ADCONIP 2014 Hiroshima

Applying Gaussian Process Models to On-line Enhancement of PID Control Design for Nonlinear Processes

Lester Lik Teck Chan, Tao Chen, and Junghui Chen*

***Abstract*— The inherent time-varying nonlinearity and complexity usually exist in chemical processes. The design of the control structure should be properly adjusted based on the current state. However, the design is based on the model of the process, so it is directly affected by the quality of the model. Because the Gaussian process (GP) model provides predictive distribution of the output and an estimate of the variance of the predicted output, PID tuning control based on the GP model is proposed in this study. The variance information is used for control to yield safety-performance trade-off. In addition, it provides a mean for the selection of data to improve the model at the successive control stage. This allows the controlled process to converge from the performance and safety trade-off to an optimal performance when the model is accurate because of the reduced model uncertainty. A case study on pH neutralization has been carried out to illustrate the applicability of the proposed method in PID tuning.**

I. Introduction

Despite the advent of many complicated control theories and techniques, more than 95% of the control loops based on proportional-Integral-Derivative (PID) controllers are still being used in the majority of industrial processes because of the ease of its use and relative higher cost of more advanced control systems. Nevertheless, the PID algorithm may have difficulty dealing with highly nonlinear and time varying chemical processes. In the past, several schemes of self-tuning PID controllers were proposed. Recent work includes fuzzy-self tuning PID control of the operation temperatures in a two-staged membrane separation process [1] and self-tuning PID control of the jacketed batch polystyrene reactor using the genetic algorithm [2]. The integrity of the model is, therefore, very important and information on the model based prediction can be invaluable. The construction of nonlinear models is difficult and there is a lack of necessary trust in the model [3]. Chen and Huang (2004) has applied the linearized neural network based model to the tuning of the PID controllers [4] whereas an approach based on a lazy learning identification has been proposed [5]. The Gaussian process (GP) model provides a natural way to evaluate the variance prediction in addition to the predictive value which gives the information on the reliability of the prediction. GP models have been increasingly applied to different nonlinear dynamic systems. Azman and Kocijan (2007) applied GP to the modeling of the bio-system and nonlinear systems [6] whereas Lundgren and Sjoberg (2003) used the GP model for linear and nonlinear model validation [7]. Ni et al. (2012) proposed a recursive GP to adapt to the process drifted in both sample-wise and block-wise manners with a filter to preprocess the output target, improving the accuracy of the prediction [8]. GPs have several key properties, including fewer optimizing parameters and the capabilities to provide predictive variance and indicate the reliability of the output in the local stochastic. The objective of this work is to use Gaussian model to provide variance information for the control method. This way, a trade-off between safety and control performance can be achieved. Moreover, the variance information also facilitates the selection of data which is carried out to improve the models at the successive stage and in return results in the control scheme converging from the performance and safety trade-off to a true optimal performance as the model uncertainty is reduced. In comparison with the conventional PID tuning, this work provides a direct evaluation of the predictive variance using GP. This added information consequently facilitates the identification of the region for improvement.

L. L. T. Chan is with Department of Chemical Engineering, Chung-Yuan Christian University, Chung-Li, Taiwan 320, Republic of China.

T. Chen is with Department of Chemical and Process Engineering, University of Surrey, Guildford, GU2 7XH, UK

J. Chen is with Department of Chemical Engineering, Chung-Yuan Christian University, Chung-Li, Taiwan 320, Republic of China (fax: 886-3-265-4199; e-mail: jason@ wavenet.cycu.edu.tw).

II. Problem Statement

Fig. 1 shows the PID control scheme considered in this study. The process output is y_t, and u_t is the controller input.

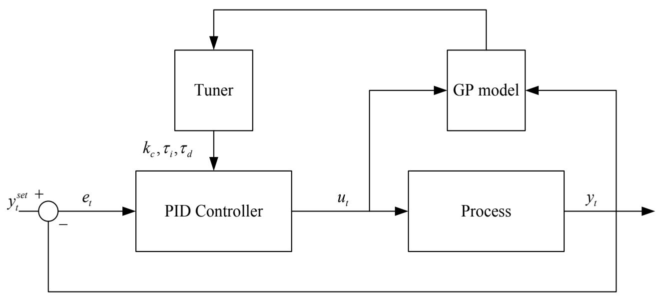


Fig. 1. The GP model based PID control scheme

The discrete and velocity form of a PID controller can be expressed as

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

where $\mathbf{k}_t = \begin{bmatrix} k_t^0 & k_t^1 & k_t^2 \end{bmatrix}^T$, and

