Guidance Notes On

Verification and Validation of Models, Simulations, and Digital Twins

ABS

November 2024

GUIDANCE NOTES ON

VERIFICATION AND VALIDATION OF MODELS, SIMULATIONS, AND DIGITAL TWINS
NOVEMBER 2024

American Bureau of Shipping
Incorporated by Act of Legislature of
the State of New York 1862

ABS Plaza
1701 City Plaza Drive
Spring, TX 77389 USA

Foreword

In recent years, there has been an increase in the development of digital solutions by Owners to support decision-making for the design, construction, operation, and maintenance of vessels and offshore structures. These digital solutions are often comprised of models and simulations. Additionally, increased connectivity and asset-specific data are enabling the configuration of interconnected models and simulations to create digital twins as part of these digital solution offerings. In many instances, these new digital solutions have no service history in the proposed application or environment, and it is often recommended or required that sufficient documentation of these digital solution's credibility be completed and provided so that decision-makers can make informed decisions about their use for specific intended purposes.

These Guidance Notes provide a recommended framework for the verification and validation of models, simulations, and digital twins. This framework is derived from existing industry codes and best practices and provides guidance and best practices for a robust verification and validation program which consists of established policies, processes, and documentation, that can be tailored to each individual use case rather than a prescriptive set of requirements. Additionally, these Guidance Notes introduce the concept of model, simulation, and/or digital twin criticality which acknowledges that the risk level of different model, simulation, and digital twin applications varies and thus the corresponding risk mitigation offered by a verification and validation approach can be tailored to the application risk level. The results of the verification and validation processes are then considered as part of a more extensive credibility assessment.

The framework for the verification and validation of models, simulations, and digital twins presented in these Guidance Notes can be used to support the application of the ABS *Rules for Alternative Arrangements, Novel Concepts and New Technologies (Part 1D)*, the *Guide for Smart Functions for Marine Vessels and Offshore Units*, and the *Requirements for Autonomous and Remote Control Functions*. An illustrative example of the documentation resulting from the implementation of this framework is presented in the Appendices.

This document also provides a foundational ontology for models, simulations, and digital twins to generalize and characterize discussion of these topics. This is further supported by a glossary provided in the Appendices consisting of standard terms and definitions, along with alternative terms that may also be utilized by the industry with the same intended meaning.

ABS welcomes comments and suggestions for improvement of these Guidance Notes. Comments or suggestions can be sent electronically to rsd@eagle.org

Terms of Use

GUIDANCE NOTES ON

VERIFICATION AND VALIDATION OF MODELS, SIMULATIONS, AND DIGITAL TWINS

CONTENTS

1 Introduction

1.1 Purpose

The purpose of these Guidance Notes is to:

i) Establish an ontology and common lexicon for models, simulations and digital twins and their development lifecycles.

ii) Introduce a risk-informed approach which categorizes the criticality of the model, simulation, and digital twin respective to their intended use.

iii) Provide a recommended framework for the verification and validation of models, simulations, and digital twins which includes guidance for establishing the organization, processes, and documentation for a robust verification and validation program tailored based on the evaluated criticality.

iv) Present an approach to support the decision makers in assessing the credibility and acceptability of models, simulations, and digital twins for an intended use case or application.

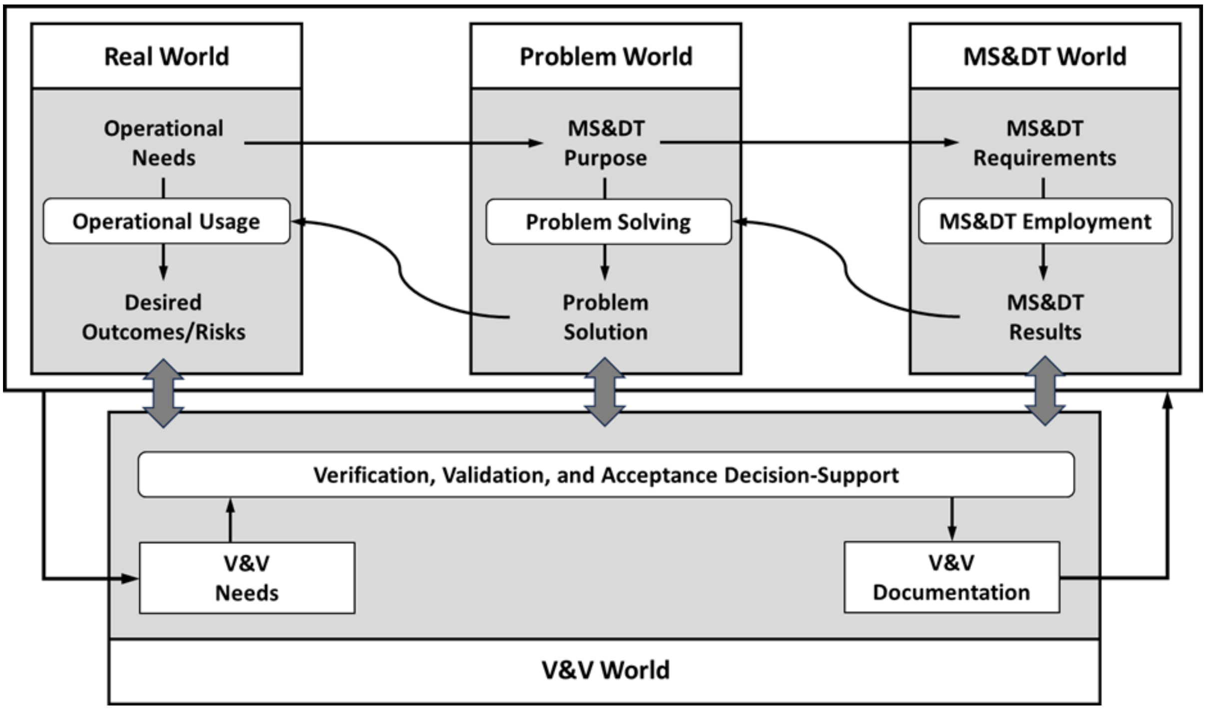
The technical guidance presented herein consists of a conceptual framework, implementation best practices, and guidance for tailoring verification and validation (V&V) activities for specific applications. This framework is rooted in established standards and other related best practice documentation. It is important to note that V&V is a risk mitigation approach that should align with the use case(s) for which it is being employed. Therefore, this document presents an overall framework and recommended best practices for the development of a robust V&V program which consists of established policies, processes, and documentation, that can be tailored to each individual use case rather than a prescriptive set of requirements. Associated activities such as accreditation, certification, and test and evaluation are outside the scope of this methodology.

1.3 Model, Simulation, and Digital Twin Problem Solving Approach

The basic premise of the conceptual framework presented in these Guidance Notes is that models, simulations, and digital twins (MS&DT) are part of a problem-solving process. MS&DT are developed to fulfill a specific purpose in solving one or more problems and then employed by their end users in operation to obtain desired outcomes. The degree of success of such MS&DT in obtaining the desired outcomes of the end user depends on how well they are specified, designed, developed, integrated, tested, used, and supported. When the MS&DT problem-solving process is properly executed, the resulting solution should satisfy the original needs with a minimal level of risk. Risk here is defined as the probability of realizing unsuccessful or undesirable results using MS&DT.

Generalizing this premise into a generic process view of MS&DT problem solving, a multiple world view is presented which categorizes the elements of the problem-solving process that occur in the Real World, the Problem World, and the MS&DT World [6]. The process begins with an operational need which exists in the Real World. This operational need is transformed into a MS&DT purpose statement in the Problem World. Based on the MS&DT purpose, requirements for the MS&DT are defined and the resulting system is then developed and employed in the MS&DT World. This MS&DT-based solution is transferred back to the Real World where it is then used in the operational environment to achieve the targeted outcomes. This multiple world view of MS&DT problem solving is illustrated in the top half of Section 1/Figure 1.

FIGURE 1
Interfacing of Multiple Worlds View of Problem Solving and V&V World [6]



The intent of this generic problem-solving process flow is to serve as a common basis in which V&V for MS&DT (e.g. concepts, principles, processes, products, and techniques) can be understood, developed, and applied. Note that this conceptual framework is not intended to be prescriptive or replace alternative, more detailed MS&DT life cycle and process implementations such as the Federation Development and Execution Process (FEDEP) or Distributed Simulation Engineering and Execution Process (DSEEP). Such concrete implementations can be considered as tailored instances of this generalized view.

1.5 Verification and Validation Framework

The key challenge that exists with the use of MS&DT in a problem-solving approach is that it is not possible to demonstrate with absolute certainty that the MS&DT system or results will meet the Real World needs prior to its actual use. Consequently, there is always a possibility that the MS&DT-based solution is not successful when used in the Real World. Such a failure could result in an undesirable impact (i.e., a realized risk) on the operational environment. Therefore, a MS&DT system or result is only acceptable if the authority responsible for employing the MS&DT has sufficient confidence that the use of the MS&DT system or result satisfies the Real World needs without posing unacceptable risks.

In these Guidance Notes, the organizational role that has the authority to make this decision is referred to as the MS&DT Technical Authority. The MS&DT Technical Authority is the official role responsible for oversight and the final acceptance decision regarding the use of the MS&DT for its intended use case. This decision should be based on the credibility of the MS&DT based in part on the outcomes of the V&V activities. The MS&DT Technical Authority role may be a single person within the organization, or there may be many MS&DT Technical Authority individuals assigned to various MS&DTs or specific use cases utilizing one or more MS&DTs (i.e. projects or programs). This role may also overlap with other organization roles involved in the development of the MS&DT and/or the performance of V&V activities.

Note that the MS&DT acceptability decision is relative to different MS&DT Technical Authorities, such that what is acceptable to one MS&DT Technical Authority for one use case may not be acceptable for another MS&DT Technical Authority employing the same MS&DT for a different use case. The MS&DT Technical Authority's decision-making process therefore requires appropriate evidence-based arguments to justify their acceptance decision. Thus, the framework described in these Guidance Notes provides best practices to collect, generate, maintain, and reason with a body of evidence in support of the MS&DT Technical Authority's acceptance decision-making process.

More specifically, evidence is gathered to answer two key questions:

1) "*Did we build the MS&DT system right (verification)?*": The response to this first question helps to establish the MS&DT Technical Authority's confidence in whether the MS&DT system or result is built right (i.e., MS&DT correctness).

2) "*Did we build the right MS&DT system (validation)?*": The response to this second question helps to establish the MS&DT Technical Authority's confidence as to whether they have built or procured the right MS&DT system or result for the intended uses (i.e., MS&DT validity).

V&V activities can be considered a specific problem domain of the MS&DT problem solving approach with its own needs, objectives, and issues. This domain can be referred to as the V&V World [6] and is illustrated in the bottom half of Section 1/Figure 1.

The V&V World groups the products, processes, and organizational aspects that are needed to develop the evidence required to support the MS&DT Technical Authority in their acceptance decision procedure(s). Note that in this framework, the acceptance decision is always the responsibility of the MS&DT Technical Authority, and the decision procedure(s) may involve trade-off aspects beyond the V&V effort scope contained herein.

1.7 Scope and Overview

These Guidance Notes introduce the conceptual framework, implementation best practices, and guidance for tailoring verification and validation (V&V) activities for specific applications. This guidance is presented for instances where V&V activities are to be performed by the stakeholder responsible for MS&DT development and independent V&V activities performed by a stakeholder not directly involved in the development of the MS&DT. The remaining sections in these Guidance Notes provide an overview of MS&DT, considerations for the management of V&V, recommended practices for performing V&V and methods for establishing the criticality and credibility of MS&DT use. The ordering of these topics within the Guidance Notes is indicated in Section 1/Figure 2. Following the completion of the activities outlined in each section, the user will be able to answer the questions on the right half of the figure and be able to generate the supporting documentation that demonstrates how those questions were addressed.

Definitions for models, simulations, and digital twins are provided as background prior to outlining the recommended V&V process. An ontology for digital twins is presented in Section 2 to help establish the relationships between the entities that compose the digital twin. Model life cycle phases are discussed to help identify when V&V needs will occur.

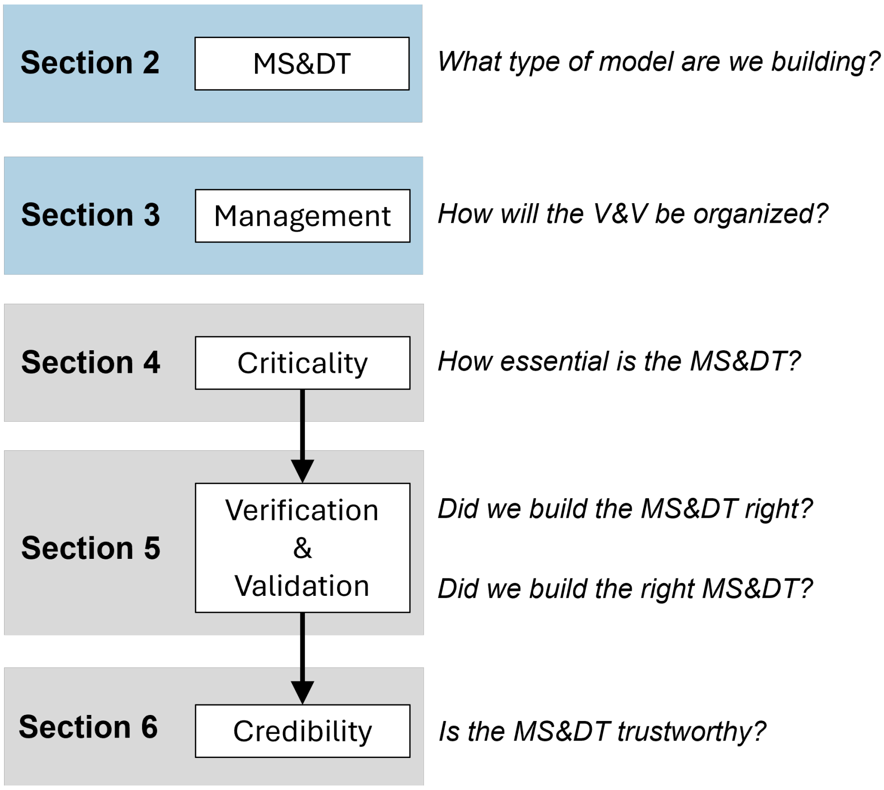
Management of V&V activities is necessary to ensure that they are conducted in an efficient and effective manner. To guide the management of V&V activities, focus is placed on establishing the context of the V&V effort, identifying stakeholders, planning at different organizational levels, and establishing the V&V timeframe.

The V&V process begins with establishing the criticality of the MS&DT as it helps define the required rigor of subsequent V&V efforts. The criticality of MS&DT is a function of the degree to which it influences decisions and the nature of the consequences of those decisions.

The core V&V processes follow a claim-evidence-reasoning approach whereby a final acceptance claim for the MS&DT can be made if there is sufficient supporting evidence to state with adequate certainty that the MS&DT satisfies the established acceptance goal. In general, the acceptance goal is decomposed to specify acceptance criteria and plan the individual V&V tasks required to generate the evidence solutions. The items of evidence are subsequently compiled to support the acceptance claim. This process is accomplished in three distinct stages: 1) concept validation, 2) verification, and 3) results validation.

The final activity is to assess the credibility of the MS&DT which is important to communicate to the users of the MS&DTs. By categorizing the credibility of the MS&DT, decision makers have a reference for weighing the level of trust they place on the MS&DT.

FIGURE 2
Overview of Guidance Notes Content



Appendices provide references to other V&V resources and standards, a glossary of common terms related to MS&DT and V&V, guidance for V&V activity documentation, and a representative case study.

SECTION 2

Models, Simulations & Digital Twins

1 Introduction

Models, simulations, and digital twins each address problems posed by operational needs in the Real World in their own way. The purpose of this section is to establish a common understanding of models, simulations, and digital twins in the context of these Guidance Notes and establish their distinguishing features. Definitions for models, simulations and digital twins are provided in Appendix 2 as these terms may have different meanings in other industries.

3 Models and Simulations

Models and simulations have a longer history of use in industry and are generally well understood terms. This section also introduces the concepts of model federations and surrogate models that have become more common as models and simulations have grown in complexity and their use has increased. Finally, the life cycle of models and simulations is discussed to provide context for the timing of V&V needs.

3.1 Models

These Guidance Notes define a model as a representation (physical, mathematical, or otherwise logical) of reality that captures key characteristics or behaviors of the selected system, entity, phenomenon, or process, including the data that is incorporated into the representation. Models are an abstraction of reality which require assumptions and simplifications for a variety of reasons, including but not limited to:

a) An exact representation is not practical because:

- *1)* Exact knowledge about the real-world system is incomplete and uncertain.
- *2)* Details are not sufficiently characterized to be included in the model.
- *3)* All possible variations of the subject real-world system cannot be reasonably included.
- *4)* The model would exceed the limits of the computational platform.

b) An exact representation is not desirable because:

- *1)* Added fidelity adds cost and complexity.
- *2)* Adding unnecessary detail detracts from focus of the analysis.

c) An exact representation is unwieldy because:

- *1)* The real-world system of interest may be extremely small and scaling the model up makes it more readily understood.
- *2)* The real-world system of interest may be extremely large and scaling the model down makes it more readily understood.

The decisions regarding how reality is to be abstracted for the creation of the model should be aligned with the model's intended purpose. This is relevant to V&V activities since models and simulations are not exact representations of reality and therefore do not produce exact or perfectly representative results. The limitations and imperfections of the model must be clearly evaluated and understood by the users of the MS&DT to support their development and acceptance of its use in its intended application.

3.3 Simulations

A simulation is defined as the imitation of the operation of a Real-World process or system over time. Simulations require the use of models, and the simulation represents the evolution of the model(s) over time. Simulations enable the behavior of the Real-World process or system to be predicted and are vital when the transient response is desired, not just a steady-state solution. The use of simulations is also valuable for complex systems where system-level behavior is governed by several models and may involve multiple domains.

3.5 Model Federations

In many instances, a model may be constructed from multiple sub-models which is referred to in this document as a model federation, where the individual sub-models and the integrated sub-models are all considered to be "models" per the definition in Section 2/3.1. There are several reasons to construct more complex models this way, including taking advantage of existing models and the benefits of modularity. Model federations may also be the result of the existing technical infrastructure and standardization which necessitates or has resulted in the siloing of models and corresponding data.

Model federations consist of both stored representations and computational representations [29]:

- *Stored representation*: Structured information (in databases, Computer-Aided Drafting (CAD), Computer-Aided Manufacturing (CAM), Building Information Modeling (BIM), Geographic Information System (GIS), point clouds, IoT streams and history, etc.) representing states and attributes of entities and processes at one or more times.
- *Computational representation*: Typically representing the functional relationship between a set of inputs and outputs contingent on a set of reference data.

The interaction between sub-models in a model federation can be unidirectional, or a complex multi-path network of interactions. In either case, the sub-models and the integrated model federation both should be clearly documented and tested.

3.7 Surrogate Models

In the domain of computational models and simulations, the term surrogate model refers to models constructed in a manner similar to that used to construct empirical models, where data from observations are used as the basis for approximating the relationships between independent and dependent variables. For surrogate models, the observations are replaced by data obtained from another model that the surrogate model is intended to replace. Surrogate models are typically developed as a more computationally efficient approach than the original model. However, surrogate models not only take on all the assumptions and simplifications of the models they are based on, but also incorporate additional limitations from their specific implementations that need to be considered.

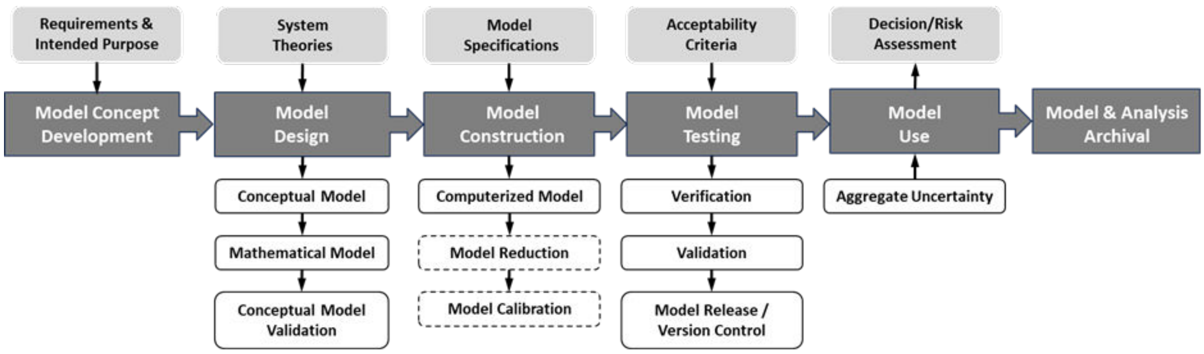
3.9 Model and Simulation Lifecycle

Like a project lifecycle, the development and use of models and simulations can be understood to follow a lifecycle as well. The principal purpose of establishing a model and simulation lifecycle is to support the management of the timeline of activities that occur during the model and simulation lifecycle.

Note that while the phases of the model and simulation lifecycle are similar to the phases of the project or operational system lifecycle, the model and simulation lifecycle phases rarely occur in parallel with the operational system lifecycle phases. Because models and simulations can inform decisions in any phase of an operational system lifecycle, an entire model and simulation lifecycle may occur within a single operational system lifecycle phase.

While there are several variations of the model and simulation lifecycle, these Guidance Notes align with the lifecycle phases proposed by the NASA Handbook for Models and Simulations [1].

FIGURE 1
Model and Simulation Lifecycle Phases



A brief description of each phase is given in Section 2/Table 1. It should be noted that although the phases are presented here in discrete, sequential order, in practice the distinction between phases and the execution of the corresponding activities may not be as clear. These activities often occur in iterative cycles both within each phase and between phases.

Additionally, while the Model & Analysis Archival phase is placed at the end of the lifecycle, the expectation is that the model and simulation, as well as any key artifacts, are expected to be archived throughout development and use. The need for this is even more evident in instances where the model and simulation lifecycle is ended before the completion of the project lifecycle. Its position at the end of the life cycle phases is to maintain consistency with the presentation of the lifecycle phases and to emphasize that the completion of activities within the development and use phases should conclude with the archival of the necessary artifacts.

TABLE 1
Model and Simulation Lifecycle Phase Descriptions [1]

PHASE	*NAME*	*DESCRIPTION*
A	Model Concept Development	This phase focuses on translating the model and simulation purpose from the Problem World to the MS&DT World. The scope of the Real World to be modeled is defined and the process of gathering the relevant information about the scoped Real World is started.
B	Model Design	This phase is a typically iterative process for generating the requirements and specifications for the model construction. System theories are used to abstract from the Real World to form a conceptual model as well as a sometimes-accompanying mathematical model. The resulting conceptual model should be validated prior to beginning model construction.
C	Model Construction	This phase encompasses the activities associated with the creation of a usable computerized model as defined by its requirements, specifications, and intended purpose. Some optional activities in this phase may include model reduction and model calibration.

PHASE	*NAME*	*DESCRIPTION*
D	Model Testing	This phase entails the gathering of evidence to support the claims that the model fulfills its requirements and intended purpose. Specific activities for model verification and model result validation are performed here. This phase also includes model release and version control, establishing the controlled version of the model and associated documentation to support its use.
E	Model Use	This phase includes the assessment by the MS&DT Technical Authority as to the acceptability of the model for an intended use. After approval, this phase also includes the activities of integrating the model into the process or operational system, running the model, gathering and post-processing the output and assessing and reporting the results.
F	Model & Analysis Archival	This phase contains activities associated with storing and managing the model and simulation and its resultant artifacts from the life-cycle phases.

5 Digital Twins

The digital twin concept is more recent than traditional models and simulations. The notion of a digital twin serving as a counter part to a physical asset throughout its lifecycle is of interest to the marine industry as marine and offshore assets are expected to have service lives measured in decades. This section provides a digital twin definition and presents an ontology to help establish the relationships between the entities that compose the digital twin.

5.1 Definition

Digital twins are defined in this document as "*a virtual representation of a physical asset, along with its environment and processes, comprised of integrated models that are updated through the exchange of information.*"

The definition necessitates that a digital twin is comprised of three principal elements:

- *Physical asset:* the digital twin should be representative of a specific physical asset that exists in the physical world. The scope of the physical world that is represented by the digital twin is based on the intended use of the digital twin and may be as simple as a single component or as expansive as a full asset with all its integrated systems and sub-systems.
- *Virtual representation*: the virtual representation is comprised of models and visualizations that are combined in meaningful ways to achieve specific outcomes or support specific decision-making. Models are defined here as either stored representations (structured information representing states and attributes) or computational representations (functional relationships between a set of inputs and outputs).
- *Information exchange*: a digital twin requires that information is exchanged periodically from the physical asset to the virtual representation and the inverse. The physical-to-virtual connection allows data collected from the physical asset to be used to update the states maintained in the virtual representation. The virtual-to-physical connection is the process that results in the transfer of information from the virtual representation back to the physical asset.

