Noname manuscript No.
(will be inserted by the editor)

A Tutorial on Bayesian Parameter Estimation in the Presence of Model Inadequacy and Data Uncertainty

Amir Shahmoradi[1†]

October 28, 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 (inverse problem) and prediction (forward problem) involve multiple steps of data analysis, hypothesis formation, model construction, parameter estimation, model validation, and finally 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 uncertainty: aleatoric vs. epistemic, as well as their relashionships 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.

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 explains various potential causal relationships between different characteristics of data, and, finally, testing the predictions of the proposed model against observational data collected 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, 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*, also known as *model calibration*

[1] Center For Computational Oncology,
Institute for Computational Engineering and Sciences,
Department of Aerospace Engineering and Engineering Mechanics,
The University of Texas at Austin, TX 78712
E-mail: amir@physics.utexas.edu
[†] Peter O'Donnell, Jr. Fellow

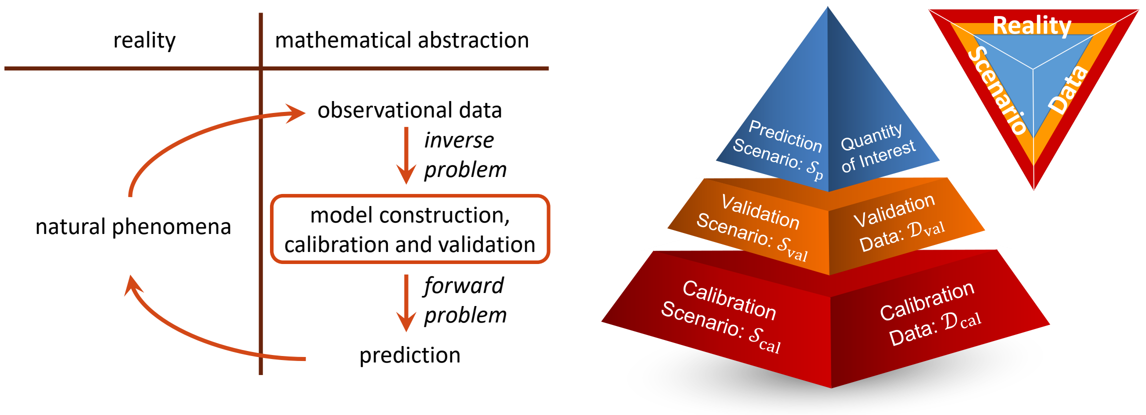
Table 1: Nomenclature and Definitions of Symbols in This Manuscript

$\boldsymbol{R}$	The reality or truth representing one event, without any observational bias or uncertainty.
$\mathcal{R}$	The set containing the truth $\boldsymbol{R}$ for each individual event that is observed.
$\boldsymbol{R}_{\mathrm{i}}$	The reality or truth, $\boldsymbol{R}$, representing the ith event in event-set $\mathcal{R}$.
$\mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}$	A subset of $\mathcal{R}$ on which the ith event, $\boldsymbol{R}_{\mathrm{i}}$, depends. It can be and often is a null set.
$\boldsymbol{D}$	All data about an event *as observed*, which is subject to epistemic uncertainty, unlike $\boldsymbol{R}$.
$\mathcal{D}$	The set of all observations, each of which corresponds to one unique event, $\boldsymbol{R}_{\mathrm{i}} \in \mathcal{R}$.
$\boldsymbol{D}_{\mathrm{i}}$	All observational data, $\boldsymbol{D}$, representing the ith observation in dataset $\mathcal{D}$.
$\mathcal{D}_{\mathrm{i}}$	A subset of $\mathcal{D}$ on which the ith observation, $\boldsymbol{D}_{\mathrm{i}}$, depends. It can be and often is a null set.
$\boldsymbol{U}$	A stochastic variable representing difference between $\boldsymbol{R}$ and the output of $\boldsymbol{M}_{\mathrm{phys}}$.
$\mathcal{U}$	The set of all $\boldsymbol{U}_{\mathrm{i}}$, each of which corresponds to one event $\boldsymbol{R}_{\mathrm{i}} \in \mathcal{R}$.
$\boldsymbol{R}^*$	One possible realization of $\boldsymbol{R}$, given $\boldsymbol{D}$ and the corresponding noise model, $\boldsymbol{M}_{\mathrm{nois}}$.
$\mathcal{R}^*$	One possible realization of $\mathcal{R}$, given $\mathcal{D}$ and the set of noise models, $\mathcal{M}_{\mathrm{nois}}$.
$\mathcal{R}^*_{\boldsymbol{R}_{\mathrm{i}}}$	One possible realization of $\mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}$.
$\boldsymbol{\mathcal{R}}^*$	The (super)set containing all possible realizations $\mathcal{R}^*$ of the set $\mathcal{R}$.
$\boldsymbol{M}_{\mathrm{phys}}$	The physical model hypothesized to hold for the collection of events in $\mathcal{R}$.
$\boldsymbol{M}_{\mathrm{inad}}$	The statistical/physical model hypothesized to describe the aleatoric uncertainty in $\mathcal{R}$.
$\boldsymbol{M}_{\mathrm{nois}}$	The statistical model hypothesized to describe the observed epistemic uncertainty in $\mathcal{D}$.
$\mathcal{M}_{\mathrm{nois}}$	The set of statistical models $\boldsymbol{M}_{\mathrm{nois,i}}$ corresponding to each of observation $\boldsymbol{D}_{\mathrm{i}} \in \mathcal{D}$.
$\mathbb{R}$	The set of real numbers.
$\boldsymbol{\theta}_{\mathrm{phys}}$	The vector of the parameters of the physical model $\boldsymbol{M}_{\mathrm{phys}}$.
$\boldsymbol{\theta}_{\mathrm{inad}}$	The vector of the parameters of the aleatoric uncertainty model $\boldsymbol{M}_{\mathrm{inad}}$.
$\boldsymbol{\theta}_{\mathrm{nois}}$	The vector of the parameters of the epistemic uncertainty model $\boldsymbol{M}_{\mathrm{nois}}$.
$\boldsymbol{\Theta}$	The parameter space of a model, $\boldsymbol{M}_{\mathrm{phys}} : \boldsymbol{\Theta}_{\mathrm{phys}}$, $\boldsymbol{M}_{\mathrm{inad}} : \boldsymbol{\Theta}_{\mathrm{inad}}$, $\boldsymbol{M}_{\mathrm{nois}} : \boldsymbol{\Theta}_{\mathrm{nois}}$.
$\boldsymbol{\Omega}_{\mathcal{R}}$	The observational sampling space to which each event in dataset belongs: $\boldsymbol{R} \in \mathcal{R} \subset \boldsymbol{\Omega}_{\mathcal{R}}$.
n_{do}	Number of data observations (events) in $\mathcal{R}$ (or equivalently, in $\mathcal{D}$).
n_{dv}	Number of data variables (attributes) by which each event is characterized; length of $\boldsymbol{R}$ & $\boldsymbol{D}$.
n_{pp}	Number of parameters of the physical model; length of $\boldsymbol{\theta}_{\mathrm{phys}}$; dimension of $\boldsymbol{\Theta}_{\mathrm{phys}}$.
n_{pi}	Number of parameters of the aleatoric model; length of $\boldsymbol{\theta}_{\mathrm{inad}}$; dimension of $\boldsymbol{\Theta}_{\mathrm{inad}}$.
n_{pn}	Number of parameters of the epistemic model; length of $\boldsymbol{\theta}_{\mathrm{nois}}$; dimension of $\boldsymbol{\Theta}_{\mathrm{nois}}$.
$\pi(\cdot)$	The Probability Density Function (PDF) of a statistical model.
$\mathcal{L}(\cdot)$	The likelihood function of the parameters of a statistical model.
$\mathcal{I}_{(\cdot)}$	The prior knowledge about the subscript entity, $(\cdot)$.

(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 [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 scientific literature.

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).



(a) The scientific methodology. **(b)** The prediction pyramid.

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 observed phenomena, which is never known to the observer. The two front faces represent the observed data ($\mathcal{D}$) which results from the convolution of the truth/reality $\boldsymbol{R}$ with various sources of uncertainty, and the scenarios ($\mathcal{S}$) under which data is collected and physical model is constructed and constrained [53, 54, 59, 61].

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, 67, 81], or in the discretization and numerical computations of the inverse and forward problems [1, 5, 6, 9, 15, 42, 52, 55–58, 64, 67].

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 uncertainties, also named *structural variability* or *risk* [20, 62] in engineering literature, are presumed to stem from inherent *unpredictable* variabilities and randomness in observational data, and are therefore thought as 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). 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 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 sought 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, 80]. Thus, from a historical perspective, aleatoric uncertainty is solely defined within the framework of frequentist statistics [17].

