
Addressing Uncertainty in MultiSector Dynamics Research

Release v0.1.0

Patrick M. Reed, ...

May 28, 2021

CONTENTS

1	Introduction	1
2	Diagnostic Modeling Overview and Perspectives	5
2.1	Overview of model diagnostics	5
2.2	Perspectives on diagnostic model evaluation	7
3	Indices and tables	9

INTRODUCTION

This guidance text has been developed in support of the Integrated Multisector Multiscale Modeling (IM3) Science Focus Area’s objective to formally integrate uncertainty into its research tasks. IM3 is focused on innovative modeling to explore how human and natural system landscapes in the United States co-evolve in response to short-term shocks and long-term influences. The project’s challenging scope is to advance our ability to study the interactions between energy, water, land, and urban systems, at scales ranging from local (~1km) to the contiguous United States, while consistently addressing influences such as population change, technology change, heat waves, and drought. Uncertainty and careful model-driven scientific insights are central to IM3’s key MultiSector Dynamics (MSD) science objectives shown below.

IM3 key MSD science objectives include:

Develop flexible, open-source, and integrated modeling capabilities that capture the structure, dynamic behavior, and emergent properties of the multiscale interactions within and between human and natural systems.

Use these capabilities to study the evolution, vulnerability, and resilience of interacting human and natural systems and landscapes from local to continental scales, including their responses to the compounding effects of long-term influences and short-term shocks.

Understand the implications of uncertainty in data, observations, models, and model coupling approaches for projections of human-natural system dynamics.

Addressing the objectives above poses a strong transdisciplinary challenge that heavily depends on a diversity of models and, more specifically, a consistent framing for making model-based science inferences. The term transdisciplinary science as used here formally implies a deep integration of disciplines to aid our hypothesis driven understanding of coupled human-natural systems—bridging differences in theory, hypothesis generation, modeling, and modes of inference (National Research Council, 2014). The IM3 MSD research focus and questions require a deep integration across disciplines, where new modes of analysis can emerge that rapidly synthesize and exploit advances for making decision relevant insights that at minimum acknowledge uncertainty and more ideally promote a rigorous quantitative mapping of its effects on the generality of claimed scientific insights. More broadly, diverse scientific disciplines engaged in the science of coupled human-natural systems, ranging from natural sciences to engineering and economics, employ a diversity of numerical computer models to study and understand their underlying systems of focus. The utility of these computer models hinges on their ability to represent the underlying real systems with sufficient fidelity and enable the inference of novel insights. This is particularly challenging in the case of coupled human-natural systems where there exists a multitude of interdependent human and natural processes taking place that could potentially be represented. These processes usually translate into modeled representations that are highly complex, non-linear, and exhibit strong interactions and threshold behaviors (Elsawah et al., 2020; Haimes, 2018; Helbing, 2013). Model complexity and detail have also been increasing as a result of our improving understanding of these processes, the availability of data, and the rapid growth in computing power (Saltelli et al., 2019). As model complexity grows, modelers need to specify a lot more information than before: additional model inputs and relationships as more processes are represented, higher resolution data as more observations are collected, new coupling relationships and interactions as models are put together to answer multisector questions (e.g., the land-water-energy nexus). Typically, not all of this information is well known, nor is the impact of these many uncertainties on model outputs well understood. It is further especially difficult to distinguish the effects of individual as well as interacting sources of uncertainty when modeling coupled systems with multisector and multiscale dynamics (Wirtz and Nowak, 2017).

