Auto-Switch Gaussian Process Regression-based Probabilistic Soft Sensors for Industrial Multi-Grade Processes with Transitions

5

Yi Liu¹, Tao Chen², Junghui Chen^{3*}

10 ¹Engineering Research Center of Process Equipment and Remanufacturing, Ministry
of Education, Institute of Process Equipment and Control Engineering, Zhejiang
University of Technology, Hangzhou, 310014, People's Republic of China

15 ²Department of Chemical and Process Engineering, University of Surrey, Guildford
GU2 7XH, UK

³Department of Chemical Engineering, Chung-Yuan Christian University, Chung-Li,
Taiwan, 320, Republic of China

20

Revision Submitted to Industrial & Engineering Chemistry Research

August, 2014

25

Abstract: Prediction uncertainty has rarely been integrated into traditional soft sensors in industrial processes. In this work, a novel auto-switch probabilistic soft sensor modeling method is proposed for online quality prediction of a whole industrial multi-grade process with several steady-state grades and transitional modes. Several single Gaussian process regression (GPR) models are first constructed for each steady-state grade. A new index is proposed to evaluate each GPR-based steady-state grade model. For the online prediction of a new sample, a prediction variance-based Bayesian method is proposed to explore the reliability of existing GPR-based steady-state models. The prediction can be achieved using the related steady-state GPR model if its reliability using this model is large enough. Otherwise, the query sample can be treated as in transitional modes and a local GPR model in a just-in-time manner is online built. Moreover, to improve the efficiency, detailed implementation steps of the auto-switch GPR soft sensors for a whole multi-grade process are developed. The superiority of the proposed method is demonstrated and compared with other soft sensors in an industrial process in Taiwan in terms of online quality prediction.

Key words: Bayesian inference; Gaussian process regression; multi-grade process; transition; prediction variance; just-in-time learning;

1. Introduction

With the rapid development of computer and communication technologies, process data have become widely available in many chemical plants. As a result, increasing data-driven soft sensor modeling methods have been applied to inferring/predicting product qualities that are difficult to measure online.¹⁻²⁹ Existing common methods include partial least squares (PLS) and other multivariate regression approaches,^{5,6,8-10} various neural networks (NN),¹¹⁻¹⁵ fuzzy systems,¹⁶ support vector regression (SVR) and least squares SVR (LSSVR),¹⁷⁻²¹ and Gaussian process regression (GPR).²²⁻²⁷ One main advantage of these soft sensor models is that they can generally be developed in a relatively simple manner without substantial understanding of the phenomenology involved.

In recent years, multi-grade processes have played a significant role in the polymer and fine chemical industries. Generally, frequent change of operating conditions is required in the production of multi-grade products. Nevertheless, product qualities, such as melt index (MI), are often evaluated

off-line and infrequently. Consequently, grade changeover typically is a manual operation in many industrial polymer plants, and it results in relatively large settling time and off-grade materials.²⁸⁻³² To overcome the problem, various soft sensors have been applied to quality prediction in multi-grade processes,^{8,16,20,25,28,29,32} in which the products are commonly short-lived and of small volume with a limited set of modeling samples. In addition, it is common to find that there are nonlinear relationships between product qualities and the operating conditions.²⁸⁻³² Therefore, SVR and GPR-based nonlinear soft sensors have attracted more attentions recently.¹⁷⁻²⁷ Compared with SVR, one main advantage of the GPR-based model is that it can simultaneously provide the probabilistic information for its prediction.²²

In process modeling and control areas, the multiple model approach is suitable for dealing with the nonlinear process with a wide operating range. It has received a great deal of attention in recent years because of its success in converting complex problems into simpler sub-problems.³³⁻³⁷ Generally, there are several steady-state grades and corresponding transitional modes between these grades in a whole multi-grade process. However, compared with the most existing soft sensors focusing on multiple steady-state grade processes, only a few methods have been applied to transitional modes.^{28,29,32} Ge et al.²⁵ proposed a weighted GPR model and showed more accurate performance than a single global GPR model for MI prediction in the polypropylene production process. Nevertheless, transitional modes were not considered in their work. Compared to steady-state grades, the prediction performance of soft sensors generally degrades during transitional modes because of different operating conditions and fewer modeling samples.^{28,29} Recently, Yu²⁷ proposed a mixture GPR soft sensor model for a better description of shifting in multi-mode processes though the transient dynamics between the two operating modes were neglected.

In industrial practice, only using a single model is insufficient to capture all the process characteristics, especially those in transitions. In our recent work, an integrated just-in-time LSSVR soft sensor was developed for the whole multi-grade process with transitions.³² MI in transitional modes can be better predicted using just-in-time-based local LSSVR models. Nevertheless, as aforementioned, the prediction variance cannot be provided using the SVR/LSSVR-based related models. To our best knowledge, the combination of just-in-time learning and the GPR modeling

method has rarely been reported, especially for the purpose of soft sensor development. Therefore, a just-in-time GPR (JGPR) nonlinear soft sensor for the online quality prediction is proposed in this paper. As for the online prediction of a query sample, its prediction variance is also evaluated. It can provide additional information for local models.

5

Prediction uncertainty has rarely been integrated into traditional soft sensors for industrial processes. As for a whole multi-grade process of several steady-state grades and corresponding transitions, it is important to evaluate the status of a sample. Additionally, it is critical to judge which model is the most suitable for online prediction of the current query sample. This is also important mainly because each soft sensor has its reliable domain. However, this issue was less investigated in the previous GPR-based soft sensors. To this end, the prediction variance of GPR is further utilized to assess the uncertainty of soft sensors in different modes. Furthermore, the modeling strategy of prediction variance-based auto-switch soft sensors for the whole multi-grade process is proposed in this work.

10

The remainder of this paper is organized as follows. The GPR-based modeling method is first described in Section 2. After the description of a whole multi-grade process, GPR and JGPR-based soft sensors for steady-state grades and transitions are proposed and analyzed in this section, respectively. The prediction variance-based Bayesian inference for the GPR model selection is presented in Section 3. Also, detailed implementation of the proposed auto-switch soft sensors in the whole multi-grade process is developed in this section. The proposed method is evaluated by the MI prediction in an industrial process in Taiwan in Section 4. It is also compared with other related approaches. Finally, the conclusion is drawn in Section 5.

20

2. GPR and JGPR soft sensor models for multi-grade processes

2.1. Basic GPR-based soft sensor models

25

Generally, development of a soft sensor using the GPR framework can be described as a problem whose aim is to learn a model f that approximates a training set $\{\mathbf{S}\} = \{\mathbf{X}, \mathbf{Y}\}$, where $\{\mathbf{X}\} = \{\mathbf{x}_i\}_{i=1}^N$ and $\{\mathbf{Y}\} = \{y_i\}_{i=1}^N$ are the input and output datasets with N samples, respectively. A GPR model

provides a prediction of the output variable for an input sample through Bayesian inference. For an output variable of $\mathbf{Y} = (y_1, \dots, y_N)^T$, the GPR model is the regression function with a Gaussian prior distribution and zero mean, or in a discrete form²²⁻²⁴

$$\mathbf{Y} = (y_1, \dots, y_N)^T \sim G(0, \mathbf{C}) \quad (1)$$

5

where $\mathbf{C}$ is the $N \times N$ covariance matrix with the ij -th element defined by the covariance function, $\mathbf{C}_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$. A common covariance function can be defined as²²

$$C(\mathbf{x}_i, \mathbf{x}_j) = a_0 + a_1 \sum_{d=1}^D x_{id} x_{jd} + v_0 \exp\left(-\sum_{d=1}^D w_d (x_{id} - x_{jd})^2\right) + \delta_{ij} b \quad (2)$$

10 where x_{id} is the d th component of the vector $\mathbf{x}_i$. $\delta_{ij} = 1$ if $i = j$; otherwise, it is equal to zero.

