SANDIA REPORT

SAND2013-7424
Unlimited Release
Printed September 2013

Roll-up of Validation Results to a Target Application

Richard G. Hills

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.

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

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

SAND 2013-7424
Unlimited Release
Printed September 2013

Roll-up of Validation Results to a Target Application

Richard G. Hills
V&V, UQ, Credibility Processes
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87195-0828

Abstract

Suites of experiments are preformed over a validation hierarchy to test computational simulation models for complex applications. Experiments within the hierarchy can be performed at different conditions and configurations than those for an intended application, with each experiment testing only part of the physics relevant for the application. The purpose of the present work is to develop methodology to roll-up validation results to an application, and to assess the impact the validation hierarchy design has on the roll-up results. The roll-up is accomplished through the development of a meta-model that relates validation measurements throughout a hierarchy to the desired response quantities for the target application. The meta-model is developed using the computation simulation models for the experiments and the application. The meta-model approach is applied to a series of example transport problems that represent complete and incomplete coverage of the physics of the target application by the validation experiments.

ACKNOWLEDGMENTS

The author would like to thank Timothy Trucano and Kenneth Hu of Sandia National Laboratories for their thorough review of this report. The author would also like to thank Martin Pilch of Sandia National Laboratories for many hours of discussion concerning uncertainty quantification of model predictions.

CONTENTS

FIGURES

TABLES

EXECUTIVE SUMMARY

Suites of experiments are performed over a validation hierarchy to test computational simulation models for complex applications. Experiments within the hierarchy can be performed at different conditions than those for an intended application, with each experiment designed to test only part of the physics relevant for the application. The experiments may utilize idealized representations of component geometries, with each experiment returning measurement types (i.e. temperature, pressure, flux, first arrival times) that may be different from the response quantity of interest for the application. How does one relate or "roll-up" results at various levels in a validation hierarchy to predictions and uncertainty for the anticipated application? Issues associated with the roll-up of hierarchical results to an application prediction include properly weighting of individual experimental results to best represent an the application, assessing whether the suite of experiments adequately tests the anticipated physics of the application, and characterizing the additional uncertainty in an application prediction due to lack of coverage of the application physics by the physics addressed by the validation experiments.

The purpose of the present work is to develop methodology to roll-up validation results to an application, and to assess the impact the design of the validation hierarchy has on the rolled-up results. This is accomplished through the development of a meta-model that relates the validation measurements to the application response quantity of interest. The meta-model possesses the following features: The meta-model accommodates the presence of computational model parameter uncertainty in both the computational models for the validation experiments and the target application, as well as measurement uncertainty in the validation data. The meta-model accommodates possible incomplete physics coverage of the application by the validation suite. The meta-model allows different validation measurement and application response quantity types. The meta-model addresses the impact of validation experiments performed at conditions different from those of the anticipated application.

For the present development, the meta-models will be constructed as weighted combinations of the computational based representations of the validation experiments over computational neighborhoods around each measurement location, time, and type. Specifically sampling techniques are used over these neighborhoods to characterize the dependence of the computational based validation and application models on the important computational model parameters and independent variables so that the meta-models best represent the behavior of the computational application model for the response quantity of interest. Two approaches are used and compared to evaluate the meta-model weights. The first is based on an objective function defined to explicitly accommodate the trade-off between the ability of the meta-model to resolve the target application model, and the sensitivity of the meta-model to computational parameter and measurement uncertainty. The second approach is based on partial least squares regression (PLSR). The trade-off between resolution (fidelity) and sensitivity for the PLSR approach is addressed through the number of latent variables utilized for the regression.

The methodology is applied to a series of example transport problems that represent complete and incomplete coverage of the physics of the target application by the validation experiments. The methodology estimates the uncertainty that is introduced due to the lack of coverage of the

application physics, due to experiments performed at different conditions than those of the application, and due to uncertainties in the validation exercise (computational model parameter and measurement uncertainty). Relative assessment of the two approaches is accomplished through comparison of meta-model results to the original computational target application model results, and through a sensitivity analysis. The results indicate that the partial least squares approach is superior for the examples considered. The examples also illustrate many of the difficulties associated with the roll-up of validation experimental results to the application, as well as some of the limitations of the present methodology.

1. BACKGROUND

1.1. Model Validation

AIAA (1998) and ASME (2006) define model validation as "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended use of the model." Model validation is based on the direct comparison between experimental data and model predictions (typically computational simulations) of the data, relative to the uncertainty in the validation exercise (ASME, 2009).

Comparisons of prediction to measurement provide the most direct evidence of the ability of a model to simulate the "real" (that is, observed) physics. These differences are only estimates as both the measurements and the model predictions contain uncertainty. The goal in validation is to characterize the difference between model prediction and "true" behavior (that is, the true value for the measurand which would be observable only if there was no measurement error present), and to characterize the uncertainty in this difference.

Generally, suites of experiments are preformed over a defined hierarchy of experiments for complex applications. This hierarchy is often referred to as a validation hierarchy or validation pyramid. There are often three types of experiments in a validation hierarchy: material characterization experiments, ensemble validation experiments, and accreditation experiments. Data from material characterization experiments are used to calibrate constitutive models, are generally less expensive to perform, and produce more observational data (i.e., over multiple material samples). Ensemble validation experiments represent suites of experiments designed to test a computational model's ability to represent various aspects of the physics relevant to the application. They generally do not represent the full complexity of the target application of the computational simulation. These experiments may or may not provide sufficient data to characterize variability across similar tests. These experiments often are more expensive than material characterization experiments, resulting in less data. Accreditation tests can involve sub-system or full system testing with application hardware tested under conditions more closely representing the design conditions or regulatory requirements of the target application. Such experiments are typically expensive, resulting in limited data. Figure 1 represents this hierarchy of experiments. Material characterization experiments generally use geometrically simple material samples and are ideally performed over the range of environmental conditions (e.g. temperature range) expected for the target application. Ensemble validation experiments represent more geometric and physical complexity, but are often not performed over the full range of environmental conditions expected for the target application. The lack of full range coverage may be due to the inability of laboratory experiments to test at the target application conditions, or due to the expense of performing validation experiments over the range of conditions. Because fewer accreditation experiments can be performed due to their expense, and because they are typically performed for a limited set of conditions, they often do not represent the entire design space of the application.

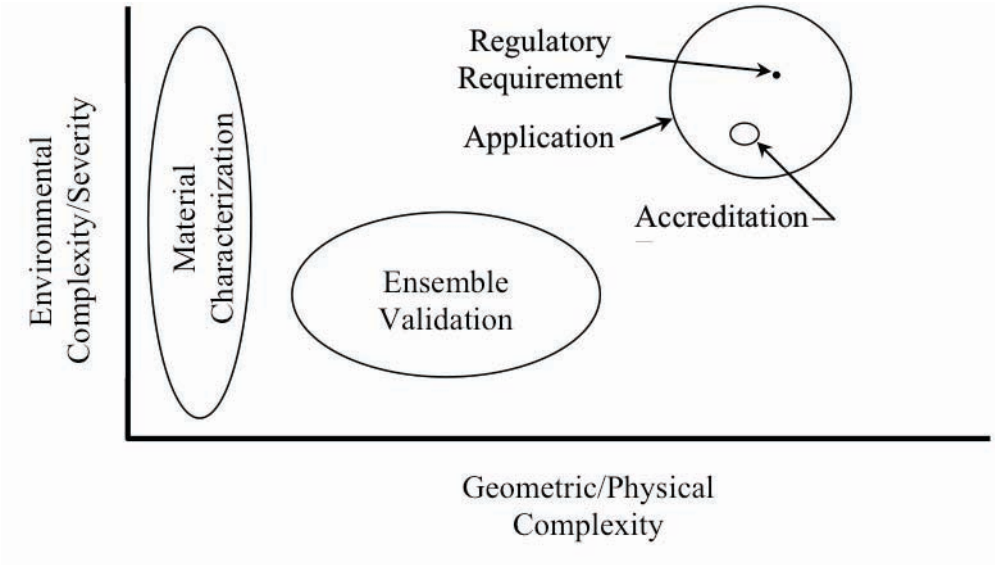

Figure 1: Experimental Hierarchy (Hills et al., 2008)

Methodology is required to assess the impact of differences observed between validation experiment results and computational model predictions of the experimental results, and computational predictions of the target application response quantities of interests. The primary focus of the present research is to investigate methodology to roll-up validation results and their associated uncertainties to the predicted response quantity of interest for a target application.

1.2. Sources and Types of Uncertainty Considered

The sources of uncertainties considered here are

- Model parameter uncertainty for the target application – uncertainty in the correct values of the model parameters for an application
- Validation uncertainty including
  - Data uncertainty for the experiments
  - Parameter uncertainty for the models of the experiments
  - Uncertainty in the projection or roll-up of observed differences to the target application due to lack of coverage of the target application by the validation experiments, or due to experimental conditions different that those for the target application.
- Expert opinion uncertainty as an assessment of adequacy of model based on experience.

Note that solution verification uncertainty is not considered. Solution verification uncertainty is that uncertainty in prediction associated with the lack of mesh convergence of finite difference or finite element algorithms. This type of uncertainty can be incorporated into the present methodology through the use of the methods defined in ASME (2009).

An important component in the uncertainty quantification process, when the target application model is utilized for risk assessment, is to identify and separate the sources of epistemic and aleatory uncertainty. Sources of aleatory uncertainty in model predictions include material properties that reflect random effects, such as heterogeneities associated with grain structure, pores, and impurities. While the uncertainty associated with these randomly induced quantities can be classified as aleatory, understand that the uncertainty in the parameters that appear in probability density functions characterizing this variability are epistemic as their uncertainty can be reduced by acquiring additional characterization data.

Sources for aleatory uncertainty in experimental data include electronic noise, and uncertainty associated with counts per unit time for a stochastic-based measurement (i.e. the number of neutron counts measured by a neutron probe). Sources of epistemic uncertainty include bias due to calibration errors, and bias in sensor readings due to installation and operation (i.e., thermal couple lead heat loses and contact resistance).

In the present work, aleatory and epistemic uncertainties are considered separately through the second probability method. This method is a sampling based uncertainty quantification (UQ) method in which aleatory uncertainty is propagated through a computational simulation, for a realization of the epistemic parameters. This process is repeated for other realizations of the epistemic parameters, resulting in families of PDF's or CDF's, with each PDF/CDF corresponding to a single realization of the epistemic parameters. This approach directly supports the evaluation of *Probability of Frequency* (PoF) of failure used for safety and other computational analysis at Sandia National Laboratories (Pilch et al., 2006, Diegart et. al, 2007). The second probability method is illustrated through an expert opinion example.

1.3. Roll-Up of Uncertainties

Computational model form uncertainties, as characterized by validation exercises, include both the observed differences in the model form error, and the uncertainty in these differences due to uncertainty in the data and in the model simulations. The roll-up of these observed differences and associated uncertainties to a target application requires some model for the relationship between observations from the validation experiments and predictions at the target level. The construction of this model for the relationship presents a difficult challenge. If the underlying physics models and the computational simulation models associated with the target application and validation experiments are approximately correct, one can use these models and simulations to characterize the relationships between different points in the hierarchy. If these models are insufficiently to provide useful approximations, then the use of these models to develop relationships between points in the hierarchy is only of vague value. Consider the following scenarios.

- **Scenario 1**: A component of physics that significantly affects behavior of the response quantity of interest at the conditions of the target application is unknown to the analyst and not included in the model, and the validation experiments are at conditions or configured so as to not be affected by this physics. In this case, the model for the target application will not be valid, and the analyst will not know that the model is not valid. The occurrence of this case is sometimes identified through unexpected failures during surveillance, day-to-day operations of the system, and certification tests.

- **Scenario 2**: Validation experiments are sensitive to physics important to the application, but the computational models do not accurately simulate this physics. While these models are not accurate, the validation experiments suggest that these models do represent the dominant physics. In this case, the computational models may be sufficient to define approximate transfer or mapping functions of the observed differences between the computational models and experimental data to the application conditions.
- **Scenario 3**: Physics or coupling of physics that is important for the application is thought to be known and is modeled, but validation experiments to test the model for this physics or coupling are not available. This may be the case where one simply cannot perform experiments due to economics, time constraints, or for environments that cannot be simulated in the lab or in field experiments. In this case, the computational model can be used to perform a sensitivity analysis to evaluate the impact on the uncertainties in the target application predictions due to the lack of coverage by the validation experiments.
- **Scenario 4**: The important physics have been approximately incorporated in the model for the application with validation data that provides good coverage of the important physics. The major uncertainty in this case is the models have been tested at conditions somewhat different from the application. As a result, the model errors and the validation exercise uncertainties must be interpolated or extrapolated to the configuration and conditions of the target application to determine their impact on the response quantity of interest.

In the following analysis, we assume that the simulation model for the application represents our best knowledge of the system. We use this knowledge and the results from the validation experiments to roll-up the validation results to the target application. This roll-up will potentially be very unreliable if the scenario discussed under the first scenario exists. Scenarios 2 through 4 represent the presence of at least some knowledge of the important physics, either as approximately represented in the simulation model, or represented in the behavior of the validation experiments.

1.3.1. Methods to Roll-up Model Form Uncertainty Characterized by Validation Experiments

A common approach in computational modeling, given observed validation differences, is to assume the model is valid if the differences are sufficiently small, and then apply the model to the application without considering the impact that these differences may have on the uncertainty in the application prediction. The choice of what ‘sufficiently small’ is arbitrary and the assumption that small differences for the validation experiments imply small differences for an application may not be appropriate. For example, should one consider the model as valid if the experimental data lie within one statistical standard deviation of the model predictions, where the standard deviation corresponds to the combined uncertainty in the validation exercise (i.e., due to model parameter, data, and numerical convergence uncertainty)? Are results that are within two standard deviations sufficient to declare a model valid? This approach also does not consider the additional uncertainties associated with the lack of coverage of the target application by the validation experiments.

Several approaches that do consider the impact of observed differences from validation experiments on a target application prediction have been developed. These approaches use the existing simulation models as a basis, and 1) expand the uncertainties in the model parameters to encompass the model validation data when the differences are larger than the estimated uncertainty in the validation exercise, 2) develop a calibrated correction of the computational simulation model that assumes some functional form over the range of conditions spanned by the application and validation experiments, or 3) uses the simulation model to develop a meta-model to map the observed differences and uncertainties from the validation experiments to the application. An alternative approach to rolling up differences directly is to develop a Bayesian network and roll-up probabilities rather than differences. The development and application of a Bayesian network to a validation hierarchy to roll-up probabilities for different validation measurement and target application response quantity types is in it's early stages of development (Mahadevan, 2011) and is not further discussed here.

Approach 1: Use Model only if Validation Results are Adequate.

Method and assumptions:

- Assume that the physics of the application is represented by the validation experiments. This includes the two-way interaction effects between different types of physics that are important to an application.
- Assume that the conditions of the validation experiments are sufficiently close to that of the application.
- If the differences between the measurements and the model predictions for the validation experiments are sufficiently small, use the application model as is.

Comments:

- Requires judgment as to what is sufficiently close and sufficiently small are (bullets 2 and 3).
- The conservative principles of the underlying physics (i.e. conservation of energy, momentum, and mass) are maintained.
- The uncertainty in application predictions is not increased due to observed differences from the validation exercises.
- The approach does not provide methodology to account for lack of coverage of the important application physics by the validation experiments, nor does it account for the impact of validation experiments performed at different conditions than those of the application.

Approach 2: Capture of Validation Differences through Model Parameters

Method and assumptions:

- Assume that the dominant physics is adequately approximated in the model, but other important (but not dominant) physics, or second order effects in the dominant physics, may not be included. Assume that these important but not modeled effects are captured by the validation experiments.
- Expand the characterized uncertainty in the model parameters such that the validation differences lie within the resulting prediction uncertainty (e.g. to lie within ±2 standard

deviations of the resulting prediction uncertainty, or some other range characterizing uncertainty) for the models of the validation experiments. The parameters chosen for uncertainty expansion should be those that are important for the application.
- Use these expanded uncertainties for the parameters in performing UQ for the application.

Comments:
- Conceptually simple and uses the existing simulation model
- The conservative principles of the underlying physics (i.e. conservation of energy, momentum, and mass) are maintained as modifications are often made to the parameters that represent properties or constitutive relationships rather than the conservation principles.
- When the differences are large relative to the uncertainty in the validation exercise, the choice of which model parameters to target for an expanded uncertainty range can be somewhat arbitrary, especially when considering experiments from a hierarchy of experiments. The relationship between the physics for an application to that across the hierarchy of validation experiments is not always fully understood.
- Expansion of the uncertainty ranges in the parameters are often well beyond physically realistic values for these parameters when model form error is present. This is a red flag that there is something wrong with either the computational simulation or the experimental data.
- The method does not explicitly account for the inability of the validation experiments to cover or represent the application since the expansion of uncertainty is based on the modeled physics of the validation experiments, and not on the physics that is missing from the suite of validation experiments that may be important to the application.

Development and application of this method is discussed in Romero (2006, 2007, 2008)

Approach 3: Calibrated Correction Model

Method and assumptions:
- Estimate an additive correction model for the behavior of the observed validation differences over the range of conditions (i.e. correction model arguments should including both the model parameters and independent variables that vary over this range) of the validation experiments and the application. Polynomial or Gaussian Process correction models are sometimes used for the correction model (Higdon et al., 2008). Assume that the original model for the application, corrected by this model, approximately represents the true physics.
- Use the validation data to estimate the regression coefficients and the uncertainties in the coefficients for the correction model. Conventional regression or Bayesian calibration is typically used.
- Use the original target application model to simulate the application, and use the correction model to correct the simulation results. Include both the simulation uncertainties and the additional uncertainties associated with the regression in the roll-up of uncertainties to the target application.

Comments:

- Choice of the correction model can be arbitrary and the quality of the results can depend on whether the target application represents an interpolation or an extrapolation from the conditions of the experiments.
- Unless the correction model is carefully chosen to be consistent with the underlying conservation principles represented by the governing mathematical equations, the corrected results will not satisfy these principles (i.e., conservation of mass, momentum, and energy). This lack of consistency is typically the case when Gaussian Process or polynomial regression correction models are used.
- The method is not applicable when the type of experimental data varies over the validation hierarchy and the target application. For example, some validation experiments may collect temperature measurements at one set of conditions, pressures at another set of conditions, and chemical concentrations at a third set of conditions, whereas the application quantity of interest may be one or none of these quantities.
- While conventional regression does not explicitly account for the lack of ability of the validation experiments to represent the application, regression techniques can provide an estimate of the increased uncertainty due to extrapolation, based on the assumed form of the regression model.
- Both Bayesian and conventional estimation methods implicitly address the ability of the data from the validation experiments to 'inform' the regression process. If the regression model contains parameters that cannot be well calibrated by the data, then the estimated uncertainties in these parameters will be large. As a result, the estimated uncertainties in extrapolation may be large if the extrapolation is sensitive to these parameters. However, one must realize that one is mapping true model form error on a simplistic, non-physics based, regression model. As a result, the estimation of uncertainties in extrapolated predictions of such regression models do not represent the uncertainties in the model form errors for the target application simulation model, but represent the uncertainties associated with the estimation process itself.

Higdon et al. (2008) present examples of Bayesian calibration using Gaussian Process models.

Approach 4: Meta-Model Approach

Method and assumptions:

- Assume the model for the target application and validation experiments is approximately correct. Dominant physics is correctly included, but other important but not dominant physics or second order effects may not be included, but is represented by the validation experiments.
- Use the behavior of the models for the validation experiments, in the neighborhood of the conditions represented by the experiments; and the model for target application, in the neighborhood of the conditions for the target application, to develop a meta-model for the relation of validation experiments differences to differences in the desired response quantities for the target application. This meta-model can be constructed as a linear combination of functions or sampled behavior, where the functions or sampled behavior

represents the non-linear dependence of the models on the arguments (parameters and independent variables) deemed important to the experiments and the application.

- Evaluate the weights (i.e. regression model coefficients) in this meta-model to "best" represent the target application. Because the target application is often not adequately resolved by the validation experiments, robust methodology must be used to develop the weights and to assess the impact of this lack of resolution.
- Use the meta-model to map observed differences from the validation experiments to the conditions of the application.
- Use the meta-model to map uncertainties associated with the validation exercises, including the additional uncertainties associated with lack of coverage of the target application by the hierarchy of experiments to the application

Comments:

- Once the results for the neighborhood behavior of the validation and application models have been evaluated, the mappings and the assessment of the mappings can be post processed using computationally inexpensive algorithms.
- Approximately preserves the conservative principles incorporated in the physics based computation model for the application used to develop the meta-modal, as the method simulates the behavior of the physics based model.
- Applicable to a hierarchy of experiments and application for which the measured and the desired applications quantities of interest can be different, or for which important physics can change from validation experiment to validation experiment.

The development and application of an earlier version of the meta-model approach is provided by Hills and Leslie (2003) and Hamilton and Hills (2010a, 2010b). A sampling based approach will be further developed here.

Summary of Roll-Up Approaches

The approach used to roll-up validation experiment uncertainties to an application, when the measurement and response variable types are different, is a new and very challenging area of research. We expect that the methods that gain favor in the future will have many features that are similar to the methods discussed above.

1.3.2. Capture of Model Form Error Characterized through Expert Opinion

Roll-up of model form error, when characterized by a subject matter expert, requires adequate specification of the expert's opinion (Meyer and Booker, 1991) in a form useful for the roll-up process. To maintain consistency with the second probability method, uncertainties in expert opinion will be characterized through a probability-based (i.e., degree of belief) assessment for the example presented here. The following assumptions are made.

- The expert opinion is based on experimental or other qualified experience with the physical items and physics that are being simulated by a computational model.
- The opinion is applicable somewhere in the validation hierarchy, from the partial physics level at conditions different from the target application, to coupled physics levels at conditions closer to the target application.