Given the challenge and opportunity posed by the disciplinary diversity of IM3, we utilized a team-wide survey to allow the project's membership to provide their views on how their areas typically address uncertainty, emphasizing key literature examples and domain-specific reviews. Our synthesis of this survey information in Figure 1 summarizes the team's perspectives, enabling a summary of the commonalities and differences for how different disciplinary areas are typically addressing uncertainty. Figure 1 highlights the non-trivial challenge posed by seeking to carefully consider uncertainty across an MSD focused transdisciplinary team. There are significant differences across the team's contributing disciplines in terms of the methodological approaches and tools used in the treatment of uncertainty. The horizontal axis of the figure represents a conceptual continuum of methodological approaches, ranging from deterministic (no uncertainty) modeling to the theoretical case of fully engaging in modeling all sources of uncertainty. The vertical axis of plot maps the analysis tools that are used in the disciplines' literature, spanning error-driven historical analyses to full uncertainty quantification. Given that Figure 1 is a conceptual illustration, the mapping of each discipline's boundaries is not meant to imply exactness. They encompass the scope of feedback attained in the team-wide survey responses. The color circles designate specific sources of uncertainty that could be considered. Within the mapped disciplinary approaches, the color circles distinguish those sources of uncertainty that are addressed in the bodies of literature reported by respondents. Note the complete absence of grey circles designating that at present few if any studies report results for understanding how model coupling relationships shape uncertainty. We can briefly distinguish the key terms of uncertainty quantification (UQ) and uncertainty characterization (UC). UQ refers to the formal focus on the full specification of likelihoods as well as distributional forms necessary to infer the joint probabilistic response across all modeled factors of interest (Cooke, 1991). Alternatively, uncertainty characterization as defined here, refers to exploratory modeling of alternative hypotheses for the co-evolutionary dynamics of influences, stressors, as well as path dependent changes in the form and function of modelled systems (Moallemi et al., 2020a; Walker et al., 2003). Uncertain factors are any model component which is affected by uncertainty: inputs, resolution levels, coupling relationships, model relationships and parameters. When a model has been established as a sufficiently accurate representation of the system some of these factors may reflect elements of the real-world system that the model represents (for example, a population level parameter would reflect a sufficiently accurate representation of the population level in the system under study). As discussed in later sections, the choice of UQ or UC depends on the specific goals of studies, the availability of data, the types of uncertainties (e.g., well-characterized or deep), the complexity of underlying models as well as the computational limits. Deep uncertainty (as opposed to well-characterized) refers to situations where expert opinions consulted on a decision do not know or cannot agree on system boundaries, or the outcomes of interest and their relative importance, or the prior probability distribution for the various uncertain factors present (Kwakkel et al., 2016; W. E. Walker et al., 2013).

At present, there is no singular guide for confronting the computational and conceptual challenges of the multi-model, transdisciplinary workflows that characterize ambitious projects such as IM3 (Saltelli et al., 2015). The primary aim of this text is to begin to address this gap and provide guidance for facing these challenges. Chapter 2 provides an overview of diagnostic modeling and the different perspectives for how we should evaluate our models. Chapter 3 the basic methods and concepts for sensitivity analysis. Chapter 4 delves into more technical applications of sensitivity analysis to support diagnostic model evaluation and exploratory modeling. Chapter 5 transitions to an overview of the key concepts and tools for UQ. Chapter 6 transitions to the use of UQ to capture risks and extremes in MSD systems. Chapter 7 provides concluding remarks across the UC and UQ topics covered in this text. The appendices of this text include a glossary of the key concepts as well as example test cases and scripts to showcase various UC and UQ related tools.