$\boldsymbol{\theta} = [a_0, a_1, v_0, w_1, \dots, w_D, b]^T$ is the hyper-parameters vector defining the covariance function. Generally, the hyper-parameters must be non-negative to ensure that the covariance matrix is non-negative definite. As depicted in Eq. (2), the first two terms denote a constant bias and a linear correlation term, respectively. The exponential term takes into account the potentially strong correlation between the

15 outputs for nearby inputs. Additionally, the term b captures the random error effect. By combining both linear and nonlinear terms in the covariance function, the GPR model is capable of handling both linear and nonlinear processes.²²⁻²⁴ Other forms of covariance functions can be referred to Rasmussen and Williams.²²

The training of a GPR model can determine the values of the hyper-parameters $\boldsymbol{\theta}$. Adopting a Bayesian approach, the hyper-parameters $\boldsymbol{\theta}$ can be estimated by maximization of the following log-likelihood function:²²

$$L(\boldsymbol{\theta}) = -\frac{1}{2} \log(\det(\mathbf{C})) - \frac{1}{2} \mathbf{Y}^T \mathbf{C}^{-1} \mathbf{Y} - \frac{N}{2} \log(2\pi) \quad (3)$$

5

This optimization problem can be solved using the derivative of the log-likelihood with respect to each hyper-parameter given as²²

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = -\frac{1}{2} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \boldsymbol{\theta}} \right) + \frac{1}{2} \mathbf{Y}^T \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \boldsymbol{\theta}} \mathbf{C}^{-1} \mathbf{Y} \quad (4)$$

10 where $\frac{\partial \mathbf{C}}{\partial \boldsymbol{\theta}}$ can be obtained from the covariance function. Detailed implementations for training a GPR model can be referred to Rasmussen and Williams.²² It should also be noted that the main computational load for training a GPR model is about $O(N^3)$, which is feasible for a moderate size of training data sets (less than several thousands) on a conventional computer. For much larger data sets, sparse training strategies may be required to reduce the computational burden. A comprehensive
15 investigation of sparse GPR is beyond the scope of this work and it can be referred to the literature.^{38,39}

Finally, the GPR model can be obtained once $\boldsymbol{\theta}$ is determined. As for a new test sample $\mathbf{x}_t$, the predicted output of y_t is also Gaussian with mean ($\hat{y}_t$) and variance ($\sigma_{\hat{y}_t}^2$), calculated as follows:²²

$$\hat{y}_t = \mathbf{k}_t^T \mathbf{C}^{-1} \mathbf{Y} \quad (5)$$

$$\sigma_{y_i}^2 = k_i - \mathbf{k}_i^T \mathbf{C}^{-1} \mathbf{k}_i \quad (6)$$

where $\mathbf{k}_i = [C(\mathbf{x}_i, \mathbf{x}_1), C(\mathbf{x}_i, \mathbf{x}_2), \dots, C(\mathbf{x}_i, \mathbf{x}_N)]^T$ is the covariance vector between the new input and the training samples, and $k_i = C(\mathbf{x}_i, \mathbf{x}_i)$ is the covariance of the new input. In summary, the vector $\mathbf{k}_i^T \mathbf{C}^{-1}$ denotes a smoothing term which weights the training outputs to make a prediction for the new input sample $\mathbf{x}_i$.²² In addition, Eq. (6) provides a confidence level on the model prediction, which is an appealing property of the GPR method.

2.2. GPR-based soft sensor for steady-state grades

A simplified flowchart of a nonlinear continuous stirred tank reactor (CSTR) process with multi-grade characteristics is shown in Fig. 1. It is common to use the CSTR process to produce products with different grades in the polymerization industry. By changing the operating conditions of the reactor or by using a different type of catalyst, the product produced from the process can be switched from one grade to another.^{8,12,28,30,31} Generally, the collected modeling data for a whole multi-grade process consists of G operation grades with related sets represented by $\{\mathbf{S}_g\}_{g=1, \dots, G} = \{\mathbf{X}_g, \mathbf{Y}_g\}_{g=1, \dots, G}$, where $\mathbf{X}_g = \{\mathbf{x}_{g,i} \in R^m\}_{i=1, \dots, n_g}$ and $\mathbf{Y}_g = \{y_{g,i} \in R^r\}_{i=1, \dots, n_g}$ are the input and output dataset of the g th operation grade with n_g samples, respectively. For simplicity, all transitional samples between steady-state grades are noted as $\mathbf{S}_t = \{\mathbf{X}_t, \mathbf{Y}_t\}$ with n_t samples. When the operating conditions are changing from “steady-state grade i ” to “steady-state grade j ”, the samples collected in this period are noted as transitional data. Due to the time delay analysis of product qualities and the complicated dynamics of operating inputs, there may be some overlaps between different steady-state grades and transitions.^{8,28,32} As simply illustrated in Fig. 1, the CSTR process has three

steady-state grades and corresponding transitional modes. The number of modeling samples for each steady-state grade is different. Additionally, there are often fewer transitional samples than steady-state samples.

5 (Fig. 1 goes here.)

Generally, each steady-state grade S_g has its special operating conditions and characteristics different from other grades. Consequently, for better description of process characteristics, several single GPR models, denoted as $GPR_{M_g}, g=1, \dots, G$, can be constructed offline for different steady-state grades using aforementioned formulations, i.e., Eqs. (1)~(6). In this work, the modes of training samples in the multi-grade production process are known based on the lab analysis results. Therefore, several single GPR models can be simply established. As for a more complex process with several unknown modes, the multiple model approaches should be applied to dividing the modeling samples into several sub-classes as a preprocessing step.³³⁻³⁷

15

As for a training sample $\mathbf{x}_i$ using the g th GPR model, its relative-root prediction variance (RPV) $v_{M_g, \mathbf{x}_i}$ can be first defined as follows:

$$v_{M_g, \mathbf{x}_i} = \frac{\sqrt{\sigma_{\hat{y}_i}^2}}{y_i} \times 100\% = \frac{\sigma_{\hat{y}_i}}{y_i} \times 100\%, \quad i = 1, \dots, n_g \quad (7)$$

20 This item can be utilized to describe the performance of relative prediction uncertainty for a training sample in this steady-state grade. As for a trained model, if the values of two samples (y_i and y_j) are

almost the same, a smaller value of $v_{M_g, \mathbf{x}_i}$ means a smaller prediction uncertainty. Consequently, for the training samples in a special steady-state grade, a small value of $v_{M_g, \mathbf{x}_i}$ means better prediction performance than a large one mainly because the values of $y_i, i = 1, \dots, n_g$ in this grade are in a certain range.

5

Moreover, an evaluated item of RPV for all the training samples in each special steady-state grade is proposed and defined as follows:

$$E_{M_g} = \frac{1}{n_g} \sum_{i=1}^{n_g} v_{M_g, \mathbf{x}_i}, g = 1, \dots, G \quad (8)$$

10 The item of E_{M_g} denotes the mean value of RPV ($v_{M_g, \mathbf{x}_i}, i = 1, \dots, n_g$). This new item can provide additional information for evaluation of the accuracy of these trained GPR models $\text{GPR}_{M_g}, g = 1, \dots, G$ for related steady-state grades. Generally, a smaller value of E_{M_g} means the better prediction performance with the smaller uncertainty for $\text{GPR}_{M_g}, g = 1, \dots, G$. Consequently, after several GPR models are trained, one can simply judge which steady-state grade in a whole

15 multi-grade process can be modeled using a more reliable GPR model.