$$k_t^0 = k_c\left(1+\frac{\Delta t}{2\tau_i}+\frac{\tau_d}{\Delta t}\right),\quad k_t^1 = -k_c\left(1+\frac{2\tau_d}{\Delta t}\right),$$
$$k_t^2 = k_c\frac{\tau_d}{\Delta t} \tag{2}$$

k_c, τ_i, τ_d are the proportional gain, integral time constant and the derivative time constant, respectively. $\mathbf{e}_t = \left[e_t \quad e_{t-1} \quad e_{t-2}\right]^T$ and $e_t = y_t^{set} - y_t$.

The parameters of the PID controller are adjusted by a tuner based on an identified process model. In this study, a GP regression model provides a mean value prediction as well as the variance prediction. The variance prediction can be viewed as the information about the model confidence. As a result, in contrast to traditional applied control, the issue of control system robustness should be considered. This issue has a major impact on the applicability of the controller scheme. For instance, a process out of control or with too aggressive control can pose a threat to the safety of a plant. The information on the confidence of the prediction enables a trade-off between designed performance and safety to be considered. Based on this information, the PID controller can be tuned based on the requirement of the process.

When the model is not accurate, there is a trade-off between performance and safety. To achieve the best performance without compromising the safety, one straightforward way is to reduce the model uncertainty. Over the course of operation, new data are available to enrich the model and provide a better prediction. In this work, the GP model is used for identification. With the information on the uncertainties of the prediction provided by the model, a simple heuristic method is used to select the data for the improvement of the model. Moreover, a tuning scheme with the objective function based on the uncertainty information is also proposed to provide satisfactory control.

III. Progressing Gaussian Process Models

A. Gaussian Process Models

A GP model is a collection of random variables which have a joint Gaussian distribution. Given data $\mathbf{D} = (\mathbf{X}, \mathbf{y})$, the inference on y_{t+1} can be readily obtained since the joint density $P(y_{t+1}, \mathbf{y})$ is also Gaussian, the posterior distribution is given by

$$P(y_{t+1}|\mathbf{y}) = \frac{1}{Z}\exp\left[-\frac{(y_{t+1}-\mu_{t+1})^2}{2\sigma^2_{\mu_{t+1}}}\right] \tag{3}$$

where

$$\mu_{t+1} = \mathbf{g}^T(\mathbf{x}_{t+1})\mathbf{C}^{-1}\mathbf{y} \tag{4}$$

$$\sigma^2_{t+1} = C(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}) - 2\mathbf{g}^T(\mathbf{x}_{t+1})\mathbf{C}^{-1}\mathbf{g}^T(\mathbf{x}_{t+1}) \tag{5}$$

μ_{t+1} is the mean prediction at $\mathbf{x}_{N+1}$ and σ^2_{t+1} is the standard deviation of this prediction and $\mathbf{g}^T(\mathbf{x}_{t+1}) = \left[C(\mathbf{x}_{t+1}, \mathbf{x}_1) \quad \cdots \quad C(\mathbf{x}_{t+1}, \mathbf{x}_t)\right]$. $\mathbf{C}$ is the covariance matrix of the data defined by the parameterized covariance function $C_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$. The vector $\mathbf{g}^T(\mathbf{x}_{t+1})\mathbf{C}^{-1}$ can be viewed as a smoothing term which weights the training outputs to make a prediction for the new input vector $\mathbf{x}_{t+1}$. Eq (5) provides a confidence level on the model prediction as the higher variance value indicates the region of the input vector contains few data or is corrupted by noise. GP is fully defined by its mean and variance. The covariance function is non-trivial and a common choice is

$$C(\mathbf{x}_i, \mathbf{x}_j) = a_0 + a_1\sum_{k=1}^{d} x_{ik}x_{jk} + v_0\exp\left(-\sum_{k=1}^{d} w_k(x_{ik}-x_{jk})\right) + s\delta_{ij} \tag{6}$$

where $a_0, a_1, d, w_k, s, \delta_{ij}$ and v_0 are the hyper-parameters to be determined. Hyper-parameter v_0 controls the overall scale of the local correlation, a_1 allows a different distance measure in each input dimension, and s is the estimate of the noise variance. The hyper-parameters can be obtained by the maximization of the log-likelihood. [6]

B. Progressing GP model update

The tuning and the performance of the proposed scheme will depend on the accuracy of the identified model. The model may be improved by using new data. Without data selection, the new data may not improve prediction and they may cause a waste on computation. To this end, the GP model provides a straightforward way of selecting a criterion.

