How NOT To Make the Joint Extended Kalman Filter Fail with Unstructured Mechanistic Models

Cristovao Freitas Iglesias Jr[1,*] **and Miodrag Bolic**[1]

[1]School of Electrical Engineering and Computer Science (EECS), University of Ottawa, Ottawa, ON K1N 6N5, Canada
[*]cfrei096@uottawa.ca

ABSTRACT

Unstructured Mechanistic Model (UMM) allows for modeling the macro-scale of a phenomenon without known mechanisms. This is extremely useful in biomanufacturing because using UMM in the Joint estimation of states and parameters with Extended Kalman Filter (JEKF) can enable the real-time monitoring of bioprocess with unknown mechanisms. However, the UMM commonly used in biomanufacturing contains Ordinary Differential Equations (ODEs) with unshared parameters, weak variables, and weak terms. When such a UMM is coupled with an initial state error covariance matrix $\mathbf{P}(t=0)$ and a process error covariance matrix **Q** with uncorrelated elements, along with just one measured state variable, the Joint Extended Kalman Filter (JEKF) fails to estimate the unshared parameters and state simultaneously. This is because the Kalman gain corresponding to the unshared parameter remains constant and equals zero. In this work, we formally describe this failure case, present the proof of JEKF failure, and propose an approach called SANTO to side-steps this failure case. The SANTO approach consists of adding a small quantity to state error covariance between the measured state variable and unshared parameter in the initial **P**(t=0) of the matrix Ricatti differential equation to compute the predicted error covariance matrix of state and prevent the Kalman gain from being zero. Our theoretical and empirical results showed the potential of SANTO with synthetic and real datasets.

Introduction

The Extended Kalman Filter (EKF) is a recursive Bayesian filter[1,2]. This Nonlinear State Estimator (NSE) is a commonly used technique for estimating the state of a nonlinear system using a state-space model, first-order linearization, and linear estimation theory. It is composed of a process model and a measurement model, along with error covariance matrices of the process (**Q**), measurement (**R**), and state (**P**)[3,4]. EKF, beyond state estimate, is also used for parameter estimation (parameter evolution[5]) of nonlinear systems (process model) considering a single joint state variable vector which includes both states and parameters of the process model[6–8]. This approach is called **J**oint estimation of states and parameters with **E**xtended **K**alman **F**ilter (JEKF). The joint estimation problem is motivated by the need to correct the prediction of a process model regarding state variables and to update the process model by evolving its parameters based on the corrections done[8]. A process model should be estimated (evolved) for different conditions of the same application. For example, in biomanufacturing, the parameters of a process model for monitoring a cell culture should change for each new condition. We can use a general set of parameters at the beginning of the process, but we need to evolve them during the process to improve the predictions about the states of the cell culture. So, JEKF uses each measurement as soon as it becomes available to correct both the predictions and parameters of a process model[8]. The first discussions and applications of the JEKF approach started in the sixties (60's) for the estimation of linear systems (in which there is a bilinear relation between the states and parameters)[6–10]. However, JEKF is still very popular with several new applications in different areas[5,11–21], and with unsolved problems[22,23]. Furthermore, JEKF has been established as the least expensive nonlinear estimator for moderate-size systems in terms of computational cost because the practical implementation of adaptive controllers using micro-controllers (or/and minicomputers or/and microprocessor) requires numerically economical and robust algorithms such as the JEKF[11,24]. An important area of application of JEKF is biomanufacturing, that is, the production of biological products from living cells[25–27]. The reason is that the JEKF with Mechanistic Model (MM) as a process model can enable real-time monitoring of critical process parameters (CPP) or critical quality attributes (CQA) that are difficult to be measured directly or that can only be measured at low sampling frequency in a bioprocess[25,28]. There are two types of MM: Structured Mechanistic Model (SMM) and Unstructured Mechanistic Model (UMM)[29]. When we have knowledge about the bioprocess, we can use an SMM with JEKF. On the other hand, when we do not know it, we can use a UMM with JEKF because the UMM allows modeling the macro-scale of a phenomenon. It is a mass-balance equation system with few parameters and variables and less complexity than SMM[30,31].

The UMM used in biomanufacturing typically consists of ODEs with unshared parameters, weak variables, and weak terms. However, this characteristic of UMM in biomanufacturing, along with the use of P(t=0) and Q with uncorrelated elements, as well as the presence of only one measured state variable, represents a failure case where the JEKF cannot estimate the unshared parameters and the state simultaneously. There are a lot of new bioprocesses without prior knowledge about them in the literature that the biopharmaceutical industry aims to monitor, such as the recombinant Adeno-Associated Virus (rAAV) production[32]. Therefore, enabling JEKF to side-step the failure case described above may allow the industry to perform biomanufacturing with real-time monitoring of bioprocess with unknown mechanisms. Consequently, this skill can help the biopharmaceutical industry to achieve biomanufacturing 4.0, becoming more agile and intelligent to enhance product quality, optimize operations, and reduce costs[26,27,33,34]. Although the biopharmaceutical industry was valued at USD 239.8 billion in 2019 and is estimated to grow at an annual rate of over 13%, it faces significant challenges in achieving desired productivity and product quality consistently[35].

In this work, we presented the common conditions in biomanufacturing that represents a failure case where JEKF fails to perform unshared parameter evolution of UMM, and we proposed a solution to side-step this failure case, called SANTO, that consists of a Specific initiAl coNdiTiOn (SANTO) for the Matrix Ricatti Differential Equation (MRDE). Our solution is inspired by the regularization technique to avoid singularity issues in EKF. However, instead of adding a small quantity to the diagonal elements of the state error covariance matrix $\mathbf{P}$[36], we only add a small quantity to state error covariance between the Measured State Variable (MSV) and an Unshared Parameter (UP) in P(t=0) for the MRDE. The proposed approach can avoid the JEKF failure by preventing the Kalman gain from being zero in the entire process, an unrealistic situation that would mean that the predictions of the UMM (used as a process model) are perfect. Our theoretical and empirical results demonstrated the effectiveness of SANTO, which was assessed using synthetic and real datasets. Furthermore, SANTO outperformed a recent approach of literature[19] by achieving better estimations with faster parameters evolution and requiring less adjustment. The code and data used in this work are available on Additional information Section to facilitate reproducibility. Our contributions can be summarized as follows:

- The proof of JEKF failure as an unshared parameter estimator with the biomanufacturing conditions representing a failure case. To our knowledge, this is the first work to formally report this failure case regarding JEKF.
- An approach to avoid the JEKF failure that enables using JEKF with UMM for real-time bioprocess monitoring. This is helpful in the macro-scale modeling of a phenomenon with UMM where the underlying process mechanism is not fully understood.

