Journal of Process Control 21 (2011) 585–601

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Identification of process and measurement noise covariance for state and parameter estimation using extended Kalman filter

Vinay A. Bavdekar [a], Anjali P. Deshpande [b], Sachin C. Patwardhan [a,*]

[a] Department of Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra 400076, India
[b] Systems and Control Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

ARTICLE INFO

Article history:
Received 12 August 2010
Received in revised form
24 December 2010
Accepted 3 January 2011
Available online 4 March 2011

Keywords:
Nonlinear state estimation
Extended Kalman filter
Covariance estimation
Maximum likelihood estimates
Expectation maximisation algorithm

ABSTRACT

The performance of Bayesian state estimators, such as the extended Kalman filter (EKF), is dependent on the accurate characterisation of the uncertainties in the state dynamics and in the measurements. The parameters of the noise densities associated with these uncertainties are, however, often treated as 'tuning parameters' and adjusted in an ad hoc manner while carrying out state and parameter estimation. In this work, two approaches are developed for constructing the maximum likelihood estimates (MLE) of the state and measurement noise covariance matrices from operating input–output data when the states and/or parameters are estimated using the EKF. The unmeasured disturbances affecting the process are either modelled as *unstructured noise* affecting all the states or as *structured noise* entering the process predominantly through known, but unmeasured inputs. The first approach is based on direct optimisation of the ML objective function constructed by using the innovation sequence generated from the EKF. The second approach – the extended EM algorithm – is a derivative-free method, that uses the joint likelihood function of the complete data, i.e. states and measurements, to compute the next iterate of the decision variables for the optimisation problem. The efficacy of the proposed approaches is demonstrated on a benchmark continuous fermenter system. The simulation results reveal that both the proposed approaches generate fairly accurate estimates of the noise covariances. Experimental studies on a benchmark laboratory scale heater-mixer setup demonstrate a marked improvement in the predictions of the EKF that uses the covariance estimates obtained from the proposed approaches.

1. Introduction

To operate a chemical process profitably, in a competitive economic environment, it becomes necessary to monitor and tightly control quality variables associated with the process. However, critical quality variables such as product concentrations in reactor/distillation column output streams, molecular weight distribution of a polymer melt or biomass concentration in a fermenter, are difficult to measure on-line. Even when these variables are measured through lab assays, such measurements are infrequent and typically available at irregular intervals. Moreover, even when online measurements of quality variables is feasible, it can prove to be a prohibitively costly option. In such a scenario, dynamic model based estimation of unmeasured states, or a soft sensor, is an attractive alternative for process monitoring and control. With the availability of high speed computers at relatively low costs, process monitoring and advanced multivariable control, which is based on on-line use of mechanistic models, are becoming feasible options in the recent years. Since a mechanistic model is representative of the physical states of the plant, it can be used to deduce information of trajectories of internal unmeasured (or irregularly measured) states at regular and faster rate.

However, even when a reliable mechanistic model for the system under consideration is available, using it for online monitoring and control is not an easy task. Process plants are continuously affected by unmeasured disturbances of different kind and the measurements are corrupted with noise. Thus, it becomes necessary to develop state estimators that systematically fuse noisy measurements with mechanistic models to generate state estimates. The state observers can be developed either through deterministic approaches [1] or through Bayesian approaches [2]. Bayesian approaches provide a systematic approach to handle the unknown inputs affecting the states and the measurement noise. To generate reliable estimates of the unmeasured or infrequently measured states it is important to develop a reasonably accurate characterisation of these unmeasurable signals. This is a critical aspect in the development of Bayesian state estimators. An incorrect choice of noise characteristics leads to deterioration in the performance of the state estimator and, in the worst case, the estimator may diverge [3].

A commonly used assumption in the development of Bayesian state estimation is that the state and measurement disturbances

* Corresponding author. Tel.: +91 22 25767211; fax: +91 22 25726895.
E-mail address: sachinp@iitb.ac.in (S.C. Patwardhan).

0959-1524/$ – see front matter
doi:10.1016/j.jprocont.2011.01.001

are additive, zero mean and have a Gaussian distribution. Thus, the problem of noise characterisation is reduced to estimation of the covariances of these multi-variate Gaussian distributions. The Kalman filter (KF) is probably the most widely used Bayesian state estimator. The KF generates the maximum likelihood estimates for a linear dynamic system subjected to additive process and measurement noises with a multivariate Gaussian distribution. The problem of estimating the state and the measurement noise covariances has been well studied in the literature on Kalman filtering. The methods for estimating the state and measurement noise covariances for the KF can be broadly classified into four categories-covariance matching, correlation techniques, Bayesian and maximum likelihood methods.

The covariance matching technique [4], is an adaptive algorithm in which the estimates of the process noise covariance (**Q**) and measurement noise covariance (**R**) are computed at every sampling instant. Using the past data of the state prediction trajectories, the sample covariance of the state prediction error is computed either cumulatively over the entire data in the past or over a moving window in time. Estimates of **Q** are constructed using the sample covariance of the state prediction error and estimates of **R** are generated using the sample covariance of the innovation sequence. The main disadvantage with this method is that it does not ensure that the positive definiteness of the **Q** and **R**, and ad hoc procedures have to be artificially introduced to maintain the positive definiteness. Moreover, Odelson et al. [5] have shown that the covariance matching method yields biased estimates of the covariances.

Correlation techniques estimate the covariance matrices by making use of the sample auto-correlations between the innovations. A least square estimate of **Q** and **R** is generated by exploiting the recurrent relations between the estimation error covariance and innovation covariance. Mehra [6] has also proposed a blockwise recursive scheme for online update of **Q** and **R**. For a large dimensional system with few measurements, this approach can prove to be restrictive as the number of unknowns in **Q** is required to be less than $n \times r$, where, n is the number of states and r is the number of measurements. Bèlanger [7] extended this method for estimating covariances of non-stationary processes. However, these methods are known to yield biased estimates of the covariance matrices [8]. Recently, Odelson et al. [5,8] have developed an autocovariance least-squares method to estimate **Q** and **R** for the KF. The algorithm defines a lagged autocovariance function between the measurements, which is used to develop a linear least squares formulation to estimate **Q** and **R**. It may be noted that the autocovariance function, and, in turn, the least square estimate, is a function of user-defined lags.

Bayesian methods, and, in particular, the maximum likelihood based methods formulated the covariance estimation problem as maximisation of the likelihood function associated with the innovations. The gradient-based numerical optimisation schemes were used to estimate the covariance matrices [10]. These methods use gradient-based numerical optimisation schemes to estimate the covariance matrices, **Q** and **R**. The expectation-maximisation (EM) algorithm, developed by Dempster et al. [9], is an iterative procedure that uses the complete data likelihood function to compute the maximum likelihood estimates of the model parameters. The advantage of the maximum likelihood approaches is that these can be applied even when measurements are available at irregular intervals. Shumway and Stoffer [10] and Raghavan et al. [11] provide a framework to identify state space models along with the state and measurement noise covariances for linear systems using the EM algorithm in the presence of irregularly sampled data. For a linear system, with additive zero-mean Gaussian white noise sequence, Shumay and Stoffer [10] show that the algorithm yields analytical expressions for the iterates of the parameters, thus avoiding explicit computations of the gradient. These methods were not preferred earlier as they require heavy computations. However, with high speed computing becoming easily available, these algorithms can be used for off-line estimation of the covariances.

Most of the work reported in literature is for estimating the densities of the state and measurement noise in the case of linear dynamic systems when a linear filtering technique (that is KF) is employed. However, since majority of the chemical processes exhibit nonlinear dynamics, it is essential to use nonlinear observers for estimating the states of the system. There is very little work reported in the literature on the estimation of state and measurement noise densities for nonlinear Bayesian observers. In practice, nonlinear filters are often tuned using heuristic approaches. However, tuning the filter using heuristics is not straightforward and can lead to degradation in the performance of the filter.

In this work the extended Kalman filter (EKF), which is arguably the most widely used Bayesian estimator for nonlinear systems, is chosen for carrying out state estimation. EKF can be viewed as an approximate and computationally tractable sub-optimal solution to the sequential Bayesian estimation problem under the simplifying assumption that estimation error densities can be approximated as Gaussian distributions. Valappil and Georgakis [12] present a systematic method to estimate the process noise covariance as arising from uncertainty in model parameters. Their approach assumes a structured uncertainty in the model, which can be modelled as variations in the parameters. A priori knowledge about the model parameter variances is combined with the state estimator using the parameter sensitivity matrices. The second approach is based on Monte Carlo simulations. Monte Carlo samples of the parameters are generated using a priori knowledge of the uncertainties associated with the parameters. While these two approaches simplify the process of tuning the EKF, in practice, the parameter error covariance matrix may change with time or it may not be known fully. In addition, they also assume that the measurement noise covariance, **R**, is known or can be obtained from the measurement data. Recently, Goodwin and Agüero [13] have proposed a modification to the EM algorithm, in which the expectation step is carried out using EKF, for identification of a nonlinear black-box model. A by product of this algorithm is the estimates of covariance matrices **Q** and **R** associated with the EKF. However, since the focus is to identify a black-box model, the resulting noise densities are related to the innovation sequences rather than physically meaningful noise sources. Chitralekha et al. [14] have demonstrated the use of an extended version of the EM algorithm for estimation of model parameters of a nonlinear mechanistic model using a batch of data available from the plant. They also present results for recursive state and parameter estimation using the conventional state augmentation approach, whereby the parameters are assumed to follow a random walk model. Their work, however, assumes that the covariance matrices of the state, parameter and measurement noise are known *a priori*.

Thus, while characterising the disturbance and measurement noise densities is an important aspect of implementing EKF, not much literature is available that deals with this issue. Among the methods available in the context of linear filtering, the MLE based methods appear to be promising candidates for dealing with the noise density estimation problem associated with nonlinear systems. In this work, two approaches are developed based on the MLE framework for estimation of the noise covariances from the operating data when the state estimation is carried out using EKF. The first approach, hencforth referred to as the direct optimisation approach, maximises the log likelihood function associated with the innovations sequence generated by the EKF. The EM algorithm has been successfully used for the estimation of covariance matrices of state and measurement noise of a linear system. In this work, modifications in the expectation step are proposed, so as to use the

EM approach to estimate **Q** and **R** matrices associated with an EKF formulation for simultaneous state and parameter estimation. The resulting approach is henceforth referred to as extended EM algorithm. The efficacy of the proposed approaches is demonstrated by simulating a benchmark continuous fermenter system. The effectiveness of the two approaches is also demonstrated by carrying out simultaneous state and parameter estimation on a benchmark heater-mixer experimental setup [15].

The organisation of the paper is as follows. The assumptions involved in the problem formulation, the EKF algorithm, formulation of the problem of estimating **Q** and **R** and the direct optimisation solution strategy are described in Section 2. The proposed extended EM-algorithm is derived and the advantages and limitations of the direct optimisation approach and extended EM algorithm are presented in Section 3. In Section 4, the results obtained by applying the two approaches on a benchmark continuous fermenter are presented. The results of experimental studies on the benchmark heater-mixer setup are presented in Section 5.

2. The covariance estimation problem

To generate reliable estimates using EKF, it is important to know the distributions of the unmeasured disturbances and measurement noise with a reasonable accuracy. In this work, it is proposed to identify the parameters of these distributions from the operating data. The details of the proposed approach are presented in this section.

2.1. Process simulation and model for state estimation

Consider a continuously operated process, which can be described by the following set of ordinary differential equations

$$\begin{aligned}\frac{d\mathbf{z}}{dt} &= \mathbf{f}(\mathbf{z}, \mathbf{m}, \mathbf{d}, \boldsymbol{\delta}, \mathbf{p}, \mathbf{t}) \\ \mathbf{y}(t) &= \mathbf{h}(\mathbf{z}) + \mathbf{v}(t)\end{aligned} \tag{1}$$

where $\mathbf{z} \in \mathbb{R}^n$ denotes the system states, $\mathbf{m} \in \mathbb{R}^m$ denotes the manipulated inputs, $\mathbf{d} \in \mathbb{R}^{d_u}$ denotes the unmeasured disturbances, $\boldsymbol{\delta} \in \mathbb{R}^{d_m}$ denotes the measured disturbances, $\mathbf{p} \in \mathbb{R}^p$, the parameters, $\mathbf{y} \in \mathbb{R}^r$ denotes the measured outputs and $\mathbf{v} \in \mathbb{R}^r$ denotes the measurement noise. The process simulations and the modelling for the state estimation are carried out under the following simplifying assumptions:

Assumption 1. The measurement noise can be modelled as a zero mean white noise processes with Gaussian distribution, i.e. $\mathbf{v}(t) \sim \mathcal{N}(\bar{0}, \mathbf{R})$, where **R** represents the covariance matrix.

Assumption 2. The manipulated inputs are piecewise constant over the sampling interval, i.e.

$$\mathbf{m}(t) = \mathbf{m}_k \quad \text{for} \quad t_k \le t < t_{k+1} = t_k + T$$

where T represents the sampling interval. Further, the true value of the manipulated inputs (**m**) is related to the known/computed value of the manipulated inputs (**u**) as follows

$$\mathbf{m}_k = \mathbf{u}_k + \mathbf{w}_{u,k}$$

where $\mathbf{w}_{u,k} \in \mathbb{R}^m$ denotes an unknown disturbance in manipulated inputs such that $\mathbf{w}_{u,k} \sim \mathcal{N}(\bar{0}, \mathbf{Q}_u)$.

Assumption 3. The choice of the sampling interval is small enough so that the variation of the unmeasured disturbances can be adequately approximated using the piecewise constant functions of the form

$$\mathbf{d}(t) = \bar{\mathbf{d}} + \mathbf{w}_{d,k} \quad \text{for} \quad t_k \le t < t_{k+1} = t_k + T$$

$\mathbf{w}_{d,k} \in \mathbb{R}^{d_u}$ denotes a disturbance in the unmeasured disturbance such that $\mathbf{w}_{d,k} \sim \mathcal{N}(\bar{0}, \mathbf{Q}_d)$ and $\bar{\mathbf{d}}$ represents the mean or the steady state value of the unmeasured disturbance at some desired operating point. Also, the variation of the measured disturbances and model parameters can be adequately approximated using the piecewise constant functions of the form

$$\left\{ \begin{array}{l} \boldsymbol{\delta}(t) = \boldsymbol{\delta}_k \\ \mathbf{p}(t) = \mathbf{p}_k \end{array} \right\} \quad \text{for} \quad t_k \le t < t_{k+1} = t_k + T$$

Under Assumptions 1–3, a discrete time representation of the true system dynamics can be arrived at as follows:

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \int_{kT}^{(k+1)T} \mathbf{f}[\mathbf{z}(\tau), \mathbf{m}_k, \mathbf{d}_k, \boldsymbol{\delta}_k, \mathbf{p}_k] d\tau = \mathbf{F}(\mathbf{z}_k, \mathbf{m}_k, \mathbf{d}_k, \boldsymbol{\delta}_k, \mathbf{p}_k) \tag{2}$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{z}_k) + \mathbf{v}_k \tag{3}$$

where $\mathbf{v}_k = \mathbf{v}(kT)$ represents measurement noise at kth sampling instant.