2.3. JGPR-based local model development for transitions

As for complicated multi-grade processes, the direct application of a global or fixed model may not be enough mainly because specifying the structure of a global/fixed model is often difficult. Another

20 limitation is that it is difficult to quickly update a global/fixed model when the process dynamics are moved away from the nominal operating area. Additionally, as for industrial transitional modes, the

training samples are so limited and process characteristics are more complicated than the steady-state grades.^{28,29,32} Therefore, it is unsuitable and more difficult to establish a special GPR model for transitions beforehand.

5 To alleviate these problems and construct the local models automatically, the just-in-time (JIT) method has been developed as an attractive alternative to nonlinear chemical process modeling and control.⁴⁰⁻⁴⁹ Several JIT-based nonlinear soft sensors have been proposed previously.⁴²⁻⁴⁶ However, they are all deterministic models (e.g., JIT-based SVR/LSSVR models simply denoted as JSVR and JLSSVR, respectively⁴⁶) rather than probabilistic ones. In literature, few probabilistic soft sensors have
10 been applied to multi-grade processes with transitional modes. In this section, a JGPR-based online modeling method is proposed for better description of those transitional modes. Generally, for a query sample $\mathbf{x}_q$, there are three main steps to construct a JGPR model online:

JGPR Step 1: Find out relevant samples to form a similar set $\mathbf{S}_{\text{sim}}$ in the database $\mathbf{S}$ using some
15 defined similarity criteria.

JGPR Step 2: Online construct a JGPR model $f_{\text{JGPR}}(\mathbf{x}_q)$ using the relevant set $\mathbf{S}_{\text{sim}}$ and the aforementioned formulations, i.e., Eqs. (1)~(6).

JGPR Step 3: Obtain the predicted output of y_q , i.e., the mean ($\hat{y}_q$) and its variance ($\sigma_{\hat{y}_q}^2$), for the current query sample $\mathbf{x}_q$, and then discard the JGPR model $f_{\text{JGPR}}(\mathbf{x}_q)$.

20

This is the basic framework of JGPR online modeling approach. As for a new query sample, a new JGPR model can be built with the same three-step procedures. Generally, the Euclidean distance-based

similarity is the most commonly utilized index.^{40,41} The similarity factor (SF) s_{qi} between the query sample $\mathbf{x}_q$ and the sample $\mathbf{x}_i$ in the dataset is defined below:^{40,41}

$$s_{qi} = \exp(-d_{qi}) = \exp(-\|\mathbf{x}_i - \mathbf{x}_q\|), i = 1, \dots, N \quad (9)$$

5 where d_{qi} is the distance similarity between $\mathbf{x}_q$ and $\mathbf{x}_i$ in the dataset. The value of s_{qi} is bounded between 0 and 1 and when s_{qi} approaches to 1, $\mathbf{x}_q$ resembles $\mathbf{x}_i$ closely.

Using the similarity criterion in Eq. (9), the n ($n_{\min} \leq n \leq n_{\max}$) similar samples should be selected for constructing a JGPR model. Generally, the $n_{\max}$ similar samples can be ranked according to the degree
10 of the similarity. A cumulative similarity factor (CSF) S_{qn} can be adopted as follows:⁴⁶

$$S_{qn} = \frac{\sum_{i=1}^n s_{qi}}{\sum_{i=1}^{n_{\max}} s_{qi}}, n \leq n_{\max} \quad (10)$$

which denotes the cumulative similarity of the n most similar samples compared to the relevant set $\mathbf{S}_{\text{sim}}$. The CSF index can simply compute the cumulative similarity and then it can determine the n most
15 similar samples.⁴⁶ Traditionally, it is difficult to determine the range of $[n_{\min}, n_{\max}]$ beforehand. Especially for a process with multiple grades, the relevant sets of a query sample for different grades are different. As an alternative approach, the search of $[n_{\min}, n_{\max}]$ can be simply substituted by the choice of S_{qn} ; e.g., $S_{qn} = 0.8$. This means 80 % of the most similar samples have been selected. Consequently, construction of the relevant set $\mathbf{S}_{\text{sim}}$ using the CSF index is a practical and efficient
20 method.³²

The comparisons of JGPR and JLSSVR/JSVR online modeling methods are listed in Table 1. Compared with JLSSVR/JSVR online modeling methods, two advantages of the JGPR-based soft sensor can be obtained. One is that the probabilistic information can be provided for its prediction. The

other, the modeling procedures of JGPR through Bayesian inference, is simpler and more straightforward. As for JLSSVR/JSVR modeling methods, the kernel function and the related parameters should be carefully chosen. This is not an easy task, especially for those complicated industrial processes.

5

(Table 1 goes here.)

3. Auto-switch soft sensors for a whole multi-grade process with transitions

As for a whole multi-grade process, several GPR-based steady-state models can be built in a relatively simple way. For a query sample $\mathbf{x}_q$, it is important to evaluate to which model it is most suitable. To account for this method, in Section 3.1, a prediction variance-based Bayesian method is proposed to explore the reliability of existing GPR-based steady-state models. In Section 3.2, detailed implemented steps of the auto-switch soft sensors for a multi-grade process with transitions are developed. The prediction can be obtained using the corresponding steady-state GPR model if its probability using this special model is large enough. Otherwise, $\mathbf{x}_q$ is considered located in transitional modes and a JGPR-based local model is built online.

3.1. Prediction variance-based Bayesian inference for GPR model evaluation

In the polymer industry, a multi-grade process is often characterized by several steady-state grades and related overlappings between them.^{28,29,32} To evaluate to which local GPR model the query sample $\mathbf{x}_q$ more suitably belongs, a prediction variance-based Bayesian inference is proposed to determine the probability of $\mathbf{x}_q$ adopting each $\text{GPR}_{M_g}, g=1, \dots, G$ model, i.e., $P(M_g | \mathbf{x}_q), g=1, \dots, G$, which is formulated as follows:

$$P(M_g | \mathbf{x}_q) = \frac{P(\mathbf{x}_q | M_g)P(M_g)}{P(\mathbf{x}_q)} = \frac{P(\mathbf{x}_q | M_g)P(M_g)}{\sum_{g=1}^G [P(\mathbf{x}_q | M_g)P(M_g)]}, \quad g=1, \dots, G \quad (11)$$

25

where $P(M_g), g=1, \dots, G$, and $P(\mathbf{x}_q | M_g), g=1, \dots, G$ are the prior probability and the conditional

probability, respectively. To obtain the posterior probability value, these two terms at the right side of Eq. (11) should be calculated. Without any process or expert knowledge, the prior probability for each local GPR model suitable for its operation mode can be simply defined as:^{25,32}

$$P(M_g) = \frac{n_g}{n}, \quad g = 1, \dots, G \quad (12)$$

5

To determine the other terms in Eq. (11), first, the RPV item (defined in Eq. (7)) of this query sample for each steady-state GPR model can be further modified as follows:

$$\hat{v}_{M_g, \mathbf{x}_q} = \frac{\sigma_{\hat{y}_{q,g}}}{\hat{y}_{q,g}}, \quad g = 1, \dots, G \quad (13)$$

10 where the actual value of y_q is unknown, so it is substituted by its prediction using the related GPR _{M_g} , $g = 1, \dots, G$ models, noted as $\hat{y}_{q,g}$. The value of $\sigma_{\hat{y}_{q,g}}$ is relatively large if the query sample $\mathbf{x}_q$ is predicted using an unsuitable model. Consequently, a larger value of $\hat{v}_{M_g, \mathbf{x}_q}$ means a larger uncertainty by adopting this special GPR model for online prediction. Due to this reason, without loss of generality, the conditional probability $P(\mathbf{x}_q | M_g)$ can be defined based on an inverse relationship

15 of $\hat{v}_{M_g, \mathbf{x}_q}$.

$$P(\mathbf{x}_q | M_g) = \frac{1}{\hat{v}_{M_g, \mathbf{x}_q}}, \quad g = 1, \dots, G \quad (14)$$

Consequently, Eq. (11) becomes

$$P(M_g | \mathbf{x}_q) = \frac{n_g}{\hat{v}_{M_g, \mathbf{x}_q} \sum_{g=1}^G (n_g / \hat{v}_{M_g, \mathbf{x}_q})}, \quad g = 1, \dots, G \quad (15)$$

20

Based on the probabilistic analysis approach, using the prediction variance of steady-state grade

models can determine the probability of $\mathbf{x}_q$ for each individual GPR model suitable in its operational mode. Note that how to treat the transitional modes using probabilistic soft sensors has not been considered before. It is relatively simple to establish the steady-state models while predicting a query sample using suitable models is difficult, especially those samples in the transitional or the new modes with complicated characteristics. Therefore, the steady-state GPR models can be trusted if they have a larger probability; i.e. more reliability. Otherwise, a JGPR model can be built online using the available information.

