Bayesian Hierarchical Linear Modeling of Profile Data with Applications to Quality Control of Nanomanufacturing

Jianguo Wu, Yuhang Liu, Shiyu Zhou

Abstract— In this article, we propose to use the multiple ultrasonic attenuation profiles measured at different locations of a specimen to infer the microstructure of metal-matrix nanocomposites. We present a general framework to connect the profile data with both explanatory variables and product quality parameters for quality inference. More specifically, a hierarchical linear model with level-2 variance heterogeneity is proposed to model the relationship between ultrasonic attenuation profiles, ultrasonic frequency, and the microstructural parameters of the nanocomposites. An integrated Bayesian framework for model estimation, model selection, and inference of the microstructural parameters is proposed through blocked Gibbs sampling, intrinsic Bayes factor, and importance sampling. The effectiveness of the proposed approach is illustrated through case studies.

Note to Practitioners— This paper was motivated by the problem of inspecting and controlling the microstructural quality of aluminum alloy based nanocomposites using ultrasonic attenuation profiles. One critical quality issue in the nanocomposite manufacturing is the uniformity and completeness of the nanoparticle dispersion which directly influence the homogeneity and grain refinement of nanocomposites. The standard quality inspection technique is to use microscopic images of microstructures, which are costly and time-consuming to obtain. Multiple attenuation profiles measured at sampled locations of the specimen surface contain rich information about the microstructural quality (e.g., grain size, homogeneity) and thus can be used for quality inference and control. This paper proposes a general multi-level regression model to characterize the relationship between the functional response variable (e.g., attenuation), explanatory variables (e.g., frequency), and the microstructural parameters of products (e.g., grain size) for quality inference. An efficient general Monte Carlo simulation approach under the Bayesian framework is developed to estimate the model parameters, to select the optimal models, and to infer the microstructural quality parameters for inspection. The proposed methodology is shown to be very effective in quality inspection through numerical and case study. It may have broad applications where the quality can be sufficiently characterized by multiple profiles.

Index Terms—Hierarchical linear model, Variance heterogeneity, Monte Carlo Markov chain, Gibbs sampling, Model selection, Metal-matrix nanocomposites, Quality control, Profile monitoring

I. INTRODUCTION

There is a rapidly growing demand for structural components of high performance lightweight materials in the automotive, aerospace, and many other industries. A206-Al3O2 metal matrix nanocomposites (MMNCs) are promising lightweight metallic materials, which can be fabricated by dispersing Al3O2 nanoparticles into molten A206 alloys (93.5%-95% Al, 4.2%-5.0% Cu) in the ultrasonic cavitation based casting technologies [1, 2]. Well-dispersed Al3O2 nanoparticles in the molten A206 alloy can enhance the nucleation of both α-Al primary phase and β-Al3Cu intermetallic phase, which leads to significantly improved mechanical properties and castability, e.g., high strength, ductility, long fatigue life, and hot tearing resistance [3]. One critical quality issue in the nanocomposites manufacturing is the uniformity and completeness of the nanoparticle dispersion. Due to their high surface energy, large surface-to-volume ratio and poor wettability in molten metal, nanoparticles tend to agglomerate and cluster together, which greatly limits their effectiveness in microstructural morphology modification and property improvement [4, 5].

The conventional quality inspection method for nanocomposites is based on microscopic images, which are costly and time-consuming to obtain. To facilitate a scale-up production, it is highly desirable to develop a fast-yet-effective quality inspection and control technique. Recently Wu et al [6] investigated the feasibility of relating the ultrasonic attenuation profiles measured using spectral ratio technique [7] with the microstructures of A206-Al3O2 MMNCs for quality inspection. It was found that specimen with more homogeneous microstructure (i.e., smaller grain size, thinner and less continuous Al3Cu intermetallic phase, and well-dispersed Al3O2 nanoparticles) has lower between-curve variation of attenuation profiles measured at multiple sampled locations. Fig. 1 shows the attenuation measurement process and attenuation profiles for two fabricated nanocomposites specimens, where the specimen of Fig. 1(c) has a more
homogeneous microstructure than that of Fig. 1(b). The phenomenon of more homogeneous microstructures resulting in less variation has also been observed by Liu et al [8] through microstructural modeling and wave propagation simulation approach. Therefore, the microstructural quality of MMNCs can be characterized by multiple attenuation profiles measured at different locations. Quality characterization based on multiple profiles measured at randomly sampled locations has also been studied in low-E glass manufacturing, where the uniformity of coating on the glass surface could be captured by multiple spectral reflectance profiles [9].

Profile data has been widely used for quality monitoring and control of manufacturing processes where products quality can be sufficiently characterized by profiles [10-12]. Various mathematical models have been proposed to describe profile data, including (1) simple or multiple linear profiles (e.g., polynomials) where a single or multiple explanatory variables are used to describe the behavior of the response variable, (2) parametric nonlinear profiles, (3) nonparametric nonlinear profiles (e.g., splines, wavelets), (4) multivariate linear profiles where a set of response variables are linear functions of explanatory variables, and (5) binary response profiles. In most of the existing profile data modeling approaches, the covariance matrix of coefficients is assumed to be unchanged, and only a single profile is modeled. Very limited research has been conducted to model the covariance matrix or the between-profile variations. To characterize the uniformity of a product such as the nanocomposites, however, a single profile is often not enough, and multiple profiles of the same response variable measured at multiple sampled locations of a product have to be used and modeled.

To fill the gap, we propose a hierarchical linear model (HLM), or more specifically a two-level model, with heterogeneous level-2 variances to infer the microstructural quality in the nanocomposites manufacturing. There are three parts in the proposed model: (1) the profile is modeled as a linear function of explanatory variables in level-1; (2) the coefficients in level-1 are modeled as linear functions of microstructural parameters in level-2 with diagonal residual covariance matrix; (3) the residual variances in level-2 are modeled as log-linear function of microstructural parameters. The model with only the first two parts is a hierarchical linear model, which is a variant term for multilevel model or for what are broadly called linear mixed-effects (LME) model. In this paper, we use the term HLM instead of LME to differentiate the new model from the common LME model used in [9]. The third part is an embedded variance regression model to characterize the heterogeneity of the variance of coefficients across different microstructural parameters. Compared with the traditional LME model used in profile monitoring [9, 13], our model has the advantage of directly relating the microstructural parameters with profiles for quality inference.

HLM has been widely used to model hierarchically structured data in the biomedical and social research [14, 15]. Extensions of the HLM with heterogeneous within-profile noise variances (e.g., residual variance in level-1 model) have also been intensively studied [16-18]. However, there is very limited work on modeling heterogeneous variances for random effects. For the standard HLM, The model parameters can be estimated using two general methods, maximum likelihood (ML) and restricted maximum likelihood (REML) [19]. However, these methods cannot be directly applied to the proposed model, as the addition of the third part of the model makes the optimization much more complicated. The remaining challenges of the proposed model are model selection (e.g., determining the degree of polynomial, which coefficient is random), and inference of the microstructural parameters of new profiles in real-time monitoring. In this paper, we propose to use Markov chain Monte Carlo (MCMC) approach under the Bayesian framework, which can not only estimate the model parameters, but also perform model

![Fig. 1. Attenuation measurement and attenuation profiles measured at sampled locations: (a) spectral ratio technique for attenuation profile measurement; (b) A206 nanocomposites with 1wt% of AlO3 nanoparticles, (c) A206 nanocomposites with 5wt% of AlO3 nanoparticles.](image)
selection and inference of the microstructural parameters efficiently and accurately.

