



# The Kalman Filter for the Supervision of Cultivation Processes



Abdolrahim Yousefi-Darani, Olivier Paquet-Durand, and Bernd Hitzmann

## Contents

- 1 Introduction
- 2 Kalman Filtering Theory and Its Non-linear Extensions
  - 2.1 The Kalman Filter
  - 2.2 Continuous-Discrete Extended Kalman Filter
  - 2.3 Other Non-linear Extensions of the Kalman Filter
- 3 Application of Kalman Filters in Bioprocess Monitoring
  - 3.1 Type of Kalman Filter
  - 3.2 Microorganism
  - 3.3 Cultivation Mode
  - 3.4 Bioprocess Phase
  - 3.5 Measurement Device
  - 3.6 Process Model
- 4 An Extended Kalman Filter for the Monitoring of a Yeast Cultivation
  - 4.1 The Cultivation Process
  - 4.2 EKF Algorithm
  - 4.3 Online Ethanol Measurements
  - 4.4 Offline Measurements
  - 4.5 State Equations of the Cultivation Process
  - 4.6 Results
- 5 Conclusion
- Appendix
- References

**Abstract** In the era of technology and digitalization, the process industries are undergoing a digital transformation. The available process models, advance sensor technologies, enhanced computational power and a broad set of data analytical

---

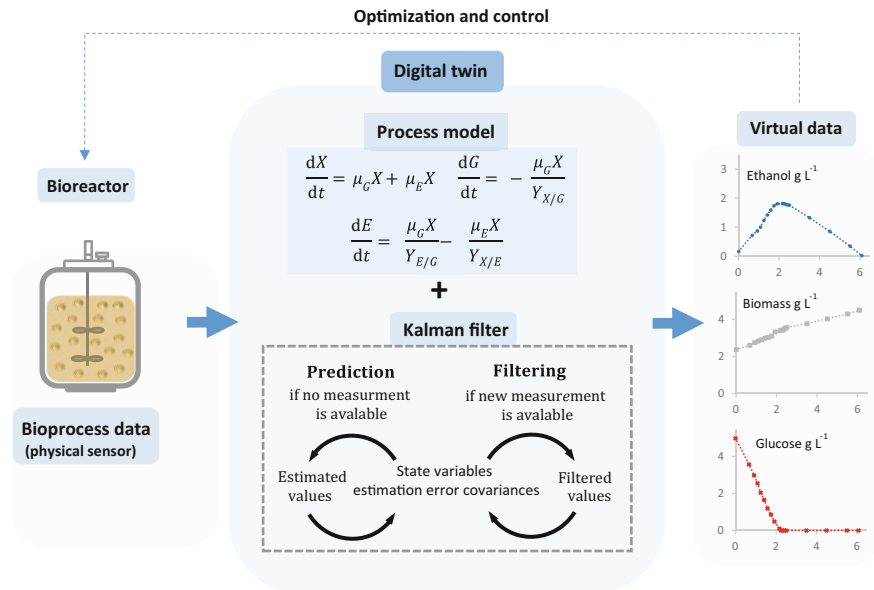
A. Yousefi-Darani (✉), O. Paquet-Durand, and B. Hitzmann  
Department of Process Analytics and Cereal Science, Institute of Food Science and Biotechnology, University of Hohenheim, Stuttgart, Germany  
e-mail: [rahim@uni-hohenheim.de](mailto:rahim@uni-hohenheim.de)

techniques enable solid bases for digital transformation in the biopharmaceutical industry.

Among various data analytical techniques, the Kalman filter and its non-linear extensions are powerful tools for prediction of reliable process information. The combination of the Kalman filter with a virtual representation of the bioprocess, called digital twin, can provide real-time available process information. Incorporation of such variables in process operation can provide improved control performance with enhanced productivity.

In this chapter the linear discrete Kalman filter, the extended Kalman filter and the unscented Kalman filters are described and a brief overview of applications of the Kalman filter and its non-linear extensions to bioreactors are presented. Furthermore, in a case study an example of the digital twin of the baker's yeast batch cultivation process is presented.

**Graphical Abstract** A digital twin of a bioreactor mirrors the processes of the real bioreactor. It contains the physical parts, the process model and prediction algorithm to predict the bioprocess variables. These values could be used for optimization and control of the process.



**Keywords** Bioprocess supervision, Cultivation, Digital twin, Estimation, Kalman filter

## Abbreviations

<b>A</b>	State transition matrix
<b>B</b>	Process input transition matrix
<b>C</b>	Measurement model
<b>CKF</b>	Cubature Kalman filter
<b>EKF</b>	Extended Kalman filter
<b>EnKF</b>	Ensemble Kalman filter
<b>F</b>	Jacoby matrix of $f()$
<b><math>f()</math></b>	Non-linear function describing the process change
<b>FIA</b>	Flow injection analysis
<b>H</b>	Jacoby matrix of measurement model
<b><math>h()</math></b>	Measurement model
<b>KF</b>	Kalman filter
<b>P</b>	Estimation error covariance matrix
<b><math>\mathbf{p}</math></b>	Model parameter vector for estimation
<b>Q</b>	Process noise covariance matrix
<b>R</b>	Measurement noise covariance matrix
<b>t</b>	Time
<b>UKF</b>	Unscented Kalman filter
<b><math>\mathbf{v}</math></b>	Measurement noise vector
<b><math>\mathbf{w}</math></b>	Process noise vector
<b><math>\mathbf{x}</math></b>	State variables vector
<b><math>\mathbf{x}(t)</math></b>	State variable at continuous time $k$
<b><math>\mathbf{x}_{[k]}</math></b>	State variable at discrete time $k$
<b><math>\mathbf{x}_{f,[k]}</math></b>	Filtered state variable at discrete time $k$
<b><math>\mathbf{z}</math></b>	Measurement vector

## 1 Introduction

Bioprocesses are described as biological systems that are non-linear, complex and unsteady; thus development of precise control systems in order to achieve robust product quality and productivity can be challenging. The control of these processes can be significantly improved by online process monitoring followed by corrective actions. In this context, bioprocess digital twins are helpful tools.

Digital twins are virtual representations of the production process which enable pre-emptive process control by using online data to predict the process outcome in advance. They convert the physical process to a smart process and thus achieve the ultimate goal of the digital transformation. This enables unprecedented possibilities for timely and automated intervention to provide critical decision support during process development [1].

Digital twins mainly consist of a mathematical model which describes the dynamic behaviour observed in a biochemical reactor and a prediction or self-learning algorithm which estimates the cellular component concentrations and the process parameters that cannot be described mechanistically [2, 3].

Bioprocess mathematical models may generally be categorized into algebraic equations and dynamic models. Algebraic equations are developed from mass and component balances, from mass or heat transfer laws or even from elemental balances. Dynamic models usually consist of dynamic balances of conserved quantities in combination with kinetics to describe rate expressions as functions of the state variables. Detailed description of mathematical modelling of bioprocesses is covered by previous authors in greater details than space allows here [4–7]. The goal of this chapter is to highlight state estimation methods with a specific focus on the Kalman filter and its non-linear extensions.

For linear systems, the Luenberger observer and the Kalman filter, whose 60th anniversary occurred in 2020 [8], are the most applied methods for estimating parameters and process variables that cannot be measured directly. In the area of non-linear systems, particle filtering (PF), high gain observers, non-linear extensions of the Kalman filter such as the extended Kalman filter (EKF) and the unscented Kalman filter (UKF) and many others have been proposed. However, due to the simple structure and low computational effort of non-linear extensions of the Kalman filter, these methods have gained more interest, and many research studies have been dedicated to the implementation of such filters for state and parameter estimation in bioprocess technologies. The main objective of this chapter is to discuss the applications of different Kalman filter algorithms in bioprocess technologies. Therefore, this chapter is organized as follows: in the next section, a brief overview of the Kalman filtering theory and its non-linear extensions will be discussed. Applications of the Kalman filter for the supervision of cultivation processes will be given in the third section, followed by a case study evaluating the implementation of an extended Kalman filter for developing a digital twin of the baker's yeast batch cultivation process. In the last section, a conclusion is presented.

## 2 Kalman Filtering Theory and Its Non-linear Extensions

The Kalman filter is a set of mathematical equations that provides an efficient computational solution of the least-squares method when the considered system is linear and the uncertainties are modelled by Gaussian random variables. When the system state dynamics is non-linear, then certain linearization methods are applied. The most prominent of these algorithms are the extended Kalman filter (EKF) and the unscented Kalman filter (UKF), invented independently by several research groups. Different extensions of the Kalman filters differ in the way the estimation error is calculated. A brief overview of these methods are as follows.

## 2.1 The Kalman Filter

The Kalman filter is used to provide optimal estimates of unmeasured states for time varying linear systems in the presence of noise by combining information from a process mathematical model with online process measurements. The process model defines the evaluation of the state from time  $k-1$  to time  $k$  as:

$$\mathbf{x}_{[k]} = \mathbf{A}\mathbf{x}_{[k-1]} + \mathbf{B}\mathbf{u}_{[k-1]} + \mathbf{w}_{[k-1]} \quad (1)$$

where  $\mathbf{x}$  is the state vector,  $\mathbf{u}$  is the process input and  $\mathbf{w}$  is the Gaussian process noise vector that is assumed to be zero-mean with the covariance  $\mathbf{Q}$ . Matrix  $\mathbf{A}$  relates the state at the previous time step  $k-1$  to the state at the current step  $k$ , matrix  $\mathbf{B}$  relates the control input to the state variables  $\mathbf{x}$ .