3.2. Auto-switch GPR modeling strategy

In this section, the main implementations to the whole multi-grade process can be formulated using the aforementioned probabilistic modeling methods and Bayesian analysis. The method is simply denoted as AGPR (auto-switch GPR) because it can select its most reliable model for a test sample with unknown modes.

First, a threshold η ($\eta > 0$) is defined to evaluate to which steady-state GPR model the query sample $\mathbf{x}_q$ is most suitable. If Eq. (16) is satisfied, assign $\mathbf{x}_q$ to the most suitable model with the largest value of probability, i.e., $\arg \max_{\text{GPR}_{M_g}, g=1, \dots, G} P(M_g | \mathbf{x}_q)$, for prediction,

$$\max_{g=1, \dots, G} P(M_g | \mathbf{x}_q) > \eta \quad (16)$$

Otherwise, $\mathbf{x}_q$ should be treated as located in some transitional modes or out of any steady-state grade to an extent. In this situation, using existing GPR models for prediction may introduce larger variance. Alternatively, a JGPR model should be constructed online for prediction using the most similar samples around $\mathbf{x}_q$.

The threshold η denotes a trade-off parameter of the soft sensor model switch. Generally, in industrial processes, its range, $0.6 \leq \eta < 1$, can be chosen because the selected $\text{GPR}_{M_g}, g=1, \dots, G$ model can be trusted if its probability is at least larger than 0.6. In this range, if the value of η is

smaller, more query samples are predicted using its suitable GPR_{M_g} model. On the contrary, if the value of η is very large, e.g. $\eta = 0.9$, most of the query samples are predicted using the JGPR model. Note that Eq. (16) is always violated once $\eta = 1$ (or $\eta \rightarrow 1$). In this situation, for every query sample $\mathbf{x}_q$, a local JGPR model is online constructed.

5

As for the online construction of a JGPR model, the computational load may be cumbersome if the candidate database $\mathbf{S}$ for search becomes larger. Actually, for multi-grade processes, the number of samples for each grade is different; e.g., $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_3$ often have uneven sizes.^{28,29,32} Therefore, once a JGPR model is applied (i.e., if Eq. (16) is violated), another small threshold $\eta_{\min}$ ($0 < \eta_{\min} \ll \eta$) can be defined to determine which model $\mathbf{x}_q$ is most unsuitable. If Eq. (17) is satisfied, $\mathbf{x}_q$ has less similarity to the related steady-state grade with the smallest value of probability.

10

$$\min_{g=1, \dots, G} P(M_g | \mathbf{x}_q) \ll \eta_{\min} \quad (17)$$

In this situation, for online construction of a JGPR model, the candidate set $\mathbf{S}$ for search becomes smaller because the most dissimilar steady-state grade has been excluded. Consequently, the computational load can be reduced.

15

Remark 1: In many production processes, the intended grade of the product is generally known. However, the transitional modes often require drastic and simultaneous changes in different operating inputs, resulting in relatively long settling times. As aforementioned, because of the time delay of the product quality analysis, the actual modes of some overlapping samples are unknown. Consequently, it is important to select or construct a suitable prediction model for the query sample. Moreover, there are several steady-state grades in an industrial production process. Eq. (17) should be an alternative to exclude the most dissimilar samples so as to better determine the candidate set.

20

25

In summary, the proposed AGPR soft sensor modeling method for the whole multi-grade process is illustrated in Fig. 2. The step-by-step procedures of the AGPR approach are summarized as below.

AGPR Step 1: The process input and output samples of the whole multi-grade process, including several steady-state grades $\{\mathbf{S}_g\}_{g=1,\dots,G} = \{\mathbf{X}_g, \mathbf{Y}_g\}_{g=1,\dots,G}$ and the transitional set $\mathbf{S}_t = \{\mathbf{X}_t, \mathbf{Y}_t\}$, are collected.

AGPR Step 2: Several single GPR models, denoted as $\text{GPR}_{M_g}, g=1,\dots,G$, are trained offline for different grades using Eqs. (1)~(6). The item of E_{M_g} denoting the mean value of RPV for $\text{GPR}_{M_g}, g=1,\dots,G$ can also be obtained using Eq. (8).

AGPR Step 3: For a query sample $\mathbf{x}_q$, its probability can be determined using Eq. (15) when using each steady-state GPR model (i.e., $P(M_g | \mathbf{x}_q), g=1,\dots,G$).

AGPR Step 4: If Eq. (16) is satisfied, $\mathbf{x}_q$ is assigned to the most reliable steady-state grade model with the largest value of probability; i.e., $\max_{g=1,\dots,G} P(M_g | \mathbf{x}_q) > \eta$, for online prediction.

AGPR Step 5: Otherwise, a JGPR model is online constructed for prediction using the most similar samples around $\mathbf{x}_q$. If meanwhile Eq. (17) is satisfied, the candidate set becomes smaller because the most dissimilar steady-state grade can be excluded.

AGPR Step 6: For online construction of a JGPR model, first find out the relevant set $\mathbf{S}_{\text{sim}}$ using Eqs. (9) and (10). Finally, the predicted output of y_q , i.e., the mean ($\hat{y}_q$) and its variance ($\sigma_{\hat{y}_q}^2$), can be simultaneously obtained.

(Fig. 2 goes here.)

In the above modeling steps, a query sample $\mathbf{x}_q$ can automatically select/construct the most suitable model based on its probability. This probability is mainly calculated using the prediction variances of GPR-based steady-state grade models. Compared with SVR-based deterministic models, prediction variances provide useful information for better description of the model reliability. The proposed AGPR method is distinguished with the integrated approach denoted as IJ-LSSVR³² in two folds. One is that AGPR is a probabilistic soft sensor while the latter is not. The other is that AGPR assigns a query sample to its reliable model. However, the IJ-LSSVR method assigns the query sample to its spatial grade of multivariables.³²

4. Application to an industrial polyethylene process

The proposed AGPR method can be considered as an auto-switch soft sensor with probabilistic prediction information for complicated multi-grade processes with transitional modes. In this section, the AGPR soft sensor modeling method is explored by online predicting MI of an industrial polyethylene production process in *Formosa Plastics Corporation* in Taiwan. All the data samples have been collected from daily process records and the corresponding laboratory analysis. The product quality in steady-state grades is commonly analyzed in the lab once a day. When the grade of the production changes, the product quality is sampled for assay about 3~4 times a day.³² Consequently, without online analyzers for MI, off-grade products and materials would be produced inevitably in industrial polyethylene production processes.