The remainder of the paper is organized as follows. In Section II the new HLM with heterogeneous level-2 variances is formulated. The MCMC estimation of model parameters and model selection are given in Section III and Section IV respectively. Section V evaluates the performance of model selection and estimation through numerical simulations. Section VI presents the case study where the proposed model is applied to the ultrasonic attenuation profiles of MMNCs. The conclusions and discussions are given in Section VII.

II. TWO-LEVEL HIERARCHICAL LINEAR MODEL WITH HETEROGENEOUS LEVEL-2 VARIANCES

\[ y_{ij} = H_1(x)H_2(\theta_i)\beta + H_1(x)\xi_{ij} + \epsilon_{ij} \]  

(4)

where \( y_{ij} \) is the profile error term, \( \beta \) is a \( q \times 1 \) vector of coefficients, \( \xi_{ij} \) is the error term, which is a random vector following i.i.d. p-dimensional Gaussian distribution for each \( i \):

\[ \xi_{ij} \sim N(0, \Sigma) \]  

(3)

Submodel (2) is used to model the dependence of coefficients in submodel (1) on the microstructural parameter \( \theta \) by the mean term \( H_2(\theta_i)\beta \) and to account for variation among profiles of the same specimen by the error term \( \xi_{ij} \). Combining (1) and (2) we obtained the general LME model as

\[ \Sigma_i = \text{diag}(\sigma_{\theta 1}^2(\theta_i), ..., \sigma_{\theta q}^2(\theta_i)) \]

(5)

where \( \Sigma_i \) is the variance heterogeneity for the profile error term, \( \Sigma_i \) is a \( r \times 1 \) vector of 2s, \( \delta_{d, d} \) is a \( r \times 1 \) vector of coefficients, and \( \delta_{d, d} \sim N(0, \sigma_{\delta d}^2) \). This part is used to model the heterogeneity of residual variance in (2). We select the log-linear model here as it is commonly used in variance function regression or heteroscedastic regression [20, 21]. Note that we ignore the correlation components in the covariance matrix \( \Sigma_i \), which is a common way to reduce the model complexity in Bayesian hierarchical models [9, 22]. If we consider all the correlation components, there will be \( p(p + 1)/2 \) submodels in the variance heterogeneity. Although ignoring correlations may give a biased estimate of covariance matrix, it can significantly simplify the model and avoid the errors caused by estimating a large number of parameters.

In the proposed model, level-1 is to model each individual profile or within-profile variations, level-2 is to model both the model heterogeneity across different specimens and the between-profile variations within each specimen, and the log-linear model is to capture the residual variance heterogeneity of level-2 model. Note that our work share some similarity with the work proposed by Castillo et al [23] to optimize the shape of profiles with both controllable factors and noise factors. The main differences between these two are that in the model proposed here all design parameters are controllable and the extra variance heterogeneity model in the second level is considered. Therefore in our model multiple profiles at each design point \( \theta_i \) are needed for model estimation and the model estimation is much more complicated.

After the new model is proposed, the remaining issues are how to efficiently estimate the model parameters and how to accurately select the right models among a set of candidate ones. In the model estimation, the parameters of interest include the fixed effects \( \beta \), within-profile error term variance \( \sigma_{\epsilon i}^2 \), variance component regression coefficients \( \{\beta_d, d = 1, ..., p\} \)
and error term variances \( \{ \sigma_d^2, d = 1, \ldots, p \} \). Denote \( \psi = \{ \beta, \sigma^2, \{ \gamma_d, \{ \sigma_d^2 \} \} \). The likelihood function for the model can be expressed by integrating out the nuisance parameters, i.e., all unobservable random effects \( \xi = \{ \xi_{ij}, i = 1, \ldots, m, j = 1, \ldots, l \} \) and variance components \( \sigma^2 = \{ \sigma_j^2(\theta_i), i = 1, \ldots, m, d = 1, \ldots, p \} \), as

\[
L(\psi|Y) = \int f(Y|\psi, \xi, \sigma^2) f(\xi|\sigma^2, \psi) f(\sigma^2|\psi) d\sigma^2 d\xi
\]

where \( Y \) is the vector of all observations, \( Y = (y_{11}, y_{12}, \ldots, y_{1d}, \ldots, y_{m1}) \). Equation (6) involves high dimensional integration and is not analytically tractable, which makes the maximum likelihood estimation very challenging. In the research, we propose to estimate the model parameters under the Bayesian framework. The posterior distribution of the models parameters are approximated using blocked Gibbs sampling method, which will be given in detail in the following section. The second issue is model selection, where the predictor variables, or the degrees of polynomials if polynomial regression is used, for all three submodels have to be determined. Section 4 will present it in detail.

III. BAYESIAN MODEL ESTIMATION USING BLOCKED GIBBS SAMPLER

A. Specification of Priors

In the Bayesian analysis of the proposed model, the priors for the mean parameters \( \beta \), and \( \{ \gamma_d, d = 1, \ldots, p \} \), and variance parameters \( \sigma^2 \) and \( \{ \sigma_d^2, d = 1, \ldots, p \} \) need to be specified. For the mean parameters, the normal priors and noninformative priors are most commonly used in the Bayesian linear regression [24]. The normal priors often provide the benefit of conjugacy in the simple linear regression or conditional conjugacy, i.e., conjugate prior conditioning on other model parameters, in the hierarchical linear regression. However, in most cases the prior information beyond the data is not available, and thus the noninformative prior is more preferred, which provides both objectiveness and convenience in Bayesian analysis. In this research, we specify noninformative priors for \( \beta \) and \( \{ \gamma_d, d = 1, \ldots, p \} \) as

\[
\pi(\beta) \propto 1
\]

\[
\pi(\gamma_d) \propto 1, d = 1, \ldots, p
\]

For the variance components, there is a lot of literature discussing how to select appropriate priors [9, 25, 26]. Two types of priors have been widely used, the noninformative prior of the form

\[
\pi(\sigma^2) \propto (\sigma^2)^{-(\alpha + 1)}
\]

and the weakly-informative inverse gamma prior

\[
\pi(\sigma^2) \propto IG(\omega, \omega)
\]

In the noninformative prior, \( \alpha = 0 \) corresponds to a uniform prior on \( \log \sigma \), i.e., \( \pi(\log \sigma) \propto 1 \), or equivalently \( \pi(\sigma) \propto 1/\sigma \). \( \alpha = -1/2 \) corresponds to a uniform prior on \( \sigma \), i.e., \( \pi(\sigma) \propto 1 \). \( \alpha = -1 \) corresponds to a uniform prior on \( \sigma^2 \). For the weakly-informative prior, the inverse-gamma distribution is within the conditionally conjugate family, with \( \omega \) set to a low value, e.g., 1, 0.1 or 0.001. Zeng et al [9] used the weakly-informative prior for the variance components of the random effects in LME model to facilitate the computation in model selection. However, Gelman [26] showed that the inferences become very sensitive to \( \omega \) for datasets in which low values of random effects variance are possible, and the prior distribution hardly looks noninformative. In this paper, we specify noninformative priors for both \( \sigma^2 \) and \( \{ \sigma_d^2, d = 1, \ldots, p \} \) for convenience and objectiveness:

\[
\pi(\sigma^2) \propto 1
\]

\[
\pi(\sigma_d^2) \propto 1, d = 1, \ldots, p
\]

