This article was downloaded by: *[Texas A&M University]*
On: *28 March 2011*
Access details: *Access Details: [subscription number 933004029]*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

IIE Transactions
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713772245

Bayesian hierarchical model for combining misaligned two-resolution metrology data

Haifeng (Heidi) Xia[a]; Yu Ding[b]; Bani K. Mallick[c]
[a] The Dow Chemical Company, Freeport, TX, USA [b] Department of Industrial & Systems Engineering, Texas A&M University, College Station, TX, USA [c] Department of Statistics, Texas A&M University, College Station, TX, USA

Online publication date: 13 February 2011

To cite this Article Xia (Heidi) , Haifeng , Ding, Yu and Mallick, Bani K.(2011) 'Bayesian hierarchical model for combining misaligned two-resolution metrology data', IIE Transactions, 43: 4, 242 — 258
To link to this Article: DOI: 10.1080/0740817X.2010.521804
URL: http://dx.doi.org/10.1080/0740817X.2010.521804

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

IIE Transactions (2011) **43**, 242–258
Copyright © "IIE"
ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2010.521804

Bayesian hierarchical model for combining misaligned two-resolution metrology data

HAIFENG (HEIDI) XIA[1,*], YU DING[2] and BANI K. MALLICK[3]

[1] *The Dow Chemical Company, 2301 North Brazosport Blvd. B-1603, Freeport, TX 77541, USA*
E-mail: hxia@dow.com
[2] *Department of Industrial & Systems Engineering, and*
[3] *Department of Statistics, Texas A&M University, College Station, TX 77843, USA*

Received May 2009 and accepted June 2010

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

This article presents a Bayesian hierarchical model to combine misaligned two-resolution metrology data for inspecting the geometric quality of manufactured parts. High-resolution data points are scarce and scatter over the surface being measured, while low-resolution data are pervasive but less accurate and less precise. Combining the two datasets should produce better predictions than using a single dataset. One challenge in combining them is the misalignment existing between data from different resolutions. This article attempts to address this issue and make improved predictions. The proposed method improves on the methods of using a single dataset or a combined prediction that does not address the misalignment problem. Improvements of 24% to 74% are demonstrated both for simulated data of circles and datasets obtained for a milled sinewave surface measured by two coordinate measuring machines of different resolutions.

Keywords: Coordinate measuring machine, Gaussian process, misalignment, multi-resolution measurements

1. Introduction

To ensure dimensional quality of manufactured products, a crucial step is to take metrology data (i.e., coordinate measurements) of the geometric features and then check their compliance with tolerance specifications. Traditionally, a Coordinate Measuring Machine (CMM) with a mechanical touch probe is used (Fig. 1(a)), due to its accuracy and versatility in measuring complicated geometries. Recently, a CMM with an optical/laser sensor probe (hereafter referred to as an OCMM, Fig. 1(b)) has been introduced into industry practice as a complement, sometimes a replacement, of the traditional CMM. An OCMM takes measurements by forming an image consisting of the laser light reflected from the part's surface.

The resolution of an OCMM is typically much lower than that of a CMM. By resolution, we refer to the smallest spatial distance that a measuring device can distinguish. A high-resolution device can distinguish two closely positioned points and pick up fine spatial features on a product surface. It therefore attains greater accuracy (i.e., smaller bias) and better precision (i.e., smaller variability) in its measurements than its low-resolution counterpart. According to Shen *et al*. (2000), a CMM can have a resolution as fine as 0.5 μm, whereas an OCMM typically has a resolution on the order of 10 μm. Even though the high-resolution CMM is capable of measuring surfaces at much finer scales, doing so is very time consuming. A common practice is to take measurements of a few locations scattered over the product surface. On the other hand, the OCMM can scan the entire surface of a product much faster than a mechanical CMM, but each measurement is of a lower resolution (i.e., higher inaccuracy and uncertainty). As such, when both metrology devices are used, one would have a large set of low-resolution data and a much smaller set of high-resolution data. The scattered points (dark) and the densely arranged points (gray) in Fig. 1 illustrate the two-resolution metrology data.

The datasets of different resolutions complement each other in terms of the information needed to predict the coordinates of unmeasured locations and to reconstruct the product surface. The low-resolution data points capture the local and global shape of the product surface nicely because of its high measurement density, but the measurements themselves are not accurate, whereas the high-resolution points, due to their scarcity, may not capture the product shape, but each measurement reflects the true yet unknown surface much better. An integrated analysis combining the information from multi-resolution data sources should be able to produce better predictions than merely using a single-resolution dataset.

*Corresponding author

0740-817X © 2011 "IIE"

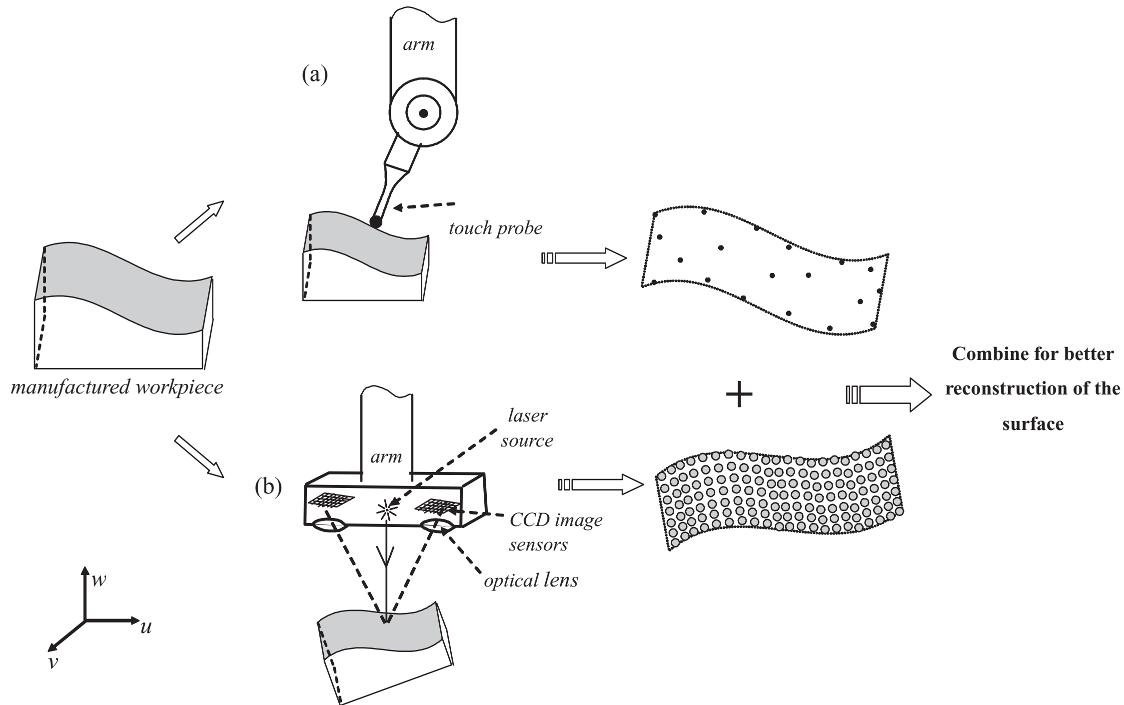

Fig. 1. Two-resolution data from a CMM using a touch probe and laser scanner.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

A unique challenge in combining data for the metrology application is the misalignment existing between the data points of different resolutions. The input, denoted by $\mathbf{x}$ in this article, cannot provide a unique reference in the metrology application to the physical point where the measurement is taken. For example, the same $\mathbf{x} = (1.0, 1.0, 1.0)$ in two respective datasets could correspond to two completely different points on the actual part. On the other hand, the same physical point could have two drastically different input $\mathbf{x}$ values in the two datasets; e.g., in one set, it is $\mathbf{x} = (0.0, 0.0, 2.0)$, while in the other set, it is $\mathbf{x} = (10.0, 0.0, 0.0)$. This misalignment happens when two metrology devices measure the same part but the data are not collected with respect to a common coordinate system. Moreover, the part may undergo a rotation between the two measuring tasks and thus have different orientations while being positioned. Even if the part is measured on a single measuring platform carrying both a touch probe and a laser probe, independent calibration is needed for each probe and the coordinates of measured points are contaminated by measurement noise, still resulting in misalignment between the datasets. This misalignment between the datasets makes it difficult to decide how the data points correspond to each other. Our research attempts to develop a method capable of handling the misaligned metrology datasets and making sound predictions of the product surface.

This article is structured as follows. Section 2 reviews the existing work relevant to combining data from different resolutions. Section 3 presents the important components in the proposed Bayesian hierarchical model, including the low-resolution model, the linkage model that connects the low- and high-resolution data, and the approach to handle the misalignment problem. Section 4 demonstrates the advantages of the proposed method over three alternative methods. Section 5 summarizes the article and discusses future research.

2. Related work

Using multiple-resolution metrology devices is a relatively new idea. Thus, research on combining the multiple datasets of different resolutions is rather limited. There have been reports of CMMs carrying multiple types of sensors (Chen and Lin, 1997; Motavalli *et al.*, 1998; Shen *et al.*, 2000; Carbone *et al.*, 2001), which typically include a mechanical touch probe and a vision system. A vision system has a much lower degree of resolution (between 100 and 200 μm, according to Shen *et al.* (2000)), which is close to, or sometimes larger than, the magnitude of manufacturing errors. On the other hand, a vision system is highly efficient in capturing the global picture of the object under measurement. Thus, the goal of having both the mechanical probe and the vision system is as follows: use the vision system to locate the object and generate a rough product contour and then establish a sampling plan based on the product contour so that the touch probe can measure the product with little human guidance or intervention. The information from different sources is used sequentially. When predicting the product surface for quality assurance, only the high-resolution measurements (from the mechanical probe) are used. The objective of this line of research is to

improve the efficiency of a measurement procedure. Our research focuses on combining information from different sources for better predictions.

Sophisticated methodologies have been developed for the problem of synthesizing spatial data collected at different scales and resolutions (Gotway and Young, 2002; Ferreira *et al.*, 2005; Wikle and Berliner, 2005), of calibrating (deterministic) computer simulation models of different accuracies, and of calibrating computer simulations with physical measurements (Kennedy and O'Hagan, 2000, 2001; Higdon *et al.*, 2004; Reese *et al.*, 2004; Qian *et al.*, 2006; Qian and Wu, 2008). One popular approach is to establish a single-resolution data model (typically for low-resolution or low-accuracy data) and a linkage model linking data from different resolutions. The linkage model assumes that each high-resolution response can be predicted by the *corresponding* low-resolution one with a scale change and a location shift.

The misalignment problem as described in Section 1 is not a major concern in computer experiments. Although computer simulation codes can be run at varying accuracy, the input **x** values are always precisely designed, and they are used as the unique reference to a response, so that the corresponding relationship of responses can be easily identified across different resolutions. The misalignment problem is not discussed in the problems of synthesizing spatial data either, because the inputs for those problems are geographic locations, the measurements of which, although contaminated by noise, are at least one order of magnitude more accurate than that of a spatial response. As a result, it is reasonable to assume that the inputs in spatial problems are also precisely known, the same as in the computer experiments.

In the literature, the majority of the matching and alignment methods were developed for image analysis and registration applications. Brown (1992) and Zitova and Flusser (2003) offered a comprehensive review of the methods in image registration. Matching two objects or images refers to finding correspondences between points or features of one object and those of the other, which is also called *labeling*. After establishing the correspondences between two datasets, one can mark the matching relationship between them as labels. The data then become *labeled* points. Given labeled points, the subsequent step is to estimate the transformation model to bring the two objects or images together. Statistical shape analysis (for example, Dryden and Mardia (1998)) provides sophisticated methodologies that address the problem of estimating the transformation model and assessing the shape differences between different images, based on labeled points.

The problem of matching the metrology data points is to match *unlabeled* points. To find the matching correspondence of two sets of unlabeled data points, there are two major approaches in the image registration literature. One relies on the feature properties associated with a point or a graphic pattern formed by a set of points. For example, Ton and Jain (1989) and Zitova and Flusser (2003) used curvatures calculated to match two sets of points. Cheeseman *et al.* (1996) have a different research objective but several pieces of their method do share commonality with ours, especially the registration part and the initial composite using a compositing kernel (their Section 4.3). However, their method uses images that are all taken from roughly the same direction and, as a result, the alignment portion is not particularly challenging; thus, their procedure focuses on the estimation of the transformation parameters between the images.