As prior requirements, the reliable sensor measurements and data collections play important roles in process modeling.² After simply preprocessing the modeling set with a 3-sigma criterion, most of the missing values and outliers have been removed. Finally, more than 300 samples of three steady-state grades and related transitions collected in a product line from 2009 to 2011 are investigated in this study.³² Half of them are for training and the rest are for testing. The number of test samples for each steady-state grade and transitions are 15 (S_1), 22 (S_2), 81 (S_3) and 37 (S_t), respectively. The simulation environment in this case is MatLab V2009b with CPU main frequency 2.3 GHz and 4 GB memory.

Two common performance indices, including root-mean-square error (RMSE) and relative RMSE (simply noted as RE) can be adopted to assess the prediction performance. Compared to RMSE, RE is more suitable to multi-grade processes because the values of MI in each grade are different.³² Consequently, the performance index of RE is considered as follows,

$$RE = \sqrt{\sum_{q=1}^l \left(\frac{\hat{y}_q - y_q}{y_q} \right)^2 / l} \quad (18)$$

where $\hat{y}_q$ denotes the prediction term of y_q and l is the number of test samples. Additionally, the item of E_{M_g} (i.e., the mean value of RPV) is also investigated to evaluate the trained GPR-based soft

sensor models. As for test samples, the item of E_{M_g} is modified as E_M below

$$E_M = \frac{1}{l} \sum_{q=1}^l v_{M, \mathbf{x}_q} = \frac{1}{l} \sum_{q=1}^l \frac{\sigma_{\hat{y}_q}}{y_q} \times 100\% \quad (19)$$

where $v_{M, \mathbf{x}_q}$ denotes the RPV value of $\mathbf{x}_q$ using corresponding soft sensor models, including AGPR
5 and JGPR models.

4.1. Analysis of offline trained GPR models

In this section, the offline modeling and analysis of GPR models are investigated. Three GPR models,
noted as $\text{GPR}_{M_g}, g = 1, \dots, 3$, have been trained for related steady-state grades. They are shown in Figs.

10 3a-3c, respectively. The upper line and lower line show $\hat{y}_i + \sigma_{\hat{y}_i}$ and $\hat{y}_i - \sigma_{\hat{y}_i}$, respectively.

Additionally, as shown in Fig. 3d, a single GPR model noted as GPR_{M_t} has been trained to fit the
transitional samples. RE (defined in Eq.(18)), the range of RPV values and E_{M_g} (defined in Eqs. (7)
and (8)) indices of the trained offline GPR models for three steady-state grades and related transitional
modes are listed in Table 2.

15

(Fig. 3a goes here.)

(Fig. 3b goes here.)

(Fig. 3c goes here.)

(Fig. 3d goes here.)

20

(Table 2 goes here.)

From the fitting results shown in Figs. 3a-3c and the RE values in Table 2, one can see that the
 GPR_{M_1} model achieves smaller training errors than GPR_{M_2} and GPR_{M_3} . However, a trained soft
sensor model with limited samples does not always guarantee a good prediction performance in its

application, especially to those complicated industrial processes with uncertainties. The proposed RPV and E_{M_g} items can provide additional information for evaluation of soft sensors. As shown in Figs. 3a-3c and Table 2, the ranges of RPV values of $\text{GPR}_{M_g}, g=1, \dots, 3$, for three steady-state grades are relatively narrow. This means the samples in each steady-state grade can be trained using a single GPR model. Among them, the range of RPV values of GPR_{M_3} is the most narrow. Moreover, the E_{M_3} index of GPR_{M_3} is the smallest of all. This indicates the uncertainty of GPR_{M_3} is the smallest. It should also be noted that the size of $\mathbf{S}_3$ is much larger than $\mathbf{S}_1$ and $\mathbf{S}_2$. Compared to $\mathbf{S}_3$, the modeling information for $\mathbf{S}_1$ and $\mathbf{S}_2$ is not enough. Therefore, the GPR_{M_3} model is more reliable and confident than the other models.

At first glance, the fitting results for transitional samples shown in Fig. 3d are not bad. However, the RE index for the transitional modes is very large. Additionally, the range of RPV values for GPR_{M_t} is very wide, indicating these samples have lots of different characteristics. The E_{M_t} value in the transitional model is much larger than those in steady-state models. These results validate that the characteristics of transitional samples are more complicated than those of the steady-state grades. Consequently, it is not suitable to train a single GPR model for these transitional samples. Unlike the traditional soft sensors, e.g., PLS, NN, SVR/LSSVR,⁵⁻²¹ the GPR-based soft sensor can provide probabilistic information for its prediction. Moreover, as for soft sensor modeling of multi-grade processes in the transitional modes, the RPV and E_{M_g} items can provide additional information to evaluate the model uncertainty. These properties are attractive for process modeling.

4.2. AGPR online prediction results

In this section, AGPR is applied to online MI prediction in the industrial polyethylene process with multiple grades and the related transitions are investigated. As aforementioned, three steady-state GPR models have been trained for the corresponding product grades, denoted as $\text{GPR}_{M_g}, g=1, \dots, 3$. For online prediction of a query sample, the AGPR can auto-select its suitable model.

There are two parameters for AGPR, η and $\eta_{\min}$, respectively. The value of η is a threshold that

determines which model the query sample $\mathbf{x}_q$ adopts, e.g., $\text{GPR}_{M_g}, g = 1, \dots, 3$ or JGPR. The range of η can be chosen as $0.6 \leq \eta < 1$ for industrial processes. Generally, a smaller value of η suggests that the GPR_{M_g} model be more suitable for prediction. On the contrary, a larger value of η means the JGPR model performs better. Consequently, a larger value of η requires more JGPR models with heavier computational load while a smaller value of η calls for GPR_{M_g} with less computational load. For most of the query samples, JGPR models will be chosen for prediction if the value of η approaches 1. In this situation, AGPR and JGPR can be considered as the same method. The main effect of $\eta_{\min}$ is to reduce the computational load of the related JGPR models. Compared to η , the range of $\eta_{\min}$ is relatively narrow. Generally, the value of $\eta_{\min}$ can be chosen as a very small one, e.g., $0.1 \leq \eta_{\min} \leq 0.2$. The detailed results and discussions can be found in Supporting Information.

For all the test samples, online MI prediction results for $\mathbf{S}_1$ (Numbers 1–15), $\mathbf{S}_2$ (Numbers 16–37), $\mathbf{S}_3$ (Numbers 38–118), and $\mathbf{S}_t$ (Numbers 119–155) using the AGPR soft sensor modeling method with $\eta = 0.7$ and $\eta_{\min} = 0.2$ are shown in Fig. 4. In order to clearly demonstrate the prediction results in Fig. 4, the test samples are rearranged based on each mode, including steady-state grades or transitions, instead of the production time. The upper line and the lower line denote $\hat{y}_q + \sigma_{\hat{y}_q}$ and $\hat{y}_q - \sigma_{\hat{y}_q}$, respectively. Detailed implementation procedures of AGPR are shown in Figs. 5a and 5b, respectively. The maximal probability and minimal probability of the query samples for steady-state grade models are both plotted in Fig. 5a. For a query sample, the suitable steady-state model $\text{GPR}_{M_g}, g = 1, \dots, 3$ for prediction can be selected if its maximal probability is larger than $\eta = 0.7$. Otherwise, the JGPR model shown in Fig. 5b is online constructed.

