Intelligent State Estimation for Online Optimizing Control of a Reactor System exhibiting Input Multiplicity

Giriraj Bagla [*] **Jayaram Valluru** [**] **Anjali P. Deshpande** [***]
Sachin C. Patwardhan [a *]

[*] *Dept. of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai, India ([a] e-mail: sachinp@iitb.ac.in)*
[**] *Dept. of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada*
[***] *Sr. Member IEEE Mumbai, India*

Abstract: Combinations of real-time optimization (RTO) and model predictive control (MPC) have been widely employed in the process industry for tracking the economic optimum in the face of drifting disturbances and parameters. Online update of model parameters is a critical step in the implementation of RTO. In this work, an intelligent state and parameter estimation approach is developed by combining a fault diagnosis approach with a moving window-based online state and parameter estimator. The estimation of unmeasured disturbance(s)/ parameter(s)/ sensor bias(es) is carried out only when required and triggered by the fault identification scheme. Thus, the subset of parameters/faults that are being estimated online can change with time. This can avoid difficulties that arise due to the observability condition. The intelligent state and parameter estimator is further combined with an online optimizing control scheme consisting of integrated frequent RTO and adaptive MPC. The integrated scheme has embedded intelligence to auto-correct models used for estimation, control, and optimization and to decide whether the detected changes require the invocation of RTO. The efficacy of the proposed scheme is investigated using a benchmark CSTR system that exhibits input multiplicity behavior. The optimum operating point of this system is sensitive to mean shifts in unmeasured disturbances or system parameters. The proposed approach successfully isolates the parameter/ unmeasured disturbance/ sensor bias that has undergone abrupt change and tracks the shifting economic optimum without significant delays. Thus, the proposed integrated approach has the ability to handle normal, off-normal, and abnormal operating envelopes of the system.

Keywords: Real-Time Optimization, Nonlinear Model Predictive Control, Simultaneous State and Parameter Estimation, Sensor Bias, Observer-Based Fault Diagnosis

1. INTRODUCTION

Combinations of real time optimization (RTO) and model predictive control (MPC) have been widely employed in the process industry for tracking economic optimum of key processing units in the face of drifting disturbances and parameters (Engell (2007); Darby et al. (2011)). The conventional approach is to employ a steady state mechanistic model for carrying out RTO. The model parameters are updated once every few hours and the model is then used to locate the economic optimum. The MPC layer is entrusted with the task of moving the system to the new optimum. This conventional approach results in long wait time before the new economic optimum can be reached. Thus, many frequent RTO schemes have been proposed in the literature with the aim of reducing the wait time (Engell (2007); Darby et al. (2011)). Maintenance of the model used for RTO using transient data is the key step in the frequent RTO schemes. Updating parameters of a dynamic model online using transient data and using the steady state version of this model for carrying out frequent RTO has been proposed recently. Krishnamoorthy et al. (2018) and Valluru and Patwardhan (2019) propose to carry out simultaneous state and parameter estimation using extended Kalman filter (EKF) for updating the model used in the RTO layer. Thus, the model update is carried out on-line as and when the parameters or disturbances change. This approach eliminates the need to wait for system to reach a steady state before the model update step is carried out.

Update of the steady state mechanistic model using recent operating data is a critical step in implementation of any RTO scheme. Darby et al. (2011) observe that "*the model accuracy requirements are higher for RTO than MPC*" since the model is used for locating the economic optimum. They proceed to state that "*the model update step is non-trivial and requires process experience and engineering judgement. The starting point is to consider which parameters/adjustments are affected by unmeasured disturbances and which subset of parameters/adjustments ensure a consistent model. Parameters that change on a slower time scale may not be appropriate to update every execution* *in order to minimize the influence of measurement noise and unmeasured disturbances.*" Systematic errors in the measurements (or sensor biases) is another issue that needs to be dealt while carrying out the model update step (Engell (2007)). Thus, the model used for RTO layer is relatively harder to maintain and requires a higher level of expertise (Darby et al. (2011)).

Mismatch between the models used for RTO and MPC layers is another major issue that needs to be addressed in a conventional RTO scheme. Typically a linear black box model is used to formulate the MPC scheme. Thus, the models used for RTO and MPC are significantly different and this can result in difficulties in realizing RTO recommendations (Engell (2007)). Recently, Valluru and Patwardhan (2019) have developed a novel online optimizing control scheme which employs a single dynamic mechanistic model for implementing RTO and adaptive NMPC. Estimates of the drifting unmeasured disturbances/parameters generated using simultaneous state and parameter estimation schemes are used to update the steady state model used for frequent RTO as well as the dynamic model used for predictions in the NMPC scheme. This approach eliminates difficulties that arise due to model mismatch between RTO and MPC layer. However, they proposes to update only a fixed subset of parameters/ unmeasured disturbances simultaneously with the states. In practice, some parameters, unmeasured disturbances or soft faults (such as sensor or actuator biases) that lie outside this chosen fixed subset can also change. Since a fixed subset of parameters is updated all the time, such a scenario leads to a model plant mismatch that can deteriorate the performance of the integrated online optimizing control scheme.

The conventional approach to simultaneous estimation of states and parameter/ unmeasured disturbances/ soft faults is to assume the random walk model for representing parameter/ disturbance/ fault dynamics. The parameter/ disturbance/ fault variations are treated as additional states and estimated simultaneously with the system states (Venkatasubramanian et al. (2003) ; Patwardhan et al. (2012)). When the system dynamics are augmented with the random walk model, observability or estimability of these additional states needs to be considered. For linear systems, the application of Hautus lemma reveals that the number of parameters that can be estimated simultaneously with the states cannot exceed the number of measurements (Muske and Badgwell (2002)). This restriction may not strictly hold for a nonlinear system. However, all parameters that are susceptible to change cannot be estimated as additional states with a reasonable accuracy Liu et al. (2021).

Automating the task of finding the "active subset of parameters/ disturbances" that need to be adjusted and isolating or accommodating biased sensors while carrying out the model update can alleviate difficulties associated with the model maintenance, and, in turn, maintenance of the integrated RTO and adaptive NMPC scheme. The problem of isolating root causes of an abnormal behavior from the transient data using a dynamic model has been well investigated in the fault diagnosis literature (Venkatasubramanian et al. (2003)). Thus, a possible approach to achieve automated selection of active set of parameters is to employ an observer based fault diagnosis approach to dynamically identify the subset of actively changing parameters/ disturbances/ biases in a moving time window and update parameters/disturbances only from this subset. Deshpande et al. (2009) have developed a nonlinear state estimation strategy with embedded intelligence to diagnose the root causes of the plant model mismatch by analyzing the innovation sequence generated by EKF. The subset of active faults is isolated the model is auto-corrected on-line so as to accommodate the isolated faults. To carry out the task of fault diagnosis, a nonlinear version of the generalized likelihood ratio (GLR) based fault diagnosis and identification (FDI) scheme (NLGLR) has been developed. Their approach can deal with sequential as well as simultaneous occurrences of multiple parameter changes, soft faults and sensor/ actuator failures. Thus, the subset of parameters that are being estimated online is changed with time and this can avoid difficulties that can arise due to the observability condition.

