Advanced Engineering Informatics 29 (2015) 1096–1109

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

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

Romain Pasquier *, Ian F.C. Smith

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

ARTICLE INFO

Article history:
Received 21 October 2014
Received in revised form 8 July 2015
Accepted 30 July 2015
Available online 15 August 2015

Keywords:
Systematic error
Extrapolation
Modeling uncertainties
Bayesian inference
Model falsification
Model class selection

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 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.

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

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

E-mail address: rpasquie@gmail.com (R. Pasquier).

http://dx.doi.org/10.1016/j.aei.2015.07.007
1474-0346/

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 weighting 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 $\mathbf{y} = [y_1, \ldots, y_{n_m}]^T$ 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 $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_{n_p}]^T$ is a vector of n_p parameters having uncertain values and defined on the parameter domain $\boldsymbol{\Theta} \subseteq \mathbb{R}^{n_p}$. The inference of the parameter values of the model class G is based on Bayes' Theorem of conditional probability:

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

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