

# Multilevel Bayesian Parameter Estimation in the Presence of Model Inadequacy and Data Uncertainty

Amir Shahmoradi <sup>1,2,3,\*</sup>

November 30, 2017

**Abstract** Model inadequacy and measurement uncertainty are two of the most confounding aspects of inference and prediction in quantitative sciences. The process of scientific inference (the inverse problem) and prediction (the forward problem) involve multiple steps of data analysis, hypothesis formation, model construction, parameter estimation, model validation, and finally, the prediction of the quantity of interest. This article seeks to clarify the concepts of model inadequacy and bias, measurement uncertainty, and the two traditional classes of uncertainty: aleatoric versus epistemic, as well as their relationships with each other in the process of scientific inference. Starting from basic principles of probability, we build and explain a hierarchical Bayesian framework to quantitatively deal with model inadequacy and noise in data. The methodology can be readily applied to many common inference and prediction problems in science, engineering, and statistics.

---

<sup>1</sup> Center For Computational Oncology, The University of Texas at Austin, TX 78712

<sup>2</sup> Institute for Computational Engineering and Sciences, The University of Texas at Austin, TX 78712

<sup>3</sup> Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, TX 78712

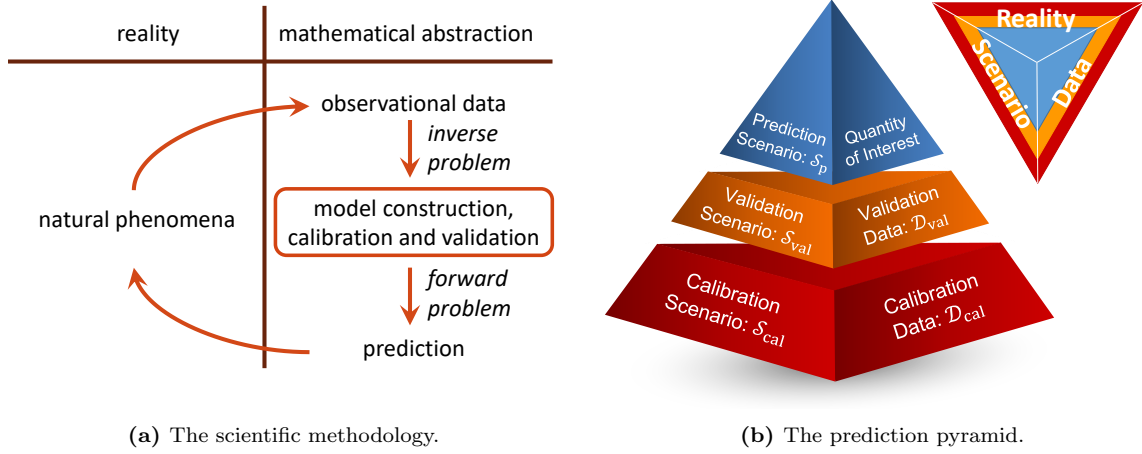
\* Peter O'Donnell, Jr. Fellow, E-mail: amir@physics.utexas.edu

## Contents

1	Introduction . . . . .	3
1.1	Aleatoric vs. Epistemic Uncertainty . . . . .	4
1.2	Model Inadequacy vs. Aleatoric Uncertainty . . . . .	5
1.3	The Goal of This Paper . . . . .	6
2	Modeling the Truth . . . . .	7
2.1	The Relationship Between Observational Data and the Truth . . . . .	7
2.2	The Forward Problem . . . . .	8
2.3	The Inverse Problem . . . . .	9
3	Modeling the Truth in the Presence of Model Inadequacy . . . . .	9
3.1	The Construction of Inadequacy Model . . . . .	10
3.2	The Likelihood Function in the Presence of Model Inadequacy . . . . .	11
3.3	Popular Choices of Inadequacy Model . . . . .	12
4	Modeling the Truth, Confounded with Noise (Measurement Error) . . . . .	12
4.1	Noise Models Are Fundamentally Different From Inadequacy Models . . . . .	14
4.2	The Likelihood Function in the Presence of Data Uncertainty . . . . .	15
5	Hierarchical Modeling of the Truth, in the Presence of Model Inadequacy and Data Uncertainty . . . . .	17
5.1	Statement of the Hierarchical Bayesian Inverse Problem . . . . .	18
5.2	General Solution . . . . .	18

**Table 1:** Nomenclature and Definitions of Symbols in This Manuscript

$\mathbf{R}$	The reality or truth representing one event, without any observational bias or uncertainty.
$\mathcal{R}$	The set containing the truth $\mathbf{R}$ for each individual event that is observed.
$\Omega_{\mathbf{R}}$	The observational sampling space to which each event in the dataset belongs: $\mathbf{R} \in \mathcal{R} \subset \Omega_{\mathbf{R}}$ .
$\mathbf{R}_i$	The reality or truth, $\mathbf{R}$ , about the $i$ th event in event-set $\mathcal{R}$ .
$\mathcal{R}_{\mathbf{R}_i}$	A subset of $\mathcal{R}$ on which the $i$ th event, $\mathbf{R}_i$ , depends. It can be and often is a null set.
$\mathbf{D}$	All data about an event <i>as observed</i> , which is subject to measurement error, unlike $\mathbf{R}$ .
$\mathcal{D}$	The set of all observations, each of which corresponds to one unique event, $\mathbf{R}_i \in \mathcal{R}$ .
$\mathbf{D}_i$	All observational data, $\mathbf{D}$ , about the $i$ th observation in dataset $\mathcal{D}$ .
$\mathbf{U}$	A stochastic variable representing difference between $\mathbf{R}$ and the output, $\mathbf{R}'$ , of $\mathbf{M}_{\text{phys}}$ .
$\mathcal{U}$	The set of all $\mathbf{U}_i$ , each of which corresponds to one event $\mathbf{R}_i \in \mathcal{R}$ .
$\mathbf{R}^*$	One possible realization of $\mathbf{R}$ , given $\mathbf{D}$ and the corresponding noise model, $\mathbf{M}_{\text{nois}}$ .
$\mathcal{R}^*$	One possible realization of $\mathcal{R}$ , given $\mathcal{D}$ and the set of noise models, $\mathcal{M}_{\text{nois}}$ .
$\mathcal{R}_{\mathbf{R}_i}^*$	One possible realization of $\mathcal{R}_{\mathbf{R}_i}$ .
$\Omega_{\mathcal{R}^*}$	The (super)set containing all possible realizations, $\mathcal{R}^*$ , of the set $\mathcal{R}$ .
$\mathbf{M}_{\text{phys}}$	The physical model hypothesized to hold for the collection of events in $\mathcal{R}$ .
$\mathbf{M}_{\text{inad}}$	The statistical physics-based model that quantifies the inadequacy of $\mathbf{M}_{\text{phys}}$ in describing $\mathcal{R}$ .
$\mathbf{M}_{\text{nois}}$	The statistical model that quantifies the experimental measurement uncertainty (noise) in $\mathcal{D}$ .
$\mathcal{M}_{\text{nois}}$	The set of statistical models $\mathbf{M}_{\text{nois},i}$ corresponding to each observation $\mathbf{D}_i \in \mathcal{D}$ .
$\boldsymbol{\theta}_{\text{phys}}$	The vector of parameters of the physical model $\mathbf{M}_{\text{phys}}$ .
$\boldsymbol{\theta}_{\text{inad}}$	The vector of parameters of the inadequacy model $\mathbf{M}_{\text{inad}}$ .
$\boldsymbol{\theta}_{\text{nois}}$	The vector of parameters of the noise model $\mathbf{M}_{\text{nois}}$ .
$\Theta(\cdot)$	The parameter space of a model, $(\cdot)$ – e.g., $\mathbf{M}_{\text{phys}} : \Theta_{\text{phys}}$ , $\mathbf{M}_{\text{inad}} : \Theta_{\text{inad}}$ , $\mathbf{M}_{\text{nois}} : \Theta_{\text{nois}}$ .
$n_{\text{do}}$	The number of data observations in $\mathcal{D}$ (or equivalently, the number of events in $\mathcal{R}$ ).
$n_{\text{da}}$	The number of data attributes by which each event is characterized; length of $\mathbf{R}$ & $\mathbf{D}$ .
$n_{\text{pp}}$	The number of parameters of the physical model; length of $\boldsymbol{\theta}_{\text{phys}}$ ; dimension of $\Theta_{\text{phys}}$ .
$n_{\text{pi}}$	The number of parameters of the inadequacy model; length of $\boldsymbol{\theta}_{\text{inad}}$ ; dimension of $\Theta_{\text{inad}}$ .
$n_{\text{pn}}$	The number of parameters of the noise model; length of $\boldsymbol{\theta}_{\text{nois}}$ ; dimension of $\Theta_{\text{nois}}$ .
$\mathbb{R}$	The set of real numbers.
$(\cdot)'$	The output of $\mathbf{M}_{\text{phys}}$ , which may or may not be identical to the input entity, $(\cdot)$ , to $\mathbf{M}_{\text{phys}}$ .
$\pi(\cdot)$	The Probability Density Function (PDF) of a statistical model.
$\mathcal{L}(\cdot)$	The likelihood function of the parameters of a model.
$\mathcal{I}(\cdot)$	The prior knowledge about the subscript entity, $(\cdot)$ .



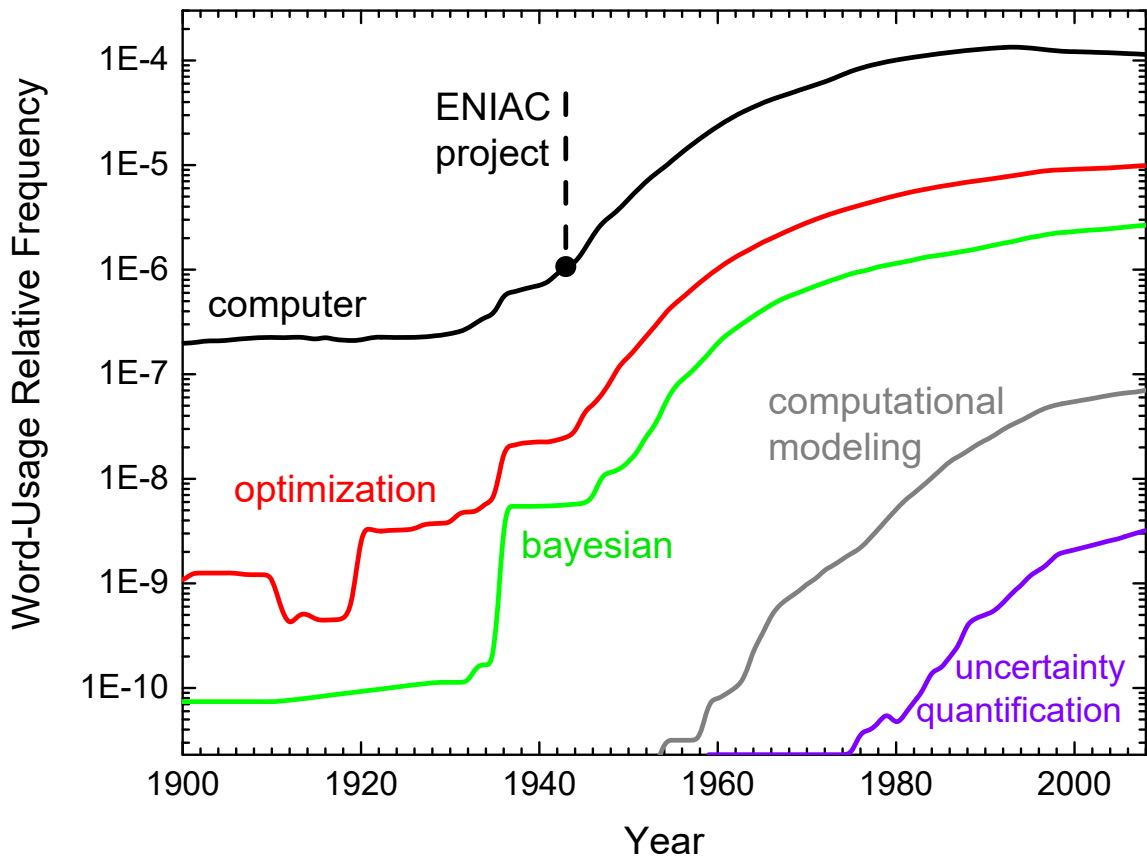
**Fig. 1: (a) The steps of scientific methodology**, involving data collection, hypothesis formulation, construction of a mathematical model and objective function, which is subsequently optimized to constrain the parameters of the model, a process known as *inversion* or *inverse problem*. Once validated, the model can be used to make predictions about the quantity of interest (*forward problem*). **(b) The prediction pyramid**, depicting the three hierarchical levels of predictive inference from bottom to top: Calibration, Validation, and Prediction of the Quantity of Interest (QoI). The rear face of the tetrahedron represents reality (truth),  $\mathcal{R}$ , about the set of observed phenomena, which is never known to the observer. The right-front face of the tetrahedron represents the observational data,  $\mathcal{D}$ , which results from the convolution of the truth/reality,  $\mathcal{R}$ , with various forms of measurement uncertainty. The left-front face represents the scenarios,  $\mathcal{S}$ , under which data is collected, as well as the set of models that are hypothesized to describe the unknown truth,  $\mathcal{R}$  [53, 54, 59, 61].