The process model is paired with the measurement model that describes the relationship between the state and the measurement at the current time step  $k$  as:

$$\mathbf{z}_{[k]} = \mathbf{C}\mathbf{x}_{[k]} + \mathbf{v}_{[k]} \quad (2)$$

where  $\mathbf{z}$  is the measurement vector and  $\mathbf{v}$  is the Gaussian measurement noise vector which is assumed to be zero-mean with the covariance  $\mathbf{R}$ . Matrix  $\mathbf{C}$  relates the state to the measurement  $\mathbf{z}_{[k]}$ . Since the measurements does not exhaustively inform on the current situation of the process, the KF aims to provide an estimate of the process state at time  $k$ , given the initial state of  $\mathbf{x}_0$ , the measurements and the information of the system.

The Kalman filter algorithm consists of two steps which are summarized as follows:

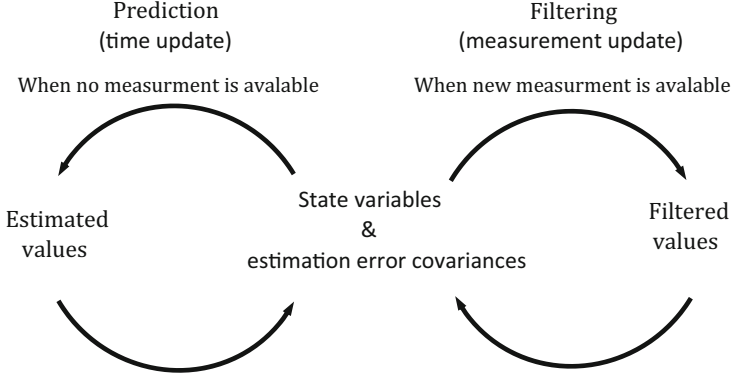
- *Prediction step (time update)*: Using the initial condition, the process model is used to predict the state variables and the estimation error covariance's until the first measurement is available.

$$\mathbf{x}_{[k]} = \mathbf{A}\mathbf{x}_{[k-1]} + \mathbf{B}\mathbf{u}_{[k-1]} \quad (3)$$

$$\mathbf{P}_{[k]} = \mathbf{A}\mathbf{P}_{[k-1]}\mathbf{A}^T + \mathbf{Q} \quad (4)$$

In the above equations,  $\mathbf{x}_{[k]}$  is the state variables estimate at time  $k$  which is deduced from a previous estimation of the state  $\mathbf{x}_{[k-1]}$  at time  $k-1$ . The new term  $\mathbf{P}$  is called the state error covariance matrix which encrypts the error covariance of the predicted state values.  $\mathbf{P}_{[k]}$  is the new prediction error covariance matrix at time  $k$  and  $\mathbf{P}_{[k-1]}$  is the previous estimated error covariance matrix at time  $k-1$ . Whenever a measurement is available, a correction step is performed:

- *Correction step (measurement update)*: In this step the predicted model estimates are combined with the measured values to provide corrected estimates.



**Fig. 1** The flow chart of the Kalman filter algorithm

$$\mathbf{x}_{f,[k]} = \mathbf{x}_{[k]} + K_{[k]}(\mathbf{z}_{[k]} - C\mathbf{x}_{[k]}) \quad (5)$$

$$P_{f,[k]} = P_{[k]}(1 - K_{[k]}C)^2 + K^2R \quad (6)$$

$$K_{[k]} = P_{[k]}C^T(R + CP_{[k]}C^T)^{-1} \quad (7)$$

The measurement prediction error, reflects the discrepancy between the true measurements  $\mathbf{z}_{[k]}$  and the predicted measurements  $C\mathbf{x}_{[k]}$ . The difference of both is multiplied by the so called Kalman gain and used to update the estimated state variables. Therefore the filtered state variables  $\mathbf{x}_{f,[k]}$  are obtained. In the similar manner, the filtered estimation error covariance  $P_{f,[k]}$  is obtained.  $K_{[k]}$  is chosen to minimizes the estimated error covariance

$$\frac{dP_f}{dK} = 0 \quad (8)$$

The measurement error variance must be compared with the estimation error variance to see how the filter is acting. For this purpose, a very rough treatment is necessary:

If  $R \ll CP_{[k]}C^T$  then  $K \approx C^{-1}$  and  $\mathbf{x}_{f,[k]} \approx C^{-1}\mathbf{z}_{[k]}$ ; so the filtered is almost determined by the measured.

If  $R \gg CP_{[k]}C^T$  then  $\mathbf{x}_{f,[k]} \approx \mathbf{x}_{[k]}$ ; the filtered value is almost the estimated one and no influence of the measurement will be obtained.

With the filtered values as initial condition the simulation of the process as well as the estimation error covariance's can be carried out until the next measurement is obtained and everything repeats again. The flow chart of the Kalman filter algorithm is presented in Fig. 1.

## 2.2 Continuous-Discrete Extended Kalman Filter

As described in the previous section, the Kalman filter addresses the general problem of trying to estimate the state of a process that is governed by a linear differential equation system. In non-linear dynamic systems, the process model or the measurement model cannot be determined with multiplication of vectors and matrices. For such systems, a linearization should be performed. The linearization can be performed by different methods. The essential difference among different versions of the Kalman filters (extended Kalman filter, unscented Kalman filter and ensemble Kalman filter) consists in how they calculate the estimation error. A Kalman filter that linearizes about the current mean and covariance is referred to as an extended Kalman filter (EKF). A non-linear dynamic system can be described by the following differential equation:

$$\frac{dx(t)}{dt} = f(x(t), u(t)) + w(t) \quad (9)$$

With discrete measurements that are:

$$z_{[k]} = h[x(t_{[k]})] + v_{[k]} \quad (10)$$

The differential equation provide the continuous part, the measurements are the discrete part, where  $f$  is a non-linear function of the state variables  $x$  and the control input  $u$ . The non-linear function  $h$  in the measurement equation relates the current state to the measurement  $z_{[k]}$ .  $w$  and  $v$  are, respectively, the process noise vector and the measurement noise vector. These noises are assumed to be zero mean, white, and independent of each other, with respective covariance matrices  $Q$  and  $R$ .

To calculate the estimation error covariance matrix, the following differential equations have to be solved in parallel to the state differential equation.

$$\frac{dP(t)}{dt} = F(t)P(t) + P(t)F^T(t) + Q \quad (11)$$

Here the Jacobian matrix is used, which is given by the following equation:

$$F = \left. \frac{\partial f}{\partial x} \right|_{x(t), u(t)} \quad (12)$$

The filtering is performed as follows:

$$K_{[k]} = P(t_k)H^T(t_k)[H(t_k)P(t_k)H^T(t_k) + R]^{-1} \quad (13)$$

$$\mathbf{x}_f(t_{[k]}) = \mathbf{x}(t_{[k]}) + K_{[k]} [z_{[k]} - h[\mathbf{x}(t_{[k]})]] \quad (14)$$

$$P_f(t_{[k]}) = [I - K_{[k]}H_{[k]}]P(t_{[k]})[I - K_{[k]}H_{[k]}]^T + K_{[k]}RK_{[k]}^T \quad (15)$$

where  $H_{[k]}$  is the Jacoby matrix of  $h[\cdot]$ :

$$H_{[k]} = \left. \frac{\partial h}{\partial x} \right|_{x_{[k]}} \quad (16)$$

Correspondingly to the KF algorithm, the EKF algorithm consists of two main parts including prediction step and the correction step.

As mentioned above, the basic framework for the EKF involves state estimation of a non-linear dynamic system. However, in some cases, prediction of  $\mathbf{x}_k$  requires coupling both state estimation and parameter estimation [9]. Here a process model parameter  $p(t)$  is considered to be time dependent and can be estimated by adding the parameter as an additional state variable whose differential equation is then given as

$$\frac{dp(t)}{dt} = 0 \quad (17)$$

At every time step, the current estimate of the parameter  $p(t)$  is used in the measurement filter. In the joint estimation method, model state variables and model parameters are included in a single joint state vector. Parameter estimation evolves in time along with state estimation, as observations are assimilated [10].

Other alternatives for parameter estimation with the KF include calibrating parameters outside the KF calculation with an outer optimisation routine [11–13], and parameter estimation in steady-state KF calculations where observations are climatological averages over the entire time period of interest [14], but in both of these two approaches the parameter estimation part of the calculation considers all observations at once rather than sequentially.

### 2.3 Other Non-linear Extensions of the Kalman Filter

As mentioned previously, when the system is non-linear and can be well approximated by linearization, then the EKF is a good option for state estimation; however EKF is not optimal if the system is highly non-linear, this is because only the mean is propagated through the non-linearity [15]. The unscented Kalman filter (UKF) is another non-linear extension of the Kalman filter which is a discrete time filtering algorithm. The UKF utilizes the unscented transformation for computing approximate solutions to the filtering problems.

A general framework for state estimation based on the UKF for this state space model is presented as follows:



In the first step, the initial values for the state and covariance estimation have to be set. Following this, the recursive estimation is performed by the prediction and correction steps. Within the prediction step, a priori state and covariance estimation utilizing the process model is performed. Using the unscented transformation, a set of sigma points are chosen. These sigma points characterize the current probability density function. Each point from the sigma matrix is propagated through the process model to calculate the estimations of state variables and the error covariance. Following this, a correction step is performed when a measurement is received. This leads to the estimations of the filtered state variables and the filtered error covariance by calculating the Kalman gain.

The UKF has been used in various fields for non-linear state estimations. However a couple of alternative approaches have emerged over the last few years, namely, the ensemble Kalman filter (EnKF) and the cubature Kalman filter (CKF) which are widely used when the process model is of extremely high order and non-linear, the initial states are highly uncertain and a large number of measurements are available [16, 17].

Similar to the UKF, the EnKF and CKF select a set of sample points (sigma points) in order to deal with the non-linearity of the system. In high-dimension systems, the weights of the sigma points in the UKF are prone to be negative, leading to low estimation accuracy.

In EnKF the error covariances are estimated approximately using an ensemble of model forecasts. The main concept behind the formulation of the EnKF is that if the dynamical model is expressed as a stochastic differential equation, the prediction error statistics, which are described by the Fokker–Plank equation, can be estimated using ensemble integrations, and the error covariance matrices can be calculated by integrating the ensemble of model states [16].

