MIT Open Access Articles

A generalized Bayesian approach to model calibration

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Tohme, Tony et al., "A generalized Bayesian approach to model calibration." Reliability Engineering & System Safety 204 (December 2020): 107141 doi. 10.1016/j.ress.2020.107141
©2020 Authors

As Published: <https://dx.doi.org/10.1016/j.ress.2020.107141>

Publisher: Elsevier BV

Persistent URL: <https://hdl.handle.net/1721.1/128133>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-NonCommercial-NoDerivs License

A GENERALIZED BAYESIAN APPROACH TO MODEL CALIBRATION*

Tony Tohme[†], Kevin Vanslette, Kamal Youcef-Toumi

Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139
{tohme, kvanslet, youcef}@mit.edu

ABSTRACT

In model development, model calibration and validation play complementary roles toward learning reliable models. In this article, we expand the Bayesian Validation Metric framework to a general calibration and validation framework by inverting the validation mathematics into a generalized Bayesian method for model calibration and regression. We perform Bayesian regression based on a user's definition of model-data agreement. This allows for model selection on any type of data distribution, unlike Bayesian and standard regression techniques, that "fail" in some cases. We show that our tool is capable of representing and combining least squares, likelihood-based, and Bayesian calibration techniques in a single framework while being able to generalize aspects of these methods. This tool also offers new insights into the interpretation of the predictive envelopes (also known as confidence bands) while giving the analyst more control over these envelopes. We demonstrate the validity of our method by providing three numerical examples to calibrate different models, including a model for energy dissipation in lap joints under impact loading. By calibrating models with respect to the validation metrics one desires a model to ultimately pass, reliability and safety metrics may be integrated into and automatically adopted by the model in the calibration phase.

Keywords Calibration and Validation · Reliability and Safety · Regression · Bayesian Validation Metric · Bayesian Model Testing · Bayesian Probability Theory · Inference

1 Introduction

The increasing complexity of engineering systems over the years has required a thorough understanding and assessment of their behaviour for their development, and the need for a framework that is capable of quantifying their reliability has become essential. One of the main objectives of reliability engineering is to identify and reduce the likelihood (or frequency) of failures. Given that most of the systems that constitute an essential part of our life (e.g. automotive, medical, energy, nuclear, etc.) are safety-critical systems, any failure in those systems can be harmful to people and the environment [1]. Hence, reliability, safety, and risk analysis [2, 3] are fundamental for designing, modeling, and testing any modern engineering system [4], and essential for risk-informed decision-making [5].

Parametric modeling is an integral part of reliability engineering and plays a critical role in designing, analysing and predicting the behavior of complex engineering systems [6]. When performing model-based reliability analysis, the uncertainty in the model parameters greatly affects the accuracy of the analysis. This uncertainty can be reduced by merging prior knowledge about the model parameters with available data describing the relation between the system inputs and outputs. Uncertainty can be aleatoric (natural variability) and/or epistemic (lack of knowledge) [6, 7, 8]. In this paper, we are interested in formulating a probabilistic framework – by expanding the Bayesian Validation Metric (BVM) framework [9] – for updating the model parameters as well as estimating and quantifying their uncertainty and

*This work was supported by the Center for Complex Systems (CCS) at King Abdulaziz City for Science and Technology (KACST) and the Massachusetts Institute of Technology (MIT).

[†]Corresponding Author.

the propagation of this uncertainty to the model response [10]. Our framework and methodology can be applied to a broad range of reliability engineering problems (e.g. the energy dissipation model [11, 12], as illustrated in Section 4.2.3, heat transfer models for fire insulation panels [13], energy models [14], dynamic thermal models [15], etc.) and deals with uncertainty through the use of confidence intervals and probability distributions [16].

Engineering systems are often represented and described by computational models in order to make predictions about the behavior of the system. Most of the time, models possess parameters that cannot be directly measured, and instead they are inferred based on experimental data of relevant inputs and outputs, in a process known as model calibration [17, 18, 19, 20]. Model calibration is the process of estimating and adjusting model parameters to obtain a model representation of the system (or data) of interest while satisfying a specific criterion (objective function). In model development, model calibration comes in the stage prior to the validation stage [21], and it usually consists of estimating model parameters given a set of observed input-output data. Then, validation is performed on a different independent data set called the validation data set. In Bayesian calibration, one generates a posterior distribution for the model parameters given some prior distribution of those parameters and the data available for calibration. In what follows, we refer to Bayesian calibration as Bayesian regression.

Representing and understanding data through learning models (by estimating their parameters) has always been a central problem in engineering and science. Least squares (or standard regression) [22], likelihood-based [23, 24], and Bayesian regression methods [25, 26, 27, 28, 29] are often used for model parameter estimation. Nonprobabilistic methods, such as parametric model regression, nonparametric neural networks, and support vector machines (SVM) [30] are able to tackle these types of problems efficiently. In Bayesian probability theory [31, 32, 33], Bayesian model testing and maximum likelihood methods provide probabilistic features (i.e. mean, covariance, distribution) for the parameters we aim to estimate, based on prior knowledge (i.e. prior distribution) and the uncertainty of the data. Bayesian model testing, which uses Bayesian parameter regression, was shown to be successful for signal detection, light sensor characterization [34], exoplanet detection [35], extra-solar planet detection [36], laser peening process [28], time series [37], astronomical data analyses [38], and cosmology and particle physics [39].

We believe that the efficacy of parametric Bayesian regression, likelihood-based, and standard regression methods can be improved. Bayesian regression methods calculate the Bayesian evidence, which is the probability the model could have produced the observed, usually over noisy or uncertain, data. If this probability is nonzero, one can proceed to calculate posterior model parameter probabilities using Bayes' Theorem. In practice, there are models and parameters that may be of interest to the user that Bayesian regression fails to regress and produce posterior parameter distributions – Figure 1. For some of the instances that Bayesian regression fails to provide a solution, standard regression methods may actually succeed, but usually with some measure of expected error. How this error can translate into parameter and model uncertainty in the presence of certain or uncertain data is a problem that is largely omitted in the literature except for a few analytic cases. In addition, standard regression methods cannot be used for model selection in which one could easily include their prior knowledge in a principled way.

Figure 1: Illustrative example of theoretical success and failure cases of Bayesian regression. In blue is a deterministic linear model ($y = ax + b$) and in red are the data probability distributions that may come from epistemic uncertainty, (constrained) aleatoric uncertainty, or from their combination.

Figure 1a shows normally distributed data (infinite tails data distribution). In this case, parametric Bayesian regression finds a linear model that sits in low probability regions of the data. Figure 1b shows uniformly distributed data (truncated data distribution). In this case, Bayesian regression cannot find a linear model solution because no linear model can pass through each data distribution simultaneously – the model given the data is regarded as impossible. Standard regression methods can provide linear model solutions here despite the model lying in a zero probability region of the data. Although this solution may be considered “wrong” because it is not supported by the data, it successfully provides useful information to the modeler (an increasing trend). The fact that, for the same model, the solution given by the calibration method can differ from method to method supports the search for a framework for their joint representation so they may be compared more concretely.

The Bayesian Validation Metric (BVM) was shown to be a general model validation and testing tool in [9]. The BVM is capable of representing all of the standard validation metrics (square error [29], reliability [11], improved reliability [12], probability of agreement [40], frequentist [41], area [20], statistical hypothesis testing [42], Bayesian model testing [32, 36, 42, 43], etc.) as special cases. Using BVM model testing, the BVM selects models according to whatever definition of model-data agreement the modeler might find useful. It was found that the BVM is able to generalize the Bayesian model testing framework, which allowed the problem of model validation to be expressed in a single framework.

In this article we represent least squares, likelihood-based, and Bayesian regression (or calibration) methods by expanding the general validation framework (BVM) into a general calibration and validation framework. Our method uses the BVM to guide the regression of a model in a flexible way. Several of our examples use generalizations of the improved reliability metric and thus reliability is automatically taken into account in our model solutions. Our method gives us better control over the predictive envelopes of the model under question, which can be used to improve model reliability and safety. By learning model parameters with the BVM, we are able to estimate and construct model parameters distributions for any type of data distribution (Gaussian, Uniform, Completely Certain, etc.), which addresses the concerns raised in Figure 1. This construction gives us additional insight into the meaning of the predictive envelopes of Bayesian regression methods.

We have found that a subset of our method shares mathematical features with Approximate Bayesian Computation (ABC) methods, which are also known as likelihood-free techniques [44, 45]. ABC methods are used strictly as an approximation method for nearly computationally intractable likelihoods in Bayesian regression. While our method gains this feature in some cases, our method’s intention is not to approximate Bayesian regression, but instead to generalize it for the purpose of robust and flexible model calibration.

Our method is able to regress models over a multitude of different data distributions by using likelihoods that are modified by user’s choice of a useful definition of agreement between the data and the model – leaning on the BVM formalism. This “choice” allows the user to program safety requirements into the model learning process if they desire. The nature of the BVM formalism forces one to express the model and data assumptions explicitly and thus, our method leads to improved model transparency and safety. In our examples we show how such a procedure leads to a model that better represents the uncertain data at hand than Bayesian and standard regression techniques. This naturally improves the model’s reliability and safety in the presence of uncertain data.

The remainder of the article is organized as follows. In Section 2, we discuss the research contribution and motivation, review Bayesian regression and model testing as well as the Bayesian Validation Metric (BVM). We then move to Section 3 where we derive our theoretical solutions for BVM regression and model selection (or learning), under different types of data distributions (or uncertainty) and according to user-specified definitions of model-data agreement. In Section 4, we illustrate the proposed method and provide three numerical examples for model calibration including a model used to predict the energy dissipation in lap joints under dynamic loading.

2 Contribution and Motivation

In this section, we will outline the research contribution and motivation. We will also introduce the notation adopted in this paper, review Bayesian regression methods, discuss their advantages and disadvantages, and review the Bayesian Validation Metric (BVM).

2.1 Contribution

In our earlier work [9], we introduced and constructed the Bayesian Validation Metric (BVM) as a general model validation and testing tool. The BVM was shown to represent and generalize all of the well-known validation metrics (e.g. reliability metric, etc.). The BVM was also shown to generalize Bayesian Model Testing and allow for model selection based on arbitrary definitions of model-data agreement. In this paper, we extend this work and present further development of the BVM framework that results in generalizing Bayesian methods for regression and model learning. The resulting framework can then be used in both the calibration and validation stages of model development. Estimating and quantifying predictive uncertainty in least squares (or standard regression) [22], likelihood-based [23, 24], and Bayesian regression [25, 26, 27, 28, 29] methods is a very challenging task. In addition, these methods face several problems (and fail) when the data they are dealing with have bounded probability distributions or are deterministic (completely certain).

