Preprints of the 23rd IFAC World Congress
Busan, Republic of Korea, August 23-28, 2026

Recursive Identification of EIV-ARX Models for Time Varying SISO Processes

Deepanjhan Das * **Shankar Narasimhan** **

* *Department of Chemical Engineering, Indian Institute of Technology Madras, (e-mail: deepanjhan.iitm22@gmail.com)*
** *Department of Chemical Engineering, Indian Institute of Technology Madras, (e-mail: naras@iitm.ac.in)*

Abstract: This paper proposes a recursive algorithm, rARX-DIPCA, for identifying errors-in-variables autoregressive models with exogenous input (EIV-ARX), for tracking time-varying SISO processes. Building on a recently developed recursive iterative PCA method, the proposed algorithm recursively updates model parameters and noise variances as new measurements arrive, without storing historical data beyond a specified lag window. The method enables real-time adaptation to sensor degradation, and changes in model coefficients. The algorithm simultaneously identifies process order, time delay, and noise variances while maintaining computational efficiency through online covariance updates. Simulation studies on benchmark systems demonstrate effective tracking performance and practical applicability.

Keywords: Time/parameter varying system identification, Linear system identification, ARX processes, Errors-in-variables, Recursive identification

1. INTRODUCTION

Linear dynamic model identification is fundamental to modern process system engineering, with applications spanning from soft sensing (Cao et al., 2018), process optimization, model predictive control and fault diagnosis (Prakash et al., 2005). Among various model structures, the autoregressive with exogenous input (ARX) model for single input and single output (SISO) systems, is preferred due to its mathematical tractability and ability to approximate complex dynamics.

Unlike the classical framework, errors-in-variables (EIV) modeling assumes that both exogenous input and output measurements are corrupted by noise, as illustrated in Fig. 1. The EIV modeling framework is particularly relevant when the input and output data are directly obtained from some operating processes rather than through designed experiments. From a theoretical standpoint, this framework shares deep structural connections with dynamic generalized factor analysis, where latent true variables must be similarly estimated from noisy observation streams (Picci et al., 2023).

Several approaches have been proposed for identifying EIV-ARX models for SISO systems. These include methods based on bias compensation (Ikenoue et al., 2005), bias elimination (Zheng, 2002), and the dynamic Frisch scheme (Diversi et al., 2010). All of these methods obtain consistent estimates of the ARX model parameters along with the input and output noise variances. However, they assume the process order to be known or specified. The Instrument Variable (IV) approach (Söderström and Stoica, 2002) consistently estimates the model parameters without the need to estimate the noise variances. Recently Maurya et al. (2022) proposed a modified dynamic iterative

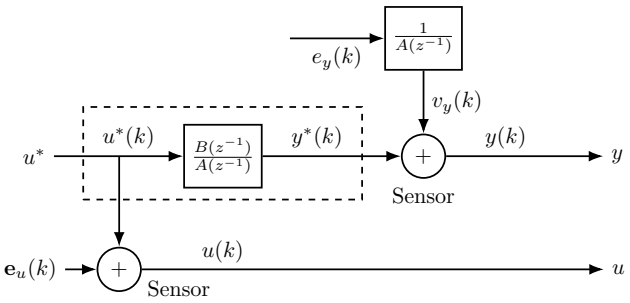


Fig. 1. Linear dynamic EIV-ARX model architecture

PCA approach that jointly estimates the process order, time delay, noise variances, and model parameters without requiring prior knowledge from the user. This technique builds on the basic Koopmans-Levin approach (Fernando and Nicholson, 1985), by including an iterative procedure to estimate noise variances, and a theoretical criterion for estimating the order of the process based on analysis of the eigenvalues of the covariance matrix.

Most industrial processes exhibit time-varying characteristics due to catalyst deactivation, equipment fouling, or changing operating conditions. For such systems, Zhao et al. (2018) addressed the specific challenge of robust identification of output error (OE) ARX models with time-varying time-delays using a Variational Bayesian (VB) approach, whereas Zhang et al. (2022) developed an expectation-maximization (EM) framework to identify EIV-ARX models with time-varying time-delays. Their works demonstrate the importance of adapting to time-varying system parameters, though their approaches remain in batch processing domain and assume an a priori knowledge of process order. Batch methods become

©2026 the authors. Accepted by IFAC for publication under a Creative Commons Licence CC-BY-NC-ND

computationally prohibitive for long data sequences, and cannot adapt to more general time-varying processes.

The need for adaptive identification has motivated the developments of recursive approaches, such as continuous adaptation of ordinary least squares (OLS) regression models, PCA based model (Li et al., 2000), reduced-order models (Prakash and Huang, 2025). However, these methods make simplistic assumptions about the noise variances that eliminates the requirement of knowing changes in noise variances to perform model updates. Most recently, Pradeep and Narasimhan (2025) introduced a recursive iterative PCA (RIPCA) framework for updating static (steady-state) EIV models and unknown, heteroskedastic noise variances online without storing past data.

In this paper, we extend the RIPCA method by drawing ideas from the modified dynamic iterative PCA approach (Maurya et al., 2022) appropriately to handle dynamic structure and colored noise characteristics of EIV-ARX models. We thereby provide an efficient recursive identification approach that is capable of continually adapting to changes in heteroskedastic noise variances, and coefficients of EIV-ARX models.