Assumption 4. The effect of unmeasured inputs/disturbances on the state dynamics is additive and can approximated using either of the following approaches:

(a) State disturbance with unknown source:

$$\mathbf{z}_{k+1} = \mathbf{F}(\mathbf{z}_k, \mathbf{u}_k, \bar{\mathbf{d}}, \boldsymbol{\delta}_k, \mathbf{p}_k) + \mathbf{w}_{s,k} \tag{4}$$

Here, $\mathbf{w}_{s,k} \in \mathbb{R}^n$ represents a zero mean white noise process with Gaussian distribution, i.e. $\mathbf{w}_{s,k} \sim \mathcal{N}(\bar{0}, \mathbf{Q}_s)$.

(b) Unknown disturbance in the manipulated inputs

$$\mathbf{z}_{k+1} = \mathbf{F}(\mathbf{z}_k, \mathbf{u}_k, \bar{\mathbf{d}}, \boldsymbol{\delta}_k, \mathbf{p}_k) + \Gamma_{u,k}\mathbf{w}_{u,k} = \mathbf{F}(\mathbf{z}_k, \mathbf{u}_k) + \Gamma_{u,k}\mathbf{w}_{u,k} \tag{5}$$

where $\Gamma_{u,k} = [\partial \mathbf{F}/\partial \mathbf{u}]_{(\mathbf{z}_k, \mathbf{u}_k, \bar{\mathbf{d}}, \boldsymbol{\delta}_k, \mathbf{p}_k)}$.

(c) Unmeasured disturbance with known sources[12]

$$\mathbf{z}_{k+1} = \mathbf{F}(\mathbf{z}_k, \mathbf{u}_k, \bar{\mathbf{d}}, \boldsymbol{\delta}_k, \mathbf{p}_k) + \Gamma_{d,k}\mathbf{w}_{d,k} = \mathbf{F}(\mathbf{z}_k, \mathbf{u}_k) + \Gamma_{d,k}\mathbf{w}_{d,k} \tag{6}$$

where $\Gamma_{d,k} = [\partial F/\partial \mathbf{d}]_{(\mathbf{z}_k, \mathbf{u}_k, \bar{\mathbf{d}}, \boldsymbol{\delta}_k, \mathbf{p}_k)}$.

Assumption 5. The variation of the measured disturbances and model parameters can be modelled as an integrated white noise process

$$\boldsymbol{\delta}_{k+1} = \boldsymbol{\delta}_k + \mathbf{w}_{\delta,k} \tag{7}$$

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{w}_{p,k} \tag{8}$$

where $\mathbf{w}_{\delta,k} \in \mathbb{R}^{d_m}$ represents the disturbance that affects the measured disturbances such that $\mathbf{w}_{\delta,k} \sim \mathcal{N}(0, \mathbf{Q}_\delta)$. The disturbances in the parameters are represented as $\mathbf{w}_{p,k} \in \mathbb{R}^p$ such that $\mathbf{w}_{p,k} \sim \mathcal{N}(0, \mathbf{Q}_p)$.

Defining an augmented state vector

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{z}_k^T & \boldsymbol{\delta}_k^T & \mathbf{p}_k^T \end{bmatrix}^T$$

the model used for state estimation (Eqs. (4), (7) and (8)) can be represented in a compact form

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k, \bar{\mathbf{d}}) + \Gamma_{w,k}\mathbf{w}_k \tag{9}$$

where the signal $\mathbf{w}_k \in \mathbb{R}^{n_w}$ denotes the appropriate form of the *state disturbance* such that $\mathbf{w}_k \sim \mathcal{N}(\bar{0}, \mathbf{Q})$ where matrix $\Gamma_{w,k}$ depends on the choice of unmeasured disturbance representation. To simplify the notation, the argument $\bar{\mathbf{d}}$ in Eq. (9) is dropped and the model used in the state estimation is represented as

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k) + \Gamma_w\mathbf{w}_k \tag{10}$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k \tag{11}$$

It is further assumed that $\mathbf{w}_k$ and $\mathbf{v}_k$ are independent, identically distributed random variables and are mutually uncorrelated.

2.2. Extended Kalman filter

The above model is used to construct the predicted ($\hat{x}_{k|k-1}$) and the filtered ($\hat{x}_{k|k}$) estimates of the states using the linearised extended Kalman filter (EKF) approach [16]. The EKF algorithm consists of two step namely, prediction and measurement update. The prediction step is given by

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{F}(\hat{\mathbf{x}}_{k|k}, \mathbf{u}_k) \tag{12}$$

$$\mathbf{P}_{k+1|k} = \boldsymbol{\Phi}_k \mathbf{P}_{k|k} \boldsymbol{\Phi}_k^T + \hat{\boldsymbol{\Gamma}}_w \mathbf{Q}_w \hat{\boldsymbol{\Gamma}}_w^T \tag{13}$$

where $\Phi_k = [\partial \mathbf{F}/\partial \mathbf{x}]_{(\hat{\mathbf{x}}_{k|k}, \mathbf{u}_k, \bar{\mathbf{d}})}$ and

$$\hat{\boldsymbol{\Gamma}}_w = \mathbf{I} \quad \text{or} \quad \hat{\boldsymbol{\Gamma}}_w = \left[\frac{\partial \mathbf{F}}{\partial \mathbf{u}}\right]_{(\hat{\mathbf{x}}_{k|k}, \mathbf{u}_k, \bar{\mathbf{d}})} \quad \text{or} \quad \hat{\boldsymbol{\Gamma}}_w = \left[\frac{\partial \mathbf{F}}{\partial \mathbf{d}}\right]_{(\hat{\mathbf{x}}_{k|k}, \mathbf{u}_k, \bar{\mathbf{d}})}$$

The measurement update or correction step is given by

$$\mathbf{e}_k = \mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) \tag{14}$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{C}_k^T [\mathbf{C}_k \mathbf{P}_{k|k-1} \mathbf{C}_k^T + \mathbf{R}]^{-1} \tag{15}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k \mathbf{e}_k \tag{16}$$

$$\mathbf{P}_{k|k} = [\mathbf{I} - \mathbf{K}_k \mathbf{C}_k] \mathbf{P}_{k|k-1} \tag{17}$$

where $\mathbf{C}_k = [\partial \mathbf{h}/\partial \mathbf{x}]_{\mathbf{x}=\hat{x}_{k|k-1}}$. It may be noted that the EKF algorithm has been developed with the following assumptions about the conditional probability densities of the states [17]

$$\begin{aligned} & p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) \approx \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \\ & p(\mathbf{x}_k|\mathbf{y}_{1:k}) \approx \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}) \\ & \hat{\mathbf{x}}_0 \sim \mathcal{N}(\bar{0}, \mathbf{P}_0) \end{aligned} \tag{18}$$

It may be noted that the above estimation algorithm can be easily modified for the case where the measurements are available at irregular sampling intervals, which are multiples of the smallest sampling time T. In the multi-rate case, the measurement equation changes to

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k$$

where the dimension of the time varying function vector $\mathbf{h}_k(\cdot)$ varies between between 0 and r. The matrix $\mathbf{C}_k$ appearing in the Kalman gain computation can be computed as $\mathbf{C}_k = \mathbf{D}_k \mathbf{H}_k$, where $\mathbf{H}_k = [\partial \mathbf{h}/\partial \mathbf{x}]_{\mathbf{x}=\hat{\mathbf{x}}_{k|k-1}}$. If all the measurements are available at a sampling instant k then, $\mathbf{D}_k = \mathbf{I}_{r\times r}$ and $\mathbf{R}_k = \mathbf{R}$. At an instant when only part of the measurement vector is available, $\mathbf{D}_k$ is a permutation matrix that takes into account only the measurements that are available at the kth instant. In the latter case, the matrix $\mathbf{R}$ in Eq. (15) is replaced by $\mathbf{R}_k = \mathbf{D}_k \mathbf{R} \mathbf{D}_k^T$.

2.3. Problem formulation

While carrying out the exercise for estimating the noise density function, it is assumed that the system under consideration is perturbed deliberately by introducing perturbations in the manipulated inputs in the control relevant frequency range. Let $Z_N = \{Y_N, U_N\}$ denote the data collected during this exercise where $Y_N = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$ and $U_N = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\}$, represent the output and input data sets, respectively, and N represents the number of data points. The noise density function parameters, $\mathbf{Q}$ and $\mathbf{R}$, are estimated using this operating input–output data under the maximum likelihood (ML) framework. Given the data set Z_N, the ML framework maximises the following log of the of likelihood function of the observed data [18].

$$\begin{aligned} & \max_{\Theta} \log L(\Theta|Z_N) = \log p(Z_N|\Theta) \\ & \qquad \text{subject to} \quad \Theta \in S \end{aligned} \tag{19}$$

where $\Theta \equiv (\mathbf{Q}, \mathbf{R}) \in \mathbb{R}^{\phi}$ represent the set of parameters to be estimated, S represents a subset of $\mathbb{R}^{\phi}$ and $p(Z_N|\Theta)$ represents a suitable conditional probability density function (*pdf*) defined using the data set Z_N. To construct such conditional *pdf* in context of the EKF, the following additional assumption is made:

Assumption 6. The innovations sequence, $\mathbf{e}_k$, generated by the EKF (Eq. (14)), is a zero mean Gaussian white noise sequence with the covariance given by

$$p(\mathbf{e}_k|\Theta) = \mathcal{N}(\bar{0}, \Sigma_k) \tag{20}$$

$$\Sigma_k = \mathbf{C}_k \mathbf{P}_{k|k-1} \mathbf{C}_k^T + \mathbf{R} \tag{21}$$

where $\mathbf{P}_{k|k-1}$ is function of $\mathbf{Q}$ through Eq. (13).

With this additional assumption, the likelihood function in Eq. (19) can be written in terms of the probability density functions of the innovation sequence [11] as follows:

$$p(Z_N|\Theta) \approx \prod_{i=1}^{N} p(\mathbf{e}_i|\Theta)$$

and the problem of estimating $\Theta = (\mathbf{Q}, \mathbf{R})$ can be formulated as an optimisation problem that minimises the negative of the log-likelihood function of the innovations

$$\hat{\Theta}_{ML} = arg \min_{\Theta} \left[-\log p(Z_N|\Theta)\right] \tag{22}$$

$$\hat{\Theta}_{ML} \approx arg \min_{(\mathbf{Q},\mathbf{R})} \left[\sum_{i=1}^{N} \log(\det(\Sigma_i)) + \mathbf{e}_i^T \Sigma_i^{-1} \mathbf{e}_i\right] \tag{23}$$

$$\text{Subject to} \tag{24}$$

$$\text{Model equations (12)–(17)} \tag{25}$$

$$\mathbf{Q} > 0 \quad \text{and} \quad \mathbf{R} > 0 \tag{26}$$

Eq. (26) imposes the constraints that the matrices $\mathbf{Q}$ and $\mathbf{R}$ are positive definite. In fact, for a discrete linear system subjected to Gaussian state and measurement noise, Eq. (23) gives exact expression for the likelihood function [19]. The above constrained optimisation problem referred to as **P1** in the rest of the text.

It may be noted that the optimisation formulation is identical in the case of irregularly sampled multi-rate systems. The only difference is in the estimation of the innovations, $\mathbf{e}_k$, and its covariance, Σ_k which are computed as follows

$$\mathbf{e}_k = \mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1})$$

$$\Sigma_k = \mathbf{C}_k \mathbf{P}_{k|k-1} \mathbf{C}_k^T + \mathbf{R}_k$$

In problem **P1**, the decision variables are the covariance matrices $\mathbf{Q}$ and $\mathbf{R}$. If the decision variables are chosen as $\mathbf{Q}^{1/2}$ and $\mathbf{R}^{1/2}$, then the constraints for maintaining positive definiteness of the matrices can, in principle, be relaxed. However, to avoid ill-conditioning of the optimisation problem during the intermediate gradient steps, it may still be necesary to incorporate suitable upper and lower bounds on the elements of $\mathbf{Q}^{1/2}$ and $\mathbf{R}^{1/2}$.

2.4. Direct solution strategy and parametrisation of ($\mathbf{Q}$,$\mathbf{R}$)

One possibility is to solve the resulting nonlinear optimisation problem by direct application of a standard constrained nonlinear programming strategy such as sequential quadratic programming. It may be noted that, even when the dynamic model is linear, the resulting optimisation problem is nonlinear with respect to the decision variables. The parametrisation of **Q** and **R** has an important bearing on the number of decision variables to be identified. If they are parametrised as full matrices, then the total number of decision variables in the optimisation problem are $[(n(n+1)/2)+(r(r+1)/2)]$. Similarly, a square-root or U-D factorisation of **Q** and **R** also leads to the same number of variables to be identified. Thus, the number of variables to be identified are of the order of (n^2+r^2). Thus, for high order systems, the resulting optimisation problem becomes highly complex in terms of computations. The large number of decision variables also implies that a data set with large N is necessary to reduce the variance errors in the estimates. To reduce the excitation period necessary to generate the identification data set, it can be further assumed that the individual elements of vectors $\mathbf{w}_k$ and $\mathbf{v}_k$ are uncorrelated with each other. This is a reasonable assumption particularly when the noise sources are arising from independent physical variables. With this additional assumption, the decision variables, **Q** and **R** can be parametrised as diagonal matrices. This parametrisation reduces the number of decision variables to $n+r$. Further, it is easy to impose the positivity constraints on the decision variables. Thus, with this modification, the optimisation problem **P1** is reduced to the following problem (**P2**)

$$(\hat{\mathbf{Q}}, \hat{\mathbf{R}}) = arg \min_{(\mathbf{Q},\mathbf{R})} \left[\sum_{i=1}^{N} \log(\det(\Sigma_i)) + \mathbf{e}_i^T \Sigma_i^{-1} \mathbf{e}_i \right] \tag{27}$$

Subject to

$$\text{Model equations (12)–(17)} \tag{29}$$

$$\mathbf{Q} = \mathbf{diag}\begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_{n_w} \end{bmatrix} \tag{30}$$

$$\mathbf{R} = \mathbf{diag}\begin{bmatrix} \gamma_1 & \gamma_2 & \ldots & \gamma_r \end{bmatrix} \tag{31}$$

$$0 < a_L \le \alpha_i \le a_H \quad \text{for} \quad i = 1, 2, \ldots n_w \tag{32}$$

$$0 < b_L \le \gamma_i \le b_H \quad \text{for} \quad i = 1, 2, \ldots r \tag{33}$$

The upper and the lower bounds are imposed to ensure that the optimisation problem does not become numerically ill conditioned during the iterations and the positive definiteness of the matrices is maintained. Such bounds can be arrived at by taking into consideration the nature of the physical variable involved.

*2.5. Choice of initial guess for **Q** and **R***

To solve the above optimisation problem, an initial guess of **Q** and **R** needs to be given to the optimisation algorithm. It is relatively easy to give an initial guess for the measurement error covariance matrix, **R**. Typically, the operating manual of a measuring instrument provides information on the precision of the instrument under some ideal conditions, which can be used as a basis to select the initial guess of the matrix **R**. However, the precision of the instrument is likely to deteriorate with time, for a variety of reasons. Thus, an initial guess, which is suitably higher than the precision reported in the user manual can be used as an initial guess for **R**.

