OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

OPERATIONAL TEST
AND EVALUATION

JAN 17 2017

MEMORANDUM FOR COMMANDING GENERAL, ARMY TEST AND EVALUATION COMMAND
COMMANDER, OPERATIONAL TEST AND EVALUATION FORCE
COMMANDER, AIR FORCE OPERATIONAL TEST AND EVALUATION CENTER
DIRECTOR, MARINE CORPS OPERATIONAL TEST AND EVALUATION ACTIVITY
COMMANDER, JOINT INTEROPERABILITY TEST COMMAND

SUBJECT: Clarifications on Guidance on the Validation of Models and Simulation used in Operational Test and Live Fire Assessments

The purpose of a rigorous verification, validation, and accreditation process is to ensure the modeling and simulation (M&S) works and to what degree it represents reality from the perspective of the intended use of the model. As I have previously emphasized (see memo dated 14 March 2016) the validation strategy should focus on the quantitative comparison of live data and M&S outcomes using statistical methods. In addition to those quantitative comparisons, a comprehensive strategy should assess M&S output across the entire operational domain for which the M&S will be accredited. Statistical analysis should be used to conduct sensitivity analysis and subject matter experts should review outcomes for consistency with reality.

In other words, not only should the simulation runs that match live test conditions be considered in the accreditation, but additionally, M&S runs that span the entire operational domain of the M&S should be used in the accreditation decision. This larger set of M&S runs may consist of the actual runs that will be used for the evaluation of effectiveness, suitability, lethality, or survivability (aka runs for the record). However, they may also differ from the final evaluation runs if updates are necessary based on the comparison to live data or other problems identified in the review of the M&S outcomes.

There will always be cases when live data for quantitative comparisons are unavailable (e.g., a new threat for which a surrogate has not yet been developed). In those instances, a well-reasoned and cautious approach should be taken to determine what, if any information, may be gleaned from M&S. In some instances, the absence of live data may prevent the accreditation of the M&S for use in the operational space. In other instances, it may reasonable to conclude that performance in one area of the operational space extends into a nearby operational space, where no live data are available. In the latter case, it is critical that the limitations of the M&S are understood and the uncertainty in the results quantified to the extent possible. Empirical models (a.k.a., emulators or meta-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. In the operational space where no data are available, the results of the M&S should be discussed in the context of limitations.

Validation (and therefore the data required for validation) is a sequential process. There is no cookbook approach; and in each phase there are numerous correct methodologies for determining what information should be collected and how it should be analyzed to support the validation and accreditation.

At a minimum the information that should be included in a rigorous validation and accreditation are:

- A strategically selected set of test points designed to compare the M&S and live data. These points should be selected using the principles of experimental design to span as much of the operational space as possible within the constraints of what is feasible to conduct in live testing. See my guidance on design of experiments (DOE) and M&S for more information on selecting a set of comparison data.

- Analysis methodologies to empirically model the live and M&S outcomes and test for statistical differences between the two outcomes.

- A robust design for the M&S that systematically covers the range of operationally realistic inputs over which the model will be accredited. Space-filling design methodologies are preferred because they not only maximize opportunities for problem detection, but also support the development of statistical emulators that can be compared to live data and assist in quantifying uncertainty in the M&S.

All of the above should be complete including the analysis of all outcomes before the final accreditation of the M&S. I would also like to re-emphasize that validation is a complex process and there are many important elements not discussed here, including documentation review, face validation, subject matter expert (SME) evaluation, and comparison to other models. While these processes typically are not statistical in nature, they provide useful information and can continue to be used in conjunction with statistical modeling.

Additionally, it is important to quantify uncertainties associated with the M&S to the extent possible. I welcome the inclusion of additional statistical methods outside of the limited required information highlighted above to include Monte-Carlo simulation and probabilistic methods that combine live data with M&S-based emulators. These additional tools can support a variety of important activities including uncertainty analysis and error propagation, sensitivity analysis, and data assimilation.

I have observed progress in the past year in the level of rigor employed in M&S validation. My office has started a website that highlights good examples and captures resources that we will continue to update as we continue to improve our methods for validation and

accreditation. My point of contact on this matter is Dr. Catherine Warner. She can be reached at (703) 697-3655.

J. M. Gilmore

J. Michael Gilmore
Director