The remainder of this paper is structured as follows. In Section 2, we provide a formal description of the EIV identification problem of SISO ARX processes and the modified dynamic iterative PCA method. This forms the basis of the proposed recursive identification procedure of EIV-ARX models for time varying processes, which we present in Section 3. In Section 4, we demonstrate the effectiveness of the proposed algorithm in adapting to slow sensor degradation and changing process parameters using two simulation studies. Section 5 addresses the concluding remarks and potential future extensions.

2. FOUNDATIONS

2.1 Dynamic Iterative PCA for identifying EIV-ARX models of SISO processes

In this section, we briefly discuss the dynamic iterative PCA (DIPCA) method for identifying an EIV-ARX model, since this forms the basis for the proposed recursive EIV-ARX model identification approach.

The deterministic linear time-invariant (LTI) SISO discrete dynamical model between noise-free input (u^*) and output (y^*) variables can in general be described as

$$y^*(k) = \sum_{i=1}^{n_y} a_i y^*(k-i) + \sum_{j=D}^{n_u} b_j u^*(k-j) \tag{1}$$

where D, n_u, n_y are input-output delay, input and output orders, respectively. The process order is defined as $\eta = \max(n_u, n_y)$. In the case of an EIV-ARX process, as shown in Fig. 1, where $A(z^{-1}) = 1 - \sum_{i=1}^{n_y} a_i z^{-1}$ and $B(z^{-1}) = \sum_{j=D}^{n_u} b_j z^{-1}$, both the input and output measurements contain additive noise which are modeled as follows:

$$y(k) = y^*(k) + v_y(k); \quad e_y(k) \sim \mathcal{N}\left(0, \sigma^2_{e_y}\right) \tag{2a}$$

$$v_y(k) = \sum_{i=1}^{n_y} a_i v_y(k-i) + e_y(k); \tag{2b}$$

$$u(k) = u^*(k) + e_u(k); \quad e_u(k) \sim \mathcal{N}\left(0, \sigma^2_{e_u}\right) \tag{2c}$$

where e_y and e_u are assumed to be mutually independent. The comprehensive identification problem requires estimates of parameters $D, n_u, n_y, \sigma_{e_u}$, σ_{e_y}, $\{a_i\}_{i=1}^{n_y}$, and $\{b_j\}_{j=D}^{n_u}$ to be obtained from N noisy measured samples.

Equation (1) can be recast as

$$\boldsymbol{\theta}^\intercal \mathbf{z}^*_\eta(k) = 0 \tag{3}$$

where $\boldsymbol{\theta}$ contains the model parameters and $\mathbf{z}^*_\eta(k)$ contains η lagged noise-free variables defined as follows.

$$\boldsymbol{\theta}(k) = [1 \ a_1 \ \dots \ a_\eta \ b_1 \ \dots \ b_\eta]^\intercal \tag{4a}$$

$$\mathbf{z}^*_\eta(k) = [y^*(k) \ \dots \ y^*(k-\eta) \ u^*(k) \ \dots \ u^*(k-\eta)]^\intercal \tag{4b}$$

Since η and the true values are unknown, a lagged data matrix $\mathbf{Z}_L$ is constructed using a sufficiently large lag window $L \geq \eta$, to identify the model parameters.

$$\mathbf{z}_L(k) = [y(k) \ \dots \ y(k-L) \ u(k) \ \dots \ u(k-L)]^\intercal \tag{5a}$$

$$\mathbf{Z}_L = [\mathbf{z}_L(L+1) \ \mathbf{z}_L(L+2) \ \dots \ \mathbf{z}_L(N)]^\intercal \tag{5b}$$

Equations (1) to (5) can be written in compact form as

$$\mathbf{A}_d \left(\mathbf{Z}^*_L\right)^\intercal = [\mathbf{A}_y \ \mathbf{B}_u] \left(\mathbf{Z}^*_L\right)^\intercal = \mathbf{0}_{d \times (N-L)} \tag{6a}$$

$$\text{subject to :} \quad \mathbf{Z}_L = \mathbf{Z}^*_L + \mathbf{E}_L \tag{6b}$$

where each row of the constraint matrix $\mathbf{A}_d : d \times 2(L+1)$ is a time-shifted version of the model coefficient vector $\boldsymbol{\theta}$. The use of a lag window greater than η results in $d = L - \eta + 1$ linearly independent equations relating the true lagged variables $\mathbf{z}^*_L$.