B. Blocked Gibbs Sampling for Posterior Estimation

Under the Bayesian framework, the model estimation is to calculate the posterior distribution of the model parameters conditioning on the observations. Once the posterior is obtained, we can either use the mean or median of the posterior as the point estimates of model parameters, or directly use the posterior distribution for future model selection and inference. In this research, the posterior distribution of interest is

\[
P(\psi|Y) = P(\beta, \sigma^2, \{ \gamma_d \}, \{ \sigma_d^2 \}|Y)
\]

\[
= \pi(\beta, \sigma^2, \{ \gamma_d \}, \{ \sigma_d^2 \})L(\beta, \sigma^2, \{ \gamma_d \}, \{ \sigma_d^2 \}|Y)
\]

where \( L(\beta, \sigma^2, \{ \gamma_d \}, \{ \sigma_d^2 \}|Y) \) is the likelihood function with full expression shown in Equation (6). As the nuisance parameters, i.e. random effects \( \xi = \{ \xi_{ij}, i = 1, \ldots, m, j = 1, \ldots, l \} \) and variance components \( \sigma^2 = \{ \sigma^2(\theta_i), i = 1, \ldots, m, d = 1, \ldots, p \} \) are not observable, the joint posterior distribution including all nuisance parameters need to be found and the posterior of interest can be obtained by marginalizing out all nuisance parameters. The joint posterior distribution is expressed as

\[
P(\beta, \sigma^2, \{ \gamma_d \}, \{ \sigma_d^2 \}, \xi, \sigma^2|Y)
\]

\[
\propto \pi(\beta, \sigma^2, \{ \gamma_d \}, \{ \sigma_d^2 \}) f(\sigma^2|\{ \gamma_d \}, \{ \sigma_d^2 \}) f(\xi|\sigma^2)
\]

\[
\times f(Y|\beta, \sigma^2, \{ \gamma_d \}, \{ \sigma_d^2 \}, \xi)
\]

Since the joint posterior cannot be directly sampled, Markov Chain Monte Carlo (MCMC) simulation has to be used. Gibbs sampling [9, 24, 25] is one of the most popular MCMC methods to estimate hierarchical models. It generates sequence of random samples that approximately follow the target posterior distribution when the direct sampling is difficult. The basic idea is to repeatedly replace the value of each component with a sample from its distribution conditioning on the current values of all other components. In this paper, we propose to use
the blocked Gibbs sampler [27], a more efficient version of
Gibbs sampler, where the variables are grouped into blocks,
and each entire block is sampled together from its joint
conditional distribution given the other components. For the
standard Bayesian LME model all conditional distributions
can be directly sampled [9]. However, due to the log-linear
heterogeneity variance regression, the nuisance parameters \( \sigma^2_i \)
in our model cannot be directly sampled through the
conditional distribution. To overcome this problem we
developed a Metropolis-Hastings [28] algorithm to sample \( \sigma^2_i \)
in the Gibbs sampling process.

In the sampling procedure, the parameters including those of
interest and nuisance parameters can be divided into 4 groups:

**G1:** The fixed effects \( \beta \) and within-profile variance of random
error \( \sigma^2_e \)

**G2:** The random effects \( \xi = \{ \xi_{ij}, i = 1, ..., m, j = 1, ..., l \} \)

**G3:** The variance components \( \sigma^2_i = \{ \sigma^2_d(\theta_j), i = 1, ..., m, d = 1, ..., p \} \)

**G4:** The variance heterogeneity regression coefficients \( \{ \sigma^2_d, d = 1, ..., p \} \) and variance of random error \( \{ \sigma^2_d, d = 1, ..., p \} \)

The blocked Gibbs sampling procedure can be summarized
using the following steps:

**Step 1:** Sampling G1 parameters from their conditional
posterior distribution

\[
P(\beta, \sigma^2_e | \{ y_d \}, \{ \sigma^2_d \}, \xi, \sigma^2_i, Y) = P(\beta, \sigma^2_e | \xi, Y)
\]

**Step 2:** Sampling G2 parameters from their conditional
posterior distribution

\[
P(\xi | \beta, \sigma^2_e, \{ y_d \}, \{ \sigma^2_d \}, \sigma^2_i, Y) = P(\xi | \beta, \sigma^2_e, \sigma^2_i, Y)
\]

**Step 3:** Sampling G3 parameters from their conditional
posterior distribution

\[
P(\sigma^2_i | \beta, \sigma^2_e, \{ y_d \}, \{ \sigma^2_d \}, \xi, Y) = P(\sigma^2_i | \{ y_d \}, \{ \sigma^2_d \}, \xi)
\]

**Step 4:** Sampling G4 parameters from their conditional
posterior distribution

\[
P(\{ y_d \}, \{ \sigma^2_d \} | \beta, \sigma^2_e, \sigma^2_i, Y) = P(\{ y_d \}, \{ \sigma^2_d \} | \sigma^2_i)
\]

By iteratively drawing samples from conditional posterior
distribution in the above four steps, a sequence of samples will
be obtained, which constitutes a Markov chain with the
stationary distribution following the joint posterior distribution
of interest. Note that in Step 1 and Step 4 the regression
coefficient and the random error variance are sampled from the
joint conditional posterior distribution, which is more efficient
than sampling from each one individually, e.g., sampling from
\( P(\beta | \sigma^2_e, \xi, Y) \) and \( P(\sigma^2_e | \beta, \xi, Y) \). The following subsection
presents the detailed conditional posterior distribution for
blocked Gibbs sampling.

C. Conditional Posterior Distribution for Gibbs Sampling

This subsection will show the conditional posterior
distributions corresponding to the four steps in Subsection 3.2
for Gibbs sampling. The Metropolis-Hastings algorithm used in
Step 3 will also be proposed.

(1) \( P(\beta, \sigma^2_e | \{ y_d \}, \{ \sigma^2_d \}, \xi, \sigma^2_i, Y) = P(\beta, \sigma^2_e | \xi, Y) \)

Let \( H \) be the stack of \( \{ H_1(\chi)H_2(\theta_j) \} \), \( \Xi \) be the stack of
\( \{ H_1(\chi) \} \), and \( E \) be the stack of \( \{ e_i \} \).

\[
H = \begin{bmatrix}
H_1(\chi)H_2(\theta_1) \\
H_1(\chi)H_2(\theta_2) \\
\vdots \\
H_1(\chi)H_2(\theta_m)
\end{bmatrix}, \Xi = \begin{bmatrix}
H_1(\chi)\xi_{11} \\
H_1(\chi)\xi_{21} \\
\vdots \\
H_1(\chi)\xi_{ml}
\end{bmatrix}, E = \begin{bmatrix}
e_{11} \\
e_{12} \\
\vdots \\
e_{ml}
\end{bmatrix}
\]

then

\[
Y - \Xi = H\beta + E \tag{13}
\]

where \( \pi(\beta) \propto 1, E \sim N(0, \sigma^2_e I_{m \times p}) \) and \( \pi(\sigma^2_e) \propto 1 \). Given
\( \{ \xi_{ij}, i = 1, ..., m, j = 1, ..., l \} \) or \( \Xi \), Equation (13) is a simple
linear model. The joint conditional posterior distribution can be
written as

\[
P(\beta, \sigma^2_e | \xi, Y) = P(\sigma^2_e | \xi, Y)P(\beta | \xi, \sigma^2_e, Y)
\]

