

Robust system identification and model predictions in the presence of systematic uncertainty

Romain Pasquier^{a,*}, Ian F. C. Smith^a

^a*Applied Computing and Mechanics Laboratory (IMAC), School of Architecture, Civil and Environmental Engineering (ENAC), Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland*

Abstract

Context: Model-based data-interpretation techniques are increasingly used to improve the knowledge of complex system behavior. Physics-based models that are identified using measurement data are generally used for extrapolation to predict system behavior under other actions. In order to obtain accurate and reliable extrapolations, model-parameter identification needs to be robust in terms of variations of systematic modeling uncertainty introduced when modeling complex systems. Approaches such as Bayesian inference are widely used for system identification. More recently, error-domain model falsification (EDMF) has been shown to be useful for situations where little information is available to define the probability density function (PDF) of modeling errors. Model falsification is a discrete population methodology that is particularly suited to knowledge intensive tasks in open worlds, where uncertainty cannot be precisely defined.

Objective: This paper compares conventional uses of approaches such as Bayesian inference and EDMF in terms of parameter-identification robustness and extrapolation accuracy.

Method: Using Bayesian inference, three scenarios of conventional assumptions related to inclusion of modeling errors are evaluated for several model classes of a simple beam. These scenarios are compared with results obtained using EDMF. Bayesian model class selection is used to study the benefit of posterior model averaging on the accuracy of extrapolations. Finally, ease of representation and modification of knowledge is illustrated using an example of a full-scale bridge.

Results: This study shows that EDMF leads to robust identification and more accurate predictions than conventional applications of Bayesian inference in the presence of systematic uncertainty. These results are illustrated with a full-scale bridge. This example shows that the engineering knowledge necessary to perform parameter identification and remaining-fatigue-life predictions of a complex civil structure is easily represented by the EDMF methodology.

Conclusion: Model classes describing complex systems should include two components: (1) unknown physical parameters that are identified using measurements; (2) conservative modeling error estimations

*Corresponding author. Address: EPFL ENAC IIC IMAC, Station 18, CH-1015 Lausanne, Switzerland. Phone number: +41 21 693 2498

Email address: rpasquie@gmail.com (Romain Pasquier)

that cannot be represented only as uncertainties related to physical parameters. In order to obtain accurate predictions, both components need to be included in the model-class definition. This study indicates that Bayesian model class selection may lead to over-confidence in certain model classes, resulting in biased extrapolation.

Keywords: systematic error, extrapolation, modeling uncertainties, Bayesian inference, model falsification, model class selection.

1. Introduction

System identification involves taking advantage of measurement data to improve the understanding of system behavior. In order to achieve this task, physics-based models can be employed to help interpret measurement data. Such models are used to predict system behavior at unmeasured locations and for other actions. For example, vibration data from a bridge may be used to infer uncertain physical parameter values such as stiffness values that are then used to predict fatigue lives.

Parameter identification and predictions are sensitive to systematic modeling errors that are induced by idealizations of real systems. Systematic errors arise due to simplifications and omissions in the modeling process and usually reflect spatially interdependency between measurement locations. This type of error is called model inadequacy in [26], model bias in [2], model discrepancy in [10] and *modeling error* in [20]. The last designation is used in this paper.

In complex systems, data interpretation is ambiguous: multiple models are able to represent measured behavior. Techniques such as residual minimization, maximum likelihood estimates and maximum *a-posteriori* estimates should be avoided when systematic errors are present in the model, since they lead to the identification of a single optimal model that is intrinsically imperfect due to parameter-value compensation [2, 3, 20, 31].

Techniques such as probabilistic Bayesian inference are able to accommodate populations of solutions. Bayesian inference determines the full posterior distribution of the uncertain parameter values by the construction of a likelihood function describing the probability of observations given a set of parameter values. In this way, this approach identifies model parameter values that are compatible with the measurement data and all these values are then used to predict system behavior.

Underestimating modeling uncertainty (i.e. either mean value or variance) during data interpretation may lead to biased parameter identification and thus to inaccurate predictions. Moreover, the convergence of the parameter values to the true values may become even more biased as the number of measurements increases [10, 20]. Nevertheless, it is possible to identify biased parameter values and still obtain accurate predictions when predicting inside the domain of experimentation. This type of prediction is called interpolation [24]. However, learning the correct values of physical parameters is important for the understanding of the true

behavior of the system and also for improving confidence in model extrapolation [10, 16]. Extrapolation values are predictions out of the measurement context, such as fatigue life in the example of the beginning of this paper.

In Bayesian inference, the common assumption is that modeling and measurement uncertainties are adequately described by independent zero-mean Gaussian distributions [5, 17, 27]. Most applications integrate the prediction-error variance as a parameter during the identification process [1, 6, 11, 32, 44] and some assign an arbitrary value to the variance [4, 15, 18, 41]. These applications lead to correct parameter identification since the assumptions made for the probability density functions (PDF) of prediction errors are compatible with assumptions related to model-class fidelity to the real system. Also, in situations where systematic errors are absent, using the current Bayesian scheme for establishing the predictive distribution leads to correct interpolation and extrapolation [3–5, 32, 46, 47]. Behmanesh et al. [7] includes mean values, the variance and correlation values of modeling uncertainty as updating parameters. However, it is shown that this approach leads to biased identification in the presence of systematic errors. Except for [7], there are few applications of Bayesian inference involving systematic errors and few studies have evaluated the validity of such assumptions through comparisons with other approaches.

The complexity of a model class is often only defined by the type and the number of parameters that require identification. However, the complexity depends also on the level of detail that is achievable and thus, depends on the modeling errors. Bayesian model class selection can be used to identify an optimal model class among a set of model classes that returns the best trade-off between data fitting and model-class complexity [6, 14, 30, 44, 45]. When several model classes are plausible, all of them are used by weighing each model-class prediction according to their plausibility in order to obtain robust predictions. Bayesian model class selection was also used to identify the best correlation model [40]. Another application involved the selection of the best prediction-error variance model [18]. However, the best model class led to a biased posterior PDF because of the presence of systematic modeling errors that were not characterized in the model-class definition. In addition, there has been little discussion of situations involving Bayesian model class selection where every model class is biased among the set of possible model classes.

Goulet and Smith [20] proposed an approach that is suitable when little is known about modeling errors. This approach, called error-domain model falsification (EDMF), combines estimated PDFs of each source of modeling and measurement error and determines conservative probabilistic thresholds that are used to falsify inadequate models. Modeling errors are estimated using engineering judgment and field observations. They have shown that this approach leads to robust parameter identification in the presence of systematic errors without precise knowledge of the dependencies between modeling errors. Goulet and Smith [20] also demonstrated that the assumption of independence in the common definition of uncertainties in Bayesian inference may bias the posterior distribution of parameter values in the presence of systematic errors. This last observation has also been noted by Simoen et al. [40]. However, the effects of systematic modeling errors

on interpolations and extrapolations were not studied.

This paper builds on the work by Goulet and Smith [20] through comparing results for predictions. Robustness of parameter-value identification and accuracy of interpolations and extrapolations are studied for several model classes of a simple beam. Using Bayesian inference for data interpretation, three scenarios are evaluated: (1) modeling errors are not included in the data-interpretation process; (2) modeling errors are described by Gaussian PDFs; (3) the variance of the prediction-error uncertainties is parametrized and is part of the set of parameters that are identified using the Bayesian framework. These scenarios are compared with results obtained using EDMF. Finally, Bayesian model class selection is used to study the benefit of posterior model averaging on the accuracy of extrapolations and is compared with extrapolations obtained using EDMF.

Sections 2 and 3 present an overview of Bayesian inference, Bayesian model class selection and error-domain model falsification. Section 4 illustrates the comparison between these data-interpretation techniques by an example involving a simply supported beam.

2. Bayesian inference

Bayesian inference uses information obtained from measurement data to update prior knowledge of the system through the identification of parameter values. Let $\mathbf{y} = [y_1, \dots, y_{n_m}]^T$ be a vector of measurement data from a physical system where n_m is the number of measurements. Then, let G be a possible model class describing the system and $\mathbf{g}(\boldsymbol{\theta})$ a vector of model predictions where $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_{n_p}]^T$ is a vector of n_p parameters having uncertain values and defined on the parameter domain $\Theta \subseteq \mathbb{R}^{n_p}$. The inference of the parameter values of the model class G is based on Bayes' Theorem of conditional probability:

$$p(\boldsymbol{\theta}|\mathbf{y}, G) = \frac{p(\mathbf{y}|\boldsymbol{\theta}, G)p(\boldsymbol{\theta}|G)}{p(\mathbf{y}|G)} \quad (1)$$

where $p(\boldsymbol{\theta}|\mathbf{y}, G)$ is the posterior PDF given the measurement data \mathbf{y} and the model class G , $p(\boldsymbol{\theta}|G)$ is the user-defined prior PDF or prior knowledge of the uncertain parameter values, $p(\mathbf{y}|\boldsymbol{\theta}, G)$ is the likelihood function and the denominator $P(\mathbf{y}|G)$ is the evidence for the model class given by measurement data \mathbf{y} . This term is used as a normalizing constant in Eq. (1) and is also important for model class selection, which is presented in Section 2.1. The prior knowledge indicates the initial user's judgment of the plausibility of the uncertain parameter values before data are taken into account. The likelihood function expresses the probability of observing measurement data from the model class having a specific set of parameters. This gives a measure of data-fit of the model. This approach updates the prior knowledge of the uncertain model parameters by leveraging of the information gained by the measured values. This also creates a mapping between the error domain $\Xi \subseteq \mathbb{R}^{n_m}$ that refers to the residuals of the differences between measured and predicted values $\boldsymbol{\epsilon}_o = [\epsilon_{o,1}, \dots, \epsilon_{o,n_m}]^T$ and the parameter domain Θ [20]. The usual formulation of the

likelihood function is based on a Gaussian PDF:

$$p(\mathbf{y}|\boldsymbol{\theta}) \propto \text{const} \cdot \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{g}(\boldsymbol{\theta}))^T \Sigma^{-1} (\mathbf{y} - \mathbf{g}(\boldsymbol{\theta}))\right] \quad (2)$$

where $\Sigma \in \mathbb{R}^{n_m \times n_m}$ is a covariance matrix composed of the variances and the correlation coefficients for each measured location of the system.