For the EIV-ARX model, the covariance matrix of the noise corrupting the lagged data vector $\mathbf{z}_L(k)$ denoted as $\boldsymbol{\Sigma}_{\mathbf{e}_L}$ is given by

$$\boldsymbol{\Sigma}_{\mathbf{e}_L} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{v}_y} & \mathbf{0} \\ \mathbf{0} & \sigma^2_{e_u} \mathbf{I}_{L+1} \end{bmatrix} \tag{7}$$

where $\boldsymbol{\Sigma}_{\mathbf{v}_y}$ is a Toeplitz matrix with elements $\gamma_v(\ell)$, which is the ACVF of $v_y(\ell)$ at lag ℓ. Given σ_{e_y} and $\{a_i\}_{i=1}^{L}$, the ACVF is estimated by solving the Yule-Walker equations

$$\gamma_v(\ell) - \sum_{i=1}^{L} a_i \gamma_v(\ell - i) = \begin{cases} \sigma^2_{e_y}, & \ell = 0 \\ 0, & \ell > 0 \end{cases} \tag{8}$$

Maurya et al. (2022) developed a method for identifying an EIV-ARX model of a SISO process based on modification of the dynamic iterative PCA approach (Maurya et al., 2018), which is capable of automatically estimating the noise variances, order and delay of the process, and ARX model coefficients. The method applies PCA to the lagged data matrix after scaling it using the Cholesky factor of corresponding noise covariance matrix $\boldsymbol{\Sigma}_{\mathbf{e}_L}$. The scaled lagged data matrix is therefore given by

$$\mathbf{Z}_{L,s} = \mathbf{Z}_L \mathbf{L}^{-1}_{\mathbf{e}_L} \tag{9}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{e}_L} = \mathbf{L}_{\mathbf{e}_L} \mathbf{L}^\intercal_{\mathbf{e}_L} \tag{10}$$

It can be shown that the smallest d singular values of the scaled data matrix $\mathbf{Z}_{L,s}$ should all be equal to unity, in the limit as N tends to infinity. This property is exploited to estimate the number of constraints (and consequently estimate the process order) by applying a hypothesis test to assess the equality of the smallest singular values. The constraint matrix $\mathbf{A}_d$ is estimated from the right singular vectors $\mathbf{V}_d$ corresponding to the d unity singular values as

$$\hat{\mathbf{A}}_d = \mathbf{V}^\intercal_d \mathbf{L}^{-1}_{\mathbf{e}_L} \tag{11}$$

In order to estimate the unknown noise variances, the above model estimation is combined with a method for estimating the noise variances within an iterative scheme. The two noise variances $\sigma^2_{e_y}$ and $\sigma^2_{e_u}$ are estimated by minimizing the following joint likelihood function of the lagged constraint residuals

$$\min_{\sigma^2_{e_y},\, \sigma^2_{e_u}} \Bigg[(N-L)\, \log |\mathbf{\Sigma}_{\mathbf{r}_L}| \; + \; \sum_{k=L+1}^{N} \mathbf{r}_L(k)^\intercal \left(\mathbf{\Sigma}_{\mathbf{r}_L}\right)^{-1} \mathbf{r}_L(k) \Bigg] \tag{12}$$

where, $\mathbf{r}_L(k) = \hat{\mathbf{A}}_d \mathbf{z}_L(k)$ and $\mathbf{\Sigma}_{\mathbf{r}_L} = \hat{\mathbf{A}}_d \hat{\mathbf{\Sigma}}_{\mathbf{e}_L} \hat{\mathbf{A}}_d^\intercal$. It may be noted from (7) and (8), that estimates of the ARX model coefficients are required for computing $\mathbf{\Sigma}_{\mathbf{e}_L}$. Due to rotational ambiguity, $\hat{\mathbf{A}}_d$, estimated using (11) is related to $\mathbf{A}_d$ via a non-singular rotation matrix $\mathbf{R}$ as $\mathbf{A}_d = \mathbf{R}\hat{\mathbf{A}}_d$. Maurya et al. (2022) have provided a formal procedure to uniquely determine $\mathbf{R}$ for obtaining the estimates of the ARX model coefficients, by exploiting the fact that the zeros and unity elements of the matrix $\mathbf{A}_d$ are known. The overall algorithm consists of three nested iterative loops. An outer loop where the number of constraints d is guessed and decremented in each iteration until the smallest d singular values corresponding to the scaled data matrix $\mathbf{Z}_{L,s}$ are equal, an inner loop in which the model and noise variances are alternatively estimated, and an innermost loop to solve the minimization problem defined by (12). The reader is referred to Maurya et al. (2022) for a detailed description of the overall modified dynamic iterative PCA algorithm for identifying an EIV-ARX model from a given sample of measurements.

3. PROPOSED METHODOLOGY

In this section, we describe the proposed recursive estimation method to identify time-varying EIV-ARX processes by extending the modified DIPCA method described in the preceding section. Recently a recursive iterative PCA method (RIPCA) was developed to identify time varying constraint models for steady state processes (Pradeep and Narasimhan, 2025). In a recursive setting, the full data matrix is not stored or available. However, the first and second order statistics of the data can be recursively updated online. Furthermore, the eigenvectors of the covariance matrix of data span the same subspace as the right singular vectors of the data matrix, which are required for estimating the constraint matrix. We propose to use a similar strategy to develop the recursive version of EIV-ARX model identification. Two modifications are made to the DIPCA algorithm described in the preceding section so that it only uses the covariance matrix of lagged measurements, rather than the lagged data matrix.