It can be easily proven that [24]

\[
\sigma^2_e(\xi, Y) \sim IG\left(\frac{mnl - 2}{2}, \frac{(Y - H\hat{\beta} - \Xi)'(Y - H\hat{\beta} - \Xi)}{2}\right)
\]

\[
\hat{\beta} | (\sigma^2_e, \xi, Y) \sim N(\hat{\beta}, \sigma^2_e (H' H)^{-1})
\]

where

\[
\hat{\beta} = (H' H)^{-1} H' (Y - \Xi)
\]

(2) \( P(\xi | \beta, \sigma^2_e, \{ y_d \}, \{ \sigma^2_d \}, \sigma^2_i, Y) = P(\xi | \beta, \sigma^2_e, \sigma^2_i, Y) \)

Given all other parameters, the random effects \( \{ \xi_{ij}, i = 1, ..., m, j = 1, ..., l \} \) are independent. Therefore they can be
sampled individually. The distribution for each component \( \xi_{ij} \)
is

\[
P(\xi_{ij} | \beta, \sigma^2_e, \sigma^2_i, Y) = P(\xi_{ij} | \beta, \sigma^2_e, \Sigma_i, y_{ij})
\]
\[ \propto \pi(\xi_{ij} | \Sigma_i)P(y_{ij} | \beta, \sigma^2_{\varepsilon}, \xi_{ij}) \]
\[ = N(\xi_{ij} | 0, \Sigma_i) \cdot N(y_{ij} | H_1(x)H_2(\theta_i)\beta + H_1(x)\xi_{ij}, \sigma^2_{\varepsilon}I_n) \]

It can be shown that the conditional posterior distribution of \( \xi_{ij} \) follows multivariate normal distribution [5]:
\[ \xi_{ij} | \beta, \sigma^2_{\varepsilon}, \Sigma_i, y_{ij} \sim N(\tilde{\xi}_{ij}, \Sigma_i), i = 1, \ldots, m, j = 1, \ldots, l \quad (15) \]
where
\[
\tilde{\xi}_{ij} = \left[ H'_1H_1 + \sigma^2_{\varepsilon}\Sigma_i^{-1}\right]^{-1}\left( H'_1(y_{ij} - H_1(x)H_2(\theta_i)\beta) \right)
\]
\[ \Sigma_i = \sigma^2_{\varepsilon}\left[ H'_1H_1 + \sigma^2_{\varepsilon}\Sigma_i^{-1}\right]^{-1} \]

(3) \[ P(\sigma^2_{\varepsilon} | \beta, \sigma^2_{\varepsilon}, \{y_d\}, \{\sigma^2_{\delta_d}\}, \xi, Y) = \mathcal{P}(\sigma^2_{\varepsilon} | \{y_d\}, \{\sigma^2_{\delta_d}\}, \xi) \]

Given all other parameters, the variance components \( \sigma^2_{\varepsilon} = \{\sigma^2_{\varepsilon}(\theta_i), i = 1, \ldots, m, d = 1, \ldots, p \} \) are independent, which can be sampled individually. In this research, we sample \( p \) components \( \{\sigma^2_{\varepsilon}(\theta_i), d = 1, \ldots, p \} \) simultaneously each time for the purpose of convenience. Let \( \eta_{id} = \log(\sigma^2_{\varepsilon}(\theta_i)) \), \( \eta_i = (\eta_{i1}, \ldots, \eta_{ip})', H_3(\theta_i) = I_p \otimes h^T_3(\theta_i), \theta = (\theta_1, \ldots, \theta_p)' \), then
\[ \eta_i = H_3(\theta_i)\gamma + \delta \]
where \( \delta = (\delta_1, \ldots, \delta_p)' \sim N(0, \text{diag}((\sigma^2_{\delta_d}, d = 1, \ldots, p))) \). The conditional posterior of \( \eta_i \) is
\[ \mathcal{P}(\eta_i | [\xi_{ij}, j = 1, 2, \ldots, l], \{\sigma^2_{\delta_d}\}, Y) \]
\[ \propto \mathcal{P}(\eta_i | [\sigma^2_{\delta_d}], Y)P([\xi_{ij}, j = 1, 2, \ldots, l] | \eta_i) \quad (16) \]

where
\[ \mathcal{P}(\eta_i | [\xi_{ij}, j = 1, 2, \ldots, p], y) \sim N \left( \frac{1}{\sigma_{\delta_d}} \text{diag}(\{\sigma^2_{\delta_d}, d = 1, \ldots, p\}) \right) \]
and
\[ P([\xi_{ij}, j = 1, 2, \ldots, l] | \eta_i) \]
\[ \propto \left( \prod_{d=1}^{p} \exp(\eta_{id}) \right)^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \sum_{j=1}^{l} \xi_{ij} \Sigma_i^{-1} \xi_{ij} \right) \]

Since the conditional posterior in (16) cannot be sampled directly, we propose the Metropolis-Hastings algorithm as follows. At each iteration \( t \) in the Gibbs sampling, we propose a new value for \( \eta_i \) using a symmetric normal proposal distribution centered at the current value as
\[ \eta_i^* = \eta_i^{(t-1)} + N(0, \Delta^2 I_p) \quad (17) \]
where \( \eta_i^{(t-1)} \) is \( \eta_i \) at step \( t - 1 \), \( \Delta \) is the step-width of the random walk or the standard deviation of the proposal distribution.
The Metropolis-Hastings acceptance ratio is given by
\[ r_a = \min \left\{ 1, \frac{P(\eta_i | [\sigma^2_{\delta_d}]^{(t-1)}, \gamma^{(t-1)})P([\xi_{ij}^{(t)}, j = 1, 2, \ldots, l], [\eta_i^{(t-1)}])}{P(\eta_i^{(t-1)} | [\sigma^2_{\delta_d}]^{(t-1)}, \gamma^{(t-1)})P([\xi_{ij}^{(t)}, j = 1, 2, \ldots, l], [\eta_i^{(t-1)}])} \right\} \quad (18) \]
Set \( \eta_i^{(t)} = \eta_i^* \) with probability \( r_a \), otherwise set \( \eta_i^{(t)} = \eta_i^{(t-1)} \).

**Algorithm 1: Blocked Gibbs Sampling Algorithm for Model Estimation**

- Select the iteration number \( N_g \).
- Specify step-width \( \Delta \).
- Specify the initial value \((\beta^{(0)}, (\sigma^2_{\varepsilon})^{(0)}, (\xi^{(0)}), (\sigma^2_{\varepsilon})^{(0)}, y^{(0)}, (\sigma^2_{\delta_d}, d = 1, \ldots, p))^{(0)} \).