Besides the mean of the predictive distribution, the GP model also provides the variance that indicates the uncertainty in the model. 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 in the region of high uncertainty. A simple heuristic method is proposed. Based on the predictive variance of the output, only the data in close proximity of the highly uncertain area are admitted. The data points are selected based on the distance from the point with high variance. When a particular point has higher uncertainty, the neighboring data can be included to provide richer information for the point, so the model performance can be improved. Consequently, the overall control performance can be enhanced. The admitted data, y_{adm} will satisfy the criterion

$$y_{adm} \in \left[y - w\sigma \quad y + w\sigma\right] \tag{7}$$

where w is the present value that determines the range of the allowable data (Fig. 2). In Fig. 2, the solid line represents the actual process output while the dotted line represents the GP model. The grey shaded band represents $\mu \pm 2\sigma$ and gives an indication of the uncertainty in prediction of y. The yellow band is the admissible region, meaning that the output data within this range and the corresponding input data will be used to update the model subsequently.

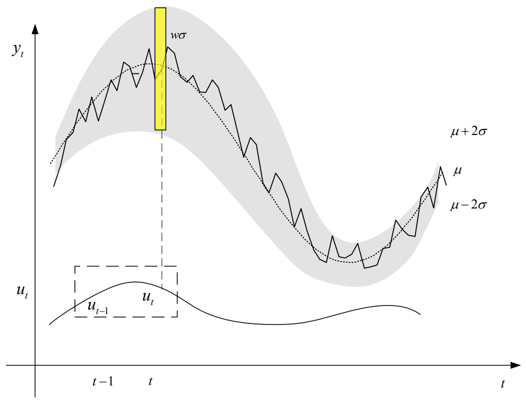


Fig. 2 The data admission region

At the successive control stage, the model may be updated with new data and the control action can be evaluated based on the updated model. Let a particular admissible output data at time t be y_{new} and the corresponding input data which consist of the current value and a past value be $\mathbf{x}_{new}$ (Fig. 2).

$$\mathbf{x}_{new} = \begin{bmatrix} u_t & u_{t-1} & y_t \end{bmatrix}^T \tag{8}$$

An operation that adds the new data to the initial data set is defined as

$$\mathbf{X}^+ \leftarrow \begin{bmatrix} \mathbf{X} & \mathbf{x}_{new} \end{bmatrix} \tag{9}$$

$$\mathbf{y}^+ \leftarrow \begin{bmatrix} \mathbf{y} & y_{new} \end{bmatrix} \tag{10}$$

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

IV. Gaussian Process Model Based PID Tuning

The identified GP model can be directly applied to tuning PID controllers. By minimizing an objective function, the optimal control action can be obtained. This solution can be obtained by the gradient method with the necessary gradient calculation. The objective function for minimum variance control is

$$J = E\left\{\left(y_{t+1}^r - \mu_{t+1}\right)^2\right\} + \lambda \Delta u_t^2 \tag{10}$$

where y_{t+1}^r is a reference target. The relative importance between the reference tracking and aggressiveness is determined by λ. Because $\mathrm{Var}\{y\} = E\{y^2\} - E^2\{y\}$ and GP provides uncertainty in terms of prediction variance, the objective function can be written as

$$J = \left(y_{t+1}^r - \mu_{t+1}\right)^2 + \sigma_{t+1}^2 + \lambda \Delta u_t^2 \tag{11}$$

PID is tuned by minimizing the objective function J with the aim to design a control action that will minimize the difference between the process output and the desired output, and the variance of the controller output with consideration to the model uncertainty. Compared to conventional PID tuning, the variance prediction term is included with the mean prediction. This means that the optimization takes into account this variance information, resulting in a more robust control system.

