Reconfigured Piecewise Linear Regression Tree for Multistage Manufacturing Process Control

Ran Jin and Jianjun Shi*

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology,

765 Ferst Drive NW, Atlanta, GA 30332, USA, E-mail: Jianjun.shi@isye.gatech.edu

Abstract

In a multistage manufacturing process, massive observational data are obtained from the measurements of the product quality features, process variables, and material properties. Those data have temporal and spatial relationships and may have nonlinear data structures. It is a challenge task to model the variation and its propagation from those observational data and further use the model for feedforward control purposes. This paper proposes a methodology of feedforward control based on piecewise linear models in the aforementioned circumstance. An engineering-driven reconfiguration method for piecewise linear regression trees is proposed. The model complexity is further reduced by merging the leaf nodes with the constraint of the control accuracy requirement. A case study in a multistage wafer manufacturing process is conducted to illustrate the procedure and effectiveness of the proposed method.

Key Words: Automatic Process Control, Engineering-driven Reconfiguration, Manufacturing, Piecewise Linear Regression Tree

1. Introduction

A multistage manufacturing process (MMP) refers to a manufacturing system consisting of multiple units, stations, or operations to finish a final product. In most cases, the final product quality of a MMP is determined by complex interactions among multiple stages. The quality characteristics of one stage are not only influenced by the local variations at that stage but also by the propagated variations from upstream stages. A MMP presents significant challenges, as well as opportunities, for quality engineering research. Two of the common challenges are how to model the variation and its propagations along the production stages, and how to further use the model to reduce the final product variation.

Various methodologies have been developed for modeling and control of system variability in MMPs. The feedforward control is one of the commonly adopted methodologies for such purposes. There are three typical feedforward control strategies reported in the literature based on the models used to represent a MMP.

One methodology is called Stream of Variation (SoV) based on a state space model (Jin and Shi, 1999; Shi, 2006). A SoV model is typically obtained from engineering knowledge, such as design information and physical laws of the process. Studies of feedforward control under the SoV framework includes the adjustment of the fixture position and the tool path in a machining process (Djurdjanovic and Zhu, 2005), and variation reduction in an assembly process when taking the controllability and measurement noises into account (Izquierdo *et al.*, 2007). In recent years, a new control strategy is developed based on a one-step ahead optimal criterion. The control actions are updated iteratively as the operations move on (Jiao and Djurdjanovic, 2010). The control performance of this type of approaches depends on the validity and accuracy of the state space model. The SoV based feedforward control may not be applicable (1) if the SoV model cannot be obtained based on the physics and engineering knowledge due to the system complexity; and (2) there are strong nonlinear relationships among process variables and quality variables in a complex MMP. In this situation, an effective data-driven modeling method is desirable to address nonlinear properties of the observational data.

Other methodologies are developed based on regression models, such as Robust Parameter Design (RPD) based feedforward control (Joseph, 2003) and DOE-based automatic process control (APC) (Jin and Ding, 2004). DOE-based APC determines the control actions by minimizing the predicted control objective function from a global regression model. The certainty equivalence control or cautious control strategies are employed in the APC context (Jin and Ding, 2004). Recently, Zhong *et al.,* (2010) has also investigated the impacts of model uncertainties and sensing errors on the control performances. The DOE-based APC approach yields better performance for variability reduction than the traditional RPD does. However, the DOE-based APC approach has two limitations: (1) the global regression model predicts the final quality variables when information at all stages are known. Thus, it cannot be used to control at an intermediate stage when only its upstream stage information is available; (2) The single regression model strategy can not address complex situations in a MMP when the data structure is nonlinear.

With abundant observational data available in a modern MMP, there are timely information provided about the process variables, material properties, and intermediate quality measures. With the help of these data, data mining techniques can be used to model the interrelationships among those variables. The regression tree models are one of effective approaches to model nonlinear data structure with high prediction accuracy and explicit interpretation of predictors. Therefore, the regression tree models are adopted in this paper to model the variation and its propagations in MMPs.

There are three typical methods to model a regression tree, which are greedy search, Bayesian tree, and statistical test. In general, the greedy search approaches are biased in splitting variable selection and computational intensive, such as AID algorithm (Morgan and Sonquist, 1963) and Classification and Regression Tree (CART) (Breiman *et al.*, 1984). To improve the computation efficiency, Bayesian tree is developed by proposing the priors distributions for both tree structure and parameters (Chipman *et al.*, 1998, 2002; Dennison *et al.*, 2002). The MCMC method is used to determine the posterior distributions. Another type of approaches use statistical tests to determine splitting variables, such as Smoothed and Unsmoothed Piecewise-polynomial Regression Trees (SUPPORT) (Chaudhuri *et al.*, 1994) and Generalized, Unbiased Interaction Detection and Estimation (GUIDE) (Loh, 2002; Kim *et al.*, 2007). In these approaches, the residuals of piecewise models are tested with better computational efficiency.

In this paper, piecewise linear regression trees (PLRTs) estimated by GUIDE are adopted to model MMPs for process control. The reasons for selection of the PLRTs from GUIDE are: (1) A PLRT from GUIDE has a better prediction accuracy for nonlinear data structure than a global regression model (Loh, 2002; Kim *et al.*, 2007; Loh *et al.*, 2007). (2) The interpretation of the PLRT is explicit. The predictors in the tree structure are explained as important factors under different scenarios or splitting conditions. (3) GUIDE has several superior properties over other estimation methods. For example, both categorical and continuous predictors can be assigned to different roles, such as splitting only, regression only, or both.

It also alleviates the selection bias and investigates the local pair-wise interactions. Therefore, it is an effective way to link the process, material property, and quality variables in MMPs.

A PLRT from GUIDE performs well for quality "*prediction*" in MMPs but not for "*variation reduction*". There are two major limitations that prohibit using a PLRT directly in feedforward control for variation reduction: (1) In a MMP, the temporal orders are determined by the design of a manufacturing system. However, the splitting order in PLRTs is prioritized according to the data structure and nonlinear relationships. Therefore, the splitting order in PLRTs may not reveal the same temporal sequence of a MMP. Thus, it is not feasible to select the potential models for the prediction of the final product quality at an intermediate stage based on the data only available in the upstream stages, since the downstream variables may be needed to make the prediction. This limitation results in that a control or adjustment decision cannot be made at an intermediate stage to reduce process variation in a MMP. (2) A PLRT model is usually used to predict a single response. Examples of multiple responses can be found in Segal (1992), Larsen and Speckman (2004), and Lee (2006), but not in a nested structure, i.e., one response becomes a predictor to another response. In a variation reduction problem, an intermediate quality variable may be a response as well as a predictor to the downstream process. In a typical MMP, multiple variables need to be predicted for quality control purposes. However, it is difficult to evaluate the splitting conditions from multiple trees, which limits the capability to make a control or adjustment decision to achieve optimal performance of multivariate responses.

This paper develops a unified modeling and control methodology for MMP based on a reconfigured PLRT model. The engineering design knowledge is used to reconfigure the model to an engineering complied, yet statistical equivalent model for feedforward control purposes. Furthermore, the model complexity is reduced by merging the splitting structures while satisfying the specified control accuracy requirement. Finally, a control strategy with an intermediate variable adjustment based on this reconfigured PLRT is proposed to reduce the variation of quality variables at the final stage.

The rest of the paper is organized as follows. In Section 2, we propose the methodology for modeling and feedforward control strategy. In Section 3, we use a multistage wafer manufacturing process

(MWMP) to illustrate the procedure of modeling and control. Finally, the conclusion is made in Section 4.

2. Reconfigured PLRT and feedforward control methodology

2.1. *Overview of the Proposed Methodology in Modeling and Control*

The proposed method to model and control a MMP with reconfigured PLRT is an engineering knowledge enhanced statistical method, as illustrated in Fig. 1.

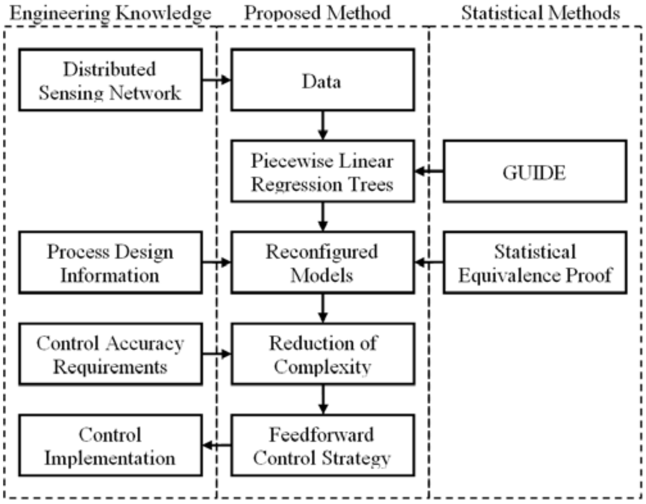

Engineering Knowledge
Proposed Method
Statistical Methods
Distributed Sensing Network
Data
Piecewise Linear Regression Trees
GUIDE
Process Design Information
Reconfigured Models
Statistical Equivalence Proof
Control Accuracy Requirements
Reduction of Complexity
Control Implementation
Feedforward Control Strategy

Fig. 1. Overview of proposed methodology

In Fig. 1, the observational data of the process, material property, and quality variables are measured from a MMP. Based on these data, PLRTs are estimated by using GUIDE to predict all intermediate and final quality variables. Then the tree models will be reconfigured to an engineering complied structure with a statistically equivalent property. Based on the final quality specifications of the MMP, the reconfigured PLRT model structure is further adjusted to find the simplest model that satisfies the accuracy requirements. In the reconfigured PLRT, a group of potential prediction models are used to predict the final product quality, as the multistage operations move from the upstream stages to the downstream stages. Therefore, a feedforward control strategy with intermediate process variable adjustment is used to take advantages of the temporally ordered layers in predicting quality variables.