The cubature Kalman filter uses the spherical–radial cubature rule to generate some weighted sampling points to approximate integral in Bayesian estimation. A brief overview of the unscented Kalman filtering and sigma point filtering in general are given by van der Merwe [18].

### 3 Application of Kalman Filters in Bioprocess Monitoring

Here 41 recent published articles [19–60] in the period of 1991–2020 on application of the Kalman filter and its extensions for state and parameter estimation in bioprocesses are discussed. Due to space limitation, only some of the reported articles are presented in Table 1. The table is organized by classifying the articles into different categories, which include the type of the Kalman filter and the applied process model, the type of microorganism and the cultivation process mode, the measured process variable(s) and the objective of the filtering algorithm. This table would help understanding how the Kalman filter was explored chronologically to date. It should be mentioned that in some works more than one Kalman filter

**Table 1** Extended Kalman filter application for cultivation processes

Estimator/ application type	Cultivation type/ microorganism	Process model	Objective	Measured state	Reference
Extended Kalman filter/experimental application	Batch cultivation/ <i>E. coli</i>	Dissolved oxygen mass balance	Noise filtering from dissolved oxygen measurements	Dissolved oxygen	Lee et al. [19]
Extended Kalman filter/experimental application	Fed-batch cultivation/ <i>S. cerevisiae</i>	Material balance equation with Monod growth rate kinetics	Parameter estimation and substrate prediction	Glucose concentration with FIA	Hitzman et al. [32]
Kalman filter/experimental application	Batch cultivation/ <i>S. cerevisiae</i>	Ideal stirred tank reactor model with Monod growth kinetics (glucose and ethanol as limiting substrates)	Noise filtering from predicted bioprocess variables	Biomass, glucose, and ethanol (with ultrasonic velocity)	Cha and Hitzmann [36]
Extended Kalman filter/experimental application	Fed-batch cultivation/ <i>S. cerevisiae</i>	A model for an ideal stirred tank reactor in combination with Monod growth kinetics	Noise filtering from predicted glucose	Glucose concentration with flow injection analyses (FIA)	Arndt and Hitzmann [37]
Extended Kalman filter/simulation	Fed-batch cultivation/ <i>S. cerevisiae</i>	Cybernetic model of Jones and Kompala	Filtering out noise from the feed stream	Dilution rate or the gas-liquid mass transfer coefficient for oxygen	Patnaik [39]
Extended Kalman filter/simulation	Fed-batch cultivation/ <i>E. coli</i>	General dynamic model of bio-reactors with Monod growth kinetics	Parameter estimation and biomass prediction	Dissolved and exhaust oxygen and carbon dioxide	Rocha et al. [40]
Extended Kalman filter/experimental application	Fed-batch cultivation/ <i>Bordetella pertussis</i>	A model with two parameters which are calculated using separate experiments	Estimation of specific growth rate, biomass, and oxygen mass transfer	Dissolved oxygen	Soons et al. [42]

(continued)

**Table 1** (continued)

Estimator/ application type	Cultivation type/ microorganism	Process model	Objective	Measured state	Reference
Unscented Kalman filter/experimental application	Fed-batch cultivation/ hybridoma cell culture	Overflow metabolism model	Noise reduction from predicted values	Predicted specific uptake and production rate	Henry et al. [41]
Extended Kalman filter/experimental application	Fed-batch cultivation/ <i>S. cerevisiae</i>	Ideal stirred tank reactor model with Monod growth kinetics	Parameter, biomass, and glucose prediction	Glucose concentration with FIA	Klockow et al. [43]
Extended Kalman filter/experimental application	Fed-batch cultivation/ <i>E. coli</i>	General dynamic model of bio-reactors with Monod growth kinetics	Estimation of biomass, glucose, and acetate	Dissolved oxygen and carbon dioxide	Veloso et al. [44]
Unscented Kalman filter/ simulation	Fed-batch cultivation/ <i>S. cerevisiae</i>	Mass balance of substrate and biomass in the head-space with Monod growth kinetics	Estimation of biomass and substrate concentrations	Dissolved oxygen and carbon dioxide	Jianlin et al. [46]
Unscented Kalman filter/ simulation	Fed-batch/ hybridoma cell	Material balance equation with Monod growth kinetic	Prediction of acetate and glucose concentration	Biomass and dissolved oxygen	Dewasme et al. [48]
Extended Kalman filter/ simulation	Batch cultivation/ <i>S. cerevisiae</i>	Unstructured model for alcoholic fermentation with immobilized cells using Monod growth kinetics	Estimation of product, substrate, and biomass concentrations	Glucose and ethanol	Popova et al. [49]
Extended Kalman filter/experimental application	Fed-batch cultivation/ <i>S. cerevisiae</i>	Mass balance of substrate and biomass in the head-space with Monod	Estimation of substrate and biomass concentrations	Substrate and biomass concentration with NIR spectrometer	Krämer and King [54]

(continued)

**Table 1** (continued)

Estimator/ application type	Cultivation type/ microorganism	Process model	Objective	Measured state	Reference
		growth kinetics			
Unscented Kalman filter/experimental application	Fed-batch cultivation/ <i>S. cerevisiae</i>	Mass balance of substrate and biomass with Monod growth kinetics	Biomass and specific biomass growth rate estimation	Oxygen uptake and CO <sub>2</sub> formation rate	Simutis and Lübert [55]
Sigma point Kalman filter/experimental application	Fed-batch cultivation/ <i>S. cerevisiae</i>	Mass balance of substrate and biomass in the head-space with Monod growth kinetics	Estimation of substrate and biomass concentrations	Substrate and biomass concentration with NIR spectrometer	Krämer and King [57]
Extended Kalman filter/ simulation	Fed-batch cultivation/ <i>S. cerevisiae</i>	Material balance equation with Monod growth rate kinetics	Ethanol prediction and state estimation	Temperature, do and substrate concentration	Lisci and Tronci et al. [60]

algorithm are examined. More detailed description of each category for all publications is presented in the following part of this section.

### 3.1 Type of Kalman Filter

According to the type of Kalman filter algorithm, the literature presented indicates there exist a considerable number of articles on implementation of EKF for state and parameter estimation. More than 60% of the applications (28 articles) have implemented EKF algorithms for their process. This is due to the fact that the cultivation process of microorganisms is a complex non-linear biochemical process and the EKF is a well-known state estimation method for non-linear systems. The linear Kalman filter which is almost exclusively used for state estimation in linear systems have also been used by some authors (3 articles). Although the EKF shows good prediction results and is widely used in literature, it presents some disadvantages. It is reliable for systems which are almost linear on the time scale of the update intervals; it requires the calculation of Jacobians at each time step, which may be difficult to obtain for higher order systems; it does linear approximations of the system at a given time instant, which may introduce errors in the estimation, leading then the state to diverge over time [9, 15]. For instance, in continuous or fed-batch

cultivations, despite continuous supply by a feed, the substrate concentration can drop to zero as the cell takes it up very fast. In such cultivations, linearization in the time and measurement update can lead to significant inaccuracies in the process, while the EKF assumes a certain probability for substrate concentrations below zero, even though this is physically impossible [54]. Therefore in recent years, application of other non-linear extensions of the Kalman filter is used. For example, Fernandes et al. [54] have implemented an UKF algorithm in order to estimate glucose and glutamine from biomass, lactate and ammonia measurement during fed-batch cultivation of hybridoma cells. The predictions were compared to the ones obtained with an EKF; they have reported the UKF achieves better level of accuracy. Krämer and King [57] have implemented a UKF in fed-batch cultivation of *S. cerevisiae* for noise filtering from predicted biomass values with NIR spectrometer. In another study, the same authors [54] have implemented an EKF for the same process. The authors have reported accurate predicted values in both studies; however there is no comparison between the two methods. Other types of the non-linear Kalman filtering method have also been reported in literature. Zhao et al. [53] have implemented a CKF for incorporating delayed measurements of biomass, substrate, and product concentration in fed-batch cultivation for penicillin production. Bavdekar et al. [47] have implemented an EnKF for overcoming delayed measurements of biomass, substrate and ethanol concentration in fed-batch cultivation of *S. cerevisiae*. Addressing the same delay problem Klockow et al. [43] complemented a ring buffer by an EKF and got satisfied results.

In order to indicate which Kalman filter extension describes the process better, numerical simulation runs are required. According to this perspective, a closer look to the presented articles indicates that most studies (31 articles) had relied on practical applications and simulation studies have been reported only 12 times.

## 3.2 Microorganism

Regarding the type of microorganism, the articles show that the majority of the research has focused on applying the Kalman filter or its extensions for state or parameter estimation during the cultivation of *S. cerevisiae* (19 articles) and *E. coli* (7 articles). The importance of these microorganisms for the biopharmaceutical industry is widely recognized, as *E. coli* and *S. cerevisiae* are the most important host microorganism used to produce recombinant proteins [58]. In addition, *S. cerevisiae* is also widely used for the production of the baker's yeast as well as wine and beer. Only a few articles demonstrate state estimation in the cultivation process of other microorganisms. For instance, some authors have implemented state estimation methods for prediction of substrate and product concentration during cultivations of *Candida utilis* [30], *Penicillium chrysogenum* [46, 53] and *Kluyveromyces marxianus* [34].

### 3.3 *Cultivation Mode*

From an operational point of view, cultivation of microorganisms can be performed in batch, fed-batch and continuous modes. In fed-batch cultivation modes, set point control of the substrate concentration by manipulating the input flow rate is a matter of particular economic and scientific interest. In order to have an efficient control system, sufficient knowledge about the process state variables is required, which can be achieved by the state estimation methods such as the Kalman filter or its extensions. Therefore, previous studies have almost exclusively focused on the application of state estimation methods for fed-batch cultivations (34 publications). However, online monitoring and estimation of state variables in batch cultivations is also crucial in order to monitor the state and if necessary may improve it to achieve high productivity over the process. For instance, controlling the level of dissolved oxygen (DO) in the fermentation broth, effects the rate of microbial metabolism. Accordingly, Lee et al. [19] have implemented an EKF for noise filtering of dissolved oxygen measurements which were used for controlling the DO levels in batch cultivation of *E. coli*. This approach and, more generally, online monitoring and state estimation of variables in batch cultivations remain briefly addressed in the literature.