- The expected behavior (maximum temperatures, failure pressures, etc.) can be characterized by the expert in quantitative but somewhat general terms. For example, the expert may suggest that based on past experience, the maximum temperature one would expect a device to reach is between T_1 and T_2 and the time that the device reaches this temperature is between t_1 and t_2. The uncertainty associated with the actual maximum temperature and time is epistemic as one could reduce these uncertainties by performing experiments on the devices. To put these quantities in a probabilistic framework, a plausibility or degree of belief probability distribution is assigned to the quantities addressed (e.g. maximum temperature and time). Meyer and Booker (1991) discuss issues with subject matter expert elicitations of distributions and provide recommendations on how to improve the elicitation process.
- For the present application, an expert derived probability distribution was used to generate "samples of expert measurements" and differences between these measurements and the modeled response are roll-up to the target application using a second probability method. The second probability method is used to separate the components of prediction uncertainty that is due to expert opinion from the other sources of uncertainty. An example of this approach is provided in Chapter 4.

1.4. Meta-Model Approach and Present Extension

The ability to resolve target application model behavior for some response quantity of interest utilizing models representing validation experimental measurements suffers from many of the issues associated with inverse problems. One often must accept the trade-off that occurs between the ability to represent the target application response quantity of interest accurately and the sensitivity of that representation to uncertainty in the validation experiments model parameters and data. Hamilton and Hills (2010a) provide methodology to provide quantitative guidance of how well validation experiments can represent a response quantity for an intended or target application model. They use physics based computational models for the validation experimental measurements and for the targeted application response quantity to develop this assessment for the application, based on the idea that the model for the target application represents the 'best knowledge' of the application. Specifically, a meta-model is developed utilizing the models for existing or potential validation experiments to best represent the dependence of the target application response quantity to important model parameters and independent variables. These meta-models provide useful quantitative information on 1) how well the experiments resolve or represent the intended application, 2) the impact of uncertainty in the validation experiment model parameters on the assessment, and 3) the trade-off between items 1) and 2). The developed meta-models are based on equivalence of first or second order Taylor Series expansions of individual validation experiment models over a suite of models, to the corresponding expansion of the target application model. The associated trade-off analysis is based on singular value decomposition (SVD). While this approach in constructing the meta-models is straight forward, the approach suffers from the limitations associated with the use of lower order Taylor Series expansions, such as the requirement to approximate derivatives that represent behavior only at the center point and not in some extended neighborhood around this point. The use of SVD to investigate the trade-off between the representation of the target application and sensitivity of the meta-model to parameter uncertainty can result in a poorly

defined trade-off curve, as just a few points along the curve are available corresponding to non-zero singular values for the corresponding pseudo-inverse.

Hamilton and Hills (2010b) extend the above approach through a two term objective function that explicitly considers the trade-off between the sensitivity of the meta-model to model parameter uncertainty (first term), and the ability of the meta-model to represent the target application model over some neighborhood (second term). A parameter is introduced that controls the relative weight of each of the two terms during the optimization process. Two measures are used to measure the closeness (i.e., ability to represent) of the meta-model to the target application model. The first looks at the sum of squares of the differences in the first order sensitivity of the target application response to the important arguments. The second uses Box-Behnken sampling (Box and Behnken, 1960) to sample quadratic behavior of validation experimental and application models to important model arguments and utilizes the maximum difference between the meta-model and application model over the sampled results. This second approach requires non-linear optimization of the weights defining the meta-model, which proves to be a non-trivial and CPU intensive process.

The present work focuses on the roll-up of model bias and uncertainty to the target application and not on the trade-off that exists between sensitivity of a meta-model to uncertainties in the model parameters and the ability of the meta-model to represent the target application model. As will be shown, one can look at a single point in the trade-off curve for this roll-up. The present work takes two approaches to developing the weights of the meta-model for this optimal location in the trade-off. The first is based on a modification of the objective function discussed above. The second is based on Partial Least Squares (Abdi, 2010). In the present development, Latin Hypercube Sampling (LHS) is used to characterize the behavior of the predictive models in the neighborhood of the validation experiment and target application conditions as defined by their model parameters and independent variables for both approaches. This provides a better representation of the nonlinear behavior of the models over the sampled neighborhoods than the previous approaches used by Hamilton and Hills (2010a, 2010b). In addition, two post assessment measures are provided to help characterize how well the validation suite can represent the physics of the target application, and the identification of the sources of lack of coverage of the physics, given the assumptions made in developing the meta-models.

2. THEORY

2.1. Background and Development

The approaches developed here to estimate the weights for the meta-model are based on the following assumptions:

- The models of the validation experiments and the model of the target application represent our 'best knowledge' of the corresponding physics.
- The primary source for information on model form error, due to incomplete or incorrect representation of the physics, is obtained from differences between model predictions and the experimental observations observed from the model validation experiments.
- The dominant physics for the application is included in the CompSim. The physics that is not included is less dominant for the application. For example, errors may exist in the constitutive models associated with transport phenomena, but the dominant transport mechanism, as based on conservation of mass, momentum, and energy, is approximately correct.

In a loose sense, the present approach develops a meta-model to *transport* differences between experiment and model prediction to the conditions of the target application. The meta-model is also used to assess uncertainties associated with the incomplete coverage of the target application by the validation experiments.

The first step is to define a metric that measures distance between a validation experiment and a target application in some sense. This metric will then be used to aggregate results from the suite of validation experiments to 'best' represent the target application. For illustrative purposes, consider the case of a propagating wave as shown in Figure 2. The two arguments in the figure may represent position and time, or represent two spatial directions (i.e., x, y) for a particular time. The target application point of interest, and some neighborhood about this point is shown in blue, while the conditions of the validation experiments are shown in red. Note that the validation experiment that best represents the dependence of the quantity of interest on the arguments as represented by the shape of the neighborhoods of the application, is the region shown without a center point (i.e., the red neighborhood on the ridge of the wave). Also note that this neighborhood is not the one with the shortest Euclidian distance (in terms of the arguments) to the target application. For this reason, Euclidian distance is not used here. Rather a distance metric that is based on differences in behavior of models at different conditions in terms of the behavior of the models about the nominal conditions (i.e., the points in the neighborhoods) is used, as developed below.

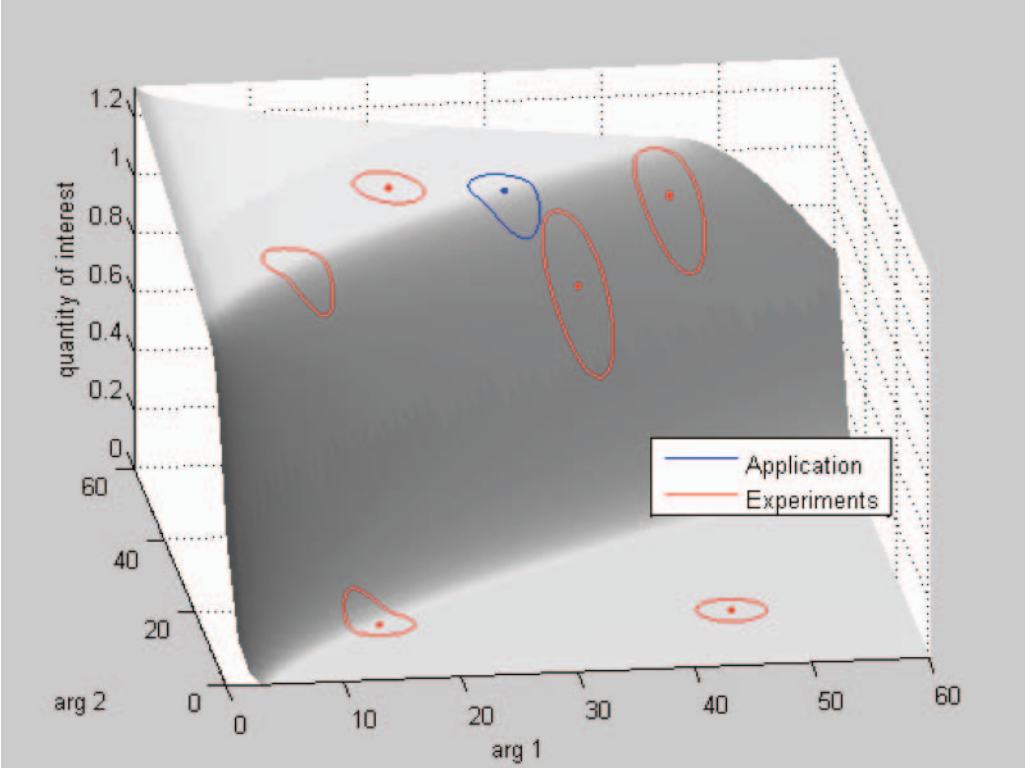

Figure 2. Validation Experiments and Target Application

Following Hills and Hamilton (2010a, 2010b), denote the model or models for the target application by

$$d = g(\mathbf{x},\boldsymbol{\alpha}) \tag{1}$$

where d is the response quantity of interest, g is the corresponding model for the application, $\boldsymbol{\alpha}$ is a vector of model parameters (e.g., thermal conductivity, specific heat), and $\mathbf{x}$ is the vector of independent variables (e.g., time and position). The parameters and independent variables are those for which the application model is significantly dependent over neighborhoods around the nominal values (blue dot in Figure 2). Insignificant model parameters can be set to nominal or mean values for the analysis and don't need to be included in $\boldsymbol{\alpha}$.

An analogous expression for the model of the validation measurements is:

$$\boldsymbol{\gamma}=\mathbf{f}(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{\alpha}_\mathbf{f}) \tag{2}$$

where $\boldsymbol{\gamma}$ represents the vector of validation model outputs simulating measurements for a suite of experiments. These models may represent measurements from suites of experiments using different experimental apparatus to test the model for different classes or combinations of physics, or similar models evaluated at different values of the independent variables and the model parameters, or the same models for repeated measurements at the same nominal conditions. $\mathbf{f}$ represents a vector of models for the suite of validation experiments simulated

measurements, each model evaluated at the conditions of the experiments. $\boldsymbol{\alpha}$ represents the vector of model parameters important to the anticipated application as utilized by Eq. (1). These parameters are analogous to the 'traveling parameters in Romero (2008), and the linking parameters in Mahadevan (2011). The vector $\boldsymbol{\alpha}_{\mathbf{f}}$ represents those parameters in **f** whose effects are significant for the validation models, but not for the anticipated target application model. **x** is the vector of independent variables. If one or more of the models represented by the vector **f** are not a function of the same arguments, then these arguments should appear as dummy arguments in (**x**, $\boldsymbol{\alpha}$) so that the total number and identity of elements in each of these vectors is the same across all models in **f** (hence the notation in (2)). The sensitivity of these models to the dummy arguments will be zero. While components in **x** and $\boldsymbol{\alpha}$ must represent the same arguments, the numerical values for these arguments need not be the same across components of **f** or g.

Hills and Leslie (2003), and Hamilton and Hills (2010a, 20010b) use two approaches to characterize behavior of **f** and g. The first approach uses gradients with respect to the **x**, $\boldsymbol{\alpha}$, evaluated at the nominal or mean experimental conditions for **x**, $\boldsymbol{\alpha}$, for each experimental measurement. The second approach uses sampling based on Box-Behnkin (Box and Behnken, 1960) experimental design. The first approach captures linear behavior, and the second captures quadratic behavior of the models over the neighborhoods of the mean or nominal values for the arguments. In contrast, Latin Hypercube Sampling (LHS) is used here to characterize this behavior where sampling for each f and g are coordinated as follows.

$$g\left(\mathbf{x}_g + \Delta\mathbf{x}_{\boldsymbol{i}}, \boldsymbol{\alpha}_g + \Delta\boldsymbol{\alpha}_{\boldsymbol{i}}\right),\ f_j\left(\mathbf{x}_{\boldsymbol{j}} + \Delta\mathbf{x}_{\boldsymbol{i}}, \boldsymbol{\alpha}_j + \Delta\boldsymbol{\alpha}_{\boldsymbol{i}}, \boldsymbol{\alpha}_{f_j}\right);\ \ i = 1, \ldots, m;\ \ j = 1, \ldots, n \quad (3)$$

Note that the change in the arguments for each of the m samples is equal for all models. This effectively requires that we sample over equally sized neighborhoods (i.e. ranges of $\Delta\mathbf{x}$, $\Delta\boldsymbol{\alpha}$**)**. The sampled results are used to develop a meta-model to relate results from suites of validation experiments (i.e., the multiple red neighborhoods of Figure 1) to the target application (blue neighborhood). The basic approach is to define the meta-model as a weighted linear combination of the validation measurement models as characterized by the samples with the weights chosen such that the meta-model approximates the dependence of the decision variable, d, for the target application model on its arguments in some neighborhood of the nominal parameter values of each individual model. Specifically, we choose the weights of the n measurement models such that the differences or residuals between the left and right hand sizes of the following equation is minimized in some sense as defined in the following sections.

$$g\left(\mathbf{x}_g + \Delta\mathbf{x}_{\boldsymbol{i}}, \boldsymbol{\alpha}_g + \Delta\boldsymbol{\alpha}_{\boldsymbol{i}}\right) - \langle g \rangle$$

$$\cong \sum_{j=1}^{n} w_j \left\{ f_j\left(\mathbf{x}_{\boldsymbol{j}} + \Delta\mathbf{x}_{\boldsymbol{i}}, \boldsymbol{\alpha}_j + \Delta\boldsymbol{\alpha}_{\boldsymbol{i}}, \boldsymbol{\alpha}_{f_j}\right) - \langle f_j \rangle \right\};\ \ i = 1, \ldots, m \quad (4)$$

where $\langle g \rangle$ and $\langle f_j \rangle$ represent expected (mean) values of g and f_j over the m sampled values of i. What is the significance of taking linear combinations of measurement model differences to define a meta-model?

- One can think of the largest linearly independent subset of the sampled validation model vectors as a basis for the target application model. The difference between the right and left hand sides of Eq. (4) allows one to evaluate how well the sampled validation model

vectors can represent the dependence of the non-linear target application on changes in its arguments. If the experiments do not contain the physics as represented by a specific parameter (say emissivity) for the application, then there is no linear combination of f_j that can represent g with zero residuals (differences between left and right sides of Eq. (4).

- If the models for a measurement and for a target application are the same models taken at the same conditions, then the behavior (dependence on the arguments around the nominal arguments) will be the same. If, as the conditions diverge and the models develop different dependence on the arguments, the ability of the measurement models to represent the target application model for the response quantity of interest may decrease.
- The evaluation of the weights is based on the sampled response of the models over a neighborhood with a neighborhood size corresponding to the uncertainty in the model parameters and the target application. Thus one requires the meta-model to be applicable only over a restricted neighborhood in the parameter domain of the model.
- If the weights w_j over several j have the same sign, then this has the effect of stacking of (i.e., adding or averaging) the corresponding measurement models, which reduces the sensitivity of the resulting meta-model to random uncertainty in the model parameters for the corresponding validation models.
- In contrast, if the weights have different signs for several of the measurement models, then one is subtracting effects. Examples of this behavior can result from application variable types that are different from the measurements types. For example, if the application variable is a flux, but the measurements are temperatures, one should expect that the meta-model would contain differences in the temperature models to represent gradients. Another example is for the case that measurement models represents physics that the target application does not. In this case, the meta-model may contain weights of different signs that subtract the effect of this unrepresented physics from the meta-model.

Denote

$$g_{\boldsymbol{i}} = g\left(\mathbf{x}_g + \Delta\mathbf{x}_i, \boldsymbol{\alpha}_g + \Delta\boldsymbol{\alpha}_i\right) - \langle g \rangle\,; \;\; i = 1 \ldots m \tag{5}$$

$$f_{ij} = f_j\left(\mathbf{x}_j + \Delta\mathbf{x}_{\boldsymbol{i}}, \boldsymbol{\alpha}_j + \Delta\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_{f_j}\right) - \langle f_j \rangle; \;\; i = 1 \ldots m, \;\; j = 1 \ldots n \tag{6}$$

Equations (4) through (6) can be written in matrix form as

$$\mathbf{g} \cong \mathbf{F}\mathbf{w} \tag{7}$$

where

$$\mathbf{g} = \begin{bmatrix} \mathrm{g}_1 \\ \vdots \\ g_m \end{bmatrix}; \; \mathbf{F} = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{bmatrix}; \;\; \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \tag{8a, 8b, 8c}$$

Note that **g** represents the sampled behavior around mean conditions of the target application as represented by the blue dot in Figure 1, and column j of **F** represents the sampled behavior for

the model of the validation experiment around the mean conditions (red dots of Figure 1) for measurement *j*. The weighting vector **w** represents the weights to apply to the measurement model vectors to 'best represent' the target application.

There is no requirement that the measurement models represent the same physical quantity, or that the target application model corresponds to those of any of the measurement models. If a target application model depends on an argument that is not reflected in any of the measurement models, then the measurement models will not show any dependence on this argument, leading to a potential decrease in the rank of **F**. Also note that there may be more measurements than samples ($n > m$), or more samples than measurements ($m > n$). Thus **F** can be (and often is) ill-conditioned or singular. Because it is not expected that a suite of validation experiments fully covers the physics and conditions of the application, this ill-conditioning tends to be the rule rather than the exception. A byproduct of ill-conditioned systems in the presence of data and model parameter uncertainty is their solution requires a trade-off between the ability to solve the system accurately and the sensitivity of the resulting solution to uncertainty. The optimum solution will be a balance between the maximum resolution of the meta-model and the sensitivity of the meta-model to uncertainty, as discussed in Hamilton and Hills (2010b). Two approaches are taken to evaluate the weights accounting for this required balance. One is based on the objective function used by Hamilton and Hills (2010b). The second in based on partial least squares (Abdi, 2010).

2.1.1. Objective Function Method

The objective function used here to evaluate the weights is (note that **g**, **w** and **F** used here are the transposes of those used in Hamilton and Hills, 2010b):

$$\mathrm{L} = \theta\, \mathbf{w}^{\mathrm{T}}\, \mathbf{var}(\mathbf{f} - \boldsymbol{\gamma})\, \mathbf{w} + (1-\theta)(\mathbf{g} - \mathbf{w}\,\mathbf{F})^{\mathrm{T}}(\mathbf{g} - \mathbf{w}\,\mathbf{F}); \quad 0 \le \theta \le 1 \qquad (9)$$

The variance of $\mathbf{f} - \boldsymbol{\gamma}$ (the reader is reminded that $\boldsymbol{\gamma}$ is the vector of validation model outputs simulating the validation measurements, with the variance of $\boldsymbol{\gamma}$ representing the uncertainty in the measurements) represents the uncertainty in the difference $\mathbf{f} - \boldsymbol{\gamma}$ due to parameter and measurement uncertainty. Evaluation of this variance is discussed in a later section. Hamilton and Hills do not include the measurement uncertainty in $\boldsymbol{\gamma}$, as their focus was on how well a set of validation experiments resolve the target application in the absence of measurement uncertainty. In the present case, we wish to use these weights to project observed validation measured/predicted differences from validation experiments to a target application response quantity. The measurement uncertainties are included in Eq. (9) so that the resulting meta-model predictions are not overly sensitive to measurement uncertainty. Note that choosing different values for θ results in a trade-off between that ability of the meta-model to represent the target application, and the sensitivity of this the meta-model to uncertainty in the validation differences $\mathbf{f} - \boldsymbol{\gamma}$ over the sampling neighborhoods.

While the minimization of the objective function (9) with respect to the weights can be evaluated through the usual procedure of taking the gradient of the objective function with respect to the weights, setting the results to zero, and solving the resulting system of equation, this approach suffers from the ill-conditioning, especially for small values of θ. The term corresponding to $\mathbf{F}^{\mathrm{T}}\mathbf{F}$

in Eq. (9) (after expansion of the terms in the parenthesis) squares the condition number of **F**, aggravating the ill-conditioning of the system. To avoid this squaring of the condition number, first decompose the variance matrices into upper/lower triangular form using Cholesky decomposition.

$$\text{var}(\mathbf{f}-\boldsymbol{\gamma}) = \mathbf{R}^{\mathrm{T}}\mathbf{R} \tag{10}$$

Define the following quantities:

$$c = \sqrt{\frac{\theta}{1-\theta}} \tag{11}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{F} \\ c\mathbf{R} \end{bmatrix};\ \mathbf{b} = \begin{bmatrix} \mathbf{g} \\ \mathbf{0} \end{bmatrix} \tag{12a, b}$$

where **0** is a *n* by 1 vector of zeros. Note that the least squares solution of

$$\mathbf{A}\mathbf{w} = \mathbf{b} \tag{13}$$

requires the minimization of

$$r^2 = (\mathbf{A}\mathbf{w} - \mathbf{b})^{\mathrm{T}}\,(\mathbf{A}\mathbf{w} - \mathbf{b}) \tag{14}$$

or

$$r^2 = \mathbf{w}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{w} - 2\,\mathbf{b}^{\mathrm{T}}\,\mathbf{A}\mathbf{w} + \mathbf{b}^{\mathrm{T}}\mathbf{b} \tag{15}$$

Using Eqs. (12) in (15) gives

$$r^2 = \mathbf{w}^{\mathrm{T}}\begin{bmatrix} \mathbf{F} \\ c\mathbf{R} \end{bmatrix}^{\mathrm{T}}\begin{bmatrix} \mathbf{F} \\ c\mathbf{R} \end{bmatrix}\mathbf{w} - 2\begin{bmatrix} \mathbf{g} \\ \mathbf{0} \end{bmatrix}^{\mathrm{T}}\begin{bmatrix} \mathbf{F} \\ c\mathbf{R} \end{bmatrix}\mathbf{w} + \begin{bmatrix} \mathbf{g} \\ \mathbf{0} \end{bmatrix}^{\mathrm{T}}\begin{bmatrix} \mathbf{g} \\ \mathbf{0} \end{bmatrix} \tag{16}$$

$$r^2 = \mathbf{w}^{\mathrm{T}}\,(\,\mathbf{F}^{\mathrm{T}}\mathbf{F} + c^2\mathbf{R}^{\mathrm{T}}\mathbf{R})\mathbf{w} - 2\,\mathbf{g}^{\mathrm{T}}\,\mathbf{F}\mathbf{w} + \mathbf{g}^{\mathrm{T}}\mathbf{g} \tag{17}$$

$$r^2 = c^2\mathbf{w}^{\mathrm{T}}\,\mathbf{var}(\mathbf{f}-\boldsymbol{\gamma})\mathbf{w} + (\mathbf{g}-\mathbf{F}\mathbf{w})^{\mathrm{T}}(\mathbf{g}-\mathbf{F}\mathbf{w}) \tag{18}$$

Comparing Eq. (18) to Eq. (9) and noting Eq. (11), we see that

$$\mathrm{L} = (1-\theta)\,r^2 \tag{19}$$

Thus, solving Eq. (13) in a least squares sense is equivalent to minimizing Eq. (9) for a given value of θ.

