I&EC research
Industrial & Engineering Chemistry Research

Article

Cite This: Ind. Eng. Chem. Res. 2019, 58, 11224−11233

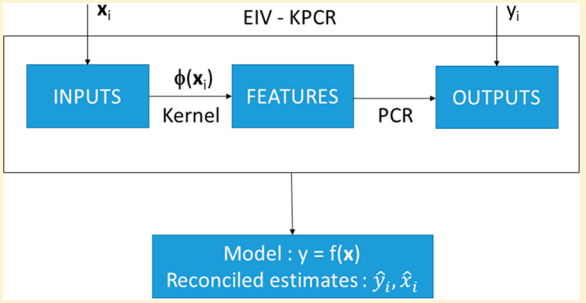
pubs.acs.org/IECR

Nonlinear Model Identification and Data Reconciliation Using Kernel Principal Component Regression

Karthikeyan Marimuthu[#] and Shankar Narasimhan*

Department of Chemical Engineering, IIT Madras, Chennai, 600036, India

ABSTRACT: Reconciliation of process data is an important preprocessing technique, the main purpose of which is to obtain accurate estimates of variables and model parameters. Reconciliation requires a process model which is generally developed using first principles. For many complex processes, the development of such models is difficult and time-consuming. In this work we propose a novel alternative method for steady state data reconciliation of nonlinear processes which does not require a functional model between variables to be specified *a priori*. A nonlinear model relating the variables is developed from a given data set, while simultaneously obtaining accurate estimates of the measured variables. The method we propose combines concepts drawn from Kernel Principal Components Regression with an error-in-variables model parameter estimation technique. Simulation studies demonstrate that the proposed approach is able to improve the accuracy of measured variables. The identified nonlinear model is also useful for reconciling future measurements of the process.



1. INTRODUCTION

In modern chemical plants, a large number of variables such as temperature, flow rates, pressures, and levels are measured and recorded for the purposes of process monitoring, optimization, and control. These measurements are inevitably corrupted by random errors caused by several sources such as power fluctuations and changes in ambient conditions. The measured data usually do not satisfy the mass and energy balances of the process and have to be reconciled before undertaking a simulation, optimization, and control exercise. The purpose of data reconciliation (DR) is to ameliorate the effect of random noise in measurements, and derive estimates that are more accurate and consistent with the mass, energy balances, and other constraints of the process. While DR deals with random errors in measurements, the companion technique of gross error detection deals with errors caused by nonrandom events such as instrument biases and malfunctions, unsuspected leaks, and so forth.

For performing data reconciliation, the process model equations relating the variables have to be specified. These equations are generally derived from first-principles using physical conservation laws, constitutive relations, and equilibrium constraints, etc. Furthermore, the accuracies of measurements specified as the variances of errors in the measurements are also assumed to be known. The techniques of data reconciliation using known process models have been described in several books,[1−3] and have also been applied in many industries such as refineries,[4] petrochemical plants, and mineral beneficiation plants.[5] However, for some process units such as a fluid catalytic cracker, it may be difficult or virtually impossible to derive the functional relationships between variables. Even if it is possible to develop a first-principles model, significant effort may be required to assemble the correlations for estimating physical and thermophysical properties. Moreover, the correlations themselves may be of limited accuracy, and the inaccuracies inherent in these correlations are usually ignored in a reconciliation exercise based on first-principles model. For such a process, an interesting question that can be posed is whether it is possible to identify a nonlinear model of the process from measured data of the process, while simultaneously reconciling the data set. We use the term reconciliation here to imply that the estimates of the measured variables obtained must be more accurate than their corresponding measured values. The identified nonlinear model should also be useful for reconciling new measurements of the process.

The genesis for the problem comes from recent papers[6,7] which show that for a linearly constrained process (such as a flow process), Principal Component Analysis (PCA) can be used to simultaneously identify the linear model from measurements and also derive the corresponding reconciled estimates. In these papers, it was shown that the linear constraints identified using PCA and the steady state flow balance constraints derived from first-principles are equivalent bases for the same row space. Furthermore, the denoised estimates derived using PCA are identical to the reconciled estimates with respect to the identified constraints. In summary, PCA can be interpreted as a procedure to simultaneously identify the process model and reconciled

Special Issue: Sirish Shah Festschrift

Received: December 9, 2018
Revised: May 23, 2019
Accepted: May 23, 2019
Published: May 23, 2019

DOI: 10.1021/acs.iecr.8b06095
Ind. Eng. Chem. Res. 2019, 58, 11224−11233

estimates from data for a linearly constrained steady state process. The purpose of this paper is to develop an analogous procedure for nonlinearly constrained steady state processes.

The problem being addressed in this paper can also be viewed from the perspective of identifying nonlinear regression models from input−output data. One approach is to approximate the nonlinear regression functions relating the input and output variables using a weighted combination of basis functions such as orthonormal polynomials or cubic splines, etc. and estimating the weight parameters from data.[8] In recent years, neural networks, support vector regression, and kernel principal component regression (KPCR) were being used especially by the data science community.[9] All these identification methods use an ordinary least-squares formulation. In other words, the objective function used in all these methods accounts only for the errors in the dependent variables and implicitly assumes that the measurements of independent variables are free of error. Thus, in their present form, these nonlinear model identification methods cannot be used to simultaneously reconcile all the measurements. To simultaneously identify a nonlinear model and obtain reconciled estimates with respect to this model, an *errors-in-variables* or *total least-squares* objective function must be used. In this formulation, measurements of both the input and output variables are assumed to contain errors, and estimates of all variables are obtained along with model parameters. The approach proposed in this work can also be regarded as an EIV nonlinear model identification method.

Valko and Vajda[10] proposed an EIV approach for estimating the parameters of a *known* nonlinear functional model. Their method is an iterative procedure, which includes a nonlinear data reconciliation step in an inner loop, embedded within a maximum likelihood estimation of the model parameters in an outer loop. The covariance matrix of errors in the measurements is assumed to be known, and the reconciled estimates of the variables are also simultaneously obtained. Similar approaches for simultaneous parameter estimation and nonlinear data reconciliation have been proposed by other researchers.[11−13] It may be noted that in all these approaches the functional relationship between variables must be known. In contrast, the approach proposed in this paper simultaneously identifies a nonlinear model and obtains reconciled estimates from noisy measurements under the EIV framework, when the functional model between the variables is not given *a priori*.

In this work, we choose to identify the nonlinear model in terms of Kernel functions. We combine concepts used in KPCR along with the iterative EIV approach of Valko and Vajda[10] to derive a method to solve our proposed problem. Simulation studies are used to demonstrate that the reconciled estimates are more accurate as compared to the measurements by comparing them to the true values used in simulation. We also demonstrate that the model identified from data can be regarded as an approximate representation of the process constraints, by showing that the reconciled estimates satisfy the true process constraints more closely as compared to the measured values.

The paper is organized as follows. In section 2, KPCR method is described in detail. Section 3 provides a description of the method of Valko and Vajda[10] which is used for simultaneous reconciliation of measurements and estimation of parameters of a specified nonlinear model. The proposed method which is derived by combining the KPCR method with Valko and Vajda's approach is described in section 4. Simulation studies on two examples are used to demonstrate the efficacy of our proposed solution in section 5. We close out with conclusions and suggestions for future extensions in section 6.

2. KERNEL PRINCIPAL COMPONENT REGRESSION

The kernel principal component regression[14] (KPCR) approach is a method for developing a nonlinear regression model between specified dependent and independent variables from measured data. In this data based approach, the functional form of the nonlinear model between variables is not given and is derived from the data. The method makes use of the Kernel Principal Components Analysis (KPCA) technique,[15] which is described below.

