Hierarchical Bayes for the Explicit Estimation of Model Prediction Errors

James-A. Goulet
Assistant professor, Dept. of Civil, Geological and Mining Engineering, Polytechnique Montreal

ABSTRACT: An extensive research effort is dedicated to Bayesian estimation methods for analyzing the empirical behaviour of structures. State-of-the-art structural identification methods currently quantify model uncertainties by estimating hyper-parameters for the prediction-error prior. This paper exposes that this uncertainty quantification procedure does not fully recognize the *epistemic* nature of model prediction errors, because their posterior probability density function (PDF) is not explicitly estimated and their interaction with model parameters are not considered. This paper presents a Hierarchical Bayes formulation for estimating the joint posterior PDF of model parameters and prediction errors. This Hierarchical Bayes approach allows capturing the dependencies between unknown model parameters and unknown prediction errors; it offers a more accurate picture of the structural behaviour than when estimating the prior hyper-parameters alone. The application of this method to large-scale structures requires an adequate model the for the prediction-error prior, which remains a case-specific challenge.

An extensive research effort is dedicated to Bayesian estimation methods for analyzing the empirical behaviour of structures. One of the goals of structural identification is to infer model parameters from on-site observations. Another is to employ the updated knowledge about parameters to make better predictions for unobserved quantities. Since the work of Beck and Katafygiotis (1998) in the field of structural identification, Bayesian estimation has been increasing in popularity for the tasks of inferring the properties of structures (Zhang et al. (2011); Papadimitriou et al. (2001); Au and Zhang (2016)), and for detecting damages in the context of structural health monitoring (Yuen et al. (2006); Simoen et al. (2012)). More recently the focus has shifted to *Hierarchical Bayes* formulations (Behmanesh et al. (2015); Huang et al. (2017); Nagel and Sudret (2016)), where Bayesian estimation is employed to identify the joint posterior probability density function (PDF) for model parameters and for the hyper-parameters of the prediction-errors prior PDF. A *prediction error* is defined as the discrepancy between a model prediction and the unknown true system's response. One common aspect of current *Hierarchical Bayes* formulations found in the literature is that prediction errors are quantified through the estimation of the hyper-parameters of the prediction-error prior PDF. This approach is limited because prediction errors are not random; they are unknown deterministic values.

Similar Bayesian estimation approaches were proposed by Brynjarsdóttir and O'Hagan (2014) and by Ling et al. (2014), where the posterior PDF for prediction errors are explicitly estimated. The field of structural identification can build on these approaches for improving the way model parameters and prediction errors are estimated. For this task, two aspects remain to be addressed: (1) the formalization of the approach in a hierarchical manner and (2) the explicit separation of prediction errors into its prior and posterior estimates.

Authors such as Simoen et al. (2013a,b) and Goulet et al. (2014) have already demonstrated that model simplifications introduce dependencies in the model prediction errors between different prediction locations. Simoen et al. (2013a) underline the need to include these dependencies in Bayesian estimation. Current research in structural identification does not fully recognize the *epistemic* nature of prediction errors, because its posterior PDF is not explicitly estimated and its interaction with model parameters is not considered.

This paper presents a Hierarchical Bayes formulation for estimating the joint posterior PDF of model parameters and prediction errors. Section 1 introduces the nomenclature for the prior, posterior, hyper-prior, and hyper-posterior employed in the Hierarchical Bayes formulation. Section 2 describes the common Hierarchical Bayes formulation employed in structural identification. Section 3 presents the new hierarchical formulation proposed in this paper that allows the joint estimation of model parameters and prediction errors. Finally, in Section 4, an example illustrates the benefits obtained by estimating the joint posterior of model parameters and prediction errors.

1. Nomenclature for the prior, posterior, hyper-prior, and hyper-posterior

The observed data is separated into two parts, the attributes x_i, and the corresponding structural response observations y_i. For the sake of simplicity, in this paper, the attributes x_i only refer to *observation locations*. The joint set of attributes and system responses is $\mathscr{D} = \{(x_i, y_i), \forall i = 1 : \mathtt{D}\} \equiv \{(\mathscr{D}_x, \mathscr{D}_y)\}$. The variables $\mathbf{y} = [y_1 : y_{\mathtt{D}}]^\intercal \in \mathbb{R}^{\mathtt{D}}$ describe the *observed structural responses*. These observations are modelled by the sum of *model predictions* $\mathbf{g}(\mathbf{p}, \mathbf{x}) \in \mathbb{R}^{\mathtt{D}}$ that are a function of attributes $\mathbf{x} = [x_1 : x_{\mathtt{D}}]^\intercal \in \mathbb{R}^{\mathtt{D}}$ and *model parameters* $\mathbf{p} = [p_1 : p_{\mathtt{P}}]^\intercal \in \mathbb{R}^{\mathtt{P}}$, *prediction errors* $\mathbf{w}(\mathbf{x}) = [w(x_1) : w(x_{\mathtt{D}})]^\intercal \in \mathbb{R}^{\mathtt{D}}$ that are also a function of attributes $\mathbf{x}$, and *measurement errors* $\mathbf{v} = [v_1 : v_{\mathtt{D}}]^\intercal \in \mathbb{R}^{\mathtt{D}}$. Note that contrarily to model prediction and prediction errors, measurements errors are assumed to be independent of the attribute $\mathbf{x}$. The model prediction $\mathbf{g}(\mathbf{p}, \mathbf{x})$ depends on the known attributes $\mathbf{x}$ describing the prediction location and unknown model parameters $\mathbf{p}$ describing physical properties of the system. The observed structural responses are described by the equation

$$\mathbf{y} = \mathbf{g}(\mathbf{p}, \mathbf{x}) + \mathbf{w}(\mathbf{x}) + \mathbf{v}. \tag{1}$$

