRL framework, quantitative descriptions of all three pillars are required. This best practice will discuss fidelity while quantitative descriptions of both referent authority and scope will be discussed in detail in an upcoming best practice, “Elements of a Mathematical Framework for Model Readiness Levels,” expected in 2022.

## Introduction

This best practice provides a detailed, quantitative approach for evaluating the fidelity of a model with a referent, which when incorporated into the MRL framework will provide a complete description of a model’s validity and help to address shortcomings of current validation practices. This best practice will first discuss the key dimensions of fidelity: accuracy, repeatability, and resolution. Next, it will discuss how these dimensions are critical in constructing a rigorous metric to evaluate the fidelity. Two metrics will be outlined which evaluate the accuracy and compare the variability of the model and referent. These two metrics combine to produce an overall metric for fidelity that is straightforward and intuitive, with a good ability to distinguish between high and low fidelity cases, as demonstrated with a toy problem. Finally, this best practice addresses the utility of this metric in the diverse M&S environment of the DoD and how it can be incorporated into a MRL to assess a model's readiness to be used and trusted in the DE paradigm.

**Dimensions of Fidelity: Accuracy, Repeatability, and Resolution**

Fidelity is defined in terms of three dimensions: accuracy, repeatability, and resolution (Ahner et al. 2021). 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). Repeatability measures the similarity of the results obtained from the same model (or referent) under the same input conditions (Ahner et al. 2021). Resolution is the degree of granularity with which a parameter or variable can be determined (Pace, 2015).

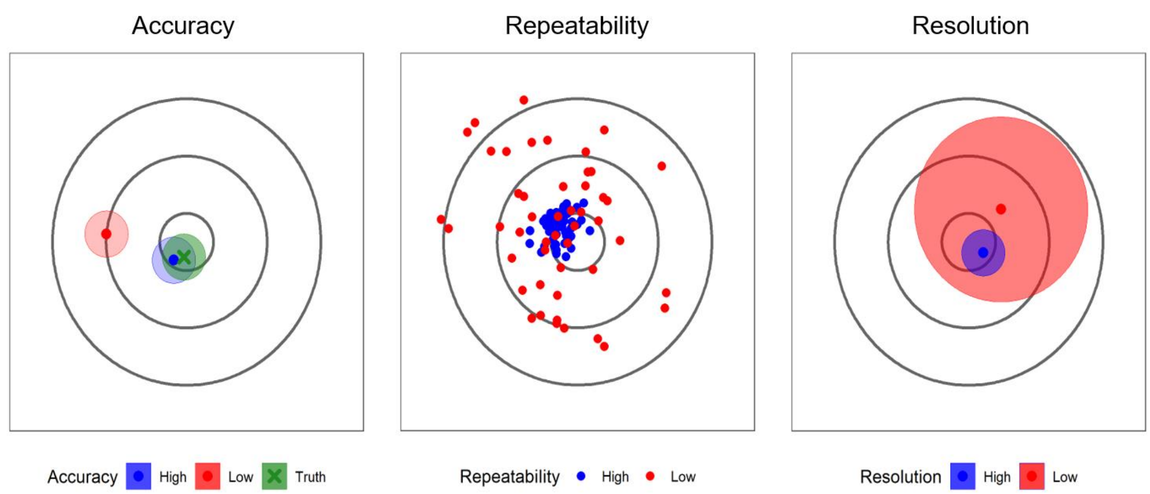
For a model to be consistent with a referent, its responses must be acceptably close to those observed in the referent. Accuracy identifies how closely the means of the two data sets align, however it fails to capture any information about how two distributions may align. The spread of the distribution is critical in an assessment of fidelity since a model should capture any real-world stochastic behavior of a system such that it can be trusted as a stand-in for reality. Together, repeatability and resolution characterize the system variability and describe the spread of the distribution; both require careful treatment for the development of a fidelity metric.

In the field of uncertainty quantification, repeatability and resolution are also known as aleatory and epistemic uncertainties, respectively. Aleatory uncertainty arises from an inherent randomness in the properties or behavior of the system under study (Helton, 2011). Epistemic uncertainty derives from a lack of knowledge about the appropriate value to use for a quantity that is assumed to have a fixed value in the context of a particular analysis (Helton, 2011).

The three dimensions of fidelity are illustrated in Figure 1.