In our metrology application, the high-resolution data points are scarce due to time and cost constraints. Therefore, it may not be possible to obtain a sufficient amount of feature properties for matching. In any case, graphic patterns obtained by using the scarce high-resolution data are different from those using the dense low-resolution data. It again makes the subsequent matching difficult. The high-resolution points function more like a set of anchor points, and their contribution in the prediction is to help calibrate the shape-revealing low-resolution points with the underlying true surface.

The other approach of point matching is based on the invariance property of inter-point distances. The underlying principle is that using either dataset, the relative distance between the same two points should be approximately the same (Ranade and Rosenfeld, 1980; Ton and Jain, 1989). The distance-based approach is applicable to the metrology application since the distances measured at different resolutions differ from one another only by a small amount caused by the measurement errors. Thus, we will later utilize the distance-based approach to assist our matching effort.

3. Bayesian hierarchical model

We devise a hierarchical predictive model as such: at the lower level, a Gaussian Process (GP) model developed by Xia *et al.* (2008) is used for representing the low-resolution data; at the upper level, a *neighborhood linkage model* is established to link each high-resolution output as a kernel regression of all the low-resolution information in its neighborhood. Different from the linkage models used in the computer experiments, our model extends their *one-to-one* linkage to a *one-to-many* linkage. The reason we make this extension is that despite the alignment efforts, there is no guarantee of finding the perfect match when distortion is present (Ton and Jain, 1989; Mitra and Murthy, 1991).

Our linkage model is conditioned on a matching scheme found between the two-resolution datasets. In order to find the matching schemes, we devise a heuristic matching algorithm based on the invariance property of inter-point distance, as mentioned in Section 2. This algorithm can produce multiple probable match candidates. Conditioned on each of them, the hierarchical predictive model can make

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

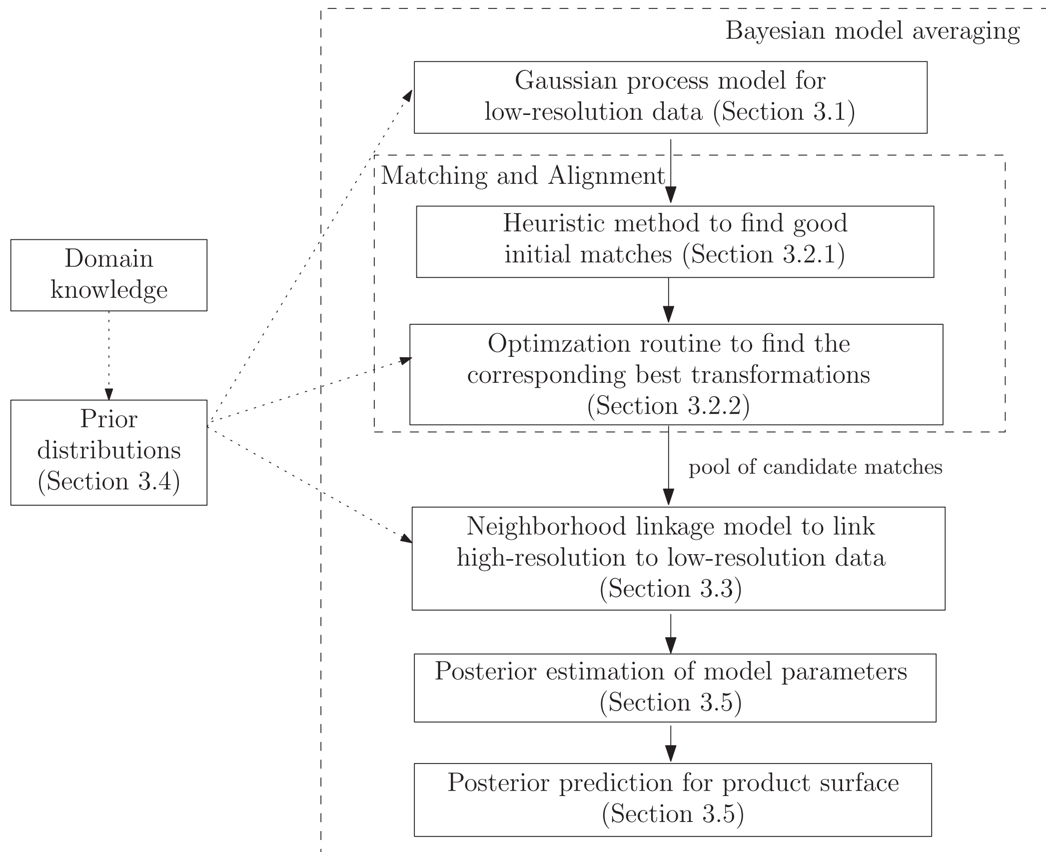

Fig. 2. Overall framework of the proposed Bayesian hierarchical model.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

a prediction of the part's surface. The final prediction is a weighted average of the individual predictions over the pool of all the match candidates. The weights are assigned according to how well the observed data support each match. Figure 2 summarizes the overall framework of the proposed method.

In the following sections, we will first discuss individual model components in Sections 3.1 to 3.4, i.e., the low-resolution model, the alignment procedure, the linkage model, and the choices of the prior, respectively. Then, we will explain in Section 3.5 how these individual components are incorporated into a Bayesian framework to make the final prediction.

3.1. *Low-resolution data model*

Xia *et al.* (2008) presented a GP model for single-resolution data of a manufactured part. This GP model is adopted for the low-resolution data in this article.

To explain how the coordinate measurements of a manufactured product are taken, consider the two-dimensional circular-shaped part in Fig. 3. The spatial coordinate used by a metrology device is denoted by (u, v, w). We avoid using (x, y, z) as the coordinate variables because those symbols will be used in the model for different meanings. A metrology device is usually controlled by a computer. The device takes an input location through the computer interface from an operator, denoted by $\mathbf{x}_i = (u_i, v_i, w_i)^{\mathrm{T}}$; then it directs its touch probe or laser beam to travel in a certain direction, denoted by $\mathbf{p}_i$, to the object under measurement; finally, it retrieves the coordinate information, denoted by $\mathbf{a}_i$. The directional vector $\mathbf{p}_i$, associated with each measurement taken, is either specified by the operator or calculated automatically by the computer.

Additional complexities are involved due to manufacturing errors and soft-fixturing process. Figure 3 shows three contours of a round part, of which the solid line represents the actual part surface, and the other two contours are explained in the following paragraphs. Before a metrology device performs the actual measuring task, it will first undergo a soft-fixturing process (Hulting, 1995), which is that the machine takes a few measurements from the actual surface and uses them to estimate where the part is located. Since the actual dimension and shape of the part are unknown initially, the measurement machine must decide the location of the part assuming that the part has the perfect design form and design dimensions; e.g., a perfectly round circle for the example in Fig. 3 with the design value r_0 as its radius. Then, the machine translates its origin $(0, 0)$ to the center of the part. As such, the machine establishes a known geometry, e.g., the circle centered at $(0, 0)$ with radius r_0 in Fig. 3, to serve as the reference for the subsequent

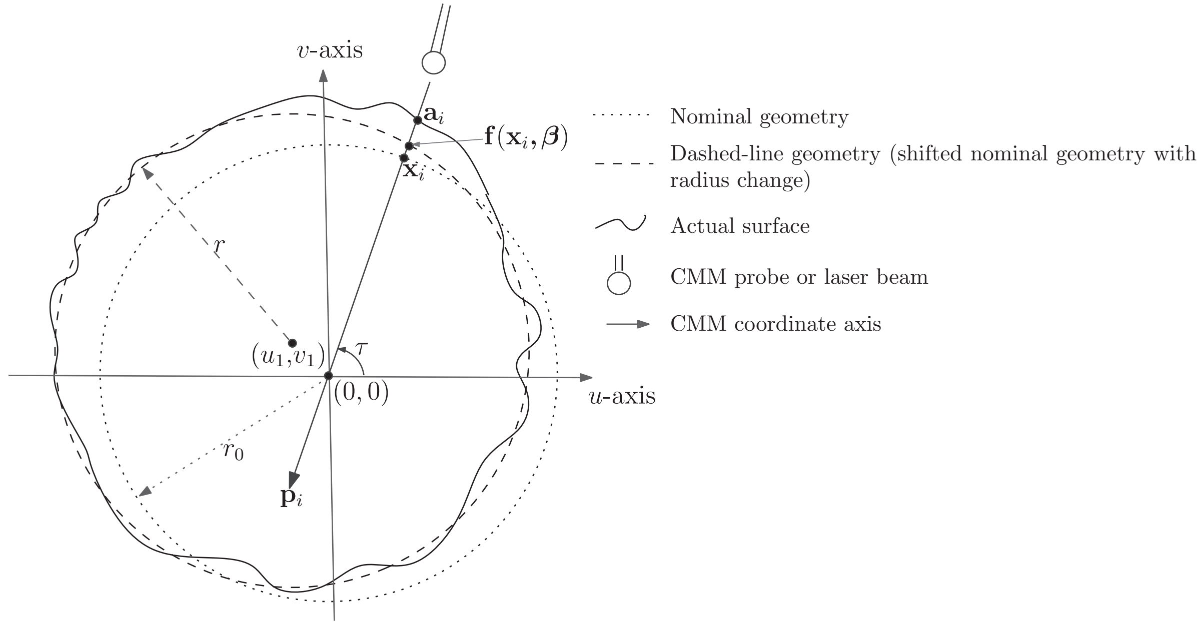

Fig. 3. Illustration of the coordinate measuring process.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

measuring process. It is denoted by the dotted line and is often called the *nominal geometry*. When the metrology device takes an input position $\mathbf{x}_i$, it considers this point to be on the nominal geometry. After the measuring action takes place, the response that a machine returns, $\mathbf{a}_i$, is on the actual surface, by which a laser beam is reflected or a probe is stopped.

The response $\mathbf{a}_i$ is typically a vector, for example, for a three-dimensional feature, $\mathbf{a}_i = (a_{u_i}, a_{v_i}, a_{w_i})^{\mathrm{T}}$. Using it as the response leads to a multivariate GP model. To simplify the response, researchers (Hulting, 1997; Xia *et al.*, 2008) chose to project the value of $\mathbf{a}_i$ (after the alignment in Section 3.2) onto the measuring direction $\mathbf{p}_i$. The resulting response $y(\mathbf{x}_i) \equiv \mathbf{a}_i^{\mathrm{T}}\mathbf{p}_i$ is a scalar function of $\mathbf{x}_i$.

Xia *et al.* (2008) used a GP model for the single-resolution data with the form of

$$y_l(\mathbf{x}_i) = \eta_l(\mathbf{x}_i) + \epsilon_l, \qquad i = 1, \ldots, m_l. \tag{1}$$

Here a subscript l is associated with each term because the model is used for the low-resolution data, although the same model can be used for the high-resolution data as well. In Equation (1), $y_l(\mathbf{x}_i)$ is the ith low-resolution observation, m_l is the number of low-resolution observations, ϵ_l is the random error, dominated by measurement noises and modeled as being independent and identically distributed (i.i.d.) $N(0, \sigma_l^2)$, and $\eta_l(\cdot)$ is the low-resolution version of the actual surface. Please note that $\eta_l(\cdot)$ is not exactly the true surface because being in a low resolution, the measurements provide a "blurred" view, rather than a sharp reflection, of the actual surface.

The low-resolution surface $\eta_l(\cdot)$ is modeled as a Gaussian process with the ith mean component being $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}_l)^{\mathrm{T}}\mathbf{p}_i$ and the covariance function $\mathrm{cov}(\eta_l(\mathbf{x}_i), \eta_l(\mathbf{x}_j)) = \kappa_l^2 R(\mathbf{x}_i, \mathbf{x}_j)$, where κ_l^2 and $R(\mathbf{x}_i, \mathbf{x}_j)$ are the variance and the correlation function, respectively. Here $\mathbf{f}(\cdot, \boldsymbol{\beta}_l)$ is another imaginary geometry, illustrated by the dashed line in Fig. 3. For convenience, we will refer to it as the *dashed-line geometry* later. This geometry is different from the nominal geometry (the dotted line) by a possible dimension difference and a location difference. The dashed-line geometry incorporates the *dimension* errors of the manufactured part. For the round part, r is the actual radius of the part and can be different from the design radius r_0. The differences between the solid-line surface and the dashed-line geometry are the *form or geometric* errors of manufacturing. Figure 4 summarizes the three types of errors.