Let $\tilde{\mathbf{x}}(i)\epsilon\mathbb{R}^m$ be a column vector of measurements of m variables at the sampling index i. If we have n such measurement vectors, then we can construct the sample matrix $\tilde{\mathbf{X}}\epsilon\mathbb{R}^{m\times n}$ as follows.

$$\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}(1)\ \tilde{\mathbf{x}}(2)\ \cdots\ \tilde{\mathbf{x}}(n)] \tag{1}$$

In KPCA, the variables $\mathbf{x}$ are transformed to a high-dimensional feature space F by using nonlinear mapping functions, schematically represented as follows.

$$\mathrm{x} \xrightarrow{\Phi(\mathrm{x})} \mathrm{z} \tag{2}$$

where $\mathbf{z}$ is the transformed vector in feature space. The dimension of the feature space, N, is determined by the dimension of the vector of mapping functions, $\mathbf{\Phi}(\mathbf{x})$. Let $\tilde{\mathbf{z}}(i) = \mathbf{\Phi}(\tilde{\mathbf{x}}(i))\epsilon\mathbb{R}^N$ be a column vector of N transformed variables at the sampling index i. Since we have n sample measurements of $\tilde{\mathbf{x}}$, the number of samples in F is n. Hence, we can construct the sample matrix $\tilde{\mathbf{Z}}\epsilon\mathbb{R}^{N\times n}$ in feature space F as follows

$$\tilde{\mathbf{Z}} = [\tilde{\mathbf{z}}(1)\ \tilde{\mathbf{z}}(2)\ \cdots\ \tilde{\mathbf{z}}(n)] \tag{3}$$

The purpose of performing a nonlinear transformation of the variables is to work with linear regression techniques in feature space F instead of nonlinear regression techniques in original variables space.

To use the above approach, we need to choose the nonlinear mapping functions, $\mathbf{\Phi}(\mathbf{x})$. Furthermore, the sample matrix $\tilde{\mathbf{Z}}$ has to be computed, which can be computationally demanding especially since a high-dimensional feature space is usually chosen to ensure that a good approximation of the nonlinear relations can be obtained. Kernel functions can be used to avoid explicit knowledge of the mapping functions. More importantly, the computation of the sample covariance matrix in feature space can also be avoided. Kernels allow inner products in the feature space to be computed using inner products in the original variable space since they satisfy Mercer's condition.[16] Several different Kernel functions have been proposed in the literature. Among them, the most popular ones are polynomial kernel of order p which is given by

$$K_p(\mathbf{x}(i), \mathbf{x}(j)) = (1 + \langle \mathbf{x}(i), \mathbf{x}(j)\rangle)^p \tag{4}$$

and Gaussian kernel with width W:

$$K_G(\mathbf{x}(i), \mathbf{x}(j)) = \mathrm{e}^{-\|\mathbf{x}(i)-\mathbf{x}(j)\|^2/W} \tag{5}$$

DOI: 10.1021/acs.iecr.8b06095
Ind. Eng. Chem. Res. 2019, 58, 11224−11233

It may be noted that a finite dimensional feature space is obtained for a polynomial kernel choice, which depends on the order p of the kernel, whereas for a Gaussian kernel choice, the feature space is infinite dimensional.

Typically, the Kernel is chosen so that the dimension of the feature space is greater than the number of samples. Therefore, the rank of the sample matrix $\tilde{\mathbf{Z}}$ is equal to number of samples n. A lower dimensional subspace of the feature space can be estimated using PCA. For this purpose, it is necessary to determine the eigenvectors of the covariance matrix among variables given by

$$\mathbf{C}_Z = \frac{1}{N}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^{\mathbf{T}} \tag{6}$$

However, we wish to avoid computation of the sample matrix $\tilde{\mathbf{Z}}$ as well as its covariance matrix $\mathbf{C}_Z$. The Kernel property can be exploited to determine the eigenvectors of $\mathbf{C}_Z$ without explicitly computing it. For this purpose we first construct the covariance matrix in feature space *among samples*, $\mathbf{C}_S$, defined by

$$\mathbf{C}_S = \frac{1}{N}\tilde{\mathbf{Z}}^{\mathbf{T}}\tilde{\mathbf{Z}} \tag{7}$$

It can be verified that the elements of $\mathbf{C}_S$ are given by

$$[\mathbf{C}_S]_{ij} = \langle \mathbf{\Phi}(\tilde{\mathbf{x}}(i)), \mathbf{\Phi}(\tilde{\mathbf{x}}(j))\rangle = K_p(\tilde{\mathbf{x}}(i), \tilde{\mathbf{x}}(j)) \tag{8}$$

Clearly the covariance matrix $\mathbf{C}_S$ can be constructed using inner products of samples in original variable space. $\mathbf{C}_S$ is also denoted as the Kernel Matrix and it is represented as $\mathbf{K}$. The eigenvectors of $\mathbf{C}_S$ can be determined, and from these the eigenvectors of $\mathbf{C}_Z$ can be computed using the following relation.[15]

$$\mathbf{U} = \tilde{\mathbf{Z}}\mathbf{V}(\mathbf{\Lambda})^{-0.5} \tag{9}$$

where $\mathbf{U}$ is the eigenvectors of $\mathbf{C}_Z$ corresponding to n nonzero eigenvalues, $\mathbf{V}$ is the eigenvectors of $\mathbf{C}_S$, and $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues. The dimension of the subspace of independent variables in feature space can be obtained by examining the eigenvalues similar to PCA or by using one of the several criteria discussed by Valle et al.[17] Following the arguments presented by Narasimhan and Shah,[6] the transpose of eigenvectors corresponding to the smallest singular vectors can be regarded as a basis for the linear relations between variables in feature space.

It may be noted that in order to compute the eigenvectors of $\mathbf{C}_Z$ using eq 9, the sample matrix $\tilde{\mathbf{Z}}$ in feature space is required, which we wish to avoid computing. In the description that follows, we show that in order to determine the regression model between dependent and independent variables using KPCR, the computation of $\tilde{\mathbf{Z}}$ can be avoided.

Using KPCR, we wish to develop a nonlinear regression model between a dependent variable y and the vector of variables $\mathbf{x}$. Let $\tilde{y}_i$ $i = 1 \cdots n$ be a sample of n measurements of the dependent variable (for simplicity we consider only a single dependent variable). In general, a nonlinear model can be obtained by fitting a linear model between the dependent variable y and the variables in feature space, $\mathbf{z}$, defined by nonlinear transformation in eq 2. (The form of the model is linear in the dependent variable in conformity with KPCR models. The extension of this model to include nonlinear features of the dependent variable must be explored in the future.) This can be written as

$$y = \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{z} \tag{10}$$

where $\boldsymbol{\gamma}$ is the vector of regression coefficients. In KPCR, it is assumed that the feature variables are not linearly independent, and that the true feature variables lie in a lower dimensional subspace $\mathbb{R}^p \subset \mathbb{R}^N$ where $p < N$. Principal components analysis (PCA) is applied to the feature variables to determine the dimension p and a basis for the lower dimensional subspace $\mathbb{R}^p$. If the eigenvectors of $\mathbf{C}_Z$ corresponding to the largest p eigenvalues is given by the columns of $\mathbf{U}_1$, then the first p principal components (which define a basis for $\mathbb{R}^p$) are given by $\mathbf{U}_1^{\mathrm{T}}\mathbf{z}$. In KPCR, a linear regression model between the outputs and chosen principal components is identified as

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{U}_1^{\mathrm{T}}\mathbf{z} \tag{11}$$

where $\mathbf{w}$ is the vector of regression coefficients to be estimated. The regression coefficients are estimated using ordinary least-squares regression (OLS) and are given by

$$\hat{\mathbf{w}} = (\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\tilde{\mathbf{y}} \tag{12}$$

The matrix $\mathbf{B}$ in the above equation is given by

$$\mathbf{B} = \tilde{\mathbf{Z}}^{\mathrm{T}}\mathbf{U}_1 = \tilde{\mathbf{Z}}^{\mathrm{T}}\tilde{\mathbf{Z}}\mathbf{V}_1(\mathbf{\Lambda}_1)^{-0.5} = \mathbf{K}\mathbf{V}_1(\mathbf{\Lambda}_1)^{-0.5} \tag{13}$$

It may be noted that the matrix $\mathbf{B}$ can be computed using the Kernel matrix and eigenvectors of $\mathbf{C}_S$, which does not require the knowledge of the nonlinear mapping functions or the computation of the sample matrix $\tilde{\mathbf{Z}}$. This is the most important utility of using Kernel functions.