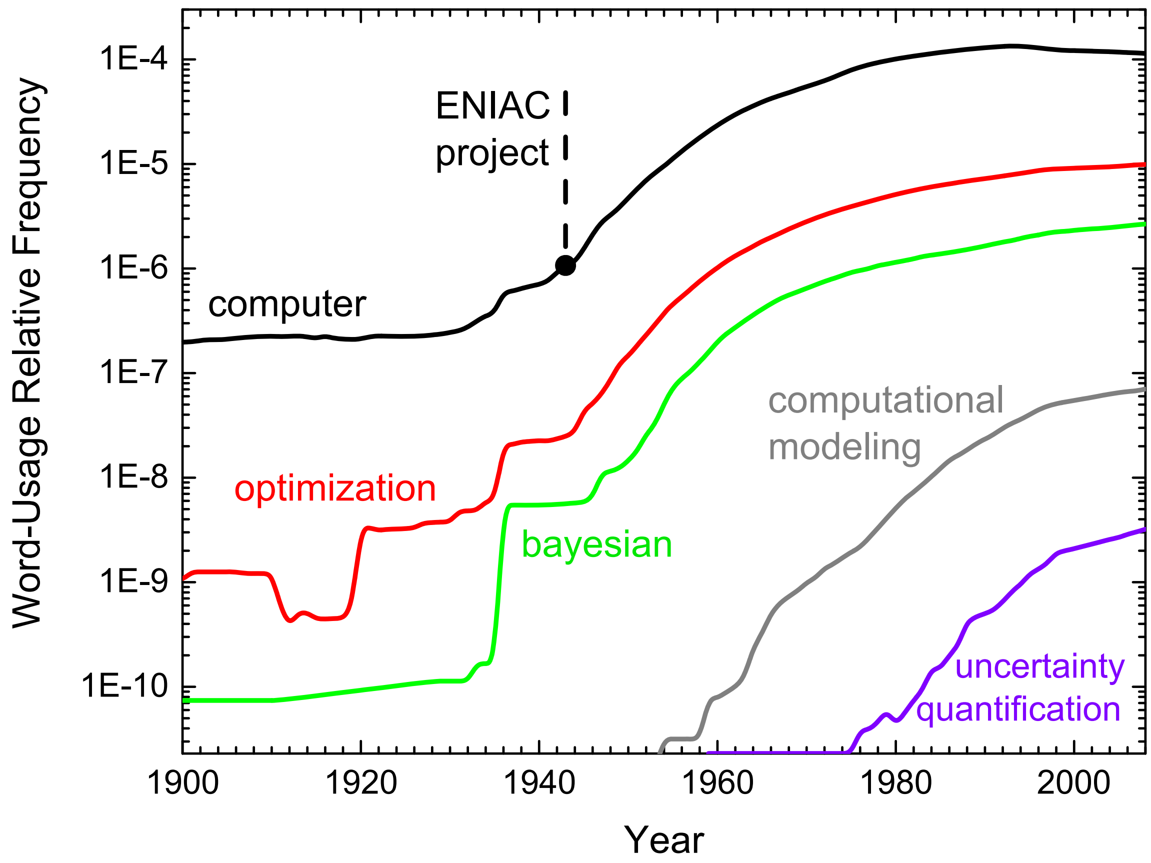


Fig. 2: A word-usage relative frequency plot [49], illustrating the exponential growth of computer technology in the mid 20th century, as well as developments in the fields of deterministic and stochastic optimization techniques, which ultimately led to the emergence of 'computational modeling' as the 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.

We remark that in a pure Bayesian system of logical probability [39], which is the viewpoint 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 with 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.

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 were 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 input data to the experiments or the experimental process itself are almost always incomplete. This lack of the complete detailed knowledge of the problem under study, often leads to the development of mathematical models that are inadequate in correctly describing the underlying physics of the problem.

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 [68] 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 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 scientific inference. This is in agreement with 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

Consider a set of $n_{\rm do}$ physical events $\mathcal{R} = \{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_{n_{\rm do}}\}$, where each event, $\boldsymbol{R}_{\rm i} \in \mathcal{R}$, is characterized by $n_{\rm dv}$ variables (i.e., observable quantities), describing different characteristics of the events. Thus, $\boldsymbol{R}_{\rm i}$ is a vector of $n_{\rm dv}$ elements representing a single event in the $n_{\rm dv}$-dimensional observational sampling space $\boldsymbol{\Omega}_{\mathcal{R}} \subset \mathbb{R}^{n_{\rm dv}}$. For the moment, suppose we live in an ideal world where natural phenomena are observed and perceived exactly and accurately, without any possible bias, inadequacy, or contamination with noise. Therefore, the observed physical events represent the *reality* as they truly occur in nature (and henceforth represented here by the symbol $\boldsymbol{R}$).

Although frequently independent of each other, the characteristics of each event, $\boldsymbol{R}_{\rm i}$, or its occurrence could also depend on any subset, $\mathcal{R}_{\boldsymbol{R}_{\rm i}} \subset \mathcal{R}$, or all of the other events. In general, there may also exist interdependencies between the characteristics of each event (i.e., the elements of each vector $\boldsymbol{R}_{\rm i}$). 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 sometimes be divided into a set of independent variables $\boldsymbol{R}_{\rm ind}$ on which the rest of the event's attributes (i.e., the dependent/response variables, $\boldsymbol{R}_{\rm dep}$), depend. In other words,

$$\boldsymbol{R}_{\rm i} = \left\{\boldsymbol{R}_{\rm i,ind}, \boldsymbol{R}_{\rm i,dep}\big(\boldsymbol{R}_{\rm i,ind}\big)\right\} \ . \tag{1}$$

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, 71, 78, 79], or modeling the energetics and occurrence rates of astrophysical phenomena as a function of their distance from the earth [69, 70, 72, 74, 76, 77].

Now, suppose we formulate a hypothesis regarding the set of events $\mathcal{R}$. This hypothesis can be cast in the form of a mathematical model $\boldsymbol{M}_{\mathrm{phys}}$ with the subscript *phys* emphasizing the type of the model (i.e., a physics-based model). This physical model $\boldsymbol{M}_{\mathrm{phys}}$ can be thought of as a collection of mathematical operators (e.g., algebraic, differential, integral, ...) that takes as input, a set of n_{pp} physical parameters represented by the vector $\boldsymbol{\theta}_{\mathrm{phys}} \in \boldsymbol{\Theta}_{\mathrm{phys}} \subset \mathbb{R}^{n_{\mathrm{pp}}}$. It then acts on some or all characteristics of a single event, $\boldsymbol{R}_{\mathrm{i}} \in \mathcal{R}$, to generates an output response corresponding to some other attributes of the event $\boldsymbol{R}_{\mathrm{i}}$.

Generally, each observational event, $\boldsymbol{R}_{\mathrm{i}}$, might be collected under a specific set of conditions collectively known as the physical scenario, $\mathcal{S}_{\mathrm{phys}}$. The scenario, $\mathcal{S}_{\mathrm{phys}}$, 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 $\boldsymbol{M}_{\mathrm{phys}}$, and its parameters $\boldsymbol{\theta}_{\mathrm{phys}}$. These features typically describe and depend on the experimental/observational setup for data collection [53]. For example, $\mathcal{S}_{\mathrm{phys}}$ could describe the limitations of a detector that was used for data collection [69, 70, 72–77, 77]. Thus, although $\mathcal{S}_{\mathrm{phys}}$ is independent of $\mathcal{R}$, the reverse is not true in general.

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 $\boldsymbol{\theta}_{\mathrm{phys}} = \widehat{\boldsymbol{\theta}}_{\mathrm{phys}} \in \boldsymbol{\Theta}_{\mathrm{phys}}$ for which,

$$\boldsymbol{R}_{\mathrm{i}} - \boldsymbol{M}_{\mathrm{phys}}\big(\boldsymbol{R}_{\mathrm{i}}, \mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}, \boldsymbol{\theta}_{\mathrm{phys}}, \mathcal{S}_{\mathrm{phys}}\big) = \mathbf{0} \quad \forall\ \boldsymbol{R}_{\mathrm{i}} \in \mathcal{R}\ , \tag{2}$$

where $\mathbf{0}$ is the null vector of length n_{dv}. Here, $\mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}$, defined in Table 1, explicitly appears as an input to the physical model, $\boldsymbol{M}_{\mathrm{phys}}$, in order to highlight the potential dependence of the event, $\boldsymbol{R}_{\mathrm{i}}$, on a subset, $\mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}$, of other events in the event set, $\mathcal{R}$. Thus, given a perfect model and an ideal dataset with no measurement error, the problem of inference is reduced to solving the system of n_{do} equations of the form (2) to obtain the feasible values, $\widehat{\boldsymbol{\theta}}_{\mathrm{phys}}$, of the set of parameters of the physical model satisfying (2).