In this work, we propose to combine the fault diagnosis based state estimation approach developed by Deshpande et al. (2009) with the integrated RTO and adaptive NMPC scheme developed Valluru et al. (2017) with the aim embedding intelligence to find the active subset of changing parameters. It is assumed that the parameters/ disturbances/ faults change infrequently and at a significantly slower rate than the system dynamics. The NLGLR approach is used to isolate a changing parameter/ disturbance and subsequently its magnitude is estimated using the moving window estimator developed by Valluru et al. (2017). If the estimates saturate, then the parameter estimation is stopped and the saturated estimate is used for state estimation subsequently. Thus, estimation of parameter(s) is carried out only when required and triggered by the fault identification scheme. The resulting intelligent state estimator is then combined with integrated frequent RTO and NMPC scheme proposed by Valluru and Patwardhan (2019). Thus, the proposed integrated scheme has embedded intelligence (i) to correct the state estimator, the prediction model used in NMPC and the steady state model used in frequent RTO using the fault location and magnitude estimates and (ii) to decide whether the isolated changes require invocation of RTO. The efficacy of the proposed integrated frequent RTO and adaptive NMPC scheme with intelligent moving window estimator is demonstrated by conducting simulation studies on a benchmark CSTR system, which exhibits input multiplicity and change in the sign of steady state.

This paper is organized in four sections. Details of the proposed Integrated scheme are presented in the Section 2. The third section presents the results of the simulation case study. The conclusions reached through analysis of the simulation results are discussed in the last section.

2. ONLINE OPTIMIZING CONTROL INTEGRATED WITH FDI

The proposed scheme has four components: (i) Fault diagnosis using NLGLR, (ii) state and parameter estimator, (iii) real time optimization, and (iv) nonlinear model predictive control. A schematic diagram of the proposed online optimizing control scheme integrated with fault

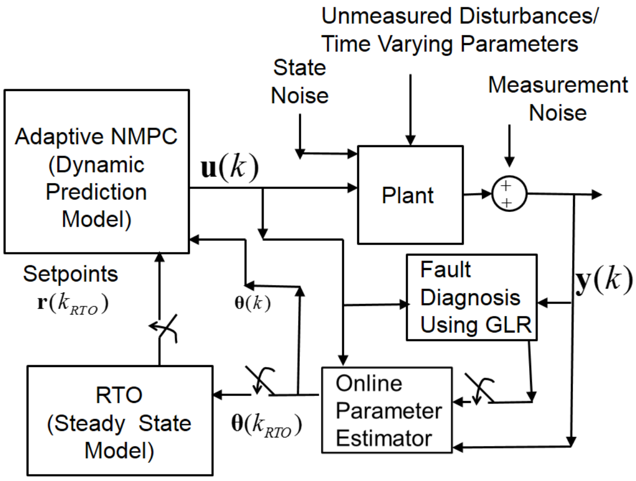

Fig. 1. Schematic diagram of proposed integrated approach

diagnosis and identification scheme is shown in Figure (1). These components and their interactions are discussed in this section.

2.1 Dynamic Process Model

Dynamics of the system *under consideration* is represented by a set of nonlinear ODEs as follows

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) \quad (1)$$

$$\mathbf{y}_T(t) = \mathbf{g}(\mathbf{x}(t))$$

where, $\mathbf{x} \in \mathbb{R}^n$ represents state variables, $\mathbf{u} \in \mathbb{R}^m$ represents manipulated inputs, $\boldsymbol{\theta} \in \mathbb{R}^p$ represents the set of model parameters/ unmeasured disturbances, and $\mathbf{y}_T \in \mathbb{R}^r$ represents the true values of the measured outputs. The operators $\mathbf{f}(.)$ and $\mathbf{g}(.)$ have dimension of $(n \times 1)$ and $(r \times 1)$, respectively, and are assumed to be known. It is assumed that the variation of parameters ($\boldsymbol{\theta}$) occur at a significantly slower rate or at a lower frequency than the rates at which states and manipulated inputs change. It is further assumed that these variations occur as infrequent abrupt jumps from their nominal values. For the purpose of carrying out state estimation and NMPC, the set of ODEs represented by eq. (1) are discretized under the assumption that the manipulated inputs are piecewise constant, i.e. $\mathbf{u}(t) = \mathbf{u}_k$ for $kT_s \leq t < (k+1)T_s$ where T_s represents the sampling interval. The discrete form of the model is represented as follows

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta}) \quad (2)$$

$$\mathbf{F}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta}) \equiv \mathbf{x}_k + \int_{kT_s}^{(k+1)T_s} \mathbf{f}(\mathbf{x}(\tau), \mathbf{u}_k, \boldsymbol{\theta}) d\tau$$

The unknown inputs influencing the state dynamics are often modelled as additive white noise in the state dynamics Valluru et al. (2017). Thus, (2) is further modified as

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta}) + \mathbf{w}_k \quad (3)$$

where $\mathbf{w}_k \in \mathbb{R}^n$ and $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}_{n\times1}, \mathbf{Q})$. Also, the measurements are assumed to be corrupted with zero mean white noise

$$\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k) + \mathbf{v}_k \quad (4)$$

where $\mathbf{v}_k \in \mathbb{R}^r$ and $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}_{r\times1}, \mathbf{R})$. This model is used for developing state estimation, fault diagnosis, NMPC and RTO components.

2.2 State Estimation under Fault Free Conditions

Under the fault free conditions, $\boldsymbol{\theta}$ is assumed to be at some nominal value, say $\bar{\boldsymbol{\theta}}$. Under this assumption, states estimation is carried out using EKF

- **Prediction step:**

$$\widehat{\mathbf{x}}_{k|k-1} = \mathbf{F}[\widehat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_{k-1}, \bar{\boldsymbol{\theta}}] \quad (5)$$

$$\mathbf{P}_{k|k-1} = \boldsymbol{\Phi}_k \mathbf{P}_{k-1|k-1} \boldsymbol{\Phi}_k^T + \mathbf{Q} \quad (6)$$

$$\boldsymbol{\Phi}_k = \exp\left[\mathbf{A}_k T_s\right] \quad ; \quad \mathbf{A}_k = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_{(.)} \quad (7)$$

where $(.) \equiv (\widehat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_{k-1}, \bar{\boldsymbol{\theta}})$.