### 3.4 *Bioprocess Phase*

Mixing of medium and pre-cultures are performed during upstream processing phase and separation and purification of the product from biomass is performed during the downstream processing phase. In order to optimize cell growth and maximize the product yield, online monitoring and a tight control is required during both phases. The presented articles show there have been numerous studies to investigate the application of state estimation methods during the cultivation phase (39 papers). However, the articles indicate that only two authors had examined the application of Kalman filtering methods for state and variable estimation in downstream processing. For efficient and robust process development in the downstream processing phase, knowledge of the location and concentration of the product and key contaminants is also crucial. Holwill et al. [28] have used a low technology detection system involving the measurement of rate of change of absorbance at a single wavelength after addition of reagent to a representative sample stream. This provided online data detailing the performance of a continuous precipitation process. This information as well as a mathematical model which describes the fractional protein perception were fed into a control algorithm which was programmed to maintain predefined set points by feedback control through adjustments to the overall feed saturation. The Kalman filter was used for estimating the parameters of the model. Feidl et al. [59] developed a state estimation procedure for estimation

of antibody concentration by combining information coming from kinetic model and a Raman analyser, in the frame of an extended Kalman filter approach (EKF).

### 3.5 Measurement Device

An overview of measurement devices that are appropriate for the operation of bioprocesses is presented by Sonnleitner [61]. More specific details of different types of sensors and their measurement principles can be found in literature [62, 63]. The literature presented indicate that in *E. coli* cultivation, most authors have employed DO and CO<sub>2</sub> measurements from the exit gas or glucose measurements using flow injection analysis as the measurement in the Kalman filter algorithm. On the other hand, in *S. cerevisiae* cultivations, besides DO, CO<sub>2</sub> and glucose measurements, biomass measurements have also been widely applied. For example, Dewasme et al. [48] applied biomass measurements for their KF during an *E. coli* cultivation.

### 3.6 Process Model

According to the articles presented, the general mass balance equations are the most common mathematical approach used for describing the process in state observing algorithms. An overview of typical models applied to bioprocesses is presented by Chhatre [64]. A wide variety of growth kinetics are developed for modelling of particular bioprocesses. The Monod growth model [65] is the most applied method for calculating the growth kinetics of microorganisms; it corresponds to a rational function in which the specific growth rate  $\mu$  is only a function of a single limiting substrate concentration and is subjected to substrate saturation when  $S \gg K_s$ .

$$\mu = \mu_{\max} \frac{S}{K_s + S} \quad (18)$$

where  $\mu_{\max}$  is the maximum specific growth rate,  $K_s$  is the Monod half-saturation constant, and  $S$  is the concentration of the limiting substrate. In the mentioned articles, all of the authors, which were growing *S. cerevisiae* and *E. coli*, have implemented the Monod growth kinetics. A modified Monod model was applied by Patnaik [35, 38] which is described in detail by Henson and Seborg [66] or Jones and Kompala [67]. Application of other methods for calculating the growth kinetics such as the Contois growth model [68] has also been reported. A feature of the Contois growth model is that growth rate depends upon the concentrations of both substrate and cell mass with the consequence that an inhibition is present at high cell concentrations. This growth kinetic has been implemented in a process model describing the growth behaviour of *Penicillium chrysogenum* in fed-batch

cultivations. A modified Contois model was applied by Jianlin et al. [48] and Zhao et al. [53] in an UKF and CKF algorithm for biomass and substrate prediction, respectively. The growth rate can also be represented by artificial neural networks. However this kind of models is not applied often in combination with a KF. Zorzetto and Wilson [27] have applied a hybrid model in an EKF algorithm which is based on the theory of limited respiratory with using artificial neural network for predicting the growth rates during fed-batch cultivation of *S. cerevisiae*.

Most of the process models which are reported in literature and are used in the Kalman filter algorithms are considered to be ideal stirred tank reactors, whereas production-scale operations are corrupted by noise. This problem is more severe in large-scale operations than in laboratory-scale fermentations [35]. This can describe why all applications of state estimation methods presented in Table 1 are performed in laboratory-scale bioreactors (most cultivations are performed in a 2–5 L bioreactor and one cultivation [57] have been performed in a 22 L bioreactor).

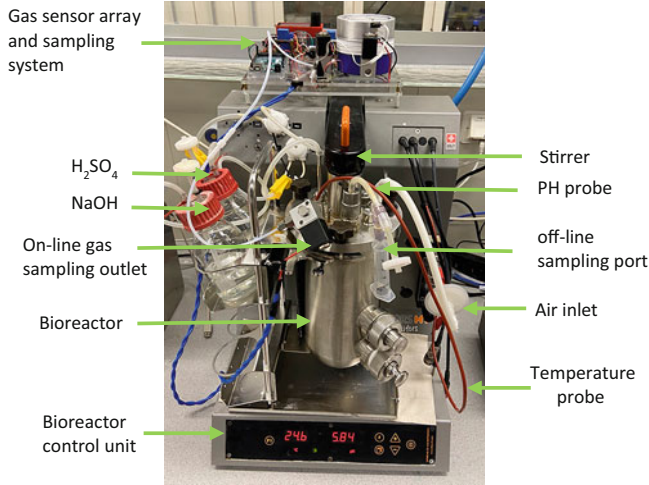
## 4 An Extended Kalman Filter for the Monitoring of a Yeast Cultivation

The integration of gas sensor array data in a non-linear state estimator has not been discussed previously in the literature. Yousefi-Darani et al. [69] have designed and implemented a model-based calibrated gas sensor array for online measurement of ethanol concentration in batch cultivation with the yeast *S. cerevisiae*. However the predicted values are only available every 5 min. Therefore in this work, in order to have continuous values of ethanol concentration as well as the values of biomass, glucose and the maximal growth rates, we have implemented an EKF. In addition, the whole estimation producer could be considered as a digital twin of the baker's yeast batch cultivation process, which could be used for process optimization and control.

### 4.1 The Cultivation Process

The cultivation of *Saccharomyces cerevisiae* (fresh baker's yeast, Oma's Ur-Hefe) was carried out in a 2.5 L bioreactor (Minifors, Infors HT, Bottmingen, Switzerland) with a vessel of stainless steel working volume of 1.35 L equipped with a temperature (set point of 30°C) and pH (set point pH = 5) control unit. The aeration and agitation rates were kept constant at 3.5 L min<sup>-1</sup> and 500 rpm, respectively. For the pre-culture, 5 g of the baker's yeast was suspended into 100 mL medium containing 0.34 g L<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.42 g L<sup>-1</sup> CaCl<sub>2</sub>·2H<sub>2</sub>O, 4.5 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.9 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>HPO<sub>4</sub>, 0.9 g L<sup>-1</sup> KCl. The inoculation was performed after 10 min of shaking. The same medium supplemented with glucose to a final concentration of





**Fig. 2** Overview of the experimental setup

5 g L<sup>-1</sup> as well as 1 mL L<sup>-1</sup> trace elements solution (0.015 g L<sup>-1</sup> FeCl<sub>3</sub>·6H<sub>2</sub>O, 9 mg L<sup>-1</sup> ZnSO<sub>4</sub>·7H<sub>2</sub>O, 10.5 mg L<sup>-1</sup> MnSO<sub>4</sub>·2H<sub>2</sub>O, and 2.4 mg L<sup>-1</sup> CuSO<sub>4</sub> 5H<sub>2</sub>O) and 1 mL L<sup>-1</sup> vitamin solution (0.06 g L<sup>-1</sup> myoinositol, 0.03 g L<sup>-1</sup> Ca-pantothenate, 6 mg L<sup>-1</sup> thiamine HCl, 1.5 mg L<sup>-1</sup> pyridoxine HCl, and 0.03 mg L<sup>-1</sup> biotin) was used for the cultivation. The experimental setup is presented in Fig. 2.

## 4.2 EKF Algorithm

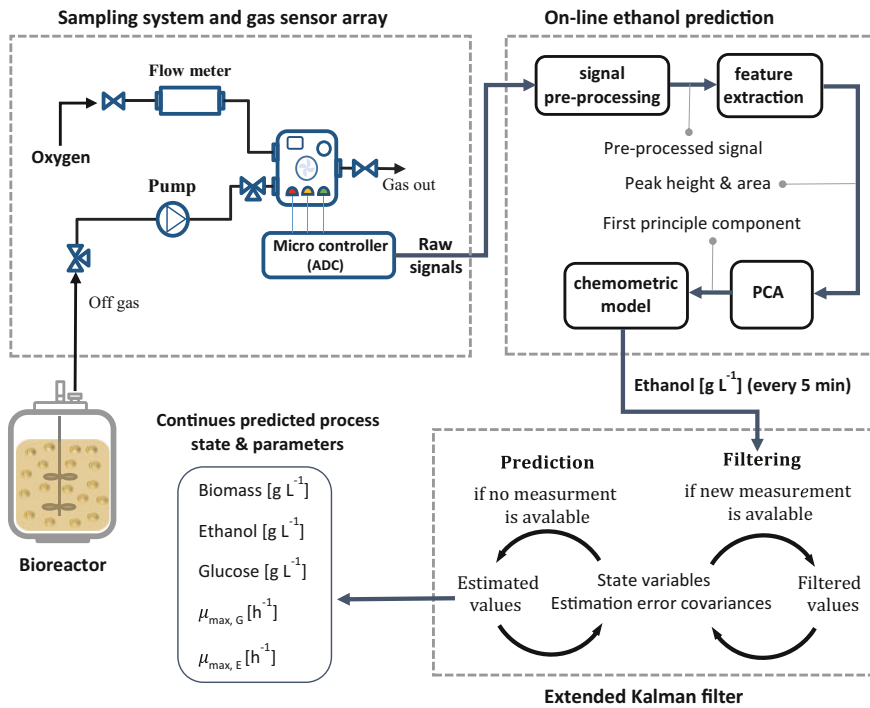
The EKF uses discrete measurements of ethanol from the gas sensor array and estimates continuous online values of ethanol, biomass and glucose concentrations as well as the maximal growth rates in *S. cerevisiae* batch cultivation. A detailed description of the working principle of the EKF is presented in Sect. 2.2.