The control actions are iteratively determined by solving optimization problems with product and process constraints, which are conducted to improve the final product quality in the MMP.

2.2. *Engineering-driven Reconfiguration of PLRTs*

The engineering-driven reconfiguration ensures the feasibility of PLRTs in a feedforward control strategy. The advantage of PLRTs in prediction accuracy is also preserved in control because the reconfiguration does not re-estimate the local models.

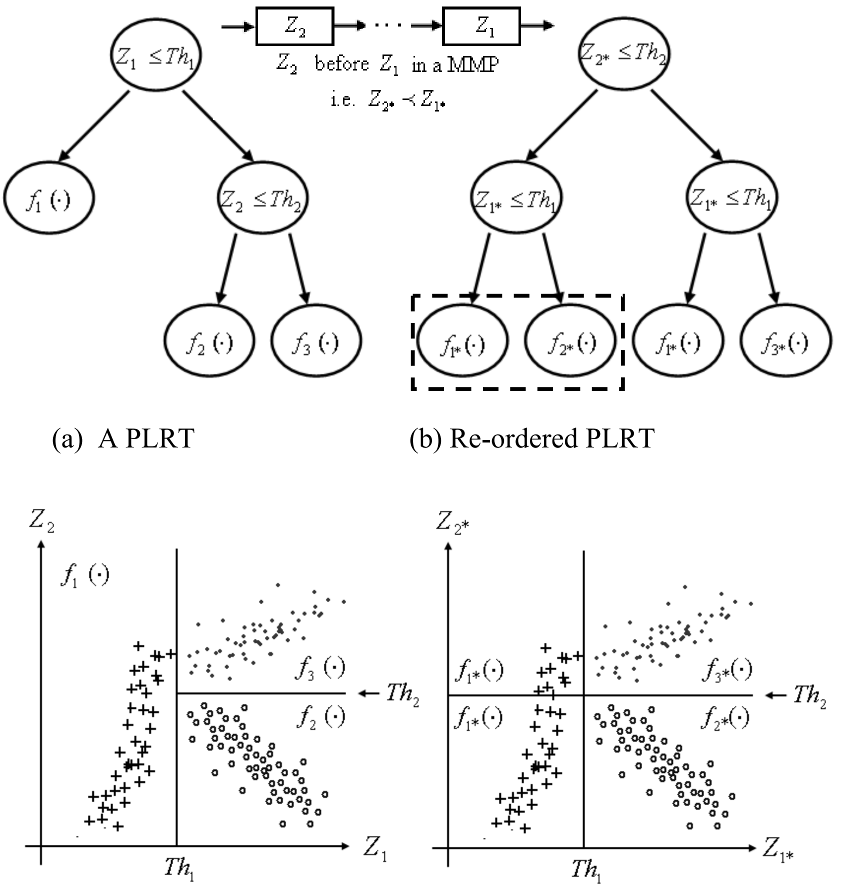
2.2.1. Multistage Manufacturing Process Modeled by PLRTs

PLRTs model the nonlinear data by partition and local fitting. Fig. 2 (a) shows an example of a PLRT estimated from GUIDE, which consists of three leaf nodes. In this tree structure, $Z_i \ (i = 1,2)$ are splitting variables; $Th_i \ (i = 1,2)$ are splitting boundaries; and $f_i \ (\cdot) \ (i = 1,2,3)$ are local regression models. When the splitting condition holds, the tree goes to the left branch. The sample space of the PLRT is illustrated in Fig. 2 (c), where $f_i \ (\cdot) \ (i = 1,2,3)$ are marked in their corresponding sub-regions.

Table 1. Variable notations

$\mathrm{Y(k)} \in \Re^{m_{\mathrm{k}} \times 1}$	: Quality variables with noise at the k-th stage
$\mathbf{Y}(0)$	: Initial quality vector before entering the manufacturing process
$\mathbf{U}_{\mathrm{k}} \in \Re^{r_{\mathrm{k}} \times 1}$	: Continuous online controllable variables at the k-th stage
u_{lk}	: The l-th variable at the k-th stage, which can be adjusted during the operations at the k-th stage
$\mathbf{X}_{\mathrm{k}} \in \Re^{n_{\mathrm{k}} \times 1}$	: Offline setting variables at the k-th stage
x_{lk}	: The l-th variable at the k-th stage, which can be adjusted between the (k-1)-th stage and the k-th stage
$\mathbf{M} \in \Re^{t \times 1}$	: Material property variables independent of stages

In a typical layout of MMP shown in Fig. 3, a stage is defined as a series of operations applied to a product to complete a manufacturing task. The intermediate quality variables are measured at each stage for modeling. A discrete part or a batch of products is processed. In this MMP, the variables can be classified as quality variables, process variables, and material property variables. Based on the controllability and variable types, variables are further classified and summarized in Table 1.



(a) A PLRT (b) Re-ordered PLRT

(c) Sample space before re-ordering (d) Sample space after re-ordering

Fig. 2. Re-ordered model from a PLRT at one stage

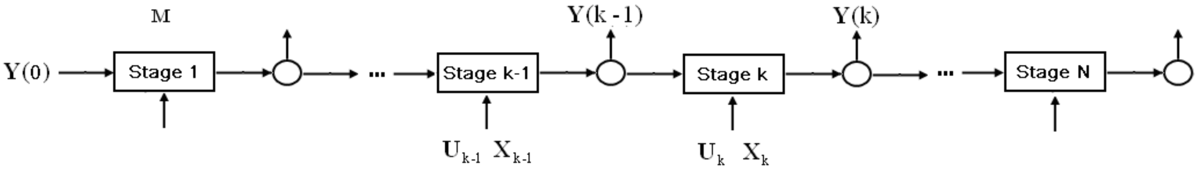


Fig. 3. A typical layout of a MMP

To model the variable relationship, a PLRT is adopted by conducting regression of the quality variables on their upstream variables. A general form of the model with *T* leaf nodes and *L* distinct splitting variables:

$$y = f(\boldsymbol{\eta}) = \sum_{i=1}^{T} f_i\ (\boldsymbol{\eta}_i) I(g_i\,(Z_1,...,Z_L)) \tag{1}$$

In this model, y could be any quality variable at any stage; if y is a quality variable at the k-th stage, then $\boldsymbol{\eta} = \{\mathbf{Y}(0), \mathbf{Y}(\mathrm{k}_1), \mathbf{U}_{\mathrm{k}_2}, \mathbf{X}_{\mathrm{k}_2}, \mathbf{M}\}$ ($\mathrm{k}_1 = 1,2,\ldots,\mathrm{k}-1$; $\mathrm{k}_2 = 1,2,\ldots,\mathrm{k}$) represents the known information at the k-th stage; $f_i(\cdot)$ and $\boldsymbol{\eta}_i$ represents the local models and the covariates in the *i*-th leaf node; $I(\cdot)$ is an indicator function, which is 1 if $g_i(\cdot)$ is non-negative, or 0 otherwise; $g_i(\cdot)$ is the combination of conditions leading to the *i*-th leaf node; and $Z_1, \ldots, Z_L$ are splitting variables for the tree structure. Furthermore, the $I(g_i(\cdot))$ can be decomposed as a product of the indicator functions of the individual splitting variables, i.e., $I(g_i(Z_1, \ldots, Z_L)) = \prod_{k=1}^{L} I(g_{i,k}(Z_k))$, where $g_{i,k}(\cdot)$ is the splitting condition of the *k*-th variable for the *i*-th leaf node. For example, in Fig. 2 (a), the splitting conditions leading to $f_2(\cdot)$ are $Z_1 > Th_1$ and $Z_2 \le Th_2$, which can be written as $I(g(Z_1, Z_2)) = I(Z_1 - Th_1)I(Th_2 - Z_2)$.

In the PLRT model estimation, there are three important issues to be addressed: splitting variable selection, splitting boundary estimation, and tree structure determination. In this paper, we follow the procedures in GUIDE, which recursively partitions the sample space, selects the splitting variables by contingence table test, and determines the splitting boundaries by minimizing the prediction errors. When a large tree grows, the 10-fold cross validation error is minimized to prune the tree structure. There are comprehensive discussion on splitting variable selection, splitting boundary estimation and pruning in the literature (Loh, 2002; Kim *et al.*, 2007; Loh *et al.*, 2007), which will not be repeated in this paper. This paper uses those methods to estimate a PLRT model from observational data. This estimated PLRT model will be used as a basis for later model reconfiguration and feedforward control design.

To explain the relationship of nodes in the tree structure, the *layer of nodes* in a tree is defined.

Definition 1. *The i-th layer of nodes*: The i-th layer of nodes in a tree is a set of nodes with depth i, i.e., the nodes which have (i-1) splits from the root of the tree, including leaf nodes and splitting nodes.

Definition 1 is illustrated with Fig. 2(a). There are three layers because the deepest leaf node from the node is reached by two splitting from the root of the tree: The splitting node of $Z_1 \le Th_1$ is the root node, which forms the first layer of the tree; Leaf node $f_1(\cdot)$ and splitting node of $Z_2 \le Th_2$ form the second layer of the tree; Leaf nodes $f_2(\cdot)$ and $f_3(\cdot)$ form the third layer of the tree.