In addition to the principal elements presented above, digital twins are considered to be dynamic entities, in that the update of the virtual representation resulting from the information exchange is expected to be periodic over the time in which the digital twin is employed. However, the frequency of these periodic updates is based on the intended use case and its decision interval, which is the time from when the data is received to the time in which a decision must be made.

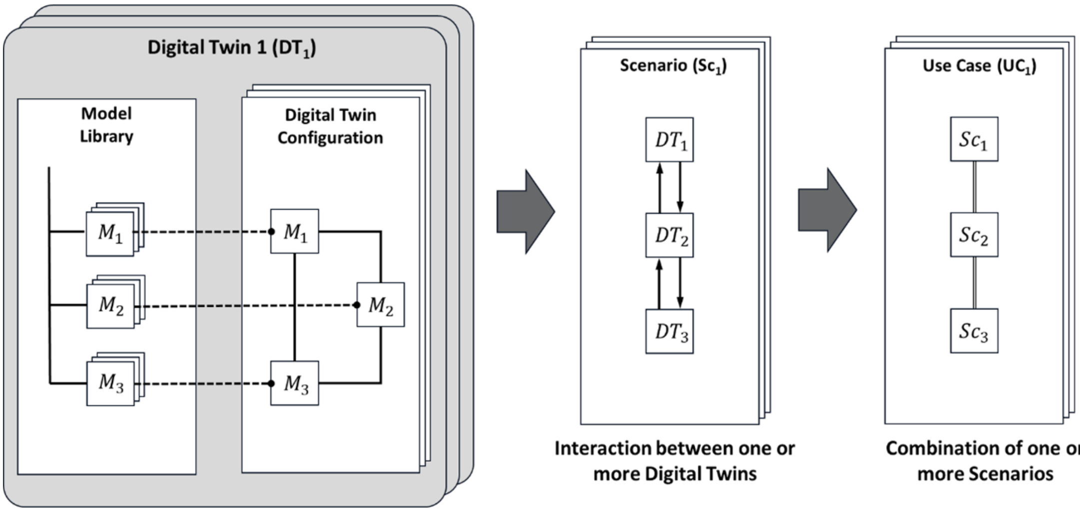
As noted in the first principal element, a specific physical asset is required which exists in the real world from which information is being exchanged with its virtual representation counterpart. However, it is acknowledged that the models and simulations utilized within the virtual representation may exist and be integrated prior to the existence of the physical asset (such as those used during asset design). This federated set of models that exist prior to any information exchange with a specific physical asset may be referred to as a "digital twin prototype." Once the physical asset exists and begins exchanging information with the virtual representation, that virtual representation becomes a "digital twin instance" associated with the specific physical asset.

5.3 Digital Twin Ontology

This section introduces a framework for describing the elements of the digital twin virtual representation in a format that supports the execution of V&V activities. This is necessary as it is practically impossible to verify complex system models that accumulate vast amounts of engineering data without well-defined structures and processes to manage this information in a structured way.

Section 2/Figure 2 provides a top-level view of the framework defining three main entities: a set of digital twins, a set of digital twin scenarios, and a set of use cases [28].

FIGURE 2
Top Level Digital Twin Framework [28]

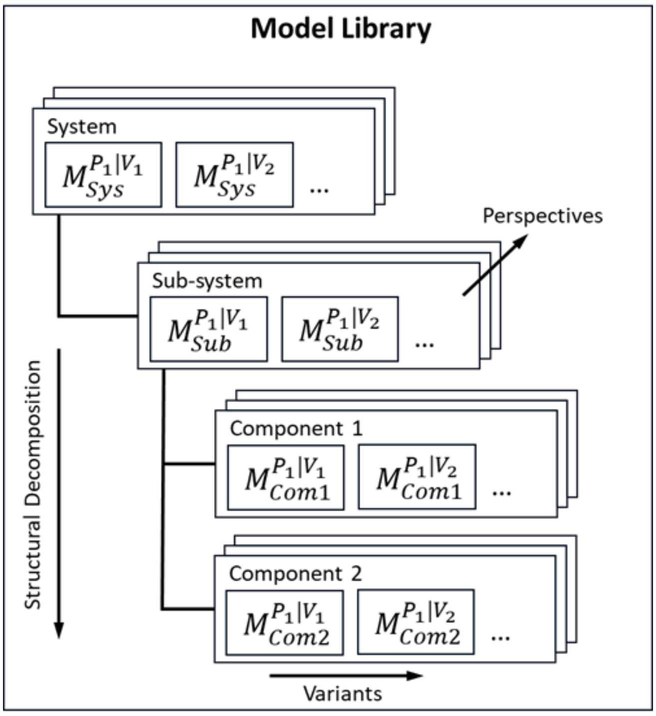


In this framework, digital twins are described as being composed of two components: a model library and a collection of digital twin configurations.

The model library shown in Section2/Figure 3 provides the basic model elements for the asset and its associated environment and processes to be represented by the digital twin [28]. It is useful to organize the model library into categories, typically one category for software components and another for hardware components. The models for hardware components can be further categorized across three dimensions:

- *Structural hierarchy*: This dimension provides alignment of models in the library with physical system hierarchies, such as structural decomposition. This dimension of the model library acknowledges the natural input/output relationships that exist between systems and the associated sub-systems and components.
- *Model variant*: This dimension refers to the possible variation of model fidelities, details, assumptions, etc. that may exist when establishing a conceptual model for the physical world.
- *Model perspective*: This dimension accounts for modelling the same system but considering different physical domains, e.g., mechanics, thermodynamics, hydrodynamics, etc.

FIGURE 3
Digital Twin Model Library [28]



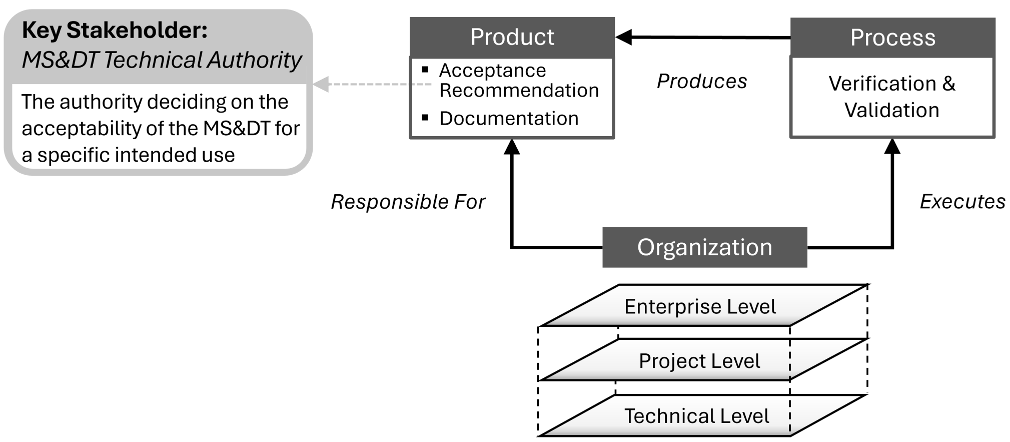
The digital twin configurations are then comprised of one or more models from the model library and can considered to be a model federation as discussed in Section 2/3.5.

The scope of reality encompassed by the digital twin configuration is generally established by the use case or intended purpose of the digital twin and the technical infrastructure that supports the digital twin. In some instances, the limitations of the technical infrastructure necessitates that multiple digital twins are integrated together in what is defined herein as a digital twin scenario. As this is due to technical limitations or convenience, the digital twin scenarios can be considered just another level of model federation. As with the digital twin configurations, the individual sub-models of the federation should be properly evaluated and tested, along with the integrated model federation.

1 General

Recognizing that V&V is a process that must be performed by one or more individuals, the need for some level of management of V&V activities is necessary and a basic implementation framework must be in place. A generic implementation framework is outlined in Section 3/Figure 1 that considers the V&V product, process, and organization responsible for its execution [6]. The organization is the roles played by persons, organizational units, or dedicated organizations during V&V. This section focuses on the management of the organizational entities involved in the V&V. The role of management and planning is to create an implementation framework with robust policies and processes that produce the desired V&V product while fulfilling other objectives such as budget and schedule. A more efficient implementation framework will reduce V&V cost and/or time. Quality benefits may be realized when those savings are reapplied within the V&V project and enable consideration of additional MS&DT configurations, scenarios, and/or use cases.

FIGURE 1
V&V Implementation Framework [6]

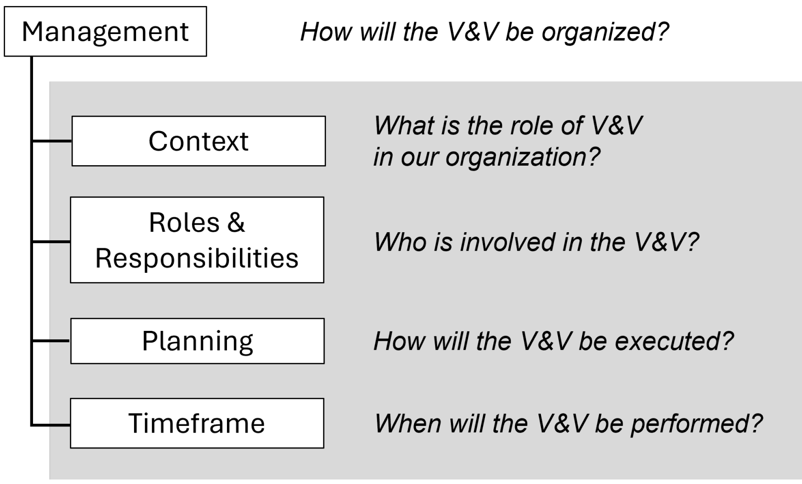


All organizations have some established structure that governs the roles and responsibilities of their employees. In the context of these Guidance Notes a generic hierarchical structure is assumed where V&V activities can be performed at the three levels as listed below. The central focus of this section is the project level, as the V&V of many MS&DT are expected to require management and decision-making above the technical level. The planning of technical level activities is addressed in Section 5/3 within the context of the claim-evidence-reasoning V&V framework that is detailed in Section 5. Enterprise level activities are expected to be driven by the business needs of an organization and can be viewed from a perspective that considers their supporting role at the project level.

- *Enterprise Level* - Organizational elements recommended to set the strategic and enabling capabilities to support the business environment required for V&V activities.
- *Project Level* - Organizational elements recommended to cover the managerial aspects needed to support the technical execution.
- *Technical Level* - Organizational elements recommended to support the engineering aspects of the V&V effort. Specifically, this level is responsible for developing and delivering the acceptance recommendation to the MS&DT Technical Authority responsible for approval of the MS&DT use.

To address the need for management of V&V activities, this section considers the four factors listed in Section 3/Figure 2 that are relevant to any project. Establishing the context of the V&V effort helps identify the specific organizational entities that will participate in the V&V project. The stakeholders involved must be specified and their specific roles and responsibilities regarding the V&V project must be assigned. Planning is required to organize the execution of the project at the technical level and provide a framework for communication between the various organizational levels. The timing of V&V activities must be determined, particularly with respect to the broader lifecycle for the MS&DT.

FIGURE 2
Factors for the Management of V&V Activities



3 Context

Understanding the role of V&V within one's organization is an important first step for establishing the context for a specific V&V need. Does our organization develop MS&DT or are they acquired from a different organization either internal or external? Is V&V performed by the MS&DT development team or by a separate group that was not part of the development process? What level of V&V rigor is applied to MS&DT? Considering these types of questions is a starting point for understanding the role and purpose of V&V. The remainder of this section focuses on identifying where V&V activities are performed within the existing organizational entities and structure and determining the desired level of independence for V&V.

3.1 Existing Organizational Entities and Structure

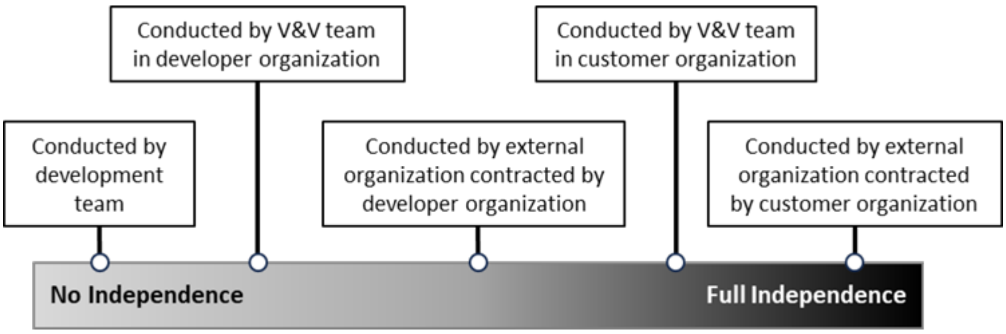
An examination of existing organizational entities and structure is important for understanding how and where the management concepts presented in these Guidance Notes for V&V relate to one's organization. Two aspects to consider are the organizational structure and decision-making authority.

- *Organizational Structure* – Several types of organizational structures exist such as functional, geographic, product/market, or process-based structures. For some structures all V&V activities may fall within the same division, while others may require cross-divisional collaboration. The involvement of multiple organizational divisions may require additional effort to ensure resource availability for V&V. Some organizations may have dedicated V&V teams while others may not.
- *Decision-Making Authority* – The designation of decision-making authority within an organization will influence the number of organizational levels involved in V&V. A vertical organizational structure will often involve more organizational levels than a flat structure. Knowledge of which roles have the appropriate decision-making authority will enable certain V&V roles to be assigned. Within the context of the V&V project, the MS&DT Technical Authority is responsible for making the final decision on whether the MS&DT may be employed.

3.3 Level of Independence

Independent V&V is often sought to help ensure that the outcome is objective. It is recognized that independence for V&V entities and stakeholders in activities can be considered as a gradient which can be selected accordingly to match the V&V needs. Some examples of the varying levels of independence are shown in Section 3/Figure 3. The level of independence of these organizational entities in their roles during the execution of V&V activities is highly dependent on the MS&DT Technical Authority acceptance decision needs and the complexity of the MS&DT system. The level of independence for the V&V activities will also factor into the credibility assessment.

FIGURE 3
Levels of Independence in V&V Activities



It is noted that other factors may contribute to the desire for independent V&V. Independent V&V may be required for policy or regulatory compliance. Also, third party V&V may be sought when the organization lacks personnel with specific skills, resources, or tools necessary for the V&V effort. The disadvantages of greater independence often include higher costs and/or duration. The additional information exchanges between the MS&DT developers and V&V team may introduce inefficiencies and/or place limitations on the timing of data exchanges.

5 Roles and Responsibilities

Identifying the stakeholders for a V&V effort is an essential management activity. There may be one or more stakeholders who support or are interested in the V&V effort, and the exact number of stakeholders for a specific V&V effort will vary on a case-by-case basis. By identifying the stakeholders, managers can begin the process of assigning personnel to specific roles and communicate their responsibilities regarding the V&V effort. This section presents several categories of stakeholders that may be involved in a V&V effort.

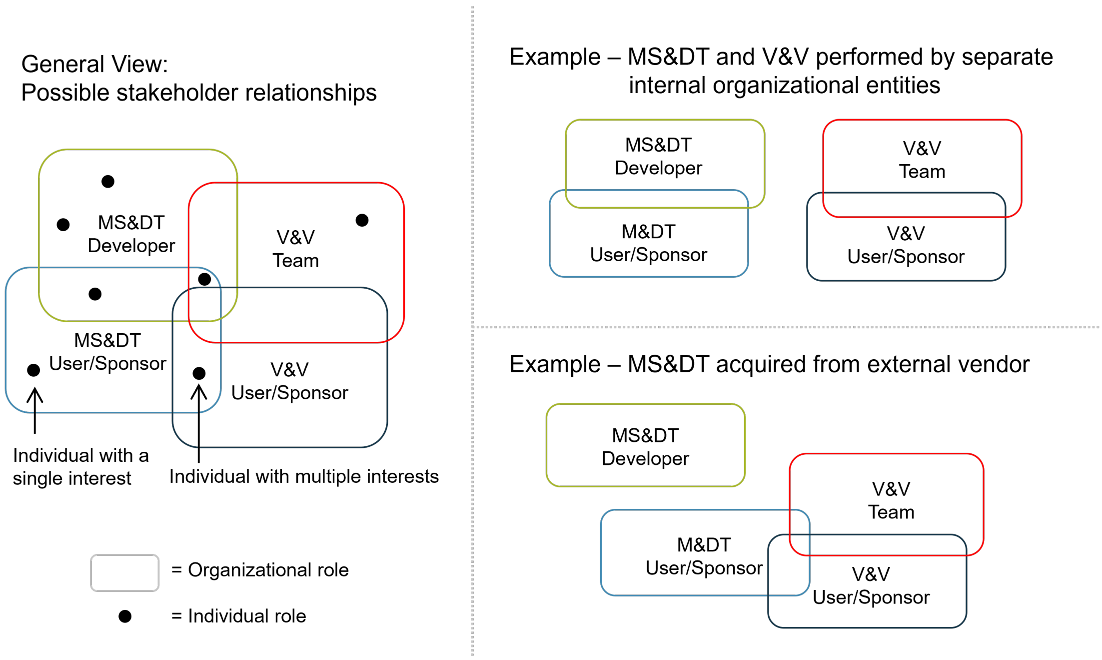
5.1 Organizational Roles

Due to the synergy between the MS&DT development process and the corresponding V&V process it is useful to distinguish between the following five stakeholders.

- *MS&DT Technical Authority*: this is the authority ultimately responsible for deciding on the acceptability of the MS&DT for a specific intended use. To this end, the results of the V&V activities are considered along with other factors in their determination for the acceptance and approval of use of the MS&DT for a specified use.
- *V&V User/Sponsor*: this is the authority responsible for determining if the acceptability criteria for the MS&DT for its intended use are met and for making the final acceptance claim based on the evidence gathered from the V&V activities. This role is also responsible for specifying V&V requirements.
- *V&V Team*: this role encompasses one or more individuals responsible for the execution of V&V activities. The extent of the V&V team is dependent on the complexity of the V&V activities and the maturity of the organization.
- *MS&DT User/Sponsor*: this is the end user of the MS&DT system or results. This role can also encompass the sponsor of the MS&DT who is interested in the outcome of the use of the MS&DT. This role may also encompass the MS&DT Technical Authority.
- *MS&DT Developer*: this is the developer of the MS&DT system or results.

While it is important to understand the distinctions among these four roles, it is not essential for all roles to be manifested in a V&V effort. In addition, there is the possibility for overlap such that a single individual might hold several stakeholder interests. Some of these variations are illustrated in Section 3/Figure 4. The general view on the left half of the figure demonstrates how individual personnel may have one or more stakeholder interests. When no overlap between certain stakeholders exists, it is possible to redraw the stakeholder boundaries to illustrate the exact scenario for the V&V effort.

FIGURE 4
Macro View of V&V Roles and Responsibilities



This macroscopic view of the key stakeholders for a V&V effort is useful for establishing the role of relevant organizational entities whether internal or external. The involvement of external organizations in the V&V effort will require additional types of management activities such as the management of agreements. Accreditation authorities are a category of external stakeholders that may be involved in the V&V effort.

- *Accreditation Authority*: this is a third-party role responsible for official certification that a MS&DT is acceptable for use for a specific purpose.

5.3 Individual Roles

Each group of stakeholders is comprised of one or more individuals that serve various roles. The roles presented here are recommended for the V&V framework provided in these Guidance Notes [6]. These roles are associated with the V&V Team and their function is to provide the MS&DT Technical Authority with an acceptance recommendation. Depending on the size of the V&V effort, individual personnel may hold one or more of these roles on the V&V Team or several individuals may be required to fulfill a specific role.

- *V&V Enterprise Manager*: this role is responsible for enabling V&V projects through the management of resources and the management of agreements.
- *V&V Project Manager*: this role is responsible for the management of the V&V project by such that that the specified outputs are produced and other terms of the agreement are met.
- *Acceptance Leader*: this role is responsible for establishing the acceptance criteria and assessing the V&V results to provide an acceptance recommendation.
- *V&V Team Leader*: this role is responsible for developing the V&V plan and integrating the results.
- *V&V Implementer*: this role is responsible for executing V&V tasks and generating the requested data.
- *Subject Matter Expert (SME)*: this is an expert in the MS&DT technical area(s) of interest that supports the V&V team.

7 Planning

This section highlights the types of V&V planning activities anticipated within the generic three-tiered organizational structure introduced in Section 3/1 and the relevant roles associated with those activities. The enterprise level focuses on the strategic management and planning of resources. The project level addresses the need to plan how decisions, risks, and information will be managed during V&V. The technical level considers planning of the engineering aspects of the V&V effort. While many organizations are hierarchical in nature and planning decisions are cascaded downward, it is expected that the stakeholders at each organizational level will have some degree of autonomy in the planning of their own activities. The presented framework is expected to be tailored to individual V&V projects, existing organizational structures, and established organizational V&V practices. The intent of this section is to help guide the development of organizational structure, policy, and processes that promote the execution of V&V activities in an organized and effective manner.

7.1 Enterprise Level

Organizational elements recommended to set the strategic and enabling capabilities to support the business environment required for V&V activities.

The individual roles involved at this level include:

i) MS&DT Technical Authority,

ii) V&V User/Sponsor,

iii) V&V Enterprise Manager

The Enterprise Manager is responsible for producing policies that will govern V&V activities at the project and technical levels. Organizational V&V standards for quality, documentation, and MS&DT lifecycle management are examples of policies set in place by the V&V Enterprise Manager. Those policies should be tailored to ensure that the organization's strategic interests are met. To retain knowledge gained from V&V efforts, data governance policies are also needed. These policies will help ensure that lessons learned from past V&V projects are easily accessible.

Both personnel and infrastructure resources are required to perform V&V for MS&DT. The infrastructure resources required for V&V may include hardware, software, services, tools, and facilities. Those resources are often limited and may need to be allocated across multiple V&V projects, or when resources are not dedicated entirely to V&V activities they must be shared with other operational functions. An Enterprise Manager has the role of confirming that a pool of personnel with the necessary skills and experience to fulfill one or more V&V roles is available and that the required infrastructure resources are available or readily acquired.

7.3 Project Level

Organizational elements recommended to cover the managerial aspects needed to support the technical execution.

The individual roles involved at this level include:

i) MS&DT Technical Authority,

ii) V&V User/Sponsor,

iii) V&V Project Manager

As the initiator of the V&V effort, the V&V User/Sponsor is involved at the project level. The V&V User/Sponsor is responsible for transforming their needs into a set of requirements for the V&V project. The V&V User/Sponsor also contributes to the development of any agreements required for the V&V project.

Project level planning by the Project Manager is required to determine how the V&V project should be organized to meet the requirements of the V&V agreement established between the V&V User/Sponsor and V&V Team. Components of the project level planning include identifying the technical-level processes that must be instantiated, the corresponding roles required, and the assignment of personnel to each role. A Project Manager is responsible for establishing and updating the project schedule, confirming personnel and resource availability, starting the project, monitoring it, and closing it. The Project Manager must establish the procedures for communication during the project and synchronize project tasks. The communication procedures should indicate what information is stored, where it is stored, and how it is retrieved.

During the project the role of the Project Manager is to support the technical team implementing the V&V processes. When decisions are required, the Project Manager can gather relevant information and involve relevant stakeholders to make informed decisions. The Project Manager should also seek to identify risks, assess them, and develop risk mitigation plans when necessary. To help improve quality and gather data to support future project planning it is recommended that the Project Manager measure and record project performance.

7.5 Technical Level

Organizational elements recommended to support the engineering aspects of the V&V effort. Specifically, this level is responsible for developing and delivering the acceptance recommendation.

The individual roles involved at this level include:

i) V&V Team Leader,

ii) Acceptance Leader,

iii) V&V Implementor,

iv) SME

Organizational elements that support the engineering aspects of the V&V effort are associated with the technical level. Specifically, this level is responsible for developing and delivering the V&V product: the acceptance recommendation for the subject MS&DT. This final acceptance recommendation is the culmination of the execution of several technical level processes. The Acceptance Leader is responsible for planning acceptance criteria for the technical level processes to ensure that the acceptance recommendation will support the needs of the MS&DT Technical Authority. After weighing the information gathered during V&V against the acceptance criteria, the Acceptance Leader forms the acceptance recommendation.

The complexity of many MS&DT may necessitate the participation of several V&V Implementors at the technical level. The V&V Team Leader's role is to plan the technical level processes to generate the evidence necessary to demonstrate that the acceptance criteria have been met. The V&V Team Leader also delegates ownership of tasks to the V&V Implementors. It is the role of the V&V Implementor(s) to execute the individual technical processes and create any intermediate information artifacts required to support other technical processes or project documentation. Subject Matter Experts (SMEs) provide additional expertise and experience on specific subjects or processes as required to complement the knowledge of the V&V Implementors. More details regarding the roles and responsibilities at the technical level are available in Section 5/3 where they are discussed in relationship to the V&V framework presented in Section 5.

