

M&S for Test and Evaluation - Guidance

The Modeling and Simulation (M&S) sections of the TEMP should address how M&S will be employed in the overall test strategy and how the M&S will be verified, validated, and accredited (VV&A). Specifically, the TEMP should list any M&S expected to be used, the intended use, the data requirements, the test objectives to be addressed and/or how test scenarios will be supplemented with M&S, the planned VV&A effort, and who will conduct the VV&A effort ([DoDI 5000.61](#)). The TEMP should list any specific test events required for VV&A of the M&S. The resources for the VV&A test events should be included in Part IV.

DOT&E requires all OT&E and LFT&E test agencies to accredit models used to support OT&E and LFT&E. The accrediting test agency will establish the acceptability criteria for M&S use, and the accreditation must be based on a verification and validation approach that is tailored for the specific intended use of the model or simulation. This means that the OTA will conduct their own assessment to accredit M&S for their use in OT. DOT&E must review and concur with the OTA's accreditation plan before the plan is executed.

Testers should focus the validation on the full system or environment being evaluated in addition to validating sub-components of the model. While validating every subcomponent that comprises the full system is ideal, it is more important to ensure that the integration of all the components and the environment together adequately represents the real-world system relevant to the intended use of the model.

Before using M&S results to support an operational test or live fire evaluation, it is essential to understand and characterize the usefulness and limitations of the M&S capability. Thus, any M&S used to support OT&E should not be accredited until a rigorous comparison of live data to the model's predictions is done (if possible), and those predictions are found to have replicated live results with sufficient accuracy for the intended evaluation in the intended domain. In addition to these direct quantitative comparisons between live data and M&S output, a comprehensive strategy should also assess M&S output across the entire operational domain for which the M&S will be accredited.

Validation of M&S should include the same [rigorous statistical and analytical principles](#) that are used to design live tests. The principles and techniques that comprise statistical test design methodologies, including design of experiments and other formal statistical tests, should be employed as part of the process of determining what and how much live data are needed for a live versus simulation comparison, and in the process of determining how well the models/simulations reflect reality. Empirical models should be used to understand M&S outcomes across the operational space and assist in the uncertainty quantification in areas where there are no live data. All results, but especially those in the operational space where no live data is available, should be discussed in the context of limitations.

Statistical techniques (e.g. hypothesis testing and regression techniques) should also be used to rigorously compare live data to simulated output. This methodology can and should be supported by

other elements of validation, including face validation, documentation review, SME evaluation, and comparison to other models. If there are extraordinary circumstances prohibiting these statistical principles from being used, the reasons why they are not used must be clearly articulated and alternative approaches used to justify validation and accreditation must be cogently explained.

As stated in the [March 14, 2016](#) and [January 17, 2017](#) memorandums, a discussion of the following elements is required:

- [Quantitative mission-focused measures](#)
- The range of conditions over which the M&S will be validated
- The plan for collecting the necessary live and simulation data for M&S validation, including both the points to be used for direct comparison, and those used to evaluate the entire M&S space of interest
- An analysis of statistical risk
- The validation methodology

This information can be presented directly in the TEMP, or the TEMP can reference other relevant documents, such as validation or accreditation plans. If the necessary detail is not yet available while the TEMP is being drafted, the information may be conveyed to DOT&E via a standalone M&S concept briefing (similar to a Test Concept) or in an Operational Test Plan, so long as the overall strategy is socialized with DOT&E as early as possible.

In addition to applying rigorous statistical and analytical principles, some other important criteria for M&S accreditation are:

- Documentation which summarizes the purpose, development background, assumptions, and application domains and provides a complete and accurate description of M&S capabilities and limitations.
- Sound approaches for M&S capability acquisition, validation, and use. M&S capabilities used for T&E should be planned and resourced early. The M&S capabilities to be used, the T&E aspects of the system evaluation that these M&S capabilities will address, and the approach for assessing credibility of these models and simulations should all be described in the TEMP.

Establishing M&S Credibility for T&E: Additional Details and Definitions

Under DoDI 5000.61, each M&S capability must complete a verification, validation, and accreditation (VV&A) process to establish its credibility for a specific intended use. Some M&S capabilities associated with T&E have special validation requirements. For example, to validate that a non-US forces or threat weapon is appropriately represented in a model, the Director, Defense Intelligence Agency is the final validation authority for oversight systems. DOT&E, through the T&E Threat Resource Activity (TETRA), is the approval authority for threat representation validation reports used for T&E. OTAs accredit threat representation models for use in OT. The Defense Acquisition Guidebook, Section 9.7.3, Validation of Threat Representations (targets,

threat simulators, or M&S) provides guidance and references on validating M&S capabilities associated with threats and targets.

Existing M&S capabilities previously accredited for other applications must complete another VV&A process and be accredited for each new intended use. However, previous VV&A may simplify the process because the previous efforts have been documented and the new VV&A effort typically can focus on the changes.

Verification determines whether the M&S accurately represents the developer's specifications. The M&S is expected to add two numbers; does it add two numbers? Validation determines whether the model is an accurate representation of specific aspects of the real world or threat system. The M&S is expected to add two numbers; does it provide the correct sum? Accreditation is the official certification that the M&S and its associated data are acceptable for an intended use.

For accreditation, the intended use is important because an M&S capability useful in one application may not be useful in another due to limitations inherent in the M&S capability, existing validation data, or a prior VV&A process. The accreditation will explicitly state the intended use, such as: "The Big Weapon Model will be used to estimate the miss distance between the weapon and the target in support of developmental test DT-II." It also should acknowledge any significant limitations: "The Big Weapon Model does not include threat countermeasures, and consequently all scenarios are simulated in a clear environment."

The scope of the accreditation effort and VV&A process are functions of how each M&S capability will be used. For example, high level or conceptual models are often used early in a program (e.g., a spreadsheet model used to estimate system performance) that require limited data for validation and accreditation. Frequently, M&S capabilities used in prior similar programs can be used and pre-existing VV&A artifacts and analysis can simplify or streamline the VV&A process for the new application. At the other extreme are high-fidelity models an evaluator might use to assess measures of effectiveness, suitability, or survivability; these must undergo a rigorous VV&A process. In general, the more important the M&S results are to the final evaluation, the more rigorous the VV&A process must be. Wherever possible, design of experiments techniques should be leveraged to ensure that test data supporting the VV&A clearly defines the performance envelope of the model or simulation, and corresponding statistical analysis techniques should be employed to analyze the data and identify factors that influence the validity of the M&S.

Some common pitfalls in using M&S for T&E that should be avoided are:

- Faulty assumptions in developing or using M&S such as assuming independence between events that actually have some type of dependency or relationship.
- Using M&S results outside their validation domain which are uncharacterized and include unknown uncertainties.

- Improper use of data for M&S development or validation such as relying solely on heart-of-the-envelope performance data or using specification values instead of actual performance data when the latter is available.
- Averaging validation results across conditions rather than discussing where the M&S is valid and where it isn't.

References

[DODI 5000.61: DoD Modeling and Simulation \(M&S\) Verification, Validation, and Accreditation \(VV&A\), 9 December 2009](#)

[Guidance on the Validation of Models and Simulation used in Operational Test and Live Fire Assessments, 14 March 2016](#)

[Clarifications on Guidance on the Validation of Models and Simulations used in Operational Test and Live Fire Assessments, 17 January 2017](#)