On prediction error correlation in Bayesian model updating

Ellen Simoen^{a,*}, Costas Papadimitriou^b, Geert Lombaert^a

^aKU Leuven, Department of Civil Engineering, Kasteelpark Arenberg 40 box 2448, B-3001 Leuven, Belgium ^bUniversity of Thessaly, Department of Mechanical Engineering, Pedion Areos, 38334 Volos, Greece

Abstract

In Bayesian model updating, probability density functions of model parameters are updated accounting both for the information contained in the data and for uncertainties present in the measurements and model predictions, requiring a probabilistic model for the error between predictions and observations. Most often, a zero-mean uncorrelated Gaussian prediction error is assumed, although in many engineering applications prediction errors will show non-negligible spatial and/or temporal correlation (e.g. when densely populated sensor grids are used). In this paper, the effect of prediction error correlation on the results of the Bayesian model updating scheme is studied, and it is investigated how the challenging task of selecting a suitable prediction error correlation structure can be addressed appropriately. In two illustrative applications, it is demonstrated that Bayesian model class selection can be effectively applied to this end, ensuring more realistic modeling and corresponding Bayesian model updating results.

Keywords: Bayesian inference, uncertainty quantification, parameter estimation, model updating, prediction error correlation, model class selection, modeling error

1. Introduction

Model updating (often also referred to as *parameter estimation*) is a problem commonly encountered in many science and engineering fields. Generally speaking, model updating aims to reconstruct or calibrate unknown functions or properties which appear as parameters in numerical models, based on observed behavior of the system of interest. For instance, in structural engineering, finite element (FE) model updating [1, 2] is often applied for structural damage assessment, where damage in civil structures is identified by calibrating stiffness parameters based on observed modal characteristics such as eigenfrequencies and mode shapes [3].

The deterministic model updating process consists of solving an inverse problem, i.e. finding the optimal parameters of a model such that the best possible fit is obtained between the model output and the observed data. This is usually accomplished by formulating the problem as a constrained optimization problem, where the objective is to minimize the discrepancy between computed and measured data. In many cases, however, this optimization problem is ill-posed, meaning that uniqueness, stability and even existence of the solution of the inverse problem cannot be guaranteed. This is a non-negligible issue in model updating, as the model predictions and measured data are always subject to errors and limitations. In the context of model updating,

^{*}Corresponding author. E-mail: ellen.simoen@bwk.kuleuven.be

Postprint submitted to Journal of Sound and Vibration

Published version:E. Simoen, C. Papadimitriou, and G. Lombaert.On prediction error cor-relation in Bayesian model updating.Journal of Sound and Vibration, 332(18):4136–4152, 2013.http://dx.doi.org/10.1016/j.jsv.2013.03.019

these errors are often categorized into *modeling errors* on the one hand, i.e. errors related to the model predictions, and *measurement errors* on the other hand. Accounting for these uncertainties and studying their effect on the results of the FE model updating scheme is therefore an important and indispensable undertaking.

One possible approach to incorporate uncertainty on the observations and model predictions into the FE model updating process is to adopt a probabilistic scheme based on Bayesian inference [4–9]. Using the well-known Bayes theorem, a prior PDF reflecting the prior knowledge about the parameters is transformed into a posterior PDF, accounting both for uncertainty in the prior information as well as for uncertainty in the experimental data and FE model predictions. This transformation is done through the so-called likelihood function, which reflects how well the FE model can explain the observed data. The likelihood function is computed using the probabilistic model of the prediction error (i.e. the discrepancy between model predictions and observations).

Effective application of the Bayesian FE model updating technique in practice requires (1) the selection of a suitable joint prior PDF and (2) the selection of a suitable likelihood function or prediction error model. The first of these topics has been documented extensively in literature [10–12]; however, much less attention has been given to the latter, as it is usually assumed that the probabilistic model of the prediction error is known. In most cases, an *uncorrelated* zero-mean Gaussian error is adopted, e.g. in [13–21], in correspondence with the Principle of Maximum Entropy [22, 23]. However, in many structural engineering applications, data and corresponding prediction errors are most likely correlated in space (when dense sensor grids are used, e.g. for optical measurement techniques), in time (when high sampling frequencies are used), or both in space and in time [24]. This means the assumption of an uncorrelated prediction error is not always realistic. In this paper, it is investigated how correlation can be properly accounted for in Bayesian model updating.

Section 2 provides a brief overview of the Bayesian inference scheme for parameter estimation, and recalls some asymptotic expressions for the posterior PDF. In section 3, it is investigated how prediction error correlation can be accounted for in the Bayesian scheme, and how correlation affects the results of the Bayesian inference methodology. Furthermore, attention is given to the non-trivial task of selecting a suitable correlation model, as in most cases very little information is available on the specific correlation structure of the modeling error. It is demonstrated how Bayesian model class selection and the estimation of error parameters as well as model parameters can be employed to this end. In section 4, Bayesian model class selection is successfully applied to a simple linear regression analysis; in Section 5, a more realistic structural mechanics application is considered, where the method is shown to be equally effective.

2. Bayesian parameter estimation

This section presents the Bayesian updating methodology, starting with some preliminary specifications of the basic framework concerning model classes and uncertainties.

2.1. Uncertainty in parameter estimation

In general terms, a model $\mathcal{M}_M(\theta_M)$ belonging to the model class \mathcal{M}_M provides a mapping from the parameters θ_M to an output vector $\mathbf{G}_M(\theta_M) \in \mathbb{R}^N$ through the transfer operator \mathbf{G}_M :

$$\mathcal{M}_{\mathrm{M}}(\theta_{\mathrm{M}}): \quad \theta_{\mathrm{M}} \mapsto \mathbf{G}_{\mathrm{M}}(\theta_{\mathrm{M}})$$
 (1)

In the ideal case, the model output $\mathbf{G}_{M}(\theta_{M})$ corresponds perfectly to the true system output \mathbf{d} , i.e. $\mathbf{G}_{M}(\theta_{M}) = \mathbf{d}$. This is the main starting point for deterministic parameter identification, where the objective is to determine the model parameters θ_{M} for a given set of observed system outputs \mathbf{d} . However, the equality $\mathbf{G}_{M}(\theta_{M}) = \mathbf{d}$ is only valid when it is assumed that the underlying fundamental physics of the system are fully known. This is of course never the case, as no model is capable of perfectly representing the behavior of the true physical system. A modeling error η_{G} is therefore always present, and can be described as the discrepancy between the model predictions $\mathbf{G}_{M}(\theta_{M})$ and the true system output \mathbf{d} , i.e. $\eta_{G} = \mathbf{d} - \mathbf{G}_{M}(\theta_{M})$.

As the true system output has to be measured and processed experimentally, the data **d** are always subject to measurement error, resulting in a discrepancy between the true system output **d** and the actually observed data $\tilde{\mathbf{d}}$. This difference is defined as the measurement error $\eta_{\rm D} = \tilde{\mathbf{d}} - \mathbf{d}$. Eliminating the unknown true system output **d** from the error equations and collecting both errors on the right hand side of the equation yields:

$$\tilde{\mathbf{d}} - \mathbf{G}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{M}}) = \boldsymbol{\eta}_{\mathrm{G}} + \boldsymbol{\eta}_{\mathrm{D}} = \boldsymbol{\eta}$$
⁽²⁾

The sum of both errors is equal to the total observed prediction error η , defined as the difference between the model predictions and the observed quantities. The above expressions serve as a starting point for the Bayesian uncertainty quantification method.

2.2. Bayesian inference methodology

The general principle of Bayesian parameter estimation is that the model parameters $\theta_{M} \in \mathbb{R}^{N_{M}}$ that parameterize model class \mathcal{M}_{M} are modeled as random variables, i.e. probability density functions (PDFs) are appointed to these parameters, which are updated in the inference scheme based on the available information. Measurement and modeling uncertainty are taken into account by modeling the respective errors as random variables as well: PDFs are appointed to η_{G} and η_{D} , which are parametrized by parameters $\theta_{G} \in \mathbb{R}^{N_{G}}$ and $\theta_{D} \in \mathbb{R}^{N_{D}}$. These parameters are added to the structural model parameters θ_{M} to form the general model parameter set $\theta = \{\theta_{M}, \theta_{G}, \theta_{D}\}^{T} \in \mathbb{R}^{N_{\theta}}$. This in fact corresponds to adding two probabilistic model classes to the structural model class \mathcal{M}_{M} to form a *joint* model class $\mathcal{M} = \mathcal{M}_{M} \times \mathcal{M}_{G} \times \mathcal{M}_{D}$, parametrized by θ . It has to be noted here that introducing a probabilistic model for the errors is only one of several possible approaches for probabilistic modeling of the uncertainties; alternatively, one could revert to non-parametric approaches [25, 26] acting directly on the operators of the model (e.g. making use of random matrix theory [27, 28]), or *generalized* probabilistic approaches [29] that combine parametric and non-parametric approaches.

To express the updated probabilities of the unknown parameters θ , given some observations $\tilde{\mathbf{d}}$ and a certain joint model class \mathcal{M} , Bayes' theorem is used:

$$p(\boldsymbol{\theta} \mid \tilde{\mathbf{d}}, \mathcal{M}) = c \ p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}, \mathcal{M}) \ p(\boldsymbol{\theta} \mid \mathcal{M})$$
(3)

where $p(\theta|\tilde{\mathbf{d}}, \mathcal{M})$ is the updated or posterior PDF of the model parameters given the measured data $\tilde{\mathbf{d}}$ and the assumed model class \mathcal{M} ; c is a normalizing constant that ensures the posterior PDF integrates to one; $p(\tilde{\mathbf{d}}|\theta, \mathcal{M})$ is the PDF of the observed data given the parameters θ ; and $p(\theta|\mathcal{M})$ is the initial or prior PDF of the parameters. In the following, the explicit dependence on the model class \mathcal{M} is omitted in order to simplify the notations.

