SANDIA REPORT

SAND2011-0084
Unlimited Release
Printed January 2011

Nuclear Energy Advanced Modeling and Simulation Waste Integrated Performance and Safety Codes (NEAMS Waste IPSC)

Verification and Validation Plan (Version 1)

H. Carter Edwards, Jose G. Arguello Jr., Roscoe A. Bartlett, Julie F. Bouchard, Geoff Freeze, Robert Howard, Patrick Knupp, Marjorie T. McCornack, Peter A. Schultz, Angel Urbina, and Yifeng Wang

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2011-0084
Unlimited Release
Printed January 2011

Nuclear Energy Advanced Modeling and Simulation Waste Integrated Performance and Safety Codes (NEAMS Waste IPSC) Verification and Validation Plan (Version 1)

H. Carter Edwards, Jose G. Arguello Jr., Roscoe A. Bartlett, Julie F. Bouchard,
Geoff Freeze, Patrick Knupp, Marjorie T. McCornack, Peter A. Schultz, Angel Urbina,
and Yifeng Wang

Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185

and

Robert Howard
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831

Abstract

The objective of the U.S. Department of Energy Office of Nuclear Energy Advanced Modeling and Simulation Waste Integrated Performance and Safety Codes (NEAMS Waste IPSC) is to provide an integrated suite of computational modeling and simulation (M&S) capabilities to quantitatively assess the long-term performance of waste forms in the engineered and geologic environments of a radioactive-waste storage facility or disposal repository. To meet this objective, NEAMS Waste IPSC M&S capabilities will be applied to challenging spatial domains, temporal domains, multiphysics couplings, and multiscale couplings. A strategic verification and validation (V&V) goal is to establish evidence-based metrics for the level of confidence in M&S codes and capabilities. Because it is economically impractical to apply the maximum V&V rigor to each and every M&S capability, M&S capabilities will be ranked for their impact on the performance assessments of various components of the repository systems. Those M&S capabilities with greater impact will require a greater level of confidence and a correspondingly greater investment in V&V. This report includes five major components: (1) a background summary of the NEAMS Waste IPSC to emphasize M&S challenges; (2) the conceptual foundation for verification, validation, and confidence assessment of NEAMS Waste IPSC M&S capabilities; (3) specifications for the planned verification, validation, and confidence-assessment practices; (4) specifications for the planned evidence information management system; and (5) a path forward for the incremental implementation of this V&V plan.

Acknowledgments

The authors thank Rhonda K. Reinert for her careful review and editing of this document.

Contents

Figures

Tables

Nomenclature

ADM	analysis data management
CCDF	complementary, cumulative distribution functions
CM	configuration management
DIME	Diplomatic, Informational, Military, and Economic
DOE	Department of Energy
EBS	engineered barrier system
ECT	Enabling Computational Technologies
EPA	Environmental Protection Agency
EVIM	evidence management
FEP	feature, event, and process
FMM	Fundamental Methods and Models
IET	integral-effects test
IPSC	Integrated Performance and Safety Codes
LSN	Licensing Support Network
M&S	modeling and simulation
MMS	method of manufactured solutions
NEA	Nuclear Energy Agency
NEAMS	Nuclear Energy Advanced Modeling and Simulation
NRC	Nuclear Regulatory Commission
PCMM	Predictive Capability Maturity Model
PDE	partial differential equation
PIRT	Phenomena Identification Ranking Table
PMESII	Political, Military Economic, Information, and Infrastructure
RM	requirements management
RTS	Regression Test Suite
Sandia	Sandia National Laboratories
SET	separate-effect test
SQE	software quality engineering
SRQ	system response quantity
THCM	thermal-hydrological-chemical-mechanical
THCMBR	thermal-hydrological-chemical-mechanical-biological-radiological
UQ	uncertainty quantification
V&V	verification and validation
V&V and UQ	verification and validation and uncertainty quantification
VU	verification and validation and uncertainty quantification (referred to as “V&V and UQ” in this document)
WIPP	Waste Isolation Pilot Plant

1 Introduction

The objective of the U.S. Department of Energy (DOE) Office of Nuclear Energy Advanced Modeling and Simulation Waste Integrated Performance and Safety Codes (NEAMS Waste IPSC) program element is to provide an integrated suite of computational modeling and simulation (M&S) capabilities to assess quantitatively the long-term performance of waste forms in the engineered and geologic environments of a radioactive-waste storage facility or disposal repository. The NEAMS Waste IPSC will include numerical models and codes for characterizing material properties, for coupling phenomena involved in the degradation of waste forms and transport of radionuclides, and for computationally efficient performance assessments of proposed nuclear-waste disposal systems. The ultimate goal is to support predictive simulation-based, risk-informed decision making about the management of future U.S. nuclear waste.

The objective of the NEAMS Waste IPSC will be fulfilled by acquiring and developing M&S capabilities. All NEAMS IPSC program elements must establish a defensible level of confidence in their M&S capabilities. This level of confidence must be commensurate with the risks associated with the intended uses of the codes. The foundation for assessing the level of confidence is based upon the rigor and results from verification, validation, and uncertainty quantification (UQ) activities.

Strategic verification and validation (V&V) goals of this program element are to establish evidence-based metrics for specifying the level of confidence in M&S capabilities, to assess M&S capabilities according to these metrics, and to provide M&S assessment results that support risk-informed decision-making processes. Because it is economically impractical to apply the maximum V&V rigor to each and every M&S capability, the M&S capabilities for phenomena will be ranked according to (1) their impact on the performance of waste forms and systems and (2) the risks associated with how the M&S capabilities are used. Those M&S capabilities for phenomena with greater impact (e.g., radionuclide transport) or higher risk (e.g., assessing a proposed disposal site near a populated area) will require a greater level of confidence and a correspondingly greater investment in V&V. The difference between required and assessed levels of confidence in M&S capabilities will influence how program resources for V&V, UQ, and M&S are prioritized and allocated.

This report presents the NEAMS Waste IPSC plan for defining and establishing confidence in the acquired and developed M&S capabilities. The primary emphasis of this plan is to implement a set of practices for establishing this confidence. These practices include verification, validation, and uncertainty quantification (V&V and UQ) activities, assessments of M&S capabilities, and maintenance and communication of the resulting supporting evidence. The secondary emphasis of the plan is to identify the systems and software quality engineering (SQE) practices and data/information management tools required to implement this plan.

The primary location for implementing the V&V plan will be at Sandia National Laboratories (Sandia). The NEAMS Waste IPSC team at Sandia will serve as the systems integrator of the existing and to-be-developed M&S capabilities (codes, data, and evidence). The Sandia team will manage the M&S capabilities and evidence according to the practices set forth in this plan and will manage this evidence within a data/information management system. The NEAMS Waste IPSC team will perform M&S development, UQ, and experimental work as needed and in collaboration with other DOE programs. The V&V plan will be implemented as the M&S capabilities are acquired and developed. Note that implementation of critical practices and information management systems will be incremental and limited by allocated resources. Furthermore, implementation will necessarily be dependent upon the practices and data/information management systems of the suppliers and users of the M&S capabilities. In effect, the implementation of the plan must be agile to accommodate currently unknown resources, suppliers, and users.

1.1 Audiences of the Plan

This plan is intended to address the needs of different types of audiences, including implementers of the plan, M&S capability developers, experimentalists, users, and NEAMS program management. The anticipated needs for these various groups are summarized below

Implementers of the V&V plan: This document provides implementers of the practices and supporting information management systems with a high-level, holistic view of the background, concepts, practices, and information management needs for V&V within the NEAMS Waste IPSC. Implementers include the people who will deploy the V&V and UQ processes, set up the quality environment and the evidence information management (EVIM) system, import codes and data, run the various tests, introduce test and peer-reviewed results into evidence, and manage the evidence.

Developers of M&S capabilities: This group includes the developers of conceptual, mathematical, and numerical models as well as developers of the code. Developers need to understand the V&V lifecycle and practices and will use this plan to guide detailed planning and analysis required during M&S capability development. Developers will need to plan and implement the associated progression of integration testing, code verification, solution verification, model validation, and UQ that accompany development of the M&S capabilities. Such work requires an understanding of the requirements and expectations for evidence and data and information management.

Experimentalists: Currently, the NEAMS Waste IPSC does not plan to generate or commission experiments or field observations. As such, this program element will rely upon data acquired from external sources to satisfy requirements for development, calibration, and validation. For those who have conducted experiments relevant to the NEAMS Waste IPSC, this plan will identify the specific requirements for documenting the operational details of experiments and their results.

Users of NEAMS Waste IPSC: Those engineers and analysts in agencies or companies who use the NEAMS Waste IPSC to design waste forms and waste repositories and to choose geologic settings need to understand the V&V lifecycle and practices to assess and communicate the level of confidence in the results of their analyses. They also need to provide feedback and support regarding the potential need for more focused, refined, or reprioritized testing and validation activities as their understanding of physical systems evolves through use of the modeling tools.

NEAMS program management: NEAMS program management will be able to use this plan for two main purposes: (1) to appreciate the NEAMS Waste IPSC V&V and UQ challenges and complexities and (2) to prioritize program resources for the implementation of this plan.

1.2 Organization of the Plan

As explained above, this plan addresses the nuclear-waste M&S capabilities that will be acquired or developed for the NEAMS Waste IPSC. These M&S capabilities will have varying modeling, V&V, and UQ requirements depending upon the intended use of a particular M&S capability. As such, the plan provides a framework or template from which M&S capability-specific V&V requirements and plans will be generated over the lifetime of the NEAMS Waste IPSC. The plan is structured to meet the needs of different types of audiences in their various roles as part of the collaborative NEAMS Waste IPSC effort.

Section 2 presents the major topics that are part of the NEAMS Waste IPSC problem domain and provides the context for the significant modeling, simulation, verification, validation, and UQ challenges inherent in this domain.

Section 3 summarizes key concepts for (1) V&V and UQ of collections of complex M&S capabilities and (2) the NEAMS Waste IPSC M&S strategy for satisfying V&V and UQ needs in the challenging problem domain. This summary is intended to provide a sufficient foundation for understanding the practices and information management required for V&V and UQ of the NEAMS Waste IPSC. This section also summarizes the NEAMS Waste IPSC strategy for meeting these challenges.

Section 4 describes the V&V and UQ practices that will be implemented and applied *as required* to M&S capabilities. The practices applied to each specific M&S capability will vary according to the V&V and UQ requirements and plans for that M&S capability. The V&V and UQ practices described in Section 4 are intended to address M&S capabilities with V&V and UQ requirements for the maximum assessed level of confidence.

The application of V&V and UQ practices to M&S capabilities will produce evidence for establishing confidence in those M&S capabilities. A significant goal of the NEAMS Waste IPSC is to manage this V&V and UQ evidence. The goals and conceptual architecture for managing V&V and UQ evidence are presented in Section 5.

Section 6 presents a path forward for implementing this V&V plan. The plan has been introduced early in the NEAMS Waste IPSC program to facilitate early implementation of V&V processes and information management. Early implementation is essential to establish a V&V culture within the program so that V&V and UQ requirements can be met and high-confidence M&S capabilities can be deployed when needed.

2 Background

Section 2 presents the major topics that are part of the NEAMS Waste IPSC problem domain and provides the context for the significant modeling, simulation, verification, validation, and UQ challenges inherent in this domain.

2.1 Aspects of the NEAMS Waste IPSC

We begin the discussion by addressing a number of different aspects of the NEAMS Waste IPSC that must be considered to meet the intended uses of the codes. These aspects cover the M&S domain, spatial and temporal domains, requirements as set forth by a relevant challenge problem, the importance of UQ in establishing confidence in simulations of the NEAMS IPSC, and the particular scales that will be employed to model the various phenomena and used to quantify uncertainties. We also address the need for coupling physics phenomena and the critical importance of assessing and managing the V&V and UQ evidence.

M&S domain: Figure 2-1 illustrates the complex M&S domain of the NEAMS Waste IPSC. As defined by the U.S. Nuclear Regulatory Commission (NRC) in 10 CRF 63.2, the term *waste form* means the radioactive waste materials and any encapsulating or stabilizing matrix. The term *waste package* refers to the waste form and any containers, shielding, packing, and other absorbent materials immediately surrounding an individual container (NRC 2001).

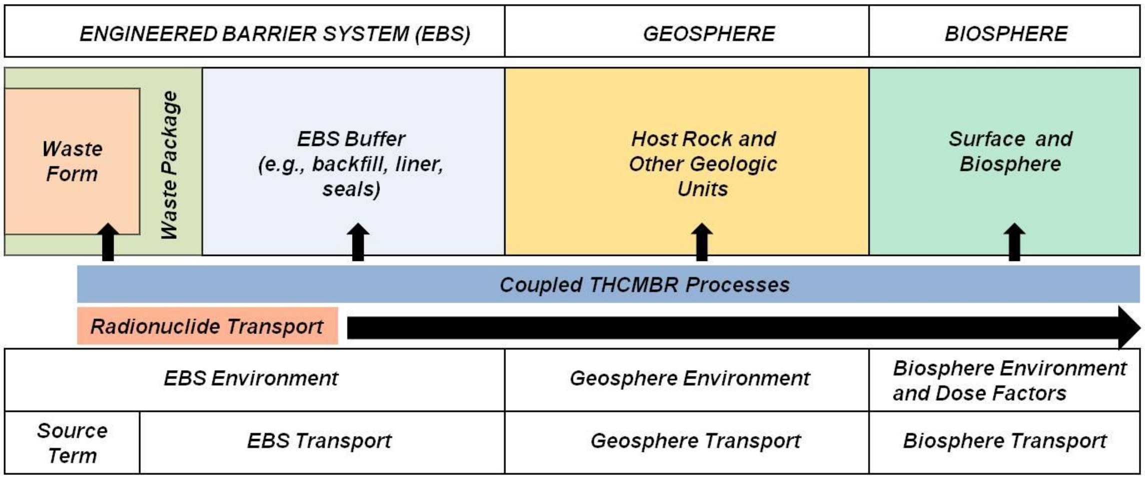


Figure 2-1. Nuclear-waste M&S domain.

The top half of Figure 2-1 denotes the physical domains of a generalized waste-disposal system—beginning with a waste package buried at some location and considering the environment through which radionuclides may be transported over time. The three physical domains, i.e., environments, in the figure are the near-field environment referred to as the ―engineered barrier system” (EBS), the geosphere, and the biosphere. The EBS has multiple subdomains to consider, such as the waste form, the waste package, and possible liners, seals, and backfill material. Because each of these subdomains has a diverse set of potential materials, designs, and applicable phenomena that need to be modeled, the M&S capabilities must be flexible to support a range of NEAMS Waste IPSC analyses. Furthermore, experiments and precise field observations are not possible over the large spatial and temporal scales of the physical problem to be simulated. Consequently, M&S capabilities are necessary for extrapolating well beyond the range for which *direct* comparison with field observations and experimental measurements is possible.

The bottom half of Figure 2-1 summarizes the phenomena associated with each of these physical domains and subdomains. These phenomena include the coupled thermal-hydrological-chemical-mechanical-biological-radiological (THCMBR) processes that describe (1) waste form degradation and subsequent release of radionuclides, (2) reactive transport through the near-field environment, (3) radionuclide transport through the geosphere, and (4) radionuclide transport and health effects in the biosphere. In addition to their direct effects on radionuclide transport, the coupled THCMBR processes also influence the physical and chemical environments in the EBS, geosphere, and biosphere, which in turn affect radionuclide transport. For example, it is possible that some radionuclides could chemically react or bond with material in the environment and seal up the hole in which a waste form was buried. While fully coupled processes are crucial in the near-field environment, these processes are likely to decline in importance as the transport extends into the geosphere and beyond. Nonetheless, the simulation capabilities must incorporate the level of coupling necessary to model reactive transport in every zone.

Spatial and temporal domains: Each phenomenon will have spatial and temporal domains of concern. The *temporal domain* covers the M&S functions from nanoseconds to a million years or more. For example, we might simulate a material at the molecular level for one microsecond, or simulate an entire repository including the surrounding geologic formation for 100,000 years. There are numerous *spatial domains* to consider. For example, we may simulate corrosion at a waste form's material boundary at a micrometer scale or simulate the geologic setting at a kilometer scale.

Challenge problem and related phenomena-modeling requirements: An anticipated performance assessment of a waste form in the environment, as represented in Figure 2-1, for the NEAMS Waste IPSC has been documented as a *challenge problem*. To basically test the evolution and implementation of the NEAMS Waste IPSC, our team defined the scope of such a problem in which one particular waste form, one particular engineered environment, and one particular geologic setting was selected. For this configuration, the “challenge” will be to apply our set of V&V and UQ processes on the M&S capabilities (as defined in this document) to demonstrate proof-of-concept and progress towards IPSC goals (Freeze, Arguello, Howard, McNeish, Schultz, and Wang 2010). This challenge problem includes a timeline of intermediate subproblem milestones leading up to the challenge problem itself.

The diversity of potential disposal concepts, geologic settings, EBSs, and waste forms leads to a broad set of phenomena-modeling requirements. These requirements will be expressed in the context of a features, events, and processes (FEPs) analysis. Simply put, “FEPs” refer to identified phenomena. The ranking of phenomena-modeling requirements will be driven by the relative impact of the identified phenomena on the performance of the repository system. The collection of ranked FEPs is also known as a Phenomena Identification Ranking Table (PIRT).

UQ and the three scales of M&S: Analyses will be required to quantify uncertainties for *the* quantity of interest: a predicted history for the migration of radionuclides into the biosphere. It is a significant challenge to establish confidence in UQ results produced by M&S capabilities that extrapolate well beyond an achievable domain of validation. The strategy to meet this challenge is based upon three scales of M&S, where confidence in higher-resolution capabilities provides a technical basis for establishing confidence in faster-running but lower-resolution capabilities. The three scales of M&S, from highest to lowest resolution, are the atomistic or subgrid scale, the continuum coupled-multiphysics scale, and the performance assessment scale. Subgrid-scale M&S will be based upon first principles and validated against experimental data. Continuum-scale multiphysics M&S capabilities will be run for larger spatial and temporal domains than subgrid-scale capabilities and will be assessed against their counterparts at the subgrid scale. M&S at the performance assessment scale will be fast-running for analyses at the maximum required spatial and temporal domains and will be assessed against analogous capabilities at the continuum scale. Uncertainties will be quantified at each scale and provide a basis for UQ at the next scale.

M&S capability coupling: The required flexibility for NEAMS Waste IPSC analyses will be supported through the user-specified coupling of M&S capabilities. This coupling of capabilities will be supported through workflows in which a sequence of codes is run with inputs to codes obtained from the outputs of other codes. The coupling of capabilities will also be supported through coupled multiphysics simulations in which a user integrates multiple physics (and chemistry) models into a single simulation code. An example of such multiphysics coupling would be necessary when simulating the effects of, say, rust on a container placed in a moist acidic environment, where the chemistry effects would be critical, as opposed to placing the container in a dry environment. The coupling of M&S capabilities introduces an additional challenge for UQ in that the uncertainties from the component capabilities must be aggregated to produce a system-level uncertainty for the analyses.

2.2 Intended Uses

The NEAMS Waste IPSC has several intended uses. The following are the primary intended uses by the main users (or customers):

- Performing analyses to support decision making and prioritization of disposal alternatives
- Designing waste forms and engineered environments
- Performing analyses to support licensing for a selected disposal alternative

- Providing a working example for meeting the requirements of the NEAMS Waste IPSC Challenge Problem (see Section 2.4)

There will be multiple users for the NEAMS Waste IPSC in the foreseeable future. In the next few years, development efforts and challenge-problem milestones could provide insights on the modeling of disposal systems to the Nuclear Energy Used Fuel Disposition Campaign. This campaign may, in turn, provide information to the Secretary of Energy's Blue Ribbon Commission on America's Nuclear Future. The NEAMS Waste IPSC will also be used by the Nuclear Energy Waste Form Campaign in evaluating the interplay between waste-form durability and performance for various waste forms and disposal-system environments. In the next 5 to 10 years, the capabilities of the NEAMS Waste IPSC will be needed by the Nuclear Energy Used Fuel Disposition Campaign to inform implementation of the Blue Ribbon Commission's recommendations and evaluate the relative performance and long-term safety of alternative radioactive-waste disposal or storage concepts and designs. The NEAMS Waste IPSC will also inform the Nuclear Energy Waste Form Campaign about the potential benefits of high-performing waste forms for selected waste streams in specific disposal-system environments. In that same time frame, simulations enabled by the NEAMS Waste IPSC capabilities may provide input and insights to the NRC as that agency considers revisions to the federal regulations governing the disposal of radioactive waste. The NEAMS Waste IPSC will be needed by the DOE to support site selection and to prepare a defensible technical basis, i.e., a performance assessment, for a license application for selected disposal alternatives.

The entire NEAMS Waste IPSC will be subjected to a high level of scrutiny by various stakeholders. The public will be the first and foremost of stakeholders, although indirectly, through the policy makers, regulators, licensing authority, advocates, and interveners. It is anticipated that policy makers and regulators will use the NEAMS Waste IPSC to support decision making and prioritization among various options for waste disposal. A second set of stakeholders will be the scientific community (including those contributing to the development of the NEAMS Waste IPSC) who will be asked to critique and evaluate the scientific adequacy and merit of the product. It is anticipated that the scientific community will use the M&S capabilities to design waste forms, engineered environments, and long-term disposal systems. Finally, another important set of stakeholders will be the users of the NEAMS Waste IPSC to run analyses to support licensing. It is anticipated that these stakeholders will be interested in system-level performance analyses with quantified uncertainties that will place strong demands on the tool set.

2.3 Scope of M&S Capabilities

The NEAMS Waste IPSC will enable simulations for a range of candidate waste-form materials, disposal concepts and designs, engineered barriers, and geologic environments over a broad range of length scales (angstroms to kilometers) in the spatial domain and time periods (nanoseconds to a million years) that make up the temporal domain. Simulations of the rate of radionuclide (or other hazardous constituent) release and subsequent reactive transport from the waste form through an engineered barrier and geologic environment into the biosphere will require coupled THCMBR (thermal-hydrological-chemical-mechanical-biological-radiological) processes. Modeling these processes will entail incorporating capabilities for (a) generating chemical reactions and chemical kinetics properties of materials at the subgrid scale, (b) upscaling into effective constitutive equations at the continuum scale, (c) developing continuum-scale codes to simulate coupled THCMBR release and transport phenomena, and (d) abstracting computationally efficient performance-assessment models to assess quantitatively the performance of disposal systems.

2.3.1 Potential Waste Form Types and Disposal Concepts

In collaboration with the Nuclear Energy Used Fuel Disposition Campaign, domain experts from the NEAMS Waste IPSC identified a set of seven potential categories of waste form types (Table 2-1) and eight potential categories of disposal concept/geologic settings (Table 2-2) to define the expected range (based on current knowledge) of disposal-system concepts, designs, settings, and conditions. The high-level waste form types in Table 2-1 are basically differentiated by the stabilizing matrix, or treatment, used on the waste.

Table 2-1. Categories of Potential Waste Form Types

Category Number	Waste Form Type	Description
1	Used Nuclear Fuel	Includes commercial spent nuclear fuel, DOE spent nuclear fuel, high-temperature gas-cooled reactor, greater than Class C
2	High-Level Waste Borosilicate Glass	Includes the borosilicate waste forms that have been or are being produced to solidify defense high-level wastes at Savannah River, West Valley, and Hanford. Current and future, e.g., no minor actinides
3	High-Level Waste Glass Ceramic	Includes waste form materials with ceramic phases embedded in a glass matrix. Current (glass-bonded sodalite) and future from electrochemical processing
4	High-Level Waste Advanced Ceramic	For example, Synroc
5	High-Level Waste Metal Alloy	Includes metal alloy waste forms from electrochemical or aqueous reprocessing
6	Lower Than High-Level Waste	Includes Class A, B, and C waste and greater than Class C waste
7	Other	Captures waste forms not included in the other categories and/or future waste forms, such as molten salt and

Category Number	Waste Form Type	Description
		electro-chemical refining waste

Table 2-2. Categories of Potential Disposal Concepts and Geologic Settings

Category Number	Disposal Concept	Description of Geologic Setting
1	Surface Storage	Long-term interim storage at reactors or at centralized sites
2	Near-Surface Disposal	For example, Lower Than High-Level-Waste disposal sites
3	Mined Geologic Disposal (Hard Rock, Unsaturated)	Granite/crystalline or tuff
4	Mined Geologic Disposal (Hard Rock, Saturated)	Granite/crystalline or tuff
5	Mined Geologic Disposal (Clay/Shale, Saturated)	Clay/shale
6	Mined Geologic Disposal (Salt, Saturated)	Salt
7	Deep Borehole Disposal	Granite/crystalline
8	Other	Subseabed, carbonate formations, etc.