In mathematical modeling, this process is widely known as *inverse problem* or *model calibration*, as illustrated in Figure 1a. When the physical model perfectly describes the truth, it is labeled as a *correctly-specified model.* In some problems, there might exist multiple or even uncountably infinite number of $\widehat{\boldsymbol{\theta}}_{\mathrm{phys}}$ that satisfy (2). 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 under similar conditions, yields six inherently different possible outcomes. A more relevant example to engineering is a set of stress-strain data obtained from a heterogenous 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 intrinsic variabilities and heterogeneities in observational data, where the physical model is incapable of taking them into account, 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 of events already in $\mathcal{D}$.

Regardless of the origins of model inadequacy – whether it be wrong physical model, or insufficiently-detailed data – here we assume that such discrepancies between model predictions and data can and often do exist. However the effects of model inadequacy has to be treated by stochastic models, since the deterministic physical model is incapable of taking them into account due to any of the two aforementioned reasons.

When there is intrinsic heterogeneity in the observed events with respect to the physical model at hand, the model is said to be *inadequate* or *misspecified*, and the equality in (2) does not hold anymore, rather we have in general,

$$\boldsymbol{R}_{\mathrm{i}} - \boldsymbol{M}_{\mathrm{phys}}\big(\boldsymbol{R}_{\mathrm{i}}, \mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}, \boldsymbol{\theta}_{\mathrm{phys}}, \mathcal{S}_{\mathrm{phys}}\big) = \boldsymbol{U}_{\mathrm{i}} \ , \tag{3}$$

where $\boldsymbol{U}_{\mathrm{i}}$ is a vector of length n_{dv}, some elements of which, corresponding to the response variables $\boldsymbol{R}_{\mathrm{i,dep}} \subset \boldsymbol{R}_{\mathrm{i}}$, are not anymore deterministic but random stochastic variables. In other words, for the same control variables $\boldsymbol{R}_{\mathrm{ind}} \subset \boldsymbol{R} = \{\boldsymbol{R}_{\mathrm{ind}}, \boldsymbol{R}_{\mathrm{dep}}\}$, there can be a finite or infinite number of possible values for the response variables $\boldsymbol{R}_{\mathrm{dep}}$. But the physical model is only capable of fitting for some representative deterministic average output response, given the input. Therefore, the set,

$$\mathcal{U} = \{\boldsymbol{U}_1, \ldots, \boldsymbol{U}_{\mathrm{n_{do}}}\} \ , \tag{4}$$

is a collection of deterministic and stochastic variables whose behavior has to be determined by yet another model, $\boldsymbol{M}_{\mathrm{inad}}$, of statistical and physical origin (as opposed to deterministic physical origin of $\boldsymbol{M}_{\mathrm{phys}}$), that in general depends explicitly on the proposed physical model, $\boldsymbol{M}_{\mathrm{phys}}$ and less commonly, may also depend on a subset of $\boldsymbol{R}_{\mathrm{i}}$ or a subset, $\mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}$, of the entire dataset, on which the occurrence of $\boldsymbol{R}_{\mathrm{i}}$ depends,

$$\boldsymbol{U}_{\mathrm{i}} \sim \boldsymbol{M}_{\mathrm{inad}}\big(\boldsymbol{R}_{\mathrm{i}}, \mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}, \boldsymbol{\theta}_{\mathrm{inad}}, \boldsymbol{M}_{\mathrm{phys}}(\boldsymbol{R}_{\mathrm{i}}, \mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}, \boldsymbol{\theta}_{\mathrm{phys}}, \mathcal{S}_{\mathrm{phys}})\big) \ . \tag{5}$$

where $\boldsymbol{\theta}_{\mathrm{inad}}$ represents the set of n_{pi} parameters of $\boldsymbol{M}_{\mathrm{inad}}$. Here, the subscript *inad* stands for *inadequacy*. The explicit presence of $\mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}$ as input to $\boldsymbol{M}_{\mathrm{inad}}$ is again to emphasize the potential dependence of $\boldsymbol{R}_{\mathrm{i}}$ on other events in $\mathcal{R}$.

The special case where $\boldsymbol{M}_{\mathrm{inad}}$ depends explicitly on $\boldsymbol{R}_{\mathrm{i,ind}} \subset \boldsymbol{R}_{\mathrm{i}}$, is known as *heteroscedasticity* in statistical literature [63]. The occurrence frequency of $\boldsymbol{U}_{\mathrm{i}}$ is then proportional to the multivariate probability density function (PDF), $\pi(\cdot)$, of the inadequacy model $\boldsymbol{M}_{\mathrm{inad}}$,

$$\mathrm{PDF}\big(\boldsymbol{U}_{\mathrm{i}}\big) \stackrel{\mathrm{def}}{=} \pi\big(\boldsymbol{U}_{\mathrm{i}}\big|\boldsymbol{\theta}_{\mathrm{inad}}, \boldsymbol{M}_{\mathrm{inad}}\big(\boldsymbol{R}_{\mathrm{i}}, \mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}, \boldsymbol{\theta}_{\mathrm{inad}}, \boldsymbol{M}_{\mathrm{phys}}(\boldsymbol{R}_{\mathrm{i}}, \mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}, \boldsymbol{\theta}_{\mathrm{phys}}, \mathcal{S}_{\mathrm{phys}})\big) \tag{6}$$

$$\stackrel{\mathrm{def}}{=} \pi\big(\boldsymbol{U}_{\mathrm{i}}\big|\boldsymbol{\theta}_{\mathrm{pi}}, \boldsymbol{M}_{\mathrm{pi}}\big) \ , \tag{7}$$

where,

$$\boldsymbol{M}_{\mathrm{pi}} = \{\boldsymbol{M}_{\mathrm{phys}}, \boldsymbol{M}_{\mathrm{inad}}\} \ , \tag{8}$$

$$\boldsymbol{\theta}_{\mathrm{pi}} = \{\boldsymbol{\theta}_{\mathrm{phys}}, \boldsymbol{\theta}_{\mathrm{inad}}\} \ , \tag{9}$$

represent the combined physical and inadequacy models for the observed dataset, and the combined set of the parameters of the two models, respectively.

3.1 The Likelihood Function for Inadequacy Model

In the presence of model inadequacy, the inference problem cannot be solved deterministically anymore as in §2, rather one has to first derive the likelihood (i.e., the unnormalized probability) function, $\mathcal{L}(\boldsymbol{\theta}_{\mathrm{pi}})$, of $\boldsymbol{\theta}_{\mathrm{pi}}$ being the correct set of parameters of $\boldsymbol{M}_{\mathrm{pi}}$, in the light of available data, $\mathcal{R}$,

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}_{\mathrm{pi}};\mathcal{U}) \equiv & \ \pi(\mathcal{U}|\boldsymbol{\theta}_{\mathrm{pi}}, \boldsymbol{M}_{\mathrm{pi}}) & (10) \\ \stackrel{\text{i.i.d.}}{=} & \prod_{\mathrm{i}=1}^{n_{\mathrm{do}}} \pi\big(\boldsymbol{U}_{\mathrm{i}}|\boldsymbol{\theta}_{\mathrm{inad}}, \boldsymbol{M}_{\mathrm{inad}}\big(\boldsymbol{R}_{\mathrm{i}}, \boldsymbol{\theta}_{\mathrm{inad}}, \boldsymbol{M}_{\mathrm{phys}}(\boldsymbol{R}_{\mathrm{i}}, \boldsymbol{\theta}_{\mathrm{phys}}, \mathcal{S}_{\mathrm{phys}})\big)\big) \ . & (11) \end{aligned}$$

This likelihood function is equivalent to the joint PDF of observing the entire set $\mathcal{U}$ (or equivalently, $\mathcal{R}$), for the given value of $\boldsymbol{\theta}_{\mathrm{pi}}$. This is the PDF term that appears on the right-hand-side of (10).