The EKF was implemented using the software Matlab<sup>®</sup> 2019a (version 9.6.0); the “Symbolic Math” toolbox (version 8.3) was used to calculate the estimation error covariance differential equation matrix (25 equations). For all calculations, a normal office PC (Intel Core<sup>®</sup> i5 8,500 with 8 GiB of RAM) with Window 10 was used. For the simulation, the system of in total 30 (5 + 25) differential equations was solved numerically using the explicit, Runge–Kutta-based ode45 method from Matlab. The Matlab code can be found in the appendix.

### 4.3 Online Ethanol Measurements

The online ethanol measurements were performed in a self-developed system equipped with commercially available metal oxide semiconductor (MOS) gas sensors (TGS 822, TGS 813 and MQ3). The sensors were located in a measuring chamber with a volume of 250 mL and operated in two cycles: a measurement cycle and a washing cycle. During the measurement cycle, the headspace gas was pumped into the measurement chamber for 10 s at a flow rate of  $400 \text{ mL min}^{-1}$  with a diaphragm pump (Schwarzer Precision, Essen, Germany). Then the chamber was flushed by pure oxygen for regeneration. A peak-shaped measurement signal is obtained, which was evaluated by using a chemometric model, which is described in detail in the literature [69]. Therefore, every 5 min a new ethanol measurement value is used by the Kalman filter. Figure 3 presents a schematic diagram of the online ethanol measurement system and the EKF for continuous state variables and parameter estimation.

Note that the EKF was carried out after the experiments were performed. The results, however, carry over to a true online application where the data is not analysed or modified in retrospect.



**Fig. 3** Schematic diagram of the online ethanol measurement system and the EKF for continuous state variables and parameter estimation

#### 4.4 Offline Measurements

For offline analysis, samples were regularly taken from the bioreactor and placed in pre-weighed and pre-dried micro centrifuge tubes. For biomass determination, the sample without supernatant were dried for 24 h at 103°C and after cooling for 30 min weighed. Using the filtrated supernatant (pore size filter, 0.45 µm, polypropylene membrane, VWR, Darmstadt, Germany), glucose and ethanol were determined by HPLC (ProStar, Variant, Walnut Creek, CA, USA); injection of 20 µL into a Rezex ROA-organic acid H+ (8%) column (Phenomenex, Aschaffenburg, Germany) and operated at 70°C with 5 mM H<sub>2</sub>SO<sub>4</sub> as an eluent at 0.6 mL min<sup>-1</sup> flow rate; software Galaxie™ Chromatography (Varian, Walnut Creek, CA, USA). The offline values were not used during the estimation of the state variables and are only taken to show that the estimates are accurate.

#### 4.5 State Equations of the Cultivation Process

As bioreactor an ideal stirred tank reactor was assumed. As state variables, the biomass, glucose and ethanol concentrations as well as the maximal specific growth rate on glucose and ethanol were applied. Therefore, the following state equations are obtained:

$$\frac{d}{dt} \begin{bmatrix} X \\ G \\ E \\ \mu_{max,G} \\ \mu_{max,E} \end{bmatrix} = \begin{bmatrix} (\mu_G + \mu_E)X \\ -\frac{\mu_G}{Y_{GX}}X \\ \left(\frac{\mu_G}{Y_{GE}} - \frac{\mu_E}{Y_{EX}}\right)X \\ 0 \\ 0 \end{bmatrix} \quad (19)$$

where  $\mu_G$  and  $\mu_E$  are given as

$$\mu_G = \frac{\mu_{max,G} \cdot G}{K_G + G} \quad (20)$$

$$\mu_E = \frac{\mu_{max,E} \cdot E}{K_E + E} \cdot \left(1 - \frac{\mu_G}{\mu_{max,G}}\right)^2 \quad (21)$$

As one can see from the state equation, the Kalman filter is used to estimate the maximum specific growth rate on glucose  $\mu_{max,G}$  and on ethanol  $\mu_{max,E}$ . The importance of the specific growth rate for the assessment of a cultivation is discussed by Galvanauskas et al. [70].

**Table 2** Parameter values used for the simulation model

Parameter	Value	Description
$K_G$	$0.1 \text{ gL}^{-1}$	Monod constant glucose
$K_E$	$0.1 \text{ gL}^{-1}$	Monod constant ethanol
$Y_{GX}$	$0.17 \text{ gg}^{-1}$	Conversion factor glucose to biomass
$Y_{GE}$	$0.46 \text{ gg}^{-1}$	Conversion factor glucose to ethanol
$Y_{EX}$	$0.6 \text{ gg}^{-1}$	Conversion factor ethanol to biomass

**Table 3** Initial conditions for the extended Kalman filter

Parameter	Value	Description
$X_t = 0$	$2.4 \text{ gL}^{-1}$	Initial biomass concentration
$G_t = 0$	$5.0 \text{ gL}^{-1}$	Initial glucose concentration
$E_t = 0$	$0.1 \text{ gL}^{-1}$	Initial ethanol concentration
$\mu_{max, G}$	$0.14 \text{ h}^{-1}$	Initial maximal growth rate on glucose
$\mu_{max, E}$	$0.07 \text{ h}^{-1}$	Initial maximal growth rate on ethanol
$P_t = 0$	$\begin{pmatrix} 0.02 \text{ g}^2\text{L}^2 & 0 & 0 & 0 & 0 \\ 0 & 0.02 \text{ g}^2\text{L}^2 & 0 & 0 & 0 \\ 0 & 0 & 0.02 \text{ g}^2\text{L}^{-2} & 0 & 0 \\ 0 & 0 & 0 & 0.02 \text{ h}^{-2} & 0 \\ 0 & 0 & 0 & 0 & 0.02 \text{ h}^{-2} \end{pmatrix}$	Initial estimation error covariance matrix

The extension to the ordinary Monod model for  $\mu_E$  is applied, so that the transformation from glucose consumption to ethanol consumption is modelled. In Tables 2, 3, and 4 the parameters of the model as well as the initial values for the state equations and the initial values of the estimation error covariance are presented.

The Matlab code as well as the measured off- and online data of this example can be found in the appendix.

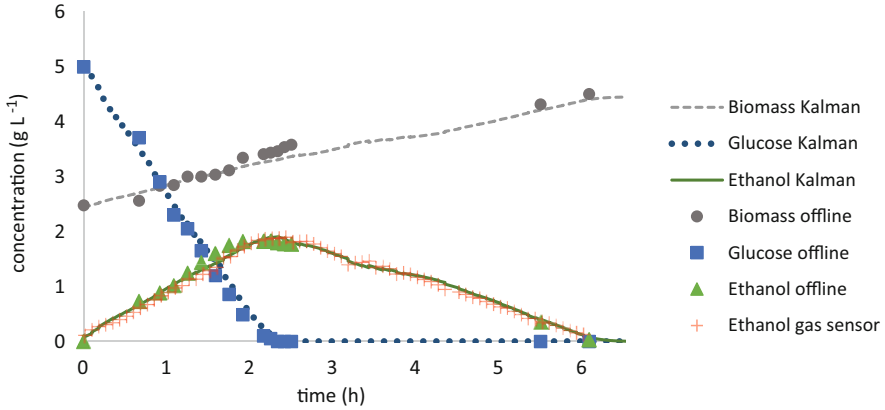
## 4.6 Results

In Fig. 4 the online and offline measured values of ethanol, the offline measured values of biomass and glucose as well as all the Kalman filter estimated values of all three bioprocess variables can be seen.

Figure 4 indicates the typical diauxic growth pattern of baker's yeast on glucose is obtained. First the glucose is consumed and biomass and ethanol are produced,

Table 4 Estimated measurement noise and process noise as well as measurement model for the EKF

Parameter	Value	Description
$R$	$0.0225 \text{ g}^2\text{L}^{-2}$	Measurement noise variance
$Q$	$\begin{pmatrix} 0.001 \text{ g}^2\text{L}^2\text{h}^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0.001 \text{ g}^2\text{L}^2\text{h}^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0.001 \text{ g}^2\text{L}^{-2}\text{h}^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0.005 \text{ h}^{-3} & 0 \\ 0 & 0 & 0 & 0 & 0.005 \text{ h}^{-3} \end{pmatrix}$	Process noise covariance matrix
$z$	$(0 \ 0 \ 1 \ 0 \ 0)$	Measurement matrix (just ethanol is measured)



**Fig. 4** Online and offline values for biomass, glucose and ethanol as well as EKF estimates for these values

**Table 5** Prediction error of EKF values compared to offline measurements

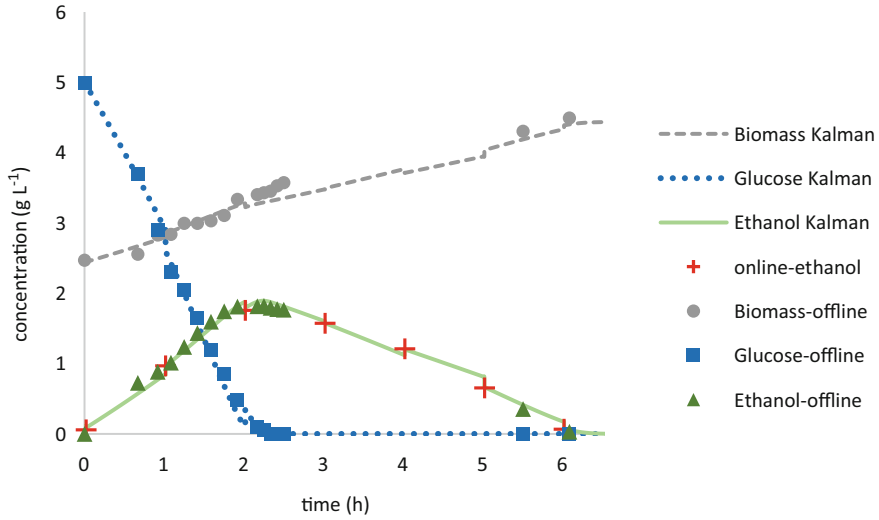
	Glucose	Ethanol	Biomass
RMSEP	0.12 g L <sup>-1</sup>	0.14 g L <sup>-1</sup>	0.12 g L <sup>-1</sup>
Error	5.6%	2.8%	6.2%
R <sup>2</sup>	0.96	0.99	0.97

then ethanol is converted to biomass. The offline measurements and its corresponding estimated values fit quite well together as can be seen in Table 5.