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

$$\frac{\partial J}{\partial \mathbf{k}_t} = \frac{\partial J}{\partial u_t}\frac{\partial u_t}{\partial \mathbf{k}_t} \tag{12}$$

where

$$\frac{\partial J}{\partial u_t} = 2\left(\mu_{t+1} - y_{t+1}^r\right)\frac{\partial \mu_{t+1}}{\partial u_t} + \frac{\partial \sigma_{t+1}^2}{\partial u_t} + 2\lambda \Delta u_t \tag{13}$$

and

$$\frac{\partial u_t}{\partial \mathbf{k}_t} = \mathbf{e}_t \tag{14}$$

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

$$\frac{\partial \mu_{t+1}}{\partial u_t} = \frac{\partial \mathbf{g}^T(\mathbf{x}_{t+1})}{\partial u_t}\mathbf{C}^{-1}\mathbf{y} \tag{15}$$

and

$$\frac{\partial \sigma_{t+1}^2}{\partial u_t} = \frac{\partial C(\mathbf{x}_{t+1}, \mathbf{x}_{t+1})}{\partial u_t} - 2\mathbf{g}^T\mathbf{C}^{-1}\frac{\partial \mathbf{g}}{\partial u_t} \tag{16}$$

where $\mathbf{x}_{t+1} = [u_t \ \cdots \ u_{t-m+1} \ y_t \ \cdots \ y_{t-l+1}]^T$. The subscripts, m and l, refer to the past m inputs and the past l outputs respectively.

Furthermore, the derivative of the covariance function in Eq. (6) is

$$\frac{\partial C(\mathbf{x}_{t+1},\mathbf{x}_i)}{\partial u_t} = a_1 x_{i1} + v_0 \exp\left(-\sum_{k=1}^{d} w_k (x_{ik} - x_{t+1,k})^2\right) \times 2 w_1 (x_{i1} - u_t) \quad (17)$$

with

$$\frac{\partial C(\mathbf{x}_{t+1},\mathbf{x}_{t+1})}{\partial u_t} = a_1 u_t \quad (18)$$

and

$$\frac{\partial \mathbf{g}}{\partial u_t} = \begin{bmatrix} \frac{\partial C(\mathbf{x}_{t+1},\mathbf{x}_1)}{\partial u_t} & \cdots & \frac{\partial C(\mathbf{x}_{t+1},\mathbf{x}_n)}{\partial u_t} \end{bmatrix}^T \quad (19)$$

Calculating each term using Eqs.(13)-(19), the Jacobian matrix is obtained and the optimal control action that minimizes Eq.(11) can be evaluated. The control action can be determined by evaluating the minimal of Eq.(12) and the corresponding PID parameters are obtained. This can be directly applied to control of the process in order to achieve the desired output.

V. CASE ~~CASED~~ STUDY

pH neutralization is fairly common among many chemical processes and in industries such as wastewater treatment. The model equations in this study are taken from Nahas et al.[9]. It is a pH continuous stirred-tank reactor (CSTR) system as shown in Fig. 3. The three input streams are acid (HNO_3), buffer ($NaHCO_3$) and base (NaOH) respectively.

The system can be described by two reaction invariants, three nonlinear ordinary differential equations and one nonlinear algebraic equation.

Charge balance

$$W_a \equiv [H^+] - [OH^-] - [HCO_3^-] - 2[CO_3^=] \quad (20)$$

Carbonate ion balance

$$W_b \equiv [H_2CO_3] + [HCO_3^-] + [CO_3^=]$$

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

$$\frac{dW_{a4}}{dt} = \frac{1}{Ah}\left[(W_{a1} - W_{a4})q_1 + (W_{a2} - W_{a4})q_2 + (W_{a3} - W_{a4})q_3\right] \quad (21)$$

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

$$W_{a4} + 10^{14\text{-pH}} + W_{b4}\frac{1 + 2\times 10^{\text{pH}-\text{pK}_2}}{1 + 10^{\text{pK}_1-\text{pH}} + 10^{\text{pH}-\text{pK}_2}} - 10^{\text{-pH}} = 0 \quad (22)$$

where h is the liquid level, W_{a4} and W_{b4} are reaction invariants of the effluent streams, with q_1, q_2 and q_3 being the acid, buffer and the base flow rate respectively. The nominal conditions and parameters are listed in Table 1