The prior PDF $p(\theta)$ reflects the probability of model parameters θ in the absence of measurement results. In many cases, the prior PDF is chosen based on engineering judgment or on

computational tractability (e.g. conjugate priors [30]). Except in cases where a large amount of data is at hand, the prior PDF most often has a significant influence on the Bayesian updating results. A wide range of methods has been developed to obtain "objective" prior PDFs based on the given prior information. One of the most commonly used approaches in this respect is the method based on the Maximum Entropy Principle [22, 23], which determines the prior PDF that, given the current state of prior information, results in maximum entropy.

The PDF of the experimental data $p(\tilde{\mathbf{d}}|\theta)$ can be interpreted as a measure of how good a model succeeds in explaining the observations $\tilde{\mathbf{d}}$. As this PDF reflects the likelihood of observing the data $\tilde{\mathbf{d}}$ when the model is parameterized by θ , it is also referred to as the *likelihood* function $L(\theta|\tilde{\mathbf{d}})$. Since the data set $\tilde{\mathbf{d}}$ is fixed, this function in fact no longer represents a conditional PDF, and can be denoted as $L(\theta; \tilde{\mathbf{d}})$; in the following, however, the common notation of $L(\theta|\tilde{\mathbf{d}})$ is pertained. The likelihood function reflects the contribution of the measured data $\tilde{\mathbf{d}}$ in the determination of the updated PDF of the model parameters, and may be determined according to the Total Probability Theorem in terms of the probabilistic models of the measurement and modeling errors:

$$p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}) \equiv L(\boldsymbol{\theta} \mid \tilde{\mathbf{d}}) = \int_{\mathbb{R}^N} p_{\tilde{\mathbf{d}}}(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}, \mathbf{d}) \ p_{\mathbf{d}}(\mathbf{d} \mid \boldsymbol{\theta}) \ d\mathbf{d}$$
(4)

$$= \int_{\mathbb{R}^N} p_{\eta_{\mathrm{D}}}(\tilde{\mathbf{d}} - \mathbf{d} \mid \boldsymbol{\theta}_{\mathrm{D}}) p_{\mathbf{d}}(\mathbf{d} \mid \boldsymbol{\theta}) d\mathbf{d}$$
(5)

$$= \int_{\mathbb{R}^N} p_{\eta_{\mathrm{D}}}(\tilde{\mathbf{d}} - \mathbf{d} \mid \boldsymbol{\theta}_{\mathrm{D}}) p_{\eta_{\mathrm{G}}}(\mathbf{d} - \mathbf{G}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{M}}) \mid \boldsymbol{\theta}_{\mathrm{G}}) d\mathbf{d}$$
(6)

where $p_{\eta_{\rm D}}(\tilde{\mathbf{d}} - \mathbf{d}|\theta_{\rm D})$ corresponds to the probability of obtaining a measurement error $\eta_{\rm D}$, given the PDF of $\eta_{\rm D}$ parameterized by $\theta_{\rm D}$, and where $p_{\eta_{\rm G}}(\mathbf{d} - \mathbf{G}_{\rm M}(\theta_{\rm M})|\theta_{\rm G})$ represents the probability of obtaining a modeling error $\eta_{\rm G}$ when the PDF of $\eta_{\rm G}$ is known and parameterized by $\theta_{\rm G}$. Here, it is implicitly assumed that the modeling error and measurement error are independent variables.

The above equations show that the likelihood function can be computed as the convolution of the PDFs of the measurement and modeling error. When no information is available on the individual errors, as is most often the case, the likelihood function can be simply constructed using the probabilistic model of the total prediction error η , parameterized by θ_{η} :

$$p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}) \equiv L(\boldsymbol{\theta} \mid \tilde{\mathbf{d}}) = p(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{\boldsymbol{\eta}})$$
(7)

When the prior PDF and likelihood function are determined, Eq. (3) allows for the updating of the PDFs of the model parameters based on experimental observations of the system. For most practical applications where multiple parameters are involved, computing the joint and marginal PDFs requires solving high-dimensional integrals, therefore use is often made of approximative measures (Section 2.3) or sampling methods such as the Markov Chain Monte Carlo (MCMC) method [31] and its derivatives, e.g. Delayed-Rejection Adaptive Metropolis-Hastings MCMC [32], the adaptive MCMC method [33] and Transitional MCMC [34]. For some particular cases, the posterior PDF can be determined analytically; in Appendix A, the posterior PDF is elaborated for a linear prediction model and a Gaussian prediction error and Gaussian prior PDF.

2.3. Asymptotic approximation of the posterior PDF

Provided sufficient data are at hand, asymptotic expressions for the posterior PDF form a costeffective alternative to computationally demanding methods (e.g. MCMC sampling). Moreover, the approximations can be employed for validation of results obtained by other approaches or as a way to perform an initial reconnaissance of the posterior updating results.

When a large amount of data is available, the posterior PDF can be approximated asymptotically by a Gaussian PDF centered at the Maximum A Posteriori (or MAP) point and characterized by covariance matrix $\hat{\Sigma}_{\theta,po}$ [4, 35]:

$$p(\boldsymbol{\theta} \mid \tilde{\mathbf{d}}) \approx \mathcal{N}(\hat{\boldsymbol{\theta}}^{\text{MAP}}, \hat{\boldsymbol{\Sigma}}_{\theta, \text{po}})$$
 (8)

The MAP estimate is defined as the parameter set that maximizes the posterior PDF, or, equivalently, minimizes the negative log posterior PDF:

$$\hat{\theta}^{\text{MAP}} = \arg\min_{\theta} \left\{ -\log L(\theta \mid \tilde{\mathbf{d}}) - \log p(\theta) \right\} = \arg\min_{\theta} J_{\text{MAP}}$$
(9)

Analogously, the Maximum Likelihood (or ML) estimate and Maximum A Priori (or MAPr) estimate are defined as the parameter sets that minimize the ML and prior objective functions, $J_{\text{ML}} = -\log L(\theta \mid \tilde{\mathbf{d}})$ and $J_{\text{MAPr}} = -\log p(\theta)$, respectively.

The approximate covariance matrix $\hat{\Sigma}_{\theta,po}$ is computed as the inverse Hessian of the MAP objective function J_{MAP} , evaluated in the MAP point:

$$\hat{\Sigma}_{\theta, \text{po}}^{-1} = \nabla_{\theta}^2 J_{\text{MAP}}(\hat{\theta}^{\text{MAP}}) \tag{10}$$

In many optimization algorithms, the Hessian in Eq. (10) is computed as a by-product in the optimization of expression (9).

2.4. Relation to information entropy

The posterior information entropy is often used as a measure of the resulting uncertainty in the Bayesian estimates of the model parameters. It is defined as:

$$h_{\rm po} = \mathbb{E}[-\log p(\theta \mid \hat{\mathbf{d}})]_{\rm po} \tag{11}$$

where $\mathbb{E}[\cdot]_{po}$ denotes the expected value with respect to the posterior PDF. When the asymptotic expression in Eq. (8) is valid, the posterior entropy may be approximated as:

$$h_{\rm po} \approx \frac{1}{2} \log \left[(2\pi e)^{N_{\theta}} \det \hat{\Sigma}_{\theta,\rm po} \right]$$
(12)

3. The prediction error correlation model

In most practical applications, very little information is available concerning the probabilistic model representing the prediction error, therefore an uncorrelated Gaussian prediction error is usually adopted in correspondence with the principle of Maximum Entropy [22]. It can sometimes be suspected, however, that correlation is present in the prediction error, e.g. spatial correlation in the case of densely spaced sensor grids or temporal correlation when high sampling frequencies are used to obtain experimental data. In those cases, it is important to account for this correlation, as it can be expected that correlated data do not provide the same information content compared to uncorrelated data [36].

In this section, it is first investigated how correlation in the prediction error affects the results of the Bayesian inference methodology, in general and more specifically for a Gaussian prediction error. Subsequently, attention is given to the challenging task of selecting a suitable correlation structure.

3.1. Influence of correlation on the Maximum Likelihood estimate

When a Gaussian prediction error is assumed, the likelihood function in Eq. (7) simplifies to:

$$L(\boldsymbol{\theta} \mid \tilde{\mathbf{d}}) = K^{-1} \exp\left[-\frac{1}{2}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\boldsymbol{\eta}\right]$$
(13)

where $K = (\sqrt{2\pi})^N (\det \Sigma_\eta)^{1/2}$. Improved insight into this expression for the likelihood can be obtained by reformulating the covariance matrix using an eigenvalue decomposition as follows:

$$\Sigma_{\eta} = \mathbf{V} \mathbf{D} \mathbf{V}^{\mathrm{T}} \tag{14}$$

where **D** is a diagonal matrix containing the *N* eigenvalues λ_j of Σ_{η} as its diagonal elements, and where **V** is an orthogonal matrix containing the corresponding eigenvectors. The eigenvalues are equal to the squares of the singular values of the covariance matrix, i.e. $\lambda_j = s_j^2$. The likelihood function in Eq. (13) can now be reformulated as:

$$L(\boldsymbol{\theta} \mid \tilde{\mathbf{d}}) = \prod_{j=1}^{N} k_j^{-1} \exp\left[-\frac{1}{2} \frac{(\boldsymbol{\eta}^{\mathrm{T}} \mathbf{V}_j)^2}{s_j^2}\right]$$
(15)

with $k_j = \sqrt{2\pi}s_j$ and corresponding ML objective function:

$$J_{\rm ML} = -\log L(\boldsymbol{\theta} \mid \tilde{\mathbf{d}}) = \sum_{j=1}^{N} \left[\log k_j + \frac{1}{2} \frac{(\boldsymbol{\eta}^{\rm T} \mathbf{V}_j)^2}{s_j^2} \right]$$
(16)

where $\log k_j = \frac{1}{2} \log(2\pi) + \log s_j$, and where \mathbf{V}_j denotes the *j*-th column of \mathbf{V} or the *j*-th eigenvector of Σ_{η} . It is clear that the likelihood function can be interpreted as the product of *N* independent 1-dimensional zero-mean Gaussian densities of $\eta^T \mathbf{V}_j$ with variances $s_j^2 = \lambda_j$. In the ML objective function – which is a generalized least squares objective function – the eigenvalues of the covariance matrix are the weighting factors appointed to the projections $\eta^T \mathbf{V}_j$ of the prediction error $\boldsymbol{\eta}$ onto the eigenvectors \mathbf{V}_j .

