Qiang Zhou
Department of Industrial and Systems Engineering,
University of Wisconsin,
Madison, 1513 University Avenue,
Madison, WI 53706
e-mail: qzhou3@wisc.edu

Peter Z. G. Qian
Department of Statistics,
1300 University Avenue,
Madison, WI 53706
e-mail: peterq@stat.wisc.edu

Shiyu Zhou[1]
Department of Industrial and Systems Engineering,
University of Wisconsin,
Madison, 1513 University Avenue,
Madison, WI 53706
e-mail: szhou@engr.wisc.edu

Surrogate Modeling of Multistage Assembly Processes Using Integrated Emulation

For the design and optimization of multistage assembly processes, a computationally cheap mathematical model that links design parameters with the final product dimensional quality is highly desirable. We propose a systematic approach to building a surrogate model of simulations of multistage assembly processes. At the heart of this approach is a multiple-input-multiple-output surrogate modeling framework that uses a recently developed integrated emulation technique. The unique feature of this technique is that the surrogate models for multiple outputs are fitted simultaneously. The corresponding experimental design issues are also addressed. The proposed method provides good prediction accuracy and requires minimal physical knowledge of the underlying system. The effectiveness of the method is demonstrated through a case study.
[DOI: 10.1115/1.4005440]

Keywords: surrogate modeling, integrated emulation, multistage assembly process, dimensional quality

1 Introduction

As a critical stage in manufacturing, assembly process joins individual parts together to form functional products. Dimensional quality of the assembled products is a crucial factor influencing the final product functionality and customer satisfaction. Therefore, dimensional variation reduction has long been an important issue in the design of product assembly systems. A well-designed assembly system should achieve dimensional variation reduction at a minimal cost in order to provide the products with greater competitiveness in the market.

To design an assembly process with good dimensional quality, a model to link the dimensional quality of the final product with various design and process parameters, such as nominal locations of fixture locators, is essential. For an assembly process, such a model can be used to evaluate various design alternatives and predict final dimensional quality of the product. This type of model can be included as an evaluating core into an optimization loop to find the optimal design parameters. Existing methods for building such models can be roughly categorized into three types: (1) computer simulation models, (2) physical models based on kinematic analysis, and (3) surrogate models based on computer simulation.

Commercially available computer softwares, such as 3DCS ANALYST and VIS VSA, can be used to build complex assembly models and provide results close to reality. Monte Carlo simulation is often used to estimate variations of dimensional errors. Running such computer models can be time-consuming given the complex nature of multistage assembly systems. Furthermore, the simulation model is built for a specific design configuration while optimization requires easy reconfiguration of the model so that a number of design alternatives can be evaluated. This can be difficult in practice for simulation models that only allow manual adjustment. For the above reasons, computer simulation models are primarily used for design validation purpose and not ideal for design optimization.

The second type of model relates product dimensional quality with design parameters and process errors through kinematic analysis of part–part and part-fixture interactions and has been studied extensively in the literature [1–6]. Since a multistage assembly process consists of multiple stations and involves a large number of design parameters, the physical models are first built at the station level, and then linked together. Closed-form mathematical formulas are derived and the model can be evaluated quickly. As these models are derived based on physical laws, the entire assembly process has to be thoroughly studied. In a multistage assembly process, large numbers of design parameters and the complexity of the process make the model building process very time-consuming, tedious, and error-prone. Therefore, in practice, these models are often used under fixed fixture layouts for tolerance analysis/synthesis and fault diagnosis (e.g., Refs. [7–9]), with relatively few other applications such as fixture layout design (e.g., Refs. [10,11]) and active control [12].

The third type of model, i.e., a surrogate model based on computer simulation, is situated in the middle of the two types discussed above. First, a set of carefully designed computer simulations are conducted to obtain training data. A surrogate model built based on the training data is then used as a predictor for untried designs. This method has several advantages over the first two types of models. First of all, it is conceptually simple and very flexible; the modeling complexity does not increase much for complicated locating schemes and it can be implemented whenever training data can be obtained from computer simulation, irrespective of the complexity of the underlying process. Second, it does not require detailed analytical derivations of the underlying physical system. Third, surrogate models are expressed in closed-form mathematical formulas and can be evaluated quickly. Because of these advantages, surrogate modeling has drawn some attention recently. Kim and Ding developed a data-mining methods guided approach for fixture layout design using feature functions as surrogate of the objective function [13]. Huang et al. proposed a yield surrogate model based approach for tolerance synthesis in multistation manufacturing systems under fixed fixture layout [7]. The surrogate models in their work are not built for directly linking design parameters with final product dimensional variations. Loose et al. developed a hybrid approach to building a surrogate model for assembly processes [14]. In their method, the mathematical structure of the model is first derived based on physical principles, and then the model parameters are fitted using the training dataset obtained through computer simulation. This method is limited to 3-2-1 fixture layout and, like the second type of model, the model derivation is lengthy, tedious and

[1]Corresponding author.

Contributed by the Design Automation Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received August 15, 2010; final manuscript received August 6, 2011; published online January 4, 2012. Assoc. Editor: Michael Kokkolaras.

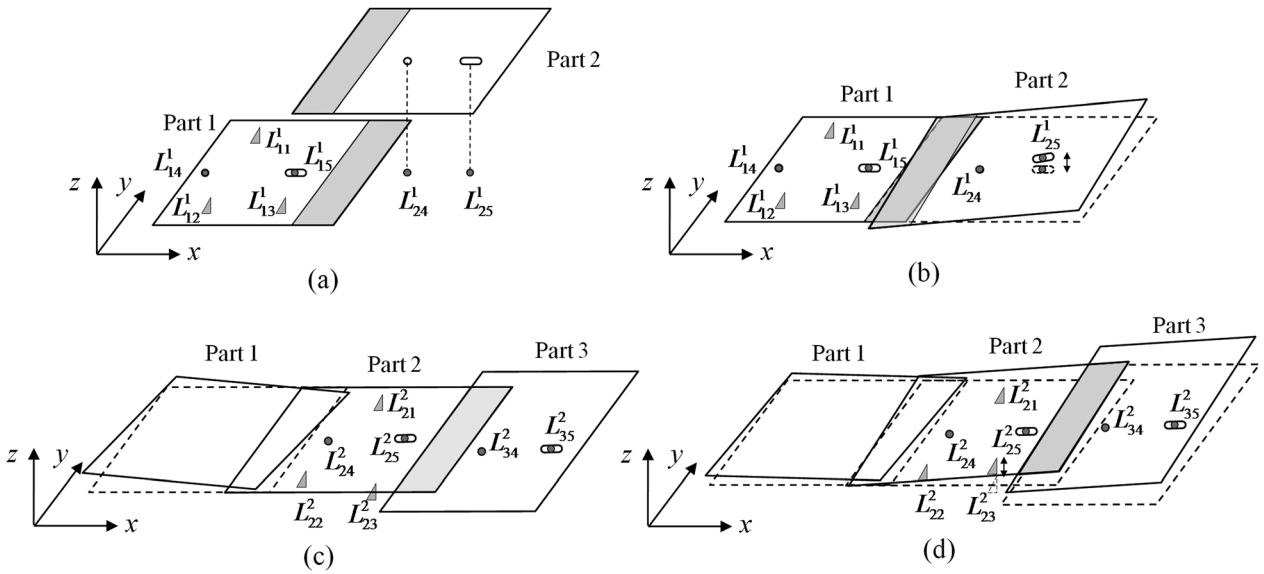

Fig. 1 Illustration of dimensional variation propagation in a multistage assembly system

error-prone. Beyond these research works that are specifically targeted on assembly process modeling, it is also possible to apply generic surrogate modeling techniques to the multistage assembly process. For example, Kriging has been one of the most popular surrogate modeling techniques applied to many engineering problems [15–18]. However, the conventional Kriging model deals with univariate output and thus a large number of Kriging models have to be built for the multistage assembly process because of the large number of quality measurements. The multivariate nature of assembly processes calls for a more viable approach for surrogate modeling.

To fill this gap, we propose a generic multi-input and multi-output surrogate modeling method for multistage assembly processes. The method directly links the fixture layout design with the dimensional quality of final product. The generic model structure as used in Kriging is adopted and thus minimal physical derivation is needed. It can also be applied to various assembly process configurations as long as their computer simulation models are available. A recently developed integrated emulation method based on Gaussian process [19], denoted as gaussian process model for qualitative and quantitative factors (GPQQ), is used to capture the relationships among multiple inputs and multiple outputs simultaneously. GPQQ allows the information contained in the training data to be extracted to a larger extend by sharing information across different response levels, which lead to improved prediction accuracy. Several specific issues in the design of experiments are addressed. This approach will be compared with a conventional Kriging approach and validated in a case study.

The proposed model has a mathematical expression explicitly relating product dimensional quality with both design parameters (fixture layout) and process errors (locator deviations). Besides design optimization, it can also be used for a number of other purposes related with propagation of process variations such as: (1) tolerance synthesis for allocating tolerance to different error sources (e.g., Refs. [7,8]; (2) fault diagnosis to identify root causes (e.g., Ref. [9]); (3) optimal sensor distribution for maximizing system diagnosability at minimal cost (e.g., Ref. [20]).

The remainder of this paper is organized as follows. In Sec. 2, we will give a detailed description of the problem. Section 3 describes the surrogate modeling approach for rigid body assembly processes. The modeling for a multistage assembly system is presented in Sec. 4. Section 5 shows a case study to validate the proposed method. Section 6 concludes the paper.