State-of-the-art as reported by IM3 teams

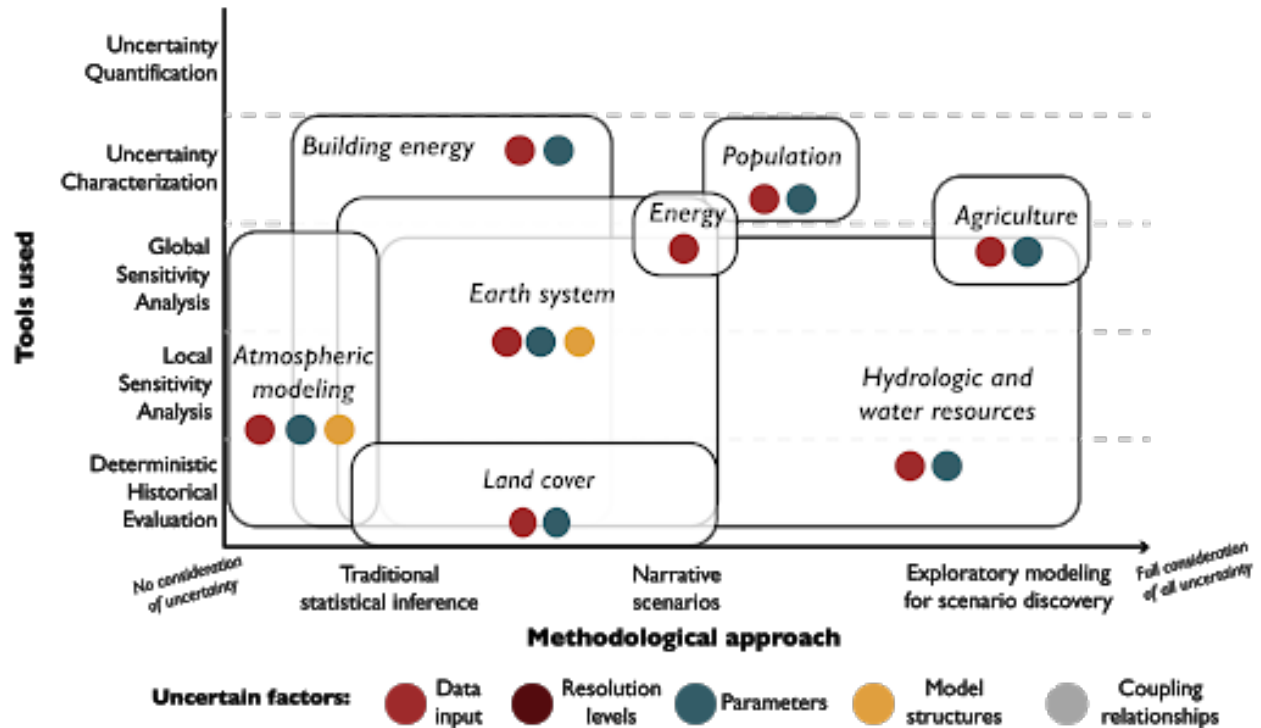


Fig. 1: State-of-the-art in different modeling communities, as reported in the survey distributed to IM3 teams. Deterministic Historical Evaluation: model evaluation under fully determined conditions defined using historical observations; Local Sensitivity Analysis: model evaluation performed by varying uncertain factors around specific reference values; Global Sensitivity Analysis: model evaluation performed by varying uncertain factors throughout their entire feasible value space; Uncertainty Characterization: model evaluation under alternative factor hypotheses to explore their implications for model output uncertainty; Uncertainty Quantification: representation of model output uncertainty using probability distributions; Traditional statistical inference: use of analysis results to describe deterministic or probabilistic outcomes resulting from the presence of uncertainty; Narrative scenarios: use of a limited decision-relevant number of scenarios to describe (sets of) changing system outcomes; Exploratory modeling for scenario discovery: use of large ensembles of uncertain conditions to discover decision-relevant combinations of uncertain factors

DIAGNOSTIC MODELING OVERVIEW AND PERSPECTIVES

This text prescribes a formal model diagnostic approach to IM3 computational experimentation that is a deliberative and iterative combination of state-of-the-art UC and global sensitivity analysis techniques that progresses from observed history-based fidelity evaluations to forward looking resilience and vulnerability inferences (Gupta et al., 2008; Hadjimichael et al., 2020).

