1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

**PID Based Nonlinear Processes Control Model Uncertainty
Improvement by Using Gaussian Process Model**

Lester Lik Teck Chan^a, Tao Chen^b and Junhui Chen^{a*}

^aDepartment of Chemical Engineering, Chung-Yuan Christian University, Chung-Li,
Taiwan 320, Republic of China

^bDepartment of Chemical and Process Engineering, University of Surrey, Guildford GU2
7XH, UK

Revision Submitted to Journal of Process Control

October, 2015

* Corresponding author: Fax: +886-3-2654199; E-mail: jason@wavenet.cycu.edu.tw

Abstract

Proportional-integral-derivative (PID) controller design based on the Gaussian process (GP) model is proposed in this study. The GP model, defined by its mean and covariance function, provides predictive variance in addition to the predicted mean. GP model highlights areas where prediction quality is poor, due to the lack of data, by indicating the higher variance around the predicted mean. The variance information is taken into account in the PID controller design and is used for the selection of data to improve the model at the successive stage. This results in a trade-off between safety and the performance due to the controller avoiding the region with large variance at the cost of not tracking the set point to ensure process safety. The proposed direct method evaluates the PID controller design by the gradient calculation. In order to reduce computation the characteristic of the instantaneous linearized GP model is extracted for a linearized framework of PID controller design. Two case studies on continuous and batch processes were carried out to illustrate the applicability of the proposed method.

Key words: approximation; Gaussian process; model update; PID control

1. Introduction

Conventional PID controllers have been the most preferred controllers due to their simple structure and hardware requirement. Moreover, PID controllers are the basic component in modern multi-level hierarchy of control, which function as regulatory controllers. The highly nonlinear and time varying chemical processes can, however, pose a great challenge to the PID controller design. Early works in self-tuning control include design via pole placement and method based on generalized minimum variance control. Recently, Romero et al. [1] presented an auto tuning algorithm for PI and PID controllers to minimize the load disturbance integral error. Leva and Maggio [2] used constrained optimization to find tuning rules for both the PID parameters and the noise filter. Alfaro et al. [3] exploited constrained optimization to determine the PID parameters. Zhang et al. [4] applied the $\mathcal{H}_\infty$ PID control for disturbance attenuation whereas Garpinger et al. [5]

investigated the tuning in terms of performance and robustness trade-offs. Fuzzy self-tuning PID control of the operation temperatures in a two-staged membrane separation process [6] and self-tuning PID control of jacketed batch polystyrene reactors using genetic algorithms [7] were proposed. Also, Lee and Teng [8] used the fuzzy neural network to develop a formula for PID controller design. Kansha et al. [9] on the other hand, presented a self-tuning PID controller based on the Lyapunov approach. Furthermore, auto-tuning of PID controllers as well as robust design of PID controller were proposed [10-12]. Nevertheless, when changes in the process or the environment occur, manual check on the adequacy of the model is required as the control performance relies on the model.

The integrity of the model is, therefore, very important and information on the model uncertainty can be invaluable. The construction of nonlinear models presents a difficulty if there is a lack of necessary trust in the model [13]. Chen and Huang [14] applied the linearized neural network based model to the tuning of PID controllers whereas an approach based on a lazy learning identification was proposed [15]. However, they did not consider the model uncertainty when tuning the controllers. To take into account the model uncertainty Gaussian Process (GP) model can be used. The GP model provides predictive variance which indicates the reliability of the prediction in the local stochastic region while traditional nonlinear models do not. If the predictive variance is low, the confidence level of the model prediction is high; on the other hand, high predictive variance indicates that the model is highly uncertain. GP model have been increasingly applied to different nonlinear dynamic systems. The GP model approach to curve fitting was first introduced by O'Hagan and Kingman [16] and was later compared to widely used models by Rasmussen [17], leading to the rapid expansion of the research into GP model since then. Rasmussen applied the technique to process modeling by estimating the state transition function for nonlinear processes. Azman and Kocijan [18] applied GP to the modeling of the bio-system whereas Lundgren and Sjoberg [19] used the GP model for linear and nonlinear model validation. di Sciascio and Amicarelli [20] developed a biomass concentration estimator for a batch biotechnological process based on the GP model. Ni et al. [21] proposed a recursive GP and adapted it to the process drift in both

sample-wise and block-wise manners with a filter. Their method preprocessed the data and improved the accuracy of the prediction. Chan et al. [22] proposed a recursive update of GP with selective data. The GP model has also been utilized with other tools. Yan et al. [23] applied the GP model to migration so that they could adapt a base model from an old process to a new but similar one. Likar and Kocijan [24] applied the GP model to predictive control in a gas-liquid separation plant. In their work, it was shown that the process safety can be assured if the plant is operated only within the known region.

In this work a method that takes into account the model uncertainty in the PID controller design based on the GP model is proposed. Using the variance information a trade-off between safety and performance can be achieved. This trade-off occurs due to the controller avoiding the region with large variance, i.e. highly uncertain region, at the cost of not tracking the set point to ensure process safety. The identified GP can be directly applied to the PID controller design by gradient method. However, the optimization problem requires the computation of the inversion of the $N \times N$ covariance matrix. There is an associated computation cost of $O(N^3)$ which poses a considerable load on the computation. To this end, similar to the linear control design theory, linearization of nonlinear models is often used in the control field to alleviate computation load in the design of controllers for nonlinear systems. By extracting the characteristic of the instantaneous linearized GP model, the linearized framework can be applied without modification. A heuristic method that allows a quick way to select new data to improve the model at the successive stage is proposed. This is facilitated by the predictive variance which gives a direct indication about whether the data in the region is lacking. With improving model the control scheme transits from performance and safety trade-off to an optimal performance as the model uncertainty is reduced. The rest of the article is arranged as follows. Section 2 presents a problem statement of this work and Section 3 details the GP model with the heuristic method to improve model uncertainty. This is followed by the direct GP based and the approximate GP based PID controller design in Section 4. Section 5 illustrates the applicability of the proposed tuning method based on two different process dynamic problems - continuous and batch - and the article ends with concluding remarks in Section 6.