The organizational composition at the technical level is influenced by the desired level of independence for V&V activities. When MS&DT Developers also serve as V&V Implementors, management is required to ensure that the needs of both stakeholders are satisfied.

9 Timeframe

The extent, rigor, and timeframe of a V&V effort is driven by the V&V needs, which are traceable to the MS&DT Technical Authority's acceptance decision procedure(s) needs. Depending on these needs, the V&V effort could span the whole MS&DT lifecycle or only a specific phase of the lifecycle and could focus on one specific or multiple (intermediate) MS&DT products. Each case may require separate acceptance criteria with its own scope and development timeline. This section presents two classical types of V&V that can be identified based on their time frames: post-hoc V&V and concurrent V&V. These Guidance Notes support both time frames but are not limited to only these distinct types. The V&V activities can be post-hoc, concurrent, iterative, recursive, or even be a recurrent effort in the case where legacy MS&DT products or results are updated or reused for a different intended use. The purpose of identifying the general timeframe for V&V activities is to enable the execution of more detailed planning activities such as scheduling the availability of personnel and resources.

9.1 Post-Hoc Verification and Validation

With post-hoc V&V, the V&V activities are conducted in retrospect on a MS&DT system after development or on MS&DT results after MS&DT system employment. Post-hoc V&V is commonly encountered and may be the only alternative when the MS&DT is acquired from an external organization and/or an independent third party is used to perform the V&V. Post-hoc V&V can reduce the number of iterations performed for some V&V activities and may be selected when there is a higher level of confidence that significant flaws requiring extensive rework will not be identified. For example, a MS&DT variant developed from a previously accepted MS&DT may be a candidate for a post-hoc approach. The creation of MS&DT using available models from a library is another such case that may be suited for post-hoc V&V.

9.3 Concurrent Verification and Validation

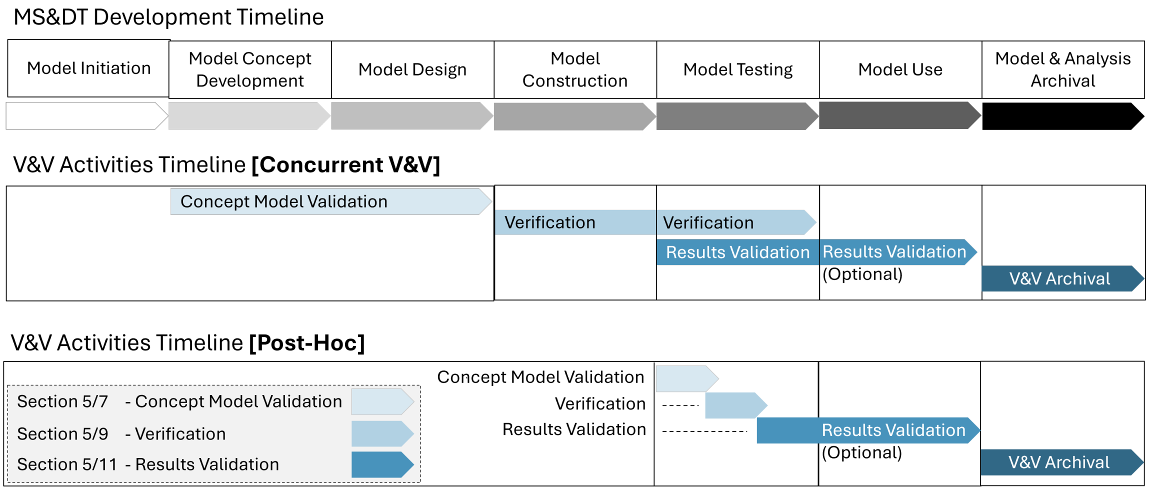
Concurrent V&V involves conducting V&V activities concurrently throughout the whole MS&DT life-cycle to manage and improve the quality of newly developed MS&DT systems or results. Concurrent V&V enables flaws in the MS&DT to be identified earlier which enables corrections to be made sooner and at lower cost. Concurrent V&V is readily supported when the V&V is performed by the MS&DT development team, however configuration management is important to ensure that the V&V tasks use the most up-to-date MS&DT version. When dedicated V&V teams or external organizations perform the V&V activities policies and procedures to govern the necessary information exchanges among parties should be put in place. Those policies and procedures should clarify when exchanges occur, how they will be executed, and what content will be shared.

9.5 Verification and Validation Timeframe Relative to MS&DT Development

Knowledge of the timing of the MS&DT lifecycle phases is important for planning the timing of V&V activities. A graphical comparison between the post-hoc and concurrent V&V approaches and their synchronization with the MS&DT development timeline is presented in Section 3/Figure 5. The MS&DT development timeline follows the MS&DT lifecycle phases described in Section 2/3.9. As indicated, the post-hoc V&V activities don't begin until after the MS&DT has been constructed while concurrent V&V begins during MS&DT concept development. The three V&V activities introduced in Section 3/Figure 5 are as follows: i) concept model validation, ii) verification, and iii) results validation, are discussed in detail in Section 5 of these Guidance Notes. Recommendations regarding the documentation of V&V activities are presented in Appendix 3.

The timelines in Section 3/Figure 5 do not account for any iteration in the V&V or MS&DT development process. With post-hoc V&V, issues identified during V&V will necessitate the iteration of one or more MS&DT lifecycle phases followed by subsequent iteration of affected V&V activities. With concurrent V&V, an iteration in MS&DT development prior to the completion of the MS&DT could trigger an iteration in the concurrent V&V process. The project-level risks associated with iteration for both MS&DT development and V&V should be weighed when selecting the appropriate approach.

FIGURE 5
Comparative view of Concurrent and Post-Hoc V&V Timeframes



SECTION 4
Criticality Categorization

1 General

These Guidance Notes present V&V activities as a risk mitigation approach. The use of any MS&DT has inherent risk due to simplifications and assumptions introduced during their development. In addition, the level of risk associated with MS&DT use will vary for different usage scenarios. For example, use of a MS&DT on a vessel carrying hazardous material is expected to introduce a higher risk level than use of the same MS&DT on a vessel with non-hazardous cargo. The criticality of a MS&DT use is based on the consequences of the MS&DT not providing a correct result and the degree to which it influences the decision-making process. The use of MS&DT in more critical applications is expected to demand a higher level of risk mitigation. The MS&DT Technical Authority is responsible for assessing the criticality of the intended use for the MS&DT. The purpose of this section is to provide guidance for the MS&DT Technical Authority in evaluating and communicating the criticality of the intended use to which the MS&DT results are to be applied.

The following points include guidance on when criticality categorization should be performed and assessing V&V needs when more critical MS&DT uses are encountered.

- It is recommended that a criticality assessment be performed for each MS&DT use case. Numerous factors may influence criticality therefore each MS&DT use should be specifically assessed. A new MS&DT use case may introduce new factors that impact criticality that had not been part of criticality assessments for prior use cases.
- When a new MS&DT use is encountered that has a higher criticality than prior MS&DT uses, previous V&V efforts may be found to provide insufficient risk mitigation and additional evidence may need to be gathered to support an acceptance claim for the more critical MS&DT use.
- Consideration of the highest anticipated criticality for a MS&DT at the time of its initial V&V may help prevent the need to generate additional evidence when future MS&DT usage scenarios are assessed.

3 Assessment of MS&DT Criticality

An MS&DT criticality assessment considers:

1) Consequences of the MS&DT-based decision on human safety, safety of the asset, and/or threat to the environment.

2) The degree to which the MS&DT influences the decision-making process.

This section provides a MS&DT criticality matrix that can be used as proactive guidance for the MS&DT Technical Authority or other V&V stakeholders to mitigate potential risks as early in the MS&DT lifecycle as possible. The criticality determination can also serve as a basis for determining how rigorously the V&V effort should follow the guidance in the subsequent sections.

3.1 Consequence of Decision

The consequence of the decision can be classified based on the potential detrimental impact the decision may have on the real-world system. These detrimental impacts may include human safety, safety of the asset, and threats to the environment.

The categories and qualifying terms for assessment of consequences are situationally dependent and should ultimately be determined and tailored by the stakeholder(s). The sample set of categories for consequence of decisions influenced by MS&DT results are provided below. These sample descriptors are based on the levels defined for system categories detailed in 4-9-3/Table 1 of the ABS Marine Vessel Rules [20] and relate to the potential extent of damage that may be caused by a failure of the decision-making process reliant on MS&DT results due to underperformance or failure of the MS&DT. This sample categorization of low (L), medium (M), and high (H) consequence level is shown in Section 4/Table 1.

TABLE 1
Decision Consequence Levels (Sample)

Decision Consequence Level	*Effects of Failure of Decision-Making Process*
L	Failure will not lead to dangerous situations for human safety, safety of the asset, and/or threat to the environment.
M	Failure could eventually lead to dangerous situations for human safety, safety of the asset, and/or threat to the environment.
H	Failure could immediately lead to dangerous situations for human safety, safety of the asset, and/or threat to the environment.

Other decision consequences that may be considered in decision consequence categorization can include but may not be limited to detrimental impacts to operational status, performance capability, schedules, cost, and/or mission success. The user may include such criteria to evaluate the Consequence Levels to augment or replace the criteria in Section 4/Table 1, based on each application of the MS&DT.

3.3 Influence of MS&DT on Decision

The influence dimension represents the degree to which the MS&DT results impact the decision under consideration. This may be dependent on the level of autonomy of the MS&DT result in the decision-making process and the amount of other information available when making the decision. A sample approach for incorporating both of these elements in determining the MS&DT influence level are given in Section 4/Table 2 and Section 4/Table 3.

TABLE 2
MS&DT Result Integration and Supplemental Information Availability (Sample)

MS&DT Result Integration Level in Decision-Making Process (RIL)	
0	Informative MS&DT results
1	Decision recommendations based on MS&DT results (Human-in-the-loop)
2	Fully autonomous decision-making based on MS&DT results

Supplementary Information Availability (SIA)	
0	No data from actual or similar system. --AND-- No other credible MS&DT or analysis data.
1	Limited data from actual or similar system used in decision-making apart from MS&DT. --OR-- Credible results from another MS&DT are available and used in decision-making.
2	Ample data from actual or similar system used in decision-making apart from MS&DT. --AND-- Credible results from another MS&DT are available and used in decision-making.

The MS&DT influence level is defined by the formula below and represented by the three levels defined in Section 4/Table 3.

$$2 * RIL + (2 - SIA) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

TABLE 3
MS&DT Influence Levels (Sample)

MS&DT Influence Level	$2 * RIL + (2 - \mathrm{SIA})$	*Description*
L	0, 1	MS&DT results have a negligible or minor influence
M	2, 3	MS&DT results have a moderate influence
H	4 and above	MS&DT results have a controlling or significant influence

3.5 Criticality Matrix

The criticality matrix is to be utilized based on both (1) the level of consequence of decision and (2) the influence of the MS&DT results on the decision as described in Section 4/3.1 and Section 4/3.3 respectively, and as shown in Section 4/Table 4.

The criticality matrix can be used to assign a criticality level (L: Low, M: Medium, H: High) for the MS&DT results as part of the decision-making process.

TABLE 4
Criticality Matrix

MS&DT Influence Level	*Decision Consequence Level*		
	L	***M***	***H***
L	L	L	M
M	M	M	H
H	M	H	H

The assessed criticality level can then be utilized to codify the measure of control and management of the MS&DT development and operations.

- MS&DT-based decisions that are classified as High are clear candidates for performing standardized and rigorous processes to establish and document the necessary evidence to assess the credibility of the MS&DT results for the intended use and the use of these Guidance Notes for guiding verification and validation activities is strongly encouraged.

- MS&DT-based decisions that are classified as Medium may or may not require the same level of scrutiny and effort as those categorized as High and the use of these Guidance Notes for guiding verification and validation activities is at the discretion of the V&V User/Sponsor.
- For MS&DT-based decisions classified as Low criticality, there is not a critical driving force for them to rigorously follow the guidance in this document, however these Guidance Notes may still be used as a guide for good practices in MS&DT verification and validation.

An organization's V&V policies may provide for tailoring of the V&V processes and resulting documentation based on the determined criticality.

While the assessment of MS&DT criticality provides insight into the level of rigor that needs to be applied during V&V, other factors can influence the breadth of V&V activities and the total level of resources and time required. For example, a complex MS&DT with numerous features is expected to require more V&V effort than a simpler solution. Similarly, a MS&DT use case associated with a more controlled environment with few anticipated risks may require fewer evaluation points during V&V compared to a MS&DT use case associated with a complex environment with numerous risks of varying severity and likelihood. The criticality level combined with additional knowledge about a specific MS&DT use can influence the allocation of V&V resources and how they are prioritized.

SECTION 5

Verification and Validation Framework

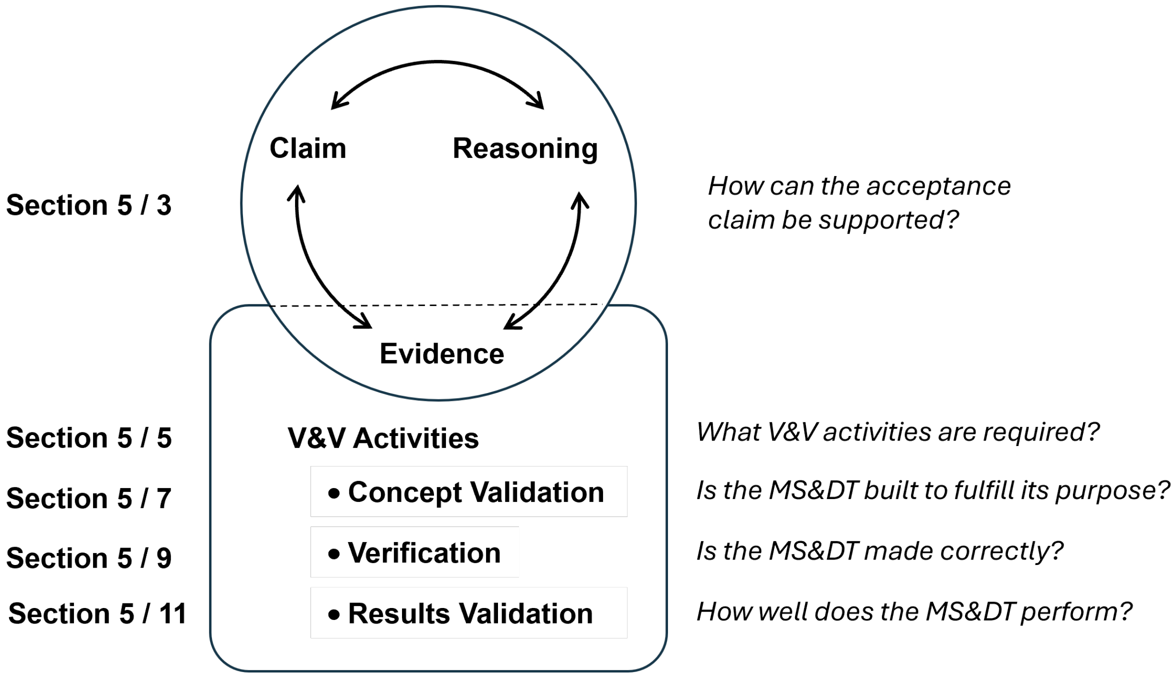
1 General

This section provides the philosophy, general concepts, and recommended best practice processes for conducting V&V activities on a MS&DT. This section has been developed based on published contributions from many other organizations concerned with the V&V and use of MS&DTs which are detailed in Appendix 1.

The objective of the V&V framework presented in this section is to enable an acceptance claim to be made based on evidence and documentation from the V&V activities that supports the MS&DT Technical Authority's decision whether the MS&DT is suitable for its intended use. An overview of the V&V framework illustrated in Section 5/Figure 1.

In the V&V framework, a claims-evidence-reasoning strategy is used to help govern the V&V activities such that they produce the evidence needed to support the acceptance claim. Prior to conducting the V&V activities, reasoning is applied to establish acceptance criteria. Provided that the evidence gathered adequately satisfies the established criteria, an affirming acceptance claim can be made. If the evidence resulting from the V&V activities indicates that some or all acceptance criteria have not been satisfied, it may require iterating through the claims-evidence flow with revised evidence solutions or revised acceptance criteria. Another alternative in this situation is for the MS&DT Technical Authority to issue an acceptance claim, but with limitations or caveats on the MS&DT use.

FIGURE 1
V&V Framework



A key component of the incorporation of the MS&DT V&V into the MS&DT Technical Authority decision-making process is the documentation of the relevant activities, assessments of V&V results, and any other supporting evidence. Relevant evidence to be included in documentation includes the MS&DT intended use statement, requirements, and V&V acceptance criteria, conceptual model validation, solution and code verification, and result validation comparisons. More details on best practices for the documentation of V&V activities can be found in Appendix 3.

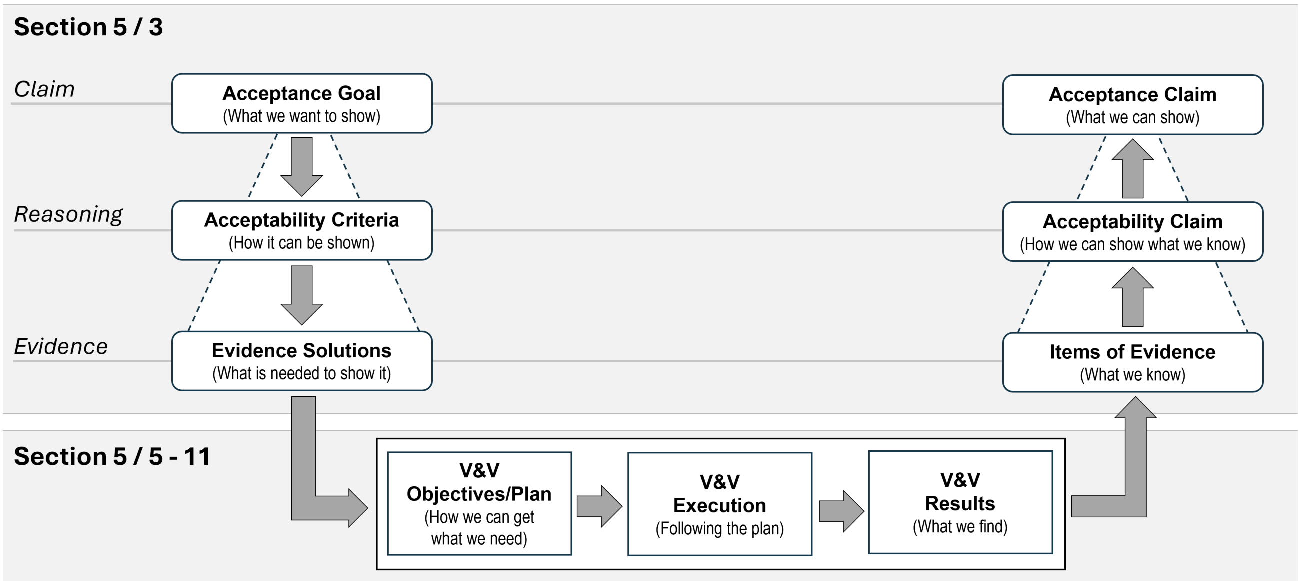
3 Claims-Evidence-Reasoning Approach

Leveraging elements of the claim-reasoning-evidence framework used in safety, security and assurance cases, a V&V goal-claim network can be established to create a structured framework for reasoning about V&V evidence as it relates to an acceptance decision [6]. This flow is shown in Section 5/Figure 2.

If this effort is done concurrent with the MS&DT development, then the criticality of the MS&DT, as discussed in Section 4, should influence the formality and rigor of this process. It begins with an acceptance goal, which aligns with the MS&DT's specific intended use, that is used to establish the formal acceptance criteria. The acceptance criteria are used to form a basis for the evidence solutions which include the data to be used, tests/experiments to be run, SMEs to incorporate, and other sources for referent data. These evidence solutions are then exercised in the V&V activities to generate items of evidence. The items of evidence are used to support the arguments made regarding acceptability claims, which are the statements made on whether the acceptance criteria have been met or not. The acceptability claims then become the basis for supporting an overall MS&DT acceptance claim. Note that this process is iterative both overall and within each step of the transition from acceptance goal to acceptance claim.

In instances where the V&V is performed post-hoc or as independent V&V, the iterative nature of this process becomes more challenging, therefore it is recommended that V&V be performed concurrently with the MS&DT development process whenever possible. It should also be noted that an acceptance recommendation made by the MS&DT Technical Authority is likely dependent on a number of other credibility factors in addition to V&V, which is discussed in Section 6.

FIGURE 2
Goal-Claim Flowchart [6]



The processes on the left and right sides of Section 5/Figure 2 can generally be thought of as inverses of one another. The development of the acceptance goal, acceptability criteria, and evidence solutions is a top-down approach whereby the necessary reasoning is developed to identify criteria that support the top-level acceptance goal and subsequent identification of evidence solutions that support the acceptability criteria. Conversely, the process of evaluating the items of evidence to create acceptability claims that support the final acceptance claim is considered a bottom-up process. While guided by the reasoning framework established during the development of the acceptability criteria, when generating the acceptability claims additional arguments are often required due to imperfections in the items of evidence. It may be necessary to identify trends, compare and contrast data sets, or perform other types of data analysis to account for various uncertainties. Uncertainties are introduced through the inability of the MS&DT to fully represent the physical system and variations in the physical system itself such as variabilities in geometry, physical properties, etc. Part of credibility categorization which is detailed in Section 6 is to account for this gap between the idealized acceptance goal and the acceptance claim with its inherent uncertainty.

The vertical alignment of the left and right sides of the idealized workflow in Section 5/Figure 2 is also done to emphasize that the claim-evidence-reasoning framework is hierarchical in the sense that many pieces of evidence are required to build acceptability claims at one or more levels which ultimately support the final acceptance claim. The remainder of this subsection focuses on the three tiers of the goal-claim framework in the top half of Section 5/Figure 2. As indicated in the left side of Section 5/Figure 2, formulating the claim-evidence-reasoning strategy is considered prework to the execution of the core V&V activities in the middle of Section 5/Figure 2. The evaluation of the evidence and development of the acceptance claim on the right side of Section 5/Figure 2 are indicated as post work since those processes rely on the V&V results as inputs. Regarding the middle portion of Section 5/Figure 2, Sections 5/7-11 provides guidance on the core verification and validation activities introduced in Section 5/5.

3.1 Establishing Claims, Reasoning, and Evidence

In a claim-reasoning-evidence approach it is important that the claims, evidence and supporting reasoning are well prepared to ensure that the claims are credible. The following definitions for each of the three components in the framework include brief descriptions of attributes associated with high-quality claims, evidence, and reasoning.

- *Claim* – A claim is a statement that satisfies the original acceptance requirements. A well-formed claim makes it clear to the reader what the criteria was without having to look at the acceptance requirements.
- *Evidence* – Evidence is the data that supports the claim. Good evidence directly supports the claim, is traceable, and cites specific data or describes trends in data when necessary.
- *Reasoning* – Reasoning is the justification that connects the evidence to the claim. Good reasoning has easy to follow logic, uses sound scientific principles, and is strongly tied to the claim.

The creation of high-quality claims, evidence, and reasoning, while important for the final acceptance claim, also promotes efficient communication of information among members of the V&V team and other stakeholders. The organization of claims, reasoning, and evidence in a hierarchical manner supports the communication of outcomes to stakeholders at different organizational levels. For example, a stakeholder at the executive level may have interest in a claim that corporate quality policies were followed as evidenced by check sheets confirming procedures were followed. Likewise, a technical level claim that the resolution for a specific computation was sufficient as evidenced by successful grid convergence tests would be appropriate for the engineer completing the same quality check sheet.