The meta-model residuals and the sensitivities of the meta-model to the model parameters and data uncertainty can be used to estimate variances.

$$r_1^2 = \frac{(\mathbf{g}-\mathbf{F}\mathbf{w})^{\mathrm{T}}(\mathbf{g}-\mathbf{F}\mathbf{w})}{m-1} \tag{20}$$

$$r_2^2 = \mathbf{w}^{\mathrm{T}}\,\mathbf{var}(\mathbf{f}-\boldsymbol{\gamma})\mathbf{w} \tag{21}$$

Equation (20) is the variance of differences (i.e., residuals between the meta-model and the application model) as estimated from the *m* samples, and Eq. (21) is the variance of meta-model due to uncertainties in the validation differences.

The choice of θ or c

The choice of optimization parameter θ, or the ratio c (Eq. 11), is a trade-off between minimizing the residuals, Eq. (20), or minimizing the variance in the reconstructed meta-model, Eq. (21). In general, this choice is arbitrary, but we choose to use the value of θ or c that minimizes the total variance of the meta-model prediction. That is, we choose the value of θ or c that minimizes

$$r_{total}^2 = r_1^2 + r_2^2 \tag{22}$$

Comparing Eq. (22) to (18), (20) and (21), we see that to minimize r_{total}^2

$$c = \sqrt{m-1} \tag{23}$$

Using this value in Eq. (12a) results in the linear system of equations given by Eq. (13) that can be solved using conventional techniques. The MatLab (2010) linear equation solver '\' is used here.

2.1.2. Partial Least Squares Regression

The second approach to estimate the weights are based on partial least squares regression. Partial Least Squares (PLS) (Abdi, 2010) regression is considered a second-generation regression technique that has several advantages over simple regression.

- PLS is applicable when the number of unknowns exceeds the number of equations. In the following examples, the number of weights (i.e., the number of measurements) exceeds the number of application responses (i.e., number of LHS samples).
- In contrast to the objective function procedure defined above, and to standard regression, PLS utilizes a full matrix of desired application responses **G**. For the present case, the columns of **G** correspond to the different prediction times, while the rows correspond to the different LHS samples. This advantage is manifested when the observed responses contain noise (i.e., analogous to measurement noise). In the present case, the only "noise" that is present in the predicted target application response is algorithmically induced noise, which is small compared to other sources of uncertainty. As a result, the present application does not significantly benefit from the ability to uses a full matrix of target application responses.
- PLS develops latent or hidden variables that best characterize the behavior of the predictors (i.e. the measurement models) and response variables (the target application). PLS utilizes an intermediate step to estimate the latent variable responses through weighted linear combinations of predictor variables (represented by the columns in **F**) so

that each linear combination is chosen to maximize the correlation between the results of the meta-model and the response matrix **G** through a step-wise procedure. The latent variables have an indirect relation to the model arguments for the present application. For the case of validation models linear in the arguments, the latent variables will correspond to linear combinations of the model arguments (i.e., perturbations in these arguments as used here) such that these perturbations best capture that behavior of the application model **G** to the arguments. Note that for many engineering applications, the number of model parameters may number in the hundreds, but the number of model parameters whose variability has a significant impact on a prediction may number in the tens or less. For example, if radiation heat transfer is the dominant transport phenomena, then the model parameters associated with radiation heat transfer may have the most significant impact on the quantities of interest. Because the number of significant model arguments is not always large, the number of latent variables required to capture the response is not large.

- Although the measurement model approach is linear in the transformed vectors, the method can model non-linear effects (see Haenlein and Kaplan, 2004, who discusses the closely related PLS structural modeling approach) by utilizing more latent variables than the number of important model parameters in the numerical models.
- Because the method utilizes a step-wise procedure, the method does not suffer from the confounding effects that can exist with non-step wise procedures for regression when two predictor variables have highly correlated (e.g. canceling) effects on the response of the meta-model.

As with most forms of regression, the PLS method can suffer from over-fitting the data if the number of latent variables is too large. The result of over-fitting is a prediction that is very sensitive to measurement noise. To insure that this is not the case for the present application, an outer loop step-wise procedure (distinct from the step-wise procedure used by the PLS algorithm) is used to select the number of latent variables or components, n_{comp}, utilized. Introduce the following variables and matrices.

$$g_{ik} = g_k\left(\mathbf{x}_g + \Delta\mathbf{x}_i, \boldsymbol{\alpha}_g + \Delta\boldsymbol{\alpha}_i\right);\ \ i = 1 \ldots m,\ \ k = 1 \ldots p \tag{24}$$

$$\mathbf{F}_{noise} = \mathbf{F} + \mathbf{H};\ \ \mathbf{G} = \begin{bmatrix} g_{11} & \cdots & g_{1p} \\ \vdots & \ddots & \vdots \\ g_{m1} & \cdots & g_{mp} \end{bmatrix} \tag{25a, 25b}$$

$$\mathbf{H} = \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ \vdots & \ddots & \vdots \\ h_{m1} & \cdots & h_{mn} \end{bmatrix} \tag{25c}$$

p is the number of prediction times for **g**, and **H** is an mxn matrix with each row containing a random vector whose generating function corresponds to a multivariate distribution (a Gaussian distribution is used here) with a zero mean and variance defined by $\mathbf{var}(\boldsymbol{\gamma})$, representing the measurement uncertainty.

The PLS method is first used to evaluate a vector **β** with $n_{comp} = 1$ for the regression equation

$$[\mathbf{1} \quad \mathbf{F}]\boldsymbol{\beta} \cong \mathbf{G} \tag{26}$$

The MatLab (2010) function *plsregress* is used. The model is now evaluated utilizing $\mathbf{F}_{noise}$ and compared to the **G** to evaluate the impact of noise on the regression.

$$[\mathbf{1} \quad \mathbf{F}_{noise}]\boldsymbol{\beta} \cong \mathbf{G}_{noise} \tag{27}$$

$$\mathbf{D} = \mathbf{G} - \mathbf{G}_{noise} \tag{28}$$

$$e^2 = \sum_k^p \sum_i^m \mathbf{D}_{ik} \tag{29}$$

Equations (26) – (29) are then re-evaluated at $n_{comp} = n_{comp} + 1$. This step-wise procedure is repeated until the new e^2 is larger than the previous value. The value for n_{comp} that results in the minimum for e^2 is used to define the final value of the weights **β**. As was the case for the objective function approach, this approach addresses the trade-off that exists between the ability to resolve the target application model using the validation models and the sensitivity of this reconstruction to measurement uncertainty.

2.2. Evaluation of $\mathbf{var}(f - \gamma)$ for the Objective Function Method

The variance of the difference between the model predictions and the experimental measurements is given by

$$\mathbf{var}(\mathbf{f} - \boldsymbol{\gamma}) = \mathbf{var}(\mathbf{f}) + \mathbf{var}(\boldsymbol{\gamma}) + \mathbf{2cov}(\mathbf{f}, \boldsymbol{\gamma}) \tag{30}$$

If the model predictions are independent of the measurements, the covariance term is zero. Hills and Leslie (2003) assume independence and use the following equation

$$\mathbf{var}(\mathbf{f} - \boldsymbol{\gamma}) = \mathbf{var}(\mathbf{f}) + \mathbf{var}(\boldsymbol{\gamma}) \tag{31}$$

Note that the $\mathbf{var}(\mathbf{f})$ represents the variance in **f** due to parameter uncertainty in the models for the validation experiments. The neighborhood size (range of uncertain parameter values) and corresponding distributions for the validation model parameters will generally be different from those of the target application. As a result, one should evaluate this variance directly for the conditions of the experiments rather than based on the LHS sampling neighborhoods used to construct the meta-model. One can do this through a first order sensitivity analysis as used by Hamilton and Hills (2010a), or by evaluating the variance of sampled values of **f** where the validation models are based on sampled values of the model parameters that represent the parameter distributions of the validation experiments. In the present work, the second approach was applied. The samples developed in the following section were re-used to do this. While the sampling approach requires additional model evaluations, the results are more reliable for the assessment of the present technique.

2.3. Projection of Validation Differences and Projection Uncertainty

The resulting meta-models can now be used to project observed differences from the validation experiments to the target application. Denote g_γ as the meta-model projection of the measurements differences and use Eq. (4) for the objective function approach to evaluate the corresponding results.

$$g_\gamma \cong \langle g \rangle + \sum_{j=1}^{n} w_j \left\{ \gamma_j - \langle f_j \rangle \right\}; \tag{32}$$

If the PLS approach is used, the corresponding equation is

$$g_\gamma \cong \beta_0 + \sum_{j=1}^{n} \beta_j \left\{ \gamma_j - \langle f_j \rangle \right\}; \tag{33}$$

The projected uncertainty will have contributions due to the uncertainty in the model parameters for each validation experiment, the uncertainty in the validation measurements, and the uncertainty in the model parameters for the target application. Sampling can be used to characterize the effect of these uncertainties as follows.

- Generate n_r samples of the model parameters for each validation experiment and the application using the specified parameter distributions for the validation experiments and the application. Note that if two sets of data are from the same validation experiment (i.e., measurement sets taken at different spatial or temporal locations for the same experiment) then the sampled model parameters for each realization should be the same.

- Evaluate the model responses for each of the n_r samples.

- Simulate the effect of additive measurement noise on the validation experiments by randomly sampling such noise from the specified distributions for the measurement uncertainty and then add the sampled noise to the model responses for the validation experiments.

- Use the validation responses, including measurement noise, in Eqs. (32) or (33) to generate n_r samples of the projected validation results for each time.

- Calculate the differences that exist between the application model results and the meta-model projections for each of the n_r samples across all prediction times/locations. This population of differences approximates the uncertainty that one would obtain if all models contain the correct physics; and given the uncertainties in the model parameters and measurements, and the uncertainties associated with the inability of the meta-model to fully represent the target application model.

- Sample these differences with replacement (bootstrapping) and add these boot-strapped differences to the mean values of the meta-model results. Mean values are used because the meta-model was constructed in terms of changes about the means. To maintain the correlation structure between measurements, one should bootstrap an entire response (i.e., all prediction times/locations) as a single sample.

- Evaluate the inter-quartile distances (i.e., locations of the 25 and 75 percentiles of the sampled output distributions) to characterize the uncertainty in the meta-model results.

The number of samples used for the examples in the present study is $n_r = 100$. Note that the above procedure accounts for the additional uncertainties that are associated with the independence that may exist in the parameters for the various validation experiments and the target application. The procedure also uses the distributions for the model parameters for the validation experiments and the application, which may be different than the multi-normal distributions used to develop the meta-model. In contrast, the meta-model was developed by using equal perturbations in the model parameter values, across the suit of validation experiments and the target application; to identify a causal relationship between the models of the validation experiments and the application. The projected uncertainty, as characterized by the above steps, reflects the additional uncertainty associated with using the meta-model for validation experiment and application models that may have different correlation structures and different distributions for the model parameters, than what was used to develop the meta-model.

2.4. Expert Opinion

Often, one must rely on expert opinion to characterize expected differences between experimental results and model predictions. For example, a subject matter expert may state that the difference between a model prediction and a physical observation is expected to be in some range. For the case of a time series of measurements, (i.e., temperature measurements made at some location as a function of time), the expert may state that over some interval of time, the response or the difference between the response and a corresponding model prediction is in some range. For the case of a peaked response, the expert may claim that the difference in time between the peak of a hypothetical measured response and the peak of the predicted response is characterized by some interval, and the difference in the size of the peaks (i.e., maximum predicted and measured temperatures) is also characterized by some interval. Note that both of these uncertainties (time lag between predicted and measured peaks, and differences in maximum value between predicted and observed peaks) are epistemic in the sense that the uncertainties in both, in concept, be reduced through the development of model validation experiments for the physics in question.

In the present work, we represent this epistemic uncertainty through probability density functions that represent plausibility or degree-of-belief (rather than frequency as would be appropriate for aleatory uncertainty). We can make no distinction between epistemic and aleatory uncertainty, and roll-up the effects of these uncertainties concurrently to the target application; or we can use a second probability method in which a meta-model is developed for each epistemic realization of the expert opinion. The development of a meta-model is a post-processing step and requires minimal CPU resources. As a result, a second probability method can be practically applied for the case of expert generated measurements without added sample runs of the underlying computation models representing the physics. We apply the second probability method here for the case of expert opinion.

3. MODELS FOR THE EXAMPLE APPLICATIONS

To illustrate the previously defined methodology, consider the following equation and the initial and boundary conditions:

$$\frac{\partial u}{\partial t} = d\frac{\partial^2 u}{\partial x^2} + c\frac{\partial u^p}{\partial x} \tag{34}$$

$$u(x,0) = \begin{cases} 0.5; & x < 1 \\ x - 0.5; & 1 \le x < 2 \\ 1.5; & 2 \le x < 3 \\ 4.5 - x; & 3 \le x < 4 \\ 0.5; & 4 \le x \end{cases} \tag{35}$$

$$u(0) = u(20) \tag{36}$$

Note that periodic boundary conditions are used for u. When $d \neq 0$, $c = 0$, Eq. (34) represents the diffusion equation; $d = 0$, $c \neq 0$, $p = 1$, the convective equation; $d = 0$, $c \neq 0$, $p = 2$, Burgers equation, and $d \neq 0$, $c \neq 0$, $p = 2$, the diffusive Burgers equation. For the present examples, the true physics of the target application are taken to be correctly represented by the diffusive Burgers equation. Various forms of Eq. (34) will be used to represent the true physics and the models of the physics for the validation experiments. The validation experimental measurements are time history observations of $u(x_{meas}, t)$ for independent experiments at two different measurement locations x_{meas}. The correlation introduced if the two measurement locations are measured from the same experiment can be easily addressed by the present methodology by incorporating this correlation into the variance matrix all of the model parameters for the experiments. Three forms of Eq. (34) are considered, as summarized by the parameter values and their standard deviations listed in Table 1. The parameter p is considered a model specification parameter and takes on the values 1 or 2 with no uncertainty.

Note that Type A1 corresponds to u for the linear convective-diffusive equation, Type A2 corresponds to u for Burgers equation, and Type A3 corresponds to u for the diffusive Burgers equation. Type B3 corresponds to flux per unit area for the diffusive Burgers equation. This flux is given by

$$\text{flux} = cu^p - d\frac{\partial u}{\partial \text{x}} \tag{37}$$

To illustrate the use of these physics/model types, consider the results shown in Figure 3. The string A1A1:A3A3 in the upper plot has the following meaning. The A1A1 string before the colon represents the type of the models used to predict the measurement for the time histories at two locations, the A3A3 string after the colon represents the ‘true physics’ of these two time

Table 1. Physics/Measurements/Predictions

Type: measured/predicted quantity	d	σ_d	c	σ_c	p
Each Validation Model					
A1: u	0.05	0.05d	0.8	0.05c	$p = 1$
A2: u	0	0	0.8	0.05c	$p = 2$
A3: u	0.05	0.05d	0.8	0.05c	$p = 2$
Application					
A3: u	0.05	0.1d	0.8	0.1c	$p = 2$
B3: flux	0.05	0.1d	0.8	0.1c	$p = 2$

histories. The first time history represents the time history observed as the 'wave' passes through $x = 4$, and the second represents the time history observed at $x = 6$. For this example, the measurement models were calibrated or tuned using the data. Note that even though the values for d and c where calibrated to each set of experimental data, the model cannot capture the experimental results, suggesting model form error (i.e. A1 vs. A3). The lower plot illustrates the effect of using the calibrated model to predict the response at $x = 8$. The string A1:A3 indicates that the model used to predict the response is A1, but the true physics of the response is A3 (i.e., model form error). Note that the model for the experiments and the target application (see Table 1) does not posses the non-linear physics induced by p=2, while this physics is present in the experimental results and in the true target application. What can we say about the uncertainty in the prediction of the application response, given the observed differences in the predicted and measured response for the validation experiments? The answer to the above question is the focus of the present work.

The methodology developed here requires that LHS samples be obtained in neighborhoods of the conditions of the validation experiments and the target application. We find that the use of multivariate normally distributed LHS sampling has the advantage of weighting model behavior closer to the nominal conditions of the arguments, due to increased number of samples near the nominal conditions, than obtained using uniformly distributed LHS. This distribution is used only to construct the meta-model. The actual distributions for the uncertain model parameters should be used for the projection of the validation uncertainties to the application, as will be demonstrated later. The samples and thus the size of the sampling neighborhoods for the various models must be the same for the development of the meta-model to insure a causal relationship is maintained in the development of the meta-model. The neighborhood size used for meta-model

development, as characterized variance matrix for the multivariate normal sampling, is given by the sum of variance matrices for the arguments of the individual validation experiments and the target application. Choosing the neighborhood size for those arguments that do not have uncertainty (i.e., p), or for the independent variables, requires judgment. One approach is to choose a range of the independent variables and fixed parameters that results in similar changes in response quantity predictions as those induced by the uncertainty in the parameters. This approach was not used here. Here we arbitrarily choose the neighborhoods on the parameter p and the independent variables x, t as characterized by the standard deviations of Table 2. All parameters are assumed to be uncorrelated (i.e., a diagonal variance matrix), unless model calibration is used. In this case, the variance matrix for each calibrated model parameter set is estimated using a sensitivity analysis as presented later in the examples. The measurement uncertainty is also listed in Table 2.

Table 2. Fixed Argument Neighborhood Sizes and Measurement Uncertainty

neighborhood σ_p	**neighborhood σ_x**	**neighborhood σ_t**	**measurement σ**
Each Validation Model			
$0.05p$	0.05	0.05	0.02
Application			
$0.10p$	0.05	0.05	

All computations are performed using an operator splitting, super-bee flux corrected algorithm for the diffusive Bergers' equation (see Hills et. al., 1994, for a version of the algorithm that uses a different flux limiter). One hundred LHS samples with rejection were used for the meta-model construction neighborhood. Samples were rejected when they resulted in negative values for any of the arguments, occasionally resulting in less than 100 samples. MatLab (2010) was used to develop the algorithms and plot the results.

The experimental data was simulated by randomly generating non-calibrated parameter vectors for use in the true physic model for the experiments, and adding uncorrelated measurements noise (normally distributed) as characterized by the standard deviations listed in Tables 1 and 2.

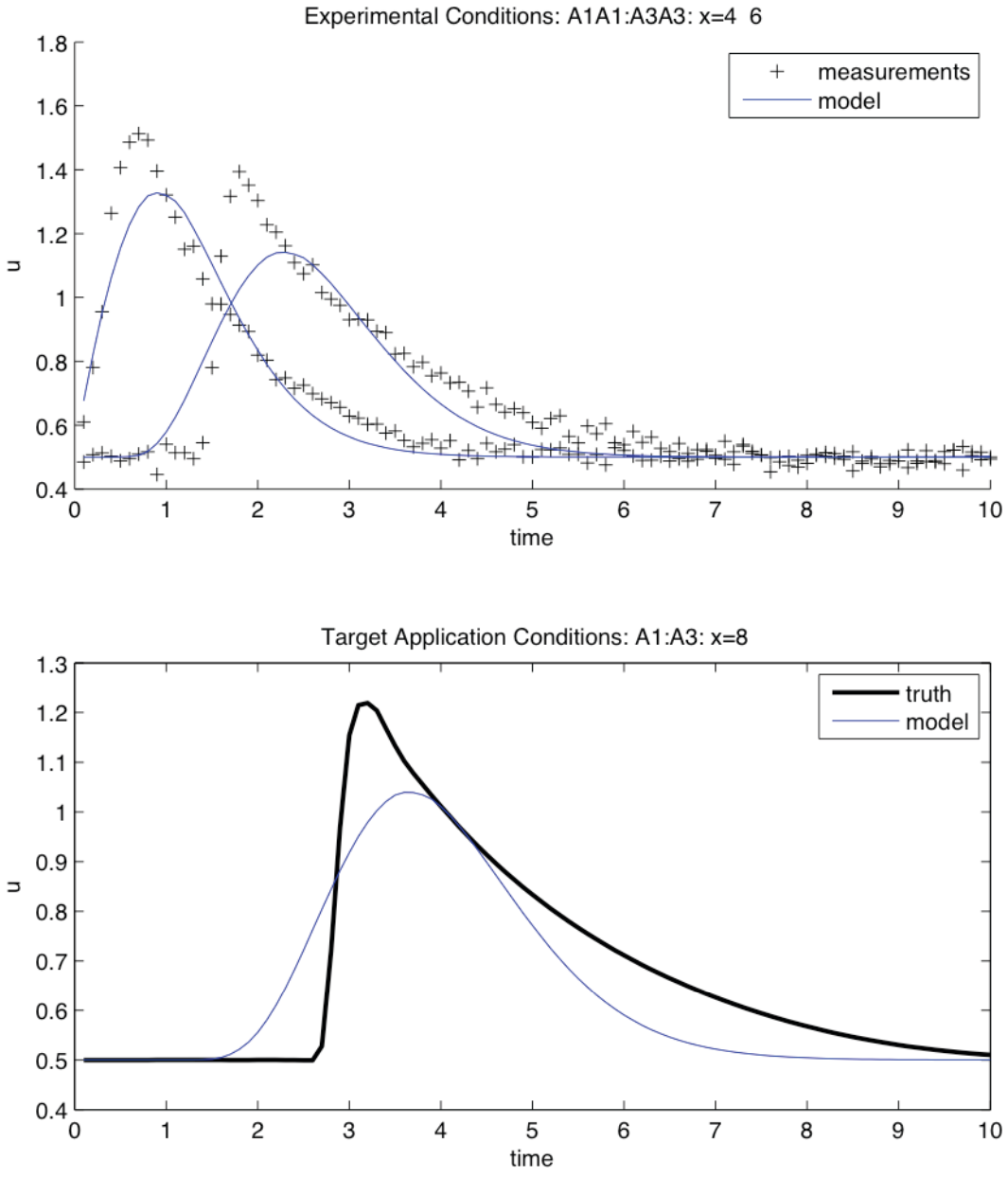

Figure 3. Example Problem: Calibrated Model (Tuned)

4. EXAMPLES OF METHODOLOGY AND ASSESSMENT

A series of example problems are presented to illustrate the application of the method under various scenarios. These examples are designed to show the features and limitations of the methodology. The examples are present through a series of figures, with discussions provided for each. We start with a favorable case.