- **Kalman Update Step:**

$$\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{L}_k \mathbf{e}_k \quad (8)$$

$$\mathbf{C}_k = \left[\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\right]_{(\widehat{\mathbf{x}}_{k|k-1})} \quad (9)$$

$$\mathbf{L}_k = \mathbf{P}_{k|k-1} \mathbf{C}_k^T \left[\mathbf{V}_k\right]^{-1} \quad (10)$$

$$\mathbf{e}_k = \mathbf{y}_k - \mathbf{g}(\widehat{\mathbf{x}}_{k|k-1}) \quad (11)$$

$$\mathbf{V}_k = \mathbf{C}_k \mathbf{P}_{k|k-1} \mathbf{C}_k^T + \mathbf{R} \quad (12)$$

$$\mathbf{P}_{k|k} = \left[\mathbf{I} - \mathbf{L}_k \mathbf{C}_k\right] \mathbf{P}_{k|k-1} \quad (13)$$

Under the normal operating conditions, it is assumed that the innovation sequence has Gaussian distribution with zero mean and covariance $\mathbf{V}_k$, i.e. $\mathbf{e}_k \sim \mathcal{N}\left(\bar{\mathbf{0}}, \mathbf{V}_k\right)$.

2.3 Fault Detection

It is assumed that only one fault occurs at a time instant and there is a sufficient time gap between the occurrence of two faults. In this work, two faults can occur sequentially in time with a sufficient time gap between them. Thus, an abrupt change in i^{th} parameter occurring at instant k_{0i} is represented as follows

$$\boldsymbol{\theta}_k = \bar{\boldsymbol{\theta}} + \Delta\theta_i \boldsymbol{\xi}^{(i)} \sigma(k - k_{0i}) \quad (14)$$

where, $\bar{\boldsymbol{\theta}}$ represents nominal value of the parameter, $\Delta\theta_i$ represents value of step change in i^{th} parameter, $\boldsymbol{\xi}^{(i)}$ is fault location vector with i^{th} element equal to one and all other elements are zero and $\sigma(k-k_{0i})$ is unit step function defined as follows:

$$\sigma(k - k_{0i}) = \left\{ \begin{array}{l} 0 \text{ if } k < k_{0i} \\ 1 \text{ if } k \geq k_{0i} \end{array} \right\} \quad (15)$$

where i can take a value from set $\{1, 2, .., p\}$. A measurement fault in j^{th} sensor occurring at instant k_{0j} is represented as follows:

$$\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k) + \mathbf{v}_k + \Delta b_j \boldsymbol{\xi}^{(j)} \sigma(k - k_{0j}) \quad (16)$$

where, Δb_j is the fault magnitude in j^{th} sensor where j can take a value from set $\{1, 2, .., r\}$.

When an abrupt change occurs in a parameter or a fault occurs in measurement, the innovation sequence does not remain zero mean. A fault detection test (FDT) is applied at every sampling instant to detect the occurrence of a fault. The test statistic for FDT is chosen as (Deshpande et al. (2009))

$$\boldsymbol{\epsilon}_k = \mathbf{e}_k^T \left[\mathbf{V}_k\right]^{-1} \mathbf{e}_k \tag{17}$$

$\boldsymbol{\epsilon}_k$ follows a chi-square distribution with r degrees of freedom, which is used to determine the threshold, γ_d, for fault detection by choosing a suitable level of significance, say α_d. Once the null hypothesis (i.e., there is no change) is rejected at instant $k \geq k_{0i}$, we set the detection time, t_d, as $t_d = k$. This indicates a possibility of occurrence of a fault. The fault is further confirmed using a set of innovations over window $k \in [t_d, t_d + N_g - 1]$, where, N_g is NLGLR window length. Test statistic for fault confirmation is as follows:

$$\boldsymbol{\epsilon}_k(t_d, N_g) = \sum_{k=t_d}^{t_d+N_g-1} \mathbf{e}_k^T \left[\mathbf{V}_k\right]^{-1} \mathbf{e}_k \tag{18}$$

Test statistic given by eq. (18) also follows chi-square distribution with $r \times N_g$ degrees of freedom. If the value of $\epsilon_k(t_d, N_d)$ crosses the threshold value for the chosen level of significance, α_c, then occurrence of a fault is confirmed. Here, we select α_c smaller than α_d to reduce the risk of false alarms. The NLGLR window length N_g is treated as a tuning parameter. If we choose smaller N_g then the risk of detecting false alarms increases. However, choosing large N_g will give delay in estimation and system will be running in degraded mode. Once the occurrence of a fault is confirmed, then the nonlinear GLR scheme is used to isolate the fault.

2.4 Fault Identification using Nonlinear GLR

A version of NLGLR method for nonlinear system developed by Deshpande et al. (2009) is summarized here. In this scheme, a separate model of the form eq. (14, 16) is developed for every hypothesized fault and EKF is applied on each model in the time interval $[t_d, t_d + N_g - 1]$ assuming fault has occurred at time t_d. Optimization problem is posted to find out the best fault model, which gives the minimum value of negative of the log-likelihood function. A sample formulation under the assumption that an abrupt change has occurred in i^{th} parameter is given as:

$$\Delta\widehat{\theta}_{i,0} = \arg \min_{\Delta\theta_i} J\left(\Delta\theta_i\right) \tag{19}$$

$$J\left(\Delta\theta_i\right) = \left[\sum_{j=t_d}^{t_d+N_g-1} \left[\mathbf{e}_j^T \mathbf{V}_j^{-1} \mathbf{e}_j + \log\left|\mathbf{V}_j\right|\right]\right]$$

subject to:

$$\widetilde{\mathbf{x}}_{j|j-1} = \mathbf{F}[\widetilde{\mathbf{x}}_{j-1|j-1}, \mathbf{u}_{j-1}, \bar{\boldsymbol{\theta}} + \Delta\theta_i \boldsymbol{\xi}^{(i)}] \tag{20}$$

$$\widetilde{\mathbf{P}}_{j|j-1} = \boldsymbol{\Phi}_j \widetilde{\mathbf{P}}_{j-1|j-1} \boldsymbol{\Phi}_j^T + \mathbf{Q} \tag{21}$$

$$\mathbf{e}_j = \mathbf{y}_j - \mathbf{g}(\widetilde{\mathbf{x}}_{j|j-1}) \tag{22}$$

$$\mathbf{V}_j = \mathbf{C}_j \widetilde{\mathbf{P}}_{j|j-1} \mathbf{C}_j^T + \mathbf{R} \tag{23}$$

$$\mathbf{L}_j = \widetilde{\mathbf{P}}_{j|j-1} \mathbf{C}_j^T \left[\mathbf{V}_j\right]^{-1} \tag{24}$$

$$\widetilde{\mathbf{x}}_{j|j} = \widetilde{\mathbf{x}}_{j|j-1} + \mathbf{L}_j \mathbf{e}_j \tag{25}$$

$$\widetilde{\mathbf{P}}_{j|j} = \left[\mathbf{I} - \mathbf{L}_j \mathbf{C}_j\right] \widetilde{\mathbf{P}}_{j|j-1} \tag{26}$$

$$\Delta\theta_i \in \left[\Delta\theta_{i,\min}, \Delta\theta_{i,\max}\right] \tag{27}$$

$$j = t_d, t_d + 1, ..., t_d + N_g - 1$$

The optimization problem is initialized with

$$\widetilde{\mathbf{x}}_{t_d-1|t_d-1} = \widehat{\mathbf{x}}_{t_d-1|t_d-1} \text{ and } \widetilde{\mathbf{P}}_{t_d-1|t_d-1} = \mathbf{P}_{t_d-1|t_d-1}$$