2 Problem Formulation

A typical multistage assembly process is illustrated in Fig. 1. To limit the scope, only rigid body assembly processes are considered in this paper. However, the methods developed could be applied to compliant assembly cases as well through simple extensions, as long as a computer simulation is available. In this figure, a solid line represents the actual position of a part while a dashed line represents its nominal position. Define a 3×1 vector $L^k_{p,i}$ as the ith locator on the pth part at the kth station, where $i = 1, \ldots, 5$, $p = 1, 2, 3$, and $k = 1, 2$. The sequence of this assembly process is as follows. First, Part 1 is constrained by pins L^1_{11}, L^1_{12}, L^1_{13}, a four-way pin L^1_{14}, and a two-way pin L^1_{15} at Station 1. Then, Part 2 joins Part 1 and is constrained by a mating feature, a four-way pin L^1_{24} and a two-way pin L^1_{25}. The two parts are welded together to form a subassembly and then transferred to Station 2. At Station 2, the subassembly is constrained by pins L^2_{21}, L^2_{22}, L^2_{23}, a four-way pin L^2_{24}, and a two-way pin L^2_{25}. Then, Part 3 joins the subassembly and is constrained by a mating feature, a four-way pin L^2_{34}, and a two-way pin L^2_{35}. Finally, Part 3 and the subassembly is welded together to form the final assembly.

To establish a model linking design parameters with product dimensional quality, we need two sets of coordinates: a fixed world coordinate system, WCS and a local coordinate system, LCS that is attached to each part. On each part, a reference point is chosen as the origin of its LCS. In this way, the spatial deviation of a part in the WCS with respect to its nominal location, which is a direct measure of the dimensional quality of the final assembly, can then be represented by the deviation of its LCS relative to WCS. Under the assumption of frictionless point contact of locators to the workpiece and following [21], we consider a generic locating scheme for a 3D rigid body with six locators. Let $\boldsymbol{l}_i$ be the position vector of ith locator in WCS, then $\mathbf{L} = [\boldsymbol{l}_1^T, \boldsymbol{l}_2^T, \ldots, \boldsymbol{l}_6^T]^T$ is an 18×1 vector containing the information of the fixture layout. Define the spatial location of the workpiece as a 6×1 vector $\mathbf{q}_0$ in WCS, which contains three translational and three rotational parameters of the workpiece in the three-dimensional space. The detailed definition of $\mathbf{q}_0$ can be found in Ref. [21]. Then, spatial deviation of the workpiece, $\delta\mathbf{q}_0$, is a nonlinear function of $\mathbf{L}$ and locator deviations $\delta\mathbf{L}$

$$\delta\mathbf{q}_0 = \boldsymbol{h}(\mathbf{L}, \delta\mathbf{L}) \tag{1}$$

The workpiece deviation is linked with fixture layout and locator deviations by the above function describing the part-fixture interaction. Based on Eq. (1), a surrogate model can be built to emulate function $\boldsymbol{h}(\cdot,\cdot)$ using input $\mathbf{L}, \delta\mathbf{L}$, and output $\delta\mathbf{q}_0$ from computer simulations. Equation (1) implies that the product dimensional quality $\delta\mathbf{q}_0$ is only influenced by the nominal locations and the deviations of the fixture locators. In reality, the geometric accuracy of the part itself will also influence the final dimensional quality. However, it is known that in most assembly processes, the part inaccuracy can be viewed as equivalent to certain fixture locator deviation [5]. For this reason, we will only consider $\mathbf{L}$ and $\delta\mathbf{L}$ as the independent variables in Eq. (1).

Under the usual assumption that the locator deviations $\delta\mathbf{L}$ are much smaller compared to size of the workpiece, Eq. (1) can be simplified to the following linear form [21]

$$\delta\mathbf{q}_0 = \tilde{\boldsymbol{h}}(\mathbf{L})\delta\mathbf{L} = -\mathbf{J}^{-1}\boldsymbol{\Phi}_{\mathbf{L}}\delta\mathbf{L} \tag{2}$$

where $\tilde{\boldsymbol{h}}(\mathbf{L})$ is a 6×18 matrix whose entries are complicated nonlinear functions of $\mathbf{L}$ and $\tilde{\boldsymbol{h}}(\mathbf{L})$ can be further decomposed as $-\mathbf{J}^{-1}\boldsymbol{\Phi}_{\mathbf{L}}$. The matrix $\boldsymbol{\Phi}_{\mathbf{L}}$ contains outgoing norms of all the surface points where locators make contact with the workpiece and $\mathbf{J}$ is known as the Jacobian matrix, which contains information of the fixture layout under nominal condition and characterizes the kinematics of the workpiece-fixture system. When the workpiece is deterministically located, the Jacobian matrix is nonsingular [21]. The matrix $\boldsymbol{\Phi}_{\mathbf{L}}$ acts as a projection matrix that converts the positional error of a locator to the translational error along the contact point's normal direction.

Comparing with Eq. (1), the special structure in Eq. (2) simplifies the surrogate modeling task significantly because we can build a surrogate model for function $\tilde{\boldsymbol{h}}(\cdot)$ depending on $\mathbf{L}$ only; then the surrogate model for $\delta\mathbf{q}_0$ can be obtained by simply multiplying $\tilde{\boldsymbol{h}}(\mathbf{L})$ with $\delta\mathbf{L}$. Furthermore, we know $\tilde{\boldsymbol{h}}(\mathbf{L}) = -\mathbf{J}^{-1}\boldsymbol{\Phi}_{\mathbf{L}}$ and $\boldsymbol{\Phi}_{\mathbf{L}}$ only depends on the norm directions of the contacting points between the part and locators, rather than specific locations of the locators. In other words, if the locators fall on the same planes in two design configurations, then the $\boldsymbol{\Phi}_{\mathbf{L}}$ matrices of these two design configurations are the same even the positions of these locators are different within the planes. Thus, in most practical design scenarios, we can treat $\boldsymbol{\Phi}_{\mathbf{L}}$ as constant and we need to focus only on the $\mathbf{J}$ matrix in the surrogate model building. Based on computer simulations, the surrogate modeling procedure for a single part using Eq. (2) can be briefly described as follows: (i) Generate carefully designed inputs $\mathbf{L}$ and $\delta\mathbf{L}$; (ii) Run computer simulations to obtain output, which is the workpiece deviation $\delta\mathbf{q}_0$; (iii) Calculate entry values of $\mathbf{J}^{-1}$ through Eq. (2); (iv) build surrogate models for $\mathbf{J}^{-1}$. The above modeling procedure can be illustrated using Fig. 2.

Once surrogate models for individual parts have been built using the above procedure, we can link all these submodels together to obtain the system level surrogate model through analysis of part–part interactions and station–station interactions. There are several challenges in building surrogate models for variation propagation in multistage assembly systems. First of all, the number of fixture locators in an assembly system can be very large. For example, a typical car body assembly can consist of 200–250 sheet metal parts assembled in 60–100 assembly stations with 1700–2100 locators [1,22]. Simultaneous modeling of all design parameters is not feasible and hence a divide-and-conquer procedure must be used to keep the problem at a manageable level. Second, the surrogate modeling of Eq. (2) is a problem with multiple responses (all entries in $\mathbf{J}^{-1}$ need to be modeled), unlike the conventional surrogate modeling problems which are usually univariate. Third, a carefully designed experiment needs to be developed to take advantage of the linear structure in Eq. (2) when modeling the multi-input-multi-output system.

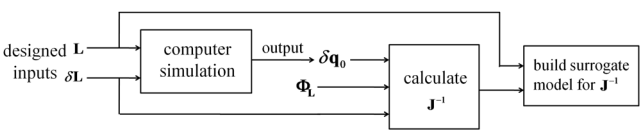

Fig. 2 Simulation based surrogate modeling procedure

In Secs. 4 and 5, we will show how the proposed surrogate modeling procedure can overcome these difficulties and how to link submodels together.

3 Surrogate Modeling for a Single Stage Assembly Process

The locating problem of a rigid body workpiece serves as the foundation of modeling a multistage assembly process. In this section, we will develop a surrogate modeling approach for modeling a single workpiece-fixture system.

3.1 Kriging Modeling for a Single Workpiece-Fixture System. As mentioned earlier, the Jacobian matrix is a function of locator nominal locations $\mathbf{L}$. Hence, the inverse Jacobian matrix also depends on $\mathbf{L}$, i.e., each entry in the 6×6 $\mathbf{J}^{-1}$ matrix is a scalar function of $\mathbf{L}$. Define $\delta\mathbf{U} = \boldsymbol{\Phi}_{\mathbf{L}}\delta\mathbf{L}$, then $\delta\mathbf{U}$ is a 6×1 vector containing deviations of locating points along the normal directions of their corresponding contact points. To take advantage of the linear relationship $\delta\mathbf{q}_0 = -\mathbf{J}^{-1}\delta\mathbf{U}$ in the computer simulation, for each designed fixture layout, we let $\delta\mathbf{U} = \tau\boldsymbol{e}_i$ ($i=1$, 2, …, 6), where τ is a small scalar representing the magnitude of the locator deviation along contact point normal direction, $\boldsymbol{e}_i$ is a 6×1 vector with ith element being 1 and all others being 0. This allows us to obtain the ith column of $\mathbf{J}^{-1}$, and hence the entire $\mathbf{J}^{-1}$ matrix by varying $\boldsymbol{e}_i$. Given a fixture layout $\mathbf{L}$, we can obtain its $\mathbf{J}^{-1}$ matrix using this method. If we generate an n-run design through Latin hypercube sampling (LHS, [23]) for all design parameters in $\mathbf{L}$, denoted as $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_n$, then their corresponding $\mathbf{J}^{-1}$ matrices $\mathbf{J}_1^{-1}, \mathbf{J}_2^{-1}, \ldots, \mathbf{J}_n^{-1}$ can be obtained through computer simulation. Let the ith row jth column entry in $\mathbf{J}^{-1}$ be denoted as $g_{ij}(\mathbf{L})$, which is a scalar function of $\mathbf{L}$ and $i, j = 1, 2, \ldots, 6$. Then each $g_{ij}(\mathbf{L})$ can be modeled through Kriging based on the N design runs, hence building a surrogate model for $\mathbf{J}^{-1} = \{g_{ij}(\mathbf{L})\}$.