All the terms in Equation 1 are considered as deterministic quantities because for a given set of observations $\mathscr{D}$, quantities $\mathbf{x}$, $\mathbf{y}$, $\mathbf{p}$, $\mathbf{w}(\mathbf{x})$, and $\mathbf{v}$ are not varying, yet values for $\mathbf{p}$, $\mathbf{w}(\mathbf{x})$, and $\mathbf{v}$ remain unknown.

The role of Bayesian probabilities is to describe our knowledge of unknown variables using a *probability density function* (PDF). Before obtaining observations $\mathscr{D}$, our knowledge is described by the *prior* PDF; after obtaining $\mathscr{D}$, our knowledge is described by the *posterior* PDF. Our prior knowledge is described by a PDF, which is itself defined by a set of unknown parameters. For example, the prior knowledge for model parameters is described by the random variable $\mathbf{P}$, which is described by the prior PDF $f(\mathbf{p}|\mathbf{z}_p)$. "$|\mathbf{z}_p$" denotes the dependence on a set of *hyper-parameters* i.e. parameters of the prior PDF. Similarly, $f(\mathbf{p}|\mathscr{D}, \mathbf{z}_p)$ describes our posterior knowledge for $\mathbf{P}$ conditioned on the set of observations $\mathscr{D}$ and deterministic hyper-parameter values $\mathbf{z}_p$. When hyper-parameters $\mathbf{z}_p$ are unknown, the random variable $\mathbf{Z}_p$ is described by a *hyper-prior* $f(\mathbf{z}_p)$ and a *hyper-posterior* $f(\mathbf{z}_p|\mathscr{D})$, i.e. the prior PDF and posterior PDF of hyper-parameters. Unlike the parameter values $\mathbf{p}$ for which true values may exist, hyper-parameters $\mathbf{z}_p$ are parameters of our prior knowledge for which no true value exists. When hyper-parameters $\mathbf{z}_p$ are assumed to be fixed constants that are not learnt from data, $f(\mathbf{p}|\mathbf{z}_p) = f(\mathbf{p})$.

The concept of *prior*, *posterior*, *hyper-prior* and *hyper-posterior* not only applies to unknown parameters $\mathbf{p}$, but also for prediction errors $\mathbf{w}(\mathbf{x})$, and measurement errors $\mathbf{v}$. Note that, unlike for hyper-parameters $\mathbf{z}_p$ and $\mathbf{z}_w$, a true value exists for $\mathbf{z}_v$. The true value for $\mathbf{z}_v$ typically corresponds to the statistical precision of the measuring instruments employed to obtain $\mathbf{y}$.

In practical situations, it is common that *uniform* or *non-informative* prior PDFs are employed Gelman et al. (2014). In that case, the prior does not depend on any hyper-parameters and it removes the need for a hyper-prior and hyper-posterior. Similarly, if the hyper-parameter values are assumed to be known, it removes the need for a hyper-prior and hyper-posterior. In this paper, we assume that $\mathbf{z}_p$ are known constants; It simplifies the notation and allows to focus the attention on model parameters $\mathbf{p}$ and prediction errors $\mathbf{w}(\mathbf{x})$.

2. Hierarchical Bayes for Structural Identification

The central idea of Hierarchical Bayes is to infer simultaneously the joint posterior for model parameters which are included in the likelihood function, as well as the hyper-parameters for the prior distributions Murphy (2012). Given the hypothesis that model ($\mathbf{W}(\mathbf{x})$) and measurement ($\mathbf{V}$) errors are zero-mean Gaussians,

$$\begin{aligned}\mathbf{W}(\mathbf{x}) &\sim \mathcal{N}(\mathbf{w};\mathbf{0},\mathbf{S}_w(\mathbf{z}_w(\mathbf{x})))\\ \mathbf{V} &\sim \mathcal{N}(\mathbf{v};\mathbf{0},\mathbf{S}_v(\mathbf{z}_v))\end{aligned}$$

the likelihood of observations $\mathbf{y}$ conditioned on attributes $\mathbf{x}$, model parameters $\mathbf{p}$, and model and measurement error hyper-parameters $\mathbf{z}_w(\mathbf{x}),\mathbf{z}_v$, is described by

$$\begin{aligned}f(\mathbf{y}|\mathbf{p},\mathbf{z}_w,\mathbf{z}_v,\mathbf{x}) &\sim \mathbf{g}(\mathbf{p},\mathbf{x})+\mathbf{W}(\mathbf{x})+\mathbf{V}\\ &= \mathcal{N}(\mathbf{y};\mathbf{g}(\mathbf{p},\mathbf{x}),\mathbf{S}_w(\mathbf{z}_w(\mathbf{x}))+\mathbf{S}_v(\mathbf{z}_v))\end{aligned} \tag{2}$$

so that the likelihood of the set of observations $\mathcal{D}_y$ is

$$\begin{aligned}&f(\mathcal{D}_y|\mathbf{p},\mathbf{z}_w,\mathbf{z}_v,\mathcal{D}_x) =\\ &\quad\mathcal{N}(\mathcal{D}_y;\mathbf{g}(\mathbf{p},\mathcal{D}_x),\mathbf{S}_w(\mathbf{z}_w(\mathcal{D}_x))+\mathbf{S}_v(\mathbf{z}_v)).\end{aligned}$$