**Figure 1**

*Dimensions of Fidelity*



For the purposes of this work, repeatability can be represented by the standard deviation of the data set. The standard deviation is a measure of the spread of a data set and for a normal distribution completely characterizes the shape of the distribution. For more complex distributions, higher order moments such as skew and kurtosis are needed to completely characterize the shape of the distribution and therefore characterize the repeatability. In the majority of DoD cases, however, data availability is limited such that using the standard deviation alone is sufficient to identify if a model has captured the real-life stochastic behavior of a system.

Resolution must also be quantified for the construction of a fidelity metric. Epistemic uncertainty can be modeled as a uniform distribution: a distribution with equal probability everywhere between two bounds. The resolution $\delta$ can be defined as the distance between these two bounds. Distributions other than the uniform distribution may in practice be more suitable for specific applications, however, using the uniform distribution provides a general approach for computing the fidelity in the majority of cases.

Since repeatability and resolution are both measures of the system variability, a single aggregate measure of variability is helpful to define. The total variance can be found by adding the variance due to the repeatability and the resolution, since variance from independent sources is additive. For repeatability, the variance is $s^2$, where $s$ is the standard deviation. For resolution, the variance of a uniform distribution is $\delta^2/12$, where $\delta$ is the resolution, or the distance between the two bounds. Equation 1 defines $s^*$, the resolution-modified standard deviation, providing a measure for the total variability.

$$s^* = \sqrt{s^2 + \frac{\delta^2}{12}} \tag{1}$$

Together, the accuracy, repeatability, and resolution can quantify the level of agreement of the model with the referent in statistical terms.

**Developing a Fidelity Metric**

To construct a rigorous metric for assessing a model's fidelity in comparison to an appropriate referent, one must consider the desired properties of such a metric. For a model to have high fidelity with a referent, a model must have realistic representations of both the accuracy and variability. If a behavior is highly variable in reality, a high fidelity model will capture this variable behavior, allowing users to trust the model as a stand-in for reality.

Additionally, the metric should be broadly applicable to many potential models and referents, since not all models have the same properties. For example, a deterministic model does not have a standard deviation as it is perfectly repeatable. The fidelity metric must be flexible in how it assesses variability in order to determine the model-referent fidelity in any case considered (e.g., stochastic and deterministic referents in tandem). The use of the resolution-modified standard deviation $s^*$ enables flexibility in the fidelity assessment as it provides a single measure encapsulating the variability from all sources. When a standard deviation does not exist, $s^* = \delta/\sqrt{12}$ represents the total variability of the system.

The fidelity metric should be bounded between zero and one, where zero denotes no fidelity with a referent and one denotes perfect fidelity with a referent. These bounds allow fidelity to be a weighting factor in the computation of a MRL. This approach identifies metrics for accuracy and variability where each is bounded between zero and one. The accuracy metric should equal one when the means are identical and approach zero as the means diverge. Similarly, the variability metric should equal one when the model and referent have the same resolution-modified standard deviation and approach zero as the resolution-modified standard deviations of the model and referent diverge.

For consistency, the fidelity metric must provide similar assessments for comparable distributions in a variety of different contexts. This can be accomplished through the use of a non-dimensional metric for fidelity, where the scale of a given situation will have no impact on the computed fidelity.

A fidelity metric must have a strong ability to discriminate between cases where the model and referent agree, and cases where the model and referent disagree. With limited data availability, determining whether or not a model agrees with a referent becomes challenging. Since data is often limited in the high cost testing environment in the DoD, this capability of the fidelity metric will be examined closely.

Finally, the fidelity metric should be an intuitive measure: users of the metric should naturally understand the fidelity rating assigned. Additionally, multiple parties who may be using a model should understand and concur on what the fidelity means for the system in question. An intuitive metric must still be calculated using objective measures; therefore, user-determined weights should be avoided as they introduce subjectivity.

To adequately capture both dimensions of accuracy and variability, this best practice proposes two metrics which assess these dimensions independently.

### *Accuracy Metric*

The accuracy metric $f_a$ is defined in Equation 2 and assesses the model's accuracy with respect to the referent.

$$f_a = e^{-\frac{1}{2}\left(\frac{\bar{x}_m - \bar{x}_r}{s_r^*}\right)^2} \tag{2}$$

In Equation 2, $\bar{x}_m$ and $\bar{x}_r$ are the sample means of the model and referent, respectively; $s_r^*$ is the resolution-modified standard deviation for the referent.