Unlike the measurement noise covariance, specifying an initial guess for matrix **Q** is relatively difficult. For *unstructured noise*, an expected variation of each of state variable and parameter, as percentage of its steady state value, can form a basis for specifying the initial guess for the diagonal elements of matrix **Q**. When the noise is *structured*, and modelled to enter through inputs, the expected variation in the inputs, around the steady state operating conditions, can be used as a basis to specify the initial guess of the diagonal elements of **Q**.

3. Solution using extended expectation maximisation (EM) algorithm

The main difficulty in the direct optimisation based approach to solving the problem **P1** is computational efforts involved in computing the gradients of the objective function with respect to **Q** and **R**. In special circumstances, when $\Gamma_w = \mathbf{I}_n$, the well known expectation maximisation (EM) algorithm in the statistics literature [10] can be modified and extended to be used for identification of **Q** and **R**. In this section, the details of the proposed extended EM algorithm are presented.

3.1. Review of the EM algorithm

In the context of state estimation, the EM algorithm has been developed for estimation of model parameters (Φ, Γ, **C**, **Q**, **R**, $\boldsymbol{\mu}_0$, $\mathbf{P}_0$) of a linear perturbation state space model of the form

$$\mathbf{x}_{k+1} = \Phi\mathbf{x}_k + \Gamma\mathbf{u}_k + \mathbf{w}_k \tag{34}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k \tag{35}$$

from Z_N. Here, $\boldsymbol{\mu}_0 \sim \mathcal{N}(\bar{\mathbf{0}}, \mathbf{P}_0)$ is the initial state of the system. Unlike the optimisation formulation **P1**, which maximises the log likelihood function of the innovation sequence, the EM approach deals with the problem of maximizing the conditional density function of the *complete data set* that includes the process states. Consider the conditional joint probability density function, $p(Z_N, X_N|\Theta)$, of the *complete data set*, $[X_N, Z_N]$, where the $X_N = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ represents the set containing true states of the system. The probability density function of the complete data, can be factorised as

$$p(Z_N, X_N|\Theta) = p(Z_N|\Theta)p(X_N|Z_N, \Theta) \tag{36}$$

Using the definitions of the likelihood functions of $p(Z_N, X_N|\Theta) = L(\Theta|Z_N, X_N)$ and $p(Z_N|\Theta) = L(\Theta|Z_N)$, Eq. (36) can be expressed as follows

$$L(\Theta|Z_N, X_N) = L(\Theta|Z_N)p(X_N|Z_N, \Theta) \tag{37}$$

and the corresponding log-likelihood function can be written as

$$\log L(\Theta|Z_N) = \log L(\Theta|Z_N, X_N) - \log p(X_N|Z_N, \Theta) \tag{38}$$

In practice, however, the set X_N is not available and the resulting maximisation problem is not tractable by the direct optimisation based approach. To solve this problem, Shumway and Stoffer [10] have proposed an iterative approach, called expectation maximisation (EM), which works with the conditional expectation of $\log L(\Theta|Z_N)$ given a parameter guess $\Theta^{(j)}$, i.e.

$$\log L(\Theta|Z_N) = \mathcal{Q}(\Theta, \Theta^{(j)}) - \mathcal{H}(\Theta, \Theta^{(j)}) \tag{39}$$

where

$$\begin{aligned} \mathcal{Q}(\Theta, \Theta^{(j)}) &\equiv E[\log p(Z_N, X_N|\Theta)|Z_N, \Theta^{(j)}] \\ \mathcal{H}(\Theta, \Theta^{(j)}) &\equiv E[\log p(X_N|Z_N, \Theta)|Z_N, \Theta^{(j)}] \end{aligned} \tag{40}$$

For the discrete linear state space model (34) and (35), the joint probability density function of the *complete data set* (X_N, Z_N) can be represented as follows

$$\begin{aligned} p(X_N, Z_N|\Theta) \; &= \frac{1}{2\pi|\mathbf{P}_0|^{1/2}} \exp\left[-\frac{(\mathbf{x}_0-\mu_0)^T \mathbf{P}_0^{-1} (\mathbf{x}_0-\mu_0)}{2} \right] \\ &\times \prod_{i=1}^{N} \frac{1}{(2\pi)^{n/2}|\mathbf{Q}|^{1/2}} \exp\left[-\frac{\mathbf{w}_i^T \mathbf{Q}^{-1} \mathbf{w}_i}{2} \right] \\ &\times \prod_{i=0}^{N} \frac{1}{(2\pi)^{p/2}|\mathbf{R}|^{1/2}} \exp\left[-\frac{\mathbf{v}_i^T \mathbf{R}^{-1} \mathbf{v}_i}{2} \right] \end{aligned} \tag{41}$$

where

$$\mathbf{w}_i = \mathbf{x}_{i+1} - (\Phi\mathbf{x}_i + \Gamma\mathbf{u}_i) \quad (42)$$

$$\mathbf{v}_i = \mathbf{y}_i - \mathbf{C}\mathbf{x}_i \quad (43)$$

and the conditional expectation of $\log L(\Theta|Z_N, X_N)$ given a parameter guess $\Theta^{(j)}$ assumes the following form

$$\begin{aligned}\log L(\Theta|Z_N, X_N) = {} & \log(|\mathbf{P}_0|) + N\log(|\mathbf{Q}|) + N\log(|R|) \\ & + E[tr\{\mathbf{P}_0^{-1}(\mathbf{x}_0 - \mu_0)(\mathbf{x}_0 - \mu_0)^T\}|Z_N] \\ & + E\left\{\sum_{i=1}^{N}[tr\{\mathbf{Q}^{-1}(\mathbf{w}_i\mathbf{w}_i^T)\}|Z_N]\right\} \\ & + E\left\{\sum_{i=0}^{N}[tr\{\mathbf{R}^{-1}(\mathbf{v}_i\mathbf{v}_i^T)\}|Z_N]\right\}\end{aligned} \quad (44)$$

Given a parameter guess $\Theta^{(j-1)}$, let $\hat{X}_N(\Theta^{(j-1)}, Z_N)$ denote the set of smoothed estimates of the states that are generated using the Kalman smoother [20]

$$\hat{X}_N(\Theta^{(j-1)}, Z_N) = \{\hat{x}_{1|N}^{(j-1)}, \ldots, \hat{x}_{N|N}^{(j-1)}\}$$

By EM approach, the problem of maximising $p(X_N, Z_N|\Theta)$ with respect to Θ is solved iteratively in the following steps

- *Expectation (E-step)*: This step involves finding the expected value of the complete data log likelihood function Eq. (B.2), given the observed data set, Z_N and the previously estimated parameter vector, $\Theta^{(j-1)}$, i.e. using $\hat{X}_N(\Theta^{(j-1)}, Z_N)$. Thus, the conditional expectations of $\mathbf{x}_0$, $\mathbf{P}_0$, $\{\mathbf{w}_j\mathbf{w}_j^T\}$, and $\{\mathbf{v}_j\mathbf{v}_j^T\}$ are computed using the smoothed estimates, $\{\hat{x}_{k|N}\}$.
- *Maximisation (M-step)*: This step involves maximising the log likelihood function

$$p[\hat{X}_N(\Theta^{(j-1)}, Z_N), Z_N|\Theta]$$

with respect to the parameter vector Θ, to generate the next iterate $\Theta^{(j)}$.

The parameters estimated in each iteration depend on the observed data, the smoothed state estimates and covariances obtained in the E-step. The iterations are terminated when the change in objective function between two successive iterations is within the specified tolerance. For every successive iteration step of the EM algorithm, it can be shown that the expression

$$\mathcal{H}(\Theta, \Theta^{(j)}) - \mathcal{H}(\Theta^{(j)}, \Theta^{(j)}) \le 0. \quad (45)$$

This is a result of the Jensen's inequality [21]. Since the log function is concave in the region $(0, \infty)$, it follows that $E[\log(x)] \le \log[E(x)]$. For any Θ, we have

$$\begin{aligned}\mathcal{H}(\Theta, \Theta^{(j)}) & = E[\log p(X_N|Z_N, \Theta)|Z_N, \Theta^{(j)}] \\ & \le \log E[p(X_N|Z_N, \Theta)|Z_N, \Theta^{(j)}] \\ & = \mathcal{H}(\Theta^{(j)}, \Theta^{(j)})\end{aligned} \quad (46)$$

Therefore, from Eq. (46), $\mathcal{H}(\Theta, \Theta^{(j)}) - \mathcal{H}(\Theta^{(j)}, \Theta^{(j)}) \le 0$. A detailed discussion on the convergence properties can be found in McLachlan and Krishnan [21] and Wu [22]. Thus, the maximising the function in Eq. (19) is equivalent to choosing $\Theta^{(j+1)}$ such that it maximises $Q(\Theta, \Theta^{(j)})$ conditioned on the observed data or, equivalently, $Q(\Theta^{(j+1)}, \Theta^{(j)}) - \mathcal{Q}(\Theta^{(j)}, \Theta^{(j)}) \ge 0$. The algorithm guarantees decrease in the negative log-likelihood function with successive EM steps.

Though this approach has been developed for the estimation of model parameters (Φ, Γ, $\mathbf{C}$, $\mathbf{Q}$, $\mathbf{R}$, $\boldsymbol{\mu}_0$, $\mathbf{P}_0$) from Z_N it can be easily modified for the scenario when the model matrices (Φ, Γ, $\mathbf{C}$) are known *a priori* and only ($\mathbf{Q}$, $\mathbf{R}$, $\boldsymbol{\mu}_0$, $\mathbf{P}_0$) are unknown.

3.2. Extended EM-algorithm

If it is desired to estimate $(\mu_0, \mathbf{P}_0, \mathbf{Q}, \mathbf{R})$ in the context of EKF, then it becomes necessary to modify the conventional EM algorithm. In this work, it is proposed to extend the EM approach as follows:

- *Expectation (E-step)*: Given an iterate of the parameter $\Theta^{(j-1)}$, the data set $\hat{X}_N(\Theta^{(j-1)}, Z_N)$ is generated using the extended Kalman smoother (EKS). The details of the EKS are given in Appendix A. The computation of $\log L(\Theta|Z_N, X_N)$ using Eq. (44) requires estimation of the conditional expectations of $\{\mathbf{w}_i\mathbf{w}_i^T\}$, and $\{\mathbf{v}_i\mathbf{v}_i^T\}$ using the smoothed estimates, $\{\hat{x}_{k|N}\}$. For the discrete linear system given by Eqs. (34) and (35), it is possible to obtain closed form expressions for these terms in terms of the smoothed estimates. However, for the nonlinear system given by Eqs. (10) and (11), it is not possible to obtain such closed form expressions. The difficulty arises in calculating the conditional expectations of the terms

$$\{\mathbf{w}_i\mathbf{w}_i^T\} = \{[\mathbf{x}_i - \mathbf{F}(\mathbf{w}_{i-1}, \mathbf{u}_{i-1})][\mathbf{x}_i - \mathbf{F}(\mathbf{x}_{i-1}, \mathbf{u}_{i-1})]^T\}$$

and

$$\{\mathbf{v}_i\mathbf{v}_i^T\} = \{[\mathbf{y}_i - \mathbf{h}(\mathbf{x}_i)][\mathbf{y}_i - \mathbf{h}(\mathbf{x}_i)]^T\}$$

In the present work, it is proposed to use approximate expressions for $F(\mathbf{x}_{i-1}, \mathbf{u}_{i-1})$ and $\mathbf{h}(\mathbf{x}_i)$, which are developed using the Taylor series approximation in the neighbourhood of the smoothed estimates, as follows

$$\mathbf{x}_i - \mathbf{F}(\mathbf{x}_{i-1}, \mathbf{u}_{i-1}) \approx \mathbf{x}_i - \left[\mathbf{F}(\hat{\mathbf{x}}_{i-1|N}, \mathbf{u}_{k-1}) + \left[\frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right]_{\mathbf{x}=\hat{\mathbf{x}}_{i-1|N}}(\mathbf{x}_{i-1} - \hat{\mathbf{x}}_{i-1|N})\right] \quad (47)$$

$$\mathbf{y}_i - \mathbf{h}(\mathbf{x}_i) \approx \mathbf{y}_i - \left[\mathbf{h}(\hat{\mathbf{x}}_{i|N}) + \left[\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right]_{\mathbf{x}=\hat{\mathbf{x}}_{i|N}}(\mathbf{x}_i - \hat{\mathbf{x}}_{i|N})\right] \quad (48)$$

Thus, at the jth iteration, the expectation of $-\log L(\Theta|Z_N, X_N)$ conditioned to Z_N and $\Theta^{(j-1)}$ is given as

$$\begin{aligned}& E[-\log L(\Theta|Z_N, X_N)|Z_N, \Theta^{(j-1)}] \\ & = \log|\mathbf{P}_0| + N\log|\mathbf{Q}| + N\log|\mathbf{R}| \\ & + tr[\mathbf{P}_0^{-1}(\mathbf{P}_{0|N}^{(j-1)} + (\mathbf{x}_{0|N}^{(j-1)} - \mu_0)(\mathbf{x}_{0|N}^{(j-1)} - \mu_0)^T)] \\ & + tr[\mathbf{Q}^{-1}(\mathbf{B}_1^{(j-1)} - \mathbf{B}_2^{(j-1)} - (\mathbf{B}_2^{(j-1)})^T + \mathbf{B}_3^{(j-1)})] \\ & + tr[\mathbf{R}^{-1}(\mathbf{B}_4^{(j-1)} - \mathbf{B}_5^{(j-1)} - (\mathbf{B}_5^{(j-1)})^T + \mathbf{B}_6^{(j-1)})]\end{aligned} \quad (49)$$

where matrices $\mathbf{B}_1^{(j-1)}$, $\mathbf{B}_2^{(j-1)}$, $\mathbf{B}_3^{(j-1)}$, $\mathbf{B}_4^{(j-1)}$, $\mathbf{B}_5^{(j-1)}$ and $\mathbf{B}_6^{(j-1)}$ are functions of the observed data and the smoothed state and covariance estimates as follows

$$\mathbf{B}_1 = \sum_{i=1}^{N}[\mathbf{P}_{i|N} + \hat{\mathbf{x}}_{i|N}\hat{\mathbf{x}}_{i|N}^T] \quad (50)$$

$$\mathbf{B}_2 = \sum_{i=1}^{N}[\mathbf{P}_{i,i-1|N}\Phi_{i-1}^T + \hat{\mathbf{x}}_{i|N}\mathbf{F}^T(\hat{\mathbf{x}}_{i-1|N}, \mathbf{u}_{i-1})] \quad (51)$$

$$\mathbf{B}_3 = \sum_{i=1}^{N} \mathbf{F}(\hat{\mathbf{x}}_{i-1|N}, \mathbf{u}_{i-1})\mathbf{F}^T(\hat{\mathbf{x}}_{i-1|N}, \mathbf{u}_{i-1}) + \Phi_{i-1|N}\mathbf{P}_{i-1|N}\Phi_{i-1|N}^T \quad (52)$$

$$\mathbf{B}_4 = \sum_{i=1}^{N} [\mathbf{y}_i\mathbf{y}_i^T] \quad (53)$$

$$\mathbf{B}_5 = \sum_{i=1}^{N} [\mathbf{y}_i\mathbf{h}^T(\hat{\mathbf{x}}_{i|N})] \quad (54)$$

$$\mathbf{B}_6 = \sum_{i=1}^{N} \mathbf{C}_i[\mathbf{P}_{i|N} + \hat{\mathbf{x}}_{i|N}\hat{\mathbf{x}}_{i|N}^T]\mathbf{C}_i^T + \mathbf{h}(\hat{\mathbf{x}}_{i|N})\mathbf{h}^T(\hat{\mathbf{x}}_{i|N}) \quad (55)$$

and are computed at using $\hat{X}_N(\Theta^{(j-1)}, Z_N)$. The detailed derivations are given in Appendix B. It may be noted that the term $P_{i,i-1|N}$ is called the lag-one covariance smoother and is given by

$$\begin{aligned} \mathbf{P}_{N,N-1|N} &= \{\mathbf{I} - \mathbf{K}_N\mathbf{C}_N\}\Phi_{N-1}\mathbf{P}_{N-1|N-1} \quad \forall k = N \\ \mathbf{P}_{i,i-1|N} &= \mathbf{P}_{i|i}\mathbf{J}_{i-1}^T + \mathbf{J}_i\{\mathbf{P}_{i+1,k|N} - \Phi_i\mathbf{P}_{i|i}\}\mathbf{J}_{i-1}^T \\ &\forall i = (N-1, N-2, \ldots, 1) \end{aligned} \quad (56)$$