In the constraint model $\mathbf{A}_d$ estimation step, the eigenvectors of the sample covariance matrix of the scaled lagged data are computed using eigen-decomposition, instead of using the singular value decomposition (SVD) of the scaled lagged data matrix. The sample covariance matrix of scaled lagged data at any time instant k is given by

$$\mathbf{S}_{\mathbf{Z}_{L,s},k} = \frac{1}{N_k} \mathbf{L}_{\mathbf{e}_L}^{-\intercal} \mathbf{Z}_{L,k}^\intercal \mathbf{Z}_{L,k} \mathbf{L}_{\mathbf{e}_L}^{-1} \tag{13}$$

where, N_k is the number of samples till the k'th instant.

The second modification is to recast the optimization problem defined in (12) to update the noise variances $\hat{\sigma}^2_{e_y,k-1}$, and $\hat{\sigma}^2_{e_u,k-1}$, so that it also requires only the sample covariance matrix of lagged measurements rather than the data matrix. By applying the cyclic property of the trace operator in (12) (Pradeep and Narasimhan, 2025), the optimization problem can be rewritten as

$$\hat{\mathbf{\Sigma}}_{\mathbf{e}_L,k} = \arg \min_{\sigma^2_{e_y},\, \sigma^2_{e_u}} \Bigg[(N-L)\, \log \left|\mathbf{\Sigma}_{\mathbf{r}_{L,k}}\right| \; + \; (N-L)\, \mathrm{Tr}\left(\left(\mathbf{\Sigma}_{\mathbf{r}_{L,k}}\right)^{-1} \hat{\mathbf{A}}_{d,k} \mathbf{S}_{\mathbf{Z}_L,k} \hat{\mathbf{A}}_{d,k}^\intercal \right) \Bigg] \tag{14}$$

where, $\mathbf{\Sigma}_{\mathbf{r}_{L,k}} = \hat{\mathbf{A}}_{d,k} \hat{\mathbf{\Sigma}}_{\mathbf{e}_L,k-1} \hat{\mathbf{A}}_{d,k}^\intercal$. It may be noted that the above optimization problem only consists of the noise variances of input and output measurements as decision variables, since the noise covariance matrix of lagged data $\hat{\mathbf{\Sigma}}_{\mathbf{e}_L,k}$ can be computed from these noise variances using (7) and (8). Similar to DIPCA, the above two steps are iteratively solved to estimate the constraint model $\hat{\mathbf{A}}_{d,k}$, and noise variances $\hat{\sigma}^2_{e_y,k}$, $\hat{\sigma}^2_{e_u,k}$.

The sample covariance matrix of the lagged data $\mathbf{S}_{\mathbf{Z}_L,k}$ corresponding to the k'th instant can be computed in a recursive manner when new measurements are received at any time instant k. This only requires (a) the measurements corresponding to the preceding L time instances, and (b) the sample covariance matrix of lagged measurements at the preceding time instant, i.e., $\mathbf{S}_{\mathbf{Z}_L,k-1}$ to be stored. The recursive equation for computing the sample covariance matrix of lagged data is given by

$$\mathbf{S}_{\mathbf{Z}_L,k} = \frac{N_{k-1}}{N_k} \mathbf{S}_{\mathbf{Z}_L,k-1} + \frac{1}{N_k} \mathbf{z}_L(k) \mathbf{z}_L^\intercal(k) \tag{15}$$

where N_{k-1} is the number of samples obtained before the k'th instant, and $N_k = N_{k-1} + 1$.

In the above description, the proposed method is applied at every time instant to re-estimate all the model parameters. In practice, while the covariance matrix can be updated at every time instant, the model re-identification can be triggered on need basis based on monitoring the model prediction errors. This can significantly reduce the computational burden. Secondly, the covariance matrix update using (15), as a part of the baseline recursive algorithm presented here, assigns equal weight to all measurements. A forgetting factor can optionally be introduced as a hyperparameter by practitioners to give more weight to recent measurements as described by Li et al. (2000), which would result in faster convergence to the new system parameters. The proposed method, referred to as rARX-DIPCA is summarized in Algorithm 1.

4. APPLICATIONS

We now demonstrate the performance of rARX-DIPCA through two simulation studies based on two systems, drawn from Maurya et al. (2022), to illustrate its application to modeling time-varying systems in two specific engineering scenarios. In each of the simulation studies, we compute the following two metrics after each sample is received for evaluating the algorithm's performance

(1) The absolute difference between the estimated and true process orders, computed as $|\hat{\eta} - \eta|$.

Algorithm 1: Recursive adaptation of time-varying EIV-ARX processes using the rARX-DIPCA algorithm

Input: The data $\mathbf{z}_L(k)$, and lag L

Prior Information: From $(k-1)$'th instant, $\mathbf{S}_{\mathbf{Z}_L,k-1}$, and N_{k-1}, as well as the estimates of $\hat{\mathbf{a}}_L = \{\hat{a}_j\}_{j=1}^{L}$, $\hat{\sigma}_{e_y,k-1}$, $\hat{\sigma}_{e_u,k-1}$

Configuration Parameters: Max iterations $i_{max} > 0$, stopping tolerance $\varepsilon_\lambda > 0$

Output: Updated estimates of $\hat{\mathbf{a}}_{\hat{\eta}_k} = \{\hat{a}_j\}_{j=1}^{\hat{\eta}_k}$, $\hat{\mathbf{b}}_{\hat{\eta}_k} = \{\hat{b}_j\}_{j=0}^{\hat{\eta}_k}$, $\hat{\sigma}_{e_y,k}$, and $\hat{\sigma}_{e_u,k}$

