Abstract—Soft sensors are used to infer the quality variable from easy-to-measure process variables. The conventional static soft sensor is incapable of handling the dynamic of processes. For data-based soft sensor development, with abundance of the raw sensor data, the problem of variable correlations and large number of sample are encountered. This work presents a latent variable model (LVM) based active learning strategy to select representative data for efficient development of the dynamic soft sensor model. In order to carry out data selection the uncertainty information is provided by Gaussian process (GP) model. The LVM with auxiliary GP model is developed under a dynamic framework which is suitable for dynamic processes. A forward-update scheme for updating the soft sensor model in advance is proposed so that the soft sensor is able to reflect the current status of the process and to improve the soft sensor model without waiting for the quality measurements. The proposed method is applied to an industrial fluid catalytic cracking process data.

I. INTRODUCTION

In process industrial plants the important quality variables are often difficult to be measured due to reasons such as high cost, low reliability or low sampling rate of the hardware sensor. To overcome this problem, soft sensors can be used to estimate quality variables indirectly online. Traditionally in the soft sensor development, it is assumed that the process is in static states with a small portion of process data adopted and the quality variables are dependent on the process variables sampled at the particular time interval [1]. The static assumption of the conventional soft sensor is incapable of handling the dynamic processes. To deal with the problem, the historical data of the dynamic process is taken into account and the resultant soft sensor is able to extract dynamic information. For data-based soft sensor development, the raw sensor data collected from the process database can contain measurements of up to several years. When big data are available in the database and the whole data set are used by practitioners the computational load can be excessive. Therefore the problem of correlated variables and large number of sample need to be tackled.

The projection methods are often used to address correlations in the variables. Multivariate statistical methods including principal component analysis (PCA) [2] partial least squares (PLS) [3] have been developed over the past two decades. Nevertheless the process data characteristics including nonlinear relationships and data correlations has resulted in many extension methods based on PCA and PLS to overcome the drawbacks of the aforementioned methods [4]. In the projection methods variables in the high dimensional space is reduced into a lower dimensional space called the latent space. The latent variable model (LVM) can thus be a basis for the soft sensor development of handling correlation of the variables.

On the other hand, sample selection is used to reduce the number of samples. Traditionally, trial-and-error approaches were used to find samples from the database. These approaches are time consuming due to random sampling and the data selected may not be representative of the system. In order to overcome such problem an active learning strategy should be applied to obtain representative data for efficient development of the soft sensor model. The active learning is an approach that can successfully handle the regression problem with big data. It aims to maximize the estimation performance of the regression model while minimizing the number of samples. The selection of the representative data from the samples is therefore a crucial task in the active learning method. The representative data that is selected can improve the current model for better performance. The active learning approach has been applied to the enhancement of monitoring of process [5]. In order to carry out data selection the uncertainty information provided by the model can be utilized [6] but conventional LVM does not provide such information. Gaussian process (GP) model provides a description of the uncertainty of the model. Based on the probabilistic structure of the GP model a selection criterion using the uncertainty information is proposed for the soft sensor modeling for dynamic processes.

As industrial process can be subject to changes in operation, the dynamic LVM with the auxiliary GP model which can predict the uncertainty under a dynamic framework is developed. A forward-update scheme is proposed to update the soft sensor model in advance to early reflect the current status of the process and to quickly improve the soft sensor model without waiting for the quality measurements. The article is organized as follows. Section II provides a problem formulation, followed by details of the proposed dynamic soft sensors with forward-update scheme in Section III. The industrial case study in Section IV demonstrates the features of proposed method and Section V ends the contribution with some concluding remarks.