The test samples in $\mathbf{S}_1$ show different characteristics from the training samples in Fig. 3a. Alternatively, JGPR models can be a suitable choice if the related steady-state models are not reliable enough. Additionally, for the transitional samples, neither the existing steady-state grade model can predict accurately. They are much more difficult to predict than the samples in steady-state grades. More than half of the transitional samples during grade changeover suggest that JGPR models should be selected for prediction. Details about the performance comparisons of MI prediction using AGPR

and JGPR methods are tabulated in Table 3. As shown in Tables 2 and 3, the training and testing accuracy of S_3 are almost the same (RE = 8.0 % and RE = 6.8 %, respectively). The prediction results of this grade are suitable for the industrial polyethylene production process. The same model is persistently selected (in the range of numbers 38–118) mainly due to two reasons. First, the size of S_3 is much larger than S_1 and S_2 . Compared to S_1 and S_2 , the modeling information for S_3 is relatively enough. Second, as analyzed in Section 4.1, GPR_{M_3} is more reliable than the others. Consequently, the maximal probability is large enough for the test samples in S_3 to adopt GPR_{M_3} .

(Fig. 4 goes here.)

(Fig. 5a goes here.)

(Fig. 5b goes here.)

(Table 3 goes here.)

In our previous study, the IJ-LSSVR method has shown better prediction performance than JLSSVR, LSSVR, weighted LSSVR, and weighted PLS soft sensor modeling methods.³² Consequently, AGPR is compared with the IJ-LSSVR method. To further demonstrate its modeling ability, AGPR is compared with other probabilistic soft sensors for MI prediction. Recently, some GPR-based mixture models for the multi-mode processes have been proposed.^{25,27} The results have shown better performance than a single global model and corresponding steady-state models.^{25,27} In this section, for comparison, a mixture GPR (MGPR) modeling method for multi-grade processes is also proposed using the probability analysis aforementioned. It is formulated below.

$$\hat{y}_q = \sum_{g=1}^3 P(M_g | \mathbf{x}_q) GPR_{M_g} \quad (20)$$

Details about the performance comparisons of MI prediction using MGPR and IJ-LSSVR methods

are also tabulated in Table 3. The MGPR model for $\mathbf{x}_q$ can be obtained by combining $GPR_{M_g}, g = 1, \dots, 3$ models. It can be considered as multi-nonlinear models for multi-grade processes.

Among these three probabilistic nonlinear soft sensors, the RE index in Table 3 show that AGPR can achieve better prediction performance than the other two methods, especially for transitional samples and those overlapping samples. As analyzed before, only using JGPR is not the best choice for complicated multi-grade processes. The MGPR approach is an alternative approach for multi-mode modeling, expert for the transitional modes or those samples showing different characteristics from the training samples (e.g., the test samples in S_1). The trained GPR steady-state models have shown that only GPR_{M_3} is reliable. Consequently, as shown in Table 3, MGPR can achieve better performance for S_3 than the other methods. However, for the whole process, MGPR is inferior to AGPR and JGPR methods mainly because the mixture strategy only combines the predictions of each steady-state model using the probability regardless of the reliability of related steady-state models.

The RPV values of all the query samples of three probabilistic nonlinear soft sensors are shown in Fig. 6. Correspondingly, the RPV values of the transitional samples are much larger than those steady-state grades. The E_M index of the three methods in Table 3 shows that the prediction results of AGPR is more reliable than JGPR and MGPR. From the RE index, AGPR is only a little superior to the IJ-LSSVR method. However, the IJ-LSSVR method cannot provide probabilistic information for its prediction. Consequently, it is difficult to know whether the prediction is good or bad before the lab analysis results are available. The proposed RPV and E_M indices are important for soft sensors because the probabilistic information can help operators/engineers utilize the prediction in a better way. Therefore, from all the obtained results, the proposed AGPR method shows better and more reliable prediction performance than the other soft sensors in terms of online MI prediction for the whole polyethylene process.

(Fig. 6 goes here.)

The comparison of the prediction performance using AGPR soft sensors with the distance and angle criterion are listed in Table S1 in Supporting Information. For the industrial data, the soft sensor models are not accurate enough to make good predictions due to several reasons. First, the industrial data samples are error-in-variables actually. Second, the modeling samples are not enough to capture all the process characteristics. Third, the process characteristics may be time-varying and have some

uncertainties. It should also be noted that the recursive methods can be adopted to update the GPR steady-state models and then enhance the prediction performance since new samples can be added to the data set gradually. Additionally, for larger modeling sets, sparse training strategies for GPR can be introduced to reduce the overall computational burden. However, these two issues are beyond the main scope of this study and they would not be investigated here. The related algorithms can be referred to recursive GPR and sparse GPR methods.^{38,39,50,51}

5. Conclusion

This paper addresses the topic of developing reliable soft sensors for complicated multi-grade processes. An auto-switch probabilistic soft sensor modeling method for online quality prediction of a whole multi-grade process in transitional modes is proposed. First, several offline steady-state models are built and analyzed using a novel performance index. Additionally, the JGPR-based local online modeling method is proposed. Furthermore, the AGPR method using Bayesian inference of uncertainty of steady-state models is presented for suitable prediction of the query samples in different modes.

Unlike the traditional soft sensors in industrial processes, prediction uncertainty is explored and integrated into the AGPR method to enhance the prediction reliability. The new performance index for the evaluation of the trained GPR-based soft sensors is explored. The superiority of AGPR is demonstrated through an online MI prediction of an industrial polyethylene process with multiple grades in Taiwan. Compared with other approaches, better and more reliable prediction of AGPR has been obtained.

Supporting Information

The results of the corresponding computational time of AGPR with different values of η and $\eta_{\min}$ are shown in Fig. S1. The effect of η on the MI prediction performance (the RE index) of AGPR with different values of $\eta_{\min}$ for all the query samples is shown in Fig. S2. The comparisons of the prediction performance using AGPR soft sensors with the distance and angle criterion are listed in Table S1. The detailed discussions can be found in Supporting Information. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Acknowledgment

The authors would like to gratefully acknowledge National Science Council, R.O.C. for the financial support. Also, they would like to thank *Formosa Plastics Corporation*, which provides the industrial data for the case study.

5

Abbreviations

AGPR = auto-switch Gaussian process regression

CSF = cumulative similarity factor

CSTR = continuous stirred tank reactor

10 GPR = Gaussian process regression

IJ-LSSVR = integrated just-in-time least squares support vector regression

JGPR = just-in-time Gaussian process regression

JIT = just-in-time

JLSSVR = just-in-time least squares support vector regression

15 JSVR = just-in-time support vector regression

LSSVR = least squares support vector regression

MI = melt index

NN = neural networks

PCA = principal component analysis

20 PLS = partial least squares

RE = relative root-mean-square error

RMSE = root-mean-square error

RPV = relative-root prediction variance

SF = similarity factor

25 SVR = support vector regression

References

- (1) Fortuna, L.; Graziani, S.; Rizzo, A.; Xibilia, M. G. *Soft Sensors for Monitoring and Control of Industrial Processes*; Springer-Verlag: New York, 2007.
- 30 (2) Kadlec, P.; Gabrys, B.; Strandt, S. Data-driven soft sensors in the process industry. *Comput. Chem. Eng.* **2009**, *33*, 795–814.