The i.i.d. equality in (11) 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 inference problem is now to find $\boldsymbol{\theta}_{\mathrm{pi}} = \{\boldsymbol{\theta}_{\mathrm{phys}}, \boldsymbol{\theta}_{\mathrm{inad}}\}$ such that the joint probability of observing all events in $\mathcal{R}$ together, combined with any prior knowledge about the parameters of the model, $\boldsymbol{M}_{\mathrm{pi}} = \{\boldsymbol{M}_{\mathrm{phys}}, \boldsymbol{M}_{\mathrm{inad}}\}$, is maximized. From the Bayes rule it follows,

$$\pi(\boldsymbol{\theta}_{\mathrm{pi}}|\mathcal{U}, \boldsymbol{M}_{\mathrm{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\mathrm{pi}}}) = \frac{\pi(\mathcal{U}|\boldsymbol{\theta}_{\mathrm{pi}}, \boldsymbol{M}_{\mathrm{pi}})\pi(\boldsymbol{\theta}_{\mathrm{pi}}|\boldsymbol{M}_{\mathrm{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\mathrm{pi}}})}{\pi(\mathcal{U}|\boldsymbol{M}_{\mathrm{pi}})} \ , \tag{12}$$

or equivalently,

$$\pi(\boldsymbol{\theta}_{\mathrm{pi}}|\mathcal{R}, \boldsymbol{M}_{\mathrm{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\mathrm{pi}}}) = \frac{\pi(\mathcal{R}|\boldsymbol{\theta}_{\mathrm{pi}}, \boldsymbol{M}_{\mathrm{pi}})\pi(\boldsymbol{\theta}_{\mathrm{pi}}|\boldsymbol{M}_{\mathrm{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\mathrm{pi}}})}{\pi(\mathcal{R}|\boldsymbol{M}_{\mathrm{pi}})} \ , \tag{13}$$

where $\mathcal{I}_{\boldsymbol{\theta}_{\mathrm{pi}}}$ represents any prior knowledge about all unknown parameters, $\boldsymbol{\theta}_{\mathrm{pi}}$, of the physical and inadequacy models together, $\boldsymbol{M}_{\mathrm{pi}}$. The subject of prior PDF construction, $\pi(\boldsymbol{\theta}_{\mathrm{pi}}|\boldsymbol{M}_{\mathrm{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\mathrm{pi}}})$, 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.

3.2 Popular Choices of Inadequacy Model

Multivariate Normal (MVN) distribution is undoubtedly the most popular and widely-used inadequacy model in scientific inference, although it 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, $\boldsymbol{M}_{\rm inad}$, for every scientific inference problem. Since $\boldsymbol{M}_{\rm inad}$ represents the inadequacy of the deterministic physics-based 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. Therefore, measurement uncertainty is 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 $\boldsymbol{R}$ 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 is 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 experimental measurements.

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

$$\boldsymbol{D}_{\rm i} \sim \boldsymbol{M}_{\rm nois,i}(\boldsymbol{R}_{\rm i}, \boldsymbol{\theta}_{\rm nois,i}) \ , \tag{14}$$

represent the ith observational data in the dataset, $\mathcal{D} = \{\boldsymbol{D}_1, \ldots, \boldsymbol{D}_{\rm n_{do}}\}$, consisting of $n_{\rm do}$ observations. Here, $\boldsymbol{M}_{\rm nois,i}$ represents the stochastic noise model for the ith event, $\boldsymbol{R}_{\rm i}$. Each observation, $\boldsymbol{D}_{\rm i} \in \mathcal{D}$, results from a complex convolution of the truth, $\boldsymbol{R}_{\rm i} \in \mathcal{R}$, with various types of noise in the measurement process, all of which are summarized in $\boldsymbol{M}_{\rm nois,i}$.

The noise model, $\boldsymbol{M}_{\text{nois,i}}$, 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 (14) and throughout this section, is used to indicate that the corresponding object specifically belongs to or is defined in relation with the ith event, $\boldsymbol{R}_{\text{i}}$. The subscript *pn* in $n_{\text{pn,i}}$ stands for *p*arameters of the *n*oise model.

A typical example of such inference problem, involving distinct noise models corresponding to each observation in dataset, can be found in the field of Astronomy and Cosmology, where observational data is frequently gathered by multiple instruments of different measurement accuracies [70, 73, 76, 77].

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

$$\text{PDF}\big(\boldsymbol{D}_{\text{i}}\big) \stackrel{\text{def}}{=} \pi\big(\boldsymbol{D}_{\text{i}}\big|\boldsymbol{R}_{\text{i}}, \boldsymbol{\theta}_{\text{nois,i}}, \boldsymbol{M}_{\text{nois,i}}\big) \tag{15}$$

There are multiple reasons that render (15) useless by itself, even when the exact mathematical form of $\boldsymbol{M}_{\text{nois,i}}$ is known:

- Firstly, this formulation requires us to know the truth $\boldsymbol{R}_{\text{i}}$ as input to $\boldsymbol{M}_{\text{nois,i}}$. In practice, we never know the truth.
- Secondly, the observer/experimenter can never obtain multiple realizations of $\boldsymbol{D}_{\text{i}}$ for the exact same truth $\boldsymbol{R}_{\text{i}}$. What an experimenter perceives about an event $\boldsymbol{R}_{\text{i}}$, is a single observation of it, $\boldsymbol{D}_{\text{i}}$, which is the result of the convolution of $\boldsymbol{R}_{\text{i}}$ with noise. From a Bayesian perspective, such experiments can never be repeated under the exact same conditions to obtain multiple observations $\boldsymbol{D}_{\text{i}}$ for the same $\boldsymbol{R}_{\text{i}}$.
- Thirdly, the input quantity to the physical model $\boldsymbol{M}_{\text{phys}}$ in (2) is $\boldsymbol{R}_{\text{i}}$, not $\boldsymbol{D}_{\text{i}}$.

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

This question has a straightforward answer using the Bayesian inversion method applied to (15), such that the probability of $\boldsymbol{R}^*{}_{\text{i}}$ being the truth about the ith event could be written as,

$$\begin{aligned}\pi\big(\boldsymbol{R}^*{}_{\text{i}} = \boldsymbol{R}_{\text{i}} \;\big|\; \boldsymbol{D}_{\text{i}}, \boldsymbol{\theta}_{\text{nois,i}}, \boldsymbol{M}_{\text{nois,i}}, \mathcal{I}_{\boldsymbol{R}_{\text{i}}}\big) &= \\ &\frac{\pi\big(\boldsymbol{D}_{\text{i}}\big|\boldsymbol{R}^*{}_{\text{i}} = \boldsymbol{R}_{\text{i}}, \boldsymbol{\theta}_{\text{nois,i}}, \boldsymbol{M}_{\text{nois,i}}\big)\ \pi\big(\boldsymbol{R}^*{}_{\text{i}} = \boldsymbol{R}_{\text{i}}\big|\mathcal{I}_{\boldsymbol{R}_{\text{i}}}\big)}{\pi\big(\boldsymbol{D}_{\text{i}}\big|\boldsymbol{\theta}_{\text{nois,i}}, \boldsymbol{M}_{\text{nois,i}}\big)}\ ,\end{aligned} \tag{16}$$

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 $\boldsymbol{R}^*{}_{\text{i}}$ which is equivalent to (15), and $\mathcal{I}_{\boldsymbol{R}_{\text{i}}}$ represents the experimenter's prior knowledge about $\boldsymbol{R}_{\text{i}}$. 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 $\boldsymbol{D}_{\text{i}}$ averaged over all possible values $\boldsymbol{R}^*{}_{\text{i}}$ for the truth $\boldsymbol{R}_{\text{i}}$.

In practice, the state of experimenter's prior knowledge, $\mathcal{I}_{\boldsymbol{R}_{\text{i}}}$, 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, $\pi\big(\boldsymbol{R}^*{}_{\text{i}} = \boldsymbol{R}_{\text{i}}\big|\mathcal{I}_{\boldsymbol{R}_{\text{i}}}\big)$ 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 [70, 77].

4.1 Noise Models Are Fundamentally Different From Inadequacy Models

Unlike the case of model inadequacy in §3 where the parameters, $\boldsymbol{\theta}_{\mathrm{inad}}$, of the inadequacy model, $\boldsymbol{M}_{\mathrm{inad}}$, were a priori unknown and were to be inferred from data, here it is essential to know the mathematical form of $\boldsymbol{M}_{\mathrm{nois,i}}$ (i.e., the PDF in (15)) for each observation $\boldsymbol{D}_{\mathrm{i}}$, as well as the values of the parameters, $\boldsymbol{\theta}_{\mathrm{nois,i}}$, of $\boldsymbol{M}_{\mathrm{nois,i}}$. 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 (16), and subsequently solve (2).