For an uncorrelated prediction error, the eigenvectors are unit vectors, which means the error components in the ML objective function are simply squared, and weighted by their appointed variances. For a correlated error, however, the error components are first recombined through the eigenvectors, and then weighted by the eigenvalues of the covariance matrix, indicating that introducing correlation leads to a recombination of the available data.

3.2. Influence of correlation on the Maximum A Posteriori estimate

As shown above, including correlation changes the weighting of the residuals and how the residuals are constructed. This also implies that, for the posterior PDF, the weighting of the data (represented by the likelihood function) is altered compared to the prior information. This can be illustrated by considering the objective function that should be minimized to obtain the MAP estimate in Eq. (9).

Assuming a zero-mean Gaussian prediction error with N = 2 data points and a prediction error covariance matrix as follows:

$$\Sigma_{\eta} = \sigma_{\eta}^{2} \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{\eta}^{2}(1+c) & 0 \\ 0 & \sigma_{\eta}^{2}(1-c) \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^{\mathrm{T}}$$
(17)

leads to the following MAP objective function:

$$J_{\text{MAP}} = J_{\text{ML}} + J_{\text{MAPr}} = \log k_1 + \frac{(\eta_1 + \eta_2)^2}{4\sigma_{\eta}^2(1+c)} + \log k_2 + \frac{(-\eta_1 + \eta_2)^2}{4\sigma_{\eta}^2(1-c)} + J_{\text{MAPr}}$$
(18)

$$= J_{\text{ML,uncorr}} + \log(1 - c^2) + \frac{c^2(\eta_1^2 + \eta_2^2) - 2c\eta_1\eta_2}{2\sigma_n^2(1 - c^2)} + J_{\text{MAPr}}$$
(19)

where $J_{ML,uncorr}$ denotes the ML objective function for a covariance matrix as in Eq. (17) with c = 0. This expression clearly shows the difference in weighting of the data residuals caused by the introduction of the correlation. Note that introducing correlation does not necessarily lead to a lower weight on the likelihood function, as the sign of the third term in the above expression depends on the signs and magnitude of the errors and the correlation coefficient.

3.3. Influence of correlation on the posterior covariance matrix

Suppose again a Gaussian prediction error and prior PDF, and a linear prediction model such that $G_M(\theta) = J_G \theta$. Elaborating the posterior PDF for this case and using the eigenvalue decomposition in Eq. (14) leads to the following expression for the inverse posterior covariance matrix of the parameters (Appendix A):

$$\boldsymbol{\Sigma}_{\theta, \text{po}}^{-1} = \mathbf{J}_{\text{G}}^{\text{T}} \boldsymbol{\Sigma}_{\eta}^{-1} \mathbf{J}_{\text{G}} + \boldsymbol{\Sigma}_{\theta, \text{pr}}^{-1} = \left(\mathbf{V}^{\text{T}} \mathbf{J}_{\text{G}} \right)^{\text{T}} \mathbf{D}^{-1} \left(\mathbf{V}^{\text{T}} \mathbf{J}_{\text{G}} \right) + \boldsymbol{\Sigma}_{\theta, \text{pr}}^{-1} = \sum_{j=1}^{N} \frac{\left(\mathbf{J}_{\text{G}}^{\text{T}} \mathbf{V}_{j} \right)^{2}}{s_{j}^{2}} + \boldsymbol{\Sigma}_{\theta, \text{pr}}^{-1}$$
(20)

It is interesting to note that, for this particular case, the observed data does not enter into this expression, i.e. the posterior covariance matrix is completely determined by the prior covariance, the prediction error covariance and the sensitivities in J_G .

For a single-parameter example, with a correlation matrix as defined in Eq. (17), and a linear prediction model characterized by $\mathbf{J}_{G} = [g_1, g_2]^T$, the posterior parameter variance can be determined according to Eq. (20) as:

$$\sigma_{\theta,\text{po}}^{2} = \left[\frac{(g_{1} + g_{2})^{2}}{2\sigma_{\eta}^{2}(1 + c)} + \frac{(-g_{1} + g_{2})^{2}}{2\sigma_{\eta}^{2}(1 - c)} + \frac{1}{\sigma_{\theta,\text{pr}}^{2}}\right]^{-1}$$
(21)

In order to investigate the influence of the correlation coefficient *c*, the prediction model is taken equal to $\mathbf{J}_{\rm G} = [1, -0.5]^{\rm T}$, and the variances σ_{η}^2 and $\sigma_{\theta,\rm pr}^2$ are set to 0.4² and 1.4², respectively; in Figure 1 the corresponding posterior variance $\sigma_{\theta,\rm po}^2$ is plotted as a function of *c*. This figure indicates that the posterior variance shows a distinct maximum, which is located at $c = g_2/g_1 = -0.5$ in this case where $g_1 > g_2$. When $g_1 < g_2$, the maximum would be located at $c = g_1/g_2$; when $g_1 = g_2$, no distinct maximum is present as $\sigma_{\theta,\rm po}^2$ becomes a linear function of *c*. As the posterior parameter entropy is here directly proportional to the posterior variance (see Eq. (12)), this maximum denotes the maximum entropy point. Note that the maximum is only located at $c = g_1$



Figure 1: Posterior variance of θ as a function of the correlation coefficient *c*. The dashed line denotes the posterior variance for c = 0.

0 (corresponding to an uncorrelated prediction error) when one of the sensitivities becomes zero. These observations imply that adopting an uncorrelated likelihood function (which corresponds to maximum entropy with respect to the *prediction error*) does not automatically lead to the highest variance regarding the posterior model parameter values. When correlation is present in the prediction errors, it is therefore not always conservative (with respect to the posterior estimates) to assume an uncorrelated error.

The posterior variance is found to be larger compared to the uncorrelated case for values of *c* in the following range:

$$0 \le |c| \le \frac{2|g_1g_2|}{g_1^2 + g_2^2} \tag{22}$$

where *c* should have the same sign as the product g_1g_2 . This condition suggests that quantities with sensitivities with opposite signs should be negatively correlated (and vice versa) in order to obtain higher posterior entropy. In this respect, it is interesting to note that, although one would intuitively expect that a correlated prediction error leads to lower information content and higher uncertainties, this appears to be not at all the case. Knowledge on the correlation between data points or errors may prove more informative and subsequently may result in a lower posterior uncertainty. This can also be observed in the limit case of full correlation (i.e. $c \rightarrow 1$ or -1), where zero posterior parameter variance is found due to a zero singular value in Eq. (17). This corresponds to a known zero variance for one of the linear prediction error combinations, meaning that the parameter can be determined unambiguously.

All these findings indicate that prediction error correlation not only affects the location of the ML and MAP estimates, but moreover has an important effect on the posterior uncertainty of the parameters of interest. This implies that the selection of a suitable prediction error correlation model is a non-trivial and challenging task, especially in the common case where little or no information is available on the specific nature of the error. In the next subsection, it is elaborated how Bayesian model class selection can be used as an effective tool in selecting an appropriate correlation model class for the prediction error, based on the available information.

3.4. Accounting for correlation in the prediction error

When it is suspected that the prediction errors are correlated and one wishes to account for it in the Bayesian scheme, the challenge remains in selecting a suitable correlation structure, especially as usually very little is known regarding the correlation. Then, it may be opted to simply assume a certain correlation model; in practice, correlation functions of the exponential family are often used [8, 37, 38]. Provided that sufficient observational data are available, as well as some basic knowledge regarding the correlation structure, error parameters characterizing the correlation model can be included in the Bayesian scheme in order to estimate them simultaneously with the other model parameters. In this paper, it is suggested to apply Bayesian model class selection [39] in order to compare different possible correlation model classes and select the model that best suits the data at hand, thus avoiding the need for information regarding the correlation structure. In this section, a brief overview of the model class selection approach is given.

3.4.1. Bayesian model class selection

The Bayesian inference scheme can be applied at model class level in order to assess the plausibility of several alternative prediction error model classes $\mathcal{M}_{\eta,i}$, or more generally model classes \mathcal{M}_i , using observed data [39, 40]. For a set \mathcal{M} of N_C candidate model classes \mathcal{M}_i , the posterior probability of each model class \mathcal{M}_i is given by Bayes' theorem as:

$$P(\mathcal{M}_i \mid \tilde{\mathbf{d}}, \mathcal{M}) = \frac{p(\tilde{\mathbf{d}} \mid \mathcal{M}_i) P(\mathcal{M}_i \mid \mathcal{M})}{p(\tilde{\mathbf{d}} \mid \mathcal{M})}$$
(23)

where $P(\mathcal{M}_i|\mathcal{M})$ is the prior probability of each model class \mathcal{M}_i . The factor $p(\mathbf{\tilde{d}} | \mathcal{M}_i)$ denotes the model class likelihood or the *evidence* for the model class \mathcal{M}_i provided by the data $\mathbf{\tilde{d}}$, and can be determined based on the Total Probability Theorem as:

$$p(\tilde{\mathbf{d}} \mid \mathcal{M}_i) = \int_{D_{\mathcal{M}_i}} p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}_i, \mathcal{M}_i) \ p(\boldsymbol{\theta}_i \mid \mathcal{M}_i) \ d\boldsymbol{\theta}_i$$
(24)

where θ_i is the parameter vector in a parameter space $D_{\mathcal{M}_i}$ that defines each model in \mathcal{M}_i . Note that the evidence $p(\mathbf{\tilde{d}} \mid \mathcal{M}_i)$ is equal to the reciprocal of the normalizing constant *c* in the general Bayes' theorem in Eq (3).