2.2.2. Reconfiguration of Trees

The engineering knowledge of MMPs used for the reconfiguration is the temporal order and the inherent relationships among the variables, i.e., the quality at the current stage is only influenced by the upstream stages rather than the downstream stages. When there is insufficient Markov property of the quality variables, prediction by all upstream variables may also improve the prediction accuracy, comparing to the modeling by only regressing on the quality at last stage.

Assuming there are L splitting variables, these splitting variables belong to certain stages of the MMP with temporal order. This paper uses notations "$\prec$", "~" or "$\prec\sim$" of variables marked by * in the superscript to describe the temporal order. Table 2 summarizes the temporal relationship of these variables, and Z_{i*} ($i = 1,2;$) is used for denoting Z_i in a temporal order. In MMPs, such a kind of temporal order of the quality and process variables at the (k-1)-th stage and the k-th stage can be presented as: $\mathbf{X}_{(k-1)*} \prec\sim \mathbf{U}_{(k-1)*} \prec \mathbf{Y}((k-1)*) \prec \mathbf{X}_{k*} \prec\sim \mathbf{U}_{k*} \prec \mathbf{Y}(k*)$.

With the temporal order of the splitting variable, the original PLRT is re-ordered into a temporally complied tree, which is defined below for further analysis.

Definition 2. *Temporally complied tree:* A tree is temporally complied if the splitting variables in the tree is temporally ordered, which is defined by the MMP layout, i.e., if $Z_{i*} \prec\sim Z_{j*}$, then Z_{i*} is in a closer layer or the same layer as the root compared to the location of Z_{j*}.

The reconfigured PLRT should have three appealing properties for the feedforward control purpose: (1) the reconfigured PLRT should be a temporally complied tree; (2) several PLRTs are estimated to predict the intermediate and final quality, which should be combined into a single decision structure; and (3) the reconfigured PLRT should be statistical equivalent to the PLRT models with high prediction accuracy.

The reconfiguration of PLRTs consists of two steps: (1) each PLRT is reconfigured according to the temporal order of the splitting variables, called *re-ordering*; and (2) a group of PLRTs is combined as a reconfigured PLRT called *combining*.

Table 2. Notations of temporal orders

$Z_{1*} \prec Z_{2*}$: Z_1 is temporally prior to Z_2 ;
$Z_{1*} \sim Z_{2*}$: Z_1 and Z_2 have the same temporal order;
$Z_{1*} \prec\sim Z_{2*}$: Z_1 is temporally prior or the same as Z_2 .

2.2.2.1 Re-ordering

Assuming the splitting order in a PLRT is not consistent with the temporal order as $Z_{1*} \prec\sim Z_{2*} \prec\sim \ldots \prec\sim Z_{L*}$, the procedure to re-order a PLRT is proposed in the Algorithm 1 in Table 3.

In Algorithm 1, all splitting variables Z_i ($\forall i$) are considered in partitioning the regions in step 2; $g_i^j(\cdot)$ are the decomposed sub-regions of $g_i(\cdot)$, where D_i is the total number of sub-regions considering all possible splits of Z_i ($\forall i$). In step 3, if the Merge Condition I (defined below) is satisfied, the sub-regions will be merged; otherwise, no further merging is needed.

Table 3. The algorithm for the re-ordering

```
Algorithm 1.
Step 1. Convert the PLRT to a summation of $f_i(\cdot)$ and $g_i(\cdot)$ as Eq. (1)
Step 2. Partition the region of $g_i(\cdot)$ w.r.t all splitting variables into the decomposed sub-regions
        $g_i^j(\cdot)$ ($j = 1,\ldots,D_i$), i.e., $y = \sum_{i=1}^{T} f_i(\boldsymbol{\eta}_i) I(g_i(Z_1,\ldots,Z_L)) = \sum_{i=1}^{T}\sum_{j=1}^{D_i} f_i(\boldsymbol{\eta}_i) I(g_i^j(Z_1,\ldots,Z_L))$
Step 3. Merge the sub-regions $g_i^j(\cdot)$ and $f_i(\boldsymbol{\eta}_i)$ for $Z_i$ ($i = 1,\ldots,L$) from $Z_{L*}$ to $Z_{1*}$, if the
        Merge Condition I is satisfied. The final re-ordered model is
        $y^* = \sum_{i=1}^{T*} f_{i*}(\boldsymbol{\eta}_{i*}) I(g_{i*}(Z_{1*},\ldots,Z_{L*}))$.
Step 4. Formulate the layers into temporal complied tree based on the re-ordered model
```

The Merge Condition I for Z_i in any two decomposed sub-regions j_1 and j_2 in leaf nodes i_1 and i_2 is: $g_{i_1,k}^{j_1}(Z_k) = g_{i_2,k}^{j_2}(Z_k)$ ($\forall k \neq i$) and $f_{i_1}(\boldsymbol{\eta}_{i_1})$ is the same model as $f_{i_2}(\boldsymbol{\eta}_{i_2})$. Here the splitting conditions of the decomposed regions are $I(g_{i_1,i}^{j_1}(Z_i))\prod_{\forall k\neq i} I(g_{i_1,k}^{j_1}(Z_k))$ and $I(g_{i_2,i}^{j_2}(Z_i))\prod_{\forall k\neq i} I(g_{i_2,k}^{j_2}(Z_k))$. $f_{i_1}(\mathbf{\eta}_{i_1})$ and $f_{i_2}(\mathbf{\eta}_{i_2})$ are the associated local regression models. After the merging process, the splitting condition for the newly merged leaf node is $\prod_{\forall k\neq i} I(g_{i_1,k}^{j_1}(Z_k))$ (or $\prod_{\forall k\neq i} I(g_{i_2,k}^{j_2}(Z_k))$).

To illustrate the Merge Condition I, the tree in Fig. 2 (a) is re-ordered as an example. Following the procedure of Algorithm 1, there will be four partitioned sub-regions as shown in Fig. 2 (d) after step 2. In step 3, assuming $Z_{2*} \succ Z_{1*}$ Z_{1*} should be merged first. Considering the merge in the dashed rectangular in Fig. 2 (b), their splitting conditions are $I(Th_1 - Z_{1*})I(Th_2 - Z_{2*})$ and $I(Z_{1*} - Th_1)I(Th_2 - Z_{2*})$. In this example, $I(g_{1,2}^{1}(Z_{2*})) = I(g_{2,2}^{1}(Z_{2*})) = I(Th_2 - Z_{2*})$, but $f_{1*}(\cdot)$ and $f_{2*}(\cdot)$ are not the same. Therefore, the Merge Condition I is not satisfied and these two leaf nodes cannot be merged. Once the re-ordered model is obtained, we can formulate Z_{2*} in the first layer, then Z_{1*} in the second layer.

Statement 1. *Statistical equivalence in re-ordering*: The original PLRT is statistically equivalent to the re-ordered temporally complied tree in prediction, i.e., $y = y^*$.

The proof of Statement 1 is in Appendix 1. To illustrate the equivalence, Fig. 2 (c) and Fig. 2 (d) are compared. By given a new sample, the local prediction models $f_i(\cdot)$ $(i = 1,2,3)$ in Fig. 2(c) and $f_{i*}(\cdot)$ $(i = 1,2,3)$ in Fig. 2 (d) are identical, since the re-ordering does not re-estimate the local regression models.

2.2.2.2 Combining

After re-ordering, multiple PLRTs are combined as a single reconfigured tree to predict multiple quality variables. If there are N_1 re-ordered PLRTs, with T_n^* leaf nodes and L_n^* splitting variables in the *n*-th tree ($n = 1,2,...,\mathrm{N}_1$), the general form of these re-ordered models are denoted as

$$y_n^* = \sum_{i=1}^{T_n^*} f_{i*}^n(\boldsymbol{\eta}_{i*}^n) I(g_{i*}^n(Z_{1*}^n,...,Z_{L_n^*}^n)) \qquad (2)$$

where all notations are similarly denoted as Eq. (1) except "n" for the n-th tree. Furthermore, $Z_{1*},...,Z_{L*}$ are the splitting variables in all these trees, with temporal order $Z_{1*} \prec\sim Z_{2*} \prec\sim ... \prec\sim Z_{L*}$. The procedure to combine the re-ordered models is proposed in the Algorithm 2 in Table 4.

Table 4. The algorithm for the combining

Algorithm 2.

Step 1. Obtain the re-ordered structure for the models in the form of Eq. (2)

Step 2. Decompose the $g_i^n(\cdot)$ into $g_i^{n,j}(\cdot)$ using the same approach of Step 2 in Algorithm 1 considering all splitting variables in different PLRTs, i.e.,

$$y_n^* = \sum_{i=1}^{T_n^*} f_{i*}^n(\boldsymbol{\eta}_{i*}^n) I(g_{i*}^n(Z_{1*}^n,...,Z_{L_n^*}^n)) = \sum_{i=1}^{T_n^*}\sum_{j=1}^{D_{n,i}^*} f_{i*}^n(\boldsymbol{\eta}_{i*}^n) I(g_{i*}^{n,j}(Z_{1*},...,Z_{L^*}))$$

Step 3. Merge the decomposed sub-regions $g_i^{n,j}(\cdot)$ using the similar procedure of Step 3 in Algorithm 1 if the Merge Condition II is satisfied. The final combined model is

$$y_n^* = \sum_{i=1}^{T^*} f_{i*}^n(\boldsymbol{\eta}_{i*}^n) I(g_{i*}^{comb}(Z_{1^*},...,Z_{L^*}))\,(n=1,2,...,\mathrm{N}_1).$$

Step 4. Formulate the layers into temporal complied tree based on the combined model

In Algorithm 2, all splitting variables in these re-ordered trees are considered in the decomposition in step 2. $D_{n,i}^*$ is the total number of decomposed sub-regions considering all possible splits of Z_{i*} ($\forall i$) from the i-th leaf node in the n-th tree. In step 3, if the Merge Condition II is satisfied, the sub-regions will be merged, and a group of N_1 regression models for multiple responses is formed. Otherwise, no further merging is needed.