The six categories in Table 2-1 and the seven categories in Table 2-2 result in 42 combinations (ignoring the placeholder ―Other" categories) of waste form types and disposal concepts and geologic settings. These combinations broadly define the range of potential alternative disposal-system designs that might need to be evaluated using the NEAMS Waste IPSC. It should be noted, however, that any single alternative disposal-system design could incorporate important subsystem designs. In other words, selecting one group from each table and pairing the two groups is just the beginning of the design and modeling decisions that must be considered. For example, different combinations of waste emplacement geometry, thermal loading, engineered-component (waste package, backfill, etc.) design and materials, and chemical conditions such as reducing or oxidizing may further differentiate the range of technical capabilities required of the NEAMS Waste IPSC. As technologies and sociopolitical drivers evolve, these categories of waste form types and of disposal concepts and geological settings will also evolve. See Freeze, Mariner, Houseworth, and Cunnane (2010) for further details about the categories listed in the two tables.

2.3.2 Generalized Waste Form and Disposal System

The conceptualization of a generic disposal system shown previously in Figure 2-1 includes components, domains, and phenomena common to most of the 42 disposal-system alternatives discussed above. For the different sources of waste and selections of waste forms, the ability to develop constitutive relations for the continuum models will be established. Such relations are typically expressed as mathematical formulas that describe, for example, materials and their responses to thermal, chemical, or other stimuli. The evaluation of the release of radionuclides will, in principle, depend on mechanistic, atomic processes extending into the subcontinuum domain. The capability to determine the governing equations for the release of radionuclides from glass, ceramic, or metallic waste forms, potentially including used nuclear fuel, will require incorporating methods of evaluating the chemical reactivity and diffusivity of these materials as a function of ambient conditions and the material environment. While not a primary goal of the NEAMS Waste IPSC, the design of waste forms for specific performance targets, i.e., materials-by-design, of the waste form in the environment would be partially enabled by the establishment of this capability. The quantitative assessment of waste form performance in chosen disposal scenarios would produce meaningful figures of merit, i.e., numbers one can trust, for the desired qualities in a waste material design, and the tools incorporated into the assessment process could equally well be used for materials design.

To facilitate an evaluation of a specific disposal-system design or a comparison of disposal system alternatives, the NEAMS Waste IPSC will provide M&S capabilities for several system and subsystem performance metrics as a function of time. The principal disposal performance metric is human health effects in the biosphere, e.g., annual dose of radiation exposure. Other disposal performance metrics that may be computed are radionuclide mass flux (which is the rate at which radionuclides are passing a particular point or through a particular area of concern) and cumulative release of radionuclides (which is the total amount of radionuclides released to the biosphere and across intermediate domain boundaries that have been selected for taking measurements); radionuclide mass in place (which is the amount of radionuclides within specified domain boundaries and remaining in the waste form); and spatial distributions (i.e., measurements taken at different locations of interest) of the various physical and chemical properties of radionuclides and materials, like backfill and geologic materials. Such properties may include pH, temperature, fluid saturation, chemical concentrations, material dissolution, and precipitation rates. Note that the *full* domain boundary would be where the cumulative release of radionuclides is zero.

2.3.3 Phenomena-Modeling Requirements

As introduced in Section 2.1, a PIRT is used to specify the requirements for phenomena that will be modeled by the NEAMS Waste IPSC. The PIRT also identifies progress toward implementation and UQ of the identified phenomena. The phenomena identified in the PIRT are also termed FEPs (features, events, and processes) of the waste form and system. The PIRT, because it contains rankings for the relative impact of phenomena and the status of the modeling, drives the focus for additional model development, V&V, and UQ. A PIRT will need to be created and maintained for the NEAMS Waste IPSC as part of this V&V plan. Importantly, the phenomena-modeling requirements are actually the start of the evidence chain in this plan.

As mentioned above, the relative impact of phenomena is an important consideration for requirements. For example, if we believe a steel casing of a waste package will rust long before the enclosed glass waste form will corrode, then modeling corrosion of the waste form will be more important than modeling rust of the steel casing.

Additional detail regarding the technical scope of the NEAMS Waste IPSC can be identified through an analysis of generic disposal-system FEPs. Such an analysis includes identification, evaluation, and implementation of the particular FEP. Identification can be done at a coarse level of detail that is potentially applicable to a broad range of disposal concepts, designs, and geologic settings. Evaluation and implementation of a FEP must consider the specific disposal-system alternative and also the performance metrics. Implementation of a FEP is further dependent on the scale of the M&S capability. Importantly, a FEP analysis is an iterative process that evolves as new information becomes available. Updates to the knowledge base of FEPs, as users create FEPs for new models, are likely to revise the ranking for phenomena-modeling requirements and require corresponding revision of the NEAMS Waste IPSC's PIRT.

The Nuclear Energy Agency (NEA) has compiled an international database that contains independent disposal-system FEPs from several countries (NEA 1999). Additional information describing the FEP analysis methodology, and specific FEP identification for the Yucca Mountain project, is available in reports from that work (BSC 2005). In collaboration with the Nuclear Energy Used Fuel Disposition Campaign, a preliminary set of FEPs that are potentially relevant to the scope of the NEAMS Waste IPSC were identified. These FEPs were developed based on the NEA's FEPs list, the Yucca Mountain project's FEPs list, and the PIRT list developed for the NEAMS Waste IPSC as part of FY09 planning (NEAMS Waste Forms Team 2009). The resulting set of 204 preliminary NEAMS Waste FEPs are listed in Appendix A of the document titled *NEAMS Waste IPSC Challenge Problems* (Freeze, Arguello, et al. 2010). Each FEP is defined by a high-level description, with additional detail given under associated processes. The FEPs are organized in accordance with a FEP categorization scheme that is similar to the NEA categorization scheme (NEA 1999, Section 3).

The NEAMS Waste IPSC categorization scheme, illustrated schematically in Figure 2-2, is based on a set of generic domains (features) that are likely to be present in most disposal-system concepts. Note that the generic features in Figure 2-2 are subsystem components, i.e., additional details, of the three physical domains (EBS, geosphere, and biosphere) shown previously in Figure 2-1. Figure 2-2 also illustrates how each of the generic features can be acted upon by events, i.e., External Factors, and/or THCMBR processes and indicates FEP categories that control the performance-assessment model calculations, i.e., Assessment Basis and Radionuclide Exposure.

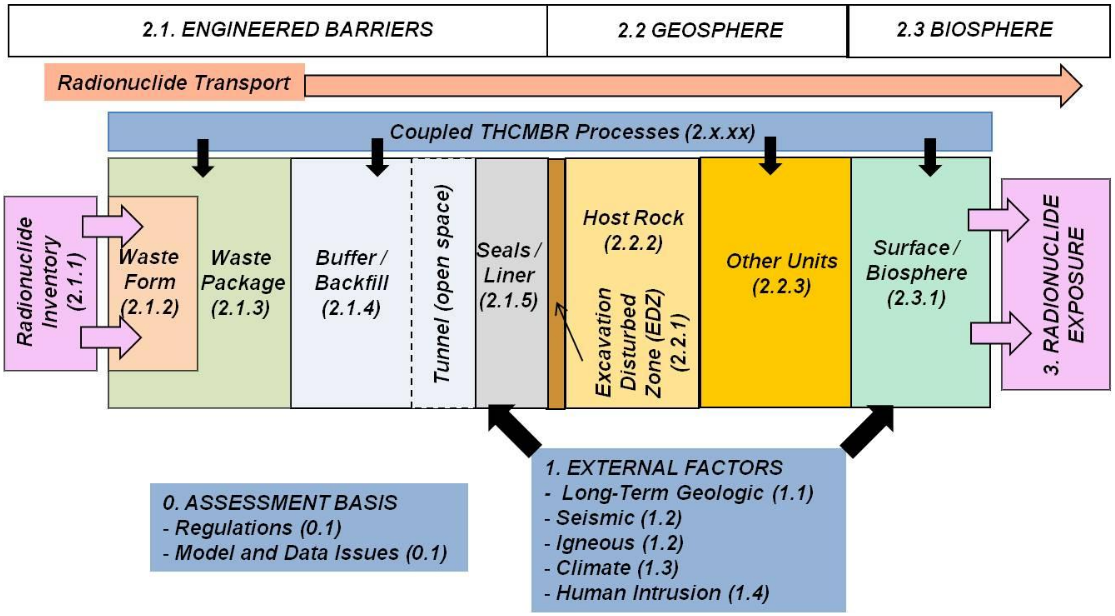


Figure 2-2. Categorization of phenomena (FEPs) for a generalized nuclear-waste disposal system.

The numbers in Figure 2-2 that identify the various components, e.g., ―0. Assessment Basis" and ―Radionuclide Inventory (2.1.1)," follow the section-numbering convention used in NEA (1999). The three-level numbers designate specific categories of FEPs.

In summary, the technical scope of the NEAMS Waste IPSC must be broad enough to represent the range of processes (and associated time- and length-scales) captured by the 204 FEPs for a range of disposal-system alternatives encompassed by the 42 combinations of disposal concepts and geologic settings and types of waste forms.

2.4 Challenge Problem

The NEAMS Waste IPSC challenge problem and a sequence of challenge problem milestones were defined in Freeze, Arguello, et al. (2010) to demonstrate progress in the development of M&S capabilities, the deployment of frameworks, and the implementation of V&V practices. The challenge problem and associated milestones summarized below are a subset of the full scope of the NEAMS Waste IPSC previously described in Section 2.3.

2.4.1 Technical Scope

The challenge problem focuses on a specific type of waste form from Table 2-1, i.e., high-level-waste borosilicate glass, and a specific disposal concept and geologic setting from Table 2-2, i.e., mined geologic disposal in salt. The problem includes the coupled THCMBR processes that describe (1) waste form degradation and the associated mobilization of radionuclides, i.e., the radionuclide source term, and (2) radionuclide transport through the EBS. The challenge problem also includes the effects of coupled

THCMBR processes on the physical and chemical environment in the EBS. The performance measures for the challenge problem include radionuclide mass flux and cumulative release (out of the EBS and across subdomain boundaries); radionuclide mass in place (within the waste form, waste package, and EBS); and spatial distributions of various physical and chemical properties in the EBS, e.g., pH, temperature, fluid saturation, and glass degradation rate.

Development of the NEAMS Waste IPSC to the point where it can be applied to this challenge problem is expected to take several years. For this reason, the process models and couplings supporting the challenge problem will be developed in phases, with additional capabilities and/or couplings being added during each successive phase. The capabilities added during each successive development phase generally build upon, and are coupled to, the previous capabilities. The development phases are defined by a set of sequential challenge milestones that collectively compose the full challenge problem. Progress towards these milestones will provide an indication of intermediate progress toward completion of the entire challenge problem.

The specification of the challenge problem's milestones was derived from the key conceptual model components of the challenge problem. THCMBR phenomenological models will be developed, as noted below.

- Models for the source term will include degradation of the borosilicate glass waste form, degradation of the waste package, and radionuclide solubility.
- Models of the EBS environment will include thermal, fluid chemistry, biology, and mechanical degradation of EBS components.
- Models of EBS transport will include the phenomena of advection, dispersion, and sorption of dissolved radionuclides.

As part of the development effort, the THCMBR phenomenological models will be coupled within and between domains.

The conceptual model components will be developed at both the continuum scale and the performance assessment scale. Where appropriate, the continuum models and/or the performance assessment models will be supported by subcontinuum constitutive relationships. The development of conceptual models at the subgrid scale will be supported by the NEAMS Fundamental Methods and Models (FMM) supporting element and by the Nuclear Energy Waste Form Campaign.

Five specific challenge milestones are identified in Table 2-3 to provide the sequential development of the key conceptual model components listed above at multiple scales of modeling. The table also identifies the degree of coupling (strong or weak) required for each of the phenomenological models, i.e., source term, EBS environment, and EBS transport.

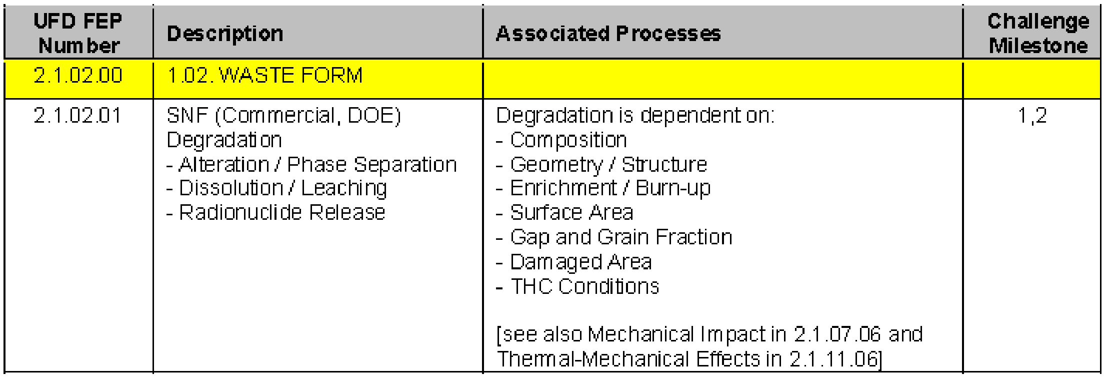
Table 2-3. Milestones for NEAMS Waste IPSC Challenge Problem

No.	Milestone	Source Term	EBS Environment	EBS Transport
1	Chemical Equilibrium Calculation (for concentrated electrolyte solutions)	T=C	T=C	N/A
2	Waste Form and Waste Package Degradation (for high-level waste borosilicate glass)	T=H=C=M	T=H=C	N/A
3	Tunnel Closure (salt creep)	N/A	T=M-H	N/A
4	Heat and Fluid Movement in the EBS (in a salt repository)	N/A	T=H=C	N/A
5	Radionuclide Mobilization and Transport in the EBS (in a salt repository)	T=H=C=M	T=H=C=M	T=H=C=M

Note: The letters T, H, C, and M signify the phenomenological processes, i.e., thermal, hydrological, chemical, and mechanical, respectively. Strong multiphysics coupling is represented by “=”; weak multiphysics coupling is represented by “-“.

Interpreting row 3 in Table 2-3, we see that a disposal scenario will be developed in which the waste form and package will be put in a tunnel, which will be a salt formation. One of the modeling effects is that the salt is going to fold in around the waste form. Models of the source term and EBS transport for this milestone will not be needed, as indicated by ―N/A” in the respective columns. For the EBS environment, though, the thermal and mechanical models must be developed and need to be strongly coupled together, with the hydrological model weakly coupled to them, as indicated by ―T=M-H.”

Each of the challenge milestones is discussed in more detail in Section 3 of *NEAMS Waste IPSC Challenge Problems* (Freeze, Argüello, et al. 2010). The detailed descriptions include discussions of physical processes and couplings, model implementation, code capabilities to be demonstrated, V&V needs, and data needs. The set of FEPs to be addressed by this challenge problem, i.e., the total set of FEPs to be addressed by the five challenge milestones, is a subset of the current set of FEPs defined in that document. It should also be noted that the subset of FEPs relevant to this challenge problem are only addressed for one of the 42 disposal-system alternatives, i.e., high-level-waste borosilicate glass waste form in a mined geologic disposal system in salt. However, the conceptual models and couplings developed for this challenge problem will provide insights and can be readily extended to other disposal-system alternatives. Figure 2-3 shows a FEP entry for the NEAMS Waste IPSC challenge problem.



UFD FEP Number	Description	Associated Processes	Challenge Milestone
2.1.02.00	1.02. WASTE FORM		
2.1.02.01	SNF (Commercial, DOE) Degradation - Alteration / Phase Separation - Dissolution / Leaching - Radionuclide Release	Degradation is dependent on: - Composition - Geometry / Structure - Enrichment / Burn-up - Surface Area - Gap and Grain Fraction - Damaged Area - THC Conditions [see also Mechanical Impact in 2.1.07.06 and Thermal-Mechanical Effects in 2.1.11.06]	1,2

Figure 2-3. Sample FEP entry for challenge problem (Source: Freeze, Arguello, et al. [2010]).

Finally, it should be noted that just as the disposal alternatives and the FEPs may evolve, so may the milestones, or at least specific details of the milestones. For example, this challenge problem does not include radionuclide transport through the geosphere, or radionuclide transport and health effects in the biosphere. However these processes could be incorporated into the challenge problem in a simplified form if necessary.

2.4.2 Computational Scope

The technical scope of the NEAMS Waste IPSC is supported by and integrated with computational capabilities. The computational capabilities and associated considerations relevant to the NEAMS Waste IPSC are as follows:

- V&V: code verification, solution verification, model validation (over the range of disposal-system alternatives, time- and length-scales), upscaling from subcontinuum-scale processes to continuum-scale processes, availability of validation data and natural analogs

- UQ: treatment of aleatory uncertainty, treatment of epistemic uncertainty, propagation of uncertainty through the model, model calibration, sensitivity analysis, availability of input data

- Framework architecture: THCMBR multiphysics coupling, upscaling coupling, analysis workflow (including input pre- and postprocessing and visualization)

- Software engineering environment: software quality, CM (configuration management), data management, software integration

The final integration of the technical and computational scope will produce a NEAMS Waste IPSC that (1) is based on mechanistic and predictive constitutive relationships derived from subcontinuum processes; (2) employs fully coupled THCMBR process models; (3) incorporates robust V&V and UQ (VU)-assessed) approaches to verification validation, and the quantification of uncertainty; (4) implements the best practices of SQE (software quality engineering); and (5) functions in a high-performance computing environment. These advanced M&S techniques will produce a NEAMS Waste IPSC that can be used to simulate the wide range of system designs and conditions demanded by our stakeholders and users.

As noted above, the computational scope (V&V, UQ, SQE tools, and frameworks) will be developed in parallel with the technical, i.e., phenomenological, scope for the NEAMS Waste IPSC. A similar approach will be followed for developing the computational scope for the challenge problem and the associated challenge milestones. However, because the full integration of the technical and computational scope will not occur until after this challenge problem has been completed, the challenge milestones will only include partial implementations of some of the computational capabilities, with a focus on the more interrelated capabilities such as upscaling, UQ, and coupling.

For each of the detailed challenge-problem milestone descriptions in Freeze, Arguello, et al. (2010), there are explicit specifications related to UQ, coupling, and, where relevant, to upscaling. There are also explicit specifications of V&V and data needs. During each of the sequential milestone phases, the state and integration of the computational capabilities with the technical scope will be evaluated. Advancements in computational capabilities will be made with each successive milestone, as listed in Table 2-3.

2.5 Probabilistic Performance Assessments

The intent of a probabilistic performance assessment of a nuclear-waste disposal system is to provide stakeholders with a risk-informed decision analysis regarding the performance of the disposal system. Such an assessment is designed to answer four key questions related to the waste isolation capability of the disposal system (Helton 1999; Pilch, Trucano, and Helton 2006):

I Scenario identification – What can happen?
II Likelihood of scenarios – How likely is it to happen?
III Consequences of scenarios – What are the consequences if it does happen?
IV Credibility – How much confidence do we have in the answers to the first three questions?

The performance of a disposal system is generally described with prespecified metrics, such as the peak or cumulative dose that a hypothetical receptor can potentially receive within a regulatory time frame, typically ranging from 10,000 to 1 million years. The dose is estimated by accounting for the release rates of all radionuclides at a specified disposal-system boundary and the associated health effects of the released radioisotopes. In a license application, the licensee must compare the projected dose with a standard defined by a regulatory agency. Figure 2-4 shows an example of the output from a probabilistic performance assessment from the Waste Isolation Pilot Plant (WIPP) for comparison.

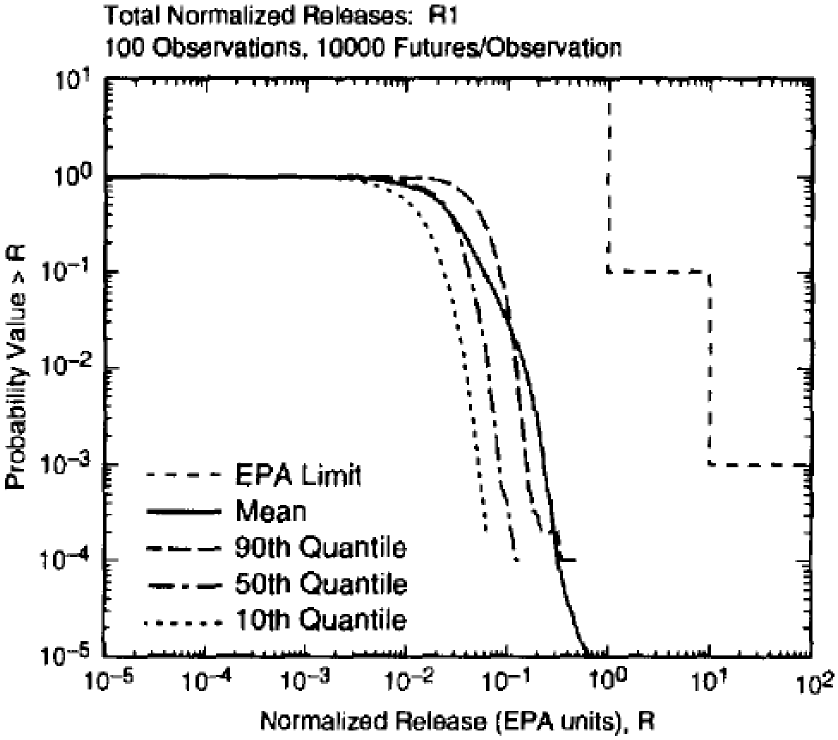


Figure 2-4. Example from Helton (1999) comparing normalized release with Environmental Protection Agency (EPA) limit for WIPP.

The safety margin of a disposal system is defined as the difference between the regulatory standard, denoted by the dashed –EPA Limit” line in Figure 2-4 and the projected value of the performance metrics, denoted by the solid –Mean” line in the figure. A licensee is also generally required to quantify the uncertainty associated with the projected safety margin. The safety margin analysis and UQ are thus an integral part of the performance assessment of a nuclear-waste disposal system. The EPA, for instance, specifically dictates the performance assessment for WIPP in 40 CFR 194. The following are examples of these requirements:

a) The results of performance assessments shall be assembled into –complementary, cumulative distribution functions” (CCDFs) that represent the probability of exceeding various levels of cumulative release caused by all significant processes and events.

(b) Probability distributions for uncertain disposal system parameter values used in performance assessments shall be developed and documented in any compliance application.

(c) Computational techniques, which draw random samples from across the entire range of the probability distributions developed pursuant to paragraph (b) of this section, shall be used in generating CCDFs and shall be documented in any compliance application. (EPA 1998)

It is worth noting that in items (b) and (c) above, the EPA specifically requires the careful evaluation of probability distributions for uncertain model input parameters and the appropriate propagation of uncertainties through performance assessment calculations. Performance assessments for the NEAMS codes will also have to produce data that meet regulatory agency requirements.

For the Yucca Mountain project, the NRC (2001) in 10 CFR 63 defines a similar methodology, although specific details and standards differ from WIPP. It is anticipated that similar methodologies will be adopted for future performance assessments of radioactive-waste disposal systems.

3 Key Concepts

Section 3 presents the key concepts for establishing confidence in the M&S capabilities of the NEAMS Waste IPSC to support risk-informed decision making. As part of the discussion, we define V&V and UQ components, practices, and processes that are critical to understanding and implementing these complex capabilities. A key process in developing confidence in the M&S capabilities is UQ, which involves determining the effect of model uncertainties on system response quantities (SRQs) of interest. The UQ process is often coupled with sensitivity analysis to understand the relationship and importance of changes in model input parameters to the SRQs of interest. Our strategy for addressing model uncertainties is to perform M&S at the three M&S scales (subgrid, continuum, and performance assessment) and to quantify uncertainties at each level based upon quantified uncertainties at the next higher level of resolution. Managing and tracing the evidence obtained throughout the development lifecycle of M&S capabilities is also a critical factor in establishing confidence of the NEAMS Waste IPSC.

3.1 Establishing Confidence in M&S Capabilities

Simulating the performance of waste forms in long-term disposal repositories or waste-storage facilities will involve the use of complex models implemented in computer codes. As stated in Section 1, establishing confidence in the M&S capabilities involves (1) verification of codes and validation of the models implemented in those codes and (2) maintenance and communication of the supporting evidence. Proof that long-term disposal repositories or waste-storage facilities will conform with the disposal facility objectives is not achievable in the ordinary sense of the word due to the uncertainties inherent in the understanding of the evolution of the geologic setting, biosphere, and EBS (engineered barrier system). Our understanding, for example, of how materials in nature behave over extremely long time periods is limited. For such long-term performance forecasts or predictions, what can be provided and what is necessary for decision makers, regulators, and other stakeholders is ―reasonable expectation” that the outcome will conform with the facility objectives, making allowances for the time period, hazards, and uncertainties involved. Simulating performance will involve the use of complex models that are supported by limited data from field and laboratory tests, site-specific monitoring, and natural analog studies that may be supplemented with prevalent expert judgment. In this context, natural analog studies are basically observational investigations in nature to acquire knowledge about phenomena of interest that is analogous to what may be learned in experiments. Performance simulations should not exclude important parameters from assessments and analyses simply because these parameters are difficult to quantify precisely to a high degree of confidence.