1

2 **2. Problem statement**

3 Fig. 1 shows the PID control scheme considered in this study. The process variable is y_t
 4 and u_t is the manipulated variable. The PID controller algorithm is

$$u_t = \bar{u} + k_c \left[e_t + \frac{1}{\tau_I} \int_0^t e_t + \tau_D \frac{de_t}{dt} \right] \quad (1)$$

5 where the error, $e_t = y_t^{set} - y_t$. k_c , τ_I and τ_D are the proportional gain, integral time
 6 constant and the derivative time constant, respectively. Using trapezoidal approximation
 7 $\int_0^t e_t \approx \sum_{i=1}^n e_{i\Delta t} \Delta t$ and differential approximation $\frac{de_t}{dt} \approx \frac{e_t - e_{t-1}}{\Delta t}$, the velocity form of the
 8 discrete PID controller can thus be expressed as,

$$\Delta u_t = u_t - u_{t-1} = k_c \underbrace{\left[1 + \frac{1}{\tau_I} + \frac{\tau_D}{\Delta t} \right]}_{k_t^0} e_t + k_c \underbrace{\left[-1 - 2 \frac{\tau_D}{\Delta t} \right]}_{k_t^1} e_{t-1} + k_c \underbrace{\frac{\tau_D}{\Delta t}}_{k_t^2} e_{t-2} \quad (2)$$

9 Eq. (2) can be expressed as

$$\Delta u_t = k_t^0 e_t + k_t^1 e_{t-1} + k_t^2 e_{t-2} = \mathbf{e}_t^T \mathbf{k}_t \quad (3)$$

10 where $\mathbf{k}_t = \begin{bmatrix} k_t^0 & k_t^1 & k_t^2 \end{bmatrix}^T$. $\mathbf{e}_t = \begin{bmatrix} e_t & e_{t-1} & e_{t-2} \end{bmatrix}^T$ and

$$k_t^0 = k_c \left(1 + \frac{\Delta t}{2\tau_I} + \frac{\tau_D}{\Delta t} \right), \quad k_t^1 = -k_c \left(1 + \frac{2\tau_D}{\Delta t} \right), \quad k_t^2 = k_c \frac{\tau_D}{\Delta t} \quad (4)$$

11 The parameters of the PID controller are adjusted by a tuner based on an identified GP
 12 model. A tuning scheme with the objective function based on the uncertainty information
 13 is proposed for controller design as the GP model provides a mean value prediction as
 14 well as the variance prediction. The GP model is a Bayesian modeling framework and is
 15 conditioned over a Gaussian prior. The posterior is also a Gaussian distribution with

mean and variance and the predictive variance can be viewed as the information about the model confidence. As a result, in contrast to traditional model based control, the issue of control system robustness can be considered. This issue has a major impact on the applicability of the controller scheme. For instance, an out-of-control or too aggressive control situation can pose a threat to the process safety. The information on the confidence of the prediction enables a trade-off between performance and safety to be considered.

When the identified GP model is not accurate, the trade-off between performance and safety occurs due to the controller avoiding the region with large variance at the cost of not tracking the set point to ensure process safety. Over the course of operation, new data are available which may provide information to enrich the model for a better prediction. With the information on the uncertainties of the prediction provided by the GP model, a simple heuristic method is used to select the data for the improvement of the model. As a result, the model uncertainty can be reduced and the controller design is able to achieve the improved performance without compromising the process safety.

3. GP model with improvement of model uncertainty

3.1 GP model

A GP model is a collection of random variables which have a joint distribution.

$$P(\mathbf{y}|\mathbf{C}, \mathbf{X}) = \frac{1}{Z} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{y} - \boldsymbol{\mu})\right) \quad (5)$$

for any collection of inputs $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_t\}$ and output $\mathbf{y} = \{y_1, \dots, y_t\}$, with mean vector $\boldsymbol{\mu}$. $\mathbf{X}$ refers to the input vectors of Gaussian process model and not the manipulated variable. For dynamic modeling $\mathbf{x}_t$ may contain past manipulated variables and the past process variables respectively

$$\mathbf{x}_t = [u_{t-1} \quad \cdots \quad u_{t-m+1} \quad y_{t-1} \quad \cdots \quad y_{t-l+1}]^T \quad (6)$$

1 where the subscript m and l refer to the past m manipulated variables and the past l
 2 process variables respectively. Z is an appropriate normalizing constant. $\mathbf{C}$ is the
 3 covariance matrix of the data defined by the parameterized covariance function
 4 $C_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$. The nature of $\mathbf{C}$ and hence the covariance function is crucial to the
 5 whole GP approach. Eq.(5) expresses the correlations between different points in the
 6 input space.

7 Given data $\mathbf{D} = (\mathbf{X}, \mathbf{y})$, the inference on y_{t+1} can be readily obtained since the joint
 8 density $P(y_{t+1}, \mathbf{y})$ is also Gaussian; thus, by Bayes theorem, the conditional distribution
 9 is

$$P(y_{t+1} | \mathbf{y}) = \frac{P(y_{t+1}, \mathbf{y})}{P(\mathbf{y})} \quad (7)$$

10 is also Gaussian. The posterior distribution is given by

$$P(y_{t+1} | \mathbf{y}) = \frac{1}{Z} \exp \left[-\frac{(y_{t+1} - \mu_{t+1})^2}{2\sigma_{\mu_{t+1}}^2} \right] \quad (8)$$

11 where

$$\mu_{t+1} = \mathbf{g}^T(\mathbf{x}_{t+1}) \mathbf{C}^{-1} \mathbf{y} \quad (9)$$

$$\sigma_{t+1}^2 = C(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}) - \mathbf{g}^T(\mathbf{x}_{t+1}) \mathbf{C}^{-1} \mathbf{g}(\mathbf{x}_{t+1}) \quad (10)$$

12 $\mathbf{g}(\mathbf{x}_{t+1})^T = [C(\mathbf{x}_{t+1}, \mathbf{x}_1) \quad \cdots \quad C(\mathbf{x}_{t+1}, \mathbf{x}_t)]$, μ_{t+1} is the mean prediction at $\mathbf{x}_{N+1}$ and σ_{t+1}^2 is
 13 the standard deviation of this prediction. The vector $\mathbf{g}^T(\mathbf{x}_{t+1}) \mathbf{C}^{-1}$ can be viewed as a
 14 smoothing term which weights the training outputs to make a prediction for the new input
 15 vector, $\mathbf{x}_{t+1}$. Eq.(10) provides a confidence level on the model prediction as the higher

1 variance value indicates that the region of the input vector contains few data or is
2 corrupted by noise.