The accuracy metric uses the negative exponential function to achieve the desired behavior of decay towards zero as the model and referent disagree more strongly. Additionally, the negative exponential maps the domain $[0, \infty)$ to $(0,1]$, bounding the metric between zero and one.

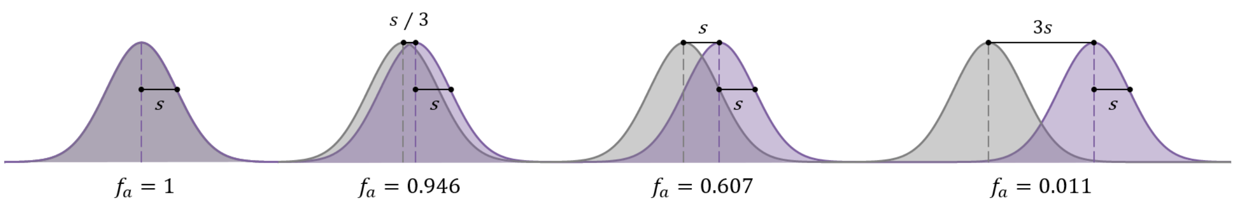
The accuracy metric assesses the model by computing the difference in means between the model and referent. The difference in means is normalized and made dimensionless by the resolution-modified standard deviation of the referent. Normalization achieves consistency when assessing scenarios of vastly different scales. For example, when $\bar{x}_m = 110$, $\bar{x}_r = 100$, and $s_r^* = 10$, the same fidelity is obtained as when $\bar{x}_m = 11$, $\bar{x}_r = 10$, and $s_r^* = 1$. The referent is used for normalization as the referent has some degree of authority in representing reality, as opposed to the model, which must inherit that authority from the referent. The use of the referent for normalization means that a fidelity assessment against a referent with high variability will have a less strict requirement for accuracy than for a comparison against a low variability referent. The model variability has no effect on the severity of the accuracy assessment.

The structure of the accuracy metric closely matches the form of the probability density function (PDF) for a normal distribution. Thus, the accuracy metric can be interpreted as the relative likelihood outputted by the normal distribution PDF, where the height of the curve is normalized to one.

The accuracy metric provides an intuitive measure for fidelity, which is observed through sample fidelity calculations. Figure 2 shows that a difference in means equal to one resolution-modified standard deviation returns a fidelity score of 0.607 while a difference of three resolution-modified standard deviations returns a score of 0.011. For illustration, Figure 2 depicts normal distributions with standard deviation $s$ and neglects resolution in the computation of the metric.

**Figure 2**

*Visualizing the Accuracy Metric*




***Variability Metric***

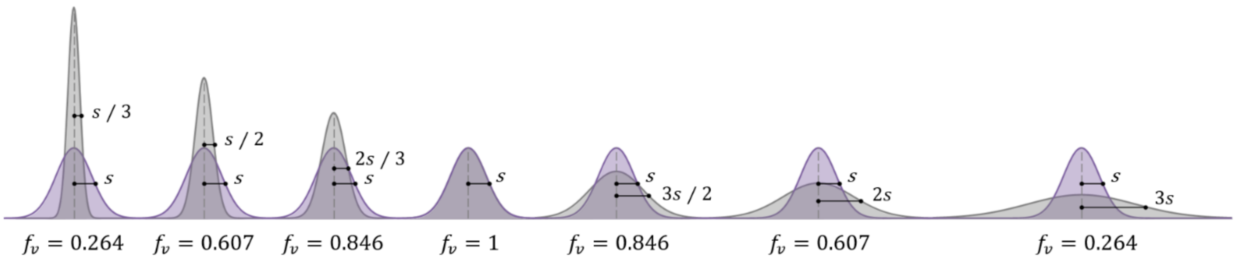
The variability metric $f_v$ is defined in Equation 3 and assesses the similarity in variability between the model and the referent.

$$f_v = e^{-\frac{(s_m^* - s_r^*)^2}{s_m^* s_r^*}} \tag{3}$$

Taking a closer look at the variability metric, the similarity in the variability of the model and referent primarily depends on the difference in the resolution-modified standard deviations. The difference in variability is normalized by the product of the resolution-modified standard deviations of the model and referent. To obtain a high value for the variability metric, the total variability of the model must match the referent, meaning the standard deviations should match as well as the resolutions. Particularly when the standard deviation is zero, models must have a realistic resolution identifying a meaningful level of precision to which the model can be trusted. The variability metric provides an intuitive measure for fidelity, which can be observed through sample fidelity calculations, pictured in Figure 3.