The root mean squared error of prediction (RMSEP) of glucose is 0.12 g L<sup>-1</sup>. The ethanol offline values during glucose consumption are mostly higher than the online measured and the predicted ones; in overall their RMSEP is 0.14 g L<sup>-1</sup>. All ethanol online measurements seems to be a little bit shifted in time compared to the offline values, which might indicate the time delay due to gas transport from the fermentation broth through the headspace of the reactor to the measurement system. The biomass has a RMSEP of 0.12 g L<sup>-1</sup>, but the highest deviation can be seen shortly after ethanol is used as substrate. The values shortly before ethanol consumption might not be predicted accurately, because the model describing the switching from glucose to ethanol might be suboptimal.

In order to investigate the influence of the measurement frequency on the performance of the EKF, we decreased the measurement frequency of the online ethanol measurements to one per hour. The results of the estimated values with the EKF are presented in Fig. 5.

Still the overall behaviour of the estimated values is the same. However, the sampling frequency has an influence on the corrections of the estimated state during filtering. Larger step changes are observed in the estimated values whenever a new measurement is available. However, even if the sampling frequency is changed to one per hour, the overall behaviour is predicted well.



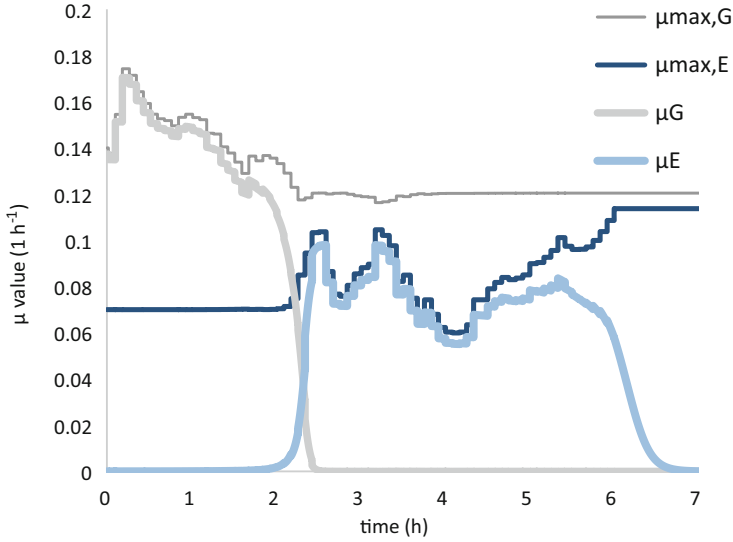
**Fig. 5** Online (every 1 h) and offline values for biomass, glucose and ethanol as well as EKF estimates for these values

Obviously with a higher sampling frequency, these step changes are smaller. Nevertheless, with a 5 min sampling time, the EKF was able to follow the true states of the system with a reasonably small error. More detailed information about the influence of the sampling frequency on the accuracy of the Kalman filter estimates can be found in literature [71, 72].

The EKF was also used for predicting the specific growth rates and their maximum values.

In Fig. 6 the estimated maximum specific growth rates with respect to glucose  $\mu_{max, G}$  and ethanol  $\mu_{max, E}$  as well as specific growth rates itself ( $\mu_G$  and  $\mu_E$  for glucose and ethanol respectively) are presented.

After inoculation, the specific growth rate and its maximum value with respect to glucose are increasing from  $0.14 \text{ h}^{-1}$  to more than  $0.18 \text{ h}^{-1}$ . However shortly thereafter they decrease again. This indicates the high sensitivity of the estimation values due to the measurement noise variance  $R$  and the process noise variance with respect to  $\mu_{max, G}$ , which is  $Q$  [4]. The smaller the  $R$  and the higher the  $Q$  [4], the more the estimated values will rely upon the measurements and as a consequence the filtered values might be changed, if the measured and estimated values deviate from each other. The more glucose is consumed, the larger will be the difference of  $\mu_{max, G}$  and  $\mu_G$ , due to the Monod growth kinetics. If the glucose is almost depleted, the extension to the Monod model on ethanol contributes to increasing growth on ethanol. Shortly after 2 h cultivation time, the transition from glucose to ethanol as substrate takes place. The maximum specific growth rate on ethanol  $\mu_{max, E}$ , which has not changed during the growth on glucose starts to increase. According to the typical Monod behaviour, before ethanol is depleted, due to the low substrate concentration,  $\mu_{max, E}$  should be almost constant while  $\mu_E$  should be increasing.



**Fig. 6** Estimated maximum specific growth rates with respect to glucose  $\mu_{\max, G}$  and ethanol  $\mu_{\max, E}$  as well as the specific growth rates ( $\mu_G$  and  $\mu_E$  for glucose and ethanol, respectively)

However this is not observed in Fig. 6 which is due to the fluctuation of the measured and estimated ethanol concentration.

## 5 Conclusion

In this chapter, the working principles as well as an overview of Kalman filter applications for state and parameter estimation in bioprocesses has been presented. Regarding the type of the Kalman filter, since most biotechnical processes are non-linear, non-linear versions of the Kalman filter, specifically the EKF, are the most applied algorithm among other extensions of the Kalman filter. However the UKF is getting attention in recent years. The results in literature indicate that the UKF algorithms deliver more accurate estimates of the parameters and state variables compared to EKF algorithms.

In spite of the apparent success of Kalman filters for state and parameter estimation in lab-scale bioreactors, the integration of Kalman filters into industrial systems is not very widespread while most of the process models mentioned in literature consider noise-free ideal fermentations, whereas production-scale operations are corrupted by concentration gradients and disturbance. Accordingly, more efforts are required towards performing simulation studies in order to model and validate proper mathematical models associated with complex non-ideal bioprocesses.

Despite the numerous examples on state estimation methods for biotechnological processes in literature, the research on implementing Kalman filters for state



estimation in downstream processing remain rather limited. The advancement in state and parameter estimation methods in downstream processes leads to better knowledge of the location and concentration of the product and key contaminants, which are essential for process optimization and control.

So far most of the Kalman filter algorithms are implemented for monitoring fed-batch cultivations; however more attention is required for real-time implementation of the Kalman filter algorithms for controlling the feed rate and substrate production in these cultivations. Further efforts are also required towards implementation of state estimation methods in batch and continuous cultivations.

From the presented literature, it could be concluded that the non-linear extensions of the Kalman filter are powerful tools for state estimation in bioprocesses; therefore they could be used for digitalization of bioprocesses. Accordingly, in a case study, a digital twin of the baker's yeast batch fermentation process was developed by using a dynamic non-linear model of the process as well as an EKF algorithm. The proposed method gives the possibility to predict glucose, ethanol and biomass concentrations simultaneously from the only available infrequent online measurements of ethanol concentration. The accuracy of the estimated biomass and substrate production are in line with other studies which have also implemented an EKF algorithm for monitoring the baker's yeast cultivation [32, 49]. However, in our application the maximal specific growth rates on glucose and ethanol are also estimated. As a consequence, the rapid and precise estimation of these variables could increase the overall knowledge integration in the digital twin of the process.

Overall, the unique advantage of online monitoring and in general digital twins of bioprocesses is that they could play critical roles in bioprocess development such as supporting problem solving in manufacturing, reducing effort in setting up a control strategy and accelerating process performance by taking corrective actions automatically and in real time.