In the KPCR method, errors in the independent variables are not explicitly modeled or accounted. The effect of the nonlinear transformation of measured data that contain noise is also unknown, although the use of PCA in feature space to obtain a lower dimensional subspace may implicitly reduce the noise in the transformed independent variables. Furthermore, the use of OLS to estimate the regression parameters is optimal only if the measurements of independent variables do not contain noise. Thus, the KPCR method can essentially be regarded as non-EIV approach for developing a nonlinear regression model. Our goal is to retain the advantages of the KPCR approach, but eliminate its deficiencies, by combining it with the EIV parameter estimation technique described below.

3. PARAMETER ESTIMATION FOR THE EIV MODEL

Valko and Vajda[10] developed a method for estimating the parameters of a nonlinear model given measurements of variables, all of which contain error. The reconciled estimates of variables are also simultaneously obtained. This method assumes that the functional form of the nonlinear model relating the variables is known. Moreover, the covariance matrix of errors in the measurements is assumed to be known. Let the nonlinear relationship between a dependent variable y and independent variables $\mathbf{x}$ be given by

$$f(y, \mathbf{x}, \mathbf{w}) = 0 \tag{14}$$

In the above equation, it is assumed that the parameters $\mathbf{w}$ are unknown and have to be estimated from noisy measurements of y and $\mathbf{x}$.

Equation 14 can also be represented as

$$f(\mathbf{\Psi}, \mathbf{w}) = 0 \tag{15}$$

where $\mathbf{\Psi} = [y\ \mathbf{x}]^T$

DOI: 10.1021/acs.iecr.8b06095
Ind. Eng. Chem. Res. 2019, 58, 11224−11233

Let $\boldsymbol{\Psi}(i) = [\tilde{y}\ \tilde{\mathbf{x}}(i)]^{\mathrm{T}}$ be the ith sample of measurements of the variables $\boldsymbol{\Psi}$ containing error in all variables. The measurement model can therefore be written as

$$\tilde{\boldsymbol{\Psi}}(i) = \boldsymbol{\Psi}(i) + \boldsymbol{\varepsilon}(i) \quad i = 1 \cdots n \tag{16}$$

where $\boldsymbol{\Psi}(i)$ is the true value of the variables, and $\boldsymbol{\varepsilon}(i)$ is the errors in the corresponding measurements. The measurement errors are assumed to follow a Gaussian distribution with zero mean and known positive definite error covariance matrix $\boldsymbol{\Sigma}$ (same for all samples). The errors in different samples are also assumed to be mutually independent and identically distributed, that is,

$$E[\boldsymbol{\varepsilon}(i)\boldsymbol{\varepsilon}^{\mathrm{T}}(j)] = 0 \quad \forall\ i \neq j \tag{17}$$

$$\boldsymbol{\varepsilon}(i) \sim N(0, \boldsymbol{\Sigma}) \tag{18}$$

Under the above assumptions, the maximum likelihood estimates of the parameters and the variables are obtained by minimizing the following objective function

$$\min_{\mathbf{w}, \boldsymbol{\Psi}(i)} J = \sum_{i=1}^{n} (\boldsymbol{\Psi}(i) - \tilde{\boldsymbol{\Psi}}(i))^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Psi}(i) - \tilde{\boldsymbol{\Psi}}(i)) \tag{19}$$

subject to the model constraints

$$f(\boldsymbol{\Psi}(i), \mathbf{w}) = 0, \quad i = 1 \cdots n \tag{20}$$

It can be verified that for some specified estimate of the parameters $\hat{\mathbf{w}}$, the above optimization problem can be decomposed into n decoupled constrained nonlinear problems given by

$$\min_{\boldsymbol{\Psi}(i)} J_i = (\boldsymbol{\Psi}(i) - \tilde{\boldsymbol{\Psi}}(i))^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Psi}(i) - \tilde{\boldsymbol{\Psi}}(i)) \tag{21}$$

subject to the model constraints

$$f(\boldsymbol{\Psi}(i), \hat{\mathbf{w}}) = 0 \tag{22}$$

The above optimization problem for determining the estimates for the variables $\boldsymbol{\Psi}(i)$, corresponding to the ith sample is the same as the nonlinear data reconciliation problem.[2] One approach to solve this nonlinear data reconciliation problem is through successive linearization of the nonlinear constraints leading to a linear data reconciliation problem at every iteration k given by

$$\min_{\boldsymbol{\Psi}(i)} J_i = (\boldsymbol{\Psi}(i) - \tilde{\boldsymbol{\Psi}}(i))^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Psi}(i) - \tilde{\boldsymbol{\Psi}}(i)) \tag{23}$$

subject to

$$\mathbf{A}\boldsymbol{\Psi}(i) = -f(\hat{\boldsymbol{\Psi}}^{k}(i), \hat{\mathbf{w}}) + \mathbf{A}\hat{\boldsymbol{\Psi}}^{k}(i) \tag{24}$$

where $\mathbf{A}$ is the Jacobian matrix given by $\mathbf{A} = \frac{\partial f}{\partial \boldsymbol{\Psi}}$ evaluated at the current estimates $\hat{\boldsymbol{\Psi}}^{k}(i)$. The new estimates are obtained using the solution of the linear DR problem and are given by

$$\hat{\boldsymbol{\Psi}}^{k+1}(i) = \tilde{\boldsymbol{\Psi}}(i) - \boldsymbol{\Sigma}\mathbf{A}^{\mathrm{T}}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A})^{-1}\{\mathbf{A}(\tilde{\boldsymbol{\Psi}}(i) - \hat{\boldsymbol{\Psi}}^{k}(i)) + f(\hat{\boldsymbol{\Psi}}^{k}(i), \hat{\mathbf{w}})\} \tag{25}$$

Let us denote the optimal estimates of $\boldsymbol{\Psi}(i)$ obtained by solving the nonlinear data reconciliation problems as $\hat{\boldsymbol{\Psi}}^{*}(i)$. The reconciled estimates are used to update the estimates of the parameters $\hat{\mathbf{w}}$ in the outer loop of an iterative scheme by minimizing the following nonlinear function.

$$\min_{\mathbf{w}} S(\mathbf{w}) = \sum_{i=1}^{n} [\mathbf{A}_i(\tilde{\boldsymbol{\Psi}}(i) - \hat{\boldsymbol{\Psi}}^{*}(i)) + f(\hat{\boldsymbol{\Psi}}^{*}(i), \mathbf{w})]^{\mathrm{T}} (\mathbf{A}\boldsymbol{\Sigma}\mathbf{A})^{-1} [\mathbf{A}_i(\tilde{\boldsymbol{\Psi}}(i) - \hat{\boldsymbol{\Psi}}^{*}(i)) + f(\hat{\boldsymbol{\Psi}}^{*}(i), \mathbf{w})] \tag{26}$$

It may be noted that both the function f and its gradient $\mathbf{A}_i$ with respect to the variables are dependent on the parameters and are evaluated at the current fixed estimates $\hat{\boldsymbol{\Psi}}^{*}(i)$ and the estimates $\hat{\mathbf{w}}$ of the parameters obtained at each iteration of the above function minimization. The above procedure may be viewed as an alternating nonlinear least-squares approach for simultaneous estimation of the parameters $\mathbf{w}$ and the reconciled estimates of $\boldsymbol{\Psi}(i)$. In the first step, optimal values of the reconciled estimates are obtained for fixed estimates of the parameters $\mathbf{w}$, by minimizing eq 19 subject to eq 20. This is equivalent to minimizing eq 21 subject to eq 22 for each sample, separately. In the second step, updated estimates of the parameters $\mathbf{w}$ are obtained for fixed estimates of $\hat{\boldsymbol{\Psi}}^{*}(i)$ obtained from step 1, by minimizing eq 19 subject to eq 20. The second step is solved using successive linearization of the nonlinear constraints, where eq 26 is the function to be minimized at each iteration of the successive linearization. The two steps are iterated until convergence.