A similar optimization problem is formulated for a bias in j^{th} sensor. Let $\mathcal{F} \equiv \{f_l : l = 1, 2, .., p + r\}$ denote a set of all hypothesized faults where

$$\mathcal{F} = \left\{\left\{\Delta\theta_i : i = 1, 2, ..., p\right\}, \left\{\Delta b_j : j = 1, 2, ..., r\right\}\right\}$$

An optimization problem is solved for each hypothesized fault, i.e., for $i = 1, 2, ...p + r$. The fault is isolated by finding the minimum value of $J\left(f_i\right)$ over $i = 1, 2, ...p + r$, i.e.

$$f = \arg \min_{f_i \in \mathcal{F}} J\left(f_i\right)$$

and the corresponding optimum $\Delta\widehat{\theta}_{f,i}$ or $\Delta\widehat{b}_{f,j}$ is treated as the initial estimate of fault that has occurred at time instant t_d.

Remark 1. In this work, it is assumed that only a single fault occurs at a time instant. In practice, however, multiple faults can occur simultaneously at the same instant. Deshpande et al. (2009) have shown that of Akaike Information Criterion (AIC) can be used for fault isolation when occurrence of multiple simultaneous faults are hypothesized together with single faults. Thus, the proposed scheme can be modified to accommodate occurrences of simultaneous faults using AIC.

2.5 Fault Magnitude Refinement using Moving Window EKF

NLGLR method gives an initial estimate of the fault. Further, to improve that estimate a window based maximum likelihood parameter estimation sachem developed by Valluru et al. (2017) is applied. This scheme simultaneously gives estimates of states also. After fault identification, at current time instant k, we consider recent past input-output data in time interval $\tau_k \equiv [k - N_{ml}, k]$ where, N_{ml} is length of estimation window. Assuming a parameter fault has occurred, a sample formulation of the moving window state and parameter estimation problem is as follows (Valluru et al. (2017)):

$$\Delta\widehat{\theta}_{f,k} = \arg \min_{\Delta\theta_f} \left[\sum_{j=k-N_{ml}+1}^{k} \left[\mathbf{e}_j^T \mathbf{V}_j^{-1} \mathbf{e}_j + \log\left|\mathbf{V}_j\right|\right]\right] \tag{28}$$

subject to

$$\widehat{\mathbf{x}}_{j|j-1} = \mathbf{F}[\widehat{\mathbf{x}}_{j-1|j-1}, \mathbf{u}_{j-1}, \bar{\boldsymbol{\theta}} + \Delta\theta_f \boldsymbol{\xi}^{(f)}] \tag{29}$$

$$\mathbf{P}_{j|j-1} = \boldsymbol{\Phi}_j \mathbf{P}_{j-1|j-1} \boldsymbol{\Phi}_j^T + \mathbf{Q} \tag{30}$$

$$\mathbf{e}_j = \mathbf{y}_j - \mathbf{g}(\widehat{\mathbf{x}}_{j|j-1}) \tag{31}$$

$$\mathbf{V}_j = \mathbf{C}_j \mathbf{P}_{j|j-1} \mathbf{C}_j^T + \mathbf{R} \tag{32}$$

$$\mathbf{L}_j = \mathbf{P}_{j|j-1} \mathbf{C}_j^T \left[\mathbf{V}_j\right]^{-1} \tag{33}$$

$$\widehat{\mathbf{x}}_{j|j} = \widehat{\mathbf{x}}_{j|j-1} + \mathbf{L}_j \mathbf{e}_j \tag{34}$$

$$\mathbf{P}_{j|j} = \left[\mathbf{I} - \mathbf{L}_j \mathbf{C}_j\right] \mathbf{P}_{j|j-1} \tag{35}$$

$$\Delta\theta_f \in \left[\Delta\theta_{f,\min}, \Delta\theta_{f,\max}\right] \tag{36}$$

$$j = k - N_{ml} + 1, k - N_{ml} + 2, ..., k$$

When a bias in sensor is detected, the refinement problem is formulated w.r.t. Δb_f with the measurement model in the optimization problem expressed as follows

$$\mathbf{e}_j = \mathbf{y}_j - \left\{\mathbf{g}(\widehat{\mathbf{x}}_{j|j-1}) + \Delta b_f\right\} \tag{37}$$

while the state predictions are carried out as

$$\widehat{\mathbf{x}}_{j|j-1} = \mathbf{F}[\widehat{\mathbf{x}}_{j-1|j-1}, \mathbf{u}_{j-1}, \bar{\boldsymbol{\theta}}] \tag{38}$$

Once we observe that estimates of $\Delta\theta_f$ (or Δb_f) obtained from the moving window scheme are not changing significantly, we stop estimating the fault magnitude and revert back to the normal EKF given by eq. (5-13). A hypothesis test is carried out to decide when to stop the estimation of the fault magnitude. To check whether the estimate has saturated, we consider two sets of recent data of estimated parameter fault in the time window $[k-2q+1, k]$ i.e.,

$$\Theta_1 = \{\Delta\widehat{\theta}_{f,k-2q+1}.....\Delta\widehat{\theta}_{f,k-q}\}$$

and

$$\Theta_2 = \{\Delta\widehat{\theta}_{f,k-q+1}.....\Delta\widehat{\theta}_{f,k}\}$$

and we construct a hypothesis test as follows:

$$\mathbf{H}_0 : \mu_1 = \mu_2 \text{ or } \mathbf{H}_1 : \mu_1 \neq \mu_2 \quad (39)$$

where μ_1 and μ_2 are expected values of data sets Θ_1 and Θ_2, respectively. A t-test is statistics with $2q-2$ degrees of freedom is given as

$$T = (\bar{\Theta}_p - \bar{\Theta}_q)/(\sqrt{(s^2 \times (2/q))}) \quad (40)$$

where $s^2 = (s_1^2 + s_2^2)/2$, s_1^2 and s_2^2 are sample variance of Θ_1 and Θ_2 respectively. Given a significance level α_t. The null hypothesis is accepted if $|T| < t_{\alpha_t/2,2q-2}$. If $\mathbf{H}_0$ is accepted at instant, say $k = t_f$, then we stop estimation of $\Delta\theta_f$ and change $\bar{\boldsymbol{\theta}}$ appearing in eq. (5) under the *normal conditions* to $\bar{\boldsymbol{\theta}}+\Delta\theta_{f,t_f}\boldsymbol{\xi}^{(f)}$ while carrying out state estimation under the new 'fault free' condition. If it is a fault in say f^{th} sensor, then "corrected measurements", i.e. $\mathbf{y}_k - b_{f,t_f}\boldsymbol{\xi}^{(f)}$, are sent to the state estimator subsequent to stopping the magnitude refinement. This allows us to detect and identify a fault that may occur in another parameter/sensor subsequently.