Related work

The main limitation of JEKF is not guaranteed convergence in some cases as reported by[6,24,37]. A solution to deal with the convergence problems of JEKF is to use recurrent derivatives[6,37]. However, a theoretical justification for that was not provided[8]. On the other hand, it was reported that the cause of divergence in JEKF is linked to the linearization of the coupled system and not due to the lack of recurrent derivatives[24]. Furthermore, there are certain cases where the JEKF may not be able to estimate the parameters and the state simultaneously, such as, singularity issues[36]. However, so far, the failure case (biomanufacturing conditions) where JEKF fails as an unshared parameter estimator has not been formally reported. Recently, the JEKF was applied for monitoring rAAV production[19]. In developing this application, the authors dealt with a situation that resembles to be the failure case reported here. Because they reported the use of a simple UMM, $\mathbf{P}$(t=0), and $\mathbf{Q}$ with uncorrelated elements and a second linear operator as an approach to enable Kalman gain ($\mathbf{K}$) and $\mathbf{P}$ to be updated with prior error covariances with regards to the UMM parameters, their results showed the unshared parameter evolution with convergence. However, the authors did not describe the problem in detail and did not present a theoretical justification for the approach used (second linear operator), despite clearly stating that the work is an initial study and fairly reporting the need for validation. We named this approach KPH2 due to using a second linear operator to enable $\mathbf{K}$ and $\mathbf{P}$ to be updated and because we can refer to this approach in our experimental evaluation for comparison purposes with our proposed approach. A description of KPH2 and possible interpretation can be found in the supplementary information Section S7.

Background

Unstructured Mechanistic Model (UMM)

Unstructured Mechanistic Models (or Unstructured Mechanistic Kinetic Models) are models of the temporal evolution of a bioprocess[38]. They are based on first-principle mechanisms that drive the bioprocess under consideration[35]. Examples of bioprocesses are i) the production of therapeutic monoclonal antibodies (mAbs), which is projected to bring in USD 300 billion by 2025[35], and ii) the rAAV production that is a viral vector technology for gene therapy considered the safest and most effective way to repair single-gene abnormalities in non-dividing cells[19,32]. It is essential to point out that despite UMM being the most suitable option to describe the dynamic behavior of bioprocesses and be considered a crucial foundation for DT development,

its industrial use is still in its early stages[29,38,39]. The UMM are important because they allow for the macro-scale modeling of the bioreactor's functionality and can provide insight into the upstream process's underlying macro-scale phenomena. For example, this kind of model can be used to depict the dynamics of the cell density, viability, nutrient/metabolite concentrations, and product titer[40–42]. Therefore, UMM is the most suitable option for explaining observed phenomena, predicting process behavior, and analyzing intrinsic bioprocess characteristics like controllability[35].

The main difference between UMM and SMM is that SMM is more complex than UMM because it provides details about the intracellular environment of a homogenous cell population. Therefore, developing SMM for a specific bioprocess requires extensive domain knowledge and substantial effort[35,40]. SMM is unsuitable for dynamic control of bioprocess in bioreactors used commonly in biomanufacturing because many of the variables used in SMM cannot be manipulated directly[35]. SMM is most suited for cell-line development, in which cells' genome-level properties are changed to produce desired process behavior[35].

It is essential to point out that a simple UMM has limited predictive power and is insufficient to process state estimation. Moreover, it is improbable that a single set of parameter values enables a kinetic model to satisfy several data sets collected under distinct operating circumstances[43]. The Kalman filter approach is commonly implemented with UMM [44] to improve prediction accuracy and generate predictions between sampling instances. Among several data analysis methods, the Kalman filter and its non-linear extensions, such as the extended Kalman filter, are effective tools for predicting the values of unobserved states. Examples of UMM used in biomanufacturing can be found in supplementary information Section S1.

Continuous-Discrete Extended Kalman Filter

This section gives an overview of the Continuous-Discrete EKF (CD-EKF) algorithm. A detailed description of CD-EKF can be found in supplementary information Section S2. The EKF requires a state–space model to perform estimation on the state variables of a process (nonlinear system) present in a state variable vector $\psi(t)$[1,43,45]. A state-space model consists of process and measurement (observation) models[46]. EKF linearizes the nonlinear system (state–space model) by calculating the Jacobians of the nonlinear process and measurement models based on the first-order Taylor series expansion in order to analytically propagate the Gaussian random-variable representation[8,20,43].

Process Model: An UMM can be used as the process model of EKF. The state variables vector to be used by the EKF is composed of the state variables of the UMM (observed and unobserved) and the state variables vector is defined as:

$$\psi(t) = [x_1, x_2, ..., x_n]^T. \tag{1}$$

Subsequently, the process model is represented as

$$\frac{d\psi(t)}{dt} = \phi(\psi(t), t, \theta) + \omega(t), \tag{2}$$

where ϕ denotes non-linear functions of the state variables in $\psi(t)$, which corresponds to an UMM. The process model is formulated in a continuous time t and the white process noise vector is represented by $\omega \sim \mathcal{N}(0, \mathbf{Q})$, with zero mean and error covariance matrix of process model represented by $\mathbf{Q}$.

Measurement Model: The measurement model is treated as a discrete system and defined as

$$\mathbf{Z}_k = h(\psi(t_k)) + v. \tag{3}$$

The non-linear function h in the measurement model relates the current state variables to the measurements $\mathbf{Z}_k$. The white measurement noise vector is represented by $v \sim \mathcal{N}(0, \mathbf{R})$, with zero mean and measurement noise variance represented by $\mathbf{R}$. When some state variables can be measured directly, we have a simple case and h can be a linear model. If h is linear, we have $h(\psi(t_k)) = \mathbf{H}\psi(t_k)$[20,45,47]. Where the matrix $\mathbf{H}$ is a linear operator (row vector) that matches the states variables of $\psi(t_k)$ to the measured variables $\mathbf{Z}_k$ that are obtained at a discrete instance k [20,47]. Consequently, the measurement model (3) can be rewritten as

$$\mathbf{Z}_k = \mathbf{H}\psi(t_k) + v. \tag{4}$$

The EKF algorithm is implemented through a state variables vector $\psi(t)$, initial condition, prediction step (time update) and correction step (measurement update)[1,20,21,45,47].

Initialization step: The initial condition is composed of the initial mean $\hat{\psi}_0 = E[\psi_0]$, and initial error covariance matrix $\mathbf{P}_0 = \mathbf{P}(t=0) = E[(\psi_0 - \hat{\psi}_0)(\psi_0 - \hat{\psi}_0)^T]$ of state variables vector in addition to the error covariance matrices of the process $\mathbf{Q}$ and measurement $\mathbf{R}$[8].