As noted by the NRC (1999) in NUREG-1636, the use of the term *prediction*, in the context of the performance of a radioactive-waste disposal facility, must come with caveats because in practice what can be provided is ―an estimate of performance under stipulated future conditions under which a hypothetical repository has to perform” (NUREG-1636 p.3). The actual performance thus may never be known. In this regard, establishing confidence in M&S capabilities includes the appropriate characterization of uncertainties (aleatory, epistemic, and variability) associated with the physical system being modeled and also demonstration that the computational tools when used appropriately represent the processes of interests, calculate the quantities of interests, and propagate the uncertainties.

Levels of Confidence and Rigor

The level of confidence in an M&S capability is dependent upon the level of rigor to which V&V is performed. For example, the use of expert judgment in code verification is generally considered to be a lower level of rigor than the use of mesh refinement studies. Higher levels of V&V rigor are a necessary condition for higher levels of confidence in M&S capabilities.

Early identification of the required level of rigor commensurate with the required confidence in the M&S capabilities is important. For example the Predictive Capability Maturity Model (PCMM) for computational M&S defines the following four levels of rigor (from lowest 0 to highest 3) tied to the intended use of the M&S capability (Oberkampf, Pilch, and Trucano 2007). A mapping of the levels of rigor and risk with examples of the intended use is given in Table 3-1.

Table 3-1. Levels in the PCMM

Rigor Level	Risk Level	Example Usage
0	Low consequence	Scoping studies
1	Moderate consequence	Design support
2	High consequence	Qualification support
3	Highest consequence	Qualification or certification decisions

Lower levels of rigor generally require less effort and have lower costs but result in lower confidence. Ideally, one chooses the level of rigor and then allocates the resources needed to attain that level, regardless of cost. In practice, the attainable level of rigor is constrained by budget pressures, resulting in lower confidence and higher risks.

For many M&S capabilities, the required level of rigor is anticipated to be high as these capabilities will be used for simulation-based decision making such as licensing. Effectively, if high-consequence decisions are being made, the tools used to support such decisions should be applied with a correspondingly high level of rigor. The V&V and UQ (verification and validation and uncertainty quantification) practices in this plan are focused on attaining the highest level of rigor.

3.2 V&V Practices

V&V practices are critical in establishing confidence in M&S capabilities. The concept of V&V as applied to M&S capabilities is refined to define components of M&S

capabilities and specific activities for verification and for validation. The discipline of V&V for M&S is an area of active research and the definitions and activities are evolving. For the current version of this V&V plan, the following definitions are used.

- Verification: *A process of assuring that the implementation of a mathematical model, in the form of a computer code, is free of coding errors, and that the numerical schemes used are within the bounds of required accuracy. The process consists of following established QA procedures during the development of the code, comparison of the code with analytic solutions, and comparison with results from other codes.* (NRC 1999 – NUREG-1636 Appendix C)

- (Model) Validation: *In the regulatory process, it should be noted that a model: (1) may need more or less validation depending on its importance to compliance demonstrations; and (2) is said to be sufficiently validated when it can be used for its intended purpose with some degree of confidence. An example is a flow model used to estimate inflow into geologic repository emplacement drifts is sufficiently validated when it is determined that the calculated inflow for plausible scenarios is within the range of data uncertainties-the validation process may employ theoretical arguments, peer review, laboratory data, field data, and data from natural analogs.* (NRC 1999 – NUREG-1636 Appendix C)

3.2.1 Components of an M&S Capability

Modeling and simulating a physical system requires conceptual modeling, mathematical modeling, numerical modeling, and software development. As illustrated in Figure 3-1, there is a progression associated with verifying and validating an M&S capability that consists of integration testing, code verification, solution verification, and model validation. Each of these V&V practices (also referred to herein as activities and/or processes) may be performed at a level of rigor that is appropriate for the level of confidence required for the M&S capability and that can be accomplished with the available resources.

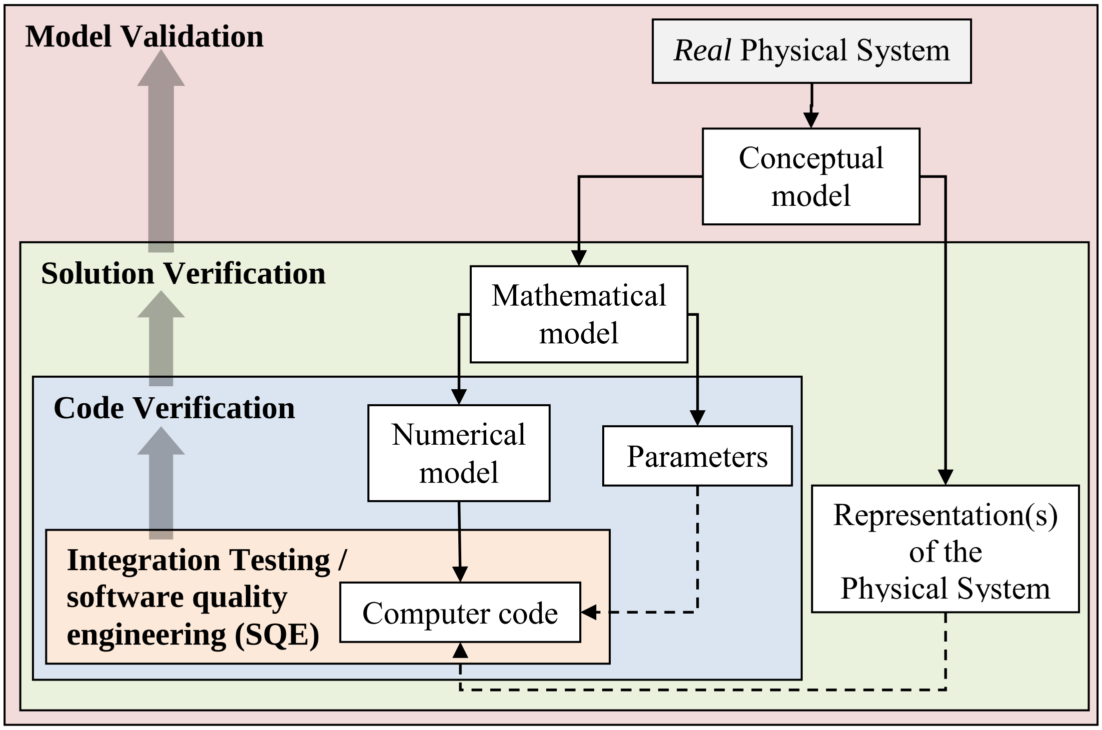


Figure 3-1. Components of M&S capability and associated V&V practices.

Essentially, Figure 3-1 should be interpreted starting at the upper right-hand side. The development of M&S capabilities begins with the real physical system, in the form of a conceptual model, and progresses downward as increasingly more complex models must be produced until the models are effectively implemented in computer code. At that point, the V&V practices (or activities) on the left-hand side of the figure come into play, with the order of these practices typically occurring from bottom to top. These V&V practices are effectively the proof that the type of code or model with which each is associated at the particular level (but not levels above the activity in the figure) is good. For example, integration testing is performed to prove that the computer code is good, but that activity does not in any way give us confidence that the code solves the model; code verification, the next activity in the sequence, must be performed for that confidence to be determined.

The M&S components represented in Figure 3-1 in the white boxes are defined as follows.

- Conceptual model: *A representation of the behavior of a real-world process, phenomenon, or object as an aggregation of scientific concepts, so as to enable predictions about its behavior. Such a model consists of concepts related to geometrical elements of the object (size and shape); dimensionality (1-, 2-, or 3-D); time dependence (steady-state or transient); applicable conservation principles (mass, momentum, energy); applicable constitutive relations; significant processes; boundary conditions; and initial conditions.* (NRC 1999 – NUREG-1636 Appendix C)

- Mathematical Model: *A representation of a conceptual model of a system, subsystem, or component through the use of mathematics. Mathematical models can be mechanistic, in which the causal relations are based on physical conservation principles and constitutive equations. In empirical models, causal relations are based entirely on observations.* (NRC 1999 – NUREG-1636 Appendix C)

- Numerical model: *An approximate representation of a mathematical model that is constructed using a numerical description method such as finite volumes, finite differences, or finite elements. A numerical model is typically represented by a series of program statements that are executed on a computer.* (NRC 2003 – NUREG-1804 Glossary)

- Computer code: *An implementation of a mathematical model on a digital computer generally in a higher-order computer language such as FORTRAN or C.* (NRC 1999 – NUREG-1636 Appendix C)

- Representation of the physical system: The physical or informational characterization of the system being analyzed or the specification of the geometric features of the system. Multiple representations of a physical system may be defined for different mathematical or numerical modeling methods or fidelities, geometric fidelities, or other representation fidelities. Computer-aided design (CAD) drawings could be representations of the physical system.

- Parameters: An implementation of a mathematical model typically consists of computer code and data parameters, e.g., material properties, that are input to that computer code.

3.2.2 Progression of V&V Practices in the Quality Environment

The V&V practices discussed here are listed above in Figure 3-1. Three of these practices (integration testing, code verification, and solution verification) are identified in this V&V plan as within the broad scope of verification. These practices are included because of their prominence in verifying previously developed M&S capabilities. It is assumed that existing M&S capabilities will be imported from their development environments into the NEAMS Waste IPSC quality environment where the V&V practices defined in this plan are applied (see Section 4.1 for a description of these environments). It is expected that these and other verification practices are also applied throughout development of M&S capabilities within their respective development environments.

- Integration testing: The focus of integration testing is to establish confidence that the computer code will run and is free of fundamental coding errors. Additional SQE practices, such as design reviews, add to this confidence. It is expected that integration testing is performed throughout the development process and reproduced when an M&S capability is imported into the quality environment.

- Code verification: The focus of code verification is to establish confidence that the computer code, numerical model, and solution algorithm correctly implement the mathematical model. For example, a mathematical model may be a set of partial differential equations (PDEs), and the numerical model and its associated solution algorithm are based upon a finite element method and iterative nonlinear solver.

- Solution verification: The focus of solution verification is to establish confidence in the correctness in the numerical solution to the mathematical model applied to a given problem. The solution must be free of errors in the problem setup, execution, and postprocessing activities. Errors due to the numerical approximation techniques and numerical solution algorithms should be small and with identified bounds for the solution's quantities of interest. A posteriori error estimates are used to quantify these numerical modeling errors.

- Validation: The focus of validation is to establish confidence that the application of an M&S capability to a given problem domain sufficiently matches experimental results and observations of the real world. Experimental results and observations can include both measurement errors and aleatory uncertainties in the physical properties of the components involved in the experiment. Thus a measure of sufficiency must account for numerical-modeling error bounds, measurement-error bounds, and parameter uncertainties.

The above high-level V&V practice definitions are derived from the report *Predictive Capability Maturity Model for Computational Modeling and Simulation* by Oberkampf, Pilch, and Trucano (2007) and the article ―Verification and Validation in Computational Engineering and Science: Basic Concepts" by Babuska and Oden (2004). Note that there is not a unanimous opinion in the V&V community on the scope or boundaries between integration testing, code verification, and solution verification.

3.3 Uncertainty and Sensitivity Analysis

Confidence in the results computed by the NEAMS Waste IPSC requires clearly identifying, modeling, and quantifying uncertainties for quantities of interest; e.g., migration of radionuclides into the biosphere. Two sources of uncertainty are addressed: uncertainties associated with the parameters of a model and uncertainties associated with the abstraction of a model. The relative importance of uncertainties in models and parameters is determined by analyzing the impact of these uncertainties on the quantities of interest.

- Uncertainty: *Alternative definitions exist for classifying the different types of uncertainty. Generally, there are two types of uncertainty present in any calculation. These are: (1) stochastic (or aleatory) uncertainty caused by the random variability in a process or phenomenon; and (2) state-of knowledge (or epistemic) uncertainty, which results from a lack of complete information about physical phenomena. State-of knowledge uncertainty may be further divided into* *(i) parameter uncertainty, which results from imperfect knowledge about the inputs to analytical models; (ii) model uncertainty, which is caused by imperfect models of physical systems, resulting from simplifying assumptions or an incomplete identification of the system modeled; or (iii) completeness uncertainty, which refers to the uncertainty as to whether all the significant physical phenomena, relationships (coupling), and events have been considered.* (NRC 1999 – NUREG-1636 Appendix C)

- Sensitivity analysis: *Because of the complexity of the systems comprising a geologic repository, it is not usually possible to develop exact analytical expressions for the relationship between repository performance (measures) and the input parameters used to formulate mathematical models. To gain this understanding, quantitative (statistical) evaluations are used to describe the change in a performance measure corresponding to a change in the value or probability distribution of a model parameter. Sensitivity analyses are used to rank parameters according to the sensitivity of the performance measure to the parameters.* (NRC 1999 – NUREG-1636 Appendix C)

3.3.1 Quantification of Uncertainties

As illustrated in Figure 3-2, UQ is the process in M&S of determining the effect of input uncertainties on SRQs of interest. In this illustration, input uncertainties to a simulation model are characterized using a probability density function. When input uncertainties are propagated through the simulation model, a probability density function is obtained for the output quantities of interest, denoted by measures 1 and 2 in the figure.

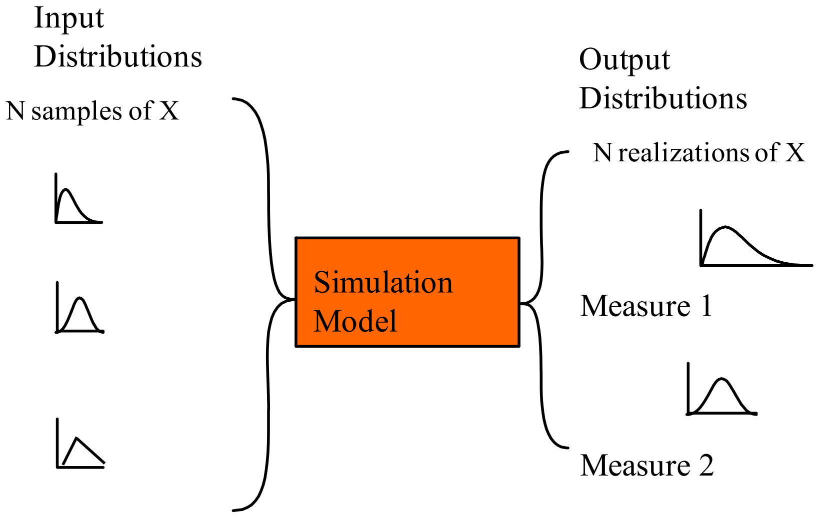


Figure 3-2. Fundamental concept behind UQ.

Further discussion and examples about how uncertainties are propagated through a simulation model are provided by Nelson, Stewart, Unal, and Williams (2010) in their work for the NEAMS Verification, Validation, and Uncertainty Quantification program element, referred to as ―VU.” Two types of uncertainties (aleatory and epistemic) are discussed next.

Aleatory Uncertainties

Aleatory uncertainties are caused by the random variability in a process or phenomenon. Probabilistic methods are commonly used to compute response distribution statistics in the characterization and propagation of probability distribution specifications for aleatory uncertainties. There are three standard approaches for propagating aleatory uncertainties: (1) sampling-based methods such as Monte Carlo and Latin hypercube sampling; (2) reliability-based methods such as the first-order reliability method, the second-order reliability method, and the advanced mean value method; and (3) methods based on stochastic expansions such as polynomial chaos expansions and stochastic collocation.

There are fundamental theoretical differences between sampling methods, reliability methods, and stochastic expansion methods. Sampling methods are the simplest to implement and are very robust, repeatable, and easy to trace when propagation is performed across many simulation codes. However, sampling methods can be quite expensive and potentially inadequate to resolve tail statistics for response functions of interest. Reliability methods transform the uncertainty problem into an optimization problem. Reliability methods can be extremely efficient at finding particular percentiles of a response function, e.g., calculating the response value that corresponds to a 99% reliability. However, repeated optimizations have to be performed if an entire cumulative distribution function is being constructed on the output. Finally, stochastic expansion methods represent a stochastic response measure as a sum of basis functions, where the basis functions are orthogonal polynomials (for polynomial chaos expansions) or Lagrange interpolants (for stochastic collocation).

Epistemic Uncertainties

Epistemic uncertainties result from a lack of complete information about the physical phenomena being modeled. Epistemic uncertainties include the accuracy with which a mathematical model describes the *real* physical system and the accuracy of the numerical solution computed for the mathematical model. For epistemic uncertainties, data are generally too sparse to support objective (frequency-based) probabilistic input descriptions, leading to either subjective probabilistic descriptions, e.g., assumed priors in Bayesian analysis, or nonprobabilistic methods based on interval specifications. Even if we had infinite amounts of data on a random variable, the choice of a probability density function to describe its probabilistic behavior is for the most part subjective. Since there is always the possibility that several probability density functions will model the data equally well, the ultimate choice is dictated by the application and by expert opinion. The limited amount of data permits a wide range or families of distribution to be ―plausible” representations of the given information. This is the big challenge with modeling epistemic uncertainties.

3.3.2 Sensitivity Analysis

A sensitivity analysis supports the quantification of uncertainties by determining the degree to which SRQs of interest, i.e., outputs from a simulation, are impacted by changes in parameters that are input to a simulation. For example, if an SRQ of interest in a simulation was the rate at which dissolved radionuclides flow through a geosphere, we might use a sensitivity analysis to determine the degree to which this rate was sensitive to the porosity of the host rock, where porosity was an input parameter to the model for the type of rock. A sensitivity analysis can be conducted using a variety of methods, ranging from very accurate quantitative methods for local (linear) sensitivities to largely qualitative methods based on expert opinion and prior experience on related efforts.

UQ and sensitivity analysis are coupled in that uncertainties in SRQs of interest are impacted by the product of a model's sensitivity to a parameter and the uncertainty in that parameter. Note, however, that an SRQ of interest can be relatively sensitive to a particular parameter, but if the uncertainty or variability in that parameter is small, then its potential for degrading M&S results will be correspondingly small. A parameter may have a relatively large uncertainty, but if the SRQ of interest is insensitive to that parameter, then its potential for degrading M&S results will also be correspondingly small. Here, we might consider that the flow rate in our above example might also be dependent on (or sensitive to) the chemical composition of the rock, an input parameter with a relatively large uncertainty. But in this hypothetical situation, the porosity of the rock may matter, as opposed to its chemical composition.

As discussed below, a sensitivity analysis can be qualitative or quantitative.

Qualitative Sensitivity Analysis

Qualitative sensitivity analysis methods, e.g., based on expert opinion, identify and characterize plausible sources of uncertainty and sensitivity. Such methods are useful in ranking phenomena in the PIRT and prioritizing M&S development, V&V, and UQ tasks. In the NEAMS Waste IPSC domain, there are too many potential parameters to consider performing an exhaustive quantitative sensitivity analysis. As such, the qualitative sensitivity analysis will be used to ―down select" or choose the subset of parameters for quantitative sensitivity analysis. After the models have been written and the codes developed, parameters selected by qualitative sensitivity analysis may be subjected to quantitative sensitivity analysis. The results from quantitative sensitivity analysis for these parameters will be used to verify the qualitative analyses.

Quantitative Sensitivity Analysis

Quantitative sensitivity analysis provides ―hard" numerical coefficients, bounds, or probability density functions for SRQs of interest for selected parameters. Confidence in these quantitative measures requires that sufficient V&V be applied to the sensitivity analysis. Quantitative sensitivity analysis can be divided into two broad categories: local, i.e., linearized, sensitivities and global sensitivities, which are most likely nonlinear. Local in this context means in the vicinity of a single point in the subdomain of interest, spatially, temporally, and in the parameter space. Global in this context means the whole problem domain and parameter space.

Accurate and efficient computation of linear sensitivities requires mathematical models that are sufficiently smooth (so linear sensitivities exist) and numerical models that support embedded forward and adjoint sensitivity methods throughout the analysis workflow. With a small number of parameters. directional finite differences may give adequate quantitative estimates of linear sensitivities. However, because forward finite-difference sensitivities are known not to scale to large numbers of parameters, embedded adjoint methods need to be considered. Computing adjoint sensitivities is very challenging in a complex, comprehensive simulation code, but this approach is critical for computing linear sensitivities in an affordable manner when there is a large number of parameters (vanBloemen Waanders, Bartlett, Collis, et al. 2005).

In contrast with a local (linear) sensitivity analysis, a global (nonlinear) sensitivity analysis is much more expensive, less scalable, and provides much less confidence on the sensitivities. There are a number of types of global sensitivity approaches, most of which are sampling based. Sampling-based global sensitivity analysis approaches require large (or massive) numbers of evaluations of the simulation. Practitioners will often use reduced-surrogate, i.e., reduced resolution, models derived from continuum-scale models to perform the sensitivity analysis. Such an analysis, however, is expected to have less accuracy and greater uncertainty in the global sensitivity estimates. Alternatively, a semiglobal optimization method, which would reduce the problem instead of reducing the resolution, could be used to try to find the ―worst" points in the domain of the parameters to assess worst-case scenarios and the nonlinear sensitivities there.

3.4 Three Scales of M&S

The NEAMS Waste IPSC will use three scales of M&S, as illustrated in Figure 3-3, to address requirements for validation, UQ, and sensitivity analysis. Subgrid-scale analyses will be used in conjunction with experimental data to characterize material properties and mechanistic processes. It is anticipated that the NEAMS FMM program element will provide many required subgrid-scale simulation capabilities. Results of coordinated subgrid-scale simulations and experimental investigations will be used to develop and verify continuum-scale models. Continuum-scale models will be integrated as necessary to analyze coupled phenomena, i.e., THCMBR (thermal-hydrological-chemical-mechanical-biological-radiological). Capabilities for M&S are abstracted from the continuum-scale simulations to be ―robust and fast" for the performance assessment of large numbers of waste forms, engineered barriers, and geologic settings for the intended, large spatial and temporal scales.

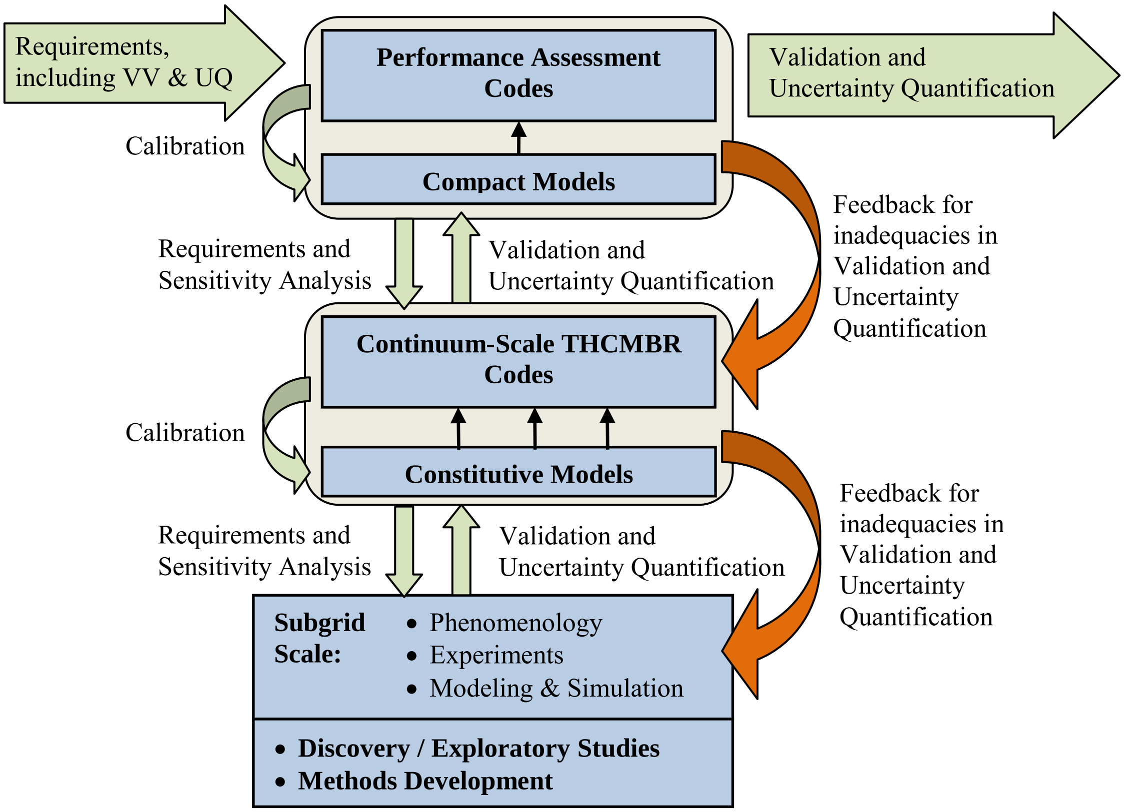


Figure 3-3. Three scales of M&S with requirements, sensitivity analysis, UQ, and validation relationships.