- *Maximization (M-Step)*: Taking the partial derivative of $E[-\log L(\Theta|Z_N, X_N)|Z_N, \Theta^{(j-1)}]$ with respect to **Q** and **R**, and setting it to zero yields the following relation

$$\mathbf{Q}^{(j)} = \frac{1}{N}[\mathbf{B}_1^{(j-1)} - \mathbf{B}_2^{(j-1)} - (\mathbf{B}_2^{(j-1)})^T + \mathbf{B}_3^{(j-1)}] \quad (57)$$

$$\mathbf{R}^{(j)} = \frac{1}{N}[\mathbf{B}_4^{(j-1)} - \mathbf{B}_5^{(j-1)} - (\mathbf{B}_5^{(j-1)})^T + \mathbf{B}_6^{(j-1)}] \quad (58)$$

Since the mean and covariance of the initial state cannot be identified simultaneously, they are conventionally fixed as follows [10]

$$\boldsymbol{\mu}_0^{(j)} = \hat{\mathbf{x}}_{0|N}^{(j-1)} \quad \mathbf{P}_0^{(j)} = \mathbf{P}_{0|N}^{(j-1)} \quad (59)$$

The iterations are terminated when the following criterion is satisfied

$$\frac{\left|\psi^{(j)} - \psi^{(j-1)}\right|}{\left|\psi^{(j)}\right|} < \varepsilon$$

where ε is a small positive number and

$$\psi^{(j)} = -\left[\sum_{k=1}^{N} \log\det(\Sigma_k^{(j)}) + \mathbf{e}_k^T(\Sigma_k^{-1})^{(j)}\mathbf{e}_k\right]$$

Thus, for the case when the noise can be treated as zero-mean additive white noise with a Gaussian distribution, explicit analytical expressions of the derivatives are available in the M-step. The modified EM approach listed in this section is referred to as extended EM algorithm in the rest of the text.

3.3. Extended EM algorithm for multi-rate data

Let $\{c_1, c_2, \ldots, c_{N_1}\}$ be a subsequence the sampling instances of $\{0, 1, \ldots N\}$ when only a part of the measurements are available and $\{l_1, l_2, \ldots, l_{N_2}\}$ be the subsequence of the remaining instances when all measurements are available. In the E-step of the extended EM algorithm, the smoothed estimates of the states and covariances obtained from the multi-rate EKF and EKS are used. The expressions for $E(\mathbf{v}_i\mathbf{v}_i^T|Z_N)$ change as follows

$$\begin{aligned} \sum_{i=1}^{N} E(\mathbf{v}_i\mathbf{v}_i^T|Z_N) &= \sum_{i=1}^{N} E[(\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i))(\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i))^T|Z_N] \\ &= \sum_{i=1}^{N_2}\left[(\mathbf{y}_{l_i} - \mathbf{h}_{l_i}(\hat{\mathbf{x}}_{l_i|N}))(\mathbf{y}_{l_i} - \mathbf{h}_{l_i}(\hat{\mathbf{x}}_{l_i|N}))^T + \mathbf{C}_{l_i}\mathbf{P}_{l_i|N}\mathbf{C}_{l_i}^T\right] + \sum_{i=1}^{N_1}\mathbf{R}_{c_i} \end{aligned} \quad (60)$$

where $\mathbf{R}_{c_i}$ is a matrix, which is computed as follows

- if the sth and tth elements of the measurement vector are available at instant j_i

$$\mathbf{R}_{c_i}(s,t) = (\mathbf{y}_{c_i,s} - \mathbf{h}_{c_i,s}(\hat{x}_{j_i|N}))(\mathbf{y}_{c_i,t} - \mathbf{h}_{c_i,t}(\hat{x}_{c_i|N}))^T + \mathbf{C}_{c_i,s}\mathbf{P}_{c_i|N}\mathbf{C}_{c_i,t}^T$$

- if either of the two is not available

$$\mathbf{R}_{c_i}(s,t) = \mathbf{R}^{(j-1)}(s,t)$$

The superscript $(j-1)$ implies that the value of the (s,t) element of **R** that is obtained in the previous iteration of the extended EM-algorithm.

3.4. Comparison of the gradient-based optimisation approach and EM-based approach

The conventional optimisation approach uses gradient-based methods to estimate the parameters $\Theta = (\mathbf{Q}, \mathbf{R})$. This involves computation of the gradients of the objective function and the constraints with respect to Θ. Owing to the nonlinear nature of the MLE objective function (Eq. (23)) with respect to Θ, the evaluation of the gradients might become computationally intensive. The extended EM-algorithm, on the other hand, is a derivative free algorithm as the algorithm does not involve computation of gradients of the objective function. It has been shown that the objective function in the EM algorithm, decreases with every iteration. The convergence properties [22] guarantee convergence to a local optimum for unimodal likelihood functions.

The extended EM-algorithm computes the next iterate of Θ based on the full data likelihood function (Eq. (41)), which is computed using the weighted two norms of the state and measurement residuals. If the process noise is modelled as per Assumption 4(a) in Section 2.1, ($\Gamma_{w,k} = \mathbf{I}_n$, Eq. (10)), the two-norm of $\mathbf{w}_k$ is weighted with the inverse of its covariance, **Q**. The existence of $\mathbf{Q}^{-1}$ is guaranteed, as **Q** is positive definite. If the process noise is modelled to enter through other known sources, as per Assumption 4(b) and (c), ($\Gamma_{w,k} \neq \mathbf{I}_n$), then the two-norm of the residuals is weighted with $(\Gamma_{w,k}\mathbf{Q}_w\Gamma_{w,k}^T)^{-1}$. It is not possible to uniquely identify $\mathbf{Q}_w$ from $(\Gamma_{w,k}\mathbf{Q}_w\Gamma_{w,k}^T)$ as the matrix $\Gamma_{w,k}$ varies with time. Hence, the use of the extended EM-algorithm is limited to cases where the noise is modelled as per Assumption 4(a). On the other hand, gradient-based optimisation algorithm computes the next iterate of Θ directly from the derivative of the objective function (Eq. (23)) and does not require the computation of the weighted two norm of $\mathbf{w}_k$. Therefore, this approach can be used to estimate the covariance of noise entering the process through inputs apart from additive noise.

4. Simulation study

In this section, application of the proposed methods to estimate the noise covariances is demonstrated on a continuous fermenter [23], which is a benchmark nonlinear system. The model equations are described as follows

$$\begin{aligned} \dot{X} &= -DX + \mu X \\ \dot{S} &= D(S_f - S) - \frac{1}{Y_{X|S}}\mu X \\ \dot{P} &= -DP + (\alpha\mu + \beta)X \end{aligned} \tag{61}$$

where X represents the yeast(cell-mass) concentration, S represents the glucose(substrate) concentration and P represents the alcohol(product) concentration. μ represents the specific growth rate, $Y_{X|S}$ is the cell-biomass yield, α and β are yield parameters for the product. The specific growth rate model is assumed to exhibit both substrate and product inhibition:

$$\mu = \frac{\mu_m(1 - (P/P_m))S}{K_m + S + (S^2/K_i)} \tag{62}$$

where μ_m is the maximum specific growth rate, P_m is the product saturation constant, K_m is the substrate saturation constant and K_i is the substrate inhibition constant. The substrate concentration, S and product concentration P are assumed to be available as measurements. The dilution rate, D, and feed substrate concentration, S_f, are available as manipulated inputs. The nominal parameters and operating conditions are given in the work by Henson and Seborg [23].

The following two scenarios were simulated—(a) the process noise with an unknown source, and (b) unknown disturbances in the manipulated inputs. To simulate the process dynamics, the sampling time (T) was chosen as 0.25 h. Input–output data was generated by subjecting the manipulated inputs to a pseudo-random binary signal (*PRBS*). The amplitude of perturbation in D is 0.015 h^{-1} (in the frequency range $\begin{bmatrix} 0 & 0.02\omega_N \end{bmatrix}$, where ω_N is the Nyquist frequency) and in S_f it is 2 g/l (in the frequency range $\begin{bmatrix} 0 & 0.016\omega_N \end{bmatrix}$). The data is collected for 500 h of running of the fermenter, which corresponds to 2000 samples of data. For scenario (a), a zero-mean Gaussian white noise was added to the states and for scenario (b), a zero-mean Gaussian white noise was added to the manipulated inputs to simulate the effect of process noise. To simulate the effect of measurement noise, a zero-mean Gaussian white noise signal was added to the measurements. For both the scenarios, the true values of the covariances used are reported in Tables 1 and 5.

To simulate the irregularly sampled multi-rate scenario, it was assumed that the measurements of product concentration, P, are available at every sampling instant, while the measurements of substrate concentration, S, are available at irregular sampling intervals, which are integer multiples of the sampling time, T. The maximum delay between any two successive samples of S is three.

4.1. Case 1: state disturbance with unknown source

The estimates of **Q** and **R** were obtained using both the proposed approaches, i.e. the extended EM algorithm and the direct optimisation (using *fmincon* function in Matlab). The criteria used for termination of iterations were chosen to be identical for both the approaches. The initial guess of the process noise covariance is chosen as discussed in Section 2.5 and is reported in Table 1. It may be noted that the initial guesses of **Q** and **R** are chosen significantly different from the true values. To give a better insight into the choice of the initial guesses, the standard deviations of the true process noise (i.e. square root of the diagonal elements of **Q**) and initial guesses for the standard deviations as a percentage of the steady state of the process are given in Table 2.

Table 1
Unknown process noise case: true values, initial guess and estimated values of **Q** and **R**.

	True Covariance		Initial Guess
Q	$10^{-3}\times \mathbf{diag}\begin{bmatrix} 11.1 & 0.7 & 139.2 \end{bmatrix}$	$\mathbf{Q}^{(0)}$	$10^{-3}\times \mathbf{diag}\begin{bmatrix} 1200 & 60 & 7500 \end{bmatrix}$
R	$10^{-3}\times \mathbf{diag}\begin{bmatrix} 5.6 & 15.6 \end{bmatrix}$	$\mathbf{R}^{(0)}$	$10^{-3}\times \mathbf{diag}\begin{bmatrix} 10 & 30 \end{bmatrix}$
	Estimated Covariances		
	Single Rate		**Multi Rate**
	Extended EM Algorithm		
$\widehat{\mathbf{Q}}$	$10^{-3}\times\begin{bmatrix} 11.2 & -0.4 & 0.9 \\ -0.4 & 0.6 & -1.4 \\ 0.9 & -1.4 & 151.0 \end{bmatrix}$		$10^{-3}\times\begin{bmatrix} 10.7 & -0.2 & 0.6 \\ -0.2 & 0.7 & -1.2 \\ 0.6 & -1.2 & 150.6 \end{bmatrix}$
$\widehat{\mathbf{R}}$	$10^{-3}\times\begin{bmatrix} 5.6 & 0 \\ 0 & 15.2 \end{bmatrix}$		$10^{-3}\times\begin{bmatrix} 5.5 & 0 \\ 0 & 15.4 \end{bmatrix}$
	Direct Optimisation		
$\widehat{\mathbf{Q}}$	$10^{-3}\times \mathrm{diag}\begin{bmatrix} 12.3 & 0.4 & 143.2 \end{bmatrix}$		$10^{-3}\times \mathrm{diag}\begin{bmatrix} 12.5 & 0.3 & 142.9 \end{bmatrix}$
$\widehat{\mathbf{R}}$	$10^{-3}\times \mathrm{diag}\begin{bmatrix} 5.7 & 20.3 \end{bmatrix}$		$10^{-3}\times \mathrm{diag}\begin{bmatrix} 5.7 & 20.5 \end{bmatrix}$

For the single rate measurements case, the estimated optimum values of **Q** and **R** matrices obtained using both the approaches, are presented in Table 1. The estimates of **Q**, obtained from both the algorithms, are close to the true values within an acceptable range, while the estimates of **R** match the true values closely. Fig. 1(a) reports the change in log likelihood function (Eq. (23)) as a function of the iteration count in the E-step. This figure also plots the value of the log likelihood function obtained when the true values of **Q** and **R** are used in EKF, which is referred to as the *ideal value* of the log likelihood function in the remaining text. It may be noted that the estimates of the log likelihood function generated by the extended EM algorithm monotonically converge to its ideal value for the regularly sampled data case. Fig. 1(b) reports the change in the value of the log likelihood with every iteration using the direct optimisation approach. In this case too, the value of the log likelihood function converges to its ideal value.

Using the matrices **Q** and **R** that were identified from data, the performance of the EKF was also evaluated on a separate validation data set, using the sum of squared values of the estimation errors (SSE) as a metric for comparison. In Table 3, the SSE values of the EKF using the estimated covariance matrices, are compared with

Table 2
Standard deviations of true process noise and initial guess as % of steady state.