## Appendix

Extended Kalman filter Matlab code: Online state prediction of batch yeast cultivations based on ethanol gas sensors.

```
%Initialization
clear; close all; clc;
sympref('AbbreviateOutput', false);
%Variable and parameter definition
%Symbols for symbolic math calculations
syms G E X P t                      real
syms Y_gx Y_ge Y_ex mu1 mu2 K_M_G K_M_E    real
%Variables / Parameters
initX = [2.5; 6; 0.2; 0.15; 0.08];      % initial state (Biomass,
% Glucose, Ethanol)
```

```

initP = diag([0.1,0.02, 0.02,0.2,0.02]); % initial process estimation
% covariance matrix
init = [initX; initP(:)]; % combined initial value
%vector
% for the odesolver
H = [0 0 1 0 0] % observation matrix
H = 1x5
0 0 1 0 0
Q = diag([0.001,0.001,0.001,0.001,0.001]) % process noise covariance
%matrix

R = 0.05 % measurement noise
%covariance matrix
K1 = 0.1; % Monod konstant glucose
K2 = 0.1; % Monod konstant ethanol

%estimated parameter values
Ygx = 0.15; % Yield glucose -> biomass
Yge = 0.34; % Yield glucose -> ethanol
Yex = 0.43; % Yield ethanol -> biomass
%Process model
%Monod terms
mue1 = mu1*G / (G+K_M_G);
mue2 = mu2*E / (E+K_M_E) * (1 - mue1/mu1);
%Model OD
dS = sym(X * [...
    (mue1 + mue2) ;... % Biomass
    -mue1/Y_gx ;... % Glucose
    (mue1/Y_gx*Y_ge - mue2/Y_ex) ;... % Ethanol
    0; % mue1
    0; % mue2
    1]);
%Jacobian of Model with respect to state variables
F = jacobian(dS, [X,G,E,mu1,mu2])
P matrix
P = sym('P', [5,5])
dP = F * P + P*F'+Q
%Simulation / State prediction and filtering
%Replace all symbolic parameters with their respective numeric values
F = subs(F, [Y_gx Y_ge Y_ex K_M_G K_M_E], [Ygx Yge Yex K1 K2]);
dS = subs(dS, [Y_gx Y_ge Y_ex K_M_G K_M_E], [Ygx Yge Yex K1 K2]);
dP = subs(dP, [Y_gx Y_ge Y_ex K_M_G K_M_E], [Ygx Yge Yex K1 K2]);
%Assemble all differential equations into a vector of 12 elements
%(3x state, 9x P)
OdeSys = matlabFunction([dS(:);dP(:)], 'Vars', {t, [X; G; E; mu1; mu2;
P(:)]});
%load measurement values from file:

load BC2_eth_pred.mat % Ethanol sensor
%measurements
load BC2.mat % Offline values for

%Simulate the process from one ethanol gas measurement time to the next:
t0 = 0;

```

## The Kalman Filter for the Supervision of Cultivation Processes

```

MC = zeros(0,5);           %store filtered states in these variables
SimState = zeros(0,5);
SimTime = [];
for i = 1:numel(timeE)
    tspan = [t0 timeE(i)];
    [T,state] = ode45(OdeSys, tspan, init);    % simulate / solve model

    PS      = state(end,1:5)';                % predicted state

    MS      = ME(i);                          % measured state
    P       = reshape(state(end,6:end),5,5);  % process covariance
                                                % matrix
    K       = P*H' / (H*P*H' + R);            % kalman gain matrix
    FS      = PS + K * (MS-PS(3));             % filtered state
    Pfilt   = P - K*H*P';                     % filtered process
                                                % covariance matrix
    init    = [FS; Pfilt(:)];                 % new initial condition
    t0      = timeE(i);                       % new starting time for
                                                % next iteration

% Save intermediate states for plotting
    MC      = [MC;FS'];
    state(end,1:3) = NaN;
    SimState = [SimState; state(:,1:5)];
    SimTime  = [SimTime; T];
end
%Results
%Plot the results in a presentable figure and save file to disk
f = figure("Position", [0,0,1600,640]);
subplot(1,2,1);
h = plot([0,time'], [initX(1:3)';M], 'l', 'MarkerSize',20); % Plot
%measurements
set(h, {'color'}, {'r'; 'g'; 'b'}); hold on;
h = plot(SimTime,SimState(:,1:3));                % Plot simulated values
set(h, {'color'}, {'r'; 'g'; 'b'});
plot(timeE,ME, 'b', 'MarkerSize',8); hold off; % Plot ethanol gas sensor
%values
ax = gca;
ax.FontSize = 14;
ax.FontName = 'Times';
ax.Position = [.05 .1 .4 .85];
ax.ActivePositionProperty = 'outerposition';
ax.GridLineStyle = ':';
ax.GridAlpha = .7;
xlabel('time $/h$', 'interpreter', 'Latex', "FontSize",16);
ylabel('concentration $\frac{g}{L}$', 'interpreter', 'Latex', "FontSize",16); ylim([0 8]);
$, 'interpreter', 'Latex', "FontSize",16);
grid on; box off; grid(gca, 'minor');
legend('Biomass offline', 'Glucose offline', 'Ethanol offline', 'Biomass
Kalman', 'Glucose Kalman', 'Ethanol Kalman', 'Ethanol gas
sensor', 'interpreter', 'Latex', "FontSize",12, "Color", [.9 .9 1 .9]);
subplot(1,2,2);
h = plot(SimTime,SimState(:,4:5));                % Plot mu values over time
ax = gca;

```

```

ax.FontSize = 14;
ax.FontName = 'Times';
ax.Position = [.55 .1 .4 .85];
ax.ActivePositionProperty = 'outerposition';
ax.GridLineStyle = ':';
ax.GridAlpha = .7;
ytickformat('%0.2f')
set(h, {'color'}, {'r'; 'k'});
xlabel('time $/h$', 'interpreter', 'Latex', 'FontSize', 16);
ylabel('$\mu$ value $\frac{1}{h}$', 'interpreter', 'Latex', 'FontSize', 16);
grid on; box off; grid(gca, 'minor');
legend('$\mu_1$', '$\mu_2$', 'interpreter', 'Latex', 'FontSize', 12, 'Color', [.9 .9 1]);
annotation('arrow', [.55 .97], [.1 .1])
annotation('arrow', [.05 .47], [.1 .1])
annotation('arrow', [.05 .05], [.1 .98])
annotation('arrow', [.55 .55], [.1 .98])
saveas(f, 'KalmanPred.svg', 'svg'); % save copy of figure to file

%Calculate Errors
SimTime = SimTime + ((1:numel(SimTime))*1e-10)';
SimValues = interp1(SimTime, SimState(:, 1:3), time);
SSE = sum((SimValues - M).^2);
RMSE = sqrt(SSE/numel(time));
SQT = sum((M - mean(M)).^2);
RSq = 1 - SSE./SQT;
T1 = table('Size', [3, 3], 'VariableTypes',
{'double', 'double', 'double'}, 'VariableNames',
{'Biomass', 'Glucose', 'Ethanol'}, 'RowNames', {'RMSEP', 'Error', 'R^2'});
T1(1, :) = num2cell(RMSE);
T1(2, :) = num2cell(SSE./ (max(M) - min(M)) * 100);
T1(3, :) = num2cell(RSq)

```

## References

1. Zobel-Roos S, Schmidt A, Mestmäcker F, Mouellef M, Huter M, Uhlenbrock L, Kornecki M, Lohmann L, Ditz R, Strube J (2019) Accelerating biologics manufacturing by modeling or: is approval under the QbD and PAT approaches demanded by authorities acceptable without a digital-twin? *PRO* 7(2):94
2. Kuchemüller KB, Pörtner R, Möller J (2020) Digital twins and their role in model-assisted design of experiments. In: *Advances in biochemical engineering/biotechnology*. Springer, Berlin. [https://doi.org/10.1007/10\\_2020\\_136](https://doi.org/10.1007/10_2020_136)
3. Nargund S, Guenther K, Mauch K (2019) The move toward biopharma 4.0: In silico biotechnology develops “smart” processes that benefit biomanufacturing through digital twins. *Genet Eng Biotechnol News* 39(6):53–55

4. Luttmann R, Bracewell DG, Cornelissen G, Gernaey KV, Glassey J, Hass VC, Kaiser C, Preusse C, Striedner G, Mandenius C-F (2012) Soft sensors in bioprocessing: a status report and recommendations. *Biotechnol J* 7:1040
5. Schügerl K, Bellgardt KH (2012) *Bioreaction engineering: modeling and control*. Springer, Berlin
6. Schügerl K (2001) Progress in monitoring, modeling and control of bioprocesses during the last 20 years. *J Biotechnol* 85(2):149–173
7. Narayanan H, Luna MF, von Stosch M, Bournazou MNC, Polotti G, Morbidelli M, Butté A, Sokolov M (2019) Bioprocessing in the digital age - the role of process models. *Biotechnol J* 761. <https://doi.org/10.1002/biot.201900172>
8. Kalman RE (1960) A new approach to linear filtering and prediction problems. *Trans ASME J Basic Eng* 82:S.35–S.45
9. Wan EA, van der Merwe R (2000) The unscented Kalman filter for nonlinear estimation. In: *Proceedings of the IEEE 2000 adaptive systems for signal processing, communications, and control symposium* (Cat. No. 00EX373). IEEE, pp 153–158
10. Matthews M (1990) A state-space approach to adaptive nonlinear filtering using recurrent neural networks. In: *Proceedings IASTED Internat. Symp. artificial intelligence application and neural networks*
11. Boulet G, Kerr Y, Chehbouni A, Kalma JD (2002) Deriving catchment-scale water and energy balance parameters using data assimilation based on extended Kalman filtering. *Hydrol Sci J* 47 (3):449–467
12. Krämer S, Grum M, Verworn HR, Redder A (2005) Runoff modelling using radar data and flow measurements in a stochastic state space approach. *Water Sci Technol* 52(5):1–8
13. Williams M, Schwarz PA, Law BE, Irvine J, Kurpius MR (2005) An improved analysis of forest carbon dynamics using data assimilation. *Glob Chang Biol* 11(1):89–105
14. Annan JD, Hargreaves JC, Edwards NR, Marsh R (2005) Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Model* 8 (1–2):135–154
15. Julier SJ, Uhlmann JK (1997) New extension of the Kalman filter to nonlinear systems. In: *Signal processing, sensor fusion, and target recognition VI*, vol 3068. International Society for Optics and Photonics, pp 182–193
16. Evensen G (1994) Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys* 99(C5):10.143–10.162
17. Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble Kalman filter technique. *Mon Weather Rev* 126(3):796–811
18. van der Merwe R (2004) *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models*. Doctoral dissertation, OGI School of Science and Engineering at OHSU
19. Lee SC, Hwang YB, Chang HN, Chang YK (1991) Adaptive control of dissolved oxygen concentration in a bioreactor. *Biotechnol Bioeng* 37(7):597–607
20. Ghouli M, Dardenne M, Fonteix C, Marc A (1991) Extended Kalman filtering technique for the on-line control of OKT3 hybridoma cultures. *Biotechnol Tech* 5(5):367–370
21. Dubach AC, Märkl H (1992) Application of an extended kalman filter method for monitoring high density cultivation of *Escherichia coli*. *J Ferment Bioeng* 73(5):396–402
22. Gudi R, Shah S (1993) The role of adaptive multirate Kalman filter as a software sensor and its application to a bioreactor. *IFAC Proc* 26(2):249–254
23. Gudi R, Gray I, Shah S (1993) Multi-rate estimation and monitoring of process variables in a bioreactor. In: *Proceedings of IEEE international conference on control and applications*, IEEE
24. Albiol J, Robusté J, Casas C, Poch M (1993) Biomass estimation in plant cell cultures using an extended Kalman filter. *Biotechnol Prog* 9(2):174–178
25. Gudi RD, Shah SL, Gray MR (1995) Adaptive multirate state and parameter estimation strategies with application to a bioreactor. *AIChE J* 41(11):2451–2464
26. Petrova M, Georgieva O, Patarinska T (1995) State and time delay estimation of continuous microorganisms cultivation. *Bioprocess Eng* 12(1–2):103–107