4. SIMULTANEOUS NONLINEAR MODEL IDENTIFICATION AND RECONCILIATION APPROACH

Our proposed approach (named as the EIV-KPCR approach) for developing a nonlinear model from data for the EIV case combines the methods described in the preceding two sections. It is assumed that the error variances in all the variables are known *a priori*. We also assume that the measured data used in developing the model is free of gross errors such as biases. The effect of such biases on the proposed approach and the method to address them is described at the end of this section. The functional form of the regression model obtained using the KPCR method is given by eq 11. This can be combined with eq 9 for the first p eigenvectors and the Kernel property of eq 8, to express the functional model in terms of Kernel functions as

$$y = g(\mathbf{x}, \mathbf{w}) = \sum_{k=1}^{p} w_k \lambda_k^{-0.5} \sum_{j=1}^{n} \nu_{kj} K(\tilde{\mathbf{x}}(j), \mathbf{x}) \tag{27}$$

where v_{kj} is the jth element of the kth eigenvector. Unlike that of KPCR, in our approach the parameters $\mathbf{w}$ in the above model are estimated along with reconciled estimates of all variables using the method described in the preceding section. It may be noted that by adopting this approach the error norm function that is minimized corresponds to a total least-squares formulation, given by eq 19.

Equation 27 can also be represented as

$$y = g(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{n} a_j K(\tilde{\mathbf{x}}(j), \mathbf{x}) \tag{28}$$

where

$$a_j = \sum_{k=1}^{p} w_k \lambda_k^{-0.5} \nu_{kj} \quad \forall\ j = 1 \cdots n \tag{29}$$

Equation 29 expresses the nonlinear model being identified as a weighted sum of Kernels over all the sample points.

DOI: 10.1021/acs.iecr.8b06095
Ind. Eng. Chem. Res. 2019, 58, 11224−11233

However, if we attempt to estimate all of the coefficients a_j, then there will be as many parameters as the measurements, which implies that the model will fit all the measurements exactly. By using the KPCR model form, we have parametrized the coefficients in terms of a reduced number of parameters w_k which is equal to the number of principal components (PCs) chosen. Clearly, the number of PCs chosen should be less than the number of samples to enable estimation of the reduced set of parameters as well as the estimates of the independent and dependent variables corresponding to each measurement. The eigenvalues and eigenvectors corresponding to the chosen principal components can be determined by applying KPCR to the measured set of independent variables. An initial estimate of the parameter vector $\mathbf{w}$ is also obtained in this process. The algorithm for determining the nonlinear model parameters along with estimates of the variables for the EIV case can be summarized as follows.

Step 1. Develop a KPCR model using the training data that are given and optimize the kernel parameters and number of principal components (using cross validation as explained in the examples presented later). Obtain the eigenvalues and eigenvectors corresponding to the PCs chosen. Also obtain an initial estimate of the model parameters $\mathbf{w}$.

Step 2. For the given estimate of the parameters, perform nonlinear data reconciliation for each sample to obtain the corresponding reconciled estimates of the independent and dependent variables. This requires the minimization of the function given by eq 21 such that the estimates satisfy the model given by eq 22.

Step 3. The reconciled estimates for all samples are used to re-evaluate the Kernels and obtain updated estimates of the parameters. If the estimates for the parameters have not converged we return to Step 2 and repeat the procedure.

In order to minimize eq 21, the gradient of the functional model eq 27 with respect to the variables is required. This is given by

$$\mathbf{A} = \left[\frac{\partial\left\{ y - \sum_{k=1}^{p} w_k \lambda_k^{-0.5} \sum_{j=1}^{n} \nu_{kj} K(\tilde{\mathbf{x}}(j), \mathbf{x}) \right\}}{\partial y} \quad \frac{\partial\left\{ y - \sum_{k=1}^{p} w_k \lambda_k^{-0.5} \sum_{j=1}^{n} \nu_{kj} K(\tilde{\mathbf{x}}(j), \mathbf{x}) \right\}}{\partial \mathbf{x}} \right] \tag{30}$$

$$\mathbf{A} = \left[1 - \frac{\partial\left\{ \sum_{k=1}^{p} w_k \lambda_k^{-0.5} \sum_{j=1}^{n} \nu_{kj} K(\tilde{\mathbf{x}}(j), \mathbf{x}) \right\}}{\partial \mathbf{x}} \right] \tag{31}$$

As we have already stated, a main feature of the KPCR model is that the exact nonlinear transformation function is not required. It can be shown that the derivative of kernel matrix elements with respect to the variables can also be computed using kernels without requiring explicit knowledge of the nonlinear transformation function. The derivatives for Gaussian and Polynomial kernels which we have used in our work are provided in the Appendix.

In the above derivation, it has been assumed that the data are mean-centered (otherwise, a constant offset parameter has to be estimated in the regression model). This implies that the Kernel matrix has to be computed for mean centered data (also referred to as the centered Kernel matrix). The centered Kernel matrix can be obtained from the uncentered Kernel matrix without the need for computing the sample matrix in feature space as described in the Appendix.

It may be noted that along with the identified model, the above approach also gives estimates of all the measured variables. These estimates will also be consistent with the identified model. If these estimates are more accurate than the measurements, then they can be regarded as reconciled estimates, despite the fact that the identified model is not exactly equivalent to the physical mass and energy balances, but represents only an approximation of the physical relationships that relate the variables. The identified model can be used for reconciling future process measurements. For this purpose, the entire data set used for developing the identified model (along with the parameters of the identified model) should also be retained for evaluating the Kernels. Furthermore, it may be noted that unlike first-principles model based data reconciliation, estimates of unmeasured variables cannot be obtained, since no knowledge of unmeasured variables is captured by the identified model.

We have assumed that the measurements used to identify the model and reconciled estimates do not contain systematic biases. However, if measurements of some of the variables in the sample contain biases, and the magnitude of the biases are constant and do not vary with the samples, then the model and reconciled estimates will not be affected since these constant biases will be included in the identified model as constant offset parameters. It may be noted that centered Kernels are used in our approach for identifying the model, and therefore the model implicitly includes offset parameters. On the other hand, if a bias is present only in few of the samples, then these samples have to be treated as outliers. Such outlier samples can be identified and removed using residual analysis in an iterative procedure, similar to the approach used in regression for dealing with outlier samples. This is left as a future extension of the proposed approach.

5. SIMULATION RESULTS AND DISCUSSION

5.1. Example 1: Reconciliation of Saturated Pressure and Temperature Data. The first example we use to demonstrate our proposed method is the development of a model (correlation) for estimating saturation pressure as a function of temperature for a pure component. For the purpose of developing this model, temperature is chosen as the independent variable (x) and pressure is chosen as the dependent variable (y). It may, however, be noted that both the temperature and pressure are measured quantities, and hence they both can contain error. Although, the Antoine equation is known to provide a good estimation of saturation pressures of most pure components, we use this example to evaluate the quality of the model developed using our proposed approach without knowledge of the functional form. For developing a model using KPCR or the proposed EIV-KPCR method, the Kernel function parameters (Kernel width for a Gaussian Kernel or the order of the polynomial for a Polynomial Kernel) and number of principal components have to be chosen. These are referred to as the metaparameters of the model and their choice can affect the model performance. The optimal choice of these metaparameters which results in the best model performance is also discussed as part of the simulation study.