The requirements called out in Figure 3-3 refer to the phenomenological requirements from the PIRT. These requirements are specified at the coarsest scale, the performance assessment scale, and specify the output desired for the problem in the applicable units of that scale. The straight down-arrows labeled ―Requirements and Sensitivity Analysis” denote how the requirements for assessing a particular waste form and waste disposal scenario can be met. If, for example, there are existing codes at the performance assessment scale that meet the requirements, these codes can be used. Otherwise, the requirements that are deemed sensitive flow down to the next (effectively higher-resolution) scale, where the same question is posed. If there are existing continuum-scale models that meet the requirements, those models are used. The requirements flow down to the subgrid scale where possible work must be done if no existing models at the other two scales exist. Validation and UQ of M&S capabilities at each level of this M&S resolution hierarchy are based upon validation and UQ results from the higher-resolution scale as summarized in Table 3-2. Evidence and metrics supporting V&V and UQ of M&S capabilities are necessarily derived from V&V and UQ evidence and metrics from corresponding higher-scale capabilities.

Table 3-2. Summary of Three Scales of M&S

Scale of M&S	Uses	V&V and UQ Basis
Performance assessment (surrogate) simulation capabilities (coarsest resolution)	Performance assessment and design analysis	Based on continuum-scale physics capabilities, V&V, and UQ
Continuum coupled-physics models and simulations	Investigate coupled multiphysics	Based on subgrid-scale capabilities, V&V, and UQ
Subgrid-scale models and simulations (finest resolution)	Characterize material properties and mechanistic processes	Based on experiments and first-principle fundamental methods and models

The strategy for addressing uncertainties in mathematical models is to use model abstractions at the three scales of M&S (subgrid, continuum, and performance assessment) and to quantify uncertainties at each level based upon quantified uncertainties at the next higher level of resolution. Upscaling or propagating subgrid-scale M&S parameters and uncertainties into the continuum scale is, in general, an area of active research and often dependent on the specific model.

Figure 3-3 illustrates the passage of information through the model resolution hierarchy. The information in UQ is passed upward in the model resolution hierarchy to propagate uncertainties from finer-resolution models to coarser-resolution models. In theory, the uncertainties are going to be smaller at the subgrid scale for a given phenomena. The uncertainties at the subgrid scale have to be mapped or propagated to the constitutive models at the continuum scale, At the continuum scale, however, the uncertainty bounds on the constitutive models are expected to be larger and less precise than they are at the subgrid scale. Uncertainties propagated from the continuum models to models at the performance assessment scale are expected to exhibit even larger uncertainty bounds on the phenomena of interest. In contrast to UQ, information in a sensitivity analysis is passed downward in the model resolution hierarchy to refine accuracy and uncertainty requirements for the modeled phenomena. A sensitivity analysis may reveal that certain phenomena are less important and needed only imprecisely, or it may reveal phenomena that have greater impact and must subsequently be improved for greater accuracy and smaller uncertainties.

Rarely is there an exact one-to-one correspondence between parameters computed directly at a subgrid scale and parameters used as input at the next scale. Furthermore, model parameters will be calibrated to better match experimental and observational data, which can correct for systematic errors introduced by the abstractions used in a simplified physical model. These abstractions and calibration of the model parameters will need to be accounted for when uncertainties are propagated by model developers from the finer to coarser scales of M&S. In addition, propagating uncertainties from one scale of models to another scale of models can be numerically challenging because the models at the different scales may not be mappable to each other, i.e., the models may employ different mathematical formulations for the same phenomena. Under such conditions, existing approaches to UQ may need to be refined to propagate the uncertainties in the model resolution hierarchy effectively.

Subgrid-scale models are frequently ―first principles” or deterministic, in the sense that few or no free parameters are available to manipulate. As such, these subgrid-scale models are not calibrated. The fundamental accuracy of these models is limited by the accuracy of the conceptual or mathematical model, such as a specific flavor of density functional in quantum simulations or a particular form of interatomic potential in an atomistic molecular dynamics simulation. When validation is being performed, it is necessary to have an estimate of the sensitivity of the simulation results to the chosen representation of the physical system. This representation includes how many atoms or molecules of different types are in the simulated domain and how they are physically distributed (in clumps, for example, or evenly spread out). Uncertainties in the numerical results arise due to the construction of numerical models, such as integration grids and reciprocal space sampling for solid-state quantum code or as length and time scales for dynamical simulations. It is important to determine and document whether the simulation is sensitive to random variations of the representation of the physical system (such as rearranging the atoms or molecules in different ways) and to develop estimates of the uncertainties in the solution that result from those variations. Sensitivity to random variations of the representation of the physical system will be problematic.

3.5 Evidence Management and Traceability

The management and traceability of verification, validation, and UQ evidence is critical to support risk-informed decisions based upon the NEAMS Waste IPSC. This evidence will not be developed or obtained at one time. It will be accumulated and change over the multidecade lifetime of the NEAMS Waste IPSC. For example, a model at the subgrid scale could become a part of the management system and later be found to be inferior to another newer model. Determining how and where that older model impacted other models would be critical in replacing the older model with the newer model and depend heavily on how well traceability in the management system was implemented. The evidence will also involve complex couplings among M&S capabilities and will be assessed to establish a level of confidence.

For accumulated evidence to be useful, it must be managed in a way that it can be efficiently queried and reported. Hardcopy output does not provide an expeditious means of either querying or reporting, and electronic file keeping is only marginally better than hardcopy output unless some sort of indexing has been applied to the information in the files to speed up retrieval. Thus, we plan to develop an EVIM (evidence information management) system to capture, manage, query, and report V&V and UQ evidence. High-level requirements for this system are discussed next.

3.5.1 Version Identification

Each managed item of evidence needs a version identifier to clearly distinguish it from older or newer versions of the same item. Normally, a version identifier will be a version number, a date, or a combination of the two. A basic version identifier for an evolving model (model X) might be version 1.5. Mapping to versions is a particularly important dimension of traceability because items being traced, such as requirements, software codes, and tests, commonly evolve over time, which means new versions are created and the new versions create new versions of results. Version 1.6 of model X, for example, might contain only slight adjustments that are improvements to version 1.5. Nevertheless, version 1.5 of model X would need to be retained in the EVIM system because it is evidence that supported something that occurred at a previous time. Results and other supporting evidence related to a particular version identifier need to be adequately connected, i.e., cross-referenced, to their source; otherwise, the trace is incomplete or inaccurate. For example, it would be necessary to link the results of a validation exercise performed of the specific version of model X that was used to produce those results. Furthermore, the versions of the codes and tests must be accessible evidence to reproduce and confirm the test results.

3.5.2 Evidence Traceability

One simple example of V&V evidence traceability would be to trace the requirements represented in the PIRT to the M&S code and then to the V&V results for that code. Included in the trace would be the version of the PIRT, the version of the code, and the date the tests that generated the V&V results was executed. This example in actuality is simple because the requirements in the PIRT could potentially trace to any number of M&S codes and from there to the V&V results associated with those codes.

Complex M&S capabilities will be formed by integrating other simpler M&S capabilities. Assessment of the integrated M&S capabilities will depend upon assessments for the component capabilities as well as assessment of the integrated assembly. The integration of M&S capabilities defines a hierarchical structure with its own traceability needs. This type of hierarchical mapping frequently is referred to as a bill of material. Figure 3-4 shows an example of a conceptual hierarchical structure.

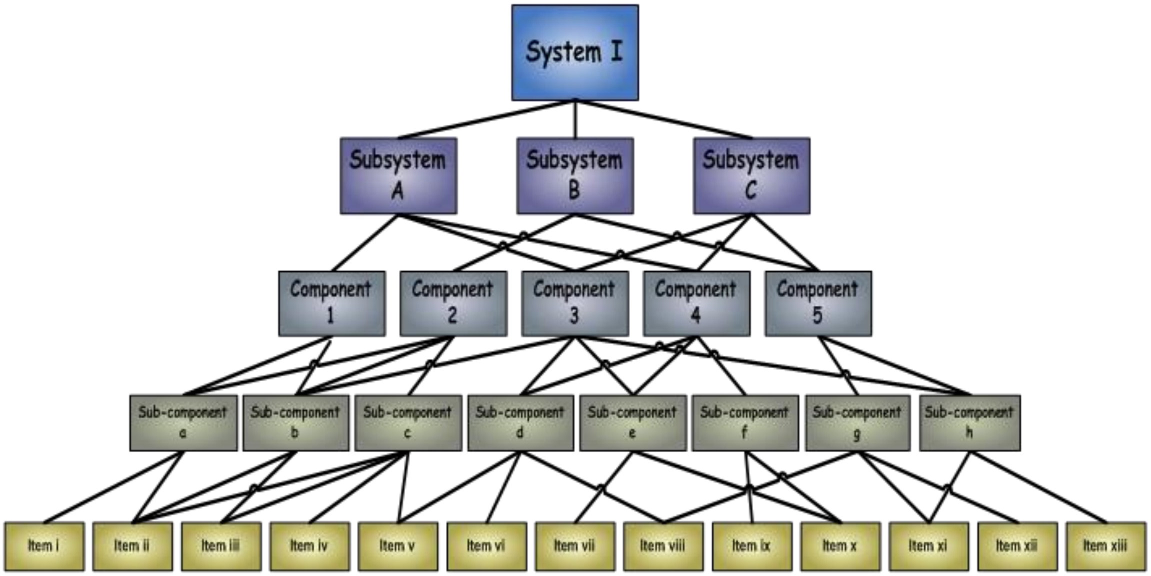


Figure 3-4. Hierarchical integration of components.

In a hierarchical integration structure, tracing can go in two directions, forward and backward. A forward trace begins at the top level or at a middle level and goes down. A backward trace begins at a middle level or a bottom level and goes up. Generally, a forward trace is used for an impact analysis to answer the question, What is the impact if I make a change at this level? For example, using Figure 3-4, the impact analysis of changing component 4 would include evaluating the effect on subcomponents d, e, and f and on items v, vi, vii, viii, ix, and x. A backward trace, on the other hand, is a traceability analysis. It generally answers questions like the following: Why do I have this item? How did this item originate? Do I need this item? A backward trace also can be used to provide roll-up information. For example, verification of higher-level requirements may depend on verification of lower-level requirements, so when verification is completed at a lower level, it is rolled up to the level directly above.

3.5.3 M&S Capability Coupling Frameworks

Three M&S capability coupling frameworks are planned to generate evidence and tracing information for the NEAMS Waste IPSC, as defined by the NEAMS Waste Forms Team (2009).

1. An *analysis workflow framework* for managing, executing, tracking, and reproducing the sequence of steps for an M&S analysis activity

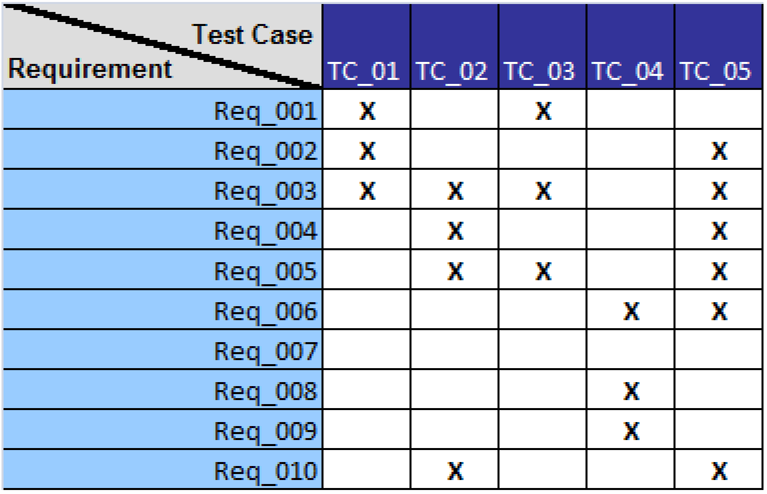
2. An *M&S code coupling framework* for integrating M&S code components into a single simulation executable code that solves coupled M&S equations

3. A *multiscale coupling M&S database* that supports traceability among THCMBR models and parameters of the same phenomena at the different scales of M&S

These frameworks will be used in V&V and UQ activities, and the evidence and tracing information generated by these frameworks will be captured and managed by the EVIM system.

3.5.4 Examples of Tracing Reports

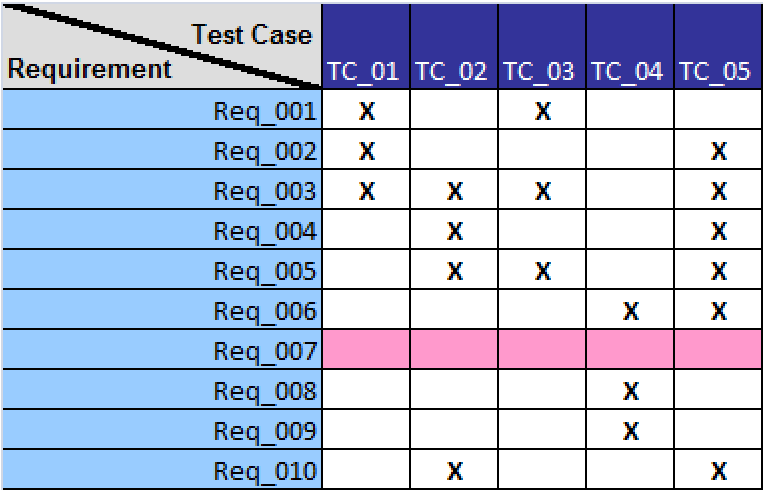
Two fairly standard reports for any kind of traceability are the traceability report and the gap report. The traceability report shows the thread, i.e., chain of evidence, through the interrelated entities and results. There are many different formats that can be used for a traceability report. One format is just a listing of the elements of the thread. Another format is a matrix that shows Xs for the intersection between row and column data. Figure 3-5 shows an example of a very simple traceability matrix that maps requirements to test cases (TC_01 through TC_05). This example could be the answer to a query that asks, What test cases show that these 10 requirements have been adequately met?



Test Case Requirement	TC_01	TC_02	TC_03	TC_04	TC_05
Req_001	X		X		
Req_002	X				X
Req_003	X	X	X		X
Req_004		X			X
Req_005		X	X		X
Req_006				X	X
Req_007					
Req_008				X	
Req_009				X	
Req_010		X			X

Figure 3-5. Example traceability matrix.

A gap analysis report shows where there are holes in the traceability. An example gap analysis report would identify the tests that still need to be run to complete the V&V of a requirement. Figure 3-6 shows a simple example of a gap analysis. In this example, the gap is that Req_007 has no test cases mapped to it, as highlighted by the blank pink row.



Test Case / Requirement	TC_01	TC_02	TC_03	TC_04	TC_05
Req_001	X		X		
Req_002	X				X
Req_003	X	X	X		X
Req_004		X			X
Req_005		X	X		X
Req_006				X	X
Req_007					
Req_008				X	
Req_009				X	
Req_010		X			X

Figure 3-6. Example gap analysis.

4 V&V and UQ Practices

The NEAMS Waste IPSC is composed of an evolving collection of M&S capabilities. Subsets of these capabilities will be used in a variety of analyses that will require varying levels of confidence. Section 4 focuses on the portion of the M&S capability lifecycle that most directly impacts assessing confidence in an M&S capability: the V&V and UQ practices illustrated in Figure 4-1. The arrows in the figure denote the flow, interrelationships, and dependencies of the practices. Each box in the figure specifies the major section number where the practice is discussed. The box in the lower-left corner and the dotted box in the middle of the flow are enabling (or foundational) practices that are prerequisites to the practices performed for V&V and UQ planning and assessment.

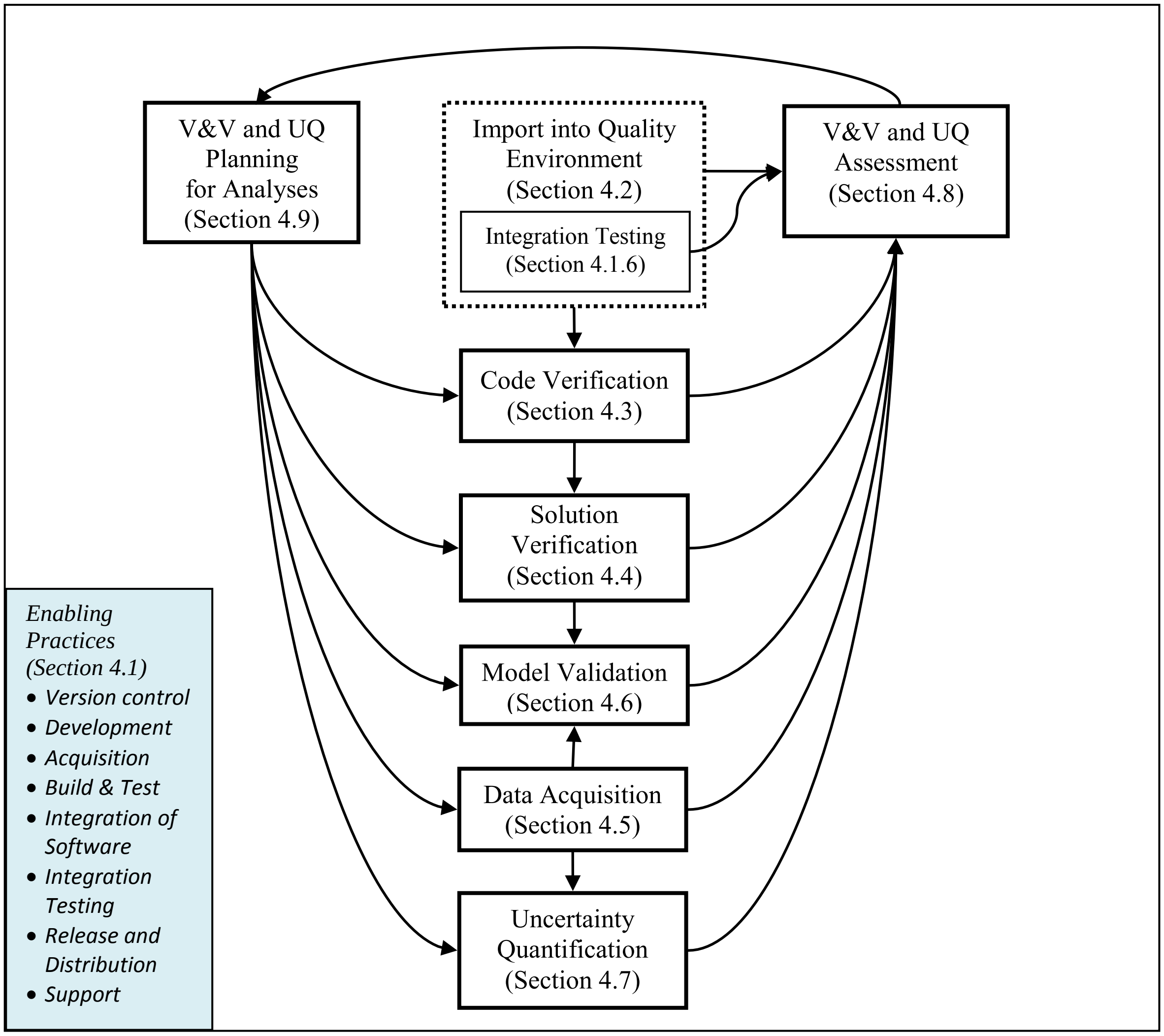


Figure 4-1. Subsets and flow of V&V and UQ practices in M&S capability lifecycle.

The subsets and flow of the V&V and UQ practices in the M&S lifecycle depicted in Figure 4-1 assume the V&V and UQ activities are performed on codes and data that have been imported into a *quality environment*. The NEAMS Waste IPSC quality environment is where developed and acquired code and data are imported for testing and assessment, as shown in Figure 4-2. Codes are imported and integrated into the quality environment for V&V and UQ assessment *prior to use for analyses or as the basis for development and V&V of coarser-scale codes*. Modifications to code and data residing in a quality environment are limited to fixing defects. V&V and UQ activities should be performed on codes and data during development; however, any resulting V&V and UQ evidence must be reproducible in the quality environment.

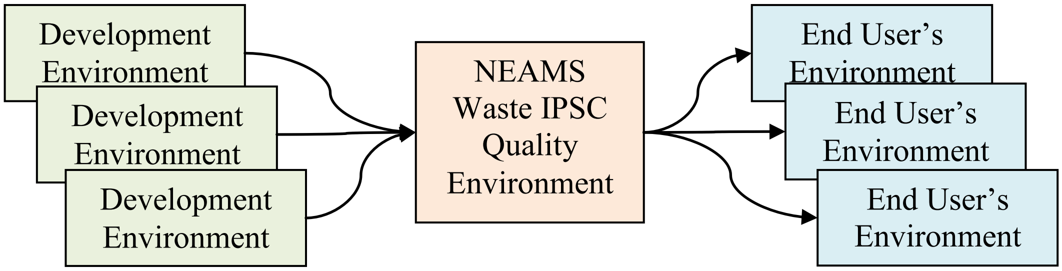


Figure 4-2. Flow of M&S capabilities from development projects through quality environment to end-user environments.

As illustrated in Figure 4-2, codes for the NEAMS Waste IPSC may be developed and integrated in many development environments. Codes at all M&S scales would be developed, tested, integrated, and updated within a CM (configuration management) system in a ―development environment” at Sandia or elsewhere. However, there is a single NEAMS Waste IPSC quality environment that is located at Sandia initially and that will be managed by the NEAMS Waste IPSC team. This quality environment will serve as a clearinghouse from which users in geographically different places will be pulling and copying codes and models from the quality environment into their end-user environments for use in analyses. V&V and UQ must occur in a development environment prior to codes being imported into the centralized quality environment. For this reason, the expectations for enabling (or foundational) practices discussed in Section 4.1 are relevant for codes at all M&S scales in all development environments.

Codes at the subgrid, continuum, and performance assessment M&S scales (discussed in Section 3.4) are imported and integrated into the quality environment for V&V and UQ assessment. V&V and UQ activities follow V&V and UQ plans. These plans are developed to satisfy analysis requirements and are not conducted in an ad hoc manner. V&V and UQ planning and assessment include steps for code verification, solution verification, model validation, and uncertainty assessment as described in Sections 4.3 through 4.7. These V&V and UQ practices will be applied to (1) address additions and revisions to M&S capabilities or the computational environment and (2) assess and address gaps between the measured and required degree of confidence in M&S capabilities. The results from a V&V and UQ gap assessment are used by project management to prioritize, plan, and execute required V&V and UQ activities.

4.1 Expectations for Enabling Practices

V&V and UQ practices and assessment are dependent upon M&S capability lifecycle practices that are included in the lower left-hand box of Figure 4-1, presented previously. Section 4.1 addresses these M&S capability lifecycle practices that must be in place in any development environment before actual V&V and UQ practices are implemented. Expectations for these enabling practices are summarized below.

4.1.1 Version Control

M&S capabilities change over time in response to changing requirements, ongoing development, and resolution of defects. When V&V and UQ activities are performed on an M&S capability, the resulting V&V and UQ evidence must be traceable to the particular version of the M&S capability. Furthermore, that version must be retrievable to support defect analysis and reproduction of the V&V and UQ activities. Version control is mandatory in the NEAMS Waste IPSC quality environment, where M&S capabilities are tested and assessed in preparation for release and distribution to customers.

4.1.2 Development

The term ―development” is used to describe the efforts of creating new or extending existing M&S capabilities to meet requirements and other expectations. Developed M&S software is unlikely to withstand the scrutiny of an assessment unless development is conducted under a valid software engineering process that generates a body of evidence to substantiate the correctness of the software. Examples of this body of evidence can include traceability to requirements, documentation, reviews performed, tests, and test results.

4.1.3 Acquisition

Acquisition applies to obtaining self-contained individual software components from an external development team (e.g., commercial vendor, other DOE program, or open-source software repository) into a development environment. An acquisition process must include obtaining, managing, and verifying a body of evidence to substantiate the correctness of the software. Just as will occur in the quality environment, development environments may acquire software components from external sources.

4.1.4 Build and Test

The building and testing of M&S software must be standardized and repeatable. The build tools (such as Cmake), the test tools (such as Ctest) and the process for using these tools must be fully automated and portable to a wide variety of platforms. The build and test tools and all of their inputs must be under version control to reproduce prior versions of M&S executable codes and tests reliably and to rerun the tests. Testing tools and processes must support multiple types of tests, including but not limited to unit testing, integration testing, system testing, and verification testing. The build and test tools must reliably report on the success or failure of building and testing codes.

4.1.5 Integration of Software

It is expected in the NEAMS Waste IPSC quality environment that software packages will be acquired from multiple development sources and environments. Development environments may also obtain code and data from multiple development sources and environments. Integration requires building and testing to confirm that the acquired or reacquired software is compatible with other dependent software packages. Incompatibilities must be resolved. The recommended software engineering practice of continuous integration supports early detection so that incompatibilities do not accumulate. There are numerous processes, practices, procedures, and supporting tools that need to be put in place to conduct development using continuous integration reliably (Bartlett 2009; Beck 2005; Beck 2003; Duvall, Matyas, and Glover 2007; Poppendieck and Poppendieck 2007).