27. Zorzetto L, Wilson J (1996) Monitoring bioprocesses using hybrid models and an extended Kalman filter. *Comput Chem Eng* 20:S689–S694
28. Holwill IJ, Chard SJ, Flanagan MT, Hoare M (1997) A Kalman filter algorithm and monitoring apparatus for at-line control of fractional protein precipitation. *Biotechnol Bioeng* 53(1):58–70
29. Hrnčič P, Náhlík J, Havlena V (1998) State estimation of Baker's yeast fed-batch cultivation by extended Kalman filter using alternative models. *IFAC Proc* 31(11):601–606
30. Ganovski L, Bliznakova M, Patarinska T (1999) State estimation of a Uricase production process with *Candida utilis*. *Bioprocess Eng* 21(3):273–277
31. Bogaerts P (1999) A hybrid asymptotic-Kalman observer for bioprocesses. *Bioprocess Eng* 20(3):249–255
32. Hitzmann B, Broxtermann O, Cha Y-L, Sobieh O, Stärk E, Scheper T (2000) The control of glucose concentration during yeast fed-batch cultivation using a fast measurement complemented by an extended Kalman filter. *Bioprocess Eng* 23(4):337–341
33. Arndt M, Hitzmann B (2001) Feed forward/feedback control of glucose concentration during cultivation of *Escherichia coli*. *IFAC Proc* 34(5):403–407
34. Longhi L, Marcon S, Trierweiler J, Secchi A (2002) State estimation of an experimental bioreactor using the extended Kalman filtering technology. *IFAC Proc* 35(1):379–382
35. Patnaik P (2003) On the performances of noise filters in the restoration of oscillatory behavior in continuous yeast cultures. *Biotechnol Lett* 25(9):681–685
36. Cha Y-L, Hitzmann B (2004) Ultrasonic measurements and its evaluation for the monitoring of *Saccharomyces cerevisiae* cultivation. *Bioautomation* 1:16–29
37. Arndt M, Hitzmann B (2004) Kalman filter based glucose control at small set points during fed-batch cultivation of *Saccharomyces cerevisiae*. *Biotechnol Prog* 20(1):377–383
38. Arndt M, Kleist S, Miksch G, Friehs K, Flaschel E, Trierweiler J, Hitzmann B (2005) A feedforward–feedback substrate controller based on a Kalman filter for a fed-batch cultivation of *Escherichia coli* producing phytase. *Comput Chem Eng* 29(5):1113–1120
39. Patnaik PR (2005) The extended Kalman filter as a noise modulator for continuous yeast cultures under monotonic, oscillating and chaotic conditions. *Chem Eng J* 108(1–2):91–99
40. Rocha I, Veloso AC, Ferreira E (2006) Design of estimators for specific growth rate control in a fed-batch *E. coli* fermentation
41. Henry O, Kamen A, Perrier M (2007) Monitoring the physiological state of mammalian cell perfusion processes by on-line estimation of intracellular fluxes. *J Process Control* 17(3):241–251
42. Soons Z, Shi J, Van der Pol L, Van Straten G, Van Boxtel A (2007) Biomass growth and kLa estimation using online and offline measurements. *IFAC Proc* 40(4):85–90
43. Klockow C, Hüll D, Hitzmann B (2008) Model based substrate set point control of yeast cultivation processes based on FIA measurements. *Anal Chim Acta* 623(1):30–37
44. Veloso AC, Rocha I, Ferreira E (2009) Monitoring of fed-batch *E. coli* fermentations with software sensors. *Bioprocess Biosyst Eng* 32(3):381–388
45. Jianlin W, Liqiang Z, Tao Y (2010) On-line estimation in fed-batch fermentation process using state space model and unscented Kalman filter. *Chin J Chem Eng* 18(2):258–264
46. Jianlin W, Xuying F, Liqiang Z, Tao Y (2010) Unscented transformation based robust kalman filter and its applications in fermentation process. *Chin J Chem Eng* 18(3):412–418
47. Bavdekar VA, Prakash J, Patwardhan SC, Shah SL (2011) Moving window ensemble Kalman filter for delayed and multi-rate measurements. *IFAC Proc* 44(1):11997–12002
48. Dewasme L, Goffaux G, Hantson A-L, Wouwer AV (2013) Experimental validation of an Extended Kalman Filter estimating acetate concentration in *E coli* cultures. *J Process Control* 23(2013):148–157
49. Popova S, Ignatova M, Lyubenova V (2013) State and parameters estimation by extended Kalman filter for studying inhomogeneous dynamics in industrial bioreactors
50. Sbarciog M, Coutinho D, Wouwer AV (2014) A simple output-feedback strategy for the control of perfused mammalian cell cultures. *Control Eng Pract* 32:123–135

51. Fernandes S, Richelle A, Amribt Z, Dewasme L, Bogaerts P, Wouwer AV (2015) Extended and unscented Kalman filter design for hybridoma cell fed-batch and continuous cultures. IFAC-Papers 48(8):1108–1113
52. Dewasme L, Fernandes S, Amribt Z, Santos LO, Bogaerts P, Wouwer AV (2015) State estimation and predictive control of fed-batch cultures of hybridoma cells. J Process Control 30:50–57
53. Zhao L, Wang J, Yu T, Chen K, Liu T (2015) Nonlinear state estimation for fermentation process using cubature Kalman filter to incorporate delayed measurements. Chin J Chem Eng 23(11):1801–1810
54. Krämer D, King R (2016) On-line monitoring of substrates and biomass using near-infrared spectroscopy and model-based state estimation for enzyme production by *S. cerevisiae*. IFAC-Papers 49(7):609–614
55. Simutis R, Lübbert A (2017) Hybrid approach to state estimation for bioprocess control. Bioengineering 4(1):21
56. Krishna VV, Pappa N, Rani SJV (2018) Implementation of embedded soft sensor for bioreactor on Zynq processing system. In: 2018 international conference on recent trends in electrical, control and communication (RTECC), IEEE
57. Krämer D, King R (2019) A hybrid approach for bioprocess state estimation using NIR spectroscopy and a sigma-point Kalman filter. J Process Control 82:91–104
58. Ritschel TK, Boiroux D, Nielsen MK, Huusom JK, Jørgensen SB, Jørgensen JB (2019) The extended Kalman filter for nonlinear state estimation in a U-loop bioreactor. In: 2019 IEEE conference on control technology and applications (CCTA), IEEE
59. Feidl F, Garbellini S, Luna MF, Vogg S, Souquet J, Broly H, Butté A (2019) Combining mechanistic modeling and Raman spectroscopy for monitoring antibody chromatographic purification. PRO 7(10):683
60. Lisci S, Grosso M, Tronci S (2020) A geometric observer-assisted approach to tailor state estimation in a bioreactor for ethanol production. PRO 8(4):480
61. Sonnleitner B (2013) Automated measurement and monitoring of bioprocesses: key elements of the M<sup>3</sup>C strategy. In: Measurement, monitoring, modelling and control of bioprocesses. Springer, Berlin, pp 1–33
62. Biechele P, Busse C, Solle D, Scheper T, Reardon K (2015) Sensor systems for bioprocess monitoring. Eng Life Sci 15(5):469–488
63. Vojinović V, Cabral JMS, Fonseca LP (2006) Real-time bioprocess monitoring: part I: in situ sensors. Sensors Actuators B Chem 114(2):1083–1091
64. Chhatre S (2012) Modelling approaches for bio-manufacturing operations. In: Measurement, monitoring, modelling and control of bioprocesses. Springer, Berlin, pp 85–107
65. Monod J (1949) The growth of bacterial cultures. Annu Rev Microbiol 3(1):371–394
66. Henson MA, Seborg DE (1992) Nonlinear control strategies for continuous fermenters. Chem Eng Sci 47(4):821–835
67. Jones KD, Kompala DS (1999) Cybernetic model of the growth dynamics of *Saccharomyces cerevisiae* in batch and continuous cultures. J Biotechnol 71(1–3):105–131
69. Contois DE (1959) Kinetics of bacterial growth: relationship between population density and specific growth rate of continuous cultures. Microbiology 21(1):40–50
69. Yousefi-Darani A, Paquet-Durand O, Babor M, Hitzmann B (2020) Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in *Saccharomyces cerevisiae* batch cultivation. Biosyst Eng 198(2020):198–209
70. Galvanauskas V, Simutis R, Levišauskas D, Urniežius R (2019) Practical solutions for specific growth rate control systems in industrial bioreactors. PRO 7(10):693
71. Oisiović RM, Cruz SL (2000) State estimation of batch distillation columns using an extended Kalman filter. Chem Eng Sci 55(20):4667–4680
72. Hashemi R, Engell S (2016) Effect of sampling rate on the divergence of the extended Kalman filter for a continuous polymerization reactor in comparison with particle filtering. IFAC-Papers 49(7):365–370