2.6 Real Time Optimization

Once NLGLR identifies change in a parameter, there is a need to find out new optimum set-points. Thus, during time interval $k \in [t_d, +N_g, t_f]$, RTO block is invoked; it is invoked over a shifting window of N_{rto} samples. Let $\Delta\widehat{\theta}_{f,k}$ for $k = k_{RTO}$ represent the parameter estimate at the instant when RTO is invoked. Then, the following optimization problem is formulated at RTO layer as:

$$\min_{\mathbf{u_s}} J_E(\mathbf{y}_c, \mathbf{u_s}) \quad (41)$$

subject to

$$\mathbf{f}(\mathbf{x}_s, \mathbf{u}_s, \bar{\boldsymbol{\theta}}+\Delta\widehat{\theta}_{f,k_{RTO}}\boldsymbol{\xi}^{(f)}) = \bar{\mathbf{0}}_{n\times 1} \quad (42)$$

$$\mathbf{y}_c = \mathbf{g}_c(\mathbf{x}_s)$$

$$\mathbf{u_s}_{\min} \leq \mathbf{u_s} \leq \mathbf{u_s}_{\max} \quad (43)$$

$$\mathbf{y_c}_{\min} \leq \mathbf{y_c} \leq \mathbf{y_c}_{\max} \quad (44)$$

where J_E represents a suitable economic objective function for the process, $\mathbf{y}_c$ represents the controlled outputs. Solution of the above optimization problem yields the optimal setpoints $\mathbf{r}_k = \mathbf{y}_c$, which is given to the NMPC component. If a sensor fault is identified by NLGLR, optimum set-points of the system do not change and invocation of RTO is not needed. Thus, RTO is invoked only when NLGLR identifies a fault in parameter.

2.7 Adaptive NMPC Formulation

At current k^{th} time instant, the following optimization problem is solved over a prediction horizon $[k, k+N_p]$

$$\min_{\{\mathbf{u}_{k|k},..,\mathbf{u}_{k+N_p-1|k}\}} \sum_{j=0}^{N_p-1} \left\{\mathbf{E}^T_{k+j+1|k}\mathbf{W}_E\mathbf{E}_{k+j+1|k} + J_{\Delta u}\right\} \quad (45)$$

$$J_{\Delta u} = \Delta\mathbf{u}^T_{k+j|k}\mathbf{W}_{\Delta u}\Delta\mathbf{u}_{k+j|k}$$

subject to,

$$\mathbf{z}_{k+j+1|k} = \mathbf{F}[\mathbf{z}_{k+j|k}, \mathbf{u}_{k+j|k}, \boldsymbol{\theta}] + \mathbf{L}_k\mathbf{e}_{k,f} \quad (46)$$

$$\widehat{\mathbf{y}}_{c,k+j+1|k} = \mathbf{g}_c(\mathbf{z}_{k+j+1|k}) + \boldsymbol{\varepsilon}_{k,f} \quad (47)$$

$$\mathbf{u}_{\min} \leq \mathbf{u}_{k+j|k} \leq \mathbf{u}_{\max} \quad (48)$$

$$\text{for} \quad j = 0, 1, ..., N_p - 1$$

$$\mathbf{E}_{k+j|k} = \hat{\mathbf{y}}_{c,k+j|k} - \mathbf{r}_k \quad \text{for} \quad j = 1, ..., N_p \quad (49)$$

$$\Delta\mathbf{u}_{k+j|k} = \mathbf{u}_{k+j|k} - \mathbf{u}_{k+j-1|k} \quad (50)$$

$$\text{for} \quad j = 1, 2, ...N_p - 1$$

$$\Delta\mathbf{u}_{k|k} = \mathbf{u}_{k|k} - \mathbf{u}_{k-1} \quad (51)$$

$$\mathbf{z}_{k|k} = \widehat{\mathbf{x}}_{k|k}$$

where, N_p is the prediction horizon. $\mathbf{W}_E$ and $\mathbf{W}_{\Delta u}$ are weighting matrices. Here, $\mathbf{e}_{k,f}$ and $\boldsymbol{\varepsilon}_{k,f}$ represent filtered innovation signal and filtered estimation error sequence which are computed as follows:

$$\mathbf{e}_{k,f} = \alpha\mathbf{e}_{k-1,f} + (1-\alpha)\mathbf{e}_k \quad (52)$$

$$\boldsymbol{\varepsilon}_{k,f} = \alpha\boldsymbol{\varepsilon}_{k-1,f} + (1-\alpha)\boldsymbol{\varepsilon}_{c,k} \quad (53)$$

where $\mathbf{e}_k = \mathbf{y}_k - \mathbf{g}(\widehat{\mathbf{x}}_{k|k-1})$ and $\boldsymbol{\varepsilon}_{c,k} = \mathbf{y}_{c,k} - \mathbf{g}_c(\widehat{\mathbf{x}}_{k|k})$. Here, $\alpha \in (0, 1)$ is a tuning parameter. These signals are used to compensate the model predictions for the model plant mismatch. The parameter vector $\boldsymbol{\theta}$ in the prediction equation (46) is adapted on-line as follows:

$$\boldsymbol{\theta} = \left\{ \begin{array}{c} \bar{\boldsymbol{\theta}} \text{ for } k < t_d + N_g \\ \bar{\boldsymbol{\theta}}+\Delta\theta_{f,k}\boldsymbol{\xi}^{(f)} \text{ for } k \in [t_d, +N_g, t_f] \\ \bar{\boldsymbol{\theta}}+\Delta\theta_{f,t_f}\boldsymbol{\xi}^{(f)} \text{ for } k \geq t_f \end{array} \right\} \quad (54)$$

This makes the NMPC formulation adaptive and sensitive to the parameter variations. Also, note that if a sensor bas is isolated, the "corrected measurements" and "corrected controlled outputs" are used for computing signals $\mathbf{e}_k$ and $\boldsymbol{\varepsilon}_{c,k}$.