4.1. Complete Physics Coverage

Figure 4 illustrates (figures provided at the end of this chapter) the application of the methodology for a case where the full physics is included in all of the models and in the experiments. The response quantity of interest is the flux (Eq. (39)) passing through the x=8 plane, and the validation measurements are the observations of u at the x=4 and 6 planes. Black curves are used to represent the true physics of the experiments and application, blue curves represent the underlying models evaluated at the nominal conditions for the parameters, and red curves represent the projection of the validation models to the application (i.e., the meta-model). The distances between the dashed lines represent inter-quartile distance, which result from the sampling over both the true target application and the reconstructed target application results. Inter-quartile results are used as they better reflect the spread of distributions when the distributions are not symmetric about the mean. The estimation of 95% confidence levels would be less precise due to the lack of sufficient samples. The red dashed curves represent the projected 25% and 75% quartiles of the expected uncertainty in the projection due to the uncertainties in the model parameters, data, and meta-model error as developed in the theory. The solid red curve *does not* represent the median of the project results, but the projection of the observed differences between the validation model predictions and the measurements. As a result, one should not expect this solid curve to lie at some average or median distance between the red dashed quartile curves.

The results shown in Figure 4 indicate that the objective function method and the partial least squares (PLS) method both provide estimates of the projected quartiles that capture most of the mean behavior of the target application. Note that the peak of the true response cannot be fully captured by the meta-models due to the differences in validation experiment locations and the target application location, and due to parameter uncertainty and measurement noise. Note also that there is a small shift in the data from the mean model results (upper plot). This shift may simply be due to a shift in the parameters of the true experimental conditions from the nominal parameter values. One can calibrate these parameters and investigate this effect. Figure 5 shows these results. The experimental results from the two measurement locations were assumed independent. As a result, a separate calibration was performed on each set of data. The act of calibration results in correlation between the estimated parameters. A first order sensitivity analysis was used to estimate the corresponding variance matrix for each experimental location. Denote $\mathbf{X}$ as the sensitivity matrix corresponding to a validation model at the calibrated values for the parameters for a location. The MatLab (2010) routine *lsqcurvefit* was used for the calibration, and this routine also provides an estimate of the sensitivity matrix. The resulting variance matrix for the parameters can be estimated from the sensitivity matrix $\mathbf{X}$ as follows:

$$\mathbf{X} = \boldsymbol{\nabla}_{\boldsymbol{\alpha}}\mathbf{f} \tag{38}$$

$$\mathbf{A} = \text{inv}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\ \mathbf{X}^{\mathrm{T}} \tag{39}$$

$$\mathbf{A} = \text{pinv}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\ \mathbf{X}^{\mathrm{T}},\ \text{if singular} \tag{40}$$

$$\mathbf{Var}(\boldsymbol{\alpha}) = \sigma_{res}^2\ \mathbf{A}\ \mathbf{A}^{\mathrm{T}} \tag{41}$$

where σ_{res} is the estimated standard deviation of the regression residuals. For the case of calibration, the resulting calibrated parameter values and the corresponding variance matrices are used for the meta-model development and the projected validation uncertainty, rather than those presented by Table 1. The calibrated results are also used in the specification of the neighborhoods and the validation differences, but the calibration constants are not used to revise the target application model parameter values for this example. While an over-parameterized model may result in a calibration with small σ_{res}^2, the corresponding uncertainties in the estimated parameters will be large due to $\mathbf{A}\ \mathbf{A}^{\mathrm{T}}$. In other words, over-fitting data through a many parameter model will result in large uncertainties in the estimation of the parameters.

The results of this approach are shown in Figure 5. The calibrated results show less shift in projected results. Because we are *not assessing model validity*, but are projecting differences to a target application, updated values for these parameters should be used to better reflect the conditions of the experiments. *All remaining examples will be based on calibrated results. The calibration is not applied to the parameters for the target application unless noted.* The number of latent variables or components used for this example was 14. In contrast, the number of arguments used to generating the **F** and **G** matrices was 5 (*d*, *c*, *p*, *x*, *t*). More latent variables are required because the meta-model is a linear combination of validation model responses whereas the computational model is non-linear. The non-linearity of the computational models is captured, within measurement noise, by the 14 latent variables and corresponding vectors for the PLS meta-model.

A more detailed assessment of the meta-model can be performed through an assessment of the samples as shown in the upper plot of Figure 6. The scatter plots show the meta-model predictions versus the application model predictions for all of the LHS samples for all prediction times used for the meta-model estimation. Note that the PLS method shows somewhat less scatter about the 45 degree line, indicating a better approximation. The corresponding inter-quartile distances as a function of time for the corresponding samples are shown in the lower plots of Figure 6. The PLS method does a significantly better job of capturing these bounds.

Figure 7 shows first order sensitivities of the application model and meta-model to the model parameters. These sensitivities can be obtained though least squares solutions of the following equation for **S**.

$$\mathbf{S}\ \Delta\boldsymbol{\alpha} = \Delta\mathbf{g} \tag{42}$$

Here we take $\boldsymbol{\Delta\alpha}$ and $\boldsymbol{\Delta g}$ as differences from their mean values to maintain consistency with the meta-model approaches. The $\boldsymbol{\Delta\alpha}$ are those used to generate the meta-model. The least squares solution of Eq. (42) for $\mathbf{S}$ gives

$$\mathbf{S}^{\mathrm{T}} = (\Delta\boldsymbol{\alpha}\Delta\boldsymbol{\alpha}^{\mathrm{T}})^{-1}\,\Delta\boldsymbol{\alpha}\Delta\mathbf{g}^{\mathrm{T}} \tag{43}$$

The sensitivities can be scaled to represent the relative impact of parameter variability on a predicted response. The square root of the diagonal elements of the cov($\boldsymbol{\alpha}$) defines a scale for the variability of the parameters for the neighborhood that was used to build the meta-model. The corresponding scaled sensitivities are define as

$$\mathbf{S}_{\boldsymbol{scaled}} = \mathbf{S}\begin{bmatrix} \sigma_{\alpha_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{\alpha_p} \end{bmatrix} \tag{44}$$

where p is the number of model parameters for the application and the σ are corresponding square roots of the diagonal elements of cov($\boldsymbol{\alpha}$) .

Applying Eq. (44) to both the application model and the meta-model, results in 3 columns for $\mathbf{S}_{\boldsymbol{scaled}}$ for the application and the meta-models, corresponding to model parameter sensitivities in c, d, and p as a function of time. The results are shown in Figure 7. Note that the sensitivity analysis indicates that the meta-model from the PLS method better captures the scaled sensitivities of the application model. Based on the results of Figure 6 and 6, the PLS method appears to provide the superior meta-model.

Effect of Measurement Noise

The presence of measurement noise in a suite of validation experiments degrades the ability to represent the target application because the measurement noise masks some of the physical response of the tested system. To investigate this effect, one can repeat the previous example with the standard deviation characterizing measurement noise set to zero. Unfortunately, the Cholesky decomposition represented by Eq. (10) for use in the objective function method fails if the variance matrix is not computationally positive definite. While one can use an incomplete decomposition for this case, we chose to add a very small amount of noise to maintain an algorithmically positive definite variance matrix. Specifically we reduced the measurement noise standard deviation by a factor of 10^{-6} to a value of $\sigma = 2\text{x}10^{-8}$. The corresponding results are shown in Figures 7 though 9. Note that the ability of the meta-model to represent the target application improves dramatically (see Figure 9) when the PLS method is used. This is because the decrease in measurement noise allows one to use 85 terms (latent variables) rather than the 14 used when more noise is present. The uncertainty of the projected differences is larger for the PLS method. We suspect that this is due to an over-fitting of the results (i.e., 85 rather than 14 terms, where 85 corresponds to the number of non-rejected samples used to construct the meta-model) resulting in projections that are sensitive to unaccounted-for numerical noise, such as that associated with the lack of grid convergence and due to the switching effects of the flux limiter. In contrast, the objective function method does not reflect this increased sensitivity. The take-away from this exercise is that while the PLS meta-model can represent the target application

very well in the extreme case of almost no measurement noise, the results can be negatively affected if other sources of noise are present and not accounted for.

4.2. Validation Experiments with for Different Physics

In developing a validation database, one cannot always perform experiments that each represent the full physics of the application. Consider the case for which one validation experiment addresses one type of physics and a second address another, both of which address the complete physics of the application. While this approach covers the physics of the application, the validation experiments may be at different conditions than the application, and the validation experiments do not reflect the coupling that may exist between the two types of physics. This situation is reflected in the results of Figure 11 through 12. Note that the first validation experiment corresponds to the convective-diffusive equation (A1), the second validation experiment corresponds to Burgers' equation without diffusion (A2), and the application corresponds flux for Burgers' equation with diffusion (B3). Comparison of these results with those of Figures 4 through 6 indicates that there is very little degradation, if any, in the ability to represent the target application with the meta-model. The most significant difference is that 15 latent variables are used to build the meta-model for the mixed physics case whereas 14 are used for the non-mixed physics case. However, it is not clear whether this lack of sensitivity to mixed or separate effects of the meta-model is for this example only.

4.3. Known but Untested Physics

Consider the case for which the physics of the application is known and included in the model, but for which physics is not included in the validation experiments. We may expect a degradation of the ability of the experiments to represent the target application though a degradation in the performance of the meta-model. To illustrate this case, consider experiments that test only the convective-diffusive equation (A1) but not the non-linear effects of the diffusive-Burger's equation (A3 or B3). The results for this example are shown in Figures 13 through 15. The results for Figure 14 indicate that the both of the methods perform slightly worse than when all of the physics is contained in the experiments (compare Figure 14 to Figure 5). However, the results shown in Figures 14 and 15 illustrates that the performance, as reflected in the scatter plot and the quartile plots are significantly worse. The meta-models cannot capture the quartiles of the target application model as reflected in the lower plot of Figure 15. The sensitivity results shown in Figure 16 indicate that the parameter p sensitivity cannot be represented by the meta-models. The lack of coverage for p is expected since the validation experiments do not include the physics associated with this parameter. The results of Figures 14 and 15 illustrate the impact of incomplete physics coverage of the application by the validation experiments and strongly suggest that experiments to address this component of the missing physics should be developed. Not doing so will result in the inability to represent the model dependence of the target application due to the non-linear effects of p.

4.4. Unknown but Tested Physics

The next case considered is for a physics model that does not contain the correct physics for the validation experiments or the target application, but the experiments do represent the correct

physics. This example would be the case when we simply do not know that the physics is important, but we are fortunate that the correct physics is represented in the validation experiments. First consider the simpler case for which the application response variable for the application is the same as that measured for the validation experiments. Figure 17 illustrates this example. Note that the results of the validation experiments clearly indicate that some important physics is not in the validation models even though the model for the validation experiments is calibrated. The projected results are especially poor. In engineering, we often calibrate a model and apply these calibration constants to the application. Because such calibration is used to compensate for missing physics in the model, and not to simply to adjust the model to the correct conditions of the experiment, we will make a distinction and denote this type of calibration as **tuning**. To illustrate the dangers of tuning, a series of examples are presented.

4.4.1. Model Tuning

The first step often taken is to calibrate the parameters and apply the calibrated parameters to the model for the anticipated application (i.e., tuning). Figure 18 illustrates this approach. The calibrated parameters for the experiments at $x = 6$ were used for the application since this experiment is at conditions closer to the application. The results presented in Figure 18 indicate that the tuned model better captures the target application response, but the methodology does not capture the uncertainty. Note that the projected differences do provide a more correct response at intermediate and late times but possess an oscillatory behavior prior to the arrival of the true front. A possible cause for this behavior is due to the forcing of one model that possesses one type of behavior to represent another type of behavior. For the present example, the area under the $u(t)$ curve for the linear convective-diffusion at some location can be expected to the same as the area and other locations as long as the time period captures the entire moving wave at both locations (i.e. due the relations between conservation of mass and flux for the linear diffusive convection equation). In contrast, the corresponding area for the diffusive Burgers equation will not be a constant because the flux is non-linear in u. As a result, the oscillations in the project validation differences may be due to this contradictory behavior.

4.4.2. Expanded Uncertainty

How does one account for the additional uncertainties associated with the fact that the validation models do not represent the true physics of the experiments. A typical approach is to expand the uncertainty on the model parameters to reflect this lack of correct physics. For example, in Bayesian updating, lack of agreement between the measurements and the tuned model results in a heavier weighting of the prior in the update, resulting in a broader and thus more uncertain a-prior likelihood for the parameters. The choice of the prior, however, can be is very tricky when one recognizes that the model does not represent the true physics. The appropriate way to interpret the results is to not consider the resulting a-prior uncertainty as a physics-based uncertainty in the parameters, but to consider the uncertainty as a regression uncertainty in estimating the parameters for this model. Romero (2006, 2007, 2008) expands the uncertainty in selected parameters common to the validation experiments and the target application sufficiently to include the observed differences when the models are applied to the validation experiments at some level or probability (say 95%) or through an expanded interval representation of the parameter. One must choose which parameter or subset of parameters to use for this expansion.

An alternative approach is use here for expending the uncertainty. We consider the original distribution of the parameters and the tuned distribution to be equally likely. Note that this approach, like the previous discussed approaches, is simply tuning a model that is missing physics to best represent the observations, without full physical justification. The hope is that the missing physics does not significantly effect the application of the model. As in the previous case, the values for the tuned parameters will still be considered the best estimate of he tuned parameters (i.e. not the mean or median of the combined distributions). The resulting variance of two equally weighted distributions, with separation in the individual population means of the parameters, is given by

$$\mathbf{V}_{\alpha,expanded} = \left(\mathbf{V}_{\alpha} + \mathbf{V}_{\alpha,est}\right)/2 + \text{diag}\{(\langle\boldsymbol{\alpha}\rangle - \langle\boldsymbol{\alpha}_{est}\rangle)^2\}/4 \tag{45}$$

where $\langle\boldsymbol{\alpha}\rangle$ are the parameter values corresponding to those listed in Table 1, and $\langle\boldsymbol{\alpha}_{est}\rangle$ are those estimated from the tuning process. The term diag{ } indicates a diagonal matrix with the diagonal elements given by the argument vector.

The results of using the tuned parameters and the expanded uncertainty for both the development of the meta-model and the projection of the validation results to the target application are shown in Figure 19. Note that the 25% and the 75% quartiles of the PLS meta-model predictions do capture much of the true behavior of for the application, whereas the objective function meta-model quartiles do not. The PLS meta-model utilizes 25 latent variables rather than the 5 used in the previous results. This is a reflection that more latent variables are required to capture the non-linear behavior of the models over the expanded neighborhoods. A side effect of the increase in the number of latent variables is that the estimation of more latent variables requires more measurements. In other words, if significant model form error exists, more experimental data is required to capture the effect of this model form error.

The corresponding assessment and sensitivity plots are shown in Figures 19 and 20. The results of Figures 19 and 20 indicate that the meta-model does a good job of utilizing the wrong models for the physics of the validation experiments to represent the wrong physics of the application.

In the above example, the quantities measured from the validation experiments were the same as the target application response variable. Figure 22 illustrates the results when flux is predicted, but not measured. Note that the results are considerably different than those of the true application flux. For this case, one simply cannot represent the behavior of flux of *u* using a model tuned based on measurements of *u*. This illustrates a danger in using tuned models that contain the wrong physics to predict behavior at other conditions, especially when the predicted response quantity is different from that used for model tuning.

4.5. Unknown and Undetected Physics

The next case represents the worst-case scenario of physics that is unknown, not included in the models, and not represented in the validation experiments. In this case, the methodology has no information that can be used to adjust or even assess this effect. The results of Figure 23 show that the projected differences have no relation to true behavior.

4.6. Expert Opinion

The final case considers expert opinion as characterized by the following statement.

If we were to perform a hypothetical experiment corresponding to model A2 measured at x = 6, we expect the peak to be located somewhere in the time interval t = [1.0, 2.5], with a peak value in the interval u = [1.5, 1.7].

Note that this statement contains uncertainty in both the location of the peak and in the peak value. Because the expert statement represents lack of knowledge and is reducible through experimentation (epistemic uncertainty), a second probability method is used. Here we represent the plausibility of the peak time as uniform over the interval [1.0, 2.5] and the plausibility of the peak values as uniform over the interval [1.5, 1.7]. The two uncertainties are considered independent. Ten samples from these distributions are shown in the upper plot of Figure 24 with the results for each of these ten realizations of the 'expert measurements' shown in lower two plots of the figure. The validation models were calibrated once, excluding the expert data and applied to all epistemic realizations. No model tuning was performed (i.e. calibrated parameters from the validation experiments were not applied to the application).

Note that the model results for $x = 6$ are lower than those 'measurements' based on the expert opinion for this experiment (upper plot). There is considerable shift in the location of the peaks in the projected responses due to the significant uncertainty in the expert opinion realizations for a hypothetical validation experiments. The range of peak heights is less as the uncertainty in this range was 0.2. The heavy dashed lines represent the quartiles of the projected uncertainties, assuming that the model is correct and ignoring the expert input. It is clear from the results that the projected results often lie outside the quartiles, suggesting that the data sampled from expert opinion is very inconsistent with the A2 validation model results.