3.3 Acceptance Goal / Acceptance Claim

The acceptance goal when performing V&V is to establish with a reasonable amount of certainty whether a MS&DT is or is not suitable for a specific use case and is common to any V&V effort. The primary objective at the onset of V&V is to perform the initial decomposition of the overall acceptance claim into a set of high-level acceptance goals that encompass needs of the MS&DT Technical Authority, needs of the V&V User/Sponsor, needs of the MS&DT User/Sponsor, and any other relevant enterprise and project level objectives. While there are many possible ways of categorizing these various needs and objectives, the purpose of this section is to introduce one approach that seeks to ensure that the initial decomposition of the acceptance goal is comprehensive. Four high-level acceptance goals are to:

- *Meet MS&DT requirements* – The development of any MS&DT is often guided by a set of requirements that are established to ensure the end product will fulfill operational needs in the Real World. Satisfaction of the MS&DT requirements is a fundamental acceptance goal. If all MS&DT requirements are not to be met, then additional development may be required before the MS&DT can be accepted.
- *Meet V&V requirements* –Within the V&V World additional requirements may be introduced that account for the needs of the MS&DT Technical Authority. The V&V requirements may also be based on established organizational policies or be project specific. For example, involvement of specific SMEs may be required if the MS&DT achieved a certain criticality level. While these V&V requirements may not be technical in nature, they are still important and relevant to acceptance.
- *Address all relevant hazards associated with the MS&DT use case* – It is anticipated that MS&DT will be used to make design or operational decisions in the Real World that will encounter known hazards. It is important to understand how MS&DT use may impact the mitigation measures put in place to address the hazards. Part 1D-3-1 of the ABS Rules for Alternative Arrangements, Novel Concepts, and New Technologies [19] outlines several high-level Tier 1 Goals that were established to address common maritime hazards. It is recommended that if the MS&DT use impacts the fulfillment of any Tier 1 Goal, that satisfaction of the relevant Tier 1 Goal(s) be part of the acceptance goal. The purpose of this goal is to ensure that no significant hazard was overlooked in the formulation of the MS&DT requirements and the V&V requirements.
- *Satisfy business needs* – Even if a MS&DT may provide a correct technical solution for a use case, it may not be the most suitable problem solution based on other criteria introduced in the Problem World such as cost and schedule targets. It is recommended that a cursory review be done to confirm that the MS&DT use case is the right solution from a business perspective. For example, would physics-based analysis or physical testing be more appropriate? Confirming that a MS&DT use case is in line with business needs and strategy is a relevant acceptance goal.

The primary outcome of the V&V framework outlined in this section is an acceptance claim. The acceptance claim is expected to be binary, either affirming or rejecting that the MS&DT is suitable for the intended use case. However, the evidence gathered during the V&V process may reveal additional insights that may be contextually relevant to the acceptance claim, therefore it may be appropriate to amend the acceptance claim with details such as known operational limits or restrictions on use. The inclusion of these caveats therefore means that the acceptance claim may not directly mirror the acceptance goal.

Within the organizational framework outlined in Section 3, the Acceptance Leader is responsible for forming the acceptance claim. Assigning ownership of the acceptance claim to a specific individual is beneficial for confirming that a high-level perspective is maintained within the technical level.

3.5 Acceptability Criteria / Acceptability Claims

3.5.1 Generation of Acceptability Criteria

The role of acceptability criteria is to establish how the acceptance goals can be achieved. Acceptability criteria are typically identified through the hierarchical decomposition of acceptance goals so that they are transformed from being generic to specific. When this decomposition process is comprehensive in nature, a basic reasoning structure is established whereby higher-level acceptability claims can be made when all relevant lower-level acceptability criteria are met. The primary challenge with this top-down approach is knowing how many levels of decomposition are necessary and whether the decomposition at each level was comprehensive. The purpose of the criticality categorization in Section 4 is to help understand when the decomposition process can terminate.

Within the organizational framework outlined in Section 3, the Acceptance Leader is responsible for acceptability criteria being traceable to the high-level acceptance goals. The Acceptance Leader can leverage knowledge from SME's and lessons learned from prior V&V projects when determining whether the acceptance criteria are sufficiently comprehensive to support the high-level acceptance goals.

3.5.2 Standards of Compliance for Acceptability Criteria

Acceptability criteria should be formulated such that they provide the standard for compliance. Verification of compliance with acceptability criteria is straightforward when the acceptability criteria enable pairwise comparisons to be made. The use of pairwise comparisons makes the logic behind an acceptability claim unambiguous for quantifiable or Boolean data. For example, it is simple to calculate the error between empirical data and MS&DT predictions and demonstrate that the values are less than the maximum permissible error stated in the acceptability criteria. Likewise, it is straightforward to verify that the MS&DT does or does not account for relevant physical phenomena in its underlying mathematical models. Qualitative data can be quantified by establishing a rubric that enables the data to be categorized and assigned a numerical value. Section 5/11.5.1 provides guidance on the development of rubrics for quantifying qualitative data.

The criticality categorization performed in Section 4 of these Guidance Notes is intended to influence how the standards for compliance associated with the acceptability criteria are defined. More critical use cases are expected to have more stringent standards for compliance. To illustrate, consider a model that relies on a material property value. For a non-critical MS&DT use it may be sufficient to use a value from a reference book while a more critical MS&DT use may necessitate more precise values supplied by the material supplier or require physical testing. This example also serves to demonstrate that it is not necessary for all acceptability criteria to be resolved into quantifiable metrics which may be challenging; it may be sufficient to ensure that established best practices are simply followed.

Within the organizational framework outlined in Section 3, the establishment of acceptability criteria relies on activities at the project and technical levels. During the V&V process it is expected that the impact of the criticality categorization be fully communicated at the project level so that there is an understanding at the technical level where SMEs can assist in confirming that the appropriate standards for compliance are used. It is also important for managers at the project level to identify any acceptability criteria that are expected to be more influential during the subsequent credibility assessment detailed in Section 6. For those acceptability criteria the Project Manager should ensure that the work products created by the technical team fully document the development of and rationale for those criteria.

3.5.3 Acceptability Claims

The acceptability criteria are expected to have corresponding acceptability claims that document the degree to which the acceptability criteria are met. When the items of evidence fully support the standard for compliance established in the acceptability criteria, the acceptability claims are easily formulated through the reapplication of the reasoning stated in the criteria. Additional effort is required when some items of evidence fail to support the acceptability claims. For example, statistical analysis can be used to form inferences from multiple items of evidence.

Within the organizational framework outlined in Section 3, the development of acceptability claims from the items of evidence gathered during V&V is initiated at the technical level as the items of evidence are generated. V&V Implementors formulate and document acceptability claims as individual tasks are completed. When the items of evidence are found to be inadequate for making strong acceptability claims, project-level decisions are necessary. Iteration of the acceptance criteria process or replanning the evidence solutions could be undertaken.

3.7 Evidence Solutions

The evidence solutions represent the set of information necessary to ascertain compliance with all acceptability criteria. When generating the evidence solutions, it is important to distinguish if the evidence is discrete or continuous in nature. For discrete data it is possible to evaluate all data points, while for continuous data it is necessary to establish representative samples for evaluation. Regardless of the data types, the complexity and scope of many MS&DT can result in domain spaces that are too large to practically evaluate fully and careful planning of the evidence solutions to be gathered is required. The following considerations are illustrative of the types of decisions that are required when planning the evidence solutions.

- *Feature testing* – All features of an MS&DT should be tested. This testing is expected to be part of the code verification process and is integral to establishing credibility with the MS&DT user.
- *Validation domain* – The validation domain should coincide with the application domain when possible. When the application domain extends beyond the bounds of the validation domain uncertainties are introduced. The number of data points in the validation domain is also relevant as a greater number of observations can provide increased confidence in the predictive capability of the MS&DT.
- *Data relevance* – Some acceptance criteria may only be satisfied by specific data points and an understanding of the nature of the standard of compliance is necessary. For example, is it sufficient to check any value for a variable in the application domain or is it necessary to test the maximum and/or minimum possible value? A verification data point for a model that predicts hull girder bending moment is more relevant if its longitudinal location is near midship where the moment value is highest versus near the bow or stern where the moment approaches zero and is less critical.
- *Data pedigree* – Data pedigree impacts the ability to generate high quality items of evidence. It is important to know the sources for input data and their accuracy, correctness, and completeness.

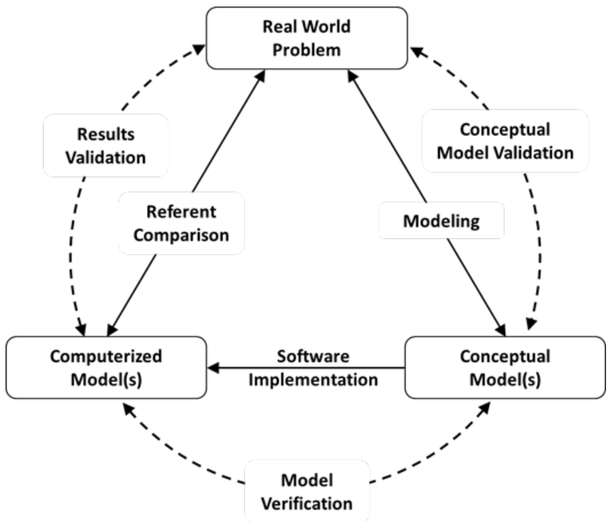
The items of evidence are the V&V results obtained for the planned evidence solutions. It is important to ensure that the items of evidence are complete, correct, produced in a consistent manner, and are sufficiently documented.

Within the organizational framework outlined in Section 3, the planning of the evidence solutions and execution of the V&V activities to create the items of evidence are performed at the technical level. The Project Manager's role is to ensure that scope of evidence solutions is within the limitations of available resources and that the items of evidence are produced on time.

5 Overview of Verification and Validation Activities

V&V activities consist of both separate verification and validation activities. At a high level, the V&V process is shown in Section 5/Figure 3. This representation is commonly referred to as the Sargent Circle and is based on a diagram developed for the Society of Computer Simulation (SCS) [5]. This diagram shows three transitional processes in which the transitions can be aligned with modeling and V&V activities.

FIGURE 3
High Level View of V&V Process [5]



The first transition in Section 5/Figure 3 is the transition from the Real World problem to a conceptual model. This corresponds with the transition from the Real World to the Digital World discussed previously when discussing problem-solving. Note that the Real World Problem represents a specific scope of reality that aligns with a problem/use case. This reality of interest could be a unit problem, component problem, subsystem, or the complete system [4]. As the scope becomes more complex and the derived digital world is comprised of a federation of models, then the V&V process shown in Section 5/Figure 3 will need to be iterated through for each of the individual units as well as the complete integrated model federation. Based on the Real World problem, a conceptual model is developed that is comprised of the model objective, model inputs/outputs, mathematical models, and assumptions/simplifications required to describe the problem of interest. The conceptual model is an abstraction of reality defined by a set of equations, modeling assumptions, and initial/boundary conditions to describe the physical phenomena of the reality being modeled.

The second transition is then from the conceptual model to a computerized model within the Digital World. The conceptual model is then implemented as a computerized model through the development of computer code to execute the model algorithms and typically requires numerical discretization of the mathematical approaches and formula established in the conceptual model.

Finally, the results of the computerized model are employed in the decision-making process that impacts the outcomes of the problem solving in the Real World.

As has been implied above and in earlier sections, the acceptance goal for a MS&DT should be aligned with a specific intended use or problem domain. This intended use establishes the outputs of interest, informs the MS&DT scope, necessary levels of model/simulation fidelity, data collection frequency, and synchronization frequency among other technical requirements that become the basis of a set of acceptance criteria. Note that acceptance criteria should be definable and measurable and preferably quantitative. The acceptance criteria should also cover the three transitional phases of the V&V process, including conceptual model validation, model verification, and results validation.

7 Conceptual Model Validation

Conceptual model validation entails the evaluation of whether the theories and assumptions underlying the conceptual model are correct and whether the model representation of the Real World problem is reasonable for the intended purpose [5]. This is accomplished by reviewing the descriptions provided for:

1) MS&DT objectives: The MS&DT objective(s) should be stated. This evidence should be assessed based on its alignment with the intended use.

2) MS&DT inputs/outputs: The inputs of the MS&DT should be identified in both their source and collection frequency. This should establish the outputs of the model which can be assessed in relation to the identified problem parameters of interest.

3) MS&DT Content: The MS&DT content describes the mathematical approaches and formulas to be used. The validity and context of these formulas must be justified with consideration to their intended use.

4) Assumptions and simplifications: Any assumptions and limitations of the MS&DT should be stated. This evidence may be used in conjunction with results validation in the determination of the applicability of the system results.

In the context of digital twins, conceptual model development is required both for the individual models comprising the digital twin configuration(s) as well as for each configuration itself. In other words, the individual models from the model library should be assessed according to the elements listed above, as well as establishing that same information for the integration of the models to form the digital twin configuration.

9 Verification

The second phase of evidence gathering is the verification assessment. The conceptual model is used to establish a set of requirements for the creation of a corresponding computerized model. Verification is the process of gathering evidence that the computerized model implementation sufficiently satisfies the requirements. Verification can be separated into two important steps: 1) code verification, and 2) calculation verification. Code verification relates to the function of the software developer(s) to produce code that is error-free, robust, secure, and reliable [4]. Calculation verification relates to the function of the MS&DT developers who use the code to obtain solutions to technical problems with sufficient accuracy [4]. These two parts of the verification process are summarized in Section 5/Table 1. It is not possible to definitively prove that software code is accurate and error-free, thus it is important to gather evidence from well-planned test cases to support the claim that the code is sufficiently accurate and error-free.

Some MS&DT may require stored representation verification as a third verification step. Stored representation verification relates to the function of the MS&DT User to correctly enter the descriptive data necessary to accurately represent the entity of interest. Verification of stored representations is important for MS&DT reuse as each scenario or variant may require unique data. Digital Twins require stored representation verification following information exchanges between the physical asset and digital representation. Consequently, Digital Twins will require periodic verification throughout their lifecycle following data updates.

TABLE 1
Verification Activity Categories [4]

Verification Category		*Focus*	*Responsibility*	*Methods*
Code Verification	Software Quality Assurance	Reliability, robustness, and security of the software	Software development team & MS&DT developer	Configuration management, static analysis, dynamic testing, formal testing
	Numerical Algorithm Verification	Correctness of the numerical algorithms implemented in the code	MS&DT developer	Analytical solutions, benchmark problems, etc.
Solution Verification	Numerical Error Estimation	Estimation of the numerical accuracy of the given solution to the governing equations	MS&DT developer MS&DT user	Grid convergence, recovery methods, Richardson extrapolation, etc.
Stored Representation Verification	Verification of geometric and physical attribute data entries	Accurate representation of design or real-world asset	MS&DT developer MS&DT user	Comparison to reference data sets

9.1 Code Verification

Code verification is necessary to ensure that the software implementation of a MS&DT functions as intended. The purpose of code verification is to identify programing errors so that they may be corrected and to ensure that numerical algorithms are implemented correctly. Guidance on code verification is available from numerous sources such as standard organizations, professional societies, government agencies, and corporate procedures. Other resources such as the ABS Guide for Integrated Software Quality Management (ISQM) [22] build upon internationally recognized standards to provide guidance for specific application domains.

The nature of the specific code verification needs for a MS&DT may vary greatly, ranging from the verification of system-level integration to individual units of code. The ISO/IEC 25010:2023 [25] standard provides a quality model that considers a broad range of characteristics that may impact MS&DT use. While code verification is commonly associated with the functional suitability characteristic, eight other characteristics are considered: performance efficiency, compatibility, interaction capability, reliability, security, maintainability, flexibility, and safety. When applying high-level, generic software quality standards, it is recommended that a review of the MS&DT acceptance criteria is performed to determine which aspects of the standard require the most focus during code verification. More detailed standards can be leveraged to guide specific code verification activities such as the IEEE/ISO/IEC 29119 [26] series that focuses specifically on software testing procedures.

The verification of code for numerical algorithms in MS&DT becomes more challenging as the level of complexity increases as it may be difficult to find suitable benchmark problems to verify calculation results, particularly with nonlinear systems. The method of manufactured solutions (MMS) is a technique that provides a way to generate non-trivial solutions that enable code to be checked for algorithm errors [8]. The solutions created via the method of manufactured solutions need not represent realistic real-world scenarios as they are designed simply as a tool to test the code. The method of manufactured solutions is part of several ASME V&V standards such as ASME V&V 20 [15].

9.3 Solution Verification

Solution verification involves the estimation of numerical accuracy of the given solution to the governing equations of a MS&DT. Unlike code verification where error can be evaluated against a known benchmark, solution verification applies to problems where an exact solution is unknown, therefore the outcome of solution verification is only an estimate of error. The primary purpose of solution verification is to confirm that the selected levels of spatial and temporal discretization, convergence tolerance and numerical precision are sufficient for the intended MS&DT use. If the estimated error calculated during solution verification is unacceptable to the MS&DT user additional fidelity may be required. Code verification as described in Section 5/9.1 is considered a prerequisite for solution verification.

Solution verification is common for MS&DT that utilize grid-based algorithms such as finite element analysis. Grid refinement studies that involve obtaining solutions at two or more grid resolutions are often used for solution verification. The results of grid refinement studies should demonstrate that discretization error asymptotically approaches zero through the use of finer grids. The grid convergence index (GCI) suggested by Roache [8] provides a consistent manner for reporting the results of grid refinement studies and is a measure of the percentage the computed value is away from the asymptotic numeric value. The GCI has been adopted in industry standards such as ASME V&V 20 [15] that detail procedures for performing solution verification.

9.5 Stored Representation Verification

Verifying the accuracy and correctness of the stored representation is imperative to MS&DT use due to their reliance on attribute and state data for calculations. Stored representations concern inputs to the MS&DT that are considered invariant. Stored representation data that are entered by a MS&DT user and must be verified such as geometric and physical properties of entities represented by a MS&DT. The stored representation of an as-designed or newly constructed physical asset should be verified against the references such as production drawings, measurements, or test results. For Digital Twins, subsequent stored representations that capture the physical asset's condition at specific times in its operational history must also be verified. For each stored digital representation, verification that the condition information collected from the physical asset was used to correctly update the MS&DT is required. It is also necessary to verify that any modifications to the physical asset relevant to the MS&DT are correctly reflected in the digital representation.

11 Results Validation

The final stage of the V&V process is results validation which entails generating evidence on the degree to which the model is an accurate representation of the Real World problem (to the degree required by the intended use) [13]. Stated in another way, results validation aims to generate evidence that supports the predictive capability of the MS&DT by comparison with real-world referent data. Agreement between the MS&DT results and the Real World referent data is measured as a function of the difference between the model output and the Real World referent data. A well-designed results validation process will consider:

1) Model parameters of interest for comparison

2) Source(s) of real-world referent data

3) Metric(s) for comparison

4) Domain of validation

In many instances, cost and feasibility limit the scope of validation testing. It is typically infeasible to validate the MS&DT over the entire problem domain and there are always unknown factors that are not accounted for in the MS&DT.

11.1 Validation Approaches

A number of best practice result validation methods are given in Section 5/Table 2:

TABLE 2
Result Validation Methods [5]

Results Validation Method	*Description*
Animation	The MS&DT's operational behavior is displayed graphically as the change over time is simulated and compared with the Real World referent.
Comparison with Benchmarks	Outputs of the MS&DT are compared with the results of other known valid models or benchmark data.
Degenerate Tests	The degeneracy of MS&DT is tested by appropriate selection of input and internal parameters.
Event Validity / Operational Graphics	Detailed MS&DT outputs are generalized to higher level, abstracted "events" or "KPIs" which are then compared with similarly abstracted data from the Real World referent.
Extreme Condition Tests	The MS&DT is tested at extreme and unlikely combination of levels of factors and checked if the corresponding results are plausible.
Face Validity	Subject matter SMEs are asked whether the MS&DT and its behavior are reasonable.
Historical Data Validation	This follows the standard train-test approach for model validation, wherein a subset of the data to be used in MS&DT development is set aside for use in model validation.
Internal Validity	Several stochastic runs are made to determine the internal stochastic variability of the MS&DT. A large amount of observed variability may cause the MS&DT's results to be questionable.
Predictive Validation	The MS&DT is used to predict the system's behavior for one or more parameters of interest, and then comparisons are made with data collected from the Real World system (e.g. field tests)
Turing Tests	Individuals knowledgeable about the system are asked if they can discriminate between the system and MS&DT outputs in a blind test.

11.3 Validation Classification

These results validation approaches can be further grouped according to whether the Real World referent data is from the system of interest and whether the approach is subjective or objective. This categorization is shown in Section 5/Table 3.

TABLE 3
Categories of Result Validation Methods [5]

	Observable System	*Unobservable System*
Subjective Approach	Animation Event Validity / Op. Graphics Turing Tests	Face Validity Extreme Condition Tests Degenerate Tests Internal Validity Comparisons w/ Benchmarks
Objective Approach	Predictive Validation Historical Data Validation	Comparisons w/ Benchmarks

11.5 Results Validation Metrics

This section discusses the use of validation metrics to assess whether a MS&DT can be considered "good enough." These metrics should align with the acceptance criteria established at the beginning of the V&V

process (Section 5/3.5). The approach of achieving a "good enough" MS&DT solution to a Real World problem is motivated by both the recognition that often an exact answer is not needed to achieve the desired outcome and business interests.

The establishment of validation metrics is challenging due to uncertainty and error in both the MS&DT results and Real World referents. MS&DT error is introduced during the modeling process and includes factors such as the idealization of physics or discretization error. Real World referents may have error due to factors such as improper instrument calibration. MS&DT uncertainty may stem from unrecognized operational scenarios or relevant physical phenomena. Real World referents may have uncertainty from factors such as an imperfect knowledge of loads, an insufficient number of measurements, or variability in physical properties. The presence of both recognized and unrecognized error and uncertainty should influence the establishment of validation metrics.

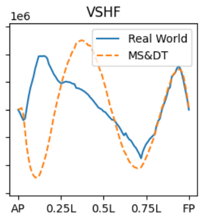
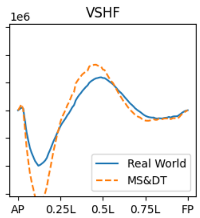
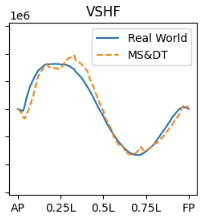
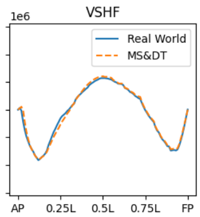
The classification of validation approaches as either subjective or objective in Section 5/11.3 is important as each class requires a different type of validation metrics. Objective validation approaches are able to assess whether requirements have been met, while subjective validation approaches assess whether expectations have been met. This section discusses the development of validation metrics for both subjective and objective validation approaches.

11.5.1 Validation Metrics for Subjective Validation Approaches

Validation metrics can be established through the use of rubrics when subjective validation approaches are used and the level of agreement between the MS&DT results and Real World referent is based on judgement. Rubrics are typically arranged in a tabular format with one axis containing criteria and the second axis containing achievement levels that indicate the degree to which the criteria have been met. The cells in the middle of the table are achievement level descriptors that relate to a specific criteria and achievement level. The descriptors should name specific discernable characteristics that can be witnessed using the selected subjective validation approach. The use of rubrics supports the documentation of subjective decisions as the level of effort to record the achievement level for each criterion is minimal.

To be effective, rubrics are recommended to have 3-5 achievement levels as the use of too many levels may reduce clarity for the user. Four achievement levels may be realized as: 1) does not meet expectation, 2) approaches expectation, 3) meets expectation, and 4) exceeds expectation. A logical progression in the descriptors is important for conveying the intent of the criteria to the user and enabling them to judge which achievement level has been met based on available evidence. When the descriptors are properly tied to the acceptance criteria, the four achievement levels listed previously enable a binary outcome from the subjective decision: 1 or 2 equating to inadequate versus 3 or 4 considered "good enough". A sample rubric is presented in Section 5/ Table 4 for a hypothetical MS&DT for predicting the vertical shear force along the length of a vessel. The rubric enables the quality of the prediction to be quantified through visual inspection with the criteria tailored to ensure poor results in the middle portion of the vessel are found unsatisfactory.

TABLE 4
Example Rubric – Vertical Shear Force Prediction

Criteria	*Does Not Meet Expectation 1*	*Approaches Expectation 2*	*Meets Expectation 3*	*Exceeds Expectation 4*
Vertical Shear Force Prediction	• Inaccurate peak value prediction • Incorrect curve profile between 0.25L and 0.75L	• Satisfactory peak value prediction between 0.25L and 0.75L • Correct curve profile between 0.25L and 0.75L	• Satisfactory peak value prediction for full vessel length • Close curve fit for majority of vessel length	• Accurate peak value prediction for full vessel length • Close curve fit for full vessel length
Example				

Holistic rubrics have one criterion and multiple achievement levels, are the most direct to apply and preferred as validation metrics. Analytic rubrics involve multiple criteria and multiple achievement levels. Use of analytic rubrics for validation is more challenging as the integration of multiple criteria introduces the need for scoring, weighting, and aggregating results from each criterion. The use of analytic rubrics should involve testing, revisions, and SME inputs prior to use.

11.5.2 Validation Metrics for Objective Validation Approaches

Objective validation approaches enable the quantification of differences between MS&DT results and Real World referents. Validation metrics for objective validation approaches may take on different forms based on the nature of the specific acceptance criterion and their hierarchy in the decomposition of the acceptance criteria. At the lowest hierarchal levels, a direct comparison between an item of evidence and the validation metric may often be made. Higher level criteria or criteria necessitating the use of multiple items of evidence may involve probabilistic approaches toward the development of validation metrics.