**Figure 3**

*Visualizing the Variability Metric*




***Fidelity Metric***

The overall metric for fidelity is given in Equation 4, where the accuracy and variability metrics are multiplied together to form a single metric for assessing the level of consistency between the model and the referent.

$$f = f_a f_v = e^{-\frac{1}{2}\left(\frac{\bar{x}_m - \bar{x}_r}{s_r^*}\right)^2} e^{-\frac{(s_m^* - s_r^*)^2}{s_m^* s_r^*}} \tag{4}$$

The multiplicative combination of metrics means that the overall fidelity metric may only be as high as the fidelity of either of its components. If the accuracy is low, but the match in variability is high, the overall fidelity metric will be low. Likewise, if the accuracy is high, but the match in variability is low, the overall fidelity metric is low. Thus, the overall metric has a strict requirement for fidelity that can distinguish good and bad matches in either dimension.

**Toy Problem**

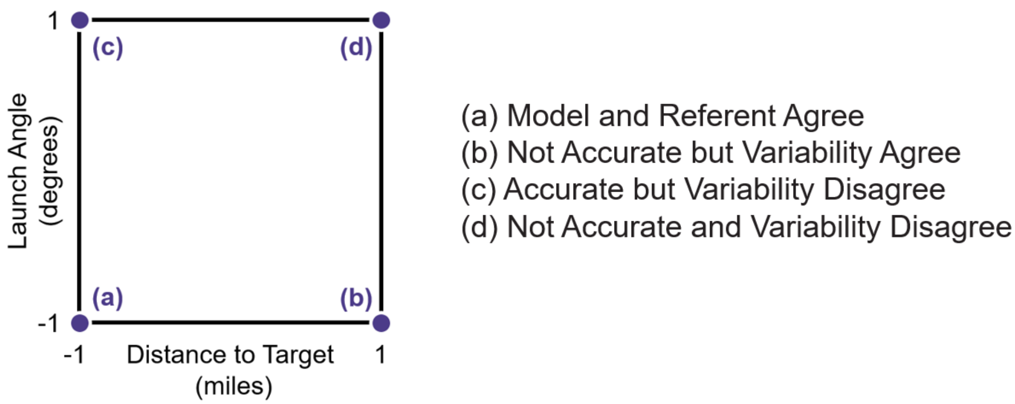
To demonstrate the utility of the fidelity metric, this section walks through its application to a toy problem to determine if a model has sufficient fidelity. Though quite simple, this example is easily extensible to much more complex factor spaces due to the fidelity metric's simple formulation and ease of automation.

Consider a simplified example of an intercontinental ballistic missile system called Next-Generation Long Range Missile, where the response of interest is the time of flight in minutes and the only two factors believed to affect this response are the distance to the target in miles and the launch angle in degrees. Model developers construct a stochastic physics-based simulation of the system. They wish to validate the model using physical data collected from a live-fire test of the system. The live-fire test had a total of 20 runs: four different factor combinations with five replicates each. The test design is pictured in Figure 4 with coded values for the factors. The model developers conducted a similar test of their simulation, but due to computing capabilities, performed 100 replicates at each of the four test points.

For illustration, this toy problem is constructed to demonstrate performance of the fidelity metric in four different potential cases of interest, where each case is observed at one of the four unique input combinations in the test. Figure 4 labels the four test points as they correspond to each of the four cases of interest.

**Figure 4**

*Toy Problem Test Design*




In this toy problem, the fidelity metric allows for the identified fidelity to vary across the factor space in both dimensions of accuracy and variability. The true but unknown parameters of both the physical system and the simulation are given in Table 1.

To compute the fidelity metric based on the collected samples, the model developers compute the mean and standard deviation of the data collected at each of the test points; see Table 2. The fidelity metric must in practice be computed from sample statistics, not from the population parameters, which are usually unknown; therefore, the sample statistics in Table 2 result in fidelities that differ from the theoretical fidelities computed from the population parameters in Table 1.