In this paper, we show how our extended BVM framework succeeds and outperforms those methods for any type of data uncertainty (or distribution). In addition, our method is successful when it comes to predictive uncertainty estimation, as we will see in our examples. Given the flexibility that the BVM framework provides when defining the agreement between the model outputs and the observed data, we show how this flexibility translates into more freedom and control over the predictive envelopes resulting from the posterior distributions of the model parameters. Finally, our method is very simple to implement, requiring few modifications to standard Markov Chain Monte Carlo (MCMC) methods or any Bayesian inference tool that calculates the evidence and produces posterior samples (as we will see, the only task required is to come up with or derive the appropriate likelihood function to be fed into the MCMC algorithm). These contributions are formulated in Sections 3 and 4.

2.2 Notation

We start by introducing the notation and language used in this paper. For an input x , we let $\hat{y}$ denote the output of a model M , i.e. $\hat{y} = M(x; \vec{\alpha})$, where $\vec{\alpha} = (\alpha_1, \dots, \alpha_m)$ represents the vector of model parameters we wish to estimate. We let $X = (x_1, \dots, x_n)$ denote a set of inputs, $\hat{Y} = (\hat{y}_1, \dots, \hat{y}_n)$ denote a set of model outputs, i.e. $\hat{Y} = M(X; \vec{\alpha})$, y denote a data point (or observed data), and Y denote a set of data points. We let $\mathcal{L}$ denote the likelihood, $\mathcal{Z}$ denote the marginal likelihood (or model evidence), and A denote model-data *agreement*.

2.3 Bayesian Regression and Model Testing

In this section, we present Bayesian regression and Bayesian model testing (BMT) while introducing some probability notations to be used throughout this paper. In Bayesian regression, rather than performing regression to learn the model parameters, one performs regression to learn the posterior probability distribution of the model parameters. That is, one estimates a set of parameters $\vec{\alpha}$ in a model (or hypothesis) $M \equiv M(x; \vec{\alpha})$ for the data D (where x is the model input). The defining equation of Bayesian regression is the learning of the posterior parameter distribution from the prior via Bayes' Rule,

$$\rho(\vec{\alpha}|M) \xrightarrow{*} \rho(\vec{\alpha}|D, M) = \frac{\rho(D|\vec{\alpha}, M) \rho(\vec{\alpha}|M)}{\rho(D|M)}, \quad (1)$$

where, for reasons that will become obvious, we have borrowed the more explicit notation from [9]. In Bayesian regression and model testing, these probabilities are named as follows:

$$\begin{aligned} \rho(\vec{\alpha}|D, M) &\equiv \mathcal{P}(\vec{\alpha}) \text{ is the posterior probability of the parameter,} \\ \rho(D|\vec{\alpha}, M) &\equiv \mathcal{L}(\vec{\alpha}) \text{ is the likelihood function,} \\ \rho(\vec{\alpha}|M) &\equiv \pi(\vec{\alpha}) \text{ is the prior probability,} \\ \rho(D|M) &\equiv \mathcal{Z} \text{ is the marginal likelihood or Bayesian evidence.} \end{aligned}$$

After learning the posterior distribution of the model parameters (given $\vec{\alpha}$ is the vector of parameters), we can evaluate the predictive distribution defined by:

$$\rho(\hat{y}|D, M) = \int_{\vec{\alpha}} \rho(\hat{y}|\vec{\alpha}, D, M) \cdot \rho(\vec{\alpha}|D, M) d\vec{\alpha}. \quad (2)$$

To perform Bayesian regression, one must calculate the Bayesian evidence, which is the marginal likelihood over $\vec{\alpha}$,

$$\mathcal{Z} = \rho(D|M) = \int_{\vec{\alpha}} \underbrace{\rho(D|\vec{\alpha}, M)}_{\mathcal{L}(\vec{\alpha})} \cdot \underbrace{\rho(\vec{\alpha}|M)}_{\pi(\vec{\alpha})} d\vec{\alpha}. \quad (3)$$

After performing regression and solving for the model parameters' values, rather than selecting the model with the lowest estimated generalization error as is done in standard regression, one instead uses BMT to select the model with the highest probability given the data. That is, for two Bayesian regressed models M_1 and M_2 , BMT uses the Bayes ratio, R , and rank the data-informed posterior model probabilities. It can be expressed in several ways using Bayes Rule,

$$R \equiv \frac{p(M_1|D)}{p(M_2|D)} = \frac{\rho(D|M_1)p(M_1)}{\rho(D|M_2)p(M_2)} = \frac{\mathcal{Z}_1 p(M_1)}{\mathcal{Z}_2 p(M_2)}.$$

If there is no reason to suspect that one model is more probable than another prior to observing the data, we may set the ratio of the prior probabilities of the model $p(M_1)/p(M_2) = 1$, *a priori*. In this case one gets,

$$R \rightarrow \frac{\mathcal{Z}_1}{\mathcal{Z}_2} \equiv K,$$

where K denotes the Bayes factor and is the ratio of model evidences. The Bayes factor is usually more accessible than R so it is usually used for model selection.

Bayesian regression has several positive and negative attributes. As a byproduct, Bayesian regression can perform model selection in a principled way that allows one to incorporate their prior knowledge into the selection process using BMT. Because Bayesian regression requires regressing probability distributions rather than just single model predictions, it can become intractable to calculate in general if the number of dimensions are large (as would standard regression if uncertainty is taken into account). Regularization in Bayesian regression is interpreted as coming from the uncertainty of the data and the uncertainty present in the prior parameters [30], which we view as being a potential drawback. If one wants to change the regularization it would require changing either of these uncertainties, or both, "artificially" because one would be tuning their prior probabilities *after* regression, which is a bit anti-Bayesian. Regularization can lead to an unnatural reduction of the posterior variances of the parameters for parametric models (a common problem with standard regression).

Further, we highlight some technical gaps found in Bayesian regression and model testing. Although almost all instances of Bayesian regression use data probability distributions that have infinite tails, truncated (or bounded) data probability density functions (pdfs) are realistic in practice too. We find that truncated data pdfs are potentially problematic for Bayesian regression if the model is deterministic. In the extreme case of completely certain data, Bayesian regression methods usually do not terminate because the Bayesian evidence is zero in (1) since there are no possible combinations of parameter values that could exactly fit the data. This problem may also arise if the data uncertainties are bounded. In principle, standard regression methods can produce a solution regardless of the form of the data pdf. In what follows, we assume we are given a set of inputs X , a set of data points $Y = D$, and a set of model outputs $\hat{Y} = M(X; \vec{\alpha})$. All the sets are n -dimensional. In addition, we assume that the n data points were collected through independent experiments. We give explicit examples of the likelihoods below (see Appendix A) given the model is deterministic:

Infinite Tail Data Distributions

Data distributions with infinite tails result in likelihoods with infinite tails in (3). Some examples of infinite tail data distributions are Gaussian, Student-t, Laplace, canonical, and Poisson. For example, Gaussian distributed data (see Figure 1a) naturally has an infinite tailed likelihood function,

$$\mathcal{L}(\vec{\alpha}) = \frac{1}{\sqrt{(2\pi)^n |\Delta|}} e^{-\frac{1}{2} (M(X; \vec{\alpha}) - D)^T \Delta^{-1} (M(X; \vec{\alpha}) - D)}, \quad (4)$$

where Δ is the covariance matrix. Since the likelihood has infinite tails, the predicted model response $M(x_j; \vec{\alpha})$ has probabilistic flexibility around its corresponding data point D_j because it is uncertain. Even far from D , Bayesian regression is capable of estimating the posterior probability distributions of the model parameters in question as they are nonzero.

Truncated Tail Data Distributions

Data distributions with truncated tails naturally lead to truncated likelihoods in (3). For example, if the uncertain data is bounded to a region and is uniformly distributed, i.e. $D_j \sim \mathcal{U}(a_j, b_j)$ (see Figure 1b), then the likelihood function is,

$$\mathcal{L}(\vec{\alpha}) = \prod_{j=1}^n \frac{\Theta(a_j \leq M(x_j; \vec{\alpha}) \leq b_j)}{b_j - a_j}, \quad (5)$$

where $\Theta(\cdot)$ is the indicator function. In other words, for the likelihood $\mathcal{L}(\vec{\alpha})$ to be nonzero, the predicted model response $M(x_j; \vec{\alpha})$ at x_j must lie within the interval $[a_j, b_j]$ for all j simultaneously. The function space defined by the model and uncertain parameters is constrained by the data. This can make the probability of estimating a regressed posterior probability distribution of the model parameters very small, and in some cases impossible, because the likelihood may evaluate to zero for almost all combinations of $\vec{\alpha}$.

This point is exaggerated if the data is completely certain or deterministic, because the likelihood function becomes

$$\mathcal{L}(\vec{\alpha}) = \delta(M(X; \vec{\alpha}) - D), \quad (6)$$

where $\delta(\cdot)$ is the Dirac delta function. In this case, the model output and observed data only agree if their values are exactly equal, i.e. $M(X; \vec{\alpha}) \rightarrow D$ for all n points, which in most cases, is only possible if we overfit the data or the model is perfect. Thus, Bayesian regression will usually fail in this case, or if it succeeds, it only produces singular posterior distributions of the model parameters (i.e. $\sigma_{\vec{\alpha}} = 0$). When Bayesian regression fails, the Bayesian evidence is zero, which, although correct (the model does not support/fit the data), may not be the most useful type of answer for the modeler. It seems reasonable that a modeler would want both the benefits of Bayesian and standard regression simultaneously.

2.4 BVM Model Testing

We present the Bayesian Validation Metric (BVM) proposed in [9]. The BVM represents model to data validation in a general way using the probability of agreement,

$$\begin{aligned} p(A|M, D, B) &= \int_{\hat{z}, z} p(A|\hat{z}, M, z, D, B) \cdot \rho(\hat{z}, z|M, D) d\hat{z} dz \\ &= \int_{\hat{z}, z} \rho(\hat{z}|M, D) \cdot \Theta(B(\hat{z}, z)) \cdot \rho(z|D) d\hat{z} dz, \end{aligned} \quad (7)$$

where $\hat{z}$ and z are the model and data comparison quantities, respectively. The “agreement kernel” $p(A|\hat{z}, M, z, D) = \Theta(B(\hat{z}, z))$ is the indicator function of a user-defined Boolean function, $B(\hat{z}, z)$, that defines the context of what is meant by “model to data agreement” by being true when $(\hat{z}, z)$ agree or false otherwise. For simplicity, we will assume $\hat{z} \rightarrow \hat{y}$ and $z \rightarrow y$ are the model output and observed data respectively.