Validation metrics that involve the direct comparison between pairs of MS&DT results and Real World referents are more readily made at higher levels of decomposition such as the component level for a modeled system as the influence of error accumulation is not as significant. For example, a validation metric may place a limit on the difference between the predicted MS&DT temperature after a heat exchanger when compared to measured referent data at that location. Often at more detailed levels, experience and established practices may guide the acceptable limits incorporated into the validation metrics without requiring detailed uncertainty quantification.

When acceptance criteria require the consideration of multiple data points, validation metrics based on statistical analysis are possible. For example, separate distributions for MS&DT results and Real World referents could be established and mean values for both distributions could be calculated and compared. Different probabilistic validation metrics are outlined in ASME VVUQ 10.2-2021 [16].

13 Model, Simulation, and Digital Twin Integration in Operational Systems

There are additional considerations for V&V activities as they relate to the various roles that MS&DTs play in system acquisition. System acquisition is a process which includes the design, development, and production of new systems or the modification to existing systems that involve redesign of the system or subsystems. MS&DTs are often an important element of the system acquisition process to reduce the time, resources, and risk required to develop, evaluate and field a system.

13.1 Roles of Models, Simulations, and Digital Twins in System Acquisition

When considering the V&V activities associated with MS&DT integration with an operational system it is necessary to recognize the five roles that MS&DT has typically been used to support system acquisition.

13.1.1 MS&DT Precedes System Development

In Role 1, the MS&DTs are used to support concept exploration and program definition. The MS&DT precedes the operational system development, and is not updated as the system is developed, such that the MS&DT loses congruence with the system as it is developed.

13.1.2 MS&DT Supports System Development

In Role 2, the MS&DTs support the system during development. In this case, the MS&DT will often be continually updated as the system matures through the development process, however the MS&DT and the system remain as separate and distinct entities. In this role, the MS&DT will often follow a Model-Test-Model paradigm where the MS&DT is used to guide system development, and the developing system's test results are used to refine the MS&DT.

13.1.3 MS&DT is Component of the System

In Role 3, the MS&DT is one or more components of the system being developed and is completely embedded within the operational system.

13.1.4 MS&DT is the System

Role 4 can be considered as a special case of Role 3, where the system itself is solely the MS&DT. In this case the system hardware consists of only the platform required to operate the MS&DT, and the software system consists of only the MS&DT.

13.1.5 MS&DT Supports System After Acquisition

In Role 5, the MS&DT is used to support the system after system acquisition, which includes system monitoring, readiness, and sustainability.

13.3 Relationship Between Verification & Validation Activities and Test & Evaluation Activities

During system development, the process by which the system is examined and validated to ensure that it meets requirements and performs as intended is known as test & evaluation. Test and evaluation (T&E) is a process to assess the performance, reliability, and safety of systems, products, or technologies. By evaluating system performance against stated requirements, the end user can gain confidence in the system produced. In this way, the T&E process is similar to the V&V process described in this section in that its primary purpose is risk reduction, but T&E and V&V activities remain separate but complementary processes. These Guidance Notes do not intend to discuss the specifics of performing T&E, but the alignment of V&V activities with T&E activities is presented.

As described in Section 5/13.1, MS&DTs may hold several roles in the system acquisition process. The overall V&V framework will be the same regardless of the MS&DT role, however the MS&DT role may inform the criticality assessment and thus the specifics of the resulting V&V activities. The role of the MS&DT in the system acquisition process will impact the relevance of MS&DT V&V activities to T&E activities. The details of the impact for V&V activities on T&E activities for each MS&DT role are provided in Section 5/Table 5.

It should also be noted that during the T&E phase of system acquisition, MS&DTs may be used develop parameters for test cases, design tests, analyze data during testing, and evaluate regions of the operation space that are otherwise not testable. However, it is important to recognize that MS&DT results are not a substitute for testing.

TABLE 5
MS&DT Roles in System Acquisition and Relevance of Verification & Validation to Test & Evaluation [9]

CASE	*MS&DT*	*OPERATIONAL SYSTEM*	*NOTES*
1 Precedes System Development	• MS&DT used for conceptual definition of operational system.	• T&E activities are independent from V&V activities.	MS&DT are developed to support concept exploration and program definition phases of acquisition. MS&DT precedes system development but is not updated as the system matures. The MS&DT loses congruence with the system being developed except for the alignment in the requirements. Any V&V activities conducted will have little relevance to T&E of the mature system.
2 Supports System Development	• MS&DT supports concept definition. • MS&DT updated during development and test.	• Acquisition supported and guided by MS&DT for performance modeling and engineering tradeoffs. • T&E activities indirectly influenced by V&V activities	MS&DT supports a system under development. The digital representation of the system precedes system development and is updated as the system matures. In this case, the real system and the MS&DT are distinct entities. The V&V activities and the T&E of the system occur in parallel. Following the Model-Test-Model paradigm, the T&E and V&V processes complement and support each other. The MS&DT is used to guide the system development and the developing system's test results are used to refine the model.
3 Component of System	• MS&DT embedded in and developed as component(s) of operational system.	• V&V activities become part of T&E effort. • V&V and T&E activities directly support each other.	The MS&DT is a subset of the system. MS&DT are totally embedded within the operational system. This integration of the V&V and T&E processes yields three key benefits: commonality and reuse of testing techniques, value of conceptual modeling, and early correction of system problems.

CASE	*MS&DT*	*OPERATIONAL SYSTEM*	*NOTES*
4 MS&DT is the System	• MS&DT is the operational system.	• V&V and T&E activities are congruent.	The system under test is itself a MS&DT. The system hardware consists solely of the computer platform(s) required to run the MS&DT and the system software consists of only the MS&DT. In this case, the relationship between V&V and T&E processes is roughly congruent, with T&E being a subset of V&V due to the V&V process requiring some additional activities, such as code verification, that are not part of the T&E process. However, where problems are identified during T&E, there may be a need to examine the code or algorithms, hence T&E is an open subset of the encompassing V&V process.
5 Not Part of System Acquisition	• MS&DT used for operational system readiness or sustainment.	• Operational system already developed. --OR-- • Operational system already completed T&E activities.	MS&DT are developed for reasons not related to system acquisition. Since there is no system being acquired and no T&E activity (or T&E activities have already been completed), there is no relationship between V&V and T&E. MS&DT used for sustainment can utilize operational data to support V&V.

15 Operations and Maintenance

The lifecycle of a MS&DT will extend beyond the initial development phase. This section discusses what policies, procedures and processes should be in place to manage the ongoing use of a MS&DT. This includes management of change (MOC) processes for updates/enhancements to the MS&DT either due to the real-world system changing, improvements to the model, or new or changing intended uses.

15.1 Changes that Impact MS&DT

An understanding of the types of changes that could impact MS&DT is important. The documentation of the potential impact of any anticipated future changes at the MS&DT development stage is recommended to guide future MOC decisions. Several types of changes that may impact MS&DT include:

- *Modifications or enhancements to MS&DT code* – Changes to the MS&DT may be initiated by feedback from users including recommended improvements or observed errors. New operational needs may also trigger MS&DT changes.
- *Implementation of new procedures* – New Real World operational procedures may alter how the MS&DT is used and may impact the underlying assumptions made during MS&DT development.
- *Change to controlled documents* – MS&DT that rely on referent data from standards or other sources may require updates to comply with changes to the controlled documents.
- *Equipment modifications* – Changes to computing hardware or software such as operating systems may require updates to the MS&DT to ensure compatibility. MS&DT that rely on sensors may require updates when the sensors are recalibrated or replaced.
- *Changes to input data sources* – Changes to databases accessed by the MS&DT or the formatting and structure of data sets may impact the MS&DT.
- *New and forthcoming regulations* – MS&DT that must comply with regulatory requirements may require updates to ensure compliance with changes.
- *Changes to training data* – MS&DT that rely on data-driven algorithms may undergo updates when new training data is available.
- *Changes or modifications to physical assets (Digital Twin)* – Modifications to the physical assets may require updates to stored representations associated with the Digital Twin.
- *Change in lifecycle phase (Digital Twin)* – A change in the lifecycle phase of the physical asset associated with a Digital Twin.

15.3 MS&DT Lifecycle Management Policy

Establishing a MS&DT lifecycle management policy will assist with MOC needs and enable changes to be executed as efficiently as possible. Having a MS&DT lifecycle management policy in essential for ensuring V&V needs are met throughout all MS&DT lifecycle phases. It is recommended that a MS&DT lifecycle management policy:

- Indicate the organizational entities and individual roles that have responsibility for maintaining the MS&DT.
- Outline any required periodic lifecycle management activities.
- Define required actions following notification of changes, needs, or policies that impact the MS&DT.
- Provide an avenue for MS&DT users to provide feedback or report problems.
- Identify what organizational entities or individual roles should be notified following updates or enhancements to the MS&DT or the policies governing its use.

15.5 Management of Change Processes for V&V

Processes for governing MOC are beneficial for preventing negative impacts on MS&DT use due to changes and confirming that V&V needs are accounted for when changes are encountered. The following five procedures are recommended to be included in MOC processes that account for V&V.

- *Recognition of change* – A procedure for recognizing change is important to ensure that changes are identified in a timely manner and not overlooked. A procedure that is open to new suggestions for change can provide opportunities for improvement.
- *Impact assessment* – An impact assessment is necessary to identify what the scope of the change is and what risks it may introduce. The impact assessment procedure should include V&V as a factor for consideration. During the impact assessment it is recommended that the change is reviewed to determine if it will alter the V&V criticality categorization. If the criticality categorization following the change is expected to be higher than before, additional V&V activities may be required.
- *Change approval* – Voluntary changes may require approval to ensure the availability of funding and resources. All changes may also require approval to ensure that impacts on other operational activities are minimized.
- *Implementation planning and execution* – An implementation plan is necessary for managing and executing MS&DT changes. Based on the impact assessment, a V&V implementation plan should be part of the broader plan. As the scope of the MS&DT change project will be reduced relative to the original development project, the V&V framework presented in this section may be tailored to suit the needs specific to the change. Concept validation may be required if there is significant change to MS&DT functionality or intended use. It is expected that any MS&DT changes will require verification to ensure that the changes have been implemented correctly. A limited amount of result validation is expected to confirm MS&DT performance. More extensive result validation may be required if the MS&DT change results in a different or expanded application domain.
- *Documentation* – Any existing documentation from the original MS&DT development and V&V should be revised to reflect the changes made to the MS&DT, updated V&V results, and record any justifications for decisions regarding the validity of prior V&V work.

17 Special Considerations

17.1 Legacy Models and Commercial-off-the-Shelf Software

There may be instances where V&V activities may need to be considered for MS&DTs where there is limited documentation or traceability to the original development requirements, source code, and verification and validation results. This can create challenges when V&V documentation is required at a later date due to a new use case or the integration of these MS&DTs into a new model federation or digital twin configuration. Examples of these instances include legacy models that were created prior to the establishment of a robust V&V framework and general-purpose software packages and commercial off-the-shelf (COTS) applications. These instances often add complexity to the V&V process as certain details of the MS&DT required for specific V&V activities may not be available.

When considering acceptance claims for MS&DTs of this type, there are several approaches that can be applied to support the claims-reasoning-evidence approach when it is not possible to apply V&V activities directly for these MS&DT entities:

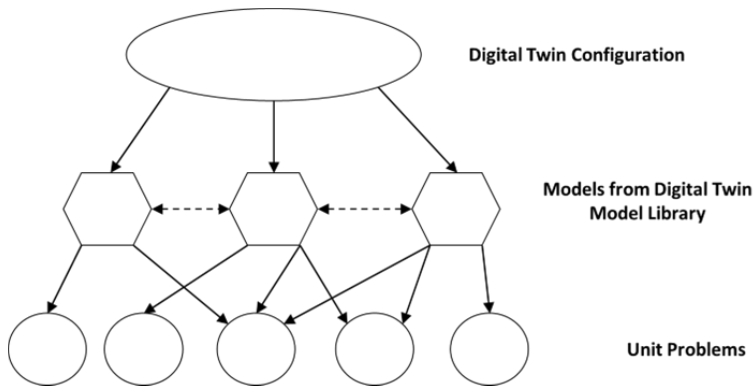
i) Where a legacy model or COTS application is employed as part of a larger integrated solution, the V&V activities at the individual MS&DT level may be minimized in favor of more robust V&V activities at the integrated MS&DT level. If the integrated MS&DT is successful in achieving the desired acceptance claim, the lower-level MS&DT can be viewed as also meeting its acceptance claim, for the specific intended purpose of the specific integration. However, if it is to be employed in another use case, additional V&V activities would be required.

ii) Where a legacy model or COTS application is intended to be employed as a stand-alone solution and access the necessary documentation and source code is unavailable, independent verification and validation techniques can be employed. These methods may include independent results validation, auditing of the developing organizations V&V processes, and consideration of the successful use of the MS&DT in other applications. However, successful application of the legacy or COTS MS&DT in other use cases should not generally be considered sufficient evidence for all future use cases by itself.

17.3 Digital Twins

Finally, there are several special considerations that arise when generating V&V activity evidence for digital twins as opposed to models and simulations. The first deals with the fact that digital twin configurations are often complex model federations, where each model must be independently verified and validated as well as the resulting model federation integration. In this way we can establish a validation hierarchy as shown in Section5/Figure 4. Thoughtful construction of this validation hierarchy is important because it will provide a path for the results validation requirements for the unit, component and sub-system models as well as defining the coupling and interactions that are expected to occur between the unit problems, component models and sub-system models. This clearly adds significant complexity and effort to the results validation process, and as mentioned previously, time and cost constraints may make it infeasible to gather detailed evidence for every element of the model federation. One approach could be to rely on a subject matter expert to estimate the uncertainties that exist within each level of the validation hierarchy and propagate these uncertainties to the model federation level. The risks associated with these approaches is that insufficient evidence on the validity of the component models may lead to a determination by the MS&DT Technical Authority that the overall digital twin is not approved for its intended use, or that it may only be used with certain restrictions.

FIGURE 4
Digital Twin Validation Hierarchy



Another special consideration associated with the V&V process for digital twins is the ongoing update process of the digital twin over time via the synchronization with data from the physical asset. This digital twin update process often results in the recalculation of both model states and model parameters. This can lead to the model moving from a solution domain that has been successfully validated to a domain that has not been fully validated. It is recommended that the acceptance decision made by the MS&DT Technical Authority be revisited on a regular interval to evaluate whether that decision still holds, or if additional V&V activities may be required to maintain that decision.

SECTION 6
Credibility Categorization

1 General

Assessing the credibility of MS&DT results is a natural part of the decision-making process and extends beyond verification and validation activities. Credibility is a measure of the level of trust that a user has in a MS&DT to provide correct results for an intended use case. Assessing MS&DT result credibility is challenging and may take on a variety of meanings depending on context. As credibility cannot be explicitly measured, a minimum set of criteria contributing to MS&DT credibility analysis are presented that can be examined more directly including verification and validation, but also incorporating additional factors is proposed.

The nine minimum credibility factors include data pedigree, verification, validation, use assessment factor, input pedigree, uncertainty characterization, result robustness, MS&DT model history, and MS&DT process management which are defined further in Section 6/Table 1 based on the NASA Handbook for Models and Simulations [1]. There may be other additional aspects relevant to the credibility assessment of a specific MS&DT that are not included in this minimum set. Tailoring the assessment to incorporate these additional factors or remove unnecessary ones is acceptable and encouraged.

The establishment of a formal credibility categorization process supports the proper application of MS&DT. Ideally, the user's perception of MS&DT quality aligns with the actual MS&DT quality that was assessed during V&V. That is, MS&DT of good quality will have high credibility with users while MS&DT of poor quality will have low credibility with users. When perceived and actual MS&DT quality are not in alignment undesirable operational outcomes will result. If a good quality MS&DT has low credibility with users, it will be underutilized. The frequent use of poor quality MS&DT due to unfounded high credibility with users can lead to the generation of incorrect results that will adversely impact decisions and outcomes in the Real World system.

The adoption of a formal credibility categorization process that requires the documentation of results is recommended for organizations that curate model libraries. When new models are developed using existing models from the library, the availability of prior credibility assessments will help guide the selection process. Prior credibility assessments can provide insights into the strengths and weaknesses of existing models. This is particularly important when existing MS&DT are used for new use cases with higher criticality than prior use cases. Review of prior credibility assessments can highlight which aspects of a MS&DT need improvement to satisfy the new use case. Does the MS&DT need to have V&V performed over a broader domain? Does the MS&DT need to be revised to provide greater fidelity?

3 Assessment of MS&DT Credibility

Assessing the credibility of MS&DT results is a natural part of the decision-making process. However, assessing MS&DT result credibility is challenging and may take on a variety of meanings depending on context. As credibility cannot be explicitly measured, these Guidance Notes present a minimum set of criteria contributing to MS&DT credibility analysis that can be examined more directly.

3.1 Credibility Factors

The nine minimum factors include data pedigree, verification, validation, use assessment factor, input pedigree, uncertainty characterization, result robustness, MS&DT history, and MS&DT process management. There may be other additional aspects relevant to the credibility assessment of a specific MS&DT that are not included in this minimum set. Tailoring the assessment to incorporate these additional factors or remove unnecessary ones is acceptable and encouraged.

TABLE 1
Credibility Factors [1]

Category	*Credibility Factor*	*Description*
MS&DT Development	Data Pedigree	Is the pedigree (and quality) of the data used to develop and validate the MS&DT adequate or acceptable?
	Verification	Were the MS&DT implemented correctly, per their requirements/ specifications?
	Validation	Did the MS&DT results compare favorably to the referent data, and how close is the referent to the real-world system.
MS&DT Operations	Use Assessment Factor	How well does the proposed MS&DT use match the permissible use(s) of the MS&DT? How similar is the proposed use of the MS&DT to previous successful uses?
	Input Pedigree	Is the pedigree (and quality) of the data used to setup and run the MS&DT adequate or acceptable?
	Uncertainty Characterization	Is the uncertainty in the current MS&DT results appropriately characterized? What are the sources of uncertainty in the results and how are they propagated through to the results of the analysis?
Supporting Evidence	Results Robustness	How thoroughly are the sensitivities of the current MS&DT results known?
	MS&DT History	How similar is the current version of the MS&DT to previous versions, and how similar is the current use of the MS&DT to previous successful uses?
	MS&DT Management	How well managed were the MS&DT processes and products?

3.3 Credibility Factor Assessment Levels

A summary of recommended level definitions for the credibility factors is provided in Section 6/Table 2 where level 2 represents higher credibility and level 0 is low credibility.

TABLE 2
Credibility Factor Assessment Levels [2]

	MS&DT Development			*MS&DT Operations*				*Supporting Evidence*	
Level	***Data Pedigree***	***Verification***	***Validation***	***Use Assessment Factor***	***Input Pedigree***	***Uncertainty Characterization***	***Results Robustness***	***MS&DT History***	***MS&DT Management***
2	All data is known & traceable to a sufficient referent. Data has acceptable accuracy, precision, and uncertainty	Formal practices applied to verify the MS&DT; all important errors satisfy requirements.	All key MS&DT outputs agree with referent data from a real-world system operating in a representative environment.	The proposed use case is similar to prior use cases and falls within the domain of prior V&V.	All input data is known and traced to the real-world system with acceptable accuracy, precision and uncertainty.	Uncertainty of results are provided quantitatively through propagation of all known uncertainty.	Sensitivities are known for most parameters.	At most, minor changes in MS&DT and use.	Controlled processes are applied and documented; process compliance is measured.
1	Some data is known and formally traceable with estimated uncertainties.	Documented practices applied to verify all MS&DT features; most important errors satisfy requirements.	Key MS&DT outputs agree with data from a sufficiently similar referent system.	The proposed use case is within the MS&DT limits but differs moderately from prior use cases.	Some input data is formally traceable with estimated uncertainties.	Sources of uncertainty identified and qualitatively assessed.	Sensitivities known for a few parameters and qualitatively assessed.	At most, moderate changes in MS&DT and MS&DT use.	Formal processes are applied and documented.
0	Some data may be known, but informally traceable.	Informal practices applied to verify some features of the model and assess errors.	Conceptual model addresses problem statement but limited to no comparison with referent data.	The proposed use case is new, but still in alignment with the domain, purpose and outputs of the MS&DT.	Some input data is known and informally traceable.	Uncertainty not identified or assessed.	Sensitivities not identified or assessed.	New MS&DT or major changes in MS&DT or major differences in MS&DT use	Informal processes applied

APPENDIX 1
References

Verification and Validation Documents

[1] NASA. (2019). "NASA-HDBK-7009A: NASA Handbook for Models and Simulations: An Implementation Guide for NASA-STD-7009"

[2] NASA. (2024). "NASA-STD-7009B: Standard for Models and Simulations".

[3] DoD. (2008). "MIL-STD-3022: Department of Defense Standard Practice – Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations".

[4] Thacker BH, Doebling SW, Hemez FM, Anderson MC, Pepin JE, Rodriguez EA. (2004). "LA-14167-MS: Concepts of Model Verification and Validation".

[5] Sargent RG. (2010). "Verification and validation of simulation models". Proceedings of the 2010 Winter Simulation Conference:166–83. https://doi.org/10.1109/WSC.2010.5679166.

[6] NATO/STO. (2015). "STO-TR-MSG-073: Generic Methodology for Verification and Validation (GM-VV) to Support Acceptance of Models, Simulations and Data".

[7] SISO. (2013). "SISO-GUIDE-001.2-2013, Generic Methodology for Verification and Validation (GM-VV) to Support Acceptance of Models, Simulations, and Data Volume 2: Implementation Guide".

[8] Roache PJ. (2009). "Fundamentals of Verification and Validation". Hermosa Publishers.

[9] Borowski M, Glasow P. (1998). "When Systems are Simulations – T&E, VV&A, or Both?". The MITRE Corporation.

Claim-Evidence-Reasoning Documents

[10] Bloomfield R, Chozos N. (2020). "CAE Mini-Guide 1: One-Page Guidance". https://claimsargumentsevidence.org/resources/the-declare-guidance/ (accessed June 12, 2024)

[11] Bloomfield R, Chozos N. (2020). "CAE Mini-Guide 4: CAE Connection Rules". https://claimsargumentsevidence.org/resources/the-declare-guidance/ (accessed June 12, 2024)

[12] Sklyar V, Kharchenko V. (2017). "Assurance Case Driven Design based on the Harmonized Framework of Safety and Security Requirements". Proceedings of the 13th International Conference on ICT in Education, Research and Industrial Applications. Kyiv, Ukraine. May 15-18, 2017. pp. 670-685.

[13] Ling Y, Mahadevan S. (2012). "Quantitative model validation techniques: new insights". Reliability Engineering & System Safety, Vol.111:217–31. https://doi.org/10.1016/j.ress.2012.11.011.

Professional Society Standards

[14] ASME. (2020). “ASME V&V 10-2019: Standard for Verification and Validation in Computational Solid Mechanics”.

[15] ASME. (2009). “ASME V&V 20-2009 (R2021): Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer”.

[16] ASME. (2022). “ASME V&V 10.2-2021: The Role of Uncertainty Quantification in Verification and Validation of Computational Solid Mechanics Models”.

[17] IEEE Standards, IEEE. (2022). "IEEE 1730-2022 Recommended Practice for Distributed Simulation Engineering and Execution Process (DSEEP)".

[18] IEEE Standards, IEEE. (2003). “IEEE 1516.3-2003 Recommended Practice for High Level Architecture (HLA) Federation Development and Execution Process (FEDEP)”.

Maritime Classification Documents

[19] ABS. (2024). “Generic Rules for Classification, Materials and Welding and Survey After Construction – Part 1D, Rules for Alternative Arrangements, Novel Concepts and New Technologies”.

[20] ABS. (2024). “Rules for the Building and Classing Marine Vessels”.

[21] ABS. (2013). “Guidance Notes on Management of Change for the Marine and Offshore Industries”.

[22] ABS. (2012). “Guide for Integrated Software Quality Management (ISQM)”.

Software Verification and Validation

[23] NASA. (2020). “NASA-HDBK-2203B: NASA Software Engineering and Software Assurance Handbook”.

[24] IEEE. (2016). “IEEE 1012-2016 - IEEE Standard for System, Software, and Hardware Verification and Validation”.

[25] ISO/IEC. (2023). “ISO/IEC 25010:2023 System and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Product Quality Model”

[26] ISO/IEC/IEEE. (2022). “ISO/IEC/IEEE 29119-1 Software and systems engineering – Software testing – Part 1: General Concepts”.

Digital Engineering Documents

[27] Office of the Under Secretary of Defense for Research and Engineering. (2019). “DoD Instruction 5000.97 Digital Engineering”. (2023). “DoD Instruction 5000.97 Digital Engineering”.