3 The covariance function is non-trivial and a common choice is

$$C(\mathbf{x}_i, \mathbf{x}_j) = a_0 + a_1 \sum_{k=1}^d x_{ik} x_{jk} + v_0 \exp\left(-\sum_{k=1}^d w_k (x_{ik} - x_{jk})^2\right) + s \delta_{ij} \quad (11)$$

4 where a_0, a_1, w_k, s and v_0 are the hyper-parameters to be determined. The hyper-
5 parameter v_0 controls the overall scale of the local correlation, a_1 allows a different
6 distance measure in each input dimension, w_k , and s is the estimate of the noise variance.
7 δ is a Kronecker delta defined as

$$\delta_{i,j} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \quad (12)$$

8 Under the Gaussian process model the prior is the Gaussian distribution. The hyper-
9 parameters can be estimated by maximization of the log-likelihood

$$L(\boldsymbol{\theta}) = -\frac{1}{2} \log(|\mathbf{C}|) - \frac{1}{2} \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y} - \frac{N}{2} \log(2\pi) \quad (13)$$

10 where $\boldsymbol{\theta} = [a_0, a_1, w_k, s, v_0]$. The optimization problem can be solved using the derivative
11 of the log-likelihood with respect to each hyper-parameter given as

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = -\frac{1}{2} \text{tr}\left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \boldsymbol{\theta}}\right) + \frac{1}{2} \mathbf{y}^T \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \boldsymbol{\theta}} \mathbf{C}^{-1} \mathbf{y} \quad (14)$$

12 3.2 Heuristic for improvement of model uncertainty

13 The closed-loop control performance depends on the tuning of the controller. In model
14 based tuning method the accuracy of the identified model plays a critical role. With
15 regard to data selection past researches [25, 26] have reported that a smaller training set
16 may produce equally or better generalization performance than a larger one containing
17 redundant data. To improve the model data selection is needed as redundant data does not

improve the prediction and result in unnecessary computation. The GP model facilitates a simple selection criterion through the predictive variance.

The predictive variance indicates the uncertainty in the model where a large variance indicates a less confident prediction while a small band implies greater confidence on the prediction. Based on the predictive variance, it is possible to select data that are at the region of high uncertainty. A simple heuristic method is thus proposed. It is a two-step procedure that involves (1) finding area of high uncertainty followed by (2) determining the range of the admissible region. This is based on the understanding of the interpretation of the predictive variance. The region of high predictive variance implies a lack of data and consequently data should be added in the region with high variance. In order to determine the spot of high certainty, a threshold value of the predictive variance, σ_{th} , is evaluated based on the magnitude of the hyper-parameter, s (Eq.(11)), which is an estimate of the process variance. It is noted that preprocessing methods can affect the threshold of predictive variance. If there is no prior knowledge then auto-scale can be used. Based on knowledge of the physical system a greater weight can be given to the more important variables. After preprocessing, the threshold value is evaluated by rescaling back to the original space. Overall, the control performance is expected to not differ because the model will be continually learned. The points at which the predictive variance exceeds this threshold value are considered high uncertainty points. They are given as follows:

$$A = \left\{ y_{adm} \mid \sigma^2 \left(y_{adm} \left(\mathbf{x}_{adm} \right) \right) > \sigma_{th}^2 \right\} \quad (15)$$

where $\mathbf{x}_{adm}$ is the corresponding input vector. Fig. 2 illustrates the determination of the high uncertainty points and a single process variable is used for the illustration purpose. The solid line represents the actual process variable while the dotted line represents the GP model prediction. The grey shaded band represents $\mu \pm 2\sigma$ and gives an indication of the uncertainty in the prediction. In Fig. 2, $y_{adm,1}$ and $y_{adm,2}$ are regarded as high uncertainty points as both exceed the threshold value. The data in close proximity of the highly uncertain points are admitted. To determine the closeness of the data, the

1 corresponding input vector to the prediction with high variance will act as a pivot. The
 2 new input vector whose Euclidean distance is within a preset value is admitted, so the
 3 new data set is

$$B = \left\{ \mathbf{x} \left\| (\mathbf{x} - \mathbf{x}_{adm,j}) \right\|^2 \leq r\sigma^2(\mathbf{x}_{adm,j}) \text{ and } \mathbf{x}_{adm,j} \in A \right\} \quad (16)$$

4 where r is a preset value that determines the range of the allowable data. According to
 5 the statistical property such as 2 or 3 standard deviation where 3 times standard deviation
 6 means 98% of value lies within the limit and 2 times standard deviation implies 95% of
 7 value lies within the limit. A starting value of $r = 3$ or $r = 2$ can be selected as all the
 8 normal input vectors are very likely to fall within the limits. Nevertheless this serves as
 9 an initial guide only; as the model is continually improved it should eventually converge
 10 to the same accuracy. When a particular point has higher uncertainty, the neighboring
 11 data can be included to provide richer information. They enhance the model prediction
 12 and consequently improve the overall control performance. Based on the established
 13 region for accepting new data, the future process variables that fall within these regions
 14 (B) are admitted, as indicated by the dashed-line box in Fig. 2.

15 At the successive control stage, the model may be updated with new data and the control
 16 action can be evaluated based on the updated model. Let a particular admissible input
 17 vector which consists of the current value and the past value be $\mathbf{x}_{new}$,

$$\mathbf{x}_{new} = [u_{t-1} \quad \cdots \quad u_{t-m+1} \quad y_{t-1} \quad \cdots \quad y_{t-l+1}]^T \quad (17)$$

18 The selection of the past manipulated variables and process variables can influence the
 19 modeling performance. A small window may capture the process changes quickly but
 20 does not contain enough information to reflect the current process operating conditions
 21 sufficiently. On the other hand, while a sufficient number of samples may represent the
 22 process operating conditions appropriately, the large window may contain useless data.
 23 The window, therefore, ideally should have enough data to appropriately describe the
 24 characteristics of the system. l and m can be chosen based on an initial prediction

1 performance test which can be done first off-line. The corresponding process variable is
 2 represented by y_{new} . Thus, in an operation, adding the new data to the initial data set is
 3 defined as

$$\mathbf{X}^+ \leftarrow [\mathbf{X} \quad \mathbf{x}_{new}] \quad (18)$$

$$\mathbf{y}^+ \leftarrow [\mathbf{y} \quad y_{new}] \quad (19)$$

4 where $\mathbf{X}^+$ and $\mathbf{y}^+$ refer to the input and output vectors after the new data are added.
 5 Using this new set of data, the model is updated with the new hyper-parameters. In the
 6 following sections, for the brevity of notations, the superscript is omitted from $\mathbf{X}^+$ and
 7 $\mathbf{y}^+$. All the data used in the derivation implies the current set.