4.1.6 Integration Testing

Integration testing establishes confidence that the components are free of fundamental coding errors, are correctly integrated, and contain interfaces between the integrated code and the code-execution environment that are operating correctly. Integration testing activities include the following:

- Developing an integration test plan
- Building and confirming that the code can be built on the given code-testing platforms
- Executing and confirming that unit tests for each code and/or module are successful
- Executing and confirming that integration tests are successful
- Rerunning integration tests as necessary to ensure that the integrated code is operating correctly

4.1.7 Release and Distribution

Release and distribution are the core activities that deliver NEAMS Waste IPSC M&S capabilities to end users. These activities are also incumbent upon development projects who provide code and data for importation into the NEAMS Waste IPSC quality environment. The key activities involved in release and distribution are as follows:

- Checking that a unique version-control tag is in place, and if not, creating a unique the tag for all release-related information in the version-control repository. This activity provides for the traceability and reproducibility of M&S capabilities that were delivered to end users.

- Performing a final round of testing, including building the release version on the consumer’s platform or a mock of it. For development projects, the consumer is the NEAMS Waste IPSC quality environment. For the NEAMS Waste IPSC team, the consumers are the end-user environments.

- Creating the release package from the final tagged files. For example, the ―tar” program can be used to compress the files.
- Distributing the officially tagged software release to consumers. For development projects, the consumer is the NEAMS Waste IPSC quality environment. For the NEAMS Waste IPSC team, the consumers are the end-user environments.

4.1.8 Support

This enabling practice is to assist users with the products they receive from others in the M&S lifecycle. The primary activities involved in user support are providing static support material (e.g., documentation, examples, and tutorials) and active user support (e.g., getting emails, phone calls, and bug reports from users and responding to individual specific requests). In the context of development environments, the users would be the NEAMS Waste IPSC team who imported M&S capabilities from the respective development environments. Similarly, the NEAMS Waste IPSC team needs to provide user support to the end-user environments.

4.2 Import into Quality Environment

This practice, which is only performed by the NEAMS IPSC team, is a prerequisite to all of the V&V and UQ practices that will be performed by the team. Developed and acquired codes and data are imported and integrated into the NEAMS Waste IPSC quality environment for confirmation of prior testing results, V&V and UQ assessment, and additional V&V and UQ testing as warranted. The V&V and UQ assessment must occur prior to the use of codes or data for analysis, or as the basis for the development and V&V and UQ of dependent, lower-resolution codes (see Section 3.6).

The following information, at a minimum, must be imported into the quality environment:

- M&S capability information including but not limited to source code, property and parameter data, code-construction information, test cases, and documentation
- Evidence of traceability to the development-source of the M&S capability, to requirements that the M&S capability is intended to satisfy, and among components of the M&S capability

Integration testing. After the NEAMS Waste IPSC team has imported codes and models into the quality environment, the team will need to perform integration testing (see Section 4.1.6). At such time, small changes to code that do not significantly affect the underlying models may be made. For example, no changes would be made to the conceptual model or the mathematical model, but it is possible that the numerical model (in addition to the code) could be changed to fix the identified error.

4.3 Code Verification

The major goal of code verification is to accumulate sufficient evidence to establish confidence that the numerical models and algorithms are implemented correctly and functioning as intended. Code verification practices focus on how correctly the numerical models and algorithms are implemented in the code, and on how accurate and reliable the numerical models and algorithms are themselves (Roache 1998).

In the V&V and UQ lifecycle depicted previously in Figure 4-1, integration testing must precede code verification to identify and remove fundamental coding mistakes before attempts are made to verify a numerical model and the algorithm chosen to solve that model. Similarly, code verification must precede solution verification to establish confidence that the numerical models and algorithms are functioning as intended before attempts are made to verify the accuracy of the numerical solution for a particular problem. Section 4.3.1 specifies the required practices for code verification. Suggestions and guidelines for performing code verification well are provided in Sections 4.3.2 and 4.3.3.

4.3.1 Code Verification Practices

A. Create and Maintain a Code Theory Manual

The manual should describe each numerical model, the mathematical model from which the particular numerical model is derived, and the algorithm chosen to solve the numerical model. Ideally, the manual should also contain a map from each numerical model and algorithm to the mathematical model.

B. Create and Maintain a Verification Test Coverage Table and/or Coverage Metrics

Coverage tables show gaps in the set of code verification tests. A simple coverage table can be created with the rows being the features (or code capabilities) and the columns being the code verification tests. Ensure there is coverage for all features supporting the intended use of the code.

C. Create a Test Plan

The test plan identifies the tests that must be passed to verify the code. The collection of tests must, at a minimum, enable one to verify that the codes' numerical models and algorithms were implemented correctly. Tests with different purposes, such as those demonstrating adequacy, must be included and demonstrate that the requirements are met.

D. Identify Testing Tools

Tests may require a set of built-in or external tools for testing. Examples of such tools are systematic mesh refinement, solution convergence analysis, and the method of manufactured solutions (MMS). The tools should be identified in the test plan at a minimum.

E. Create and Run Code Verification Tests

This practice consists of creating the test input files, writing the code for the verification tests, executing the tests, and determining whether the test results meet the acceptance criteria. As part of this testing activity, identify any exact solutions to the tests that are available. If the test results do not meet the acceptance criteria, troubleshoot the problem to determine whether the errors occurred because the test was wrong or the code being tested was wrong.

F. Maintain a Suite of Verification Tests

The test suite (in addition to integration tests) should be maintained so that code verification tests can be reproduced as the code changes. Note that the NEAMS Waste IPSC team will run the code verification test suites as a part of accepting the code into the quality environment.

4.3.2 Testing for Numerical Model and Algorithm Correctness

Ideally, the mathematical model, the numerical model, and the numerical algorithm are derived prior to their implementation in code and thus exist independently of any particular computer code. The combination of the numerical model and the algorithm chosen to solve the numerical model describes a procedure by which numerical approximations to the exact solution to the mathematical model can be produced. Sections 4.3.2.1 and 4.3.2.2 discuss different approaches for testing the correctness of this combination. Additional guidance is provided in Section 4.3.2.3. Importantly, the considerations addressed in these sections regarding implementation of the numerical model and its algorithm emphasize the continuum-scale codes and the performance assessment–scale codes that are derived from sampling the continuum codes via Monte Carlo methods. Additional research is needed for verification methodologies and practices for subgrid-scale M&S and more general M&S at the performance assessment scale.

4.3.2.1 Asymptotic Convergence for Spatial Discretizations

Many numerical models and algorithms use meshes to discretize the geometric domain. In this context, the measure of error in the numerical approximation must approach zero in the asymptotic limit as the mesh is systematically refined. Mesh refinement studies are a primary tool in code verification. In these studies, a sequence of refined meshes is created, and the solution on each is computed. The set of solutions created from the refined meshes is compared to the exact solution to determine an observed order of accuracy. If the observed order of accuracy is positive, the asymptotic numerical approximation error can be driven toward zero. Sequences of numerical solutions that exhibit the correct asymptotic behavior are said to be in the *asymptotic regime* of the numerical method.

A numerical model, the algorithm chosen to solve the numerical model, and the integration of the model and the algorithm are said to be correct if the numerical approximation error approaches zero in the asymptotic limit. Establishing the correctness of the implementation is a necessary part of code verification. In other words, it is expected that the error behavior in the asymptotic limit will be verified.

To establish that the numerical approximation error approaches zero in the asymptotic limit, a comparison needs to be made between the code's numerical results and the exact results obtained from the corresponding mathematical model. Any metrics defined for this comparison are part of the numerical modeling requirements. If it is difficult or impossible to obtain an exact solution to a particular problem, the MMS (method of manufactured solutions) can be considered as a tool for code verification (Knupp and Salari 2003).

Order-verification is a mesh refinement study in which the formal order of accuracy is known a priori. In order-verification, the observed order of accuracy of the numerical solution is compared to the formal order of accuracy to determine whether they are the same. It has been found in practice that order-verification is an effective means of identifying incorrect implementations of the numerical algorithm (Knupp, Ober, and Bond 2007; Bond, Knupp, Ober, and Bova 2007; Bond, Knupp, and Ober 2005; Bond, Knupp, and Ober 2005) .

4.3.2.2 Other Techniques

Establishing that the numerical model and the algorithm chosen to solve the model are implemented and integrated correctly can involve more than investigating the spatial discretization asymptotic error of the implementation. For example, if the code has an adaptive time-stepping algorithm, correctness may need to be established using other techniques. In addition, it is often the case in practice that the numerical model and the algorithm are not derived prior to their implementation in software and little is known concerning their asymptotic or other properties. Nonetheless, it remains important to verify that the code can produce correct numerical approximations to the solution to the mathematical model. Other potential code verification techniques are as follows: (Oberkampf and Trucano 2003)

- Analytic solutions for simplified physics
- MMS
- Ordinary differential equation (ODE) benchmark solutions
- PDE (partial differential equation) benchmark solutions
- Conservation tests
- Alternate coordinate system tests
- Symmetry tests
- Iterative convergence tests

4.3.2.3 Necessary but not Sufficient

Code verification does not demonstrate that the code *will* produce correct numerical results, only that the code *can* produce correct numerical results when it is properly used. For example, a code that has undergone extensive verification can produce incorrect numerical results if the user makes a mistake in the input. Additionally, if the mesh used by the code is too coarse or the time step is too large, the numerical solution may be inaccurate and misleading. Moreover, a code can produce no numerical solution at all if, for example, the numerical algorithm fails to satisfy an iterative convergence tolerance. These examples demonstrate the need for additional V&V and UQ practices, such as solution verification.

4.3.3 Testing for Adequacy

Establishing the correctness of the numerical model and the algorithm chosen to solve the model is a necessary part of code verification; however, this is often insufficient. The adequacy of the numerical model and algorithm for their intended use must also be demonstrated. It is from the intended use that the parameters and domain geometries of the problems to be simulated are determined. The intended use also defines the performance needs of the end user. Thus, given the computational resources (memory, computational capacity, and time limitations) of the end user, can the code produce correct numerical solutions that are in the asymptotic regime for the physical problems of interest? From this perspective, a numerical model can be inadequate, for example, if (a) its asymptotic rate of convergence is too low, (b) it is inefficient, (c) its iterative algorithm does not converge quickly enough, or (d) it is not robust. It is not difficult to find examples in M&S practice in which the numerical models and algorithms are correct but are not adequate. It is thus important to assess adequacy so that the appropriate level of confidence is assigned.

Because it is impossible to anticipate in advance all the details of the physical problems that an end user will present to the code, adequacy is often investigated with benchmark problems. A benchmark problem is a test intended to resemble an end user’s problem that is run to assess adequacy from an end user’s perspective. Although the exact solution to a benchmark problem may or may not be known, a benchmark problem can be used to estimate (1) the memory-size of the mesh needed to reach the asymptotic domain, (2) the CPU time needed, (3) the number of nonlinear iterations required, and (4) the robustness of the algorithm when applied to a problem that is close enough to (or representative of) the real problem being investigated. A benchmark problem can also be useful in ferreting out other types of algorithmic deficiencies. The NEAMS Waste IPSC Challenge Problem can be considered a benchmark problem (Freeze, Arguello, et al. (2010).

4.4 Solution Verification

The focus of solution verification is to establish confidence in the accuracy of the numerical solution to the mathematical model applied to a given problem, e.g., benchmark or validation tests. The solution must be free of errors in the problem setup, execution, and postprocessing activities. Discretization-dependent, e.g., mesh-dependent, numerical solutions must be in the asymptotic convergence regime. Errors due to the numerical approximation techniques and numerical solution algorithms must be small and within identified bounds for the solution’s SRQs of interest.

One facet of solution verification is a quantitative estimation of accuracy for the numerical solution to the mathematical model. The primary numerical errors that are estimated in solution verification are due to (1) spatial and temporal discretization of the PDEs and (2) iterative solution error resulting from a linearized solution approach to a set of nonlinear, coupled equations. The importance and difficulty of numerical error estimation has increased as the complexity of the physics and mathematical models has increased (Oberkampf, Pilch, and Trucano 2007).

In solution verification, the exact solution to the mathematical model is always unknown, and thus the exact discretization error is also unknown. If the exact solution is known, then the test problem being run through the simulation is a code verification test and not a solution verification test. Solution verification must be performed only after code verification has been applied to establish confidence that the numerical models and algorithms are correctly implemented.

Solution verification should be performed on a simulation before the results from that simulation are compared to experimental results, observations, or results from validated higher-resolution M&S capabilities. Solution verification should also be performed

- to support sensitivity analyses conducted to characterize parameter uncertainty,
- to support the development of performance assessment–scale models derived from continuum-scale models,
- in performance assessments that use continuum-scale codes within a Monte Carlo procedure, and
- when using the code to make predictions.

The practices for solution verification follow in Section 4.4.1. Performing solution verification at different scales is addressed in Section 4.4.2. Developers, members of the NEAMS Waste IPSC team, and end users may all have responsibilities to perform solution verification at some level in their respective V&V activities.

4.4.1 Solution Verification Practices

A. Review the Code's Inputs and Outputs

The primary goal of this practice is to verify that the inputs to the code were correct and ensure that the intended problem was run in the simulation. The practice involves verifying physical data input, numerical model inputs, model options, boundary and initial condition data, solution algorithm options, flow-control options, the mesh and the time-steps, and any other inputs to the particular code. The following are examples of questions that should be asked:

- Are the physical data used in the calculation traceable back to the right parameters in the database?
- Were appropriate types of termination criteria for iterative solution strategies used? For example, was the spatial discretization being automatically adapted to make sure the solution was in the convergent regime?

- Have the right model and flow-control options been selected?
- Does the mesh contain inverted elements or other problems?

In terms of verifying the output, it is important to verify that the calculation terminated properly, for example, because the iterative-solution convergence criteria were satisfied and not because the iteration limit was reached. The use of numerical solutions resulting from improperly terminated calculations is dangerous when performing automated sets of calculations and should not be accepted.

The input can be verified at several levels of rigor that are reflected in the PCMM (Predictive Capability Maturity Model): verification of the input through visual inspection by the analyst, verification of the input through visual inspection by a knowledgeable peer, or reproduction of the input by an independent party (before looking at the other’s input).

B. Analyze the Solution’s Sensitivity to Numerical Model Parameters

Parameters appearing in the numerical model can control and alter the numerical solution. Examples of such parameters are controls on the convergence of linear and nonlinear solvers, numerical damping parameters, and limiters. The sensitivity of the numerical solution to these types of parameters must be explored to ensure that the values are appropriate. For example, are the convergence criteria too tight, too loose, or “just right”? The possibility of numerical round-off error should be considered, particularly if the problem might be ill conditioned.

C. Perform Systematic Mesh Refinement or Coarsening Studies

The most effective way to determine whether the numerical solution is in the asymptotic range is to compute the solution using several levels of mesh and/or time-step resolution. Although the exact solution is unknown, the trends in the solutions at the different levels of resolution can be examined to determine whether the solutions change in a systematic way that is predicted by the expected order of accuracy.

It is recognized that, in practice, systematic mesh refinement and coarsening studies can be difficult to do properly, particularly on unstructured or highly nonuniform meshes. However, the often-offered excuse “that these studies cannot be done because the memory and central-processing-unit (CPU) limits of the computer on which the calculations are performed are exhausted” is not valid because one can always coarsen the resolution instead of refining it. The best practice is to simultaneously refine both the mesh discretization parameter and the time step. Attempts to refine the time step only, using a highly resolved spatial solution, can be misleading.

If the expected trends in the solution are observed, then one can conclude that the solution is in the asymptotic range. Moreover, if the code has been subject to rigorous code verification, one can also conclude that the sequence of numerical approximations is moving toward the correct solution to the continuum equations.

If the expected trends in the solution are not observed and the systematic mesh refinement has been properly carried out, then one is forced to conclude that the numerical solution is not in the asymptotic range. In that case, use of the numerical solution in the contexts mentioned above is highly problematic. If the decision is made to use a nonasymptotic solution anyway, the fact should be documented and identified as an epistemic uncertainty.

D. Estimate Error

Because the exact solution is unknown, the discretization error cannot be computed in solution verification. A posteriori error estimation computations are needed to quantify uncertainty in the solution due to the discretization error. These should be applied particularly to the SRQs (system response quantities) of interest, such as solution functionals. One of the most versatile methods for estimating error is Richardson's extrapolation. Error estimation by Richardson's extrapolation requires that the numerical solution be in the asymptotic range and thus requires mesh refinement studies. The method is versatile in that it applies to many different types of equations, not just elliptic PDEs. Other methods for estimating error may be used.

It is equally important to calculate error bars around the solution along with the estimated discretization error. These error bars are required to fully characterize uncertainty due to the discretization error.

E. Use Mesh Adaptivity

Some M&S codes use solution-adaptive meshing to locally refine a mesh (h-adaptation) in order to reduce the discretization error. Other solution-adaptive methods include p-refinement, r-refinement, or time-step adaptivity to reduce the discretization error. The best of these methods permit one to input a desired level of discretization error and refine the mesh accordingly. The exact discretization error resulting from these methods may not always be constrained by this limit, however. Solutions produced by these methods must still be verified via the other solution verification practices discussed in this section.

4.4.2 Multiple-Scale Considerations

The solution verification practices above emphasize continuum-scale codes, including codes with coupled multiphysics. With proper modifications to these practices, solution verification is also relevant to M&S codes that do not compute numerical approximations to the solutions to their mathematical model, but compute, up to computer round-off, exact solutions. Modifications to these practices entail a change to the definition of correctness and adequacy. Exact (up to round-off) numerical solutions are deemed *correct* if the code inputs and outputs are correct; these solutions are *adequate* if the round-off error has been characterized.

Numerical solutions produced by subgrid-scale codes can additionally be assessed by checking the correctness of the inputs and outputs and by investigating the sensitivity of the solution to the number of atoms or particles in the calculation. Solution verification for performance assessment–scale codes is an open research question, and little has been published on the topic.

4.5 Data Acquisition

Measurement data are acquired from real physical systems through controlled experiments or field observations. Measurement data are used during model development to conceptualize and parameterize constitutive models. Measurement data are used during model calibration and validation to assess accuracy through quantitative comparison with simulation results. Effective validation requires that adequate and suitably targeted measurement data be available to perform quantitatively meaningful comparisons with ―computational observables.” Thus, each data-acquisition activity must (1) identify and record sources of the acquired data and (2) assess and record the suitability of the acquired data.

The current scope of the NEAMS Waste IPSC does not include generating or commissioning experiments or field observations. This program element plans to rely upon data acquired from external sources to satisfy development, calibration, and validation requirements. Acquiring these data will require interaction and coordination with robust experimental programs. The NEAMS Waste IPSC plans to actively engage with the DOE Waste Form and Used Fuel Disposition Campaigns to identify appropriate contacts and coordinate activities.

Experimental and observational measurement data are expected to have greater credibility than data generated through simulations. It is expected that suitable measurement data for model conceptualization, calibration, or validation cannot be obtained to directly support all required M&S capabilities. The ―three scales of M&S” strategy, discussed previously in Section 3.4, addresses this expected deficiency by treating results from ―validated,” higher-resolution M&S capabilities as measurement data for the next lower-resolution M&S capability. Thus, development, calibration, and validation activities at the continuum scale can treat results from *validated* subgrid-scale, e.g., atomistic, analyses as measurement data. Similarly development, calibration, and validation activities at the performance assessment scale can treat results from *validated* continuum scale analyses as measurement data.

Data acquired from experiments, field observations, or analyses will be entered into the EVIM system. Records must be sufficiently complete to support traceability, to provide data and evidence necessary for credibility, and to enable assessment of the adequacy or gaps in the data and evidence. Section 4.5.1 presents the required practices for data acquisition. The near- and interim-term strategies for data acquisition by the NEAMS Waste IPSC team are described in Section 4.5.2.

4.5.1 Data Acquisition Practices

A. Identify Sources

The NEAMS Waste IPSC will require measurement data acquired from external sources. Potential sources of data for validation and model development data are as follows:

- Published literature and government and industrial reports
- Unpublished data and ―private communications”
- Coordination with DOE Waste Forms and Used Fuel Disposition Campaign activities
- Tabulated results (databases) maintained by independent organizations
- Leveraged collaborations, e.g., Fuel Cycle Research and Development (FCRD), the DOE Office of Basic Energy Sciences (BES), and Nuclear Energy University Programs (NEUP)

Although the acquired data may not be complete or otherwise ideally suitable, such data may be the only source of validation and model-development data available for particular activities. The source of the data must be identified and recorded to establish provenance.

B. Acquire Data

Experimental, observational, and analogous analyses-generated data will be entered into the EVIM system along with the following attributes of the acquired data.

- Provenance – who generated the data, where the data were found (literature or other source). If the data were found in multiple sources, all should be listed.
- Description of how the data were acquired, i.e., experimental inputs: samples, experimental conditions, apparatus, and experimental procedures and protocols. The description can be a reference to published literature or a report.
- Uncertainties in the data
- Description of the sources of uncertainties and uncertainty analysis. The description can be a reference to published literature or a report.
- Assessment of the quality and gaps in the data, especially those not identified by actual absences of other entries.

It is anticipated that when data are acquired and entered into the EVIM system, not all the information may be available to fully populate the above attributes. In such a case, incomplete or missing information must be noted with the acquired data. This record of missing attributes can then serve as a source for refining requirements and planning V&V and UQ activities.

C. Assess Uncertainties and Data

Acquired data are subject to data verification and validation, namely, that the data were collected correctly and the correct data were acquired. Data verification asks, for example, was the experiment conducted correctly? Data validation asks, for example, was the correct experiment performed?

Acquired data that will be used for validation need to have their uncertainties quantified. Measurement data inherently have aleatory uncertainties due to potential variability in the physical system and variability in the measurement devices and processes. Sources of aleatory uncertainties may also arise in the numerical data analysis of the raw data. Formally, numerical data analysis must be subject to the same V&V and UQ requirements of simulations method development, but these analyses are typically embedded in the experimental or observational study, not described as simulations. These analyses are a frequently neglected source of uncertainties in quoted experimental values.

In addition, measurement data often have epistemic uncertainty due to lack of knowledge about the physical system. Potential sources of epistemic uncertainty in experimental data and field observations include the following:

- Potentially faulty underlying assumptions about what is being measured
- Unknowns in the physical system, such as partially characterized samples, uncontrolled boundary conditions, and an imprecise system history
- Imperfect alignment between the physical quantity of interest and the computed quantity of interest

Records of measurement data in the EVIM system must contain information that addresses the sources of aleatory and epistemic uncertainty. Data used in validation studies should be well characterized, which includes analysis of these uncertainties.

Inputs to experiments must be documented sufficiently to enable reproducibility and assessment of procedures by independent subject-matter experts. Examples of such inputs are (a) the characterization of samples and experimental conditions, (b) a description of the experimental apparatus, and (c) experimental procedures and protocols. An experimental study should summarize the results of an ensemble of measurements sufficiently well to provide statistically meaningful analysis of the measurement and sample uncertainties. The data analysis embedded in the experimental study must be documented, including its assumptions and uncertainties. Frequently, this documentation will be incomplete in preexisting literature or reports, or it may be available only in private communications or unpublished data. For this reason, the V&V activities must be coordinated with ongoing experimental efforts to obtain complete and applicable data to satisfy IPSC validation and model development requirements.

D. Address Gaps in Data Uncertainties

Quantitative validation is dependent on accurate measurement data and quantitative descriptions of the uncertainty in the data. Measurement data can provide direct input to M&S parameters, and the associated uncertainties can support validation and uncertainty analyses. Often, data from a source will be incomplete or inadequate, and further work is required to address such gaps. Possible mitigation strategies for gaps in measurement data uncertainties include the following:

- *Extracting estimates of uncertainties from data synthesized from an ensemble of sources*. An ensemble of data acquired from multiple sources may reflect possible variability in samples and experiments as compared to an ensemble acquired from a single source. Such a multiple-source ensemble can be helpful in identifying bias introduced by samples, equipment, or procedures in the experiments. Synthesized uncertainty estimates must be conservative. The process by which the estimates were synthesized needs to be documented.

- *Eliciting estimates of uncertainties from subject matter experts.* For experimental studies in which uncertainties are not documented, expected uncertainties and errors may be estimated by subject matter experts, provided that those experimental studies adequately describe their experiments. The sources, i.e., subject matter experts, and their rationale for the estimates need to be documented.

- *Identifying needs for additional experiments*. This mitigation strategy requires (1) determining inadequacies in the experimental data to support validation, (2) updating the PIRT to raise the priority of a phenomenon for which more experimental data are needed, and (3) identifying needs for experimental support to more precisely evaluate experimental uncertainties.

4.5.2 Near-Term and Interim Data Acquisition

Development of a formal NEAMS Waste IPSC V&V EVIM system is a major component of this V&V plan; however, this system has not been designed or implemented. Nonetheless, the measurement data will be acquired to support NEAMS Waste IPSC development activities, including validation and UQ. It is expected that such data acquisition will include the capture of sufficient information to satisfy these data acquisition practices and populate the required EVIM records. Near-term data acquisition requires deployment of an interim (ad hoc) information management solution to capture acquired data and records for entry into the planned EVIM system.