## 1 Introduction

The process of scientific inference involves the collection of experimental data from observations of a set of natural phenomena, the analysis and reduction of the collected dataset, the formulation of a hypothesis (i.e., development of a physics-based mathematical model) that attempts to explain various potential causal relationships between different characteristics of data, and, finally, testing the predictions of the proposed model against new observational data by performing new experiments (Figure 1a).

In the majority of scientific problems, the proposed physical model involves a set of parameters that have to be tuned in order to best describe the available data. For example, Einstein’s famous equation of mass-energy equivalence,  $E = mc^2$ , relates the mass,  $m$ , of any material to an equivalent amount of energy,  $E$ , via an a-priori unknown constant,  $c$ , that is the speed of light, which has to be determined by experimental data. The process of inferring the parameters of the physical model is commonly known as *inversion* or an *inverse problem*, which could also be regarded as *model calibration* (Figure 1b).

Once the parameters of a physical model are constrained, the proposed physical model has to be verified and its predictions validated against a new independent dataset. Extensive literature already exists on the topic of model verification and validation [e.g., 3, 4, 7, 8, 24, 44, 51, 56, 60, 66] as well as on decision theory [for elegant reviews from a Bayesian perspective, see 39, 40, 47]. The validated model can be then used to make predictions of the Quantities of Interest (QoI), the precise physical features of the response of the system targeted in the simulation. This prediction step is commonly known as the *forward problem* in the scientific literature.



**Fig. 2:** A word-usage relative frequency plot [49], illustrating the exponential growth of computer technology in the mid 20th century, as well as developments in the fields of deterministic and stochastic optimization techniques, which ultimately led to the emergence of ‘computational modeling’ as the third pillar of science [61]. Advances in the computational methods and technology also led to the gradual popularity of Bayesian techniques in mathematical modeling toward the end of the 20th century, as well as the emergence of *uncertainty quantification* as a new field of science. Note that the positive-slope linear behavior on this semi-logarithmic plot implies exponential growth.

The process of scientific inference described above, although straightforward at a first glance, is severely complicated by the presence of many sources of uncertainty in multiple levels of data acquisition and model construction, as well as the inverse and forward problems. In fact, the significance of the effects of uncertainty in data and modeling has led to the emergence of a new field of science within the past three decades, specifically dedicated to *Uncertainty Quantification* (Figure 2).

### 1.1 Aleatoric vs. Epistemic Uncertainty

Different classes and sources of uncertainty have been already identified and extensively studied at different levels of the scientific inference process, for example, in data acquisition and model construction [2, 16, 20, 34, 68, 82], or in the discretization and numerical computations of the inverse and forward problems [1, 5, 6, 9, 15, 42, 52, 55–58, 64, 68].

Uncertainty in data acquisition and model construction has been traditionally divided into two categories of *epistemic* and *aleatoric* (*aleatory*) [11, 12, 19, 20, 23, 28, 62]. Aleatoric un-

certainties are presumed to stem from inherent *unpredictable* variabilities and randomness in observational data and are therefore thought to be irreducible. For example, the experiment of throwing a die could be considered as an experiment with aleatoric uncertainty in its outcome (but note that this statement is incorrect within the Bayesian framework. See §1.2 for clarification). This form of uncertainty is sometimes also called *structural variability* or *risk* [20, 62] in engineering literature. To the contrary, epistemic uncertainties represent any lack of knowledge about data/experiment that can be potentially acquired in future, for example, the measurement errors in an experiment.

There is a long history of confusion and disagreement in the scientific literature about the exact definition and extent, or even the existence of these two forms of uncertainties [38, 39]. The origin of this century-long debate can be uncovered in the correspondence that is generally assumed between the two classes of aleatoric vs. epistemic uncertainties and the two prominent interpretations of probability: frequentist vs. Bayesian, respectively [17, 62, 81]. Thus, from a historical perspective, aleatoric uncertainty is solely defined within the framework of frequentist statistics [17].

We remark that in a pure Bayesian system of logical probability [39], which is the view we adopt in this work, all uncertainty is epistemic [17]: randomness is strictly a means to express a lack of knowledge. By contrast, what is often meant by aleatoric uncertainty in contemporary scientific literature appears to conform well to the concept of *model inadequacy*, which is further described below in §1.2.

Of course, on scales relevant to Quantum Mechanics [21], one may argue that Heisenberg’s Uncertainty Principle [29] dictates an inherent uncertainty in Nature, setting a hard limit on the extent of human knowledge. This strict epistemological limitation on human knowledge would, therefore, resemble aleatoric uncertainty in Natural phenomena at the ontological level. However, counter-arguments have been put forth by prominent physicists in the 20th century against this Copenhagen interpretation of Quantum Mechanics [13, 22, 38]. Regardless of the validity of Copenhagen interpretation and the Heisenberg Uncertainty Principle, the quantum mechanical limitations imposed on human knowledge can be considered irrelevant to virtually all practical modeling problems beyond the subatomic scales of Quantum Mechanics (See [67] for some illustrative discussions on this topic).

## 1.2 Model Inadequacy vs. Aleatoric Uncertainty

In an ideal and deterministic world, where there is neither model imperfection nor uncertainty in computation or data, one would naturally expect the physical model to perfectly describe observational data. This is, however, never the case in virtually all real-world inference problems. In reality, all models are imperfect or wrong (echoing the famous statement of E. Box that “all models are wrong but some can be useful” [14]) and none can provide a full description of data. This model imperfection is widely known in the literature as *model discrepancy* or *model inadequacy* [19, 53], the possible remedies of which have been already extensively studied [2, 16, 42, 48, 50].

Model inadequacy is often confused with aleatoric uncertainty since both can have identical effects on scientific inference. From a Bayesian perspective, one can argue that any type of intrinsic unexplained variability observed in natural phenomena is a result of our limited knowledge/data or a consequence of an imperfect physical model for the observed phenomenon.

Consider as an example, the experiment of throwing an unbiased die repeatedly under ‘similar conditions’. What do we really mean here by *similar conditions*? Indeed, if the experimenter had *complete knowledge* of the conditions under which the die was thrown, there would be no intrinsic unexplained randomness in the experiment’s output; that is, one would be able to predict exactly the outcome of each die-throwing experiment.

Therefore, our ignorance of the many details of input data to a sequence of experiments and the conditions under which the experiments were performed, manifest themselves in the form of an inherent variability in the experimental output. In other words, no two experiments can be truly considered as identical replicates of each other, because our knowledge of the experimental setup as well as the input data to the experiments is almost always incomplete. This lack of a complete detailed knowledge of the problem under study often leads to the development of mathematical models that are inadequate in describing the underlying physics of the problem correctly.

In sum, model inadequacy appears to be frequently confused with aleatoric uncertainty in contemporary scientific literature. From a Bayesian viewpoint, aleatoric uncertainty does not exist. As a result, aleatoric uncertainty is sometimes implicitly redefined as the class of uncertainties for which there is no foreseeable possibility of elimination or reduction *at the time of inference* [20], although it may be reduced with the arrival of new models, experimental designs, or more detailed experimental data in future (see [69] for an elegant historical/philosophical review).

### 1.3 The Goal of This Paper

Regardless of the terminology used for uncertainty classification, the truth is always convolved with uncertainties that are either due to measurement errors or incomplete (insufficiently-detailed) data. This lack of knowledge, in turn, leads to the development of imperfect physical models whose predictions are inadequate for a complete description of the observed data. As a result, new physically-inspired stochastic and/or deterministic models are needed to further describe the inadequacy of the physical models.

Description of a general framework for incorporating different sources of uncertainty, in particular, the measurement error and model inadequacy in the process of scientific inference, seems to be lacking in the current scientific and engineering literature. Most of the few resources available on this topic focus on special cases where the errors involved in the problem take simple Gaussian forms and are assumed to be additive [2, 16, 18, 30, 31, 53, 65].

In the following sections, we consider the effects of model inadequacy (or as sometimes referred to it by ‘aleatoric uncertainty’ in the literature), as well as the effects of noise and

measurement error in experimental data, on parameter estimation and predictive inference. Although we have argued, and assume throughout the rest of this paper, that all uncertainties are epistemic (i.e., due to lack of knowledge), we recognize and show in the following sections that model inadequacy and measurement errors require fundamentally different treatments in the process of scientific inference. This is in agreement with the general consensus in the literature [25]. A complete description of all the variables used in this manuscript is given in Table 1.

## 2 Modeling the Truth

### 2.1 The Relationship Between Observational Data and the Truth

Consider a set of  $n_{\text{do}}$  observations,

$$\mathcal{D} = \{\mathbf{D}_1, \dots, \mathbf{D}_{n_{\text{do}}}\} , \quad (1)$$

collected about a set of natural phenomena. This observational dataset is a result of the convolution of the unknown reality with various forms of uncertainties in the measurement process. Let  $\mathbf{R}_i$  represent the reality corresponding to the  $i$ th observation,  $\mathbf{D}_i$ , in dataset  $\mathcal{D}$ . Let,

$$\mathcal{R} = \{\mathbf{R}_1, \dots, \mathbf{R}_{n_{\text{do}}}\} , \quad (2)$$

represent the collection of  $n_{\text{do}}$  realities, corresponding to the observational dataset,  $\mathcal{D}$ , as illustrated in Figure 3.

Each observation,  $\mathbf{D}_i \in \mathcal{D}$ , and the corresponding reality to it,  $\mathbf{R}_i \in \mathcal{R}$ , is characterized by  $n_{\text{da}}$  variables (i.e., observable quantities), describing different characteristics of the events. Thus,  $\mathbf{R}_i$  is a vector of  $n_{\text{da}}$  elements representing a single event in the  $n_{\text{da}}$ -dimensional observational sampling space  $\Omega_{\mathbf{R}} \subset \mathbb{R}^{n_{\text{da}}}$ .

For the moment (and throughout §2 and 3), suppose we live in an ideal world where natural phenomena are observed and measured exactly and accurately, without any possible bias, inadequacy, or contamination with noise. Therefore, the observational dataset,  $\mathcal{D}$ , represents the reality,  $\mathcal{R}$ , exactly and deterministically. Hence, we will only consider modeling of the truth in this section, dealing only with  $\mathcal{R}$  and its members,  $\mathbf{R}_i$ .

Although frequently independent of each other, the attributes of each event,  $\mathbf{R}_i$ , or its occurrence could also depend on any subset,  $\mathcal{R}_{\mathbf{R}_i} \subset \mathcal{R}$ , of the other events. In general, there may also exist interdependencies between the attributes of each event (i.e., the elements of each vector  $\mathbf{R}_i$ ). A well-known generic problem of this type in engineering and natural sciences is regression, where the experimenter/observer has control over some characteristics of the events. These characteristics serve as input to the experiment and result in some experimental output that represent the response characteristics of the corresponding events.