In most applications of Bayesian inference, independent zero-mean Gaussian distributions are assumed to describe the uncertainties due to measurement and modeling errors. In addition to the independence assumption, the same variance is assumed for each model prediction leading to a diagonal covariance matrix with all non-zero terms being equal. In such examples, the uncertainty is only evaluated through the variance σ_ϵ^2 . The posterior distribution of uncertain parameters can then be used to predict future behavior of the system. The procedure aims at deriving the predictive distribution of a system output based on the updated model for various types of inputs, such as load configurations. Let \mathbf{q} be the system outputs to be predicted, the predictive distribution obtained from the updated model class G is given by Eq. (3).

$$p(\mathbf{q}|\mathbf{y}, G) = \int_{\Theta} p(\mathbf{q}|\boldsymbol{\theta}, G) p(\boldsymbol{\theta}|\mathbf{y}, G) d\boldsymbol{\theta} \quad (3)$$

This methodology is employed in [1, 3, 5, 44] and is used in many applications for updating structural reliability outputs, for example, [4, 32, 46].

2.1. Bayesian model class selection

Several model classes are often a priori plausible representations of the real system. In this case, Bayesian model class selection is employed for determining the most plausible model class among the set of model classes. This selection technique returns the best trade-off between the degree of model data fitting (fidelity to data) and model-class complexity. These competing aspects are included in the calculation of the evidence (Eq. 5). As a result, this approach penalizes more complex model classes in order to select model classes that are robust to small imperfections, which enforces the principle of model parsimony (Ockham's razor) [28, 45]. Bayesian model class selection is used to determine the most plausible model class among the set of model classes G_k where $k \in \{1, 2, \dots, n_c\}$. This approach rates the plausibility of each model class conditionally on the measurement data. This is achieved by applying Bayes' Theorem at the model-class level:

$$P(G_k|\mathbf{y}) = \frac{p(\mathbf{y}|G_k) P(G_k)}{p(\mathbf{y})} \quad (4)$$

where $P(G_k|\mathbf{y})$ is the posterior plausibility of the model class G_k and $P(G_k)$ the prior plausibility that is usually chosen to be equal for each model class. The term $p(\mathbf{y}|G_k)$ represents the evidence, which is the same term as the denominator of Eq. (1). This is the term that plays the main role in Bayesian model class selection since it expresses the likelihood of obtaining the measurement data by selecting model class G_k .

The denominator $p(\mathbf{y}) = \sum_{k=1}^{n_c} p(\mathbf{y}|G_k) P(G_k)$ is the normalization of the plausibilities and is obtained by the law of total probability.

The evidence is obtained also by the law of total probability at the model-parameter level by considering the likelihood and the prior probability of Eq. (1):

$$p(\mathbf{y}|G_k) = \int_{\Theta_k} p(\mathbf{y}|\boldsymbol{\theta}_k, G_k) p(\boldsymbol{\theta}_k|G_k) d\boldsymbol{\theta}_k \quad (5)$$

where $\boldsymbol{\theta}_k$ is the parameter vector for each model class G_k . The updated plausibility of Eq. (4) can be difficult to compute because of the two high dimensional integrals, the evidence $p(\mathbf{y}|G_k)$ and the normalization of the plausibilities $p(\mathbf{y})$ (denominator of Eq. 4). In such situations, Yuen [45] proposed Eq. (6) as an alternative calculation.

$$P(G_k|\mathbf{y}) = \frac{\exp(\ln p(\mathbf{y}|G_k) - M)}{\sum_{k=1}^{n_c} \exp(\ln p(\mathbf{y}|G_k) - M)} \quad (6)$$

In Eq. (6), $M = \max_k \ln p(\mathbf{y}|G_k)$ is the maximum log-evidence among each model class. This expression does not modify the relative plausibility of each model class, as far as all model classes have the same prior probability.

The predictive distribution of each model class can be weighted by their plausibility in order to account for the uncertainty at the model-class level for predictions:

$$p(\mathbf{q}|\mathbf{y}) = \sum_{k=1}^{n_c} p(\mathbf{q}|G_k, \mathbf{y}) P(G_k|\mathbf{y}) \quad (7)$$

This expression is called posterior model averaging. Bayesian model class selection has been widely used in many applications [6, 14, 30, 44] for identifying the optimal model class among possible model classes and for obtaining the predictive distribution by posterior model averaging [12, 13]. More related to modeling-error assumptions, Simoen et al. [40] proposed the use of Bayesian model class selection to determine an adequate correlation model of the prediction-error uncertainty. Goller et al. [18] proposed a methodology based on Bayesian model class selection to identify the most plausible model class from a set of model classes differing from the value of the uncertainty variance.

Solving the Bayesian framework may be difficult for high-dimensional parameter domains when using simple sampling methods such as uniform sampling or Monte Carlo simulation. This difficulty has motivated techniques such as Markov Chain Monte Carlo simulations for sampling posterior PDFs, predictive distributions and evidence in a more efficient manner. In this paper, MCMC is based on the Metropolis-Hastings algorithm. An overview of this method is presented in [28, 44].

3. Error-domain model falsification

Similarly to Bayesian inference, error-domain model falsification (EDMF) considers a set of model predictions $\mathbf{g}_k(\boldsymbol{\theta}_k)$ of a model class G_k having n_p unknown physical parameters $\boldsymbol{\theta}_k = [\theta_{k1}, \theta_{k2}, \dots, \theta_{kn_p}]^T$. The

vector $\boldsymbol{\theta}_k$ represents the parameter uncertainty. For complex systems, it is not feasible to parametrize every phenomenon. Therefore, this approach also includes modeling error $\epsilon_{\text{model},k}^*$ that is induced by inevitable assumptions, simplifications and omissions made during the modeling process. Examples of sources of modeling error are idealized support conditions, geometric variability of the structure, load amplitude and position, Bernoulli-beam hypothesis, constitutive law of materials, no dynamic magnification, etc. For finite-element (FE) models, there are also mesh refinement and interpolation, the element-type choices, the presence of singularities, etc.

By including both model parameter uncertainty and modeling error at each measurement location $i \in \{1, 2, \dots, n_m\}$, the true value of the system is approximated. This true value, \mathcal{Q}_i , is also approximated with measurements y_i and measurement error $\epsilon_{\text{measure},i}^*$. Eq. (8) summarizes these relations.

$$g_{ki}(\boldsymbol{\theta}_k^*) - \epsilon_{\text{model},ki}^* = \mathcal{Q}_i = y_i - \epsilon_{\text{measure},i}^* \quad (8)$$

The modeling error values $\epsilon_{\text{model},ki}^*$ that compensate model-prediction values in order to obtain the true values depend on the choice of the model class G_k . In real situations, neither modeling and measurement error values ϵ_{ki}^* , nor true values \mathcal{Q}_i are known with certainty.

These errors can only be evaluated using probability density functions and thus treated as random variables $U_{\text{model},ki}$ and $U_{\text{measure},i}$. The difference between modeling and measurement uncertainties determine the random variables $U_{c,ki}$ that describe the observed residuals $\epsilon_{c,ki}$ (i.e. the differences between predicted and measured values). The PDF $f_{U_{c,ki}}(\epsilon_{c,ki})$ represents the probability of the continuous random variables $U_{c,ki}$. An instance of the parameter set $\boldsymbol{\theta}_k$ and of a model class G_k , is falsified, if for any measurement location $i \in \{1, 2, \dots, n_m\}$, the difference between predicted and measured values lies outside the interval defined by the threshold bounds $[T_{\text{low},ki}, T_{\text{high},ki}]$ (Eq. 10) such that Eq. (9) is not satisfied:

$$\forall i \in \{1, \dots, n_m\} : T_{\text{low},ki} \leq g_{ki}(\boldsymbol{\theta}_k^*) - y_i \leq T_{\text{high},ki} \quad (9)$$

Inversely, a model instance is accepted if this difference is inside these bounds at each location. Since evaluating many values of $g_{ki}(\boldsymbol{\theta}_k^*)$ often requires multiple numerical simulations, EDMF is only feasible with modern computing technology. Advances in massively parallel computers particularly in the context of cloud computation significantly reduce computation time of tasks involving polynomial time complexity. Discrete population approaches such as EDMF show much promise for a range of abductive engineering tasks.

Threshold bounds define the shortest sets of intervals including a target probability $\phi_d \in [0, 1]$ that the right model is not falsely discarded. They are determined by satisfying Eq. (10).

$$\forall i \in \{1, \dots, n_m\} : \phi_d^{1/n_m} = \int_{T_{\text{low},ki}}^{T_{\text{high},ki}} f_{U_{c,ki}}(\epsilon_{c,ki}) d\epsilon_{c,ki} \quad (10)$$

This way of calculating these bounds uses the Šidák correction [39]. This leads to determination of conservative threshold bounds for each measurement location regardless of the values of correlations between uncertainties [25].

Practically, the user defines an initial set of model instances $G_k(\boldsymbol{\theta}_k)$ having many combinations of values for $\boldsymbol{\theta}_k$ that are obtained based on engineering heuristics. A target probability of identification ϕ_d is also defined, thereby explicitly setting the level of false-rejection error that is acceptable. After falsification of the model instances that are incompatible with the measured values given the modeling and measurement uncertainties, a subset of candidate models $\boldsymbol{\theta}_k^*$ remains. Since more precise probabilities cannot be assumed in full-scale cases, each candidate model is taken to be equally likely to be the right model. Thus, every candidate model is used for prediction of $\mathbf{q} = [q_{k1}, q_{k2}, \dots, q_{kn_r}]$ at n_r locations of the model. From Eq. (8), the expression for the distributions q_{kj} can be deduced for the identified model class G_k . Eq. (11) presents the combination of the random variables $U_{\boldsymbol{\theta}_k^*}$ describing the candidate models and $U_{\text{model},kj}$ describing the modeling error:

$$q_{kj} = g_{kj}(U_{\boldsymbol{\theta}_k^*}) - U_{\text{model},kj}, \quad j \in \{1, 2, \dots, n_r\} \quad (11)$$

where q_{kj} is a random variable representing the predicted value obtained for a random candidate-model instance and a random sample of modeling uncertainty. Thus, distributions of q_{kj} are obtained by the combination of distributions of candidate-model predictions with distributions of modeling error for each prediction location j . Lower and higher prediction thresholds are then calculated for each distribution q_{kj} using a similar procedure as in Eq. (10). The prediction thresholds $[q_{\text{low},kj}, q_{\text{high},kj}]$ represent the shortest sets of intervals including the target probability of prediction ϕ_p . Also, these bounds are determined using the Šidák correction such that for each location j , the prediction thresholds are defined based on a probability ϕ_p^{1/n_r} .