Prediction step: In this step, the a priori predictions represented by the predicted mean $\hat{\psi}(t_{k/k-1})$ and predicted error covariance matrix $\mathbf{P}(t_{k|k-1})$ of state variables vector $\psi(t)$ are obtained respectively by numerically integrating $\phi(\psi(t), t, \theta)$ from discrete time t_{k1} to t_k the following equation

$$\hat{\psi}(t_{k/k-1}) = \hat{\psi}(t_{k-1}) + \int_{t_{k-1}}^{t_k} \phi(\hat{\psi}(t))dt \Big|_{\hat{\psi}(t_{k-1})} \tag{5}$$

and solving the MRDE to predict the state error covariance matrix[4,48]

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{J}_t^{\phi}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{J}_t^{\phi T} + \mathbf{Q} \tag{6}$$

from t_{k-1} to t_k, where a new measurement is obtained at time k[4,49,50]. The Equation 6 is basically a matrix of ODEs, and the matrix of ODEs solutions obtained from t_{k-1} to t_k represent each error covariance of the system state.

Correction step: In this step, the results of the prediction step ($\hat{\psi}(t_{k/k-1})$ and $\mathbf{P}(t_{k|k-1})$) are combined with the measured value $\mathbf{Z}_k$ and Kalman gain ($\mathbf{K}_k$) to provide the estimated mean $\hat{\psi}(t_{k/k})$ and estimated error covariance matrix $\mathbf{P}(t_{k|k})$ of state variables vector using the following equations:

$$\mathbf{K}_k = \mathbf{P}(t_{k|k-1})\mathbf{H}^T(\mathbf{H}\mathbf{P}(t_{k|k-1})\mathbf{H}^T + \mathbf{R})^{-1} \tag{7}$$

$$\hat{\psi}(t_{k/k}) = \hat{\psi}(t_{k/k-1}) + \mathbf{K}_k(\mathbf{Z}_k - \mathbf{H}\hat{\psi}(t_{k/k-1})) \tag{8}$$

$$\mathbf{P}(t_{k|k}) = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}(t_{k|k-1}) \tag{9}$$

The Kalman gain is a scaling factor (ratio) to estimate the state variables by setting a value between the predicted state and measured state[4,51]. The $\mathbf{K}_k$ chooses a value along the residual range ($\mathbf{Z}_k$ - $\mathbf{H}\hat{\psi}(t_{k/k-1})$)[8,51]. $\mathbf{K}_k$ enables to set a value for $\hat{\psi}(t_{k/k})$ between the $\hat{\psi}(t_{k/k-1})$ (prediction) and $\mathbf{Z}_k$ (measurement) using Equation 8, and update the belief regards the state variables based on how certain we are regards the measurement using the Equation 9[51]. The Kalman gain is computed as a ratio of prior and measurement uncertainty available; see Equation 7. The one dimensional form Equation 7 is the following $K = P/(P+R)$[51]. It is important to point out that linear operator $\mathbf{H}$ matches the states variables of $\psi(t_k)$ to the measured variables $\mathbf{Z}_k$ that are obtained at a discrete instance.

Using the estimated mean $\hat{\psi}(t_{k/k})$ and estimated error covariance matrix $\mathbf{P}(t_{k|k})$ state variables vector as an initial condition, we can return to the prediction step until the next measurement be obtained and everything repeated again.

JEKF

JEKF is a Bayesian filter-based joint estimation approach where the states x_i and parameters θ of a process model are concatenated into a single joint state vector[52]. Then, the state variables vector ($\psi(t) = [x_1, x_2, ..., x_n]^T$) is considered as extended/augmented as following,

$$\psi(t) = [x_1, x_2, ..., x_n, \theta_1, ..., \theta_n]^T. \tag{10}$$

To be more specific, we consider the problem of learning both the states x_i and parameters θ of a discrete-time nonlinear dynamical system (such as the UMM described in supplementary information Section S1) used as a process model. In JEKF, the system states $\mathbf{x}$ and the set of model parameters θ for the dynamical system are simultaneously corrected based only on the observed noisy signal Z_k. It is essential to point out that we consider JEKF as an approach for parameter evolution[5]. It cannot guarantee convergence in some cases[6]. However, it can guarantee the evolution of the parameters based on the following equation[5]

$$\theta(t_k) = \theta(t_{k-1}) + noise, \tag{11}$$

where the parameters are defined as random variables with perturbation (noise) added at each time step. This parameter evolution can be enough to update the process model parameters when we are near the optimal parameters regarding a specific condition. In this paper, when we say parameter estimation, we are referring to parameter evolution.

Theoretical Analysis

This Section presents the theoretical analysis of the JEKF failure to perform unshared parameter evolution with a UMM and one possible approach (SANTO) for this problem.

JEKF failure

First, we present the conditions where JEKF fails to estimate (parameter evolution) the unshared parameters of a UMM. Next, we present the theoretical proof of the failure. However, before starting the analysis, we formally define unshared parameters and weak and strong terms/variables of an ODE as follows:

- **Unshared parameters:** They are parameters used only in one term of an ODE and not used by other ODEs of the same UMM. See the example in supplementary information Section S3.1.
- **Weak and Strong term of an ODE:** A *weak term* is a term of an ODE with a low percentage of variables of state variable vector, and a *"strong term"* is one with a high percentage of variables of state variable vector. See the example in supplementary information Section S3.2.
- **Weak and Strong variable of an ODE:** A *weak variable* is a variable used only in the first member of an ODE in UMM, and a *strong variable* is a variable used in the first member and different terms of the second member of an ODE. Furthermore, it is used in the second member of other ODEs of the same UMM. See the example in supplementary information Section S3.3.

Failure Case: Biomanufacturing conditions

The following conditions are prevalent in biomanufacturing and should be taken into consideration while developing JEKF applications for this area:

- **P and Q with uncorrelated elements.** In case of the limited amount of data, it is very common to assume **P** and **Q** with uncorrelated elements in EKF applications[19–21,47]. This assumption means that the error covariance matrices **P** and **Q** are diagonal, with the diagonal elements being the noise variances ($P_{i,i} \neq 0$ and $Q_{i,i} \neq 0$) and off-diagonal elements equal to zero ($P_{i,j} = 0$ and $Q_{i,j} = 0$). The **Q** constant and with uncorrelated elements is used only to build the MRDE, and the **P** with uncorrelated elements can be used to build an MRDE and as an initial condition of MRDE (the initial predicted state error covariance **P**(t=0)).