Kriging models the response as a regression part for global trend plus a Gaussian random process for local perturbation [18], i.e.,

$$g(\mathbf{L}) = \boldsymbol{f}^T(\mathbf{L})\boldsymbol{\beta} + \varepsilon(\mathbf{L}) \tag{3}$$

where $\boldsymbol{f}(\mathbf{L})$ is a set of prespecified regression functions, $\boldsymbol{\beta}$ is a vector containing corresponding regression parameters, and the residual $\varepsilon(\mathbf{L})$ is assumed to follow a stationary Gaussian random process with zero mean. A correlation structure is defined for the Gaussian random process that assumes input points nearer to each other have higher correlations. Parameters in the model are often estimated using maximum likelihood method. Kriging has been implemented in various readily available packages such as the MATLAB toolbox DACE [24].

It needs to be pointed out that although $\mathbf{L}$ has 18 free parameters, the actual number is usually smaller in practice because some of the less interested parameters may be predetermined and some others may be inherently invariant in the system configuration.

3.2 Integrated Emulation for a Single Workpiece-Fixture System. The surrogate modeling procedure described in Sec. 3.1 is straightforward but not favorable in this application. First of all, the method is very messy because a separate Kriging model has to be built for each entry in $\mathbf{J}^{-1}$, thus resulting in a large number of fitted models. Second, individual fitting of each entry in $\mathbf{J}^{-1}$ ignores the similarities among them, which in fact can be used to improve prediction performance. Therefore, a better way is to use

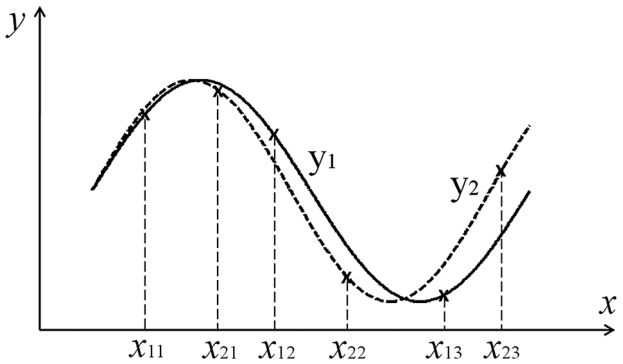

Fig. 3 Illustration of integrated emulation approach

an integrated emulation technique for building a surrogate model of $\mathbf{J}^{-1}$.

For a multivariate system, instead of modeling each response separately, a multivariate emulation technique can model all the responses from different levels as a combined dataset. By incorporating the cross-correlations between different response levels into the model, the multivariate emulation approach can improve the prediction accuracy by borrowing information from other levels.

The idea of integrated emulation can be demonstrated through the simple example in Fig. 3. In this figure, y_1 and y_2 are two unknown functions that bear some similarity and defined in the same design space x. In order to emulate the functions, we evaluate the two functions at training sites x_{11}, x_{12}, x_{13} and x_{21}, x_{22}, x_{23}, respectively. Following a conventional surrogate modeling procedure, we can build a surrogate model for each function based on its responses at the three design points. However, if we consider the similarity between the two functions, modeling of y_1 can use information not only from its own responses at x_{11}, x_{12}, x_{13}, but also from responses of y_2 at x_{21}, x_{22}, x_{23} with some penalty depending on their degree of similarity. In this way, we borrow information from y_2 to help construct a better model for y_1. Similarly, information can be borrowed from y_1 when modeling y_2. This integrated approach essentially increases the number of observations for a function without actually requiring more observations and hence improve the model prediction accuracy. Higher similarity between the two functions will give the integrated model better prediction accuracy.

3.2.1 Brief Introduction to the GPQQ Method. As introduced earlier, the GPQQ method will be used for surrogate modeling due to its high efficiency and ease of implementation. Before using this method to emulate $\mathbf{J}^{-1}$ in Eq. (2), it is necessary to briefly introduce it first.

In Ref. [19], the method is presented as an integrated emulation method for computer experiments with both qualitative and quantitative factors. Modeling a multivariate response system can be treated as a special case for the GPQQ method if we add an artificial qualitative factor to indicate different response levels (referred to as *categories* in the paper). For example, to emulate the two functions in Fig. 3 using an integrated approach, apart from the quantitative input x, we add a qualitative factor with two levels, each corresponding to a different function.

Assume there are m response levels (m functions defined over the same input space) and let $\boldsymbol{w} = (\mathbf{L}^T, c)^T$ $(c = 1, \ldots, m)$ denote the input at the cth level with fixture layout $\mathbf{L}$. Similar to that of Kriging in Eq. (3), the scalar response y of the simulation model at an input value $\boldsymbol{w}$ is modeled as

$$y(\boldsymbol{w}) = \boldsymbol{f}^T(\boldsymbol{w})\boldsymbol{\beta} + \varepsilon(\boldsymbol{w}) \tag{4}$$

where $\boldsymbol{f}(\boldsymbol{w}) = (f_1(\boldsymbol{w}), \ldots, f_p(\boldsymbol{w}))^T$ is a set of p user-specified regression functions, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a vector of unknown coefficients and the residual $\varepsilon(\boldsymbol{w})$ is assumed to be a stationary Gaussian process with mean 0 and variance σ^2. For any two input values $\boldsymbol{w}_j = (\mathbf{L}_j^T, c_j)^T (j = 1, 2; c_j = 1, \ldots, m)$, the correlation between $\varepsilon(\boldsymbol{w}_1)$ and $\varepsilon(\boldsymbol{w}_2)$ is defined to be

$$\begin{aligned} \mathrm{cor}(\varepsilon(\boldsymbol{w}_1), \varepsilon(\boldsymbol{w}_2)) &= \tau_{c1,c2} K(\mathbf{L}_1, \mathbf{L}_2) \\ &= \tau_{c1,c2} \exp\left\{ -\sum_{i=1}^{I} \phi_i (l_{1i} - l_{2i})^2 \right\} \end{aligned} \tag{5}$$

where $\tau_{c1,c2} (|\tau_{c1,c2}| \le 1)$ is the cross-correlation between two response levels c_1 and c_2, $K(\mathbf{L}_1, \mathbf{L}_2)$ is the spatial correlation between two fixture layout designs, and ϕ_i are roughness parameters. Essentially, $\tau_{c1,c2}$ serves as a penalty term for borrowing training samples from c_2 into the modeling of c_1, and vice versa. As an extreme case, if response levels c_1 and c_2 are exactly the same (i.e., they always produce the same value with the same input $\mathbf{L}$), we have $\tau_{c1,c2} = 1$. Define an $m \times m$ matrix $\mathbf{T}$ whose rth row and sth column entry is $\tau_{r,s}$, $r, s = 1, 2, \ldots, m$. Then $\mathbf{T}$ must be a positive definite matrix with unit diagonal elements so that Eq. (5) can be a valid correlation function. To guarantee this property, a parametric construction of $\mathbf{T}$ has been used in the GPQQ method. Details of the parametric construction can be found in Ref. [19]. Note that the correlation function defined in Eq. (5) implies response surfaces at different levels exhibit similar behavior in terms of spatial correlation. While there are some other possible ways to define the correlation function, the proposed form is simple and effective, with clear physical meanings of the parameters involved; its effectiveness has been demonstrated in the literature [19,25].

Specifically, the artificial qualitative factor c is the index for entries in $\mathbf{J}^{-1}$, hence $m = 36$ in this case. By using the GPQQ method, entries in the $\mathbf{J}^{-1}$ matrix can be collectively modeled together and this model is expected to provide better prediction than separately modeling each entry as described in Sec. 3.1.

3.2.2 Experimental Design and Model Fitting. Since the GPQQ method improves its prediction accuracy of one entry in $\mathbf{J}^{-1}$ by borrowing strength from other entries with different input $\mathbf{L}$, the design sites for different entries need to be different so that we can get maximum benefit from this method. To understand this, let us consider the design where all entries in $\mathbf{J}^{-1}$ are evaluated at a common set of $\mathbf{L}$ inputs. In this case, the design sites for the ith entry would be $(\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{in}) = ((\mathbf{L}_1, i), \ldots, (\mathbf{L}_n, i))$, $i = 1, 2, \ldots, 36$. Note that the n designs of $\mathbf{L}$ are the same for all 36 entries and there are totally $36n$ design sites yielding $36n$ responses. For the simple example illustrated in Fig. 3, such a design implies $x_{11} = x_{21}$, $x_{12} = x_{22}$, and $x_{13} = x_{23}$. Under this design structure, the correlation matrix $\mathbf{R}$ (whose ith row jth column entry is the correlation between ith and jth inputs as defined in Eq. (5)) in the GPQQ model is the Kronecker product (denoted by $\otimes$) of the cross-correlation matrix and spatial correlation matrix

$$\mathbf{R} = \mathbf{T} \otimes \mathbf{H} \tag{6}$$

where $\mathbf{H}$ is the $n \times n$ spatial correlation matrix whose (j_1, j_2)th entry is $K(\mathbf{L}_{j1}, \mathbf{L}_{j2})$ and $\mathbf{T}$ is the $m \times m$ cross-correlation matrix whose (r, s)th entry is $\tau_{r,s}$. As these three matrices are all positive definite, from the property of Kronecker product, we have

$$\mathbf{R}^{-1} = \mathrm{T}^{-1} \otimes \mathbf{H}^{-1} \tag{7}$$

For the GPQQ model, the empirical best linear unbiased predictor (EBLUP) of y for any untried input value $\boldsymbol{w}_0$ is