Figure 3 exaggerates the differences for the purpose of illustration. Typically, the actual solid-line surface closely follows the dashed-line geometry and the differences are no more than a few hundreds of micrometers. The closeness between the dashed-line geometry and the actual surface motivates the selection of $\mathbf{f}(\cdot, \boldsymbol{\beta}_l)$ as the mean component for the GP model. The function format of $\mathbf{f}(\cdot, \boldsymbol{\beta}_l)$ is known from the design process of the part. Thus, the dashed-line geometry is simply parameterized by $\boldsymbol{\beta}_l$, including the dimension and location parameters. For the round part example in Fig. 3, $\boldsymbol{\beta}_l = (r, u_1, v_1)$, where (u_1, v_1) are the coordinates of the part's center. Naturally, $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}_l)$ represents the point on the dashed-line geometry corresponding to $\mathbf{x}_i$ on the nominal geometry. A three-step procedure to compute the $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}_l)$ for a general geometry can be found in Xia *et al.* (2008).

The correlation function $R(\mathbf{x}_i, \mathbf{x}_j)$ is modeled as a product of Gaussian correlation functions (Santner *et al.*, 2003)

$$R(\mathbf{x}_i, \mathbf{x}_j) = \prod_{k=1}^{d} \exp\{-\nu_k (x_{ki} - x_{kj})^2\}, \tag{2}$$

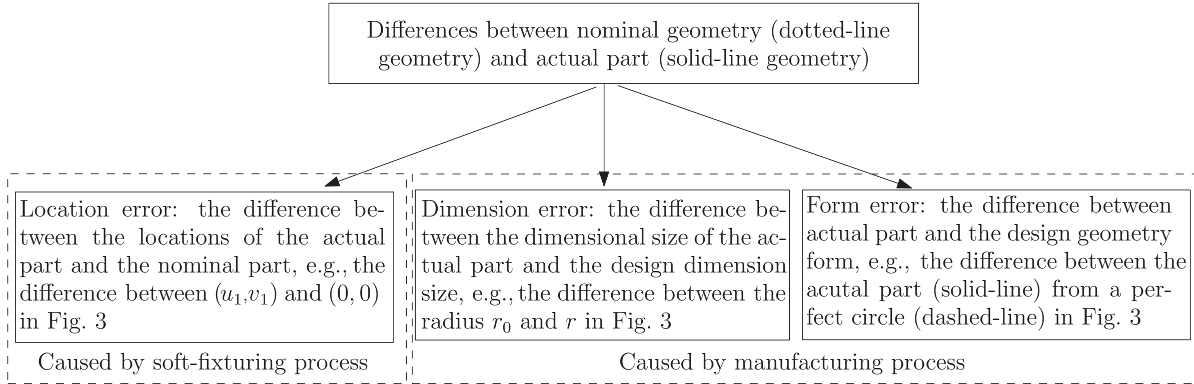

Fig. 4. Three different types of errors.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

where d is the dimension of the input variables, i.e., $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{di})$, and $\boldsymbol{\nu}_l = (\nu_1, \nu_2, \ldots, \nu_d)$ are the scale parameters controlling how fast the correlation decays as the between-input distance increases in each dimension. For the metrology applications, $d = 1, 2$, or 3. For this GP model, we summarize the model parameters in $\boldsymbol{\theta}_l = (\boldsymbol{\beta}_l, \kappa_l^2, \sigma_l^2, \boldsymbol{\nu}_l)$.

Denote by $\mathbf{X}_l = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m_l})^{\mathrm{T}}$ the location matrix where low-resolution observations are made, by $\boldsymbol{\eta}_l = (\eta_l(\mathbf{x}_1), \eta_l(\mathbf{x}_2), \ldots, \eta_l(\mathbf{x}_{m_l}))^{\mathrm{T}}$ the low-resolution surface and by $\mathbf{y}_l = (y_l(\mathbf{x}_1), y_l(\mathbf{x}_2), \ldots, y_l(\mathbf{x}_{m_l}))^{\mathrm{T}}$ the low-resolution data. Conditioned on $\boldsymbol{\theta}_l$, the joint distribution of $\boldsymbol{\eta}_l$ and $\mathbf{y}_l$ is as follows:

$$p(\boldsymbol{\eta}_l, \mathbf{y}_l \mid \boldsymbol{\theta}_l) = N\left(\begin{bmatrix}\mathbf{g}(\mathbf{X}_l, \boldsymbol{\beta}_l)\\ \mathbf{g}(\mathbf{X}_l, \boldsymbol{\beta}_l)\end{bmatrix}, \begin{bmatrix}\kappa_l^2\mathbf{R}_l & \kappa_l^2\mathbf{R}_l\\ \kappa_l^2\mathbf{R}_l^T & \boldsymbol{\Sigma}_l\end{bmatrix}\right), \quad (3)$$

where $\mathbf{g}(\mathbf{X}_l, \boldsymbol{\beta}_l)$ is an $m_l \times 1$ column vector, whose ith element is $g(\mathbf{x}_i, \boldsymbol{\beta}_l) = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}_l)^T\mathbf{p}_i$; $\mathbf{R}_l$ is an $m_l \times m_l$ correlation matrix whose (i, j)th element is defined according to Equation (2); and $\boldsymbol{\Sigma}_l = \kappa_l^2\mathbf{R}_l + \sigma_l^2\mathbf{I}$.

Using the above joint distribution, the distribution of $\boldsymbol{\eta}_l$ conditioned on the observations of $\mathbf{y}_l$ can be obtained as

$$p(\boldsymbol{\eta}_l|\mathbf{y}_l, \boldsymbol{\theta}_l) = N\big(\mathbf{g}(\mathbf{X}_l, \boldsymbol{\beta}_l) - \kappa_l^2\mathbf{R}_l\boldsymbol{\Sigma}_l^{-1}(\mathbf{y}_l - \mathbf{g}(\mathbf{X}_l, \boldsymbol{\beta}_l)), \kappa_l^2\mathbf{R}_l - \kappa_l^4\mathbf{R}_l\boldsymbol{\Sigma}_l^{-1}\mathbf{R}_l^T\big). \quad (4)$$

This equation implies that $\boldsymbol{\eta}_l$ works also as a filtered version of the low-resolution data, which are free of the random measurement errors. Later, the filtered low-resolution data $\eta_l(\mathbf{x}_i)$, instead of the unfiltered $y_l(\mathbf{x}_i)$, for $i = 1, \ldots, m_l$, will be used when establishing a link to the high-resolution data.

3.2. *Align the two-resolution metrology data*

Suppose that, before projecting onto $\mathbf{p}$, we have a set of high-resolution data $D_h = \{\mathbf{a}_i^h : i = 1, \ldots, m_h\}$ and a set of low-resolution data $D_l = \{\mathbf{a}_j^l : \ j = 1, \ldots, m_l\}$, where $m_h \ll m_l$. Both sets of data points are arbitrarily labeled for identification. The goal is to match a subset of the low-resolution data points to the entire set of high-resolution data points and then find the corresponding transformation.

Matching two datasets refers to establishing the correspondence between the two datasets. For example, a found match could be $\mathbf{a}_1^h \rightarrow \mathbf{a}_5^l, \mathbf{a}_2^h \rightarrow \mathbf{a}_{11}^l, \mathbf{a}_3^h \rightarrow \mathbf{a}_{64}^l, \ldots, \mathbf{a}_{20}^h \rightarrow \mathbf{a}_{496}^l$, as in Fig. 5(a). Transformation refers to the rigid body transformation between the low-resolution dataset and the high-resolution dataset, as illustrated in Fig. 5(b). The parameters of the transformation (i.e., three for translation and three for rotation) can be computed using the established correspondence. In other words, the alignment procedure normally goes through two steps: first, we find the correspondence between the two datasets, i.e., a procedure called matching; and second, we compute the transformation parameters based on the found match.

3.2.1. *Heuristic matching*

We develop a heuristic matching algorithm utilizing the invariance property of inter-point distances explained in Section 2, which was initially introduced for image registration (Ranade and Rosenfeld, 1980). Mathematically, it means the following: if $(\mathbf{a}_i^h, \mathbf{a}_j^l)$ is a pair of matching points, then for every other point $\mathbf{a}_k^h$ in D_h, there should be a corresponding point $\mathbf{a}_s^l$ such that the distance $||\mathbf{a}_i^h - \mathbf{a}_k^h||$ "almost" equals to the distance $||\mathbf{a}_j^l - \mathbf{a}_s^l||$. Here we use "almost" because both datasets are noisy, so an exact equality is difficult to attain. During implementation, this "almost equality" is implemented by setting a threshold for the allowed difference. Before performing the heuristic matching, we first estimate the nominal surfaces from both datasets and align the nominal surfaces. In this way, the two datasets are made close to each other so that the heuristic algorithm can work effectively. The heuristic algorithm proceeds as follows.

Repeat the following two steps for $j = 1, 2, \ldots, m_l$,

Step 1. Let the first data point $\mathbf{a}_1^h$ in D_h be paired with the jth data point $\mathbf{a}_j^l$ in D_l.

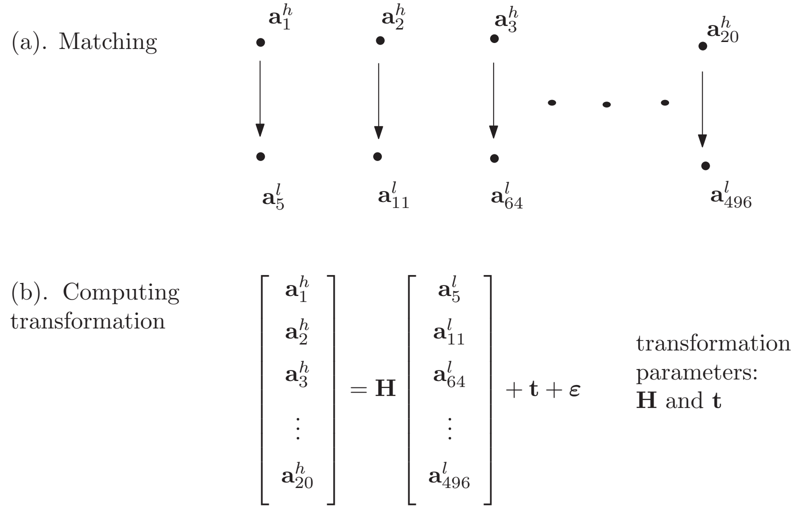

Fig. 5. Two steps to align two-resolution data.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

Step 2. For $k = 2, 3, \ldots, m_h$, calculate the inter-point distance between $\mathbf{a}_k^h$ and $\mathbf{a}_1^h$, and then find a point in D_l, e.g., the sth low-resolution point, so that the difference between $\|\mathbf{a}_s^l - \mathbf{a}_j^l\|^2$ and $\|\mathbf{a}_k^h - \mathbf{a}_1^h\|^2$ is less than some threshold ϖ. If such a low-resolution point is found, $\mathbf{a}_k^h$ and $\mathbf{a}_s^l$ will be considered to be a pair of matching points. When there is more than one low-resolution point satisfying the above requirement, select the one giving the smallest difference between $\|\mathbf{a}_s^l - \mathbf{a}_j^l\|^2$ and $\|\mathbf{a}_k^h - \mathbf{a}_1^h\|^2$. After going through all the m_h high-resolution points, if all of them find their matching points in the low-resolution data, we consider a match identified and call the set of the matching pairs a *consistent match*, a term first coined in the image registration literature (Mitra and Murthy, 1991). If not all the m_h high-resolution points find their matching point in the low-resolution data, we say no consistent match exists for $\mathbf{a}_j^l$, go back to Step 1, and let $\mathbf{a}_1^h$ be paired with the next low-resolution point.