There are situations where all the initial model instances are falsified provided that a sufficient number of measurements is used as presented in [20]. This means that erroneous assumptions are made in the model-class definition. In such situations, the estimation of modeling and measurement uncertainties is incompatible with the model class and the complete falsification avoids making a wrong parameter-value identification. In addition, model-class falsification can help explore possible model classes of systems. When several model classes are possible, the model class is selected based on the performance of its intended use, for example, the precision of predictions.

Falsification is a well-known scientific concept that was popularized in 1930's by Karl Popper [37]. He argued that data can only be used to falsify models. Since then, authors such as Tarantola [42] and Beven [8] recognized the importance of this concept for system identification. Beven and Binley [9] proposed a methodology based on model falsification that is intended to overcome limitations of traditional approaches in the field of environmental sciences.

Examples of applications of EDMF where uncertainties are estimated based on engineering experience can be found in [22, 23, 34]. Candidate models have also been used to predict the remaining-fatigue-life of critical details in [35], wind flow around buildings [43] and leaks in water supply networks [21, 29]. Model falsification has also been applied to sensor configuration [19, 33, 38].

In most applications of system identification, it is commonly assumed that the PDF of modeling and measurement uncertainty is a Gaussian distribution that is centered on zero. This means that models that fits the measurements better (zero residual) are more likely than the others. This assumption is justified when the model class that is used describes with fidelity the real system behavior, when modeling errors are aleatory and when there are enough measurements to provide an approximation of the distribution. In such cases, the likelihood of a model depends on the value of the variance that is used to describe the uncertainty PDF. However, in the presence of systematic modeling errors, the predictions of the model having the correct parameter values are biased compared with the measurements and these biases are spatially interdependent. Thus, in order to identify correct parameter values, a more robust approach involves i) estimating these biases using non-zero-mean PDFs whose sources and forms are determined using engineering judgment and ii) defining probabilistic threshold bounds based on the combination of each source.

4. Illustrative example

This example studies a simply supported beam that is partially fixed on the left side by a rotational spring as shown in Figure 1. This beam has a rectangular cross-section with a moment of inertia I and a

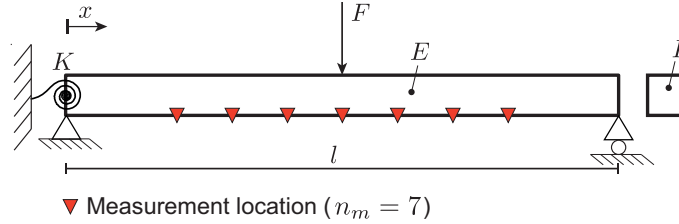


Figure 1: True beam configuration with measurement locations for $n_m = 7$.

length l . Its other characteristics are the Young's modulus E and the spring stiffness K . Values for these characteristics are given in Table 1. The beam deflection $v(x)$ for a single load F applied at midspan can be computed using Eq. (12).

$$v(x) = \begin{cases} \frac{Fx(18EI^2 + 9Kl^2x - x^2(24EI + 11Kl))}{96EI(3EI + Kl)} & \text{if } x \leq l/2 \\ \frac{F(x-l)(2l^2(3EI + Kl) + (x^2 - 2xl)(24EI + 5Kl))}{96EI(3EI + Kl)} & \text{if } x > l/2 \end{cases} \quad (12)$$

Displacement measurements are simulated by adding a measurement error following an independent zero-mean Gaussian distribution of standard deviation $\sigma_{\text{measure}} = 0.01 \text{ mm}$ to the real displacement $v(x)$. Measurement locations are uniformly distributed between the coordinate $x = x_s = 1000 \text{ mm}$ and $x = l - x_s$. The coordinate x_i of each measurement y_i is defined by:

$$x_i = x_s + (l - 2x_s) \frac{i}{n_m + 1} \quad (13)$$

The measurements are simulated for various number of measurements $n_m \in \{1, 3, 7, 15\}$. By increasing n_m in this way, each higher measurement set includes the locations of the lower sets.

In this example, simulated measurements are used to identify the structural behavior of the beam. This knowledge is then used to predict the displacement at midspan under the same load (interpolation) and predict the tensile strain at the bottom of the beam at the coordinate $x = 2250 \text{ mm}$ under the distributed load $q = 5 \text{ kN/m}$ as shown in Figure 2. The strain prediction is calculated using Eq. (14) where h is the

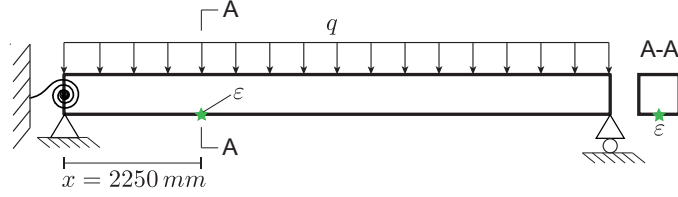


Figure 2: Strain prediction at the bottom of the beam cross-section at $x = 2250 \text{ mm}$. This configuration is used for model extrapolation.

height of the beam cross-section that has a true value of 303 mm .

$$\varepsilon(x) = -\frac{q(l-x)(Kl(l-4x) - 12EIx)}{8EI(3EI + Kl)}h/2 \quad (14)$$

In order to study the effect of systematic modeling errors on the parameter identification and model predictions, four model classes are built:

- The *true model class* G_0 is composed of the four parameters $\boldsymbol{\theta}_0 = [\theta_{01}, \theta_{02}, \theta_{03}, \theta_{04}]^T = [E, K, I, l]^T$ and thus is built by parameterizing every uncertain characteristic of the system.
- *Model class* G_1 involves the assumption of a simply supported beam with a pinned connection on the left side and approximated values for the moment of inertia I and the length l obtained by measuring the dimensions of the beam cross-section and span. This model class accounts for the uncertainty associated with the Young's modulus value and uses it as a parameter $\theta_{11} = E$ in the system-identification process.
- *Model class* G_2 is defined through recognition of the partial connection on the left side of the beam and the uncertainty associated with the moment of inertia. For this model class, the parameter vector is $\boldsymbol{\theta}_2 = [\theta_{21}, \theta_{22}]^T = [K, I]^T$. The Young's modulus and the length are set to their approximated values, see Table 1.
- *Model class* G_3 , similarly to Model class G_1 includes a pinned connection on the left side of the beam, and the moment of inertia and the length are taken as parameters. For this model class, the parameter vector is $\boldsymbol{\theta}_3 = [\theta_{31}, \theta_{32}, \theta_{33}]^T = [E, I, l]^T$.

Table 1: Real beam characteristics and model-class description

	$E[GPa]$	$K[\log(Nmm/rad)]$	$I[mm^4] \times 10^8$	$l[mm]$	$F[kN]$
Real beam	70	9.8	7.01	8920	12
Initial parameter ranges	40 – 100	8 – 12	6.48 – 7.02	8910 – 9090	–
True model class G_0	θ_{01}	θ_{02}	θ_{03}	θ_{04}	12
Model class G_1	θ_{11}	$\rightarrow 0$ (pinned)	6.75	9000	12
Model class G_2	59	θ_{21}	θ_{22}	9000	12
Model class G_3	θ_{31}	$\rightarrow 0$ (pinned)	θ_{32}	θ_{33}	12

For each model class G_k , the applied load F is identical. A summary of model-class parameters and approximated values is given in Table 1.

The purpose of this illustrative example is to compare Bayesian inference results with error-domain model falsification results when performing parameter identification, interpolation and extrapolation. Model predictions $\mathbf{g}_{ki}(\boldsymbol{\theta}_k)$ are determined for each model class using Eq. (12) at the measurement coordinate x_i using approximated values given in Table 1. Uncertain parameters have initial ranges of values estimated using engineering judgment.

For Bayesian inference, these ranges define uniform prior probabilities of the uncertain parameters that are used to solve Eq. (1). The integral of the normalization of this equation may be difficult to compute for high-dimensional parameter domain. For each model-class identification, one million Markov Chain samples are used and the convergence of the posterior PDF is verified. Three scenarios are assumed for the estimation of the modeling errors:

- *Scenario I* involves no recognition of modeling errors and includes only the measurement error during parameter identification. The uncertainty distribution (Eq. 2) is thus centered on zero with a covariance matrix having $\sigma_\epsilon^2 = \sigma_{\text{measure}}^2$ for the diagonal elements. For prediction, only the parameter uncertainty is included since erroneously, no modeling uncertainty is assumed.
- *Scenario II* involves recognition of simplifications and omissions present in the model and evaluates the modeling errors by independent Gaussian distributions. Mean and standard deviation values of the modeling uncertainty are estimated using engineering judgment (see Table 2). In this scenario, the likelihood function then includes both measurement and modeling uncertainties for parameter identification. For prediction, modeling uncertainties are included based on Eq. (3) using estimated values given in Table 2.
- *Scenario III* involves an assumption of independent zero-mean Gaussian distribution for the modeling and measurement errors and parametrizes the variance σ_ϵ^2 of the likelihood function that becomes an additional parameter to identify. In this scenario, the prior PDF of the variance is uniform and varies

between 0 and 100% of the measured value at each measurement location. The identified variance values are then assumed to be estimates of the modeling error and are included in interpolation and extrapolation calculations.

Table 2: Modeling error estimations by independent Gaussian distributions for parameter identification and prediction.

	Identification		Interpolation		Extrapolation	
	mean [%] ^a	std [%] ^a	mean [mm]	std [mm]	mean [$\mu\epsilon$]	std [$\mu\epsilon$]
Model class G_0	0	0	0	0	0	0
Model class G_1	28	4.7	0.86	0.14	30.3	3.46
Model class G_2	34	5.4	1.04	0.16	28.6	4.53
Model class G_3	22	4	0.67	0.12	27.7	3.36

^aThis percentage refers to the measured value at each measurement location.