The Merge Condition II for Z_{i*} *in two decomposed sub-regions* j_1 *and* j_2 *in leaf nodes* i_1 *and* i_2 is: $I(g_{i_1^*,k^*}^{n,j_1}(Z_{k*}^n)) = I(g_{i_2^*,k^*}^{n,j_2}(Z_{k*}^n))$, ($\forall k^* \neq i^*$) and $f_{i_1^*}^n(\boldsymbol{\eta}_{i_1^*}^n)$ is the same model as $f_{i_2^*}^n(\boldsymbol{\eta}_{i_2^*}^n)$. Here the splitting conditions of these two decomposed sub-regions are $I(g_{i_1^*,i^*}^{n,j_1}(Z_{i*}^n)) \prod_{\forall k^* \neq i^*} I(g_{i_1^*,k^*}^{n,j_1}(Z_{k*}^n))$ and $I(g_{i_2^*,i^*}^{n,j_2}(Z_{i*}^n)) \prod_{\forall k^* \neq i^*} I(g_{i_2^*,k^*}^{n,j_2}(Z_{k*}^n))$. $f_{i_1^*}^n(\boldsymbol{\eta}_{i_1^*}^n)$ and $f_{i_2^*}^n(\boldsymbol{\eta}_{i_2^*}^n)$ are the associated local models. After the merging process, the splitting condition for the newly merged leaf node is $\prod_{\forall k^* \neq i^*} I(g^{n,j_1}_{i_1^*,k^*}(Z^n_{k^*}))$ (or $\prod_{\forall k^* \neq i^*} I(g^{n,j_2}_{i_2^*,k^*}(Z^n_{k^*}))$).

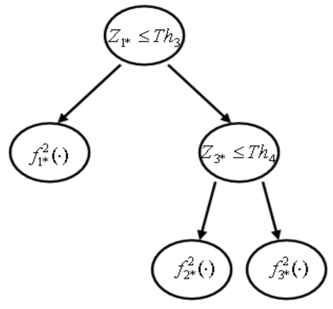
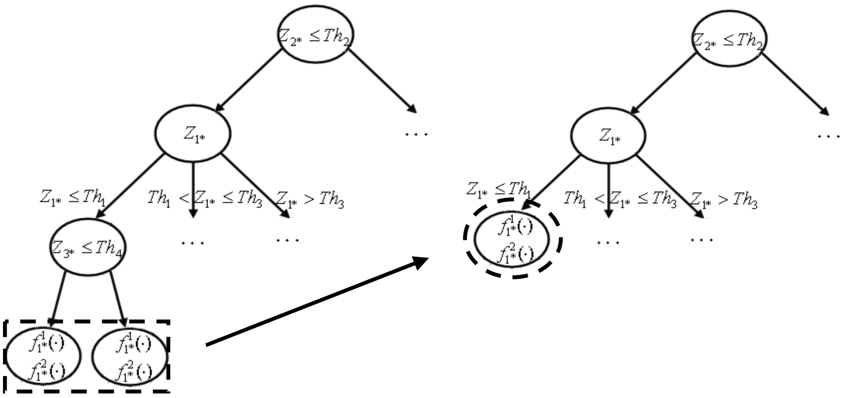


Fig. 4. Another re-ordered PLRT



(a) Before merge (b) After merge

Fig. 5. Merging leaf nodes in combining

To illustrate the Merge Condition II, two trees in Fig. 2 (b) and Fig. 4 are combined as an example, assuming $Th_1 < Th_3$. In this case, the local models $f_{i*}(\cdot)$ ($i = 1,2,3$) in Fig. 2 (b) becomes $f^1_{i*}(\cdot)$ ($i = 1,2,3$) to distinguish the models in Fig. 4. There are three distinct splitting variables in these trees: Z_{1*}, Z_{2*}, and Z_{3*}. Following the procedure of Algorithm 2, all possible splits are generated in step 2. In step 3, assuming $Z_{2*} \prec Z_{1*} \prec Z_{3*}$, Z_{3*} should be merged first. Considering the merger of two leaf nodes that are marked by the dashed rectangular in Fig. 5 (a), the splitting conditions are

$I(Th_1 - Z_{1*})I(Th_2 - Z_{2*})I(Th_4 - Z_{3*})$ and $I(Th_1 - Z_{1*})I(Th_2 - Z_{2*})I(Z_{3*} - Th_4)$. In this example, $I(g_{1,1}^{n,1}(Z_{1*})) = I(g_{2,1}^{n,1}(Z_{1*})) = I(Th_1 - Z_{1*})$ for n=1, 2, and $I(g_{1,2}^{1,1}(Z_{2*})) = I(g_{2,2}^{1,1}(Z_{2*})) = I(Th_2 - Z_{2*})$. The local models are also identical. Therefore, the Merge Condition II is satisfied and these two leaf nodes should be merged, shown in Fig. 5 (b).

Statement 2. *Statistical equivalence in combining*: A group of re-ordered models from PLRTs is combined into a single statistically equivalent model using Algorithm 2.

The proof of Statement 2 is shown in Appendix 2. To illustrate the equivalence, the local models in the re-ordered trees (Fig. 2 (b) and Fig. 4) are compared with the reconfigured tree (Fig. 5 (b)). For example, if $Z_{1*} \le Th_1$, $Z_{2*} \le Th_2$ and $Z_{3*} > Th_4$, the local models for prediction are $f_{1*}^1(\cdot)$ and $f_{1*}^2(\cdot)$, which are the same as the models with the same splitting conditions, circled by dashed circle in Fig. 5 (b).

After the reconfiguration, the splitting variables are re-ordered into different layers, which map to the temporal order of the manufacturing stages, as shown in Fig. 6. The splitting conditions are combined, which lead to different model groups to predict the intermediate and final quality variables stage-by-stage. This reconfigured PLRT is preferred over the original PLRT for the purpose of the feedforward control.

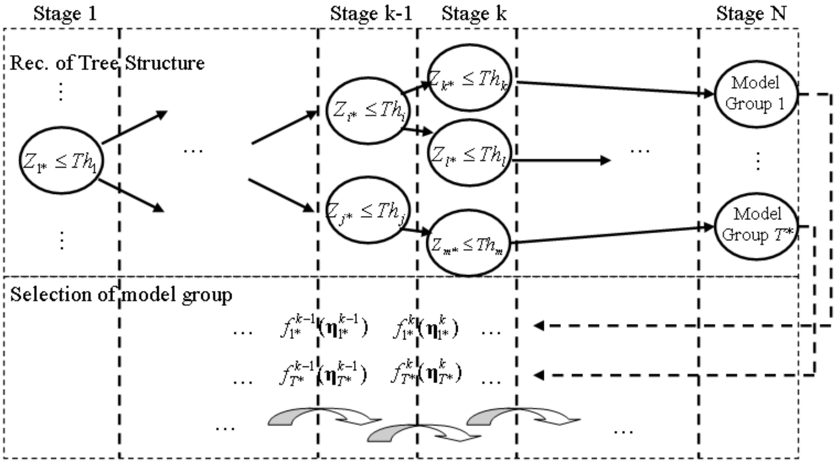


Fig. 6. Reconfigured PLRT for a MMP

2.3. *Reconfigured model complexity and control accuracy*

The PLRTs from GUIDE are pruned by cross validation to minimize the predicted SSE (Loh, 2002). After the reconfiguration, the reconfigured model yields the best prediction accuracy due to the statistical equivalency. However, the reconfigured PLRT may be very complex with many leaf nodes and many potential local models, which increases computational efforts in the control optimization. On the other hand, there is an engineering tolerance for the controlled objectives, which can be further transferred to the needs of the model precision used in the feedforward control. In other words, the model used for control purpose may not have the same level of high precision requirement as the prediction obtained from the original PLRT. Therefore, the model complexity can be reduced, while the model still satisfies the control accuracy requirements. The reduction of the model complexity is achieved by assuming that there are limited numbers of variables having nonlinear relationship with the response. Detail discussions on how to further simplify the reconfigured PLRT with fewer leaf nodes is provided below.