The BVM model testing framework was shown to generalize BMT where the probability of agreement plays the role of the evidence,

$$\mathcal{Z}(B) = p(A|M, D, B) = \int_{\vec{\alpha}} \underbrace{p(A|\vec{\alpha}, M, D, B)}_{\mathcal{L}(\vec{\alpha}, B)} \cdot \underbrace{\rho(\vec{\alpha}|M) d\vec{\alpha}}_{\pi(\vec{\alpha})}. \quad (8)$$

where $\mathcal{Z}(B)$ and $\mathcal{L}(\vec{\alpha}, B)$ are the BVM evidence and likelihood, respectively, that have been modified by a user’s definition of model-data agreement B . Analogous to the Bayesian model testing framework, we can perform BVM model testing between two models M_1 and M_2 using the probability of agreement defined above as follows

$$R(B) \equiv \frac{p(M_1|A, D, B)}{p(M_2|A, D, B)} = \frac{p(A|M_1, D, B) p(M_1|D, B)}{p(A|M_2, D, B) p(M_2|D, B)} = \frac{\mathcal{Z}_1(B) p(M_1|D, B)}{\mathcal{Z}_2(B) p(M_2|D, B)},$$

where $p(M_1|D, B)/p(M_2|D, B)$ is the ratio of prior probabilities of M_1 and M_2 , which can often be set to unity, i.e. $p(M_1|D, B)/p(M_2|D, B) = 1$. In this case, we get

$$R(B) \rightarrow \frac{p(A|M_1, D, B)}{p(A|M_2, D, B)} = \frac{\mathcal{Z}_1(B)}{\mathcal{Z}_2(B)} = K(B) \quad (9)$$

where $R(B)$ denotes the BVM ratio and $K(B)$ denotes the BVM factor, which is analogous to Bayes factor.

2.5 Example: Improved Reliability Metric

The reliability metric discussed in [11] is defined as the probability that the mean of the model prediction is within a tolerance ϵ of the mean of the data. This metric was later expanded in [12] to consider tolerances between each of the model and data point pairs, $|y_j - \hat{y}_j| \leq \epsilon_j$ for $j = 1, \dots, n$, rather than comparing their means.

Consider a set of inputs X , a set of model outputs $\hat{Y}$ and a set of observed data points Y . Assume that all the sets are n -dimensional and that the data were collected through independent experiments. The improved reliability metric r_i is,

$$r_i = \int_Y \rho(Y|D) \left(\int_{Y-\epsilon}^{Y+\epsilon} \rho(\hat{Y}|M) d\hat{Y} \right) dY, \quad (10)$$

which can always be rewritten as,

$$r_i = \int_{\hat{Y}, Y} \rho(\hat{Y}|M) \cdot \Theta(|\hat{Y} - Y| \leq \epsilon) \cdot \rho(Y|D) d\hat{Y} dY. \quad (11)$$

This equation may be identified as a special case of the BVM (7) when,

$$\Theta(B(\hat{z}, z)) \rightarrow \Theta(B(\hat{Y}, Y)) = \Theta(|\hat{Y} - Y| \leq \epsilon) = \prod_{j=1}^n \Theta(|\hat{y}_j - y_j| \leq \epsilon_j), \quad (12)$$

and where $\hat{z} = \hat{Y}$, $z = Y$ [9]. Thus, this agreement kernel is based on the ϵ -Boolean, as we call it in later sections. From (8) (see Appendix B for details), the BVM in this case is,

$$\mathcal{Z}(B) = p(A|M, D, B) = \int_{\vec{\alpha}} \left(\int_Y \Theta(|M(X; \vec{\alpha}) - Y| \leq \epsilon) \cdot \rho(Y|D) dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}. \quad (13)$$

Thus, the choice of the Boolean agreement function results in representing the different validation metrics in terms of the BVM [9]. The ϵ -Boolean participates in several of our BVM regression examples in the following sections and thus reliability is automatically regressed into our model solutions through the improved reliability metric.

3 Generalized Bayesian Regression via the BVM

This section introduces BVM regression, which generalizes Bayesian and standard regression. This method has the ability to produce posterior parameter distributions and predictive envelopes for any data distribution, include prior knowledge about model parameters (if there is any), and regularize parameter solutions in a way that parameter uncertainty increases rather than decreases.

BVM regression consists of learning the posterior of a set of parameters $\vec{\alpha}$, given the agreement A and the Boolean function B , from the prior via Bayes' Rule,

$$\rho(\vec{\alpha}|M) \xrightarrow{*} \mathcal{P}(\vec{\alpha}|A) \equiv \rho(\vec{\alpha}|A, M, D, B) = \frac{p(A|\vec{\alpha}, M, D, B) \rho(\vec{\alpha}|M)}{p(A|M, D, B)}. \quad (14)$$

After learning the posterior distribution of the model parameters, we can evaluate the predictive distribution defined by:

$$p(\hat{y}|A, M, B) = \int_{\vec{\alpha}} p(\hat{y}|\vec{\alpha}, A, M, B) \cdot \rho(\vec{\alpha}|A, M, D, B) d\vec{\alpha}. \quad (15)$$

Performing BVM regression requires evaluating the BVM probability of agreement. At the beginning of Appendix B, we give a derivation showing that (8) can be written as,

$$\mathcal{Z}(B) = p(A|M, D, B) = \int_{\vec{\alpha}} \left(\int_Y \Theta(B(M(X; \vec{\alpha}), Y)) \cdot \rho(Y|D) dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}, \quad (16)$$

which is analogous to (3) in form,³ and where the comparison values are $\hat{z} = \hat{Y} = M(X; \vec{\alpha})$ and $z = Y$ (we assume we are dealing with a set of inputs, model outputs (i.e. a set of n points on a model curve), and n observed data points, where all the data points were collected through independent experiments).

³In terms of the BVM, the Bayesian evidence in BMT, i.e. Equation (3), may be interpreted as the probability that the uncertain data and model output are exactly equal, i.e. $\rho(D|M) \equiv \rho(\hat{Y} \equiv Y|M, D) \equiv \rho(A|M, D)$ which is Equation (25) derived in Appendix A.

BVM regression can reproduce Bayesian regression, standard regression, and likelihood-based methods as special cases. When the data and model outputs must be exactly equal to agree with one another (i.e. $\delta(\hat{Y} - Y)$), the BVM produces BMT as a special case and the regression solutions are given in Appendix A. Typical likelihood-based methods follow from the same “exactly equal” definition of model-data agreement. We find that the Boolean function $B_{S,R}(\hat{Y}(\vec{\alpha}^*), Y)$ that reproduces standard regression is defined to be true iff $\vec{\alpha}^* = \underset{\vec{\alpha}}{\operatorname{argmin}} \mathcal{E}(M(X; \vec{\alpha}), Y)$ for some objective function $\mathcal{E}(\cdot)$. This only gives nonsingular posterior parameter distributions and predictive model envelopes if the data is uncertain and/or if $\mathcal{E}(M(X; \vec{\alpha}), Y)$ does not have a unique global minimum.

If the objective function is convex, then we have a single minimum which results in one vector of parameters $\vec{\alpha}^*$ that makes $B_{S,R}(\hat{Y}(\vec{\alpha}^*), Y)$ true. However, when the cost function is non-convex, then multiple parameter vectors $\vec{\alpha}^*$ corresponding to different local minima, lead to a true $B_{S,R}(\hat{Y}(\vec{\alpha}^*), Y)$ and may be accepted due to the approximate nature of non-convex optimization methods. This results in multiple regressed solutions for the regression problem and approximates the posterior parameters’ distribution $\rho(\vec{\alpha}|A)$ (analogous to the accepted parameter samples in MCMC simulation). Marginalizing leads to the predictive posterior model distribution $p(\hat{y}|A, M, B)$ as in (15). Finding the predictive model output average is analogous to the results obtained in the ensemble methods in machine learning [46]. Because the BVM can reproduce these special cases and generate new ones by extending, combining, and modulating Boolean agreement functions, BVM regression may be seen as a generalized regression method.

Due to the flexibility of the BVM framework, there are many possible definitions of agreement that the user can define. Table 1 below contains some of these definitions.

	Agreement Boolean Function
ϵ -Boolean	True iff $ y_j - M(x_j; \vec{\alpha}) \leq \epsilon_j \forall j$
(γ, ϵ, ℓ) -Boolean	True iff $ y_j - M(x_j; \vec{\alpha}) \leq \ell \epsilon_j \forall j$ and $\frac{1}{n} \sum_j \Theta(y_j - M(x_j; \vec{\alpha}) \leq \epsilon_j) \geq \gamma\%$
$\langle \epsilon \rangle$ -Boolean	True iff $\frac{1}{n} \sum_j y_j - M(x_j; \vec{\alpha}) \leq \langle \epsilon \rangle$
$(\langle \epsilon \rangle, \hat{\alpha})$ -Boolean	True iff $\frac{1}{n} \sum_j y_j - M(x_j; \vec{\alpha}) \leq \langle \epsilon \rangle$ and $0.91 \leq \frac{1}{n} \sum_j \Theta(y_j \in [-c_{\hat{\alpha}}, c_{\hat{\alpha}}]_j) \leq 0.99$

Table 1: Some examples of agreement Boolean functions.

To address the concerns we raised about Bayesian and standard regression depicted in Figure 1, consider using the ϵ -Boolean with the agreement kernel,

$$\Theta(B(M(x_j; \vec{\alpha}), y_j)) = \begin{cases} 1, & \text{if } |y_j - M(x_j; \vec{\alpha})| \leq \epsilon_j \\ 0, & \text{otherwise} \end{cases}$$

for all $j = 1, \dots, n$, where ϵ_j may be adjusted and tuned to impose more or less strict agreement conditions which may be used by the modeler to enforce reliability in some region or to be more tolerant of training errors at instance x_j . For simplicity, We assume that $\epsilon_j = \epsilon$ for all j . Note that this is (12) in Section 2.5. In other words, we use the special case of the BVM, the improved reliability metric with evidence derived in (13), to derive our theoretical solutions.⁴ Utilizing this BVM definition allows us to solve the truncated tail data distributions problem in Bayesian regression in a simple way – details in Appendix B.⁵

Truncated Tail Data Distributions Solution Summary

Let the data be known to have the truncated pdf $y_j \sim \mathcal{U}(a_j, b_j)$, for $j = 1, \dots, n$. By using the ϵ -Boolean, we introduce leniency into the regression in that it no longer needs to exactly pass through all intervals $[a_j, b_j]$ simultaneously to count as a “fit”. This produces likelihood functions such as,

$$\mathcal{L}(\vec{\alpha}, B) = \prod_{j=1}^n \frac{u_j - l_j}{b_j - a_j}, \quad (17)$$

where l_j and u_j are defined by the boundaries of the intersection of the data uncertainty and the model’s tolerance ϵ ,

⁴Note that by choosing to adopt a different agreement kernel (or Boolean function) as in Section 4.2.2, we generalize the results derived above; this is the power of the BVM.