II. PROBLEM DEFINITION

The LVM method is an effective tool for data compression and extraction. Nevertheless, the sample selection is crucial to constructing models for good prediction performance. In order to present the idea clearly, without loss of generality, the kernel-PLS (KPLS), which is a type of LVM structure, is used...
throughout the rest of this article. It should be noted the method described can be easily used in other LVM structures. The nonlinear system that is used for the illustration purpose is given as

\[
y_i(x) = -0.85\cos(3x_{k-1})x_k\exp(-(0.8x_k - 0.4)^2)
\]  
(1)

\[y_i(x) \text{ is a function of } x_{k-1} \text{ and } x_k, \text{ where } x_k \text{ is a sinusoidal input}\]

\[x_k = 3\cos(kw_0)\]

(2)

Eq.(1) represents a single input single output function, for brevity, but can be generalized to a multiple input multiple output system. \(k\) is the time and \(w_0\) is the frequency of oscillation. In this illustration \(x_k\) is sampled every time point whereas \(y_i\) is sampled at an interval of 2 time points. Figure 1 shows the predictions from a KPLS model and the actual relationship among \(y_i\), \(x_{k-1}\) and \(x_k\) in a 3-dimensional plot.

The red and green curves represent the actual values and the predictions respectively whereas the blue dots represent the training data used. Some of the prediction results, for example, for point 7 and 17, are poor. The differences between the actual values and the predictions are marked by the black dotted lines and show large deviations from the actual values. The regions in close proximity of point 7 and 17 are characterized by high nonlinearity and no data is selected for model training and thus result in poor performance. In reality the actual system cannot be known in advance and there is no way to find out about the accuracy of prediction based on the prediction alone. In this research work, the representative data can be selected through the analysis of the uncertainty in the model.

The design of a good training data is therefore a critical issue which is the focus of this work. To this end the active data selection is implemented for the construction of the soft sensor model. The active learning structure for this study is shown in Figure 2. An initial training data set is selected for the constructing a preliminary LVM model. As the LVM does not provide uncertainty information the latent variables is used to construct a GP based complementary model for obtaining the uncertainty information that allows for analysis. If the model satisfies a predefined level of confidence (based on the uncertainty), the soft modeling is completed. Otherwise, based on the uncertainty analysis (defined in later section), new data are acquired from the database to improve the model. Once the model is built, the soft sensor is put into on-line applications (indicated by the blue-line in Figure 2). In many chemical processes, the sampling rates of the process variables and the quality variables are unequal, because it is difficult to acquire measurements at a high frequency for quality variables, such as the composition, density, and molecular weight distribution. The forward-update of the model the prediction from the soft sensor model is corrected in advance without waiting for the quality measurements. Thus, the uncertainty is also continuously checked. When the soft sensor predictive variance is not within the threshold, it is updated using information from the database. Also, the new available quality measurements with time going would add into the database. Under this active learning process architecture several issues should be addressed, (1) the construction of the auxiliary model of latent variable, (2) the definition of the selection criterion for active data selection, (3) the stopping criterion, (4) the adaptive criterion for forward update.

III. DYNAMIC LVM-GP ACTIVE LEARNING

The dynamic LVM with auxiliary GP is detailed in this section. The idea is to augment the LVM with the uncertainty provided by the GP in the latent space and thus enables the analysis of the uncertainty in the LVM. The active learning criterion for this model is defined for optimal data selection. In addition the forward-update scheme is detailed in this section.

A. Preliminaries:

1) Gaussian process model

The GP model method for nonlinear process regression is to learn a model \(f\) that approximates a training set \(\{X,y\}\), where \(X\) and \(y\) are, respectively, the input and output datasets with \(N\) samples. A GP provides a prediction of the output variable for an input sample through Bayesian inference. For a single output variable \(y = (y_1, \ldots, y_N)^T\), the GP is the regression function with a Gaussian prior distribution and zero mean, or in a discrete form [7]

\[y = (y_1, \ldots, y_N)^T \sim G(0, \mathbf{K})\]

(3)

where \(\mathbf{K}\) is the \(N \times N\) covariance matrix with the \(lm\)-th element defined by the covariance function, \(K_{lm} = k(x_l, x_m)\). A common covariance function can be defined as