We record each consistent match found between the two datasets using a matching matrix $\mathbf{M} = (M_{ij})$, where

$$M_{ij} = \begin{cases} 1 & \text{if } \mathbf{a}_j^l \text{ is matched to } \mathbf{a}_i^h, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Note that $\sum_{i,j} M_{ij} = m_h$. Also, only one low-resolution point will match each high-resolution data point; i.e., $\sum_j M_{ij} \leq 1$.

The heuristic matching procedure eventually produces K consistent matches $\{\mathbf{M}_k, k = 1, \ldots, K\}$, where K is typically much smaller than m_l. We will then find their corresponding transformation matrices or vectors and feed them to the Bayesian hierarchical model that combines the two-resolution data for prediction (in Section 3.5).

Given the two-resolution data, the number of consistent matches K found by the heuristic procedure depends on the distance threshold ϖ. A small ϖ will lead to a fewer number of consistent matches. As a result, the Bayesian hierarchical model will run over a fewer number of matches, and the subsequent inferences will incur less computation. For the examples in Section 4, we set ϖ to be the average inter-point distance in the low-resolution dataset. With this ϖ, for the two examples in Section 4, the heuristic matching produces $K = 15$ or 30 consistent matches for the circular surfaces and $K = 2$ or 1 consistent matches for the sinewave surface, respectively. In the case where unevenly spaced measurement data are used, one needs to relax the distance threshold ϖ to be anisotropic, meaning that different ϖ values are used for each direction of interest, and each ϖ can be set as the average inter-point distance along a specific direction.

3.2.2. *Optimization to calculate the transformation parameters*

After matching the two datasets, we can calculate the transformation parameters (i.e., three rotation parameters in $\mathbf{H}$ and three translation parameters in $\mathbf{t}$) by solving the following optimization, which minimizes the sum of the squared distances between the corresponding points in the two datasets over the possible rigid body transformations:

$$\min_{\substack{t_u, t_v, t_w \\ \rho_u, \rho_v, \rho_w}} \sum_{\substack{i=1:m_h \\ j:M_{ij}=1}} \| \mathbf{a}_i^h - \left(\mathbf{H}(\rho_u, \rho_v, \rho_w)\mathbf{a}_j^l + (t_u, t_v, t_w)^{\mathrm{T}}\right) \|^2 \tag{6}$$

subject to

$$\mathbf{H}(\rho_u, \rho_v, \rho_w) = \begin{bmatrix} \cos(\rho_u)\cos(\rho_v) & \cos(\rho_u)\sin(\rho_v)\sin(\rho_w) - \sin(\rho_u)\cos(\rho_w) & \cos(\rho_u)\sin(\rho_v)\cos(\rho_w) + \sin(\rho_u)\sin(\rho_w) \\ \sin(\rho_u)\cos(\rho_v) & \sin(\rho_u)\sin(\rho_v)\sin(\rho_w) + \cos(\rho_u)\cos(\rho_w) & \sin(\rho_u)\sin(\rho_v)\cos(\rho_w) - \cos(\rho_u)\sin(\rho_w) \\ -\sin(\rho_v) & \cos(\rho_v)\sin(\rho_w) & \cos(\rho_v)\cos(\rho_w) \end{bmatrix},$$

$$t_u, t_v, t_w \in \mathbb{R};$$

$$\rho_u, \rho_v, \rho_w \in [0, 2\pi);$$

where ρ_u, ρ_v, ρ_w are the rotations around, and t_u, t_v, t_w are the translations along, the u-, v-, and w-axes, respectively.

3.3. *Neighborhood linkage model*

Combining the metrology data of different resolutions for prediction is fulfilled by using a neighborhood linkage model. The linkage model connects each high-resolution data point to all the low-resolution response $\eta_l(\cdot)$ in its neighborhood as follows:

$$y_h(\mathbf{x}_i) = \alpha_1 \sum_{j=1,\ldots,m_l} K(\mathbf{x}_i, \mathbf{x}_j)\, \eta_l(\mathbf{x}_j) + \alpha_0 + e, \tag{7}$$

where $y_h(\mathbf{x}_i)$ is the univariate high-resolution response defined in the same way as the low-resolution one, i.e., $y_h(\mathbf{x}_i) = (\mathbf{a}_i^h)^{\mathrm{T}}\mathbf{p}_j$; $\boldsymbol{\alpha} = (\alpha_1, \alpha_0)$ are the scale and location coefficients, respectively; $K(\cdot, \cdot)$ is a kernel function; e is the residual, assumed to be i.i.d. $N(0, \sigma_e^2)$.

We use a tri-cube kernel function but generalize it by having different kernel widths $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ associated with different axes. The resulting kernel function is as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Psi\left(\sum_{k=1}^{d}\left(\frac{x_{ki} - x_{kj}}{\lambda_k}\right)^2\right), \tag{8}$$

with

$$\Psi(t) = \begin{cases} (1 - t^{3/2})^3 & \text{if } t \le 1; \\ 0 & \text{if } t > 1. \end{cases} \tag{9}$$

This kernel function defines the neighborhood of a high-resolution data point. Only the low-resolution data points within the neighborhood are linked with the high-resolution data point. The size of the neighborhood is controlled by $\boldsymbol{\lambda}$ and will be estimated using the data.

Utilizing the linkage model, we can predict the high-resolution response $y_h(\mathbf{x}_0)$ at any given location $\mathbf{x}_0$. Given the filtered low-resolution responses $\boldsymbol{\eta}_l$ at m_l locations, and conditioned on the model parameters $\boldsymbol{\theta}_h = (\boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda})$ and a match $\mathbf{M}_k$, $y_h(\mathbf{x}_0)$ has the following distribution:

$$(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) \sim N\big(\mathbf{F}_{\lambda}(\mathbf{x}_0)\boldsymbol{\alpha}, \sigma_e^2\big), \tag{10}$$

where $\mathbf{F}_{\lambda}(\mathbf{x}_0)$ is a row vector, defined as $(\sum_{i=1,\ldots,m_l} K(\mathbf{x}_0, \mathbf{x}_i)\ \eta_l(\mathbf{x}_i), 1)$. Denote by $\mathbf{X}_h = (\mathbf{x}_1, \ldots, \mathbf{x}_{m_h})^{\mathrm{T}}$ the locations of high-resolution responses. The distribution in Equation (10) can be extended to the multivariate case, where $\mathbf{y}_h = (y_h(\mathbf{x}_l), \ldots, y_h(\mathbf{x}_{m_h}))^{\mathrm{T}}$ such that

$$\big(\mathbf{y}_h|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}\big) \sim N\big(\mathbf{F}_{\boldsymbol{\lambda}}\boldsymbol{\alpha}, \sigma_e^2\mathbf{I}\big), \tag{11}$$

where $\mathbf{F}_{\boldsymbol{\lambda}}$ abbreviates $\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{X}_h)$, an $m_h \times 2$ matrix whose jth row is defined as $(\sum_{i=1,\ldots,m_l} K(\mathbf{x}_j, \mathbf{x}_i)\ \eta_l(\mathbf{x}_i), 1)$, and $\mathbf{x}_i$ is the ith row of $\mathbf{X}_l$ and $\mathbf{x}_j$ is the jth row of $\mathbf{X}_h$.

3.4. *Bayesian priors for the predictive model*

This subsection discusses the choices of prior for the parameters in the hierarchical predictive model, and the next section will derive the Bayesian inference and prediction. We would like to note that both sections and the appendices associated with the sections are developed in a manner similar to Qian and Wu (2008).

The parameters consist of two parts: $(\boldsymbol{\theta}_l, \boldsymbol{\theta}_h)$, where $\boldsymbol{\theta}_l = (\boldsymbol{\beta}_l, \sigma_l^2, \kappa_l^2, \boldsymbol{\nu}_l)$ are the parameters involved in the low-resolution data model, and $\boldsymbol{\theta}_h = (\boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda})$ are the parameters involved in the neighborhood linkage model. They can also be grouped into three categories: the mean-component parameters $(\boldsymbol{\beta}_l, \boldsymbol{\alpha})$, the variance parameters $(\sigma_l^2, \kappa_l^2, \sigma_e^2)$, and the parameters $(\boldsymbol{\nu}_l, \boldsymbol{\lambda})$ in the correlation function and the kernel function, respectively.

Regarding the GP model for the low-resolution data, we follow the common practice (e.g., Qian and Wu (2008)) and use a normal distribution for the priors of the mean parameter $\boldsymbol{\beta}_l$, inverse-gamma distributions for the priors on the variance parameters (σ_l^2, κ_l^2), and gamma distributions for the priors on the correlation parameter $\boldsymbol{\nu}_l$.

The linkage model is essentially a linear regression model once the kernel width $\boldsymbol{\lambda}$ is given. We choose the non-informative priors for $\boldsymbol{\alpha}$ and σ_e^2 to reflect our limited knowledge regarding how the two-resolution metrology data are related. An accompanying benefit of using the non-informative priors is that the resulting posterior distributions of $\boldsymbol{\alpha}$ and σ_e^2 are in closed forms, which helps speed up the computation in the subsequent Bayesian inference. The prior distribution for the kernel width $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ is chosen to be uniform over a range. Let λ_i follow a discrete distribution taking values $j \times c$ for $j = 1, \ldots, \lambda_0$, where c is the size of the increment, and λ_0 is a positive, typically large, integer. Assuming $\boldsymbol{\lambda}$ a discrete distribution helps simplify the computation of the subsequent Bayesian inference as well.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

Furthermore, the joint prior distribution is assumed to be the product of the prior distributions of individual parameters as follows:

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\beta}_l)p(\sigma_l)p(\kappa_l)p(\boldsymbol{\alpha})p(\sigma_e)p(\boldsymbol{\nu}_l)p(\boldsymbol{\lambda}), \quad (12)$$

where

$$\begin{aligned}
&\boldsymbol{\beta}_l \sim N(\boldsymbol{\mu}_l, \mathbf{Q}_l)\\
&\sigma_l^2 \sim \text{Inv-Gamma}(a_1, b_1),\\
&\kappa_l^2 \sim \text{Inv-Gamma}(a_2, b_2),\\
&\nu_i \sim \text{Gamma}(a_3, b_3) \qquad \text{for } i = 1, \ldots, d,\\
&p(\boldsymbol{\alpha}) \propto 1,\\
&p(\sigma_e^2) \propto \sigma_e^{-2},\\
&p(\lambda_i = j \times c) = 1/\lambda_0 \text{ for } j = 1, 2, \ldots, \lambda_0; \;\; i = 1, \ldots, d.
\end{aligned}$$

The covariance matrix $\mathbf{Q}_l$ is a diagonal matrix, whose ith diagonal elements q_i^2 denote the variance for the ith element of $\boldsymbol{\beta}_l$. Appendix A includes more discussions on how to choose the parameters in the aforementioned prior distributions.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

3.5. *Bayesian inference and prediction*

The ultimate goal of combining the two-resolution data is to predict the response $y_h(\mathbf{x}_0)$ at any input location $\mathbf{x}_0$, given the observed data $\mathbf{y}_h$ and $\mathbf{y}_l$, i.e., $p(y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l)$.

After presenting in the previous sections the low-resolution model, the linkage model, and the alignment procedure, we can in principle make predictions conditioned on a given matching scheme $\mathbf{M}$ of the two datasets. In other words, we can obtain $p(y_h(\mathbf{x}_0)|\mathbf{M}, \mathbf{y}_h, \mathbf{y}_l)$, and the details are shown in Equation (14). As such, $p(y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l)$ can then be obtained by integrating over the distribution of $\mathbf{M}$:

$$p(y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l) = \int_{\mathbf{M}} p(y_h(\mathbf{x}_0)|\mathbf{M}, \mathbf{y}_h, \mathbf{y}_l) \times p(\mathbf{M}|\mathbf{y}_h, \mathbf{y}_l)d\mathbf{M}.$$

However, it is practically difficult to solve the above integration. Thus, we recommend using a summation over a discrete number of candidate matches to approximate the integration. Suppose the alignment procedure in Section 3.2 produces K matches, together with their corresponding transformation parameters. Here we denote both a match and its corresponding transformation by $\{\mathbf{M}_k, k = 1, 2, \ldots, K\}$. Using the K distinct matches, the above integration is approximated by

$$p(y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h) = \sum_{k=1}^{K} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_l, \mathbf{y}_h) \times p(\mathbf{M}_k|\mathbf{y}_l, \mathbf{y}_h). \quad (13)$$

where the first term $p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_l, \mathbf{y}_h)$ is the individual prediction conditioned on a given match, and the second term $p(\mathbf{M}_k|\mathbf{y}_h, \mathbf{y}_l)$ is the posterior probability of a matching scheme, representing how much a match is consistent with the observed data. It puts different weights on the individual predictions conditioned on the corresponding match and thus accounts for the uncertainty in the matches between the two-resolution datasets.