In a generic form, the model prediction error covariance matrix is parameterized by $\mathbf{z}_w(\mathbf{x})$ so that

$$\begin{aligned}&[\mathbf{S}_w(\mathbf{z}_w(\mathbf{x}))]_{ij} = s_w(\mathbf{x}_i)\cdot s_w(\mathbf{x}_j)\cdot\rho(\mathbf{x}_i,\mathbf{x}_j),\text{ where}\\ &\mathbf{z}_w(\mathbf{x}_i) = [s_w(\mathbf{x}_i),\rho(\mathbf{x}_i,\mathbf{x}_j)]^\intercal,\ \forall\{i\wedge ij : i>j\}\end{aligned}$$

where $s_w(\mathbf{x}_i)$ are the prediction errors standard deviations for an attribute $\mathbf{x}_i$ and $\rho(\mathbf{x}_i,\mathbf{x}_j)$ the correlation coefficients defined for a pair of attributes $\mathbf{x}_i,\mathbf{x}_j$. Observation errors are typically independent from one to another, so the covariance matrix $\mathbf{S}_v(\mathbf{z}_v)=\mathbf{s}_v\cdot\mathbf{I}$, which is parametrized by $\mathbf{z}_v=[s_{v,1}:s_{v,\mathtt{D}}]^\intercal$ and is diagonal. Given the hypothesis that the prior knowledge for parameters $\mathbf{p}$, $\mathbf{z}_w(\mathbf{x})\equiv\mathbf{z}_w$, and $\mathbf{z}_v$ are independent from each other, their joint posterior PDF is obtained using the Bayes conditional probability

$$\overbrace{f(\mathbf{p},\mathbf{z}_w,\mathbf{z}_v|\mathcal{D})}^{\text{posterior}} = \frac{\overbrace{f(\mathcal{D}_y|\mathbf{p},\mathbf{z}_w,\mathbf{z}_v,\mathcal{D}_x)}^{\text{likelihood}}\cdot\overbrace{f(\mathbf{p})}^{\text{prior}}\cdot\overbrace{f(\mathbf{z}_w)\cdot f(\mathbf{z}_v)}^{\text{hyper-prior}}}{\underbrace{f(\mathcal{D}_y)}_{\text{normalization cte.}}}. \tag{3}$$

Note that $f(\mathbf{p})$ describes the prior PDF for model parameters and $f(\mathbf{z}_w)\cdot f(\mathbf{z}_v)$, the joint hyper-prior PDF. $f(\mathbf{p},\mathbf{z}_w,\mathbf{z}_v|\mathcal{D})$ is the joint posterior PDF for the model parameters, the prediction-errors hyper-parameters and measurements-errors hyper-parameters.

The main limitation of this Hierarchical Bayes approach is that it only allows estimating the posterior for model parameters, as well as for hyper-parameters. Hyper-parameters are parameters of the prior, consequently, they provide no information about the posterior distribution of the actual prediction errors or measurement errors; it only provides information about our prior knowledge. Estimating the posterior PDF for model prediction errors is essential because there are non-linear dependencies between model parameter values and prediction errors. If we consider prediction errors as *epistemic* uncertainties, we should explicitly estimate the posterior PDF for these quantities. The formulation suited for this task is presented in the following section.

3. Joint Estimation of Model Parameters and Prediction Errors

Let us consider that prediction errors $\mathbf{w}(\mathbf{x})$ are unknown constants, as describes in Section 1. In that case, we estimate the posterior PDF of $\mathbf{w}(\mathbf{x})$ conditioned on data by reformulating the likelihood defined in the previous section. Here, the likelihood function explicitly includes the prediction errors term $\mathbf{w}(\mathbf{x})$ so that

$$\begin{aligned}f(\mathbf{y}|\mathbf{p},\mathbf{z}_v,\mathbf{w}(\mathbf{x}),\mathbf{x}) &\sim \mathbf{g}(\mathbf{p},\mathbf{x})+\mathbf{w}(\mathbf{x})+\mathbf{V}\\ &= \mathcal{N}(\mathbf{y};\mathbf{g}(\mathbf{p},\mathbf{x})+\mathbf{w}(\mathbf{x}),\mathbf{S}_v(\mathbf{z}_v)).\end{aligned} \tag{4}$$