As usually equal prior model class probabilities are adopted, it suffices to compute the evidence values for all model classes and ranking them accordingly. Then, the most probable model class – according to the available data – corresponds to the model class with the highest evidence value. The actual computation of the evidence values poses, however, a challenging problem in most practical applications, as it entails computing a complex and usually high-dimensional integral. To overcome this problem, asymptotic approximations can be applied [35, 41], or, alternatively, methods based on stochastic simulation may be used, e.g. approaches using MCMC samples of the posterior PDF [42, 43], or multi-level MCMC methods such as TMCMC [19, 34] or nested sampling [44].

3.4.2. Model parsimony

It can be shown [45, 46] that applying a Bayesian inference scheme at the model class level automatically enforces model parsimony. Using Bayes' theorem and the fact that the posterior PDF integrates to one, the logarithm of the evidence in Eq. (24) can be reformulated and ex-

panded as:

$$\log p(\tilde{\mathbf{d}} \mid \mathcal{M}_i) = \text{LE} = \mathbb{E} \left[\log p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}_i, \mathcal{M}_i) \right]_{\text{po}} - \mathbb{E} \left[\log \frac{p(\boldsymbol{\theta}_i \mid \tilde{\mathbf{d}}, \mathcal{M}_i)}{p(\boldsymbol{\theta}_i \mid \mathcal{M}_i)} \right]_{\text{po}}$$
(25)

$$= LE_{data} - D_{KL} = LE_{data} + LE_{occam}$$
(26)

The first term in the above log evidence (LE) expression is the posterior mean value of the log likelihood function, termed LE_{data}, which gives a measure for the average data fit for the model class \mathcal{M}_i . The second term is the Kullback-Leibler divergence or relative entropy D_{KL} between the prior and the posterior PDF. This term gives a measure for the difference between the prior and posterior PDF, i.e. it is a measure for the information that is gained from the observations $\tilde{\mathbf{d}}$. It can be shown [39, 45] that this term is always non-negative and increases when the number of model parameters in the model class increases. Therefore, this term penalizes more complex models that extract more information from the data, thus eliminating the need for ad-hoc penalty terms. In the following, the negative of this term is denoted as the Occam term LE_{occam}, as it enforces the well-known Occam simplicity principle.

For a linear prediction model and a Gaussian prediction error and prior PDF, it can be verified (Appendix B) that the following expressions for the data fit term LE_{data} and the Occam term LE_{occam} are valid:

$$LE_{data} = \mathbb{E}\left[\log p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}_{i}, \mathcal{M}_{i})\right]_{po} = \log\left[p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}_{po,i}, \mathcal{M}_{i})\right] - \frac{1}{2}\operatorname{Tr}\left\{\mathbf{I} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}, pr}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\theta}, po}\right\}$$
(27)

$$LE_{occam} = -\mathbb{E}\left[\log\frac{p(\theta_i \mid \mathbf{d}, \mathcal{M}_i)}{p(\theta_i \mid \mathcal{M}_i)}\right]_{po} = \log p(\theta_{po,i} \mid \mathcal{M}_i) - \frac{1}{2}\operatorname{Tr}\left\{\Sigma_{\theta, pr}^{-1}\Sigma_{\theta, po}\right\} + h_{po}$$
(28)

where $\theta_{po,i}$ is the posterior expected value $\mathbb{E}[\theta_i]_{po}$ of the parameter set θ_i , which can in this case be computed according to Eq. (A.1); h_{po} denotes the information entropy of the posterior PDF, which in this case can be computed using Eq. (12). In many applications, the trace terms in the above expressions are of negligible magnitude compared to the other terms.

In cases where it can be assumed that the posterior PDF is close to Gaussian (Section 2.3), the approximative expressions in Eqs. (9) and (10) may be used to approximate the posterior mean value and covariance matrix. These quantities can be introduced into the expressions (27) and (28) above in order to approximate the terms LE_{data} and LE_{occam} .

3.4.3. Applying model class selection for the prediction error model

In general, it can be stated that when a limited amount of information is available regarding the prediction error correlation, it is desirable to make as few assumptions as possible in order to avoid biased results. Therefore, it is suggested to select a substantial set of alternative prediction error correlation model classes $\mathcal{M}_{\eta,i}$ based on the information at hand. When some information is available on the prediction error, these model classes may for instance be very similar but of increasing complexity; however, when very little is known regarding the prediction error, a range of possible model classes with a wide variety of properties may be selected.

In order to select the most probable model class from the set of alternative model classes, the log evidence values LE are computed for all model classes, and by comparing the associated data fit LE_{data} and Occam values LE_{occam} . As mentioned above, the data fit term yields insight into how well the joint model class is able to fit the data, whereas the Occam term provides a measure

for how much information was gained from the data, which roughly corresponds to how much the uncertainty has decreased through the data.

The efficacy of the Bayesian model class selection approach for differentiating between prediction error model classes will be demonstrated below in two applications: a very simple analytical example allows for improved insight into the workings of the approach and its associated difficulties, whereas a second example concerns a more realistic engineering application.

4. Analytical example: linear regression problem

In this section, the proposed approach is applied to a simple linear regression problem that consists of estimating two parameters, α and β , using some measured data $\tilde{\mathbf{y}}$ obtained at N observation points \mathbf{x} and where the prediction model is given as:

$$\mathbf{G}(\boldsymbol{\theta}_{\mathrm{M}}) = \mathbf{J}_{\mathrm{G}}\boldsymbol{\theta}_{\mathrm{M}} = \begin{bmatrix} \mathbf{x} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$
(29)

4.1. Simulated data

The data $\tilde{\mathbf{y}}$ are simulated by superimposing a correlated error $\boldsymbol{\eta}^*$ on the exact data $\mathbf{y} = \mathbf{J}_G \boldsymbol{\theta}_M^*$, where the true parameter values α^* and β^* are equal to 0.5 and 1.5, respectively. The prediction error $\boldsymbol{\eta}^*$ is assumed to have a zero-mean Gaussian distribution, characterized by a covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^*$ defined as $(\sigma^*)^2 \mathbf{C}_{\boldsymbol{\eta}}^*$ where $\sigma^* = 0.15$. An exponential correlation function is assumed for the correlation matrix $\mathbf{C}_{\boldsymbol{\eta}}^*$, such that:

$$[\mathbf{C}_{\eta}^{*}]_{ij} = \exp\left(-\frac{\Delta_{ij}}{\ell^{*}}\right)$$
(30)

In this equation, Δ_{ij} represents the distance between two observation points *i* and *j*, here equal to $|x_i - x_j|$, and where ℓ is the correlation length, with true value $\ell^* = 0.8$.

With these expressions, $N_D = 50$ sets of N = 20 data points with *x*-values equally spaced in the interval (1, 5) are simulated. In order to assess the influence of data availability, the computations are performed for an increasing number of data sets. For illustration purposes, the data sets $\tilde{\mathbf{y}}$ are shown in Figure 2, as well as the true model $\mathbf{G}_{M}(\theta_{M}^{*})$.



Figure 2: All data sets used in the analysis, where large dots denote the first data set. The solid line reflects the true model $G_M(\theta_M^*)$.

4.2. Prediction error model classes

Four alternative zero-mean Gaussian model classes are selected for the prediction error correlation structure: an uncorrelated model class (A), a model class with an exponential correlation function (B), a model class with a spherical correlation function (C) and a model class with an exponentially damped cosine correlation function (D). The covariance matrices of these four model classes are of increasing complexity, and are described as follows:

$$[\boldsymbol{\Sigma}_{\eta}^{A}]_{ij} = (\theta_{\eta,1}^{A})^{2} \mathbf{I}$$
(31)

$$[\boldsymbol{\Sigma}_{\eta}^{B}]_{ij} = (\theta_{\eta,1}^{B})^{2} \exp\left(-\frac{\Delta_{ij}}{\theta_{\eta,2}^{B}}\right)$$
(32)

$$[\boldsymbol{\Sigma}_{\eta}^{C}]_{ij} = (\theta_{\eta,1}^{C})^{2} \left[1 - \frac{3}{2} \left(\frac{\Delta_{ij}}{\theta_{\eta,2}^{C}} \right) + \frac{1}{2} \left(\frac{\Delta_{ij}}{\theta_{\eta,2}^{C}} \right)^{3} \right]$$
(33)

$$[\mathbf{\Sigma}_{\eta}^{D}]_{ij} = (\theta_{\eta,1}^{D})^{2} \exp\left(-\frac{\Delta_{ij}}{\theta_{\eta,2}^{D}}\right) \cos(\theta_{\eta,3}^{D} \Delta_{ij})$$
(34)

In these equations, parameters $\theta_{\eta,1}$ are the prediction error standard deviations, $\theta_{\eta,2}$ denote correlation lengths and $\theta_{\eta,3}$ corresponds to the correlation wavenumber that determines where the zeros of the cosine in Eq. (34) are located. The true model class corresponds to model class B with an imposed standard deviation of $\theta_{n,1}^B = 0.15$ and a correlation length $\theta_{n,2}^B = 0.8$.

In the following, the Bayesian inference scheme is applied to estimate the model parameters $\theta_{\rm M} = \{\alpha, \beta\}^{\rm T}$ and the prediction error parameters θ_{η} ; moreover, the most suitable correlation model will be searched by applying Bayesian model class selection. In order to demonstrate the importance of choosing a suitable correlation model, the prediction error parameters are taken fixed in a first stage.

4.3. Case I: Bayesian inference for model parameters only

In Eqs. (31) to (34), all prediction error parameters are assumed to be fixed and equal to the following values: $\theta_{\eta,1} = 0.15$ for all model classes, $\theta_{\eta,2}^B = \theta_{\eta,2}^D = 0.8$, $\theta_{\eta,2}^C = 4$ and $\theta_{\eta,3}^D = 2\pi$. The resulting correlation functions (Figure 3) are quite different in nature: correlation models B and C model only positive correlations, whereas model D includes negative correlations as well.