Another difference with the case of inadequacy model is that, the presence of noise in data does not necessarily invalidate (2). As long as the physical model $\boldsymbol{M}_{\mathrm{phys}}$ is perfectly capable of describing $\mathcal{R}$ and the sampling space of $\boldsymbol{M}_{\mathrm{nois,i}}$ (i.e., domain of $\boldsymbol{R}^{*}{}_{\mathrm{i}}$) is accurately defined such that,

$$\pi\big(\boldsymbol{R}_{\mathrm{i}}\big|\boldsymbol{D}_{\mathrm{i}},\boldsymbol{\theta}_{\mathrm{nois,i}},\boldsymbol{M}_{\mathrm{nois,i}}\big) \neq 0 \quad \forall\ \boldsymbol{R}_{\mathrm{i}} \in \mathcal{R}\ , \tag{17}$$

then, there will be at least one set of physical parameters $\boldsymbol{\theta}_{\mathrm{phys}}$ satisfying (2),

$$\boldsymbol{R}^{*}{}_{\mathrm{i}} - \boldsymbol{M}_{\mathrm{phys}}\big(\boldsymbol{R}^{*}{}_{\mathrm{i}},\mathcal{R}^{*}_{\boldsymbol{R}_{\mathrm{i}}},\boldsymbol{\theta}_{\mathrm{phys}},\mathcal{S}_{\mathrm{phys}}\big) = \boldsymbol{0} \quad \forall\ \boldsymbol{R}^{*}{}_{\mathrm{i}} \in \mathcal{R}^{*}\ , \tag{18}$$

where $\mathcal{R}^{*}_{\boldsymbol{R}_{\mathrm{i}}}$ is a possible realization of $\mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}$ in (2) as defined in Table 1. Similar to (2) and (5), its explicit appearance here is again to emphasize that $\boldsymbol{R}_{\mathrm{i}}$ might in general depend on a subset, $\mathcal{R}_{\boldsymbol{R}_{\mathrm{i}}}$, of other events in $\mathcal{R}$.

The symbol $\mathcal{R}^{*}$ in (18) represents a set of possible realizations, $\boldsymbol{R}^{*}{}_{\mathrm{i}}$, of each of the events, $\boldsymbol{R}_{\mathrm{i}} \in \mathcal{R}$, that together satisfy (18) for a specific set of parameters values, $\boldsymbol{\theta}_{\mathrm{phys}}$,

$$\mathcal{R}^{*} = \big\{\boldsymbol{R}^{*}{}_{1},\cdots,\boldsymbol{R}^{*}{}_{\mathrm{n_{do}}}\big\}\ . \tag{19}$$

Therefore $\mathcal{R}^{*}$ represents a possible realization of $\mathcal{R}$. Frequently however, there can be uncountably infinite number of $\boldsymbol{\theta}_{\mathrm{phys}}$ that satisfy (2) depending on the given set of noise models and their 'a priori known' parameters,

$$\mathcal{M}_{\mathrm{nois}} = \{\boldsymbol{M}_{\mathrm{nois},1},\ldots,\boldsymbol{M}_{\mathrm{nois,n_{do}}}\}\ , \tag{20}$$

$$\Theta_{\mathrm{nois}} = \{\boldsymbol{\theta}_{\mathrm{nois},1},\ldots,\boldsymbol{\theta}_{\mathrm{nois,n_{do}}}\}\ . \tag{21}$$

This results from the fact that the truth set, $\mathcal{R}$, is not known (and will never be known). Therefore, we have a range of possibilities, $\mathcal{R}^{*}$, whose probabilities of its individual members, $\boldsymbol{R}^{*}{}_{\mathrm{i}}$, being the truth, $\boldsymbol{R}_{\mathrm{i}}$, can be computed via (16). Thus, the set of $\boldsymbol{\theta}_{\mathrm{phys}}$ satisfying (18) often forms a manifold in the n_{pp}-dimensional space of the physical parameters, $\boldsymbol{\theta}_{\mathrm{phys}}$. Each feasible $\boldsymbol{\theta}_{\mathrm{phys}}$ residing on this manifold is associated with at least one possible realization, $\mathcal{R}^{*}$, of the truth set, $\mathcal{R}$.

One can define the probability of $\mathcal{R}^{*}$ being the truth as the joint probability of all individual realizations, $\boldsymbol{R}^{*}{}_{\mathrm{i}}$, being the truth $\boldsymbol{R}_{\mathrm{i}}$. Since the effects of the effects of noise models on each

$\boldsymbol{R}_{\rm i} \in \mathcal{R}$ are independent of each other, the joint probability density function of $\mathcal{R}^*$ takes the simple form,

$$\begin{aligned}
\mathrm{PDF}\big(\mathcal{R}^* = \mathcal{R}\big) &\equiv \pi\big(\mathcal{R}^* = \mathcal{R} \big| \mathcal{D}, \mathit{\Theta}_{\rm nois}, \mathcal{M}_{\rm nois}, \mathcal{I}_{\mathcal{R}}\big) && (22)\\
&= \frac{\pi\big(\mathcal{D} \big| \mathcal{R}^* = \mathcal{R}, \mathit{\Theta}_{\rm nois}, \mathcal{M}_{\rm nois}\big)\ \pi\big(\mathcal{R}^* = \mathcal{R} \big| \mathcal{I}_{\mathcal{R}}\big)}{\pi\big(\mathcal{D} \big| \mathit{\Theta}_{\rm nois}, \mathcal{M}_{\rm nois}, \mathcal{I}_{\mathcal{R}}\big)} && (23)\\
&\overset{\rm i.i.d.}{=} \prod_{\rm i=1}^{n_{\rm do}} \pi\big(\boldsymbol{R}^*{}_{\rm i} = \boldsymbol{R}_{\rm i} \big| \boldsymbol{D}_{\rm i}, \boldsymbol{\theta}_{\rm nois,i}, \boldsymbol{M}_{\rm nois,i}, \mathcal{I}_{\boldsymbol{R}_{\rm i}}\big) && (24)\\
&\overset{\rm i.i.d.}{=} \prod_{\rm i=1}^{n_{\rm do}} \frac{\pi\big(\boldsymbol{D}_{\rm i} \big| \boldsymbol{R}^*{}_{\rm i} = \boldsymbol{R}_{\rm i}, \boldsymbol{\theta}_{\rm nois,i}, \boldsymbol{M}_{\rm nois,i}\big)\ \pi\big(\boldsymbol{R}^*{}_{\rm i} = \boldsymbol{R}_{\rm i} \big| \mathcal{I}_{\boldsymbol{R}_{\rm i}}\big)}{\pi\big(\boldsymbol{D}_{\rm i} \big| \boldsymbol{\theta}_{\rm nois,i}, \boldsymbol{M}_{\rm nois,i}, \mathcal{I}_{\boldsymbol{R}_{\rm i}}\big)}\ . && (25)
\end{aligned}$$

Since there can be potentially infinitely many possibilities, $\boldsymbol{R}^*{}_{\rm i}$, corresponding to each event, $\boldsymbol{R}_{\rm i}$, the set $\mathcal{R}^*$ is not unique. Therefore, one can construct a superset $\boldsymbol{\mathcal{R}}^*$ consisting of all possible combinatorial realizations of the set of events $\mathcal{R}$,

$$\boldsymbol{\mathcal{R}}^* = \big\{\mathcal{R}^*{}_1, \ldots, \mathcal{R}^*{}_{\rm j}, \ldots\big\}\ , \tag{26}$$

where *only and only* one realization $\mathcal{R}^*{}_{\rm j} \in \boldsymbol{\mathcal{R}}^*$ corresponds to the truth set $\mathcal{R}$, although each member of $\boldsymbol{\mathcal{R}}^*$, $\mathcal{R}^*{}_{\rm j}$ satisfies (18) for some specific parameter values, $\boldsymbol{\theta}_{\rm phys}$. The superset $\boldsymbol{\mathcal{R}}^*$ can be of finite size, or countably/uncountably infinite, depending on the type of the measurement uncertainty involved in a problem. When all noise models, $\boldsymbol{M}_{\rm nois,i} \in \mathcal{M}_{\rm nois}$, for the specific problem under study give rise to sets of finite possibilities, $\boldsymbol{R}^*{}_{\rm i}$, for each event, $\boldsymbol{R}_{\rm i} \in \mathcal{R}$, then the size of $\boldsymbol{\mathcal{R}}^*$ would be also finite. Otherwise, if any or all noise models are countably/uncountably infinite, then $\boldsymbol{\mathcal{R}}^*$ would be also countably/uncountably infinite size.