It is anticipated that the DOE Waste Form and Used Fuel Disposition Campaigns and other program elements of NEAMS, including FMM, will also wish to make use of a centralized database for experimental data involving waste form assessment. The development of the EVIM system could therefore involve consultations with multiple partners, such as the NEAMS VU, ECT, and Capability Transfer (CT) cross-cutting elements.

4.6 UQ

A number of challenges for applying current UQ methods to the NEAMS Waste IPSC have been identified:

- At the subgrid scale, uncertainties in material properties and mechanistic processes must be quantified and propagated.

- Upscaling from subgrid-scale models to continuum models (or coupled multiphysics continuum to surrogate performance assessment models) must address the uncertainties introduced from upscaling.
- Uncertainties from coupled M&S capabilities must be integrated into a single uncertainty estimate for the integrated M&S capability.

The NEAMS Waste IPSC is taking a comprehensive view of uncertainty to include traditional categories, such as parametric, data, model form, and numerical, as well as new uncertainties that result from the upscaling and downscaling between models and from calibration used to characterize unknown uncertainties. The resulting UQ framework should be based on a proper computation and mathematical analysis and should address a number of a number of challenges, including how to (1) integrate all forms of uncertainty quantitatively, (2) propagate uncertainty between multiphysics models, (3) estimate uncertain model inputs based on data for model outputs, and (4) reduce uncertainty adaptively in model and system responses.

4.6.1 UQ Practices

A. Propagate Uncertainties

When sufficient data are available for characterizing aleatory uncertainties, probabilistic methods can be used to compute a probability density function for the response of the system based on input probability distribution specifications. For epistemic uncertainties, data are generally too sparse to support objective probabilistic input descriptions, leading to either subjective probabilistic descriptions, such as assumed priors in Bayesian analysis, or nonprobabilistic methods based on interval specifications. Standard approaches for propagating aleatory uncertainties include Monte Carlo and Latin hypercube sampling; analytic reliability methods, such as the first-order reliability method, the second-order reliability method, and the advanced mean value method; and stochastic expansion methods, such as polynomial chaos expansions and stochastic collocation.

Approaches for treating epistemic uncertainty include interval methods, possibility theory, evidence theory, fuzzy set theory, and probability theory. At Sandia, we have focused mostly on (a) interval methods where epistemic variables are only assumed to have possible values within a set of interval bounds or (b) evidence theory.

Uncertainty propagation becomes more challenging when both aleatory and epistemic uncertainties are present. A common approach to quantifying the effect of mixed aleatory and epistemic uncertainties is to perform second-order probability analysis. This approach is typically implemented with nested sampling loops: an outer loop for epistemic uncertainties and an inner loop for aleatory uncertainties. These nested loops can be computationally expensive; however, the nested-loop approach has the advantage in that it allows the user to analyze the aleatory and epistemic uncertainties separately. A new more efficient approach for second-order probability analysis has been proposed that uses a stochastic expansion method for the inner loop and an interval optimization method for the outer loop (Helton 2009).

B. Estimate Numerical Errors

The UQ methods referenced in the ―Propagate Uncertainties" practice above assume that numerical errors do not affect the statistics computed by the UQ methods. When significant numerical errors are present and not accounted for, they will have a large effect on the accuracy of the computed uncertainties. Therefore, it is critical to have general error and uncertainty estimates that account for numerical errors. Systematic a posteriori error estimation for SRQs of interest in coupled multiphysics and multiscale simulations is an area of active research that is being pursued by the NEAMS VU program element in support of the NEAMS Waste IPSC and other NEAMS program elements.

C. Calibrate and Upscale

Uncertainties in M&S input parameters can be computed by calibrating these parameters with measurement data obtained from experiments or observations. This inverse problem is solved to determine what input parameters will cause the output parameters to match the measurement data. In the deterministic case, the problem of parameter estimation is based on inverting the model based on a set of model responses (the data) to solve for the parameters. Typically, the problem to be solved is either over- or underdetermined and hence solved by imposing some form of regularization, e.g., least squares.

Upscaling is a related parameter-estimation problem where data from a fine-scale–distributed parameter or subgrid-scale model must be projected as an input into a coarser model. Examples include porous flow heterogeneity (permeability data), model reduction, and multiscale computations where properties are upscaled with large-scale reductions. When upscaling is based on matching model responses, it becomes an inverse problem for the unknown coarse model parameters, which can be deterministic or stochastic.

4.7 Model Validation

The goal of model validation is to establish confidence that the implementation of the model for a given problem domain sufficiently matches experimental results and observations of the real world. Accumulated evidence is assessed to determine whether a model is valid for the SRQs of interest within the domain of intended use. Comparisons are expected to be statistical because experimental and observational data provide a distribution of measurements and M&S results inherently have approximations, uncertainties, and errors relative to the *real* physical system.

The full spatial, temporal, and coupled THCMBR domain of the NEAMS Waste IPSC intended use is outside the domain over which it is feasible to perform experiments or observations. For example, it would not be feasible to conduct an experiment spanning thousands of years and hundreds of kilometers. In the three scales of M&S strategy discussed previously in Section 3.4, comparison of M&S results with experimental results and field observations will occur through the hierarchy: experiments & observations → subgrid scale → continuum scale → performance assessment scale. In this hierarchy, subgrid-scale M&S capabilities are validated over a given domain through comparison with observations and experiments. Continuum-scale M&S capabilities are validated through comparison with *validated* subgrid-scale M&S capabilities. Performance assessment–scale M&S capabilities are validated through comparison with *validated* continuum-scale M&S capabilities.

Figure 4-3 depicts a view of the overall process of measuring the predictive capability of the model *and* determining whether or not system requirements are met. Comparing the M&S capability to system requirements, which are defined in the PIRT, demands a metric. The metric is used when (1) measuring the computational prediction including the associated uncertainty, (2) comparing the simulation results with uncertainties to the experimental results or field observations to identify the error (referred to herein as *prediction error*), and (3) comparing the prediction error with one or more quantitative requirements. An example of such a metric is the difference between the M&S probability density function and an experimental probability density function.

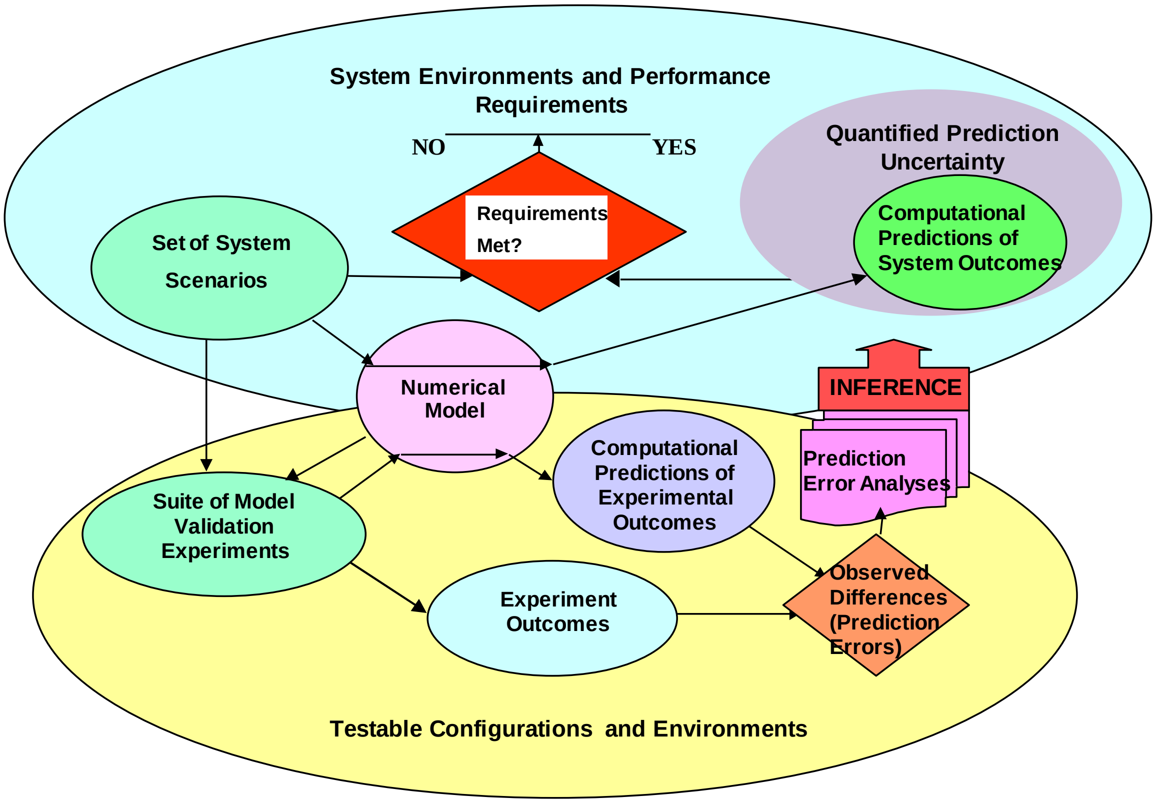


Figure 4-3. System view of process of quantification of predictive capability for complex numerical, i.e., computational, models.

Specification of one or more metrics for these three activities is not a trivial task. Metric specification includes determining the process and algorithm for the metric and a description or example of how the metric will be used in each applicable activity. If more than one metric is defined, the process for using and resolving discrepancies between the results of the metrics must be addressed.

The practices for model validation follow in Section 4.7.1, with a summary about model validation given in Section 4.7.2.

4.7.1 Model Validation Practices

The Sandia V&V planning guidelines (Pilch et al. 2000) distinguish three increasingly complex categories of model validation, along with a proposed model accreditation category. These categories reflect increased complexity of the validation activity from single phenomena through simply coupled phenomena to fully coupled phenomena of the complexity of the intended application. For the NEAMS Waste IPSC, we find it convenient to distinguish qualitatively different validation approaches that lie at opposite ends of the experimental spectrum. The first approach is *phenomenon-centric* validation, and the second approach is *application-centric* validation. Success in application-centric validation requires success in phenomenon-centric validation as a necessary precondition. A quantitative validation methodology to support the solution of both validation approaches is needed (Trucano, Easterling, Dowding, et al. 2001).

A key validation research issue is how to properly integrate these two distinct validation approaches. For example, the two levels of associated validation experiments, along with any intermediate levels, cannot be performed in isolation. Certainly for the physical phenomena underlying the application, phenomenon-centric validation is a critical precursor to application-centric validation. Each level of validation contributes to the assessment of the predictive capability of the model for the intended application. Phenomenon-centric validation must be defined by the ultimate application need for the model and the resulting demands of application-centric validation. This view has been stressed in the Sandia V&V guidelines (Pilch, Trucano, Moya, Froehlich, Hodges, and Peercy 2000). The application-centric validation process and needed metrics focus on characterizing how accurately the model predicts the complicated and coupled phenomena representative of a driving application. Effective links between phenomenon-centric and application-centric validation are important for the ultimate success of application-centric validation. Traceability between phenomenon-centric and application-centric validation are some of the validation process details that will need to be developed and refined for NEAMS.

A. Perform Phenomenon-Centric Validation

The goal of phenomenon-centric validation is to address the degree to which a model adequately represents a single physical phenomenon for the application of interest. The phenomenon itself may be well characterized experimentally or it may not. The key point is that validation experiments supporting phenomenon-centric validation are designed to isolate that particular phenomenon. Such experiments are referred to as separate-effect tests (SETs) and discussed in detail in the recent NEAMS document by Nelson, Stewart, Unal, and Williams (2010). It is important to ensure consistency between the experiment and the model so that the experiment satisfies the basic assumptions and application conditions of the model. Generally, such SETs are also designed to involve relatively simple geometries and materials. The validation process and needed metrics in this case focus on determining how accurately a model predicts the isolated phenomenon as represented by the chosen experiments and their resulting data.

B. Perform Application-Centric Validation

Application-centric validation measures the accuracy with which a model represents an intended realistic application. The applications of interest to Sandia typically have several phenomena of interest that are coupled to a greater or lesser degree. These validation experiments, referred to as integral-effects tests (IETs), typically must exercise multiphenomena, with more complex geometries and materials than the experiments that support phenomenon-centric validation. For IET examples, see Nelson, Stewart, Unal, and Williams (2010). Because IETs are harder and more expensive than SETs, we expect fewer useful data to be available for application-centric validation. Further, the data generated from IETs will be more complex than the data generated from SETs in phenomenon-centric validation (possibly exhibiting complex space-time correlations that may not be present in simpler validation problems). Similarly, the model simulation results required for comparison with IETs are also typically far more complex than those required for performing SETs in phenomenon-centric validation. Application-centric validation is a far more daunting task than phenomenon-centric validation, which is not to claim that phenomenon-centric validation is easy. Application-centric validation properly includes phenomenon-centric validation as one of its tasks.

C. Develop Validation Test Program

A validation test program is necessary for the comparison of computational predictions to system requirements. A validation test program consists of a suite of experiments and computational simulations for those experiments. The test program is constrained and shaped by the scenarios, the capabilities of the model, and the experimental capabilities. Traceability of validation tests to requirements in the PIRT and of validation tests to experiments is necessary to ensure adequate coverage of the validation tests. Some key considerations for the validation test program are alignment of parameters and prediction error (Trucano, Easterling, Dowding, et al. 2001).

Alignment of Parameters: For each experiment in the program, simulations mirroring the experiment must be performed. In other words, when an experiment is being done for the purpose of validating the code, the code must be run to try and match the simulation results with the experimental results. Major compatibility issues may need to be resolved in accomplishing this testing. We will refer to such issues as *alignment*, as in ―proper alignment of simulation results and validation experiments.” Proper alignment between experimental results and simulation results is required to meaningfully compare and analyze the differences between the two types of results.

Consider the problem where a prediction generated by a simulation is represented by Equation (4-1), where $M(x,\phi)$ represents the numerical model of the phenomenon of interest, x is a vector of model input variables, ϕ is a vector of numerical algorithm parameters, and $y*$ is the simulation results. Note that this equation holds for simulation results that have gone through solution verification.

$$y^*(x) = M(x,\phi). \tag{4-1}$$

The numerical model's input vector x describes a representation of the physical system and the environment to which that entity is subjected for purposes of both simulation and experiment. The input vector x enforces alignment of the numerical model with the experiment. Thus, vector x will include parameters for physical dimensions, initial and boundary conditions, environmental conditions, and material parameters. There would be an ―alignment" problem between the experimental conditions and the model input parameters if an experimental specimen was iron but the parameters for copper were input to the material model.

The numerical model's parameter vector ϕ contains all parameters that are necessary for performing simulations that do not influence alignment of the simulation with the experiment. In short, ϕ contains all other parameters used to define numerical algorithms and specifications, such as grid definitions and numerical algorithm specifications, that are not placed in vector x. Thus, vector ϕ may contain nonphysical material parameters to govern the underlying numerical algorithm of a model. An example of a member of ϕ is a molecular relaxation time, a parameter that has no experimental analog. Similar parameters, such as radiant heating sources, may also be involved in complex boundary conditions. Care must be taken when selecting values for the experiment-independent ϕ parameters because if these values are not chosen wisely, they can affect how the simulation runs.

The essential point of the discussion immediately above is that a validation experiment is defined in terms of the input vector x, not the parameter vector ϕ.

Prediction Error: Prediction error is the difference between the experimental results and the simulation results. Consider an experiment conducted at a specified x, having outcome $y(x)$. Recall that all the parameters in the vector x are related to the experiment.

The *prediction error* of the model at x is defined as

$$e_x = y(x) - y^*(x). \tag{4-2}$$

Note that e_x contains all of the bias and uncertainty associated with both the experiment y and the simulation results y^*. Also note that if solution verification was performed on the simulation results prior to validation, there should be no errors in y^* due to numerical algorithm parameters in vector ϕ. Evaluating model predictive capability requires a collection of experiments and simulations $\{x_i, y(x_i), y^*(x_i) : i = 1,\ldots,n\}$ and evaluating e_x for each member of the collection.

A key assumption underlying the present discussion is that there is random variation in e_x. To test the viability of this assumption that e_x can be modeled as a random variable or a random field, it is recommended that an ensemble of tests (experiments) and simulations be run and that statistical methods be used to analyze whether the error is random or systemic. Some systemic errors could be addressed by calibration, whereas others could require more investigation into where the model is inadequate.

By whatever modeling method is considered or adopted, the general objectives for analyzing the *x* data are the following: First, estimate the probability distribution of e_x at the values of *x* for which simulations and experiments are conducted. Second, estimate the probability distribution of e_x at the values of *x* pertaining to physical quantities that have not, cannot, or will not be subjected to physical testing. This estimated distribution can be applied to estimate e_x for the relevant application.

D. Analyze Validation Results

In this model validation practice, the prediction error determined by comparing experimental results with simulation results is now compared with one or more quantitative requirements, and an assessment is made regarding the pass/fail status of the applicable requirements. There are two specific concerns about model performance that should be addressed. The first concern is that there are ranges or values of *x* where the prediction error is not zero or even approximately zero, indicating possible biases in the model parameters. Sometimes these biases are addressed by calibrating the model parameters. A second concern is that the model is not adequate. One common way that the model may not be adequate is that the model is missing a necessary parameter. Another way is that the equations do not incorporate a parameter properly. Models that lack predictive capability yield high variability in e_x, the ensemble (or collection) of experimental and simulation results, even after compensation has been made for measurement errors associated with the physical experiments.

Analyzing the information available through model validation involves trying to assess both the information generated through completed experimentation and the information that might be gained through further testing and computational analysis. This evaluation is best accomplished after model deficiencies discussed in the previous paragraph have been addressed so that reasonable models for e_x are available. Comparison of this information to requirements can also help establish the needs and directions for further experimentation and computation.

4.7.2 Summary

In summary, model validation objectives include answering the following questions: Is the model adequate for the application?, What is the predictive capability of the code?, How well is that capability understood?

Confidence in computational predictions comes predominantly from comparisons of computations with field observations and experimental results. Model validation results are used to identify gaps in conceptual and mathematical models. These gaps may be addressed through calibration of existing M&S capabilities, or the gaps may require M&S capabilities with more accurate conceptual or mathematical models.

Model validation experiments range from single-physics, tightly controlled laboratory-scale experiments for a single phenomenon to a range of combined or coupled physical tests to complex and expensive system-level multiphysics tests. At each level of complexity, the intent for model validation is to provide quantitative information about the predictive capability of the model.

4.8 V&V and UQ Assessment

Each V&V and UQ practice presented in Sections 4.1 through 4.7 identifies a number of activities that must be carried out at a level of rigor commensurate with the required level of confidence in the M&S capability. The level of rigor defines the outcome to be achieved by each activity and the corresponding evidence to be produced. An assessment of this evidence is performed to confirm that a given activity was performed at a specified level of rigor and to assess an overall level of level of rigor for the V&V or UQ practice.

An important consideration is who will perform these assessments. When the assessment criteria are subjective, the assessor must have sufficient expertise in the M&S domain and independence from developers of the M&S capabilities being assessed. When the assessment criteria are completely objective, i.e., measurable, independence is less of a concern. Objective criteria will be defined as much as possible; however, it is expected that independent assessors will be needed to provide their expert opinions.

The starting point for defining assessment criteria is the PCMM (Predictive Capability Maturity Model), as described in Oberkampf, Pilch, and Trucano (2007). This model identifies four qualitative levels of M&S capability *maturity* for (1) model representation and geometric fidelity, (2) physics and material model fidelity, (3) code verification, (4) solution verification, (5) model validation, and (6) UQ and sensitivity analysis. It is planned that the NEAMS VU program element will support development and implementation of specific assessment criteria and metrics.

Practices for V&V and UQ assessment are a work in progress and thus not specified below. Instead, we present examples of various assessment criteria in Section 4.8.1 and then address the topics of metrics (Section 4.8.2) and assessment gaps (Section 4.8.3).

4.8.1 Example Assessment Criteria

Tables 4-1 through 4-5 illustrate example assessment criteria for V&V and UQ practices, activities, and expected outcomes for each of the levels of rigor. These tables are work in progress. A method of aggregating scores for the various levels of rigor in each table will be developed during implementation of this V&V plan. When completed and distributed in their final form, these five tables should be created and maintained for each of the many different codes, aggregated codes or models, numerical solutions, model validation efforts, sensitivity study, and UQ efforts in the project.

Table 4-1. Code Verification Results versus Level of Rigor

Level	Mesh Refinement Tests	Other Tests	Regression Test Suite (RTS)	Coverage Tables	Review and Documentation
0	None	A few tests in place	None	None	Little or none
1	All F/Cs at least convergent	A collection of unit and functional tests in place	In place	In progress	Theory and user manuals independently reviewed
2	All F/Cs have at least positive observed order-of-accuracy	Benchmark and nearby problems added to test suite	Tests in RTS tagged to identify whether or not they involve mesh refinement	Coverage table based exclusively on mesh refinement tests exists	Verification test suite document reviewed
3	Observed order of accuracy matches expected order of accuracy for all F/Cs	All tests in verification test suite document added to test suite and run successfully	All mesh refinement tests in coverage table also in RTS	Coverage tables complete with respect to F/Cs and updated periodically	Every test in RTS included in test setup and results document

Basically, if the required level of rigor for an M&S capability is level 3, then the relevant activities and their outcomes are easily identified in the tables. Some of the outcomes associated with lower levels of rigor may also be required at level 3 (for example, in code verification, the theory manual is needed at levels 1–3).

Given a target level of rigor, as identified by the row, the level of V&V and UQ maturity can be assessed in terms of the activities and outcomes employed by the project at any given time. A periodic comparison of current project activities and outcomes to the planned list of activities and required outcomes will identify shortcomings and generate discussion on how to address them.

The implementation of V&V and UQ practices will vary somewhat, depending upon the M&S scale of the individual code or model. For example, mesh refinement studies are not needed for codes that produce exact numerical solutions, as opposed to mesh-dependent approximate numerical solutions.

Table 4-2. Solution Verification Results versus Levels of Rigor

Level	I/O Verification	Num. Model Sensitivity	Mesh Refinement	Error Est.	Reviews	Documentation
0	Casual	Little or none	None	None	None	None
1	By Analyst	Informal investigation of sensitivity of solution to some numerical parameters	In progress or solution shown to be non-asymptotic	Implementation of error estimation methods in progress	Casual	Solution archived and retrievable
2	By Peers	Systematic investigation of sensitivity of solutions to all numerical parameters	Solution and/or SRQs might be asymptotic	Error estimates on some SRQs	By peers	A SAND or other report
3	Independently Reproduced Solution	Systematic investigation of sensitivity of all SRQs to all numerical parameters	Solution definitely asymptotic as are all SRQs	Error estimates on all SRQs, plus error bars	Independent review panel	An archive journal publication

Table 4-3. Model Validation Activities versus Levels of Rigor

Level	Model Validation Activity
0	* Judgment only * Few comparisons between numerical and experimental (or other) results
1	* Large or unknown experimental uncertainties
2	* Quantitative comparisons of predictive accuracy for some system response quantities in IETs (integral-effects tests) and SETs (separate-effect tests) Experimental uncertainties well-characterized for SETs * Peer review
3	* Quantitative comparisons of predictive accuracy for all SRQs in IETs and SETs * Experimental uncertainties well-characterized for all SETs and IETs * Independent peer review

Table 4-4. UQ/Sensitivity Analysis Activities versus Levels of Rigor

Level	UQ/Sensitivity Analysis Activity
0	* Judgment only * Deterministic calculations only * Uncertainties and sensitivities not addressed
1	* Aleatory and epistemic uncertainties propagated forward but without distinction * Informal sensitivity studies conducted * Many strong assumptions about UQ/sensitivity analysis made
2	* Aleatory uncertainties in the SRQs segregated, propagated, and identified * Quantitative sensitivity analyses conducted for most parameters * Numerical propagation errors estimated and their effect known * Some strong assumptions made * Peer review
3	* Aleatory and epistemic uncertainties comprehensively treated and properly interpreted * Comprehensive quantitative sensitivity analyses conducted for parameters and models * Numerical propagation errors demonstrated to be small * No significant UQ/sensitivity analysis assumptions made * Independent peer review

Table 4-5. Example Subgrid-Scale M&S Capability Assessment Metrics

	Level 1 Moderate Impact	**Level 3** Very High Impact	
Practice	The physical/chemical information from this subcontinuum model is required for continuum modeling, and there are other sources of information that confirm the results,	The physical/chemical information from this subcontinuum model is required for continuum modeling, and continuum models will rely exclusively on the results from this subcontinuum model.	**Metrics (Evaluation Criteria)**
CM (Configuration Management)	Tested and released versions of code are tested and found to be retrievable by *team*.	Simulations and releases are tested and found to be retrievable and *repeatable* by *independent party*.	Simulations are repeatable. Releases are repeatable. Versions of code and results of simulations and tests are retrievable.
Representation and Geometric Fidelity	V&V compares and documents results with other sources. Uncertainty quantified for representation and geometric fidelity.	Geometry and representation issues are determined to be appropriate by *independent review*. Uncertainty quantified for representation and geometric fidelity.	Geometry and representation issues are independently identified and prioritized. Fidelity tests are executed. Boundary condition tests are executed. Geometry and representation issues are determined to be appropriate by independent review. Uncertainty quantified for representation and geometric fidelity.
Model Validation	Pass all validation tests related to high-priority PIRT items. or Validation tests and results are determined to be appropriate by *team review*.	Validation approach, tests, results, and PIRT coverage are determined to be appropriate by *independent review*. Uncertainty quantified for experimentation.	Pass all validation tests related to high-priority PIRT items. Experimental uncertainties quantified. Independent review of validation approach and results.