Therefore, the attributes of a physical event can be often divided into a set of independent characteristics,  $\mathbf{R}_{\text{ind}}$ , on which the rest of the event's attributes (i.e., the dependent/response variables,  $\mathbf{R}_{\text{dep}}$ ) depend,

$$\mathbf{R}_i = \{ \mathbf{R}_i^{\text{ind}} , \mathbf{R}_i^{\text{dep}}(\mathbf{R}_i^{\text{ind}}) \} . \quad (3)$$

$\mathcal{R}$	$\mathbf{R}^{\text{ind}}$	$\mathbf{R}^{\text{dep}}$
$\mathbf{R}_1$	$R_1^1, \dots, R_1^j$	$R_1^{j+1}, \dots, R_1^{n_{\text{da}}}$
$\mathbf{R}_2$	$R_2^1, \dots, R_2^j$	$R_2^{j+1}, \dots, R_2^{n_{\text{da}}}$
$\vdots$	$\vdots$	$\vdots$
$\mathbf{R}_{n_{\text{do}}}$	$R_{n_{\text{do}}}^1, \dots, R_{n_{\text{do}}}^j$	$R_{n_{\text{do}}}^{j+1}, \dots, R_{n_{\text{do}}}^{n_{\text{da}}}$

**Table 2: The Truth Set.** The set of all physical events under study.

$\mathcal{D}$	$\mathbf{D}^{\text{ind}}$	$\mathbf{D}^{\text{dep}}$
$\mathbf{D}_1$	$D_1^1, \dots, D_1^j$	$D_1^{j+1}, \dots, D_1^{n_{\text{da}}}$
$\mathbf{D}_2$	$D_2^1, \dots, D_2^j$	$D_2^{j+1}, \dots, D_2^{n_{\text{da}}}$
$\vdots$	$\vdots$	$\vdots$
$\mathbf{D}_{n_{\text{do}}}$	$D_{n_{\text{do}}}^1, \dots, D_{n_{\text{do}}}^j$	$D_{n_{\text{do}}}^{j+1}, \dots, D_{n_{\text{do}}}^{n_{\text{da}}}$

**Table 3: The Data Set.** A result of the convolution of the truth set with measurement error.

**Fig. 3: The Experimental Measurement Process.** A schematic representation of the process of collecting observational dataset,  $\mathcal{D}$ , which is a result of the convolution of the unknown truth,  $\mathcal{R}$ , with various sources of uncertainty and measurement error during the experimental data collection process. In the absence of noise and measurement error, the set of observational data,  $\mathcal{D}$ , in Table 3 would be identical to the truth set,  $\mathcal{R}$ , in Table 2. Each physical event,  $\mathbf{R}_i$ , and the corresponding data observation,  $\mathbf{D}_i$ , is composed of a set of  $n_{\text{da}}$  event attributes, a subset of which ( $\mathbf{R}_i^{\text{dep}}$  and  $\mathbf{D}_i^{\text{dep}}$ ) are hypothesized/known to depend on the remaining  $j$  independent characteristics of the event, ( $\mathbf{R}_i^{\text{ind}}$  and  $\mathbf{D}_i^{\text{ind}}$ ). For example, in many problems, the temporal and spatial coordinates are the independent attributes of data and other physical quantities are modeled as a function of these independent attributes.

Such modeling scenarios are abundant in science and engineering [53]; for example, a set of fatigue experiments designed to measure maximum tolerable stress (the dependent variable) in a material as a function of strain (the independent variable) [10], the measurement of the growth of a malignant tumor as a function of time in a murine subject or patient [32, 59], the evolution of protein amino acid sequence as function of its structural characteristics [33, 72, 79, 80], or modeling the energetics and occurrence rates of astrophysical phenomena as a function of their distances from the earth [70, 71, 73, 75, 77, 78].

## 2.2 The Forward Problem

Now, suppose we formulate a hypothesis with regards to the set of  $n_{\text{do}}$  events in  $\mathcal{R}$ . This hypothesis can be cast in the form of a mathematical model,  $\mathbf{M}_{\text{phys}}$ , with the subscript *phys* emphasizing the type of the model (i.e., a physics-based model). This physical model can be thought of as a collection of mathematical operators (e.g., algebraic, differential, integral, ...) that takes as input, a set of  $n_{\text{pp}}$  physical parameters represented by the vector  $\boldsymbol{\theta}_{\text{phys}} \in \Theta_{\text{phys}} \subset \mathbb{R}^{n_{\text{pp}}}$ . It then acts on some or all  $n_{\text{da}}$  attributes of an independent event or a set of dependent events,  $\mathbf{R} \in \Omega_{\mathbf{R}}$ , that is input to the model and generates an output response,  $\mathbf{R}'$ , of the same length/size as  $\mathbf{R}$ , which may or may not be identical to the input  $\mathbf{R}$  (Figure 4),

$$\mathbf{M}_{\text{phys}}(\mathbf{R}, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}) = \mathbf{R}' \quad \forall \mathbf{R} \in \Omega_{\mathbf{R}}, \quad (4)$$

Given the set of events,  $\mathcal{R}$ , as in (2), the forward problem can be written as,

$$\mathbf{M}_{\text{phys}}(\mathcal{R}, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}) = \mathcal{R}' \quad \forall \mathcal{R} \in \Omega_{\mathcal{R}}. \quad (5)$$

where,

$$\mathcal{R}' = \{\mathbf{R}'_1, \dots, \mathbf{R}'_{n_{\text{do}}}\}, \quad (6)$$

is the output of the physical model,  $\mathbf{M}_{\text{phys}}$ , corresponding to the event set,  $\mathcal{R}$ . When the events in  $\mathcal{R}$  are independent of each other, the forward problem takes the more simplified form,

$$\mathbf{M}_{\text{phys}}(\mathbf{R}_i, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}) = \mathbf{R}'_i \quad \forall \mathbf{R}_i \in \mathcal{R} \subset \Omega_{\mathbf{R}}. \quad (7)$$

In general, the observational information about each event,  $\mathbf{R}_i$ , might be collected under a specific set of conditions collectively named here as the *physical scenario*,  $\mathcal{S}_{\text{phys}}$ . The scenario,





**Fig. 4: Schematic diagrams of the physical model with its input and output arguments.** The physical model takes as input a set of values for the model parameters,  $\theta_{\text{phys}}$ , as well as the set of attributes,  $\mathbf{R}$ , of an event or a collection of events. It then outputs a vector,  $\mathbf{R}'$ , of the same length/size as  $\mathbf{R}$ , whose elements may or may not be identical to the corresponding elements in  $\mathbf{R}$ , depending on  $\theta_{\text{phys}}$  and the validity of the physical model. **(a) A misspecified physical model** whose output for *at least* one event,  $\mathbf{R}_i$ , in the event set,  $\mathcal{R}$ , does not match the input vector,  $\mathbf{R}_i$ , for any possible values of  $\theta_{\text{phys}}$ . **(b) A correctly-specified physical model** for which there is at least one set of input parameter values,  $\hat{\theta}_{\text{phys}}$ , that result in an output from the physical model which is identical to the input,  $\mathbf{R}_i$ , for all  $\mathbf{R}_i \in \mathcal{R}$  [53].

$\mathcal{S}_{\text{phys}}$ , describes the set of all features of the scientific problem at hand, that can be exactly specified, that is also independent of the data,  $\mathcal{R}$ , the physical model,  $\mathbf{M}_{\text{phys}}$ , and its parameters,  $\theta_{\text{phys}}$ . These features typically describe and depend on the experimental/observational setup for data collection [53].

For example,  $\mathcal{S}_{\text{phys}}$  could describe the limitations of a detector that was used for data collection [70, 71, 73–78, 78]. Thus, although  $\mathcal{S}_{\text{phys}}$  is independent of  $\mathcal{R}$ , the observation of individual events in the dataset,  $\mathcal{D}$ , and hence the set  $\mathcal{R}$ , is not independent of  $\mathcal{S}_{\text{phys}}$  in general.

### 2.3 The Inverse Problem

If the proposed model is capable of describing the phenomenon of interest perfectly without any bias or inadequacy, then by definition, there should exist at least one set of feasible parameter values  $\theta_{\text{phys}} = \hat{\theta}_{\text{phys}} \in \Theta_{\text{phys}}$  for which,

$$\mathcal{R} - \mathbf{M}_{\text{phys}}(\mathcal{R}, \hat{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}) = \mathbf{0} , \quad (8)$$

where  $\mathbf{0}$  is the null matrix of size  $n_{\text{do}} \times n_{\text{da}}$ . In the case of independent events,

$$\mathbf{R}_i - \mathbf{M}_{\text{phys}}(\mathbf{R}_i, \hat{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}) = \mathbf{0} \quad \forall \mathbf{R}_i \in \mathcal{R} , \quad (9)$$

where  $\mathbf{0}$  is now the null vector of length  $n_{\text{da}}$ . Thus, given a perfect model and an ideal dataset with no measurement error, the problem of inference is reduced to solving (8) or a system of  $n_{\text{do}}$  equations of the form (9) to obtain the feasible values,  $\hat{\theta}_{\text{phys}}$ , of the set of parameters of the physical model satisfying (8) or (9).

In mathematical modeling, this process is widely known as *inverse problem* or *model calibration*, as illustrated in Figure 1a. When there is a  $\hat{\theta}_{\text{phys}}$  for which the physical model,  $\mathbf{M}_{\text{phys}}$ , perfectly describes the truth (by satisfying (8) or (9)), then  $\mathbf{M}_{\text{phys}}$  is said to be a *correctly-specified* or *well-specified* model [53] as illustrated in Figure 4b. In some problems, there might exist multiple or even uncountably infinite number of  $\hat{\theta}_{\text{phys}}$  that satisfy (8). In such cases, the system of equations is said to be degenerate or ill-posed in the sense of Hadamard [27].

### 3 Modeling the Truth in the Presence of Model Inadequacy

One can envisage many experimental setups in which the same input to the experiment, results in a variety of possible outcomes. For example, the experiment of throwing an unbiased die un-

der similar conditions yields six inherently different possible outcomes. A more relevant example to the field of engineering is a set of stress-strain data obtained from a heterogeneous material. In such data, a single strain could yield several different stress values in repeated identical experiments, depending on the level of heterogeneity and imperfection of the material being tested.

Such apparently unexplained variabilities and heterogeneities in observational data with respect to the predictions of the physical model, fall into the category of *model inadequacy*, or *model discrepancy*, sometimes also referred to it as *structural uncertainty* or *aleatoric uncertainty*. This is a class of uncertainty that presumably cannot be reduced with further collection of data or more accurate measurements of *the same characteristic features* already describing each event,  $\mathbf{R}_i \in \mathcal{R}$ .

Regardless of the origins of model inadequacy – whether it be a wrong physical model or insufficiently-detailed data – such discrepancies between model predictions and data almost always exist, and their behavior has to be described by yet another set of models of stochastic origin.

### 3.1 The Construction of Inadequacy Model

When there is unexplained variability in the events with respect to the predictions of the physical model, the model is said to be *inadequate* or *misspecified* [53], and the output of the physical model for some or all events,  $\mathbf{R}_i \in \mathcal{R}$ , does not match the available data. Hence, the equalities in (8) and (9) do not hold. Rather we have in general,

$$\mathcal{R} - \mathbf{M}_{\text{phys}}(\mathcal{R}, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}) = \mathcal{U} , \quad (10)$$

where  $\mathcal{U}$  is set of  $n_{\text{do}}$  vectors,  $\mathbf{U}_i$ , each of which corresponds to one event,  $\mathbf{R}_i \in \mathcal{R}$ . In the case of independent events, (10) takes the more simplified form,

$$\mathbf{R}_i - \mathbf{M}_{\text{phys}}(\mathbf{R}_i, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}) = \mathbf{U}_i \quad \forall \mathbf{R}_i \in \mathcal{R} . \quad (11)$$

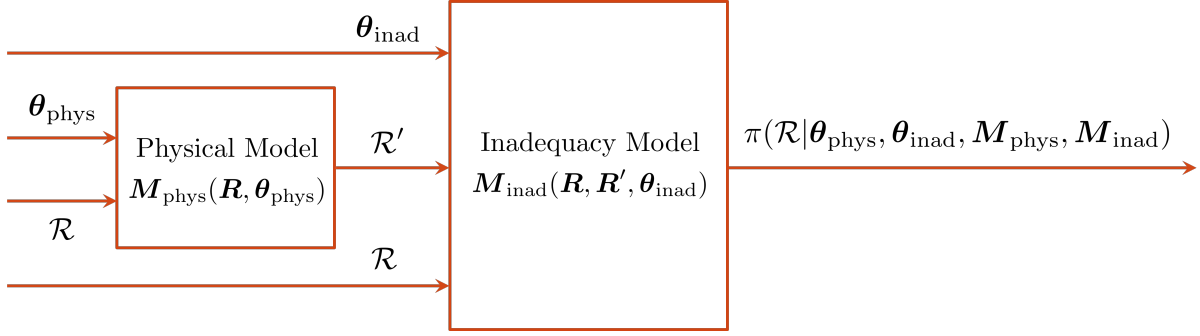
Each  $\mathbf{U}_i$  is a vector of length  $n_{\text{da}}$ , some elements of which, corresponding to the response variables,  $\mathbf{R}_i^{\text{dep}} \subset \mathbf{R}_i$ , are not anymore deterministic but random stochastic variables. In other words, for the same control variables  $\mathbf{R}^{\text{ind}} \subset \mathbf{R} = \{\mathbf{R}^{\text{dep}}, \mathbf{R}^{\text{dep}}\}$ , there can be a finite or infinite number of possible values for the response vector  $\mathbf{R}^{\text{dep}}$ . But the physical model is only capable of fitting some representative deterministic average output response to  $\mathbf{R}_i^{\text{dep}} \subset \mathbf{R}_i$ ,  $\forall \mathbf{R}_i \in \mathcal{R}$ . Therefore, the set,

$$\mathcal{U} = \{\mathbf{U}_1, \dots, \mathbf{U}_{n_{\text{do}}}\} , \quad (12)$$

is a collection (i.e., a  $n_{\text{do}} \times n_{\text{da}}$  matrix) of deterministic and stochastic variables whose behavior has to be determined by yet another model,  $\mathbf{M}_{\text{inad}}$ , of statistical and physical origin (as opposed to the deterministic physical origin of  $\mathbf{M}_{\text{phys}}$  (Figure 5). This stochastic model depends explicitly on the output of the proposed physical model,  $\mathbf{M}_{\text{phys}}$ ,

$$\mathcal{U} \sim \mathbf{M}_{\text{inad}}(\mathcal{R}, \boldsymbol{\theta}_{\text{inad}}, \mathbf{M}_{\text{phys}}(\mathcal{R}, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}})) , \quad (13)$$

where  $\boldsymbol{\theta}_{\text{inad}}$  represents the set of  $n_{\text{pi}}$  parameters of  $\mathbf{M}_{\text{inad}}$ . Here, the subscript *inad* stands for *inadequacy*. In the case of independent and identically distributed (i.i.d.) events, the above