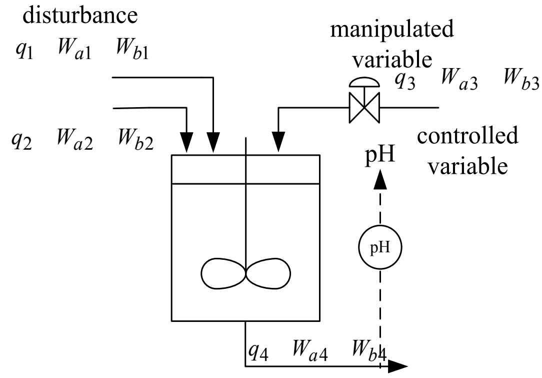


Fig. 3 pH CSTR systems

TABLE I. PARAMTERS FOR CASE STUDY

$A = 207\ \text{cm}^2$	$C_v = 8.75\ \text{ml cm}^{-1}\ \text{s}^{-1}$
$\text{pK2} = 10.25$	$W_{a1} = 3\times10^{-3}\ \text{M}$
$W_{a3} = -3.05\times10^{-3}\ \text{M}$	$W_{b3} = 5\times10^{-5}\ \text{M}$
$q_2 = 0.55\ \text{ml s}^{-1}$	$q_3 = 15.6\ \text{ml s}^{-1}$
[Buffer] = 0.03 M $NaHCO_3$	[Base] = 0.003 M NaOH
$\text{pK1} = 6.35$	
$W_{a2} = -3\times10^{-2}\ \text{M}$	
$q_1 = 16.6\ \text{ml s}^{-1}$	
[Acid] = 0.003 M HNO_3	

In this case study, the variance term used in the objective functions (Eq.(11)) will be shown useful and then the control performance will be presented using the proposed method. For the first goal, the data used for the training are obtained from the region of pH 7. The distribution of the input and output data are concentrated in the region of the proximity of pH 7 and as a result, the identification of this region will be satisfactory.

In the event that the collected data are rich enough, a comparison of the proposed method with neural network model based control [4] is made. Fig. 4 shows the result from GP PID control based on the direct application of the GP model when the data are sufficient implying a good prediction and low variance. Again, the set point is changed from pH 7 to pH 9 and then from pH 9 to pH 7. It is found that the control results are similar for all the 3 cases, including the NN based model, the GP model without the variance and the GP model with the variance. The GP model without the variance refers to an objective function that is Eq.(11) without the term, σ_{t+1}^2, on the right-hand side,

$$J = \left(y_{t+1}^r - \mu_{t+1}\right)^2 + \lambda \Delta u_t^2 \qquad (23)$$

the NN model and the GP model are similar in terms of the model performance. The control algorithm based on these models should be similar (Fig. 4). Moreover, the data used in identification are sufficient in the region of operation, so the variance is small and accordingly the confidence in the prediction is high. The GP model with the variance results in the performance similar to the model without the variance.
As the actual operation process data may be insufficient, the model built upon these insufficient data is not accurate. The consequence of using this model can have a drastic effect on the control outcome as illustrated in Fig. 5. It shows the result of the inaccurate model and a comparison between the cases of using and not using the variance term in the control objective. When the variance term is not taken into account, an unreliable prediction value is used without any additional check in the control action (Eq.(11)). It results in a large offset from the set point represented by the dash-dot line in Fig. 5. This is because the prediction of the calculated input has a large prediction error. On the other hand, if the variance is taken into account, the compromise between the performance and reliability means that the eventual results do not deviate as much while the set point is not reached. As this region has less data, a higher variance would occur. The response of the system is optimized so that it can "hold off" the system from reaching the region with high variance. Fig. 6 shows the comparison of the PID parameters for GP with and without variance. It is shown that the tuning performes better in the case of GP with ~~without~~ variance. This can be viewed as a trade-off between safety and the performance.

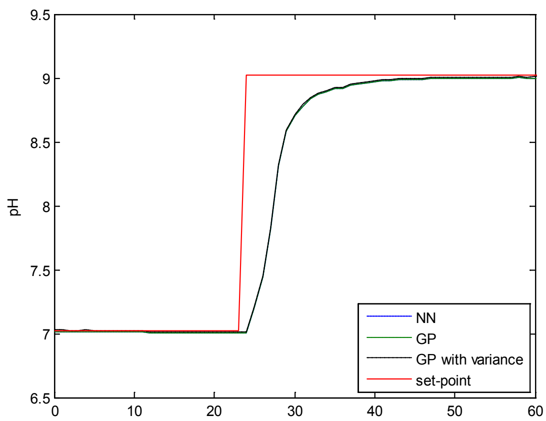


Fig. 4 Comparison of direct GP PID control with variance, direct GP PID control without variance and the neural network when the model is accurate