$$\hat{y}(\boldsymbol{w}_0) = \boldsymbol{f}^T(\boldsymbol{w}_0)\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{r}}_0^T \hat{\mathbf{R}}^{-1}(y - \mathbf{F}\hat{\boldsymbol{\beta}}) \tag{8}$$

where $\hat{\boldsymbol{r}}_0 = (\mathrm{cor}(\varepsilon(\boldsymbol{w}_0), \varepsilon(\boldsymbol{w}_{11})), \ldots, \mathrm{cor}(\varepsilon(\boldsymbol{w}_0), \varepsilon(\boldsymbol{w}_{mn})))^T$, $\mathbf{F} = (f(\boldsymbol{w}_{11}), \ldots, f(\boldsymbol{w}_{mn}))^T$, $\boldsymbol{y}$ is the $mn \times 1$ vector containing responses

at all training sites, and the hat symbol means they are estimated values from maximum likelihood estimation. Let us assume we want to predict $\boldsymbol{w}_0=(\mathbf{L}_0, i)$, which corresponds to the ith entry with untried fixture layout $\mathbf{L}_0$. Because of the special design structure, we have

$$\hat{\boldsymbol{r}}_0 = (\mathrm{cor}(\varepsilon(\boldsymbol{w}_0), \varepsilon(\boldsymbol{w}_{11})), \ldots, cor(\varepsilon(\boldsymbol{w}_0), \varepsilon(\boldsymbol{w}_{mn})))^T = \hat{\mathbf{T}}_i \otimes \hat{\boldsymbol{\gamma}}_0 \tag{9}$$

where $\hat{\mathbf{T}}_i$ is the ith column of $\hat{\mathbf{T}}$ and $\hat{\boldsymbol{\gamma}}_0 = (\hat{K}(\mathbf{L}_0, \mathbf{L}_1), \ldots, \hat{K}(\mathbf{L}_0, \mathbf{L}_n))^T$. Hence, we have

$$\begin{aligned}\hat{\boldsymbol{r}}_0^T\hat{\mathbf{R}}^{-1} &= (\hat{\mathbf{T}}_i \otimes \hat{\boldsymbol{\gamma}}_0)^T \cdot (\hat{\mathbf{T}}^{-1} \otimes \hat{\mathbf{H}}^{-1}) = (\hat{\mathbf{T}}_i^T \otimes \hat{\boldsymbol{\gamma}}_0^T) \cdot (\hat{\mathbf{T}}^{-1} \otimes \hat{\mathbf{H}}^{-1}) \\ &= (\hat{\mathbf{T}}_i^T\hat{\mathbf{T}}^{-1}) \otimes (\hat{\boldsymbol{\gamma}}_0^T\hat{\mathbf{H}}^{-1}) = \boldsymbol{e}_i^T \otimes (\hat{\boldsymbol{\gamma}}_0^T\hat{\mathbf{H}}^{-1}) \\ &= (\underbrace{\mathbf{O}^T, \ldots, \mathbf{O}^T}_{i-1}, \hat{\boldsymbol{\gamma}}_0^T\hat{\mathbf{H}}^{-1}, \mathbf{O}^T, \ldots, \mathbf{O}^T)\end{aligned} \tag{10}$$

where $\boldsymbol{e}_i$ is an $m \times 1$ vector whose ith entry is 1 and all other entries 0, and $\mathbf{O}$ is a $n \times 1$ vector with all 0 entries. If we plug the result of Eq. (10) into predictor Eq. (8), it is clear that only the training samples from the ith entry are actually used in the prediction of $y(\boldsymbol{w}_0)$, i.e., the part of $\underline{y}$ that corresponds to the nonzero part in the $1 \times mn$ vector $\hat{\boldsymbol{r}}_0^T\hat{\mathbf{R}}^{-1}$. Therefore, the prediction of $\boldsymbol{w}_0$ from the ith entry does not borrow strength from training samples of other entries. Based on this result, the GPQQ method loses its major advantage under this specific design structure and it should therefore be avoided. However, it needs to point out that this result does not mean the integrated emulation approach will produce exactly the same prediction values as Kriging since they will have different estimated values of parameters. Whether the integrated approach can provide any benefit over the univariate approach under such design is a complicated issue that extends beyond the scope of this paper. In general, to take full advantage of the GPQQ method, different entries in $\mathbf{J}^{-1}$ should be evaluated with different input $\mathbf{L}$.

Another issue in the experimental design is that the workpiece can be very sensitive to locator errors under some extreme fixture layout designs: it can have very large deviations even with small locator errors. This can be illustrated using a simple example shown in Fig. 4. A block is deterministically located by six locators as shown in the figure. The fixture layout on the right is different from that on the left by only the nominal location of locator L_3. Assume L_3 in both layouts have a deviation of the same magnitude ΔL along the z direction. The resulting deviation of workpiece on the right will be much larger than that on the left because the nominal positions of L_1, L_2, and L_3 on the right are almost collinear.

Based on the relationship in Eq. (2), such fixture layout designs can lead to extremely large entries in $\mathbf{J}^{-1}$ and cause formidable difficulty for the GPQQ method. A simple remedy is to remove them by setting limit on the magnitude of entries in $\mathbf{J}^{-1}$. Based on Eq. (2), any large entry in $\mathbf{J}^{-1}$ indicates high sensitivity of the workpiece to certain locator error. As the first three rows and last three rows in $\mathbf{J}^{-1}$ correspond to different types of errors with distinct units in $\delta\mathbf{q}_0$ (translational and rotational errors), two different limits may be set for them. In practice, they generally correspond to physically meaningless designs that will be prescreened at the experimental design stage and hence will not enter the optimization cycle. Some simple *ad hoc* prescreening rules can be found in Refs. [14,26]. An alternative, yet more complex, method to build surrogate models for response surface with nonstationarity is using a treed model, i.e., design space is partitioned into smaller regions and local models are built (e.g., Ref. [27]).

For the reasons stated above, we suggest the experimental design proceed as follows:

(1) Using LHS, generate n fixture layout designs for $\mathbf{L}$.
(2) For each fixture layout design, set $\delta\mathbf{U} = \tau\boldsymbol{e}_1$ as the locator deviation and run computer simulation to obtain workpiece deviation $\delta\mathbf{q}_0$. Obtain the first column entries of $\mathbf{J}^{-1}$ as $-\delta\mathbf{q}_0/\tau$. Check if any design has entries in $-\delta\mathbf{q}_0/\tau$ larger than the set limits. If none, go to step 4); remove them if there are (assume n' designs are removed, $n' \ll n$).
(3) Generate n' new designs based on LHS and check again following step 2). If none is removed, proceed to step 4); if there are still undesirable designs in the newly generated designs, repeat the above process until all n designs obtained contain no large entries.
(4) Repeat steps (1)–(3) to obtain the remaining five columns in $\mathbf{J}^{-1}$ by setting $\boldsymbol{e}_i$ for the ith column, a different set of n training sites should be generated each time.

The above design procedure ensures that the six entries in each row of $\mathbf{J}^{-1}$ are evaluated with different fixture layouts, and hence we can expect improved prediction accuracy from the integrated emulation approach. It should be noted that the entries in the same column are evaluated at the same n sites. This is due to the fact that under a given fixture layout, by setting $\delta\mathbf{U} = \tau\boldsymbol{e}_i$, we always obtain the entire ith column through $(\mathbf{J}^{-1})_i = -\delta\mathbf{q}_0/\tau$. Hence, by the results shown in Eq. (10), it is suggested to model each row of $\mathbf{J}^{-1}$ separately using GPQQ instead of modeling all the 36 entries together

$$\delta\mathbf{q}_0 = -\mathbf{J}^{-1}\delta\mathbf{U} = -[\mathbf{G}_1 \quad \mathbf{G}_2 \quad \mathbf{G}_3 \quad \mathbf{G}_4 \quad \mathbf{G}_5 \quad \mathbf{G}_6]^T\delta\mathbf{U} \tag{11}$$

where $\mathbf{G}_i$ is a GPQQ surrogate model for the ith row of $\mathbf{J}^{-1}$ and dependent upon locator nominal locations. Hence, the number of entries modeled by each GPQQ model is six. In this way, we can still take advantage of the GPQQ method as a better integrated predictor against separately fitted Kriging models for individual entries in the matrix. Also, comparing with modeling 36 entries altogether, modeling each row separately can significantly reduce the total number of observations used to fit the GPQQ model. This

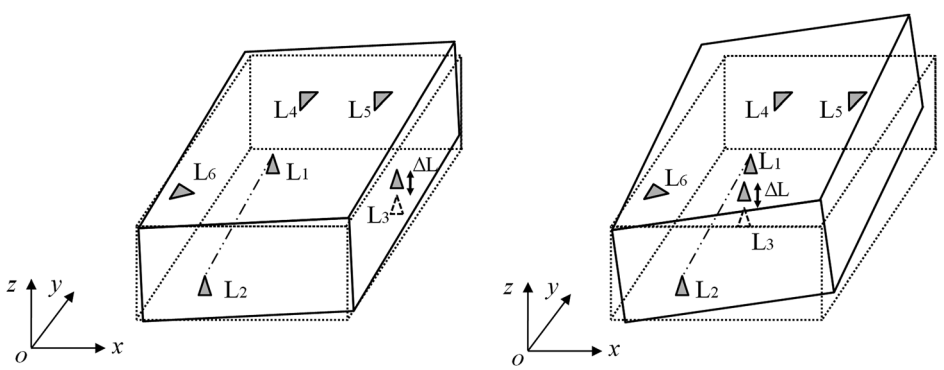

Fig. 4 Illustration of undesirable fixture layout with near singular Jacobian matrix