In a reconfigured PLRT, the control performance can be evaluated by the accumulative errors of all PLRT model errors at different stages. However, different model groups may be selected in control according to the splitting conditions. Thus, it is difficult to estimate the control accuracy for every possible path used in control. In this paper, the largest prediction variance is proposed to evaluate the control accuracy of this leaf node, shown as follows:

$$\sigma_{k,\mathrm{j}}^2 = \max_{\mathbf{U}_1,\ldots,\mathbf{U}_\mathrm{N},\mathbf{X}_1,\ldots,\mathbf{X}_\mathrm{N}} \mathrm{Var}(\mathbf{Y}(\mathrm{N})_\mathrm{j}) \tag{3}$$

$$\text{s.t.} \quad x_{\mathrm{lk}} \in \{x_{\mathrm{lk}}\},\ u_{\mathrm{lk}}^{\mathrm{L}} < u_{\mathrm{lk}} < u_{\mathrm{lk}}^{\mathrm{U}}, \forall \mathrm{l}, \forall \mathrm{k}$$

where $\sigma_{k,\mathrm{j}}^2$ is the maximum prediction variance of the j-th quality variable in the *k*-th leaf node, obtained by enumerating all control actions; $\mathbf{Y}(\mathrm{N})_\mathrm{j}$ is the predicted final quality variable; the optimization constrains are the controllability of the process variables, where $\{x_{\mathrm{lk}}\}$ is the set of all possible settings of x_{lk}, and $u_{\mathrm{lk}}^{\mathrm{L}}, u_{\mathrm{lk}}^{\mathrm{U}}$ represents the lower and upper bound of the feasible range for u_{lk}.

The control accuracy of the overall structure is evaluated by the pooled variance of these leaf nodes. Assuming there are equal numbers of products in different leaf nodes in control, thus, the pooled variance is the average of the control accuracy of all leaf nodes, shown as follows:

$$\sigma^2_{\mathrm{Rec.,j}} = \frac{1}{T}\sum_{k}^{T}\sigma^2_{k,\mathrm{j}} \tag{4}$$

where T is the number of leaf nodes in the reconfigured PLRT; and $\sigma^2_{\mathrm{Rec.,j}}$ represents the control accuracy for the j-th quality variable. In this way, the control accuracy is evaluated by $\sigma^2_{\mathrm{Rec.,j}}$.

To reduce the model complexity, the leaf nodes should be merged. With less leaf nodes, the prediction performance will be degraded because the PLRTs are pruned to minimize the predicted SSE in the cross validation. There are two issues to be addressed to balance the model complexity and the control accuracy: (1) which leaf nodes should be merged, and (2) when the merging process should be stopped?

The leaf nodes with the least important splitting structure should be merged first because it would result in the smallest decrease in the prediction accuracy. Although the control accuracy is evaluated based on the reconfigured PLRT, the temporarily complied splitting variables no longer provide information on the importance of splitting structure. Nevertheless, the original PLRTs preserve the importance of the splitting variables for prediction in splitting orders from the more significant ones to the less significant ones. Therefore, reducing the number of leaf nodes will merge the nodes in the deepest layer in the original PLRTs. The merging process is stopped when the control accuracy of the reconfigured PLRT exceeds the pre-determined control accuracy requirement. The merging process is completed in an iterative way shown in Fig. 7.

In Fig. 7, the control accuracy of the current reconfigured PLRT $\sigma^2_{\mathrm{Rec.,j}}$ is estimated first. Then different deepest leaf nodes in the original PLRTs are merged once at a time. In this way, a set of new control accuracy estimates $\sigma^2_{\mathrm{Rec.,j}}$ of the final reconfigured PLRTs is obtained. We choose the minimal $\sigma^2_{\mathrm{Rec.,j}}$ and compare it with a pre-determined threshold $\sigma^2_{\mathrm{T,j}}$ for the j-th quality variable. One concludes that the model with minimal $\sigma^2_{\text{Rec.,j}}$ is acceptable if it is smaller than $\sigma^2_{\text{T,j}}$. In this case, we reconstruct the reconfigured PLRT in the next iteration. Otherwise, the control accuracy of the current model does not satisfy the control accuracy requirement, thus the merging should be stopped. After this procedure, the reconfigured model has reached a balance between the model complexity and the control accuracy.

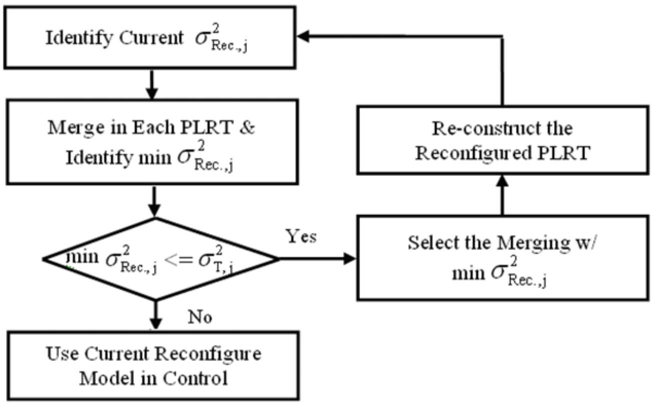

Identify Current $\sigma^2_{\text{Rec.,j}}$
Merge in Each PLRT &
Identify min $\sigma^2_{\text{Rec.,j}}$
min $\sigma^2_{\text{Rec.,j}} <= \sigma^2_{\text{T,j}}$
Yes
No
Select the Merging w/
min $\sigma^2_{\text{Rec.,j}}$
Re-construct the
Reconfigured PLRT
Use Current Reconfigure
Model in Control

Fig. 7. The procedure to reduce model complexity

2.4. *Feedforward control strategy of reconfigured PLRT*

The engineering-driven reconfiguration has made it possible to develop a feedforward control strategy by actively adjusting the process variables and compensating the quality variable for variation reduction. The overall strategy is shown in Fig. 8. The basic idea to achieve a feedforward control based on the reconfigured PLRT models is presented below.

At each controllable stage, several potential model groups are determined based on the splitting conditions. If the splitting variables are measured at previous stages or layers, a model group in a leaf is selected when the splitting conditions are satisfied. Otherwise, several branches and leaves may be selected, which form a cluster of potential model groups. In this case, the splitting conditions are formulated as constrains in the optimization problem.

The control optimization at the k-th stage is formulated as the-smaller-the-better problem:

$$\min_{u_{\mathrm{li}}, x_{\mathrm{li}}, \mathrm{i=k}, \cdots \mathrm{N}} \quad J(\mathbf{U}, \mathbf{X}) = \sum_{\mathrm{j=1}}^{\mathrm{m}} \mathrm{c}_{\mathrm{j}} \mathrm{E}(\mathbf{Y}(\mathrm{N})_{\mathrm{j}}^{2}) \tag{5}$$

$$\begin{aligned} s.t. \quad & \mathbf{Y}(\mathrm{N})_{\mathrm{j}} = f_{\omega}^{\mathrm{j}}(\boldsymbol{\eta}) \\ & h(\mathbf{Y}(\mathrm{s})_{\mathrm{j}}) < H_{\mathrm{js}} \\ & x_{\mathrm{li}} \in \{x_{\mathrm{li}}\} \\ & u_{\mathrm{li}}^{\mathrm{L}} < u_{\mathrm{li}} < u_{\mathrm{li}}^{\mathrm{U}} \\ & I(g_{\omega}(Z_1, ..., Z_L)) > 0 \\ & \mathrm{s} = 1,2,\cdots,\mathrm{N}; \mathrm{i} = \mathrm{k},\cdots,\mathrm{N}. \end{aligned}$$

where the objective function is the weighted summation of the second order moment of m predicted final quality variables; $\mathbf{Y}(\mathrm{N})_{\mathrm{j}}$ is the j-th final quality variable predicted from the k-th stage; c_{j} is the weight of the importance of the j-th quality variable. The decision variables are the process variables from the k-th stage to the N-th stage. In the constraints, $f_{\omega}^{\mathrm{j}}(\cdot)$ is a potential model group for the quality prediction determined by the splitting conditions; $h(\mathbf{Y}(\mathrm{s})_{\mathrm{j}}) < H_{\mathrm{js}}$ represents the quality specification for the j-th quality variable at the s-th stage ($\mathrm{s} = 1,2,\cdots,\mathrm{N}$); $u_{\mathrm{li}}^{\mathrm{L}} < u_{\mathrm{li}} < u_{\mathrm{li}}^{\mathrm{U}}$ and $x_{\mathrm{li}} \in \{x_{\mathrm{li}}\}$ ($\mathrm{i} = \mathrm{k},\cdots,\mathrm{N}$) represent the controllability as described in Eq. (3). The optimization problem is solved by Iterated Local Search Algorithm (Stutzle, 1998).

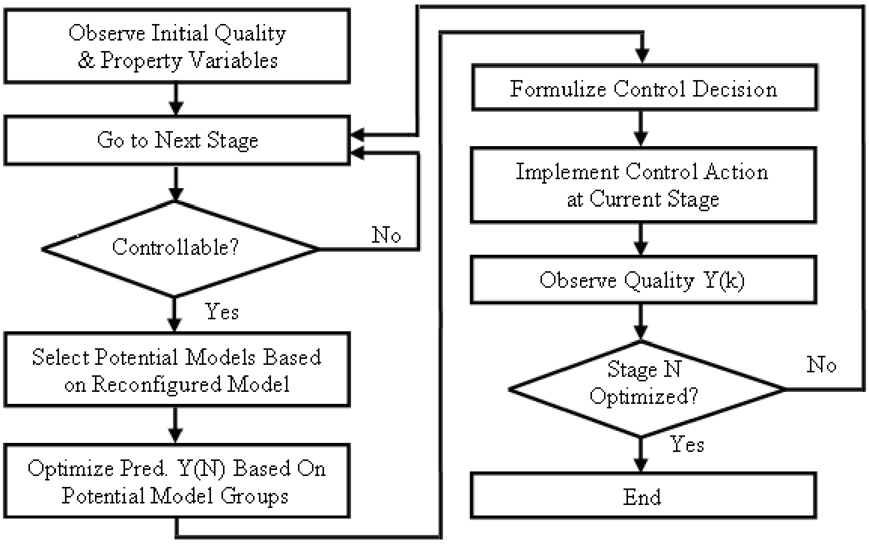

Observe Initial Quality & Property Variables
Go to Next Stage
Controllable?
No
Yes
Select Potential Models Based on Reconfigured Model
Optimize Pred. Y(N) Based On Potential Model Groups
Formulize Control Decision
Implement Control Action at Current Stage
Observe Quality Y(k)
Stage N Optimized?
No
Yes
End

Fig. 8. The overall feedforward control strategy

3. Case study

A case study in a multistage wafer manufacturing process (MWMP) is conducted to illustrate the procedure of modeling and control based on the reconfigured PLRTs. A comparison study of the feedforward control strategy based on a reconfigured PLRT and regression model groups is conducted to show the effectiveness of the proposed approach.