Given the prior distribution in Equation (12), it can be shown (please see Appendix B for details) that the first term in Equation (13), $p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_l, \mathbf{y}_h)$, can be solved through the following steps:

$$p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_l, \mathbf{y}_h) = \int_{\boldsymbol{\theta}_l} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \mathbf{y}_h)p(\boldsymbol{\eta}_l|\boldsymbol{\theta}_l, \mathbf{y}_l) \times p(\boldsymbol{\theta}_l|\mathbf{y}_l)d\boldsymbol{\theta}_l, \quad (14)$$

where $p(\boldsymbol{\eta}_l|\boldsymbol{\theta}_l, \mathbf{y}_l)$ is given in Equation (4), $p(\boldsymbol{\theta}_l|\mathbf{y}_l) \propto p(\boldsymbol{\theta}_l) \cdot p(\mathbf{y}_l|\boldsymbol{\theta}_l)$, and

$$p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \mathbf{y}_h) = \sum_{\substack{\lambda_1 = c, 2c, \ldots, \lambda_0 c\\ \vdots \\ \lambda_d = c, 2c, \ldots, \lambda_0 c}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l, \boldsymbol{\lambda}) \times p(\boldsymbol{\lambda}|\mathbf{y}_h, \boldsymbol{\eta}_l), \quad (15)$$

in which

$$(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l, \boldsymbol{\lambda}) \sim t_{m_h-2}(\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x}_0)\hat{\boldsymbol{\alpha}}, s^2(1 + \mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x_0}) \times (\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\lambda}})^{-1}\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x_0})^{\mathrm{T}})), \quad (16)$$

$$p(\boldsymbol{\lambda}|\mathbf{y}_h, \boldsymbol{\eta}_l) \propto p(\boldsymbol{\lambda})|\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}} \times \left[\frac{(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^{\mathrm{T}}(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})}{2}\right]^{-\frac{m_h}{2}+1}, \quad (17)$$

$\mathbf{F}_{\boldsymbol{\lambda}}$ and $\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x}_0)$ are defined in Equation (10) and Equation (11), respectively,

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\lambda}}\right)^{-1}\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{y}_h,$$

$$s^2 = \frac{1}{m_h - 2}(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^{\mathrm{T}}(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}}).$$

The integration of $\boldsymbol{\theta}_l$ in Equation (14) is solved numerically using a Markov Chain Monte Carlo (MCMC) algorithm.

The second term in Equation (13), the posterior probability $p(\mathbf{M}_k|\mathbf{y}_h, \mathbf{y}_l)$, can be computed by

$$p(\mathbf{M}_k|\mathbf{y}_h, \mathbf{y}_l) = \frac{p(\mathbf{y}_h|\mathbf{M}_{\mathrm{k}}, \mathbf{y}_l)p(\mathbf{M}_k)}{\sum_{k=1}^{K} p(\mathbf{y}_h|\mathbf{M}_k, \mathbf{y}_l)p(\mathbf{M}_{\mathrm{k}})}, \quad (18)$$

where

$$p(\mathbf{y}_h|\mathbf{M}_k, \mathbf{y}_l) = \int_{\boldsymbol{\theta}_l} p(\mathbf{y}_h|\mathbf{M}_k, \boldsymbol{\eta}_l)p(\boldsymbol{\eta}_l|\mathbf{y}_l, \boldsymbol{\theta}_l)d\boldsymbol{\theta}_l \quad (19)$$

$$p(\mathbf{y}_h|\mathbf{M}_k, \boldsymbol{\eta}_l) = \sum_{\substack{\lambda_1 = 1, 2, \ldots, \lambda_0\\ \vdots \\ \lambda_d = 1, 2, \ldots, \lambda_0}} (2\pi)^{-\frac{m_h}{2}+1}\left|\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\lambda}}\right|^{-\frac{1}{2}}\Gamma\left(\frac{m_h}{2} - 1\right) \times \left[\frac{(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^{\mathrm{T}}(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})}{2}\right]. \quad (20)$$

A uniform prior is used for $\mathbf{M}_k$, and the derivation of Equation (20) is included in Appendix C.

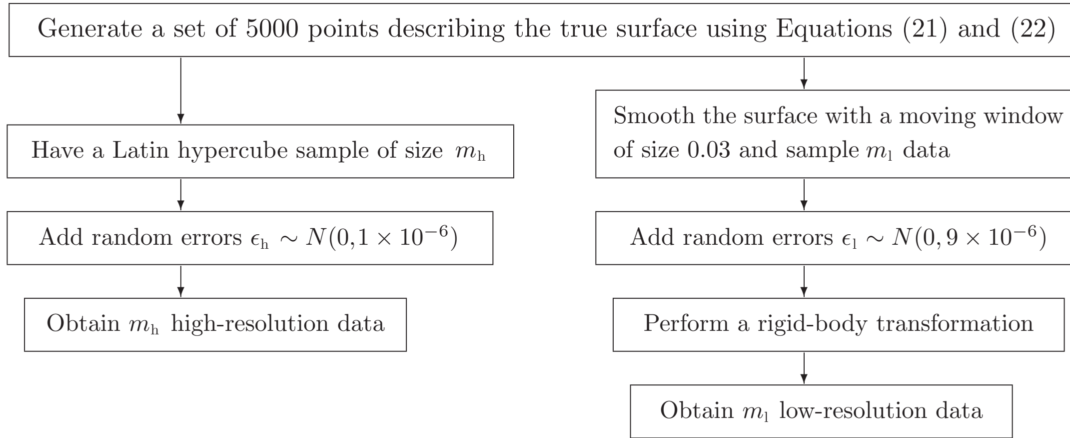

Fig. 6. Procedure to generate two-resolution data.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

4. Illustrative examples

This section presents two examples: the first example simulates two datasets of different resolutions obtained from three circular features, respectively; the second example uses measurements of a milled sinewave surface from a CMM and an OCMM as illustrated in Fig. 1. Throughout this section, the values of physical quantities are expressed in millimeters unless otherwise indicated.

4.1. *Circular features*

Circular features manufactured by turning processes are simulated using the formula proposed in Desta *et al.* (2003):

$$u = u_1 + (r + A_1 \sin(4\tau) + A_2 \cos(3\tau) + A_3 \sin(7\tau) + A_4 \cos(10\tau) + \xi) \cos\tau, \quad (21)$$

$$v = v_1 + (r + A_1 \sin(4\tau) + A_2 \cos(3\tau) + A_3 \sin(7\tau) + A_4 \cos(10\tau) + \xi) \sin\tau, \quad (22)$$

where $A_1 \sin(4\tau) + A_2 \cos(3\tau) + A_3 \sin(7\tau) + A_4 \cos(10\tau)$ represents the low-frequency manufacturing errors showing some systematic pattern; ξ represents the high-frequency, random manufacturing errors; and the other notations follow those in Fig. 3.

In the simulation, (u_1, v_1) is set to be (0.05, 0.02), and the values are chosen to be within the typical range of the soft-fixturing errors. A_1, A_2, A_3, and A_4 are chosen to be 0.03, −0.02, −0.01, and −0.008, respectively, to reflect typical manufacturing errors of a turning process (Groover, 2004, p. 85). A standard deviation of 0.002 is chosen for ξ, according to the roughness level of a turning process (Groover, 2004, p. 86). In other words, ξ is simulated by $N(0, 4 \times 10^{-6})$.

Three circular features are simulated using Equations (21) and (22) with radius $r = 41, 100$, and 150, respectively. Each of these simulated circles is treated as a true surface to be measured.

The high-resolution data are taken to be scattered over the surface, but each of them can be very close to the true surface. Thus m_h data points are first sampled from the simulated surface via a Latin hypercube sampling. Then independent random noises ϵ_h of variance of 1×10^{-6} (i.e., a standard deviation of 1 μm) are added to them.

The low-resolution data are usually the averages of the surface coordinates in a spatial neighborhood within which a low-resolution device cannot distinguish distinct points. To simulate a device of resolution roughly 30 μm, a 30 × 30 μm moving window is first used to smooth the simulated surface. A dense sample of m_l data points is then obtained, equally spaced over the entire surface. Then, independent random noises of variance of 9×10^{-6} (i.e., with a standard deviation of 3 μm) are added. Since low-resolution data may be misaligned from high-resolution data, a rigid-body transformation is also performed to the low-resolution data. The simulation procedure is summarized in Fig. 6, and the amounts of high- and low-resolution data are listed in Table 1.

The proposed Bayesian hierarchical model predicts the surface coordinates at locations where only low-resolution data are available. The resulting predictions are compared with three alternative approaches: the low-resolution data, the predictions using the high-resolution data alone (also based on a GP model), and the predictions using a multi-resolution GP model that does not address the misalignment problem. For the last one, we use the `tgp` package for `R` (Gramacy and Taddy, 2008) which is close to the Kennedy–O'Hagan method (Kennedy and O'Hagan, 2000). Since this is a simulation study, the true circular surface is known. The predicted surface from each method is compared with the true circular surface. Table 1 lists the Mean Squared Errors (MSE) of predictions, which are calculated by averaging the squared errors between the predicted and true surfaces over the different input locations.

As Table 1 shows, the proposed Bayesian hierarchical model (in the fifth column) outperforms all the other

Table 1. Evaluation of prediction (MSE in mm^2) accuracy for a circular surface

		Predicted			
Radius	*Data amount* (m_h/m_l)	*Low-resolution*	*Using high-resolution data*	*Using two-resolution data and the proposed method (data aligned)*	*Using two-resolution data and the* `tgp` *package (data not aligned)*
41	20/500	3.83×10^{-5}	6.47×10^{-4}	9.84×10^{-6}	5.35×10^{-4}
100	20/1021	3.74×10^{-5}	1.11×10^{-3}	1.09×10^{-5}	1.98×10^{-3}
150	20/1013	3.84×10^{-5}	1.76×10^{-3}	1.81×10^{-5}	2.19×10^{-3}

alternatives. For the three simulated circular features, the integrated prediction improves (in terms of the MSE values) over the observed low-resolution data by 74.31%, 70.86%, and 52.86%, respectively. We also observe that the predictions using the high-resolution data alone perform worse than the low-resolution data. We believe that this is because the amount of high-resolution data is insufficient to allow good predictions. The last column has the predictions from using the `tgp` package, which does not align the data of different resolutions. When the two-resolution data are not aligned, the low-resolution data either contribute little to the combined prediction or mislead the integrated predictions. Thus, the `tgp` package performs worse than the proposed method. This result suggests the importance of addressing the misalignment problem when combining the multi-resolution data for prediction.

4.2. *Sinewave surface*

We have a manufactured part of size $101 \times 101 \times 51$ and its top surface is milled to be in a sinewave shape. The nominal geometry is $w = \varphi_1 \sin(2\pi(\varphi_2 + u)/\varphi_3)$ where $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$ are the dimension parameters, and φ_1, φ_2 and φ_3 are called the amplitude, the phase, and the wavelength, respectively.

Two metrology devices are used to measure the sinewave surface: the CMM is a Sheffield Discovery II D-8 with a TB 20 touch probe; the OCMM is a LDI Surveyor DS-2020 with a RPS 150 laser unit. The CMM has a resolution of roughly 5 μm, whereas the OCMM has a resolution of about 50 μm. Even though this CMM does not have a resolution as high as many used in precision engineering, it serves as the high-resolution device in this particular pair. Moreover, the CMM and the OCMM have their own measuring platforms so that the part is repositioned while being measured on the second machine.