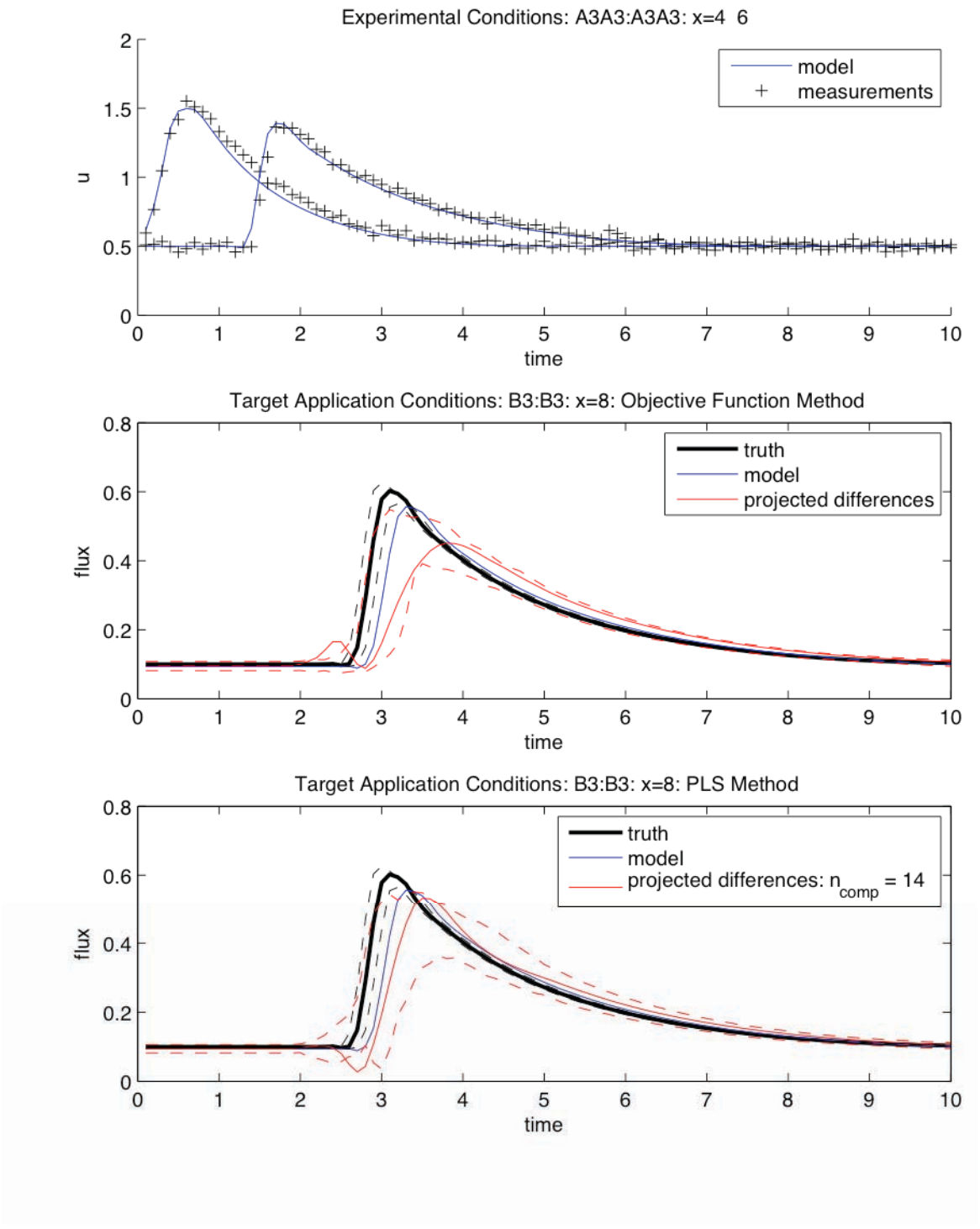

Figure 4. Fully Resolved Physics: Not Calibrated

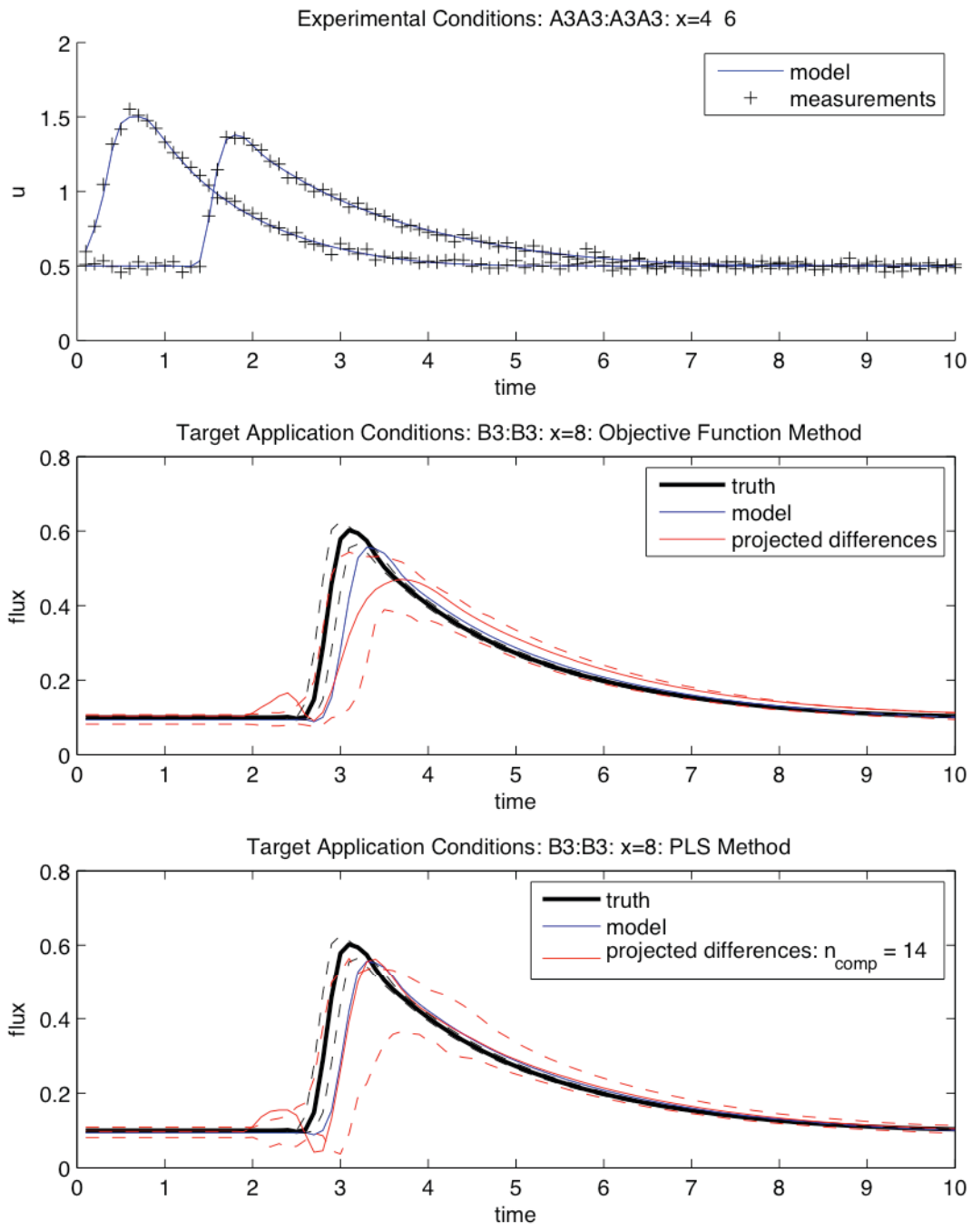

Figure 5. Fully Resolved Physics: Calibrated

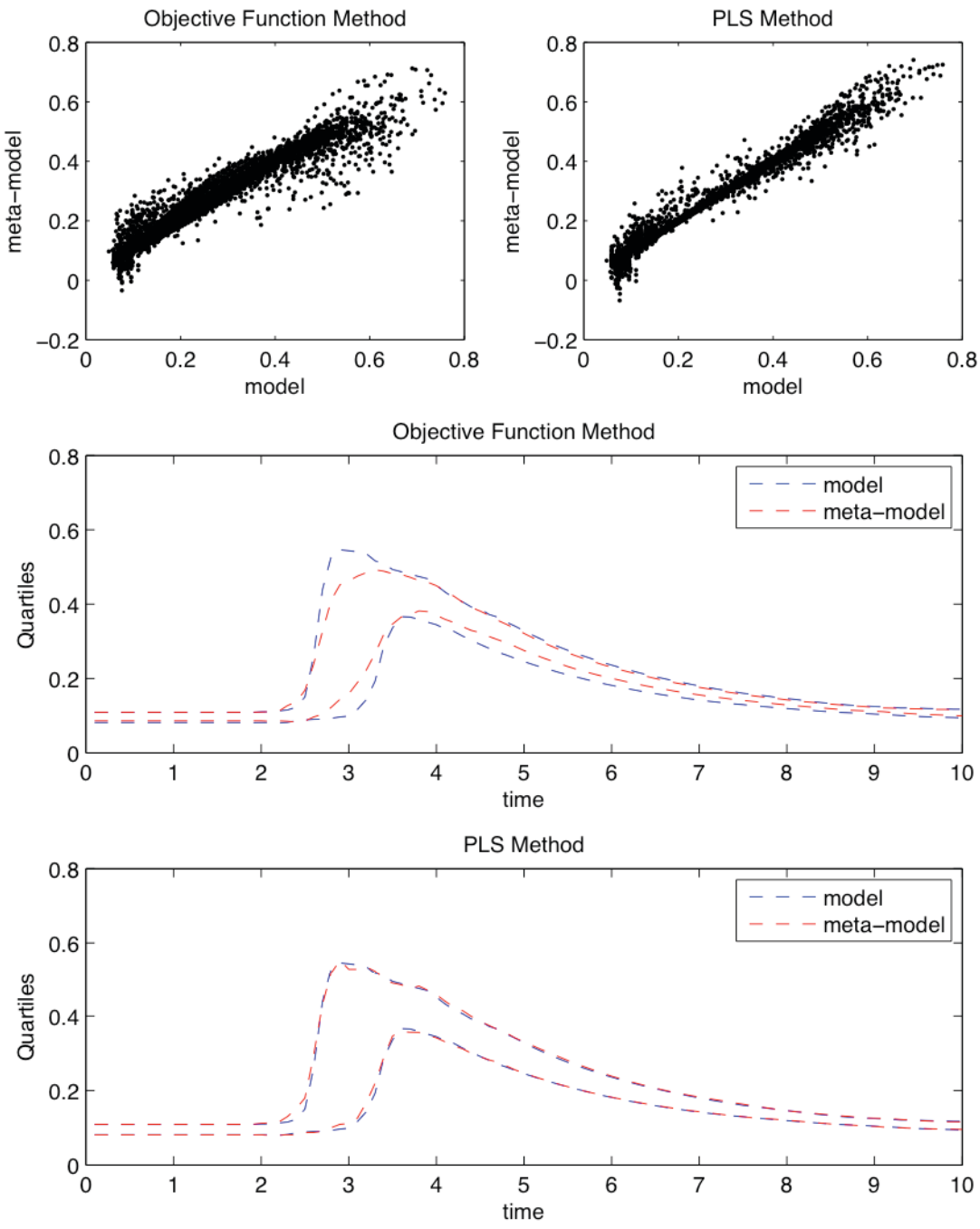

Figure 6. Fully Resolved Physics: Assessment - Calibrated

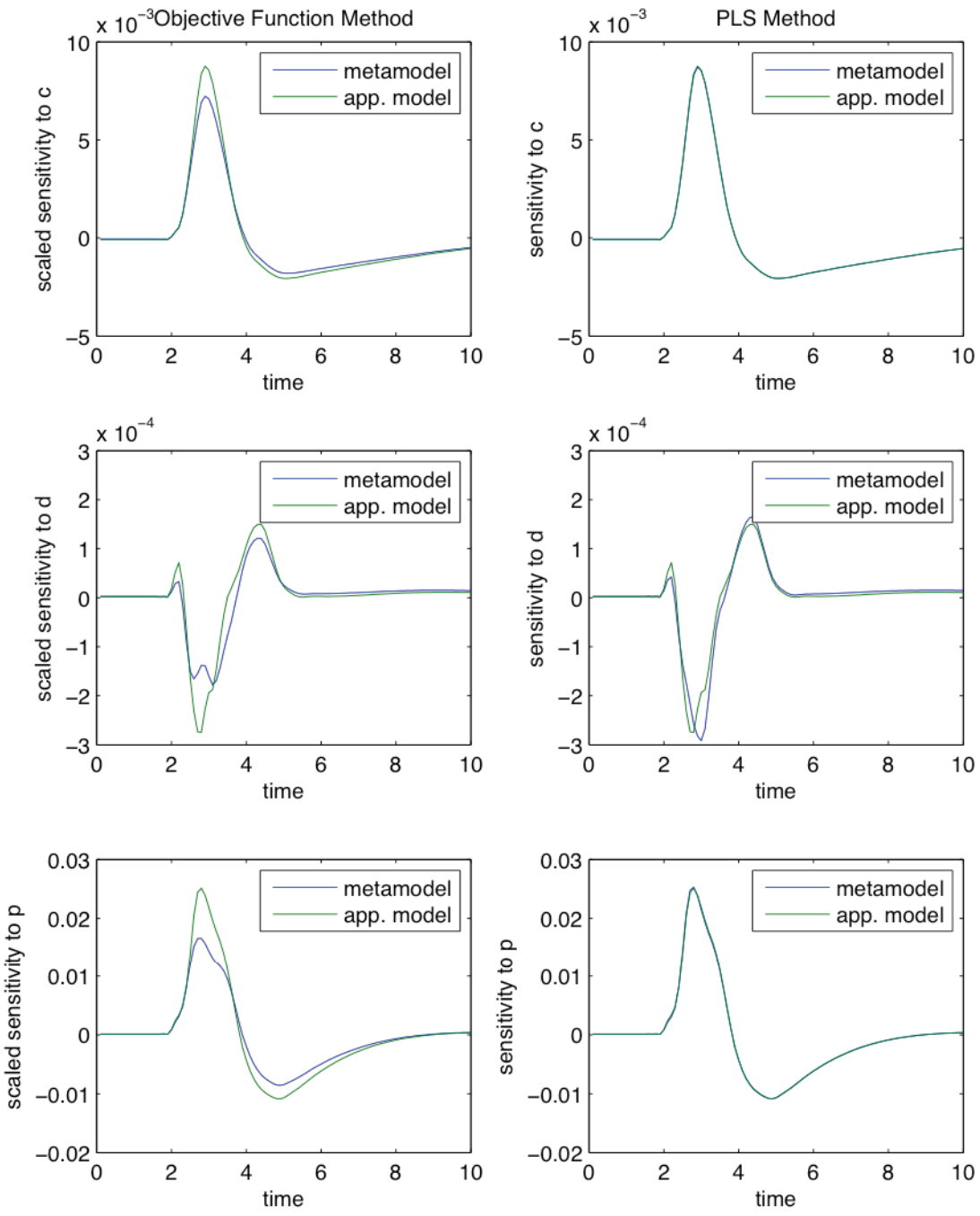

Figure 7. Fully Resolved Physics: Sensitivity - Calibrated

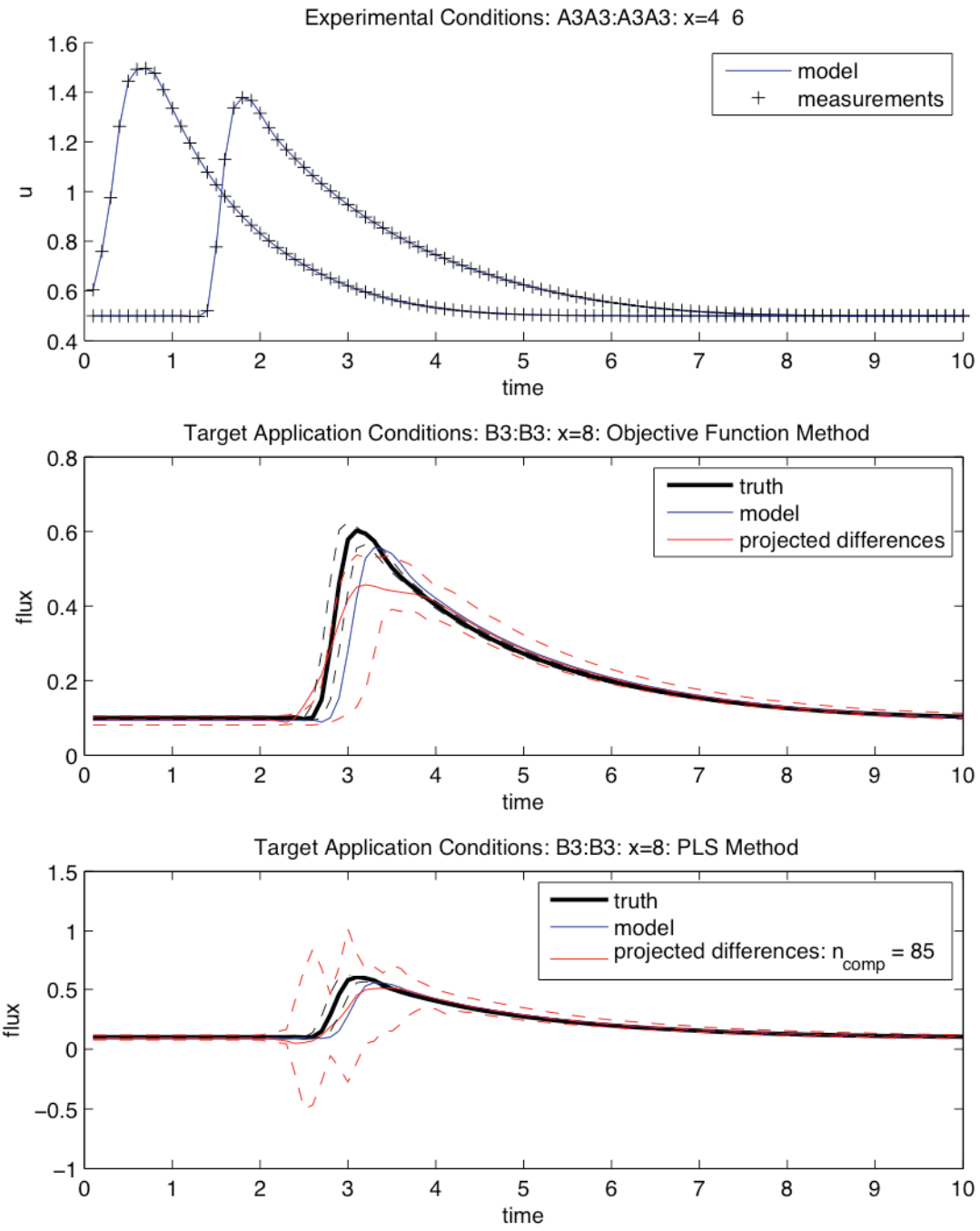

Figure 8. Fully Resolved Physics: Calibrated, Measurement uncertainty = 10^{-6} x table values (sigma = 2 x 10^{-8})

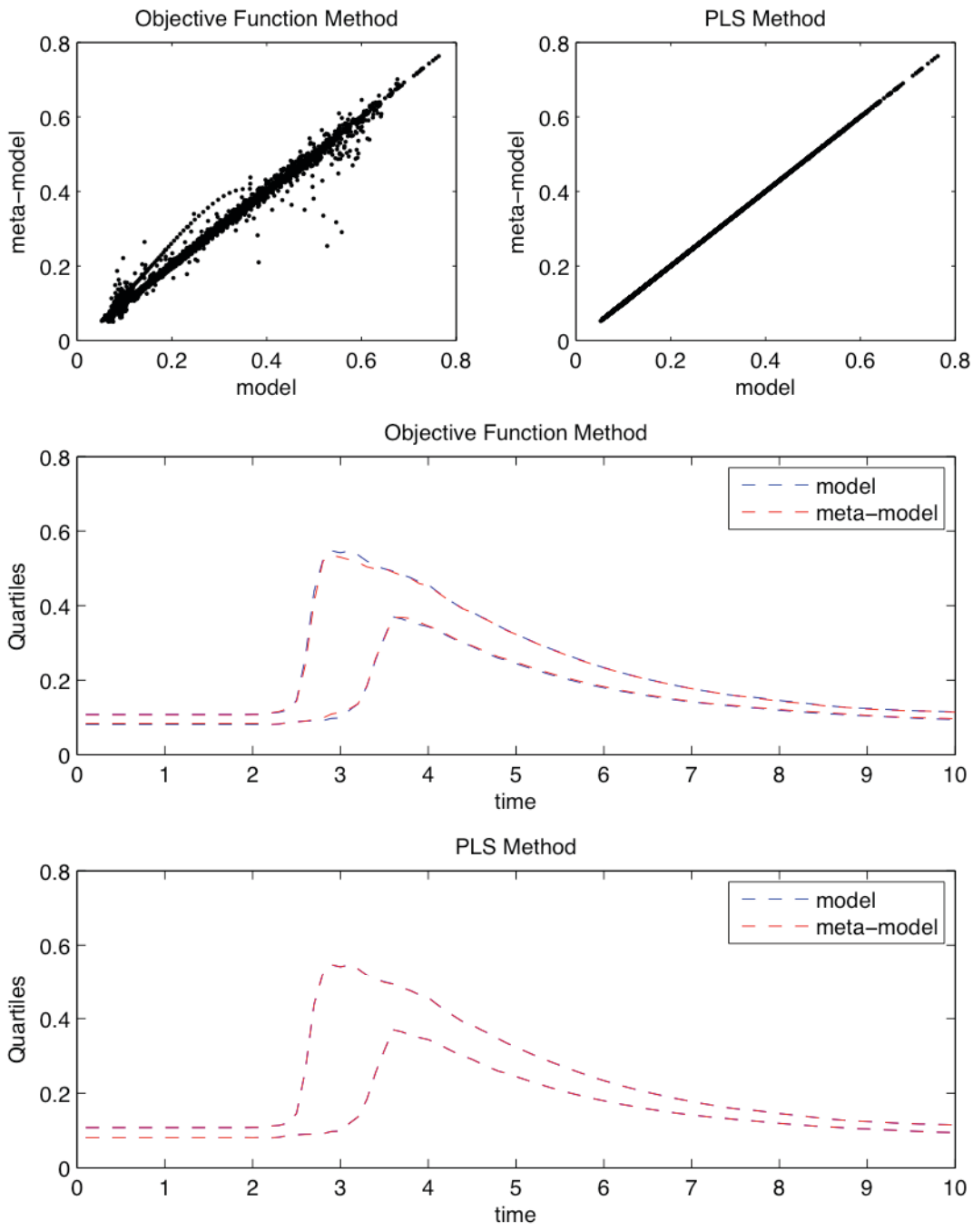

Figure 9. Fully Resolved Physics: Assessment - Calibrated, Measurement uncertainty = 10^{-6} x table values (sigma = 2 x 10^{-8})

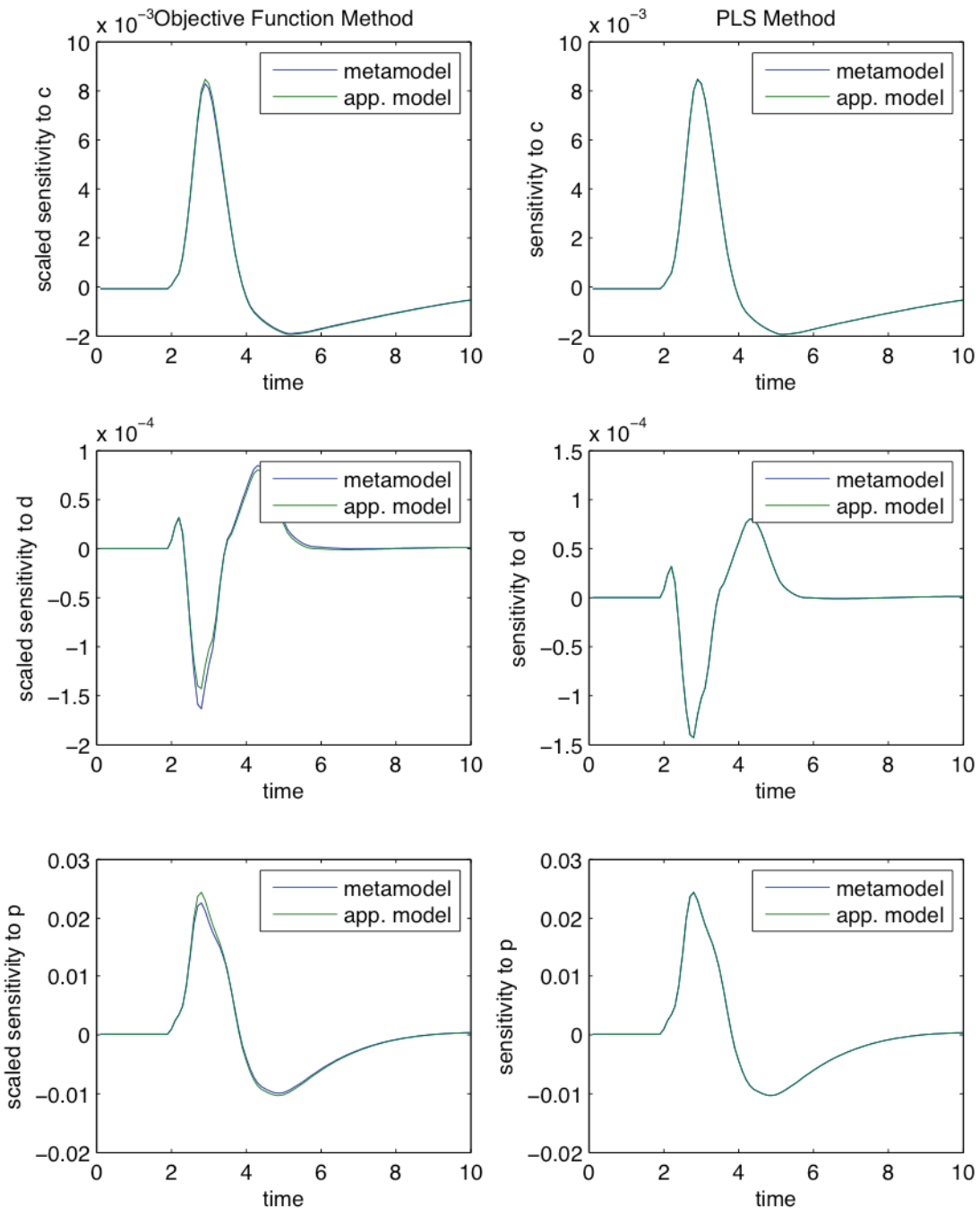

Figure 10. Fully Resolved Physics: Sensitivity - Calibrated, Measurement uncertainty = 10^{-6} x table values (sigma = 2 x 10^{-8})