8

9 **4. GP model based PID tuning**

10 The identified GP model can be directly applied to PID controller design. By minimizing
 11 an objective function, the optimal control action can be obtained. This solution can be
 12 obtained by the gradient method with the necessary gradient calculation as detailed in
 13 Section 4.1. In Section 4.2, the approximation method is presented.

14 **4.1 GP based PID tuning with the gradient method (direct method)**

15 The objective function for minimum variance control is

$$J = E \left\{ \left(y_{t+1}^r - \mu_{t+1} \right)^2 \right\} + \lambda \Delta u_t^2 \quad (19)$$

16 where y_{t+1}^r is a reference target. The relative importance between the reference tracking
 17 and the aggressiveness is determined by a weighting factor, λ . If the controller focuses
 18 exclusively on set-point tracking, it might choose to make large manipulated variable
 19 adjustments which could lead to control system instability. A large λ forces the controller
 20 to make smaller, more cautious move. It should be selected so that the controller will be

1 less sensitive to prediction inaccuracies of the process. Using the fact that
 2 $\text{Var}\{y\} = E\{y^2\} - E^2\{y\}$ and that GP provides uncertainty in terms of predictive
 3 variance, the objective function can be written as

$$J = \left(y_{t+1}^r - \mu_{t+1}\right)^2 + \sigma_{t+1}^2 + \lambda \Delta u_t^2 \quad (20)$$

4 PID is tuned by minimizing the objective function J with the aim to design a control
 5 action that will minimize the difference between the process variable and the desired
 6 value, the variance of the prediction with consideration to the model uncertainty. The
 7 optimization takes into account this variance information, resulting in a more robust
 8 control system.

9 In order to use GP to tune PID, a gradient based optimization algorithm is derived. The
 10 Jacobian is

$$\frac{\partial J}{\partial \mathbf{k}_t} = \frac{\partial J}{\partial u_t} \frac{\partial u_t}{\partial \mathbf{k}_t} \quad (21)$$

11 where

$$\frac{\partial J}{\partial u_t} = 2\left(\mu_{t+1} - y_{t+1}^r\right) \frac{\partial \mu_{t+1}}{\partial u_t} + \frac{\partial \sigma_{t+1}^2}{\partial u_t} + 2\lambda \Delta u_t \quad (22)$$

12 and

$$\frac{\partial u_t}{\partial \mathbf{k}_t} = \mathbf{e}_t \quad (23)$$

13 The partial derivative terms of Eq.(22) can be expanded as follows

$$\frac{\partial \mu_{t+1}}{\partial u_t} = \frac{\partial \mathbf{g}^T(\mathbf{x}_{t+1})}{\partial u_t} \mathbf{C}^{-1} \mathbf{y} \quad (24)$$

14 and

$$\frac{\partial \sigma_{t+1}^2}{\partial u_t} = \frac{\partial C(\mathbf{x}_{t+1}, \mathbf{x}_{t+1})}{\partial u_t} - 2\mathbf{g}^T(\mathbf{x}_{t+1})\mathbf{C}^{-1} \frac{\partial \mathbf{g}(\mathbf{x}_{t+1})}{\partial u_t} \quad (25)$$

1 where $\mathbf{x}_{t+1} = [u_t \ \cdots \ u_{t-m+1} \ y_t \ \cdots \ y_{t-l+1}]^T$. Furthermore, the derivative of the
 2 covariance function (Eq.(11)) is

$$\frac{\partial C(\mathbf{x}_{t+1}, \mathbf{x}_i)}{\partial u_t} = a_1 x_{i1} + v_0 \exp\left(-\sum_{k=1}^d w_k (x_{ik} - x_{t+1,k})^2\right) \times 2w_1 (x_{i1} - u_t) \quad (26)$$

3 where at $\mathbf{x}_i = \mathbf{x}_{t+1}$, $\frac{\partial C(\mathbf{x}_{t+1}, \mathbf{x}_{t+1})}{\partial u_t} = a_1 u_t$,

$$\frac{\partial \mathbf{g}}{\partial u_t} = \left[\frac{\partial C(\mathbf{x}_{t+1}, \mathbf{x}_1)}{\partial u_t} \quad \frac{\partial C(\mathbf{x}_{t+1}, \mathbf{x}_2)}{\partial u_t} \quad \cdots \quad \frac{\partial C(\mathbf{x}_{t+1}, \mathbf{x}_n)}{\partial u_t} \right]^T \quad (27)$$

4 Calculating each term using Eqs.(22)-(27), the Jacobian is obtained and the optimal
 5 control action that minimizes Eq.(20) can be evaluated and the corresponding PID
 6 parameters obtained. This can be directly applied to controlling the process in order to
 7 achieve the desired process variable.

8 **4.2 Instantaneous GP PID tuning**

9 The key factor that controls the properties of GP is the covariance function. An
 10 approximation method is proposed to compliment the tuning. Using the approximated
 11 function, the computational load can be decreased, so it will be more convenient for
 12 tuning the PID controller. Fig. 3 illustrates the instantaneous linearization of the GP
 13 model. The actual process is represented by a solid line while the dotted line represents
 14 the GP model and the dashed line represents the approximated model.

15 At time t , the model is approximated by the value at the previous time $t-1$. For the
 16 approximation method, the objective function is expressed as a quadratic w.r.t $\mathbf{k}_t$,

$$J = \frac{1}{2} \mathbf{k}_t^T \mathbf{M} \mathbf{k}_t + \mathbf{d}^T \mathbf{k}_t + c \quad (28)$$

- 1 To express J as a quadratic, the term $(y_{t+1}^r - \mu_{t+1})^2$ in Eq. (20) has to be quadratic w.r.t
 2 $\mathbf{k}_t$. Observing Eq. (24) and Eq. (25) implies that $\mathbf{g}$ has to be a linear function of u_t .
 3 Using Taylor expansion around the input at the previous step u_{t-1}

$$\mathbf{g}^T(\mathbf{x}_t) = \mathbf{a}_1 + \mathbf{a}_2 u_t \quad (29)$$

- 4 where $\mathbf{a}_1 = \mathbf{g}|_{u_{t-1}} - u_{t-1} \mathbf{a}_2$ and $\mathbf{a}_2 = \frac{\partial \mathbf{g}}{\partial u_t} \Big|_{u_{t-1}}$.