**Fig. 5:** A schematic illustration of the relationship between the physical model,  $\mathbf{M}_{\text{phys}}$ , and the inadequacy model,  $\mathbf{M}_{\text{inad}}$ : the statistical model that describes the inadequacy of the physical model with respect to the available data. On input,  $\mathbf{M}_{\text{inad}}$  takes a set of values for its parameters,  $\boldsymbol{\theta}_{\text{inad}}$ , the output of the physical model,  $\mathcal{R}'$ , and the truth about the collection of events,  $\mathcal{R}$ . On output,  $\mathbf{M}_{\text{inad}}$  gives the probability of the truth being  $\mathcal{R}$ , given the physical model's output,  $\mathcal{R}'$ , and the specific values of the parameters of the physical and inadequacy models:  $\{\boldsymbol{\theta}_{\text{phys}}, \boldsymbol{\theta}_{\text{inad}}\}$ .

equation can be written as,

$$U_i \sim \mathbf{M}_{\text{inad}}(\mathbf{R}_i, \boldsymbol{\theta}_{\text{inad}}, \mathbf{M}_{\text{phys}}(\mathbf{R}_i, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}})) . \quad (14)$$

The special case where the parameters of  $\mathbf{M}_{\text{inad}}$  depend explicitly on  $\mathbf{R}_i^{\text{ind}} \subset \mathbf{R}_i$ , is known as *heteroscedasticity* in the statistical literature [63].

### 3.2 The Likelihood Function in the Presence of Model Inadequacy

In the presence of model inadequacy, the inference problem cannot be solved deterministically anymore as in §2.3, rather one has to first derive the likelihood function (i.e., the unnormalized density function),  $\mathcal{L}$ , of the parameters of the physical and inadequacy models. Let,

$$\mathbf{M}_{\text{pi}} = \{\mathbf{M}_{\text{phys}}, \mathbf{M}_{\text{inad}}\} , \quad (15)$$

$$\boldsymbol{\theta}_{\text{pi}} = \{\boldsymbol{\theta}_{\text{phys}}, \boldsymbol{\theta}_{\text{inad}}\} , \quad (16)$$

represent respectively, the combined physical and inadequacy models for  $\mathcal{R}$ , and the combined set of parameters of the two models. Then, by definition,

$$\pi(\mathcal{R} | \boldsymbol{\theta}_{\text{pi}}, \mathbf{M}_{\text{pi}}) \equiv \pi(\mathcal{U} | \boldsymbol{\theta}_{\text{pi}}, \mathbf{M}_{\text{pi}}) , \quad (17)$$

represents the multivariate probability density function (PDF),  $\pi(\cdot)$ , of obtaining  $\mathcal{R}$  given  $\mathbf{M}_{\text{pi}}$  and the parameter values  $\boldsymbol{\theta}_{\text{pi}}$ , which in the case of i.i.d. events simplifies to,

$$\pi(\mathbf{R}_i | \boldsymbol{\theta}_{\text{pi}}, \mathbf{M}_{\text{pi}}) \equiv \pi(\mathbf{U}_i | \boldsymbol{\theta}_{\text{pi}}, \mathbf{M}_{\text{pi}}) \quad \forall \mathbf{R}_i \in \mathcal{R} , . \quad (18)$$

Then, the likelihood of  $\boldsymbol{\theta}_{\text{pi}}$  being the correct set of values for the parameters of  $\mathbf{M}_{\text{pi}}$ , in the light of available data,  $\mathcal{R}$ , becomes,

$$\mathcal{L}(\boldsymbol{\theta}_{\text{pi}}; \mathcal{R}) \equiv \mathcal{L}(\boldsymbol{\theta}_{\text{pi}}; \mathcal{U}) \quad (19)$$

$$\equiv \pi(\mathcal{R} | \boldsymbol{\theta}_{\text{pi}}, \mathbf{M}_{\text{pi}}) \quad (20)$$

$$\stackrel{\text{i.i.d.}}{=} \prod_{i=1}^{\text{ndo}} \pi(\mathbf{R}_i | \boldsymbol{\theta}_{\text{inad}}, \mathbf{M}_{\text{inad}}(\mathbf{R}_i, \boldsymbol{\theta}_{\text{inad}}, \mathbf{M}_{\text{phys}}(\mathbf{R}_i, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}))) , \quad (21)$$

where again, the last equality, (21), holds only on the special occasion where the observed events are independent of each other and are equally likely to occur, that is, independent and identically distributed (i.i.d.). Although the i.i.d. property represents a special case, in practice it holds for a wide range of scientific inference and modeling problems.

The problem of inference is now reduced to finding a  $\boldsymbol{\theta}_{\text{pi}} = \{\boldsymbol{\theta}_{\text{phys}}, \boldsymbol{\theta}_{\text{inad}}\}$  such that the joint probability of obtaining all events in  $\mathcal{R}$  together, combined with any prior knowledge about the parameters of the models,  $\mathbf{M}_{\text{pi}} = \{\mathbf{M}_{\text{phys}}, \mathbf{M}_{\text{inad}}\}$ , is maximized. From the Bayes rule, it follows,

$$\pi(\boldsymbol{\theta}_{\text{pi}}|\mathcal{U}, \mathbf{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}) = \frac{\pi(\mathcal{U}|\boldsymbol{\theta}_{\text{pi}}, \mathbf{M}_{\text{pi}})\pi(\boldsymbol{\theta}_{\text{pi}}|\mathbf{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}})}{\pi(\mathcal{U}|\mathbf{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}})}, \quad (22)$$

or equivalently,

$$\pi(\boldsymbol{\theta}_{\text{pi}}|\mathcal{R}, \mathbf{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}) = \frac{\pi(\mathcal{R}|\boldsymbol{\theta}_{\text{pi}}, \mathbf{M}_{\text{pi}})\pi(\boldsymbol{\theta}_{\text{pi}}|\mathbf{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}})}{\pi(\mathcal{R}|\mathbf{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}})}, \quad (23)$$

where  $\mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}$  represents any prior knowledge about all unknown parameters,  $\boldsymbol{\theta}_{\text{pi}}$ , of the physical and inadequacy models together. The subject of constructing the prior PDF,  $\pi(\boldsymbol{\theta}_{\text{pi}}|\mathbf{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}})$ , from the available knowledge in an inference problem is as old as the Bayesian probability theory itself. Over the past century, several methods such as *Jeffreys' principle of invariance under reparametrization* [41] or *Jaynes' principle of maximum entropy* [26, 35–37, 39, 39] have been developed to construct objective priors for Bayesian inference.

### 3.3 Popular Choices of Inadequacy Model

The Multivariate Normal (MVN) distribution is undoubtedly the most popular and widely-used choice of probability density function for inadequacy models in scientific inference, although it has appeared under different names, notably the Least-Squares method first introduced by Adrien Legendre in 1805 [46]. Another popular choice of inadequacy model is the Laplace distribution, more commonly known as the Least Absolute Deviation method first introduced by Pierre Laplace in 1774 [43, 45].

Despite their popularity, neither MVN nor Laplace are the most appropriate choices of inadequacy model,  $\mathbf{M}_{\text{inad}}$ , for every scientific inference problem. Since  $\mathbf{M}_{\text{inad}}$  represents the inadequacy of the physical model, in general, it has to be also inferred from the characteristics of data and the physical phenomena being investigated.

## 4 Modeling the Truth, Confounded with Noise (Measurement Error)

Independently of the model inadequacy, experimental observations are always contaminated with measurement error (or equivalently as we use hereafter, noise). Such sources of uncertainty in data are sometimes called *epistemic uncertainty*, although the scope of epistemic uncertainty goes beyond measurement error, as we discussed in §1.

Unlike model inadequacy, uncertainty due to measurement error is a result of the fundamental limitations of the measurement process and instruments. It is therefore assumed to be

reducible by gathering higher quality information about the phenomenon of interest, for example, by making more accurate measurements with more accurate devices.

The inevitable existence of noise in every real-world experiment implies that the truth,  $\mathbf{R}$ , about an event will never be known to the observer/experimenter unless the uncertainty in data is exactly and *deterministically* modeled and removed. This is, however, impossible since the effects of noise on reality are virtually never known deterministically. Under the most optimistic scenarios, an experimenter may only be able to make an educated guess on the general average stochastic effects of the noise on the experimental measurements.

In other words, what an observer/experimenter can perceive about the truth,  $\mathbf{R}$ , of an event is only the *stochastic* output,  $\mathbf{D}$ , of a complex measurement process whose input is  $\mathbf{R}$ . Therefore, a primary and major task in designing and performing an experiment is to alleviate to the extent possible, the effects of this stochastic noise component on inferences made about the truth.

Under a well-determined experimental setup and measurement process, one may be able to provide a stochastic model for the random effects of noise on the truth. Let,

$$\mathbf{D}_i \sim \mathbf{M}_{\text{nois},i}(\mathbf{R}_i, \boldsymbol{\theta}_{\text{nois},i}) , \quad (24)$$

represent the  $i$ th observational data in the dataset  $\mathcal{D} = \{\mathbf{D}_1, \dots, \mathbf{D}_{n_{\text{do}}}\}$  which consists of  $n_{\text{do}}$  observations. Here,  $\mathbf{M}_{\text{nois},i}$  represents the stochastic noise model for the  $i$ th event,  $\mathbf{R}_i$ . Each observation,  $\mathbf{D}_i \in \mathcal{D}$ , results from a complex convolution of the truth,  $\mathbf{R}_i \in \mathcal{R}$ , with various types of noise in the measurement process, all of which we hope to approximate and model by  $\mathbf{M}_{\text{nois},i}$ .

The noise model,  $\mathbf{M}_{\text{nois},i}$ , takes as input, the set of  $n_{\text{pn},i}$  parameters represented by the vector  $\boldsymbol{\theta}_{\text{nois},i}$ . Here the subscript  $i$ , wherever it appears in (24) and throughout this section, is used to indicate that the corresponding object specifically belongs to or is defined in relation with the  $i$ th event,  $\mathbf{R}_i$ . The subscript  $\text{pn}$  in  $n_{\text{pn},i}$  stands for the *p*arameters of the *n*oise model.

A typical example of such inference problem, involving distinct noise models corresponding to each  $\mathbf{R}_i \in \mathcal{R}$ , can be found in the field of Astronomy and Cosmology where observational data is frequently gathered by multiple instruments of different measurement accuracies [71, 74, 77, 78].