Notice that contrarily to Equation 2, the likelihood in Equation 4 now explicitly depends on the prediction errors $\mathbf{w}(\mathbf{x})$. Assuming that observation errors follow zero-mean Gaussians, the likelihood of data $\mathcal{D}_y$ conditioned on all parameters and hyper-parameters is

$$f(\mathscr{D}_y|\mathbf{p},\mathbf{z}_v,\mathbf{w}(\mathscr{D}_x),\mathscr{D}_x) = \\ \mathscr{N}(\mathscr{D}_y;\mathbf{g}(\mathbf{p},\mathscr{D}_x)+\mathbf{w}(\mathscr{D}_x),\mathbf{S}_v(\mathbf{z}_v)).$$

The formulation for the posterior PDF now has a prior term for the prediction-errors conditional on its hyper-parameters so that

$$\overbrace{f(\mathbf{p},\mathbf{z}_w,\mathbf{z}_v,\mathbf{w}(\mathscr{D}_x)|\mathscr{D})}^{\text{posterior}} = \frac{\overbrace{f(\mathscr{D}_y|\mathbf{p},\mathbf{z}_v,\mathbf{w}(\mathscr{D}_x),\mathscr{D}_x)}^{\text{likelihood}} \cdot \overbrace{f(\mathbf{w}(\mathscr{D}_x)|\mathbf{z}_w)\cdot f(\mathbf{p})}^{\text{prior}} \cdot \overbrace{f(\mathbf{z}_w)\cdot f(\mathbf{z}_v)}^{\text{hyper-prior}}}{f(\mathscr{D}_y)} \quad (5)$$

where $f(\mathbf{p},\mathbf{z}_w,\mathbf{z}_v,\mathbf{w}(\mathscr{D}_x)|\mathscr{D})$ is the joint posterior distribution for the model parameters, the hyper-parameters, and most importantly *prediction errors* $\mathbf{w}(\mathscr{D}_x)$. This new formulation allows the explicit quantification of the dependence between model parameters and model prediction errors at observed locations. In the case where several sets of model parameters and prediction errors $\{\mathbf{p},\mathbf{w}(\mathscr{D}_x)\}$ can equally explain the observations, then all these sets of values will end up having an equal posterior probability, given that the prior probability of each set is also equal.

The key aspect for the successful application of this Hierarchical Bayes formulation is to have a proper *model for the model prediction-errors prior PDF*, $f(\mathbf{w}(\mathscr{D}_x)|\mathbf{z}_w) \sim \mathbf{W}(\mathbf{z}_w(\mathbf{x}))$. A convenient choice already employed by many authors, e.g. Simoen et al. (2013a,b); Behmanesh et al. (2015); Huang et al. (2017); Papaioannou and Straub (2017), is to describe the prior for prediction errors using a Gaussian process. Under the assumption that our model is unbiased, i.e. the prior expected value for prediction errors is $\mathbf{0}$, this Gaussian process is expressed as

$$\begin{aligned} \mathbf{W}(\mathbf{z}_w(\mathbf{x})) &\sim \mathscr{N}(\mathbf{w};\mathbf{0},\mathbf{S}_w(\mathbf{z}_w(\mathbf{x}))), \text{ where} \\ [\mathbf{S}_w(\mathbf{z}_w(\mathbf{x}))]_{ij} &= s_w(x_i)\cdot s_w(x_j)\cdot\rho(x_i,x_j,l), \text{ and} \\ s_w(x) &= \text{fct}(x,\mathbf{z}_s), \\ \rho(x_i,x_j,l) &= \exp\left(-\frac{1}{2}\frac{(x_i-x_j)^2}{l^2}\right),\ l\geq 0, \\ \mathbf{z}_w &= [\mathbf{z}_s,l]^\intercal. \end{aligned}$$

The prediction error standard deviation $s_w(x)$ is typically attribute-dependent, so that it needs to be represented by a problem-specific function $\text{fct}(x,\mathbf{z}_s)$, where $\mathbf{z}_s$ are the function's parameters. The correlation structure can be represented by a square-exponential covariance function, parameterized by the length-scale factor l. l is the hyper-parameter describing how the correlation decreases as the distance $(x_i-x_j)^2$ increases. Note that this choice for the correlation structure is not exclusive and many others can be employed as described by Rasmussen and Williams (2006). The hyper-parameters in $\mathbf{z}_w$ needs to be learnt from data.

In order to shorten the notation, the posterior PDF in Equation 5 is summarized as $f(\mathbf{p},\mathbf{z}_w,\mathbf{z}_v,\mathbf{w}(\mathscr{D}_x)|\mathscr{D}) \sim \boldsymbol{\theta}|\mathscr{D}$. Here, the posterior PDF for model prediction errors $\mathbf{w}(\mathscr{D}_x)$ is estimated only for measured locations $\mathscr{D}_x$. However, in practical applications, one typically needs to predict the structure behaviour u_i at unobserved locations x_i so that $\mathscr{P} = \{(x_i,u_i)\ ,\forall i=1:\mathtt{P}\} \equiv \{(\mathscr{P}_x,\mathscr{P}_u)\}$. The structure behaviour at unobserved locations is expressed as a random process following

$$\{\mathbf{U}|\mathscr{D},\mathscr{P}_x\} = \mathbf{g}(\mathscr{P}_x,\{\boldsymbol{\theta}|\mathscr{D}\}) + \mathbf{W}(\mathscr{P}_x,\{\boldsymbol{\theta}|\mathscr{D}\}) \quad (6)$$