Digital Twins

[28] Dahmen U, Osterloh T, Rossmann J. (2019). “Operational Modeling for Digital Twins Using the Formal Simulation Model Configuration Framework”. International Journal of Simulation – Systems, Science and Technology – IJSST V20. https://doi.org/10.5013/IJSSST.A.20.06.07.

[29] Digital Twin Consortium. (2024). “The DTC Glossary”. URL: https://www.digitaltwinconsortium.org/glossary/glossary/ (accessed September 18, 2024).

APPENDIX 2
Definitions

Verification & Validation

V&V Core Terms	
verification	the process of determining the extent to which a model/Digital Twin is compliant with its requirements and specifications as detailed in its conceptual models, mathematical models, or other constructs. ("Did I build the thing right?").
validation	the process of determining the degree to which a model/Digital Twin is an accurate representation of the real world from the perspective of the intended use ("Did I build the right thing?").
criticality	a measure of the importance of the actions or decisions supported by a model, simulation, or digital twin. Criticality is based on the consequences of the model, simulation, or digital twin not providing a correct result and the degree to which it is integral to achieving the desired outcomes in the real-world system.
credibility	a measure of the level of trust that a user has of a model, simulation, or digital twin to provide correct results for an intended use case.
V&V Activities	
concept model validation	the evaluation of whether the theories and assumptions underlying the conceptual model are correct and reasonable for the intended purpose
code verification	relates to the function of the MS&DT Developer to produce code that is error-free, robust, secure, and reliable
solution verification	relates to the function of the MS&DT Developer who uses the code to obtain solutions to technical problems with sufficient accuracy
stored representation verification	relates to the function of the MS&DT User to correctly enter the descriptive data necessary to accurately represent the entity of interest
results validation	generating evidence on the degree to which the model is an accurate representation of the Real-World problem
Claim-Evidence-Reasoning Framework	
claim	statement that satisfies the original acceptance requirements
evidence	data that supports the claim
reasoning	justification that connects the evidence to the claim
V&V Framework elements based on Claim-Evidence-Reasoning	

acceptance goal	the V&V objective of ascertaining with a reasonable amount of certainty whether a MS&DT is or is not suitable for an intended use
acceptance claim	a statement affirming or rejecting the suitability of the MS&DT for the intended use
acceptability criteria	the standards of compliance and supported reasoning necessary to achieve the acceptance goal
acceptability claims	statements that document the degree to which the acceptability criteria are met
evidence solutions	the set of information necessary to ascertain compliance with all acceptability criteria
items of evidence	the V&V results obtained for the planned evidence solutions

Models, Simulations and Digital Twins

MS&DT Terms	
model	an idealized representation (physical, mathematical, or otherwise logical) of reality that consists of the key characteristics or behaviors of the selected system, entity, phenomenon, or process.
simulation	the imitation of the operation of a real-world process or system over time. Simulations require the use of models. The simulation represents the evolution of the model over time.
digital twin	a virtual representation of a physical asset, along with its environment and processes, comprised of integrated models that are updated through data exchange to provide decision-making support over its lifecycle. • the digital twin can extend across the digital thread and exist in all lifecycle phases: design, build and sustain. • the digital twin's scope and model fidelity are driven by desired insights and decision-making support
use case	a set of circumstances for detailing the use of something (in this case, a MS&DT) to achieve a desired outcome, which is typically associated with constraints or other requirements.
surrogate model	a model that mimics the behavior of a simulation model as closely as possible while being computationally cheaper to evaluate.
federation of models/ simulations	a system of interacting models and/or simulations, and a supporting infrastructure that are based on a common understanding of the objects portrayed in the system.
stored representation	structured information (in databases, CAD/CAM, BIM, GIS, point clouds, IoT streams and history, etc.) representing states and attributes of entities and processes at one or more times
computational representation	typically representing the functional relationship between a set of inputs and outputs contingent on a set of reference data
Multiple World Views in MS&DT Problem Solving	
real world	view that identifies the original problem based on operational needs / applies problem solutions to operations
problem world	view that identifies the purpose of the MS&DT in problem solving process / applies MS&DT results to solve problems
MS&DT world	view that identifies the MS&DT requirements / employs the MS&DT to generate results
V&V world	view that identifies V&V needs / applies the V&V framework to make an acceptance decision and supporting documentation
Organizational Dimensions for Model Libraries	

structural hierarchy	dimension that provides alignment of models in the library with physical system hierarchies, such as structural decomposition
model variant	dimension that refers to the possible variation of model fidelities, details, assumptions, etc. that may exist when establishing a conceptual model for the physical world.
model perspective	dimension that accounts for modelling the same system but considering different physical domains, e.g., mechanics, thermodynamics, hydrodynamics, etc.

V&V Management

Primary V&V Role	
MS&DT technical authority	role responsible for making the final decision on whether the MS&DT may be employed
V&V Organizational Roles	
V&V user/sponsor	organizational role responsible for specifying V&V needs and confirming the completion of V&V
V&V team	organizational role responsible for the execution of V&V activities
MS&DT user/sponsor	organizational role interested in MS&DT use and its outputs
MS&DT developer	organizational role responsible for developing the MS&DT
Organizational Levels	
enterprise level	organizational level responsible for establishing an environment that supports V&V
project level	organizational level responsible managing V&V projects to ensure their completion
technical level	organizational level responsible for supporting the engineering aspects of V&V efforts
V&V Individual Roles	
V&V enterprise manager	individual role responsible for confirming resource availability and creation of governing policies for V&V
V&V project manager	individual role responsible for the management of V&V projects
V&V leader	individual role responsible for developing the V&V plan and result integration
acceptance leader	individual role responsible for establishing the acceptance criteria and providing an acceptance recommendation
V&V implementer	individual role responsible for executing V&V tasks
SME (subject matter expert)	individual role responsible for providing technical expertise to support V&V

APPENDIX 3

Documentation

1 Introduction

The principal outputs of V&V activities are the acceptance claim and the supporting documentation which captures the MS&DT-specific inputs to the claim-reasoning-evidence process, and the resulting evidence generated from the V&V activities. The documentation generated from this process are normally prepared and used by different roles within the V&V process at different times throughout the process as well as after the completion of the process. The documentation to be created, its format, and the timing for the creation of this documentation should be tailored to align with the organization's policy and processes developed in Section 3 and Section 5 of these Guidance Notes. These policies and processes should also consider the tailoring of the scope of the documentation based on the criticality of the MS&DT based on the MS&DT criticality identified from the process outlined in Section 4. The consideration should be for providing a robust set of documentation that supports decision-making without becoming overly administratively burdensome.

This Appendix provides:

i) Descriptions of the types of documents that may be generated by the V&V process.

ii) Discussion of the factors that should be considered when aligning documentation policies with MS&DT development and V&V processes.

3 Types of Documentation for Verification and Validation

A robust V&V framework should have clear policies and processes for the generation of the format and content of documentation associated with MS&DT development and the corresponding V&V activities. As discussed in these Guidance Notes, the purpose of a V&V framework is to enable the MS&DT Technical Authority to make a decision regarding the use of a MS&DT for a specific use case or intended purpose. This decision is supported by evidence generated from V&V activities that is used to make acceptance claim(s). The purpose of documentation is to capture both the key inputs to the process which were used to plan and execute the evidence generating activities in addition to the corresponding results and recommendations.

In most implementations of the V&V framework, polices and processes will dictate that standardized approved templates are used to capture this information. However, it may also be reasonable to specify the content required to be documented and allow the MS&DT development team and V&V team to determine the most appropriate approach to capture the required information.

While the specific documentation templates will be tailored by each organization implementing the V&V framework, the types of documentation can be generally categorized as either planning documents or results documents.

3.1 Planning Documentation

Planning documentation should capture the elements of the V&V framework related to the development of the acceptance claim, its associated acceptance criteria, and the evidence generating methods. More specifically, it should capture the following as a minimum:

i) Methodology for tailoring the V&V activities to the specific use case or intended purpose.

ii) The proposed acceptance claim and associated acceptability criteria and metrics.

iii) Definition of V&V tasks (evidence generating methods) that will be employed to satisfy the acceptability criteria and metrics.

iv) Definition of the resources and other stakeholders required to perform the V&V and the required level of independence of these resources.

v) Identification of potential issues, limitations, or assumptions associated with performing the V&V activities.

3.3 Results Documentation

Results documentation should capture the evidence generated from the V&V activities and relate it to its corresponding acceptability criteria. Results documentation should show clear traceability of the reasoning linking the acceptability criteria to the corresponding results to the final acceptance claim. The goal for results documentation is to provide clear evidence to support the V&V acceptance claim and the resulting MS&DT Technical Authority's decision on MS&DT use. At a minimum, this should include:

i) Documentation of the results of each of the V&V activities down to the task level with traceability to original requirements and acceptability criteria.

ii) Documentation of MS&DT criticality, assumptions, capabilities, limitations, risks, tailoring, and impacts.

iii) Identification of unresolved issues associated with V&V implementation.

iv) Documentation of recommendation and lessons learned during V&V.

v) Actual resources expended on V&V activities.

3.5 Check Sheets

Check sheets are documents that help ensure that all steps associated with specific V&V activities have been completed. As personnel complete individual V&V tasks they can mark their progress using the check. sheet. Individual checkboxes within the check sheet may cover steps such as documenting the source of V&V evidence, confirming that the correct acceptance criteria were applied, noting whether each criterium has been met or not, and ensuring that data has been properly stored.

3.7 Tailoring of Documentation Requirements

When a section or sub-section of the V&V document required information is identified as being not applicable, best practice is to retain that section within the documentation but mark it as "This section is not applicable." which ensures completeness and clarity for future readers of the documentation. Note that this tailoring effort can be based on the criticality of the MS&DT use case as identified in Section 3 of these Guidance Notes.

5 Example Documentation Templates

There are no prescribed templates for V&V activity documentation. A robust V&V program should establish policies for documentation that align with the organization, the types of MS&DT being developed, and the role of the organization in the decision-making process. Additionally, it is also important that even with established templates, the processes allow for tailoring of the required inputs to align with the specific use case.

5.1 Department of Defense MIL-STD-3022

Department of Defense MIL-STD-3022 provides templates for documenting both V&V activities as well as accreditation activities. These documents are structured as planning documents and results documents and would align with both concurrent V&V execution as well as post-hoc V&V execution. An overview of the V&V document templates is shown in A3/Figure 1. Details on what is recommended to include in each sub-section of the document can be found in the MIL-STD-3022 reference [3].

Figure 1
MIL-STD-3022 Core V&V Documents [3]

V&V Plan	V&V Report
0. Executive Summary	0. Executive Summary
1. Problem Statement	1. Problem Statement
2. MS&DT Requirements and Acceptability Criteria	2. MS&DT Requirements and Acceptability Criteria
3. MS&DT Assumptions, Capabilities, Limitations & Risks/Impacts	3. MS&DT Assumptions, Capabilities, Limitations & Risks/Impacts
4. V&V Methodology	4. V&V Task Analysis
5. V&V Issues	5. V&V Recommendations
6. Key Participants	6. Key Participants
7. Planned V&V Resources	7. Actual V&V Resources Expended
	8. V&V Lessons Learned
Suggested Appendices	Suggested Appendices
A. MS&DT Description	A. MS&DT Description
B. MS&DT Requirements Traceability Matrix	B. MS&DT Requirements Traceability Matrix
C. Basis of Comparison	C. Basis of Comparison
D. References	D. References
E. Acronyms	E. Acronyms
F. Glossary	F. Glossary
G. V&V Programmatics	G. V&V Programmatics
H. Distribution List	H. Distribution List
I. Accreditation Plan	I. V&V Plan
	J. Test Information

5.3 Example Shipyard Templates

Several case studies were conducted at shipyards employing MS&DTs in the design, construction, and sustainment of marine assets. These case studies involved the development of robust V&V practices at these organizations based on the V&V framework outlined in these Guidance. The developed policies and processes were then applied to several example MS&DTs to demonstrate the approach. One of these case studies, along with the corresponding documentation templates is discussed in Appendix 4.

APPENDIX 4

Case Study for the Application of Verification and Validation Framework

1 Introduction

For demonstration of the V&V framework presented in these Guidance Notes, a case study was identified and evaluated. The activities presented in this case study demonstrate the practical implementation of the V&V framework with all the considerations that would bound any real-world case including criticality, tailoring, time, and budget. The case study presented involves the development of a digital twin for a small class unmanned underwater vehicle (UUV) used to support the design and testing of the vehicle control software.

This Appendix provides:

i) Overview of a portion of the model library and digital twin configurations

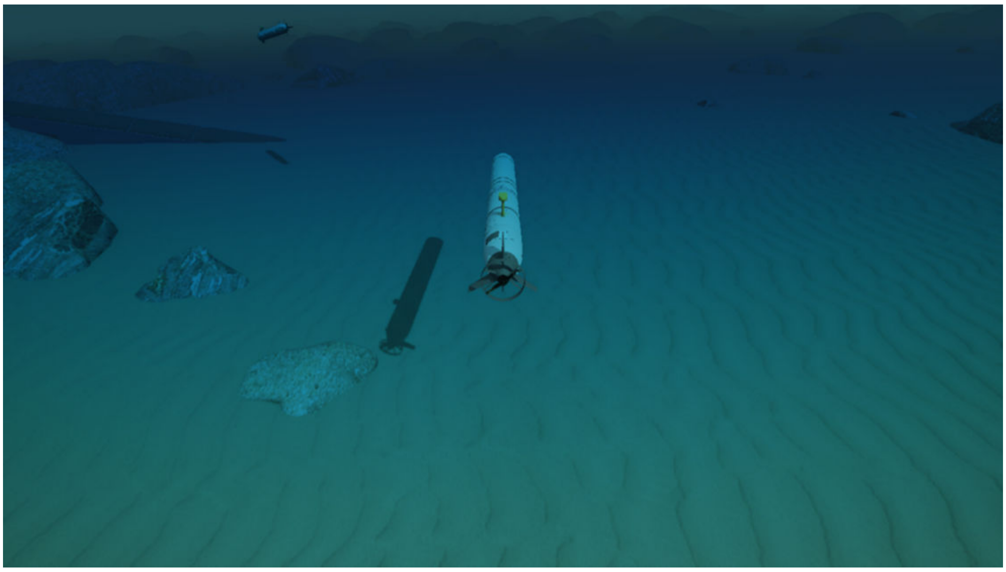
ii) Description of the organization roles and responsibilities supporting MS&DT development and V&V activities

iii) Presentation of the V&V process and resulting documentation and recommendations

3 Digital Twin Overview

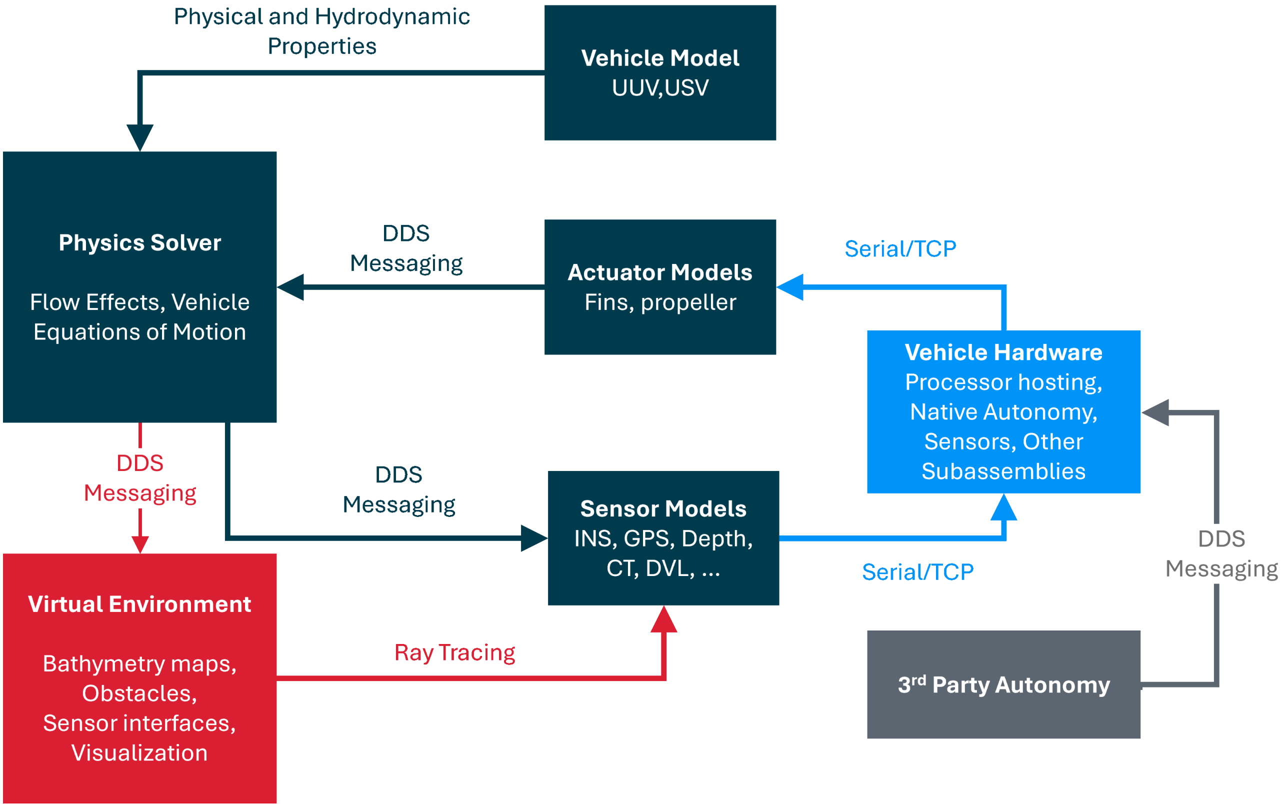
The complexities of designing robots to withstand harsh ocean environments has prompted the development of a UUV digital twin (DT) to aid in the decision-making process. The DT is a physics-based representation of a UUV, its processes, and its environment. The DT user can vary those representations to obtain useful output that would otherwise be prohibitively time consuming, expensive, or impossible to generate without first building a UUV.

FIGURE 1
Simulated Small- and Medium-class UUVs



The UUV DT is comprised of four major components: (1) the virtual world, (2) the vehicle model, (3) sensor models, and (4) the hydrodynamic physics solver. The virtual world is populated with bathymetry maps, ocean turbulence, water density profiles, sea state effects and physical obstacles in an integrated framework built on Unreal Engine and supporting processing nodes. The vehicle model is represented by the hydrodynamic hull signature and actuators populated from a suite of high-fidelity computational fluid dynamics (CFD) simulations. This vehicle is outfitted with simulated sensor models developed as mathematical representations of the sensor functions with specifications applied based on vendor data sheets and real-world data. Some of these simulated sensors include an inertial navigation system (INS), global positioning system (GPS), doppler velocity log (DVL), depth, and conductivity and temperature (CT). The simulated vehicle interacts with the virtual world through its sensors and hydrodynamic properties. The high-fidelity physics solver uses these interactions to solve a set of non-linear equations integrated in time for the dynamic vehicle state. This DT is developed in a modular framework that allows models to be used when needed and combined to represent various vehicle configurations. Each model publishes its output information onto the data distribution service (DDS) bus for any other model to subscribe to.

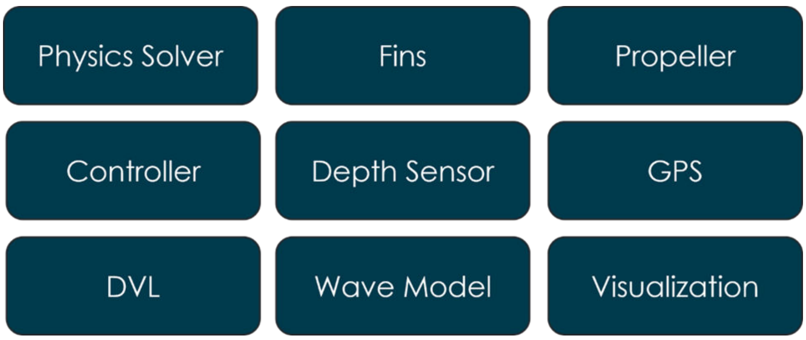
FIGURE 2
Simulation Architecture for UUV Digital Twin



3.1 Model Library

The model library is a collection of model elements that can be combined to represent the digital twin. For the UUV DT, models are essentially the discretization of the real world into its virtual, mathematical counterparts. Each model has a distinct purpose in the virtual world and when combined, forms a complete representation of the UUV.

FIGURE 3
UUV Digital Twin Model Library Examples



The fidelity of each model varies, both in software complexity and digital twin classification level. Many of the models are classified as digital twin prototypes because they do not exchange information with the corresponding real-world systems. Other models have modes to allow connections to real hardware and are an example of a complete digital twin. The overall complexity of the DT is influenced by how the models are linked together and deployed.

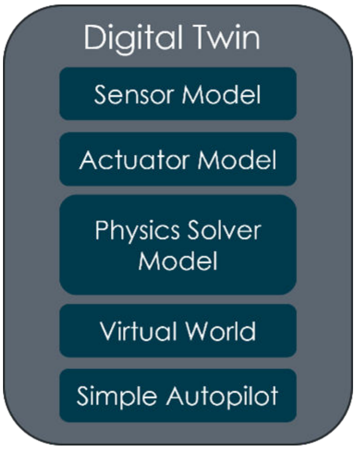
3.3 Digital Twin Configuration

A digital twin configuration is a set of models that are linked in a certain way to run a simulation. Due to the modular nature of the DT, many configurations can be formed depending on the use case. Three of the most common configurations for the UUV DT example are noted below and are referenced throughout this report.

3.3.1 Autopilot Configuration

The autopilot configuration runs the DT prototype as a self-contained software package (i.e., does not exchange information in real time with a physical asset(s)). The autopilot model is responsible for publishing actuator commands and running the vehicle control process.

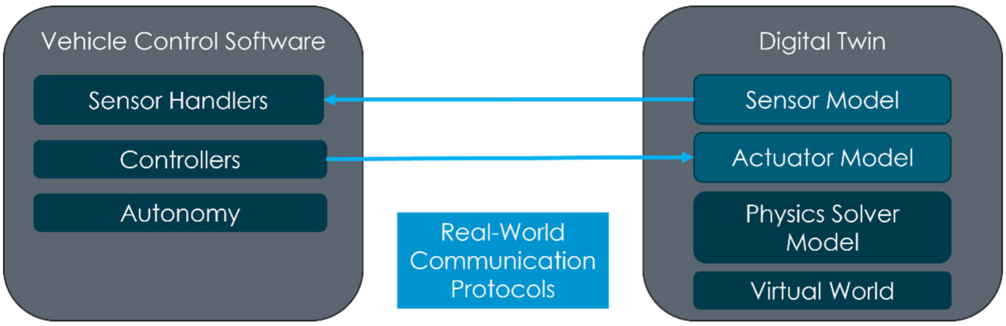
FIGURE 4
Example Models Used in a UUV Digital Twin Autopilot Configuration



3.3.2 Software in the Loop Configuration

The software-in-the-loop (SIL) configuration links the DT to an external piece of software. In the UUV domain, this software is commonly a vehicle control program, which runs on the internal hardware of the UUV. The vehicle control software is responsible for sensor handling, control, and autonomous decision making. By linking this software with the DT, a virtual mission can be executed using the vehicle dynamics, sensors, and environment provided by the digital twin.

FIGURE 5
Example Models Used in a UUV Digital Twin SIL Configuration

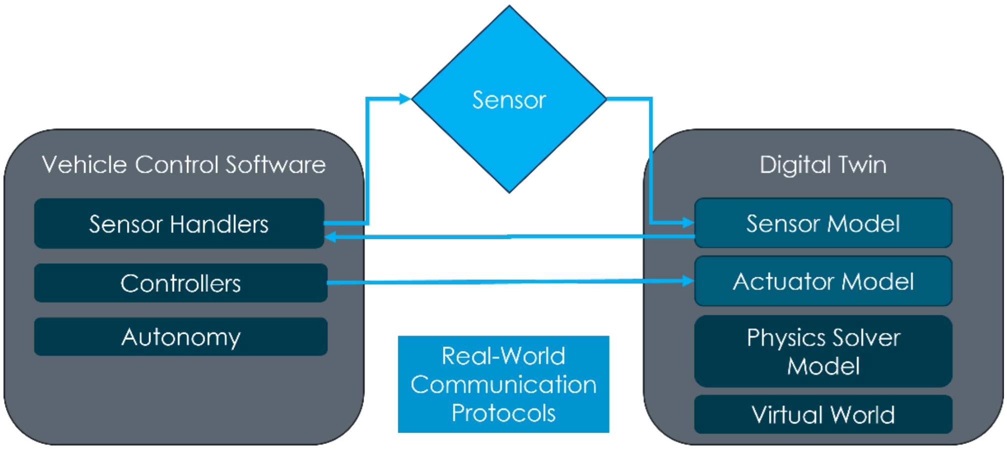


Section A4/Figure 5 shows the data flow for a representative SIL configuration. The autopilot model is disabled on the DT side, as commands originate from the controllers in the vehicle control software. Additionally, the sensor and actuator models must operate in a different mode to facilitate communication protocols with the external software.