It is assumed that the joint prior PDF of the model parameters θ_M is Gaussian, characterized as follows:

$$p(\boldsymbol{\theta}_{\mathrm{M}}) = \mathcal{N}_{\boldsymbol{\theta}_{\mathrm{M}}}(\boldsymbol{\theta}_{\mathrm{M,pr}}, \boldsymbol{\Sigma}_{\mathrm{M,pr}}) = \mathcal{N}_{\boldsymbol{\theta}_{\mathrm{M}}} \left(\begin{bmatrix} 2\\ 3 \end{bmatrix}, \begin{bmatrix} 1.5^2 & 0\\ 0 & 2^2 \end{bmatrix} \right)$$
(35)

As the prediction model is linear and the prior PDF as well as the likelihood function are Gaussian PDFs, the expressions elaborated in Appendix A are valid and can be directly implemented. The mean posterior values of the parameters and their corresponding posterior standard deviations and correlations are listed in Table 1 for the first data set (Figure 2) and all 50 data sets, the corresponding marginal PDFs for the single data set are shown in Figure 4. Multiple data sets are included through the likelihood function; it can be shown that for this particular case

	Single data set						50 data sets					
Model	$\hat{\alpha}_{po}$	\hat{eta}_{po}	$\hat{\sigma}_{\alpha,\mathrm{po}}$	$\hat{\sigma}_{\beta,\mathrm{po}}$	$\hat{c}_{\alpha\beta,\mathrm{po}}$	$\hat{\alpha}_{po}$	$\hat{eta}_{ m po}$	$\hat{\sigma}_{\alpha,\mathrm{po}}$	$\hat{\sigma}_{\beta,\mathrm{po}}$	$\hat{c}_{\alpha\beta,\mathrm{po}}$		
					$(\times 10^{-2})$					$(\times 10^{-5})$		
А	0.528	1.461	0.029	0.090	-0.241	0.503	1.505	0.004	0.013	-4.840		
В	0.515	1.491	0.052	0.173	-0.799	0.506	1.502	0.007	0.025	-16.12		
С	0.502	1.515	0.055	0.189	-0.875	0.509	1.501	0.008	0.027	-17.67		
D	0.534	1.444	0.025	0.077	-0.179	0.503	1.502	0.004	0.011	-3.594		

Table 1: Results of the Bayesian parameter estimation: mean posterior values $\hat{\alpha}_{po}$ and $\hat{\beta}_{po}$, their corresponding standard deviations $\hat{\sigma}$ and correlation values $\hat{c}_{\alpha\beta,po}$.

the Gaussian likelihood function for N_D data sets is given as:

$$L(\boldsymbol{\theta} \mid \tilde{\mathbf{d}}) = \prod_{k=1}^{N_D} L(\boldsymbol{\theta} \mid \tilde{\mathbf{d}}_k) = K^{-1} \exp\left[\sum_{k=1}^{N_D} -\frac{1}{2} \boldsymbol{\eta}_k^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} \boldsymbol{\eta}_k\right]$$
(36)

It is easily verified that considering $N_D = 50$ data sets therefore in fact corresponds to performing 50 consecutive Bayesian updating routines, where each time the current posterior is taken as prior and updated using an additional data set. Note that this is not equivalent to performing a single Bayesian updating routine with $50 \times 20 = 1000$ data points; this would moreover require adapting the prediction error correlation models in Eqs. (31)–(34).

Examining the posterior mean values, it is clear that all values lie reasonably close to the true values, and that the accuracy of the estimates improves for a higher number of data sets, as expected. When comparing the posterior standard deviations for the correlated models B, C and D with the uncorrelated model A, it is clear that models B and C result in wider, very similar, PDFs, whereas model D yields smaller uncertainties compared to the uncorrelated case. This illustrates immediately the importance of the correlation structure. In the case where no information is available on the specific nature of the correlation, simply assuming model D would in this case lead to non-conservative posterior estimates.

Bayesian model class selection is applied in order to distinguish between the alternative models: for each model class, the log evidence value LE is computed as the sum of the terms LE_{data}



Figure 3: Case I: Correlation functions for fixed model classes A, B, C and D.



Figure 4: Normalized marginal posterior PDFs for (a) α and (b) β , for all four models and for a single data set.

	Si	ingle data	a set	50 data sets			
Model	LE	LE _{data}	LEoccam	LE	LE _{data}	LEoccam	
А	9.478	17.30	-7.821	498.4	510.1	-11.73	
В	15.56	21.90	-6.335	904.6	<u>914.9</u>	-10.24	
С	11.33	17.40	-6.070	452.6	462.6	-9.973	
D	13.55	21.76	-8.212	550.3	562.4	-12.12	

Table 2: Log evidence values (LE), and corresponding data and occam terms for all models.

and LE_{occam} (according to Eqs. (27) and (28)), all listed in Table 2. Investigating the values for a single data set, it is immediately clear that the highest log evidence value is associated with the correct model B. When the data fit and Occam terms are compared for models B and C, it is found that even though the spherical model C yields a higher Occam factor, the better data fit provided by the exponential model B proves dominant.

Comparison of the other values also shows that even though model D has a similar data fit as model B, it is penalized because of the higher information content obtained from the data (resulting in lower posterior uncertainty). This illustrates that the Bayesian model class selection scheme selects the model that results in a high data fit combined with relatively high posterior uncertainty (i.e. relatively low resolution). It should be noted that out of the 50 randomly simulated data sets, the correct correlation model is selected 45 out of 50 times based on the LE value, and 35 out of 50 times based on the LE_{data}-value.

Of course, as more data sets are added, the log Occam term will be increasingly overwhelmed by the amount of information contained in the data fit term. In Table 2, it is immediately clear that the correct model is selected solely based on the information contained in the data. In realistic applications, however, there is rarely more than a single data set available, which means the investigation of the log Occam terms becomes increasingly important.

4.4. Case II: Bayesian inference for model parameters and error parameters

In this subsection, it is assumed that the prediction error parameters in Eqs. (31) to (34) are unknown; they are included in the Bayesian inference scheme and are estimated together with the two model parameters α and β . After the Bayesian estimation, a model class selection routine is performed in order to infer the most probable model class. Note that the prediction error model classes are of increasing complexity: model class A is parameterized by a single parameter (the error variance), model classes B and C are characterized by two prediction error parameters and model class D has three parameters. This has to be taken into account when comparing e.g. log Occam terms.

In the Bayesian scheme, Gamma-distributed prior PDFs are appointed to all prediction error parameters, in correspondence with the Principle of Maximum Entropy for second-order positive variables [22, 47]; the appointed shape and scale factors are given in Table 3. The model parameters α and β are appointed the same priors as for case I (Eq. (35)). The data are simulated as before with model class B, with an imposed variance of $(0.15)^2$ and a correlation length of 0.8. The asymptotic expressions are used to compute the MAP estimates and posterior covariance matrices (Table 4); these values are then employed to determine the log evidence values LE, log data fit values LE_{data} and log Occam terms LE_{occam}, also listed in Table 4.

	$\theta^{A}_{\eta,1}$	$\theta^B_{\eta,1}$	$\theta^B_{\eta,2}$	$\theta_{\eta,1}^C$	$\theta_{\eta,2}^C$	$\theta^{D}_{\eta,1}$	$ heta_{\eta,2}^D$	$\theta^{D}_{\eta,3}$
t	1.24	1.24	6.25	1.56	44.4	1.56	4.00	4.00
k	2.13	2.13	0.19	0.88	0.14	0.89	0.17	0.50
MAPr	0.50	0.50	1.00	0.50	6.00	0.5	0.50	1.50
COV	0.90	0.90	0.40	0.80	0.15	0.80	0.50	0.50

Table 3: Appointed shape factors k and scale factors t characterizing the prior distributions of the prediction error parameters for Case II, and corresponding Maximum A Priori points (MAPr = t(k - 1)) and coefficients of variation (COV = $\sqrt{1/k}$).

Single data set										
Model class	α	β	$\theta_{\eta,1}$	$\theta_{\eta,2}$	$\theta_{\eta,3}$	LE	LE _{data}	LE _{occam}		
А	0.529	1.460	0.072	-	-	9.103	22.92	-13.81		
В	0.524	1.471	0.079	0.263	-	10.02	21.56	-11.55		
С	0.502	1.517	0.263	5.900	-	11.56	18.87	-7.315		
D	0.525	1.468	0.080	0.337	2.208	10.82	24.61	-13.79		
50 data sets										
Model class	α	β	$\theta_{\eta,1}$	$\theta_{\eta,2}$	$\theta_{\eta,3}$	LE	LE _{data}	LE _{occam}		
А	0.503	1.505	0.145	-	-	492.2	509.7	-17.48		
В	0.506	1.502	0.147	0.745	-	895.4	913.3	-17.87		
С	0.509	1.501	0.254	4.000	-	852.4	869.7	-17.37		
D	0.506	1.501	0.146	0.756	0.793	894.7	913.2	-18.50		

Table 4: MAP values for all parameters and log evidence values for all model classes, for a single and 50 data sets.

Examining the values for a single data set, an unexpected result emerges at first sight: the spherical model class C is distinctly preferred over the other model classes, mainly due to the high associated log Occam value which signifies that relatively little information was obtained from the data using this model class. Most likely, this is caused by the fact that the spherical correlation function can never approximate the true model, as it is constrained in the parameter $\theta_{\eta,2}^C$ which should always be larger than or equal to the maximum Δ_{ij} . This is confirmed when the MAP correlation models are compared with the true sample correlation (Figure 5a), and is also reflected in the relatively low data fit value for this model class.

Moreover, based on the data fit values for a single data set, it appears that the cosine model D is selected, whereas previously it was found that this model was not preferred. This is due to the fact that the parameters of the correlation model are now modified such that the available

information is best fitted. It appears that correlation model D now is able to fit the true correlation structure almost equally well as the true correlation model B, an observation that is even more prominent when 50 data sets are used and which is confirmed in Figure 5b. It is also clear that when more data sets are used, the correct correlation model class B is eventually preferred, as model class D is penalized slightly more through the Occam factor due to the higher number of parameters representing this model class. Note also that despite the unsuitable correlation structure, model class C succeeds relatively well in achieving a high data fit for 50 data sets, especially compared to the uncorrelated model A. This can be explained by the presence of the variance parameter $\theta_{n,1}^C$, which is increased to compensate for the improper correlation structure.