3. SIMULATION STUDIES

A CSTR system exhibiting input multiplicity is chosen to demonstrate the efficacy of the proposed approach. The system dynamics is represented by the following set of nonlinear ODE's (Deshpande et al. (2009)):

$$\frac{dC_A}{dt} = \frac{F_i}{hA_c}(C_{Ai} - C_A) - K_1C_A + K_2C_B \quad (55)$$

$$\frac{dC_B}{dt} = -\frac{F_i}{hA_c}C_B + K_1C_A - K_2C_B \quad (56)$$

$$\frac{dT}{dt} = \frac{1}{hA_c}F_i(T_i - T) + \frac{-H_r}{\rho C_p}(K_1C_A - K_2C_B) \quad (57)$$

$$\frac{dh}{dt} = \frac{1}{A_c}(F_i - k\sqrt{h}) \quad (58)$$

$$K_1 = k_f \exp(\frac{-E_f}{T});\ K_2 = k_b \exp(\frac{-E_b}{T})$$

A reversible first order reaction takes place in the CSTR system where A is converting to product B ($A \rightleftharpoons B$). Concentration of B (C_B), level (h) and Temperature (T) are assumed to be measured states. Concentration of A (C_A) is an unmeasured state. C_B, and h are controlled

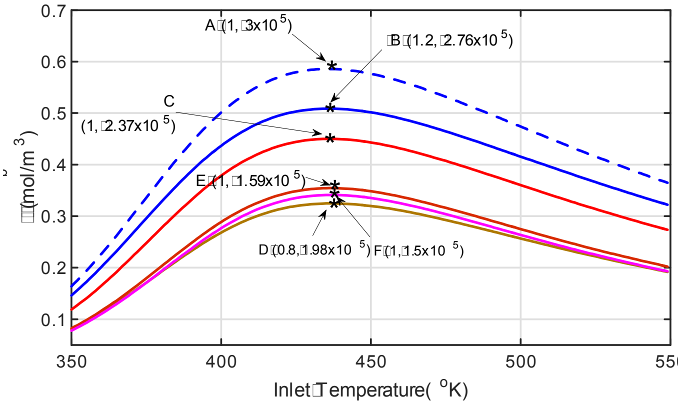

Fig. 2. CSTR system: steady state operating points (C_{Ai}, k_f)

outputs. Inlet flow rate (F_i) and inlet Temperature (T_i) are treated as the manipulated inputs. Forward reaction rate constant (k_f) and inlet concentration (C_{Ai}) are considered as time varying set of parameters. Nominal parameters and initial optimal steady state conditions can be found in Li and Biegler (1988). The steady state characteristic of this system is shown in Figure (2). This figure shows steady state optimal profile of C_B w.r.t. change in inlet Temperature for different set of parameters values of C_{Ai} and k_f. We can see the peak point shifts when parameter set (C_{Ai}, k_f) changes. The objective of controller is to maintain the system at the shifting peak point. It is difficult to control at the peak because steady state gain changes its sign across the peak. The Sampling time interval (T_s) is taken 0.1 min.

In this case study, we have hypothesized five different faults as: (i) fault in unmeasured disturbance (C_{Ai}), (ii) fault in parameter (k_f), (iii) biases in all three sensors, i.e.

$$\mathcal{F} = \{\Delta C_{Ai}, \Delta k_f, \{\Delta b_j : j = 1, 2, 3\}\}$$

NLGLR and estimation window length are $N_g = 40$ and $N_{ml} = 20$, respectively. Levels of significance for fault detection and confirmation are taken 0.0005 and 0.0001, respectively. Step changes of -16.66% and $+16.66\%$ are introduced in k_f at 100^{th} and 800^{th} time instant respectively. Similarly, step changes of $+20\%$ and -20% are introduced in C_{Ai} at 400^{th} and 1200^{th} time instant respectively. The nominal steady state condition for plant and estimator are

$$\mathbf{x_s} = \left[\, 0.4912 \;\, 0.5088 \;\, 438.4889 \;\, 0.1600 \,\right]^T$$

and $\mathbf{u_s} = \left[\, 1 \;\, 435.945 \,\right]^T$. State noise and measurement noise covariances are chosen as:

$$\mathbf{Q} = diag\left[\, 0.0025^2 \;\, 0.0025^2 \;\, 0.0750^2 \;\, 0.0010^2 \,\right]$$

$$\mathbf{R} = diag\left[\, 0.0025^2 \;\, 0.1^2 \;\, 0.001^2 \,\right]$$

Estimator is initialized with $\widehat{\mathbf{x}}_{0|0} = \mathbf{x_s}$ and $P_{0|0} = 5 \times Q$. MPC controller tuning parameters, Error weighting matrix ($\mathbf{W}_E$) and Input change weighting matrix ($\mathbf{W}_{\Delta u}$) are as follow:

$$\mathbf{W}_E = 10^2 \times \mathbf{I}_{2\times 2} \text{ and } \mathbf{W}_{\Delta u} = diag\left[\, 1 \;\, 0.1 \,\right]$$

Prediction horizon (N_p) and control horizon (N_c) are selected 25 and 4, respectively. RTO is invoked at every 10 sampling instant after fault confirmation ($N_{rto} = 10$). The objective function at RTO layer is chosen as $J_E =$

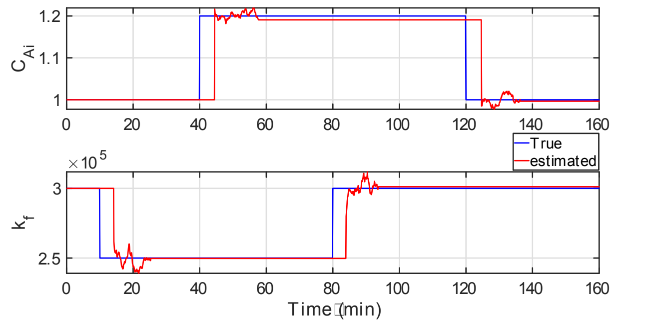

Fig. 3. CSTR System: Inlet concentration(C_{Ai}) fault and kinetic rate constant (k_f) fault estimation