4.8.2 Metrics

Assessed levels of rigor for V&V and UQ practices define a collection of metrics. It may prove useful to define aggregated or ―rolled up” metrics from these levels-of-rigor metrics. For example, it may be useful to define an overall level of rigor across all the solution verification activities for a particular numerical solution. One approach would be to define the aggregate solution verification metric to be the minimum current level of rigor across all the activities. Aggregate metrics could span multiple V&V and UQ practices.

The set of V&V and UQ assessment metrics is expected to grow and change as the V&V and UQ practices are implemented and improved over the lifetime of the NEAMS Waste IPSC. The set of metrics provided previously in Tables 4-1 through 4.5 represents the starting point of this work.

The EVIM system is required to keep track of all the VU-assessment tables, associated evidence, assessment metrics, and linkages. The EVIM system must have sufficient flexibility to periodically extend the kinds of evidence managed and the metrics associated with that evidence. It is expected that many assessment metrics early in the NEAMS Waste IPSC program element will be qualitative, and as the state-of-the-art in V&V and UQ progresses, new quantitative metrics will be defined.

4.8.3 Assessment Gaps

Gaps between assessed and required levels of rigor will be used to plan V&V and UQ activities required to close identified gaps. As these activities progress, new evidence will be generated, and the V&V and UQ assessment will be revised to incorporate the new evidence. M&S capabilities will be revised, extended, and ported to new computational environments. Whenever an M&S capability is changed, it must be reassessed to update the corresponding V&V and UQ evidence or to confirm that this evidence has not changed.

4.9 V&V and UQ Planning

Analyses performed with the NEAMS Waste IPSC will require confidence in M&S capabilities commensurate to the risks associated with decisions that the analyses will support. Confidence requirements are quantified in terms of the required level of rigor for the M&S capabilities. V&V and UQ plans are developed and carried out to ensure that the M&S capabilities meet these confidence requirements.

4.9.1 V&V and UQ Planning Practices

A. Analyze Confidence Requirements and Gaps

Planned applications of the NEAMS Waste IPSC are analyzed to determine which M&S capabilities will be used and the required level of confidence commensurate with the intended use of the results. These requirements are quantified in terms of the V&V and UQ assessment criteria and level of rigor. A gap analysis can be used to compare the required versus actual V&V and UQ assessment values for required M&S capabilities. In this case, there would be a gap, say, if the required level of rigor for a particular capability was 3, but the assessed level of rigor for that capability was 2.

Gaps between required and actual V&V and UQ assessment values are closed by either carrying out V&V and UQ activities to meet the confidence requirements or relaxing the confidence requirements. The required level of rigor for V&V and UQ can only be relaxed if the end users accept a correspondingly greater level of risk in intended use of their M&S results.

B. Develop V&V and UQ Plan

Required V&V and UQ activities are identified to close gaps. These required activities should be incorporated into the project or program plans and define the V&V and UQ plan for the end users' analyses. For example, participants in the NEAMS Waste IPSC program element need to be mindful of the level of rigor for the various V&V and UQ practices, and based on the intended use of the codes, clearly document the effort and resources required to achieve the necessary level of confidence in their M&S capabilities. The required V&V and UQ activities could include additional data acquisition, new small-scale phenomena-centric validation experiments, major application-centric experimental programs, and/or the development of entirely new M&S capabilities.

5 Management of Evidence Information

Throughout this document, we have stressed the need for and development of the EVIM system as a central electronic repository for the NEAMS Waste IPSC. Generating the evidence information from V&V and UQ activities, as discussed in Section 4, is important but not sufficient. We must also be able to maintain (or manage) that information and be able to immediately access it when necessary. In Section 5, we present a goal-based definition of the EVIM system's scope, introduce preliminary architectural designs that portray the most likely interacting system elements, and highlight examples of other similar systems that can inform our construction of the EVIM system. Current plans are to acquire and configure the EVIM system to meet these goals.

5.1 EVIM System Scope

NEAMS Waste IPSC V&V and UQ activities will produce evidence to obtain confidence in M&S capabilities. This evidence must be formally maintained, traceable through the scales of M&S and coupling of M&S capabilities, and support effective communication to the consumers of this evidence. A significant goal of the NEAMS Waste IPSC is to manage V&V and UQ evidence resulting from V&V and UQ activities. An EVIM system will be deployed to provide timely, searchable, and minable access to this evidence as well as traceability to evidence integrated from multiple sources and supporting multiple formats of information.

Table 5-1 summarizes the scope definition for the EVIM system. The seven goals identified in this table will be used to guide deployment of the EVIM system. The table also provides a concept of how the system might be used.

Table 5-1. Summary of the EVIM System Scope

Scope Item	Value
Business Case	Stakeholders who make decisions based on the results of simulations want to know why they should have confidence in those results, but manually searching and analyzing the extensive evidence that is collected can be ineffective and human-resource intensive.
Need	Provide timely, searchable, and minable access to M&S V&V evidence and traceability integrated from multiple different sources and supporting multiple formats of information.
Goal I	Develop an electronic repository of M&S V&V evidence and traceability. **<u>Objectives</u>** 1. Permanently record the following information: a. M&S V&V metrics and results b. UQ results c. simulation inputs and outputs d. simulation results e. assessment metrics and results f. analysis and review results g. analysis and review sign-offs h. versions of the M&S code and data i. code requirement or capability specifications 2. Permanently record traceability a. of M&S V&V evidence to simulation runs, b. of M&S V&V evidence to versions of code and data sets, c. of M&S V&V evidence to requirements, d. of model and data file versions to simulation runs, e. from requirements to code and data sets, f. between scales of M&S and composition hierarchies, and g. of simulation runs to hardware specifications such as operating system and platform. 3. Support multiple diverse formats of information, such as images, video, text, Word, and Portable Document Formats (PDFs). 4. Integrate data from different sources and in different formats.

Scope Item	Value
Goal II	Allow for timely capture and update of information described in **Goal I**. **Objectives** 1. Allow for the capture of M&S V&V evidence as it is generated. 2. Define the format needed for the M&S V&V evidence to be captured in the electronic format. 3. Provide an electronic means to enter the evidence. 4. Make the evidence capture straightforward. 5. Keep the data up-to-date. **6.** Prevent the buildup of a large backlog of evidence data that need to be stored.
Goal III	Enable derived V&V metrics for M&S capabilities coupled by composition hierarchies and interscale relationships. **Objectives** 1. Support extensible specifications for V&V metrics that are derived from V&V information present in component members of the hierarchy or higher-resolution members of the interscale coupling. 2. Provide automated roll-up or update of derived V&V metrics when components' V&V information is inserted or modified. 3. Update the assessment status when components' information is inserted or modified.

Scope Item	Value
Goal IV	Establish and maintain data quality. **Objectives** 1. Assess the quality of captured and updated data and information. Examples of data quality attributes are accuracy, comprehensiveness, timeliness, consistency, relevancy, review history, and validity of sources. 2. Identify duplicate data. 3. Enforce data-naming standards. 4. Ensure data comply with data set consistency standards for data types, measuring units, accuracy, and precision. Consistency in data sets is needed to support meaningful search results and metrics. Ensuring compliance with data set consistency standards occurs either through clean up as part of the acquisition process or by enforcing consistency standards on the data sets before accepting the data.
Goal V	Allow for searching of the M&S V&V evidence and traceability. **Objectives** 1. Provide a user interface to the electronic repository that allows users to query and extract the M&S V&V evidence and traceability information. 2. Allow the user to find M&S V&V evidence based on the following: a. requirement or capability b. code c. data set d. simulation run 3. Provide a reporting capability that generates a. traceability reports and b. gap analysis reports.

Scope Item	Value
Goal VI	Provide an estimated level of confidence assessment, including uncertainty, for a prediction or prediction capability. **<u>Objectives</u>** 1. Provide an expert-systems capability to compile evidence into an estimated level-of-confidence assessment (a long-term research and development [R&D] objective). 2. Allow for data mining of the M&S V&V evidence and traceability. 3. Allow the results to be displayed in either summary or detail levels.
Goal VII	Establish an operational production infrastructure. **<u>Objectives</u>** 1. Define and set up security controls (user authentication, file, and data access control). 2. Provide for system maintenance (regular backups, software and hardware patches, support personnel). 3. Provide for system and technology upgrades. 4. Log changes and historical activity for the life of the system. 5. Develop training. 6. Develop a help desk.

Scope Item	Value
Operational Concept	The EVIM system might be used routinely in the following ways: • A creator of M&S V&V evidence links code, data, and documentations from databases and CM (configuration management) systems to specific models and phenomena via the PIRT (Phenomena Identification Ranking Table). • A creator of M&S V&V evidence specifies the M&S capability links and derived metrics for composition coupling and interscale coupling. • An analyst develops an analysis workflow with specifications of M&S capabilities to be used, workflow couplings to be performed, and intended level of rigor or confidence. • An analyst queries the V&V evidence related to the analysis workflow. • An analyst runs a simulation and records the appropriate V&V evidence. • A user obtains evidence reports by selecting a capability. • A user obtains evidence reports by selecting a code and data set. • A user generates a traceability report that shows the relationships of the evidence. • A user generates a gap analysis report that identifies gaps in the V&V requirements.
Assumptions	The data and information coming into the EVIM system are from many diverse sources and companies. The EVIM system will be developed in stages.
Constraints	[TBD]

5.2 Architectural Design for the EVIM System

5.2.1 Interfaces to the EVIM System

The architectural design of the EVIM system is under development. Figure 5-1 and Figure 5-2 show high-level conceptualizations of the EVIM system's interfaces and workflow. Figure 5-2 is a context diagram of the EVIM system that shows the interfaces to other systems or external entities. Sections 5.2.1.1 through 5.2.1.5 describe in general the interfaces between the EVIM system and the external entities shown in Figure 5-2. The details about the interfaces, such as connections, protocols, data specifications, data transfer frequency, and data ingestion methods, will be determined as part of the design and development of the system, and consequently, are beyond the scope of this document. As the EVIM system evolves, the interface details should be specified in other documentation, including interface control documents, design documents, process definition documents, and data description documents.

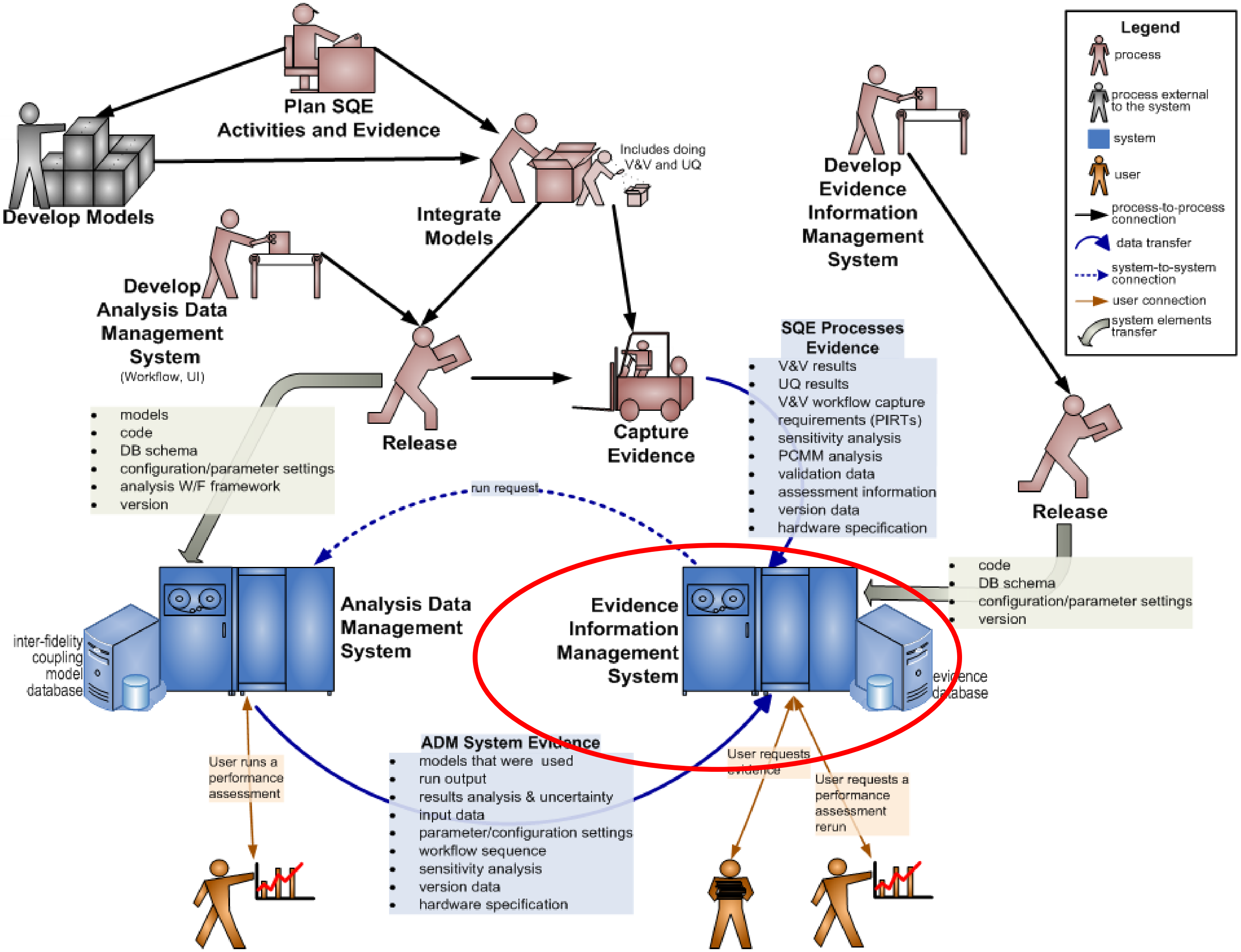


Figure 5-1. NEAMS Waste IPSC conceptual workflow.

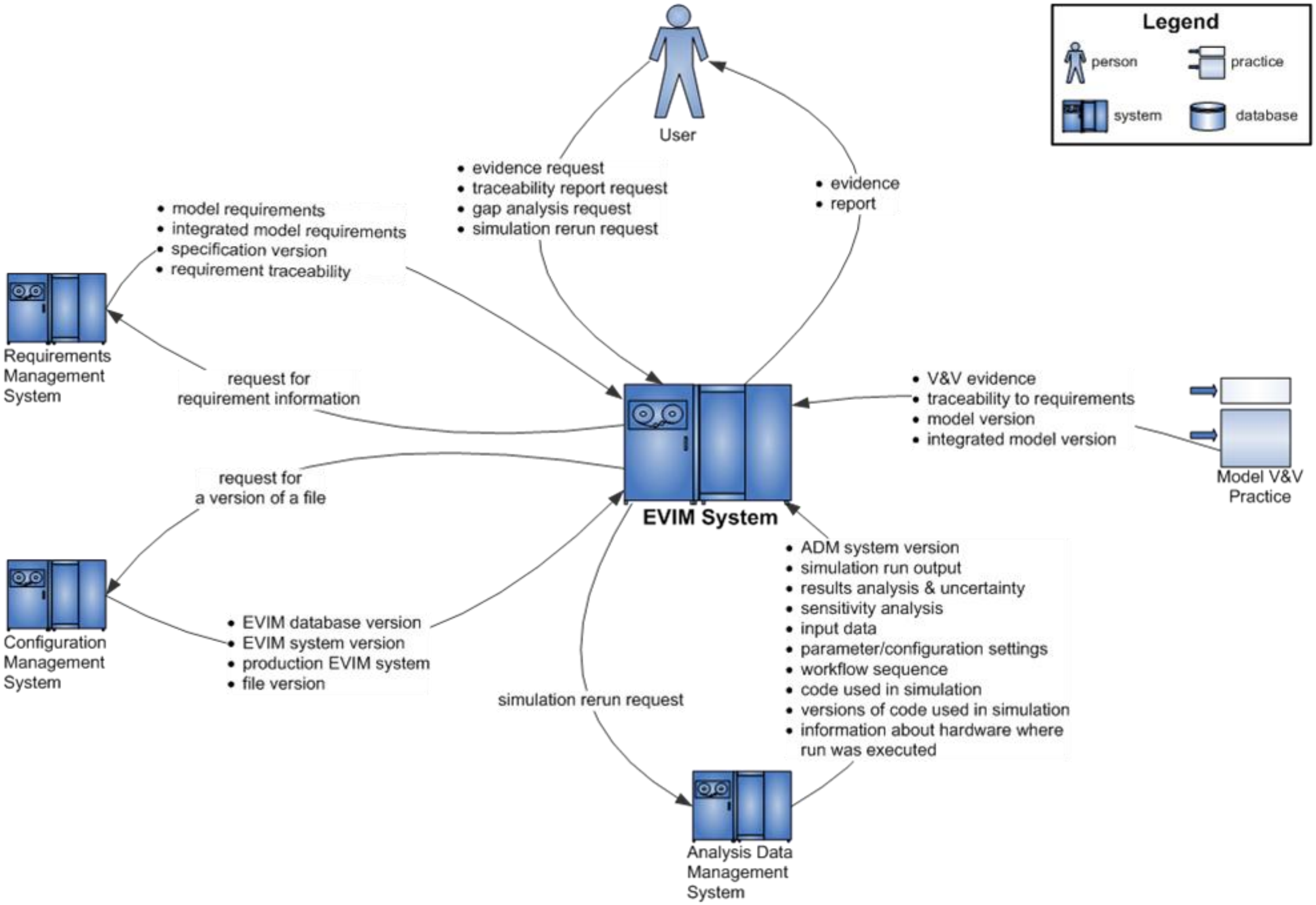


Figure 5-2. EVIM system context diagram.

5.2.1.1 Users

Users of the EVIM system will be querying the system to retrieve V&V evidence and traceability about the M&S capabilities. The querying and display of results will be accessed through a user interface that will support different querying criteria and forms of displaying the results (see Table 5-1: Goal V). The query results may be displayed within the user interface—such as in a graph, diagram, or table—or may be displayed in the form of an online or hardcopy report.

One aspect of traceability in the EVIM system is mapping M&S V&V evidence to analysis workflows (see Table 5-1: Goal I, Objective 2). For example, while viewing the M&S evidence and traceability, a user may wish to rerun an analysis workflow. Accordingly, the EVIM system must contain sufficient information to rerun the analysis. This rerun process would basically involve the following steps: (a) retrieving and building the required version of the M&S codes via the CM (configuration management) system, (b) retrieving the workflow and code inputs, and (c) rerunning the analysis workflow through the analysis data management (ADM) system.

The reliability and usability of rerunning analysis workflows could be improved through automation of this rerunning process. In this approach, as a user views the information through the EVIM user interface, the EVIM system would initiate the rerun process, access the results, and present those results to the user.

5.2.1.2 Model V&V Practice

The EVIM system provides the repository for M&S V&V evidence and traceability produced from applying the V&V practices identified in this document. The type of information that will be input to the system is the M&S V&V evidence and traceability as described in Table 5-1: Goal II.

5.2.1.3 Configuration Management System

The CM system is the set of tools and processes defined and deployed to control and administer the configuration of the EVIM system through its development and maintenance lifecycles. The CM system includes tools and processes for maintaining version control of the hardware, software, database, and configuration files; building the software; and releasing and distributing the system. In general, the interface between the EVIM system and the CM system is one way—the CM system provides information to the EVIM system. The key elements provided by the CM system to the EVIM system will be (1) a released version of the EVIM system itself with its associated database and (2) the version information that accompanies the release.

There may be cases where a user of the EVIM system may want to view specific versions of files that are kept under CM control in the CM system. In those cases, the EVIM system will send a request to the CM system for versions of the files, and the CM system will send the file versions back to the EVIM system. A specific example of this interface would be if a user of the EVIM system is looking at the evidence and wishes to see the modeling code that generated the results. As the CM system will be the repository for the software versions, the EVIM system must be capable of requesting the code from the CM system so that the user can view the code.

Depending on the design of the EVIM system, it is possible that the CM system could be tightly coupled with the EVIM system. In such a case, the CM system would be viewed as part of the EVIM system rather than as an external entity.

5.2.1.4 Analysis Data Management System

Simulations and analysis are run under the auspices of the ADM system. Information about simulation runs is part of the EVIM system as described in Table 5-1: Goal I. Thus, after a simulation is run, the ADM system will send the simulation run information to the EVIM system. The simulation run information includes the setup for the simulation run, such as input data, parameters, configuration settings, and workflow sequence, and also the results from the run, such as the run output, results analysis and uncertainty, and sensitivity analysis. The simulation run information also includes configuration information, such as the codes and the version of the codes used for the simulation run, the version of the ADM system that was used, and the hardware specification denoting where the simulation was run. All of the simulation run information supports traceability and reproducibility of the M&S V&V evidence.

As described in Section 5.2.1.1, a user may request that a simulation be rerun. That request will be processed through the user interface to the EVIM system. The input that is required when a user interacts directly with the ADM system to request a simulation run will also be required when the request comes through the EVIM system. When the user enters the rerun request, the EVIM system will then send the request to the ADM system. After the simulation is rerun, the simulation run information will be sent to the EVIM system as would occur in any other simulation run.

5.2.1.5 Requirements Management System

The RM system is the set of tools and processes for identifying, documenting, tracing, prioritizing, agreeing to, modifying, and communicating requirements (and their modifications) to relevant stakeholders.

In general, the interface between the RM system and the EVIM system is one way. The RM system provides requirements and requirement traceability to the EVIM system. Both the RM system and the EVIM system keep requirement and requirement-traceability information, but the EVIM system has a broader scope. Consequently, the RM system may keep copious amounts of information about requirements, but it will supply to the EVIM system only the requirement and requirement-traceability information that is needed for tracing the M&S V&V evidence to requirements and for querying about traceability of requirements. The RM system also informs the EVIM system of the version of the requirements specification that is represented by the requirements.

There may be cases where the EVIM user needs more-detailed requirements information than is kept in the EVIM system. In those cases, the EVIM system will send a request to the RM system, and the RM system will send the requested information back to the EVIM system to display to the user.

Depending on the design of the EVIM and RM systems, it is possible that the RM system could be tightly coupled with the EVIM system. In such a configuration, the RM system would be viewed as part of the EVIM system rather than as an external entity.

5.2.2 Database and Software Components of the EVIM System

5.2.2.1 Evidence Database

The evidence database addresses most of the objectives in Goal I. The evidence database will contain data sets for specific V&V evidence. The design of the database will identify the data sets, data items, attributes, and values of those attributes. The data sets and data items below are examples of the types of evidence identified by Goal I that must be maintained:

- V&V metrics and results
- UQ results
- simulation inputs and outputs
- assessment metrics and results
- analysis and review results
- analysis and review sign-offs

Figure 5-3 lists the type of content that might be included in each of these data sets.

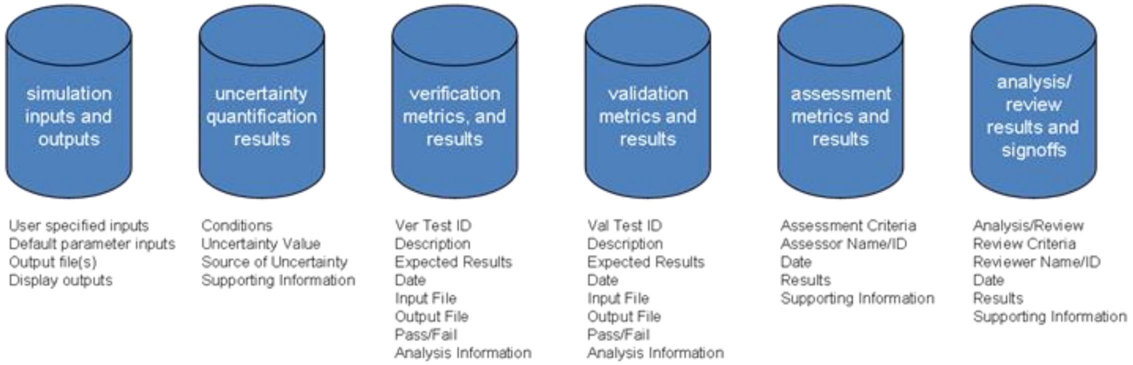


Figure 5-3. Evidence database data sets (example).