When measuring with the OCMM, it results in a low-resolution dataset of $m_l = 1560$ (i.e., 40×39) points, which are evenly spaced over the surface with approximately 2.54 μm inter-point distance in both u and v axes; see Fig. 7. The CMM measures the same surface and obtains a high-resolution dataset of also 1560 points. A large portion of the 1560 high-resolution data points, after removing those m_h points used for model building, is reserved as the

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

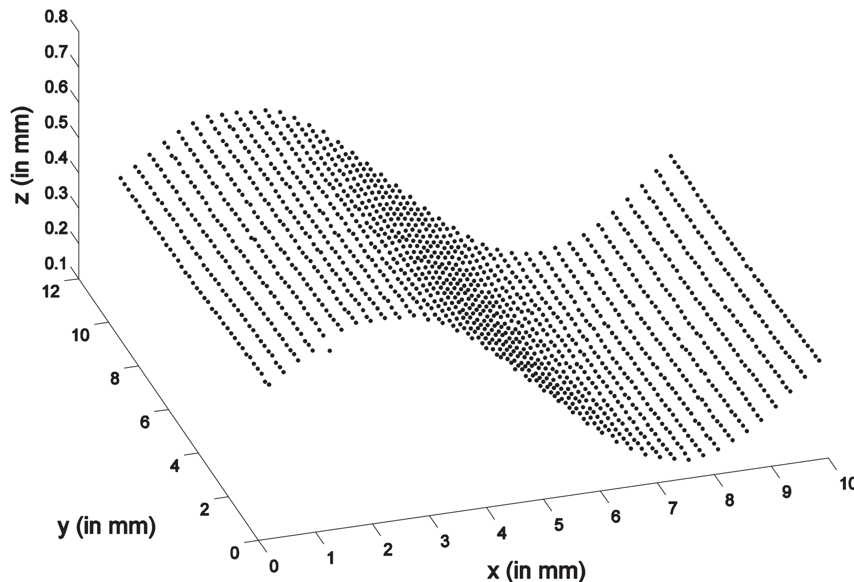

Fig. 7. Three-dimensional view of the low-resolution data, illustrating the sinewave surface.

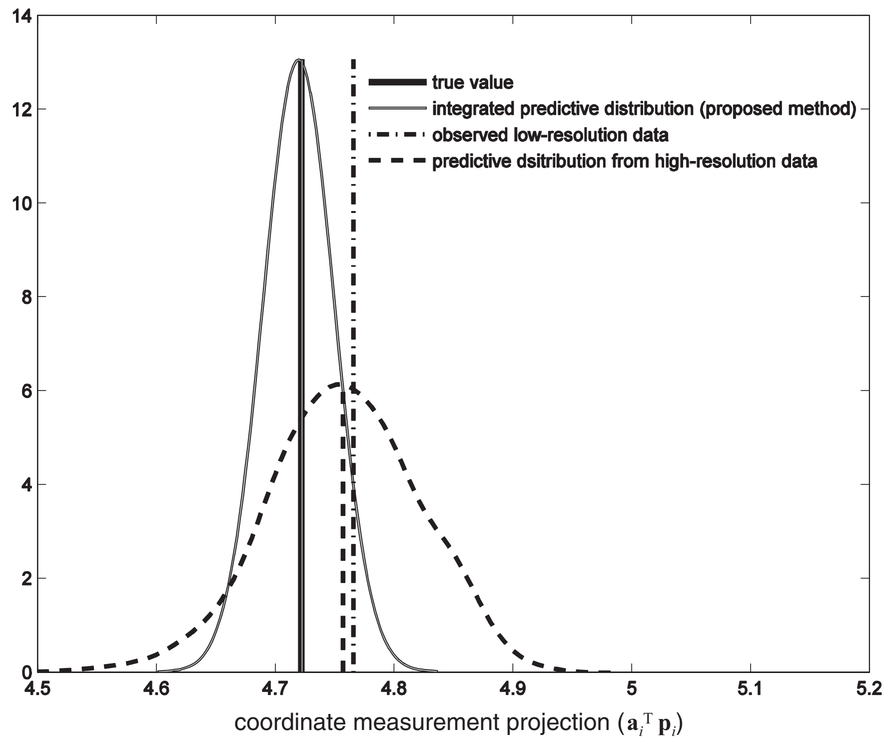

Fig. 8. Integrated prediction reduces bias and uncertainty.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

representation of the "true" surface and will be used as a benchmark for assessing the prediction quality.

Consider two cases where the high-resolution observations are of size $m_h = 20$ and 40, respectively. These observations are chosen via Latin hypercube sampling from the 1560 high-resolution data. For each case, the proposed Bayesian hierarchical model is used to make a prediction combining the m_h high-resolution observations and the $m_l = 1560$ low-resolution observations.

The position and orientation of the part on a metrology measuring platform are unknown. It may be different from its nominal position by translation **t** and rotation **H**. For this reason, the coordinates of a point on the dashed-line geometry $\mathbf{f}(\cdot, \boldsymbol{\beta}_l)$ is $\mathbf{H} \cdot (u_i, v_i, w_i)^{\mathrm{T}} + \mathbf{t}$, where (u_i, v_i, w_i) is a point on the nominal geometry, satisfying the constraint $w_i = \varphi_1 \sin(2\pi(\varphi_2 + u_i)/\varphi_3)$.

As in the simulated examples, the proposed Bayesian hierarchical model is used to predict the part surface at locations where only the low-resolution data are available. The predictions using the proposed method are compared with the observed low-resolution data, the predictions using the m_h (=20 or 40) high-resolution data alone, and the predictions using the `tgp` package.

Figure 8 shows the prediction results for one location on the part to illustrate the benefit of combining the multi-resolution information. In Fig. 8, the "true" value (i.e., the black solid line) is actually a reserved high-resolution measurement. The gray solid curve represents the integrated predictive distribution using the proposed multi-resolution method, and the gray line in the middle is the predicted median. The dashed-dotted line represents the observed low-resolution data. For this particular point, the integrated prediction almost coincides with the true value and is an improvement on the low-resolution data. The dashed line and the dashed curve denote the predicted median and the predictive distribution using the m_h high-resolution observations. It is clear that the predictive distribution using the multi-resolution data results in a much narrower distribution (i.e., a smaller uncertainty) than the prediction distribution using the high-resolution data alone. In summary, the integrated prediction produces the best prediction, less biased (in terms of the distance between the predictive median and the true value) and with reduced uncertainty.

If the two-resolution datasets are combined to make predictions without properly aligning the datasets, they may not produce the desired benefit. Still using the prediction on the above chosen location, Fig. 9 compares the predictive distribution between the proposed method and the multi-resolution GP method in the `tgp` package. The proposed method achieves a significantly better prediction than the `tgp` package.

We use all of the aforementioned methods to make predictions over the product surface and compare the predictions with the reserved high-resolution data. Table 2 summarizes the MSE values of the predictions. The results confirm the understanding garnered from the simulated example. For both high-resolution data sizes ($m_h =$ 20 or 40), the proposed Bayesian hierarchical model shows

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

Table 2. Comparison of the predictions (MSE in mm^2) for the sinewave surface

		Predicted		
Data amount (m_h/m_l)	*Low-resolution*	*Using high-resolution data*	*Using two-resolution data and the proposed method (data aligned)*	*Using two-resolution data and the* `tgp` *package (data not aligned)*
20/1560	1.18×10^{-3}	1.04×10^{-2}	7.85×10^{-4}	3.24×10^{-2}
40/1560	1.05×10^{-3}	2.44×10^{-3}	6.10×10^{-4}	3.60×10^{-3}

significant improvements from using the low-resolution observations. When $m_h = 20$ and $m_l = 1560$, the proposed method improves the prediction of low-resolution data by 25.24%. When $m_h = 40$ and $m_l = 1560$, the proposed method improves the prediction of low-resolution data by 30.86%. When the number of high-resolution data increases from 20 to 40, the prediction using the proposed Bayesian hierarchical method improves. The improvement, however, is not proportional to the increase in the amount of high-resolution data. Studying what is the most beneficial data ratio between the low- and high-resolution data is worth further exploration.

Recall that a discrete uniform distribution $U(1, \lambda_0)$ is assigned as the prior distribution of the kernel width $\boldsymbol{\lambda}$. In this example, we choose $\lambda_0 = 30$ and $c = 2.54$, the average inter-point distance in the low-resolution data. Figure 10 shows the marginal posterior distributions of the kernel widths in the u- and v-axes, respectively. The kernel width along the u-axis λ_1 is 2.54, suggesting that only the nearest low-resolution data point in the u-direction has a strong connection with each high-resolution data point. For the v-axis, the kernel width λ_2 of 7.62 has the highest probability, and the probabilities at the 11 kernel widths have a noticeable non-zero mass. This makes good sense for the sinewave surface in Fig. 1, which is easier to align along the u-direction than along the v-direction because of a much larger surface slope in the u-axis than in the v-axis.

The Bayesian hierarchical method assigns posterior weights to different kernel widths according to how much they are consistent with the data. As shown in Fig. 10, the kernel widths greater than 27.94 (or 11 inter-point distances) barely have any posterior weight and will not contribute much to the posterior inference in this application. This is not a surprise because after all the alignment efforts, the remaining misalignment is presumed to be small. This implies that one can choose a big upper bound λ_0 so that

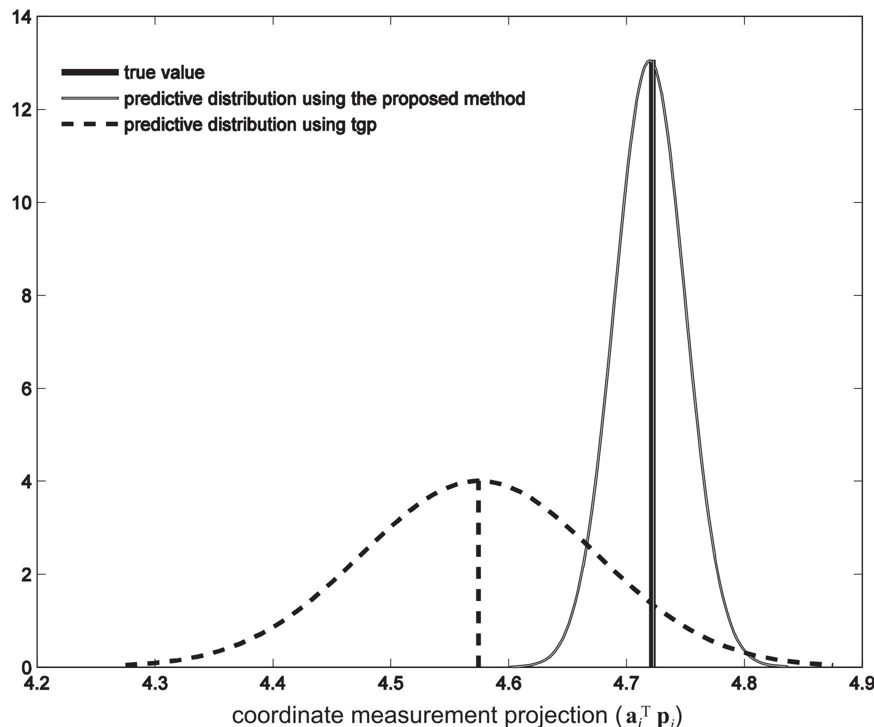

Fig. 9. Integrated prediction using misaligned data could lead to worse results.

Table 3. Sensitivity analysis of the choices of c ($c = i \times 2.54$ mn).

	i				
	1	*0.75*	*0.5*	*0.25*	*0.1*
MSE	7.853×10^{-4}	7.782×10^{-4}	7.772×10^{-4}	7.818×10^{-4}	7.822×10^{-4}

the posterior inference of the prediction is not sensitive to the prior choice for $\boldsymbol{\lambda}$. In this example where λ_0 is chosen to be 30 (i.e., the kernel width is 76.20), much greater than 11, we believe that the posterior inference on $\boldsymbol{\lambda}$ largely reflects the information from the data.

We also conducted a sensitivity analysis of the integrated predictions under different increments of c (using $m_h = 20$ and $m_l = 1560$). The results are summarized in Table 3. It shows that the integrated predictions are insensitive to the value of c when c takes a value less than or equal to the average inter-point distance of the low-resolution data.

5. Concluding remarks

We present a Bayesian hierarchical model for combining misaligned two-resolution metrology measurements for the purpose of predicting a product surface. The current model is not a fully Bayesian model because a heuristic matching and alignment algorithm is used for finding the set of the consistent matches. Our final prediction is approximated by averaging over the set of consistent matches. A fully Bayesian approach needs the development of a Bayesian alignment model. Green and Mardia (2006) developed a Bayesian alignment model for a different problem but its application to the metrology problem does not appear straightforward. Several critical issues still remain to be solved, including the effectiveness and efficiency of solving the Bayesian alignment model and the integration of the alignment model with the other models into a Bayesian hierarchical modeling framework.