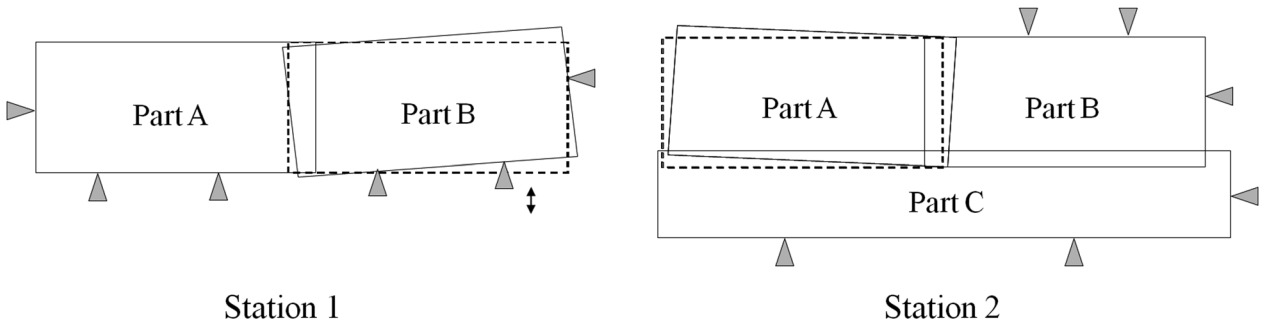

Fig. 5 Illustration of reorientation error

helps maintain numerical stability of the method because matrix inversion of correlation matrix **R** is involved in the model fitting.

Note that the method used for generating design points, LHS, can handle only rectangular design regions. In practice, the design region may be irregular such as door panel frames. Constructing space-filling designs for computer experiments in irregular regions is often a difficult task and limited work has been done. Stinstra et al. and Trosset developed algorithms to construct maximin designs in nonrectangular regions [28,29]. Chuang and Huang constructed uniform designs in any convex design region [30]. One simple solution to construct designs in irregular regions is to discretize the region into grids. Another issue associated with experimental design is to choose a suitable number n so that the response surface can be well approximated with reasonable amount of computation. The adequate number of simulation runs for model fitting depends on the complexity of the response surface, which is determined by the configuration of the workpiece-fixture system and hence unique to all systems. We suggest n be determined by cross-validation with a preset error limit such as 5%.

4 Linking Submodels for Multistage Assembly Process

In this section, we will show how to link submodels for individual parts together. Without loss of generality, we assume parts are assembled together in a sequential way at each station. In each operation, a new part will be added to an existing part/subassembly that is completely constrained by its fixture. The new part can either be constrained by fixture or the mating feature from an existing part, or both. The two parts will then be permanently joined together as a new subassembly. This subassembly can be joined by another new part in the similar way. This process can continue until all parts have been assembled at one station. At the end of one station, all the assembled parts have been permanently joined together to form a subassembly. Then, the subassembly is transferred to the next station to be assembled with more coming parts. In this process, two types of errors dominate the part dimensional variation: fixture-induced error and reorientation-induced error. Here, we focus on the deviation of an arbitrary part P at the kth station. If Part P is a newly joined part at the kth station, then we can model its deviation by simply following the procedures introduced in Sec. 3. If it has already joined the assembly at a previous station and transferred to the kth station as a part contained within a subassembly, its accumulated part deviation at the kth station, denoted as $\Delta\mathbf{q}^{\mathrm{P}}(k)$ in WCS, can be modeled through the following relationship [2,6]

$$\Delta\mathbf{q}^{P}(k) = \Delta\mathbf{q}^{P}(k-1) + \mathbf{E}_{\mathrm{F}}^{P}(k) + \mathbf{E}_{\mathrm{R}}^{P}(k) \tag{12}$$

where $\mathbf{E}_{\mathrm{F}}^{P}(k)$ is the fixture-induced error and $\mathbf{E}_{\mathrm{R}}^{P}(k)$ the reorientation-induced error on Part P. $\mathbf{E}_{\mathrm{F}}^{P}(k)$ is caused by errors of the locators constraining the subassembly, following [6], we have

$$\mathbf{E}_{\mathrm{F}}^{P}(k) = \mathbf{Q}_{\mathrm{SA}}^{P}\mathbf{E}_{\mathrm{F}}^{\mathrm{SA}}(k) \tag{13}$$

where $\mathbf{E}_{\mathrm{F}}^{\mathrm{SA}}(k)$ is the fixture-induced error of the subassembly and $\mathbf{Q}_{\mathrm{SA}}^{P}$ is the matrix that transforms the deviation of subassembly LCS to the deviation of LCS attached to Part P. Let (x_o^P, y_o^P, z_o^P) be the origin of Part P LCS expressed in LCS of the subassembly containing Part P, then

$$\mathbf{Q}_{\mathrm{SA}}^{P} = \begin{bmatrix} \mathbf{I}_{3\times3} & \bar{\mathbf{Q}}_{SA}^{P} \\ \mathbf{O}_{3\times3} & \mathbf{I}_{3\times3} \end{bmatrix}, \qquad \bar{\mathbf{Q}}_{\mathrm{SA}}^{P} = \begin{bmatrix} 0 & z_o^P & -y_o^P \\ -z_o^P & 0 & x_o^P \\ y_o^P & -x_o^P & 0 \end{bmatrix} \tag{14}$$

where $\mathbf{I}_{3\times3}$ is a 3×3 identity matrix and $\mathbf{O}_{3\times3}$ is a 3×3 zero matrix. The fixture-induced error on the subassembly can be obtained through Eq. (2)

$$\mathbf{E}_{\mathrm{F}}^{\mathrm{SA}}(k) = -(\mathbf{J}^{\mathrm{SA}}(k))^{-1}\delta\mathbf{U}^{\mathrm{SA}} \tag{15}$$

where $\Delta\mathbf{U}^{\mathrm{SA}}$ is the deviation of subassembly locators along their contacting points' normal directions at the kth station.

$\mathbf{E}_{\mathrm{R}}^{P}(k)$ is caused by errors of the subassembly locating points at the (k-1)th station. The reorientation error occurs when the subassembly locating point errors are reset to zero after transferred to the next station. Figure 5 gives a simple example to demonstrate the concept of reorientation error. At Station 1, a locator error causes a corresponding deviation in Part B and then Part B is welded with Part A to form a subassembly. The subassembly is then transferred to Station 2 and located by another set of fixture. Even there is no locator error at Station 2, Part A still deviates from its nominal position because of the Part B location error at Station 1.

In fact, $\mathbf{E}_{\mathrm{R}}^{P}(k)$ is a special case of fixture-induced error and it can be generalized into $\mathbf{E}_{\mathrm{F}}^{P}(k)$. Define $\Delta\mathbf{D}^{\mathrm{SA}}(k-1)$ as errors of the locating points on the subassembly along normal directions at the (k-1)th station, the reorientation error can be calculated as [6]

$$\begin{aligned}\mathbf{E}_{\mathrm{R}}^{P}(k) &= \mathbf{Q}_{\mathrm{SA}}^{P}\mathbf{E}_{\mathrm{R}}^{\mathrm{SA}}(k) = \mathbf{Q}_{\mathrm{SA}}^{P}(-(\mathbf{J}^{\mathrm{SA}}(k))^{-1})(-\Delta\mathbf{D}) \\ &= \mathbf{Q}_{\mathrm{SA}}^{P}(\mathbf{J}^{\mathrm{SA}}(k))^{-1}\Delta\mathbf{D}^{\mathrm{SA}}(k-1)\end{aligned} \tag{16}$$

Plugging Eqs. (13), (15), and (16) into Eq. (12), we obtain

$$\begin{aligned}\Delta\mathbf{q}^{P}(k) &= \Delta\mathbf{q}^{P}(k-1) + \mathbf{E}_{\mathrm{F}}^{P}(k) + \mathbf{E}_{\mathrm{R}}^{P}(k) \\ &= \Delta\mathbf{q}^{P}(k-1) + \mathbf{Q}_{\mathrm{SA}}^{P}[-(\mathbf{J}^{\mathrm{SA}}(k))^{-1}(\Delta\mathbf{U}^{\mathrm{SA}} - \Delta\mathbf{D}^{\mathrm{SA}}(k-1))]\end{aligned} \tag{17}$$

In Eq. (17), $\Delta\mathbf{D}^{\mathrm{SA}}(k-1)$ can be obtained from the (k-1)th station, and we can use the procedures described in Sec. 3 to build surrogate model for $(\mathbf{J}^{\mathrm{SA}}(k))^{-1}$.