For \( t = 1: N_g \):

1. Sample \( (\sigma^2_{\varepsilon})^{(t)} \sim P(\sigma^2_{\varepsilon} | (\xi^{(t-1)}), Y) \) and \( \beta^{(t)} \sim P(\beta | (\xi^{(t-1)}, (\sigma^2_{\varepsilon})^{(t)}, Y)) \) using Eq. (14).
2. Sample \( \xi_{ij}^{(t)} \sim P(\xi_{ij} | \beta^{(t)}, (\sigma^2_{\varepsilon})^{(t)}, \Sigma_i^{(t-1)}, y_{ij}) \) for \( i = 1, \ldots, m, j = 1, \ldots, l \) using Eq. (15).
3. Metropolis-Hastings algorithm for \( (\sigma^2_{\varepsilon})^{(t)} \)

   For \( i = 1: m \):
   - Generate \( \eta_i^* \) using Eq. (17).
   - Calculate \( r_a \) based on Eq. (18).
   - Sample \( u \sim U(0,1) \); if \( u \leq r_a \), set \( (\sigma^2_{\varepsilon})^{(t)} = \exp(\eta_i) \); else set \( (\sigma^2_{\varepsilon})^{(t)} = (\sigma^2_{\varepsilon})^{(t-1)} \).
4. Sample \( (\sigma^2_{\delta_d})^{(t)} \sim P(\sigma^2_{\delta_d} | V_d^{(t)}) \) and \( y_{d}^{(t)} \sim P(y_{d} | V_d^{(t)}, (\sigma^2_{\delta_d})^{(t)}) \) using (19) for \( d = 1, \ldots, p \).

End.
(4) \( P(\{y_d\}, \{\sigma^2_{d|a}\} | \beta, \sigma^2, \sigma^2_\xi, \xi) = P(\{y_d\}, \{\sigma^2_{d|a}\} | \sigma^2_\xi) \)

Since \( (y_d, \sigma^2_{d|a}) \) is independent of \( (y_d', \sigma^2_{d'|a}) \) for \( d \neq d' \), the joint conditional posterior distribution of \( (y_d, \sigma^2_{d|a}) \), which is similar to (14), can be sampled individually. Let \( V_d = (\log \sigma^2_\theta(\theta_1), ..., \log \sigma^2_\theta(\theta_m))', H_4 \) is the stack of \( (h_3(\theta_i)^T, i = 1, ..., m) \), then similar to (14), the conditional posterior distributions follow the distribution as

\[
\sigma^2_{d|a|V_d = \text{IG}} \left( \frac{m - 2}{2}, \frac{(V_d - H_4 \hat{y}_d)'(V_d - H_4 \hat{y}_d)}{2} \right) \\
y_d|V_d, \sigma^2_{d|a} \sim N(\hat{y}_d, \sigma^2_{d|a}(H_4'H_4)^{-1})
\]

where \( \hat{y}_d = (H'_4 H_4)^{-1} H'_4 V_d \)

The overall blocked Gibbs sampling is shown in the Algorithm I. To speed up the convergence efficiency, the initial value for all the models parameters can be set using multiple-stage analysis, i.e., fitting linear regression for each profile and treat each coefficient as sampled in the level-2 model fitting, and then use the sample residual variance of level-2 model as the responses in the variance regression. After the iteration of the Gibbs sampling is finished, the obtained samples can be truncated to remove the initial bias for the posterior estimation.

IV. MODEL SELECTION AND MICROSTRUCTURAL PARAMETER INFERENCE

A. Model Selection using Intrinsic Bayes Factor

The most popular model selection methods are the information criteria based methods, such as Akaike Information Criteria (AIC; [29]) and Bayesian Information Criteria (BIC; [30]), where the criteria is to find a model that minimizes an estimate of a criterion consisting of a loss function (e.g., \(-2 \times \log\)-likelihood) and a penalty function. These methods are commonly used in linear regressions, where the penalty function is a function of model complexity, or number of parameters. However, for the model proposed here, there are both regression parameters and variance parameters at different levels, which have different relative importance in analysis. It is very challenging to incorporate the relative importance into the penalty function of the information criteria.

The Bayes factor (BF) is a very flexible model selection method that can compare models of any forms [31]. For two competing models \( M_i \) and \( M_j \), \( i \neq j \), the BF of \( M_i \) to \( M_j \) is defined as the ratio of the observed marginal densities

\[
B_{ij} = \frac{P(\{y\} | M_i)}{P(\{y\} | M_j)} = \frac{\int P(\{y\} | \psi_i, M_i) \pi(\psi_i | M_i) d\psi_i}{\int P(\{y\} | \psi_j, M_j) \pi(\psi_j | M_j) d\psi_j}
\]

where \( P(\{y\} | M_i) \) is the marginal or predictive densities of \( Y \), \( \psi_i \) is the vector of model parameters and \( \pi(\psi_i | M_i) \) is the prior density function of model parameters under model \( M_i \). It can also be interpreted as the weighted likelihood ratio of \( M_i \) to \( M_j \), with the priors being the “weighting functions”. Intuitively, higher \( B_{ij} \) indicates a stronger evidence of \( M_i \) against \( M_j \). A set of cutoff values of \( B_{ij} \) has been suggested and widely used in literature [9, 32], as shown in Table I.

<table>
<thead>
<tr>
<th>( B_{ij} )</th>
<th>2 ( \log(B_{ij}) )</th>
<th>Evidence against ( M_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–3</td>
<td>0–2</td>
<td>Barely worth mentioning</td>
</tr>
<tr>
<td>3–20</td>
<td>2–6</td>
<td>Positive</td>
</tr>
<tr>
<td>20–150</td>
<td>6–10</td>
<td>Strong</td>
</tr>
<tr>
<td>&gt;150</td>
<td>&gt;10</td>
<td>Very strong</td>
</tr>
</tbody>
</table>

Although BF is very flexible, a direct computation is very challenging, since the marginal density involves integration over the parameter space of high dimension. A natural approach to solve this issue is MCMC simulation, where two popular methods are used, the product space search and the marginal likelihood estimation method [33]. When the two models have parameters of different dimensions, however, using improper noninformative priors for all models parameters will result in indeterminate BFs, as the marginal density \( P(\{y\} | M_i) = \int P(\{y\} | \psi_i, M_i) \pi(\psi_i | M_i) d\psi_i \) is not well-defined for \( \pi(\psi_i | M_i) \propto 1 \). To see how this happens, suppose \( \pi(\psi_i | M_i) \propto 1 \) and \( \pi(\psi_j | M_j) \propto 1 \) are used as priors for \( M_i \) and \( M_j \) respectively. Then \( c_i \pi(\psi_i | M_i) \) and \( c_j \pi(\psi_j | M_j) \) can also be used as improper priors, which results in another BF \( c_i/c_j B_{ij} \).

In this paper, we propose to use the intrinsic Bayes factor (IBF) [34] for model selection.

Let \( Y(s) \) denote the training profiles and \( Y(-s) \) denote the remaining profiles for testing (the partition is based on the value \( \theta \) or based on the specimen). The basic idea of IBF is to use the training profiles \( Y(s) \) to convert the improper noninformative priors to proper posterior distributions and then to compute the BF with the remainder of the profiles \( Y(-s) \). The IBF can be expressed as

\[
I_B ij = \frac{P(\{Y(-s)\} | Y(s), M_i)}{P(\{Y(-s)\} | Y(s), M_j)} = \frac{\int P(\{Y(-s)\} | \psi_i, M_i) \pi(\psi_i | Y(s)) d\psi_i}{\int P(\{Y(-s)\} | \psi_j, M_j) \pi(\psi_j | Y(s)) d\psi_j}
\]

As we can see, the marginal density in IBF is quite similar to cross-validation techniques commonly used in the model validation. Naturally, we can partition all the profiles into several groups and calculate the IBF using each group of profiles as the testing profiles and the remainder as the training profiles. By averaging all the IBFs, we can get a more stable IBF.