⁵A complete derivation for the infinite tail Gaussian data distribution is given in Appendix B.1.

$$[l_j, u_j] = [M(x_j; \vec{\alpha}) - \epsilon, M(x_j; \vec{\alpha}) + \epsilon] \cap [a_j, b_j] \quad j = 1, \dots, n.$$

An illustration of how the BVM works with truncated data distributions is shown in Figure 2 below. For example, at instance x_2 , the interval $[l_2, u_2]$ is found by intersecting the intervals $[a_2, b_2]$ and $[M(x_2; \vec{\alpha}) - \epsilon, M(x_2; \vec{\alpha}) + \epsilon]$. Note that this applies to the instances x_j for all j . In this case, the likelihood is nonzero, resulting in a nonzero evidence (16). Thus, given this agreement definition, the probability of finding a model given the truncated data is nonzero.

Figure 2: Truncated tail data distributions solution. Using BVM regression results in a nonzero probability of finding a model given the observed truncated data.

Now, if we consider the special case when the data is completely certain, deterministic, i.e. $Y = D$, then the likelihood function is

$$\mathcal{L}(\vec{\alpha}, B) = \Theta(B(M(X; \vec{\alpha}), D)), \quad (18)$$

which can be seen as a relaxed general form of the delta function adopted in the Bayesian model testing (where $\epsilon = 0$), which implies that the model output must be within ϵ from the observed measurements in order for them to agree. An analogous ϵ -Boolean solution exists for standard regression methods which leads to nonsingular parameter distributions whether the regression is regularized or not.

Tolerant agreement as a new kind of regularization

The purpose of regularization is to better represent one's expectations of unobserved data using the chosen model or model class. Using BVM regression and nonzero agreement tolerances (e.g. $\epsilon > 0$ in the ϵ -Boolean), we can broaden the model's prediction envelope to better represent our expectations of the data. Increasing agreement tolerances naturally increases the posterior variance of the parameters, which differs from standard regularization methods and can be used to avoid conceptual issues of interpreting regularized physical parameters. It should also be noted that this is done without changing the prior distributions of the parameters nor the given probability distributions of the data. This becomes a useful feature in our first example.

4 Implementation and Examples

4.1 Implementation Procedure

Like the Bayesian evidence, the BVM evidence is computationally expensive to calculate when one has many model parameters to learn. Several approaches were adopted to solve this problem. Markov Chain Monte Carlo (MCMC) is a computational technique used for Bayesian methods that has been widely studied and improved [47, 48, 49, 50, 51, 52] as it is considered an indispensable tool for Bayesian inference. Other techniques include the Nested Sampling method [38, 53] and the MultiNest algorithm [39].

We will approximate the BVM evidence and generate the posterior model parameter distributions (for the purpose of generating model's predictive envelopes) using MCMC (with proposal or candidate samples being accepted/rejected according to Metropolis-Hastings criterion [33, 48]). The MCMC algorithm takes the following main inputs: the likelihood function $\mathcal{L}$ and the prior distribution $\pi(\vec{\alpha})$ of the model parameters, a model M and a set of input/output data points $\{X, Y\}$. The Bayesian terms $(\mathcal{Z}, \mathcal{L})$ have analogous BVM terms $(\mathcal{Z}(B), \mathcal{L}(B))$ and it is therefore

straightforward to extend the MCMC algorithm to BVM calculations to obtain posterior parameter distributions. These distributions are used to generate a model's predictive envelope. Note that the MCMC algorithm takes some additional inputs (e.g. the number of iterations, burn-in, etc.) that are usually optional or chosen according to the user's preference and the problem setting.

As we outlined above, our generalized Bayesian regression method is simple and easy to implement. Algorithm 1 summarizes the implementation procedure that we will adopt later in the illustrative numerical case study and examples.

Algorithm 1 Implementation Procedure

1. Prepare the data $\{X, Y\}$ to be used for regression/calibration.
 2. Determine the model M (to fit the data) and its parameters' prior distributions $\pi(\vec{\alpha})$.
 3. Choose the comparison quantities (usually the model output and the observed data) and an appropriate Boolean agreement function B (it is recommended to include the reliability metric inside the agreement function).
 4. Derive the likelihood function $\mathcal{L}(B)$ (after specifying the type of data uncertainty) and determine a method to compute it, either analytically or numerically.
 5. Compute the BVM evidence $\mathcal{Z}(B) \leftarrow \text{MCMC}(\{X, Y\}, \mathcal{L}(B), M, \pi(\vec{\alpha}), \text{etc.})$. This will produce the parameters' posterior distributions $\mathcal{P}(\vec{\alpha})$.
-

4.2 BVM Regression Examples

4.2.1 Numerical Example 1: Bacterial Growth Model

We consider the case study investigated in [54] using a bacterial growth model. The data is obtained by operating a continuous flow biological reactor at steady-state conditions. The observations are as follows:

x (mg/L COD)	28	55	83	110	138	225	375
y (1/h)	0.053	0.060	0.112	0.105	0.099	0.122	0.125

Table 2: The observations we aim to fit.

where y is the growth rate at substrate concentration x . We replicate the results found in [54] using the nonlinear Monod model to fit the data, i.e.,

$$\hat{y} = M(x; \vec{\alpha}) = \frac{\alpha_1 x}{\alpha_2 + x} \quad (19)$$

where α_1 is the maximum growth rate (h^{-1} : per hour), and α_2 is the saturation constant (mg/L COD: the Chemical Oxygen Demand, measured in milligrams per liter).

Following the procedure outlined in Section 4.1 is straightforward. We first prepare the data: from Table 2, we have $\{X, Y\} = \{x, y\}$. We then determine the model M which, in this case, is described by Equation (19). To determine the prior distribution of the parameters $\pi(\alpha_1)$ and $\pi(\alpha_2)$ (if no prior knowledge about the parameters is known), a good practice is to assume that the prior distribution is Gaussian around a point close to the solution resulting from applying least squares (i.e. minimizing the sum of the squares of the residuals), with some standard deviation. In our case, the least squares solution is $\alpha_1 = 0.14542$ and $\alpha_2 = 49.053$, and thus we assume $\pi(\alpha_1) = \mathcal{N}(0.17, \sigma_{\alpha_1}^2)$ and $\pi(\alpha_2) = \mathcal{N}(47.5, \sigma_{\alpha_2}^2)$, where σ_{α_1} and σ_{α_2} are determined by the user (here, we use $\sigma_{\alpha_1} = 0.025$ and $\sigma_{\alpha_2} = 3$). We then choose the comparison quantities to be the sets of model outputs and observed data, i.e. $\hat{z} \rightarrow \hat{Y}$ and $z \rightarrow Y$, and use the ϵ -Boolean agreement function (i.e. the improved reliability metric is included in our agreement definition). Regarding the likelihood function, we experiment with the different types of data uncertainty that were discussed earlier. In other words, we use the likelihoods derived in (17) and (18) corresponding to the different types of data distributions discussed above, i.e. bounded or truncated uniform data distributions, and completely certain observation points (we also consider normal or Gaussian data distributions with infinite tails). Finally, we run MCMC and find that BVM regression is capable of constructing posterior inferences for the model parameters for each type of data distributions (i.e. for each likelihood function that we use), unlike Bayesian model testing and standard regression techniques that fail at this task for truncated and completely certain data, as discussed before. We summarize the results in Table 3 below.

Data Distribution	BVM regression	Standard regression	Bayesian regression
Infinite Tail	✓	✓	✓
Truncated Tail	✓	✓	✗
Completely Certain	✓	✗	✗

Table 3: High probability of the model producing posterior parameter distributions and predictive envelopes for different types of data distributions using the three approaches. BVM regression is capable of producing posterior distributions of the model parameters for any type of data distributions.

Using the BVM regressed parameters' distributions, we can make predictions of y for new values of x , i.e., $p(\hat{y}|A, M, B)$ from (15). In addition, instead of just computing a point estimate of the fit, we also study the predictive posterior distribution of the model (also called the predictive envelope or confidence band) that results from the posterior distributions of the parameters. As an illustration of the predictive posterior distribution of our BVM regressed model, we plot the predictive envelopes of the nonlinear Monod model described in (19), along with the joint posterior distribution of the parameters α_1 and α_2 , while treating the data as completely certain and using the ϵ -Boolean function with a tolerance $\epsilon = 0.03$.

Figure 3: Predictive envelopes of the model and parameters' joint posterior distribution in the absence of data uncertainty using the BVM. As tabulated in Table 3, Bayesian regression fails to produce a candidate model solution as the data is completely certain and standard regression produces a single deterministic solution with no model uncertainty.

The blue curve shows the predicted response, which is the model fit calculated using the mean values of the parameters α_1 and α_2 in the chain. The blue shaded areas correspond to 68%, 95%, and 99.7% predictive posterior regions (by computing the model fit for a randomly selected subset of the chain). In other words, the blue regions span 1, 2, and 3 standard deviations on either side of the mean response, respectively. We will leave the interpretation of the predictive envelopes for our compound Boolean agreement function example in Section 4.2.2.

By examining the model parameters' joint posterior distribution in Figure 3, one can suspect that the value of the tolerance ϵ chosen affects the shape of the model parameters' distributions and thus the predictive envelope (or confidence regions). A smaller tolerance implies stricter agreement conditions between the model outputs and the observed data, which results in less uncertainty in the predictive posterior distributions of the model parameters and a narrower envelope. On the other hand, a larger tolerance implies more flexible agreement conditions, and results in more uncertainty in the predictive distributions, a wider envelope and a less predictive power. Thus, increasing ϵ can always result in finding a model given the data. To avoid getting very wide envelopes relative to the spread of the data, we start with a very small ϵ when running the MCMC simulation. We then keep increasing ϵ until the MCMC algorithm starts achieving a reasonably small acceptance rate for the new candidates in the chain.

Since this model has just two adaptive parameters, namely α_1 and α_2 , we can plot the prior and posterior distributions directly in parameter space. We explore the dependence between the parameters' posterior distributions and the value of

the tolerance ϵ . Figure 4 shows the results of BVM learning for the Monod model in (19) as the value of ϵ is decreases. For comparison, the optimal parameter values $\alpha_1 = 0.14542$ and $\alpha_2 = 49.053$ computed using standard regression are shown by a red star in the first row of Figure 4.