By following the procedures above, we will be able to link all the part-level submodels together and construct an overall surrogate model for the entire assembly process. The chain-like linear relationship in Eq. (12) will finally lead to an overall model structure in the following generic linear form

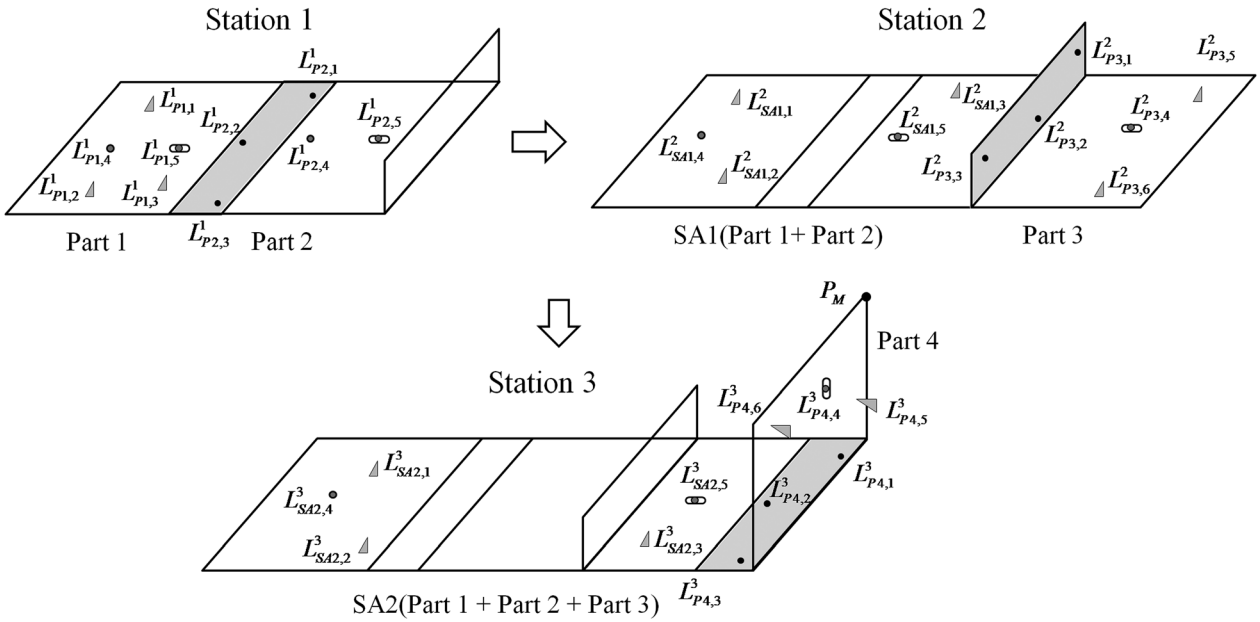

Fig. 6 Case study multistage assembly process

$$\mathbf{y} = \mathbf{\Gamma f} \tag{18}$$

where $\mathbf{y}$ is a vector consisting of part deviations at the final station, $\mathbf{\Gamma}$ is a coefficient matrix dependent upon fixture layouts and other process/product parameters at all stations, $\mathbf{f}$ is a vector containing all locator deviations. It can be seen that, under small error assumption, the final part dimensional errors have nonlinear relationships with fixture layout and linear relationships with fixture locator deviations. Based on Eq. (18), the variance of part errors can be calculated under a given fixture layout as

$$\mathrm{Var}(\mathbf{y}) = \mathbf{\Gamma}\mathrm{Var}(\mathbf{f})\mathbf{\Gamma}^T \tag{19}$$

Equation (19) shows that the error variances can be obtained directly through closed-form expressions rather than Monte Carlo simulation based results from computer simulation, which significantly reduces computational load. As discussed before, the model can be used for design optimization for fixture layout, as well as several other purposes such as tolerance analysis/synthesis, optimal sensor distribution, and fault diagnosis for multistage assembly systems.

5 Case Study

In this section, a case study is used to demonstrate the proposed method. In the case study, the proposed method will be compared with the Kriging approach introduced in Sec. 3.1 based on 3DCS computer simulation results.

5.1 Model Setup. The case study is a four-part, three station assembly process shown in Fig. 6. At the first station, Part 1 and Part 2 are assembled together to form a subassembly SA1. SA1 is then transferred to the second station and joined with Part 3 to form a larger subassembly SA2. Finally, SA2 is transferred to the third station and joined with Part 4 to form the final assembly. The shaded areas in Fig. 6 indicate active mating features during assembly process. A measurement point P_M has been added to Part 3 as it is the furthest down the assembly stream and susceptible to all upstream errors.

The involvement of mating features in this example can be solved by introducing the concept of *virtual locators*—imaginary locators that provide equivalent constraints as the mating feature. In Fig. 6 at Station 1, the matching of the mating features on Part 1 and Part 2 is equivalent to locating the mating plane of Part 2 on three locators ($L^1_{P2,1}$, $L^1_{P2,2}$, $L^1_{P2,3}$) attached to the mating plane of Part 1. These mating locators interact with Part 2 just like physical locators; the only difference is that they cannot move independently but always stay on the plane defined by Part 1. Similar ideas can be applied to Stations 2 and 3. In this way, the constraint provided by mating features can be generalized into equivalent physical locators and modeled accordingly. The deviation of a virtual locator can be calculated based on the positional error of the part it is attached to. Define $\mathbf{P} = [\mathrm{P}_x, \mathrm{P}_y, \mathrm{P}_z]^T$ as coordinates of a point expressed in the part LCS, then $\Delta\mathbf{P}$, the deviation of this point due to the part deviation, $\Delta\mathbf{q}_{\mathrm{part}}$, can be determined as [5]

$$\Delta\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & \mathrm{P}_z & -\mathrm{P}_y \\ 0 & 1 & 0 & -\mathrm{P}_z & 0 & \mathrm{P}_x \\ 0 & 0 & 1 & \mathrm{P}_y & -\mathrm{P}_x & 0 \end{bmatrix} \times \Delta\mathbf{q}_{\mathrm{part}} \tag{20}$$

Throughout the entire assembly process, there are 24 physical locators and 9 virtual locators involved. The design parameters considered in this case study are listed in Table 1 along with their design regions. Coordinates of virtual locators and some other fixed locators that are not considered as design parameters are given in Table 2.

Table 1 Design spaces for locator nominal locations

Parameter	Design region	Parameter	Design region	Parameter	Design region
$\mathrm{L}^1_{\mathrm{P1,1}}(x)$	10–30	$\mathrm{L}^2_{\mathrm{SA1,2}}(x)$	5–40	$\mathrm{L}^3_{\mathrm{SA2,1}}(y)$	25–45
$\mathrm{L}^1_{\mathrm{P1,1}}(y)$	25–45	$\mathrm{L}^2_{\mathrm{SA1,2}}(y)$	5–25	$\mathrm{L}^3_{\mathrm{SA2,2}}(x)$	5–40
$\mathrm{L}^1_{\mathrm{P1,2}}(x)$	5–20	$\mathrm{L}^2_{\mathrm{SA1,3}}(x)$	50–75	$\mathrm{L}^3_{\mathrm{SA2,2}}(y)$	5–25
$\mathrm{L}^1_{\mathrm{P1,2}}(y)$	5–25	$\mathrm{L}^2_{\mathrm{SA1,3}}(y)$	5–45	$\mathrm{L}^3_{\mathrm{SA2,3}}(x)$	90–120
$\mathrm{L}^1_{\mathrm{P1,3}}(x)$	25–40	$\mathrm{L}^2_{\mathrm{P3,4}}(x)$	90–120	$\mathrm{L}^3_{\mathrm{SA2,3}}(y)$	5–45
$\mathrm{L}^1_{\mathrm{P1,3}}(y)$	5–25	$\mathrm{L}^2_{\mathrm{P3,4}}(y)$	25–45	$\mathrm{L}^3_{\mathrm{P4,4}}(y)$	25–45
$\mathrm{L}^1_{\mathrm{P2,4}}(x)$	50–65	$\mathrm{L}^2_{\mathrm{P3,5}}(x)$	90–120	$\mathrm{L}^3_{\mathrm{P4,4}}(z)$	5–15
$\mathrm{L}^1_{\mathrm{P2,4}}(y)$	5–45	$\mathrm{L}^2_{\mathrm{P3,5}}(y)$	5–25	$\mathrm{L}^3_{\mathrm{P4,5}}(y)$	5–25
$\mathrm{L}^1_{\mathrm{P2,5}}(x)$	65–80	$\mathrm{L}^2_{\mathrm{P3,6}}(x)$	90–120	$\mathrm{L}^3_{\mathrm{P4,5}}(z)$	5–15
$\mathrm{L}^2_{\mathrm{SA1,1}}(x)$	5–40	$\mathrm{L}^2_{\mathrm{P3,6}}(y)$	5–45	$\mathrm{L}^3_{\mathrm{P4,6}}(y)$	5–45
$\mathrm{L}^2_{\mathrm{SA1,1}}(y)$	25–45	$\mathrm{L}^3_{\mathrm{SA2,1}}(x)$	5–40	$\mathrm{L}^3_{\mathrm{P4,6}}(z)$	5–15

Table 2 Coordinates for virtual locators and fixed locators

$L^1_{P1,4}$	$L^1_{P1,5}$	$L^1_{P2,1}$	$L^1_{P2,2}$	$L^1_{P2,3}$	$L^2_{SA1,4}$	$L^2_{SA1,5}$	$L^2_{P3,1}$
(5, 25, 0)	(35, 25, 0)	(49, 49, 0)	(41, 25, 0)	(49, 1, 0)	(10, 25, 0)	(60, 25, 0)	(80, 49, 9)
$L^2_{P3,2}$	$L^2_{P3,3}$	$L^3_{SA2,4}$	$L^3_{SA2,5}$	$L^3_{P4,1}$	$L^3_{P4,2}$	$L^3_{P4,3}$	
(80, 25, 1)	(80, 1, 9)	(5, 15, 0)	(105, 15, 0)	(130, 40, 15)	(130, 20, 10)	(130, 30, 5)	

The assembly process is modeled through the proposed method based on all the design parameters listed in Table 1. For each part/subassembly where a surrogate model is built, we follow procedures described in Sec. 3.2.2 with $n = 100$.

5.2 Model Derivation. In this section, the proposed method is applied to the assembly process example to build a surrogate model linking product dimensional quality with locator nominal locations and deviations.