- 5 The $\mathbf{M}$ and $\mathbf{d}$ terms in Eq.(28) can be obtained by differentiating the objective function
 6 and collecting the appropriate coefficient terms. The results after further evaluation of the
 7 partial derivatives are

$$\mathbf{d} = \begin{bmatrix} e_t & e_t & e_t \\ e_{t-1} & e_{t-1} & e_{t-1} \\ e_{t-2} & e_{t-2} & e_{t-2} \end{bmatrix} \begin{bmatrix} 2(y_{t+1}^r - \mathbf{a}_1 \mathbf{b} - \mathbf{a}_2 \mathbf{b} u_{t-1}) \mathbf{a}_2 \mathbf{b} \\ -2f \\ -4h u_{t-1} \end{bmatrix} \quad (30)$$

- 8 and

$$\mathbf{M} = \left((\mathbf{a}_2 \mathbf{b})^T (\mathbf{a}_2 \mathbf{b}) + \lambda - 4h \right) \begin{bmatrix} e_t^2 & e_t e_{t-1} & e_t e_{t-2} \\ e_{t-1} e_t & e_{t-1}^2 & e_{t-1} e_{t-2} \\ e_{t-2} e_t & e_{t-1} e_{t-2} & e_{t-2}^2 \end{bmatrix} \quad (31)$$

- 9 where $\mathbf{b} = \mathbf{C}^{-1} \mathbf{y}$, $f = [\mathbf{a}_1^T \mathbf{C}^{-1} \mathbf{a}_2 + \mathbf{a}_2^T \mathbf{C}^{-1} \mathbf{a}_1]$ and $h = \mathbf{a}_2^T \mathbf{C}^{-1} \mathbf{a}_2$. The optimal point occurs
 10 when the gradient is equal to zero; i.e. $\frac{\partial J}{\partial \mathbf{k}_t} = 0$; thus,

$$\mathbf{k} = -\mathbf{M}^{-1} \mathbf{d} \quad (32)$$

1 The optimal parameters are subjected to physical constraints, so Eq.(33) can be solved as
 2 a quadratic programming

$$\min J = \frac{1}{2} \mathbf{k}_t^T \mathbf{M} \mathbf{k}_t + \mathbf{d}^T \mathbf{k}_t \quad (34)$$

3 with the following constraints

$$\begin{aligned} k_c > 0 &\Rightarrow k_t^1 + k_t^2 < 0 \\ \tau_d > 0 &\Rightarrow (k_t^0 + k_t^1 + 2k_t^2) > 0 \\ \tau_i > 0 &\Rightarrow k_t^2 > 0 \end{aligned} \quad (35)$$

4 The optimal control action obtained using the linearization term is less expensive in terms
 5 of the calculation load. Compared to the direct GP method, which is the iterative
 6 calculation of the objective function, the approximation method is a one-time gradient
 7 calculation, so it has a faster performance.

8

9 **5. Case studies**

10 Two cases studies were carried out to show the applicability of the proposed method.
 11 They are a pH neutralization system and a fed-batch fermentation process. The pH
 12 neutralization system is a demonstration of continuous process where the process
 13 specification is not often changed. For the case of changing set points the capability of
 14 the proposed method is demonstrated with the fed-batch process fermentation process.

15 **5.1 pH neutralization system**

16 pH neutralization is fairly common among many chemical processes. The model
 17 equations in this study are obtained from Nahas et al. [27]. The pH continuous stirred-
 18 tank reactor (CSTR) system is shown in Fig. 4. The three input streams are acid (HNO_3),
 19 buffer (NaHCO_3) and base (NaOH) respectively. The system can be described by two

1 reaction invariants, three nonlinear ordinary differential equations and one nonlinear
 2 algebraic equation.

Charge balance

$$W_a \equiv [\text{H}^+] - [\text{OH}^-] - [\text{HCO}_3^-] - 2[\text{CO}_3^{=}] \quad (36)$$

Carbonate ion balance

$$W_b \equiv [\text{H}_2\text{CO}_3] + [\text{HCO}_3^-] + [\text{CO}_3^{=}]$$

$$\frac{dh}{dt} = \frac{1}{A}(q_1 + q_2 + q_3 - C_v h^{0.5})$$

$$\frac{dW_{a4}}{dt} = \frac{1}{Ah} [(W_{a1} - W_{a4})q_1 + (W_{a2} - W_{a4})q_2 + (W_{a3} - W_{a4})q_3] \quad (37)$$

$$\frac{dW_{b4}}{dt} = \frac{1}{Ah} [(W_{b1} - W_{b4})q_1 + (W_{b2} - W_{b4})q_2 + (W_{b3} - W_{b4})q_3]$$

$$W_{a4} + 10^{14-\text{pH}} + W_{b4} \frac{1 + 2 \times 10^{\text{pH}-\text{pK}_2}}{1 + 10^{\text{pK}_1-\text{pH}} + 10^{\text{pH}-\text{pK}_2}} - 10^{-\text{pH}} = 0 \quad (38)$$

3 where h is the liquid level, W_{a4} and W_{b4} are reactions invariants of the effluent streams,
 4 with q_1 , q_2 and q_3 being the acid, the buffer and the base flow rate respectively. The
 5 nominal conditions and the parameters are listed in Table 1. In pH neutralization control,
 6 automatic adaptive updating scheme was implemented by Zhang and Zhang [28]
 7 whereby a pseudo-derivative description was used to model the nonlinear behavior using
 8 a linear model. The linear model was continuously updated through an adaptive learning
 9 algorithm to ensure that the process was described as accurately as possible by a linear
 10 model. The adaptive approaches can handle the nonlinearities better than the simple
 11 linear method but is limited to areas where pH behaves fairly linearly. Moreover, the data
 12 is continually updated without selection. In contrast to the work, we present a data
 13 selection for model update and also considered the model uncertainty in the controller
 14 design. This case study is thus used to illustrate the usefulness of the variance term in the

objective functions (Eq.(20)) and then present the control performance using the proposed method.