Figure 4: Illustration of BVM Learning for the Monod model for decreasing values of ϵ . In the first row is the prior/posterior parameter distribution in (α_1, α_2) space. The data points are shown by a black circle in the second row. The first column corresponds to the situation before any data point is observed and shows a plot of the prior distribution in (α_1, α_2) space together with six samples of the model response $M(x; \vec{\alpha})$ (blue lines) in which the values of α_1 and α_2 are randomly drawn from the prior. In the second, third and fourth columns, we see the situation after running BVM regression using MCMC, with a tolerance $\epsilon = 0.03$, $\epsilon = 0.025$ and $\epsilon = 0.02$, respectively. The posterior has now been influenced by the agreement tolerance ϵ , this gives a relatively compact posterior distribution. Samples from this posterior distribution lead to the functions shown in blue in the second row.

As Figure 4 shows, the smaller the tolerance is, the narrower and sharper the posterior joint distribution of the parameters is, the closer the blue lines get to each other, and the lower the uncertainty is. This explains the shape of the predictive envelopes as was discussed before. Thus, by varying ϵ , one can tune the posterior distribution of the model response to be more or less representative of the data. We will elaborate more on this in Section 4.3. It is worth noting that for a very small value of ϵ , the posterior joint distribution of the parameters converges toward the least squares solution, highlighted by a red star in figure 4. Also note that in our example, when ϵ goes below about 0.017, no solution seems to be possible and hence the probability of finding a model given the observed data becomes zero (i.e. our method starts to behave like Bayesian regression). In this case, the analyst may choose to work with any tolerance beyond this threshold, depending on his specifications and agreement requirements.

Once we generate the posterior distributions of the model parameters and the predictive envelopes, we can measure and estimate the reliability and accuracy of our computational model using a validation metric (including the BVM). By doing so, we determine how accurate is our model representation of the real world.

4.2.2 Numerical Example 2: Toy Model

After showing how the BVM can be used to perform regression on any type of data distribution to generate posterior model parameters' distributions and predictive envelopes, we now focus on how the user can choose the Boolean function to define the model-data agreement.

In this example, we will use the compound Boolean function as presented in [9].⁶ In that case, the definition of agreement requires the model to pass an average square error threshold of $\langle \epsilon \rangle$ as well as a check for probabilistic model configuration. The latter states that $95\% \pm 4\%$ of the uncertain (data) observations should lie inside the model's $1 - \hat{\alpha} = 95\%$ confidence interval. Note that we impose the $\pm 4\%$ tolerance to prevent the scenario where all 100% of the data points lie within an overly wide confidence interval, being marked as “agreeing”. We denote this compound Boolean function by $B(\hat{Y}, Y, \langle \epsilon \rangle, \hat{\alpha})$ and it is equal to,

$$B\left(\frac{1}{n} \sum_i |\hat{y}_i - y_i| \leq \langle \epsilon \rangle\right) \wedge B\left(0.91 \leq \frac{1}{n} \sum_i \Theta(y_i \in [-c_{\hat{\alpha}}, c_{\hat{\alpha}}]_i) \leq 0.99\right), \quad (20)$$

where n is the number of data points in the set Y , and $[-c_{\hat{\alpha}}, c_{\hat{\alpha}}]_i$ is the model's 95% confidence interval at instance x_i (this is the $(\langle \epsilon \rangle, \hat{\alpha})$ -Boolean in Table 1). Note that, although this compound Boolean seems to be complex, it is relatively easy to code and implement.

The BVM probability of agreement in this case can be expressed as,

$$\mathcal{Z}(B) = p(A|M, D, B, \langle \epsilon \rangle, \hat{\alpha}) = \int_{\vec{\alpha}} \underbrace{\left(\int_Y \Theta\left(B(M(X; \vec{\alpha}), Y, \langle \epsilon \rangle, \hat{\alpha})\right) \cdot \rho(Y|D) dY \right)}_{\mathcal{L}(\vec{\alpha}, B)} \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}$$

Note that the likelihood $\mathcal{L}(\vec{\alpha}, B)$ can be expressed as an expectation value over $\rho(Y|D)$,

$$\mathcal{L}(\vec{\alpha}, B) = E\left[\Theta\left(B(M(X; \vec{\alpha}), Y, \langle \epsilon \rangle, \hat{\alpha})\right)\right] \sim \frac{1}{K} \sum_{k=1}^K \Theta\left(B(M(X; \vec{\alpha}), Y^{(k)}, \langle \epsilon \rangle, \hat{\alpha})\right), \quad (21)$$

where $Y^{(k)}$ denotes the k^{th} set of data points drawn randomly from the probability distribution of Y . This allows us to approximate the integral using a statistical method like Monte-Carlo (MC). In this example, we use MC with $K = 50$.

We implement the compound Boolean, $B(\hat{Y}, Y, \langle \epsilon \rangle, \hat{\alpha})$, and show its ability to combine and quantify the average error as well as the probabilistic model representation of the uncertain data observations. We generated data using,

$$y(x) = 1 + xe^{-\cos(10x)} + \sin(10x) + \epsilon_a(x),$$

where $\epsilon_a(x) \sim \mathcal{N}(0, 0.4^2)$ for $x \in [0, 1.5]$ and $\epsilon_a(x) \sim \mathcal{N}(0, 0.6^2)$ for $x \in [1.5, 3]$, which represents the aleatoric stochastic uncertainty due to the system's randomness. We also assume the presence of epistemic measurement uncertainty in the data [43] with an additional normal distribution $\mathcal{N}(0, 0.5^2)$ about each data point.

To solve this example, we consider the following deterministic non-linear model,

$$\hat{y} = M(x; \vec{\alpha}) = \alpha_1 + \alpha_2 xe^{-\alpha_3 \cos(\alpha_4 x)} + \alpha_5 \sin(\alpha_6 x), \quad (22)$$

where $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)$ is the vector of model parameters having normally distributed prior distributions with means $\mu_{\vec{\alpha}} = (0, 0, 0, 9, 0, 9)$ and standard deviations $\sigma_{\vec{\alpha}} = (1, 1, 1, 0.5, 1, 0.5)$. We then conduct two experimental simulations while treating the data as normally distributed (i.e. having infinite tails). We first run the MCMC algorithm for 5000 iterations with 10% burn-in using Bayesian regression and plot the results in Figure 5a. We then repeat the simulation by performing BVM regression instead (i.e. by following the procedure outlined in Section 4.1), and using the approximate likelihood $\mathcal{L}(\vec{\alpha}, B)$ from (21) with a threshold of $\langle \epsilon \rangle = 0.7$. The results are shown in Figure 5b.

⁶It should be noted that in [9], a model is simply validated according to this compound metric – here we calibrate the model with respect to it instead.

Figure 5: Comparison between Bayesian regression and BVM regression. (a) Bayesian regression under infinite tail (Gaussian) data distribution. Note that the $\pm 3\sigma$ confidence region is very narrow and standard regression method produces a nearly identical result. (b) BVM regression using the compound Boolean. In this case, the $\pm 3\sigma$ confidence region is much wider and represents the data more accurately. Note that this probabilistic model passes both agreement conditions imposed by the compound Boolean $B(\hat{Y}, Y, \langle \epsilon \rangle, \hat{\alpha})$. Starting with a very small $\langle \epsilon \rangle$ in the MCMC simulation, we tune $\langle \epsilon \rangle$ by gradually increasing its value until both elements of the compound Boolean are naturally satisfied.

The BVM regression framework offers new insights into the interpretation of the predictive envelopes of Bayesian and standard regression. It is clear in Figure 5a that the Bayesian and standard regression methods generate predictive envelopes that would not accurately predict new target points. Surprisingly, these envelopes actually quantify the uncertainties in the *least square error solution* due to the presence of data uncertainty rather than a measure of predictive uncertainty. By choosing an appropriate model-data agreement function (as in Figure 5b), predictive uncertainty estimation became possible, i.e. we were able to construct predictive envelopes that satisfy our desire in representing new target points probabilistically. In other words, using the BVM regression framework gives the user more control over the predictive envelopes and what uncertainties they represent.

4.2.3 Numerical Example 3: Energy Dissipation Model

In this example, we consider the calibration of a three-parameter Smallwood model [11, 12, 55, 56, 57] in order to predict the energy dissipation D_E per cycle at a lap joint when subjected to an impact harmonic force of amplitude F . The energy loss in the joint under one cycle of sinusoidal loading is found by integrating the area under the hysteresis curve (force vs. displacement graph) and analytically derived as

$$D_E = k_n \left(\frac{m-1}{m+1} \right) \Delta z^{m+1} \quad (23)$$

where k_n is a nonlinear stiffness, m is a nonlinear exponent, and Δz is the displacement amplitude obtained by solving

$$2F = k\Delta z - k_n\Delta z^m \quad (24)$$

where k is a linear stiffness term. The aim is to calibrate the three parameters k_n , m and k using the available input-output data, where the input corresponds to the force F and the output corresponds to the measured dissipated energy D_E . Five levels of loading (and hence five measurements) were considered in the experiment and are shown in Table 4 below.

Force F (lbf)	Energy D_E (lbf $\times$ in)
60	5.30×10^{-5}
120	2.85×10^{-4}
180	7.78×10^{-4}
240	1.55×10^{-3}
320	2.50×10^{-3}

Table 4: Smallwood model: Calibration data.

We now follow the calibration procedure outlined in Section 4.1. The calibration data is available in Table 4 and we have $\{X, Y\} = \{F, D_E\}$. The model M can be expressed as $\hat{y} = D_E = M(F; k_n, m, k)$ which is obtained by combining Equations (23) and (24) (by eliminating Δz). The prior distributions of the parameters (m, k_n, k) are assumed to be Gaussian with the following mean and standard deviation.

Parameter	m	$\log_{10}(k_n)$	k
Mean	1.20	5.61	1172700
Standard Deviation	0.09	0.40	13760

Table 5: Prior mean and standard deviation of model parameters.

Note that m has no unit, k and k_n have units of lbf/in . We choose the comparison quantities to be the sets of model outputs and observed data, i.e. $\hat{z} \rightarrow \hat{Y}$ and $z \rightarrow Y$, and consider the ϵ -Boolean agreement function (i.e. the improved reliability metric is included in our agreement definition) with a tolerance $\epsilon = 10^{-3}$. We assume the data to be completely certain (deterministic) and thus we use the likelihood function derived in (18). We run MCMC for 10000 iterations with 20% burn-in and we find that our method is able to construct the posterior inferences of the model parameters. The prior versus posterior distributions and the pairwise relationships between model parameters (using 3000 samples) are shown in Figure 6.

Figure 6: Posterior inferences of model parameters.

From the posterior uncertainties of model parameters, we show the propagation of these uncertainties to the model response. In other words, we plot the predictive response and envelopes of the energy dissipation model, as shown in Figure 7.

Figure 7: Model response and predictive envelopes.