![Figure 1 Predictions of KPLS](image1)

![Figure 2 Active learning structure for dynamic LVM](image2)
\[ k(x_i, x_n) = v_0 \exp \left( -\frac{1}{2} \sum_{d=1}^{D} w_d (x_{i,d} - x_{n,d})^2 \right) \]  
(4)

where \( x_{i,d} \) is the \( d \)th component of the vector \( x_i \). \( \mathbf{\theta} = [v_0, w_1, \cdots, w_D, b]^T \) is the hyper-parameter vector defining the covariance function. Adopting a Bayesian approach, the hyper-parameters \( \mathbf{\theta} \) can be estimated by maximization of the following log-likelihood function

\[ \log p(y|X, \mathbf{\theta}, \sigma^2_e) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \log \mathbf{K} + \frac{1}{2} \mathbf{y}^T (\mathbf{K} + \sigma^2_e \mathbf{I})^{-1} \mathbf{y} \]  
(5)

Finally, the GP can be obtained once \( \mathbf{\theta} \) is determined. For a test sample \( x_i \), the predicted output of \( y_i \) is also Gaussian with mean \( (\hat{y}_i) \) and variance \( (\hat{\sigma}_i^2) \), calculated as follows

\[ \hat{y}_i = \mathbf{k}^T (\mathbf{K} + \sigma^2_e \mathbf{I})^{-1} \mathbf{y} \]  
(6)

\[ \hat{\sigma}_i^2 = k_i + \sigma^2_e - \mathbf{k}^T (\mathbf{K} + \sigma^2_e \mathbf{I})^{-1} \mathbf{k} \]  
(7)

where \( \mathbf{k} = [k(x_i, x_1), k(x_i, x_2), \cdots, k(x_i, x_N)]^T \) is the covariance vector between the new input and the training samples, and \( k_i = k(x_i, x_i) \) is the covariance of the new input. In summary, the vector \( \mathbf{k}^T \mathbf{K}^{-1} \) denotes a smoothing term which weights the training outputs to make a prediction for the new input sample \( x_i \). In addition, Eq.(7) provides a confidence level on the model prediction, which is an appealing property of the GP.

2) Latent variable model

In general the LVM can be expressed as

\[ \mathbf{A} = \mathbf{TL}^T + \mathbf{E} \]  
\[ \mathbf{Y} = \mathbf{TG}^T + \mathbf{F} \]  
(8)

where \( \mathbf{T} \) denotes the scores matrix. \( \mathbf{A} \) is a transformation of the sampling matrix \( \mathbf{X} \) and \( \mathbf{Y} \) is the transformation of the outputs. Eq.(8) describes the relationship between the input variables to latent variables and the output variables to the latent variables respectively. The common linear latent variable models are the PCR and PLS. For nonlinear processes, kernel-based latent variable models such as kernel-PCR and kernel-PLS (KPLS) are commonly used. The kernel based method first map the nonlinear input space into a linear feature space and then to compute PCR or PLS in the feature space. For these common algorithms the transformation matrix can be expressed as

\[ \mathbf{A} = \begin{cases} \mathbf{X} & \text{for PCR, PLS} \\ \mathbf{\Phi} & \text{for KPCR, KPLS} \end{cases} \]  
(9)

The kernel can be obtained by the following inner products

\[ \mathbf{\Phi} = \mathbf{K}(x_i, x_n) = \left[ \phi(x_1), \phi(x_2), \ldots, \phi(x_n) \right] \]  
(10)

with \( \phi(x_i) \) as the kernel. The specific kernel function implicitly determines the associated mapping \( \phi(\cdot) \) and the feature space.