State	Steady State	σ of true process noise	σ of initial guess
X	7.038	1.5	15
S	2.404	1.1	10
P	24.869	1.5	11

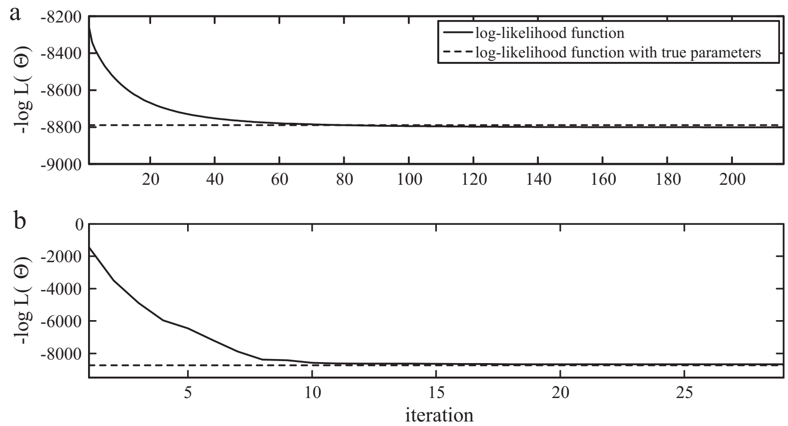

Fig. 1. Single rate case: log-likelihood function with iteration. (a) Extended EM algorithm and (b) direct optimisation method.

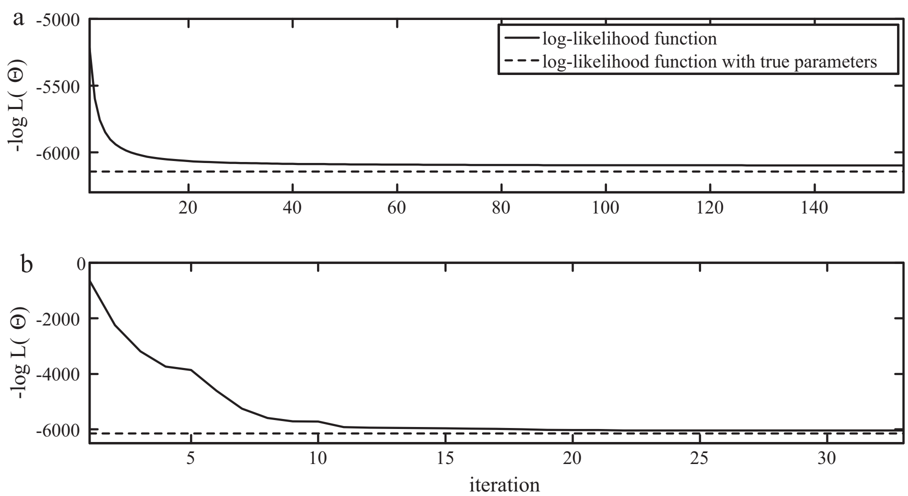

Fig. 2. Multi-rate sampling case: log-likelihood function with iteration. (a) Extended EM algorithm and (b) direct optimisation method.

the corresponding values obtained using the initial guess and the true values of the covariance matrices. When the estimated optimum values of **Q** and **R** are used, the state estimation errors are significantly smaller. The improvements obtained using the estimated optimum values are more pronounced in the case of the unmeasured state, i.e. biomass concentration (X). The autocovariance function plot of the innovations of product concentration, P (for the single-rate case, using the estimated values of **Q** and **R**)

Table 3
Unknown process noise case: SSE values of the estimation errors.

Parameters	ε_X	ε_S	ε_P
Extended EM algorithm			
Q and **R** matrices	Single	Rate	Case
Initial guess	43.158	2.802	9.815
True	36.243	1.262	7.454
Estimated (extended EM)	36.653	1.261	7.241
Estimated (direct optimisation)	36.513	1.271	7.938
Parameters	ε_X	ε_S	ε_P
Extended EM algorithm			
Q and **R** matrices	Multi	Rate	Case
Initial guess	46.916	3.237	9.815
True	37.075	1.774	7.454
Estimated (extended EM)	36.971	1.779	7.445
Estimated (direct optimisation)	41.722	2.085	7.976

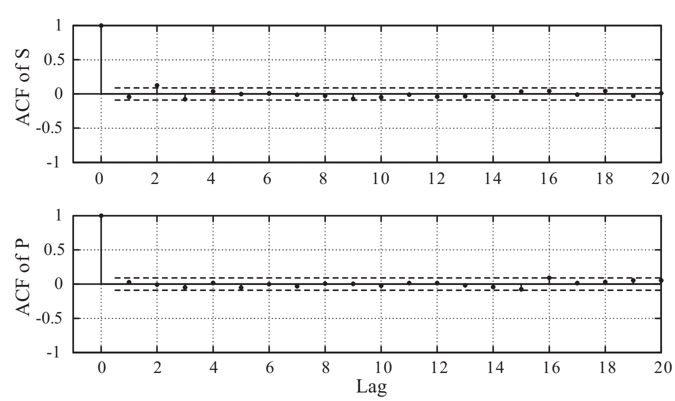

Fig. 3. The sample autocorrelation function of the innovations of S and P.

is shown in Fig. 3. The plot indicates that the innovations are an uncorrelated white noise sequence, which supports Assumption 4(a).

For the multi-rate case too, the estimates of **Q** and **R** obtained from both the approaches are accurate within an acceptable range. The values of the estimates are reported in Table 1. However, the performance of both the methods, extended EM algorithm and direct optimisation method, deteriorates for the multi-rate data case as is evident from Fig. 2, which plots the change in the

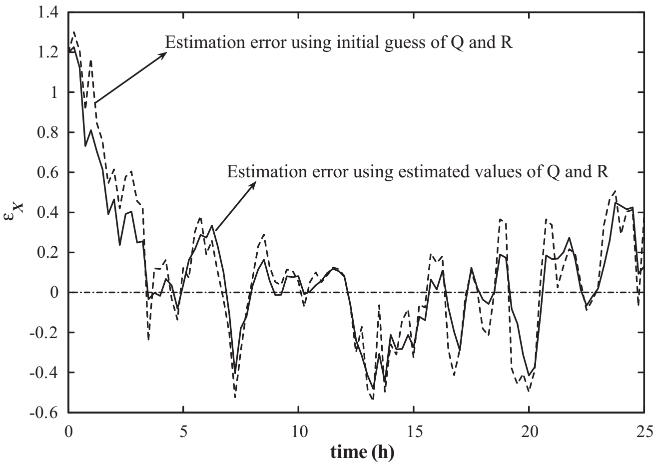

Fig. 4. Unknown source of disturbances case: comparison of initial transients of estimation errors obtained using the initial guess and the estimated parameters from extended EM algorithm for multi-rate EKF.

log-likelihood function with every iteration. The estimates of the log-likelihood function do not converge to its ideal value. The SSE values of the state estimation error, reported in Table 3, indicate that there is improvement in performance of the EKF, especially in case of the unmeasured state, biomass concentration (X) and the irregularly measured state, glucose concentration (S). A comparison of the estimation error in biomass concentration (X), for the estimates obtained using the initial guess of parameters ($\mathbf{Q}_0$,$\mathbf{R}_0$) and the estimated values of ($\mathbf{Q}$,$\mathbf{R}$) are shown in Fig. 4. The SSE values in Table 3 indicate that the performance of the EKF implemented with estimated parameters, is close to that of the EKF with true parameters.

4.2. Case 2: unknown disturbance in manipulated inputs

For this case, the process was simulated with unknown white noise disturbances in the manipulated inputs. As mentioned earlier, the use of the extended EM algorithm is restricted to cases where the process noise sources are unknown. Therefore, for this case, the estimates of $\mathbf{Q}$ and $\mathbf{R}$, can be obtained using only the direct optimisation approach. The true values and the initial guess of the input noise covariance used for initialising the optimisation is given in Table 5. The initial guess is chosen in accordance with the procedure mentioned in Section 2.5. The standard deviations of the true noise and initial guess of noise, as a percentage of the steady state of the inputs, is given in Table 4. For the single rate case, the optimum values of $\mathbf{Q}$ and $\mathbf{R}$ obtained from the direct optimisation approach are given in Table 5. As is evident from the table, the estimate of $\mathbf{Q}$ is close to the its true value.

The SSE values of the estimation errors using the EKF with estimated values of $\mathbf{Q}$ and $\mathbf{R}$ were compared with the corresponding values obtained using the true parameters on a different validation data. The comparison of the SSE values is presented in Table 6. There is a marked reduction in the SSE values of the unmeasured state, X, when the estimated values of ($\mathbf{Q}$,$\mathbf{R}$) are used.

For the case of multi-rate measurements too, the algorithm gives estimates which are close to the true values of the covariances. The estimated values of ($\mathbf{Q}$,$\mathbf{R}$) are given in Table 5. A graphical comparison of the estimation errors in X for the EKF using the initial guess ($\mathbf{Q}_0$,$\mathbf{R}_0$) and the estimated values of ($\mathbf{Q}$,$\mathbf{R}$), in the multi-rate case is shown in Fig. 5. The SSE values in Table 6 and the graphical comparison in Fig. 5 indicate that the performance of the EKF with the estimated values of ($\mathbf{Q}$,$\mathbf{R}$) is close to that with their true values. In particular, it is important to note that there is a considerable reduction in SSE of the unmeasured state X and the irregularly sampled state S. When the estimated ($\mathbf{Q}$,$\mathbf{R}$) are used, the error dynamics of the unmeasured variable, X, quickly converge to zero as compared to the error dynamics when the initial guess ($\mathbf{Q}^{(0)}$, $\mathbf{R}^{(0)}$) are used.

Table 4
Standard deviations of true input noise and initial guess as % of steady state.

Input	Steady state	σ of true noise	σ of initial guess
D	0.15	4.6	15
S_f	20	5.5	10

Table 5
True values, initial guess and estimated values of $\mathbf{Q}$ and $\mathbf{R}$ for input process noise.

	True covariance		Initial guess
$\mathbf{Q}$	diag[4.9×10^{-5} 1.21]	$\mathbf{Q}^{(0)}$	diag[5×10^{-4} 4]
$\mathbf{R}$	$10^{-3}\times$ diag[5.6 15.6]	$\mathbf{R}^{(0)}$	$10^{-3}\times$ diag[10 30]
Estimated covariances			
	Single rate		Multi rate
$\hat{\mathbf{Q}}$	diag[4.91×10^{-5} 1.201]		diag[4.88×10^{-5} 1.197]
$\hat{\mathbf{R}}$	$10^{-3}\times$ diag[5.69 48.31]		$10^{-3}\times$ diag[5.87 48.37]

Table 6
SSE values of estimation error: input noise case.

Parameters	ε_X	ε_S	ε_P
Single rate case			
Initial Guess	7.244	2.574	13.926
True	2.745	1.325	7.430
Estimated	2.772	1.323	7.194
Multi rate case			
Initial Guess	7.283	5.321	16.350
True	1.852	1.641	8.517
Estimated	1.846	1.638	8.239

5. Experimental study

The covariance matrices identification exercise was carried out using the experimental data obtained from the benchmark heater-mixer setup [24] available at Automation Laboratory, Department of Chemical Engineering, Indian Institute of Technology Bombay. The results of verification on the experimental setup are presented in this section.

5.1. The heater-mixer system

The heater-mixer setup consists of two tanks in series as shown in Fig. 6. The contents in the tank are well agitated using vari-

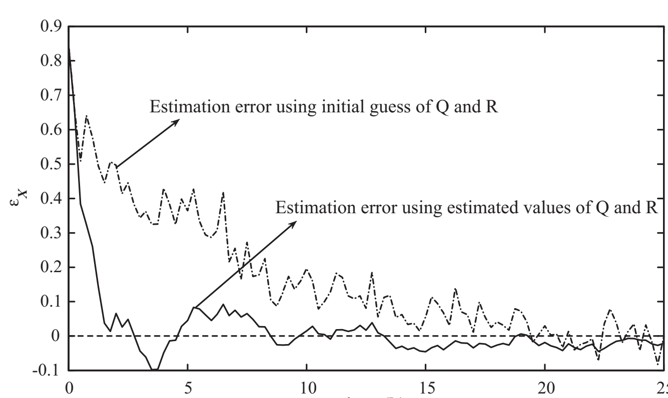

Fig. 5. Noise in inputs case: comparison of initial transients of estimation errors using the initial guess and estimated parameters for multi-rate EKF.

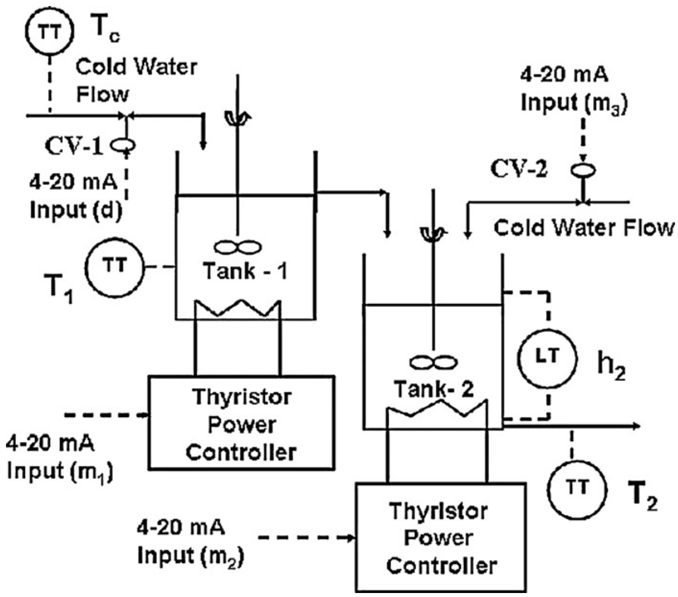

Fig. 6. Schematic of the heater-mixer setup.

able speed agitators. A cold water stream is introduced in the first tank. The water in Tank-1 is heated using 4 kW heating coil. The hot water that overflows the first tank is mixed with cold water stream entering into Tank-2. The water in the second tank is heated using another 3 kW heating coil. The heat inputs Q_1 and Q_2 to both the tanks can be manipulated by changing current inputs to the thyristor power control units, which can be varied between 4 and 20 mA. The cold water inlet flow F_1 and F_2 to both the tanks can be manipulated using pneumatic control valves. The temperatures in the two tanks (T_1 and T_2) and level in second tank (h_2) are measured variables while the heat inputs to Tank-1 (Q_1) and Tank-2 (Q_2) and cold water inlet flowrates to Tank-1 (F_1) and Tank-2 (F_2) are treated as manipulated inputs. The system is interfaced with a computer using a data acquisition module (Advantech 5000 series hardware) and combination of LabVIEW and MATLAB. A grey box model was developed for the system [15,25], which is described by the following differential equations

$$V_1\frac{dT_1}{dt} = F_1(D)(T_c - T_1) + \frac{\alpha_1 Q_1(U_1)}{\rho C_p} \tag{63}$$

$$A_2 h_2\frac{dT_2}{dt} = F_1(D)(T_1 - T_2) + F_2(U_3)(T_c - T_2) + \frac{1}{\rho C_p}[\alpha_2 Q_2(U_2) - 2\pi r_2 h_2 U(T_2 - T_a)] \tag{64}$$

$$A_2\frac{dh_2}{dt} = F_1(D) + F_2(U_3) - 10^{-3} \times k\sqrt{h_2} \tag{65}$$

The flows (F_1 and F_2) are modelled as a function of the % of the input current range (D and U_3) to the respective control valves. The heat terms (Q_1 and Q_2) are modelled as % of the input current range (U_1 and U_2) to the respective thyristor power control modules. The algebraic equations that relate the model inputs to the mA signals are given in Appendix C. The parameters α_1 and α_2 are deliberately introduced to account for changes in the heat transfer correlations due to variations in the heat loss to the atmosphere from Tank-1 and Tank-2, respectively. These parameters are modelled as integrated white noise sequences with unknown variances. The cold water temperature (T_c) is a measured disturbance for this system and its variation is also modelled as an integrate white noise sequence.

$$\alpha_{1,k+1} = \alpha_{1,k} + w_{\alpha_1,k} \tag{66}$$

$$\alpha_{2,k+1} = \alpha_{2,k} + w_{\alpha_2,k} \tag{67}$$

$$T_{c,k+1} = T_{c,k} + w_{T_c,k} \tag{68}$$

where $w_{\alpha_1,k} \sim \mathcal{N}(0, q_{\alpha_1})$, $w_{\alpha_2,k} \sim \mathcal{N}(0, q_{\alpha_2})$ and $w_{T_c,k} \sim \mathcal{N}(0, q_{T_c})$. The augmented state vector that is used for state and parameter estimation is

$$x = \begin{bmatrix} T_1 & T_2 & h_2 & T_c & \alpha_1 & \alpha_2 \end{bmatrix}^T$$

The following states are available as measurements

$$\mathbf{y} = \begin{bmatrix} T_1 & T_2 & h_2 & T_c \end{bmatrix}$$

The nominal operating conditions and model parameters are given in Table 7.