where $\mathbf{W}(\cdot)$ is a Gaussian process defined for prediction locations $\mathscr{P}_x$ conditioned on $\mathscr{D}$. The advantage of modelling the prior PDF of prediction errors as a zero-mean Gaussian process is that its conditional distribution is also Gaussian so that

$$\begin{aligned} \mathbf{W}(\mathscr{P}_x,\{\boldsymbol{\theta}|\mathscr{D}\}) &\sim \mathscr{N}(\mathbf{w}(\mathscr{P}_x);\mathbf{M}_{\mathscr{P}|\mathscr{D}},\mathbf{S}_{\mathscr{P}|\mathscr{D}}), \\ \mathbf{M}_{\mathscr{P}|\mathscr{D}} &= \mathbf{S}_{\mathscr{P}\mathscr{D}}\mathbf{S}_{\mathscr{D}\mathscr{D}}^{-1}(\mathbf{w}(\mathscr{D}_x)|\mathscr{D}) \\ \mathbf{S}_{\mathscr{P}|\mathscr{D}} &= \mathbf{S}_{\mathscr{P}\mathscr{P}} - \mathbf{S}_{\mathscr{P}\mathscr{D}}\mathbf{S}_{\mathscr{D}\mathscr{D}}^{-1}\mathbf{S}_{\mathscr{P}\mathscr{D}}^\intercal \end{aligned} \quad (7)$$

and where $\mathbf{S}_{\mathscr{P}\mathscr{P}}$ is the covariance matrix for prediction errors at unobserved locations, $\mathbf{S}_{\mathscr{D}\mathscr{D}}$ is the covariance matrix for prediction errors at observed locations, and $\mathbf{S}_{\mathscr{P}\mathscr{D}}$ is the covariance matrix between prediction errors at observed and unobserved locations. Specifically, these matrices are defined following

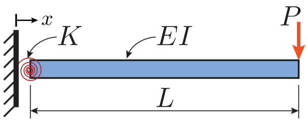
$$\begin{aligned} \mathbf{S}_{\mathscr{P}\mathscr{P}} &= \mathbf{S}_w(\{\mathbf{Z}_w|\mathscr{D}\},\mathscr{P}_x) \\ \mathbf{S}_{\mathscr{P}\mathscr{D}} &= \mathbf{S}_w(\{\mathbf{Z}_w|\mathscr{D}\},\{\mathscr{P}_x,\mathscr{D}_x\}) \\ \mathbf{S}_{\mathscr{D}\mathscr{D}} &= \mathbf{S}_w(\{\mathbf{Z}_w|\mathscr{D}\},\mathscr{D}_x) + \mathbf{s}_v\cdot\mathbf{I} \end{aligned} \quad (8)$$

where the observation uncertainties ($\mathbf{s}_v$) are only included on the diagonal of the $\mathbf{S}_{\mathscr{D}\mathscr{D}}$ matrix.

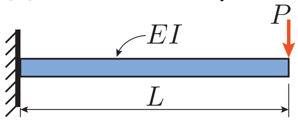
In practical applications, the posterior PDF in Eq.5 as well as the posterior PDF for prediction errors at unobserved locations in Eq.7 are not analytically tractable. A common solution to this challenge is to employ MCMC sampling methods to approximate the joint posterior Gelman et al. (2014). In this case, once samples from the posterior PDF $f(\boldsymbol{\theta}|\mathscr{D})$ are obtained, they are employed to generate realizations of the matrices in Eq.8 for all prediction locations $\mathscr{P}_x$. These matrices are then employed to generate realizations from posterior PDF for model prediction errors $\mathbf{W}(\cdot)$ in Eq.7, again, for all prediction locations $\mathscr{P}_x$.

4. Illustrative Example

This Section presents an example illustrating the potential of the hierarchical approach presented in this paper, for the field of structural identification. The structure considered is a cantilever beam. The real structure is made of an elastic beam connected to a rotational spring as presented in Figure 4a. In order to represent practical cases, simulated data is generated from the beam in Figure 4a, yet, data is interpreted using the simplified model in Figure 4b, where the rotational spring is omitted.



(a) Real structural system



(b) Simplified model

Figure 1: (a) presents the real structural system employed to generate simulated observations at location x. (b) presents the simplified model where the fixed end is assumed to be fully rigid. This simplifying assumption is introduced to represent practical situations where models contains simplifications in comparison with the system studied.

This model simplification introduces dependencies in the model prediction errors. The prior PDF for prediction errors is defined by the standard deviation function $s_w(x) = a \cdot x$, where $a : a > 0$ is a hyper-parameter to be learnt jointly with the correlation length, so that $\mathbf{z}_w = [a, l]^\intercal$. The true parameter values employed for generating simulated observations are

$$\begin{aligned} K &= 1.75 \times 10^{11}\, N/rad \\ L &= 10\, m \\ P &= 5\, kN \\ I &= 6.66 \times 10^{9}\, mm^4 \\ E &= 35\, GPa \end{aligned}$$

Measurement standard deviation is assumed to be known and is equal to $s_v = 1$ mm. Observation locations are $\mathscr{D}_x = \{5, 10\}\, m$ and observed displacements are $\mathscr{D}_y = \{-4.44, -11.28\}\, mm$. Finally, predicted displacements are sought for locations $\mathscr{P}_x = \{0, 0.2, \cdots, 10\}\, m$. The prior engineering knowledge for the model parameter and hyper-parameters are described by Gaussian PDFs truncated at 0 in order to respect the constrains on the physically possible values for E, a and l. The choice of Gaussian priors for E, a and l is made in order to facilitate the estimation of the proposal PDF using the Laplace approximation (Murphy (2012)). Table 1 summarizes the prior knowledge PDFs.