  This assumption raises two scenarios:

  1. The use of **P** with uncorrelated elements to build the MRDE (Equation 6) and **P**(t=0) with uncorrelated elements as initial condition. When **P** with uncorrelated elements is used to build the MRDE, the ODEs of MRDE are based only on noise variance of $P_{i,i}$ and $Q_{i,i}$ and elements of Jacobian $\mathbf{J}_t^{\phi}$. See the example in supplementary information Section S3.4. It is important to point out that depending on partial derivative, the ODE to predict a state error covariance can be time-invariant $\frac{dP_{i,j}(t_{k|k-1})}{dt} = 0$. See the supplementary information Section S3.2.
  2. The use of **P** with correlated elements to build the MRDE (Equation 6) and **P**(t=0) with uncorrelated elements as initial condition . This means that the ODE of MRDE can be composed of off-diagonal elements of **P**, and it can reduce the number of the time-invariant ODE to predict a state error covariance between two state variables.

- **ODEs of UMM with weak terms.** A strong term contributes more than a weak term to compute the predicted state error covariance $\mathbf{P}(t_{k|k-1})$. Many elements of Jacobian $\mathbf{J}_t^{\phi}$ result from the partial derivation of a strong term. See the example in supplementary information Section S3.2.

- **ODEs of UMM with weak variables.** In the Jacobian $\mathbf{J}_t^{\phi}$, the first-order partial derivatives of all functions with respect to a *weak variable* are equal to zero. Consequently, this variable type does not contribute to the calculations of predicted error covariance $\mathbf{P}(t_{k|k-1})$ since it will not be part of any element of MRDE to predict the state error covariance matrix $\mathbf{P}(t_{k|k-1})$. On the other hand, a *strong variable* contributes to the calculations of predicted error covariance $\mathbf{P}(t_{k|k-1})$. See the example in supplementary information Section S3.3.

- **Only one measured state variable.** In some cases (JEKF application), measuring only one state variable is possible. This measured state variable determines which column of predicted state error covariance $\mathbf{P}(t_{k|k-1})$ is used to compute the Kalman gain through $\mathbf{P}(t_{k|k-1})\mathbf{H}^T$ in the Equation 7. If this column has a row with a value equal to zero (no covariance between the measured variable and state variable represented by the row), the Kalman gain cannot be computed to the state variable defined by the row. See the example in supplementary information Section S3.5.

- **ODEs of UMM with unshared parameters.** This type of parameter is commonly used in ODE to model the dynamic of product formation in biomanufacturing[53–55]. See the example in supplementary information Section S3.1.

Lemma: Inability to Update Kalman Gain for Unshared parameters based P and Q with uncorrelated elements

Given the conditions described above, we have the following Lemma:

The Kalman gain cannot be updated (by Eq 7) for an unshared parameter that is part of a state variable vector and part of a weak term in an UMM, if the initial state error covariance matrix $\boldsymbol{P}$*(t=0) and* $\boldsymbol{Q}$ *are formed by uncorrelated elements and there is only one state variable measured.* See the proof and an example in the supplementary information Section S4.

Theorem: JEKF failure

The consequence of the Lemma described above is the following Theorem:

The JEKF (as described in previous Sections) fails to estimate an unshared parameter (parameter evolution) that is part of a state variable vector and part of a weak term in a UMM if the initial state error covariance matrix $\boldsymbol{P}$*(t=0) and* $\boldsymbol{Q}$ *are composed of uncorrelated elements, and there is only one state variable measured. This is because the Kalman gain value for the unshared parameter is equal to zero for all steps of execution of the JEKF algorithm.*

The theoretical proof and an example of this theorem can be found in the supplementary information Section S5.

SANTO: Specific initial condition for MRDE ($P_{MSV,UP}(t=0) \neq 0$ *in* $\mathbf{P}_0$)

This section presents the SANTO approach to avoid the JEKF failure described in Theorem above. The initial condition of MRDE is the initial state error covariance matrix $\mathbf{P}_0 = \mathbf{P}(t=0)$. When it has only uncorrelated elements, $\mathrm{P}_{i,j} = 0$, some initial conditions of ODEs in the MRDE are zero, and consequently, the obtained solutions from t_{k-1} to t_k for some of these ODEs are zero too. Furthermore, in the presence of the biomanufacturing conditions (failure case presented above), we have that the Kalman gain value regarding the unshared parameter (K_{UP}) and the predicted state error covariance between the unique measured state variable and the unshared parameter ($P_{MSV,UP}(t_{k|k-1})$), are zero too, $K_{UP} = 0$ and $P_{MSV,UP}(t_{k|k-1}) = 0$. Then, the K_{UP}, and $P_{MSV,UP}(t_{k|k-1})$ that compose $\mathbf{P}(t_{k|k-1})$ cannot be updated regards to the unshared parameter (see the Lemma above), and they are constant and equal to zero during the entire process execution of JEKF. It is worth noting that $P_{MSV,UP}(t_{k|k-1})$ is an element of $\mathbf{P}(t_{k|k-1})$ such as $P_{MSV,UP}(t=0)$ is an element of $\mathbf{P}(t=0)$. Furthermore, the $K_{UP} = 0$ during the entire JEKF execution reflect an unrealistic situation. This would mean that the prediction regarding the unshared parameter is perfect and does not need the influence of the measurement in the correction step of JEKF since there is no uncertainty in the prediction regarding the unshared parameter. This reflects the second intuition behind Kalman gain described in the supplementary information Section S2. However, based on prior knowledge, we know that the process model predictions regarding the unshared parameter are imperfect since we need to perform the evolution of the unshared parameter; otherwise, they would be the same during the entire process. Therefore, we need $K_{UP} \neq 0$ and $P_{MSV,UP}(t_{k|k-1}) \neq 0$.

In general, the initial condition of MRDE is $\mathbf{P}(t=0)$ with uncorrelated elements ($\mathrm{P}_{i,j} = 0$) due to the difficulty of estimating all covariances with a limited dataset. However, instead, considering all off-diagonal elements of $\mathbf{P}(t=0)$ equal zero ($\mathrm{P}_{i,j} = 0$), we can consider only the key off-diagonal element (that is the $P_{MSV,UP}(t=0)$) with an initial value different of zero ($P_{MSV,UP}(t=0) \neq 0$) to avoid the failure case. This value could be a small quantity, λ, because, based on prior knowledge described above, we know that this value should be different from zero and small enough to not significantly affect the filter's estimates but large enough to prevent the failure case. Then, with this consideration, we can have a value for the initial state error covariance between the MSV and an UP, $P_{MSV,UP}(t=0)$, and if we add it in initial state error covariance matrix $\mathbf{P}(t=0)$ with the others uncorrelated elements, we have a specific initial condition for MRDE that enables to update the K_{UP} and $P_{MSV,UP}(t_{k|k-1})$ present in $\mathbf{P}(t_{k|k-1})$ and, consequently, avoids the JEKF failure. The proof of this approach can be found in supplementary information Section S6.