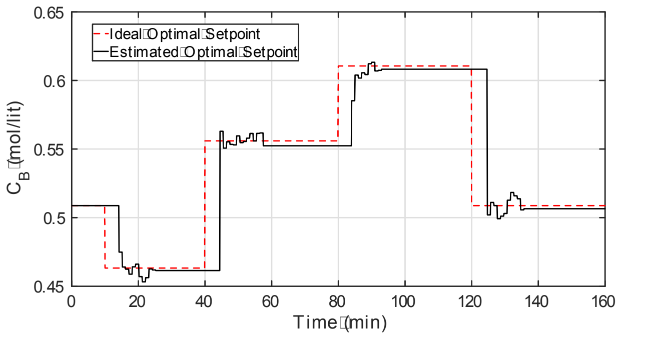

Fig. 4. CSTR System: Ideal and estimated setpoint (C_B) comparison

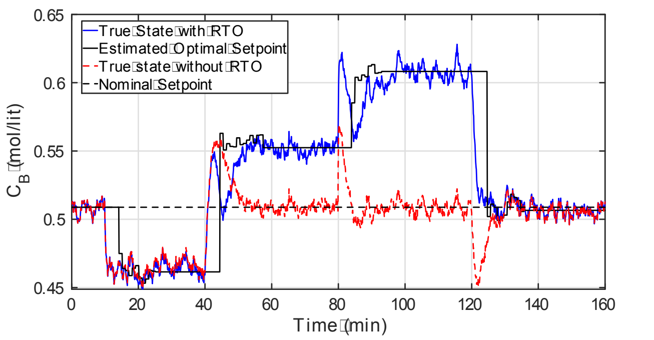

Fig. 5. CSTR System: Comparison of setpoint (C_B) tracking with and without RTO

$-C_B$, which is minimized w.r.t decision variable T_i. For estimation stopping criterion, to compare the mean of 2 data sets from recent past 80 data points with each set of $q = 40$ are used to apply the hypothesis test with significance level $\alpha_t = 0.03$.

Figure (3) shows estimation of unmeasured disturbance (C_{Ai}) and parameter (k_f), respectively. It is evident from these figures that proposed approach is able to isolate the correct fault and moving window estimation scheme is able to improve the fault magnitude estimates. However, the final value of the estimated parameter depends on stopping criterion; as a consequence, a small bias appears between the estimate and the true value. Comparison of ideal and estimated setpoint is shown in Figure (4). The ideal setpoints are calculated assuming that the parameter variation is perfectly known to RTO and RTO is invoked

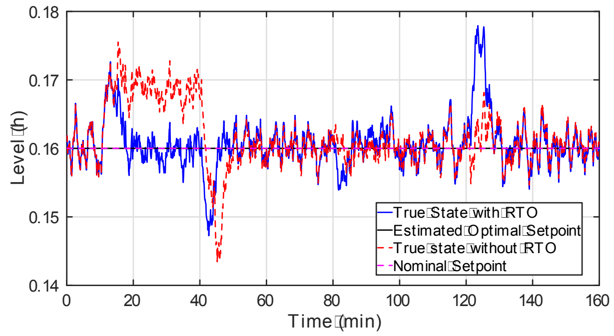

Fig. 6. CSTR System: Comparison of setpoint (h) tracking with and without RTO

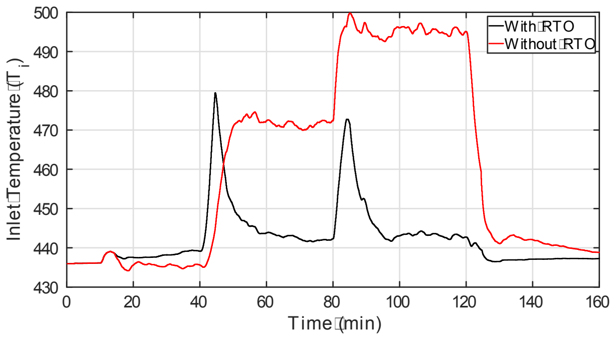

Fig. 7. CSTR System: Input Inlet Temperature profile

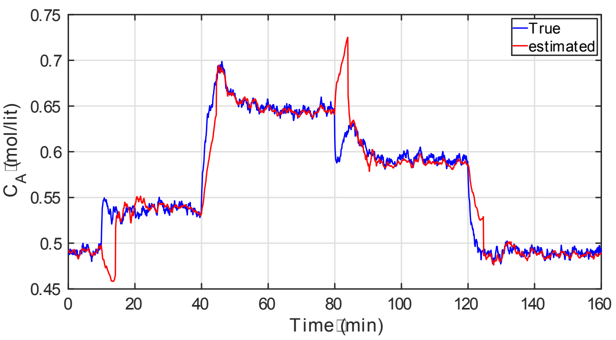

Fig. 8. CSTR System: Comparison of true and estimated state C_A

only once when parameter changes. It can be observed that estimated setpoint profile is able to closely track the ideal optimum setpoint. Figure (5) presents setpoint tracking profile of C_B. This figure also compares the tracking of C_B when setpoint is kept at nominal value (i.e., RTO is not invoked) throughout the simulation. In this case, when setpoint shifts above the nominal setpoint, we can see that the CSTR is operated under suboptimal conditions and as it tracks the nominal setpoint. On the other hand, when the nominal setpoint become infeasible, the NMPC results in an offset and also results in a sub-optimal operation. The tracking of level setpoint is shown in Figure (6). An offset is observed in the case when RTO is not invoked. Input profiles of inlet temperature obtained from the two NMPC schemes are shown in Figure (7). Comparison of true and estimated unmeasured state, C_A, obtained using the proposed intelligent state estimation scheme is presented in Figure (8). It is observed that estimated states

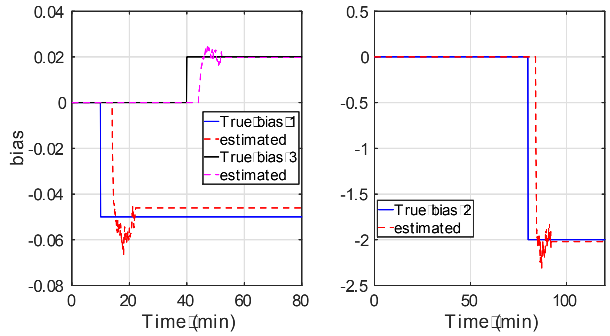

Fig. 9. CSTR system: sensor bias estimation

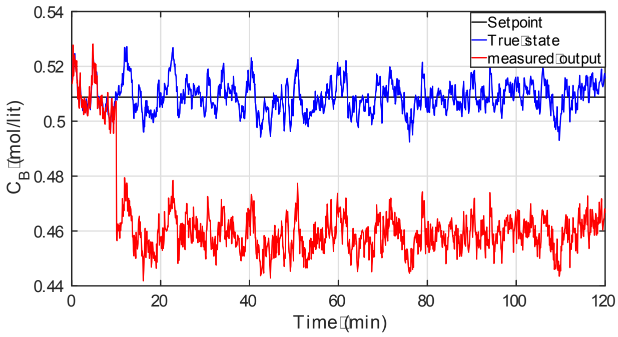

Fig. 10. CSTR system: setpoint tracking of C_B with sensor bias

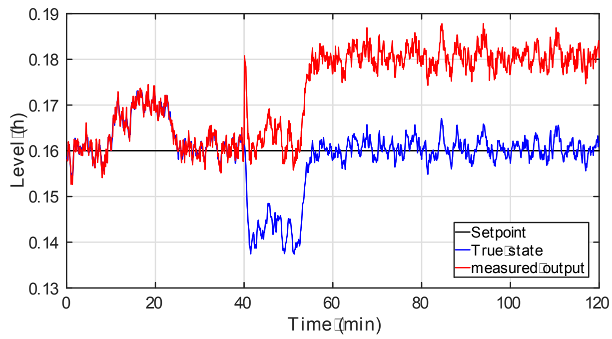

Fig. 11. CSTR system: setpoint tracking of level (h) with sensor bias

are able to track true states with a delay. NLGLR uses data over a window results in the delay in the parameter estimation, which causes a mismatch between the true and estimated states during application of NLGLR.