Compute $N_k = N_{k-1} + 1$. Initialize $i \leftarrow 1$, $\lambda^{(0)} \leftarrow 0$, $\lambda^{(i)} \leftarrow 1$, $\hat{\sigma}^{(i)}_{e_y,k} \leftarrow \hat{\sigma}_{e_y,k-1}$, $\hat{\sigma}^{(i)}_{e_u,k} \leftarrow \hat{\sigma}_{e_u,k-1}$;

Obtain initial estimate of $\hat{\mathbf{\Sigma}}^{(i)}_{\mathbf{e}_L,k}$ using $\hat{\mathbf{a}}_L$, $\hat{\sigma}^{(i)}_{e_y,k}$, and $\hat{\sigma}^{(i)}_{e_u,k}$ based on (7) and (8);

Compute $\mathbf{S}_{\mathbf{Z}_L,k}$ using $\mathbf{S}_{\mathbf{z}_L,k-1}$ and $\mathbf{z}_L(k)$ as in (15);

while $i \leq i_{max}$ **and** $\left|\left(\lambda^{(i)} - \lambda^{(i-1)}\right)/\lambda^{(i-1)}\right| > \varepsilon_\lambda$ **do**

  Scale the covariance using (13) to obtain $\mathbf{S}^{(i)}_{\mathbf{Z}_{L,s},k}$ using $\hat{\mathbf{L}}^{(i)}_{\mathbf{e}_L,k}$, the Cholesky factor of $\hat{\mathbf{\Sigma}}^{(i)}_{\mathbf{e}_L,k}$;

  $[\hat{\mathbf{V}}\ \hat{\mathbf{D}}] = \text{eig}\left(\mathbf{S}^{(i)}_{\mathbf{Z}_{L,s},k}\right)$. Obtain $\hat{d}$ by employing hypothesis test on $\hat{\mathbf{D}}$ as detailed in Pradeep and Narasimhan (2025). The estimated process order is $\hat{\eta}_k = L - \hat{d} + 1$;

  Compute $\hat{\mathbf{A}}^{(i+1)}_{\hat{d},k} = \hat{\mathbf{V}}^{\intercal}_{\hat{d}} \times \hat{\mathbf{L}}^{(i)-1}_{\mathbf{e}_L,k}$ as in (11);

  Obtain the rotation matrix $\mathbf{R}$ by leveraging the known structure of $\mathbf{A}_d$. $\hat{\mathbf{A}}^{(i+1)}_{\hat{d},k} \leftarrow \mathbf{R} \times \hat{\mathbf{A}}^{(i+1)}_{\hat{d},k}$;

  Obtain the updated estimates of $\hat{\sigma}^{(i+1)}_{e_y,k}, \hat{\sigma}^{(i+1)}_{e_u,k}$ by solving (14) using $\mathbf{S}_{\mathbf{Z}_L,k}$, $\hat{\mathbf{A}}^{(i+1)}_{\hat{d},k}$, $\hat{\sigma}^{(i)}_{e_y,k}$, and $\hat{\sigma}^{(i)}_{e_u,k}$;

  Compute $\hat{\mathbf{\Sigma}}^{(i+1)}_{\mathbf{e}_L,k}$ using $\hat{\mathbf{A}}^{(i+1)}_{\hat{d},k}$, $\hat{\sigma}^{(i+1)}_{e_y,k}$, and $\hat{\sigma}^{(i+1)}_{e_u,k}$ based on (7) and (8);

  $\lambda^{(i-1)} \leftarrow \lambda^{(i)}$, $i \leftarrow i + 1$;

  Compute $\lambda^{(i)}$ as sum of the $\hat{d}$ smallest eigenvalues contained in $\hat{\mathbf{D}}$;

end while

Set $\hat{\sigma}_{e_y,k} \leftarrow \hat{\sigma}^{(i+1)}_{e_y,k}$, $\hat{\sigma}_{e_u,k} \leftarrow \hat{\sigma}^{(i+1)}_{e_u,k}$. Extract the coefficients $\hat{\mathbf{a}}_{\hat{\eta}_k}$, and $\hat{\mathbf{b}}_{\hat{\eta}_k}$ from $\hat{\mathbf{A}}^{(i+1)}_{\hat{d},k}$;

(2) The sum of the relative difference between the estimated and true noise standard deviations for both input and output variables, computed as

$$\left|\frac{\hat{\sigma}_{e_y,k} - \sigma_{e_y}}{\sigma_{e_y}}\right| + \left|\frac{\hat{\sigma}_{e_u,k} - \sigma_{e_u}}{\sigma_{e_u}}\right| \tag{16}$$

In addition, we also compare the estimated model coefficients obtained at the final time instant with their corresponding true values.