It is essential to point out that the SANTO is inspired by the idea of regularization technique used to avoid the singularity problem in the state error covariance matrix[36,56]. However, instead of adding a small quantity (such as 1×10^{-6}) to the diagonal

elements of the state error covariance matrix **P**, such as the perturbed-P algorithm[36], we only add a small quantity (λ) to the $P_{MSV,UP}(t=0)$ in $\mathbf{P}(t=0)$ to initialize the MRDE. Furthermore, a small quantity to the $P_{MSV,UP}(t=0)$ can be defined by empirical tuning. One of the most common ways to define a small quantity is by trial and error. This involves running the filter with different values of λ and choosing the value that results in the best performance[57]. This approach can be time-consuming, but it can be useful when dealing with systems with complex dynamics or when the system is unknown.

Empirical Evaluation

In our evaluation, we have the two goals (**G1** and **G2**) that are addressed by answering three Research Questions (**RQs**) comparing three NSEs that are JEKF (classic), JEKF-SANTO and JEKF-KPH2. First, the goals are the following:

- **G1**) Experimentally test the Theorem (JEKF Failure);
- **G2**) Test whether SANTO can avoid the JEKF failure and compare its performance with KPH2.

Lastly, the research questions are the following:

- **RQ1-G1**) Is there any variation in the unshared parameter estimation done by JEKF with the biomanufacturing conditions (failure case), or are the estimations constant in the entire process?
- **RQ2-G2**) Is there any variation in the unshared parameter estimation done by SANTO and KPH2 with the biomanufacturing conditions (failure case), and which one has the best estimations (performance)?
- **RQ3-G2**) Can the SANTO simultaneously estimate more than one unshared parameter, performing better than KPH2?

Experimental Setup

Synthetic Dataset - mAb production

The Synthetic dataset (SD) has data regarding Monoclonal Antibody (mAb) productions that represent the biomanufacturing of a protein widely used as diagnostic reagents and for therapeutic purposes[58]. The SD is composed of two runs (A-SD and B-SD) with different conditions. The runs have different samples regarding the state variables viable cell density (Xv), glucose (GLC), glutamine (GLN), lactate (LAC), ammonium (AMM) and mAb. They were generated using the UMM proposed by[54] with two sets of different parameters see the supplementary information Section S1.4. Run A-SD has a maximum of mAb (titer) equal to 671.0 mg/L obtained with the parameter QmAb equal to 0.79 ($\times 10^{12}$ $g\ cells^{-1}h^{-1}$) representing a condition with pH varying from 7.1 to 6.7 and constant temperature equal to 36º C. On the other hand, run B-SD has a maximum of mAb (titer) equal to 930.9 mg/L obtained with the parameter QmAb equal to 1.00487 ($\times 10^{12}$ $g\ cells^{-1}h^{-1}$) representing the condition of pH and temperature varying from 7.1 to 7 and from 37.5º C to 35º C, respectively. The samples of the runs add up to 6720 with a sample rate of 3 minutes during 336 hours of the process. The run B-SD has the samples regarding Xv (c/mL) without and with Gaussian white noise (standard deviation of 0.5). It is important to point out that Xv samples with Gaussian white noise represent a possible online measurement with sensor errors. The details of the synthetic dataset development can be seen in supplementary information Section 8.1.

Real Dataset : AAV production

The Real dataset (RD) contains data regarding Recombinant Adeno-Associated Virus (rAAV) productions and it is described and available in[19]. rAAV is a viral vector technology for gene therapy that is considered the safest and most effective way to repair single-gene abnormalities in non-dividing cells[59]. The RD has two runs with online and offline measurements of the state variables viable cell density (Xv), glucose (GLC), glutamine (GLN), lactate (LAC), ammonium (AMM) and rAAV (titer) regarding the rAAV production in shake-flasks and in bioreactors. The run A-RD (production in shake-flasks) has only offline measurements and the run B-RD (production in bioreactor) has online measurements of Xv and offline measurements of GLC, LAC and rAAV (titer). The samples of the runs add up to 2902 with a sample rate of 1 minute during 48.3 hours of the process. The details of the real dataset development can be seen in supplementary information Section S8.2

NSEs assessment with synthetic dataset to address RQ1-G1 and RQ2-G2

All NSEs (JEKF (classic), JEKF-SANTO and JEKF-KPH2) used the UMM described in the supplementary information Section S1.4 as process model and the same initial concentration regards the state variables, see supplementary information Table S3. The NSEs were used to correct (estimate) the predictions regarding state variables (Xv and mAb) and to evolve the unshared parameter (QmAb) of the process model. This was done using the Xv samples with the noise of the run B-SD as the unique measured state variable and the parameters used to generate the run A-SD as initial parameters (see supplementary information Table S2). This situation represents a joint estimation problem where the prediction and parameter of the process model should be corrected by the NSEs based on measured state variable Xv. For example, the initial value used of QmAb is the value

of run A-SD (0.79×10^{12} $g\ cells^{-1}h^{-1}$), and it should be evolved to the value of run B-SD (1.00487×10^{12} $g\ cells^{-1}h^{-1}$) based on Xv with the noise of run B-SD. Furthermore, the Xv (without noise) and mAb samples of run B-SD were used as ground truth too. It is important to point out that the estimations were done with MRDE formed by **P** with correlated elements (MRDE-PC) and uncorrelated elements (MRDE-PU). In addition, MRDE-PC and MRDE-PU were combined with standard and specific **P**(t=0) to check the sensitivity of SANTO (regards to $P_{MSV,UP}(t=0)$) and KPH2 (regards to $P_{UP,UP}(t=0)$). The standard **P**(t=0) means that all NSE used the same **P**(t=0). On the other hand, the specific **P**(t=0) means that each NSE used a different **P**(t=0) that enables its best performance. For example, the specific **P**(t=0) for SANTO contains a specific value of $P_{MSV,UP}(t=0)$, and the specific **P**(t=0) for KPH2 includes specific value of $P_{UP,UP}(t=0)$. The specific **P**(t=0) was obtained by trial and error and a standard **Q** was used for all NSEs. In addition, the root means square error (RMSE) was used as a metric to assess the similarity between NSEs estimations and the ground truth of run B-SD. The details about the design of NSEs with SD can be found in the supplementary information Section S8.3.