2.1 Overview of model diagnostics

Model diagnostics provide a rich basis for hypothesis testing, model innovation, and improved inferences when classifying what is controlling highly consequential results (e.g., vulnerability or resilience in coupled human-natural systems). Figure 2, adapted from (Saltelli et al., 2019), presents idealized illustrations of the relationship between UC and global sensitivity analysis (GSA) for two coupled simulation models. The figure illustrates how UC can be used to address how uncertainties in various modeling decisions (data inputs, parameters, model structures, coupling relationships, and elsewhere) can be sampled and simulated to yield the empirical model output distribution(s) of interest. Monte Carlo frameworks allow us to sample and propagate (or integrate) the ensemble response of the model(s) of focus. The first step of any UC analysis is the specification of the initial input distributions as illustrated in Figure 2. The second step is to perform the Monte Carlo simulations. The question can then be raised, which of the modeling assumptions in our Monte Carlo experiment are the most responsible for the resulting output uncertainty. We can answer this question using “global sensitivity analysis” (GSA) as illustrated in Figure 2. GSA can be defined as a formal Monte Carlo sampling and analysis of modeling choices (structures, parameters, inputs) to quantify their influence on direct model outputs (or output-informed metrics). UC experiments by themselves do not explain why you get a particular uncertain outcome. The pie chart shown in Figure 2 is a conceptual representation of the results of GSA to identify those factors that are most dominantly influencing results, either individually or interactively (Saltelli et al., 2008).

UC and GSA are not independent modeling analyses. As illustrated here, any GSA requires an initial UC hypothesis in the form of statistical assumptions and representations for the modeling choices of focus (structural, parametric, and data inputs). Information from these two model diagnostic tools can then be used to inform data needs for future model runs, experiments to reduce the uncertainty present, or the simplification or enhancement of the model where necessary. Together UC and GSA provide a foundation for diagnostic exploratory modeling that has a consistent focus on the assumptions, structural model forms, alternative parameterizations, and input data sets that are used to characterize the behavioral space of one or more models.