4.1 Tracking Sensor Degradation

We consider the following second-order LTI system

$$y^*(k) = 1.5y^*(k-1) - 0.7y^*(k-2) + u^*(k-1) + 0.5u^*(k-2) \tag{17}$$

where, we choose full band random binary signal (RBS) as input u^* with sample size 2047, and generate the noise-free output y^* using (17). We generate the noisy measurements by adding noise to the true values following the ARX structure, defined in (2). To simulate degradation in sensor precision, the noise variances $\sigma^2_{e_u}, \sigma^2_{e_y}$ are kept constant for the first 250 samples equal to 0.1 and 0.2, respectively, maintaining a signal-to-noise ratio (SNR) of 10. The noise variances are subsequently gradually increased in a quadratic fashion up to the 500'th instant until $\sigma^2_{e_u} = 0.2$, and $\sigma^2_{e_y} = 0.5$ after which they are kept constant for the remaining samples. The variation in the simulated noise variances are reported in Figs. 2 and 3.

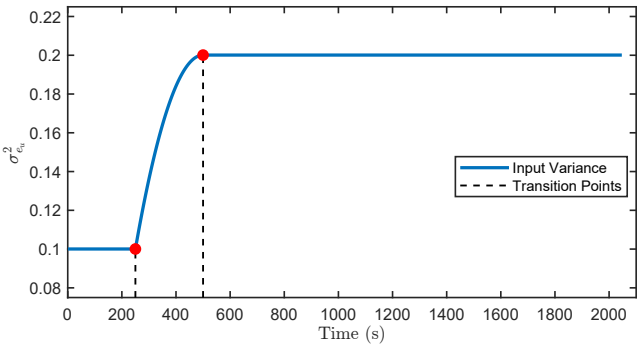


Fig. 2. Slow degradation in the input noise variance from 250'th instant up to 500'th instant.

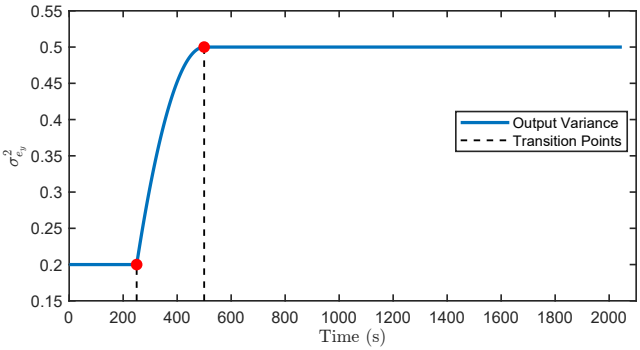


Fig. 3. Slow degradation in the output noise variance from 250'th instant up to 500'th instant.

The initial estimates required to kick start rARX-DIPCA are obtained by applying the modified dynamic iterative PCA approach discussed in Section 2.1, on the first 100 samples using a lag $L = 8$. At each subsequent time instant when we receive a new measurement sample, rARX-DIPCA is used to re-estimate the noise variances, process order, and model coefficients.

We perform 50 simulation tests and report the mean and 95% confidence intervals of the two performance metrics specified earlier, in Figs. 4 and 5, respectively.

It can be observed from Fig. 4 that the process order converges to true process order $\eta = 2$ within 160 samples. Subsequently the process order is correctly estimated even though the noise variances gradually increase from the 250'th time instant. We further observe from Fig. 5 that the error in the estimated noise variances gradually decrease until the 250'th instant, after which it initially increases due to induced changes in the sensor noise

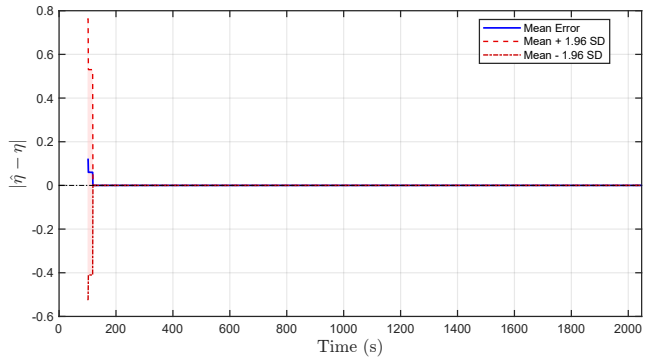


Fig. 4. Absolute difference between the estimated and true process orders.

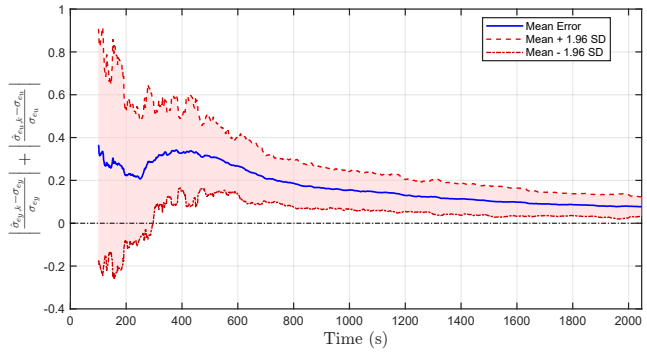


Fig. 5. Sum of the relative errors in the estimates of noise standard deviations, computed at $k = 100, \ldots, 2047$.