Table 1: Prior PDFs for parameters and hyper-parameters.

	Description	Prior/Hyper-prior
E	Young's modulus	$\mathscr{N}(E; 30, 15^2)$
a	Pred. error prior scale	$\mathscr{N}(a; 10^{-4}, 10^{-3})$
l	Correlation length	$\mathscr{N}(l; 5, 50)$
$\mathbf{w}(\mathbf{x})$	Prediction errors	$\mathscr{N}(\mathbf{w}; \mathbf{0}, \mathbf{S}_w(a, l))$

This illustrative example compares the results obtained using (1) the current Hierarchical Bayesian (Current HB) approach presented in Section 2 and using (2) the approach presented in Section 3 which estimates the joint posterior for model parameters, prediction errors and hyper-parameters (New HB). Both approaches employ the same prior structure except that the Current HB approach estimates $f(\mathbf{p}, \mathbf{z}_w, \mathbf{z}_v|\mathscr{D})$ whereas the New HB approach proposed in this paper estimates $f(\mathbf{p}, \mathbf{z}_w, \mathbf{z}_v, \mathbf{w}(\mathscr{D}_x)|\mathscr{D})$. Displacement predictions $\{\mathbf{U}|\mathscr{D}, \mathscr{P}_x\}$ for unobserved locations $\mathscr{P}_x$, are computed following Equation 6. For the New HB approach, prediction errors

$$\mathbf{W}(\mathscr{P}_x, \{\boldsymbol{\theta}|\mathscr{D}\}) \sim \mathscr{N}(\mathbf{w}(\mathscr{P}_x); \mathbf{M}_{\mathscr{P}|\mathscr{D}}, \mathbf{S}_{\mathscr{P}|\mathscr{D}}) \quad (9)$$

correspond to the posterior prediction errors described in Equation 7. For the Current HB approach, because the posterior PDF for prediction errors is not explicitly estimated, prediction errors are computed from the prior PDF so that

$$\mathbf{W}(\mathscr{P}_x, \{\boldsymbol{\theta}|\mathscr{D}\}) \sim \mathscr{N}(\mathbf{w}(\mathscr{P}_x); \mathbf{0}, \mathbf{S}_{\mathscr{P}\mathscr{P}}) \quad (10)$$

where the covariance matrix $\mathbf{S}_{\mathscr{P}\mathscr{P}}$ is estimated using the posterior PDF of hyper-parameters $\{\mathbf{Z}_w|\mathscr{D}\}$.

Joint samples $\boldsymbol{\theta}_i|\mathscr{D}$, $\forall i = 1:\mathtt{N}$ from the joint posterior PDF are taken using the Metropolis-Hasting algorithm where the proposal PDF is a multivariate Gaussian for which the covariance matrix is estimated using the Laplace approximation as described in Murphy (2012). Three parallel chains, containing a total of $\mathtt{N} = 10^5$ joint samples are taken. For each chain, the first 15 000 samples are discarded as burn-in samples; The Metropolis-Hasting acceptance rate is approximately 0.25 and the Estimated Potential Scale Reduction (EPSR) (i.e. chain stationarity metric), is below 1.005 for all parameters and hyper-parameters (Murphy (2012)). Both the acceptance rate and EPSR are within target ranges for ensuring the sampling efficiency and chains stationarity (Murphy (2012); Gelman et al. (2014)).

Figure 2 and 3 describe the posterior PDFs computed using Current HB and New HB methods respectively. Histograms located on the matrices diagonal describe the marginal posterior probability for each parameter or hyper-parameter. For the parameter E and prediction errors w_1, w_2, the solid dot represents the true value. Note that a true value does not exist for the prediction error hyper-parameters a and l. The scatter plots above the diagonal represent the pairwise posterior samples. Note that only a random set of 2 000 joint samples is presented.

For both methods, the marginal posterior for parameters E and hyper-parameters $\{a, l\}$ are identical within the range of sampling variability. However, the New HB method provides additional information compared with the Current HB; it allows to explicitly estimate the posterior distribution for prediction errors. In Figure 3, scatter plots for E–w_1 and E–w_2 display the non-linear dependence between model prediction errors and the Elastic modulus. Here, large (small) values for E leads to stiffer (softer)-than-reality models which need to be compensated by negative (positive) prediction error values. Also, the length-scale factor l has in most cases a value much greater than the length of the beam. It indicates that model prediction errors are almost linearly correlated. This is in agreement with the almost linear correlation observed between model prediction errors w_1–w_2.