3.3.3 Hardware in the Loop Configuration

The hardware-in-the-loop (HIL) configuration links the DT to both external software and external hardware. Again, different models and modes of operation are pulled into the DT configuration to support these linkages. An example single sensor HIL configuration is shown in A4/Figure 6, but this concept can be extended for vehicle subsections and entire UUV-in-the-loop configurations.

FIGURE 6
Example Models Used in a UUV Digital Twin HIL Configuration



3.5 Configuration Considered for Use Case

When performing V&V on a digital twin, the use case is an essential aspect to establish up front, as it has a large impact on whether the V&V activities provide acceptable results. Example use cases for the UUV DT include use as a testbed for developing autonomy algorithms, training autonomy algorithms, troubleshooting hardware, feasibility studies for new intended vehicle operations, or operator training. These would each pose very different V&V tests and success criteria.

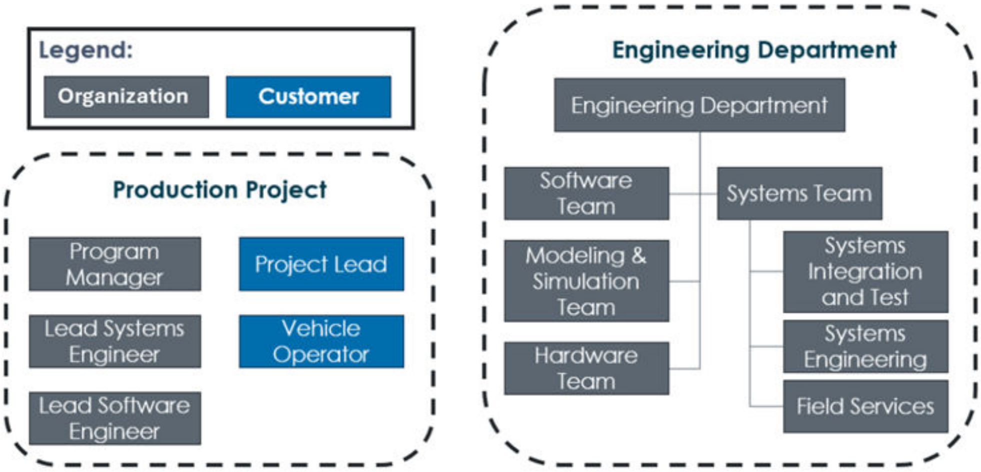
At this point, it can be helpful to return to the Real World, Problem World, MS&DT World view established in S1/1.3 to hone in on a specific use-case. A real-world problem is the risk of faulty vehicle control software when it is used the first time at sea, which could cause mission failure and testing delays. This leads to the purpose statement in the problem world, which is to implement testing on vehicle software in a simulated environment to identify and fix problems earlier in the design cycle. Lastly, this purpose statement is used to define the requirements of the modeling and simulation world: this testing requires a simulation environment with an accurate dynamic response, and it must implement all navigational sensors and actuators with the vehicle control software in the loop.

This Appendix will discuss the V&V for the digital twin of a small class UUV when used for design and testing of the vehicle control software. This DT application uses a SIL configuration introduced in Section A4/3.3.2. The V&V of individual models will be discussed in the context of the overall digital twin, as its modular nature lends itself to activities on a model-by-model basis.

5 Organizational Roles and Responsibilities

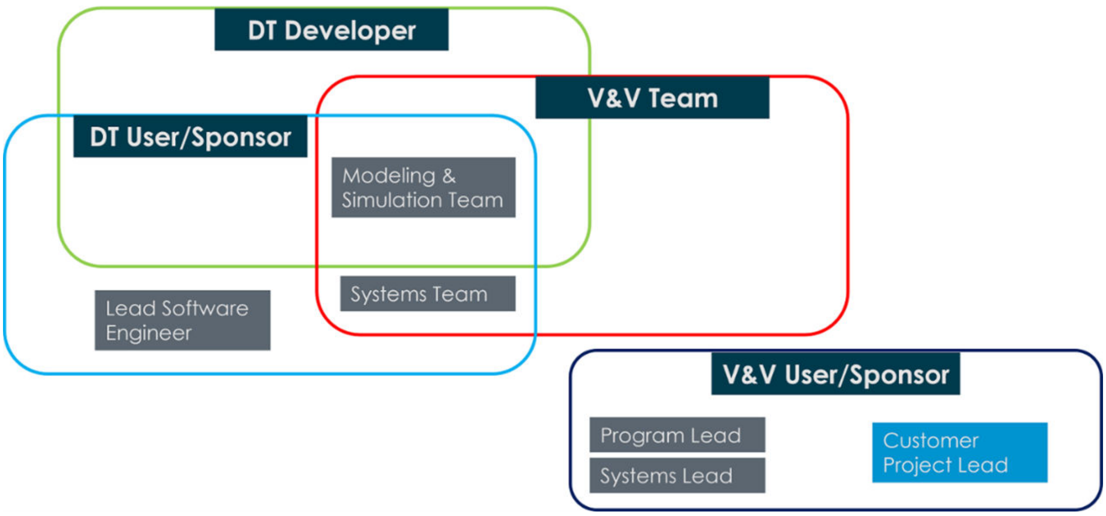
Though the V&V stakeholder roles outlined in Section 3/5 can vary per use case of the digital twin, a typical example is outlined herein. Most projects are a combination of engineering resources and a dedicated project team, which may have considerable overlap. This is depicted in the diagram of roles in A4/Figure 7.

FIGURE 7
Example Organizational Structure



For vehicle production projects, there exists a project team comprised of both organization and customer personnel (shown on the left) and the supporting technical team (right). The example DT use case introduced in Section A4/3.3 is exercised in these production projects as a testbed to augment any new vehicle software development and for final verification of vehicle software before deployment at sea. For the DT to be used and trusted for such a purpose, it first undergoes its own V&V process where the key players from the diagram are the customer project lead, organization program manager, lead software engineer, lead systems engineer, systems engineering team and the modeling and simulation team. These V&V roles for the digital twin itself are detailed in the following subsections.

FIGURE 8
V&V Organizational Roles



The project and engineering team members can also be organized into the V&V stakeholder view as shown in A4/Figure 8. This is a helpful tool to conceptualize the independence of DT V&V effort, which is low for this example digital twin with significant overlap of roles depicted in Appendix 4/Figure 8. The V&V of the digital twin is primarily conducted by the development team, with some input from other groups in the same organization. This matches the organizational structure introduced in Section 3/5.1, where the V&V effort is fully executed at the technical level with minor oversight and acceptance at the project level.

5.1 Key Roles

The key MS&DT and V&V roles are highlighted below:

5.1.1 MS&DT Technical Authority

The MS&DT Technical Authority is the stakeholder responsible for determining if a DT is acceptable for a specific use case. For the reported use case, the small class UUV customer and the program manager on the manufacturing contract have the final say on the use of the DT.

5.1.2 DT User/Sponsor

The DT User/Sponsor is the end user of the digital twin, which for this use case, is the lead software engineer on the small class UUV vehicle delivery. The DT is used real-time during the software development process as a testbed for design and testing before the physical vehicle is ready for in-water deployment.

5.1.3 V&V Team

The V&V team is responsible for the execution of the V&V activities. The modeling and simulation team conducts most of the V&V as part of the DT development process and standardized testing. The systems team also carries out V&V activities through regular use of the digital twin for various use-cases.

5.1.4 Digital Twin Developer/ Software Development Team

The Modeling & Simulation team is responsible for the development of the digital twin.

5.1.5 Subject Matter Expert (SME)

The lead systems engineer on the small class UUV contract is the SME for the use case of the DT. They have a deep understanding of the systems in place on the specific vehicle that the DT is being developed and used for, making them a valuable resource during DT development and V&V activities.

7 Verification and Validation Process

This section will cover the process implemented by the Modeling and Simulation team to complete verification and validation activities during the development and testing of the digital twin, following the framework laid out in these Guidance Notes. The result of this process is a V&V report, one per individual model and one per DT configuration use case.

7.1 New Feature Identification and Requirements

When a manufacturing contract is established, the modeling and simulation team works closely with the systems and software engineering teams to determine the digital twin requirements. New models or upgrades to existing models are identified to support the new vehicle variant. Requirements and development are tracked using Jama and Jira (requirements and workflow management respectively) tools and linked to the final V&V report created by the team.

For the test case outlined in A4/3.3, the digital twin must mimic the vehicle dynamic response, processes, and operating environment of the small class UUV. If the digital twin data is proven to accurately represent the real vehicle's motion through the environment and all feedback from sensors and processes, the software team (DT users) can use the digital twin to test vehicle control code in the virtual environment. The requirements compiled in A4/Table 1 represent a sample of the overarching goals for this use case and are the result of discussion between the DT Users and the DT developers.

TABLE 1
Small Class UUV Digital Twin Requirements for Vehicle Control Software Testing Use-Case

Requirement	*Acceptance Criteria*	*Metrics*
UUV DT sensors, actuators, and physics solver should individually satisfy their Model V&V requirements.	Model V&V complete for each sensor, actuator, and physics solver model individually. These include: depth sensor, conductivity-temperature sensor, DVL, GPS, INS, fins, propeller, motor controller, physics solver, and water profile.	See other Model V&V documents.
UUV DT shall satisfy all communication needs to vehicle software.	All necessary communication messages are present, sent to proper port, and at the proper frequency.	All annunciators in vehicle interface program indicate green/yellow status.
UUV DT shall capture similar steady state flight to actual vehicle.	Pitch response percent error as compared to sea data is lower than maximum threshold percent error required per vehicle specifications.	Pitch response from the simulated vs actual vehicle log files.
UUV DT shall achieve depth control comparable to actual vehicle.	DT depth response maintains a percent error lower than the maximum threshold as required per vehicle specifications when compared to sea data depth during steady level flight mission legs.	Depth data from the simulated vs actual vehicle log files.
UUV DT shall capture similar ascent slope to the actual vehicle.	DT depth slope is lower than maximum threshold percent error required per vehicle specifications when compared to sea data.	Depth response slope during ascent captured in the simulated vs actual vehicle log files.
UUV DT captures all necessary objectives as defined by the UUV mission file.	Visual inspection that mission track lines correlate.	Latitude and longitude data in the simulated vs actual vehicle log files.

The requirements for individual model V&V activities can differ in format from the use-case requirements. Many sensor requirements are driven by vendor specification sheets and communication protocols with vehicle control code, while actuator requirements are based on higher order modeling inputs. In the next phase of the report, the validation of both an individual sensor and an actuator will be discussed. Example templates for both the model and digital twin configuration V&V workflows can be found in Sections A4/9.1 and A4/9.2.

7.3 Criticality Determination

Before beginning the validation and verification activities, a criticality assessment is completed for each model or digital twin configuration. This determines how rigorous the V&V process should be. The assessment looks at various criteria about the model's importance in the overall digital twin configuration, such as number of other models its data impacts, the effect that model failure has on the overall outcome, and whether similar data is available or has been used for decision making in the past.

One step of the criticality assessment is the tailoring of the criticality criteria provided in Section 4 of these Guidance Notes to better fit both the model and DT configuration V&V activities for the present organization's process and use case. The tailored decision consequence level tables are shown below (A4/Table 2 and A4/Table 3). The criteria for the digital twin configuration are more use-case dependent compared to the model matrix.

TABLE 2
Tailored Consequence Level Criteria for UUV DT Configuration and Use Case

Decision Consequence Level	*Effects of Model Failure on Decision-Making Process*
L	The results from the digital twin are for informational purposes only and will not directly feed into the actual vehicle under operation at sea. *i.e., Test feasibility of a vehicle operating under new environmental conditions or mission objectives. Or Attempt to recreate a customer reported issue.*
M	The digital twin resulting in incorrect data is not probable to cause dangerous situations for human safety, safety of the asset, and/or threat to the environment but could lead to inability of vehicle to meet its intended objectives. *i.e., Testing low-risk autonomy routines or performing low-risk mission functions tests prior to factory acceptance sea testing.*
H	The digital twin resulting in incorrect data has a moderate-high probability to cause a dangerous situation for the human safety, safety of the asset, and/or threat to the environment. *i.e., The results of the DT are feeding an application that will be deployed to the customer without thorough on-vehicle testing.*

TABLE 3
Tailored Decision Consequence Level Criteria for UUV DT Model

Decision Consequence Level	*Effects of Model Failure on Decision-Making Process*
L	If the data is incorrect, a usable data set can still be achieved via the digital twin. Model does not interact with many other models (1-3)
M	If the data is incorrect, the digital twin will still execute but the data may not be accurate. Model interacts with other models (3-5)
H	If the model's data is incorrect, the digital twin will not operate. Model interacts with many other models (5+)

The resulting consequence level is combined with the results integration level and supplementary information availability level to calculate a model influence level. This is required to determine overall criticality of the model or digital twin configuration and thus determine the magnitude of V&V required. The full steps are outlined in the documents in Sections A4/9.1 and A4/9.2.

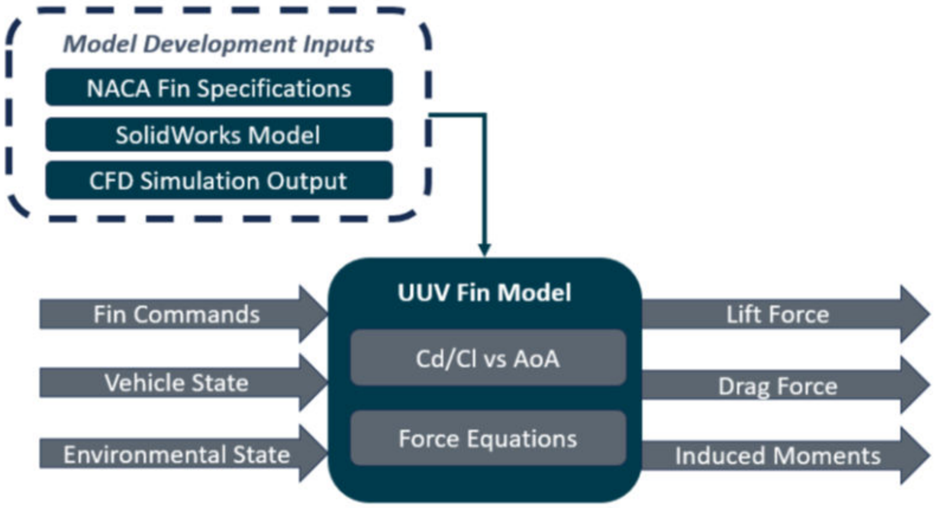
7.5 Conceptual Model Validation

Conceptual model validation answers the question of whether the theories and assumptions used to create the model are correct and if the model representation of the real-world component is reasonable for the use-case. The goal of this V&V activity is to document this information before and during model development in a standardized format, as both V&V report content and a resource for future work. The conceptual model validation also clearly bounds the problem space and use-case of each model, providing hard evidence when discussing applications of the overall digital twin. Two individual models are presented below as example cases for the conceptual model validation process: (1) actuator models and (2) doppler velocity log (DVL) sensor model.

7.5.1 Actuator Model

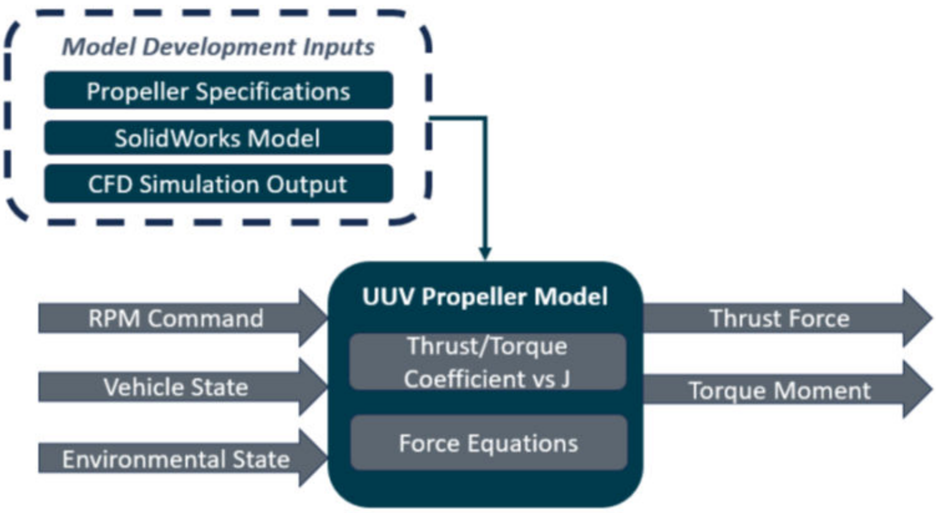
A UUV's fins and propeller are the main actuators of the UUV. The fins are a specialized model in the DT because the model content is largely driven by computational fluid dynamics (CFD) modeling completed for specific fin dimensions. This CFD output defines the lift and drag coefficients (Cd and Cl) based on angle of attack (AoA) that feed into the DT fin model, along with lift and drag force equations.

FIGURE 9
Model Inputs, Content, and Outputs for UUV Fin Model



These model development inputs are completed once and inform the developers who are writing the model. A different set of inputs is required to run the model. These include fin position commands, vehicle velocities, rates and wave elevation and density from the environmental state. The outputs from the model include lift force, drag force, and induced moments.

FIGURE 10
Model Inputs, Content, and Outputs for UUV Propeller Model

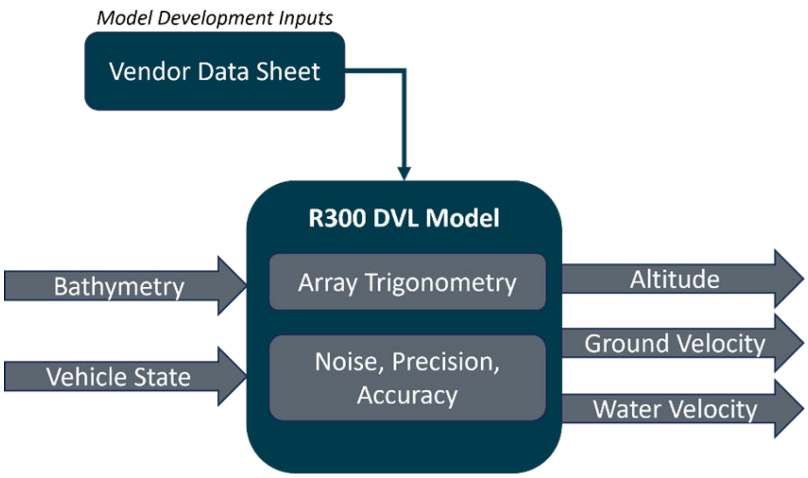


As seen in A4/Figure 9, a propeller model requires similar model development inputs, including a CAD model and CFD simulation outputs, along with the unique physical propeller specifications such as number of blades, propeller diameter, blade pitch, hub diameter and wake. It is also common to develop the model based on tank test data as opposed to CFD. The model itself operates under the same general input, processing, and output framework as the fin model. The model requires commands, vehicle state, and environmental data as input, force calculations make up the bulk of the model content, and force and moment values are output for use by other models.

7.5.2 DVL Model

A doppler velocity log (DVL) is a UUV sensor that measures water speed, speed over ground, and distance from ground (altitude) of the vehicle. The DVL is another specialized model in the DT because it interacts with the virtual environment. The virtual DVL uses ray tracing to interact with the bathymetry surface in the virtual ocean environment. This data becomes the basis for model calculations that convert the ray lengths into altitude, based off the positioning of the DVL array. The other input needed to run the DVL model is vehicle state, used for ground speed and water speed. The model applies error to these values based off the real sensor datasheet that the DVL is mimicking and outputs the noisy values.

FIGURE 11
Model Inputs, Content, and Output for the UUV DVL Model



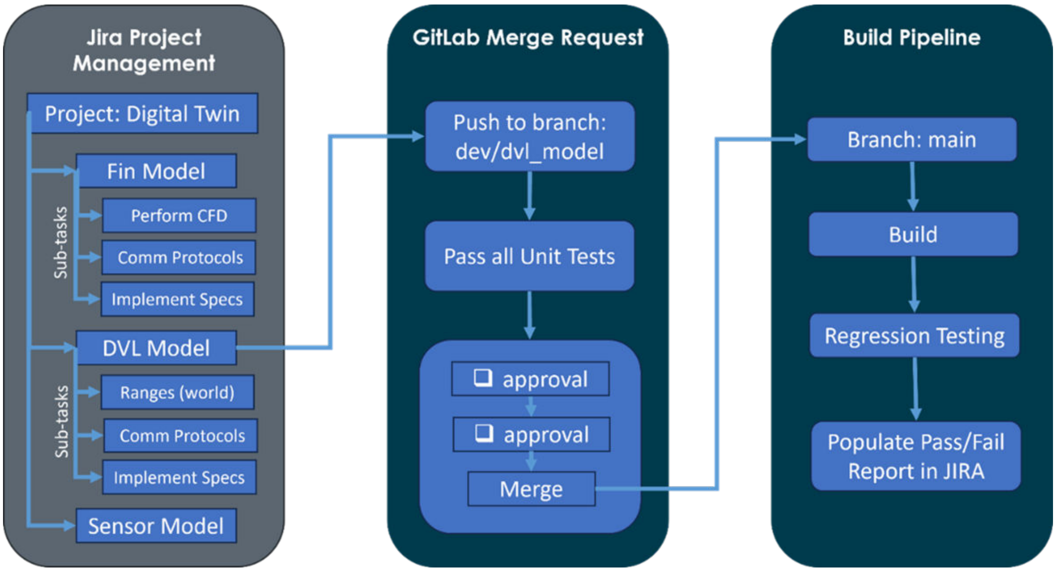
7.7 Model Verification

The next process implemented as part of the DT development and V&V framework is model verification, which determines if the model implementation sufficiently satisfies the requirements of the use case. Model verification is also documented as part of the individual model V&V process for the DT. Model verification methods fall under two categories outlined in Section 5/9 of this Guidance Note: Code Verification and Calculation Verification. The model verification practices for the UUV DT outlined in this Appendix are primarily code verification methods including both software quality assurance and numerical algorithm verification.

7.7.1 Code Verification: Software Quality Assurance

The Modeling and Simulation team is the primary digital twin developer and is responsible for conducting software quality assurance. At the beginning of the model development process, a new ticket is created in Jira and assigned. This ticket serves as a compilation platform for any information, bugs, anomalies, and testing sequences during the model development. The team also uses GitLab for configuration management of the digital twin source code. Individual branches are used by developers to make and test any new model, before merging work into the main branch. A detailed code review is completed at the merge stage, and approval is needed from two team members before any change is accepted. Additionally, tagged versions of the digital twin code base are created and saved after major changes. These Jira tickets, GitLab merge requests, and any other development work are included in the model V&V document as part of model verification (A4/Figure 11).

FIGURE 12
Model Development Workflow Example



7.7.2 Code Verification: Numerical Algorithm Verification

The Modeling and Simulation team is also responsible for numerical algorithm verification, which determines the correctness of the algorithms implemented in the model code. The verification rigor in this category scales depending on the model complexity. The UUV DVL model implements straightforward formulas and therefore automated unit tests are sufficient for verification. Unit tests are implemented for all models as baseline testing that output simple binary pass/fail reports.

The UUV fin model outlined in Section A4/7.5.1 has a more complicated formulation, therefore the developer also conducts integration tests using the autopilot configuration to verify the model. An example mission includes a sandy flat seafloor, no environmental effects, and commands the vehicle to fly a box or lawnmower pattern (A4/Figure 12). Visual inspection is used to verify that the vehicle dives, surfaces, turns left, and turns right when commanded to do so. Evidence from the unit tests and baseline autopilot missions is gathered and submitted as part of each model's V&V document.