At Station 1, two submodels are built separately for Part 1 and Part 2. Following the notations in Sec. 4, the deviations of Part 1 and Part 2 at Station 1 can be denoted as $\Delta\mathbf{q}^{P1}(1)$ and $\Delta\mathbf{q}^{P2}(1)$. Then, a submodel is built for Part 1 to link $\Delta\mathbf{q}^{P1}(1)$ with locator positions $L^1_{P1,1}(x)$, $L^1_{P1,1}(y)$, $L^1_{P1,2}(x)$, $L^1_{P1,2}(y)$, $L^1_{P1,3}(x)$, $L^1_{P1,3}(y)$, and locator deviations. Since Part 1 is exclusively constrained by physical locators, the surrogate model can be built following the procedures described in Sec. 3. Also, another submodel is built for Part 2 to link $\Delta\mathbf{q}^{P2}(1)$ with locator positions $L^1_{P2,4}(x)$, $L^1_{P2,4}(y)$, $L^1_{P2,5}(x)$, $L^1_{P2,5}(y)$, and locator deviations. Its mating plane constraint can be replaced by the virtual locators $(L^1_{P2,1}, L^1_{P2,2}, L^1_{P2,3})$. At the end of Station 1, Part 1 is permanently joined with Part 2 to form subassembly SA1, which is considered as one rigid body in later stations.

At Station 2, three submodels are built separately for Part 1, Part 2, and Part 3. When transferred from Station 1 to Station 2, the deviation of subassembly SA1 (hence Part 1 and Part 2) is caused by two sources: reorientation error and locator error. A submodel that links deviation of Part 1 $\Delta\mathbf{q}^{P1}(2)$ with locator positions $L^2_{SA1,1}(x)$, $L^2_{SA1,1}(y)$, $L^2_{SA1,2}(x)$, $L^2_{SA1,2}(y)$, $L^2_{SA1,3}(x)$, $L^2_{SA1,3}(y)$, locator deviations, and $\Delta\mathbf{q}^{P1}(1)$ can be built based on the method described in Sec. 4. Similarly, a submodel is built for Part 2 to link $\Delta\mathbf{q}^{P2}(2)$ with the same locator positions, deviations and $\Delta\mathbf{q}^{P2}(1)$. The submodel building steps for Part 3 is similar as those for Part 2 at Station 1 since they two are both newly joined parts constrained partly by mating features. This submodel links $\Delta\mathbf{q}^{P3}(2)$ with locator positions $L^2_{P3,4}(x)$, $L^2_{P3,4}(y)$, $L^2_{P3,5}(x)$, $L^2_{P3,5}(y)$, $L^2_{P3,6}(x)$, $L^2_{P3,6}(y)$, and locator deviations, where three of the locator deviations are from virtual locators. At the end of Station 2, SA1 and Part 3 are permanently joined to form subassembly SA2, which is to be transferred to Station 3.

At Station 3, four submodels are built separately for Part 1, Part 2, Part 3, and Part 4. Similarly, after transferred to Station 3, SA2 (hence Part 1–3) has both reorientation error and locator error. A submodel linking $\Delta\mathbf{q}^{P1}(3)$ with locator positions $L^3_{SA2,1}(x)$, $L^3_{SA2,1}(y)$, $L^3_{SA2,2}(x)$, $L^3_{SA2,2}(y)$, $L^3_{SA2,3}(x)$, $L^3_{SA2,3}(y)$, locator deviations, and $\Delta\mathbf{q}^{P1}(2)$ is built based on method described in Sec. 4. In a similar way, submodels for $\Delta\mathbf{q}^{P2}(3)$ and $\Delta\mathbf{q}^{P3}(3)$ are also built to link them with the same locator positions, deviations and their deviations at Station 2.The final submodel is for Part 4 to link $\Delta\mathbf{q}^{P4}(3)$ with locator positions $L^3_{P4,4}(y)$, $L^3_{P4,4}(z)$, $L^3_{P4,5}(y)$, $L^3_{P4,5}(z)$, $L^3_{P4,6}(y)$, $L^3_{P4,6}(z)$, and locator deviations, where three of the locators are virtual locators. The dimensional error of SA2 will propagate to Part 4 through the mating feature. Once we have obtained the submodel of $\Delta\mathbf{q}^{P4}(3)$, positional error of the measurement point P_M on Part 4 can be calculated using Eq. (20).

The modeling process and error flow are demonstrated in Fig. 7. Since a submodel is required for each part at each station to provide prediction for its deviation (denoted by boxes with solid lines), it can be observed from the figure that there are nine such submodels in the entire assembly process. However, the actual surrogate modeling procedure described in Sec. 3 only needs to be performed six times because a subassembly (denoted by boxes with dotted lines) is a single rigid workpiece, and submodels for the parts within the subassembly are derived from the same subassembly surrogate model. For example, at Station 2, Part 1 is included within subassembly SA1 and its model is given by

$$\Delta\mathbf{q}^{P1}(2) = \Delta\mathbf{q}^{P1}(1) + \mathbf{Q}^{P1}_{SA1}[-(\mathbf{J}^{SA1}(2))^{-1}(\Delta\mathbf{U}^{SA1} - \Delta\mathbf{D}^{SA1}(1))] \tag{21}$$

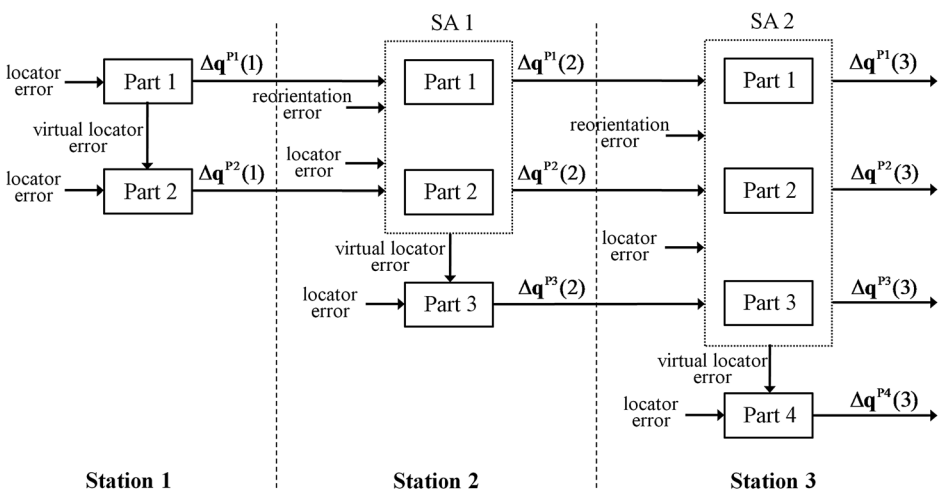

Fig. 7 Modeling structure and error flow of the assembly process

In Eq. (21), $\Delta\mathbf{q}^{P1}(1)$ is a 6×1 vector given the spatial deviation of Part 1 at Station 1 and can be obtained by Part 1's model at Station 1; $\mathbf{Q}_{SA1}^{P1}$ is a 6×6 matrix that transforms the deviation of SA1 to the deviation of Part 1; $\Delta\mathbf{U}^{SA1}$ contains locator deviations along normal directions on subassembly SA1; $\Delta\mathbf{D}^{SA}(k-1)$ contain errors of the locating points on the subassembly SA1 along normal directions at Station 1; and finally, $(\mathbf{J}^{SA1}(2))^{-1}$ is the inverse of the 6×6 Jacobian matrix that describes the kinematic relationship between subassembly SA1 and its locators. In our model, the matrix $(\mathbf{J}^{SA1}(2))^{-1}$ is expressed as

$$(\mathbf{J}^{SA1}(2))^{-1}=\left[\mathbf{G}_1^{SA1}(2)\quad\ldots\quad\mathbf{G}_6^{SA1}(2)\right]^T \tag{22}$$

where the 6×1 vector $\mathbf{G}_i^{SA1}(2)$ is a surrogate model for the ith row of $(\mathbf{J}^{SA1}(2))^{-1}$. Given fixture layout $\mathbf{L}_0$ for subassembly SA1, the jth $(j=1,\ldots,6)$ element of $\mathbf{G}_i^{SA1}(2)$ is given by the GPQQ predictor Eq. (8) with input $\boldsymbol{w}_0=(\mathbf{L}_0, j)$. Models for Part 2 at Station 2 and Parts 1–3 at Station 3 have similar forms as above.

The model is much simpler for a new part entering the assembly stream, for example, deviation of Part 3 at Station 2 is modeled by

$$\Delta\mathbf{q}^{P3}(2)=-(\mathbf{J}^{P3}(2))^{-1}\delta\mathbf{U}^{P3} \tag{23}$$

Again, models for Part 1 and 2 at Station 1 and Part 4 at Station 3 have similar forms as (23).

Based on the figure, it is also clear that the deviation of Part 4, $\Delta\mathbf{q}^{P4}(3)$, is influenced by all the upstream errors through virtual locator error from SA2. The impact of all those errors on $\Delta\mathbf{q}^{P4}(3)$ is hence captured by the model.

5.2 Model Validation. To validate the model, another 50 fixture layout designs are generated using LHS and the locator deviations are set to ± 0.1. For each of the newly generated designs, the deviation of a measuring point P_M on Part 4 is predicted using the surrogate model. Following the suggestion in Ref. [31], the departure of the estimated values from the computer simulated values is assessed using the following criteria:

$$R^2=1-\frac{[\boldsymbol{y}-\hat{\boldsymbol{y}}]^T[\boldsymbol{y}-\hat{\boldsymbol{y}}]}{[\boldsymbol{y}-\bar{y}\cdot\mathbf{1}]^T[\boldsymbol{y}-\bar{y}\cdot\mathbf{1}]} \tag{24}$$

where $\boldsymbol{y}$ is the 150×1 vector containing coordinate deviations of P_M for the 50 new designs obtained through computer simulation, $\bar{y}$ is the average value of $\boldsymbol{y}$, $\hat{\boldsymbol{y}}$ is the model predicted values of $\boldsymbol{y}$, and $\mathbf{1}$ is a 150×1 vector with all 1 s. If $\hat{\boldsymbol{y}}$ is close to $\boldsymbol{y}$, R^2 will be close to 1. Based on the validation results, the R^2 value of Kriging approach is 0.75 and that of the proposed method is 0.97. In addition, Table 3 lists the R^2 values of Part 1 and Part 2 deviations at all three stations. In this table, the $\mathbf{y}$ vector in Eq. (24) stacks the part deviation vector $\Delta\mathbf{q}$ for the 50 new designs (due to the different units of the three translational errors and three rotational errors in $\Delta\mathbf{q}$, only the three translational errors are used in the calculation).