For EDMF, the initial parameter ranges in Table 1 are divided into 60 uniformly distributed values in order to generate the initial model set. For scenario I, the modeling errors are unrecognized and threshold bounds are calculated by including only the measurement uncertainty. For scenario II and III, the modeling uncertainties are included using values given in Table 2. Modeling and measurement uncertainties are combined to obtain $U_{c,ki}$ and threshold bounds are determined using Eq. (10). After falsification of inadequate models, the candidate models are used for interpolation and extrapolation based on Eq. (11) and including the modeling uncertainties given in the interpolation and extrapolation columns of Table 2. For each scenario, target probabilities, $\phi_d = 0.95$ for identification, and $\phi_p = 0.95$ for prediction are chosen.

For Bayesian inference, the combined mean values of modeling and measurement uncertainties are included in the residual calculation $\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})$ of the likelihood function (Eq. 2) and the combined variances are included in the covariance matrix Σ , which is diagonal since for any scenario, the errors are assumed to be independent.

The next section presents the comparison of Bayesian inference and EDMF for the true model class for this illustrative example. Section 4.2 presents the same comparisons for the other model classes for the three scenarios. Bayesian model class selection is also investigated in Section 4.4. These comparisons are presented as summarized in Table 3.

4.1. True model class

The true model class G_0 is the result of the parametrization of all sources of modeling uncertainty. Since no systematic error is induced by the model predictions, the measurement uncertainty only needs to be included during data interpretation. As a result, the assumption of independent zero-mean Gaussian uncertainty for Bayesian inference is justified and parameter identification leads to accurate and precise posteriors for the four parameters. When using these posteriors to extrapolate the predictive distribution of

Table 3: Summary of comparisons by sections and model classes.

Section	Scenario	Identification	Interpolation	Extrapolation
4.1	-	-	-	G_0
4.2.1	I	G_1, G_2, G_3	G_1	G_1
4.2.2	II	G_1, G_2, G_3	G_1	G_1
4.2.3	III	G_1, G_2, G_3	G_1	G_1
4.4	I,III	-	-	G_1, G_2, G_3^a

^aBased on posterior model averaging.

the strain using Eq. (3) and (14), the results are accurate and precise for $n_m \geq 3$ if compared with the true strain value as depicted in Figure 3. This figure presents also predictions obtained with EDMF. Although EDMF predicts accurately the strain, the extrapolation is less precise than with Bayesian inference.

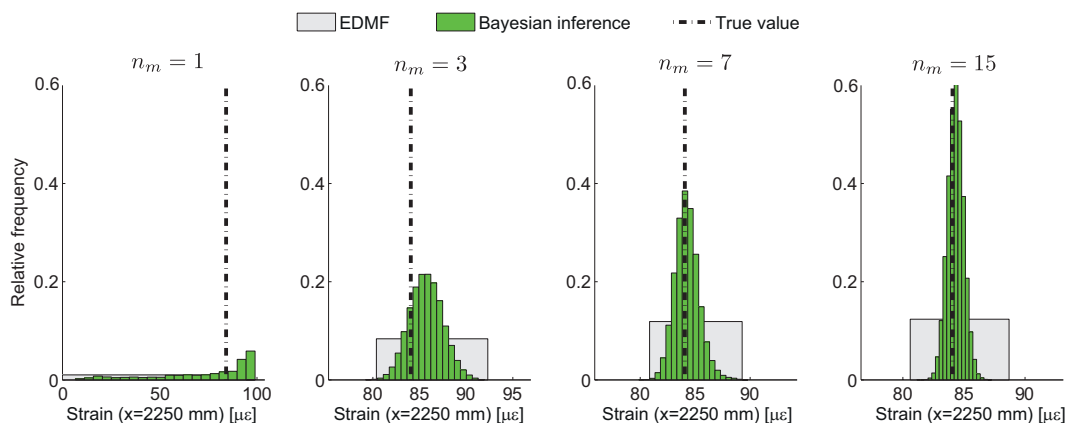


Figure 3: Example of strain predictions at $x = 2250 \text{ mm}$ that lead to accurate and reliable extrapolation of the true model. Predictions obtained using Bayesian inference and error-domain model falsification (EDMF) are compared with the true strain value where the number of measurements n_m varies from 1 to 15. In this case, where only measurement errors are accommodated, EDMF leads to predictions that are less precise than Bayesian inference.

4.2. Unknown model class

For complex systems, the true model class is generally not definable. Only imperfect model classes are defined to approximate the real behavior. In this section, the three model classes $G_k, k \in \{1, 2, 3\}$ are investigated for the three scenarios that reflect conventional assumptions made when applying Bayesian inference for system identification and results are compared with those obtained using EDMF.

4.2.1. Scenario I: unrecognized modeling errors

In this scenario, modeling uncertainties are not included in the identification framework, neither for Bayesian inference nor for EDMF. Table 4 presents a comparison of parameter identification results obtained

using Bayesian inference and EDMF for model classes G_1 to G_3 and $n_m = 7$. Since posteriors of E and K are close to a Gaussian distribution, the range of identified values is given with two standard deviations apart from the mean value of the posterior PDF. Posteriors of I and l are close to uniform distribution and thus, minimum and maximum values are displayed. This table shows that parameter values that are inferred using Bayesian framework are biased for any model class, while EDMF falsifies all initial model instances. Figure 4 describes parameter identification (Young’s modulus), interpolation and extrapolation

Table 4: Scenario I: comparison of parameter identification using Bayesian inference and using error-domain model falsification (EDMF) for $n_m = 7$. Model-class falsification is indicated as 0 CM (zero candidate models).

Parameter θ_k		$E[GPa]^a$	$K[\log(Nmm/rad)]^a$	$I[mm^4] \times 10^{8b}$	$l[mm]^b$
True values		70	9.8	7.01	8920
Initial ranges		40 – 100	8 – 12	6.48 – 7.02	8910 – 9090
G_1	Bayesian inference	89.1 – 89.7	-	-	-
	EDMF (min-max)	0 CM	-	-	-
G_2	Bayesian inference	-	10.18 – 10.21	6.99 – 7.02	-
	EDMF (min-max)	-	0 CM	0 CM	-
G_3	Bayesian inference	86.8 – 95.7	-	6.48 – 7.02	9016 – 9090
	EDMF (min-max)	0 CM	-	0 CM	0 CM

^aRanges for Bayesian inference are $\pm 2\sigma$ apart from the mean posterior.

^bRanges for Bayesian inference are minimum and maximum values of the posterior.

results obtained using Bayesian inference and EDMF for model class G_1 . Results are compared with the true values. Although similar results are obtained for model classes G_2 and G_3 , the result for the model class having a one-dimensional parameter domain is shown for convenience. This figure exhibits that when modeling uncertainty is not included, Bayesian inference leads to biased identification and predictions, while EDMF falsifies every model in the model class and thus avoids making a wrong parameter-value identification and wrong predictions. This was also observed in Goulet and Smith [20] for parameter identification. Note that model-class falsification is observed only for $n_m > 1$. When a single measurement is used, EDMF returns similar results than Bayesian inference for this scenario.

4.2.2. Scenario II: recognized modeling errors

In this case, systematic modeling errors are recognized and carefully evaluated by independent Gaussian uncertainties. Table 5 presents a comparison of the parameter identification results obtained using both approaches for the three model classes. This table shows that Bayesian inference makes correct identification of the parameter values. Here, the assumption of independence of errors amongst measurement locations is not justified since the nature of the modeling error induces spatial interdependencies (see Table 1: each model-class definition induces systematic errors). However, correct parameter-value identification is obtained

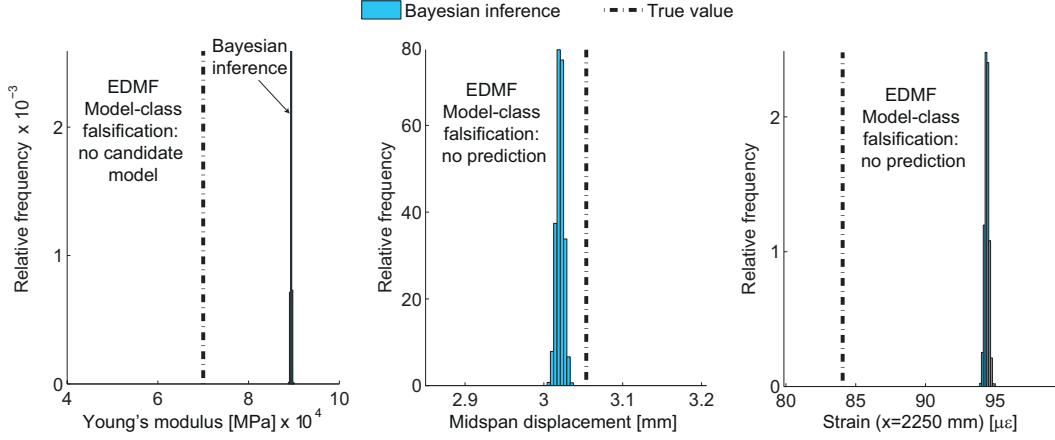


Figure 4: Scenario I: parameter identification (Young’s modulus), midspan displacement prediction (interpolation) and strain prediction at $x = 2250\text{ mm}$ (extrapolation) using model class G_1 for $n_m = 7$. Bayesian inference is compared with EDMF and the true values. Since modeling errors are not taken into account, Bayesian inference leads to biased identification and predictions. EDMF falsifies every model in the model class and thus avoids making a wrong parameter-value identification and wrong predictions.

regardless of this wrong assumption [20, 40]. In addition, EDMF also leads to robust identification of parameter values for each model class.

Figure 5 describes parameter identification (Young’s modulus), interpolation and extrapolation results obtained using Bayesian inference and EDMF for model class G_1 and $n_m = 7$.

These results are compared with the true values. Bayesian inference results in correct parameter identification as depicted also in Table 5. Interpolation and extrapolation predictive distributions are presented with inclusion of modeling uncertainties and without inclusion of them in order to show the importance of accounting for them in the prediction calculation (Eq. 3). Indeed, when they are included in predictive distribution calculations, interpolation and extrapolation are accurate, which is not the case without inclusion of modeling uncertainties. However, such results are not always obtainable due to either wrongly assumed spatial correlation values or misevaluation of the mean value of the modeling errors. For the model class G_2 , extrapolations are biased even if modeling uncertainty is included in the predictive distribution (see Figure 6).

In this figure, although the true value is near the higher prediction threshold of EDMF, the extrapolation range still includes the true value with $\phi_p = 95\%$ reliability given the uncertainty distribution, while the position of the true value is in the distribution tail of the Bayesian predictive distribution (larger than $\pm 2\sigma$ from the mean prediction). Since modeling uncertainty is always included in both identification and prediction processes (Eq. 10 and 11), EDMF is able to identify correctly the parameter value and to predict displacement and strain values accurately and reliably.