The sampling time (T) for the experiment was chosen as 5 s. To stabilise the level dynamics, the level loop in Tank-2 was closed using a PI controller. To generate data for identification, the system was perturbed over the entire range of the heat inputs (Q_1 and Q_2). A PRBS of amplitude 2 mA was introduced in D (to manipulate F_1) and in the set-point for h_2. The controller output (U_3) was available and recorded at every sampling instant. The variations in the inputs are shown in Fig. 7. The corresponding variations in the outputs of the system are shown in Fig. 8. Note that the variation in F_2 is in response to change in set-point to Tank-2 level (h_2).

5.2. The covariance estimation problem

The estimation problem is formulated as a state and parameter estimation problem. The state disturbance can be assumed to be from an unknown source or as entering through the flow inputs F_1 and F_2. In this work, both the scenarios have been considered. In order to evaluate the proposed approaches in a multi-rate measurements scenario, samples of the temperature measurements, T_1, were dropped randomly. The maximum delay between two measurements of T_1 is restricted to three sampling instants.

5.2.1. Case 1: state disturbance with unknown source

The covariances of the state, parameter and measurement noise are estimated from the perturbation data using the extended EM-algorithm as well as the direct optimisation approach. The initial guesses of $\mathbf{Q}$ (for state and parameters) and $\mathbf{R}$ are given in Table 8. The standard deviation in initial guess of process noise as a percentage of the steady state is given in Table 9. It may be noted that, for a real plant, it is difficult to know the true values of the state noise covariances for comparison. However, the initial guess used to initialise both the algorithms, is deliberately selected significantly larger from the expected range of values. This can be

Table 7
Experimental heater-mixer setup: parameters and nominal operating conditions.

Variable	Description	Value
k	Discharge coefficient	0.1
V_1	Volume of Tank-1	1.75×10^{-3} m^3
A_2	Cross sectional area of Tank-2	7.854×10^{-3} m^2
r_2	Radius of Tank-2	0.05 m
U	Heat transfer coefficient	175 W/m K
T_a	Ambient temperature	301 K
T_1	Temperature of Tank-1	58.6 °C
T_2	Temperature of Tank-2	50.8 °C
h_2	Level in Tank-2	0.38 m
U_1	Input to heater-1	10 mA
U_2	Input to heater-2	10 mA
U_3	Input to valve-2	11 mA
D	Input to valve-1	9 mA
T_c	Cold water temperature	33.8 °C

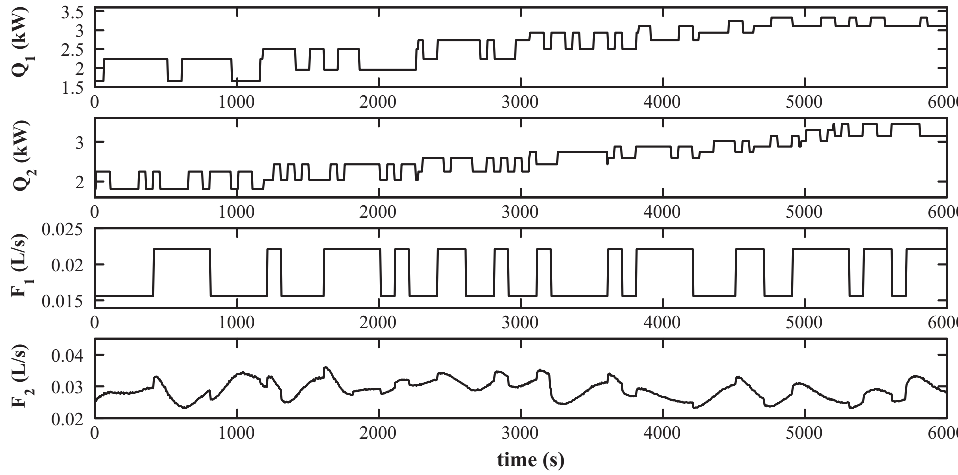

Fig. 7. Heater-mixer setup: variation of input variables.

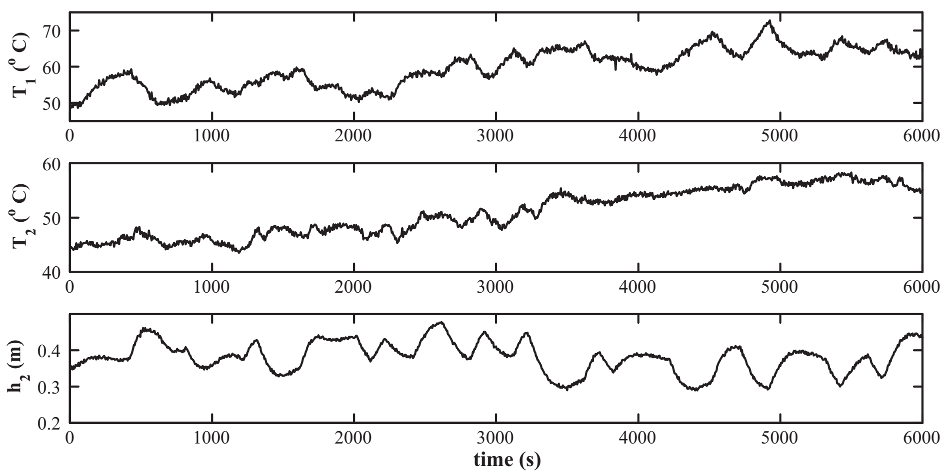

Fig. 8. Heater-mixer setup: variations in the output.

ascertained from standard deviation of the process noise obtained from the estimated values of **Q**. These values are reported in Table 9.

For the direct optimisation approach, **Q** and **R** are parametrised as diagonal matrices. The parameters α_1 and α_2 are expected to settle to some mean steady state as time progresses. This information about the steady state values of α_i is propagated to the next iteration of the optimisation algorithms. The extended EM algorithm returns the smoothed estimates of the initial values of the parameters α_1 and α_2. In the direct optimisation approach, the mean of the last 200 samples from the trajectories of α_i were used as the initial values of α_1 and α_2 for the next iteration. For the single rate case, the optimum values of **Q** and **R** obtained from both the methods, are presented in Table 8.

The sample variance of the measurement errors were estimated independently using steady state values of the measurements [25]. The variance in T_1 was 0.2172 ($^\circ$C^2), for T_2 it was 0.1351 ($^\circ$C^2) and for h_2 it was 1.156×10^{-5} (m^2). When compared with the diagonal elements of **R**, which represent the measurement error variance, the estimated variance in T_1 matches closely, while those of T_2 and h_2 are close in the order of magnitude range. The performance of the EKF was evaluated on a separate validation data set, using the identified values of **Q** and **R**. For the multi-rate case, the predictions of the EKF for Tank-1 temperature (T_1), using the estimated values of **Q** and **R** from the extended EM algorithm are shown in Fig. 9. The estimates of the Tank-2 temperature are plotted in Fig. 10, while those of the Tank-2 level are plotted in Fig. 11. The mean of the last

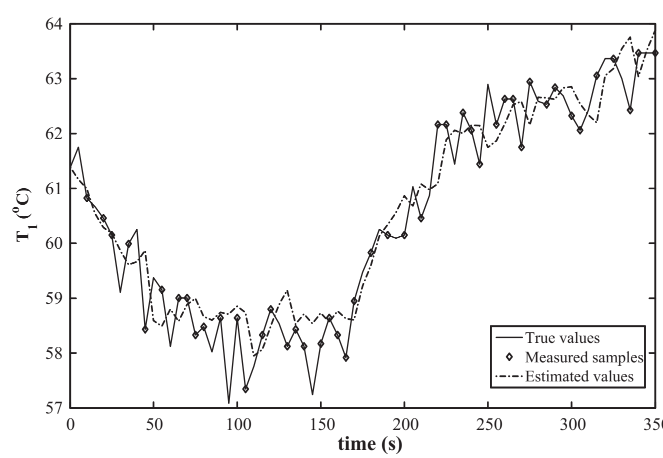

Fig. 9. Heater-mixer setup, state noise from unknown source, multi-rate case: Tank-1 temperature: comparison of the EKF predictions with actual output.

200 samples of the trajectories of α_i, generated in the final iteration step were used as the initial values of the parameters α_1 and α_2 for the validation exercise. The variations of the parameters α_1 and α_2 are shown in Fig. 12. α_1 has a mean of 0.704 and a variance of 8.215×10^{-4}, while α_2 has a mean of 0.700 and a variance of 4.055×10^{-3}. The sum of squared values of the innovations ($\mathbf{e}_k$),

Table 8
Heater-mixer setup, single rate case, state noise from unknown source: Initial guess and estimated values of **Q** and **R**.

Initial Guess of Q and R	
$\mathbf{Q}^{(0)}$	$10^{-3}\times\text{diag}\begin{bmatrix}8000 & 9000 & 1.5 & 4000 & 40 & 50\end{bmatrix}$
$\mathbf{R}^{(0)}$	$10^{-3}\times\text{diag}\begin{bmatrix}500 & 15 & 60 & 20\end{bmatrix}$
Estimated values of Q and R	
Extended EM algorithm	
$\widehat{\mathbf{Q}}$	$10^{-3}\times\begin{bmatrix}137.494 & 0.156 & -0.690 & 0.775 & 0.200 & 1.052\\ 0.156 & 44.388 & -0.018 & 0.782 & -0.133 & 0.181\\ -0.690 & -0.018 & 0.009 & 0.005 & -0.002 & -0.006\\ 0.775 & 0.782 & 0.005 & 3.564 & -0.100 & -0.098\\ 0.200 & -0.133 & -0.002 & -0.100 & 0.022 & 0.043\\ 1.052 & 0.181 & -0.006 & -0.098 & 0.043 & 0.208\end{bmatrix}$
$\widehat{\mathbf{R}}$	$10^{-3}\times\begin{bmatrix}177.598 & -32.410 & 0.212 & 70.117\\ -32.410 & 69.587 & 0.024 & -5.949\\ 0.212 & 0.024 & 0.004 & 0.040\\ 70.117 & -5.949 & 0.040 & 147.081\end{bmatrix}$
Direct Optimisation	
$\widehat{\mathbf{Q}}$	$10^{-3}\times\text{diag}\begin{bmatrix}121.9 & 54.42 & 0.004 & 3.711 & 0.002 & 0.006\end{bmatrix}$
$\widehat{\mathbf{R}}$	$10^{-3}\times\text{diag}\begin{bmatrix}234.2 & 76.937 & 0.005 & 147.6\end{bmatrix}$

presented in Table 10, show a pronounced improvement in the predictions of the EKF when the optimal values of **Q** and **R** are used, in comparison with the initial guess of **Q** and **R**.

For the multi-rate case, the estimates of **Q** and **R**, obtained from both the approaches are given in Table 11. The values of **Q** and **R** are similar to those obtained from the single rate case. For a separate validation data, with irregular samples of T_1, the SSE values of the innovations are compared in Table 10. The SSE values indicate a significant improvement in the performance of the EKF when the estimated values of **Q** and **R** are used. This is especially significant in the case of the irregularly sampled state T_1.

Table 9
Heater-mixer setup: standard deviation of initial guess and estimated value of process noise as % of the steady state.

State	Steady state	σ initial guess	σ estimated covariance
T_1	58.6 °C	5	1.6
T_2	50.8 °C	6	0.46
h_2	0.38 m	10	0.54
T_c	33.8 °C	6	0.20
α_1	1	20	0.5
α_2	1	20	1.4

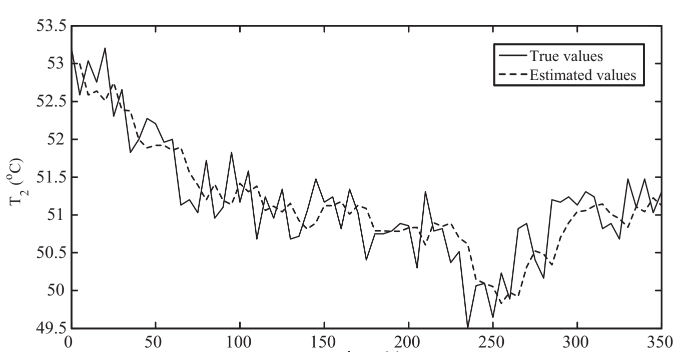

Fig. 10. Heater-mixer setup, state noise from unknown source, multi-rate case: Tank-2 temperature: comparison of the EKF predictions with actual output.

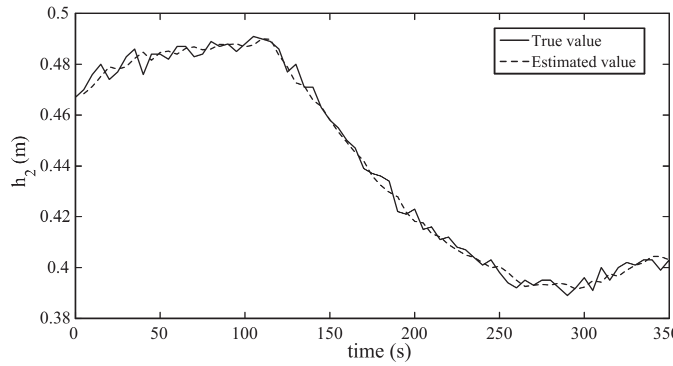

Fig. 11. Heater-mixer setup, state noise from unknown source, multi-rate case: Tank-2 level: comparison of the EKF predictions with actual output.