Figure 5: MAP correlation functions for all model classes, compared with the true average sample correlation (thick solid line), for (a) a single and (b) 50 data sets.

5. Engineering example: Bayesian FE model updating of a reinforced concrete beam

In this section, the approach proposed above in Section 3.4 is used in an engineering example where the objective is to use simulated modal data to estimate stiffness parameters of a finite element model representing a damaged reinforced concrete (RC) beam, and to estimate parameters characterizing the prediction error. In order to mimic a more realistic prediction error, the simulated data are obtained from a 3D volume element model whereas the prediction model used in the Bayesian scheme consists of a simple 2D beam element model.

The RC beam in question has a length of 6 m, a cross-section of 200×250 mm², a mass of 750 kg, and is reinforced by 6 \emptyset 16 mm rebars and vertical stirrups \emptyset 8 mm every 200 mm. As in this case spatial correlation in the prediction error is considered, the computations are performed for seven different sensor configurations with increasing sensor density along the beam.

5.1. Simulated data

The simulated data are obtained from a 3D finite element model constructed using 5250 volume elements (Figure 6); the reinforcement has only been taken into account through an adapted material density. The data consists of the first four bending modes of the beam and associated natural frequencies, computed with free-free boundary conditions. In order to simulate structural damage in the beam, the Young's modulus of a small area around one third of the length of the beam is reduced to 18.75 GPa, whereas the rest of the beam is appointed a stiffness modulus of 37.5 GPa. An uncorrelated zero-mean Gaussian error with a standard deviation of 0.1% is superimposed on these modal data in order to simulate a measurement error; the final data set is shown in Figure 7a.



Figure 6: 3D solid FE model of the RC beam. The area with lowered stiffness is shown in gray.



Figure 7: (a) Modal data simulated with the damaged volume model and (b) normalized sample correlation in the difference between the simulated data and the predictions made by the beam model.

In Figure 8, the seven different sensor configurations are shown, where the number of sensors ranges from a mere 6 sensors up to all 151 sensors. In order to ensure a valid comparison between configurations, each sensor configuration is selected such that it contains the previous configuration. In the following, the number of sensors – corresponding to the number of observed DOFs – will be denoted as N_s .

5.2. The prediction model and prediction error

To model the reinforced concrete beam, a 2D finite element model is constructed which consists of 150 beam elements and 151 nodes, with two DOFs per node (UY, ROTZ), resulting in a total of $N_d = 302$ DOFs in the FE model. The moment of inertia of the equivalent cross-section equals 1.93×10^{-4} m⁴ and the initial Young's modulus is assumed to be 37.5 GPa. Shear deformation is included in the FE model with a factor $k_y = A_s/A = 1.2$.



Figure 8: The seven different sensor configurations used in this example: gray color denotes the presence of a sensor.

The model of the RC beam is parameterized by a single parameter θ_M representing the Young's modulus of the same region of elements at around one third of the length of the beam. This parameterized 2D FE model is used as prediction model in the Bayesian scheme: it allows for the computation of a set of modal data as a function of the model parameter θ_M . In the following, it is assumed that the computed modal data consist of N_M mode shapes $\phi_r \in \mathbb{R}^{N_d}$ and associated eigenvalues $\lambda_r = (2\pi f_r)^2$ with eigenfrequencies f_r , which are the solutions of the (undamped) eigenvalue equation $\mathbf{K}(\theta_M) \Phi = \mathbf{M} \Phi \Lambda$, where $\mathbf{K}(\theta_M)$ is the FE model stiffness matrix and \mathbf{M} the mass matrix. Φ collects the eigenvectors ϕ_r that correspond to the eigenvalues λ_r on the diagonal of Λ .

To obtain an initial idea of the occurring correlation, Figure 7b shows the estimated normalized sample correlation (assuming stationarity) present in the error between the simulated data and the predictions made by the 2D beam model (for $\theta_M = 20$ GPa). This figure clearly indicates that the prediction errors are correlated – albeit to a limited degree – and that the correlation is dependent on the shape of the considered mode. More specifically, the correlation length seems to decrease as the mode order increases.

5.3. Bayesian inference and model class selection

In order to account for correlation in the prediction error in a comprehensive and correct manner, the Bayesian scheme is applied to estimate the most probable prediction error model class based on the available data. A zero-mean Gaussian prediction error is adopted, where it is assumed that the eigenfrequency discrepancies are independent from the mode shape discrepancies, meaning that the covariance matrix Σ_{η} can be constructed as the $\Sigma_{\eta} = \text{blkdiag}(\Sigma_{\lambda}, \Sigma_{\phi})$. The eigenfrequency covariance matrix is assumed to be diagonal, and parameterized as $\Sigma_{\lambda} = \theta_{\lambda}^2 \operatorname{diag}(\tilde{\lambda}_1^2, \ldots, \tilde{\lambda}_{N_m}^2)$. For the mode shapes, it is suspected that the covariance matrix for the mode shape components can be constructed as the block diagonal matrix of N_m individual covariance matrices: $\Sigma_{\phi} = \operatorname{blkdiag}(\Sigma_{\phi}^1, \ldots, \Sigma_{\phi}^{N_m})$. A set of three alternative prediction error model classes is determined for Σ_{ϕ} : an uncorrelated model class A, a model class B with an exponential correlation function and a model class C with a exponentially damped cosine correlation function. Each of these model classes is parameterized by a number of prediction error parameters as

	$\theta_{\rm M}$ [GPa]	$ heta_{\lambda}$	$\theta_{\phi,1}$	$\theta^{B,r}_{\phi,2}$	$\theta_{\phi,2}^{C,r}$	$\theta^{C,r}_{\phi,3}$
k	1.50	2.04	1.56	1.24	1.24	1.24
t	45.0	0.02	0.004	4.26	8.53	6.40
MAPr	22.50	0.02	0.002	1.00	2.00	1.50
COV	0.82	0.70	0.80	0.90	0.90	0.90

Table 5: Appointed shape factors k and scale factors t characterizing the prior gamma-distributions, and corresponding Maximum A Priori points (MAPr = t(k - 1)) and coefficients of variation (COV = $\sqrt{1/k}$).

follows:

$$\Sigma_{\phi}^{A,r} = (\theta_{\phi,1}^{A,r})^2 \mathbf{I}_{N_s}$$
(37)

$$[\mathbf{\Sigma}_{\phi}^{B,r}]_{ij} = (\theta_{\phi,1}^{B,r})^2 \exp\left(-\frac{\Delta_{ij}}{\theta_{\phi,2}^{B,r}}\right)$$
(38)

$$[\mathbf{\Sigma}_{\phi}^{C,r}]_{ij} = (\theta_{\phi,1}^{C,r})^2 \exp\left(-\frac{\Delta_{ij}}{\theta_{\phi,2}^{C,r}}\right) \cos(\theta_{\phi,3}^{C,r}\Delta_{ij})$$
(39)

where Δ_{ij} represents the distance between sensors *i* and *j*. In an effort to reduce the number of error parameters, it is assumed for each model class that the error variance parameter $\theta_{\phi,1}$ is the same over all modes, e.g. $\theta_{\phi,1}^{A,1} = \ldots = \theta_{\phi,1}^{A,N_m}$. In this application, the number of incorporated modes N_m is taken equal to 4, which means that, in total, the uncorrelated model class is parameterized by 3 parameters (θ_M , θ_λ and $\theta_{\phi,1}^A$), model class B is parameterized by 7 parameters, and model class C is parameterized by 11 parameters. Each of the parameters is assumed to be Gamma-distributed a priori; the appointed shape and scale factors are listed in Table 5. The computations are performed for all sensor configurations and for each correlation model, using Markov chain Monte Carlo (MCMC) sampling where a Metropolis-Hastings sampling algorithm is used to obtain 50 000 samples of the posterior PDF.

The MAP estimates resulting from these computations are listed for all parameters in Table 6, for sensor configurations S031 and S151. Figure 9a shows the MAP estimates of the model parameter θ_M for all sensor configurations and all alternative model classes, and Figure 9b shows the corresponding posterior standard deviations. It is clear that the MAP values of the model parameter are affected by the choice of a particular correlation model class, although the values are all similar and converge to the correct value for increasing sensor numbers.

Examining the MAP values for model class B, it appears that the MAP value of the correla-

		$\theta_{\rm M}$	$ heta_\lambda^1$	$\theta^1_{\phi,1}$	$\theta^1_{\phi,2}$	$\theta_{\phi,2}^2$	$\theta^3_{\phi,2}$	$\theta^4_{\phi,2}$	$\theta^1_{\phi,3}$	$\theta^2_{\phi,3}$	$\theta^3_{\phi,3}$	$\theta^4_{\phi,3}$
А	S031	18.64	0.044	0.0021	-	-	-	-	-	-	-	-
	S151	18.77	0.046	0.0021	-	-	-	-	-	-	-	-
В	S031	18.39	0.042	0.0024	8.19	4.15	2.44	1.54	-	-	-	-
	S151	19.29	0.028	0.0020	9.32	8.41	6.28	4.71	-	-	-	-
С	S031	19.45	0.065	0.0016	16.82	20.21	15.40	10.66	0.98	1.43	1.98	2.49
	S151	19.08	0.026	0.0019	11.14	14.67	15.25	13.95	0.96	1.40	1.93	2.46

Table 6: MAP values of all parameters, for sensor configurations S031 and S151.



Figure 9: (a) MAP estimates of θ_M and (b) posterior standard deviations of of θ_M for all sensor configurations and all alternative model classes.