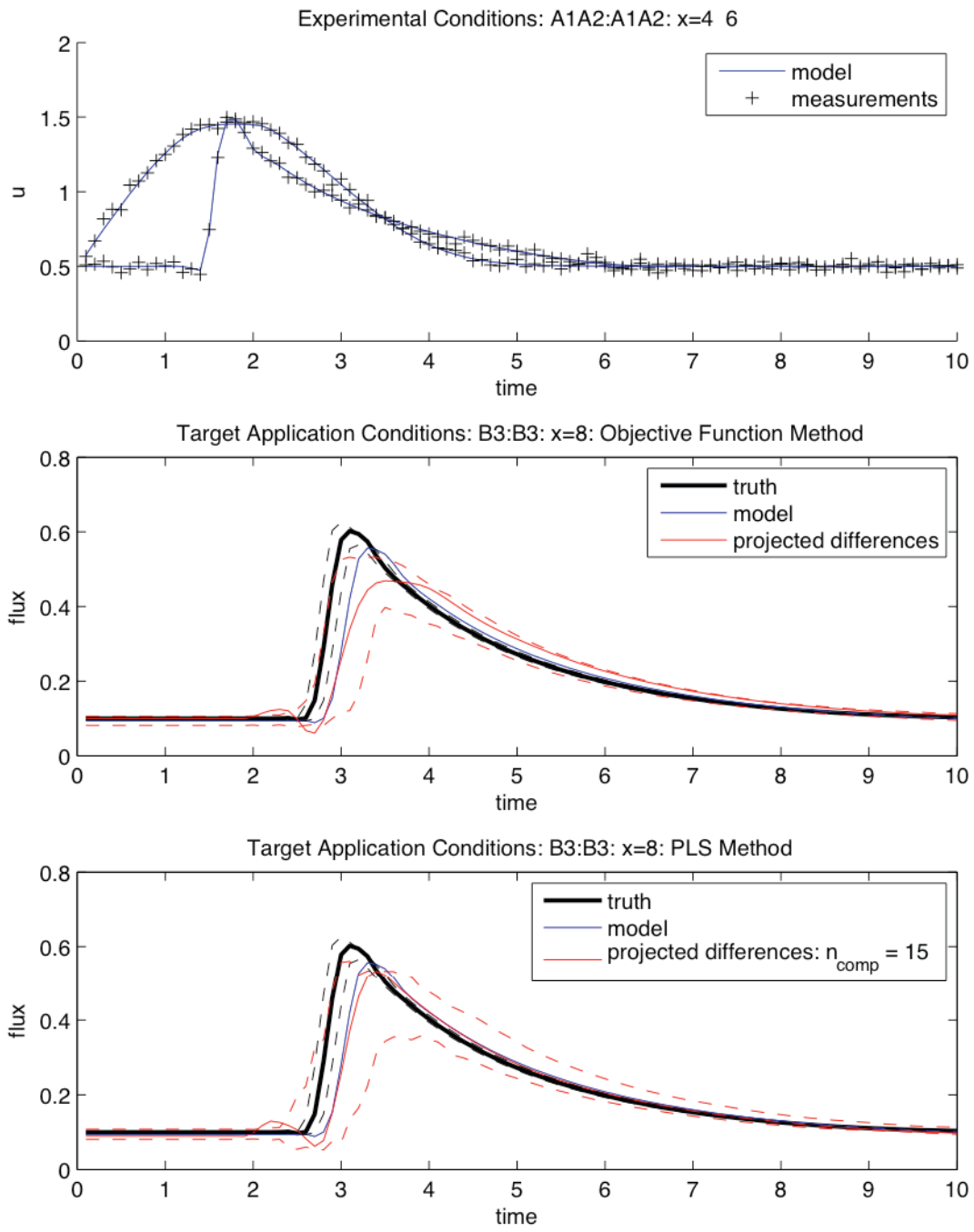

Figure 11. Mixed Physics

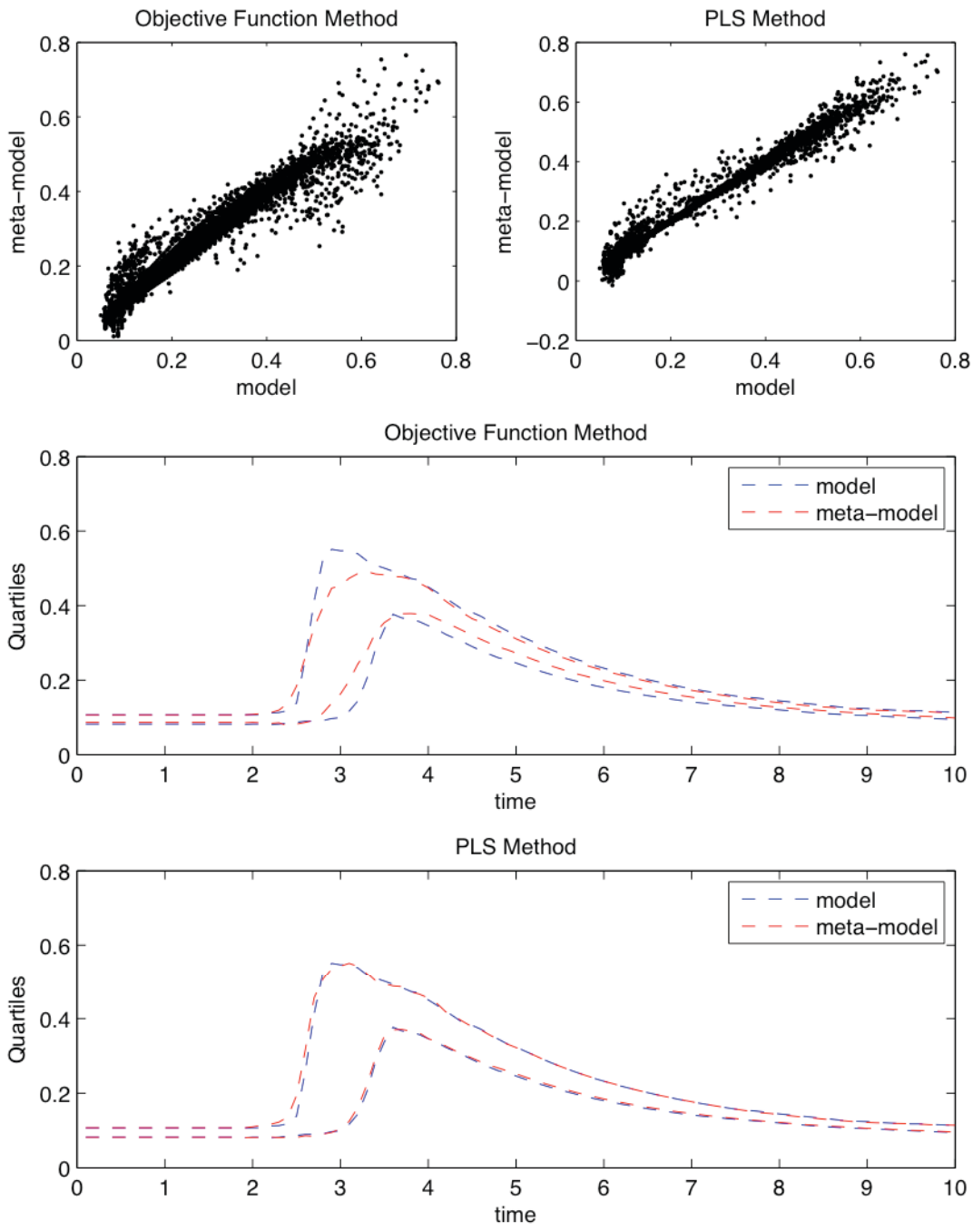

Figure 12. Mixed Physics: Assessment

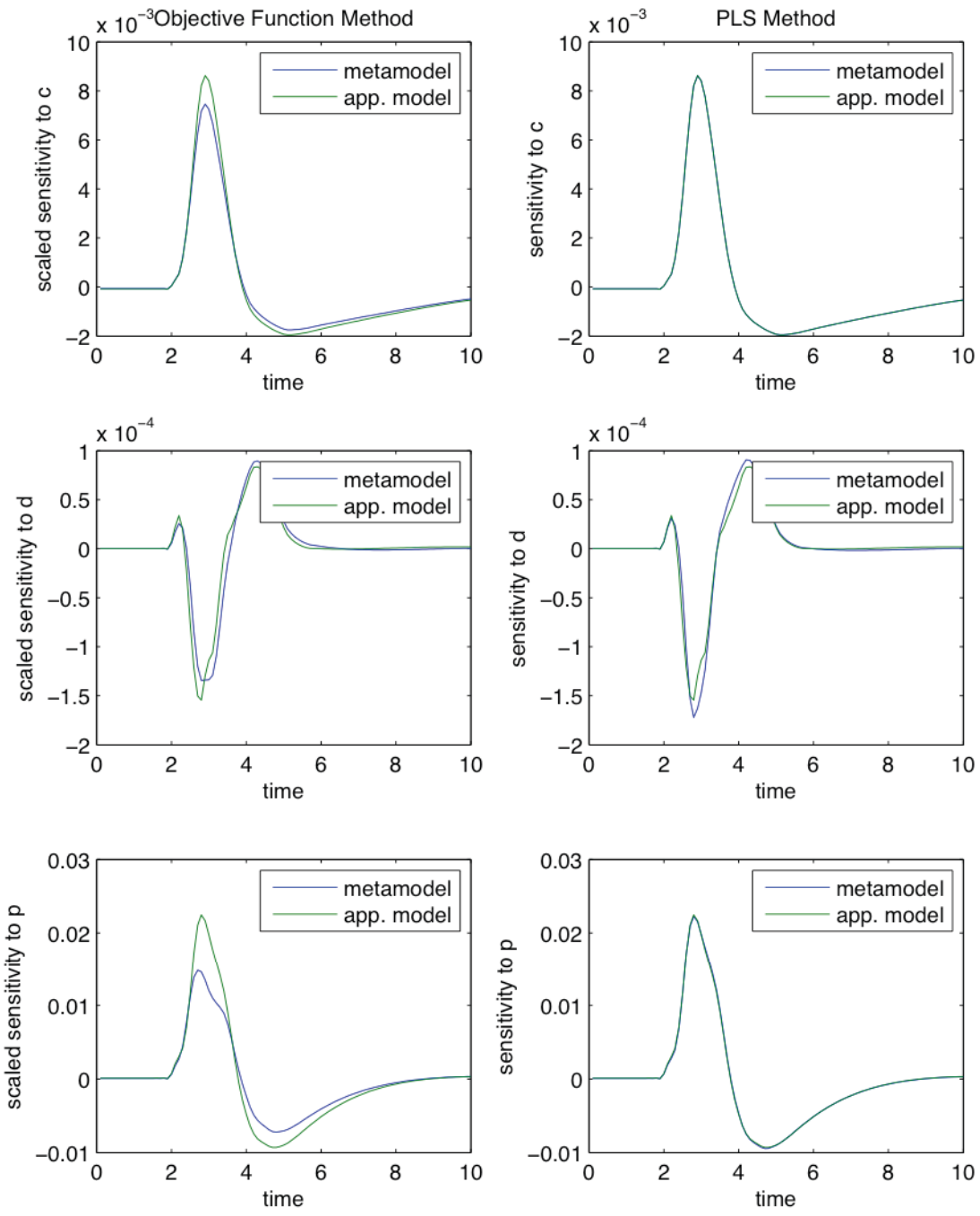

Figure 13. Mixed Physics: Sensitivity

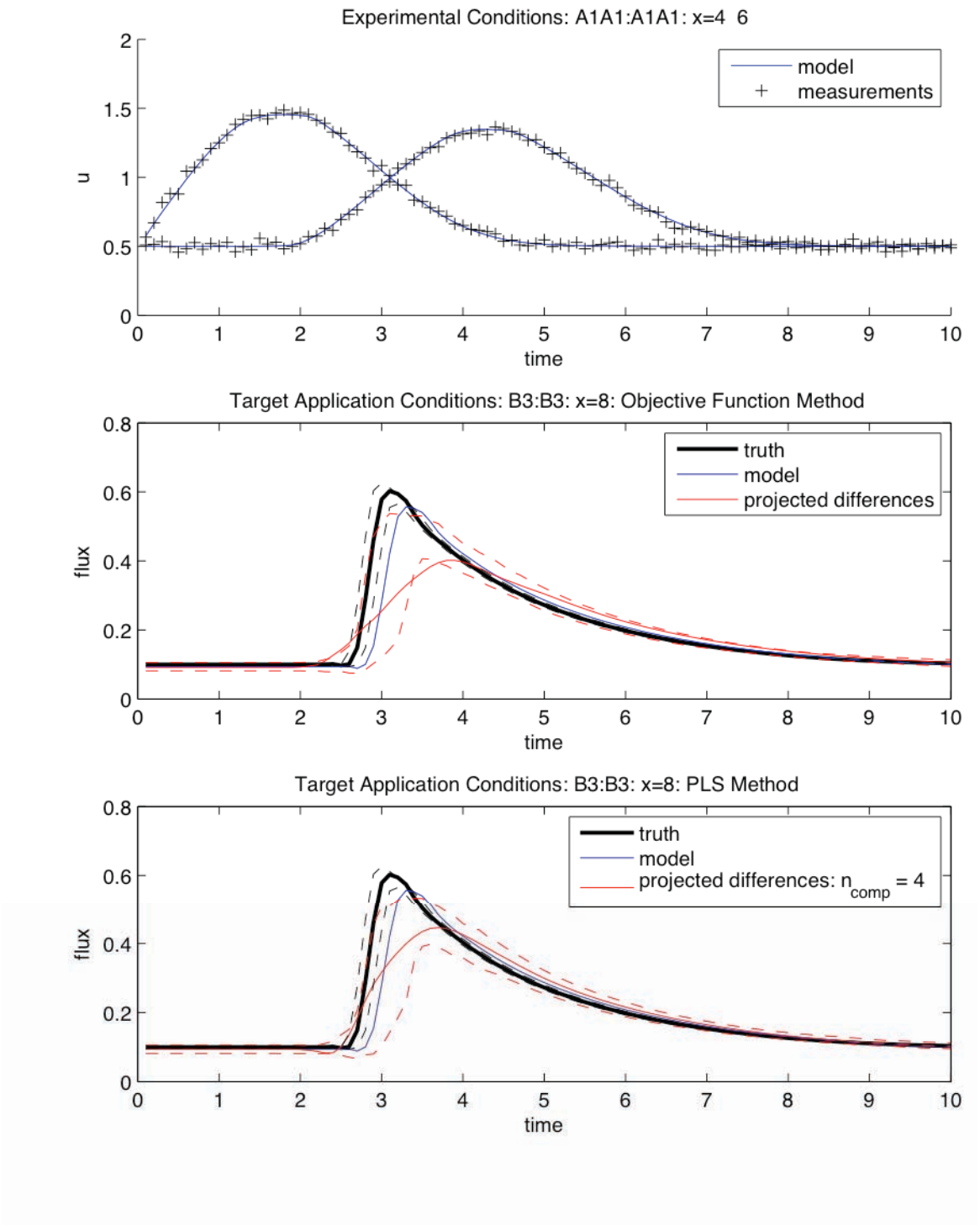

Figure 14. Known but Untested Physics

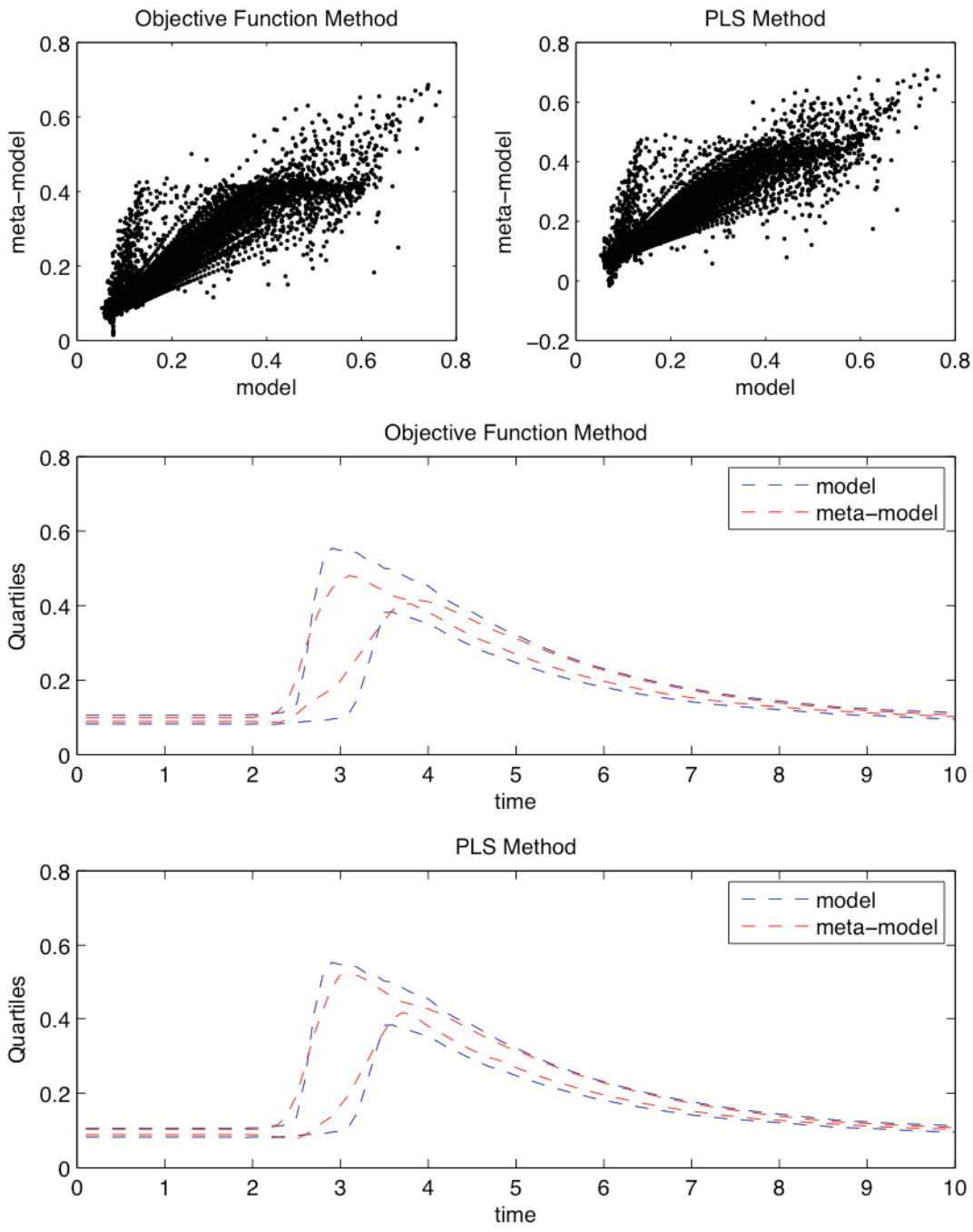

Figure 15. Known but Untested Physics: Assessment

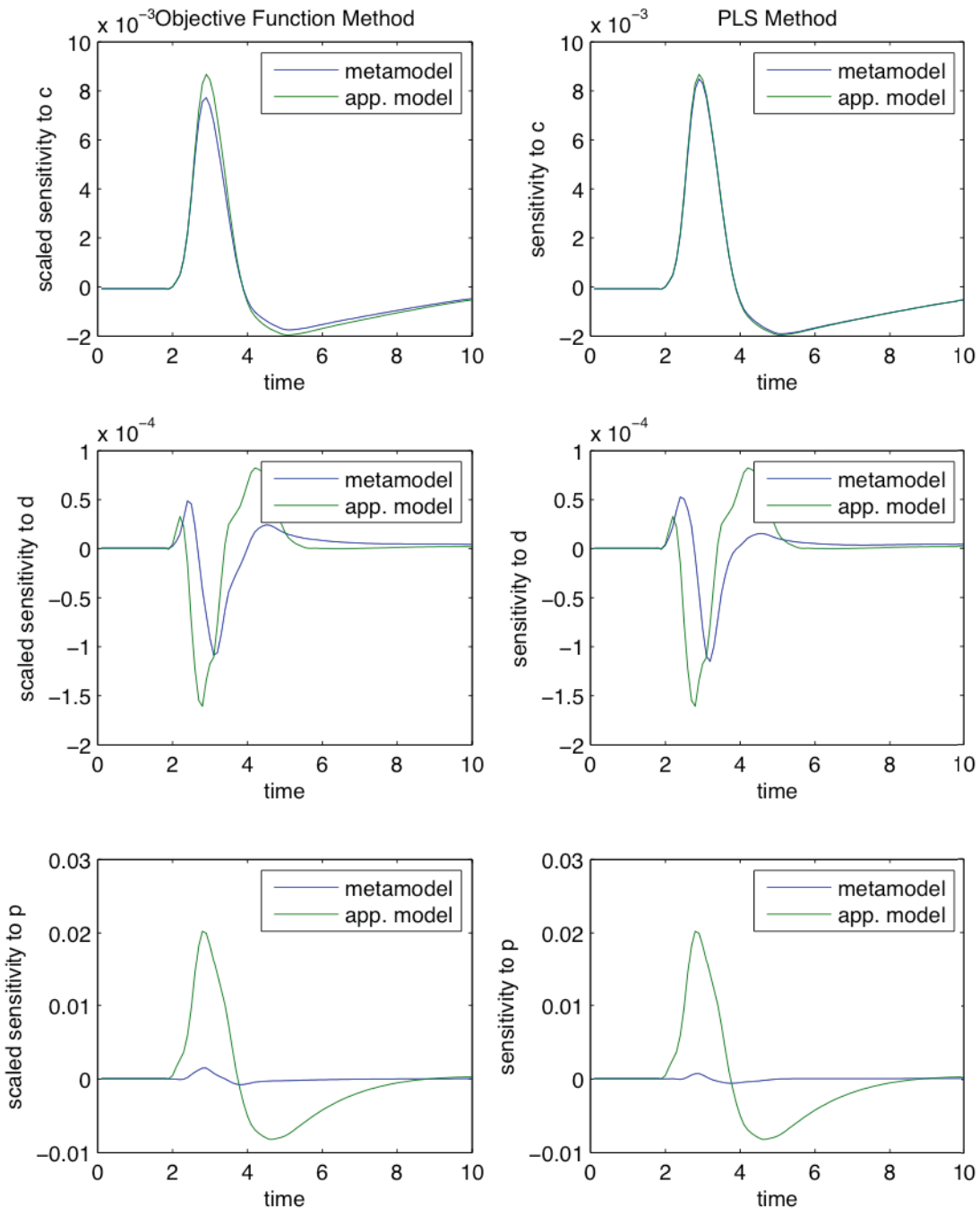

Figure 16. Known but Untested Physics: Sensitivity

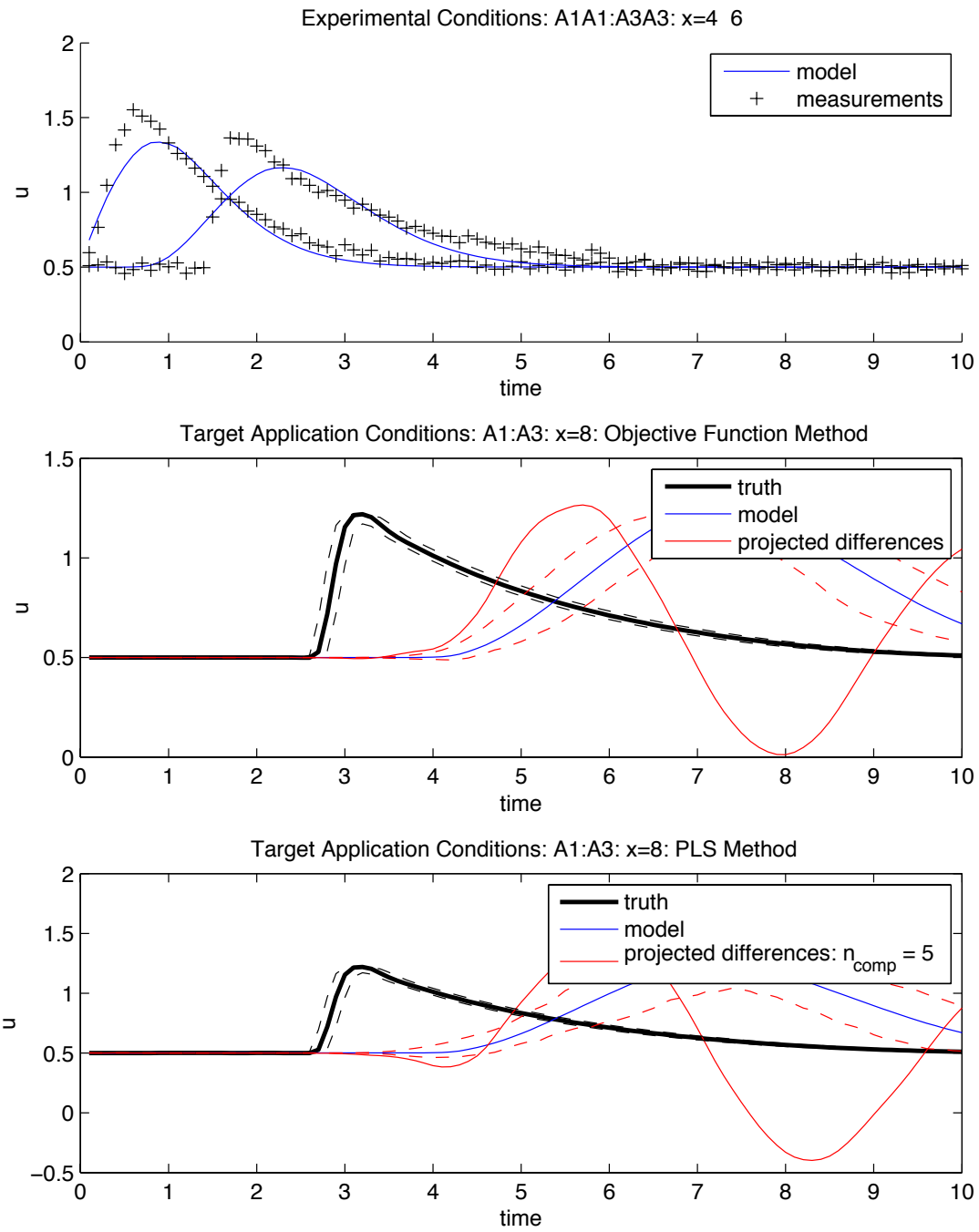

Figure 17. Unknown Tested Physics

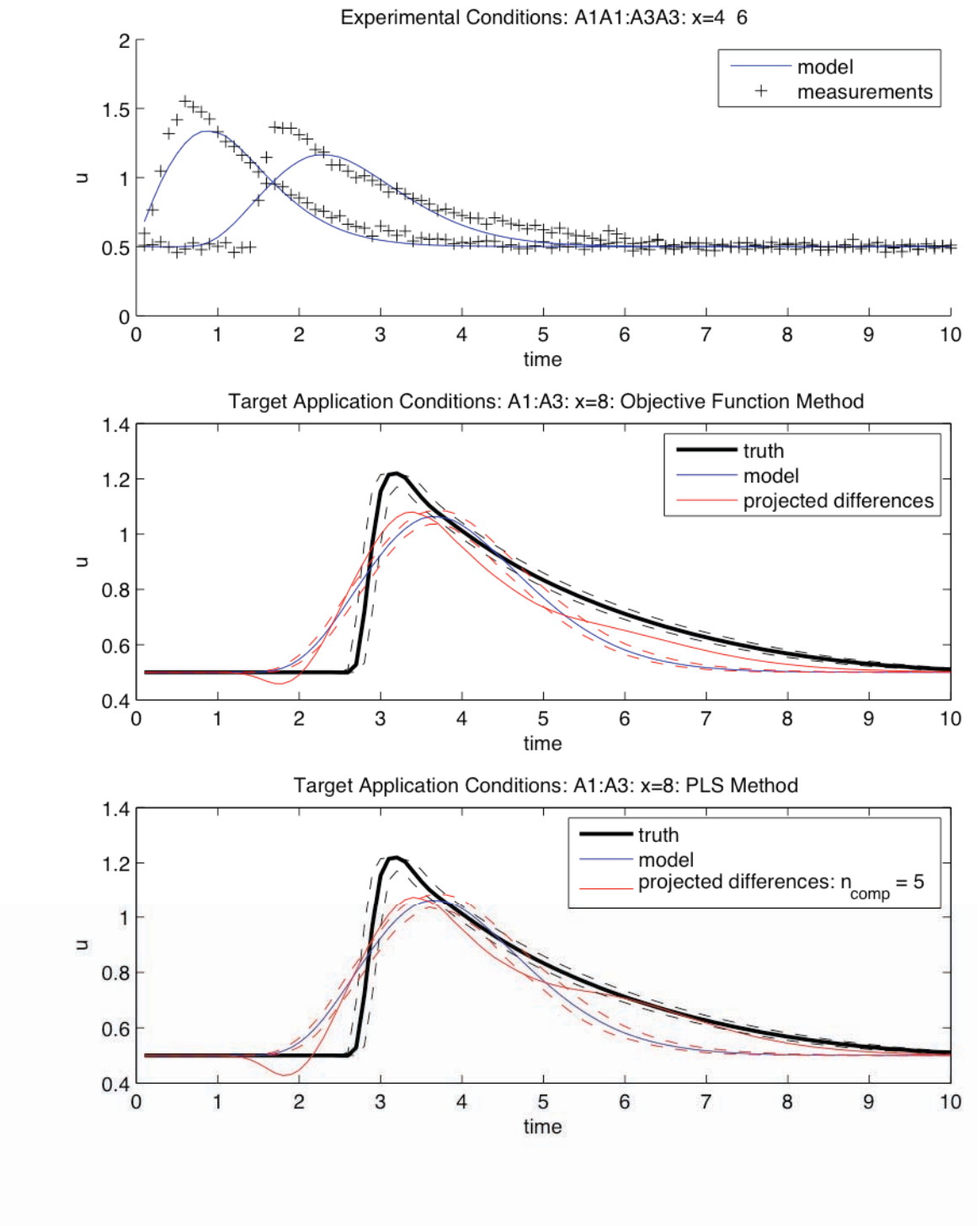

Figure 18. Unknown Tested Physics: Tuned to x=6 Results

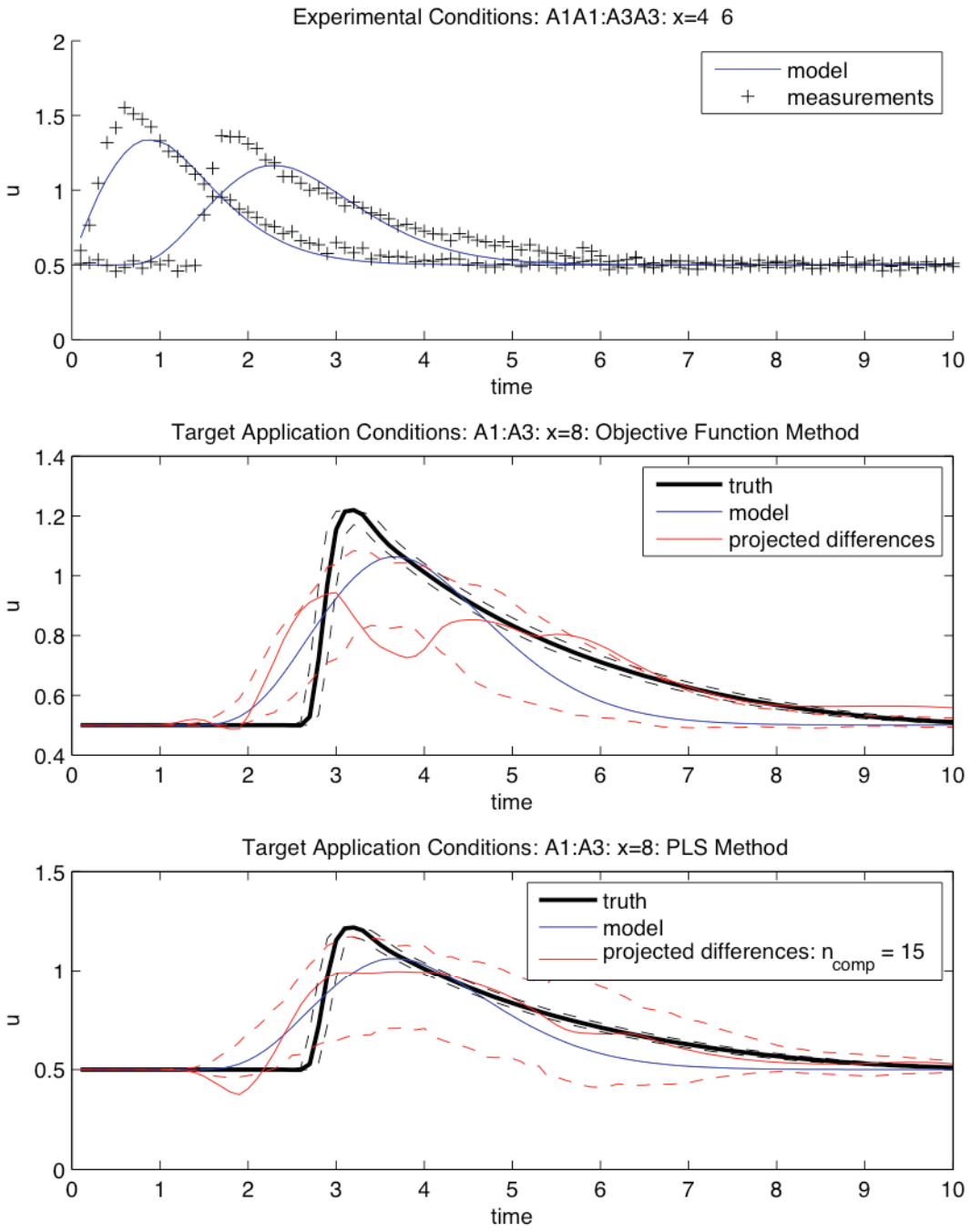

Figure 19. Unknown Tested Physics: Tuned, Expanded Uncertainty

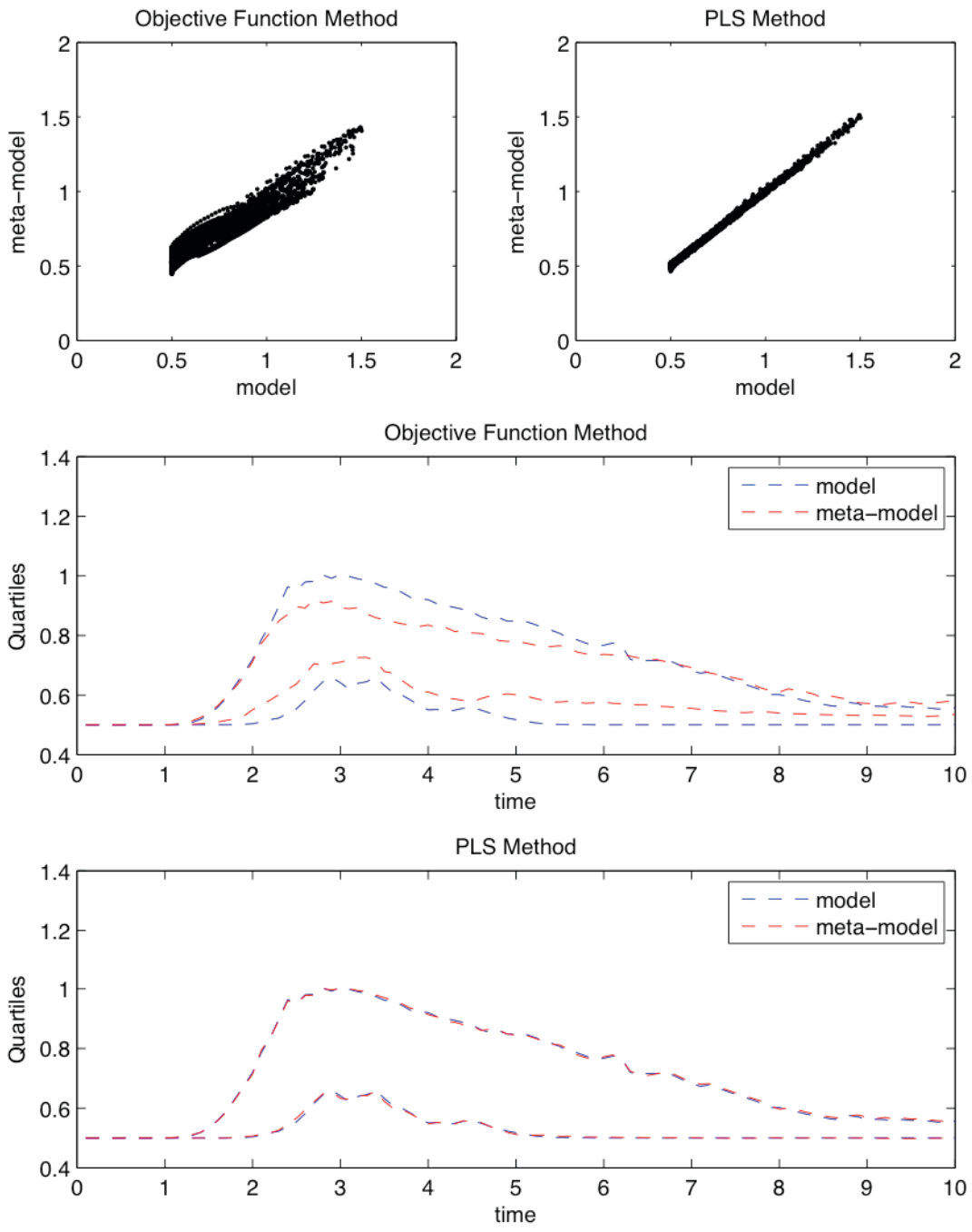

Figure 20. Unknown Tested Physics: Tuned, Expanded Uncertainty, Assessment

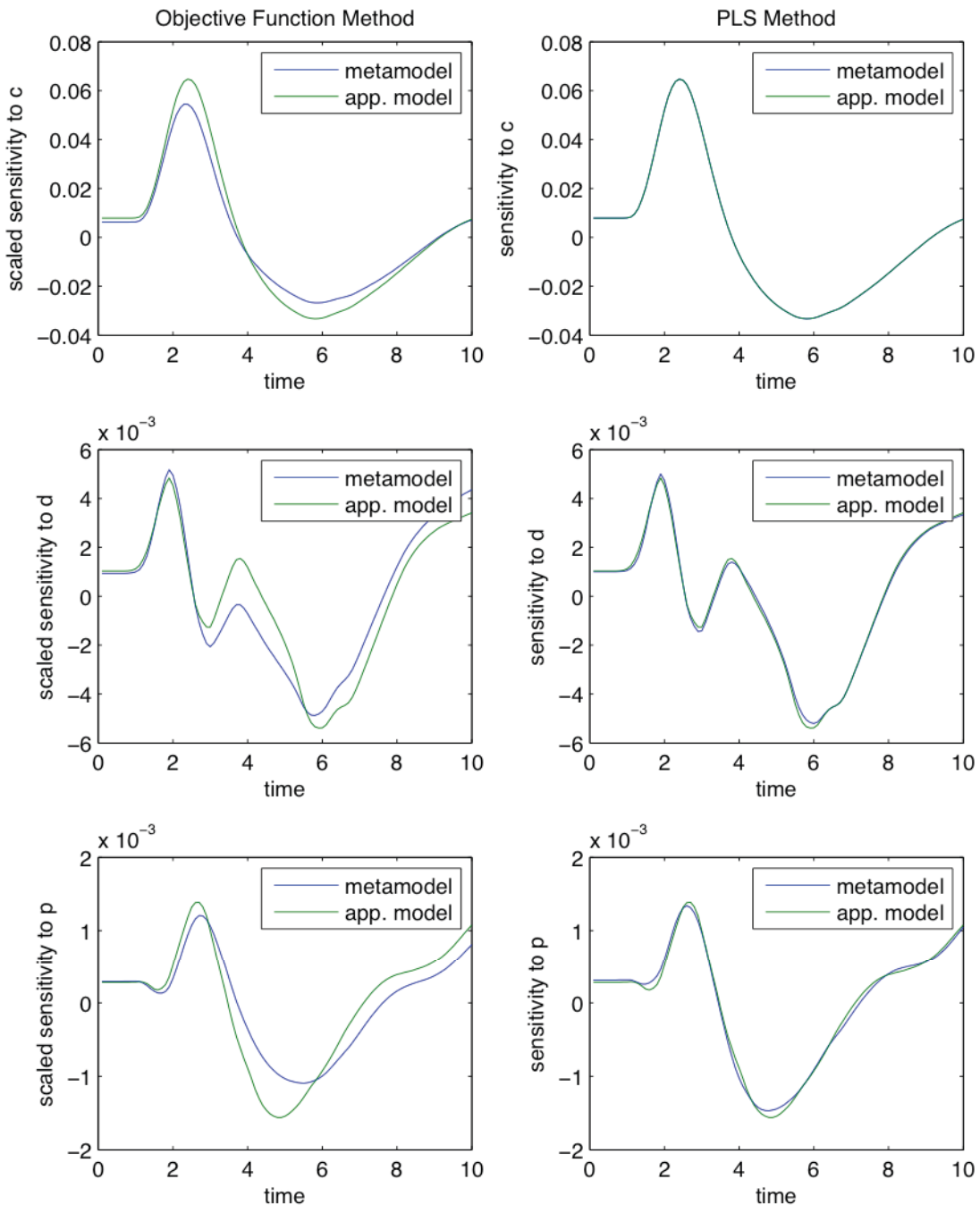

Figure 21. Unknown Tested Physics: Tuned, Expanded Uncertainty, Sensitivity

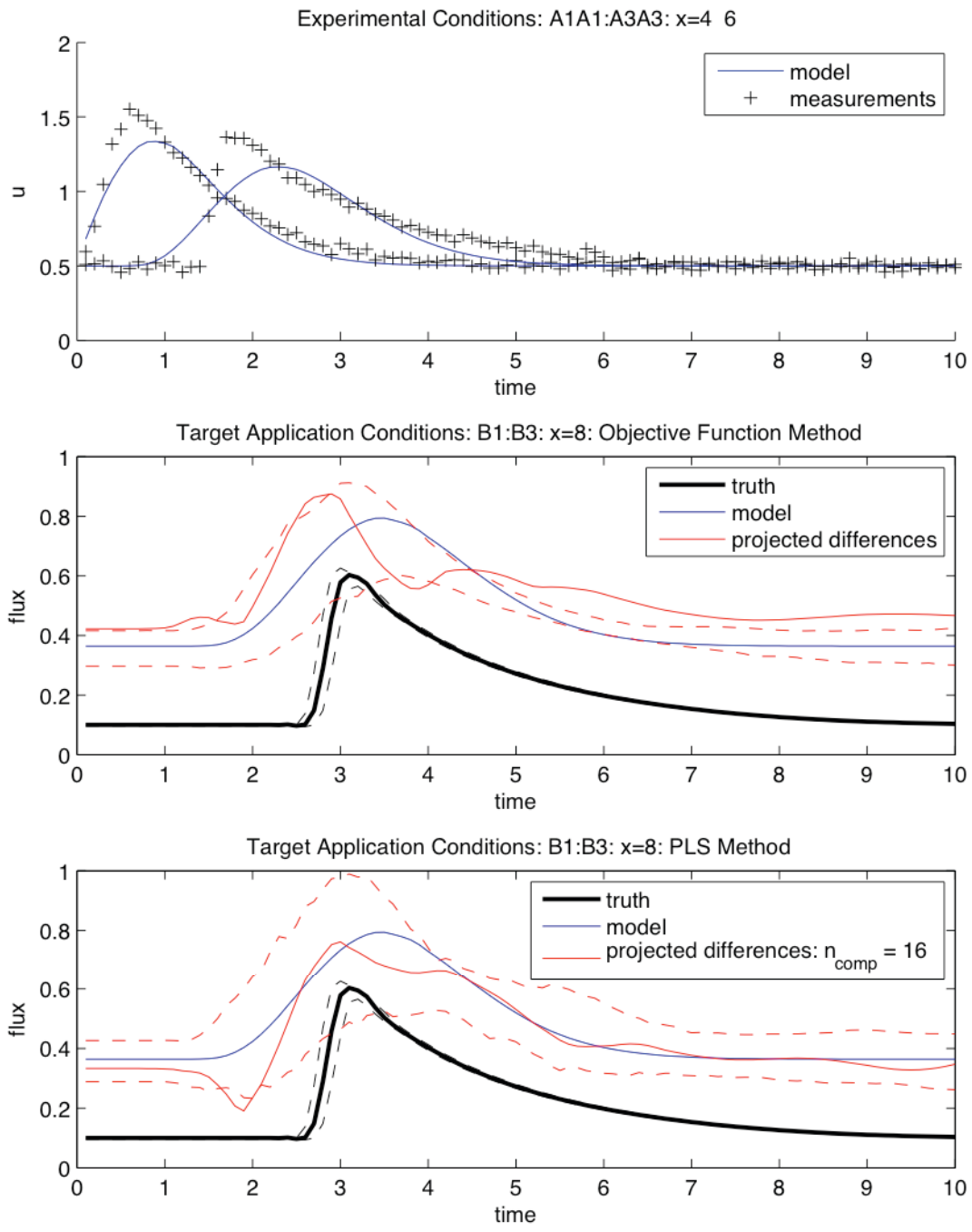

Figure 22. Unknown Tested Physics: Tuned, Expanded Uncertainty, Case 2

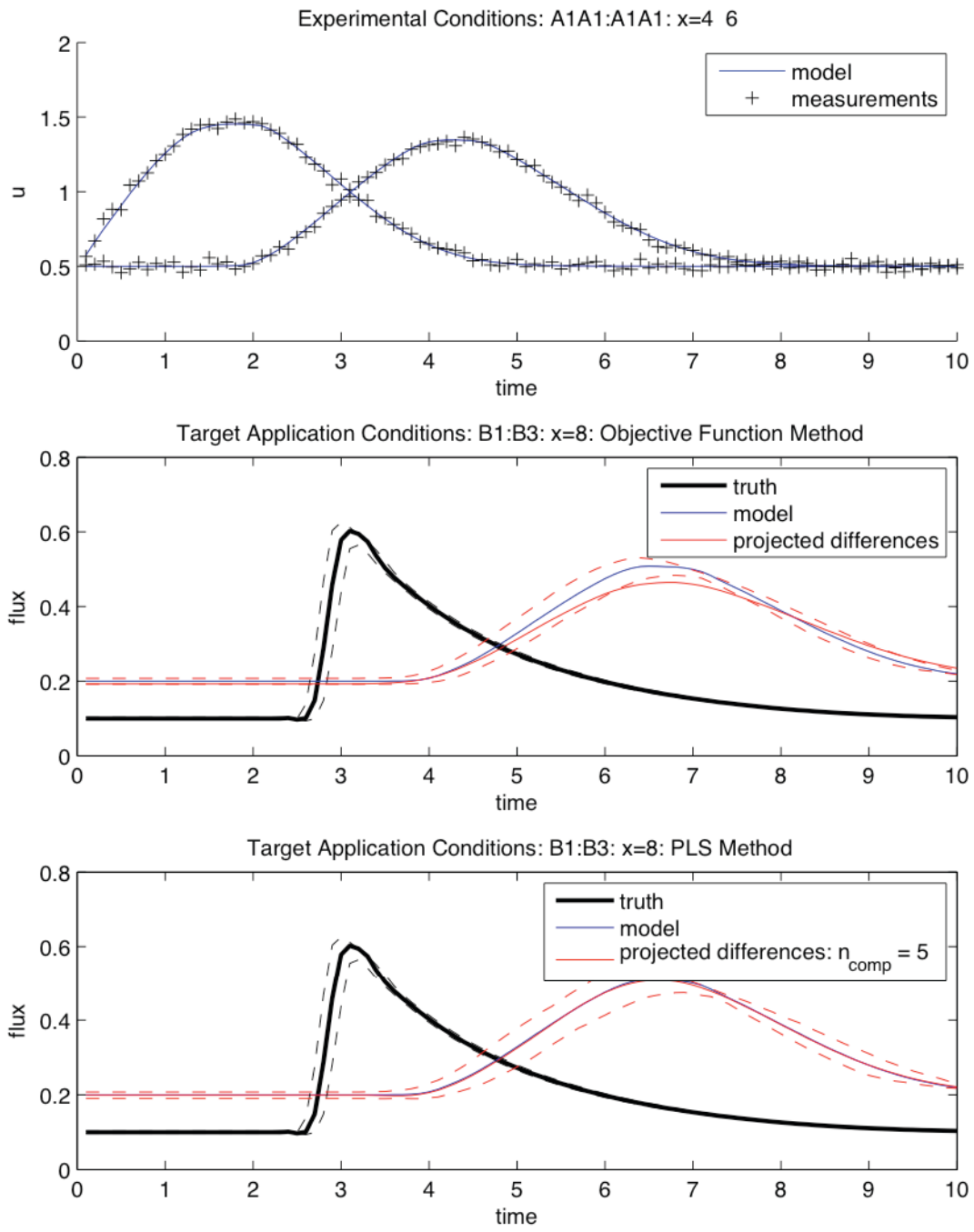

Figure 23. Unknown Untested Physics

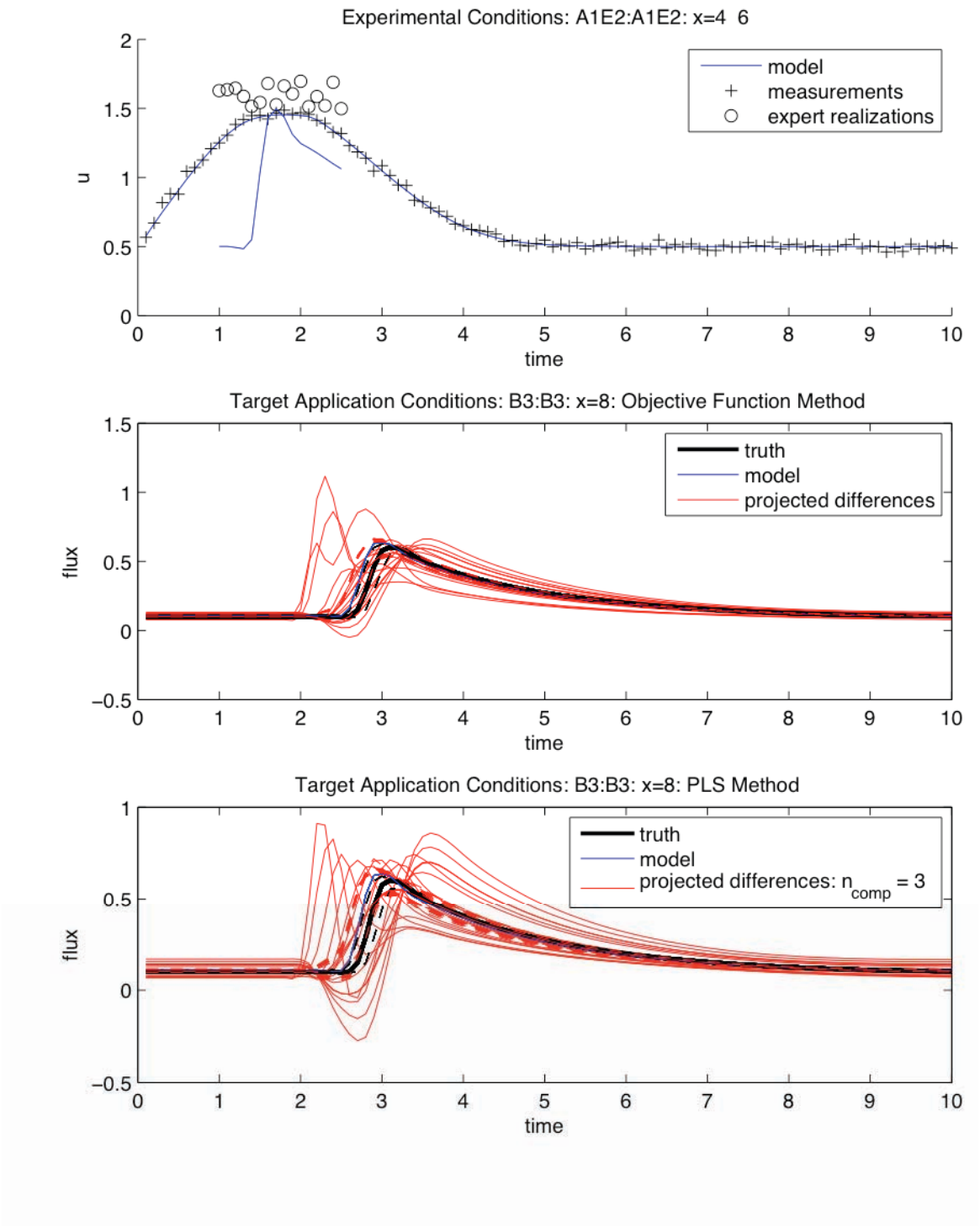

Figure 24. Expert Opinion

5. DISCUSSION

The methodology presented provides a means to assess the impact of the lack of full physics coverage of a target application by a suite of validation experiments. The methodology constructs a meta-model from models of the validation experiments to best represent the model for the target application response variable; and then assesses the ability to resolve the application model by the meta-model, and the uncertainty in the meta-model due to uncertainty in the model parameters for the models.

There is flexibility in the choice of the form of the meta-model. The meta-model was taken as a linear combination of sampled responses for the validation experimental models for the present development. Other choices can be made, such as products of sampled responses. The downside of other forms is that they may require non-linear optimization to estimate the weights, a process that is difficult if there are more unknowns (weights) than observations (the number of model runs).

Some of the limitations and concerns associated with the present methodology are listed below.

- The methodology can provide non-physical ringing or oscillatory behavior (see Figure 24) when there are considerable discrepancies between the experimental data and the validation model predictions. While other approaches, such as those surveyed earlier, generally will not suffer from this ringing for this transport application, they can suffer from the constraints placed on the target application predictions by the computational model itself. Methodology to bring in other types of expert judgment in the construction of meta-models can help. For example, the expert may state that *the projected target application result should not be oscillatory and contain only one peak*. Comparative analysis of the various approaches is required to assess the significance of this and other effects.
- The methodology cannot create knowledge out of nothing as the unknown-unknown example illustrates. This statement is true for any method. In the end, engineering judgment based on experience is required to cast a sufficiently broad net as to the types of physics that may be important and tested by the validation experiments.
- The use of the tuning and expanded uncertainty approach when the validation results indicate that the model is missing important physics, should cause discomfort in the analyst. The use of a clearly invalid model to predict target application performance should be done with significant skepticism or perhaps even cynicism. In such cases, one preferably should continue to develop the model, or at least develop an additive term that can be used to extrapolate observed differences to a prediction. The deficit model approach of Higdon et al. (2008) is an example of such an approach. However, the application of this approach for cases when the desired response quantities for the application are different from those of the validation experimental measurements, or for extrapolation beyond the support of the measurement data, provides a difficult and challenging problem; as deficit models generally do not reflect the underlying physics of the application (e.g., mass, momentum, and energy conserved).