5.1.1 GP model with rich data

The data in the region of pH 7 to pH 9 is used for the model identification where the input vector with $l = 1$ and $m = 2$ is this found to be appropriate to describe the dynamics of the process. As a result, it will be accurate for the region between pH 7 and pH 9. In the event that the model is accurate, the control results from using the GP model without considering the variance and using the GP model considering the variance are compared. The GP model without considering the variance refers to an objective function that is Eq.(20) without the term, σ_{t+1}^2 , on the right hand side. As the model is accurate, the predictive variance is small and it does not affect the control results. In Fig. 5, the tracking performance is satisfactory when the set-point is changed from pH 7 to pH 9. Fig. 6 shows the result from the approximation algorithm. The approximate GP model based control method exhibits a similar performance to the direct GP model. A comparison between the proposed method and linearized neural network model based control [14] was made, too. Fig. 6 shows that the approximate GP performs as well as the linearized neural network thereby confirming the applicability of the approximation. In terms of computation, the simulation performing on the same PC shows that the direct GP requires 7.5s while the approximate GP only takes 3.2s; i.e. a reduction of almost 60% in calculation time. The computation is similar between the linearized NN and the linearized GP but the GP has the added advantages of the predictive variance. The approximate GP is a trade-off between speed and computation load. For online application that requires fast computation, the approximate provides a speed advantage for a small penalty to the accuracy. If the data set is small, the direct method can be used. Otherwise, for the large data set the approximate GP is preferred. Thus, the small drop in the performance of the approximate GP is compensated by the increase in the speed of evaluation.

5.1.2 GP model with poor data

On the other hand, the actual operation process data may be lacking; as a result, the model obtained from this insufficient data is not accurate. As the model only contains data in the region between pH 7 and pH 9 and because of the nonlinearity of the pH process, the prediction outside this region will not be good. With prior knowledge that the model is inaccurate in some region $\lambda = 2$ is used so that the controller does not make a sudden, drastic change. The consequence of using this model can have a drastic effect on the control outcome as illustrated in Fig. 7. It shows the result when the model is inaccurate and a comparison between the cases of using and not using the variance term in the control objective. When the set point is set to pH 6, the model is inaccurate in this region as it lacks the training data in this region. When the variance term is not considered, an unreliable prediction value is used without any additional check in the control action (Eq.(20)). This results in a large offset from the set point represented by the dash-dot lines in Fig. 7 because there is a large prediction error in the calculated manipulated variable. On the other hand, the process variable that is being controlled does not deviate much because less data can be found in this region, resulting in higher uncertainty of prediction. The response to the system is optimized, so it prevents the system from reaching the region with high variance. Fig. 8 shows the comparison of the PID parameters for GP with and without variance. In Fig. 8(a), the controller neglects tuning of the region with high variance, so k_c and τ_i are fairly constant. It can be seen that the tuning is greater for the case of GP without variance consideration. This can be viewed as a trade-off between safety and the performance. Note that the performance of the GP without variance in Fig. 7 is the same as that of the linearized NN. This is because the model prediction for both is the same and the variance does not play a part.

5.1.3 Model update with new data

In this part of the case study, the effect on the control performance from the improvement to the model is demonstrated. The process is initially operated at pH 7. Subsequently, the set point is changed from pH 7 to pH 9 and then to pH 6. This change from pH 9 to pH 6 is done for a number of repetitions to show how the process will respond to the proposed method.

The process model is built from the data in the range between pH 7 and pH 9 and thus will be accurate for the proximity of the region of pH 9. Using this model, the controller is able to achieve the set-point at the desired value of pH 9. A further set-point change to pH 6 is then made and the performance deteriorates as shown in Fig. 9. This can be explained by the fact that the model is poor in such a region because there are scarce data and information available from the training set. With the availability of the new data and based on the selection, the criteria data are admitted when they fall within the uncertain region of the model. They can be used to enrich the model. Based on the proposed heuristic method, the region of pH 5.5 is greater than the threshold of 0.1. The threshold is selected based on an estimate of the variance of the process variable which is 0.1 in this case. The corresponding input vector is $[900 \ 896 \ 5.6]$ and $r = 2$ is chosen. In the simulation, 10 sets of new data are added. The red line represents the set-point; blue line represents the pH and (*) represents the admitted process variables that update the model in Fig. 9. It is noted that Figure 7 and Figure 9 show the same results of the GP PID control with variance before the model is updated. The admitted new input vectors are labelled as star (*) and the training input vectors are labelled as circle (o) as shown in Fig. 10. Fig. 11 shows that on the subsequent return to the set-point at pH 6, the control performance has improved. Note that Fig. 11 shows the result after two cycles of change from pH 9 to 6. The model is now more accurate because of using the proposed heuristic method to add data in Section 3.2. It has improved the prediction which leads to improved control.

The above description illustrates the importance of the model accuracy; in fact, all model based control methods share the same dependency on the model accuracy. However, unlike the conventional methods, the information from the GP model is used to carry out data selection; that is, not all the data will be admitted, which is advantageous in terms of computation load. The selection criterion implies that only the useful data are used for calculation and there is no extra effort expended on the data that do not contribute to the accuracy of the process model.

5.2 Fed-batch fermentation

1 In this case study, the fed-batch fermentation bioreactor which has been well studied in
 2 various applications is used [29]. The process dynamic behavior is described by the
 3 following set of differential equations,

$$\begin{aligned}\frac{dS}{dt} &= -\sigma X + \frac{u}{V}(S_{in} - S) \\ \frac{dX}{dt} &= \mu X - \frac{u}{V} X \\ \frac{dV}{dt} &= u\end{aligned}\tag{39}$$

4 where u is the feed flow rate of substrate. S is the substrate concentration; X , the
 5 biomass concentration and V , the volume of the reaction mixture. The specific growth
 6 rate is of the Haldane type,

$$\mu = \frac{\mu_m S}{K_p + S + S^2/K_i}\tag{40}$$

7 with three parameters, μ_m , K_p and K_i . The substrate consumption rate is given by

$$\sigma = \frac{\mu}{Y} + m\tag{41}$$

8 where the yield Y and maintenance factor m are assumed to be fixed. The process
 9 variable is the biomass concentration, X , and the manipulated variable is the substrate
 10 feed flow rate, u .