Figure 10: MAP estimates of (a) $\theta_{\phi,2}^{B,r}$ and (b) $\theta_{\phi,3}^{C,r}$ as a function of the number of sensors.

tion length parameter $\theta_{\phi,2}^{B,r}$ is dependent on the considered mode shape: the estimated correlation length decreases for higher order modes (Figure 10a). This is in accordance with the expectations, as the correlation length is expected to be related to the wavelength of the mode shape in question: larger wavelengths (i.e. lower order modes) imply higher correlation lengths. However, the estimated correlation length also seems to be highly dependent on the sensor configuration: it increases as the number of included sensors grows, i.e. high correlations are estimated for sensors located closely together and vice versa. This may indicate that model class B is not particularly suited to model the true correlation structure of the prediction error. Examining the MAP values of the prediction error parameters for model class C, it is clear that the values of e.g. the wavelength parameter $\theta_{\phi,3}^{C,r}$ are not sensitive to the number of sensors (Figure 10b), indicating that this model class seems to be better suited to model the true prediction error correlation. Below, it is shown how Bayesian model class selection can be applied to assess the validity of these findings, but first the posterior resolution or uncertainty reduction through the data is examined.

It is apparent from Figure 9b that the posterior uncertainty associated with the model parameter θ_M decreases steadily as the number of sensors increases, except for model class C which shows a stagnation compared to the other model classes. This is also observed for most – but



Figure 11: (a) Log evidence values and (b) log Occam values for all sensor configurations and for all correlation models.

not all – of the other parameters; in order to obtain insight into the reduction of uncertainty in all parameters, the log Occam values LE_{occam} can be examined, as it corresponds to the negative relative entropy between prior and posterior PDF, as elaborated above. The Occam values are computed using the expressions in Eqs. (27) and (28), where the posterior mean value and covariance matrix are determined based on the MCMC simulations; the obtained values are validated using the approach suggested in [43]. In Figure 11b, it is clear that both model classes A and B show decreasing log Occam values as the number of sensors increases, implying an increasing reduction of parameter uncertainty through the data. For model class C, however, a stagnation of the uncertainty reduction is observed for increasing sensor numbers. This indicates that the amount of information that can be obtained from the correlated data is bounded and that the uncertainty reduction cannot be increased unlimitedly by introducing more sensors. Intuitively, one would expect this also to occur in real applications.

When it comes to selecting the most probable model class, the log evidence values (Figure 11a) can be compared for all model classes. It appears that the correlated model classes B and C are distinctly preferred over the uncorrelated model class, where model class C is preferred overall. This can also be observed when the MAP correlation functions for S151 are plotted for model classes B and C (Figure 12); it is immediately clear that the most probable correlation functions for case C correspond much better to the initially estimated sample correlation shown in Figure 7b. The linear increase in the log evidence values reflects the linearly increasing log data fit values, which prove in this particular case to be determining in the model class selection procedure.

6. Conclusions

In this paper, the effect of correlation in the prediction error on Bayesian model updating results is studied. It is found that a correlated prediction error results in a recombination of the data and an altered weighting of prior information and observational data compared to the case of an uncorrelated prediction error. This means that correlation in the prediction error has a large influence on the posterior parameter estimates and their associated posterior uncertainties. Firstly, it is observed that correlation in the prediction error does not necessarily lead to an increase in posterior uncertainty on the parameter estimates compared to the uncorrelated case. Likewise, the



Figure 12: MAP correlation functions for S151, for (a) the exponential correlation function (model class B) and (b) the exponentially damped cosine correlation function (model class C).

assumption of an uncorrelated prediction error, in correspondence with maximum data entropy, is found not to guarantee the highest posterior uncertainty on the Bayesian parameter estimates.

This leads to the conclusion that correlation in the prediction error plays an important role in the Bayesian inference process, and should be properly accounted for when it is suspected that correlation is present in the data or corresponding prediction error. This poses major challenges, however, as in many practical applications very little information is available regarding prediction error characteristics such as the degree or type of correlation. In this paper, it is demonstrated in two illustrative applications that Bayesian model class selection can be used as an effective tool in appropriately accounting for prediction error correlation, as it allows to differentiate between several alternative model classes and select the most probable prediction error model class according to the available experimental observations. The results show that, provided sufficient data are available, the true correlation structure is accurately estimated, ensuring a more realistic joint structural-probabilistic model and corresponding Bayesian model updating results.

Appendix A. Posterior PDF for linear G_M and a Gaussian prediction error and prior PDF

Suppose the prediction model \mathbf{G}_{M} is a linear model, such that it can be written as $\mathbf{J}_{G}\boldsymbol{\theta}_{M}$. It is assumed that the probabilistic model of the prediction error $\boldsymbol{\eta} = \mathbf{J}_{G}\boldsymbol{\theta}_{M} - \tilde{\mathbf{d}}$ is known to be a zeromean Gaussian model, characterized by $\boldsymbol{\Sigma}_{\eta}$, which reduces the total parameter set to $\boldsymbol{\theta} = \boldsymbol{\theta}_{M}$. A priori, the parameter set $\boldsymbol{\theta}$ is assumed to be a Gaussian random variable characterized by a mean value $\boldsymbol{\theta}_{pr}$ and a prior covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta},pr}$. As both the prior PDF and the likelihood function are Gaussian, the posterior PDF will also have a Gaussian distribution, characterized by a mean value $\boldsymbol{\theta}_{po}$ and a covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta},po}$, which can be easily be verified to be equal to:

$$\boldsymbol{\theta}_{\mathrm{po}} = \left(\mathbf{J}_{\mathrm{G}}^{\mathrm{T}} \boldsymbol{\Sigma}_{\eta}^{-1} \mathbf{J}_{\mathrm{G}} + \boldsymbol{\Sigma}_{\theta,\mathrm{pr}}^{-1}\right)^{-1} \left(\mathbf{J}_{\mathrm{G}}^{\mathrm{T}} \boldsymbol{\Sigma}_{\eta}^{-1} \tilde{\mathbf{d}} + \boldsymbol{\Sigma}_{\theta,\mathrm{pr}}^{-1} \boldsymbol{\theta}_{\mathrm{pr}}\right)$$
(A.1)

$$\boldsymbol{\Sigma}_{\theta, \text{po}} = \left(\mathbf{J}_{\text{G}}^{\text{T}} \boldsymbol{\Sigma}_{\eta}^{-1} \mathbf{J}_{\text{G}} + \boldsymbol{\Sigma}_{\theta, \text{pr}}^{-1} \right)^{-1}$$
(A.2)

or, alternatively:

$$\boldsymbol{\theta}_{\text{po}} = \boldsymbol{\theta}_{\text{pr}} + \boldsymbol{\Sigma}_{\boldsymbol{\theta},\text{pr}} \mathbf{J}_{\text{G}}^{\text{T}} \left(\mathbf{J}_{\text{G}} \boldsymbol{\Sigma}_{\boldsymbol{\theta},\text{pr}} \mathbf{J}_{\text{G}}^{\text{T}} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}} \right)^{-1} \left(\tilde{\mathbf{d}} - \mathbf{J}_{\text{G}} \boldsymbol{\theta}_{\text{pr}} \right)$$
(A.3)

$$\boldsymbol{\Sigma}_{\theta, \text{po}} = \left[\mathbf{I} - \boldsymbol{\Sigma}_{\theta, \text{pr}} \mathbf{J}_{\text{G}}^{\text{T}} \left(\mathbf{J}_{\text{G}} \boldsymbol{\Sigma}_{\theta, \text{pr}} \mathbf{J}_{\text{G}}^{\text{T}} + \boldsymbol{\Sigma}_{\eta} \right)^{-1} \mathbf{J}_{\text{G}} \right] \boldsymbol{\Sigma}_{\theta, \text{pr}}$$
(A.4)

Note that these expressions correspond exactly to the Kalman filter equations with optimal Kalman gain [48]. It can be verified that the asymptotic expressions elaborated above in equations (9) and (10) are exact in this particular case.

Appendix B. Elaboration of the posterior expectation of the log likelihood function and of the log prior PDF

For a Gaussian prediction error and a linear prediction model, the likelihood function is given as:

$$L(\boldsymbol{\theta} \mid \tilde{\mathbf{d}}) = p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}) = \left[(2\pi)^N \det \boldsymbol{\Sigma}_{\eta} \right]^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{J}_{\mathrm{G}} \boldsymbol{\theta} - \tilde{\mathbf{d}})^{\mathrm{T}} \boldsymbol{\Sigma}_{\eta}^{-1} (\mathbf{J}_{\mathrm{G}} \boldsymbol{\theta} - \tilde{\mathbf{d}}) \right]$$
(B.1)

The posterior expectation of the log likelihood function can then be reformulated as:

$$\mathbb{E}\left[\log p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta})\right]_{\text{po}} = -\frac{1}{2}\log\left[(2\pi)^{N}\det\Sigma_{\eta}\right] - \frac{1}{2}\mathbb{E}\left[(\mathbf{J}_{\text{G}}\boldsymbol{\theta} - \tilde{\mathbf{d}})^{\text{T}}\Sigma_{\eta}^{-1}(\mathbf{J}_{\text{G}}\boldsymbol{\theta} - \tilde{\mathbf{d}})\right]_{\text{po}}$$
(B.2)

$$= -\frac{1}{2} \log \left[(2\pi)^{N} \det \Sigma_{\eta} \right] - \frac{1}{2} (\mathbb{E} [\mathbf{J}_{G} \boldsymbol{\theta}]_{po} - \tilde{\mathbf{d}})^{\mathrm{T}} \Sigma_{\eta}^{-1} (\mathbb{E} [\mathbf{J}_{G} \boldsymbol{\theta}]_{po} - \tilde{\mathbf{d}})$$
(B.3)

$$-\frac{1}{2}\mathbb{E}\left[\left(\mathbf{J}_{\mathrm{G}}\boldsymbol{\theta}-\mathbb{E}[\mathbf{J}_{\mathrm{G}}\boldsymbol{\theta}]_{\mathrm{po}}\right)^{\mathrm{T}}\boldsymbol{\Sigma}_{\eta}^{-1}\left(\mathbf{J}_{\mathrm{G}}\boldsymbol{\theta}-\mathbb{E}[\mathbf{J}_{\mathrm{G}}\boldsymbol{\theta}]_{\mathrm{po}}\right)\right]_{\mathrm{po}}$$
(B.4)

Using the notations θ_{po} and $\Sigma_{\theta,po}$ for the posterior mean value $\mathbb{E}[\theta]_{po}$ and covariance matrix $\mathbb{E}[(\theta - \theta_{po})(\theta - \theta_{po})^T]_{po}$, respectively, the above expression can be reformulated as:

$$\mathbb{E}\left[\log p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta})\right]_{\text{po}} = \log p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}_{\text{po}}) - \frac{1}{2} \mathbb{E}\left[\left(\mathbf{J}_{\text{G}}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{po}})\right)^{\text{T}} \boldsymbol{\Sigma}_{\eta}^{-1} \left(\mathbf{J}_{\text{G}}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{po}})\right)\right]_{\text{po}}$$
(B.5)

$$= \log p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}_{po}) - \frac{1}{2} \operatorname{Tr} \left\{ \left(\mathbf{J}_{G}^{T} \boldsymbol{\Sigma}_{\eta}^{-1} \mathbf{J}_{G} \right) \boldsymbol{\Sigma}_{\boldsymbol{\theta}, po} \right\}$$
(B.6)

$$= \log p(\tilde{\mathbf{d}} \mid \boldsymbol{\theta}_{po}) - \frac{1}{2} \operatorname{Tr} \left\{ \mathbf{I} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}, pr}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\theta}, po} \right\}$$
(B.7)

where use is made of the fact that the trace of a scalar is equal to that scalar, and of Eq. (A.2).