Table 5: Scenario II: comparison of parameter identification using Bayesian inference and using error-domain model falsification (EDMF) for $n_m = 7$.

Parameter θ_k	$E[GPa]^a$	$K[\log(Nmm/rad)]^a$	$I[mm^4] \times 10^{8b}$	$l[mm]^b$
True values	70	9.8	7.01	8920
Initial ranges	40 – 100	8 – 12	6.48 – 7.02	8910 – 9090
G_1	Bayesian inference	68.3 – 72.2	-	-
	EDMF (min-max)	66.4 – 76.6	-	-
G_2	Bayesian inference	-	9.79 – 9.99	6.48 – 7.02
	EDMF (min-max)	-	9.65 – 10.1	6.48 – 7.02
G_3	Bayesian inference	69.8 – 79.1	-	8910 – 9090
	EDMF (min-max)	66.7 – 84.9	-	8910 – 9090

^aRanges for Bayesian inference are $\pm 2\sigma$ apart from the mean posterior.

^bRanges for Bayesian inference are minimum and maximum values of the posterior.

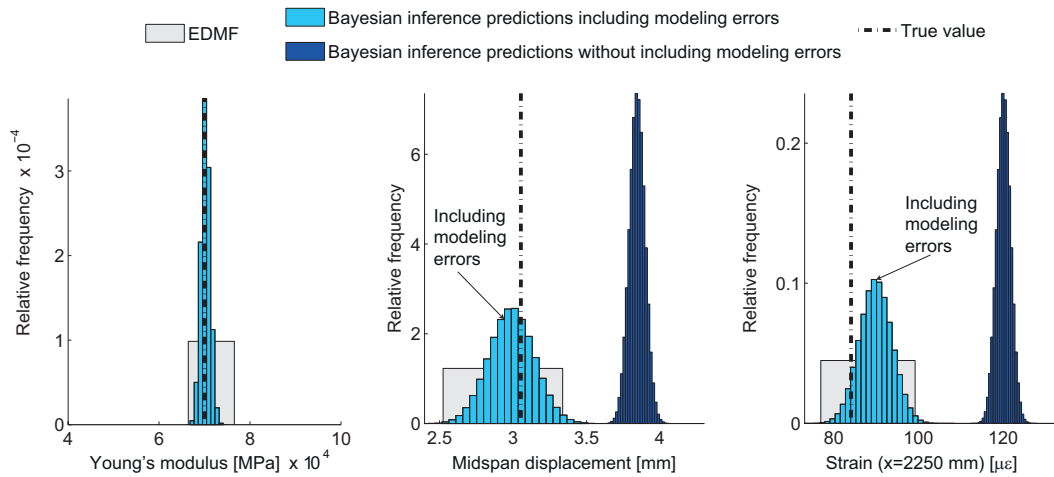


Figure 5: Scenario II: parameter identification (Young's modulus), midspan displacement prediction (interpolation) and strain prediction at $x = 2250$ mm (extrapolation) using model class G_1 for $n_m = 7$. While systematic modeling errors are estimated and included in the identification, Bayesian inference leads to better identification and extrapolation. However, such results are not always obtainable. In addition, EDMF is robust to systematic modeling errors.

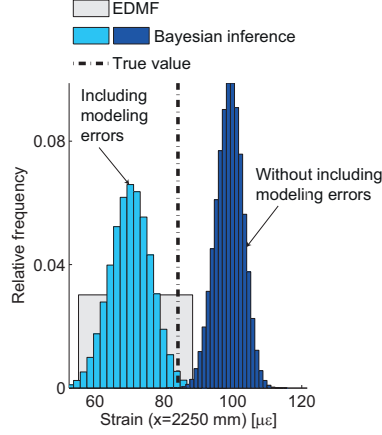


Figure 6: Scenario II: strain prediction at $x = 2250 \text{ mm}$ (extrapolation) using model class G_2 for $n_m = 7$. For this model class, Bayesian inference returns biased extrapolation due to either wrongly assumed spatial correlation values or misevaluation of the mean value of the modeling errors.

4.2.3. Scenario III: parametrized modeling uncertainties

In the third scenario, the variance of likelihood function becomes an additional parameter to be identified in Bayesian inference. Table 6 presents a comparison of the parameter identification results obtained using Bayesian inference and EDMF.

Table 6: Scenario III: comparison of parameter identification using Bayesian inference and using error-domain model falsification (EDMF) for $n_m = 7$.

Parameter θ_k		$E[\text{GPa}]^a$	$K[\log(N\text{mm}/\text{rad})]^a$	$I[\text{mm}^4] \times 10^{8b}$	$l[\text{mm}]^b$	$\sigma_\epsilon^2[\%]^c$
True values		70	9.8	7.01	8920	-
Initial ranges		40 – 100	8 – 12	6.48 – 7.02	8910 – 9090	0 – 100
G_1	Bayesian inference	87.9 – 92.0	-	-	-	(0.09; 0.13)
	EDMF (min-max)	66.4 – 76.6	-	-	-	-
G_2	Bayesian inference	-	10.1 – 10.4	6.48 – 7.02	-	(0.7; 0.69)
	EDMF (min-max)	-	9.65 – 10.1	6.48 – 7.02	-	-
G_3	Bayesian inference	84.9 – 96.3	-	6.48 – 7.02	8910 – 9090	(0.09; 0.11)
	EDMF (min-max)	66.7 – 84.9	-	6.48 – 7.02	8910 – 9090	-

^aRanges for Bayesian inference are $\pm 2\sigma$ apart from the mean posterior.

^bRanges for Bayesian inference are minimum and maximum values of the posterior.

^cValues for Bayesian inference are (mean;std) of log-normal distribution.

EDMF results, identical to Table 5, are shown a second time for comparison. As for Table 5, ranges for E and K are shown in terms of $\pm 2\sigma$ from the mean value, I and l in terms of minimum and maximum values and σ_ϵ^2 , which is close to a log-normal distribution is described by its mean and standard deviation values. This

table shows that when modeling uncertainties are parametrized and systematic errors are present, Bayesian inference results in biased identification. For model class G_2 and G_3 , although the moment of inertia is correctly identified, the posteriors of the other parameter values are biased. The identified variances do not describe correctly the modeling and measurement errors and thus, the identified parameter values cannot represent the true beam behavior.

Figure 7 describes parameter identification (Young’s modulus), interpolation and extrapolation results obtained using Bayesian inference and EDMF for model class G_1 and $n_m = 7$ and are compared with the true values. Bayesian inference leads to biased identification when compared with the true parameter

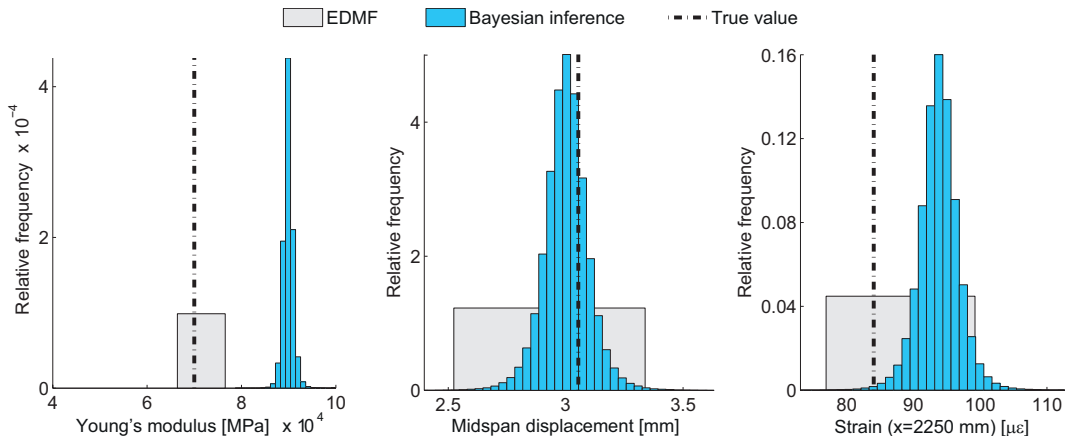


Figure 7: Scenario III: parameter identification (Young’s modulus), midspan displacement prediction (interpolation) and strain prediction at $x = 2250$ mm (extrapolation) using model class G_1 for $n_m = 7$. When parameterizing the uncertainty variance, Bayesian inference leads to biased identification and extrapolation in the presence of systematic errors. As in scenario II, EDMF leads to robust identification and accurate predictions since modeling errors are accommodated in the same way.

value. In this scenario, σ_ϵ^2 acts as a tuning parameter and compensates the effects of the other parameters leading to biased identification. As each parameter value compensates for each other to fit the measurement, interpolating with such models leads to accurate predictions. However, when extrapolating, since wrong parameter values and wrong modeling uncertainties are identified, the predictions are inaccurate. When interpolating, calculations may lead to accurate predictions even if the parameter identification is wrong. This can lead to overconfidence in model classes and thus, inaccurate subsequent extrapolations. As in scenario II, EDMF leads to robust identification and accurate predictions since modeling errors are accommodated in the same way.

4.3. Summary of scenarios I to III

Table 7 summarizes findings related to scenarios, model classes and activities such as parameter identification, interpolation and extrapolation for both data-interpretation approaches. Error-domain model falsification is robust whatever activity is performed with each model class. In the first scenario, where

Table 7: Summary of findings (\checkmark : successful, \times : not successful).

Approach	Bayesian inference									Error-domain model falsification							
Scenario	-	I			II			III			-	I			II-III		
Model class	G_0	G_1	G_2	G_3	G_1	G_2	G_3	G_1	G_2	G_3	G_0	G_1	G_2	G_3	G_1	G_2	G_3
Identification	\checkmark	\times	\times	\times	\checkmark	\checkmark	\checkmark	\times	\times	\times	\checkmark	\checkmark^a	\checkmark^a	\checkmark^a	\checkmark	\checkmark	\checkmark
Interpolation	\checkmark	\times	\times	\times	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Extrapolation	\checkmark	\times	\times	\times	\checkmark	\times	\checkmark	\times	\times	\times	\checkmark	\checkmark^a	\checkmark^a	\checkmark^a	\checkmark	\checkmark	\checkmark

^aProvided that a sufficient number of measurements is used (in this case, $n_m > 1$).

modeling errors are not recognized in the presence of systematic errors, EDMF is able to detect erroneous model-class definitions provided that more than a single measurement is used. In other scenarios where modeling errors are adequately estimated, EDMF provides robust parameter identification and accurate predictions. Pasquier et al. [36] have shown that the sensitivity to diagnosis error (i.e. making biased identification) depends on the number of measurements and the degree of misevaluation of either the mean value or the variance of uncertainty. Also, when the variance of modeling uncertainty is underestimated, model-class falsification is possible for a sufficient number of measurements that depends on the degree of underestimation. Although increasing the number of measurements helps, in case of slight underestimation of uncertainty mean value and variance, it may be possible to identify the wrong model class. Nevertheless, EDMF is capable of managing model instances that originate from several model classes.