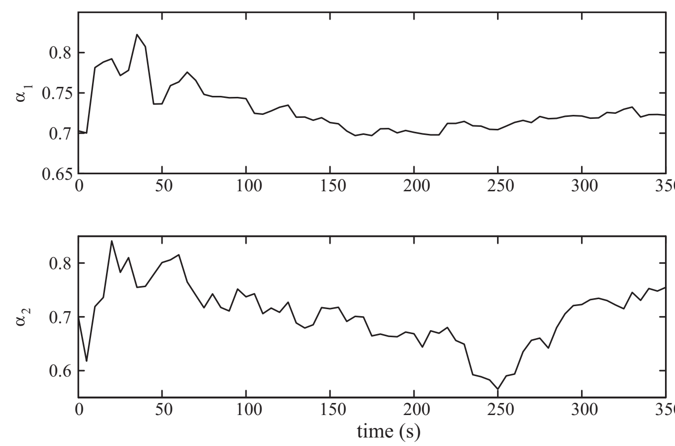

Fig. 12. Heater-mixer setup, noise from unknown source, multi-rate case: variation in α_1 and α_2, for estimated covariances from extended EM algorithm.

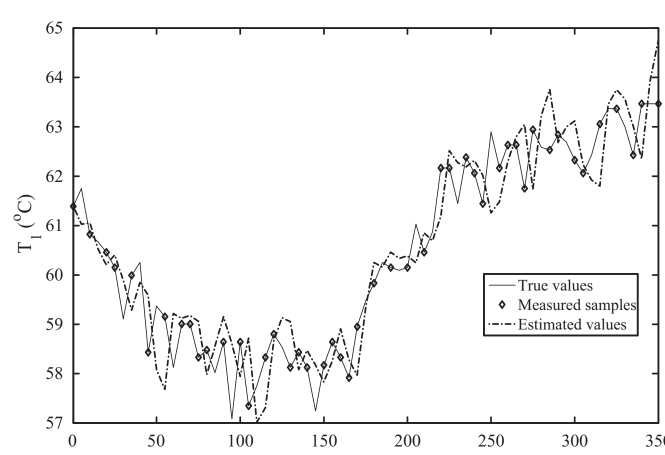

Fig. 13. Heater-mixer setup, input noise multi-rate case: Tank-1 temperature: comparison of EKF predictions with actual measurements.

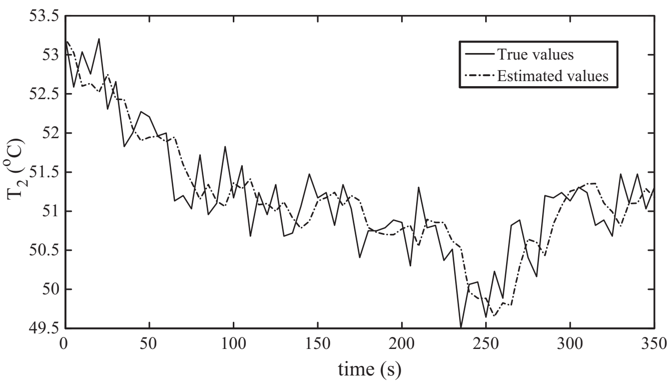

Fig. 14. Heater-mixer setup, input noise multi-rate case: Tank-2 temperature: comparison of EKF predictions with actual measurements.

5.2.2. Case 2: unknown disturbance in manipulated inputs

In this case, the noise was modelled as a white noise entering the process through the flow inputs F_1 and F_2. The variations in heat are modelled through the parameters α_1 and α_2. Thus, the covariance matrix **Q** consists the variance of the noise in F_1, F_2, T_c, α_1 and α_2. As in the previous case, the estimation problem is a state and parameter estimation problem. The input, parameter and measurement noise covariances were parametrised as diagonal matrices and estimated using the direct optimisation approach. The initial guess of **Q** and **R** and the optimum values are shown in Table 13. In the same table, the estimated values of (**Q**,**R**) are also reported. The standard deviations of the initial guess and obtained covariance of the input noise are reported in Table 12. In the case of T_1 and T_2, the estimate of **R** is close to the independently obtained sample variances of both the measurements. In the case of h_2, it is close in the order of magnitude range.

Table 10
Heater-mixer setup, state noise from unknown source: SSE values of the innovations.

	e_{T_1}	e_{T_2}	e_{h_2}
Single rate case			
Initial guess	249.944	102.045	1.465×10^{-2}
Estimated (extended EM algorithm)	162.950	57.348	5.172×10^{-3}
Estimated (direct optimisation)	171.276	58.263	4.902×10^{-3}
Multi rate case			
Initial guess	186.419	102.702	1.463×10^{-2}
Estimated (extended EM algorithm)	117.205	57.498	4.656×10^{-3}
Estimated (direct optimisation)	126.406	58.496	4.904×10^{-3}

Table 11
Heater-mixer setup, multi-rate case, state noise from unknown source: estimated values of **Q** and **R**.

Extended EM Algorithm

$$\widehat{\mathbf{Q}} \quad 10^{-3} \times \begin{bmatrix} 162.188 & -5.138 & -0.665 & 2.826 & 0.335 & 2.000 \\ -5.138 & 28.6512 & 0.0177 & 0.622 & 0.035 & 0.514 \\ -0.665 & 0.0177 & 0.007 & 0.002 & -0.002 & -0.009 \\ 2.826 & 0.622 & 0.002 & 3.4748 & -0.054 & -0.044 \\ 0.335 & 0.035 & -0.002 & -0.054 & 0.0282 & 0.079 \\ 2.000 & 0.514 & -0.009 & -0.044 & 0.079 & 0.5036 \end{bmatrix}$$

$$\widehat{\mathbf{R}} \quad 10^{-3} \times \begin{bmatrix} 142.394 & -19.911 & -0.236 & 29.062 \\ -19.911 & 113.8359 & 0.023 & -4.566 \\ -0.236 & 0.023 & 0.013 & 0.045 \\ 29.062 & -4.566 & 0.045 & 151.207 \end{bmatrix}$$

Direct Optimisation

$$\widehat{\mathbf{Q}} \quad 10^{-3}\times \operatorname{diag}\begin{bmatrix} 137.014 & 53.813 & 0.004 & 3.565 & 0.002 & 0.005 \end{bmatrix}$$

$$\widehat{\mathbf{R}} \quad 10^{-3}\times \operatorname{diag}\begin{bmatrix} 262.718 & 75.963 & 0.005 & 147.501 \end{bmatrix}$$

Table 12
Heater-mixer setup: standard deviation of initial guess and estimated value of input noise as % of the steady state.

Input	Steady state	σ initial guess	σ estimated covariance
U_2	43.75%	4	0.1
D	31.25%	6	0.02
T_c	33.8 °C	2	0.20
α_1	1	20	2
α_2	1	12	5.5

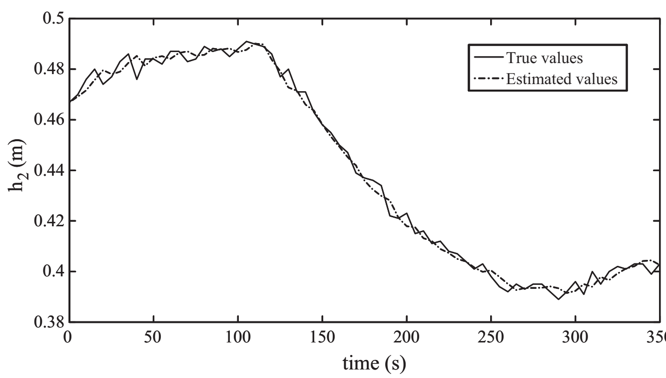

Fig. 15. Heater-mixer setup, input noise multi-rate case: Tank-2 level: comparison of EKF predictions with actual measurements.

Using the estimated values of **Q** and **R**, the performance of the EKF was evaluated on a separate validation data set. The SSE values of the innovations of the three measurements, obtained with the optimal values of **Q** and **R** are compared with the SSE values obtained with the initial guess of **Q** and **R**. The comparison in Table 14 indicates a definite improvement in the performance of the EKF when the estimated values of the covariances are used.

For the multi-rate case, the estimated values of (**Q**,**R**) are given in Table 13. The EKF predictions for Tank-1 temperature (T_1) are plotted in Fig. 13. The estimates of Tank-2 temperature are plotted in Fig. 14 while those of Tank-2 level are plotted in Fig. 15. The variations in the parameters α_1 and α_2 are shown in Fig. 16. α_1 has a mean of 0.701 and variance of 4.576×10^{-3} and α_2 has a mean of 0.710 and variance of 0.0104. In this case too, similar conclusions about the diagonal values of **R** can be drawn, as were deduced for the single rate case. The performance of the EKF using $(\mathbf{Q}^{(0)},\mathbf{R}^{(0)})$

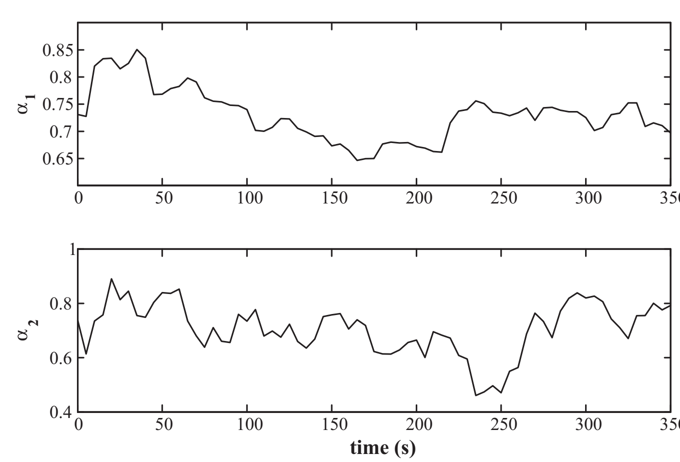

Fig. 16. Heater-mixer setup, input noise multi-rate case: variation in α_1 and α_2 using estimated values of the covariances.

Table 13
Heater-mixer setup, input noise case: initial guess and estimated values of the **Q** and **R**.

	Initial guess of **Q** and **R**
$\mathbf{Q}^{(0)}$	$10^{-3} \times \text{diag}[\,240 \quad 300 \quad 460 \quad 30 \quad 14\,]$
$\mathbf{R}^{(0)}$	$10^{-3} \times \text{diag}[\,500 \quad 15 \quad 60 \quad 20\,]$

Estimated values of Q and R

	Single rate case	Multi rate case
$\hat{\mathbf{Q}}$	$10^{-3} \times \text{diag}[\,3.590 \quad 0.035 \quad 4.058 \quad 0.385 \quad 3.304\,]$	$10^{-3} \times \text{diag}[\,3.243 \quad 0.187 \quad 3.566 \quad 0.328 \quad 2.545\,]$
$\hat{\mathbf{R}}$	$10^{-3} \times \text{diag}[\,357.26 \quad 151.07 \quad 0.006 \quad 147.71\,]$	$10^{-3} \times \text{diag}[\,244.05 \quad 91.133 \quad 0.006 \quad 147.26\,]$

Table 14
Heater-mixer setup-input noise case: comparison of the SSE values of the innovations.

	e_{T_1}	e_{T_2}	e_{h_2}
Single rate case			
Initial guess	455.845	129.469	0.029
Estimated	296.204	86.298	0.007
Multi rate case			
Initial guess	224.553	90.930	0.020
Estimated	146.046	60.469	0.005

and the estimated (**Q**,**R**) was compared and the SSE values of the innovations are reported in Table 14. There is a significant reduction in the SSE values of the irregularly sample state T1, when the estimated values of (**Q**,**R**) are used.

6. Conclusions

In this work, two approaches have been developed for constructing the maximum likelihood estimates of the process and measurement noise covariance matrices from operating data for state and parameter estimation using the extended Kalman filter. The key benefit of using the MLE approach is that the algorithms for estimation of covariances require minor modification to incorporate the occurrence of irregularly sampled data. The unmeasured disturbances affecting the state dynamics are modelled either as ‘unstructured noise’ affecting all the states or as a ‘structured noise’ entering the system dynamics through known but unmeasured input sources. To begin with, the covariance estimation problem is posed as minimising the negative of the log likelihood function of the innovations sequence. The resulting optimisation problem is solved by direct gradient-based optimisation methods. Alternatively, the expectation-maximisation (EM) algorithm is extended to deal with the covariance estimation problem. The proposed extended EM algorithm is a gradient-free approach that can be used to estimate the noise covariance if the process noise is assumed to be unstructured. The direct optimisation based method, on the other hand, can be used to estimate the covariances of noise, for the structured as well as unstructured noise cases. The efficacy of the proposed covariance estimation approach is demonstrated using simulations of a continuous fermenter process and experimental data from the bench mark heater-mixer setup. For the simulation case, it was shown that both the methods converge close to the true values of the covariances even for significantly large errors in the initial guesses of the state and noise covariance. Both the approaches generated reasonably accurate estimates of the covariances, even in the case of irregularly sampled data. For the experimental case study, the use of covariance matrices estimated from the operating data results in a marked improvement in the performance of the EKF. The estimates of the measurement error variances match closely with the sample variances of the measurement errors, that were obtained independently.

It may be noted that the estimates of the states and covariance matrices generated in the E-step involve approximations. Current research efforts, are directed towards investigating and evaluating methods that can improve these estimates.