The evidence database interfaces with other NEAMS Waste IPSC systems. An early representation of these interfaces is discussed in Section 5.2 and shown in Figure 5-2. Each piece of evidence in the evidence database is associated with a specific version of one or more codes and one or more data sets. The PIRT identifies IPSC capabilities, and PIRT items are associated with specific versions of codes and data sets. These interfaces between NEAMS Waste IPSC systems allow evidence to be selected by requirement, capability, code, and/or data set. These linkages represent the interfaces identified by Goal I that need to be maintained:

- traceability to simulation runs and results
- traceability to versions of the M&S code and data
- traceability to requirements

5.2.2.2 Developer Interface Software

The developer interface software addresses most of the objectives in Goal II, Goal III, and Goal IV and begins to address Goal VI. The developer interface software is required for developers to submit evidence into the evidence database. The first step in implementing this software will be to develop and review a detailed design.

5.2.2.3 Evolution of the EVIM System Software

Implementing the first version of the EVIM system software addresses most of the objectives in Goal V. This software will (1) allow users to retrieve evidence from the evidence database, (2) add the interface to the ADM system, and (3) provide additional evidence-database tags or links necessary to support the software. Upgrading the EVIM system software over the long term addresses the objectives in Goal VI. Significant research and design is required before the EVIM system software should be upgraded to address Goal VI.

5.3 Examples from Existing V&V Evidence Systems

In our initial survey of existing EVIM-like systems, we did not find an existing system that would provide a suitable foundation for the NEAMS Waste IPSC EVIM system. Sections 5.3.1 and 5.3.2 describe examples from existing systems that illustrate elements that are part of the scope of the EVIM system. In further developing the EVIM system, we will evaluate these and other examples to analyze EVIM requirements, better understand the EVIM design space, and apply lessons learned.

5.3.1 Yucca Mountain Project Licensing Support Network

During development of the Yucca Mountain project software, the project had an internal CM system, a records processing center, and several procedures that addressed software management and the control of electronic information captured in procedure documents. But a single point of access for documents did not exist.

The Licensing Support Network (LSN) was developed to support the Yucca Mountain project's licensing process (LSN 2010). The LSN provides a *single place to access documents* related to DOE's application for construction authorization for a high-level radioactive waste repository at Yucca Mountain, Nevada. The LSN offers both regular and advanced search capabilities. The advanced search form is shown in **Figure 5-4.**

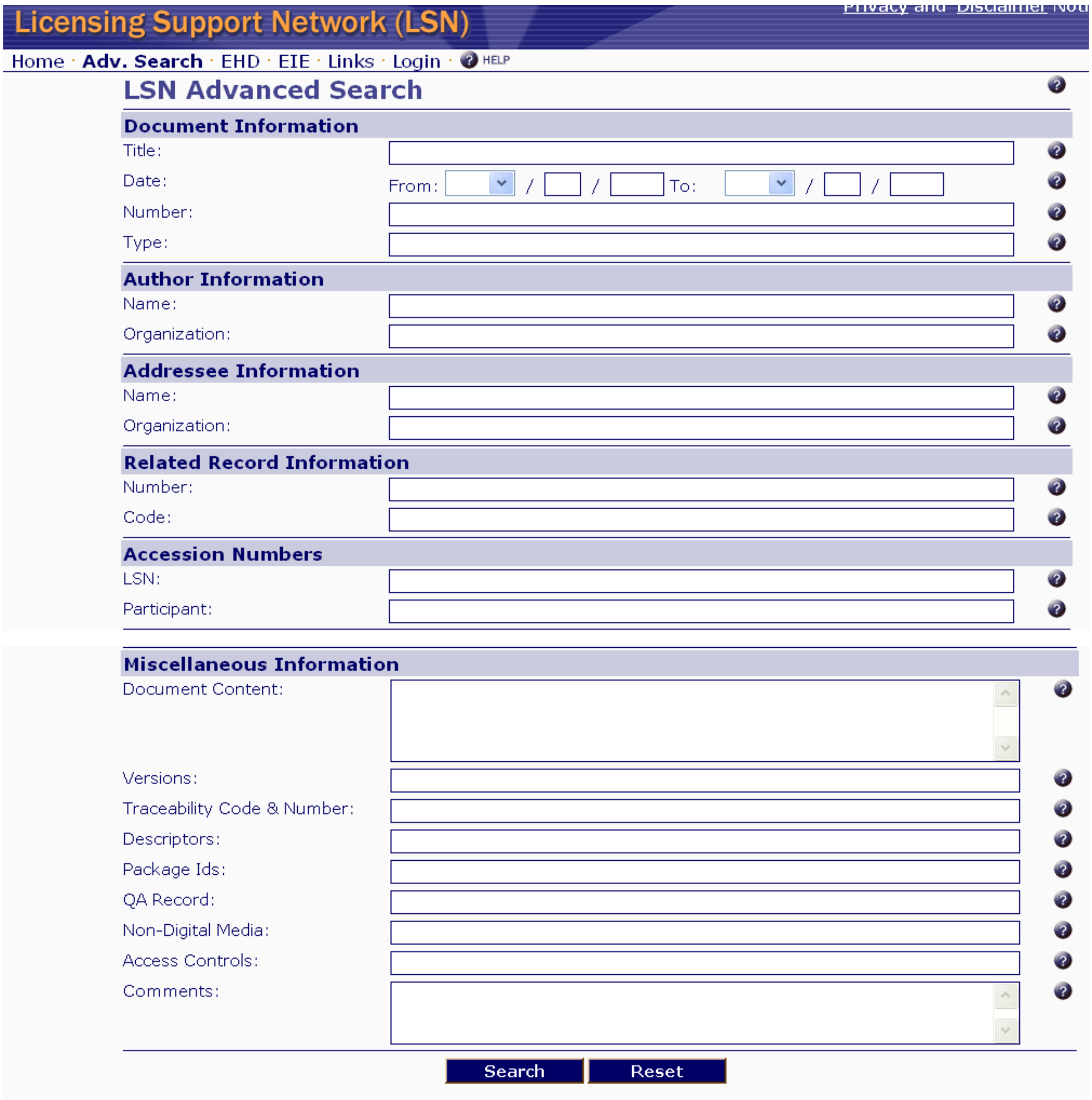
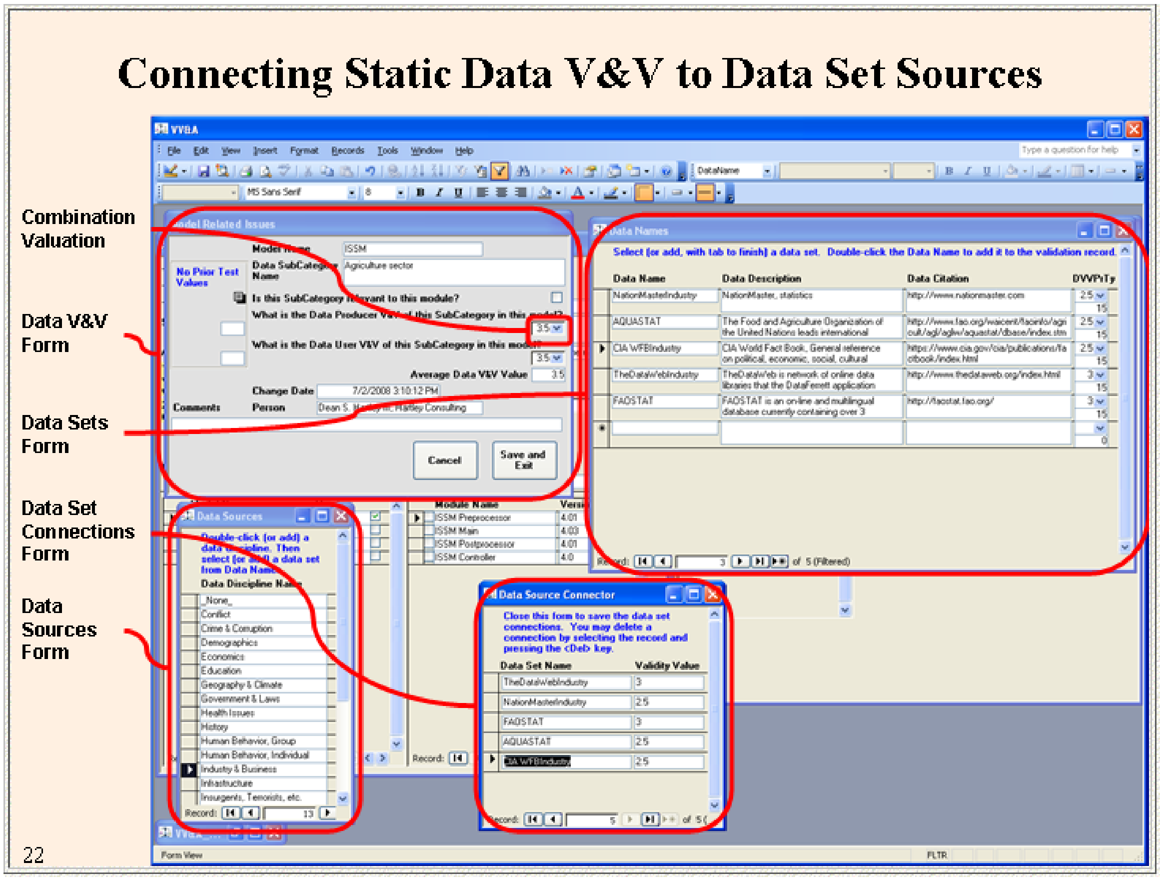


Figure 5-4. YMP LSN search interface.

5.3.2 DIME/PMESII Tool

The currently unreleased Verification, Validation, and Accreditation (VV&A) tool, referred to here as DIME/PMESII, is an example of a tool that presents *linkages of V&V results* to data sets and requirements (Hartley 2009). DIME/PMESII comes from the Department of Defense. DIME stands for Diplomatic, Informational, Military, and Economic; PMESII stands for Political, Military Economic, Information, and Infrastructure. Sandia has a demonstration version of this tool, and a presentation is available (Hartley 2009). Figure 5-5 and Figure 5-6 are screen shots from the presentation.



Connecting Static Data V&V to Data Set Sources

Model Related Issues

Model Name	ISSM
Data SubCategory Name	Agriculture sector
Is this SubCategory relevant to this module?	☐
What is the Data Producer V&V of this SubCategory in this model?	3.5
What is the Data User V&V of this SubCategory in this model?	3.5
Average Data V&V Value	3.5
Change Date	7/2/2008 3:10:12 PM
Person	Dean S. Hartley III, Hartley Consulting
Comments	

No Prior Test Values

Data Names

Select (or add, with tab to finish) a data set. Double-click the Data Name to add it to the validation record.

Data Name	Data Description	Data Citation	DVVPrTy
NationMasterIndustry	NationMaster, statistics	http://www.nationmaster.com	2.5 15
AQUASTAT	The Food and Agriculture Organization of the United Nations leads international	http://www.fao.org/waicent/faoinfo/agricult/agl/aglw/aquastat/dbase/index.stm	2.5 15
CIA WFBIndustry	CIA World Fact Book, General reference on political, economic, social, cultural	https://www.cia.gov/cia/publications/factbook/index.html	2.5 15
TheDataWebIndustry	TheDataWeb is network of online data libraries that the DataFerrett application	http://www.thedataweb.org/index.html	3 15
FAOSTAT	FAOSTAT is an on-line and multilingual database currently containing over 3	http://faostat.fao.org/	3 15
			0

Record: 3 of 5 (Filtered)

Module Name	Version
ISSM Preprocessor	4.01
ISSM Main	4.03
ISSM Postprocessor	4.01
ISSM Controller	4.0

Data Sources

Double-click (or add) a data discipline. Then select (or add) a data set from Data Names.

Data Discipline Name
None
Conflict
Crime & Corruption
Demographics
Economics
Education
Geography & Climate
Government & Laws
Health Issues
History
Human Behavior, Group
Human Behavior, Individual
Industry & Business
Infrastructure
Insurgents, Terrorists, etc.

Record: 13

Data Source Connector

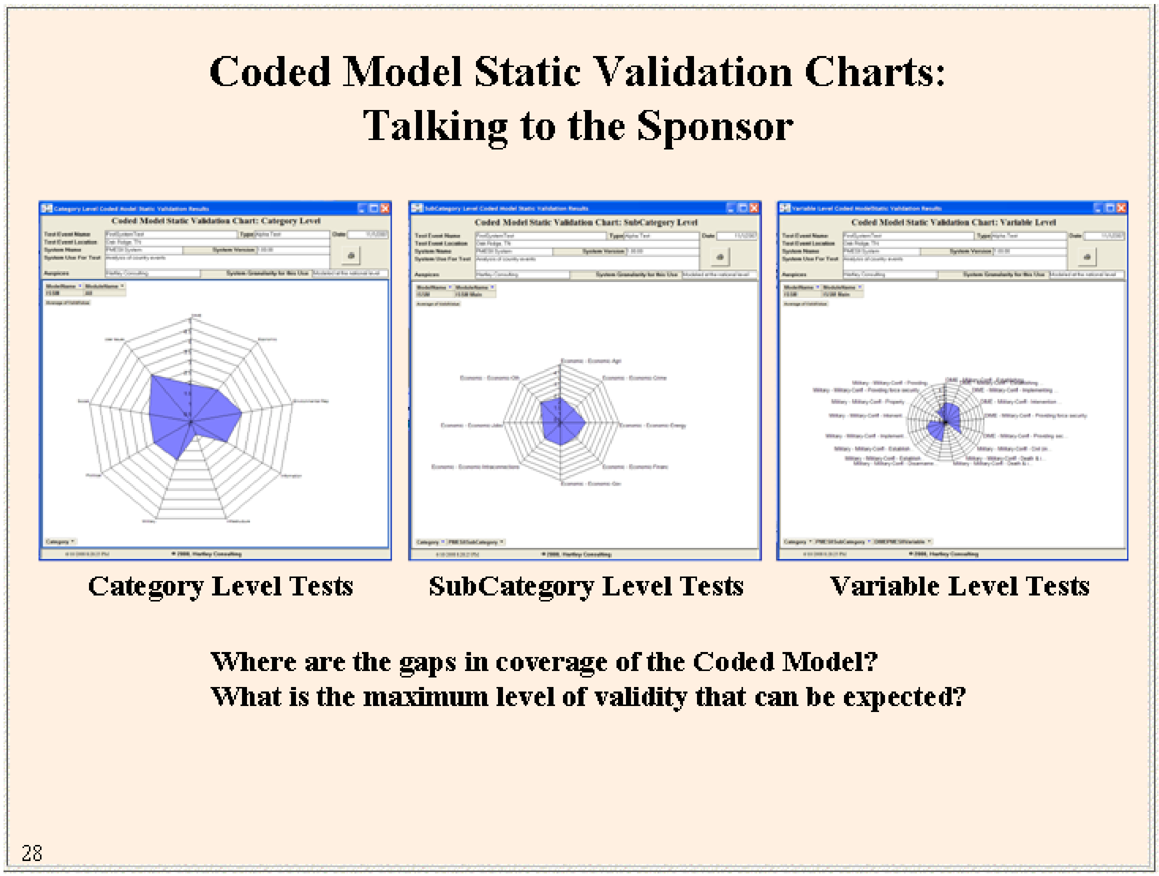
Close this form to save the data set connections. You may delete a connection by selecting the record and pressing the <Del> key.

Data Set Name	Validity Value
TheDataWebIndustry	3
NationMasterIndustry	2.5
FAOSTAT	3
AQUASTAT	2.5
CIA WFBIndustry	2.5

Record: 5 of 5

22

Figure 5-5. Screen shot from DIME/PMESII VV&A tool presentation (1 of 2).



Coded Model Static Validation Charts: Talking to the Sponsor

Category Level Tests

SubCategory Level Tests

Variable Level Tests

Where are the gaps in coverage of the Coded Model?
What is the maximum level of validity that can be expected?

Figure 5-6. Screen shot from DIME/PMESII VV&A tool presentation (2 of 2).

6 Path Forward for Implementation

Section 6 outlines the anticipated, prioritized steps to implement this V&V plan incrementally. It is critical that the practices and EVIM system presented in this plan be implemented early in the NEAMS Waste IPSC program. Early implementation will (1) institutionalize V&V and UQ within the program; and (2) allow early evaluation and improvements to the V&V and UQ assessment metrics, processes, and EVIM system. Early evaluations and improvements are likely to yield greater efficiency and effectiveness as the number of M&S capabilities and end users grows.

Implementation of the EVIM system must be extensible, flexible, and agile to incorporate anticipated revisions to the V&V and UQ assessment metrics and processes as well as changes to the enabling data management systems and processes, e.g., CM and RM.

The following implementation phases are anticipated and subject to change as the implementation of this V&V plan progresses. A schedule cannot be established at this time due to strong uncertainties in funding, resources, and availability of suitable existing processes and tools.

Phase 1
During the first phase, an initial quality environment is deployed with enabling infrastructure and SQE tools and practices. This quality environment will be the foundation for implementation of V&V and UQ processes and the EVIM system. Codes and data will be imported into the quality environment where these processes are to be applied. The V&V and UQ processes and information traceability are dependent upon CM, RM, and ADM systems. These enabling data management systems and associated processes will be included in the quality environment.

Phase 2
The second phase consists of defining an initial set of VV and UQ assessment metrics, a V&V and UQ assessment process, and code and solution verification processes. An initial EVIM system is implemented within the quality environment to support verification and assessment processes and metrics. The verification and assessment processes are applied to challenge-problem codes and analyses to evaluate the codes, quality environment, EVIM system, interfaces with enabling data management systems and processes, V&V and UQ processes, and V&V and UQ assessment metrics.

Phase 3
In the third phase, revisions are made to the set of VV and UQ assessment metrics and the V&V and UQ assessment process, and processes are defined for data acquisition, UQ and validation. In addition, the EVIM system is revised to incorporate any changes to the enabling data management systems, V&V and UQ metrics, and V&V and UQ processes. Also, during phase 3, an initial user interface for the EVIM system is implemented to support V&V and UQ practitioners and other users of the NEAMS Waste IPSC.

Phase 4

The now-complete set of V&V and UQ processes is rigorously applied in the fourth phase to challenge-problem codes and analyses to evaluate the codes, processes, and EVIM system. V&V and UQ practitioners and end-user stakeholders participate in the evaluation to assess the applicability, organization, and accessibility of the V&V and UQ evidence (data, metrics, and traceability).

Phase 5

The V&V and UQ assessment metrics, processes, EVIM system, and user interface is revised as necessary based upon results from the previous evaluation.

References

Babuska, I., and J. T. Oden. (2004). ―Verification and Validation in Computational Engineering and Science: Basic Concepts.” *Computer Methods in Applied Mechanics and Engineering* 193, nos. 36–38: 4057-4066.

Bartlett, R. (2009). ―Integration Strategies for Computational Science & Engineering Software,” Second International Workshop on Software Engineering for Computational Science and Engineering. Available from http://www.cs.sandia.gov/~rabartl/publications.html (accessed on January 5, 2011).

Bechtel SAIC Company (BSC). (2005). *The Development of the Total System Performance Assessment-License Application Features, Events, and Processes*. TDR-WIS-MD-000003 REV 02. Las Vegas, NV: Bechtel SAIC Company.

Beck, K. (2005). *Extreme Programming*. 2nd ed. Addison Wesley Professional.

Bond, R., P. Knupp, and C. Ober. (2004). ―A Manufactured Solution for Verifying CFD Boundary Conditions, Part I.” AIAA paper 2004-2629.

Bond, R., P. Knupp, and C. Ober. (2005). ―A Manufactured Solution for Verifying CFD Boundary Conditions, Part II.” AIAA paper 2005-0088.

Bond, R., P. Knupp, C. Ober, and S. Bova. (2007). ―Manufactured Solution for Computational Fluid Dynamics Boundary Condition,” *AIAA Journal* 45, no. 9: 2224–2226.

Cressie, N.A.C. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons.

DIME/PMESII VV&A Tool. Available from http://home.comcast.net/~dshartley3/VVATool/VVA.htm (accessed on December 1, 2010).

Duvall, P. M., S. Matyas, and A. Glover. (2007). *Continuous Integration: Improving Software Quality and Reducing Risk*. Addison Wesley Professional.

Freeze, G., P. Mariner, J. E. Houseworth, and J. C. Cunnane. (2010). *Used Fuel Disposition Campaign Features, Events, and Processes (FEPs): FY10 Progress Report*. SAND2010-5902. Albuquerque, NM: Sandia National Laboratories,

Freeze, G., J. G. Arguello, R, Howard, J, McNeish, P. A. Schultz, and Y. Wang. (2010). *NEAMS Waste IPSC Challenge Problems*. Forthcoming. Albuquerque, NM: Sandia National Laboratories.

Helton, J. C. (1999). ―Uncertainty and Sensitivity Analysis in Performance Assessment for the Waste Isolation Pilot Plant.” *Computer Physics Communications* 117, no. 1–2: 156–180.

Knupp, P., and K. Salari. (2003). *Verification of Computer Codes in Computational Science and Engineering*. Boca Raton, FL: Chapman & Hall/CRC.

Knupp, P., C. Ober, and R. Bond. (2007). ―Measuring Progress in Premo Order-Verification." *Engr. w/Computers* 23: 283–294.

Licensing Support Network. (LSN). (2010). www.lsnnet.gov (accessed on October 10, 2010).

NEAMS Waste Forms Team. (2009). *Waste Forms and Systems Integrated Performance and Safety Codes System Design Specification*. SAND2009-3969. Albuquerque, NM: Sandia National Laboratories.

Nelson, R., C. Unal, J. Stewart, and B. Williams. (2010). *Using Error and Uncertainty Quantification to Verify and Validate Modeling and Simulation*. LA-UR-10-06125. Los Alamos, NM: Los Alamos National Laboratory.

Nuclear Energy Agency (NEA). 1999. *An International Database of Features, Events and Processes*. Nuclear Energy Agency, Paris, France: Organization for Economic Co-operation and Development.

Oberkampf, W. L., and T. G. Trucano. (2003). *Verification, Validation, and Predictive Capability in Computational Engineering and Physics*. SAND2003-3769. Albuquerque, NM: Sandia National Laboratories.

Oberkampf, W. L., M. Pilch, and T. G. Trucano. (2007). *Predictive Capability Maturity Model for Computational Modeling and Simulation*. SAND2007-5948. Albuquerque, NM: Sandia National Laboratories.

Pilch, M., T. Trucano, J. Moya, G. Froehlich, A. Hodges, and D. Peercy (2000). *Guidelines for Sandia ASCI Verification and Validation Plans – Content and Format: Version 2.0*. SAND2000-3101. Albuquerque, NM: Sandia National Laboratories.

Pilch, M., T. G. Trucano, and J. C. Helton. (2006). *Ideas Underlying Quantification of Margins and Uncertainties (QMU): A White Paper*. SAND2006-5001. Albuquerque, NM: Sandia National Laboratories.

Poppendieck, M., and T. Poppendieck. (2006). *Implementing Lean Software Development*. Addison Wesley.

Roache, P. J. (1998). *Verification and Validation in Computational Science and Engineering*. Albuquerque, NM: Hermosa Publishers.

Trucano, T. G., R. G. Easterling, K. J. Dowding, T. L. Paez, A. Urbina, V. J. Romero, B. M. Rutherford, and R. G. Hills. (2001). *Description of the Sandia Validation Metrics Project*. SAND2001-1339. Albuquerque, NM: Sandia National Laboratories.

U.S. Environmental Protection Agency (EPA). (1998). *Criteria for the Certification and Re-Certification of the Waste Isolation Pilot Plant's Compliance with the 40 CFR Part 191 Disposal Regulations*. 40 CFR 194. Washington, DC: Office of Radiation and Indoor Air.

U.S. Nuclear Regulatory Commission (NRC). (1999). *Regulatory Perspectives on Model Validation in High-Level Radioactive Waste Management Programs: A Joint NRC/SKI White Paper*. NUREG-1636. Washington, DC: U.S. Nuclear Regulatory Commission.

U.S. Nuclear Regulatory Commission (NRC). (2001). *Disposal of High-Level Radioactive Wastes in a Geologic Repository at Yucca Mountain, Nevada*. 10 CFR 63. Washington, DC: U.S. Government Printing Office.

U.S. Nuclear Regulatory Commission (NRC).(2003). *Yucca Mountain Review Plan*. NUREG-1804, Revision 2. Washington, DC: Office of Nuclear Material and Safeguards.

van Bloemen Waanders, B. G., R. A. Bartlett, S. S. Collis, E. R. Keiter, C. C. Ober, T. M. Smith, V. Akcelik, O. Ghattas, J. C. Hill, M. Berggren, M. Heinkenschloss, and L. C. Wilcox. (2005). *Sensitivity Technologies for Large Scale Simulation*. SAND2004-6574. Albuquerque, NM: Sandia National Laboratories.

Wilks, D. S. (1995). *Statistical Methods in the Atmospheric Sciences*. San Diego, CA: Academic Press.

Distribution

(to be distributed electronically)

1	MS0899	Technical Library	9536 (electronic copy)

Sandia National Laboratories