Note that, corresponding to each set $\mathcal{R}^* \in \boldsymbol{\mathcal{R}}^*$ there should exist, by definition, at least one $\boldsymbol{\theta}_{\rm phys}$ that satisfies (2). Now, let $\boldsymbol{\mathcal{R}}^*{}_{\boldsymbol{\theta}_{\rm phys}} \subset \boldsymbol{\mathcal{R}}^* \subset \boldsymbol{\Omega}_{\mathcal{R}} \subset \mathbb{R}^{n_{\rm dv}}$ represent the set of all possible realizations, $\mathcal{R}^* \in \boldsymbol{\mathcal{R}}^*$, of $\mathcal{R}$ that satisfy (18) for a given $\boldsymbol{\theta}_{\rm phys}$. Then, the likelihood (i.e., the unnormalized probability) of $\boldsymbol{\theta}_{\rm phys}$ being the true set of parameter values for $\boldsymbol{M}_{\rm phys}$, can be written as the sum of the probabilities of all possible realizations, $\mathcal{R}^* \subset \boldsymbol{\mathcal{R}}^*{}_{\boldsymbol{\theta}_{\rm phys}}$, of the truth set, $\mathcal{R}$, that satisfy (18) for the given value of $\boldsymbol{\theta}_{\rm phys}$,

$$\begin{aligned}
\mathcal{L}\big(\boldsymbol{\theta}_{\rm phys}\big) &\equiv \int_{\boldsymbol{\mathcal{R}}^*{}_{\boldsymbol{\theta}_{\rm phys}}} \pi\big(\mathcal{R}^* = \mathcal{R} \big| \mathcal{D}, \mathit{\Theta}_{\rm nois}, \mathcal{M}_{\rm nois}\big)\ {\rm d}\mathcal{R}^* && (27)\\
&= \int_{\boldsymbol{\mathcal{R}}^* \subseteq \boldsymbol{\Omega}_{\mathcal{R}}} \mathbf{1}\big(\mathcal{R}^*\big)\ \prod_{\rm i=1}^{n_{\rm do}} \pi\big(\boldsymbol{R}^*{}_{\rm i} = \boldsymbol{R}_{\rm i} \big| \boldsymbol{D}_{\rm i}, \boldsymbol{\theta}_{\rm nois,i}, \boldsymbol{M}_{\rm nois,i}\big)\ {\rm d}\boldsymbol{R}^*{}_1 \cdots {\rm d}\boldsymbol{R}^*{}_{\rm n_{do}}\ , && (28)
\end{aligned}$$

where the second equality follows from (24), with $\mathbf{1}(\,\cdot\,)$ being an indicator function which has the value 1, only when all $\boldsymbol{R}^*{}_{\rm i} \in \mathcal{R}^*$ simultaneously satisfy (18) for the same value of $\boldsymbol{\theta}_{\rm phys}$,

$$\mathbf{1}\big(\mathcal{R}^* = \big\{\boldsymbol{R}^*{}_1, \ldots, \boldsymbol{R}^*{}_{\rm n_{do}}\big\}\big) \overset{\rm def}{=} \begin{cases} 1 & \mathcal{R}^* \in \boldsymbol{\mathcal{R}}^*{}_{\boldsymbol{\theta}_{\rm phys}}\ , \\ 0 & \text{otherwise}\ . \end{cases} \tag{29}$$

Thus, now the problem of inference is to compute and maximize the PDF of $\boldsymbol{\theta}_{\rm phys}$, which can be done by combining (27) with any prior knowledge, $\pi\big(\boldsymbol{\theta}_{\rm phys} \big| \mathcal{I}_{\boldsymbol{\theta}_{\rm phys}}, \boldsymbol{M}_{\rm phys}\big)$, about the parameters of the model using the Bayes rule,

$$\pi\big(\boldsymbol{\theta}_{\rm phys} \mid \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}, \boldsymbol{M}_{\rm phys}, \mathcal{I}_{\boldsymbol{\theta}_{\rm phys}}\big) = \frac{\left[\int_{\boldsymbol{\mathcal{R}}^*_{\boldsymbol{\theta}_{\rm phys}}} \pi\big(\mathcal{R}^* = \mathcal{R} \big| \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}\big) \, {\rm d}\mathcal{R}^*\right] \pi\big(\boldsymbol{\theta}_{\rm pi} \big| \boldsymbol{M}_{\rm phys}, \mathcal{I}_{\boldsymbol{\theta}_{\rm phys}}\big)}{\pi\big(\boldsymbol{\mathcal{R}}^* \big| \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}, \boldsymbol{M}_{\rm phys}, \mathcal{I}_{\boldsymbol{\theta}_{\rm phys}}\big)} \,, \tag{30}$$

where like (13), $\mathcal{I}_{\boldsymbol{\theta}_{\rm phys}}$ stands for all prior knowledge about the parameters of the model, and the denominator is simply a factor that properly normalizes the posterior distribution of $\boldsymbol{\theta}_{\rm phys}$ to a proper probability density function, such that the integral of (30) over the entire parameter space, $\boldsymbol{\Theta}_{\rm phys}$, is 1,

$$\pi\big(\boldsymbol{\mathcal{R}}^* \big| \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}, \boldsymbol{M}_{\rm phys}, \mathcal{I}_{\boldsymbol{\theta}_{\rm phys}}\big) = \int_{\boldsymbol{\Theta}_{\rm phys}} \left[\int_{\boldsymbol{\mathcal{R}}^*_{\boldsymbol{\theta}_{\rm phys}}} \pi\big(\mathcal{R}^* = \mathcal{R} \big| \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}\big) \, {\rm d}\mathcal{R}^*\right] \pi\big(\boldsymbol{\theta}_{\rm pi} \big| \boldsymbol{M}_{\rm phys}, \mathcal{I}_{\boldsymbol{\theta}_{\rm phys}}\big) \, {\rm d}\boldsymbol{\theta}_{\rm phys} \,. \tag{31}$$

It is clear from (22), (27), (28), (29), and (30) that the posterior PDF of the parameters of the physical model, $\pi\big(\boldsymbol{\theta}_{\rm phys} \big| \mathcal{D}, \mathcal{M}_{\rm nois}, \boldsymbol{M}_{\rm phys}, \mathcal{I}_{\boldsymbol{\theta}_{\rm phys}}\big)$, is explicitly independent of the physical model, $\boldsymbol{M}_{\rm phys}$, and its output. Instead, it is completely determined by the set of noise models, $\mathcal{M}_{\rm nois}$, of the available dataset, $\mathcal{D}$, and any probabilistic prior knowledge about the model parameters, $\boldsymbol{\theta}_{\rm phys}$.

The only influence of the physical model, $\boldsymbol{M}_{\rm phys}$, and its output on the posterior PDF of its parameters, $\boldsymbol{\theta}_{\rm phys}$, in (30) is through the definition of the domain of integration in the likelihood function of $\boldsymbol{\theta}_{\rm phys}$ in (27) or (28). This seemingly-bizarre behavior of the posterior PDF in (30) 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,
2. also 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 heterogenous with respect to the predictions of the physical model at hand.

Statement of the Hierarchical Bayesian Inverse Problem

Given,

1. the dataset $\mathcal{D} = \big\{\boldsymbol{D}_{\rm i}, \cdots, \boldsymbol{D}_{\rm n_{do}}\big\}$,
2. the corresponding set of noise models, $\mathcal{M}_{\rm nois}$, as given by (20) for each observation $\boldsymbol{D}_{\rm i} \in \mathcal{D}$,

3. a physical model, $\boldsymbol{M}_{\text{phys}}$,
4. a statistical physics-based model, $\boldsymbol{M}_{\text{inad}}$, for the inadequacy of $\boldsymbol{M}_{\text{phys}}$ 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 $\boldsymbol{\theta}_{\text{pi}} = \{\boldsymbol{\theta}_{\text{phys}}, \boldsymbol{\theta}_{\text{inad}}\}$.

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 (2) as given in (3). However, this equation takes as input, the truth, $\boldsymbol{R}_{\text{i}}$, about an observation, $\boldsymbol{D}_{\text{i}}$. Since $\boldsymbol{R}_{\text{i}}$ is unknown, we have to consider all possibilities, ${\boldsymbol{R}^*}_{\text{i}}$, for $\boldsymbol{R}_{\text{i}}$, whose PDF is given by (16).