NSEs assessment with real dataset to address RQ3-G2

The NSEs (JEKF-SANTO and JEKF-KPH2) used the UMM described in the supplementary information Section S1.5 as process model and the same initial concentration regards the state variables, see supplementary information Table S8. These two NSEs were used to correct (estimate) the predictions regards Xv, GLC, LAC and rAAV (titer) and to evolve the unshared parameters (μ_{Lac}, μ_{GLC} and μ_{rAAV}) of the process model. This was done using the Xv samples with the noise of the run B-RD as the unique measured state variable and the parameters obtained with the run A-SD as initial parameters (see supplementary information Table S9). This situation also represents a joint estimation problem where the predictions and parameters of the process model should be corrected simultaneously by the NSEs based on measured state variable Xv. However, in this case, the NSEs have to correct three different unshared parameters simultaneously based on Xv with the noise of run B-RD. Furthermore, the root means square error (RMSE) was used as a metric to assess the similarity between NSEs estimations and the ground truth of run B-SD, which are the offline measurements of GLC, LAC and rAAV (titer) of run B-RD. It is important to point out that the estimations were also done with MRDE-PC and MRDE-PU, and with standard and specific **P**(t=0). In addition, a specific **Q** was used to KPH2 beyond of the standard **Q** used in all NSEs aiming to improve the KPH2 estimations. The details about the design of NSEs with RD can be found in the supplementary information Section S8.4.

Results

The results are organized by research questions **RQ1-G1**, **RQ2-G2** and **RQ3-G2**:

Answer to RQ1-G1. In Figure 1, we reported the estimations done by all of the three NSEs in regards to Xv, mAb and QmAb of mAb production (synthetic dataset) using a MRDE-PC and a specific **P**(t=0) for SANTO, and another one for KPH2. In plot A of Figure 1, we can see that all NSEs estimated the Xv close to the ground truth. However, the JEKF-classic was not able to evolve (update) the unshared parameter QmAb, because the estimations about QmAb were constant and equal to the initial value of 0.79 ($\times 10^{12}$ $g\ cells^{-1}h^{-1}$). Consequently, the JEKF-classic estimation regards mAb were far from the ground truth (see plots B and C in Figure 1), and it had the RMSE value of 132.4 that was the highest value among all NSEs, see Table 1. The same results regarding the JEKF-classic were obtained using MRDE-PU, see supplementary Fig. S2 and Table 1.

Answer to RQ2-G2. In plots B and C of Figure 1, we can see that JEKF-SANTO and JEKF-KPH2 evolved the QmAb from the initial value to the ground truth, and consequently estimated the mAb close to the ground truth. In this case, JEKF-SANTO and JEKF-KPH2 had similar performances. These results are the opposite from the ones obtained with the JEKF-classic, and remained very similar using MRDE-PU, see supplementary Fig. S2. However, these results (Figure 1 and supplementary Fig. S2) were obtained by the use of specific **P**(t=0), because when we used a standard **P**(t=0) for SANTO and KPH2, their estimations are worse, despite that the SANTO is slightly more accurate than KPH2. The results using standard **P**(t=0) can be found in the supplementary Figures S3 and S4, and Table S14. These results (with standard and specific **P**(t=0)) show that the KPH2 is sensitive to the initial $P_{UP,UP}(t=0)$, and SANTO is sensitive to $P_{MSV,UP}(t=0)$. Since their better results were obtained with their specific **P**(t=0). The supplementary Table S5 shows specific **P**(t=0) used in KPH2, and the supplementary Table S6 shows specific **P**(t=0) used in SANTO.

Answer to RQ3-G2. In Figure 2, we show the estimations done by JEKF-SANTO and JEKF-KPH2 with regard to Xv, GLC, LAC, rAAV and the unshared parameters (μ_{GLC},μ_{LAC}, and μ_{rAAV}) of rAAV production (real dataset) using the **MRDE-PC** and the specific **P**(t=0) and standard **Q**. In the plot A of Figure 2, we can see that SANTO and KPH2 estimated the Xv inside of the noise range of the real online measurement of Xv by the capacitance probe. The next plots B, C and D show the estimation obtained regards the variables GLC, LAC and rAAV, respectively. In these plots and in the RMSE Table 2, we can see that only KPH2 made estimation far from the ground truth (red points). This is because it was not able to evolve the μ_{GLC},μ_{LAC}, and μ_{rAAV} (plots E, F and G) to values that enabling the process model to estimate GLC, LAC and rAAV near the ground truth (green line and red points in plots B, C and D). On the other hand, the proposed approach (SANTO) was able to evolve the unshared parameters simultaneously converging to values that enabled the estimation of GLC, LAC and rAAV near the ground

truth (blue line and red points in plots B, C and D). The KPH2 estimations done with MRDE-PU and specific **P**(t=0) were a little better than the ones done with MRDE-PC and specific **P**(t=0), see green lines and red points in supplementary Figure S5 and in Figure 2. Furthermore, the KPH2 had a better performance than SANTO estimating GLC. However, KPH2 estimations regards to LAC and rAAV were not better than SANTO estimations since SANTO estimation are the nearest to the ground truth (see supplementary Figure S5). The all KPH2 estimations improved only using a specific **Q** that represents the update of $Q_{Xv,Xv}$ from 0.0006 to 0.000006, see Figure 3 and RMSE Table 2. In this case, KPH2 evolved the μ_{GLC},μ_{LAC}, and μ_{rAAV} (plots E, F and G in Figure 3) to values that enabled the process model to estimate GLC, LAC and rAAV near to ground truth (green line and red points in plots B, C and D in Figure 3). Here, the SANTO was executed with MRDE-PC, specific **P**(t=0), and standard **Q**, and KPH2 was executed with a specific **Q**. However, despite these improvements, the KPH2 estimation for Xv was out of noise range at the beginning of the process (plots A in Figure 3) and this was not found in the results with SANTO. It is important to point out that SANTO converged the unshared parameters values faster than KPH2, because the KPH2 only started the convergence before 70h, see plots E, F and G in Figure 3. Furthermore, all estimations done by SANTO and KPH2 with standard **P**(t=0) were far from the ground truth (see red points in supplementary Figures S6 and S7) in the same way that happens with their estimations done with standard **P**(t=0) in synthetic dataset case, see supplementary Figures S3 and S4. This shows, one more time, that the KPH2 is sensitive to the $P_{UP,UP}(t=0)$, and SANTO is sensitive to $P_{MSV,UP}(t=0)$.