Notice that the model (mean) response fits well the observed data, and the 68% confidence band (i.e. $\pm 1\sigma$ confidence region) captures all of the data points. It is worth noting that in [55], the objective was to calibrate the nonlinear stiffness k_n , while treating the linear stiffness k and the exponent m as inputs (along with the force F). However, in our calibration process, we learned all three parameters (m , k_n , k) assuming F to be the only input. From our previous discussions, we know that the parameters' posterior inferences as well as the model's predictive envelopes are affected by the tolerance level ϵ . Therefore, the analyst can always tune ϵ to produce predictive envelopes that better represent the observed data, resulting in a more reliable model. Since in this example we are using the ϵ -Boolean agreement function, we know that the improved reliability metric [12] is automatically taken into account in the calibration process.

To validate our model and test its reliability, we simply compute the BVM probability of agreement, $\mathcal{Z}(B)$ in Equation (16), using the available data and the posterior distributions of the model parameters. It is worth noting that in this case $\mathcal{Z}(B)$ is a measure of the model reliability. Using the posterior inferences of the model parameters, we calculate $\mathcal{Z}(B)$ to be 0.93. Therefore, given that this value is a measure of probability of agreement between the model outputs and the observed data, we can conclude that our model is acceptable according to the agreement function and the assumptions under consideration. Also note that our method is suitable for model selection; one can calibrate two models and then compute the BVM ratio (in Equation 9 as described in Section 2.4) to select the more reliable model.

Finally, remember that the reliability metric can always be a part of the agreement function (by combining it with the Boolean function that the analyst decides to work with, resulting in a compound Boolean as in the previous example). Also note that the BVM framework can be used in both the calibration stage as well as the validation stage, as we have shown above (which means that the reliability of the model can be taken into account in both stages).

4.3 Discussion

As we have shown in the numerical examples above, the results are sensitive to the tolerance ϵ . This arises from the fact that ϵ represents the modeler's tolerance on the difference between the model outputs and the observed data. Thus, any other parameter that is included in the Boolean function B defined by the modeler will have an effect on the parameters' posterior distributions and the predictive envelopes. As shown in the previous examples, we can take advantage of this feature for modeling.

We analyse the effect of the tolerance ϵ as follows. Changing the agreement tolerance ϵ affects the acceptance rate in the MCMC iterations, i.e. a larger tolerance yields a larger variance of accepted "candidate" samples. The increased variance in the accepted samples produces wider posterior model parameter distributions. A smaller tolerance implies

the converse.

Parameters not regressed inside the Boolean function B (ϵ in this case) play a role similar to hyperparameters in Machine Learning. By tuning the hyperparameters, the modeler can make the predictive envelopes more representative of the data, and improve the overall performance of the model when compared to a randomly selected test (or validation) data set.

Note that in the case of the ϵ -Boolean, the user can increase ϵ indefinitely and still “regress” the model, although the predictive envelopes will be much wider than the data spread and hence less representative of the data. While widening the predictive envelopes can be useful for reliability and safety, if they are widened too much, the model can lose some of its predictive utility. To balance the trade-off between safety and utility, we regressed the model in (22) with respect to the compound Boolean (20) in Section 4.2.2. This Boolean function forced the model to be regressed in such a way that $95\% \pm 4\%$ of the (uncertain) data points lie within the model’s 95% confidence region while simultaneously satisfying the $\langle \epsilon \rangle$ requirement (for diversity we used $\langle \epsilon \rangle$ instead of ϵ although nothing prevents us from doing so in principle). As stated in the caption of Figure 5, we gradually tuned the value of $\langle \epsilon \rangle$ (hyperparameter tuning) to simultaneously satisfy both requirements imposed by the compound Boolean function while getting an accurate predictive uncertainty estimation.

5 Conclusion

In the calibration stage of model development, we can use the BVM to perform regression and model learning on data with any type of uncertainty, generate posterior parameter distributions, and model predictive envelopes, according to user-specified definitions of model-data agreement. The BVM regression framework proved its potential in offering new insights into the interpretation of the predictive envelopes of the Bayesian regression, standard regression, and likelihood-based techniques, and hence providing the analyst with more freedom and control over the predictive envelopes and their meaning. By calibrating models with respect to the validation metrics one desires a model to ultimately pass, reliability and safety metrics were integrated into our examples and automatically adopted by the model in the calibration phase. Finally, we find BVM regression to be a generalized regression and model learning tool allowing us to address several potential shortcomings in Bayesian and standard regression methods.

It is worth noting that all the work presented in this paper relies solely on parametric models. Thus, our method is limited to the cases and scenarios where an explicit expression of the model under question in terms of the parameters is known. Our future work involves expanding the BVM framework to nonparametric modeling (i.e. without making any assumptions about the model’s expression). Finally, we emphasize the importance of practicing statistical responsibility by being explicit in the definition of agreement (between the model output and the data), the assumptions made, and the criteria considered when performing model calibration.

Acknowledgments This work was supported by the Center for Complex Systems (CCS) at King Abdulaziz City for Science and Technology (KACST) and the Massachusetts Institute of Technology (MIT). We would like to thank all the researchers in the CCS.

References

- [1] Sohag Kabir and Yiannis Papadopoulos. Applications of bayesian networks and petri nets in safety, reliability, and risk assessments: A review. *Safety science*, 115:154–175, 2019.
- [2] Enrico Zio. Challenges in the vulnerability and risk analysis of critical infrastructures. *Reliability Engineering & System Safety*, 152:137–150, 2016.
- [3] Enrico Zio. The future of risk assessment. *Reliability Engineering & System Safety*, 177:176–190, 2018.
- [4] Enrico Zio. *An introduction to the basics of reliability and risk analysis*, volume 13. World scientific, 2007.
- [5] Jussi-Pekka Penttinen, Arto Niemi, Johannes Gutleber, Kari T Koskinen, Eric Coatanea, and Jouko Laitinen. An open modelling approach for availability and reliability of systems. *Reliability Engineering & System Safety*, 183:387–399, 2019.
- [6] Eric VanDerHorn and Sankaran Mahadevan. Bayesian model updating with summarized statistical and reliability data. *Reliability Engineering & System Safety*, 172:12–24, 2018.

- [7] Shankar Sankararaman and Sankaran Mahadevan. Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data. *Reliability Engineering & System Safety*, 96(7):814–824, 2011.
- [8] Saideep Nannapaneni and Sankaran Mahadevan. Reliability analysis under epistemic uncertainty. *Reliability Engineering & System Safety*, 155:9–20, 2016.
- [9] Kevin Vanslette, Tony Tohme, and Kamal Youcef-Toumi. A general model validation and testing tool. *Reliability Engineering & System Safety*, 195:106684, 2020.
- [10] Boudewijn R Haverkort and Adrianus Marinus Hendrikus Meeuwissen. Sensitivity and uncertainty analysis of markov-reward models. *IEEE Transactions on Reliability*, 44(1):147–154, 1995.
- [11] Ramesh Rebba and Sankaran Mahadevan. Computational methods for model reliability assessment. *Reliability Engineering & System Safety*, 93(8):1197 – 1207, 2008.
- [12] Shankar Sankararaman and Sankaran Mahadevan. Assessing the reliability of computational models under uncertainty. In *54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 1873, 2013.
- [13] P-R Wagner, Reto Fahrni, Michael Klippel, Andrea Frangi, and Bruno Sudret. Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels. *Engineering Structures*, 205:110063, 2020.
- [14] Adrian Chong, Weili Xu, Song Chao, and Ngoc-Tri Ngo. Continuous-time bayesian calibration of energy models using bim and energy data. *Energy and Buildings*, 194:177–190, 2019.
- [15] L Raillon and Christian Ghiaus. An efficient bayesian experimental calibration of dynamic thermal models. *Energy*, 152:818–833, 2018.
- [16] Liang Yin, Marcel AJ Smith, and Kishor S Trivedi. Uncertainty analysis in reliability modeling. In *Annual Reliability and Maintainability Symposium. 2001 Proceedings. International Symposium on Product Quality and Integrity (Cat. No. 01CH37179)*, pages 229–234. IEEE, 2001.
- [17] Joshua Mullins, You Ling, Sankaran Mahadevan, Lin Sun, and Alejandro Strachan. Separation of aleatory and epistemic uncertainty in probabilistic model validation. *Reliability Engineering & System Safety*, 147:49 – 59, 2016.
- [18] T.G. Trucano, L.P. Swiler, T. Igusa, W.L. Oberkampf, and M. Pilch. Calibration, validation, and sensitivity analysis: What’s what. *Reliability Engineering & System Safety*, 91(10):1331 – 1357, 2006. The Fourth International Conference on Sensitivity Analysis of Model Output (SAMO 2004).
- [19] Guesuk Lee, Wongon Kim, Hyunseok Oh, Byeng D Youn, and Nam H Kim. Review of statistical model calibration and validation—from the perspective of uncertainty structures. *Structural and Multidisciplinary Optimization*, pages 1–26, 2019.
- [20] Danqing Wu, Zhenzhou Lu, Yanping Wang, and Lei Cheng. Model validation and calibration based on component functions of model output. *Reliability Engineering & System Safety*, 140:59 – 70, 2015.
- [21] Shankar Sankararaman and Sankaran Mahadevan. Integration of model verification, validation, and calibration for uncertainty quantification in engineering systems. *Reliability Engineering & System Safety*, 138:194 – 209, 2015.
- [22] CJ Wild and GAF Seber. *Nonlinear regression*. New York: Wiley, 1989.
- [23] Anthony William Fairbank Edwards. *Likelihood*. CUP Archive, 1984.
- [24] Yudi Pawitan. *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press, 2001.
- [25] Thomas Leonard and John SJ Hsu. *Bayesian methods: an analysis for statisticians and interdisciplinary researchers*, volume 5. Cambridge University Press, 2001.
- [26] Peter M Lee. *Bayesian statistics*. Arnold Publication, 1997.
- [27] Alberto Malinverno and Victoria A Briggs. Expanded uncertainty quantification in inverse problems: Hierarchical bayes and empirical bayes. *Geophysics*, 69(4):1005–1016, 2004.
- [28] Inseok Park, Hemanth K. Amarchinta, and Ramana V. Grandhi. A Bayesian approach for quantification of model uncertainty. *Reliability Engineering and System Safety*, 95(7):777–785, 2010.
- [29] Chong Wang and Hermann G. Matthies. Novel model calibration method via non-probabilistic interval characterization and bayesian theory. *Reliability Engineering & System Safety*, 183:84 – 92, 2019.
- [30] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