Using the KPLS algorithm as an example, the regression function between \( \mathbf{Y} \) and \( \phi(\mathbf{X}) \) can be expressed by the following matrix forms [8],

\[ \hat{\mathbf{Y}} = \mathbf{\Phi}(\mathbf{X}) \mathbf{G} = \mathbf{T} \mathbf{T}^T \mathbf{Y} \]  
(11)

\[ \mathbf{T} = \mathbf{\Phi}(\mathbf{X}) \mathbf{R} \]  
\[ \mathbf{R} = \mathbf{\Phi}(\mathbf{X})^T \mathbf{U}(\mathbf{T}^T \mathbf{K})^{-1} \]  
(12)

where \( \mathbf{G} = \mathbf{\Phi}(\mathbf{X})^T \mathbf{U}(\mathbf{T}^T \mathbf{K})^{-1} \mathbf{T}^T \mathbf{Y} \) is the regression coefficient, and \( \hat{\mathbf{Y}} \) is the prediction of \( \mathbf{Y} \). Based the LVM structure, the dynamic LVM can be constructed for handling of the process dynamics.

B. LVM-GP model

It follows from GP formulation that for the input \( (x_n) \) with the score of component \( r (t_{sr}) \), as output,

\[ t_{sr} = f(x_n) + \epsilon_r, \quad r = 1, \ldots, R \]  
\[ \epsilon_r \sim N(0, \sigma^2_e) \]  
(13)

Based on the formulation of the GP (Eqs.(6)-(7)) the latent variable prediction, \( \hat{i}_{sr} \) and its variance, \( \hat{\sigma}_{i sr} \), can be denoted as

\[ \hat{i}_{sr} = \mathbf{k}^T (\mathbf{K} + \sigma^2_e \mathbf{I})^{-1} \mathbf{t}, \]  
\[ \hat{\sigma}^2_{isr} = k_i + \sigma^2_e - \mathbf{k}^T (\mathbf{K} + \sigma^2_e \mathbf{I})^{-1} \mathbf{k} \]  
(14)

where the subscript \( r \) is used to denote the \( r \)th latent variable. In soft sensing the variables of interest are usually the quality variables. Once the mean and variance of the latent variable are obtained based on the LVM regression, the quality variables from the LVM model are given as

\[ \mathbf{Y} = \mathbf{T}\mathbf{G}^T + \mathbf{F} \]  
\[ \mathbf{G} = [g_1, \ldots, g_Q] \in \mathbb{R}^{Q \times R} \]  
(16)

and in terms the mean and variance of quality variable,

\[ \hat{y}_{s,q} = \sum_{r=1}^{R} \hat{i}_{sr} g_{r,q}, \quad q = 1, \ldots, Q \]  
\[ \hat{\sigma}^2_{ysq} = \sum_{r=1}^{R} \sigma^2_{isr}, \quad q = 1, \ldots, Q \]  
(17)

Therefore, the latent variable prediction, \( \hat{i}_{sr} \) and its variance, \( \hat{\sigma}_{i sr} \), can be obtained from the resultant GP. The quality \( \hat{y}_{s,q} \) and its predictive variance \( \hat{\sigma}^2_{ysq} \) is obtained through the projection back to the original space. The variance of the GP is a point variance that describes the proximity of the data. Continuing the example as shown in Figure 1, the uncertainty of the resultant LVM is shown in Figure 3, in which the orange dotted line represents the limits placed at three standard deviations from the mean, i.e., \( \hat{y}_{s,q} \pm 3\hat{\sigma}_{ysq} \). The variance information at each point is represented and it can be seen that the variance at the region with data is lower and higher for the
region without data. This can be interpreted as the confidence in the prediction where a lower variance implies that the prediction is more reliable and higher variance implies a less reliable prediction.