Since  $\mathbf{D}_i$  is a random variable, there is *no* one-to-one mapping between  $\mathbf{D}_i$  and the truth  $\mathbf{R}_i$ . Given the noise model,  $\mathbf{M}_{\text{nois},i}$ , and its parameters,  $\boldsymbol{\theta}_{\text{nois},i}$ , the corresponding PDF of  $\mathbf{D}_i$  is,

$$\text{PDF}(\mathbf{D}_i) \stackrel{\text{def}}{=} \pi(\mathbf{D}_i | \mathbf{R}_i, \boldsymbol{\theta}_{\text{nois},i}, \mathbf{M}_{\text{nois},i}) \quad (25)$$

There are multiple reasons that render (25) useless by itself, even when the exact mathematical form of  $\mathbf{M}_{\text{nois},i}$  as well as its parameters,  $\boldsymbol{\theta}_{\text{nois},i}$ , are known:

- Firstly, this formulation requires us to know the truth  $\mathbf{R}_i$  as an input to  $\mathbf{M}_{\text{nois},i}$ . In practice however, we never know the truth.
- Secondly, the observer/experimenter can never obtain multiple realizations of  $\mathbf{D}_i$  for the exact same truth  $\mathbf{R}_i$ . What an experimenter gathers about an event,  $\mathbf{R}_i$ , is a single observation of it,  $\mathbf{D}_i$ , which is a result of the convolution of  $\mathbf{R}_i$  with noise. From a Bayesian perspective,

such experiments can never be repeated under the exact same conditions to obtain multiple observations,  $\mathbf{D}_i$ , for the same  $\mathbf{R}_i$ .

- Thirdly and most importantly, the input quantities to the physical model,  $\mathbf{M}_{\text{phys}}$ , in (8) and (9) are  $\mathcal{R}$  and  $\mathbf{R}_i$  respectively, not  $\mathcal{D}$  and  $\mathbf{D}_i$ .

A better formulation of the problem can be obtained by asking an appropriate question in relation to noise in data: Given a single observation,  $\mathbf{D}_i$ , obtained for an unknown event,  $\mathbf{R}_i$ , what is the probability that the underlying truth about this event is  $\mathbf{R}_i^*$ ? Here the superscript  $*$  is to emphasize that  $\mathbf{R}_i^*$  may not necessarily correspond to the truth  $\mathbf{R}_i$ .

This question has a straightforward answer using the Bayesian inversion method applied to (25), such that the probability of  $\mathbf{R}_i^*$  being the truth about the  $i$ th event could be written as,

$$\pi(\mathbf{R}_i^* | \mathbf{D}_i, \boldsymbol{\theta}_{\text{nois},i}, \mathbf{M}_{\text{nois},i}, \mathcal{I}_{\mathbf{R}_i}) = \frac{\pi(\mathbf{D}_i | \mathbf{R}_i^*, \boldsymbol{\theta}_{\text{nois},i}, \mathbf{M}_{\text{nois},i}) \pi(\mathbf{R}_i^* | \mathcal{I}_{\mathbf{R}_i})}{\pi(\mathbf{D}_i | \boldsymbol{\theta}_{\text{nois},i}, \mathbf{M}_{\text{nois},i}, \mathcal{I}_{\mathbf{R}_i})}, \quad (26)$$

where the left-hand-side of the equation is the posterior probability density function of the truth, the first term in the numerator is the likelihood function of  $\mathbf{R}_i^*$ , which is equivalent to the probability density function of  $\mathbf{D}_i$  in (25), and  $\mathcal{I}_{\mathbf{R}_i}$  represents the experimenter's prior knowledge about  $\mathbf{R}_i$ . The denominator is simply a normalization factor (the Bayesian evidence) that makes the left-hand-side a properly normalized PDF. It gives the probability of observing  $\mathbf{D}_i$  averaged over all possible values,  $\mathbf{R}_i^*$ , for the truth,  $\mathbf{R}_i$ .

In practice, the state of experimenter's prior knowledge,  $\mathcal{I}_{\mathbf{R}_i}$ , about the truth is often *complete ignorance*, meaning that all possibilities for the truth are equally probable in the eyes of the experimenter, regardless of the noise model. Therefore,  $\pi(\mathbf{R}_i^* | \mathcal{I}_{\mathbf{R}_i})$  is frequently assigned an improper (unbounded) uniform distribution [35, 36, 39], often without explicitly acknowledging it. There are, however, important exceptions to this general rule, for example, in hierarchical Bayesian inference problems [71, 78] where  $\mathbf{R}_i$  is defined only on a subset of the real numbers,  $\mathbb{R}$ .

#### 4.1 Noise Models Are Fundamentally Different From Inadequacy Models

Unlike the case of model inadequacy studied in §3 where the parameters,  $\boldsymbol{\theta}_{\text{inad}}$ , of the inadequacy model,  $\mathbf{M}_{\text{inad}}$ , were a priori unknown and were to be inferred from data, here it is essential to know the mathematical form of  $\mathbf{M}_{\text{nois},i}$  (i.e., the PDF in (25)) for each observation,  $\mathbf{D}_i$ , as well as its parameter values,  $\boldsymbol{\theta}_{\text{nois},i}$ . In other words, the noise model and its parameters are part of observational data that will have to be fed to the physical and inadequacy models. Otherwise, it would be impossible to construct (26), and subsequently solve (8).

Another difference with the case of inadequacy model is the fact that the presence of noise in data does not necessarily invalidate (8). As long as the physical model,  $\mathbf{M}_{\text{phys}}$ , is perfectly capable of describing  $\mathcal{R}$  – an ideal assumption that we make in this section (i.e., §4) – and the sampling space of  $\mathbf{R}_i^*$  is accurately defined such that,

$$\pi(\mathbf{R}_i | \mathbf{D}_i, \boldsymbol{\theta}_{\text{nois},i}, \mathbf{M}_{\text{nois},i}) \neq 0 \quad \forall \mathbf{R}_i \in \mathcal{R}, \quad (27)$$

then, there will be at least one set of physical parameters,  $\hat{\theta}_{\text{phys}}$ , for which  $\mathcal{R}^* = \mathcal{R}$  satisfies (8),

$$\mathcal{R}^* - \mathcal{R}^{*'} = \mathbf{0} , \quad (28)$$

where,

$$\mathcal{R}^{*'} = M_{\text{phys}}(\mathcal{R}^*, \theta_{\text{phys}}, \mathcal{S}_{\text{phys}}) , \quad (29)$$

or in the case of i.i.d. events,

$$\mathbf{R}_i^* - \mathbf{R}_i^{*'} = \mathbf{0} \quad \forall \mathbf{R}_i^* \in \mathcal{R}^* , \quad (30)$$

with,

$$\mathbf{R}_i^{*'} = M_{\text{phys}}(\mathbf{R}_i^*, \theta_{\text{phys}}, \mathcal{S}_{\text{phys}}) \quad \forall \mathbf{R}_i^* \in \mathcal{R}^* . \quad (31)$$

The symbol  $\mathcal{R}^*$  in (28) represents a set of possible realizations,  $\mathbf{R}_i^*$ , of each of the events,  $\mathbf{R}_i \in \mathcal{R}$ ,

$$\mathcal{R}^* = \{\mathbf{R}_1^*, \dots, \mathbf{R}_{n_{\text{do}}}^*\} . \quad (32)$$

Therefore, the set  $\mathcal{R}^*$  represents a possible realization of  $\mathcal{R}$ , which may satisfy (28) for no  $\theta_{\text{phys}}$  values, or a single value, or multiple values, or an infinite number of values of  $\theta_{\text{phys}}$ . Let,

$$\mathcal{M}_{\text{nois}} = \{M_{\text{nois},1}, \dots, M_{\text{nois},n_{\text{do}}}\} , \quad (33)$$

$$\Theta_{\text{nois}} = \{\theta_{\text{nois},1}, \dots, \theta_{\text{nois},n_{\text{do}}}\} , \quad (34)$$

denote respectively, the set of noise models corresponding to each event in  $\mathcal{R}$ , and the set of ‘a priori known’ parameters of each model,  $M_{\text{nois},i} \in \mathcal{M}_{\text{nois}}$ . Then, one can write the probability of  $\mathcal{R}^*$  being the truth,  $\mathcal{R}$ , via the Bayes rule as,

$$\pi(\mathcal{R}^* | \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) = \frac{\pi(\mathcal{D} | \mathcal{R}^*, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}) \pi(\mathcal{R}^* | \mathcal{I}_{\mathcal{R}})}{\pi(\mathcal{D} | \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}})} \quad (35)$$

$$\stackrel{\text{i.i.d.}}{=} \prod_{i=1}^{n_{\text{do}}} \pi(\mathbf{R}_i^* | D_i, \theta_{\text{nois},i}, M_{\text{nois},i}, \mathcal{I}_{\mathbf{R}_i}) \quad (36)$$

$$\stackrel{\text{i.i.d.}}{=} \prod_{i=1}^{n_{\text{do}}} \frac{\pi(D_i | \mathbf{R}_i^*, \theta_{\text{nois},i}, M_{\text{nois},i}) \pi(\mathbf{R}_i^* | \mathcal{I}_{\mathbf{R}_i})}{\pi(D_i | \theta_{\text{nois},i}, M_{\text{nois},i}, \mathcal{I}_{\mathbf{R}_i})} . \quad (37)$$

where  $\mathcal{I}_{\mathcal{R}}$  represents the experimenter’s prior knowledge about the entire set  $\mathcal{R}$ . The latter two equalities hold only under the assumption of independent and identical distribution (i.i.d.) of the experimental measurements,  $D_i \in \mathcal{D}$ .

#### 4.2 The Likelihood Function in the Presence of Data Uncertainty

In most practical inference problems, the noise model,  $M_{\text{nois},i}$ , for a given event,  $\mathbf{R}_i \in \mathcal{R}$ , has a continuous probability density function for  $\mathbf{R}_i^*$ , leading to infinitely many possibilities,  $\mathbf{R}_i^*$ , for each event,  $\mathbf{R}_i$ . This means that the set  $\mathcal{R}^*$  is not unique. Therefore, one can construct a superset,  $\Omega_{\mathcal{R}^*}$ , consisting of all possible combinatorial realizations of the set of events in  $\mathcal{R}$ ,

$$\Omega_{\mathcal{R}^*} = \{\mathcal{R}_1^*, \dots, \mathcal{R}_j^*, \dots\} , \quad (38)$$

where *only and only* one realization,  $\mathcal{R}^*_j \in \Omega_{\mathcal{R}^*}$ , corresponds to the truth set,  $\mathcal{R}$ . Many members of  $\Omega_{\mathcal{R}^*}$  might satisfy (28) for some specific parameter values,  $\theta_{\text{phys}}$ , that is not necessarily the same parameter values for which  $\mathcal{R}$  satisfies (28). To the contrary, there might exist other members of  $\Omega_{\mathcal{R}^*}$  that do not satisfy (28) for any possible parameter values,  $\theta_{\text{phys}}$ .

The superset  $\Omega_{\mathcal{R}^*}$  can be finite, or countably/uncountably infinite, depending on the type of measurement uncertainties involved in the problem. When all of the noise models,  $\mathbf{M}_{\text{nois},i} \in \mathcal{M}_{\text{nois}}$ , for the specific problem under study give rise to sets of finite possibilities,  $\mathbf{R}^*_i$ , for the corresponding events,  $\mathbf{R}_i \in \mathcal{R}$ , then the size of  $\Omega_{\mathcal{R}^*}$  would be also finite. Otherwise, if any or all of the noise models are countably/uncountably infinite, then  $\Omega_{\mathcal{R}^*}$  would be also countably/uncountably infinite.