Additionally, the model developers identify that the simulation response of time of flight should be trusted to within ±1 minute due to numerical computing precision and the epistemic uncertainty in the simulation inputs. This uncertainty corresponds to $\delta = 2$ and allows for calculation of $s_m^*$ as shown in Table 1 for the population and Table 2 for the sample. For the live-fire data, the measurement error directly contributes to the observed variability, and the sample standard deviation of the data set can be considered equivalent to $s_r^*$.

**Table 1**

*True Population Parameters and Fidelities of Physical System and Simulation*

| | Referent | | Model | | | | Fidelity | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\mu_r$ | $s_r^*$ | $\mu_m$ | $s_m^*$ | $\sigma$ | $\delta$ | $f_a$ | $f_v$ | $f$ |
| (a) Model and Referent Agree | 20 | 2.1 | 20 | 2.08 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| (b) Not Accurate but Variability Agree | 25 | 2.1 | 30 | 2.08 | 2.00 | 2.00 | 0.06 | 1.00 | 0.06 |
| (c) Accurate but Variability Disagree | 20 | 3.3 | 20 | 8.02 | 8.00 | 2.00 | 1.00 | 0.43 | 0.43 |
| (d) Not Accurate and Variability Disagree | 25 | 3.3 | 30 | 8.02 | 8.00 | 2.00 | 0.32 | 0.43 | 0.14 |

**Table 2**

*Sample Statistics and Fidelities of Physical System and Simulation*

| | Referent | | Model | | | | Fidelity | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\bar{x}_r$ | $s_r^*$ | $\bar{x}_m$ | $s_m^*$ | $s$ | $\delta$ | $f_a$ | $f_v$ | $f$ |
| (a) Model and Referent Agree | 20.54 | 3.20 | 20.22 | 2.33 | 2.26 | 2.00 | 0.99 | 0.90 | 0.90 |
| (b) Not Accurate but Variability Agree | 25.39 | 2.11 | 29.82 | 2.18 | 2.11 | 2.00 | 0.11 | 1.00 | 0.11 |
| (c) Accurate but Variability Disagree | 21.03 | 3.49 | 19.68 | 7.64 | 7.62 | 2.00 | 0.93 | 0.52 | 0.49 |
| (d) Not Accurate and Variability Disagree | 26.74 | 2.52 | 30.04 | 7.28 | 7.26 | 2.00 | 0.42 | 0.29 | 0.12 |

Comparing Tables 1 and 2, the fidelity metric accurately identifies the fidelity of the four cases, however with a small difference in the calculated fidelity due to the limited sample size collected, which reduces the accuracy of the fidelity metric. Thus, the model developers are able to identify in which regions their model has poor fidelity, and whether the model has poor accuracy or a poor match in variability in those regions. In this case, the model is not accurate at higher distances to the target and fails to match the variability when the launch angle is high. The model developers can then make changes to the model and reassess fidelity until the desired fidelity has been reached.