11 In the batch process, the growth of the biomass concentration follows a desired trajectory.
 12 However, the biomass concentration is difficult to obtain instantaneously and is often
 13 available at the end of the batch in the lab. This lack of data can yield a less satisfactory
 14 process model. The tracking performance of the batch process will depend on the model
 15 and this is particular important for the batch process where on-line measurement is not
 16 easy to obtain. [Dondo and Marques \[30\]](#) presented a numerical study on the performance
 17 degradation of a batch bioreactor optimal control strategy when the model used to
 18 compute such strategy shows wide variations from the real process. The model was

improved by conducting new experiments but there is lack of information on where the experiments should be carried out to improve the model. To this end in this work using the variance information the location of the experimental data for the model improvement can be obtained. For the case study, an initial test found that $m=3$ and $l=3$ are appropriate for describing the dynamics of the process. Compared to the pH neutralization, the fed-batch fermentation is affected greater by past dynamics. Moreover as in the pH fermentation $\lambda=2$ is used so that the controller does not make a sudden, drastic change.

If the model is not good, the performance is inevitably unsatisfactory as shown in Fig. 12(a). The large variance implies that the result is not very reliable, the training data is lacking and the model is unable to provide a good prediction, resulting in the bad performance. It is noted that in Fig. 12 the shaded region is angular as only the variance prediction is plotted at the sampling time. Because of the repetitive nature of the batch process, the data from each run can be used to improve the model. When the data from the run is used to update the model, it is found that control based on this improved model is able to make the performance better than the previous run. Fig. 12(b) shows the result of control after the model has been updated. Prior to this run, data of $t=5$ to $t=10$ are added as the variance exceeds the threshold. This addition of data has improved the model prediction for the relevant period of the operation. The control results show the mid-section of the operation has been improved. With each successive run, new data are available from the operation. Subsequently, 4 ($t=7$ to $t=10$) data sets in the end-section are added to improve the model. These data which provide the new and the useful information can be used to improve the model. With the availability of an improved model, the controller tuning becomes more accurate and further improvement in the performance is as shown in Fig. 12(c). The performance from the approximate GP is shown as (Δ) from Fig. 12(a) to Fig. 12(c). It can be seen that the performance of the approximate GP model is comparable to that of the direct GP method.

As an indication of convergence which tells the user if the model needs to be further updated the control performance P is defined the average sum of the squared error and the variance

$$P = \frac{1}{N} \left(\sum e^2 + \sum \sigma_{t+1}^2 \right) \quad (42)$$

where N is the number of sampling points and e , the error. Fig. 13 shows P against batch number. It can be seen that at each successive batch, the sum is decreasing, indicating that the error and the variance have decreased. The performance after batch no. 3 reaches a steady value, implying the model is accurate after batch no.3, and as shown in Fig 12(c), the process variable follows the reference trajectory closely. In terms of the safety, the controller will safeguard against a highly unsafe operation, so it can avoid the consequence of the process 'runaway'. In term of the performance, the proposed method is able to drive the process toward the optimal performance based on the improvement made to the process model at each successive run using historical data from the completed run. These benefits of the safe and improving operation should provide an attractive solution to many industrial nonlinear control problems.

6. Conclusion

In this work, the variance information provided by the GP model is used in the PID controller design. The objective takes into account the prediction and the variance which results in a safety-performance trade-off control due to the controller avoiding the region with large variance at the cost of not tracking the set point to ensure process safety. A heuristic selection of data is also proposed to improve the models at the successive stage to improve control performance as the model uncertainty is reduced. Furthermore, an approximation algorithm decreases the computation load when the control action is evaluated is proposed. The proposed method is applied to the pH neutralization control system and the fed-batch fermentation process. In the pH neutralization case study, the necessity of taking into account the model uncertainty represented by the predictive

variance is highlighted. When the identified GP model is not accurate the controller avoids the region with large variance at the cost of not tracking the set point. The fed-batch fermentation demonstrated the capability of the proposed method when the set-point is a time varying profile. The ability to improve the model through the addition of data is clearly shown in the fed-batch fermentation process. The identification of the lower confidence region provides a mean to overcome missing data problem whose measurement may not be readily available. Both case studies demonstrated the applicability and versatility of the proposed method to improve model uncertainty for PID controller design.

References

- [1] J.A. Romero, R. Sanchis, P. Balaguer, PI and PID auto-tuning procedure based on simplified single parameter optimization, *Journal of Process Control*, 21 (2011) 840-851.
- [2] A. Leva, M. Maggio, A systematic way to extend ideal PID tuning rules to the real structure, *Journal of Process Control*, 21 (2011) 130-136.
- [3] V.M. Alfaro, R. Vilanova, Performance and Robustness Considerations for Tuning of Proportional Integral/Proportional Integral Derivative Controllers with Two Input Filters, *Industrial & Engineering Chemistry Research*, 52 (2013) 18287-18302.
- [4] H. Zhang, Y. Shi, A.S. Mehr, Robust $\mathcal{H}^\infty$ PID control for multivariable networked control systems with disturbance/noise attenuation, *International Journal of Robust and Nonlinear Control*, 22 (2012) 183-204.
- [5] O. Garpinger, T. Hägglund, K.J. Åström, Performance and robustness trade-offs in PID control, *Journal of Process Control*, 24 (2014) 568-577.
- [6] L. Wang, W. Du, H. Wang, H. Wu, Fuzzy self-tuning PID control of the operation temperatures in a two-staged membrane separation process, *Journal of Natural Gas Chemistry*, 17 (2008) 409-414.
- [7] A. Altınten, F. Ketevanlioğlu, S. Erdoğan, H. Hapoğlu, M. Alpbaz, Self-tuning PID control of jacketed batch polystyrene reactor using genetic algorithm, *Chemical Engineering Journal*, 138 (2008) 490-497.