C. Dynamic LVM

The dynamic LVM is used in order to describe the process dynamics accurately. Figure 4 shows the input and output data sequence in the modeling of the dynamic LVM. The measurement time interval of quality variables is denoted as \( J \) whereas the inputs are available at each time unit. For \( N \) number of quality variables the total measurement time is \( NJ \). \( y_n \) refers to the \( n \)th measurement of the quality variables. Each blue arrow in Figure 4 denotes the input and output relation in which the output \( y_n \) is a related to the past process variable vectors \( x_{nJ-j+1} \) to \( x_{nJ} \) inputs.

![Figure 4 Input and out data sequence for dynamic LVM](image)

In building the soft sensor model, assumed that \( y_n \) is affected by the past \( J \) process variables vectors, and thus \( y_n \) can be denoted as

\[
y_n = f \left( \left[ x_{nJ-j+1} \ldots x_{nJ-j+j} \ldots x_{nJ} \right] \right) \quad (19)
\]

\[
x_{nJ-j+j} = \left[ x_{nJ-j+j,1} \ldots x_{nJ-j+j,D} \right], j = 1, \ldots, J \quad (20)
\]

\[
y_n = \left[ y_{n,1} \ldots y_{n,q} \ldots y_{n,Q} \right] \quad (21)
\]

where \( D \) is the total number of process variables and \( Q \) is total number of quality variables. \( J \) is selected to be much greater than the process dynamics so that the data at different windows are considered independent. As the quality variable vector \( y_n \) is related to \( J \) past process variable vectors, the data structure is handled with the time-wise unfolding. The unfolding method converts the variable data into the vector representation that can be expressed as

\[
\tilde{x}_{nJ} = \left[ x_{nJ-j+1} \ldots x_{nJ-j+j} \ldots x_{nJ} \right] \quad (22)
\]

\( \tilde{x}_{nJ} \) is the process variable vector with \( DJ \) dimension. The input data structure is denoted as \( \tilde{X} = \{ \tilde{x}_{nJ}, n = 1, \ldots, N \} \) and the corresponding quality data structure is given as \( Y = \{ y_n, n = 1, \ldots, N \} \). Using \( \tilde{X} \in \mathbb{R}^{N \times DJ} \) and \( Y \in \mathbb{R}^{N \times Q} \) the LVM can be constructed and the prediction is given as

\[
\hat{y}_k = f(\tilde{x}_k) \quad (23)
\]

where

\[
\hat{y}_k = [\hat{y}_{k,1} \ldots \hat{y}_{k,q} \ldots \hat{y}_{k,Q}] \quad (24)
\]

the corresponding input vector is represented as

\[
\tilde{x}_k = [x_{kJ-j+1} \ldots x_{kJ-j+j} \ldots x_k] \quad (25)
\]

It is noted that unlike measured quality variable, the predicted quality variable is available at the each time point \( k \) and is denoted as \( \hat{y}_k \) with the moving window structure \( \tilde{x}_k \) as depicted in Eq.(25) as the input. The LVM-GP model output quality prediction and the predictive variance can be obtained using Eqs.(17)-(18) and the quality variable prediction is a function of the past inputs according to Eq.(23).

IV. IMPROVING DYNAMIC LVM WITH ACTIVE LEARNING APPROACH

For the active learning of the model the selection criterion needs to be defined and for online adaptation the window size which determines the updating time frame of the LVM-GP needs to be considered due to the dynamic of the system. These are described in the following.

A. Offline active learning

The active learning aims to select representative data for constructing the model. The selection criterion based on the uncertainty information of the LVM-GP is proposed. The regions with a relatively large prediction variance indicate that the data are lacking and the representative data corresponds to these regions should be added to improve the model. The criterion for the data addition based on the uncertainty information can be expressed as

\[
\tilde{x}_{in} = \arg \max_{\tilde{x}_i} \left( \sum_{q=1}^{Q} \sigma_{y_q}^2 (\tilde{x}_i) \right) \quad (26)
\]