In the current study, we only used data evenly spaced over the entire product surface. The evenly spaced measurements are readily achievable using today's metrology technology. We also expect that our method does not heavily rely on the data being evenly spaced. Nevertheless, we acknowledge that it is possible to have unevenly spaced low-resolution data from scanning a high-curvature feature of

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

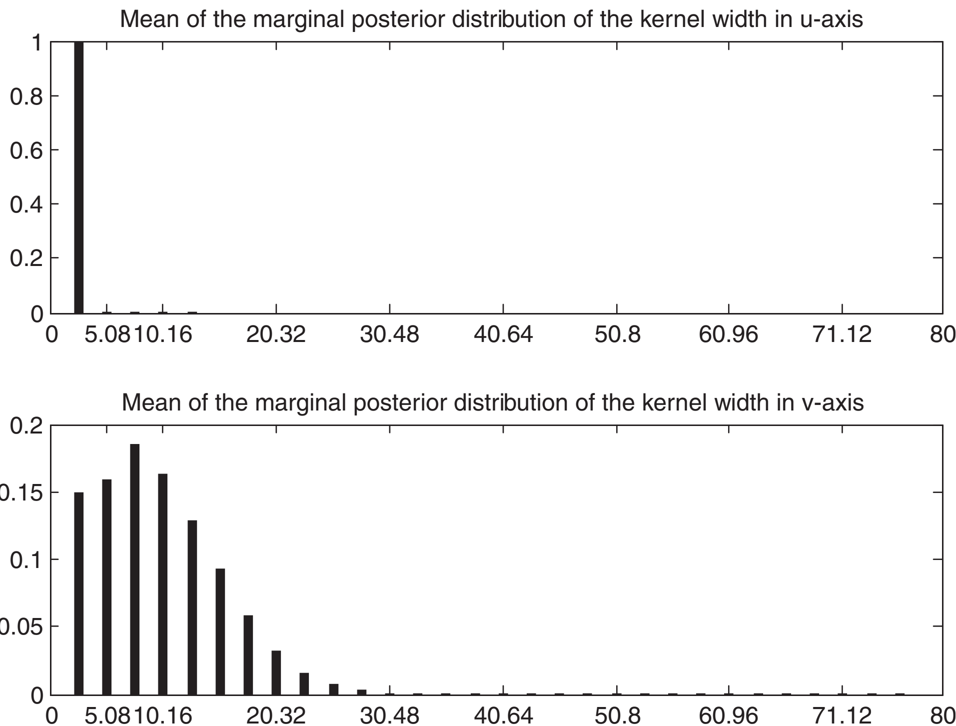

Fig. 10. Marginal posterior distributions of the kernel widths in the u-axis and v-axis.

a complicated three-dimensional part, and the effect (or the non-effect) of the data distribution on our proposed method has not yet been experimentally verified. It would be valuable in future work to conduct a comprehensive experimental study to see how the proposed method works for unequally spaced low-resolution data, different geometry complexity, and other practical considerations.

If the high-resolution data do not cover the same surface area as the low-resolution data, the matching results may become trapped in local minima, leading to an incorrect coordinate transformation and poor alignment. Thus, if engineers are concerned about fully utilizing the information in the two datasets that requires a good alignment, they should plan for it during the data collection stage, in order to ensure that both datasets cover the whole workpiece surface.

Our study uses a univariate response model, which was first developed by Hulting (1995, 1997). Using this univariate response leads to a univariate GP model for the low-resolution data. Recently, progress has been made to extend GP modeling to multivariate cases; for example, Ankenman *et al*. (2010), Santner *et al.* (2003), and Wackernage (2003). However, a multivariate GP model for metrology applications is not yet available. Since projecting the multivariate measurements $\mathbf{a}_i$ onto the approaching direction $\mathbf{p}_i$ might introduce additional deviations, it is worthwhile to explore the benefit of having a multivariate GP model for handling $\mathbf{a}_i$ directly.

The effectiveness of the integrated prediction depends on the data ratio in the multi-resolution datasets as well as on the resolution ratio of the metrology devices. The sinewave example shows that when the amount of high-resolution data is doubled (i.e., the cost of measurement is almost doubled), the prediction errors are far from being halved. Studying how to produce the most cost-effective benefit in combining multi-resolution data for prediction merits further exploration as well.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

References

Ankenman, B., Nelson, B.L. and Staum, J. (2010) Stochastic kriging for simulation metamodeling. *Operations Research*, **58**(2), 371–382.

Brown, L. (1992) A survey of image registration techniques. *ACM Computing Surveys*, **24**, 326–376.

Carbone, V., Carocci, M., Savio, E., Sansoni, G. and Chiffre, L. (2001) Combination of a vision system and a coordinate measuring machine for the reverse engineering of freeform surfaces. *International Journal of Advanced Manufacturing Technology*, **17**, 263–271.

Cheesman, P., Kanefsky, B., Kraft, R., Stutz, J. and Hanson, R. (1996) Super-resolved surface reconstruction from multiple images, in *Maximum Entropy and Bayesian Methods*, Heidberg, G.R. (ed.), Kluwer, Amesterdam, The Netherlands, pp. 293–308.

Chen, L. and Lin, G. (1997) A vision-aided reverse engineering approach to reconstructing free-form surfaces. *Robotics and Computer-Integrated Manufacturing*, **13**, 323–336.

Desta, M., Feng, H. and OuYang, D. (2003) Characterization of general systematic form errors for circular features. *International Journal of Machine Tools and Manufacture*, **43**, 1069–1078.

Dryden, I. and Mardia, K. (1998) *Statistical Shape Analysis*, John Wiley & Sons, Hoboken, NJ.

Ferreira, M., Higdon, D. and Lee, H. (2005) Multi-scale random field models. ISDS Discussion Paper 05-02, available at http://ftp.stat.duke.edu/WorkingPapers/05-02.html (accessed January 3, 2011).

Gelman, A., Carlin, J., Stein, H. and Rubin, D. (2003) *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton, FL.

Gotway, C. and Young, L. (2002) Combining incompatible spatial data. *Journal of the American Statistical Association*, **97**, 632–648.

Gramacy, R. and Taddy, M. (2008) The tgp Package, available at http://www.ams.ucsc.edu/ rbgrammacy/tgp.html (accessed January 3, 2011).

Green, P. and Mardia, K. (2006) Bayesian alignment using hierarchical models, with applications in protein bioinformatics. *Biometrika*, **93**, 235–254.

Groover, M. (2004) *Fundamentals of Modern Manufacturing—Materials, Processes and Systems*, John Wiley & Sons, New York, NY.

Higdon, D., Kennedy, M., Cavendish, J., Cafeo, J. and Ryne, R. (2004) Combining field data and computer simulations for calibration and prediction. *SIAM Journal of Scientific Computation*, **26**, 448–446.

Hulting, F. (1995) Comment: an industrial view of coordinate measurement data analysis. *Statistica Sinica*, **5**, 191–204.

Hulting, F. (1997) Discussion: statistical issues in geometric feature inspection using coordinate measuring machines. *Technometrics*, **39**, 18–20.

Kennedy, M. and O'Hagan, A. (2000) Predicting the output from a complex computer code when fast approximations are available. *Biometrica*, **87**, 1–13.

Kennedy, M. and O'Hagan, A. (2001) Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B* (*Statistical Methodology*), **63**, 425–464.

Mitra, R. and Murthy, N. (1991) Elastic, maximal matching. *Pattern Recognition*, **24**, 747–753.

Motavalli, S., Suharitdamrong, V. and Alrashdan, A. (1998) Design model generating for reverse engineering using multi-sensors. *IIE Transactions*, **30**, 357–366.

Qian, Z., Seepersad, C., Joseph, R., Allen, J. and Wu, C.F.J. (2006) Building surrogate models with detailed and approximate simulations. *ASME, Journal of Mechanical Design*, **128**, 668–677.

Qian, Z. and Wu, C.F.J. (2008) Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, **50**, 192–205.

Ranade, S. and Rosenfeld, A. (1980) Point pattern matching by relaxation. *Pattern Recognition*, **12**, 269–275.

Reese, S., Wilson, A., Hamada, M., Martz, H. and Ryan, K. (2004) Integrated analysis of computer and physical experiments. *Technometrics*, **46**, 153–164.

Santner, T., Williams, B. and Notz, W. (2003) *The Design and Analysis of Computer Experiments*, Springer, New York, NY.

Shen, T., Huang, J. and Menq, C. (2000) Multiple-sensor integration for rapid and high-precision coordinate metrology. *IEEE/ASME Transaction on Mechatronics*, **5**, 110–121.

Ton, J. and Jain, A. (1989) Registering landsat images by point matching. *IEEE Transactions on Geoscience and Remote Sensing*, **27**, 642–651.

Wackernagel, H. (2003) *Multivariate Geostatistics: An Introduction with Applications*, third edition, Springer-Verlag, Berlin.

Wikle, C. and Berliner, M. (2005) Combining information across spatial scales. *Technometrics*, **47**, 80–91.

Xia, H., Ding, Y. and Wang, J. (2008) Gaussian process method for form error assessment using coordinate measurements. *IIE Transactions*, **40**, 931–946.

Zitova, B. and Flusser, J. (2003) Image registration methods: a survey. *Image and Vision Computing*, **21**, 977–1000.

Appendices

Appendix A: Choosing parameters for Bayesian priors

This section provides additional thoughts on how to choose the parameters for the prior distributions. The $\boldsymbol{\beta}_l$ includes the parameters accounting for a part's actual location (on a measuring platform) and dimension. In Fig. 3, the dashed-line geometry $\mathbf{f}(\cdot, \boldsymbol{\beta}_l)$ deviates slightly in location from the nominal geometry. The deviation is small due to the use of the soft-fixturing process. There is also some difference in dimensions between these two geometries, but it is small because a manufacturing process normally produces the required dimension with reasonable accuracy. By treating the location of the nominal geometry to be at the origin of the coordinate system, we can assign the mean component of the prior distribution of $\boldsymbol{\beta}_l$ as $\boldsymbol{\mu}_l = (\mathbf{0}, \boldsymbol{\varphi}^*)$, where $\boldsymbol{\varphi}^*$ is the nominal dimensions of the part. The variance q_i^2 can be determined from a crude least-squares estimation of the part's location as well as the typical manufacturing process capability (Groover, 2004, p. 85) causing variability in actual part's dimensions.

For the prior distribution of σ_l^2 and κ_l^2, we choose $\alpha_1 = \alpha_2 = 1$ and $\beta_1 = \beta_2 = 1 \times 10^{-4}$ so that σ_l and κ_l mostly take values smaller than 0.2 mm (or 200 μm), which is consistent with engineers' knowledge on the error magnitudes from the manufacturing processes and the metrology devices. Also, this prior distribution gives higher probabilities to the smaller values as they correspond to a better fit of the low-resolution data to the model. The prior chosen for each correlation parameter ν_i is Gamma(0.01, 0.01), effectively a non-informative prior (Gelman *et al.*, 2003).