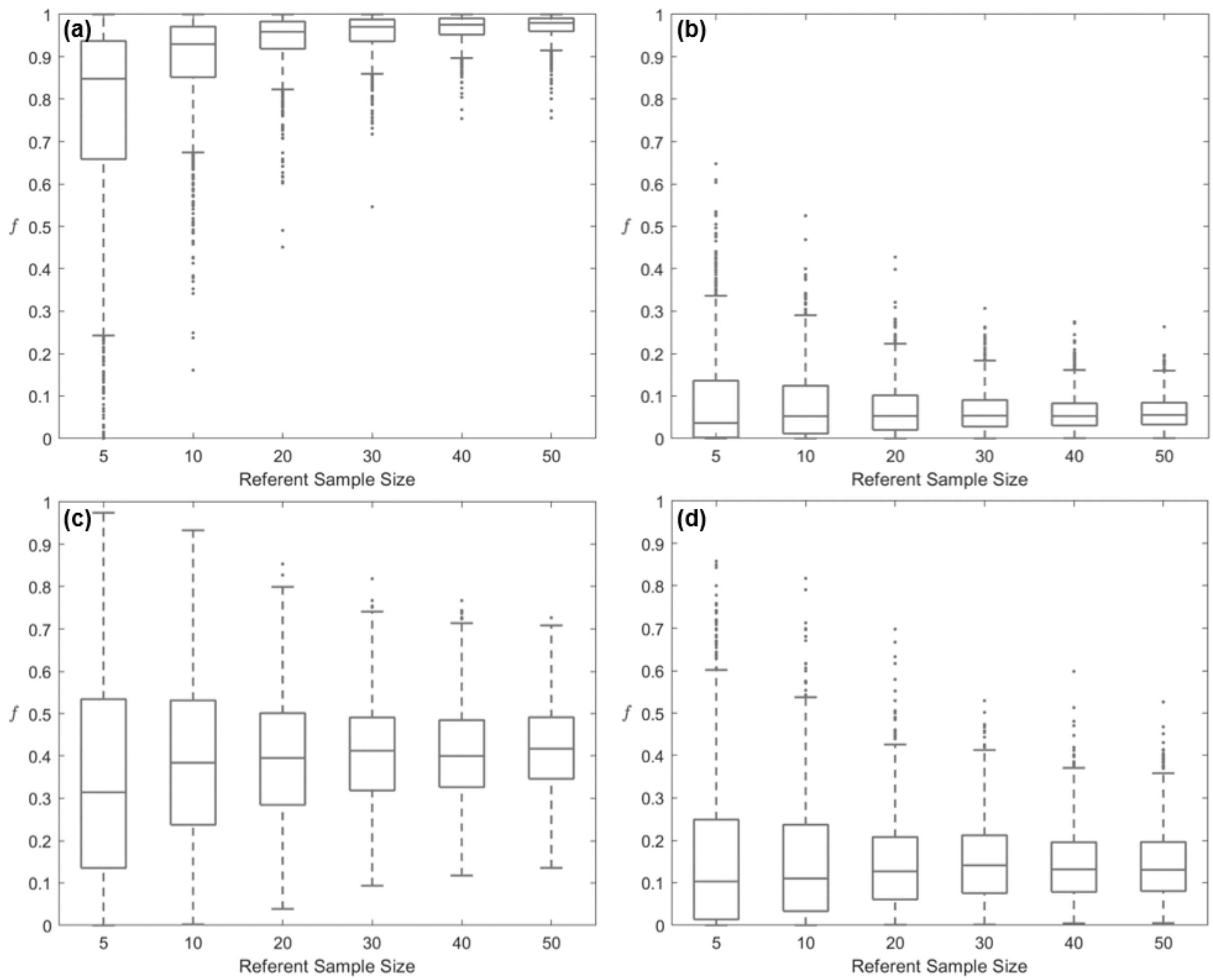
To demonstrate the effect of sample size on the metric, a simulation of 1000 different random samples was performed for increasing sample size of the referent data collected at each test point, from 5 replicates per test point to 50 replicates per test point. The model sample size was held constant at 100 replicates per test point. The referent and model data were generated using a normal distribution with the parameters in Table 1. The results are pictured in Figure 5.

In all four cases in Figure 5, the mean fidelity approaches the theoretical value computed in Table 1 as the sample size increases. The means generally begin lower and increase with increasing sample size toward the expected value, indicating that the fidelity metric scores lower on average when data is limited. Additionally, as the sample size increases, the range of fidelity values becomes narrower. The small ranges observed at high sample size indicate that the metric has a high ability to distinguish between matching and dissimilar distributions; however this ability is lower when less data is available.

To establish a comprehensive view of trust in the model and calculate an MRL, the model developers must consider the authority of the referent used for comparison and whether sufficient fidelity has been demonstrated for the entire factor region of interest. These will be discussed in the upcoming MRL best practice.

**Figure 5**

*Distributions of Fidelity Computed with Varying Referent Sample Sizes*




## Discussion

The STAT COE fidelity metric provides a rigorous and intuitive method to assess the consistency of a model and referent for use of a model in the DE environment. However, for use in the wide range of relevant DoD systems, deterministic systems and models, large factor spaces, limited data availability, and multiple referents may complicate the computation of the fidelity.

In deterministic systems, where no aleatory uncertainty is present, the standard deviation is zero. In this case, the resolution-modified standard deviation defaults to $s^* = \delta/\sqrt{12}$, where the resolution captures all of the uncertainty in the system. For deterministic referents or models, the resolution, and therefore s*, will likely be small if there is a high ability to resolve behavior through measurement or other methods. Due to the structure of the fidelity metric, this results in a stricter requirement for fidelity in deterministic cases than in stochastic cases, which may have a larger standard deviation. While this may not pose a problem in purely deterministic systems, it may pose a problem when utilizing both deterministic and stochastic referents in tandem to validate a model, as the scale of the fidelity metric may differ. Deterministic referents will often be scored with a lower fidelity than stochastic referents due to the increased strictness of the metric, making it more difficult to objectively compare fidelities among a body of different referents.