Clearly, the proposed method based on integrated emulation can provide much better prediction accuracy than individually fitted Kriging surrogate model. It can be seen from the table that the prediction accuracy of Kriging deteriorates at a much faster rate through the stations than the proposed method. Additionally, the model provides a direct way of calculating variance of part errors, as previously given in Eq. (19). In this particular example, the time needed to obtain part error variances at the final station for the proposed model is only about 1/16 of that for the DCS simulation model (assume 1000 simulation runs for variance estimation). Consider the high dimension of the design space, particularly if the scale of the assembly system increases, this can lead to significant time saving for fixture layout optimization.

Table 3 R^2 values of estimated Part 1 and Part 2 deviations at three stations

	$\Delta\mathbf{q}^{P1}(1)$	$\Delta\mathbf{q}^{P2}(1)$	$\Delta\mathbf{q}^{P1}(2)$	$\Delta\mathbf{q}^{P2}(2)$	$\Delta\mathbf{q}^{P1}(3)$	$\Delta\mathbf{q}^{P2}(3)$
R^2_{Kriging}	0.97	0.95	0.91	0.88	0.83	0.81
R^2_{proposed}	0.99	0.99	0.99	0.98	0.98	0.98

6 Conclusions

In this paper, a general framework for building surrogate models of dimensional variation propagation in a multistage assembly process has been developed. Due to the multi-input-multi-output nature of assembly processes, this method uses an integrated emulation technique to provide an efficient solution with higher prediction accuracy. A special design scheme has been proposed to take full advantage of the GPQQ emulation technique. To keep each surrogate model at a manageable scale, a submodel has been built for each part and linked together based on known linear relationships of part–part interactions and station–station interactions. This divide-and-conquer approach makes sure the complexity of model building will not increase dramatically with the scale of the assembly process. The proposed method has been compared with conventional Kriging surrogate modeling based on computer simulation results in the case study. The results show that the proposed method gives significantly better prediction accuracy.

It needs to be pointed out that the prediction power of surrogate models generally cannot be expected to exceed that of the models derived based on physical principles. However, the design optimization of assembly process generally does not need to find *the* optimal set of design parameters and it would be sufficient if the model can capture the general profile and trend of the response surface. Considering the benefits provided by surrogate modeling approaches, the method offers a good tradeoff between prediction accuracy and model complexity.

Acknowledgment

The authors thank the Associate Editor and three referees for their helpful comments that have led to improvements in this article.

References

[1] Shiu, B. W., Ceglarek, D., and Shi, J., 1996, "Multi-Stations Sheet Metal Assembly Modeling and Diagnostics," Trans NAMRI/SME, **24**, pp. 199–204.
[2] Jin, J., and Shi J., 1999, "State Space Modeling of Sheet Metal Assembly for Dimensional Control," Trans. ASME J. Manuf. Sci. Eng., **121**(4), pp. 756–762.
[3] Ding, Y., Ceglarek, D., and Shi, J., 2000, "Modeling and Diagnosis of Multistage Manufacturing Processes: Part I State Space Model," Proceedings of the 2000 Japan/USA Symposium on Flexible Automation, Ann Arbor, MI, 2000JUSFA-13146.
[4] Ceglarek, D., Huang, W., Zhou, S., Ding, Y., Kumar, R., and Zhou, Y., 2004, "Time-Based Competition in Manufacturing: Stream-of-Variation Analysis (SOVA) Methodology-Review," Int. J. Flexible Manuf. Syst., **16**(1), pp. 11–44.
[5] Huang, W., Lin, J., Bezdecny M., Kong, Z., and Ceglarek, D., 2007, "Stream-of-Variation Modeling I: A Generic 3D Variation Model for Rigid Body Assembly in Single Station Assembly Processes," *ASME 2006 International Manufacturing Science and Engineering Conference (MSEC2006)*, Ypsilanti, Michigan, pp. 661–672.
[6] Huang, W., Lin, J., Kong, Z., and Ceglarek, D., 2007, "Stream-of-Variation Modeling II: A Generic 3D Variation Model for Rigid Body Assembly in Multi Station Assembly Processes," ASME Trans. J. Manuf. Sci. Eng., **129**(4), pp. 832–842.
[7] Huang, W., Phoomboplab, T., and Ceglarek, D., 2009, "Process Capability Surrogate Model-Based Tolerance Synthesis for Multi-Station Manufacturing Systems (MMS)," IIE Trans., **41**(4), pp. 309–322.
[8] Chen, S., Wang, H., and Huang, Q., 2009, "Multistage Machining Process Design and Optimization Using Error Equivalence Method," ASME-MSEC, MSEC Paper No. 2009-84359.
[9] Kong, Z., Ceglarek, D., and Huang, W., 2008, "Multiple Fault Diagnosis Method in Multistation Assembly Processes Using Orthogonal Diagonalization Analysis," J. Manuf. Sci. Eng., **130**(1), pp. 11–14.

[10] Phoomboplab, T., and Ceglarek, D., 2008, "Process Yield Improvement Through Optimum Design of Fixture Layouts in 3D Multistation Assembly Systems," J. Manuf. Sci. Eng., **130**, p. 061005.
[11] Izquierdo, L., Hu, J., Du, H., Jin, R., Jee, H., and Shi, J., 2009, "Robust Fixture Layout Design for a Product Family Assembled in a Multistage Reconfigurable Line," J. Manuf. Sci. Eng., **131**, pp. 1–9.
[12] Zhong, J., Liu, J., and Shi, J., 2010, "Predictive Control Considering Model Uncertainty for Variation Reduction in Multistage Assembly Processes," IEEE Trans. Autom. Sci. Eng., **7**(4), pp. 724–735.
[13] Kim, P., and Ding, Y., 2005, "Optimal Engineering Design Guided by Data-Mining Methods," Technometrics, **47**(3), pp. 336–348.
[14] Loose, J. P., Chen, N., and Zhou, S., 2009, "Surrogate Modeling of Dimensional Variation Propagation in Multistage Assembly Processes," IIE Trans., **41**(10), pp. 893–904.
[15] Sacks, J., Schiller, S. B., and Welch W. J., 1989, "Designs for Computer Experiments," Technometrics, **31**, pp. 41–47.
[16] Sacks, J., William, J. W., Mitchell, T. J., and Wynn H. P., 1989, "Design and Analysis of Computer Experiments," Statist. Sci., **4**, pp. 409–423.
[17] Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D., 1991, "Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments," J. Am. Statist. Assoc., **86**, pp. 953–963.
[18] Santner, T. J., Williams, B. J., and Notz, W. I., 2003, *The Design and Analysis of Computer Experiments*, Springer, New York.
[19] Zhou, Q., Qian, P. Z. G., and Zhou, S., (2011), "A Simple Approach to Emulation for Computer Models With Qualitative and Quantitative Factors," Technometrics, **53**(3), pp. 266–273.
[20] Ding, Y., Kim, P., Ceglarek, D., and Jin, J., 2003, "Optimal Sensor Distribution for Variation Diagnosis in Multistation Assembly Processes," IEEE Trans. Rob. Autom., **19**(4), pp. 543–556.
[21] Cai, W., Hu, S. J., and Yuan, J. X., 1997, "A Variational Method of Robust Fixture Configuration Design for 3-D Workpieces," Trans. ASME J. Manuf. Sci. Eng., **119**(4/A), pp. 593–602.
[22] Ceglarek, D., and Shi, J., 1995, "Dimensional Variation Reduction for Automotive Body Assembly," Manuf. Rev., **8**(2), pp. 139–154.
[23] McKay, M. D., Beckman, R. J., and Conover, W. J., 1979, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," Technometrics, **21**, pp. 239–245.
[24] Lophaven, S. N., Nielsen, H. B., and Sondergaard, J., 2002, *Matlab Kriging toolbox DACE*, Version 2.5. http://www2.imm.dtu.dk/~hbn/dace/.
[25] Qian, P. Z. G., Wu, H., and Wu, C. F. J., 2008, "Gaussian Process Models for Computer Experiments With Qualitative and Quantitative Factors," Technometrics, **50**(3), pp. 283–396.
[26] Kim, P., and Ding Y., 2004, "Optimal Design of Fixture Layout in Multi-Station Assembly Processes," IEEE Trans. Autom. Sci. Eng., **1**(2), pp. 133–145.
[27] Gramacy, R. B., and Lee, H., 2008, "Bayesian Treed Gaussian Process Models With an Application to Computer Modeling," J. Am. Statist. Assoc., **103**(483), pp. 1119–1130.
[28] Stinstra, E., den Hertog, D., Stehouwer, P., and Vestjens, A., 2003, "Constrained Maximin Designs for Computer Experiments," Technometrics, **45**(4), pp. 340–346.
[29] Trosset, M. W., 1999, "Approximate Maximin Distance Designs," *ASA Proceedings of the Section on Physical and Engineering Sciences*, pp. 223–227.
[30] Chuang, S. C., and Hung, Y. C., 2010, "Uniform Designs Over General Input Domains With Applications to Target Region Estimation in Computer Experiments," Comput. Stat. Data Anal., **54**(1), pp. 219–232.
[31] Jin, R., Chen, W., and Simpson, T. W., 2001, "Comparative Studies of Metamodeling Techniques Under Multiple Modeling Criteria," Struct. Multidiscip. Optim., **23**(1), pp. 1–13.