DOI: 10.1021/acs.iecr.8b06095
Ind. Eng. Chem. Res. 2019, 58, 11224−11233

Measured data for saturated pressures of hexane for different temperatures are simulated using the Antoine equation given by

$$\log P^{\text{sat}} = A - \frac{B}{T + C} \tag{32}$$

where P^{sat} is in mm Hg and T is in degree Celsius.

The values of A, B, and C for n-hexane are 6.87773, 1171.53, and 224.366, respectively. For simulating the measured data, a temperature range of 10 to 70 °C is chosen, with an interval of one degree Celsius. Corresponding to this, the saturated pressures are computed using eq 33 and are found to be in the range 100−800 mmHg. The number of measured samples is equal to 61, and the standard deviation of the true pressures for this data set is 210 mmHg, while the standard deviation of the true temperatures is 18 °C. We consider the following two simulated data sets for this example.

Two simulated data sets are generated for studying the effect of Kernel parameters as well as for demonstrating the effect of errors in independent variables. In data set 1, only the measurements of pressure contain errors while temperature measurements do not contain any error (which is the usual assumption under which parameters of the Antoine equation are estimated). A random measurement error from a Gaussian distribution with mean zero and standard deviation of 2 mmHg is added to the true saturated pressures to generate the measurements. The error variance used in the simulation is comparable to the accuracy of typically available pressure sensors. The signal-to-noise ratio (SNR), which is defined as the ratio of variance in true signal to the noise variance for the simulated pressures is approximately equal to 10000.

For generating data set 2, random errors are added to both the true pressure and temperature data to simulate the measurements. The standard deviation of error added to pressure is 2 mmHg, while that for temperature measurement is 0.5 °C. The SNR in pressure and temperature measurements is 10 000, and 3000, respectively. Similar test data sets consisting of 30 samples are also generated using randomly chosen temperatures in the range 10−70 °C for evaluating the performance of the models developed.

5.1.1. Performance Metrics. The performance of the proposed EIV-KPCR approach (and other methods used in the comparative study) can be assessed based on the root-mean-square of the differences between the predicted and true pressures (RMSE) on a test data set defined by

$$\text{RMSE} = \frac{\sqrt{\sum_{i=1}^{n} (P_S^{\text{sat}}(i) - \hat{P}_S^{\text{sat}}(i))^2}}{n} \tag{33}$$

where $P_S^{\text{sat}}(i)$ is the true saturation pressure and $\hat{P}_S^{\text{sat}}(i)$ is the estimated pressure of sample i in data set S. The data set S could be either the training data set used to build the model, or a test data set used to assess the predictive capability of the identified model. However, this measure will not be useful for choosing the optimal values of the Gaussian Kernel width and number of PCs, since the true values of the variables of a data set will not be available in practice. Instead, we have to determine the optimal values of these parameters based on the root-mean-square differences between measured and predicted pressures (RM-PRESS) for a data set. We therefore also compute RM-PRESS values for a data set defined as

$$\text{RM} - \text{PRESS} = \frac{\sqrt{\sum_{i=1}^{n} (\tilde{P}_S^{\text{sat}}(i) - \hat{P}_S^{\text{sat}}(i))^2}}{n} \tag{34}$$

where $\tilde{P}_S^{sat}(i)$ is the measured saturation pressure of sample i in the test data set S.

We use RM-PRESS to determine the optimal values of the metaparameters (Kernel width and number of PCs), because this measure can be computed using the given measurements. However, it has to be noted that RM-PRESS should be computed for a test set and not the training data set, because the RM-PRESS of the training data set can be trivially reduced to zero by choosing as many PCs as the number of samples in the training set. The method of using RM-PRESS to determine the optimal values of the metaparameters of a model is also known as the cross-validation approach.[17] Once the best model is obtained, RMSE is used as the metric to assess the performance of the model for data reconciliation, because this measure indicates how close the reconciled estimates are to the true values.

5.1.2. Performance of KPCR Method. We first study the effect of the choice of the Kernel parameters and number of principal components on the quality of the model developed using KPCR, in order to determine their best choice. Although, both polynomial and Gaussian Kernels were evaluated for this example, it was found that the use of the Gaussian Kernel gave better results, and therefore the results obtained using Gaussian Kernels are reported.

For each value of the Kernel width parameter value in the range 1−800, the KPCR method was used to obtain the model for different choices of the number of principal components (PCs), and RM-PRESS values for the models were computed using the test data set. The maximum number of PCs (number of regression parameters) was limited to 20 which is equal to about 33% of the number of samples in the training set. The optimal value of the number of PCs corresponding to each value of the Kernel width parameter is the one which gives the least RM-PRESS value.

Figure 1 shows the variation of the optimal number of principal components obtained for different Kernel widths,

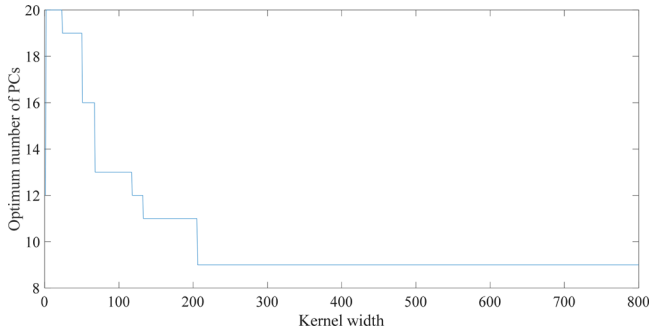


Figure 1. Optimal number of principal components for different Kernel widths for data set 1.

when KPCR is applied to training data set 1. The optimum number of PCs is found to be less than 13 above a Kernel width of 100. Figure 2 shows the variation of the minimum RM-PRESS values obtained for different choices of the Kernel width parameter (and corresponding optimal number of PCs). The least RM-PRESS value of 1.8 mmHg is obtained for a choice of the Kernel width equal to 270 and corresponding optimal number of PCs equal to 9, although the optimum RM-

DOI: 10.1021/acs.iecr.8b06095
Ind. Eng. Chem. Res. 2019, 58, 11224−11233

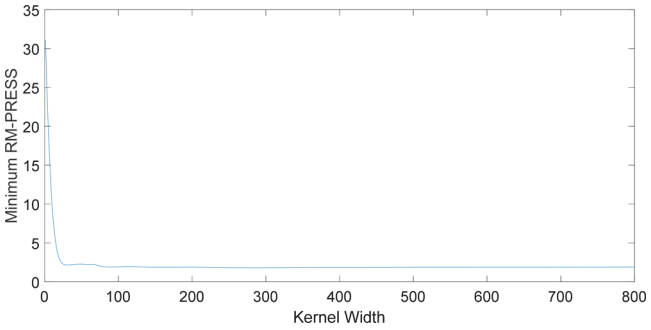


Figure 2. Minimum RM-PRESS of predicted pressures for data set 1.

PRESS does not change significantly above a Kernel width of 200.

The RMSE value of the saturated pressures predicted using the best KPCR model corresponding to a Kernel width of 270 and number of PCs equal to 9 is found to be 0.6 mmHg. Since the standard deviation of errors in the measured pressures is 2 mmHg, this implies that on an average the KPCR model is able to reduce 70% of the error in the measurements. Thus, in the absence of errors in the independent variables the standard KPCR method is able to simultaneously identify a nonlinear model and obtain reconciled estimates of the dependent variable which are better than the corresponding measurements.