When the true model class is involved with all uncertainties attached to physical parameters, Bayesian inference leads to accurate and precise identification and predictions. However, in the first scenario, Bayesian inference cannot detect erroneous assumptions made with modeling uncertainties and leads to inaccurate results for any activity and model class. In the second scenario, the inclusion of modeling uncertainties results in accurate identification and interpolations. However, the accuracy of extrapolations cannot be guaranteed for every model class. In the third scenario, the description of uncertainties by a zero-mean Gaussian distribution with parametrization of the variance of the likelihood function is not a robust methodology in the presence of systematic errors. Indeed, the parameter values identified are inaccurate for any model class and thus lead also to inaccurate extrapolations. However, interpolations are accurate since the fitted models have predictions that are valid in the domain of experimentation.

The inaccurate results obtained with Bayesian inference are the consequence of either the underestimation of either the mean value or the variance of the modeling uncertainties in the model-class definition. Since conservative estimation of modeling errors is part of the model-class definition, error-domain model falsification offers an alternative approach that is robust and reliable for identifying systems and making predictions.

4.4. Results obtained using Bayesian model class selection

In situations where various model classes are possible representations of a system, Bayesian model class selection is used either to identify the optimal model class given the measurement data or to rate model classes regarding their plausibilities. This approach is applied for model classes G_1 to G_3 for scenarios I and III by evaluating first the log-evidence for each model class for several numbers of measurements. Then, posterior model averaging is used to study the potential of this methodology for improving prediction accuracy.

Log-evidence values are determined using MCMC simulation with one million Markov Chain samples. Figure 8 presents the comparison of log-evidence values of the three model classes and the true model class obtained using Eq. (5) and calculated for n_m varying between 1 and 15. Understandably, the true model

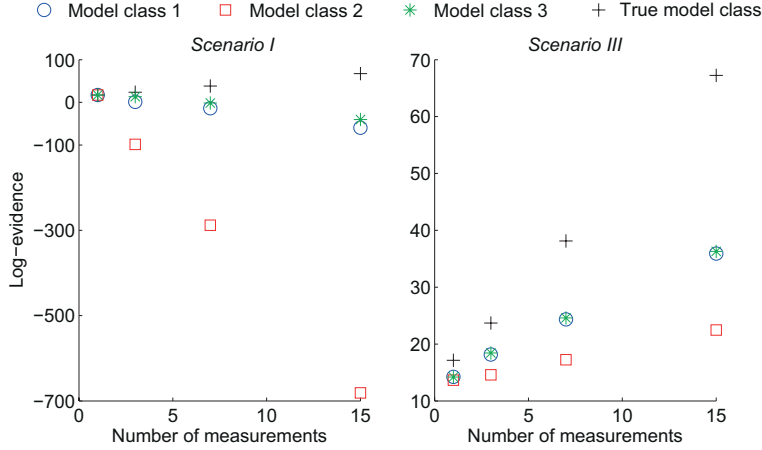


Figure 8: Comparison of log-evidence of model classes G_1 to G_3 with the true model class for scenarios I and III where the number of measurements n_m varies from 1 to 15. The true model class returns the largest log-evidence for any number of measurements. Model class G_3 is the best model class among the three whatever the number of measurements according to Bayesian model-class selection even if the log-evidence of G_3 and G_1 are very close.

class returns the largest log-evidence for any number of measurements. Model class G_3 is the best model class among the three in scenario I for any n_m . The log-evidence value of model class G_2 is very low compared with the others. This classification is due to the size of the systematic bias in each model class that is lower in G_3 than in G_1 and G_2 and that is reflected in the estimation of the modeling errors in Table 2. In scenario III, the log-evidence values of model classes G_1 and G_3 are very close. Model class G_2 has clearly the lowest log-evidence value for any number of measurement larger than 1. This is due to the identified variance parameter that is larger than for the other model classes as depicted in Table 6.

Based on the log-evidence values, the plausibilities of each model class are updated with Eq. (6). Table 8 summarizes the updated plausibility values. These values reflect the classification of log-evidence values of Figure 8. In scenario I, model class G_3 is optimal for $n_m > 1$ and in scenario III, these values cannot

depict an unambiguous optimal model class with a plausibility of 1. These plausibility values are then used to weight the predictive distribution of each model class based on Eq. (7) in order to obtain the posterior model averaging of the extrapolation values. Table 8 compares also the model classes using EDMF. In scenario I, although EDMF identifies wrong candidate models for $n_m = 1$, all model classes are falsified for $n_m > 1$, in agreement with the erroneous model-class definition. In scenario III, candidate models are correctly identified for any set of measurements since the estimated modeling uncertainties are compatible with the model-class definition.

Table 8: Model plausibilities obtained using Bayesian model class selection for n_m varying from 1 to 15.

Scenario	Bayesian plausibilities								Model-class falsification ^a							
	I				III				I				III			
	n_m	1	3	7	15	1	3	7	15	1	3	7	15	1	3	7
Model class G_1	0.34	0.00	0.00	0.00	0.40	0.43	0.44	0.41	1	0	0	0	1	1	1	1
Model class G_2	0.33	0.00	0.00	0.00	0.22	0.01	0.00	0.00	1	0	0	0	1	1	1	1
Model class G_3	0.33	1.00	1.00	1.00	0.38	0.56	0.56	0.59	1	0	0	0	1	1	1	1

^aBelow, 0 means complete model-class falsification and 1 means that candidate models are identified.

Figure 9 presents strain predictions obtained by weighting each model-class predictive distribution by their plausibility and EDMF predictions of model class G_3 for scenario I and for n_m varying from 1 to 15. Even by weighting the predictions with the model-class plausibilities, the extrapolations are inaccurate

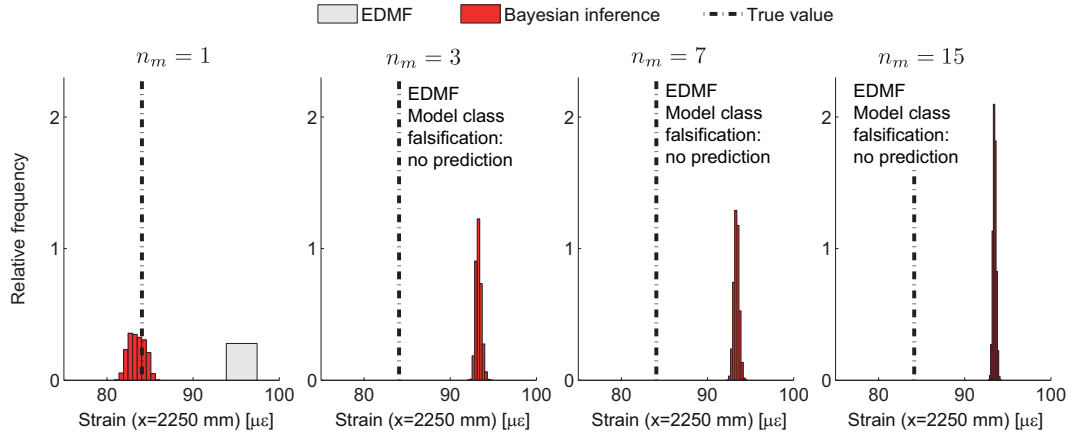


Figure 9: Comparison of strain predictions (extrapolation) obtained by weighting model-class predictive distributions by their plausibility with EDMF predictions of model class G_3 for scenario I and n_m varying from 1 to 15. Although predictions are weighted by the model-class plausibilities, the extrapolations are biased for any number of measurements larger than 1. For $n_m > 1$, EDMF falsifies the model class and avoids wrong identification and predictions.

for any number of measurements larger than 1. For $n_m = 1$, model class G_2 underestimates the strain

prediction, while G_1 and G_3 overestimate it.

When n_m increases, since all model classes identify incorrectly the parameter values for this scenario, the model class selection is unable to provide better accuracy and reliability of predictions than using a single model class. In addition, the more measurements are used, the more inaccurate the strain prediction is. With EDMF, although a single measurement is insufficient to detect the erroneous assumption of uncertainty estimation since wrong extrapolations are obtained, this approach avoids incorrect extrapolations when n_m increases.

Figure 10 presents the same comparison as in Figure 9 for scenario III. By weighting the predictions with the model class plausibilities, the extrapolations are still inaccurate for any number of measurements larger than 1. In addition, in this scenario also, when the number of measurements increases, the prediction inaccuracy increases. The extrapolations obtained using EDMF and model class G_3 are accurate and reliable for any number of measurements. Here, only the extrapolations using G_3 are shown because this is the most accurate model class among the three and also, this model class has the least bias. For this scenario, posterior model averaging is unable to help improve the accuracy and the reliability of extrapolations.