- [8] C.-H. Lee, C.-C. Teng, Calculation of PID controller parameters by using a fuzzy neural network, *ISA Transactions*, 42 (2003) 391-400.
- [9] Y. Kansha, L. Jia, M.-S. Chiu, Self-tuning PID controllers based on the Lyapunov approach, *Chemical Engineering Science*, 63 (2008) 2732-2740.
- [10] O. Arrieta, A. Visioli, R. Vilanova, PID autotuning for weighted servo/regulation control operation, *J. Process Control*, 20 (2010) 472-480.
- [11] A. Leva, S. Negro, A.V. Papadopoulos, PI/PID autotuning with contextual model parametrisation, *J. Process Control*, 20 (2010) 452-463.
- [12] R. Toscano, A simple robust PI/PID controller design via numerical optimization approach, *J. Process Control*, 15 (2005) 81-88.
- [13] J. Kocijan, R. Murray-Smith, C.E. Rasmussen, A. Girard, Gaussian process model based predictive control, in: *American Control Conference, 2004. Proceedings of the 2004, IEEE, 2004*, pp. 2214-2219.
- [14] J. Chen, T.-C. Huang, Applying neural networks to on-line updated PID controllers for nonlinear process control, *Journal of Process Control*, 14 (2004) 211-230.
- [15] T. Pan, S. Li, W.-J. Cai, Lazy Learning-Based Online Identification and Adaptive PID Control: A Case Study for CSTR Process, *Industrial & Engineering Chemistry Research*, 46 (2007) 472-480.
- [16] A. O'Hagan, J.F.C. Kingman, Curve Fitting and Optimal Design for Prediction, *Journal of the Royal Statistical Society. Series B (Methodological)*, 40 (1978) 1-42.
- [17] C.E. Rasmussen, Evaluation of Gaussian processes and other methods for non-linear regression, in, *Citeseer*, 1996.
- [18] K. Ažman, J. Kocijan, Application of Gaussian processes for black-box modelling of biosystems, *ISA Transactions*, 46 (2007) 443-457.
- [19] A. Lundgren, J. Sjöberg, Gaussian processes framework for validation of linear and nonlinear models, in: *System Identification (SYSID'03): A Proceedings Volume from the 13th IFAC Symposium on System Identification, Rotterdam, the Netherlands, 27-29 August 2003, Elsevier, 2004*, pp. 67.

- [20] F. di Sciascio, A.N. Amicarelli, Biomass estimation in batch biotechnological processes by Bayesian Gaussian process regression, *Computers & Chemical Engineering*, 32 (2008) 3264-3273.
- [21] W. Ni, K. Wang, T. Chen, W.J. Ng, S.K. Tan, GPR model with signal preprocessing and bias update for dynamic processes modeling, *Control Engineering Practice*, 20 (2012) 1281-1292.
- [22] L.L.T. Chan, Y. Liu, J. Chen, Nonlinear System Identification with Selective Recursive Gaussian Process Models, *Industrial & Engineering Chemistry Research*, 52 (2013) 18276-18286.
- [23] W. Yan, S. Hu, Y. Yang, F. Gao, T. Chen, Bayesian migration of Gaussian process regression for rapid process modeling and optimization, *Chemical Engineering Journal*, 166 (2011) 1095-1103.
- [24] B. Likar, J. Kocijan, Predictive control of a gas–liquid separation plant based on a Gaussian process model, *Computers & Chemical Engineering*, 31 (2007) 142-152.
- [25] Y.A.W. Shardt, B. Huang, Data quality assessment of routine operating data for process identification, *Computers & Chemical Engineering*, 55 (2013) 19-27.
- [26] J. Sun, G.S. Hong, Y.S. Wong, M. Rahman, Z.G. Wang, Effective training data selection in tool condition monitoring system, *International Journal of Machine Tools and Manufacture*, 46 (2006) 218-224.
- [27] E.P. Nahas, M.A. Henson, D.E. Seborg, Nonlinear internal model control strategy for neural network models, *Computers & Chemical Engineering*, 16 (1992) 1039-1057.
- [28] B. Zhang, W. Zhang, Adaptive predictive functional control of a class of nonlinear systems, *ISA Transactions*, 45 (2006) 175-183.
- [29] K.J. Versyck, J.F. Van Impe, Feed rate optimization for fed-batch bioreactors: from optimal process performance to optimal parameter estimation, *Chemical Engineering Communications*, 172 (1999) 107-124.
- [30] R.G. Dondo, D. Marqués, Optimal control of a batch bioreactor: a study on the use of an imperfect model, *Process Biochemistry*, 37 (2001) 379-385.

Table 1. Parameters for case study 1

$A = 207 \text{ cm}^2$	$C_v = 8.75 \text{ ml cm}^{-1} \text{ s}^{-1}$	$\text{pK1} = 6.35$
$\text{pK2} = 10.25$	$W_{a1} = 3 \times 10^{-3} \text{ M}$	$W_{a2} = -3 \times 10^{-2} \text{ M}$
$W_{a3} = -3.05 \times 10^{-3} \text{ M}$	$W_{b3} = 5 \times 10^{-5} \text{ M}$	$q_1 = 16.6 \text{ ml s}^{-1}$
$q_2 = 0.55 \text{ ml s}^{-1}$	$q_3 = 15.6 \text{ ml s}^{-1}$	$[\text{Acid}] = 0.003 \text{ M HNO}_3$
$[\text{Buffer}] = 0.03 \text{ M NaHCO}_3$	$[\text{Base}] = 0.003 \text{ M NaOH}$

Fig. 1. The GP model based PID control scheme.

Fig. 2. The data admission region.

Fig. 3. Instantaneous linearization of the GP model.

Fig. 4. pH CSTR system.

Fig. 5. Comparison of direct GP PID control with variance and direct GP PID control without variance when the model is accurate in the pH neutralization system.

Fig. 6. Comparison among direct GIP PID control, approximate GP PID control and linearized neural-network when the model is accurate in the pH neutralization system.

Fig. 7. Comparison between direct GP PID control with variance and direct GP PID control without variance when the model is inaccurate in the pH neutralization system.

(a)

(b)

Fig. 8. PID parameter tuning in the pH neutralization system: (a) GP with variance (b) GP without variance.

Fig. 9. GP PID control with initial data which are rich in the region of pH 8; the dotted points are pH value of high uncertainty used to update the model in the pH neutralization system.

Fig. 10. Historical data ($\circ$) and new admitted data ($*$) in the input space of the pH neutralization system

Fig. 11. GP PID control with the updated model from the admitted data in the pH neutralization system.

(a)

(b)

(c)

Fig. 12. Evolution of GP PID control in the fed-batch fermentation bioreactor: (a) Initial batch (b) Batch No. 2 (c) Batch No. 3

Fig. 13. Plot of control performance vs batch number in the fed-batch fermentation bioreactor