Although the marginal density is well-defined with proper posterior distribution \( \pi(\psi_i | Y(s)) \), the direct computation is still challenging. With the availability of posterior samples of \( \pi(\psi_i | Y(s)) \) obtained from the blocked Gibbs sampler, we can compute the IBF using MCMC approach as follows. The marginal density for \( M_i \) is written as
\[ m_i(Y(-s)|Y(s)) = \int f(Y(-s)|\psi_i(\sigma^2_i(-s)),\sigma^2_i(-s))\pi(\psi_i|Y(s))d(\psi_i(\sigma^2_i(-s))) \]

Here \((\sigma^2_i(-s))_i\) denote the variances of random effects of the testing profiles \(Y(-s)\) under the mode \(M_i\). Suppose the posterior samples of the training data \(Y(s)\) obtained in the Gibbs sampling are

\[
\left\{\left(\psi_i^{(g)}, \xi(s)^{(g)}_i, \sigma^2_i(s)^{(g)}_i\right), g = 1, 2, \ldots, G\right\}
\]

Then for \(g = 1, \ldots, G\), we can sample \((\sigma^2_i(-s))^{(g)}_i\) conditioning on \(\psi^{(g)}\) through the lognormal distribution, as shown in (5). The marginal density could be estimated by

\[
m_i(Y(-s)|Y(s)) \approx \frac{1}{G} \sum_{g=1}^{G} f(Y(-s)|\psi_i^{(g)}, (\sigma^2_i(-s))^{(g)}_i)
\]

where the profile \(y(\theta)\) given \(\psi\) and \(\sigma^2_i\) follows normal distribution based on (4):

\[ y(\theta) | \psi, \Sigma(\theta) \sim N(H_1(x)H_2(\theta)\beta, H_3(x)\Sigma(\theta)H_3(x)^T + \sigma^2 I) \]

**B. Inference on the Microstructural Parameter**

After the optimal model is selected and estimated through the Gibbs sampler and IBF, we can use it to infer the microstructural parameters \(\theta\) for quality control and diagnosis. Suppose the measured profiles for a new specimen are \(Y_{\text{new}}\), then the posterior of \(\theta_{\text{new}}\) given \(Y_{\text{new}}\) and model parameters is of interest. If we use the mean or median of the posterior distributions of \(\psi\) as the point estimate of the model parameters, denoted as \(\bar{\psi}\), then the posterior of \(\theta_{\text{new}}\) is

\[ P(\theta_{\text{new}}|Y_{\text{new}}, \bar{\psi}) \]

Alternatively, we could use all the Gibbs samples instead of the point estimate for the model parameters.

**Algorithm II. Inference of \(\theta\) using Importance Sampling**

Specify the number of samples \(N_s\)

1. Draw samples \(\theta^{(1)}, \ldots, \theta^{(N_s)}\) from \(\pi(\theta_{\text{new}})\)
2. Calculate the importance weight of each sample using (22)

\[ w^{(j)} = P(Y_{\text{new}}|\theta^{(j)}, Y), j = 1, \ldots, N_s \]

3. Approximate the expectation

\[ E_P(\theta|Y_{\text{new}}, Y)(h(\theta)) = \frac{\sum_{j=1}^{N_s} w^{(j)}h(\theta^{(j)})}{\sum_{j=1}^{N_s} w^{(j)}} \]

The posterior is expresses as

\[ P(\theta_{\text{new}}|Y_{\text{new}}, Y) \propto \pi(\theta_{\text{new}})P(Y_{\text{new}}|\theta_{\text{new}}, Y) \]

where \(\pi(\theta_{\text{new}})\) is the prior distribution for \(\theta\). To estimate this posterior, the importance sampling [35] can be applied where the prior \(\pi(\theta_{\text{new}})\) is selected as the importance distribution and \(P(Y_{\text{new}}|\theta_{\text{new}}, Y)\) is the weight function, which can be estimated using (22). The expectation of \(h(\theta_{\text{new}})\) with respect to \(P(\theta_{\text{new}}|Y_{\text{new}}, Y)\) where \(h(\theta_{\text{new}})\) is any function of \(\theta_{\text{new}}\), can be estimated using the Algorithm II.

**V. NUMERICAL STUDY FOR PERFORMANCE EVALUATION**

In this section, simulated profiles are used to evaluate the efficiency of the proposed Gibbs sampling for model estimation, intrinsic Bayes factor for model selection, and parameter inference. In total two models are used in the simulation, with one model for illustration of posterior sampling and both models for model selection and parameter inference.

**A. Simulation Setup**

The models used in the simulation with specified parameters are shown in Table II. For simplicity we assume that \(\theta\) is a scalar parameter and \(h_i(\cdot), i = 1, 2, 3\) are polynomials of degrees \(p - 1, q - 1\) and \(r - 1\) respectively.

<table>
<thead>
<tr>
<th>Table II: MODEL SETTING FOR SIMULATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1</strong></td>
</tr>
<tr>
<td>(p = 2, q = 3, r = 2)</td>
</tr>
<tr>
<td>(h_1(x) = (x, 1)^T)</td>
</tr>
<tr>
<td>(h_2(\theta) = (\theta^2, \theta, 1)^T)</td>
</tr>
<tr>
<td>(h_3(\theta) = (\theta, 1)^T)</td>
</tr>
<tr>
<td>(\beta_1 = (2, 2.2)^T)</td>
</tr>
<tr>
<td>(\beta_2 = (-1, -1, -1)^T)</td>
</tr>
<tr>
<td>(\gamma_1 = (6, -6)^T)</td>
</tr>
<tr>
<td>(\gamma_2 = (4, -4)^T)</td>
</tr>
<tr>
<td>(\sigma^2_1 = 0.01, \sigma^2_3 = 0.1)</td>
</tr>
</tbody>
</table>

Fig. 3. Illustration of the simulated profiles from Model 1 with increasing \(\theta\) from 0.1 to 0.9: (a), (b), ..., (i) corresponds to \(\theta = 0.1, 0.2, \ldots, 0.9\) respectively.

For each model, \(l = 60\) profiles are generated with \(m = 33\) equally spaced design points for \(\theta\) in \([0.1, 0.9]\), i.e., \(\theta = \)
0.1,0.125, ...,0.9, and $n = 11$ equally spaced design points for $x$ in $[2, 3]$, i.e., $x = 2, 2.1, ..., 3$. The first model will be used to show the efficiency of blocked Gibbs sampling and both models will be used to illustrate the IBF model selection and parameter inferences. Fig. 3 shows part of the simulated profiles from Model 1, where we can see obvious increase of between-curve dispersion when increasing $\theta$.

B. Results of Posterior Sampling

In the posterior sampling, we assume that the true model of the simulated profiles is given. Only the model parameters are unknown and need to be estimated. The initial values for all parameters are arbitrarily set to 1. The standard deviation of the proposal distribution is set as $\Delta = 0.1$.