Now, let  $\Omega_{\mathcal{R}^*}(\theta_{\text{phys}}) \subset \Omega_{\mathcal{R}^*}$  represent the set of all possible realizations,  $\mathcal{R}^* \in \Omega_{\mathcal{R}^*}$ , of  $\mathcal{R}$  that satisfy (28) for the given  $\theta_{\text{phys}}$ . Then, the likelihood (i.e., the unnormalized probability) of  $\theta_{\text{phys}}$  being the true set of parameter values for  $\mathbf{M}_{\text{phys}}$  can be written as the sum of the probabilities of all possible realizations,  $\mathcal{R}^* \in \Omega_{\mathcal{R}^*}(\theta_{\text{phys}})$ ,

$$\mathcal{L}(\theta_{\text{phys}}) \equiv \pi(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} | \theta_{\text{phys}}, \mathbf{M}_{\text{phys}}, \mathcal{I}_{\mathcal{R}}) \quad (39)$$

$$= \int_{\Omega_{\mathcal{R}^*}(\theta_{\text{phys}})} \pi(\mathcal{R}^* | \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) d\mathcal{R}^* \quad (40)$$

$$= \int_{\Omega_{\mathcal{R}^*}} \mathbf{1}(\mathcal{R}^*) \pi(\mathcal{R}^* | \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) d\mathcal{R}^* \quad (41)$$

$$\stackrel{\text{i.i.d.}}{=} \prod_{i=1}^{n_{\text{do}}} \pi(\mathbf{D}_i, \theta_{\text{nois},i}, \mathbf{M}_{\text{nois},i} | \theta_{\text{phys}}, \mathbf{M}_{\text{phys}}, \mathcal{I}_{\mathbf{R}_i}) \quad (42)$$

$$\stackrel{\text{i.i.d.}}{=} \prod_{i=1}^{n_{\text{do}}} \int_{\Omega_{\mathcal{R}^*}} \mathbf{1}(\mathbf{R}^*_i) \pi(\mathbf{R}^*_i | \mathbf{D}_i, \theta_{\text{nois},i}, \mathbf{M}_{\text{nois},i}, \mathcal{I}_{\mathbf{R}_i}) d\mathbf{R}^*_i, \quad (43)$$

where  $\mathbf{1}(\cdot)$  is an indicator function which has the value 1, only if  $\mathcal{R}^*$  and  $\mathbf{R}^*_i$  satisfy (28) and (30) respectively for the given value of  $\theta_{\text{phys}}$ , otherwise it is 0,

$$\mathbf{1}(\mathbf{R}^*_i) \stackrel{\text{def}}{=} \begin{cases} 1 & \mathbf{R}^{*'}_i = \mathbf{R}^*_i, \\ 0 & \text{otherwise,} \end{cases}, \quad 1 < i < n_{\text{do}}, \quad (44)$$

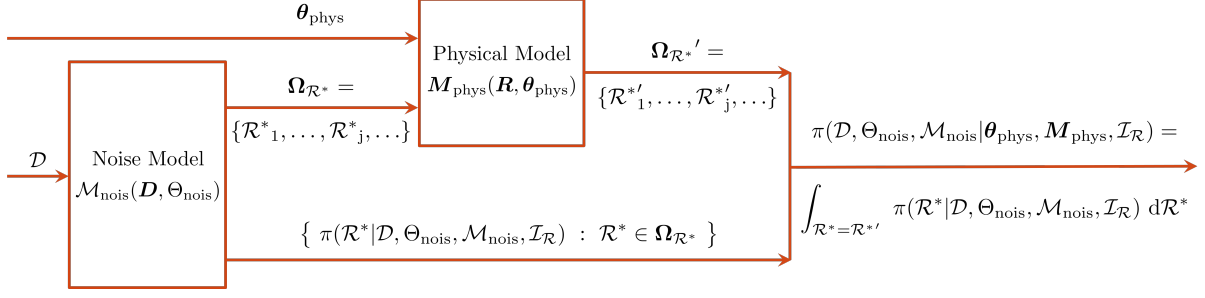
$$\mathbf{1}(\mathcal{R}^*) = \prod_{i=1}^{n_{\text{do}}} \mathbf{1}(\mathbf{R}^*_i) \quad (45)$$

$$\stackrel{\text{def}}{=} \begin{cases} 1 & \mathcal{R}^{*'} = \mathcal{R}^*, \\ 0 & \text{otherwise.} \end{cases} \quad (46)$$

where  $\mathcal{R}^{*'}$  and  $\mathbf{R}^{*'}_i$  are defined by (29) and (31). The equality in (43) holds only under the assumption of independent and identical distribution (i.i.d.) of the experimental measurements,  $\mathbf{D}_i \in \mathcal{D}$ . A schematic diagram illustrating the derivation of this likelihood function is given in Figure 6.

Thus, the problem of inference is now to compute and maximize the PDF of  $\theta_{\text{phys}}$ , which can be obtained by combining (40) with any prior knowledge,  $\mathcal{I}_{\theta_{\text{phys}}}$ , about the parameters of





**Fig. 6:** A schematic illustration of the derivation of the likelihood function when the physical model,  $\mathbf{M}_{\text{phys}}$ , has no inadequacy and the observational dataset,  $\mathcal{D}$ , is subject to noise and measurement error. In such scenario, if the set of noise models and their parameters,  $\mathcal{M}_{\text{nois}}(\Theta_{\text{nois}})$ , affecting the data are known, then one has to consider all possibilities,  $\{\mathcal{R}^*_1, \dots, \mathcal{R}^*_j, \dots\}$ , for the unknown truth,  $\mathcal{R}$ , underlying the observational data,  $\mathcal{D}$ . All such possibilities are then fed to the physical model, which yield the corresponding output set  $\{\mathcal{R}^{*'}_1, \dots, \mathcal{R}^{*'}_j, \dots\}$ . Then, the probability of observing  $\mathcal{D}$  given a specific set of values for the parameters,  $\theta_{\text{phys}}$ , of the physical model, is the sum of the probabilities of all possible realizations,  $\mathcal{R}^*$ , of the truth,  $\mathcal{R}$ , that satisfy the physical model (i.e., when  $\mathcal{R}^* = \mathcal{R}'$ ).

the model using the Bayes rule,

$$\begin{aligned} \pi(\theta_{\text{phys}} \mid \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathbf{M}_{\text{phys}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{phys}}}) \\ = \frac{\pi(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} \mid \theta_{\text{phys}}, \mathbf{M}_{\text{phys}}, \mathcal{I}_{\mathcal{R}}) \pi(\theta_{\text{phys}} \mid \mathbf{M}_{\text{phys}}, \mathcal{I}_{\theta_{\text{phys}}})}{\pi(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} \mid \mathbf{M}_{\text{phys}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{phys}}})} \end{aligned} \quad (47)$$

$$= \frac{\left[ \int_{\Omega_{\mathcal{R}^*}(\theta_{\text{phys}})} \pi(\mathcal{R}^* \mid \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) d\mathcal{R}^* \right] \pi(\theta_{\text{phys}} \mid \mathbf{M}_{\text{phys}}, \mathcal{I}_{\theta_{\text{phys}}})}{\pi(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} \mid \mathbf{M}_{\text{phys}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{phys}}})}, \quad (48)$$

where the denominator is simply a factor that properly normalizes the posterior distribution of  $\theta_{\text{phys}}$  to a proper probability density function, such that the integral of (48) over the entire parameter space,  $\Theta_{\text{phys}}$ , is 1,

$$\begin{aligned} \pi(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} \mid \mathbf{M}_{\text{phys}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{phys}}}) = \\ \int_{\Theta_{\text{phys}}} \left[ \int_{\Omega_{\mathcal{R}^*}(\theta_{\text{phys}})} \pi(\mathcal{R}^* \mid \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) d\mathcal{R}^* \right] \pi(\theta_{\text{phys}} \mid \mathbf{M}_{\text{phys}}, \mathcal{I}_{\theta_{\text{phys}}}) d\theta_{\text{phys}}. \end{aligned} \quad (49)$$

It is clear from (40), (41), (43), and (48), that the posterior PDF of the parameters of the physical model,  $\pi(\theta_{\text{phys}} \mid \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathbf{M}_{\text{phys}}, \mathcal{I}_{\theta_{\text{phys}}})$ , depends only implicitly on the physical model,  $\mathbf{M}_{\text{phys}}$ , and its parameters,  $\theta_{\text{phys}}$ . The only influence of  $\mathbf{M}_{\text{phys}}$  and  $\theta_{\text{phys}}$  on the posterior PDF of (48) is through the definition of the domain of integration in the likelihood function of  $\theta_{\text{phys}}$  in (40), (41), and (43).

This seemingly-bizarre behavior of the posterior PDF in (48) is a natural consequence of the underlying idealistic assumption that we have made in this section, that is, there is no model inadequacy and the physical model could perfectly describe  $\mathcal{R}$  if we knew it.

## 5 Hierarchical Modeling of the Truth, in the Presence of Model Inadequacy and Data Uncertainty

So far, none of the idealized scenarios in §2, §3, and §4 represent what a practical researcher confronts in modeling natural phenomena. In reality, an experimenter collects a dataset,  $\mathcal{D}$ , each event of which is,

1. contaminated with various types of noise (measurement error) and is also,
2. insufficiently detailed,

leading to the development of wrong or incomplete physical models that are inadequate in providing a full description of the available dataset,  $\mathcal{D}$ . Hence, the resulting dataset appears to be heterogeneous with respect to the predictions of the physical model at hand.

### 5.1 Statement of the Hierarchical Bayesian Inverse Problem

Given,

1. the dataset  $\mathcal{D} = \{\mathbf{D}_i, \dots, \mathbf{D}_{n_{\text{do}}}\}$ ,
2. the corresponding set of noise models,  $\mathcal{M}_{\text{nois}}$ , and their ‘a priori known’ parameters,  $\Theta_{\text{nois}}$ , as given by (33) and (34), corresponding to each observation,  $\mathbf{D}_i \in \mathcal{D}$ ,
3. a physical model,  $\mathbf{M}_{\text{phys}}$ ,
4. a stochastic physics-based model,  $\mathbf{M}_{\text{inad}}$ , for the inadequacy of  $\mathbf{M}_{\text{phys}}$  in describing  $\mathcal{D}$ ,

we seek to quantify the posterior probability density function of the combined set of unknown parameters of the two physical and inadequacy models  $\boldsymbol{\theta}_{\text{pi}} = \{\boldsymbol{\theta}_{\text{phys}}, \boldsymbol{\theta}_{\text{inad}}\}$ .

### 5.2 General Solution

Our goal can be achieved by combining the approaches already developed in the previous sections §2, §3, and §4. First, note that the presence of model inadequacy requires us to use the modified form of (8) as given in (10). However, this equation takes as input, the truth,  $\mathbf{R}$ , about the set of observations,  $\mathcal{D}$ . Since  $\mathcal{R}$  is unknown, we have to consider all possibilities,  $\mathcal{R}^*$ , for  $\mathcal{R}$ , whose PDF is given by (35).

To do so, consider for the moment, a single realization,  $\mathcal{R}^*$ , of the truth dataset,  $\mathcal{R}$ , as defined in (32). The corresponding equations to (10), (12), and (13) for  $\mathcal{R}^*$  would be then,

$$\mathcal{U}^* = \mathcal{R}^* - \mathbf{M}_{\text{phys}}(\mathcal{R}^*, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}) , \quad (50)$$

$$\mathcal{U}^* = \{\mathbf{U}^*_1, \dots, \mathbf{U}^*_{n_{\text{do}}}\} , \quad (51)$$

$$\mathcal{U}^* \sim \mathbf{M}_{\text{inad}}(\mathcal{R}^*, \boldsymbol{\theta}_{\text{inad}}, \mathbf{M}_{\text{phys}}(\mathcal{R}^*, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}})) . \quad (52)$$

Therefore, the modified set of equations corresponding to (19), (20), and (21) take the form,

$$\mathcal{L}(\boldsymbol{\theta}_{\text{pi}} ; \mathcal{R}^*, \mathbf{M}_{\text{pi}}) \equiv \mathcal{L}(\boldsymbol{\theta}_{\text{pi}} ; \mathcal{U}^*, \mathbf{M}_{\text{pi}}) \quad (53)$$

$$\equiv \pi(\mathcal{R}^* | \boldsymbol{\theta}_{\text{pi}}, \mathbf{M}_{\text{pi}}) \quad (54)$$

$$\stackrel{\text{i.i.d.}}{=} \prod_{i=1}^{n_{\text{do}}} \pi(\mathbf{R}^*_i | \boldsymbol{\theta}_{\text{inad}}, \mathbf{M}_{\text{inad}}(\mathbf{R}^*_i, \boldsymbol{\theta}_{\text{inad}}, \mathbf{M}_{\text{phys}}(\mathbf{R}^*_i, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}))) . \quad (55)$$

Similar to (22) and (23), the posterior probability density of  $\boldsymbol{\theta}_{\text{pi}}$  for a single realization,  $\mathcal{R}^*$ , of  $\mathcal{R}$  can be then computed from the Bayes rule as,

$$\pi(\boldsymbol{\theta}_{\text{pi}} | \mathcal{R}^*, \mathbf{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}) = \frac{\pi(\mathcal{R}^* | \boldsymbol{\theta}_{\text{pi}}, \mathbf{M}_{\text{pi}}) \pi(\boldsymbol{\theta}_{\text{pi}} | \mathbf{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}})}{\pi(\mathcal{R}^* | \mathbf{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}})} , \quad (56)$$

However, the set  $\mathcal{R}^*$  is only one possibility among the (potentially infinitely) many possible representations of the truth set,  $\mathcal{R}$ . Therefore, the likelihood in (54) has to be further modified

to include not one, but all possibilities,  $\mathcal{R}^* \in \Omega_{\mathcal{R}^*}$ , for the reality,  $\mathcal{R}$ .