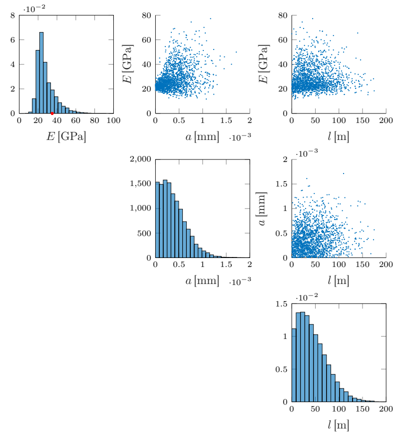


Figure 2: Posterior PDF representation for the Current Hierarchical Bayes approach (Current HB). Histograms on the diagonal represent the marginal posterior probabilities for each parameter. Each scatter plot above the diagonal represents the pairwise representation of MCMC samples. Here only a random set of 2 000 joint samples is presented. In the case of parameters E, the solid dot represents the true value.

Figure 4a and 4b present the true, observed and predicted beam deflexion obtained with Current HB and New HB respectively. The solid line represents the true deflexion and the dashed line the expected value of the predicted deflexion. Crosses represent the observed values and the coloured regions represent the ±2 standard deviation intervals for the prior and posterior prediction errors respectively. The posterior prediction error obtained with the New HB method is significantly smaller than the one obtained with the Current HB method. This is caused by the dependency between prediction errors and model parameters observed in Figure 3, where the effect of parameter values on displacements are compensated by prediction errors. The Current HB method overestimate the prediction er-

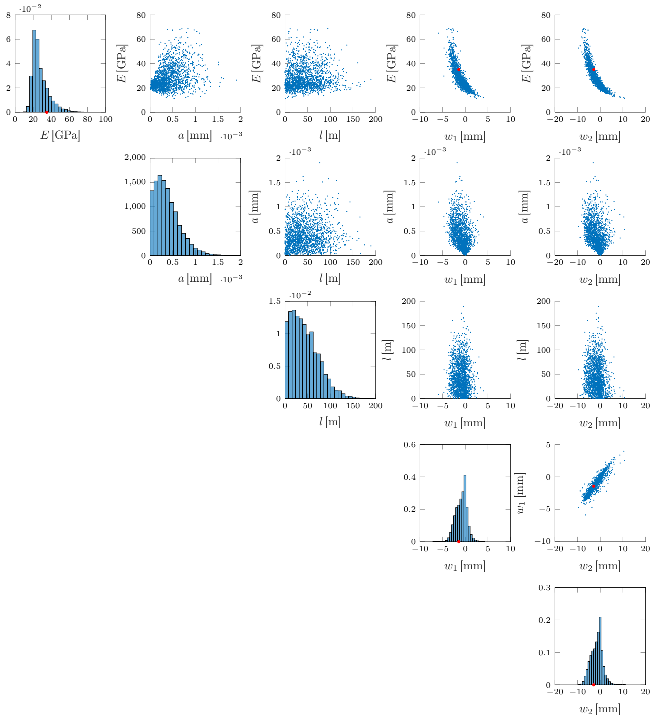
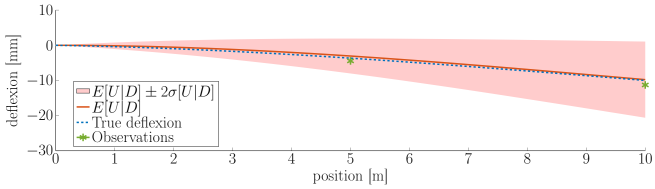
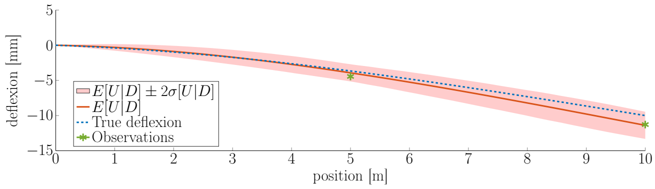


Figure 3: Posterior PDF representation for the Hierarchical Bayes approach proposed in this paper (New HB). Histograms on the diagonal represent the marginal posterior probabilities for each parameter. Each scatter plot above the diagonal represents the pairwise representation of MCMC samples. Here only a random set of 2 000 joint samples is presented. In the case of parameters E, w_1 and w_2, the solid dot represents the true value.



(a) Current Hierarchical Bayes (HB) method



(b) New Hierarchical Bayes (HB) method

Figure 4: True, observed and predicted beam deflexion. The dashed line represents the true beam deflexion and the solid line represents the expected beam deflexion conditional on observations depicted by crosses. The coloured region represents the $\pm 2\sigma$ confidence interval for the predicted displacement estimated.

ror because it only considers its prior PDF. It shows that the Current HB method can lead to an important overestimation of prediction errors because it does not explicitly consider the joint posterior distribution of model parameters and prediction errors.

For the existing HB method, MCMC sampling takes approximately five minutes, for the new HB method, it takes three times as much because tuning the proposal parameters according to the method presented by Murphy (2012) takes several steps. For more complex problems involving hundreds of parameters, more advanced sampling approaches such as Hamiltonian Monte-Carlo (Neal et al. (2011)) must be employed.