![Sample paths of several representative mean and variance parameters from blocked Gibbs sampling; the horizontal dashed lines denote the true parameters of the model.](image1)

Fig. 4. Sample paths of several representative mean and variance parameters from blocked Gibbs sampling; the horizontal dashed lines denote the true parameters of the model.

![Histograms of the parameter samples; the vertical dashed lines denote the true parameters](image2)

Fig. 5. Histograms of the parameter samples; the vertical dashed lines denote the true parameters.

Fig. 4 shows the sample paths of several representative mean and variance parameters of Model 1. As we can see, all the chains gradually move into the true values of the model parameters after about 20K iterations. We can also observe that the sequences of samples are highly correlated, i.e., requiring many iterations to forget the starting point and reach the equilibrium distribution. The step-width $\Delta$ could be increased or adjusted to reduce the correlation and speed up the convergence. Since it is not the focus, we will not discuss it here. The total computational time of the Gibbs sampling step is about 12 minutes using MATLAB running on an Intel core i5-4590 processor of 3.3GHz. The histograms of the samples in the equilibrium stage are shown in Fig. 5, where the last 10K samples of each chain are selected. As we can see, the centers of the posterior are very close to the true values.

C. Model Selection

Changing the degree of the polynomial in each submodel, or setting certain coefficients to zero with fixed degree at each level, will result in many candidate models, which makes it unrealistic to fit all models and compare them all. In application, the multiple-stage analysis (i.e., fitting the model from the first level to the last one, and using fitted parameters in current level as responses in the next level fitting) can be used to select some most likely models and then use IBF to select the best one among them. Alternatively, the forward selection strategy can be used, where one starts from the simplest model, and each time adds one variable that has the most significant improvement (i.e., increase in marginal density) to the model fitting until there is no significant improvement. For simplicity, we only compare models with different degrees to illustrate the effectiveness of IBF in model selection. In the IBF computation, the profiles with $\theta = 0.1, 0.125, ..., 0.725$ are used as training data and others with $\theta = 0.75, 0.775, ..., 0.9$ are used as testing data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dim. $(p, q, r)$</th>
<th>Log(m) Data1</th>
<th>2 Log(IBF) Data1</th>
<th>Log(m) Data2</th>
<th>2 Log(IBF) Data2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1$</td>
<td>(2,1,1)</td>
<td>17.3</td>
<td>221.2</td>
<td>92.7</td>
<td>111.2</td>
</tr>
<tr>
<td>$M_2$</td>
<td>(2,1,2)</td>
<td>67.6</td>
<td>120.6</td>
<td>108.6</td>
<td>79.4</td>
</tr>
<tr>
<td>$M_3$</td>
<td>(2,1,2)</td>
<td>58.5</td>
<td>138.8</td>
<td>101.3</td>
<td>94</td>
</tr>
<tr>
<td>$M_4$</td>
<td>(2,2,1)</td>
<td>100.8</td>
<td>54.2</td>
<td>141.4</td>
<td>13.8</td>
</tr>
<tr>
<td>$M_5$</td>
<td>(2,2,2)</td>
<td>123.9</td>
<td>8</td>
<td>148.3</td>
<td>4.0</td>
</tr>
<tr>
<td>$M_6$</td>
<td>(2,2,3)</td>
<td>123.1</td>
<td>9.6</td>
<td>126.6</td>
<td>43.4</td>
</tr>
<tr>
<td>$M_7$</td>
<td>(2,3,1)</td>
<td>105.9</td>
<td>44</td>
<td>140.0</td>
<td>16.6</td>
</tr>
<tr>
<td>$M_8$</td>
<td>(2,3,2)</td>
<td>127.9</td>
<td>-</td>
<td>148.2</td>
<td>0.2</td>
</tr>
<tr>
<td>$M_9$</td>
<td>(2,3,3)</td>
<td>127.5</td>
<td>0.8</td>
<td>147.1</td>
<td>2.4</td>
</tr>
<tr>
<td>$M_{10}$</td>
<td>(3,1,1)</td>
<td>46.0</td>
<td>163.8</td>
<td>45.7</td>
<td>205.2</td>
</tr>
<tr>
<td>$M_{11}$</td>
<td>(3,1,2)</td>
<td>38.9</td>
<td>178.0</td>
<td>52.3</td>
<td>192</td>
</tr>
<tr>
<td>$M_{12}$</td>
<td>(3,1,3)</td>
<td>26.5</td>
<td>202.8</td>
<td>41.9</td>
<td>212.8</td>
</tr>
<tr>
<td>$M_{13}$</td>
<td>(3,2,1)</td>
<td>53.3</td>
<td>149.2</td>
<td>65.2</td>
<td>166.2</td>
</tr>
<tr>
<td>$M_{14}$</td>
<td>(3,2,2)</td>
<td>64.5</td>
<td>126.8</td>
<td>64.6</td>
<td>167.4</td>
</tr>
<tr>
<td>$M_{15}$</td>
<td>(3,2,3)</td>
<td>45.4</td>
<td>165</td>
<td>59.9</td>
<td>176.8</td>
</tr>
<tr>
<td>$M_{16}$</td>
<td>(3,3,1)</td>
<td>48.9</td>
<td>158</td>
<td>67.5</td>
<td>161.6</td>
</tr>
<tr>
<td>$M_{17}$</td>
<td>(3,3,2)</td>
<td>72.8</td>
<td>110.2</td>
<td>67.5</td>
<td>161.6</td>
</tr>
<tr>
<td>$M_{18}$</td>
<td>(3,3,3)</td>
<td>53.3</td>
<td>149.2</td>
<td>62.2</td>
<td>172.2</td>
</tr>
</tbody>
</table>

Table III shows the candidate models, estimated marginal densities and the IBF for the two set of profile data (Data1 and Data2) generated from Model 1 and Model 2. As we can see, the true models for both dataset, i.e., $M_8$ for Data1 and $M_5$ for
Data2, have the highest marginal densities than all other candidate models. Almost all the IBFs of the true models to other candidate models are significant according to the recommended BF range and evidence given in Table I. Note that the IBF of $M_8$ to $M_9$ for Data1 and the IBF of $M_8$ to $M_8$ for Data2 are not significant in terms of the IBF value. However, $M_8$ is simpler than $M_9$ and $M_5$ is simpler than $M_8$, indicating that the true models $M_8$ and $M_5$ are preferable to $M_9$ and $M_8$ for Data1 and Data2 respectively. Therefore, the IBF can effectively select the best model among all candidate models.

D. Inference of the Microstructural Parameter $\theta$

$\theta = 0.4, 0.6, 0.8$ are used to generate the new data using Model 1 and Model 2 for parameter inference. 20 profiles are generated for each $\theta$. The prior distribution of $\theta$ is assumed to be uniform in the interval [0,1]. The posterior distribution of $\theta$ is estimated using the importance sampling algorithm shown in Section 4.3.

Fig. 6. Estimated posterior distribution of $\theta$: (a)-(c) for Model 1 and (d)-(f) for Model 2. The vertical dashed lines denote the true $\theta$.

Fig. 6 shows the estimated posterior distributions. We can see that the center of the posterior is very close to the true value of $\theta$, and the variance of the posterior using 20 profiles is also very small. The computational cost of the posterior estimation is about 26 minutes for total 40 samples using MATLAB under the same computer configuration as used in Section B. The computational burden may be an issue in online applications where faster inference is required. To overcome this issue, parallel computing could be applied to speed up the computation.