Given the observed dataset,  $\mathcal{D}$ , and the associated set of the noise models,  $\mathcal{M}_{\text{nois}}$ , and their parameters,  $\Theta_{\text{nois}}$ , the probability of  $\mathcal{R}^*$  being the truth,  $\mathcal{R}$ , is given by the posterior PDF in (35). Thus, combining (35) with (54) yields the modified likelihood function of the model parameters as,

$$\begin{aligned} \mathcal{L}(\theta_{\text{pi}} ; \mathcal{R}^*, \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}, \mathbf{M}_{\text{pi}}) \\ \equiv \pi(\mathcal{R}^* | \theta_{\text{pi}}, \mathbf{M}_{\text{pi}}) \times \pi(\mathcal{R}^* | \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) \end{aligned} \quad (57)$$

$$\begin{aligned} &\stackrel{\text{i.i.d.}}{=} \prod_{i=1}^{n_{\text{do}}} \pi(\mathbf{R}_i^* | \theta_{\text{inad}}, \mathbf{M}_{\text{inad}}(\mathbf{R}_i^*, \theta_{\text{inad}}, \mathbf{M}_{\text{phys}}(\mathbf{R}_i^*, \theta_{\text{phys}}, \mathcal{S}_{\text{phys}}))) \\ &\times \pi(\mathbf{R}_i^* | \mathbf{D}_i, \theta_{\text{nois},i}, \mathbf{M}_{\text{nois},i}, \mathcal{I}_{\mathbf{R}_i}) . \end{aligned} \quad (58)$$

One can then marginalize (57) over  $\mathcal{R}^*$  to obtain the marginal likelihood of the parameters of the models, given only the *known quantities*:  $\mathcal{D}$ ,  $\Theta_{\text{nois}}$ ,  $\mathcal{M}_{\text{nois}}$ ,  $\mathcal{I}_{\mathcal{R}}$ ,  $\mathbf{M}_{\text{phys}}$ ,  $\mathbf{M}_{\text{inad}}$ ,

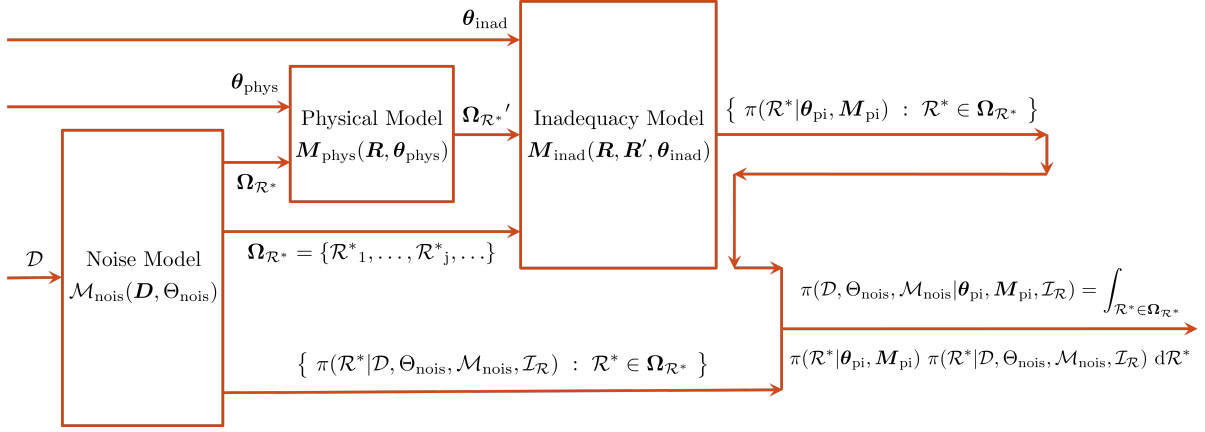
$$\begin{aligned} \mathcal{L}(\theta_{\text{pi}} ; \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}, \mathbf{M}_{\text{pi}}) \\ \equiv \pi(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} | \theta_{\text{pi}}, \mathbf{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}) \end{aligned} \quad (59)$$

$$\equiv \int_{\Omega_{\mathcal{R}^*}} \pi(\mathcal{R}^* | \theta_{\text{pi}}, \mathbf{M}_{\text{pi}}) \times \pi(\mathcal{R}^* | \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) d\mathcal{R}^* \quad (60)$$

$$\begin{aligned} &\stackrel{\text{i.i.d.}}{=} \prod_{i=1}^{n_{\text{do}}} \int_{\Omega_{\mathbf{R}}} \pi(\mathbf{R}_i^* | \theta_{\text{inad}}, \mathbf{M}_{\text{inad}}(\mathbf{R}_i^*, \theta_{\text{inad}}, \mathbf{M}_{\text{phys}}(\mathbf{R}_i^*, \theta_{\text{phys}}, \mathcal{S}_{\text{phys}}))) \\ &\times \pi(\mathbf{R}_i^* | \mathbf{D}_i, \theta_{\text{nois},i}, \mathbf{M}_{\text{nois},i}, \mathcal{I}_{\mathbf{R}_i}) d\mathbf{R}_i^* . \end{aligned} \quad (61)$$

A schematic diagram illustrating the derivation of this likelihood function is given in Figure 7. Despite their similarities, there is a fine difference between the marginalization over  $\mathcal{R}^*$  performed in the above likelihood function and the marginalization performed in the case of an ideal physical model in the presence of data uncertainty, which appears in (48). The marginalization of (57) and (58) is performed over all possible realizations of the truth,  $\Omega_{\mathcal{R}^*}$ , whether or not they satisfy (28), whereas in (40) it is assumed that the physical model has no inadequacy in describing the truth. Therefore, the marginalization in (40) is strictly performed on a subset of  $\Omega_{\mathcal{R}^*}$  that satisfy (28).

By contrast, not all possible realizations,  $\mathcal{R}^* \in \Omega_{\mathcal{R}^*}$ , of  $\mathcal{R}^*$  in the hierarchical likelihood function of (60) have to necessarily satisfy (28). For the construction of this likelihood function, we have already assumed that the physical model is inadequate in describing the truth,  $\mathcal{R}$ . Hence, the marginalization spans the entire sampling space of  $\mathcal{R}^*$ , which is  $\Omega_{\mathcal{R}^*}$ .



**Fig. 7:** A schematic illustration of the derivation of the likelihood function when the observational data,  $\mathcal{D}$ , is subject to measurement error described by the set of noise models,  $\mathcal{M}_{\text{nois}}$ , and their known parameters,  $\Theta_{\text{nois}}$ . In addition, the physical model,  $\mathbf{M}_{\text{phys}}$ , is inadequate, requiring an inadequacy model,  $\mathbf{M}_{\text{inad}}$ , with unknown parameters,  $\theta_{\text{inad}}$ , whose values have to be constrained by data. Given  $\{\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}\}$ , there exists a range of possibilities,  $\mathcal{R}^*$ , for the unknown truth,  $\mathcal{R}$ . The set  $\Omega_{\mathcal{R}^*}$  represents the collection of all such possibilities. Then, each one of these possibilities is fed to the physical model, whose output,  $\mathcal{R}^{*'} \in \Omega_{\mathcal{R}^*}'$ , is subsequently fed to the inadequacy model,  $\mathbf{M}_{\text{inad}}$ , which in turn outputs the probability of obtaining  $\mathcal{R}^*$  given  $\mathcal{R}^{*'}$  for the specific set of values of the parameters,  $\theta_{\text{pi}} = \{\theta_{\text{phys}}, \theta_{\text{inad}}\}$ , of the physical and inadequacy models,  $\mathbf{M}_{\text{pi}} = \{\mathbf{M}_{\text{phys}}, \mathbf{M}_{\text{inad}}\}$ . Simultaneously, the probability of  $\mathcal{R}^*$  being  $\mathcal{R}$  given only  $\{\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}\}$  and the prior knowledge about the truth,  $\mathcal{I}_{\mathcal{R}}$ , is also known. Therefore, the probability of  $\mathcal{R}^*$  being  $\mathcal{R}$  given all three physical, inadequacy, and noise models, as well as their parameter values, can be computed. Finally, the probability of  $\{\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}\}$  being the correct representation of the truth,  $\mathcal{R}$ , given  $\{\mathbf{M}_{\text{pi}}, \theta_{\text{pi}}, \mathcal{I}_{\mathcal{R}}\}$ , can be obtained by marginalizing the probability of  $\mathcal{R}^*$  over all possibilities,  $\mathcal{R}^* \in \Omega_{\mathcal{R}^*}$ , for the truth,  $\mathcal{R}$ .

Now, with the hierarchical likelihood function of  $\theta_{\text{pi}}$  in hand, one can use the Bayes rule to write the posterior PDF of  $\theta_{\text{pi}}$  as,

$$\begin{aligned} \pi(\theta_{\text{pi}} \mid \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathbf{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{pi}}}) \\ = \frac{\mathcal{L}(\theta_{\text{pi}} ; \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) \pi(\theta_{\text{pi}} \mid \mathbf{M}_{\text{pi}}, \mathcal{I}_{\theta_{\text{pi}}})}{\pi(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} \mid \mathbf{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{pi}}})} \end{aligned} \quad (62)$$

$$= \frac{\int_{\Omega_{\mathcal{R}^*}} \pi(\mathcal{R}^* \mid \theta_{\text{pi}}, \mathbf{M}_{\text{pi}}) \pi(\mathcal{R}^* \mid \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) d\mathcal{R}^* \pi(\theta_{\text{pi}} \mid \mathbf{M}_{\text{pi}}, \mathcal{I}_{\theta_{\text{pi}}})}{\pi(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} \mid \mathbf{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{pi}}})}, \quad (63)$$

in which the denominator is again a factor that properly normalizes the posterior distribution to a posterior PDF. It gives the probability of all possibilities,  $\Omega_{\mathcal{R}^*}$ , for the truth,  $\mathcal{R}$ , where  $\Omega_{\mathcal{R}^*}$  is fully determined by  $\{\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}\}$ ,

$$\pi(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} \mid \mathbf{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{pi}}}) \equiv \pi(\Omega_{\mathcal{R}^*} \mid \mathbf{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{pi}}}) \quad (64)$$

$$= \int_{\Theta_{\text{pi}}} \int_{\Omega_{\mathcal{R}}} \pi(\mathcal{R}^* \mid \theta_{\text{pi}}, \mathbf{M}_{\text{pi}}) \pi(\mathcal{R}^* \mid \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) \pi(\theta_{\text{pi}} \mid \mathbf{M}_{\text{pi}}, \mathcal{I}_{\theta_{\text{pi}}}) d\mathcal{R}^* d\theta_{\text{pi}}. \quad (65)$$

Plugging (35) into (63) one gets,

$$\begin{aligned} \pi(\theta_{\text{pi}} \mid \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathbf{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{pi}}}) \\ = \frac{\int_{\Omega_{\mathcal{R}}} \pi(\mathcal{D} \mid \mathcal{R}^*, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}) \pi(\mathcal{R}^* \mid \mathcal{I}_{\mathcal{R}}) \pi(\mathcal{R}^* \mid \theta_{\text{pi}}, \mathbf{M}_{\text{pi}}) d\mathcal{R}^* \pi(\theta_{\text{pi}} \mid \mathbf{M}_{\text{pi}}, \mathcal{I}_{\theta_{\text{pi}}})}{\pi(\mathcal{D} \mid \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}) \pi(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} \mid \mathbf{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{pi}}})}. \end{aligned} \quad (66)$$

Thus, in a sense, the two latter terms,  $\pi(\mathcal{R}^* \mid \mathcal{I}_{\mathcal{R}}) \pi(\mathcal{R}^* \mid \theta_{\text{pi}}, \mathbf{M}_{\text{pi}})$ , in the integrand of the numerator of (66), act like a prior probability on the likelihood of the observed dataset,

$\pi(\mathcal{D}|\mathcal{R}^*, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}})$ , and correct its value according to the physical and inadequacy models in hand, for the given set of parameter values,  $\theta_{\text{pi}}$ .

In the case of i.i.d. events, ignoring the normalization constants, the posterior PDF of (66) takes the simple form,

$$\begin{aligned} \pi(\theta_{\text{pi}} \mid \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathbf{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\theta_{\text{pi}}}) \\ \propto \pi(\theta_{\text{pi}} \mid \mathbf{M}_{\text{pi}}, \mathcal{I}_{\theta_{\text{pi}}}) \\ \times \prod_{i=1}^{n_{\text{do}}} \int_{\Omega_{\mathbf{R}}} \pi(\mathbf{D}_i \mid \mathbf{R}_i^*, \theta_{\text{nois},i}, \mathbf{M}_{\text{nois},i}) \pi(\mathbf{R}_i^* \mid \mathcal{I}_{\mathbf{R}_i}) \pi(\mathbf{R}_i^* \mid \theta_{\text{pi}}, \mathbf{M}_{\text{pi}}) d\mathbf{R}_i^* . \end{aligned} \quad (67)$$

Equations (66) and (67) describe the general form of the hierarchical (multilevel) posterior probability density function of the parameters of the physical model,  $\mathbf{M}_{\text{phys}}$ , whose inadequacy is described by the stochastic inadequacy model,  $\mathbf{M}_{\text{inad}}$ , in the presence measurement error in data, whose effects are assumed to be fully determined by a set of noise models,  $\mathcal{M}_{\text{nois}}$ , and their known parameters,  $\Theta_{\text{nois}}$ .  $\square$

## Acknowledgement

I thank J. Tinsley Oden, Ivo Babuska, and Fatemeh Bagheri for their helpful comments and valuable insights into many aspects of this manuscript.