3.1. *Wafer manufacturing processes*

A MWMP is a complex MMP involving chemical and mechanical process to transform a silicon ingot into a wafer with uniform thickness, fine surface roughness, and good overall geometric shape for future processing. The process in this case study consists of five major manufacturing stages as shown in Fig. 9, including slicing, lapping, chemical vapor deposition (CVD) of polysilicon, CVD of SiO_2, and polishing. Each stage is a combination of multiple operations with quality measured at the end of the stage.

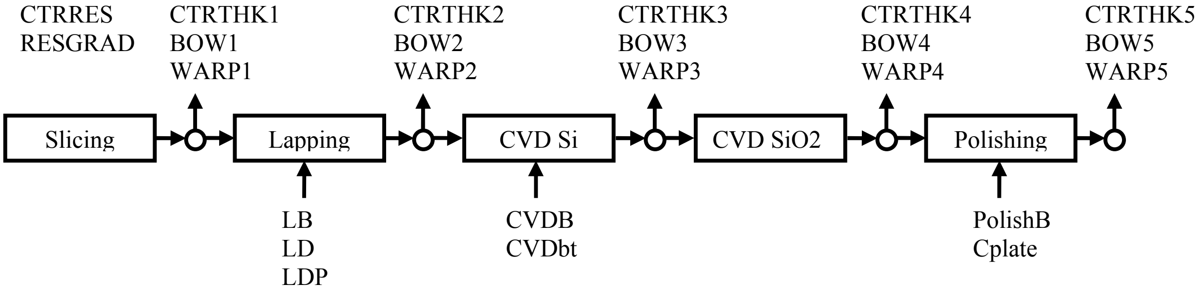


Fig. 9. A five-stage MWMP

In a MWMP, the overall geometric shape is a critical geometric quality index of a wafer. BOW and WARP of a wafer represent the overall shape of a wafer, which is used as the quality improvement objective in the case study. In general, smaller absolute values of these variables indicate better quality of the wafer.

In this case study, observational data of three types of variables (quality, process, and material property) were collected in a real production environment. Those variables are summarized in Table 5. In this table, the CTRRES represent the position of wafers in an ingot. In the case study, the central thickness of a wafer is measured in each stage, which is used in the selection of settings of downstream process parameters. Therefore, the central thickness of wafer is treated as a predictor rather than quality variables. The initial quality vector $\mathbf{Y}(0)$ in this process is assumed to be a zero vector.

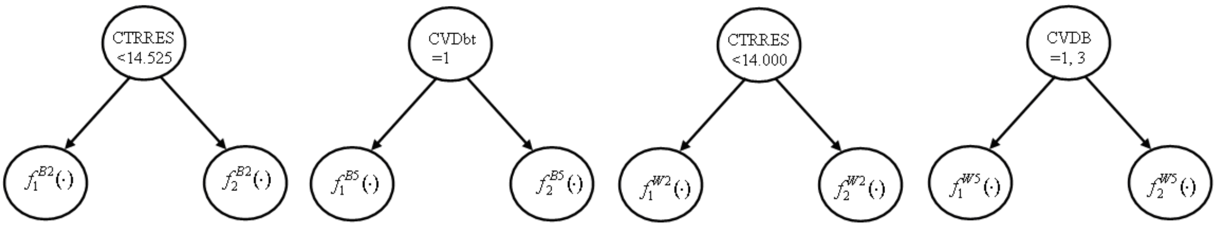
Table 5. Measured variables in the MWMP

Variable Type	Variable Name	Discrete / Continuous	Measured Stage	Physical Meaning
Process Variables	LB	Discrete	Lapping	Lapping batch, representing processing time and compressive force with 15 levels
	LD	Discrete	Lapping	Lapping disk, representing maintenance conditions of pulley discs with 5 levels
	LDP	Discrete	Lapping	Position of wafer in lapping disks with 6 levels
	CVDB	Discrete	CVD Si	CVD batch, representing different tubes and processing time with 5 levels
	CVDbt	Discrete	CVD Si	CVD boat, representing wafers' positions in CVD tube
	PolishB	Discrete	Polishing	Polishing batch, representing age of slurry and polishing pad with 12 levels
	Cplate	Discrete	Polishing	Ceramic plate, representing the alignment of ceramic plate holders with 4 levels
Material Property Variables	CTRRES	Continuous	Independent of Stages	Central resistivity of a wafer
	RESGRAD	Continuous	Independent of Stages	Resistivity gradient of a wafer
Quality Variables	BOW	Continuous	All Five Stages	Local warp at the center of a wafer
	WARP	Continuous	All Five Stages	Maximum local warp
	CTRTHK	Continuous	All Five Stages	Central thickness of wafer, used as predictors rather than responses.

In this process, some intermediate quality specifications of wafers need to be satisfied. For example, the thickness of a wafer in certain lapping batch should be within a specified range; otherwise, the wafer will be broken during the lapping. These intermediate quality specifications are formulated as constraints in the optimization problem. Overall, data of 373 wafers are obtained in production for the case study. The PLRTs are constructed based on the training data set (250 wafers) and the control performance is evaluated based on the testing data set (123 wafers).

3.2. *PLRT models of the MWMP*

The PLRTs for this MWMP are estimated and shown in Fig. 10. In Fig. 10, there are four splitting structures to predict BOW2, BOW5, WARP2, and WARP5, while the models for other quality variables are regression models without splitting structures. In each leaf node, there is a local regression model, where "B" or "W" represents the quality variable BOW or WARP respectively. The PLRTs have explicit interpretations. For example, material property CTRRES at different segments of ingot yield different prediction models to predict BOW2 (Fig. 10 (a)). This shows that the prediction of BOW2 is influenced by the material heterogeneity of wafers at the tail and the head of the ingot. Similar interpretations are obtained for BOW5, WARP2, and WARP5.



(a) Model BOW2 (b) Model BOW5 (c) Model WARP2 (d) Model WARP5

Fig. 10. PLRTs in MWMP

Since GUIDE does not consider the interactions in estimating the local regression models, the regression model of $f^{B3}(\cdot)$, $f^{B4}(\cdot)$, $f^{W3}(\cdot)$ and $f^{W4}(\cdot)$ is re-estimated considering the interactions of predictors to further reduce the predicted SSE.

3.3. *Reconfiguration of PLRT*

Based on the PLRT from GUIDE, a reconfigured PLRT is obtained in Fig. 11. In Fig. 11, the temporal order of the splitting variables is $\text{CTRRES} \prec \text{CVDB} \prec \text{CVDbt}$, which is re-ordered into different layers of the reconfigured PLRT from the root. In this example, CTRRES is split into three sub-regions in the first layer of the model, which are based on the splittings in the original models to predict BOW2 and WARP2. In the second and the third layer of the model, CVDB and CVDbt are split as the same as the original model. In this way, 12 regression model groups are generated, which will be selected by the splitting conditions. The overall structure clearly represents the sequence of manufacturing from the root to the leaf nodes, and predicts multiple intermediate and final quality variables.