variances. However, as more measurements are received, the estimates of the noise variances gradually converge to their true values. The final estimates of the noise variances obtained are $\hat{\sigma}^2_{e_y} = 0.439 \pm 0.062$ and $\hat{\sigma}^2_{e_u} = 0.189 \pm 0.046$, which are close to the simulated values. The estimated model coefficients, averaged over the 50 simulation trials along with their respective 95% CI at the end of the simulation run (2047'th time instant) are given below

$$y^*(k) = \underset{\pm(0.046)}{1.489}\, y^*(k-1) - \underset{\pm(0.047)}{0.689}\, y^*(k-2) + \underset{\pm(0.034)}{0.002}\, u^*(k) + \underset{\pm(0.050)}{1.002}\, u^*(k-1) + \underset{\pm(0.075)}{0.514}\, u^*(k-2) \quad (18)$$

The results indicate that rARX-DIPCA is able to obtain unbiased estimates of the noise variances, and the model parameters even at a low SNR. Although the modified DIPCA method is developed under the assumption that the noise variances is constant over all samples, rARX-DIPCA is able to track sensor degradation as new samples are received, due to re-estimating the noise variances at every time instant.

4.2 Tracking Changes in Operating Conditions

For this case study, we consider a second-order system with third-order input dynamics as described in the following

$$y^*(k) = 1.1y^*(k-1) - 0.7y^*(k-2) + u^*(k-2) + 0.5u^*(k-3) \quad (19)$$

A nonlinear process can be modeled using an approximate linear model with time varying model coefficients to capture the process behavior induced by changes in operating conditions. In order to assess the capability of the proposed rARX-DIPCA method for such an application, in this simulation study the model coefficients are varied. We first choose full band RBS as input u^* with sample size 6095, and true values of y^* are generated using (19) for first 400 samples, after which we vary the coefficients in quadratic fashion till 650'th instant, yielding the following updated system for the remaining instances

$$y^*(k) = y^*(k-1) - 0.6y^*(k-2) + 1.2u^*(k-2) + 0.7u^*(k-3) \quad (20)$$

Noise is added to the true values following the ARX model to generate the measurements, using noise variances $\sigma^2_{e_u} = 0.1$ and $\sigma^2_{e_y} = 0.15$, which correspond to SNR 10. Fifty simulation trials are conducted with different noise realizations, each with 6095 measurement observations.

We obtain the lagged sample covariance, along with the estimates of the input-output noise variances, process order, and model coefficients by applying the modified Dynamic IPCA considering the first 100 samples, to initialize the rARX-EIV method. At each subsequent sampling instant, upon receiving new measurements, the proposed recursive identification algorithm is applied to re-estimate all the model parameters. The performance metrics described in preceding sub-section are again used to assess the performance of the method.

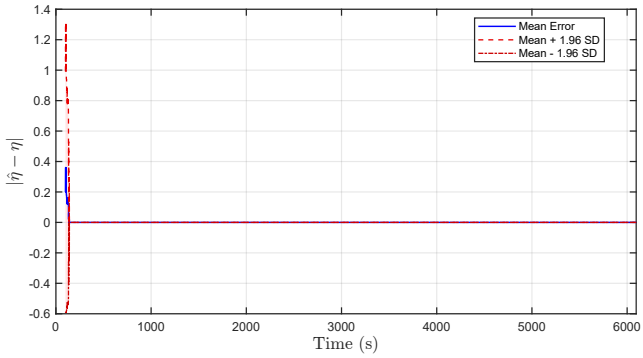


Fig. 6. Absolute difference between the estimated and true process orders.

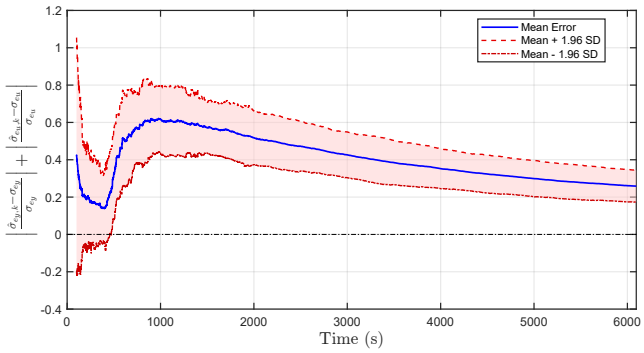


Fig. 7. Sum of the relative errors in the estimates of noise standard deviations, computed at $k = 100, \ldots, 6095$.

From Fig. 6, we observe that the estimate of the process order converges to the true value $\eta = 3$ within 200 samples. Fig. 7 shows that the error in the estimated noise variances decreases before the changes in the model coefficients are introduced. Although the noise variances are kept constant, the error in the estimated noise variances increases for a short period after the 650'th time instant when changes in the model coefficients are introduced.

This is due to the fact that the linear model identification is developed under the assumption that the entire sample of measurements used for model identification is generated from a constant time invariant linear process. Despite this, the estimated noise variances start converging to their respective true values as additional measurements corresponding to the modified process are received. The noise variance estimates obtained at the end of the simulation are $\hat{\sigma}^2_{e_u} = 0.089 \pm 0.014$, and $\hat{\sigma}^2_{e_y} = 0.173 \pm 0.027$, which indicate that they are unbiased. Since accurate estimation of the noise variances is the primary requirement for unbiased PCA-based parameter estimation, the successful tracking of these variances, as illustrated in Fig. 7, serves as strong indirect evidence of the algorithm's parameter-tracking capabilities throughout the transition period.