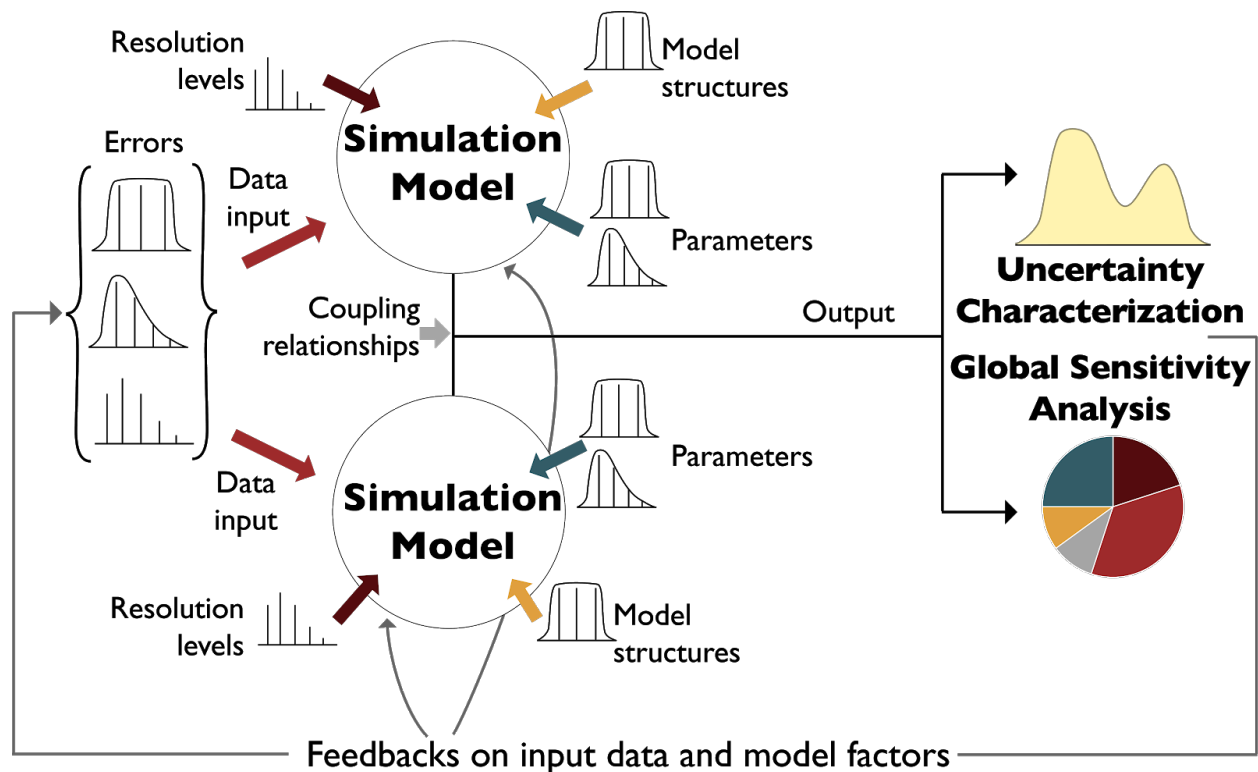


Fig. 1: Idealized uncertainty characterization and global sensitivity analysis for two coupled simulation models. Uncertainty coming from various sources (inputs, model structures, coupling relationships, and elsewhere) is propagated through the coupled model(s) to generate empirical distributions of outputs of interest (uncertainty characterization). This model output uncertainty can be decomposed to its origins, by means of sensitivity analysis. This figure has been adapted from Saltelli et al. (2019).

2.2 Perspectives on diagnostic model evaluation

When we judge or diagnose models the terms “verification and validation” are commonly used. However, their appropriateness in the context of numerical models representing complex coupled human-natural systems is questionable (Beven, 2002; Oreskes et al., 1994). The core issue relates to the fact that these systems are often not fully known or perfectly implemented when modeled. Rather, they are defined within specific system framings and boundary conditions in an evolving learning process with the goal of making continual progress towards attaining higher levels of fidelity. For example, observations used to evaluate the fidelity of parameterized processes are often measured at a finer resolution than is represented in the model and then must be scaled up for the evaluation. In other cases, numerical models may neglect or simplify system processes because the data is not available or the physical mechanisms are not fully known. If sufficient agreement between prediction and observation is not achieved, it is challenging to know whether these types of modeling choices are the cause, or if other issues, such as deficiencies in the input parameters and/or other modeling assumptions are the true cause of errors. Even if there is high agreement between prediction and observation, the model cannot necessarily be considered validated, as it is always possible that the right values were produced for the wrong reasons. For example, low error can stem from a situation where different errors in underlying assumptions or parameters cancel each other out (“compensatory errors”). Furthermore, coupled human-natural system models are often subject to “equifinality”, a situation where multiple parameterized formulations can produce similar outputs or equally acceptable representations of the observed data. There is therefore no uniquely “true” or validated model, and the common practice of selecting “the best” deterministic calibration set is more of an assumption than a finding (Beven, 1993; Beven and Binley, 1992). The situation becomes even more tenuous when observational data is limited in its scope and/or quality to be insufficient to distinguish model representations or their performance differences. These limitations on model verification undermine any purely positivist treatment of model validity: that a model should correctly and precisely represent reality to be valid. Under this perspective, closely related to empiricism, statistical tests should be used to compare the model’s output with observations and only through empirical verification can a model or theory be deemed credible. A criticism to this viewpoint (besides the aforementioned challenges for model verification) is that it reduces the justification of a model to the single criterion of predictive ability and accuracy (Barlas and Carpenter, 1990). Authors have argued that this ignores the explanatory power held in models and other procedures, which can also advance scientific knowledge (Toulmin, 1977). These views gave rise to relativist perspectives of science, which instead place more value on model utility in terms of fitness for a specific purpose or inquiry, rather than representational accuracy and predictive ability (Kleindorfer et al., 1998). This viewpoint appears to be most prevalent among practitioners seeking decision relevant insights (i.e., inspire new views vs. predict future conditions). The relativist perspective argues for the use of models as heuristics that can enhance our understanding and conceptions of system behaviors or possibilities (Eker et al., 2018). In contrast, natural sciences favor a positivist perspective, emphasizing similarity between simulation and observation even in application contexts where it is clear that projections are being made for conditions that have never been observed and the system of focus will have evolved structurally beyond the model representation being employed (e.g., decadal to centennial evolution of human-natural systems).

These differences in prevalent perspectives are mirrored in how model validation is defined by the two camps: From the relativist perspective, validation is seen as a process of incremental “confidence building” in a model as a mechanism for insight (Barlas, 1996), whereas in natural sciences validation is framed as a way to classify a model as having an acceptable representation of physical reality (Oreskes et al., 1994). Even though the relativist viewpoint does not dismiss the importance of representational accuracy, it does place it within a larger process of establishing confidence through a variety of tools. These tools, not necessarily quantitative, include communicating information between practitioners and modelers, interpreting a multitude of model outputs, and contrasting preferences and viewpoints.