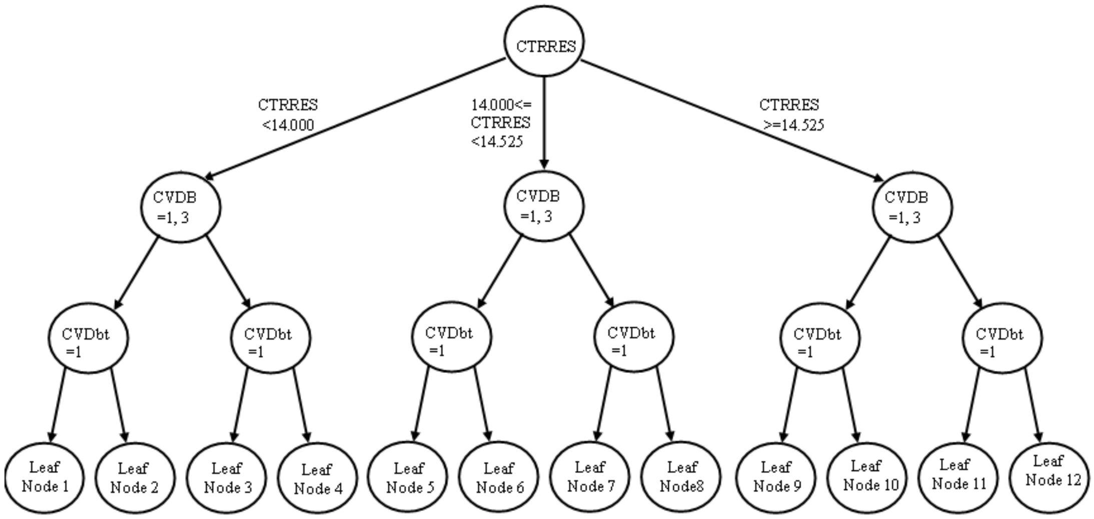
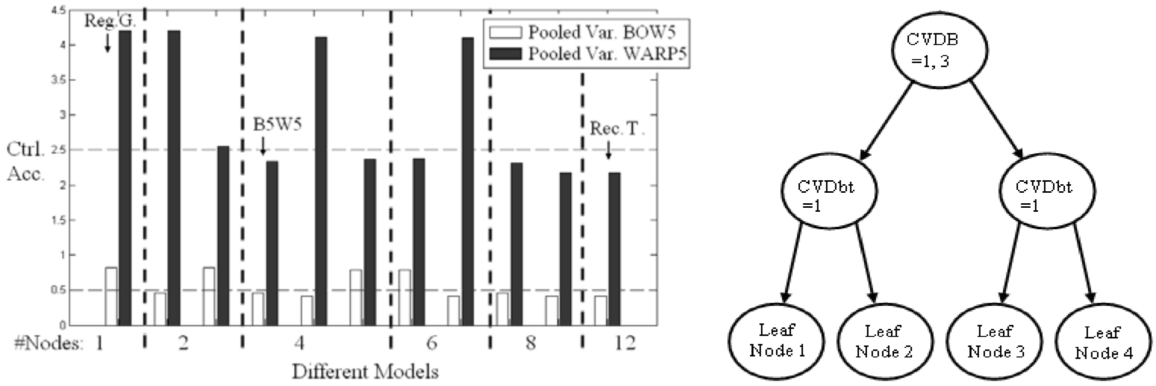


Fig. 11. Reconfigured PLRT for MWMP

3.4 *Reduce model complexity*

To reduce the model complexity, the control accuracy of BOW5 and WARP5 are evaluated in different reconfigured models. Fig. 12 (a) shows control accuracy of 11 models from regression group (Reg. G.) with the worst control accuracy to the reconfigured PLRT (Rec. T.) with the best control accuracy. The number of nodes is marked for each model. In this figure, the control accuracy varies as different model complexities are adopted. Such an analysis provides guidelines to select a model with appropriate complexity that satisfies the control accuracy requirement. In this case study, the control accuracy requirement of BOW5 and WARP5 are 0.5 and 2.5 (horizontal dashed lines). The model with splits in BOW5 and WARP5 (B5W5) has the minimal number of leaf nodes to satisfy the requirement, which has only two significant splitting variables and four leaf nodes retained for control optimization, shown in Fig. 12 (b). By comparing the "Rec. T." model, the model complexity has been significantly reduced.



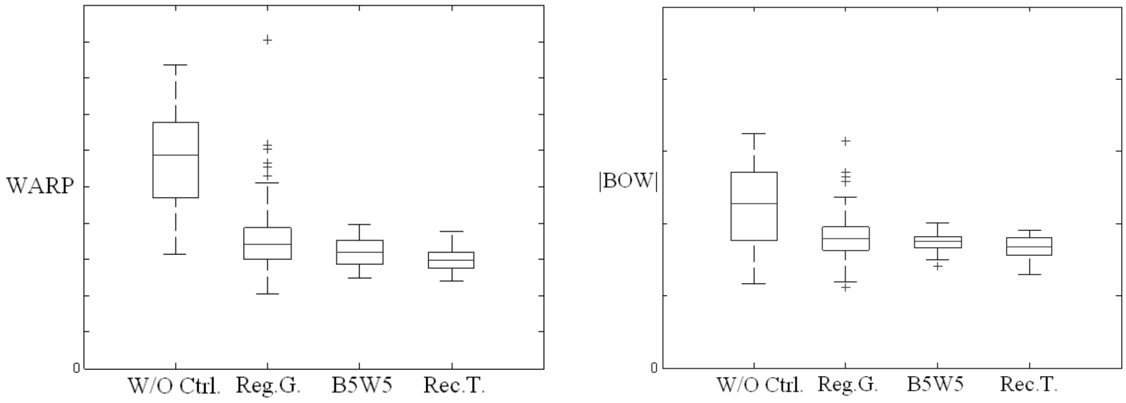
(a) Control accuracy of different models (b) Final reconfigured PLRT for control

Fig. 12. Control accuracy and model complexity

3.5. *Simulation study of feedforward control*

To compare the feedforward control strategy, a total of 50 simulation runs were conducted based on three different models: “Reg. G.”, “B5W5” and “Rec. T.”. In the simulation, the “Reg. G.” model is a global regression model without using splitting variables. The “B5W5” and “Rec. T.” models use the reconfigured PLRT models for prediction. Without loss of the generality, we set $c_j = 1$ in Eq. (5).

Figure 13 (a) shows the controlled WARP5 in one simulation run. The horizontal axis represents the performance without control and with control based on different models. The control based on the reconfigured PLRTs yields better performance in reducing mean and variance of the final quality than regression group models. Moreover, there is no significant increase in mean and variance of the controlled quality when using the “B5W5” model verses the “Rec. T.”. This indicates that there is no significant loss in control performance when merging some of the splitting structures. Fig. 13 (b) shows controlled performance of the absolute value of BOW5 with similar interpretations.



(a) WARP5 (b) |BOW5|

Fig. 13. Controlled quality performance in a simulation run

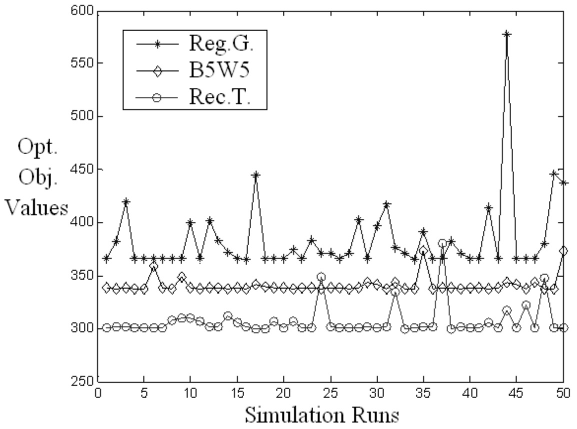


Fig. 14. Comparison of control performance based on different models

Table 6. Controlled objective values in simulations

	Reg. G.	B5W5	Rec. T.
Mean	383.80	340.68	307.05
St. Dev.	35.63	7.64	15.20

The values of the optimal objective function of 50 simulation runs are summarized in Fig. 14. The values of the optimal objective function based on "Reg. G." are larger than those based on reconfigured PLRT in most of the simulation runs, i.e., a better control performance is obtained with the reconfigured PLRT model. The "Rec. T." model has a better controlled performance than the "B5W5" model. However, a more complex model structure leads to a higher demand on computational efforts. The proposed reconfigured PLRT with reduced model complexity has less leaf nodes and sacrifices the control accuracy, but it still sufficiently meets the control requirements from an engineering perspective.

The mean and standard deviation of the optimal values are summarized in Table 6. There is an average of 11.24% and 20.00% reduction in objective value for the “B5W5” and “Rec. T.” compared to “Reg. G.”. The standard deviation of the values of the objective function is also reduced for the proposed “B5W5” model. The study indicates that the reconfigured PLRT is more effective in variation reduction than the standard regression models based on the proposed control strategy.

4. Conclusion

It is a challenging task to model the variations and their propagations in MMPs, especially when the relationships among process parameters and product quality variables are nonlinear. In this case, a PLRT model can be adopted that has high prediction accuracy and explicit interpretation in describing nonlinear data structure. However, it fails to illustrate the temporal order and inherent relationships among variables in a MMP.

This paper bridges the gap between the needs for advanced models for MMP variation reduction and the limitations of PLRT. An engineering-driven reconfiguration of the PLRT is proposed to convert the original model into an engineering compliant model. The reconfigured PLRT not only has the high prediction accuracy of the original tree structure, but also provides a feasible solution in determining the potential prediction models sequentially as the operations move from the upstream stages to the downstream stages. This sequential model selection procedure enables its capability in active compensation by implementing a feedforward control strategy. The model complexity is also reduced by analyzing the control accuracy of the models. A case study has been conducted in a real MWMP, which demonstrates better control performance by using the reconfigured PLRT model compared to that using a standard regression model.

References

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984) *Classification and Regression Trees*, Wadsworth, Belmont, CA.

Chaudhuri, P., Huang, M. C., Loh, W. Y. and Yao, R. (1994) Piecewise-polynomial regression trees. *Statistical Sinica*, **4**, 143-167.

Chipman, H., George, E. and McCulloch, R. (1998) Bayesian CART model search (with discussion). *Journal of the American Statistical Association*, **93**, 935-960.

- (2002) Bayesian Treed Models. *Machine Learning*, **48**, 299-320.

Denison, D., Adams, N., Holmes, C. and Hand, D. (2002) Bayesian partition modeling. *Computational Statistics and Data Analysis*, **38**, 475-485.

Djurdjanovic, D. and Zhu, J. (2005) Stream of variation based error compensation strategy in multistation manufacturing processes. In *Proceedings of the 2005 ASME International Mechanical Engineering Congress and Exposition*, paper IMECE2005-81550, 314-319, Orlando, FL.

Hastie, T., Tibshirani, R. and Friedman, J. (2001) *The Elements of Statistical Learning*, Springer, New York, NY.