To include deterministic models and referents, a value for the resolution must be determined. As stated, the resolution is the degree of granularity with which a parameter or variable can be determined (Pace, 2015). In physical testing, lower resolution could result from measurement error; however, measurement error can be difficult to determine in practice. In the case of physical testing, this measurement error contributes to the overall variability, and the epistemic uncertainty is confounded with the aleatory uncertainty: one cannot tell which portion of the error is due to measurement error and which portion is due to the inherent randomness of the system. Therefore, in this case, the sample standard deviation is in fact representative of the total variability and is interchangeable with $s^*$.

However, in many cases, for example in stochastic simulations or deterministic systems, a determination of resolution is required separately from the standard deviation. The method for determining this resolution will vary between different systems. For example, in a stochastic simulation, if the simulation is very precise, the resolution may be negligibly small in comparison to the inherent variability, and the resolution may not need to be determined. However, in deterministic or poor resolution cases, the resolution will non-negligibly contribute to the overall variability and uncertainty quantification methods are needed to determine the overall system variability.

In the DoD, systems operate under wide ranges of several different factor combinations, with several responses available to compare. For these systems, computing the overall fidelity between a model and a referent is not a single computation of the fidelity metric because many distributions exist to compare. The behavior of a response may vary widely with differing inputs, and a model must be validated across the entire operational region. In order to make an unbiased comparison of model and referent, the response must be under the same set of input factor conditions. At each of these input points, the fidelity is assessed to determine a cumulative fidelity across an entire region of inputs.

There are often no replicates at a single set of input conditions, which poses a particular challenge for computing the fidelity metric as no standard deviation can be computed. Given the form of the fidelity metric, when no standard deviation exists, the resolution alone contributes to the variability. However, considering a stochastic system with only a single point at a given input, the resolution alone is not representative of the entire system variability. This presents an additional challenge to the computation of an MRL.

The STAT COE is currently working to incorporate this fidelity metric into the MRL framework, particularly in the presence of limited data. This framework will be discussed in detail in the upcoming best practice on MRLs.

## Conclusion

As the DoD shifts toward a DE approach which more heavily relies on models for decision making, models must be considered trustworthy to minimize any risk introduced by the use of models in place of physical articles. MRLs provide an objective validation framework that can be quickly and repeatedly applied to the wide variety of M&S in the DoD. Additionally, MRLs aim to inject rigor into the existing subjective validation process that relies on a binary indicator of validity, which grants a model validity for its entire lifetime without reassessment. The fidelity metrics presented here provide a straightforward, intuitive method for evaluating fidelity in the diverse M&S environment of the DoD, with a high ability to distinguish between high and low fidelity cases. The fidelity metric together with MRLs will ultimately enable assessment of a model's readiness to be used and trusted in the DE paradigm, with full comprehension of the model's capabilities and risk.

**Key Definitions**

To ensure a common understanding of the subject, the following definitions are used throughout this 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).

aleatory uncertainty: uncertainty arising from an inherent randomness in the properties or behavior of the system under study (Helton, 2011).

epistemic uncertainty: uncertainty derived from a lack of knowledge about the appropriate value to use for a quantity that is assumed to have a fixed value in the context of a particular analysis (Helton, 2011).

fidelity: the level of consistency between a model and a referent, defined in the three dimensions of accuracy, repeatability, and resolution.

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 trusted representation of reality.

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.

trust: to rely on the truthfulness or accuracy of (Merriam-Webster).

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).

## References

Ahner, D. K., Jones, N., Key, M., Adams, W., Burke, S., & Weeks, C. (2021). *A Conceptual Framework for the Establishment of Model Readiness Levels.* White Paper, Scientific Test & Analysis Techniques Center of Excellence (STAT COE).

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.

Helton, J. C., & Johnson, J. D. (2011). Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty. *Reliability Engineering & System Safety*, *96*(9), 1034–1052. https://doi.org/10.1016/j.ress.2011.02.013

Merriam-Webster. (n.d.). Trust. In Merriam-Webster.com dictionary. Retrieved June 10, 2022, from https://www.merriam-webster.com/dictionary/trust

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

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

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.