On the technical side of the argument, differing views on the methodology of model validation appear as early as in the 1960’s. (Naylor and Finger, 1967) argue that model validation should not be limited to a single metric or test of performance (e.g., a single error metric), but should rather be extended to multiple tests that reflect different aspects of a model’s structure and behavior. This and similar arguments are made in literature to this day (Beven, 2018; Gupta et al., 2012, 2008; Kumar, 2011; Nearing et al., 2020) and are primarily founded on two premises. First, that even though modelers widely recognize that their models are abstractions of the truth, they still make truth claims based on traditional performance metrics that measure the divergence of their model from observation (Nearing et al., 2020). Second, that the natural systems mimicked by the models contain many processes that exhibit significant heterogeneity at various

temporal and spatial scales. This heterogeneity is lost when a single performance measure is used, as a result of the inherent loss of process information occurring when transitioning from a highly dimensional and interactive system to the dimension of a single metric (Beven, 2002). These arguments are further elaborated in section 4.1 Understanding Errors.

Multiple authors have proposed that, instead, the evaluation of several model performance signatures (characteristics) should be considered to identify model structural errors and achieve a sufficient assessment of model performance (Gupta et al., 1998). There is however a point of departure here, especially when models are used to produce inferences that can inform decisions. When agencies and practitioners use models of their systems for public decisions, those models have already met sufficient conditions for credibility (e.g., acceptable representational fidelity), but may face broader tests on their salience and legitimacy in informing negotiated decisions (Cash et al., 2003; Eker et al., 2018; White et al., 2010). This presents a new challenge to model validation, that of selecting decision-relevant performance metrics, reflective of the system's stakeholders' viewpoints, so that the most consequential uncertainties are identified and addressed (Saltelli and Funtowicz, 2014). For complex multisector models at the intersection of climatic, hydrologic, agricultural, energy, or other processes, the output space is made up of a multitude of states and variables, with very different levels of salience to the system's stakeholders and to their goals being achieved. This is further complicated when such systems are also institutionally and dynamically complex. As a result, a broader set of qualitative and quantitative performance metrics is necessary to evaluate models of such complex systems, one that embraces the plurality of value systems, agencies and perspectives present. For IM3, even though the goal is to develop better projections of future vulnerability and resilience in co-evolving human-natural systems and not to provide decision support per se, it is critical for our multisector, multiscale model evaluation processes to represent stakeholders' adaptive decision processes credibly.

As a final point, when a model is used in a projection mode, its results are also subject to additional uncertainty, as there is no guarantee that the model's functionality and predictive ability will stay the same as the baseline, where the verification and validation tests were conducted. This challenge requires an additional expansion of the scope of model evaluation: a broader set of uncertain conditions needs to be explored, spanning beyond historical observation and exploring a wide range of unprecedented conditions. This perspective on modeling, termed exploratory (Bankes, 1993), views models as computational experiments that can be used to explore vast ensembles of potential scenarios so as to identify those with consequential effects. Exploratory modeling literature explicitly orients experiments toward stakeholder consequences and decision-relevant inferences and shifts the focus from predicting future conditions to discovering which conditions lead to undesirable or desirable consequences.

This evolution in modeling perspectives can be mirrored by the IM3 family of models in a progression from evaluating models relative to observed history to advanced formalized analyses to make inferences on multisector, multiscale vulnerabilities and resilience. Exploratory modeling approaches can help fashion experiments with large numbers of alternative hypotheses on the co-evolutionary dynamics of influences, stressors, as well as path-dependent changes in the form and function of human-natural systems (Weaver et al., 2013). The aim of this text is to therefore guide the reader through the use of sensitivity analysis and uncertainty methods across these perspectives on diagnostic and exploratory modeling.

INDICES AND TABLES

- genindex
- modindex
- search