Izquierdo, L. E., Shi, J., Hu, S. J. and Wampler, C. W. (2007) Feedforward control of multistage assembly processes using programmable tooling. *Transactions of NAMRI/SME*, **35**, 295-302.

Jin, J. and Ding, Y. (2004) Online automatic process control using observable noise factors for discrete part manufacturing. *IIE Transactions*, **36**, 899-911.

Jin, J. and Shi, J. (1999) State space modeling of sheet metal assembly for dimensional control. *ASME Transactions*, *Journal of Manufacturing Science and Engineering*, **121**, 756-762.

Jiao, Y. and Djurdjanovic, D. (2010) Joint allocation of measurement points and controllable tooling machines in multistage manufacturing processes. *IIE Transactions*, **42**, 703-720.

Joseph, R. (2003) Robust parameter design with feed-forward control. *Technometrics*, **45**, 284-292.

Kim, H., Loh, W. Y., Shih, Y. and Chaudhuri, P. (2007) Visualizable and interpretable regression models with good prediction power. *IIE Transactions*, **39**, 565-579.

Larsen, D. R., and Speckman, P. L. (2004) Multivariate regression trees for analysis of abundance data. *Biometrics*, **60**, 543-549.

Lee, S. K. (2006) On classification and regression trees for multiple responses and its application. *Journal of Classification*, **23**, 1, 123-141.

Loh, W. Y. (2002) Regression trees with unbiased variable selection and interaction detection. *Statistical Sinica*, **12**, 361-386.

Loh, W. Y., Chen, C. and Zheng, W. (2007) Extrapolation errors in linear model trees. *ACM Transactions on Knowledge Discovery in Data*, **1**,1-17.

Loh, W. Y. (2007) Regression by parts: fitting visually interpretable models with GUIDE in *Handbook of Data Visualization*, Chen, C., Hardle, W. and Unwin, A., Springer-Verlag, Berlin, Germany, pp. 447-468.

Mantripragada, R. and Whitney, D.E. (1999) Modeling and controlling variation propagation in mechanical assemblies using State Transition Models. *IEEE Transactions on Robotics and Automation*, **115**, 124-140.

Morgan, J. N. and Sonquist, J. A. (1963) Problems in the analysis of survey data, and a proposal. *Journal of American Statistical Association*, **58**, 415-434.

Segal, M. R. (1992) Tree-structured methods for longitudinal data. *Journal of the American Statistical Association*, **87**, 407-418.

Shi, J. (2006) *Stream of Variation Modeling and Analysis for Multistage Manufacturing Processes*, CRC Press, New York, NY.

Shi, J., Wu, C. F., Yang, X. and Zheng, H. (2005) Design of DOE-based automatic process controller for complex manufacturing processes. In *Proceedings of NSF DMII Grantees Conference*, Scottsdale, Arizona.

Shi, J. and Apley, D. (1998) A suboptimal N-Step-Ahead cautious controller for adaptive control applications. *ASME Transactions*, *Journal of Dynamic Systems, Measurement and Control*, **120**, 419-423.

Stutzle, T. (1998) *Local Search Algorithms for Combinatorial Problems-Analysis, Improvements and New Applications*, Ph.D thesis, FB Informatik, TU Darmstadt.

Zhong, J., Shi, J. and Wu, C. F. (2010) Design of DOE-based automatic process controller with consideration of model and observation. *IEEE Transactions on Automation Science and Engineering*, **7**, 266-273.

Acknowledgement:

The authors gratefully acknowledge the support of National Science Foundation Grant (NSF CMMI-0927574).

Appendix 1: Proof of Statement 1

The temporal order of the splitting variables is assumed as $Z_{1*} \prec= Z_{2*} \prec= \ldots \prec= Z_{L*}$. In the decomposition of the sub-regions of $g_i(\cdot)$ into $g_i^j(\cdot)$

$$y = \sum_{i=1}^{T} f_i\,(\boldsymbol{\eta}_i) I(g_i(Z_1,\ldots,Z_L)) = \sum_{i=1}^{T}\sum_{j=1}^{D_i} f_i\,(\boldsymbol{\eta}_i) I(g_i^j(Z_1,\ldots,Z_L)) \qquad \text{(A1)}$$

Since the decomposed sub-regions involve all splitting variables, the temporally complied variables can be substituted into $g_i^j(\cdot)$.

$$y = \sum_{i=1}^{T}\sum_{j=1}^{D_i} f_i\,(\boldsymbol{\eta}_i) I(g_i^j(Z_{1*},\ldots,Z_{L*})) \qquad \text{(A2)}$$

Since the splitting variables are temporally complied, the tree can be re-arranged into a temporally complied tree. Based on this tree, the merge of sub-regions follows the reverse temporal order. After the merge, sub-region $g_{i*}^j(\cdot)$ is the *j*-th region defined by a subset of $\{Z_{i*}\}$ for $f_i\,(\cdot)$, and there are T^* leaf nodes left:

$$y = \sum_{i*=1}^{T*} f_{i*}(\boldsymbol{\eta}_{i*}) I(g_{i*}(Z_{1*},\ldots,Z_{L*})) = y^* \qquad \text{(A3)}$$

The original PLRT is statistically equivalent as the re-ordered model in prediction. □

Appendix 2: Proof of Statement 2

Without loss of the generality, consider the case when there are two re-ordered models to be combined together, which are $y_1^* = \sum_{i=1}^{T_1^*} f_{i^*}^1(\boldsymbol{\eta}_{i^*}^1) I(g_{i^*}^1(Z_{1^*}^1,...,Z_{L_1^*}^1))$ and $y_2^* = \sum_{i=1}^{T_2^*} f_{i^*}^2(\boldsymbol{\eta}_{i^*}^2) I(g_{i^*}^2(Z_{1^*}^2,...,Z_{L_2^*}^2))$. If $g_i^n(\cdot)$ is decomposed by all possible splits of the splitting variables in both models, then the first model is

$$y_1^* = \sum_{i=1}^{T_1^*} f_{i^*}^1(\boldsymbol{\eta}_{i^*}^1) I(g_{i^*}^1(Z_{1^*}^1,...,Z_{L_1^*}^1)) = \sum_{i=1}^{T_1^*}\sum_{j=1}^{D_{1,i}^*} f_{i^*}^1(\boldsymbol{\eta}_{i^*}^1) I(g_{i^*}^{1,j}(Z_{1^*}^1,...,Z_{L_1^*}^1, Z_{1^*}^2,...,Z_{L_2^*}^2))$$

$$= \sum_{i=1}^{T_1^*}\sum_{j=1}^{D_{1,i}^*} f_{i^*}^1(\boldsymbol{\eta}_{i^*}^1) I(g_{i^*}^{1,j}(Z_1,...,Z_{L^*})) \qquad \text{(A4)}$$

Similarly, the second model is:

$$y_2^* = \sum_{i=1}^{T_2^*}\sum_{j=1}^{D_{2,i}^*} f_{i^*}^2(\boldsymbol{\eta}_{i^*}^2) I(g_{i^*}^{2,j}(Z_1,...,Z_{L^*})) \qquad \text{(A5)}$$

Since all possible splits of the splitting variables in both models are considered,

$$g_{i^*}^{1,j}(Z_1,...,Z_{L^*}) = g_{i^*}^{2,j}(Z_1,...,Z_{L^*}) \qquad \text{(A6)}$$

By following the procedure in step 3 of Algorithm 2, these two models can be presented as:

$$y_1^* = \sum_{i^*}^{T^*} f_{i^*}^1(\boldsymbol{\eta}_{i^*}^1) I(g_{i^*}^{comb}(Z_1,...,Z_{L^*})) \qquad \text{(A7)}$$

and

$$y_2^* = \sum_{i^*}^{T^*} f_{i^*}^2(\boldsymbol{\eta}_{i^*}^2) I(g_{i^*}^{comb}(Z_1,...,Z_{L^*})) \qquad \text{(A8)}$$

where $g_{i^*}^{comb}(Z_1,...,Z_{L^*})$ in both models are the same. Therefore, the combined model is the same as the original two re-ordered models in prediction. □

Biographies

Ran Jin is a Ph.D. student in H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. He received B.Eng. in Electronics Information Engineering at the Tsinghua University, Beijing, in 2005, and M.S. in Industrial Engineering and M.A. in Statistics at the University of Michigan, Ann Arbor, in 2007 and 2009 respectively. His research interests include data mining, engineering knowledge enhanced statistical modeling of complex systems, process monitoring, diagnosis and control.

Dr. Jianjun Shi is Carolyn J. Stewart Chair Professor at H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology. Before joined Georgia Tech in 2008, he was the G. Lawton and Louise G. Johnson Professor of Engineering at the University of Michigan. He got his B.S. and M.S. in Electrical Engineering at the Beijing Institute of Technology in 1984 and 1987 respectively, and his Ph.D. in Mechanical Engineering at the University of Michigan in 1992. Professor Shi's research interests focus on the fusion of advanced statistics, signal processing, control theory, and domain knowledge to develop methodologies for modeling, monitoring, diagnosis, and control for complex systems in a data rich environment. Professor Shi is the founding chairperson of the Quality, Statistics and Reliability (QSR) Subdivision at INFORMS. He currently serves as the Focus Issue Editor of IIE Transactions on Quality and Reliability Engineering. He is a Fellow of the Institute of Industrial Engineering (IIE), a Fellow of American Society of Mechanical Engineering (ASME), and a Fellow if Institute of Operations Research and Management Science (INFORMS), and also a life member of ASA.