- (3) Kadlec, P.; Grbic, R.; Gabrys, B. Review of adaptation mechanisms for data-driven soft sensors. *Comput. Chem. Eng.* **2011**, *35*, 1–24.
- (4) Kano, M.; Ogawa, M. The state of the art in chemical process control in Japan: Good practice and questionnaire survey. *J. Process Control* **2010**, *20*, 969–982.
- 5 (5) Undey, C.; Cinar, A. Statistical monitoring of multistage, multiphase batch processes. *IEEE Control Syst. Mag.* **2002**, *22*, 40–52.
- (6) Yao, Y.; Gao, F. R. A survey on multistage/multiphase statistical modeling methods for batch processes. *Annu. Rev. Control* **2009**, *33*, 172–183.
- (7) McAuley, K. B.; MacGregor, J. F. On-line inference of polymer properties in an industrial polyethylene reactor. *AIChE J.* **1991**, *37*, 825–835.
- 10 (8) Sharmin, R.; Sundararaj, U.; Shah, S.; Griend, L. V.; Sun, Y. J.. Inferential sensors for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant. *Chem. Eng. Sci.* **2006**, *61*, 6372–6384.
- (9) Qin, S. J. Recursive PLS algorithms for adaptive data modeling. *Comput. Chem. Eng.* **1998**, *22*, 503–514.
- 15 (10) Ahmed, F.; Nazir, S.; Yeo, Y. K. A recursive PLS-based soft sensor for prediction of the melt index during grade change operations in HDPE plant. *Korean J. Chem. Eng.* **2009**, *26*, 14–20.
- (11) Zhang, J.; Morris, A. J.; Martin, E. B.; Kiparissides, C. Prediction of polymer quality in batch polymerisation reactors using robust neural networks. *Chem. Eng. J.* **1998**, *69*, 135–143.
- 20 (12) Zhang, J.; Jin, Q.; Xu, Y. M. Inferential estimation of polymer melt index using sequentially trained bootstrap aggregated neural networks. *Chem. Eng. Technol.* **2006**, *29*, 442–448.
- (13) Gonzaga, J. C. B.; Meleiro, L. A. C.; Kiang, C.; Filho, R. M. ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. *Comput. Chem. Eng.* **2009**, *33*, 43–49.
- 25 (14) Mat Noor, R. A.; Ahmad, Z.; Mat Don, M.; Uzir, M. H. Modelling and control of different types of polymerization processes using neural networks technique: A review. *Can. J. Chem. Eng.* **2010**, *88*, 1065–1084.
- (15) Lou, H. C.; Su, H. Y.; Xie, L.; Gu, Y.; Rong, G. Inferential model for industrial polypropylene melt index prediction with embedded priori knowledge and delay estimation. *Ind. Eng. Chem. Res.* **2012**, *51*, 8510–8525.
- 30 (16) Liu, J. L. On-line soft sensor for polyethylene process with multiple production grades. *Control Eng. Pract.* **2007**, *15*, 769–778.
- (17) Han, I. S.; Han, C. H.; Chung, C. B. Melt index modeling with support vector machines, partial least squares, and artificial neural networks. *J. Appl. Polym. Sci.* **2005**, *95*, 967–974.
- 35 (18) Lee, D. E.; Song, J. H.; Song, H. O.; Yoon, E. S. Weighted support vector machine for quality estimation in the polymerization process. *Ind. Eng. Chem. Res.* **2005**, *44*, 2101–2105.
- (19) Shi, J.; Liu, X. G. Melt index prediction by weighted least squares support vector machines. *J. Appl. Polym. Sci.* **2006**, *101*, 285–289.
- (20) Chitralekha, S. B.; Shah, S. L. Application of support vector regression for developing soft sensors

- for nonlinear processes. *Can. J. Chem. Eng.* **2010**, 88, 696–709.
- (21) Kaneko, H.; Funatsu, K. Development of soft sensor models based on time difference of process variables with accounting for nonlinear relationship. *Ind. Eng. Chem. Res.* **2011**, 50, 10643–10651.
- (22) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*; MIT Press, 2006.
- 5 (23) Gregorcic, G.; Lightbody, G. Gaussian process approach for modeling of nonlinear systems. *Eng. Appl. Art. Intell.* **2009**, 22, 522–533.
- (24) Chen, T.; Ren, J. H. Bagging for Gaussian process regression. *Neurocomput.* **2009**, 72, 1605–1610.
- (25) Ge, Z. Q.; Chen, T.; Song, Z. H. Quality prediction for polypropylene production process based on CLGPR model. *Control Eng. Pract.* **2011**, 19, 423–432.
- 10 (26) Ou, X. L.; Martin, E. Batch process modelling with mixtures of Gaussian processes. *Neural Comput. Applic.* **2008**, 17, 471–479.
- (27) Yu, J. Online quality prediction of nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach. *Chem. Eng. Sci.* **2012**, 82, 22–30.
- 15 (28) Kim, M.; Lee, Y. H.; Han, I. S.; Han, C. H. Clustering-based hybrid soft sensor for an industrial polypropylene process with grade changeover operation. *Ind. Eng. Chem. Res.* **2005**, 44, 334–342.
- (29) Kaneko, H.; Arakawa, M.; Funatsu, K. Novel soft sensor method for detecting completion of transition in industrial polymer processes. *Comput. Chem. Eng.* **2011**, 35, 1135–1142.
- (30) Richards, J. R.; Congalidis, J. P. Measurement and control of polymerization reactors. *Comput.*
20 *Chem. Eng.* **2006**, 30, 1447–1463.
- (31) Kiparissides, C. Challenges in particulate polymerization reactor modeling and optimization: A population balance perspective. *J. Process Control* **2006**, 16, 205–224.
- (32) Liu, Y.; Chen, J. Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes. *J. Process Control* **2013**, 23,
25 793–804.
- (33) Murray-Smith, R.; Johansen, T. A. *Multiple Model Approaches to Modeling and Control*. Taylor & Francis: London, 1997.
- (34) Ozkan, L.; Kothare, M.; Georgakis, C. Control of a solution co polymerization reactor using multi-model predictive control. *Chem. Eng. Sci.* **2003**, 36, 1207–1221.
- 30 (35) Galán, O.; Romagnoli, J. A.; Palazoglu, A. Real-time implementation of multilinear model-based control strategies—an application to a bench-scale pH neutralization reactor. *J. Process Control* **2004**, 14, 571–579.
- (36) Hosseini, S.; Fatehi, A.; Johansen, T. A.; Sedigh, A. K. Multiple model bank selection based on nonlinearity measure and H-gap metric. *J. Process Control* **2012**, 22, 1732–1742.

- (37) Hariprasad, K.; Bhartiya, S.; Gudi, R. D. A gap metric based multiple model approach for nonlinear switched systems. *J. Process Control* **2012**, *22*, 1743–1754.
- (38) Csato, L.; Oppner, M. Sparse on-line Gaussian processes. *Neural Comput.* **2002**, *14*, 641–668.
- (39) Ranganathan, A.; Yang, M. H.; Ho, J. Online sparse Gaussian process regression and its applications. *IEEE Trans. Image Process.* **2011**, *20*, 391–404.
- (40) Atkeson, C. G.; Moore, A. W.; Schaal, S. Locally weighted learning. *Artif. Intell. Rev.* **1997**, *11*, 11–73.
- (41) Bontempi, G.; Birattari, M.; Bersini, H. Lazy learning for local modeling and control design. *Int. J. Control* **1999**, *72*, 643–658.
- (42) Cheng, C.; Chiu, M. S. A new data-based methodology for nonlinear process modeling. *Chem. Eng. Sci.*, **2004**, *59*, 2801–2810.
- (43) Fujiwara, K.; Kano, M.; Hasebe, S.; Takinami, A. Soft-sensor development using correlation-based just-in-time modeling. *AIChE J.*, **2009**, *55*, 1754–1765.
- (44) Fujiwara, K.; Kano, M.; Hasebe, S. Development of correlation-based clustering method and its application to software sensing. *Chemom. Intell. Lab. Syst.* **2010**, *101*, 130–138.
- (45) Ge, Z. Q.; Song, Z. H. A comparative study of just-in-time-learning based methods for online soft sensor modeling. *Chemom. Intell. Lab. Syst.* **2010**, *104*, 306–317.
- (46) Liu, Y.; Gao, Z. L.; Li, P.; Wang, H. Q. Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes. *Ind. Eng. Chem. Res.* **2012**, *51*, 4313–4327.
- (47) Hu, Y.; Ma, H.; Shi, H. Enhanced batch process monitoring using just-in-time-learning based kernel partial least squares. *Chemom. Intell. Lab. Syst.* **2013**, *123*, 15–27.
- (48) Pan, T. H.; Li, S. Y.; Cai, W. J. Lazy learning-based online identification and adaptive PID control: A case study for CSTR process. *Ind. Eng. Chem. Res.* **2007**, *46*, 472–480.
- (49) Yang, X.; Jia, L.; Chiu, M. S. Adaptive decentralized PID controllers design using JITL modeling methodology. *J. Process Control* **2012**, *22*, 1531–1542.
- (50) Ni, W.; Tan, S.; Ng, W.; Brown, S. D. Moving-window GPR for nonlinear dynamic system modeling with dual updating and dual preprocessing. *Ind. Eng. Chem. Res.* **2012**, *51*, 6416–6428.
- (51) Chan, L. L. T.; Liu, Y.; Chen, J. Nonlinear system identification with selective recursive Gaussian process models. *Ind. Eng. Chem. Res.* **2013**, *52*, 18276–18286.