KPCR is applied to data set 2 in which the measurements of both the independent and dependent variables contain errors. Figure 3 shows the optimal number of PCs to be chosen for

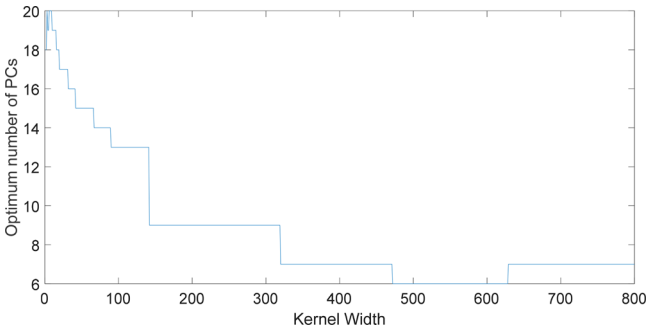


Figure 3. Optimal number of principal components for different Kernel widths for data set 2.

each Kernel width and Figure 4 shows the variation of the optimal RM-PRESS values for different Kernel widths. For a low Kernel width (around 30), the RM-PRESS in the

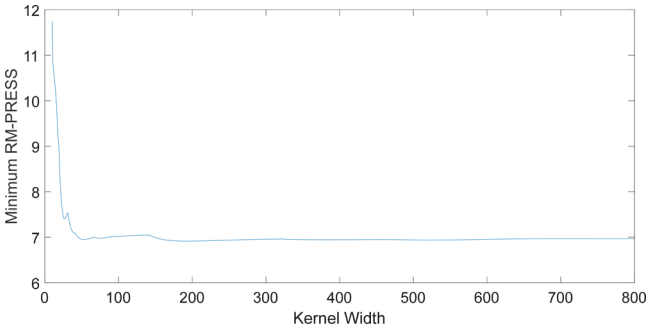


Figure 4. Minimum RM-PRESS of predicted pressures for data set 2.

predicted pressures is as high as 7.5 mmHg, and passes through a minimum as the width increases. The least value of RM-PRESS was obtained by choosing a Kernel width of 194 and 9 PCs. Corresponding to this best model, the RMSE value in the predicted pressures is found to be 6.86 mmHg. It may be noted that this is more than thrice the standard deviation of error in the measured pressures. This implies that the predictions obtained using the best KPCR model is more inaccurate as compared to the direct measurement of the pressure. This occurs because the KPCR method does not account for errors in the temperature measurements (independent variable) in developing the model.

5.1.3. Performance of Proposed EIV-KPCR Approach. The results of applying KPCR to the two data sets not only gives an indication of the appropriate parameter choices for developing a good model, but also points out to the need for taking into account the errors in independent variables while developing the model. This is brought out clearly by applying our proposed EIV-KPCR approach to data set 2. To make a fair comparison between KPCR and our proposed approach, a Kernel width of 194 and 9 PCs are chosen for both methods to build the model. Table 1 gives the RMSE in temperature and

Table 1. RMSE of Estimates Obtained Using Different Approaches for the Test Data Set

variable	measurements	KPCR-EIV	KPCR	Antoine equation
pressure (mmHg)	1.73	1.63	6.86	1.70
temp (°C)	0.4236	0.4177	0.4236	0.19

pressure estimates obtained using KPCR and the proposed methods. It may be noted, that the proposed method simultaneously reconciles both the pressure and temperature measurements while developing the model. Comparing the results of columns 3 and 4 in Table 1, we clearly observe that the proposed method is able to reduce the errors in both the temperature and pressure measurements. In contrast, the temperature estimates obtained using the KPCR method contain the same error as in the measurements (since the method assumes that these measurements are perfect), while the pressure estimates contain more error than the measurements. About 10% of the error in the pressure measurements is reduced, while 1.5% of the error in temperature measurements is reduced by the proposed method. As a further comparison, we perform nonlinear data reconciliation using the true Antoine equation relation that relates the saturated pressure and temperature and report the results in column 5 of Table 1. Since the true relationship is used for the reconciliation, this gives the maximum possible reduction in error. It is observed that the RMSE in the reconciled pressure estimates obtained using the proposed approach and true functional model are almost the same, while the error in the temperature measurements using the true relation is 50% lower than that obtained using the identified model in the proposed approach.

For the above examples, we have not reported the nine eigenvectors ($\mathbf{v}_k$) and their corresponding eigenvalues (λ_k), and the optimized estimates of the parameters ($\mathbf{w}_k$) required to characterize the identified nonlinear function, due to the large amount of data to be reported. Furthermore, this information is useful only if they go hand in hand with the training data set.

5.2. Example 2: Reconciliation of Heat Exchanger Data. The second example that we consider is the

DOI: 10.1021/acs.iecr.8b06095
Ind. Eng. Chem. Res. 2019, 58, 11224−11233

reconciliation of flow and temperature measurements for a heat exchanger. This example corresponds to a multiple input−single output (MISO) case study in contrast to example 1, which is a single input−single output (SISO) case. A counter current heat exchanger is considered with specific heat capacities of the cold and hot streams assumed to be linear functions of temperature. The flow rates of the cold and hot streams and the inlet and outlet temperatures of the two streams are assumed to be the measured variables. The heat balance for the exchanger is given by the following equation:

$$\alpha \int_{T_{h2}}^{T_{h1}} \dot{m}_h C_{ph}(T)\, dT = \int_{T_{c1}}^{T_{c2}} \dot{m}_c C_{pc}(T)\, dT \tag{35}$$

In the above equation the factor α represents the fraction of energy of the hot stream that is transferred to the cold stream. If α is equal to 1, then there is no energy loss, whereas if it is less than 1, then the remaining fraction of energy is assumed to be lost to the ambient. The nonlinear dependence of specific heat capacity on the temperature for the hot and cold streams is given by the following correlations.[18]

$$C_{ph}(T) = 0.409 + 0.00096T \ \frac{\text{kcal}}{\text{kg °C}} \tag{36}$$

$$C_{pc}(T) = 0.444 + 0.00101T \ \frac{\text{kcal}}{\text{kg °C}} \tag{37}$$

where T is in °C. Since this process has six variables related by an energy balance, it has five independent variables and one dependent variable. The outlet cold-water temperature is chosen as the dependent variable. Simulated measured data containing errors are generated as follows. The independent variables are chosen in the range as given in Table 2, and the dependent variable (cold stream outlet temperature) is computed by solving the energy balance to obtain the true values of all variables. Random errors drawn from the normal distribution with mean zero and variance as given in Table 2 are added to the true values to generate the measurements. The training data set used for developing the model consists of 50 samples, while a test data set containing 25 samples is also generated. The signal-to-noise ratio in each measurement is also given in Table 2.

Table 2. Characteristics of Simulated Data Set for Example 2

no.	stream	minimum value	maximum value	variance of noise	SNR
1	cold fluid flow rate (tons/h)	153	201	1 (tons/h)^2	450
2	hot fluid flow rate (tons/h)	106	175	1 (tons/h)^2	177
3	cold water inlet temperature (°C)	125	160	2 (°C)^2	40
4	hot fluid inlet temperature (°C)	241	338	2 (°C)^2	402
5	hot fluid outlet temperature (°C)	162	295	2 (°C)^2	441
6	cold fluid outlet temperature (°C)	144	267	2 (°C)^2	315

For this example the choice of a polynomial Kernel gave good models. This is to be expected, since the energy balance is a polynomial equation of order 3. By applying the KPCR method, it was found that the least RM-PRESS was obtained using a polynomial Kernel of order 2 and choosing seven principal components. Using these parameters, the proposed method was applied to simultaneously reconcile the measurements and obtain the nonlinear model. Table 3 shows the RMSE values of the reconciled variables for the training data obtained using the proposed method. The RMSE values of the reconciled estimates obtained using nonlinear data reconciliation based on the true functional relationship between variables for training data set are also shown in the last column of the table. The results show that the proposed method is able to provide estimates which are marginally better than the measurements. In comparison, the standard deviations of errors in reconciled estimates of temperatures obtained using the first-principles model is between 8 and 22% less than those in the corresponding measurements. Since there is only one equation between the six variables that can be exploited for reconciliation, the improvement that can be achieved for this process is limited.