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011

Appendix B: Proof of Equation (14)

Recall that the model parameters are $\theta = (\theta_l, \theta_h)$. We express $p(y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h)$ as

$$\begin{aligned} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_l, \mathbf{y}_h) &= \int_{\theta_l}\int_{\theta_h} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_l, \mathbf{y}_h, \theta_l, \theta_h) \\ &\quad \times p(\theta_l, \theta_h|\mathbf{M}_k, \mathbf{y}_l, \mathbf{y}_h) d\theta_h d\theta_l \\ &= \int_{\theta_l}\int_{\theta_h} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \eta_l, \theta_h) \\ &\quad \times p(\theta_h|\mathbf{M}_k, \eta_l(\mathbf{X}_l), \mathbf{y}_h) d\theta_h \\ &\quad \times p(\eta_l|\mathbf{y}_l, \theta_l) \times p(\theta_l|\mathbf{y}_l) d\theta_l \\ &= \int_{\theta_l} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \eta_l) \\ &\quad \times p(\eta_l|\mathbf{M}_k, \mathbf{y}_l, \theta_l) p(\theta_l|\mathbf{y}_l) d\theta_l \end{aligned} \tag{A1}$$

where

$$\begin{aligned} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \eta_l, \mathbf{y}_h) &\equiv \int_{\theta_h} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \eta_l, \theta_h) \\ &\quad \times p(\theta_h|\mathbf{M}_k, \eta_l, \mathbf{y}_h) d\theta_h \\ &= \int_{\boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \eta_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) \\ &\quad \times p(\boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}|\mathbf{M}_k, \eta_l, \mathbf{y}_h) \\ &\quad \times d\boldsymbol{\alpha}\, d\sigma_e^2\, d\boldsymbol{\lambda} \\ &= \int_{\boldsymbol{\lambda}}\int_{\sigma_e^2, \boldsymbol{\alpha}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \eta_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) \\ &\quad \times p(\boldsymbol{\alpha}, \sigma_e^2|\mathbf{M}_k, \boldsymbol{\lambda}, \mathbf{y}_h, \eta_l)\, d\boldsymbol{\alpha}\, d\sigma_e^2 \\ &\quad \times p(\boldsymbol{\lambda}|\mathbf{M}_k, \eta_l, \mathbf{y}_h) d\boldsymbol{\lambda} \end{aligned} \tag{A2}$$

In order to get the expression of $p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \eta_l, \mathbf{y}_h)$, perform the integration in (A2) in the following two steps:

(i) Integrate out $\boldsymbol{\alpha}$ and σ_e^2;
(ii) Integrate out $\boldsymbol{\lambda}$.

Step (i): integrate out $\boldsymbol{\alpha}$ and σ_e^2. We denote the inner integration in Equation (A2) by $p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \eta_l, \boldsymbol{\lambda})$, that is,

$$\begin{aligned} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \eta_l, \boldsymbol{\lambda}) &\equiv \int_{\sigma_e^2, \boldsymbol{\alpha}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \eta_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) \\ &\quad \times p(\boldsymbol{\alpha}, \sigma_e^2|\mathbf{M}_k, \boldsymbol{\lambda}, \mathbf{y}_h, \eta_l)\, d\boldsymbol{\alpha}\, d\sigma_e^2 \\ &\propto \int_{\sigma_e^2, \boldsymbol{\alpha}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \eta_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) p(\mathbf{y}_h|\mathbf{M}_k, \eta_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) \\ &\quad \times p(\boldsymbol{\alpha}, \sigma_e^2)\, d\boldsymbol{\alpha}\, d\sigma_e^2. \end{aligned}$$

Given the kernel width $\boldsymbol{\lambda}$, the linkage model can be considered as a linear regression model $\mathbf{y}_h = \mathbf{F}_{\boldsymbol{\lambda}}\boldsymbol{\alpha} + \boldsymbol{\epsilon}_h$. Recall that $\boldsymbol{\epsilon}_h \sim N(\mathbf{0}, \sigma_e^2\mathbf{I})$. Therefore,

$$\begin{gathered} \left(\mathbf{y}_h|\mathbf{M}_k, \eta_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}\right) \sim N\left(\mathbf{F}_{\boldsymbol{\lambda}}\boldsymbol{\alpha}, \sigma_e^2\mathbf{I}\right) \\ (y_h(\mathbf{x}_0)|\mathbf{M}_k, \eta_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) \sim N(\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x}_0)\boldsymbol{\alpha}, \sigma_e^2) \end{gathered}$$

These are the same results as in Equations (10) and (11). Given that the prior distribution of $\boldsymbol{\alpha}$ and σ_e^2 is $p(\boldsymbol{\alpha}, \sigma_e^2) \propto \sigma_h^{-2}$, Gelman *et al.* (2003, p. 359) stated that under this priors, the posterior predictive distribution of $y_h(\mathbf{x}_0)$, conditioned on the data and kernel width $\boldsymbol{\lambda}$, is

$$\begin{aligned} &(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \eta_l(\mathbf{X}_0), \boldsymbol{\lambda}) \\ &\quad \sim t_{m_h-2}(\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x}_0)\hat{\boldsymbol{\alpha}}, s^2(1 + \mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x}_0)(\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\lambda}})^{-1}\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x}_0)^{\mathrm{T}})). \end{aligned}$$

where $\hat{\boldsymbol{\alpha}} = (\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\lambda}})^{-1}\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{y}_h$ and $s^2 = \frac{1}{m_h-2}(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^{\mathrm{T}}(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})$. This is how Equation (14) is obtained. Consequently, after $\boldsymbol{\alpha}$ and σ_e^2 are integrated out, Equation (A2) becomes

$$\begin{aligned} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \eta_l, \mathbf{y}_h) &= \int_{\boldsymbol{\lambda}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \eta_l(\mathbf{X}_0), \boldsymbol{\lambda}) \\ &\quad \times p(\boldsymbol{\lambda}|\mathbf{M}_k, \boldsymbol{\beta}_l, \mathbf{y}_h) d\boldsymbol{\lambda}. \end{aligned} \tag{A3}$$

Step (ii), integrate out $\boldsymbol{\lambda}$. Recall that $\boldsymbol{\lambda}$ has a discrete distribution. Thus, the integration in (A3) can be written as a

summation (and Equation (15) is obtained):

$$p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \eta_l, \mathbf{y}_h) = \sum_{\substack{\lambda_1=1,2,\ldots,\lambda_0 \\ \vdots \\ \lambda_d=1,2,\ldots,\lambda_0}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \eta_l, \boldsymbol{\lambda}) \, p(\boldsymbol{\lambda}|\mathbf{M}_k, \mathbf{y}_h, \eta_l),$$

where

$$p(\boldsymbol{\lambda}|\mathbf{M}_k, \mathbf{y}_h, \eta_l) \propto p(\boldsymbol{\lambda}) p(\mathbf{y}_h|\mathbf{M}_k, \eta_l, \boldsymbol{\lambda}). \tag{A4}$$

The marginal distribution of the high-resolution data given the inputs η_l and the kernel width $\boldsymbol{\lambda}$ is as follows

$$\begin{aligned}
p(\mathbf{y}_h|\mathbf{M}_k, \eta_l, \boldsymbol{\lambda}) &= \int_{\sigma_e^2, \boldsymbol{\alpha}} p(\mathbf{y}_h, \boldsymbol{\alpha}, \sigma_e^2|\mathbf{M}_k, \eta_l, \boldsymbol{\lambda}) \, d\boldsymbol{\alpha} \, d\sigma_e^2 \\
&= \int_{\sigma_e^2, \boldsymbol{\alpha}} p(\mathbf{y}_h|\mathbf{M}_k, \eta_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) \\
&\quad \times p(\boldsymbol{\alpha}, \sigma_e^2) \, d\boldsymbol{\alpha} \, d\sigma_e^2 \\
&= \int_{\sigma_e^2} \int_{\boldsymbol{\alpha}} (2\pi)^{-\frac{m_h}{2}} (\sigma_e^2)^{-\frac{m_h}{2}} \\
&\quad \times \exp\left\{-\frac{1}{2\sigma_e^2} (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \boldsymbol{\alpha})^{\mathrm{T}} (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \boldsymbol{\alpha})\right\} \\
&\quad \times d\boldsymbol{\alpha} \, \sigma_h^{-2} d\sigma_e^2 \\
&= \int_{\sigma_e^2} \int_{\boldsymbol{\alpha}} \exp\left\{-\frac{1}{2\sigma_e^2} [(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})^{\mathrm{T}} (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}}) \right. \\
&\quad \left. + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\lambda}} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})]\right\} \, d\boldsymbol{\alpha} \\
&\quad \times (2\pi)^{-\frac{m_h}{2}} (\sigma_e^2)^{-\frac{m_h}{2}-1} d\sigma_e^2 \\
&= \int_{\sigma_e^2} \int_{\boldsymbol{\alpha}} \exp\left\{-\frac{1}{2\sigma_e^2} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\lambda}} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})\right\} \\
&\quad \times d\boldsymbol{\alpha} (2\pi)^{-\frac{m_h}{2}} \exp\left\{-\frac{1}{2\sigma_e^2} (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})^{\mathrm{T}} \right. \\
&\quad \left. \times (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})\right\} (\sigma_e^2)^{-\frac{m_h}{2}-1} d\sigma_e^2 \\
&= \int_{\sigma_e^2} 2\pi |\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}} \sigma_e^2 (2\pi)^{-\frac{m_h}{2}} \\
&\quad \times \exp\left\{-\frac{1}{2\sigma_e^2} (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})^{\mathrm{T}} (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})\right\} \\
&\quad \times (\sigma_e^2)^{-\frac{m_h}{2}-1} d\sigma_e^2 \\
&= (2\pi)^{-\frac{m_h}{2}+1} |\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}} \int_{\sigma_e^2} (\sigma_e^2)^{-(\frac{m_h}{2}-1+1)} \\
&\quad \times \exp\left\{-\frac{(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})^{\mathrm{T}} (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})}{2\sigma_e^2}\right\} d\sigma_e^2 \\
&= (2\pi)^{-\frac{m_h}{2}+1} |\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}} \Gamma\left(\frac{m_h}{2} - 1\right) \\
&\quad \times \left[\frac{(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})^{\mathrm{T}} (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})}{2}\right]^{-\frac{m_h}{2}+1}.
\end{aligned} \tag{A5}$$

Note that in the second line of this derivation, we utilize that (11) which specifies the distribution of $p(\mathbf{y}_h|\eta_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda})$. Given the above, Equation (A4) can now be written as

$$\begin{aligned}
p(\boldsymbol{\lambda}|\mathbf{y}_h, \eta_l) &\propto p(\boldsymbol{\lambda}) |\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}} \\
&\quad \times \left[\frac{(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})^{\mathrm{T}} (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})}{2}\right]^{-\frac{m_h}{2}+1}
\end{aligned}$$

This shows how Equation (14) is obtained.

Appendix C: Proof of Equation (20)

To prove Equation (20), we just integrate Equation (A6) over $\boldsymbol{\lambda}$. As $\boldsymbol{\lambda}$ takes discrete values, we end up with a summation over possible values for $\boldsymbol{\lambda}$ as in Equation (20):

$$p(\mathbf{y}_h|\mathbf{M}_k, \eta_l) = \sum_{\substack{\lambda_1=1,2,\ldots,\lambda_0 \\ \vdots \\ \lambda_d=1,2,\ldots,\lambda_0}} (2\pi)^{-\frac{m_h}{2}+1} |\mathbf{F}_{\boldsymbol{\lambda}}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}} \Gamma\left(\frac{m_h}{2} - 1\right) \times \left[\frac{(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})^{\mathrm{T}} (\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}} \hat{\boldsymbol{\alpha}})}{2}\right].$$

Biographies

Haifeng (Heidi) Xia is a senior research engineer at the Dow Chemical Company. She received her Ph.D. in Industrial and Systems Engineering from Texas A&M University in 2008. She currently works in the area of information extraction and integration, process control and diagnosis as well as design of experiment for mixtures. She serves as a reviewer for the journals *IIE Transactions* and *Technometrics*.

Yu Ding received a B.S. in Precision Engineering from the University of Science and Technology of China in 1993; an M.S. in Precision Instruments from Tsinghua University, China, in 1996; an M.S. in Mechanical Engineering from the Pennsylvania State University in 1998; and a Ph.D. in Mechanical Engineering from the University of Michigan in 2001. He is currently an Associate Professor and Holder of the Centerpoint Energy Career Development Professorship in the Department of Industrial and Systems Engineering at Texas A&M University. His research interests are in the area of quality and reliability engineering and systems informatics. He currently serves as a Department Editor of *IIE Transactions* and is a member of IEEE, INFORMS, IIE, and ASME.

Bani K. Mallick is a Professor in the Statistics department at Texas A&M University. He holds B.S. and M.S. degrees from Calcutta University. He received his Ph.D. (1994) in Statistics from the University of Connecticut, Storrs. His research interest is in Bayesian modeling and computation. He is the author of two books, four edited books, and over 100 refereed journal publications and was named an elected fellow of the American Statistical Association and Institute of Mathematical Statistics.

Downloaded By: [Texas A&M University] At: 18:24 28 March 2011