The advantages of the meta-model approach are several-fold:

- The primary advantage is the approach allows one to assess whether the suit of validation experiments, as modeled, covers the target application response quantity, as modeled. The approach also can be used to assess whether the meta-model is sensitive to the same model parameters as the application model. This information can be used to justify the need for additional experiments. The methodology addresses the impact of model parameter uncertainty in both the target application and in the validation experiments, as well as measurement uncertainty, as characterized by the variance matrix of the measurements. Covariance between the measurement, between the model parameters, and between model parameters from different experiments (not demonstrated here, but included the author's MatLab, 2010, algorithm) can be easily handled.
- The weights developed in defining the meta-model have the effect of weighting those measurements most that best reflect the response of the target application. Measurements made closer to the conditions and physics of the application tend to be weighted more. These weights can be used to define an application level validation metric that utilizes the validation results taken throughout the validation hierarchy. The use of such weights to define a system relevant validation metric is presented in Hills and Leslie (2003).
- The approach can be exercised to provide additional insight into the relationship between validation experiments and a target application. For example, one observation made for the example problems was that more measurements are needed to represent a target application if one plans to use a "not so valid model" to predict a response quantity for an application. This is due to the requirement to estimate more latent variable/vectors to capture the nonlinear behavior over a larger neighborhood as expanded due to model tuning.

Both the objective function approach and the partial least squares (PLS) approach can be used to construct the weights associated with the meta-model. An advantage of the PLS approach for the present applications is that the PLS approach better captured the response of the target application. The PLS method is available in many statistical packages, such as MatLab (2010). Both approaches are designed to handle singular system for which there are more unknowns (number of weights or validation measurements) than responses (number of LHS realizations). This proves to be an advantage as the number of model evaluations that one can afford may be less than the number of measured quantities (typically large when time responses are measured).

Rolling up validation experiment results and uncertainties to target application predictions represents one of the more difficult problems in model validation when testing models throughout a validation hierarchy. Significant research is needed in the development of such methodology. Such research should be somewhat general in the structure of the validation hierarchy, in the sense that the methodology should be applicable to experiments with different physics, different measured and predicted response types, experiments at different conditions from the application, and to hierarchies that possess incomplete coverage of the application physics.

Perhaps the most significant aspects of the present work is that this work illustrates the need to roll-up the additional uncertainties due to lack of coverage of the application by the validation

hierarchy, illustrates many of the issues associated with this roll-up, and provides potential example problems to test other methodologies.

6. REFERENCES

[1] Abdi, H. (2010), Partial Least Squares Regression and Projection on Latent Structure Regression (PLS Regression), WIREs Computational Statistics, Focus Article, www.wiley.com/wires/compstats, John Wiley & Sons, Inc.

[2] AIAA (1998), Guide for the Verification and Validation of Computational Fluid Dynamics Simulations, AIAA-G-077-1998, American Institute of Aeronautics and Astronautics, Reston, VA, p. 3.