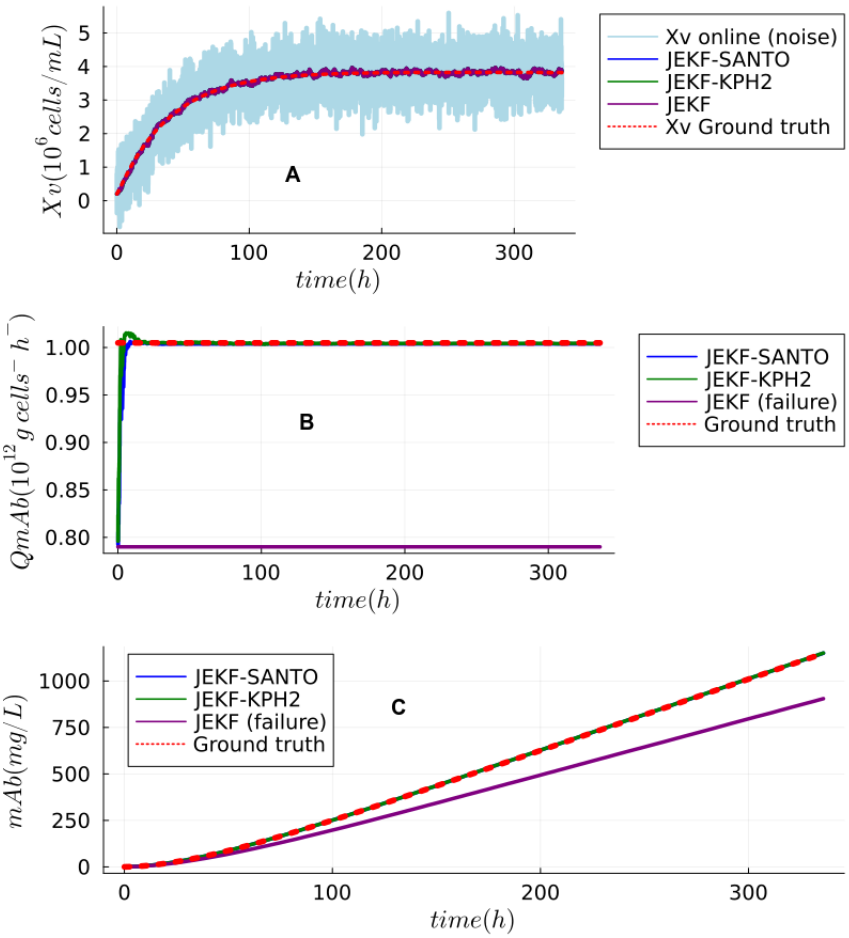

Figure 1. Experimental test of the Theorem (JEKF failure) and the JEKF-SANTO to avoid the JEKF failure with the biomanufacturing conditions (failure case). This experiment used the synthetic dataset, and plot A shows that all estimations with regard to Xv were close to the ground truth. Plots B and C show the estimations regards the unshared parameter QmAb and mAb (titer), respectively. The JEKF-SANTO and JEKF-KPH2 were able to evolve QmAb with convergence to the ground truth value, but JEKF-classic failed. It was not able to evolve the mAb. All NSEs were executed with MRDE-PC and specific **P**(t=0).

Table 1. RMSE between NSEs estimations about mAb and ground truth of synthetic dataset with specific **P**(t=0).

NSE	**RMSE (MRDE-PU)**	**RMSE (MRDE-PC)**
JEKF-SANTO	0.45	1.28
JEKF-KPH2	1.91	1.6
JEKF-Classic	132.4	132.4

Table 2. RMSE between NSEs estimations and ground truth of real dataset with MRDE-PC and specific **P**(t=0).

Ground truth	**JEKF-SANTO** (standard **Q**)	**JEKF-KPH2** (standard **Q**)	**JEKF-KPH2** (specific **Q**)
GLC	0.60	2.080	0.778
LAC	0.0195	1.656	0.2281
rAAV (titer)	0.114	1.566	0.286

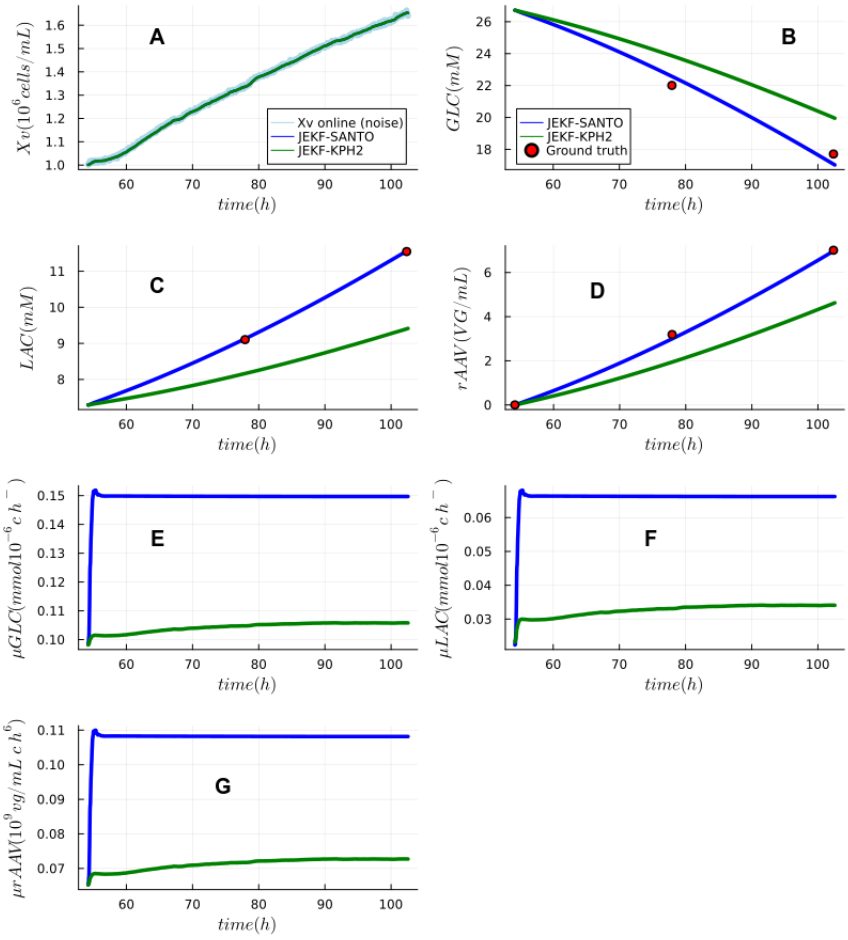

Figure 2. Simultaneous unshared parameters estimation by the JEKF-SANTO and JEKF-KPH2 with real dataset (rAAV production). Plot A shows the estimations regarding Xv, and all estimations were inside of the noise range of the real online measurement of Xv by the capacitance probe. Plots B, C and D show the estimation obtained regards the variables GLC, LAC and rAAV respectively. In these plots, we can see that KPH2 performed the worst estimation. Its estimations were far from the ground truth (red points), despite it evolved the μ_{GLC},μ_{LAC}, and μ_{rAAV} (unshared parameters) with convergence, see plots E, F and G. All NSEs were executed with MRDE-PC, specific **P**(t=0) and standard **Q**.