- [31] David J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA, 2002.
- [32] D. S. Sivia and J. Skilling. *Data Analysis - A Bayesian Tutorial*. Oxford Science Publications. Oxford University Press, 2nd edition, 2006.
- [33] Scott M. Lynch. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer-Verlag, New York, 2007.
- [34] Kevin H. Knuth, Michael Habeck, Nabin K. Malakar, Asim M. Mubeen, and Ben Placek. Bayesian evidence and model selection. *Digital Signal Processing*, 47:50–67, Dec 2015.
- [35] Ben Placek, Kevin H. Knuth, and Daniel Angerhausen. EXONEST: BAYESIAN MODEL SELECTION APPLIED TO THE DETECTION AND CHARACTERIZATION OF EXOPLANETS VIA PHOTOMETRIC VARIATIONS. *The Astrophysical Journal*, 795(2):112, oct 2014.
- [36] Ben Placek. *Bayesian detection and characterization of extra-solar planets via photometric variations*. PhD thesis, State University of New York at Albany, 2014.
- [37] Paul Troughton and Simon J. Godsill. Bayesian model selection for time series using markov chain monte carlo. In *ICASSP*, pages 3733–3736, 1997.
- [38] F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to markov chain monte carlo methods for astronomical data analyses. *Monthly Notices of the Royal Astronomical Society*, 384(2):449–463, Jan 2008.
- [39] F. Feroz, M. P. Hobson, and M. Bridges. Multinest: an efficient and robust bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4):1601–1614, Oct 2009.
- [40] Nathaniel Stevens. *Assessment and comparison of continuous measurement systems*. PhD thesis, University of Waterloo, 2014.
- [41] William L. Oberkampff and Matthew F. Barone. Measures of agreement between computation and experiment: Validation metrics. *Journal of Computational Physics*, 217(1):5 – 36, 2006. Uncertainty Quantification in Simulation Science.
- [42] Ruoxue Zhang and Sankaran Mahadevan. Bayesian methodology for reliability model acceptance. *Reliability Engineering & System Safety*, 80(1):95 – 103, 2003.
- [43] Shankar Sankararaman and Sankaran Mahadevan. Model validation under epistemic uncertainty. *Reliability Engineering & System Safety*, 96(9):1232 – 1241, 2011. Quantification of Margins and Uncertainties.
- [44] Mark A. Beaumont, Wenyang Zhang, and David J. Balding. Approximate bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.
- [45] Jean-Michel Marin, Pierre Pudlo, Christian P. Robert, and Robin J. Ryder. Approximate bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, Nov 2012.
- [46] Thomas G. Dietterich. Ensemble methods in machine learning. In *Multiple Classifier Systems*, pages 1–15, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.
- [47] Radu V. Craiu and Jeffrey S. Rosenthal. Bayesian computation via markov chain monte carlo. *Annual Review of Statistics and Its Application*, 1(1):179–201, 2014.
- [48] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
- [49] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
- [50] Radford M. Neal. An improved acceptance procedure for the hybrid monte carlo algorithm. *J. Comput. Phys.*, 111(1):194–203, March 1994.
- [51] Matthew D. Parno and Youssef M. Marzouk. Transport map accelerated markov chain monte carlo. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):645–682, Jan 2018.
- [52] Christian P. Robert, Víctor Elvira, Nick Tawn, and Changye Wu. Accelerating mcmc algorithms. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(5):e1435, 2018.
- [53] John Skilling. Nested sampling for general bayesian computation. *Bayesian Anal.*, 1(4):833–859, 12 2006.
- [54] Paul Mac Berthouex and Linfield C. Brown. *Statistics for Environmental Engineers*. Lewis Publishers, 2002.

- [55] Shankar Sankararaman. *Uncertainty quantification and integration in engineering systems*. Vanderbilt University, 2012.
- [56] David O Smallwood, Danny L Gregory, and Ronald G Coleman. A three parameter constitutive model for a joint which exhibits a power law relationship between energy loss and relative displacement. In *Proceedings of the 72nd Shock and Vibration Symposium*, pages 12–16, 2001.
- [57] Angel Urbina, Thomas L Paez, Dan L Gregory, and Brain R Resor. Probabilistic modeling of mechanical joints. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 2003, pages 464–473. Institute of Noise Control Engineering, 2003.

A Bayesian Model Testing

We derive Equations (4) – (6) mentioned in Section 2.3. In the Bayesian model testing framework, the model output and the observed data are defined to agree only if their values are exactly equal. Thus, Bayesian model testing is a special case of the BVM where the agreement kernel is equal to the kronecker delta function (exact agreement) with continuous indices, i.e. $\Theta(B(\hat{Y}, Y)) = \delta_{\hat{Y}, Y} = \prod_{i=1}^n \delta_{\hat{y}_i, y_i}$. Since Bayesian model testing deals with probability densities, we have the following expression for the probability density of agreement (7):

$$\begin{aligned} \rho(A|M, D) &= \frac{p(A|M, D)}{dA} = \frac{1}{dA} \int_{\hat{Y}, Y} \rho(\hat{Y}|M, D) \cdot \Theta(B(\hat{Y}, Y)) \cdot \rho(Y|D) d\hat{Y} dY \\ &= \int_{\hat{Y}, Y} \rho(\hat{Y}|M, D) \cdot \frac{\delta_{\hat{Y}, Y}}{dA} \cdot \rho(Y|D) d\hat{Y} dY. \end{aligned}$$

The kronecker delta $\delta_{\hat{Y}, Y}$ and the dirac delta $\delta(\hat{Y} - Y)$ functions are related as follows:

$$\frac{\delta_{\hat{Y}, Y}}{dA} = \delta(\hat{Y} - Y) = \prod_{i=1}^n \delta(\hat{y}_i - y_i).$$

Thus, the probability density of agreement becomes,

$$\begin{aligned} \rho(A|M, D) &= \int_{\hat{Y}, Y} \rho(\hat{Y}|M, D) \cdot \delta(\hat{Y} - Y) \cdot \rho(Y|D) d\hat{Y} dY \\ &= \int_{\hat{Y}, Y} \left(\int_{\vec{\alpha}} \rho(\hat{Y}|\vec{\alpha}, M) \rho(\vec{\alpha}|M) d\vec{\alpha} \right) \cdot \delta(\hat{Y} - Y) \cdot \rho(Y|D) d\hat{Y} dY \\ &= \int_{\vec{\alpha}} \left(\int_{\hat{Y}, Y} \rho(\hat{Y}|\vec{\alpha}, M) \cdot \delta(\hat{Y} - Y) \cdot \rho(Y|D) d\hat{Y} dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha} \\ &= \int_{\vec{\alpha}} \left(\int_{\hat{Y}, Y} \underbrace{\delta(\hat{Y} - M(X; \vec{\alpha}))}_{\hat{Y} = M(X; \vec{\alpha})} \cdot \delta(\hat{Y} - Y) \cdot \rho(Y|D) d\hat{Y} dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha} \\ &= \int_{\vec{\alpha}} \left(\int_Y \delta(M(X; \vec{\alpha}) - Y) \cdot \rho(Y|D) dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}. \end{aligned} \tag{25}$$

A.1 Normally Distributed Data

If we assume the data to be normally distributed, i.e. $Y \sim \mathcal{N}(D, \Delta)$, we get,

$$\rho(Y|D) = \frac{1}{\sqrt{(2\pi)^n |\Delta|}} e^{-\frac{1}{2}(Y - D)^T \Delta^{-1} (Y - D)},$$

where n is the dimension of the training data set, Δ is the covariance matrix, and D is the observed data values.

Therefore, using (25), we have,

$$\begin{aligned}
\rho(A|M, D) &= \int_{\vec{\alpha}} \left(\int_Y \delta(M(X; \vec{\alpha}) - Y) \cdot \frac{1}{\sqrt{(2\pi)^n |\Delta|}} e^{-\frac{1}{2}(Y - D)^T \Delta^{-1} (Y - D)} dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}, \\
&= \int_{\vec{\alpha}} \underbrace{\frac{1}{\sqrt{(2\pi)^n |\Delta|}} e^{-\frac{1}{2}(M(X; \vec{\alpha}) - D)^T \Delta^{-1} (M(X; \vec{\alpha}) - D)}}_{\mathcal{L}(\vec{\alpha})} \cdot \underbrace{\rho(\vec{\alpha}|M) d\vec{\alpha}}_{\pi(\vec{\alpha}) d\vec{\alpha}}.
\end{aligned}$$

Therefore, the likelihood function to be used in the MCMC algorithm is

$$\mathcal{L}(\vec{\alpha}) = \frac{1}{\sqrt{(2\pi)^n |\Delta|}} e^{-\frac{1}{2}(M(X; \vec{\alpha}) - D)^T \Delta^{-1} (M(X; \vec{\alpha}) - D)},$$

which is Equation (4) presented in Section 2.3.

A.2 Uniformly Distributed Data

We first note that

$$dY \equiv d^n Y \equiv \prod_{j=1}^n dy_j, \quad j = 1, \dots, n. \quad (\text{C.2})$$

Now, we assume the data to be uniformly distributed, i.e.

$$y_j \sim \mathcal{U}(a_j, b_j), \quad j = 1, \dots, n.$$

Then, the probability density $\rho(Y|D)$ becomes:

$$\rho(Y|D) = \prod_{j=1}^n \rho(y_j|D) = \prod_{j=1}^n \frac{\Theta(a_j \leq y_j \leq b_j)}{b_j - a_j}. \quad (\text{C.3})$$

Notice that we can generalize $\rho(y_j|D) = \prod_{j=1}^n \Theta(a_j \leq y_j \leq b_j) \rho(y_j|D)$ to any bounded probability density function (pdf). Therefore, using (25), we have,

$$\begin{aligned}
\rho(A|M, D) &= \int_{\vec{\alpha}} \left(\prod_{j=1}^n \int_{y_j} \delta(M(x_j; \vec{\alpha}) - y_j) \cdot \frac{\Theta(a_j \leq y_j \leq b_j)}{b_j - a_j} dy_j \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}, \\
&= \int_{\vec{\alpha}} \underbrace{\left(\prod_{j=1}^n \frac{\Theta(a_j \leq M(x_j; \vec{\alpha}) \leq b_j)}{b_j - a_j} \right)}_{\mathcal{L}(\vec{\alpha})} \cdot \underbrace{\rho(\vec{\alpha}|M) d\vec{\alpha}}_{\pi(\vec{\alpha}) d\vec{\alpha}}.
\end{aligned}$$

Therefore, the likelihood function to be used in the MCMC algorithm is

$$\mathcal{L}(\vec{\alpha}) = \prod_{j=1}^n \frac{\Theta(a_j \leq M(x_j; \vec{\alpha}) \leq b_j)}{b_j - a_j},$$

which is Equation (5) presented in Section 2.3.