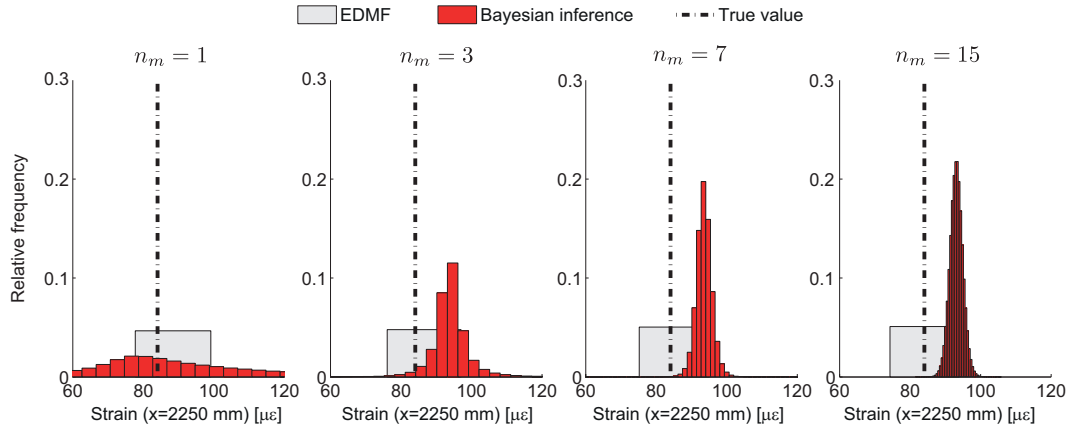


Figure 10: Comparison of strain predictions obtained by weighting model-class predictive distributions by their plausibility with EDMF predictions of model class G_3 for scenario III and n_m varying from 1 to 15. Although predictions are weighted by the model-class plausibilities, the extrapolations are biased for any number of measurements larger than 1. In addition, the prediction inaccuracy increases when additional measurements are involved. EDMF is still robust to systematic modeling errors and provide accurate identification and extrapolation of any number of measurements.

Figure 9 and 10 demonstrate that Bayesian model class selection may not be able to uncover wrong assumptions related to model classes and modeling error estimations and thus, it is not possible to determine the accuracy of extrapolations. These figures along with Table 8 demonstrate also that EDMF is robust when erroneous assumptions are made regarding modeling errors (provided that more than a single measurement is used), thereby avoiding making inaccurate extrapolations. When modeling errors are compatible with model-class definitions, EDMF provides accurate and reliable extrapolations.

Figure 11 compares the initial-model-set (IMS) strain predictions, the candidate-model-set (CMS) predictions for model class G_3 and $n_m = 7$, and the true strain value. Taking all cases, the reduction in prediction

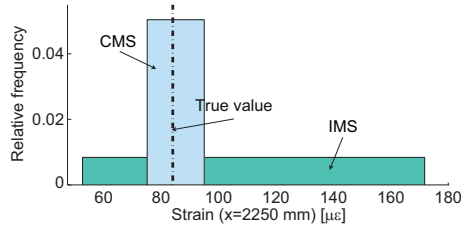


Figure 11: Strain prediction ranges for model class G_3 using the initial model set (IMS), the candidate model set (CMS) for $n_m = 7$, and the true strain value. The reduction in prediction range from IMS predictions to CMS predictions is up to 83%. EDMF is thus able to reduce significantly the uncertainty associated with the extrapolations and return reliable predictions.

range from IMS predictions to CMS predictions is up to 83%. While in general, EDMF returns less precise predictions than Bayesian inference, when compared with the initial knowledge of structural behavior (IMS), EDMF is able to significantly reduce the uncertainty associated with the extrapolations and return reliable predictions.

5. Discussion

For complex civil structures, accuracy of model predictions is important since it governs management decisions such as retrofit, repair and replacement. In such systems, physics-based models are usually imperfect and induce systematic modeling errors. The illustrative example demonstrates that erroneous assumptions regarding modeling uncertainty lead to inaccurate Bayesian predictive distributions that might consequently lead to bad decisions. Although including these uncertainties in the Bayesian framework improves the accuracy of extrapolations, predictions are inaccurate in some situations. Conversely, error-domain model falsification is robust in every scenario and thus, this approach provides accurate and reliable predictions.

The simplicity of the illustrative example allows for knowledge of true values, thus providing a comparative basis for data-interpretation approaches. In real systems, true values are unknown. However, this study is necessary for investigating the validity of such approaches. System-identification techniques that may not be robust and accurate for simple examples are unlikely to be adequate for identification of complex systems in practice.

The estimation of modeling errors, knowing the true values, is trivial. For complex systems, some modeling errors, such as geometric variability of the system, the variability of material properties and mesh-refinement uncertainty of finite-element models, can be explicitly quantified. Other sources are estimated based on engineering judgment. This requires that the nature of the modeling error is recognized. Brynjarsdóttir and O’Hagan [10] observed that modelers are aware of what physical processes are missing in the

model, and they also recommend the use of engineering judgment to estimate model discrepancy. Estimating modeling errors rather than modifying the model is meaningful in situations where improving the model is not possible and when changes would complicate the model to an extent that would either increase the number of parameters to identify or lengthen computational time excessively.

When modeling errors are not included during system identification and systematic errors are present, error-domain model falsification is able to falsify the model class and avoid making wrong predictions. By including estimated modeling errors, EDMF is able to select model classes based on their prediction precision. Even if the precision is low, predictions are accurate, and when further information is gathered, prediction precision may be increased. Thus, EDMF is able to explore possible model classes by falsifying inadequate model classes and by selecting promising ones based on prediction precision. For this exploration, model classes need to be defined as two components: (1) unknown physical parameters that are identified using measurements; and (2) conservative modeling error estimations that cannot be represented only as uncertainties related to physical parameters. This second component is the reason for the words "error domain" in EDMF. Special care is needed for the estimation of the modeling errors and it should be done independently of the identification process. Parametrization of uncertainties to increase confidence in estimated values is not always successful.

System identification of complex civil structures is an ill-posed task that is carried out in open-world conditions and that leads to multiple solutions for the structural health management of existing structures. In such complex tasks, engineers have to rely on heuristics and experience as well as data-interpretation tools such as Bayesian inference and EDMF. Engineering knowledge combines with data interpretation through assumptions in processes such as model-class building, initial ranges of parameter values and estimation of modeling and measurement errors. In addition, data interpretation needs to be flexible with respect to modification of the engineering knowledge and robust with respect to simplifying assumptions.

5.1. A knowledge intensive example

Figure 12 presents a full-scale bridge that is studied in [35]. EDMF has been used for the study of this structure using measurement data recorded from static load tests in order to identify candidate models that are then used to predict remaining fatigue life of critical connections.

A FE model is built and several sources of modeling uncertainty are identified, including rotational stiffness of truss connections, stiffness of expansion joints, material properties, geometrical properties, mesh refinement and other model simplifications. The most influential sources of uncertainty are taken as parameters and the others are represented by modeling uncertainties, estimated using engineering judgment and field heuristics, and then combined to determine $U_{c,i}$.

Sources such as Poisson's ratio, diameter and thickness of hollow sections are parametrized and evaluated through the FE model using Monte Carlo simulation. For the others, conservative estimations of the PDF of

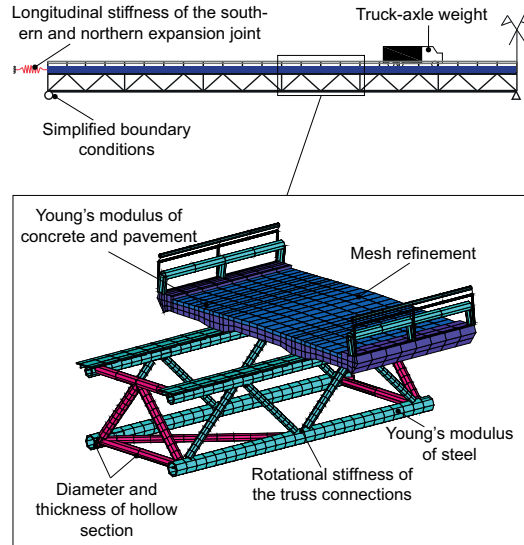


Figure 12: Example of the Aarwangen Bridge and description of modeling uncertainties induced by model simplifications. Engineering knowledge is required to build the model class and identify sources of modeling and measurement errors to accommodate in EDMF in order to identify candidate models and predict remaining fatigue life of critical connections.

each source are made based on engineering judgment. Systematic bias due to model simplifications, such as boundary conditions and truss connections, are estimated and included in the threshold-bound calculation. The engineering knowledge required to support the identification and the evaluation of remaining fatigue life of such complex civil structures evolves as information is acquired and it often necessitates modification. Table 9 summarizes the engineering knowledge that is required for the study of a structure such as the Aarwangen Bridge.

EDMF presents the advantage of estimating modeling errors and accommodating any value of correlation since thresholds represent a rectangular coverage region that is conservatively adjusted to the number of measurements using the Šidák correction [20]. However, in the case of slight underestimation of uncertainty mean value and variance, it may be possible to make a diagnosis error by falsely accepting wrong model instances instead of falsifying the entire model class. Nevertheless, EDMF is capable of managing model instances that originate from several model classes.

Common use of Bayesian inference does not account for systematic errors. In addition, it usually leads to unsafe assumptions describing modeling and measurement errors by a joint independent zero-mean Gaussian PDF. Such practices would require the parametrization of all sources of modeling uncertainty such as is illustrated for model class G_0 . However, since there may be tens of sources, this is usually not feasible for complex structures and it may also lead to an over-fitted model class. In addition, usage of Bayesian model-class selection may lead to overconfidence in a model-class choice for the Aarwangen Bridge. Thus, traditional use of Bayesian inference is not compatible with diagnosis and prognosis tasks of real case studies.

Table 9: Summary of the engineering knowledge necessary to perform parameter identification and remaining-fatigue-life predictions of a structure such as the Aarwangen Bridge. Since these tasks are knowledge intensive, ease of representation and modification are important aspects.

Required engineering knowledge	Ease of representation and modification within EDMF
Model-class building	Behavioral hypotheses are tested with measurement data (EDMF supports model class exploration through falsifying erroneous model classes)
Initial parameter ranges	Included explicitly in the initial model set
Identification of sources of errors	Ease of modification through combination of all sources of uncertainty
Estimation of modeling and measurement errors for identification	Inclusion of modeling bias due to simplifications, ease of modification and robust to unknown correlation values [20]
Estimation of modeling errors for prediction	Inclusion of modeling bias, ease of modification, particularly for sources that do not need to be accounted for explicitly in the predictions ^a and robust to unknown correlation values

^aFor example, for remaining-fatigue-life predictions, variability of diameter and thickness of hollow sections are already taken into account in the fatigue-strength curves provided by codes.

Conservative estimations of modeling errors are needed to avoid the risk of falsely identifying a wrong model class. However, if estimations are over-conservative, the performance of the identification and the predictions decreases. In the context of health management of structures such as the Aarwangen Bridge, it would be useful to more precisely quantify the modeling errors of simplified components of the model class through experimental testing and refined modeling. Since the computational demand of very detailed FE model may be prohibitive, the systematic errors made by simplifying the modeling of a repeated component may be estimated such that an appropriate performance of the identification and the predictions is obtained. Engineering compromise between model complexity and model performance is thus supported transparently.