Appendix A. The extended Kalman smoother equations

The extended Kalman smoother (an extension of the Kalman smoother [20] for nonlinear systems) calculates the smoothed estimates based on the entire set of measurements available. With initial conditions as $\hat{\mathbf{x}}_{N|N}$ and $\mathbf{P}_{N|N}$, the expressions for the smoother $\forall k = N-1, N-2, \ldots, 1$ are written as

$$\begin{aligned}\mathbf{J}_{k-1} &= \mathbf{P}_{k-1|k-1}\Phi_{k-1}^T(\mathbf{P}_{k|k-1})^{-1}\\ \hat{\mathbf{x}}_{k-1|N} &= \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{J}_{k-1}(\hat{\mathbf{x}}_{k|N} - \hat{\mathbf{x}}_{k|k-1})\\ \mathbf{P}_{k-1|N} &= \mathbf{P}_{k-1|k-1} + \mathbf{J}_{k-1}(\mathbf{P}_{k|N} - \mathbf{P}_{k|k-1})\mathbf{J}_{k-1}^T\end{aligned} \tag{A.1}$$

The extended Kalman smoother(EKS) implicitly assumes

$$p(\mathbf{x}_k|\mathbf{y}_{1:N}) \approx \mathcal{N}(\hat{\mathbf{x}}_{k|N}, \mathbf{P}_{k,N}) \tag{A.2}$$

Appendix B. Extended EM-algorithm: M-step derivations

Taking the negative of the log of Eq. (41) and ignoring the constant terms,

$$\begin{aligned}-\log L(\Theta|Z_N, X_N) = {}& \log(\left|\mathbf{P}_0\right|) + [(\mathbf{x}_0 - \mu_0)^T\mathbf{P}_0^{-1}(\mathbf{x}_0 - \mu_0)]\\ &+ N\log(\left|\mathbf{Q}\right|) + [\mathbf{w}_i^T\mathbf{Q}^{-1}\mathbf{w}_i] + N\log(\left|\mathbf{R}\right|)\\ &+ [\mathbf{v}_i^T\mathbf{R}^{-1}\mathbf{v}_i]\end{aligned} \tag{B.1}$$

Taking the expectation of the RHS conditioned on Z_N and $\Theta^{(j-1)}$, and using the identity $tr(\mathbf{x}^T\mathbf{x}) = tr(\mathbf{x}\mathbf{x}^T)$, where $tr(\cdot)$ is the trace operator

$$\begin{aligned}-\log L(\Theta|Z_N, X_N) = {}& \log(\left|\mathbf{P}_0\right|) + N\log(\left|\mathbf{Q}\right|) + N\log(\left|\mathbf{R}\right|)\\ &+ E[tr\{P_0^{-1}(\mathbf{x}_0 - \mu_0)(\mathbf{x}_0 - \mu_0)^T\}|Z_N]\\ &+ E\left\{\sum_{i=1}^{N}[tr\{\mathbf{Q}^{-1}(\mathbf{w}_i\mathbf{w}_i^T)\}|Z_N]\right\}\\ &+ E\left\{\sum_{i=0}^{N}[tr\{\mathbf{R}^{-1}(\mathbf{v}_i\mathbf{v}_i^T)\}|Z_N]\right\}\end{aligned} \tag{B.2}$$

For the term involving the initial state on the RHS

$$\begin{aligned}&E[tr\left\{\mathbf{P}_0^{-1}(\mathbf{x}_0 - \mu_0)(\mathbf{x}_0 - \mu_0)^T\right\}|Z_N]\\ &\quad = tr[\mathbf{P}_0^{-1}(\mathbf{P}_{0|N} + (\mathbf{x}_{0|N} - \mu_0)(\mathbf{x}_{0|N} - \mu_0)^T)]\end{aligned} \tag{B.3}$$

The term $[\mathbf{w}_i\mathbf{w}_i^T]$ is

$$[\mathbf{w}_i\mathbf{w}_i^T] = \{[\mathbf{x}_i - \mathbf{F}(\mathbf{x}_{i-1}, \mathbf{u}_{i-1})][\mathbf{x}_i - \mathbf{F}(\mathbf{x}_{i-1}, \mathbf{u}_{i-1})]^T\} \tag{B.4}$$

To compute the term $E[\mathbf{w}_i\mathbf{w}_i^T|Z_N]$ the RHS in Eq. (B.4) is approximated using a Taylor series expansion

$$[\mathbf{x}_i - \mathbf{F}(\mathbf{x}_{i-1},\mathbf{u}_{i-1})] \approx \mathbf{x}_i - \left[\mathbf{F}(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1}) + \left[\frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right]_{\mathbf{x}=\hat{\mathbf{x}}_{i-1|N}}(\mathbf{x}_{i-1}-\hat{\mathbf{x}}_{i-1|N})\right] \quad (B.5)$$

Using Eq. (B.5) the following expression is obtained

$$E[\mathbf{w}_i\mathbf{w}_i^T|Z_N] = E[\{\mathbf{x}_i - (\mathbf{F}(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1}) + \Phi_{i-1|N}(\mathbf{x}_{i-1}-\hat{\mathbf{x}}_{i-1|N}))\} \times \{\mathbf{x}_i - (\mathbf{F}(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1}) + \Phi_{i-1|N}(\mathbf{x}_{i-1}-\hat{\mathbf{x}}_{i-1|N}))\}^T \,|Z_N] \quad (B.6)$$

where $\Phi_{i-1|N} = [\partial \mathbf{F}/\partial \mathbf{x}]_{\mathbf{x}=\hat{\mathbf{x}}_{i-1|N}}$. Now,

$$\begin{aligned} E[\mathbf{x}_i\mathbf{x}_i^T|Z_N] &= \hat{\mathbf{x}}_{i|N}\hat{\mathbf{x}}_{i|N}^T + \mathbf{P}_{i|N} \\ E[\mathbf{x}_i\mathbf{x}_{i-1}^T|Z_N] &= \hat{\mathbf{x}}_{i|N}\hat{\mathbf{x}}_{i-1|N}^T + \mathbf{P}_{i,i-1|N} \\ E[\mathbf{x}_i\mathbf{F}^T(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1})|Z_N] &= \hat{\mathbf{x}}_{i|N}\mathbf{F}^T(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1}) \end{aligned} \quad (B.7)$$

Expanding the terms in Eq. (B.6) and taking their expectations, using Eq. (B.7)

$$\begin{aligned} E[\mathbf{w}_i\mathbf{w}_i^T|Z_N] \equiv\ & [\hat{\mathbf{x}}_{i|N}\hat{\mathbf{x}}_{i|N}^T + \mathbf{P}_{i|N}] - [\hat{\mathbf{x}}_{i|N}\mathbf{F}^T(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1}) + \mathbf{P}_{i,i-1|N}\Phi_{i-1|N}^T] \\ & + [\mathbf{F}(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1})\mathbf{F}^T(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1}) + \Phi_{i-1|N}\mathbf{P}_{i-1|N}\Phi_{i-1|N}^T] \\ & - [\Phi_{i-1|N}\mathbf{P}_{i,i-1|N}^T + \hat{\mathbf{x}}_{i|N}\mathbf{F}^T(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1})] \end{aligned} \quad (B.8)$$

Referring to Eqs. (50)–(52) and taking the summation of the expectations over the length of data in Eq. (B.8)

$$\mathbf{B}_1 = \sum_{i=1}^{N}[\hat{\mathbf{x}}_{i|N}\hat{\mathbf{x}}_{i|N}^T + \mathbf{P}_{i|N}] \quad (B.9)$$

$$\mathbf{B}_2 = \sum_{i=1}^{N}[\hat{\mathbf{x}}_{i|N}\mathbf{F}^T(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1}) + \mathbf{P}_{i,i-1|N}\Phi_{i-1|N}^T] \quad (B.10)$$

$$\mathbf{B}_3 = \sum_{i=1}^{N}[\mathbf{F}(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1})\mathbf{F}^T(\hat{\mathbf{x}}_{i-1|N},\mathbf{u}_{i-1}) + \Phi_{i-1|N}\mathbf{P}_{i-1|N}\Phi_{i-1|N}^T] \quad (B.11)$$

The term $[\mathbf{v}_i\mathbf{v}_i^T]$ is

$$[\mathbf{v}_i\mathbf{v}_i^T] = [\mathbf{y}_i - \mathbf{h}(\mathbf{x}_i)][\mathbf{y}_i - \mathbf{h}(\mathbf{x}_i)]^T \quad (B.12)$$

To compute the term $E[\mathbf{v}_i\mathbf{v}_i^T|Z_N]$, the term on the RHS of Eq. (B.12) is approximated using a Taylor series expansion

$$\mathbf{y}_i - \mathbf{h}(\mathbf{x}_i) \approx \mathbf{y}_i - \left[\mathbf{h}(\hat{\mathbf{x}}_{i|N}) + \left[\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right]_{\mathbf{x}=\hat{\mathbf{x}}_{i|N}}(\mathbf{x}_i - \hat{\mathbf{x}}_{i|N})\right] \quad (B.13)$$

Using Eq. (B.13) the following expression is obtained

$$E[\mathbf{v}_i\mathbf{v}_i^T|Z_N] \equiv E[\{\mathbf{y}_i - (\mathbf{h}(\hat{\mathbf{x}}_{i|N}) + \mathbf{C}_{i|N}(\mathbf{x}_i - \hat{\mathbf{x}}_{i|N}))\} \times \{\mathbf{y}_i - (\mathbf{h}(\hat{\mathbf{x}}_{i|N}) + \mathbf{C}_{i|N}(\mathbf{x}_i - \hat{\mathbf{x}}_{i|N}))\}^T \,|Z_N] \quad (B.14)$$

where, $\mathbf{C}_{i|N} = [\partial \mathbf{h}/\partial \mathbf{x}]_{\mathbf{x}=\hat{\mathbf{x}}_{i|N}}$. Now,

$$\begin{aligned} E[\mathbf{C}_{i|N}\mathbf{x}_i\mathbf{x}_i^T\mathbf{C}_{i|N}^T|Z_N] &= \mathbf{C}_{i|N}[\mathbf{P}_{i|N} + \hat{\mathbf{x}}_{i|N}\hat{\mathbf{x}}_{i|N}^T]\mathbf{C}_{i|N}^T \\ E\left[\mathbf{y}_i\mathbf{y}_i^T|Z_N\right] &= \mathbf{y}_i\mathbf{y}_i^T \end{aligned} \quad (B.15)$$

Expanding the terms in Eq. (B.14) and taking their expectations using Eq. (B.15)

$$E[\mathbf{v}_i\mathbf{v}_i^T|Z_N] \equiv \mathbf{y}_i\mathbf{y}_i^T - \mathbf{y}_i\mathbf{h}^T(\hat{\mathbf{x}}_{i|N}) - \mathbf{h}(\hat{\mathbf{x}}_{i|N})\mathbf{y}_i^T + \mathbf{C}_{i|N}\mathbf{P}_{i|N}\mathbf{C}_{i|N}^T + \mathbf{h}(\hat{\mathbf{x}}_{i|N})\mathbf{h}^T(\hat{\mathbf{x}}_{i|N}) \quad (B.16)$$

Referring to Eqs. (53)–(55) and taking the summations of the length of data in Eq. (B.8)

$$\mathbf{B}_4 = \sum_{i=0}^{N-1}\mathbf{y}_i\mathbf{y}_i^T \quad (B.17)$$

$$\mathbf{B}_5 = \sum_{i=0}^{N-1}[\mathbf{y}_i\mathbf{h}^T(\hat{\mathbf{x}}_{i|N})] \quad (B.18)$$

$$\mathbf{B}_6 = \sum_{i=0}^{N-1}[\mathbf{C}_{i|N}\mathbf{P}_{i|N}\mathbf{C}_{i|N}^T + \mathbf{h}(\hat{\mathbf{x}}_{i|N})\mathbf{h}^T(\hat{\mathbf{x}}_{i|N})] \quad (B.19)$$

Appendix C. Grey-box model of the Heater-mixer setup

The following algebraic equations were used to relate the mA input signal to the process input variables

(1) Flow to Tank-1

$$F_1(D) = (1.2676D^3 - 108.02D^2 + 8166D) \times 10^{-10}\ \mathrm{m^3/s} \quad (C.1)$$

(2) Flow to Tank-2

$$F_2(U_3) = (-0.1115U_3^4 + 17.2404U_3^3 - 817.6U_3^2 + 18858U_3) \times 10^{-10}\ \mathrm{m^3/s} \quad (C.2)$$

(3) Heat to Tank-1

$$Q_1(U_1) = -0.0026U_1^3 + 0.0584U_1^2 + 53.579U_1\ \mathrm{W} \quad (C.3)$$

(4) Heat to Tank-2

$$Q_2(U_2) = -0.0035U_2^3 - 0.782U_2^2 + 78.82U_2\ \mathrm{W} \quad (C.4)$$

The variables U_1, U_2, U_3 and D are expressed as percentage values of the current range (0–100%) to the final control elements. During the exercise of deriving these relations, the values of α_1 and α_2 were assumed to be 1. The details of the derivations of these equations can be found in the work by Deshpande [25].

References

[1] D. Dochain, State and parameter estimation in chemical and biochemical processes: a tutorial, J. Process Control 13 (2003) 801–818.
[2] J. Rawlings, B. Bakshi, Particle filtering and moving horizon estimation, Comput. Chem. Eng. 30 (2006) 1529–1541.
[3] R.J. Fitzgerald, Divergence of the Kalman filter, IEEE Trans. Automat. Control 16 (6) (1976) 736–747.
[4] K.A. Myers, B.D. Tapley, Adaptive sequential estimation with unknown noise statistics, IEEE Trans. Automat. Control 21 (1976) 520–523.
[5] B.J. Odelson, A. Lutz, J.B. Rawlings, The autocovariance-least squares method for estimating covariances: application to model-based control of chemical reactors, IEEE Trans. Control Syst. Technol. 14 (3) (2006) 532–540.
[6] R.K. Mehra, On the identification of variances and adaptive Kalman filtering, IEEE Trans. Automat. Control 15 (12) (1970) 175–184.
[7] P.R. Bèlanger, Estimation of noise covariances for a linear time-varying stochastic process, Automatica 10 (1974) 267–275.
[8] B.J. Odelson, M.R. Rajamani, J.B. Rawlings, A new autocovariance least-squares method for estimating noise covariances, Automatica 42 (2006) 303–308.
[9] A.P. Dempster, N.M. Laird, D. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. B 3 (1977) 1–38.
[10] R.H. Shumway, D.S. Stoffer, Time Series Analysis and its Applications, Springer-Verlag New York Inc., New York, 2000.
[11] H. Raghavan, A.K. Tangirala, R.B. Gopaluni, S.L. Shah, Identification of chemical processes with irregular output sampling, Control Eng. Pract. 14 (2006) 467–480.
[12] J. Valappil, C. Georgakis, Systematic estimation of state noise statistics for extended Kalman filters, AIChE J. 46 (2) (2000) 292–308.
[13] G.C. Goodwin, J.C. Agüero, Approximate EM algorithms for parameter and state estimation in nonlinear stochastic models, in: Proc. 44th IEEE Conference on Decision and Control, 2005, pp. 368–373.
[14] S.B. Chitralekha, J. Prakash, H. Raghavan, R.B. Gopaluni, S.L. Shah, A comparison of simultaneous state and parameter estimation schemes for a continuous fermenter, J. Process Control 20 (8) (2010) 934–943.

[15] N.F. Thornhill, S.C. Patwardhan, S.L. Shah, A continuous stirred tank heater simulation, J. Process Control 18 (3/4) (2008) 347–360.
[16] K.R. Muske, T.F. Edgar, in: M.A. Henson, D.E. Seborg (Eds.), Nonlinear State Estimation in Nonlinear Process Control, Prentice Hall, New Jersey, 1997.
[17] M.S. Arulampalam, S. Maskell, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process. 50 (2) (2002) 174–188.
[18] D.M. Himmelblau, Process Analysis by Statistical Methods, John Wiley & Sons Inc., New York, 1970.
[19] F.C. Schweppe, Evaluation of likelihood functions for Gaussian signals, IEEE Trans. Inform. Theory 11 (1) (1965) 61–70.
[20] H.E. Rauch, F. Tung, C.T. Striebel, Maximum likelihood estimates of linear dynamic systems, AIAA J. 3 (8) (1965) 1445–1450.
[21] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, John Wiley & Sons Inc., New York, 1997.
[22] C.F.J. Wu, On the convergence properties of the EM algorithm, Ann. Stat. 11 (1) (1983) 95–103.
[23] M.A. Henson, D.E. Seborg, Nonlinear control strategies for continuous fermenters, Chem. Eng. Sci. 47 (4) (1992) 821–835.
[24] M. Srinivasrao, S.C. Patwardhan, R.D. Gudi, Nonlinear predictive control of irregularly sampled multirate systems using nonlinear black box observers, J. Process Control 17 (1) (2007) 17–35.
[25] A.P. Deshpande, A unified framework for online fault identification and accommodation in nonlinear systems, Ph.D. thesis, Indian Institute of Technology Bombay, Mumbai, India, 2008.