5. Discussion

The Hierarchical Bayes approach presented in this paper allows estimating the joint posterior probability for model prediction errors as well as model parameters. It enables taking advantage of the dependencies between prediction errors and parameter values in order to obtain more precise predictions for unobserved locations. The main limitation of any Hierarchical Bayes approach is that the quality of the results directly depends on the choice for the prior PDF for the prediction errors. For simple applications similar to the toy problem in this paper, this task is easy. However, in applications relevant for real life applications, identifying a good definition the for the prior prediction-error structure is a difficult challenge which, for the moment, needs to be addressed case by case and for which no generally applicable solution exists. Moreover, in this paper, the prior knowledge for prediction errors is described by a Gaussian process. The only limitation if one want to employ any other probability distributions, is in Equation 7, which would not have a closed form solution anymore. The solution is to infer prediction errors at locations of interests using a sampling approach in the same way it is currently done for the prediction errors at observed locations.

A second limitation of the explicit estimation of prediction errors in a Hierarchical Bayes approach is that the number of parameters to be estimated increases linearly with the number of observations. This aspect is a key computational challenge because sampling in high dimensions is a difficult task. The explicit extension to non-gaussian cases and large number of observations remain open questions.

6. Conclusion

The current approach of employing a Hierarchical Bayes approach to estimate model parameters and hyper-parameters does not fully account for prediction errors. Taking account of prediction errors requires estimating the joint PDF for model parameters, hyper-parameters, and model prediction errors themselves. It allows capturing the dependencies between epistemic uncertainties related to model parameters and prediction errors. This approach better represent the epistemic nature of model prediction errors, which like model parameters are unknown constants. The real-life application of any Hierarchical Bayes method requires a good definition the for the prior prediction-error structure which remains a case-specific challenge for which no generally applicable solution exists for the moment.

7. REFERENCES

Au, S.-K. and Zhang, F.-L. (2016). "Fundamental two-stage formulation for bayesian system identification, part i: General theory." *Mechanical Systems and Signal Processing*, 66, 31–42.

Beck, J. and Katafygiotis, L. (1998). "Updating models and their uncertainties. i: Bayesian statistical framework." *Journal of Engineering Mechanics*, 124(4), 455–461.

Behmanesh, I., Moaveni, B., Lombaert, G., and Papadimitriou, C. (2015). "Hierarchical bayesian model updating for structural identification." *Mechanical Systems and Signal Processing*, 64, 360–376.

Brynjarsdóttir, J. and O'Hagan, A. (2014). "Learning about physical parameters: The importance of model discrepancy." *Inverse Problems*, 30(11), 114007.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*. CRC Press, 3 edition.

Goulet, J.-A., Texier, M., Michel, C., Smith, I., and Chouinard, L. (2014). "Quantifying the effect of modelling simplifications for structural identification of bridges." *Journal of Bridge Engineering*, 19(1), 59–71.

Huang, Y., Beck, J. L., and Li, H. (2017). "Hierarchical sparse bayesian learning for structural damage detection: Theory, computation and application." *Structural Safety*, 64, 37–53.

Ling, Y., Mullins, J., and Mahadevan, S. (2014). "Selection of model discrepancy priors in bayesian calibration." *Journal of Computational Physics*, 276, 665–680.

Murphy, K. (2012). *Machine learning: a probabilistic perspective*. The MIT Press.

Nagel, J. B. and Sudret, B. (2016). "A unified framework for multilevel uncertainty quantification in bayesian inverse problems." *Probabilistic Engineering Mechanics*, 43, 68–84.

Neal, R. M. et al. (2011). "Mcmc using hamiltonian dynamics." *Handbook of Markov Chain Monte Carlo*, 2(11).

Papadimitriou, C., Beck, J., and Katafygiotis, L. (2001). "Updating robust reliability using structural test data." *Probabilistic Engineering Mechanics*, 16(2), 103–113.

Papaioannou, I. and Straub, D. (2017). "Learning soil parameters and updating geotechnical reliability estimates under spatial variability–theory and application to shallow foundations." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 11(1), 116–128.

Rasmussen, C. and Williams, C. (2006). "Gaussian processes for machine learning." *the MIT Press*, 2(3), 4.

Simoen, E., Moaveni, B., Conte, J. P., and Lombaert, G. (2012). "Probabilistic damage assessment of a seven-story rc building slice using bayesian fe model updating." *9th National Congress on Theoretical and Applied Mechanics*, Citeseer.

Simoen, E., Papadimitriou, C., and Lombaert, G. (2013a). "On prediction error correlation in bayesian model updating." *Journal of Sound and Vibration*, 332(18), 4136—4152.

Simoen, E., Papadimitriou, C., Roeck, G. D., and Lombaert, G. (2013b). "Influence of the prediction error correlation model on bayesian FE model updating results." *Life-Cycle and Sustainability of Civil Infrastructure Systems*, 192–199.

Yuen, K., Beck, J., and Katafygiotis, L. (2006). "Unified probabilistic approach for model updating and damage detection." *Journal of Applied Mechanics-Transactions of the ASME*, 73(4), 555–564.

Zhang, E., Feissel, P., and Antoni, J. (2011). "A comprehensive Bayesian approach for model updating and quantification of modeling errors." *Probabilistic Engineering Mechanics*, 26(4), 550–560.