where \( \sigma_{y_q}^2 \) is the predicted variance of the \( q \)th quality variable. Data are added according to the order of the magnitude of the term \( \sum_{q=1}^{Q} \sigma_{y_q}^2 \) and \( \tilde{x}_{in} \) is the input vector that corresponds to the largest variance in \( \hat{y}_{in} \). The exact data match from the query of database is used. If no exact match is found in the database, the nearest value according to Euclidean distance
\[ d_i(\tilde{x}_i, \tilde{x}_e) = \sqrt{\sum_{d=1}^{D} \sum_{j=1}^{J} (x_{i,d,j} - x_{i-1,d,j})^2} \]  

is selected. \( \tilde{x}_e \) is the process variable vector in the database. Once a data is added through the selection the model can be improved. To assess if more data should be added to the model a convergence index is proposed. This can be expressed as

\[ \zeta_i = \sqrt{\frac{1}{Q} \sum_{k=0}^{Q} \left( \sum_{j=1}^{J} (\tilde{y}_{i+1,j} - \tilde{y}_{i,j}) \left( \tilde{y}_{i+1,j} - \tilde{y}_{i,j} \right)^T \right)} \]  

where \( \zeta_i \) is the model performance after the \( i \) th addition of data. Its performance is measured against that before the addition of data denoted with the superscript index \( i-1 \). Convergence is reached when \( \zeta_i \) does not change significantly after the addition of data as this indicates that no further improvement to the model is achieved with the addition of new data. The soft sensor constructed can then be operational in plant.

### B. Online forward-update

Once the LVM-GP is constructed and in operation, the uncertainty threshold is continuously checked on-line. Indeed it is possible to learn the entire system. However given that the current process is only operated at some region, learning the whole range of system would require huge computation and is inefficient. The idea is to update the model as needed online for efficient learning. At time \( k \) the criterion is expressed as

\[ \tilde{x}_{k+j} = \arg \max_{k+j} \left( \sum_{q=1}^{Q} \sigma^2_{y_{t,q,i}} \right) \]  

\( \text{st. } \sigma^2_{y_{t,q,i}} \geq \varepsilon, \quad q = 1, ..., Q \)  

where \( \tilde{x}_{k+j} \) refers to the future input that is obtained from process design or designated by operators. \( \sigma^2_{y_{t,q,i}} \) is the \( j \)-step ahead predicted variance at time \( k \) of \( q \) th quality variable and it must be above a user defined threshold \( \varepsilon \) for updating the quality variables. Data is added according to the order of the magnitude of the term \( \sum_{q=1}^{Q} \sigma^2_{y_{t,q,i}} \) and \( \tilde{x}_{k+j} \) denotes the input vector that corresponds to the largest variance at \( j \) steps in \( \hat{y}_{t,j} \) among the \( J \)-step ahead predicted variances. As a result the future predicted variance is checked in advance before the next quality measurement. The forward-update is performed when the threshold is exceeded. Similar to the offline active learning, either the exact match data is selected or the nearest value based on Eq.(27) is selected. For \( J \)-step ahead predictions the update is carried out until either the threshold \( \varepsilon \) is satisfied or all the candidates are updated. With the forward-update scheme the model is improved in advance and thus is able to provide a better prediction performance without waiting for the quality measurements.

### V. Case Study

The industrial case study is presented to shows the applicability of the proposed dynamic LVM soft sensor. The data from the fluid catalytic cracking (FCC) process from one of the plants of Chinese Petroleum Corporation (CPC) (Taiwan) is used for the case study. The FCC unit converts the heavy crude oil with high molecular weight into valuable hydrocarbon such as gasoline or propene. Figure 5 shows the flowsheet of the FCC process which consists of the de-etherizer and the debutanizer. The upstream gas mixture from cracking is first passed through the de-etherizer. The overhead is predominantly C2-products and lighter gas such as hydrogen. The heavier bottoms product (C2 / C3 / C4 / C5) are passed through two pre-treatment units and then separated by the depropanizer column. The main products at the top are propane (C3) and some ethane (C2), while the product at the bottom of tower consists of butanes (C4) and pentanes (C5). The C3H6 from the process is important in the production and is regarded as the quality variable. The concentration of C3H6, however, is difficult to measure and is conventionally obtained from lab assay. In order for instantaneous concentration the LVM soft sensor is developed to predict the concentration of the propene.