Similarly, the posterior expected value of the log prior PDF, where the prior is a Gaussian PDF characterized by covariance matrix $\Sigma_{\theta, \text{pr}}^{-1}$, can be described as:

$$\mathbb{E}\left[\log p(\boldsymbol{\theta})\right] = \log p(\boldsymbol{\theta}_{\text{po}}) - \frac{1}{2} \operatorname{Tr}\left\{\boldsymbol{\Sigma}_{\boldsymbol{\theta}, \text{pr}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\theta}, \text{po}}\right\}$$
(B.8)

 J. Mottershead, M. Friswell, Model updating in structural dynamics: a survey, Journal of Sound and Vibration 167(2) (1993) 347–375.

^[2] M. Friswell, J. Mottershead, Finite element model updating in structural dynamics, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

- [3] A. Teughels, G. De Roeck, Structural damage identification of the highway bridge Z24 by FE model updating, Journal of Sound and Vibration 278 (3) (2004) 589–610.
 - URL http://dx.doi.org/10.1016/j.jsv.2003.10.041
- [4] J. Beck, L. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, ASCE Journal of Engineering Mechanics 124 (4) (1998) 455–461.
- [5] L. Katafygiotis, J. Beck, Updating models and their uncertainties. II: Model identifiability, ASCE Journal of Engineering Mechanics 124 (4) (1998) 463–467.
- [6] G. Box, G. Tiao, Bayesian inference in statistical analysis, Addison-Wesley, 1973.
- [7] E. Jaynes, Probability Theory. The Logic of Science, Cambridge University Press, Cambridge, UK, 2003.
- [8] M. C. Kennedy, A. O. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society. Series B (Methodological) 63 (3) (2001) 425–464.
- [9] K.-V. Yuen, S.-C. Kuok, Bayesian Methods for Updating Dynamic Models, Applied Mechanics Reviews 64 (1).
- [10] J. M. Bernardo, Reference posterior distributions for Bayesian inference, Journal of the Royal Statistical Society. Series B (Methodological) 41 (2) (1979) 113–147.
- [11] E. Jaynes, Highly informative priors, Bayesian statistics 2 (1985) 329-360.
- [12] J. O. Berger, J. M. Bernardo, D. Sun, The formal definition of reference priors, The Annals of Statistics 37 (2) (2009) 905–938.
- [13] L. Katafygiotis, P. C., H. Lam, A probabilistic approach to structural model updating, Soil Dynamics and Earthquake Engineering 17 (1998) 495–507.
- [14] H. Sohn, K. Law, Bayesian probabilistic damage detection of a reinforced-concrete bridge column, Earthquake Engineering and Structural Dynamics 29 (8) (2000) 1131–1152.
- [15] J. Beck, S.-K. Au, W. Vanik, Monitoring structural health using a probabilistic measure, Computer-Aided Civil and Infrastructure Engineering 16 (2001) 1–11.
- [16] J. Ching, J. L. Beck, New Bayesian model updating algorithm applied to a structural health monitoring benchmark, Structural Health Monitoring 4 (3) (2004) 313–332.
- [17] K.-V. Yuen, J. Beck, L. Katafygiotis, Efficient model updating and health monitoring methodology using incomplete modal data without mode matching, Structural Control and Health Monitoring 13 (2006) 91–107.
- [18] K. Christodoulou, C. Papadimitriou, Structural identification based on optimally weighted modal residuals, Mechanical Systems and Signal Processing 21 (2007) 4–23.
- [19] M. Muto, J. Beck, Bayesian updating and model class selection for hysteretic structural models using stochastic simulation, Journal of Vibration and Control 14 (1–2) (2008) 7–34.
- [20] E. Ntotsios, C. Papadimitriou, P. Panetsos, G. Karaiskos, K. Perros, P. Pardikaris, Bridge health monitoring system based on vibration measurements, Bulletin of Earthquake Engineering 7 (2) (2009) 469–483.
- [21] B. Goller, G. I. Schuëller, Investigation of model uncertainties in Bayesian structural model updating, Journal of Sound and Vibration 330 (2011) 6122–6136.
- [22] E. Jaynes, Information theory and statistical mechanics, The Physical Review 106 (4) (1957) 620-630.
- [23] C. Soize, Construction of probability distributions in high dimensions using the maximum entropy principle: applications to stochastic processes, random fields and random matrices, International Journal for Numerical Methods in Engineering 75 (2008) 1583–1611.
- [24] M. C. Vuran, O. B. Akan, I. F. Akyildiz, Spatio-temporal correlation: theory and applications for wireless sensor networks, Computer Networks 45 (2004) 245–259.
- [25] C. Soize, A nonparametric model of random uncertainties for reduced matrix models in structural dynamics, Probabilistic Engineering Mechanics 15 (2000) 277–294.
- [26] C. Soize, A comprehensive overview of a non-parametric probabilistic approach of model uncertainties for predictive models in structural dynamics, Journal of Sound and Vibration 288 (2005) 623–652.
- [27] C. Soize, Random matrix theory and non-parametric model of random uncertainties in vibration analysis, Journal of Sound and Vibration 263 (4) (2003) 893–916.
- [28] M. Mehta, Random Matrices, 3rd Edition, Elsevier, San Diego, CA, 2004.
- [29] C. Soize, Generalized probabilistic approach of uncertainties in computational dynamics using random matrices and polynomial chaos decompositions, International Journal for Numerical Methods in Engineering 81 (8) (2010) 939–970.
- [30] P. Diaconis, D. Ylvisaker, Conjugate priors for exponential families, The Annals of Statistics 7 (2) (1979) 269-281.
- [31] D. Gamerman, Markov Chain Monte Carlo: stochastic simulation for Bayesian inference, Chapman & Hall, London, 1997.
- [32] H. Haario, M. Laine, A. Mira, E. Saksman, DRAM: Efficient adaptive MCMC, Statistics and Computing 16 (4) (2006) 339–354.
- [33] J. Beck, S.-K. Au, Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation, ASCE Journal of Engineering Mechanics 128 (4) (2002) 380–391.
- [34] J. Ching, Y.-C. Chen, Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class

selection, and model averaging, ASCE Journal of Engineering Mechanics 133 (7) (2007) 816-832.

- [35] C. Papadimitriou, J. Beck, L. Katafygiotis, Asymptotic expansions for reliability and moments of uncertain systems, ASCE Journal of Engineering Mechanics 123 (12) (1997) 1219–1229.
- [36] C. Papadimitriou, G. Lombaert, The effect of prediction error correlation on optimal sensor placement in structural dynamics, Mechanical Systems and Signal Processing 28 (2012) 105–127. URL http://dx.doi.org/10.1016/j.ymssp.2011.05.019
- [37] J. McFarland, S. Mahadevan, Multivariate significance testing and model calibration under uncertainty, Computer Methods in Applied Mechanics and Engineering 197 (2008) 2467–2479.
- [38] S. Cheung, T. Oliver, E. Prudencio, S. Prudhomme, R. Moser, Bayesian uncertainty analysis with applications to turbulence modeling, Reliability Engineering and System Safety 96 (2011) 1137–1149.
- [39] J. Beck, K.-V. Yuen, Model selection using response measurements: Bayesian probabilistic approach, ASCE Journal of Engineering Mechanics 130 (2) (2004) 192–203.
- [40] K.-V. Yuen, Recent developments of Bayesian model class selection and applications in civil engineering, Structural Safety 32 (5) (2010) 338–346.
- [41] J. L. Beck, Bayesian system identification based on probability logic, Structural Control and Health Monitoring 17 (7) (2010) 825–847.
- [42] S. H. Cheung, 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) (2010) 304–321.
- [43] A. Gelfand, D. Dey, Bayesian model choice: asymptotics and exact calculations, Journal of the Royal Statistical Society 56 (3) (1994) 501–514.
- [44] J. Skilling, Nested Sampling for General Bayesian Computation, Bayesian Analysis 1 (4) (2006) 833-859.
- [45] J. Beck, Bayesian updating and model class selection of deteriorating hysteretic structural models using seismic response data, in: Proceedings of the ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, Rethymno, Crete, Greece, 2007.
- [46] J. Beck, Using model classes in system identification for robust response predictions, in: Proceedings of COMP-DYN 2009, ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, Rhodos, Greece, 2009.
- [47] C. Soize, Probabilités et modélisation des incertitudes: éléments de base et concepts fondamentaux, Handed out at the séminaire de formation de l'école doctorale MODES, Paris (May 2003).
- [48] R. Kalman, A new approach to linear filtering and prediction problems, Journal of Basic Engineering, Transactions of the ASME 82D (1960) 35–45.