VI. APPLICATION TO ULTRASONIC ATTENUATION PROFILES IN NANO COMPOSITES MANUFACTURING

In this section the proposed model is applied to the ultrasonic attenuation profiles of the A206-$\text{Al}_2\text{O}_3$ nanocomposites for quality inspection. Due to high experimental cost and difficulty in fabricating nanocomposites of desired microstructural features, it is very challenging to obtain sufficient experimental data for model building. Liu et al [8] recently proposed a microstructural modeling and wave propagation simulation approach to enrich the database of microstructures and the corresponding ultrasonic attenuation profiles. A Voronoi diagram is modified to simulate the microstructures based on the micrographs and morphology modification mechanisms of $\text{Al}_2\text{O}_3$ nanoparticles, and then an elastodynamic finite integration technique VEFIT [36] is used to simulate the wave propagation. In the simulation, the attenuation profiles are measured at a fixed location of repeatedly simulated specimens, which is equivalent to measuring different locations of one specimen. The simulation approach can effectively capture the features of microstructures and generate attenuation profiles comparable to experimental results. In the microstructure generation, two key parameters are used to control the morphology, the number of cells $N$, and the percentage of Voronoi edge length left after dissolving, denoted as $\theta$.

Fig. 7. Attenuation profiles for microstructures with $\theta = (0.1, 0.2, ..., 0.9)$ from (a) to (i).

Fig. 8. Exploratory analysis for the attenuation profiles using multiple-stage analysis. The solid lines denote the simple linear regression lines.
Fig. 7 shows the attenuation profiles (20 profiles each sub-figure) of microstructures with $\theta = (0.1, 0.2, ..., 0.9)$ and the corresponding $N$'s that keeps the total amount of intermetallic phase unchanged. As we can see, the attenuation profiles linearly increase with frequency in the selected frequency range, and the between-profile variation increases with $\theta$. Fig. 8 shows the exploratory analysis of the attenuation profiles using the multiple-stage analysis. We can see that the slope, intercept and their log-variances are quite linear with $\theta$. The model selection is conducted to the data and the optimal model with $p = 2, q = 2$ and $r = 2$ is selected. The estimated parameters using the mean of the posterior samples are $\beta_{11} = 0.231$, $\beta_{12} = 0.253$, $\beta_{21} = -0.489$, $\beta_{22} = -0.396$, $\sigma_\theta^2 = 4.95 \times 10^{-4}$, $\gamma_{11} = 9.92$, $\gamma_{12} = -12.72$, $\gamma_{21} = 11.31$, $\gamma_{22} = -12.68, \sigma_\theta^2 = 0.0077$ and $\sigma_\beta^2 = 0.01$.

The estimated model is used to infer the microstructural parameters of two physically simulated microstructures ($\theta = 0.2$ and $0.8$). Fig. 9 (a) and (b) show the posterior distribution of $\theta$ with uniform prior distribution $U(0, 1)$. We can clearly see that the posteriors are centered on the true values with small variance. The established model is also applied to two fabricated specimens with attenuation profiles shown in Fig. 1. For the two specimens, the posteriors of $\theta$ are shown in Fig. 9 (c) and (d), respectively. The mean estimates of the posteriors are 0.68 and 0.46 respectively. We can see that the $\theta$ of first specimen is higher than the second one, which is consistent with the experimental result that the second specimen has smaller grain size and more homogeneous microstructure. By setting a threshold $\theta_0$ for $\theta$, the posterior can be used to estimate the probability of $\theta < \theta_0$ and use it to for quality control. Therefore the estimated posterior distribution can be used for both quality control and microstructure diagnosis in the ultrasonic attenuation based quality inspection of nanocomposites.

We also fitted a standard Bayesian HLM without considering the variance heterogeneity, i.e., the covariance matrix is constant across all microstructural parameters. For page limits, we do not put the detailed results here. We found that the estimated parameters of $\theta$ and $\sigma_\theta^2$ are very close between these two models. However, the diagonal entries of the estimated covariance matrix $\Sigma$ are higher than those of the proposed model for $\theta < 0.7$, and lower for $\theta > 0.7$. It is what we expected since the estimated variance in the model with constant variance has to compromise between large between-profile variance and small between-profile variance. In the microstructural parameter inference, the posterior modes are very close between these two models. However, the variance of the posterior distribution for the model without considering variance heterogeneity is significantly higher than the proposed model, indicating that the inference using the former is not as informative as using the latter. The addition of the variance heterogeneity model provides more information in the microstructural parameter inference and thus leads to a more accurate inference. This advantage would be more significant when $\theta$ is not sensitive to $\theta$. An extreme case is that $\theta$ is constant across all design points $\theta_1$ while only $\Sigma$ varies with $\theta$. In such case, the model without considering the between-profile variance is not able to infer the microstructural parameter anymore.

VII. CONCLUSION AND DISCUSSION

In this paper, a hierarchical linear model with level-2 variance heterogeneity is proposed to build a relationship between profiles data, the explanatory variables, and the microstructural parameters for quality inspection and control. The integrated Bayesian framework for model estimation and selection is proposed through the blocked Gibbs sampling and intrinsic Bayes factor. The inference of the microstructural parameters based on the estimated model is proposed through importance sampling. The numerical study shows that the proposed approach can effectively identify the true model, estimate the model parameters, and infer the microstructural parameters for new profiles. The proposed model is applied to the ultrasonic attenuation profiles in the manufacturing of metal-matrix nanocomposites. The results show that this approach can be effectively used for quality inference and control. Compared with standard hierarchical linear model, the addition of the variance heterogeneity provides more information in microstructural parameter inference and thus give more accurate inference.

There may still exist other issues that need to be addressed to make this model more robust in many applications. One issue is that if the profile is very complex in shape, e.g. tonnage signals in [37], simple linear model is not adequate to characterize each profile. We may need nonparametric nonlinear models, such as wavelets or B-splines in regression. Another typical issue we may face is the “curse of dimensionality”. If the dimension of the model is very high, it may require a large amount of historical profile data in model building. Besides, to select an appropriate model, it may need to compare a large number of candidate models (e.g., degree of polynomial is very high),
which is often unrealistic. One possible solution to this issue is to incorporate domain knowledge or physical models into statistical models to simplify the selection process.

REFERENCES


Yuhang Liu received the B.S. degree in mathematics in 2012 and the M.S. degree in industrial and system engineering in 2013 from University of Minnesota Twin Cities. He is currently a Ph.D. student in University of Wisconsin Madison. He is a member of the Institute for Operations Research and the Management Science (INFORMS) and SME.

Shiyu Zhou is a Professor in the Department of Industrial and Systems Engineering at the University of Wisconsin-Madison. He received his B.S. and M.S. in Mechanical Engineering from the University of Science and Technology of China in 1993 and 1996, respectively, and his master's in Industrial Engineering and Ph.D. in Mechanical Engineering from the University of Michigan in 2000. His research interests include in-process quality and productivity improvement methodologies by integrating statistics, system and control theory, and engineering knowledge. He is a recipient of a CAREER Award from the National Science Foundation and the Best Application Paper Award from IIE Transactions. He is a member of IIE, INFORMS, ASME, and SME.