To do so, consider for the moment, a single realization, $\mathcal{R}^*$, of the truth dataset, $\mathcal{R}$, as defined in (19). The corresponding equations to (3), (4), and (5) for $\mathcal{R}^*$ would be then,

$$
\begin{aligned}
{\boldsymbol{U}^*}_{\text{i}} &= {\boldsymbol{R}^*}_{\text{i}} - \boldsymbol{M}_{\text{phys}}\big({\boldsymbol{R}^*}_{\text{i}}, \mathcal{R}^*_{\boldsymbol{R}_{\text{i}}}, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}}\big) \ , && (32)\\
\mathcal{U}^* &= \{{\boldsymbol{U}^*}_1, \ldots, {\boldsymbol{U}^*}_{\text{n}_{\text{do}}}\} \ , && (33)\\
{\boldsymbol{U}^*}_{\text{i}} &\sim \boldsymbol{M}_{\text{inad}}\big({\boldsymbol{R}^*}_{\text{i}}, \mathcal{R}^*_{\boldsymbol{R}_{\text{i}}}, \boldsymbol{\theta}_{\text{inad}}, \boldsymbol{M}_{\text{phys}}({\boldsymbol{R}^*}_{\text{i}}, \mathcal{R}^*_{\boldsymbol{R}_{\text{i}}}, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}})\big) \ . && (34)
\end{aligned}
$$

Therefore, the modified equations corresponding to (6), (10), and (11) take the form,

$$
\begin{aligned}
\text{PDF}\big({\boldsymbol{U}^*}_{\text{i}}\big) &\overset{\text{def}}{=} \pi\big({\boldsymbol{U}^*}_{\text{i}} \big| \boldsymbol{\theta}_{\text{inad}}, \boldsymbol{M}_{\text{inad}}\big({\boldsymbol{R}^*}_{\text{i}}, \mathcal{R}^*_{\boldsymbol{R}_{\text{i}}}, \boldsymbol{\theta}_{\text{inad}}, \boldsymbol{M}_{\text{phys}}({\boldsymbol{R}^*}_{\text{i}}, \mathcal{R}^*_{\boldsymbol{R}_{\text{i}}}, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}})\big) && (35)\\
&\overset{\text{def}}{=} \pi\big(\boldsymbol{U}_{\text{i}} \big| \boldsymbol{\theta}_{\text{pi}}, \boldsymbol{M}_{\text{pi}}\big) \ , && (36)\\
\Rightarrow \mathcal{L}\big(\boldsymbol{\theta}_{\text{pi}}; \mathcal{U}^*\big) &\equiv \ \pi\big(\mathcal{U}^* \big| \boldsymbol{\theta}_{\text{pi}}, \boldsymbol{M}_{\text{pi}}\big) && (37)\\
&\overset{\text{i.i.d.}}{=} \prod_{\text{i}=1}^{n_{\text{do}}} \pi\big({\boldsymbol{U}^*}_{\text{i}} \big| \boldsymbol{\theta}_{\text{inad}}, \boldsymbol{M}_{\text{inad}}\big({\boldsymbol{R}^*}_{\text{i}}, \boldsymbol{\theta}_{\text{inad}}, \boldsymbol{M}_{\text{phys}}({\boldsymbol{R}^*}_{\text{i}}, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}})\big)\big) \ , && (38)\\
\text{or, } \mathcal{L}\big(\boldsymbol{\theta}_{\text{pi}}; \mathcal{R}^*\big) &\equiv \ \pi\big(\mathcal{R}^* \big| \boldsymbol{\theta}_{\text{pi}}, \boldsymbol{M}_{\text{pi}}\big) && (39)\\
&\overset{\text{i.i.d.}}{=} \prod_{\text{i}=1}^{n_{\text{do}}} \pi\big({\boldsymbol{R}^*}_{\text{i}} \big| \boldsymbol{\theta}_{\text{inad}}, \boldsymbol{M}_{\text{inad}}\big({\boldsymbol{R}^*}_{\text{i}}, \boldsymbol{\theta}_{\text{inad}}, \boldsymbol{M}_{\text{phys}}({\boldsymbol{R}^*}_{\text{i}}, \boldsymbol{\theta}_{\text{phys}}, \mathcal{S}_{\text{phys}})\big)\big) \ . && (40)
\end{aligned}
$$

Therefore, similar to (12) and (13), the posterior probability density of $\boldsymbol{\theta}_{\text{pi}}$ for a single realization, $\mathcal{R}^*$, of $\mathcal{R}$ can be computed by the Bayes rule as,

$$
\pi\big(\boldsymbol{\theta}_{\text{pi}} \big| \mathcal{U}^*, \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big) = \frac{\pi\big(\mathcal{U}^* \big| \boldsymbol{\theta}_{\text{pi}}, \boldsymbol{M}_{\text{pi}}\big) \pi\big(\boldsymbol{\theta}_{\text{pi}} \big| \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big)}{\pi\big(\mathcal{U}^* \big| \boldsymbol{M}_{\text{pi}}\big)} \ , \qquad (41)
$$

or equivalently,

$$
\pi\big(\boldsymbol{\theta}_{\text{pi}} \big| \mathcal{R}^*, \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big) = \frac{\pi\big(\mathcal{R}^* \big| \boldsymbol{\theta}_{\text{pi}}, \boldsymbol{M}_{\text{pi}}\big) \pi\big(\boldsymbol{\theta}_{\text{pi}} \big| \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big)}{\pi\big(\mathcal{R}^* \big| \boldsymbol{M}_{\text{pi}}\big)} \ , \qquad (42)
$$

However, the set $\mathcal{R}^*$ (or equivalently $\mathcal{U}^*$) is only one possibility among the (potentially infinitely) many possible representations of the truth set, $\mathcal{R}$. Therefore, the likelihood in (39) has to be further modified to include, not one, but all possibilities, $\mathcal{R}^* \in \boldsymbol{\mathcal{R}}^*$ of the reality $\mathcal{R}$.

Given the observed dataset, $\mathcal{D}$, and the associated set of the noise models, $\mathcal{M}_{\rm nois}$, and their parameters, $\Theta_{\rm nois}$, the probability of $\mathcal{R}^*$ being the truth, $\mathcal{R}$, is given by the posterior PDF in (22) and (23). Therefore, combining (23) with (39) yields the modified likelihood of the model parameters as,

$$
\begin{aligned}
&\mathcal{L}\big(\boldsymbol{\theta}_{\rm pi}\ ;\ \mathcal{R}^*, \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}, \mathcal{I}_{\mathcal{R}}\big) \\
&\quad\equiv\ \pi\big(\mathcal{R}^* \big| \boldsymbol{\theta}_{\rm pi}, \boldsymbol{M}_{\rm pi}\big)\ \times\ \pi\big(\mathcal{R}^* \big| \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}, \mathcal{I}_{\mathcal{R}}\big) && (43)\\
&\quad\overset{\text{i.i.d.}}{=}\ \prod_{\rm i=1}^{n_{\rm do}} \pi\big(\boldsymbol{R}^*{}_{\rm i} \big| \boldsymbol{\theta}_{\rm inad}, \boldsymbol{M}_{\rm inad}\big(\boldsymbol{R}^*{}_{\rm i}, \boldsymbol{\theta}_{\rm inad}, \boldsymbol{M}_{\rm phys}(\boldsymbol{R}^*{}_{\rm i}, \boldsymbol{\theta}_{\rm phys}, \mathcal{S}_{\rm phys})\big)\big) \\
&\quad\times\ \pi\big(\boldsymbol{R}^*{}_{\rm i} \big| \boldsymbol{D}_{\rm i}, \boldsymbol{\theta}_{\rm nois,i}, \boldsymbol{M}_{\rm nois,i}, \mathcal{I}_{\boldsymbol{R}_{\rm i}}\big)\ . && (44)
\end{aligned}
$$

One can then marginalize (45) over $\mathcal{R}^*$ to obtain the marginal likelihood of the parameters of the models, given only the *known quantities*: $\boldsymbol{\mathcal{R}}^*$, $\mathcal{D}$, $\Theta_{\rm nois}$, $\mathcal{M}_{\rm nois}$, $\boldsymbol{M}_{\rm phys}$, $\boldsymbol{M}_{\rm inad}$, $\mathcal{I}_{\mathcal{R}}$,

$$
\begin{aligned}
&\mathcal{L}\big(\boldsymbol{\theta}_{\rm pi}\ ;\ \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}, \mathcal{I}_{\mathcal{R}}\big) \\
&\quad\equiv\ \int_{\boldsymbol{\mathcal{R}}^*=\boldsymbol{\Omega}_{\mathcal{R}}} \pi\big(\mathcal{R}^* \big| \boldsymbol{\theta}_{\rm pi}, \boldsymbol{M}_{\rm pi}\big)\ \times\ \pi\big(\mathcal{R}^* \big| \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}, \mathcal{I}_{\mathcal{R}}\big)\ {\rm d}\mathcal{R}^* && (45)\\
&\quad\overset{\text{i.i.d.}}{=}\ \int_{\boldsymbol{\mathcal{R}}^*=\boldsymbol{\Omega}_{\mathcal{R}}} \prod_{\rm i=1}^{n_{\rm do}} \pi\big(\boldsymbol{R}^*{}_{\rm i} \big| \boldsymbol{\theta}_{\rm inad}, \boldsymbol{M}_{\rm inad}\big(\boldsymbol{R}^*{}_{\rm i}, \boldsymbol{\theta}_{\rm inad}, \boldsymbol{M}_{\rm phys}(\boldsymbol{R}^*{}_{\rm i}, \boldsymbol{\theta}_{\rm phys}, \mathcal{S}_{\rm phys})\big)\big) \\
&\quad\times\ \pi\big(\boldsymbol{R}^*{}_{\rm i} \big| \boldsymbol{D}_{\rm i}, \boldsymbol{\theta}_{\rm nois,i}, \boldsymbol{M}_{\rm nois,i}, \mathcal{I}_{\boldsymbol{R}_{\rm i}}\big)\ {\rm d}\boldsymbol{R}^*{}_1\ {\rm d}\boldsymbol{R}^*{}_2\ \cdots\ {\rm d}\boldsymbol{R}^*{}_{\rm n_{do}}\ . && (46)
\end{aligned}
$$

Despite their similarities, there is a fine difference between the marginalization over $\mathcal{R}^*$ performed in the above likelihood function and the marginalization in the case of an ideal physical model in the presence of noise, as appeared in (30) and (31) of §4. The symbol $\boldsymbol{\mathcal{R}}^*$ stands for the set of all realizations, $\mathcal{R}^*$, of the truth set $\mathcal{R}$ that satisfy (18). Therefore, the set $\boldsymbol{\mathcal{R}}^*$ is not necessarily equivalent to the entire sampling space of data, $\boldsymbol{\Omega}_{\mathcal{R}}$, as defined by $\big(\mathcal{D}, \Theta_{\rm nois}, \mathcal{M}\big)$. It is merely a subset of the sampling space of data: $\boldsymbol{\mathcal{R}}^* \subseteq \boldsymbol{\Omega}_{\mathcal{R}}$.