Table 3. RMSE of Reconciled Estimates for Training Data Set

variables	measured data	proposed method (polynomial order 2)	known functional model
cold fluid outlet temperature (°C)	1.37	1.48	1.23
cold water inlet tcemperature (°C)	1.53	1.37	1.19
hot fluid inlet temperature (°C)	1.78	1.80	1.60
hot fluid outlet htemperature (°C)	1.27	1.20	1.17
hot fluid flow rate (tons/h)	0.92	0.93	0.89
cold fluid flow rate (tons/h)	0.94	0.94	0.95

To verify whether the nonlinear model developed from data using the proposed method has any correspondence with the energy balance equation, we assessed how well the reconciled estimates obtained using the proposed method satisfy the true energy balance (eq 36). The RMSE of the constraint residual (energy imbalance) for the reconciled estimates obtained using the proposed method was 52.9 kcal/s, while the energy imbalance due to the measurement errors was 54.9 kcal/s. This indicates that the nonlinear model obtained using the proposed approach is a reasonable approximation of the energy balance equation. It may be noted that the reconciled estimates obtained using the first-principles model exactly satisfy the energy balance and hence the energy imbalance is zero.

We also assessed the performance of the identified nonlinear model for reconciling the test data. Table 4 compares the RMSE values of the reconciled estimates using the nonlinear model developed by the proposed approach with that obtained from the first-principles model on the test data. The reconciled estimates obtained using the proposed method is not significantly better or worse than the measurements, whereas the reconciled estimates obtained using the first-principles model show some improvement over the measurements.

The main benefit of simultaneously identifying the model and reconciling measurements using the proposed approach arises when the first-principles model is itself not correct (due to incorrect assumptions made in deriving it). To illustrate this,

DOI: 10.1021/acs.iecr.8b06095
Ind. Eng. Chem. Res. 2019, 58, 11224−11233

Table 4. RMSE of Reconciled Estimates for Test Data Set

variables	measured data	proposed method (polynomial order 2)	known functional model
cold fluid outlet temperature (°C)	1.47	1.78	1.33
cold water inlet temperature (°C)	1.25	1.23	0.91
hot fluid inlet temperature (°C)	1.66	1.61	1.53
hot fluid outlet temperature (°C)	1.59	1.65	1.50
hot fluid flow rate (tons/h)	1.00	1.00	1.00
cold fluid flow rate (tons/h)	0.97	0.99	0.97

we simulated a training data set consisting of 50 samples and a test data set of 25 samples for the heat exchanger example which included a 20% energy loss. In other words, the true data for every sample were generated to satisfy eq 36, with the factor α chosen equal to 0.8. If we reconcile the measurements, using a first-principles model assuming that the energy balance is satisfied without any loss (because we are unaware of the energy loss), then the RMSE values in the reconciled estimates of all variables obtained are shown in the last column of Table 5. It can be observed from the results that, on an average, the

Table 5. RMSE of Reconciled Estimates of Test Data for 20% Energy Loss in the Exchanger

variables	measured data	proposed method (polynomial order 2)	known functional model
cold fluid outlet temperature (°C)	1.18	1.14	3.22
cold water inlet temperature (°C)	1.34	1.57	3.75
hot fluid inlet temperature (°C)	1.21	1.17	2.52
hot fluid outlet temperature (°C)	1.17	1.26	2.59
hot fluid flow rate (tons/h)	0.98	0.93	1.02
cold fluid flow rate (tons/h)	1.01	1.03	1.14

standard deviations of errors in the temperature estimates are 2−3 times higher than the standard deviations of errors in the corresponding measured values, implying that reconciled estimates are worse than the measured values. This is due to the incorrect first-principles model being used. On the other hand, if we use the proposed method, then the RMSE values in the estimates obtained are shown in the third column of Table 5. These results are obtained using a polynomial Kernel of order 2 and number of PCs equal to seven (which is the best model obtained based on RM-PRESS on the training data set). The results show that the reconciled estimates are marginally better than the measured values for hot fluid inlet and cold fluid outlet temperature and marginally worse than the measured values of hot fluid outlet and cold fluid inlet temperatures. Although, reconciled estimates obtained using the identified model are no better or worse than the measured values, if we substitute these estimates in the energy balance equation, the RMSE of the imbalance is found to be 219 kcal/s. The RMSE of the true imbalance in the energy equation is 249 kcal/s. This indicates that the identified model is a good representation of the heat exchanger energy balance with energy loss.

6. SUMMARY

We have proposed a novel method for reconciling a data set while simultaneously identifying a nonlinear steady state model of the process. The method combines an EIV formulation with a Kernel Principal Component Regression modeling framework. The main advantage of the proposed approach is for complex processes for which it is difficult to develop a first-principles model. Since in the proposed approach the model is developed from data, it is likely to provide more accurate estimates, especially for those processes for which the structure may not be completely known or for which the assumptions used in developing a first-principles model may not be valid. The model identified using the proposed approach can also be used for reconciliation of future data. In this work it is assumed that variances of errors corrupting all measurements are known. The extension of the proposed approach to simultaneously estimate the error variances, nonlinear model, and reconciled estimates may be explored in the future, which will be analogous to the iterative PCA method described for linear systems in Narasimhan and Shah.[6] In the proposed approach, the training data are assumed to be free of gross errors. The proposed approach can be extended to identify and eliminate measurements containing gross errors by combining it with gross error detection strategies[18] or by using robust regression methods.[19]

APPENDIX

Gradient of Kernel Functions

The gradient of a Kernel is required for evaluating the derivatives of the functional model given by eq 32.

$$\mathbf{A}_i = \left[1 \quad \frac{\partial\left\{-\sum_{k=1}^{p}\gamma(k)\lambda^{-0.5}(k)\sum_{j=1}^{n}\nu_{kj}K(\mathbf{x},\tilde{\mathbf{x}}(j))\right\}}{\partial \mathbf{x}} \right]_{\mathbf{x}=\hat{\mathbf{x}}(i)} \tag{A1}$$

where K represents the centered Kernel function. We can expand the centered kernel in terms of un-centered kernels (K^{UC}) as follows.[5]

$$\begin{aligned} K(\mathbf{x},\mathbf{x}(j)) = K^{UC}(\mathbf{x},\tilde{\mathbf{x}}(j)) &- \frac{1}{n}\sum_{l=1}^{n}K^{UC}(\mathbf{x},\tilde{\mathbf{x}}(l)) \\ &- \frac{1}{n}\sum_{l=1}^{n}K^{UC}(\tilde{\mathbf{x}}(\mathbf{j}),\tilde{\mathbf{x}}(l)) \\ &+ \frac{1}{n}\sum_{l=1}^{n}\sum_{m=1}^{n}K^{UC}(\tilde{\mathbf{x}}(l),\tilde{\mathbf{x}}(m)) \end{aligned} \tag{A2}$$

We note that the last two terms in the above equation are independent of the variables and are dependent only on the sample points and hence are constants for a given sample set.