Figure Captions

Fig. 1. (Left) A simplified flowchart of a multi-grade process with three steady-state grades and corresponding transitional modes (Right) Data distribution of three steady-state grades and related transitions

Fig. 2. Auto-switch GPR-based soft sensor modeling flowchart for a multi-grade process with several steady-state grades and corresponding transitional modes

Fig. 3a. The offline trained GPR soft sensor model GPR_{M_1} and its RPV values (with $E_{M_1} = 3.43\%$) for steady-state grade S_1

10 **Fig. 3b.** The offline trained GPR soft sensor model GPR_{M_2} and its RPV values (with $E_{M_2} = 4.95\%$) for steady-state grade S_2

Fig. 3c. The offline trained GPR soft sensor model GPR_{M_3} and its RPV values (with $E_{M_3} = 1.33\%$) for steady-state grade S_3

15 **Fig. 3d.** The offline trained GPR soft sensor model GPR_{M_t} and its RPV values (with $E_{M_t} = 25.04\%$) for steady-state grade S_t

Fig. 4. Online prediction of MI for S_1 (Numbers 1–15), S_2 (Numbers 16–37), S_3 (Numbers 38–118), and S_t (Numbers 119–155) with the AGPR soft sensor modeling method with $\eta = 0.7$ and $\eta_{\min} = 0.2$

Fig. 5a. Implementation procedure of the AGPR soft sensor modeling method with $\eta = 0.7$ and $\eta_{\min} = 0.2$: maximal probability of the query samples for steady-state grade models

20 **Fig. 5b.** Prediction models for query samples: Model 0 means that the query samples adopt the JGPR model for prediction, and Models 1-3 denote the query samples adopt steady-state models.

Fig. 6. Comparisons of RPV values of AGPR, JGPR and MGPR soft sensors for online prediction of all the query samples (S_1 : Numbers 1–15, S_2 : Numbers 16–37, S_3 : Numbers 38–118, and S_t : Numbers 119–155)

25

Table Captions

Table 1 Comparisons between JGPR and JLSSVR/JSVR online modeling methods

Table 2 RE (defined in Eq. (18)), the range of RPV values and E_{M_g} (defined in Eqs. (7) and (8))

5 indices of the trained offline GPR models for several steady-state grades and related transitional modes

Table 3 Comparisons of prediction performances using different soft sensors for several steady-state grades and related transitional modes. (The best results with related index are bold and underlined.)

Fig. 1. (Left) A simplified flowchart of a multi-grade process with three steady-state grades and corresponding transitional modes. (Right) Data distribution of three steady-state grades and related

5

transitions

10

15

20

Fig. 2. Auto-switch GPR-based soft sensor modeling flowchart for a multi-grade process with several steady-state grades and corresponding transitional modes

5

10

15

Fig. 3a. The offline trained GPR soft sensor model GPR_{M_1} and its RPV values (with $E_{M_1} = 3.43 \%$)

for steady-state grade S_1

5

10

15

Fig. 3b. The offline trained GPR soft sensor model GPR_{M_2} and its RPV values (with $E_{M_2} = 4.95 \%$)

for steady-state grade S_2

5

10

15

Fig. 3c. The offline trained GPR soft sensor model GPR_{M_3} and its RPV values (with $E_{M_3} = 1.33\%$)

for steady-state grade S_3

5

10

15

Fig. 3d. The offline trained GPR soft sensor model GPR_{M_t} and its RPV values (with $E_{M_t} = 25.04 \%$)

for steady-state grade S_t

5

10

15

Fig. 4. Online prediction of MI for S_1 (Numbers 1–15), S_2 (Numbers 16–37), S_3 (Numbers 38–118), and S_4 (Numbers 119–155) with the AGPR soft sensor modeling method with $\eta = 0.7$ and $\eta_{\min} = 0.2$

5

10

Fig. 5a. Implementation procedure of the AGPR soft sensor modeling method with $\eta = 0.7$ and $\eta_{\min} = 0.2$: maximal probability of the query samples for steady-state grade models

5

10

Fig. 5b. Prediction models for query samples: Model 0 means that query samples adopt the JGPR model for prediction, and Models 1-3 denote that the query samples adopt steady-state models

$$\text{GPR}_{M_g}, g = 1, 2, 3, \text{ respectively}$$

5

10

Fig. 6. Comparisons of RPV values of AGPR, JGPR and MGPR soft sensors for online prediction of

5 all the query samples (S_1 : Numbers 1–15, S_2 : Numbers 16–37, S_3 : Numbers 38–118, and S_4 : Numbers
119–155)

Table 1 Comparisons between JGPR and JLSSVR/JSVR online modeling methods

	Model parameters	Determination of model parameters	Probabilistic information of prediction
JGPR	Hyper-parameters	Bayesian inference	Yes
JLSSVR/JSVR	Regularization parameter and kernel function with its parameter	Cross-validation or fast leave-one-out cross-validation	No

5

Table 2 RE (defined in Eq. (18)), the range of RPV values and E_{M_g} (defined in Eqs. (7 and 8))

indices of the trained offline GPR models for several steady-state grades and related transitional modes

Trained offline GPR models	RE (%)	E_{M_g} (%)	Range of RPV values
GPR _{M₁} for steady-state grade S ₁	2.25	3.43	2~6
GPR _{M₂} for steady-state grade S ₂	9.16	4.95	2~8
GPR _{M₃} for steady-state grade S ₃	8.00	1.33	0~3
GPR _{M_t} for transitional modes S _t	65.18	25.04	0~120

10

15

Table 3 Comparisons of prediction performance using different soft sensors for several steady-state grades and related transitional modes (The best results with related index are bold and underlined.)

Method	Mode	RE (%)	E_M (%)
AGPR with $\eta = 0.7$ and $\eta_{\min} = 0.2$	S	<u>25.2</u>	<u>11.5</u>
	S ₁	28.3	<u>10.9</u>
	S ₂	<u>24.7</u>	<u>13.0</u>
	S ₃	6.8	3.3
	S _t	<u>43.3</u>	<u>28.9</u>
JGPR with $\eta_{\min} = 0.2$	S	29.8	14.5
	S ₁	<u>18.8</u>	13.6
	S ₂	34.9	20.3
	S ₃	9.3	3.4
	S _t	51.5	35.8
MGPR	S	39.5	15.2
	S ₁	31.4	13.1
	S ₂	26.5	14.2
	S ₃	<u>6.2</u>	<u>2.0</u>
	S _t	75.1	45.5
IJ-LSSVR ³²	S	25.9	No probabilistic information for its prediction.
	S ₁	22.6	
	S ₂	32.5	
	S ₃	6.4	
	S _t	43.3	