The available online measurements of process variables include the feed temperature (TT1), pressure (PT), the condenser temperature (TT2), the reboiler temperature (TT6) and the temperature at first plate (TT3), 21st plate (TT4) and 40th plate (TT5). The composition of the overhead consists of C3H6, C3H8 and C2H4 and C2H6. The 99.5 mol\% C3H6 with small amount C3H8, C2H4 and C2H6 is specified in this process. As the C2H4 and C2H6 are only present in minute level and they are considered a single variable (C2H4-C2H6) in the development of the LVM soft sensor. Figure 6 shows the process variables and the quality variables for the first 500 time units.

The three quality variables, C3H6, C3H8 and C2H4-C2H6, are measured at each 20 time unit intervals. Consequently based on the time-unfolding in Section III, a row of \( 7 \times 20 \) process variables correspond to the quality measurements (C3H6, C3H8, C2H4 and C2H6). The process variable data from time 1-200 is used to build the initial model and there are thus 10 quality variable data for the 200 unit-time data. The prediction results using the initial set of data are shown on the left hand panel of Figure 7 and for brevity only the predictions of the main product, C3H6 are shown. It can be seen that some of predictive variances are large, for instance at \( k = 5 \), \( k = 128 \) and \( k = 165 \). To improve the soft sensor model for the dynamic system, in the offline active learning procedure, based on the proposed criterion of Eq.(26), the model is updated. The result after update is shown on the right hand side of Figure 7. It can be seen that the variance have decreased after the update indicating an improved confidence in the prediction. In this case, the convergence of the model performance is reached after 15 data points are added.

When the update is completed, the soft sensor is put in operation from time point 200 onwards. The left hand side of Figure 8 shows the prediction from the time point 200 to 220 when the soft sensor is in operation until time point 230. If no forward update is carried out, the variance from the time point
217 onwards is high and the prediction does not match the actual measurement (red triangle) at the time point 220. In the online forward update scheme, the model is updated in advance at the time point 217 and 218 using data from the query of database. The prediction with the forward update on the right hand side of Figure 8 shows that prediction after the update matches the actual measurement.

For the purpose of comparison, the root mean square errors (RMSE) of 3 strategies when the soft sensor is operational are compared. These strategies include (1) no update, (2) update every 20 time points and (3) forward-update. In this case study, RMSE is computed, depending on when the actual measurement becomes available. This is formulated as

$$\text{RMSE} = \frac{1}{(K / J - N)} \sum_{n=N+1}^{K/J} (\hat{y}_{aj} - y_{aj})^2 (\hat{y}_{aj} - y_{aj})$$ (30)

where $K$ is the final measurement time when the RMSE is computed. Table I show the RMSE computed at time points 300, 400 and 500. It can be seen that the performance of the forward-update is the best. 19 new data points are added after 500 time points of operation. The model that is updated in advance is able to reflect the status of the process and provides a better prediction performance.

![Figure 6 First 500 data for LVM process variables and quality variables](image)

![Figure 7 Prediction for time point 1-200 before and after update](image)

![Figure 8 Prediction for time point 200-220 before and after update](image)

**VI. CONCLUSION**

The soft sensor with forward-update scheme based on LVM with auxiliary GP model is developed in this work for dynamic processes. The forward-update improves the models without waiting for the quality measurements. Its applicability is demonstrated through an application to the industrial fluid catalytic cracking process data. As an extension work, the active learning criterion can be further explored to improve the active learning process. For instance, the effect of input variables on the prediction values can also be taken into account.

**REFERENCES**