Polynomial Kernel. For a polynomial Kernel of order p, the gradient is given by

DOI: 10.1021/acs.iecr.8b06095
Ind. Eng. Chem. Res. 2019, 58, 11224−11233

$$\mathbf{A}_i = \left[1 - \sum_{k=1}^{p} \gamma(k)\lambda^{-0.5}(k) \sum_{j=1}^{n} \nu_{kj} \frac{\partial}{\partial \mathbf{x}} \left((1 + \langle \mathbf{x}, \tilde{\mathbf{x}}(j) \rangle)^p - \frac{1}{n} \sum_{l=1}^{n} (1 + \langle \mathbf{x}, \tilde{\mathbf{x}}(l) \rangle)^p \right) \right]_{\mathbf{x}=\hat{\mathbf{x}}(i)} \tag{A3}$$

$$= \left[1 - \sum_{k=1}^{p} \gamma(k)\lambda^{-0.5}(k) \sum_{j=1}^{n} \nu_{kj} p \left((1 + \langle \mathbf{x}, \tilde{\mathbf{x}}(j) \rangle)^{p-1} \tilde{\mathbf{x}}(j) - \frac{1}{n} \sum_{l=1}^{n} (1 + \langle \mathbf{x}, \tilde{\mathbf{x}}(l) \rangle)^{p-1} \tilde{\mathbf{x}}(l) \right) \right]_{\mathbf{x}=\hat{\mathbf{x}}(i)} \tag{A4}$$

$$= \left[1 - \sum_{k=1}^{p} \gamma(k)\lambda^{-0.5}(k) \sum_{j=1}^{n} \nu_{kj} p \left(\tilde{\mathbf{x}}(j) K_{p-1}^{UC}(\mathbf{x}, \tilde{\mathbf{x}}(j)) - \frac{1}{n} \sum_{l=1}^{n} \tilde{\mathbf{x}}(l) K_{p-1}^{UC}(\mathbf{x}, \tilde{\mathbf{x}}(l)) \right) \right]_{\mathbf{x}=\hat{\mathbf{x}}(i)} \tag{A5}$$

Gaussian Kernel. For a Gaussian Kernel the gradient is given by

$$\mathbf{A}_i = \left[1 - \sum_{k=1}^{p} \gamma(k)\lambda^{-0.5}(k) \sum_{j=1}^{n} \nu_{kj} \frac{\partial}{\partial \mathbf{x}} \left\{ K_G^{UC}(\mathbf{x}, \tilde{\mathbf{x}}(j)) - \frac{1}{n} \sum_{l=1}^{n} K_G^{UC}(\mathbf{x}, \tilde{\mathbf{x}}(l)) \right\} \right]_{\mathbf{x}=\hat{\mathbf{x}}(i)} \tag{A6}$$

where $K_G^{UC}(\boldsymbol{x}, \tilde{\mathbf{x}}(j))$ is the uncentered Gaussian Kernel. The second term in the RHS of eq A6 denoted as $\mathbf{A_{12}}$ is given by

$$\mathbf{A}_{12} = \left[- \sum_{k=1}^{p} \gamma(k)\lambda^{-0.5}(k) \sum_{j=1}^{n} \nu_{kj} \frac{\partial}{\partial \mathbf{x}} \left\{ e^{-\|\tilde{\mathbf{x}}(j) - \mathbf{x}\|^2 / \mathbf{W}} - \frac{1}{n} \sum_{l=1}^{n} e^{-\|\tilde{\mathbf{x}}(l) - \mathbf{x}\|^2 / \mathbf{W}} \right\} \right]_{\mathbf{x}=\hat{\mathbf{x}}(i)} \tag{A7}$$

$$= \left[- \sum_{k=1}^{p} \gamma(k)\lambda^{-0.5}(k) \sum_{j=1}^{n} \nu_{kj} \left\{ \frac{2}{W} (\tilde{\mathbf{x}}(j) - \mathbf{x}) e^{-\|\tilde{\mathbf{x}}(j) - \mathbf{x}\|^2 / \mathbf{W}} - \frac{1}{n} \sum_{l=1}^{n} \frac{2}{W} (\tilde{\mathbf{x}}(l) - \mathbf{x}) e^{-\|\tilde{\mathbf{x}}(l) - \mathbf{x}\|^2 / \mathbf{W}} \right\} \right]_{\mathbf{x}=\hat{\mathbf{x}}(i)} \tag{A8}$$

$$= - \sum_{k=1}^{p} \gamma(k)\lambda^{-0.5}(k) \sum_{j=1}^{n} \nu_{kj} \left\{ \frac{2}{W} (\tilde{\mathbf{x}}(j) - \mathbf{x}) K_G^{UC}(\mathbf{x}, \tilde{\mathbf{x}}(j)) - \frac{1}{n} \sum_{l=1}^{n} \frac{2}{W} (\tilde{\mathbf{x}}(l) - \mathbf{x}) K_G^{UC}(\mathbf{x}, \tilde{\mathbf{x}}(l)) \right\}_{\mathbf{x}=\hat{\mathbf{x}}(i)} \tag{A9}$$

AUTHOR INFORMATION

Corresponding Author

*E-mail: naras@iitm.ac.in.

ORCID

Shankar Narasimhan: 0000-0002-0558-0206

Present Address

#Currently an Associate Engineer with Phillips 66, USA.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Veverka, V. V.; Madron, F. *Material and energy balancing in the process industries*; Elsevier: Netherlands, 1997.

(2) Narasimhan, S.; Jordache, C. *Data reconciliation and gross error detection: An intelligent use of process data*; Gulf Hemisphere: Houston, USA, 2000.

(3) Romagnoli, J. A.; Sanchez, M. C. *Data Processing and Reconciliation for Chemical Process Operations*; Academic Press: San Diego, USA, 2000.

(4) Bhat, S. A.; Saraf, D. N. Steady-state identification, gross error detection, and data reconciliation for industrial process units. *Ind. Eng. Chem. Res.* **2004**, *43*, 4323.

(5) Hodouin, D.; Jämsä-Jounela, S. L.; Carvalho, M. T.; Bergh, L. State of the art and challenges in mineral processing control. *Control Engineering Practice* **2001**, *9*, 995.

(6) Narasimhan, S.; Shah, S. L. Model Identification and error covariance matrix estimation from noisy data using PCA. *Control Engineering Practice* **2008**, *16*, 146.

(7) Narasimhan, S.; Bhatt, N. P. Deconstructing principal component analysis using a data reconciliation perspective. *Comput. Chem. Eng.* **2015**, *77*, 74.

(8) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning—Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, USA, 2009.

(9) Jade, A. M.; Srikanth, B.; Jayaraman, V. K.; Kulkarni, B. D.; Jog, J. P.; Priya, L. Feature extraction and denoising using kernel PCA. *Chem. Eng. Sci.* **2003**, *58*, 4441.

(10) Valko, P.; Vajda, S. An extended Marquardt type procedure for fitting error in variables models. *Comput. Chem. Eng.* **1987**, *11*, 37.

(11) Faber, R.; Arellano-Garcia, H.; Li, P.; Wozny, G. An optimization framework for parameter estimation of large-scale systems. *Chem. Eng. Process.* **2007**, *46*, 1085.

(12) Francken, J.; Maquin, D.; Ragot, J.; Bele, B. Simultaneous data reconciliation and parameter estimation. Application to a basic oxygen furnace. *IFAC Proceedings Volumes* **2009**, *42*, 86.

(13) Camara, M. M.; Soares, R. M.; Feital, T.; Anzai, T. K.; Diehl, F. C.; Thompson, H. T.; Pinto, J. C. Numerical aspects of data reconciliation in industrial applications. *Processes* **2017**, *5*, 56.

(14) Rosipal, R.; Trejo, L. J. Kernel partial least squares regression in reproducing kernel Hilbert space. *J. Mach. Learn. Res.* **2001**, *2*, 97.

(15) Wu, W.; Massart, D. L.; Jong de, S. The kernel PCA algorithm for wide data. Part I: Theory and algorithm. *Chemom. Intell. Lab. Syst.* **1997**, *36*, 165.

(16) Scholkopf, B.; Smola, A.; Muller, K. R. (1998). Nonlinear component analysis as a Kernel eigenvalue problem. *Neural Computation* **1998**, *10*, 1299.

(17) Valle, S.; Li, W.; Qin, S. J. Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods. *Ind. Eng. Chem. Res.* **1999**, *38*, 4389.

(18) Renganathan, T.; Narasimhan, S. A strategy for detection of gross errors in nonlinear processes. *Ind. Eng. Chem. Res.* **1999**, *38*, 2391.

(19) Wang, D.; Romagnoli, J. A. A framework for robust data reconciliation based on a generalized objective function. *Ind. Eng. Chem. Res.* **2003**, *42*, 3075.

DOI: 10.1021/acs.iecr.8b06095
Ind. Eng. Chem. Res. 2019, 58, 11224−11233