By contrast, not all possible realizations, $\mathcal{R}^* \in \boldsymbol{\mathcal{R}}^*$, of the truth set in the hierarchical likelihood function of (45) have to necessarily satisfy (18). We already know that the physical model can be (or is) inadequate in describing the reality, hence, the range of $\boldsymbol{\mathcal{R}}^*$ spans the entire sampling space of data: $\boldsymbol{\mathcal{R}}^* = \boldsymbol{\Omega}_{\mathcal{R}}$.

Now, with the hierarchical likelihood function of $\boldsymbol{\theta}_{\rm pi}$ at hand, one can use the Bayes formula to write the posterior PDF of $\boldsymbol{\theta}_{\rm pi}$ as,

$$
\begin{aligned}
&\pi\big(\boldsymbol{\theta}_{\rm pi}\ \big|\ \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}, \boldsymbol{M}_{\rm pi}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\boldsymbol{\theta}_{\rm pi}}\big) \\
&\quad= \frac{\mathcal{L}\big(\boldsymbol{\theta}_{\rm pi}\ ;\ \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}, \mathcal{I}_{\mathcal{R}}\big)\ \pi\big(\boldsymbol{\theta}_{\rm pi} \big| \boldsymbol{M}_{\rm pi}, \mathcal{I}_{\boldsymbol{\theta}_{\rm pi}}\big)}{\pi\big(\mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois} \big| \boldsymbol{M}_{\rm pi}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\boldsymbol{\theta}_{\rm pi}}\big)} && (47)\\
&\quad= \frac{\displaystyle\int_{\boldsymbol{\Omega}_{\mathcal{R}}} \pi\big(\mathcal{R}^* \big| \boldsymbol{\theta}_{\rm pi}, \boldsymbol{M}_{\rm pi}\big)\ \pi\big(\mathcal{R}^* \big| \mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois}, \mathcal{I}_{\mathcal{R}}\big)\ {\rm d}\mathcal{R}^*\ \pi\big(\boldsymbol{\theta}_{\rm pi} \big| \boldsymbol{M}_{\rm pi}, \mathcal{I}_{\boldsymbol{\theta}_{\rm pi}}\big)}{\pi\big(\mathcal{D}, \Theta_{\rm nois}, \mathcal{M}_{\rm nois} \big| \boldsymbol{M}_{\rm pi}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\boldsymbol{\theta}_{\rm pi}}\big)}\ , && (48)
\end{aligned}
$$

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

$$\begin{aligned}
&\pi\big(\boldsymbol{\mathcal{R}}^* \big| \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big) \\
&\equiv \pi\big(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} \big| \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big) && (49)\\
&= \int_{\boldsymbol{\Theta}_{\text{pi}}} \int_{\boldsymbol{\Omega}_{\mathcal{R}}} \pi\big(\mathcal{R}^* \big| \boldsymbol{\theta}_{\text{pi}}, \boldsymbol{M}_{\text{pi}}\big)\ \pi\big(\mathcal{R}^* \big| \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}\big)\ \pi\big(\boldsymbol{\theta}_{\text{pi}} \big| \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big)\ \mathrm{d}\mathcal{R}^*\ \mathrm{d}\boldsymbol{\theta}_{\text{pi}}\ . && (50)
\end{aligned}$$

Plugging (23) into (48) one gets,

$$\begin{aligned}
&\pi\big(\boldsymbol{\theta}_{\text{pi}} \big| \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big) \\
&= \frac{\displaystyle\int_{\boldsymbol{\Omega}_{\mathcal{R}}} \pi\big(\mathcal{D} \big| \mathcal{R}^*, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}\big)\ \pi\big(\mathcal{R}^* \big| \mathcal{I}_{\mathcal{R}}\big)\ \pi\big(\mathcal{R}^* \big| \boldsymbol{\theta}_{\text{pi}}, \boldsymbol{M}_{\text{pi}}\big)\ \mathrm{d}\mathcal{R}^*\ \pi\big(\boldsymbol{\theta}_{\text{pi}} \big| \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big)}{\pi\big(\mathcal{D} \big| \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \mathcal{I}_{\mathcal{R}}\big)\ \pi\big(\mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}} \big| \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big)}\ . && (51)
\end{aligned}$$

Thus in a sense, the two latter terms, $\pi\big(\mathcal{R}^* \big| \mathcal{I}_{\mathcal{R}}\big)\ \pi\big(\mathcal{R}^* \big| \boldsymbol{\theta}_{\text{pi}}, \boldsymbol{M}_{\text{pi}}\big)$, in the integrand in the numerator of (51), act like a prior on the likelihood of the observed dataset, $\pi\big(\mathcal{D} \big| \mathcal{R}^*, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}\big)$ and correct its value according to the physical and inadequacy models at hand, for a given set of values of their parameters, $\boldsymbol{\theta}_{\text{pi}}$.

In the case of i.i.d. events, ignoring the normalization constants, the posterior PDF of (51) takes the simple form,

$$\begin{aligned}
\pi\big(\boldsymbol{\theta}_{\text{pi}}\ &\big|\ \mathcal{D}, \Theta_{\text{nois}}, \mathcal{M}_{\text{nois}}, \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big) \\
&\propto \pi\big(\boldsymbol{\theta}_{\text{pi}} \big| \boldsymbol{M}_{\text{pi}}, \mathcal{I}_{\boldsymbol{\theta}_{\text{pi}}}\big) \\
&\times \prod_{\text{i}=1}^{n_{\text{do}}} \int_{\boldsymbol{\Omega}_{\mathcal{R}}} \pi\big(\boldsymbol{D}_{\text{i}} \big| \boldsymbol{R}^*{}_{\text{i}}, \boldsymbol{\theta}_{\text{nois,i}}, \boldsymbol{M}_{\text{nois,i}}\big)\ \pi\big(\boldsymbol{R}^*{}_{\text{i}} \big| \mathcal{I}_{\boldsymbol{R}_{\text{i}}}\big)\ \pi\big(\boldsymbol{R}^*{}_{\text{i}} \big| \boldsymbol{\theta}_{\text{pi}}, \boldsymbol{M}_{\text{pi}}\big)\ \mathrm{d}\boldsymbol{R}^*{}_{\text{i}}\ . && (52)
\end{aligned}$$

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 fittinga 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. Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. Statistical science pp 409–423
68. 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
69. Shahmoradi A (2013) Gamma-ray bursts: Energetics and prompt correlations. arXiv preprint arXiv:13081097
70. Shahmoradi A (2013) A multivariate fit luminosity function and world model for long gamma-ray bursts. The Astrophysical Journal 766(2):111
71. Shahmoradi A (2015) Dissecting the relationship between protein structure and sequence evolution. PhD thesis
72. Shahmoradi A, Nemiroff R (2009) How real detector thresholds create false standard candles. In: AIP Conference Proceedings, AIP, vol 1133, pp 425–427
73. Shahmoradi A, Nemiroff R (2011) A cosmological discriminator designed to avoid selection bias. In: Bulletin of the American Astronomical Society, vol 43
74. 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
75. Shahmoradi A, Nemiroff R (2014) Classification and energetics of cosmological gamma-ray bursts. In: American Astronomical Society Meeting Abstracts# 223, vol 223
76. 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
77. 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
78. 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
79. 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
80. Yager RR, Liu L (2008) Classic works of the Dempster-Shafer theory of belief functions, vol 219. Springer
81. Zellner A (1971) An introduction to bayesian inference in econometrics. Tech. rep.