Table 1. Comparing the model coefficient estimates of the systems defined in (19) and (20).

	Estimation before 401'th instance		Estimation over 6095 observations	
Corresponding Variable	True values	Estimated values	True values	Estimated values
$y^*(k)$	1.0	1.0	1.0	1.0
$y^*(k-1)$	1.1	1.117 $\pm(0.038)$	1.0	1.023 $\pm(0.024)$
$y^*(k-2)$	-0.7	-0.724 $\pm(0.050)$	-0.6	-0.628 $\pm(0.028)$
$y^*(k-3)$	0.0	0.017 $\pm(0.032)$	0.0	0.013 $\pm(0.019)$
$u^*(k)$	0.0	0.008 $\pm(0.025)$	0.0	0.002 $\pm(0.016)$
$u^*(k-1)$	0.0	0.011 $\pm(0.029)$	0.0	-0.0004 $\pm(0.012)$
$u^*(k-2)$	1.0	1.007 $\pm(0.025)$	1.2	1.173 $\pm(0.011)$
$u^*(k-3)$	0.5	0.487 $\pm(0.051)$	0.7	0.659 $\pm(0.033)$

The model coefficient estimates, obtained before the 401'th instant (before the changes in the model coefficients are introduced), and at the final sampling instant (i.e., $k = 6095$) are reported in table 1 along with their respective 95% confidence intervals. The results indicate that rARX-DIPCA has obtained unbiased estimates of the initial process model coefficients as defined in (19), and has also adapted to the time-varying process by providing unbiased estimates of the final process model coefficients given by (20). However, the estimated coefficients at the 6095'th instant, corresponding to the variables $u^*(k-2)$ and $u^*(k-3)$ have very small bias, which, as per the observed convergence trend, is expected to vanish as more measurements are received.

5. CONCLUSION

This paper has introduced rARX-DIPCA, a recursive identification algorithm for time-varying EIV-ARX processes without the need for complete historical data storage. Simulation studies show that the proposed method is able to adapt to changes in the noise variances due to sensor degradation, and model coefficient changes resulting from changing operating conditions. Future research directions potentially include extending the framework to handle multi-variable MIMO-ARX systems in order to address more complex industrial processes with coupled variables, incorporating time-varying delay tracking capabilities to handle systems with transport phenomena or communication latency. Theoretical convergence analysis under persistent parameter variations would strengthen the method's foundation for safety-critical applications.

REFERENCES

Cao, P., Luo, X., and Song, X. (2018). Modeling and identification for soft sensor systems based on the separation of multi-dynamic and static characteristics. *Chinese Journal of Chemical Engineering*, 26(1), 137–143.

Diversi, R., Guidorzi, R., and Soverini, U. (2010). Identification of ARX and ARARX models in the presence of input and output noises. *European Journal of Control*, 16(3), 242–255.

Fernando, K. and Nicholson, H. (1985). Identification of linear systems with input and output noise: the Koopmans-Levin method. *IEE Proceedings D (Control Theory and Applications)*, 132, 30–36.

Ikenoue, M., Kanae, S., Yang, Z.J., and Wada, K. (2005). Identification of noisy input-output system using bias-compensated least-squares method. *IFAC Proceedings Volumes*, 38(1), 803–808. 16th IFAC World Congress.

Li, W., Yue, H., Valle-Cervantes, S., and Qin, S. (2000). Recursive PCA for adaptive process monitoring. *Journal of Process Control*, 10(5), 471–486.

Maurya, D., Tangirala, A.K., and Narasimhan, S. (2018). Identification of errors-in-variables models using dynamic iterative principal component analysis. *Industrial & Engineering Chemistry Research*, 57(35), 11939–11954.

Maurya, D., Tangirala, A.K., and Narasimhan, S. (2022). Identification of errors-in-variables ARX models using modified dynamic iterative PCA. *Journal of the Franklin Institute*, 359(13), 7069–7090.

Picci, G., Falconi, L., Ferrante, A., and Zorzi, M. (2023). Hidden factor estimation in dynamic generalized factor analysis models. *Automatica*, 149, 110834.

Pradeep, M. and Narasimhan, S. (2025). Recursive iterative principal component analysis. *Computers & Chemical Engineering*, 201, 109170.

Prakash, J., Narasimhan, S., and Patwardhan, S.C. (2005). Integrating model based fault diagnosis with model predictive control. *Industrial & Engineering Chemistry Research*, 44(12), 4344–4360.

Prakash, O. and Huang, B. (2025). Real-time update of data-driven reduced and full order models with applications. *Computers & Chemical Engineering*, 194, 108923.

Söderström, T. and Stoica, P. (2002). Instrumental variable methods for system identification. *Circuits, Systems and Signal Processing*, 21(1), 1–9.

Zhang, J., Guo, F., Hao, K., Chen, L., and Huang, B. (2022). Identification of errors-in-variables ARX model with time varying time delay. *Journal of Process Control*, 115, 134–144.

Zhao, Y., Fatehi, A., and Huang, B. (2018). Robust estimation of ARX models with time varying time delays using Variational Bayesian approach. *IEEE Transactions on Cybernetics*, 48(2), 532–542.

Zheng, W.X. (2002). A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, 47(7), 1142–1147.