A.3 Completely Certain Data

If we consider the data to be completely certain, deterministic, i.e. $Y = D$, then, the probability density $\rho(Y|D)$ becomes $\rho(Y|D) = \delta(Y - D)$, and thus, using (25), we have,

$$\begin{aligned}
\rho(A|M, D) &= \int_{\vec{\alpha}} \left(\int_Y \delta(M(X; \vec{\alpha}) - Y) \cdot \delta(Y - D) dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha} \\
&= \int_{\vec{\alpha}} \underbrace{\delta(M(X; \vec{\alpha}) - D)}_{\mathcal{L}(\vec{\alpha})} \cdot \underbrace{\rho(\vec{\alpha}|M) d\vec{\alpha}}_{\pi(\vec{\alpha}) d\vec{\alpha}}.
\end{aligned}$$

Therefore, the likelihood function to be used in the MCMC algorithm is

$$\mathcal{L}(\vec{\alpha}) = \delta(M(X; \vec{\alpha}) - D),$$

which is Equation (6) presented in Section 2.3.

B BVM Model Selection

We derive Equations (16) – (18) presented in Section 3. We show how we can apply Bayesian model selection on any data distribution using the BVM probability of agreement. Starting from the original definition of probability of agreement (7), we have,

$$\begin{aligned} p(A|M, D, B) &= \int_{\hat{Y}, Y} \rho(\hat{Y}|M, D) \cdot \Theta(B(\hat{Y}, Y)) \cdot \rho(Y|D) d\hat{Y} dY \\ &= \int_{\hat{Y}, Y} \left(\int_{\vec{\alpha}} \rho(\hat{Y}|\vec{\alpha}, M) \rho(\vec{\alpha}|M) d\vec{\alpha} \right) \cdot \Theta(B(\hat{Y}, Y)) \cdot \rho(Y|D) d\hat{Y} dY \\ &= \int_{\vec{\alpha}} \left(\int_{\hat{Y}, Y} \rho(\hat{Y}|\vec{\alpha}, M) \cdot \Theta(B(\hat{Y}, Y)) \cdot \rho(Y|D) d\hat{Y} dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha} \\ &= \int_{\vec{\alpha}} \left(\int_{\hat{Y}, Y} \delta(\hat{Y} - M(X; \vec{\alpha})) \cdot \Theta(B(\hat{Y}, Y)) \cdot \rho(Y|D) d\hat{Y} dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha} \\ &= \int_{\vec{\alpha}} \left(\int_Y \Theta(B(M(X; \vec{\alpha}), Y)) \cdot \rho(Y|D) dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}, \end{aligned}$$

which is Equation (16) derived in Section 3. From (C.2), we know that

$$dY \equiv d^n Y \equiv \prod_{j=1}^n dy_j, \quad j = 1, \dots, n.$$

The probability density $\rho(Y|D)$ can be expressed as:

$$\rho(Y|D) = \prod_{j=1}^n \rho(y_j|D).$$

We also note that the compound Boolean under question can be expressed as:

$$\Theta(B(M(X; \vec{\alpha}), Y)) = \prod_{j=1}^n \Theta(B(M(x_j; \vec{\alpha}), y_j)) \quad j = 1, \dots, n.$$

Thus, we rewrite the BVM probability of agreement as follows:

$$p(A|M, D, B) = \int_{\vec{\alpha}} \left(\prod_{j=1}^n \int_{y_j} \Theta(B(M(x_j; \vec{\alpha}), y_j)) \cdot \rho(y_j|D) dy_j \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}.$$

We will use the ϵ -Boolean indicator function defined as:

$$\Theta(B(M(x_j; \vec{\alpha}), y_j)) = \begin{cases} 1, & \text{if } |y_j - M(x_j; \vec{\alpha})| \leq \epsilon \\ 0, & \text{otherwise} \end{cases}$$

where $j = 1, \dots, n$.

Then, the indicator function can be rewritten as:

$$\begin{aligned} \Theta(B(M(x_j; \vec{\alpha}), y_j)) &= \Theta(|y_j - M(x_j; \vec{\alpha})| \leq \epsilon) \\ &= \Theta(M(x_j; \vec{\alpha}) - \epsilon \leq y_j \leq M(x_j; \vec{\alpha}) + \epsilon), \end{aligned}$$

where $j = 1, \dots, n$. Therefore, the BVM probability of agreement can be expressed as:

$$p(A|M, D, B) = \int_{\vec{\alpha}} \left(\prod_{j=1}^n \int_{y_j} \Theta(M(x_j; \vec{\alpha}) - \epsilon \leq y_j \leq M(x_j; \vec{\alpha}) + \epsilon) \cdot \rho(y_j|D) dy_j \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}. \quad (\text{D.1})$$

B.1 Normally Distributed Data

Note that the Boolean $\Theta\left(M(x_j; \vec{\alpha}) - \epsilon \leq y_j \leq M(x_j; \vec{\alpha}) + \epsilon\right)$ is equal to 1 only when y_j belongs to the interval $\left[M(x_j; \vec{\alpha}) - \epsilon, M(x_j; \vec{\alpha}) + \epsilon\right]$.

Thus, the BVM probability of agreement becomes:

$$p(A|M, D, B) = \int_{\vec{\alpha}} \left(\prod_{j=1}^n \int_{M(x_j; \vec{\alpha}) - \epsilon}^{M(x_j; \vec{\alpha}) + \epsilon} \rho(y_j|D) dy_j \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}.$$

Now, we assume that the data is normally distributed, i.e.

$$y_j \sim \mathcal{N}(D_j, \sigma_j^2), \quad j = 1, \dots, n.$$

Then, the probability density $\rho(Y|D)$ becomes:

$$\rho(Y|D) = \prod_{j=1}^n \rho(y_j|D) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{y_j - D_j}{\sigma_j}\right)^2}.$$

Thus, we rewrite the BVM probability of agreement as follows:

$$\begin{aligned} p(A|M, D, B) &= \int_{\vec{\alpha}} \left(\prod_{j=1}^n \int_{M(x_j; \vec{\alpha}) - \epsilon}^{M(x_j; \vec{\alpha}) + \epsilon} \rho(y_j|D) dy_j \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha} \\ &= \int_{\vec{\alpha}} \left(\prod_{j=1}^n \int_{M(x_j; \vec{\alpha}) - \epsilon}^{M(x_j; \vec{\alpha}) + \epsilon} \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{y_j - D_j}{\sigma_j}\right)^2} dy_j \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha} \\ &= \int_{\vec{\alpha}} \underbrace{\prod_{j=1}^n \left(F\left(M(x_j; \vec{\alpha}) + \epsilon\right) - F\left(M(x_j; \vec{\alpha}) - \epsilon\right) \right)}_{\mathcal{L}(\vec{\alpha}, B)} \cdot \underbrace{\rho(\vec{\alpha}|M) d\vec{\alpha}}_{\pi(\vec{\alpha})} d\vec{\alpha}, \end{aligned}$$

where $F(x)$ is the cumulative distribution function (cdf) expressed as:

$$F(x) = \Phi\left(\frac{x - D}{\sigma}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, i.e. $\mathcal{N}(0, 1)$, and expressed as:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Therefore, the likelihood function to be used in the MCMC algorithm is

$$\mathcal{L}(\vec{\alpha}, B) = \prod_{j=1}^n \left(F\left(M(x_j; \vec{\alpha}) + \epsilon\right) - F\left(M(x_j; \vec{\alpha}) - \epsilon\right) \right).$$

B.2 Uniformly Distributed Data

If we assume the data to be uniformly distributed (C.3), then the BVM probability of agreement (D.1) becomes:

$$p(A|M, D, B) = \int_{\vec{\alpha}} \left(\prod_{j=1}^n \int_{y_j} \Theta\left(M(x_j; \vec{\alpha}) - \epsilon \leq y_j \leq M(x_j; \vec{\alpha}) + \epsilon\right) \cdot \frac{\Theta(a_j \leq y_j \leq b_j)}{b_j - a_j} dy_j \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha}.$$

Note that the product $\Theta\left(M(x_j; \vec{\alpha}) - \epsilon \leq y_j \leq M(x_j; \vec{\alpha}) + \epsilon\right) \cdot \Theta\left(a_j \leq y_j \leq b_j\right)$ is equal to 1 only when y_j belongs to both intervals $\left[M(x_j; \vec{\alpha}) - \epsilon, M(x_j; \vec{\alpha}) + \epsilon\right]$ and $\left[a_j, b_j\right]$.

Let l_j and u_j be such that

$$\left[l_j, u_j \right] = \left[M(x_j; \vec{\alpha}) - \epsilon, M(x_j; \vec{\alpha}) + \epsilon \right] \cap \left[a_j, b_j \right] \quad j = 1, \dots, n$$

Thus, the BVM probability of agreement becomes:

$$\begin{aligned} p(A|M, D, B) &= \int_{\vec{\alpha}} \left(\prod_{j=1}^n \int_{y_j} \frac{\Theta(l_j \leq y_j \leq u_j)}{b_j - a_j} dy_j \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha} \\ &= \int_{\vec{\alpha}} \left(\prod_{j=1}^n \int_{l_j}^{u_j} \frac{1}{b_j - a_j} dy_j \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha} \\ &= \int_{\vec{\alpha}} \underbrace{\left(\prod_{j=1}^n \frac{u_j - l_j}{b_j - a_j} \right)}_{\mathcal{L}(\vec{\alpha}, B)} \cdot \underbrace{\rho(\vec{\alpha}|M) d\vec{\alpha}}_{\pi(\vec{\alpha}) d\vec{\alpha}}. \end{aligned}$$

Therefore, the likelihood function to be used in the MCMC algorithm is

$$\mathcal{L}(\vec{\alpha}, B) = \prod_{j=1}^n \frac{u_j - l_j}{b_j - a_j},$$

which is Equation (17) presented in Section 3.

B.3 Completely Certain Data

If we consider the data to be completely certain, deterministic, i.e. $Y = D$, then, the probability density $\rho(Y|D)$ becomes $\rho(Y|D) = \delta(Y - D)$, and thus, using (16), we have,

$$\begin{aligned} p(A|M, D, B) &= \int_{\vec{\alpha}} \left(\int_Y \Theta\left(B(M(X; \vec{\alpha}), Y)\right) \cdot \delta(Y - D) dY \right) \cdot \rho(\vec{\alpha}|M) d\vec{\alpha} \\ &= \int_{\vec{\alpha}} \underbrace{\Theta\left(B(M(X; \vec{\alpha}), D)\right)}_{\mathcal{L}(\vec{\alpha}, B)} \cdot \underbrace{\rho(\vec{\alpha}|M) d\vec{\alpha}}_{\pi(\vec{\alpha}) d\vec{\alpha}}. \end{aligned}$$

Therefore, the likelihood function to be used in the MCMC algorithm is

$$\mathcal{L}(\vec{\alpha}, B) = \Theta\left(B(M(X; \vec{\alpha}), D)\right),$$

which is Equation (18) presented in Section 3.