Another simulation is run to show the efficacy of the proposed scheme in the presence of sensor biases. Simulation is carried out for 1200 samples. Biases in sensors are introduced sequentially as follows: a bias of magnitude -0.050 is introduced in the measurement (C_B) at 100^{th} time instant, followed by biases of magnitude -2 and 0.020 are introduced in the measurements (T and h) at 800^{th} and 400^{th} time instants, respectively. From the Figure (9), it can be observed that the proposed scheme is capable to identify and track the correct fault. When a sensor bias is isolated, there is no need to invoke RTO and the setpoints remain at their nominal values. Figures (10) and (11) show that the tracking of setpoint for C_B and

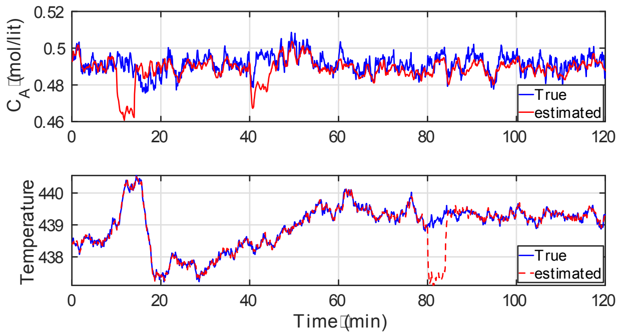

Fig. 12. CSTR System: Comparison of true and estimated state (C_A and T) with sensor bias

h, respectively. As can be seen from these figures, the proposed fault compensation scheme is able to move and maintain the true states close to their respective setpoints while an offset develops between the measured outputs and setpoints. Figure (12) presents comparison of estimated and true states for C_A and T. As can be seen from this figure, the estimated states tracks the true states except for time intervals when data is collected for application of NLGLR. A small sustained bias is observed in estimates of C_A, which can be attributed to inaccuracies in the fault magnitude estimates.

The average computation time required for implementing the integrated FDI, RTO and NMPC scheme (using a PC with 8GB RAM, Intel(R) Xeon(R) CPU E3-1226 v3 @ 3.30Ghz processor and 'fmincon' function) was 4.13 sec. When the system operated under NMPC alone between two instances of occurrence of parameter change/ sensor bias, then the average computation time is 0.46 sec.

4. CONCLUSION

This work attempts to enlarge the scope of the integrated RTO-MPC scheme developed by Valluru and Patwardhan (2019) to include a wider operating envelope of the process covering both normal and off-normal and abnormal operating scenarios. This is achieved by combining fault diagnosis (nonlinear GLR approach (Deshpande et al. (2009))) with simultaneous state and parameter estimation (moving window state and parameter estimator (Valluru et al. (2017))) with embedded intelligence to discern the impending situation and configure the state and parameter estimator to work in an efficient and parsimonious manner. The estimation of parameter(s)/bias is carried out only when required and triggered by the fault identification scheme. Thus, the subset of parameters that are being estimated online can change with time. The proposed approach employs a single model to carry out four different tasks: process monitoring, state and parameter estimation, nonlinear control, and real-time optimization. The proposed intelligent optimizing control scheme is implemented on a benchmark CSTR system exhibiting input multiplicity behavior. The set of hypothesized faults consists of abrupt changes in an unmeasured disturbance, a reaction kinetics parameter, and biases in three sensors. The optimum operating point of this system is sensitive to the mean shift in the unmeasured disturbance or the system parameter. The proposed approach successfully isolates the parameter/ unmeasured disturbance that has undergone abrupt change and uses the estimated values of the parameter to update models used in the RTO and NMPC. Thus, the shifting economic optimum is tracked without significant delays. When a sensor measurement becomes biased, the proposed scheme can correctly isolate the bias sensor and make appropriate modifications to the measurement model. Also, the proposed fault-tolerant control scheme maintains the true values of the states at the specified setpoint in the presence of sustained bias. Moreover, the proposed approach can isolate multiple changes in parameters or sensor biases that occur sequentially in time. Thus, the proposed integrated approach can be viewed as a step toward developing an autonomous, flexible, and resilient closed-loop system. This work was focused only on abrupt changes in parameters and soft faults. The work is further being extended to accommodate slowly drifting parameters and sensor/actuator failures.

REFERENCES

Darby, M.L., Nikolaou, M., Jones, J., and Nicholson, D. (2011). RTO: An overview and assessment of current practice. *Journal of Process Control*, 21(6), 874–884.

Deshpande, A.P., Patwardhan, S.C., and Narasimhan, S.S. (2009). Intelligent state estimation for fault tolerant nonlinear predictive control. *Journal of Process Control*, 19(2), 187–204.

Engell, S. (2007). Feedback control for optimal process operation. *Journal of Process Control*, 17(3), 203–219.

Krishnamoorthy, D., Foss, B., and Skogestad, S. (2018). Steady-state real-time optimization using transient measurements. *Computers and Chemical Engineering*, 115, 34–45.

Li, W.C. and Biegler, L.T. (1988). Process Control Strategies for Constrained Nonlinear Systems. *Industrial and Engineering Chemistry Research*, 27(8), 1421–1433.

Liu, J., Gnanasekar, A., Zhang, Y., Bo, S., Liu, J., Hu, J., and Zou, T. (2021). Simultaneous State and Parameter Estimation: The Role of Sensitivity Analysis. *Industrial & Engineering Chemistry Research*, 60(7), 2971–2982.

Muske, K.R. and Badgwell, T.A. (2002). Disturbance modeling for offset-free linear model predictive control. *Journal of Process Control*, 12(5), 617–632.

Patwardhan, S.C., Narasimhan, S., Jagadeesan, P., Gopaluni, B., and L. Shah, S. (2012). Nonlinear Bayesian state estimation: A review of recent developments. *Control Engineering Practice*, 20(10), 933–953.

Valluru, J., Lakhmani, P., Patwardhan, S.C., and Biegler, L.T. (2017). Development of moving window state and parameter estimators under maximum likelihood and Bayesian frameworks. *Journal of Process Control*, 60, 48–67.

Valluru, J. and Patwardhan, S.C. (2019). An Integrated Frequent RTO and Adaptive Nonlinear MPC Scheme Based on Simultaneous Bayesian State and Parameter Estimation. *Industrial & Engineering Chemistry Research*, 58(18), 7561–7578.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., and Kavuri, S.N. (2003). A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, 27(3), 293–311.