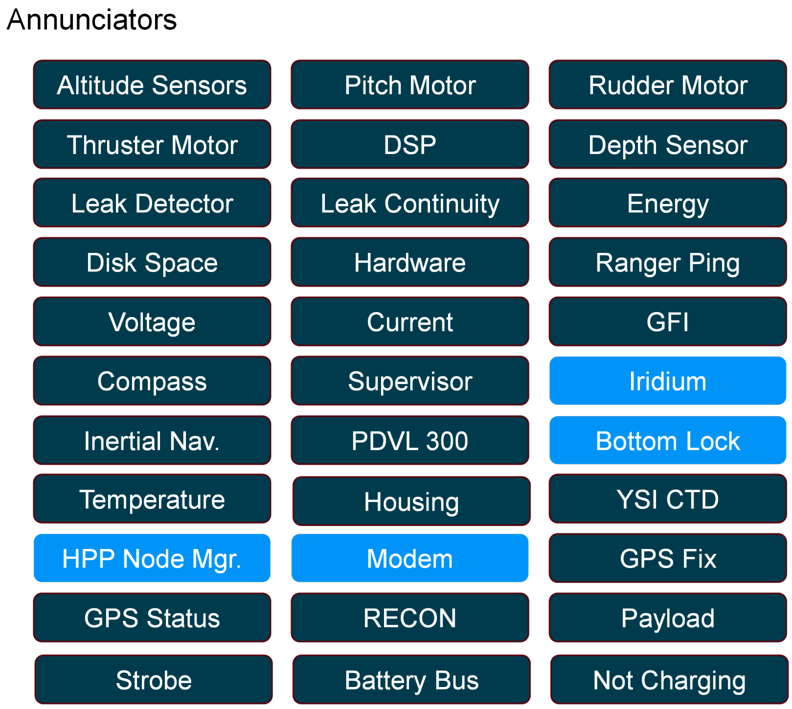
7.9 Results Validation

The last process conducted as part of the V&V activities is results validation, which determines if the digital twin is an accurate representation of the real-world problem. A key component of results validation is comparison with real-world data, so the predictive capability of the digital twin can be understood. This process is completed by the DT developer, but input may be needed from SME's and the DT users when performing analysis.

At this point, the focus shifts from the individual model level to the overall digital twin configuration. The entire digital twin must be exercised and compared to real world data to determine if it meets the use case criteria. For the example use case, the digital twin is used to test the small class UUV control code in the virtual environment. Since ample sea data exists for the UUV, a software-in-the-loop mission can be completed that mimics conditions from the real world. By comparing the outputs from the two missions, the digital twin's predictive capability is established. The full DT configuration validation is captured in a separate document shown in A4/9.2 rather than the individual model V&V shown in A4/9.1.

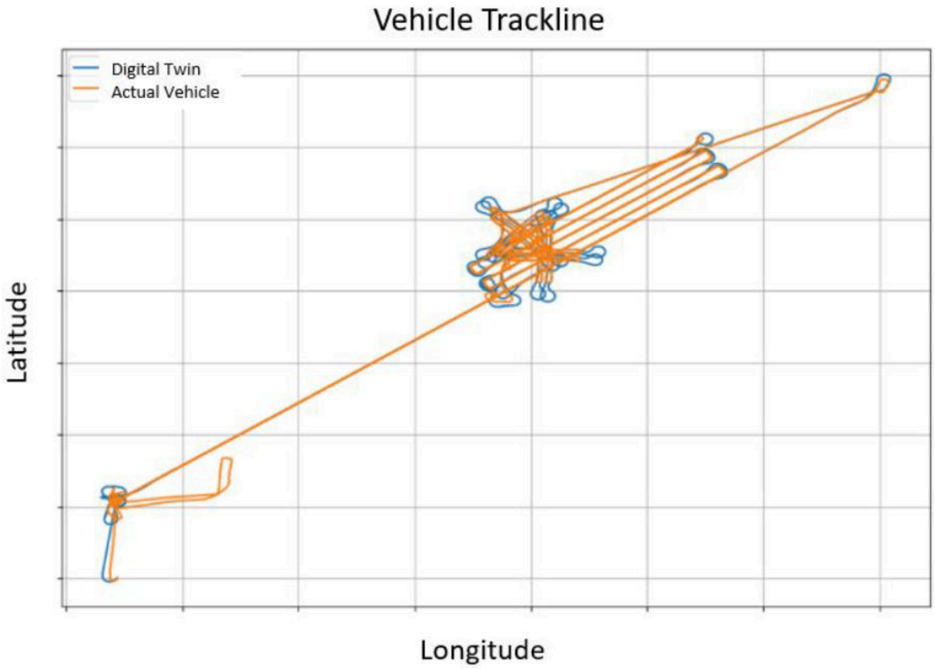
The general SIL validation process can be broken down into three major steps. The first is establishing communications between the vehicle control software and the virtual sensors and actuators running as part of the digital twin configuration (A4/Figure 13).

FIGURE 13
Example of Successful Model Communication for Fully Configured SIL Setup



During this step, any errors in the sensor messaging format can be identified and fixed. Once the vehicle control code is working with the digital twin code, the mission starts. At this point, any major dynamics discrepancies are analyzed in real time. For example, inability to dive or major pitch oscillations would merit a pause in simulation and for the developers to revisit the digital twin models to identify bugs. Once the mission is completed, detailed post mission analysis (PMA) is conducted. The PMA determines the dynamic performance of the digital twin vs the real-world referent data and drives the conclusions about the digital twin's predictive capability.

FIGURE 14
Example PMA Data Animation Comparison Between DT and Referent Vehicle



7.11 V&V Acceptance Claim and Recommendations

Lastly, the digital twin use-case requirements are revisited, and formal recommendations are documented. Based on the data generated by the verification and validation activities, an assessment is made about the use of the digital twin. Any unachievable requirements are discussed between the DT User/Sponsor and the DT Developer for the use case to identify error sources and next steps. This can often lead to updates to individual models in the DT or a reduction of the use-case scope.

9 Documentation

9.1 Example Sensor Model V&V Report

This subsection displays an example of a process sheet for model V&V of a generic depth sensor model to provide more insight into the V&V guidelines applied for UUV digital twins. Note that some information is non-specific, as a placeholder for sensitive information that would usually be populated in the report.

FIGURE 15
Example Process Sheet for Model V&V of Depth Sensor Model

1. Purpose

The purpose of this process sheet is to document the verification and validation (V&V) activities for a digital twin model. A model refers to an individual ROS2 node in the simulator code base.

2. Scope

The scope of this document focuses on V&V of a single model for one or two use-cases (more can be added depending on the purpose of the model).

3. Tools/References/Materials

3.1. Tools

3.2. References

- [1] E. VanDerHorn and S. Valluri, "Guidance on the Verification and Validation of Digital Twins," *29th Offshore Symposium, 2024.*
  Note: All the leading Overview subsections in this V&V process sheet include excerpts directly from the previous reference for easy access. It is highly recommended to read the paper in full.

3.3. Materials

A simulation asset is needed to perform model V&V. This can be one of the following sets of bulleted lists below as needed/available.

- Ubuntu machine running the simulator
- Vehicle processor board running vehicle code build with serial breakout port (as required)
- Serial (RS232) to USB cable (as required)

OR

- Simulation Cart Assembly (PN XXXXXX)

OR

- Simulator box (PN XXXXXX)

4. Procedure

4.1. Define Model + Use Case(s)

Fill in the following table and add any additional information below as needed.

Table 4: Configuration and Reference Information

	Input Data:
Model Name:	Depth Sensor
Development Branch:	dev/initials/MODSIM-XXX
Link to active_sim folder:	*(File path on company server)*
Jira Ticket(s):	MODSIM-XXX
GitLab Merge Request(s):	(Link to GitLab merge request)
Vehicle Code Version (if applicable):	v10.0.10

Additional Notes:

Define the use-case(s) for the model. Add or remove rows as needed.

Table 5: Model Use Case List

Model Use Cases	
1	The depth sensor is used to provide depth readings to the appropriate controller (Autopilot, PIL, HWIL)
2	The depth sensor is used to communicate in the required format with vehicle control code (PIL, HWIL)

4.2. Criticality Assessment Overview

4.3. Criticality Assessment

Evaluate the criticality of the results that this model produces, according to its use case.

Table 11: Criticality Assessment Results

Metric	Model Value
Decision Consequence Level (L, M, H) [Table 6]	M
Results Integration Level (0, 1, 2) [Table 7]	0
Supplementary Information Availability (0,1,2) [Table 8]	1
Model Influence Level (2 * RIL+ (2 - SIA)) [Table 9]	1
Overall Criticality (L, M, H) [Table 10]	L

Provide a brief statement on the overall criticality level:

The depth sensor model has a low calculated criticality. This corresponds with this sensor's use case - while depth is important, the same information can be tracked in other calculated messages.

4.4. Conceptual Model Validation

Evaluate whether the theories and assumptions underlying the conceptual model are correct, and whether the model representation of the real-world problem is reasonable (did I build the right thing?).

Table 12: Conceptual Model Details

	Metric	Description
A	Model Objectives	- Calculate and output depth of the simulated vehicle. - Communicate depth values in the required format.
B	Model Inputs - what inputs are needed to run this model? (Configuration files, other model output, etc.)	- vehicle state (provides z location where z=0 is the water surface) - environmental state (provides local ocean density) - input configuration file specifying depth sensor variant information
C	Collection frequency for the model inputs	- vehicle state is collected at the update rate of the digital twin software - configuration information is collected once per run
D	Model Outputs - what data does this model produce?	- measured depth
E	Data Sources for Model Development (CFD, Bollard test data, spec sheets, data extracted from research papers etc.)	- specification sheet for the depth sensor variant
F	Data Quality - Model Input Data - Model Development Data	The data quality is high. The vehicle state data is the product of a medium fidelity solver, and the sensor specification sheet is assumed to be fully tested and correct.
G	Model Content - methodology behind the model (how was reality simplified?), main formulations used in the model calculations	Pressure = density * gravity * depth
H	Model Assumptions and Simplifications - in the formulations - in the model data sources (Rows B and E)	- this model assumes a gaussian distribution for sensor noise

4.5. Model Verification Overview

4.6. Model Verification

Model verification determines if the model implementation sufficiently satisfies the requirements of the use-case (<u>did I build the thing right?</u>). For integration testing and autopilot runs, give a detailed description of the test, and link the file location for input configuration files.

Note: Not every activity can be completed for every model - leave blank activities that do not apply or add rows for other types of verification.

Table 14: Model Verification Activity Log

Verification Activity	Description	Add Evidence
Software Quality Assurance	Model code review with another team member	Completed. (Link to Gitlab merge request which contains review approvals.)
Numerical Algorithm Verification	Unit Testing	Completed. Unit tests run automatically when the sensor code is pushed to Gitlab, pass/fail status is documented in the merge request.
Numerical Algorithm Verification	Integration Testing	N/A
Numerical Algorithm Verification	Autopilot Runs	Autopilot runs were completed in a simulated environment with no ocean current, turbulence, or waves. The vehicle was commanded to dive to a specific depth and the output from the depth sensor was monitored to verify that it matched the depth of the vehicle state (within the noise spectrum). Plots documented in the sensor Jira ticket.
Numerical Algorithm Verification	Comms Check via Interface Program	Communications were established between the depth sensor model and the vehicle control software. Communication monitored in raw data view and view green annunciator in the vehicle interface program. Port:__xxxx___ Baud rate:__xxxx Annunciators Altitude Sensors / Pitch Motor / Rudder Motor Thruster Motor / DSP / Depth Sensor

4.7. Model Results Validation

A separate process sheet has been created for results validation, which compares output from the digital twin configuration that includes this new model with real-world referent data. Once completed, please link the validation report here:

See Appendix B

4.8. Model V&V Closeout

	Signature	Date
Report completed by:	Initialize	Date
Report reviewed by:	Initialize	Date

9.3 Example Digital Twin Configuration V&V Report

This subsection displays an example of the process sheet for completing V&V of a generic small class UUV digital twin configuration to provide more insight into the applied V&V guidelines. This example represents the validation efforts for the digital twin configuration repeated due to the addition of the new depth sensor model from Section A4/9.1 being incorporated into the digital twin configuration.

FIGURE 16
Example Process Sheet for Model V&V of Generic Small Class UUV Digital Twin Configuration

Process Sheet

Digital Twin Configuration Validation Process

Table of Contents

1. Purpose

The purpose of this process sheet is to document the results validation activities for a digital twin configuration. A configuration is a set of models (ROS2 nodes) that are used in tandem to run the overall digital twin. The common configurations used by developers are autopilot, processor-in-the-loop (PIL), and hardware-in-the-loop (HWIL) for a specific virtual vehicle.

Results validation determines if the digital twin configuration is an accurate representation of the real-world problem. A key component of results validation is comparison with real-world data, so the predictive capability of the digital twin can be understood.

2. Scope

The scope of this document focuses on a single configuration for an overarching use case. For example, validation of a small class UUV digital twin as a test platform for vehicle software (processor-in-the-loop).

3. Tools/References/Materials

3.1. Tools

3.2. References

- V &V Process Sheet for each model exercised in this configuration
- [1] E. VanDerHorn and S. Valluri, "Guidance on the Verification and Validation of Digital Twins," *29th Offshore Symposium, 2024*.
  Note: All the leading Overview subsections in this V&V process sheet include excerpts directly from the previous reference for easy access. It is highly recommended to read the paper in full.

3.3. Materials

A simulation asset is needed for digital twin configuration validation. This can be one of the following sets of bulleted lists below as needed/available.

- Ubuntu machine running the simulator
- Vehicle processor board running vehicle code build with serial breakout port (as required)
- Serial (RS232) to USB cable (as required)

OR

- Simulation Cart Assembly (PN XXXXXX)

OR

- Simulator box (PN XXXXXX)

4. Procedure

4.1. Define Digital Twin Configuration + Use Case

Fill in the following table and add any additional information below as needed.

Table 15: Configuration and reference Information

	Input Data:
Configuration Name:	Small-class UUV Digital Twin (with new depth sensor model) as a vehicle control software test platform.
Development Branch:	dev/Acosta/MODSIM-XXX
Link to active_sim folder:	*(File path on company server)*
Jira Ticket(s):	MODSIM-XXX

GitLab Merge Request(s):	(Link to GitLab merge request)
Vehicle Code Version:	v10.0.10

Additional Notes:

N/A

Define the use-case for the digital twin configuration.

Table 16: Model Use Case List

Configuration Use Case	
1	This configuration will be used as a test platform for small class vehicle software control logic verification.

4.2. Criticality Assessment Overview

4.3. Criticality Assessment

Evaluate the criticality of the results that this digital twin configuration produces, according to its use case.

Metric	Value
Decision Consequence Level (L, M, H) [Table 17]	M
Results Integration Level (0, 1, 2) [Table 18]	1
Supplementary Information Availability (0,1,2) [Table 19]	1
Model Influence Level (2 * RIL+ (2 - SIA)) [Table 20]	3
Overall Criticality (L, M, H) [Table 21]	M

Provide a brief statement on the overall criticality level:

This configuration's use-case has a medium criticality level. This tracks with the typical use of the configuration, which is used to test software before it heads to sea. Any failures of the digital twin at this point would be quickly rectified at sea because they would not cause catastrophic failure.

4.4. Requirements

Based on the intended use case(s), list a set of requirements for a successful model. Add rows as needed to capture requirements. The acceptance criteria defines the specific measures needed to accept a requirement. The metric will be used to evaluate the model against the criteria.

Table 22: Requirements List

Requirement	Acceptance Criteria	Metrics
UUV DT sensors, actuators, and physics solver should individually satisfy their Model V&V requirements.	Model V&V complete for each sensor, actuator, and physics solver model individually. These include: depth sensor, conductivity-temperature sensor, DVL, GPS, INS, fins, propeller, motor controller, physics solver, and water profile.	See other Model V&V documents (In particular, depth sensor Process Sheet in Appendix A was added to the configuration)
UUV DT shall satisfy all communication needs to vehicle software.	All necessary communication messages are present, sent to proper port, and at the proper frequency.	All annunciators in vehicle interface program indicate green/yellow status.
UUV DT shall capture similar steady state flight to actual vehicle.	Pitch response percent error as compared to sea data is lower than maximum threshold percent error required per vehicle specifications.	Pitch response from the simulated vs actual vehicle log files.
UUV DT shall achieve depth control comparable to actual vehicle.	DT depth response maintains a percent error lower than the maximum threshold as required per vehicle specifications when compared to sea data depth during steady level flight mission legs.	Depth data from the simulated vs actual vehicle log files.
UUV DT shall capture similar dive slope to the actual vehicle.	DT dive slope is lower than maximum threshold percent error required per vehicle specifications when compared to sea data.	Depth response slope during dive captured in the simulated vs actual vehicle log files.
UUV DT shall capture similar ascent slope to the actual vehicle.	DT depth slope is lower than maximum threshold percent error required per vehicle specifications when compared to sea data.	Depth response slope during ascent captured in the simulated vs actual vehicle log files.
UUV DT captures all necessary objectives as defined by the UUV mission file.	Visual inspection that mission track lines correlate.	Latitude and longitude data in the simulated vs actual vehicle log files.

4.5. Results Validation Overview

4.6. Results Validation

Conduct results validation. First determine the validation method to use, then complete the test plan and generate results.

Note: If several tests will be used, duplicate the rest of this section, and repeat as necessary.

Test #1

Validation method to be used: Animation, comparison with benchmarks.

Test #1 Overview: Processor-in-the-Loop test

A **comparison with benchmark test** is completed for this integrated model. A step depth mission was selected to recreate using the UUV digital twin and the results are compared with the results from running the same mission at sea. This test focuses on the newly updated depth sensor model and the model results will be analyzed during all stages of the mission, including depth goals and altitude goals.

Our test set-up also includes **animation** allowing the V&V user to verify expected behavior visually throughout mission. Due to lack of visualization on the real vehicle undergoing its mission at sea, benchmark comparison is reduced to 2D plotting.

Table 24: Validation Test Plan

	Test Plan	Add Details
A	Test Description	Processor-in-the-loop simulation of a full mission. The vehicle log file produced from the simulated mission is compared directly to log file from the at-sea mission.
B	Describe the hardware/software architecture	The simulator is launched and run from a laptop connected via ethernet and RS232 to the vehicle processor. A second laptop is connected to the network to setup and run the mission in the vehicle interface program. Fin and propeller commands flow from the vehicle processor to the simulator through the RS232/ethernet connection. The simulation then uses TCP/UDP protocols to handle these incoming commands and report back sensor status updates based on the state of the simulation.
C	Relationship to acceptability criteria	This test will provide evidence for the depth control requirement
D	Identify prerequisite conditions for test completion	Successfully complete at-sea mission and produce the vehicle log file.
E	Describe necessary test inputs	Mission and configuration file on the vehicle processor. Input configuration files in the simulation environment to properly define the specific UUV variant and the operating environment to match the sea day environment (bathymetry, ocean current/turbulence, sea state). Settings for serial communications.
F	Describe expected test outputs/results	Vehicle log file data to reflect the corresponding vehicle log from the sea test.
G	Define the test procedure	Populate simulation input configuration files to reflect the UUV characteristics and components. Complete the set up described in B. Run the virtual world and UUV DT

		nodes. Once annunciators are satisfied in the vehicle interface program, select "Run" to start the mission. Download and save the vehicle log file after mission is complete for detailed side-by-side comparison.
H	Identify assumptions/ simplifications in the test	We cannot perfectly mimic the environmental effects of the sea day.

Table 25: Validation Test Results

	Test Results	Add Details
A	Depth Control Results	*Raw data internal to customer.* Depth vs time was plotted for the DT vs the real-world test data. Root Mean Square (RMS) values were used to compare the datasets at each distinct depth step.
B	Altitude Control Results	*Raw data internal to customer.* Depth and altitude vs time were plotted for the DT vs the real-world test data. The plots were analyzed to confirm that the vehicle maintained proper altitude control while receiving input from the newly modified depth sensor. Additionally, the total water column height was tracked to ensure that the depth values were correct.
C	Compare expected to actual test results. Describe and analyze any anomalies	No anomalies present.
D	Provide an assessment about the cause of any discrepancies and a means of correcting them	N/A
E	Assess and describe how the results compare to the related acceptability criteria	These results satisfy the depth control acceptability criteria. They show that the depth is reported in a correct format and with corresponding values as compared to the depth from the sea dataset.
F	Identify location where the validation data is saved	Folder in Simulation Archive: 2022.12-Ph3 UUV FAT Validation.docx

<u>Test #2</u>

Validation method to be used: Animation, comparison with benchmarks.

Test #2 Overview: Processor-in-the-Loop test

A **comparison with benchmark test** is completed for this integrated model. A dynamics mission was selected from customer's factory acceptance test (FAT) procedure, simulated in virtual world conditions to best mimic the records from a specific sea day, and compared to the mission results from the sea day.

Our test set-up also includes **animation** allowing the V&V user to verify expected behavior visually throughout the mission. Due to lack of visualization on the real vehicle undergoing its mission at sea, benchmark comparison is reduced to 2D plotting.

See internal documents for more detailed validation plan.

Table 26: Validation Test Plan

	Test Plan	Add Details
A	Test Description	Processor-in-the-loop simulation of a full mission. The vehicle log file produced from the simulated mission is compared directly to log file from the at-sea mission.
B	Describe the hardware/software architecture	The simulator is launched and run from a laptop connected via ethernet and RS232 to the vehicle processor. A second laptop is connected to the network to setup and run the mission in the vehicle interface program. Fin and propeller commands flow from the REMUS processor to the simulator through the RS232/ethernet connection. The simulation then uses TCP/UDP protocols to handle these incoming commands and report back sensor status updates based on the state of the simulation.
C	Relationship to acceptability criteria	This test is intended to address all requirements outlined in the earlier section.
D	Identify prerequisite conditions for test completion	Successfully complete mission and produce the vehicle log file.
E	Describe necessary test inputs	Mission and configuration file on the vehicle processor. Input configuration files in the simulation environment. Settings for serial communications.
F	Describe expected test outputs/results	Vehicle log file data to reflect the following mission shown in the track line below.
G	Define the test procedure	Populate simulation input configuration files to reflect the UUV characteristics and components. Complete the set up described in B. Run the virtual world and UUV DT nodes. Once annunciators are satisfied in the vehicle interface program, select "Run" to start the mission. Download and save the vehicle log file after mission is complete for detailed side-by-side comparison.
H	Identify assumptions/ simplifications in the test	We cannot perfectly mimic the environmental effects of the sea day. Also, records were minimal from that day,

		so ocean turbulence is the only environmental effect used in addition to the bathymetry.

Table 27: Validation Test Results

	Test Results	Add Details
A	Steady State Results	*Raw data internal to customer.* Pitch vs time was plotted for the DT vs the real-world test data. RMS values were used to compare the datasets
B	Dive Results	*Raw data internal to customer.* Depth vs time was plotted for the DT vs the real-world test data during a vehicle dive from the surface. Dive time, overshoot, and settling time are compared between the datasets
C	Ascent Results	*Raw data internal to customer.* Depth vs time was plotted for the DT vs the real-world test data during a vehicle ascent from 4m. Ascent time, overshoot, and settling time are compared between the datasets
D	Depth Response	*Raw data internal to customer.* Depth vs time was plotted for the DT vs the real-world test data during a vehicle transit in altitude mode. The results were compared to determine how well the DT tracked the bathymetry
E	Compare expected to actual test results. Describe and analyze any anomalies	Overall, the results correlate well. Minor differences in the track line are due to a ~60 min vehicle loiter that was not repeated with the digital twin mission
A	Provide an assessment about the cause of any discrepancies and a means of correcting them	N/A
B	Assess and describe how the results compare to the related acceptability criteria	Meets criteria. Percent differences are reported to capture offset between DT data and real vehicle sea data. Steady state flight exhibited an average %-difference in pitch standard deviation between simulated results and at-sea results equal to XX% which is less than YY% threshold specified. Dive slope percent difference of XX% (RMS)

		Surfacing slope percent difference of XX% (RMS) Depth response is tracking between the two data sets with <X meters offset.
C	Identify location where the validation data is saved	Folder in Simulation Archive: 2022.12-Ph3 UUV FAT Validation

4.7. Recommendations

Finally, revisit the requirements from 4.4 and provide recommendations based on the results of the validation testing. Provide explanations and actions from any discrepancies.

Table 28: Requirements Conclusions

Requirement	Acceptance Criteria	Analysis
UUV DT sensors, actuators, and physics solver should individually satisfy their Model V&V requirements.	Model V&V complete for each sensor, actuator, and physics solver model individually. These include: Depth sensor, Conductivity-temperature sensor, DVL, GPS, INS, fins, propeller, motor controller, physics solver, and water profile.	Requirements achieved.
UUV DT shall satisfy all communication needs to actual vehicle software.	All necessary communication messages are present, sent to proper port, and at the proper frequency.	Requirement achieved by visual inspection.
UUV DT shall capture similar steady state flight to actual vehicle.	Pitch response percent error as compared to sea data < XX%.	Requirement achieved by numerical analysis.
UUV DT shall achieve depth control comparable to actual vehicle.	DT depth within X meters of sea data depth during comparable mission legs	Requirement achieved by numerical analysis.
UUV DT shall capture similar dive slope to the actual vehicle.	Depth slope error <XX%.	Requirement achieved by numerical analysis.
UUV DT shall capture similar ascent slope to the actual vehicle.	Depth slope error <XX%.	Requirement achieved by numerical analysis.
UUV DT captures all necessary objectives as defined by the vehicle mission file.	Visual inspection that mission track lines correlate.	Requirement achieved by visual inspection.

4.8. Configuration Validation Closeout

	Signature	Date
Report completed by:	Initialize	Date
Report reviewed by:	Initialize	Date