## References

1. Ainsworth M, Oden JT (1992) A procedure for a posteriori error estimation for hp finite element methods. *Computer Methods in Applied Mechanics and Engineering* 101(1-3):73–96
2. Arendt PD, Apley DW, Chen W (2012) Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *Journal of Mechanical Design* 134(10):100,908
3. Babuska I, Oden JT (2004) Verification and validation in computational engineering and science: basic concepts. *Computer Methods in Applied Mechanics and Engineering* 193(36):4057–4066
4. Babuška I, Oden JT (2005) The reliability of computer predictions: can they be trusted. ICES report 5
5. Babuška I, Rheinboldt WC (1978) A-posteriori error estimates for the finite element method. *International Journal for Numerical Methods in Engineering* 12(10):1597–1615
6. Babuška I, Strouboulis T (2001) *The finite element method and its reliability*. Oxford university press
7. Babuska I, Nobile F, Oden J, Tempone R (2007) Reliability, uncertainty, estimates validation and verification. In: *Transcription of I. Babuskas presentation at the workshop on the Elements of Predictability*, J. Hopkins Univ, Baltimore, MD
8. Babuška I, Nobile F, Tempone R (2008) A systematic approach to model validation based on bayesian updates and prediction related rejection criteria. *Computer Methods in Applied Mechanics and Engineering* 197(29):2517–2539
9. Babuska I, Whiteman J, Strouboulis T (2010) *Finite elements: an introduction to the method and error estimation*. Oxford University Press
10. Babuška I, Sawlan Z, Scavino M, Szabó B, Tempone R (2016) Bayesian inference and model comparison for metallic fatigue data. *Computer Methods in Applied Mechanics and Engineering* 304:171–196
11. Beck JL (2010) Bayesian system identification based on probability logic. *Structural Control and Health Monitoring* 17(7):825–847

12. Beven K (2016) Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal* 61(9):1652–1665
13. Bohm D (1952) A suggested interpretation of the quantum theory in terms of "hidden" variables. i. *Physical Review* 85(2):166
14. Box GE (1976) Science and statistics. *Journal of the American Statistical Association* 71(356):791–799
15. Brenner S, Scott R (2007) *The mathematical theory of finite element methods*, vol 15. Springer Science & Business Media
16. Brynjarsdóttir J, OHagan A (2014) Learning about physical parameters: The importance of model discrepancy. *Inverse Problems* 30(11):114,007
17. Chen W, Kesidis G, Morrison T, Oden JT, Panchal JH, Paredis C, Pennock M, Atamturktur S, Terejanu G, Yukish M (2017) Uncertainty in modeling and simulation. In: *Research Challenges in Modeling and Simulation for Engineering Complex Systems*, Springer, pp 75–86
18. Clark JS, Gelfand AE (2006) *Hierarchical modelling for the environmental sciences: statistical methods and applications*. Oxford University Press on Demand
19. Council NR, et al (2012) *Assessing the reliability of complex models: mathematical and statistical foundations of verification, validation, and uncertainty quantification*. National Academies Press
20. Der Kiureghian A, Ditlevsen O (2009) Aleatory or epistemic? does it matter? *Structural Safety* 31(2):105–112
21. Dirac PAM (1981) *The principles of quantum mechanics*. 27, Oxford university press
22. Einstein A, Podolsky B, Rosen N (1935) Can quantum-mechanical description of physical reality be considered complete? *Physical review* 47(10):777
23. Faber MH (2005) On the treatment of uncertainties and probabilities in engineering decision analysis. *Journal of Offshore Mechanics and Arctic Engineering* 127(3):243–248
24. Farrell K, Oden JT, Faghihi D (2015) A bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems. *Journal of Computational Physics* 295:189–208
25. Ferson S, Ginzburg LR (1996) Different methods are needed to propagate ignorance and variability. *Reliability Engineering & System Safety* 54(2-3):133–144
26. Gregory P (2005) *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support*. Cambridge University Press
27. Hadamard J (1902) Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin* pp 49–52
28. Haukaas T, Gardoni P (2011) Model uncertainty in finite-element analysis: Bayesian finite elements. *Journal of Engineering Mechanics* 137(8):519–526
29. Heisenberg W (1985) Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. In: *Original Scientific Papers Wissenschaftliche Originalarbeiten*, Springer, pp 478–504
30. Higdon D, Kennedy M, Cavendish JC, Cafeo JA, Ryne RD (2004) Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing* 26(2):448–466
31. Higdon D, Gattiker J, Williams B, Rightley M (2008) Computer model calibration using high-dimensional output. *Journal of the American Statistical Association* 103(482):570–583
32. Hormuth DA, Weis JA, Barnes SL, Miga MI, Rericha EC, Quaranta V, Yankeelov TE (2017) A mechanically coupled reaction–diffusion model that incorporates intra-tumoural heterogeneity to predict in vivo glioma growth. *Journal of The Royal Society Interface* 14(128):20161,010
33. Jackson EL, Shahmoradi A, Spielman SJ, Jack BR, Wilke CO (2016) Intermediate divergence levels maximize the strength of structure–sequence correlations in enzymes and viral proteins. *Protein Science* 25(7):1341–1353
34. Jaynes E (1991) Straight line fitting a bayesian solution. Unpublished manuscript, item 22
35. Jaynes ET (1957) Information theory and statistical mechanics. *Physical review* 106(4):620
36. Jaynes ET (1957) Information theory and statistical mechanics. ii. *Physical review* 108(2):171
37. Jaynes ET (1973) The well-posed problem. *Foundations of Physics* 3(4):477–492

38. Jaynes ET (1989) Clearing up mysteriesthe original goal. In: Maximum Entropy and Bayesian Methods, Springer, pp 1–27
39. Jaynes ET (2003) Probability theory: The logic of science. Cambridge university press
40. Jeffrey R (1992) Probability and the Art of Judgment. Cambridge University Press
41. Jeffreys H (1998) The theory of probability. OUP Oxford
42. Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63(3):425–464
43. Keynes J (1911) The principal averages and the laws of error which lead to them. Journal of the Royal Statistical Society 74(3):322–331
44. Knupp P, Salari K (2002) Verification of computer codes in computational science and engineering. CRC Press
45. Laplace PS (2012) Pierre-Simon Laplace Philosophical Essay on Probabilities: Translated from the fifth French edition of 1825 With Notes by the Translator, vol 13. Springer Science & Business Media
46. Legendre AM (1805) Nouvelles méthodes pour la détermination des orbites des comètes. F. Didot
47. Lindley DV (2006) Understanding uncertainty. John Wiley & Sons
48. Ling Y, Mullins J, Mahadevan S (2014) Selection of model discrepancy priors in bayesian calibration. Journal of Computational Physics 276:665–680
49. Michel JB, Shen YK, Aiden AP, Veres A, Gray MK, Pickett JP, Hoiberg D, Clancy D, Norvig P, Orwant J, et al (2011) Quantitative analysis of culture using millions of digitized books. science 331(6014):176–182
50. Morrison RE, Oliver TA, Moser RD (2016) Representing model inadequacy: A stochastic operator approach. arXiv preprint arXiv:160401651
51. Oberkampf WL, Barone MF (2006) Measures of agreement between computation and experiment: validation metrics. Journal of Computational Physics 217(1):5–36
52. Oden J (1993) Error estimation and control in computational fluid dynamics. Brunel University, Institute of Computational Mathematics
53. Oden J, Babuska I, Faghihi D (2004) Predictive computational science: Computer predictions in the presence of uncertainty. Encyclopedia of Computational Mechanics, E Stein, R de Borst, and TJR Hughes, eds, Wiley, Hoboken, NJ
54. Oden JT (2017) Foundations of predictive computational sciences. ICES Reports
55. Oden JT, Prudhomme S (2002) Estimation of modeling error in computational mechanics. Journal of Computational Physics 182(2):496–515
56. Oden JT, Prudhomme S (2011) Control of modeling error in calibration and validation processes for predictive stochastic models. International Journal for Numerical Methods in Engineering 87(1-5):262–272
57. Oden JT, Reddy JN (2012) An introduction to the mathematical theory of finite elements. Courier Corporation
58. Oden JT, Vemaganti KS (2000) Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials: I. error estimates and adaptive algorithms. Journal of Computational Physics 164(1):22–47
59. Oden JT, Prudencio EE, Hawkins-Daarud A (2013) Selection and assessment of phenomenological models of tumor growth. Mathematical Models and Methods in Applied Sciences 23(07):1309–1338
60. Oden JT, Farrell K, Faghihi D (2015) Estimation of error in observables of coarse-grained models of atomic systems. Advanced Modeling and Simulation in Engineering Sciences 2(1):5
61. Oden T, Moser R, Ghattas O (2010) Computer predictions with quantified uncertainty, part i. SIAM News 43(9):1–3
62. Paté-Cornell ME (1996) Uncertainties in risk analysis: Six levels of treatment. Reliability Engineering & System Safety 54(2-3):95–111
63. Pearson K (1904) Mathematical contributions to the theory of evolution, vol 13. Dulau and co.

64. Prudhomme S, Oden JT (2003) Computable error estimators and adaptive techniques for fluid flow problems. In: Error estimation and adaptive discretization methods in computational fluid dynamics, Springer, pp 207–268
65. Qian PZ, Wu CJ (2008) Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics* 50(2):192–204
66. Roache PJ (1998) Verification and validation in computational science and engineering, vol 895. Hermosa Albuquerque, NM
67. Rosenthal JS (2006) Struck by lightning: the curious world of probabilities. National Academies Press
68. Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. *Statistical science* pp 409–423
69. Shafer G (2008) Non-additive probabilities in the work of bernoulli and lambert. In: Classic Works of the Dempster-Shafer Theory of Belief Functions, Springer, pp 117–182
70. Shahmoradi A (2013) Gamma-ray bursts: Energetics and prompt correlations. arXiv preprint arXiv:13081097
71. Shahmoradi A (2013) A multivariate fit luminosity function and world model for long gamma-ray bursts. *The Astrophysical Journal* 766(2):111
72. Shahmoradi A (2015) Dissecting the relationship between protein structure and sequence evolution. PhD thesis
73. Shahmoradi A, Nemiroff R (2009) How real detector thresholds create false standard candles. In: AIP Conference Proceedings, AIP, vol 1133, pp 425–427
74. Shahmoradi A, Nemiroff R (2011) A cosmological discriminator designed to avoid selection bias. In: Bulletin of the American Astronomical Society, vol 43
75. Shahmoradi A, Nemiroff R (2011) The possible impact of gamma-ray burst detector thresholds on cosmological standard candles. *Monthly Notices of the Royal Astronomical Society* 411(3):1843–1856
76. Shahmoradi A, Nemiroff R (2014) Classification and energetics of cosmological gamma-ray bursts. In: American Astronomical Society Meeting Abstracts# 223, vol 223
77. Shahmoradi A, Nemiroff RJ (2010) Hardness as a spectral peak estimator for gamma-ray bursts. *Monthly Notices of the Royal Astronomical Society* 407(4):2075–2090
78. Shahmoradi A, Nemiroff RJ (2015) Short versus long gamma-ray bursts: a comprehensive study of energetics and prompt gamma-ray correlations. *Monthly Notices of the Royal Astronomical Society* 451(1):126–143
79. Shahmoradi A, Wilke CO (2016) Dissecting the roles of local packing density and longer-range effects in protein sequence evolution. *Proteins: Structure, Function, and Bioinformatics* 84(6):841–854
80. Shahmoradi A, Sydykova DK, Spielman SJ, Jackson EL, Dawson ET, Meyer AG, Wilke CO (2014) Predicting evolutionary site variability from structure in viral proteins: buriedness, packing, flexibility, and design. *Journal of molecular evolution* 79(3-4):130–142
81. Yager RR, Liu L (2008) Classic works of the Dempster-Shafer theory of belief functions, vol 219. Springer
82. Zellner A (1971) An introduction to bayesian inference in econometrics. Tech. rep.