Nevertheless, in the presence of outlier measurements, the identification of a wrong model class may not be avoided. Future work is thus needed to include robustness with respect to such aspects in the EDMF methodology.

6. Conclusion

This paper compares the impact of common assumptions that are made for inclusion of modeling uncertainty in model-based data-interpretation tasks. Comparisons are made in terms of parameter identification and accuracy of predictions. In addition, model-class selection is studied for two data-interpretation approaches, Bayesian inference and error-domain model falsification, for three model classes describing a simple beam. This study leads to the following conclusions:

- Although estimating modeling errors is a difficult task, error-domain model falsification provides robust identification and accurate predictions in the presence of systematic errors. Since EDMF results are insensitive to changes in correlations, the correlations induced by systematic errors have no consequence.
- EDMF is able to support model-class selection by detecting erroneous model-class definitions. Although complete falsification is not assured when the model class is wrong, when it does occur, model-class selection is supported.
- Robust model classes that describe complex systems include two components: (1) unknown physical parameters that are identified using measurements; and (2) conservative modeling error estimations that cannot be represented only as uncertainties related to physical parameters.
- Common assumptions involving descriptions of modeling and measurement uncertainties by independent zero-mean Gaussian distributions may lead to inaccurate predictions in the presence of systematic modeling errors, particularly when extrapolating. In addition, through evaluating the relative plausibility of available model classes, Bayesian model class selection may not be able to uncover wrong assumptions related to model classes and modeling error estimations, and thus it may be not possible to ensure robustness of parameter identification and accuracy of extrapolations.
- EDMF involves representations of knowledge that are easy to understand. This feature is most attractive when knowledge needs to be modified as it is often the case in system identification of complex systems.

Acknowledgements

This work was funded by the Swiss National Science Foundation under Contract no. 200020-144304. The authors acknowledge Dr. J.-A. Goulet for fruitful discussions.

References

- [1] C. Argyris, P. Tsopelas, and C. Papadimitriou. Bayesian uncertainty quantification in seismically isolated structures equipped with nonlinear hysteretic devices. In J. Rodellar, A. Güemes, and F. Pozo, editors, *Proceedings of the 6th World Conference on Structural Control and Monitoring (WCSCM), Barcelona, Spain*, pages 2478–2488, 2014.
- [2] S. Atamturktur, Zh. Liu, and H. Cogan, S.and Juang. Calibration of imprecise and inaccurate numerical models considering fidelity and robustness: a multi-objective optimization-based approach. *Structural and Multidisciplinary Optimization*, pages 1–13, 2014.
- [3] J.L. Beck. Bayesian system identification based on probability logic. *Structural Control and Health Monitoring*, 17(7): 825–847, 2010.

- [4] J.L. Beck and S.-K. Au. Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. *Journal of Engineering Mechanics*, 128(4):380–391, 2002.
- [5] J.L. Beck and L.S. Katafygiotis. Updating models and their uncertainties. I: Bayesian statistical framework. *Journal of Engineering Mechanics*, 124(4):455–461, 1998.
- [6] J.L. Beck and K.-V. Yuen. Model selection using response measurements: Bayesian probabilistic approach. *Journal of Engineering Mechanics*, 130(2):192–203, 2004.
- [7] I. Behmanesh, B. Moaveni, G. Lombaert, and C. Papadimitriou. Hierarchical Bayesian model updating for structural identification. *Mechanical Systems and Signal Processing*, in press, 2015.
- [8] K. Beven. A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1-2):18–36, 2006.
- [9] K. Beven and A. Binley. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3):279–298, 1992.
- [10] J. Brynjarsdóttir and A. O’Hagan. Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30(11):114007, 2014.
- [11] J.L. Cheung, S.H. and Beck. Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters. *Journal of Engineering Mechanics*, 135(4):243–255, 2009.
- [12] S.H. Cheung and J.L. Beck. Calculation of posterior probabilities for bayesian model class assessment and averaging from posterior samples based on dynamic system data. *Computer-Aided Civil and Infrastructure Engineering*, 25(5):304–321, 2010.
- [13] M. Chiachío, J. Chiachío, G. Rus, and J.L. Beck. Predicting fatigue damage in composites: A Bayesian framework. *Structural Safety*, 51:57–68, 2014.
- [14] J. Ching and Y.-C. Chen. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *Journal of Engineering Mechanics*, 133(7):816–832, 2007.
- [15] N.C. Dubbs and F.L. Moon. Comparison and implementation of multiple model structural identification methods. *Journal of Structural Engineering*, in press:04015042, 2015.
- [16] I. Farajpour and S. Atamturktur. Error and uncertainty analysis of inexact and imprecise computer models. *Journal of Computing in Civil Engineering*, 27(4):407–418, 2012.
- [17] B. Goller and G.I. Schuëller. Investigation of model uncertainties in Bayesian structural model updating. *Journal of Sound and Vibration*, 330(25):6122–6136, 2011.
- [18] B. Goller, J.L. Beck, and G.I. Schuëller. Evidence-based identification of weighting factors in Bayesian model updating using modal data. *Journal of Engineering Mechanics*, 138:430–440, 2012.
- [19] J.-A. Goulet and I.F.C. Smith. Performance-driven measurement system design for structural identification. *Journal of Computing in Civil Engineering*, 27(4):427–436, 2013.
- [20] J.-A. Goulet and I.F.C. Smith. Structural identification with systematic errors and unknown uncertainty dependencies. *Computers & Structures*, 128:251–258, 2013.

- [21] J.-A. Goulet, S. Coutu, and I.F.C. Smith. Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. *Advanced Engineering Informatics*, 27(2):261–269, 2013.
- [22] J.-A. Goulet, C. Michel, and I.F.C. Smith. Hybrid probabilities and error-domain structural identification using ambient vibration monitoring. *Mechanical Systems and Signal Processing*, 37(1-2):199–212, 2013.
- [23] J.-A. Goulet, M. Texier, C. Michel, I.F.C. Smith, and L. Chouinard. Quantifying the effects of modeling simplifications for structural identification of bridges. *Journal of Bridge Engineering*, 19(1):59–71, 2014.
- [24] F. Hemez, H.S. Atamturktur, and C. Unal. Defining predictive maturity for validated numerical simulations. *Computers & Structures*, 88:497–505, 2010.
- [25] JCGM. *Evaluation of measurement data – Supplement 2 to the “Guide to the expression of uncertainty in measurement” – Extension to any number of output quantities*, volume JCGM 102:2011. JCGM Working Group of the Expression of Uncertainty in Measurement, Sèvres, France, 2011.
- [26] M.C. Kennedy and A. O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.
- [27] H.F. Lam, H.Y. Peng, and S.K. Au. Development of a practical algorithm for Bayesian model updating of a coupled slab system utilizing field test data. *Engineering Structures*, 79:182–194, 2014.
- [28] D.J.C. Mackay. *Information theory, inference and learning algorithms*. Cambridge University Press, 2003.
- [29] G. Moser and I.F.C. Smith. Detecting leak regions through model falsification. In *20th International Workshop: Intelligent Computing in Engineering 2013, Vienna, Austria*, 2013.
- [30] M. Muto and J.L. Beck. Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *Journal of Vibration and Control*, 14(1-2):7–34, 2008.
- [31] M.B. Neumann and W. Gujer. Underestimation of uncertainty in statistical regression of environmental models: influence of model structure uncertainty. *Environmental Science & Technology*, 42(11):4037–4043, 2008.
- [32] C. Papadimitriou, J.L. Beck, and L.S. Katafygiotis. Updating robust reliability using structural test data. *Probabilistic Engineering Mechanics*, 16(2):103–113, 2001.
- [33] M. Papadopoulou, B. Raphael, I.F.C. Smith, and C. Sekhar. Hierarchical sensor placement using joint entropy and the effect of modeling error. *Entropy*, 16(9):5078–5101, 2014.
- [34] R. Pasquier and I.F.C. Smith. Sources and forms of modelling uncertainties for structural identification. In *Proceedings of 7th International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII), Torino, IT*, page in press, 2015.
- [35] R. Pasquier, J.-A. Goulet, C. Acevedo, and I.F.C. Smith. Improving fatigue evaluations of structures using in-service behavior measurement data. *Journal of Bridge Engineering*, 19 (11):04014045, 2014.
- [36] R. Pasquier, J.-A. Goulet, and I.F.C. Smith. Model-based data interpretation and diagnosis robustness. In Deodatis, Ellingwood, and Frangopol, editors, *Safety, Reliability, Risk and Life-Cycle Performance of Structures & Infrastructures. Proceedings of 11th International Conference on Structural Safety & Reliability (ICOSSAR), New York, USA*, pages 2497–2504, 2014.

- [37] K.R. Popper. *The logic of scientific discovery*. 3rd ed. New York, Routledge, 2002.
- [38] Y. Robert-Nicoud, B. Raphael, and I.F.C. Smith. Configuration of measurement systems using Shannon’s entropy function. *Computers & structures*, 83(8):599–612, 2005.
- [39] Z. Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967.
- [40] E. Simoen, C. Papadimitriou, and G. Lombaert. On prediction error correlation in Bayesian model updating. *Journal of Sound and Vibration*, 332(18):4136–4152, 2013.
- [41] D. Straub and I. Papaioannou. Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics*, 141(3):04014134, 2014.
- [42] A. Tarantola. Popper, Bayes and the inverse problem. *Nature Physics*, 2(8):492–494, 2006.
- [43] D.G. Vernay, B. Raphael, and I.F.C. Smith. Augmenting simulations of airflow around buildings using field measurements. *Advanced Engineering Informatics*, 28(4):412–424, 2014.
- [44] K. Worden and J. J. Hensman. Parameter estimation and model selection for a class of hysteretic systems using bayesian inference. *Mechanical Systems and Signal Processing*, 32:153–169, 2012.
- [45] K.-V. Yuen. *Bayesian methods for structural dynamics and civil engineering*. Wiley, 2010.
- [46] K.-V. Yuen and J.L. Beck. Reliability-based robust control for uncertain dynamical systems using feedback of incomplete noisy response measurements. *Earthquake Engineering & Structural Dynamics*, 32(5):751–770, 2003.
- [47] J. Zhang, C. Wan, and T. Sato. Advanced Markov chain Monte Carlo approach for finite element calibration under uncertainty. *Computer-Aided Civil and Infrastructure Engineering*, 28(7):522–530, 2013.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