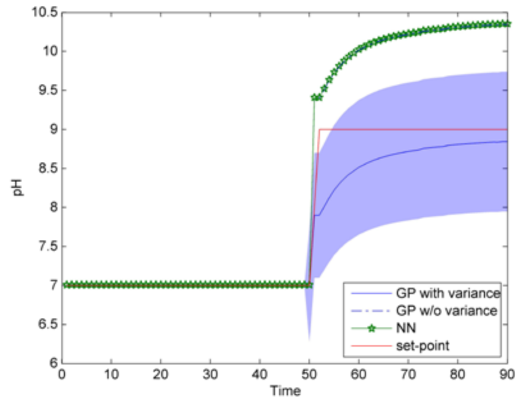
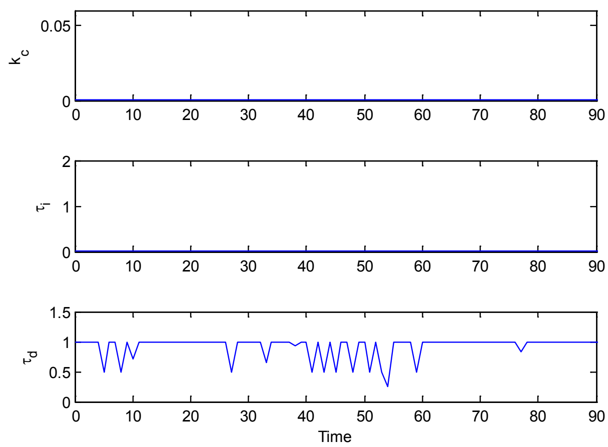


Fig. 5 Comparison of direct GP PID control with variance, direct GP PID control without variance and the neural network when the model is inaccurate.

(a)



(b)

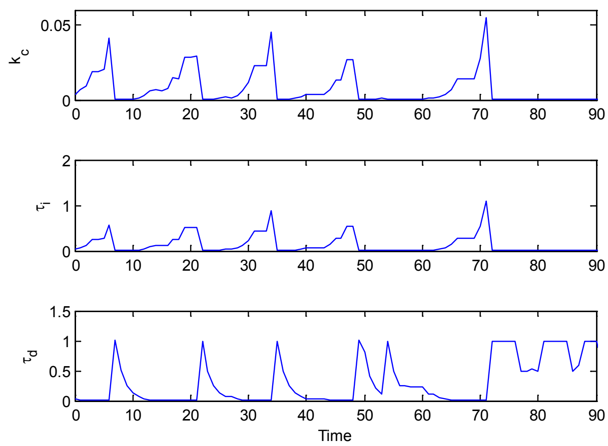


Fig. 6 PID parameter tuning: (a) GP with variance (b) GP without variance

In this case study, the proposed method is applied to a continuous process with the changing set-point to demonstrate its effect on the control performance and model improvement. The process is initially operated at pH 8. Subsequently the set point is changed from pH 8 to pH 9 and then back to pH 8. This change is done for a number of repetitions to show how the process will respond to the proposed method. Initially, the model is only accurate in the proximity of the region of pH 8. The model is updated at each change of conditions. With this model, the controller is able to achieve the set-point at the desired value of pH 8. When the set-point is changed from pH 8 to pH 9 ($t = 21 \sim 70, t = 121 \sim 170$) the control performance drops slightly as shown in Fig. 7. This performance drops a little because the information of this region is not rich. As the operation is in the close proximity to the initial data and there is a relatively small error band, the performance in this case is acceptable. A further set-point change to pH 6 is then made and the performance deteriorated more drastically. 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 data. With the availability of the new data and based on the selection criteria, data are admitted when they fall within the uncertain region of the model and can be used to enrich the model. The region of the admitted data is represented by the dotted circle in Fig. 7 . Fig. 8 shows that on subsequent return to the set point at pH 8, the control performance is improved. The model, because of the additional data, is now more accurate, so the improvement in the predicted value makes the control performance better. Improvement in the control performance is also exhibited when the set point is changed to pH 6 and further improvement to the model can help attain the desired pH.