[3] ASME (2006), Guide for Verification and Validation in Computational Solid Mechanics, ASME V&V 10-2006, The American Society of Mechanical Engineers, New York, p. 3.

[4] ASME (2009), Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer, ASME V&V 20-2009, The American Society of Mechanical Engineers, New York, NY.

[5] Box G. E. P., D. W. Behnken (1960), Some New 3 Level Designs for the Study of Quantitative Variables, Technometrics 2, pp. 455-475.

[6] Diegert, K., V. Kathleen, S. E. Klenke, G. Novotny, R. A. Paulsen Jr., A. Robert, M. Pilch, T. G. Trucano. (2007), Toward a More Rigorous Application of Margins and Uncertainties within the Nuclear Weapons Life Cycle – A Sandia Perspective, SAND2007-6219, Sandia National Laboratories, Albuquerque, NM.

[7] Haenlein M. and A. M. Kaplan (2004), A Beginner's Guide to Partial Least Squares (PLS) Analysis", Understanding Statistics – Statistical Issues in Psychology, Education, and Social Sciences, 3: 4, pp. 283-297.

[8] Hamilton, J. R. and Hills, R. G. (2010a), Relation of Validation Experiments to Applications, Numerical Heat Transfer, Part B: Fundamentals, 57: 5, pp. 307-332.

[9] Hamilton, J. R. and Hills, R. G. (2010b), 'Relation of Validation Experiments to Applications: A Nonlinear Approach, Numerical Heat Transfer, Part B: Fundamentals, 57: 6, pp. 373-395.

[10] Higdon, D., Nakhleh, C., Gattiker, J., and Ginzburg L. (2008), A Bayesian Calibration Approach to the Thermal Problem, Computer Methods in Applied Mechanics and Engineering, 197: 29-32, pp. 2457-2466.

[11] Hills, R.G., K.A. Fisher, M.R. Kirkland, and P.J. Wierenga (1994), Application of Flux Corrected Transport to the Las Cruces Trench Site, Water Resources Research, 30: 8, pp. 2377-2385.

[12] Hills, R. G., I. H. Leslie (2003), Statistical Validation of Engineering and Scientific Models: Validation Experiments to Application, SAND2003-0706, Sandia National Laboratories, Albuquerque, NM.

[13] Hills, R.G., M. Pilch, K.J. Dowding, J. Red-Horse, T.L. Paez, I. Babuška, and R. Tempone (2008), Validation Challenge Workshop, Computer Methods in Applied Mechanics and Engineering, 197: pp. 2375-2380.

[14] Mahadevan, S. (2011), Roll-Up of Multi-Level UQ Activities towards System Level QMU, Presented at Sandia National Laboratories, Albuquerque, NM, Oct. 4.

[15] MatLab (2010), MATLAB ® Mathematics, Version 7, R2010b, The MathWorks, Inc., Natick, MA.

[16] Mayer, M. A., and Booker, J. M. (1991), Eliciting and Analyzing Expert Judgment; A Practical Guide, ASA-SIAM Series on Statistics and Applied Probability, Society of Industrial and Applied Mathematics and American Statistical Association, London.
[17] Pilch, M., T. G. Trucano, J. C. Helton (2006), “Ideas Underlying Quantification of Margins and Uncertainties (QMU): A White Paper,” SAND2006-5001, Sandia National Laboratories, Albuquerque, NM.
[18] Romero V.J. (2006), Issues and Needs in Quantification of Margins and Uncertainty (QMU) for Phenomenologically Complex Coupled Systems, paper AIAA-2006-1989, 8th AIAA Non-Deterministic Analysis Conference, Newport, RI, May 1-4.
[19] Romero, V.J. (2007), Validated Model? Not So Fast. The Need for Model 'Conditioning' as an Essential Addendum to Model Validation, paper AIAA-2007-1953, 9th AIAA Non-Deterministic Methods Conference, April 23 - 26, Honolulu, HA.
[20] Romero, V.J. (2008), Type X and Y Errors and Data & Model Conditioning for Systematic Uncertainty in Model Calibration, Validation, and Extrapolation, SAE paper 2008-01-1368 for Society of Automotive Engineers 2008 World Congress, April 14-17, Detroit, MI.

DISTRIBUTION

Sandia Internal Distribution (electronic)

1	MS0346	M. Mesh	1523
1	MS0557	T. W. Simmermacher	1523
1	MS0807	K. F. Alvin	1544
1	MS0815	J. E. Johannes	1500
1	MS0828	A. R. Black	1544
1	MS0828	R. B. Bond	1541
1	MS0828	K. J. Dowding	1544
1	MS0828	A. C. Hetzler	1544
1	MS0828	R. G. Hills	1544
1	MS0828	G. E. Orient	1544
1	MS0828	V. J. Romero	1544
1	MS0828	A. Urbina	1544
1	MS0828	W. R. Witkowski	1544
1	MS0845	T. L. Galpin	1545
1	MS0845	J. Jung	1542
1	MS0845	D. E. Womble	1540
1	MS0897	T. D. Blacker	1543
1	MS1152	K. Cartwright	1352
1	MS1168	B. V. Oliver	1340
1	MS1318	B. M. Adams	1441
1	MS1318	M. S. Eldred	1441
1	MS1318	J. R. Stewart	1441
1	MS1318	T. G. Trucano	1440
1	MS1323	W. J. Rider	1443
1	MS9042	M. E. Gonzales	8250
1	MS9159	P. D. Hough	8954
1	MS0899	Technical Library	9536


Sandia National Laboratories