The above case illustrates the importance of the accuracy of the model in this proposed model; in fact, all model based control depends on the model accuracy. However, compare to conventional methods, the information from the GP model is used to carry out data selection because not all the data are admitted and advantageous in terms of computational load. The selection criterion implies that only useful data are used for calculation and no extra effort is needed on the data that does not contribute to the accuracy of the process model.

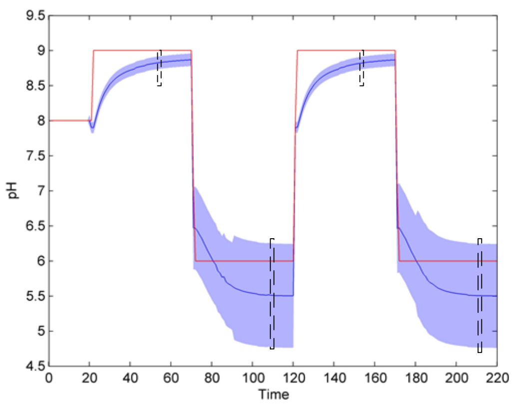


Fig. 7 GP PID control with initial data rich in the region of pH 8; data in dotted-line regions are used to update the model.

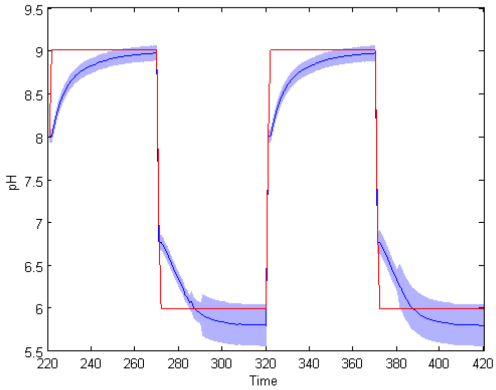


Fig. 8 GP PID control with updated model from admitted data.

VI. Conclusion

For model based control, the integrity of the model is important. A new control approach using a GP model around its current operation region is developed. The GP approach to modeling provides a prediction as well as variance on the predictive distribution. This provides a quantitative measure on the trust of model. A safety-performance trade-off control goal can be achieved based on the controller objective which takes into account the prediction and the variance. In the future, the framework would be extended to other model based control.

Acknowledgment

The authors wish to express their sincere gratitude to Chung-Yuan Christian University and National Science Council (NSC) for their financial support.

References

[1] L. Wang, W. Du, H. Wang, H. Wu, "Fuzzy self-tuning PID control of the operation temperature in a two-staged membrane separation process," J. Natural Gas. Chem. vol.17, pp.409-414, 2008.

[2] A. Altinten, F. Ketevanlioğlu, S. Erdoğan, H. Hapoğlu, M. Alpbaz, "Self-tuning PID control of jacketed batch polystyrene reactor using genetic algorithm," Chem. Eng. J. vol.138, pp.490-497, 2008.

[3] J. Kocijan, R. Murray-Smith, C.Rasmussen, A. Girard, "Gaussian process model based predictive control," American Control Conference. Boston, USA (2004)

[4] J. Chen, T.C. Huang, "Applying neural networks to on-line updated PID controllers for nonlinear process control," J. Proc. Cont. vol.14, pp.211–230, 2004.

[5] T. Pan, S. Li, W.-J. Cai, "Lazy learning based online identification and adaptive PID control: A case study for CSTR process," Ind. Eng. Chem. Res., vol.46, pp.472-480, 2007.

[6] K. Azman, J. Kocijan, "Application of Gaussian processes for black-box modeling of bio-systems," ISA Transactions, vol.46, pp.443-457, 2007.

[7] A. Lundgren, J. Sjoberg, "Gaussian processes framework for validation of linear and nonlinear models." In 13th IFAC Symposium on System Identification (SYSID 2003), Rotterdam, The Netherlands, 67-72.

[8] W. Ni, K. Wang, T. Chen, S. Tan, W. Ng, "GPR model with signal preprocessing and bias update for dynamic processes modeling," Control Eng. Pract. vol.20(12), pp.1281-1292, 2012.

[9] E.P. Nahas, M.A. Henson; D.E. Seborg, "Nonlinear internal model control strategy for neural network models," Comput. Chem. Eng. vol.1, pp.1039-1057, 1992.