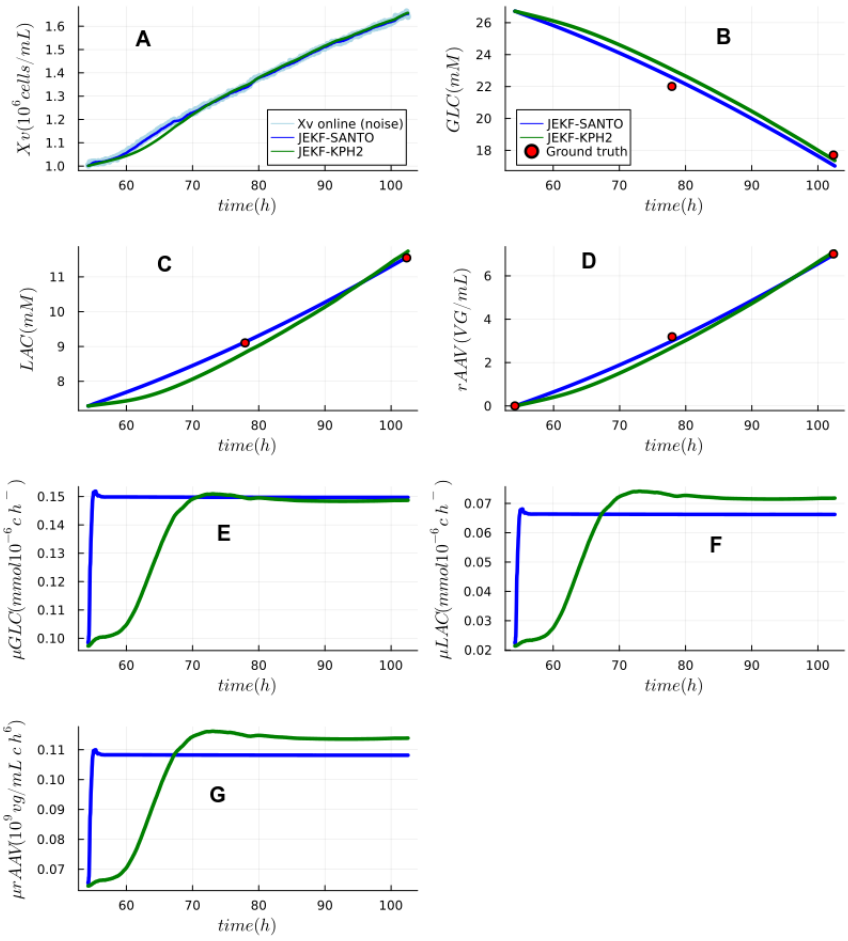

Figure 3. Simultaneous unshared parameters estimation by the JEKF-SANTO and JEKF-KPH2 with real dataset (rAAV production). Plot A shows the estimations regarding Xv, and JEKF-SANTO estimations were inside of the noise range of the real online measurement of Xv by the capacitance probe. However, JEKF-KPH2 estimations were outside at the beginning of the process. Plots B, C, and D show the estimation obtained regards the variables GLC, LAC, and rAAV, respectively. These plots show that the JEKF-SANTO had the closest estimation to the ground truth despite JEKF-KPH2 having improved performance. They evolved the μ_{GLC},μ_{LAC}, and μ_{rAAV} (unshared parameters) with convergence, see plots E, F and G. All NSEs were executed with MRDE-PC, specific **P**(t=0). However, SANTO used standard **Q** and KPH2 was executed with a specific **Q**.

Discussion

Our theoretical and empirical results showed the JEKF failure with biomanufacturing conditions. These results clearly showed that the JEKF (classic) could not estimate the unshared parameters and the state simultaneously since the Kalman gain related to the unshared parameter was constant and equal to zero in the entire process. On the other hand, the results also showed that the SANTO and KPH2 approaches could avoid the JEKF failure. However, the SANTO approach results in a more accurate estimation than KP2, while also having faster unshared parameters evolution to values that allowed the best performance of the process model. Furthermore, the results showed that both approaches are sensitive to **P**(t=0). The KPH2 is sensible to the $P_{UP,UP}(t=0)$, and the SANTO is sensible to $P_{MSV,UP}(t=0)$. However, the KPH2 was more difficult to adjust than SANTO. It required specific updates in **Q** beside a specific **P**(t=0) when executed with the real dataset. On the other hand, the SANTO approach required only a specific **P**(t=0). It is essential to point out that the SANTO approach did not change the probabilistic view of JEKF, and the minimization cost function in JEKF remained the same. Therefore, the SANTO approach can be viewed as an artifact that prevents the Kalman gain from becoming zero with the biomanufacturing conditions (failure case). In addition, the SANTO approach only addresses the failure case. It does not solve other issues, such as nonlinearity or high dimensionality, and should be used as a complementary approach.

Conclusion

In this work, firstly, we presented the common conditions in biomanufacturing that represent a failure case for the classical JEKF. Secondly, we proved that the classical JEKF, with these conditions, could not estimate the unshared parameters and the state simultaneously since the Kalman gain related to the unshared parameter is constant and equal to zero in the entire process. Lastly, we presented an approach called SANTO, which is a simple and effective way to address the failure case by adding a small quantity regarding the initial state error covariance between a measured state variable and an unshared parameter ($P_{MSV,UP}(t=0)$) in **P**(t=0) with uncorrelated elements. Our empirical evaluation showed that SANTO could evolve unshared parameters with convergence to the ground truth value of the synthetic dataset. Moreover, in the real case evaluation, SANTO overcame the approach KPH2 with better estimations (low RMSE). However, the main findings are that SANTO presented the fastest unshared parameter evolution and required less adjustment than KPH2, being more efficient than KPH2. Since SANTO had adjustment only in **P**(t=0), being that the KPH2 required adjustment in **P**(t=0) and **Q**.

References

1. Jin, X.-B., Robert Jeremiah, R. J., Su, T.-L., Bai, Y.-T. & Kong, J.-L. The new trend of state estimation: from model-driven to hybrid-driven methods. *Sensors* **21**, 2085 (2021).
2. Alexander, R., Campani, G., Dinh, S. & Lima, F. V. Challenges and opportunities on nonlinear state estimation of chemical and biochemical processes. *Processes* **8**, 1462 (2020).
3. Kalman, R. E. A new approach to linear filtering and prediction problems. (1960).
4. Jazwinski, A. Stochastic processes and filtering theory. *Math. science engineering (* (1970).
5. Aswal, N., Bhattacharya, B. & Sen, S. Joint and dual estimation of states and parameters with extended and unscented kalman filters. In *Recent Developments in Structural Health Monitoring and Assessment–Opportunities and Challenges: Bridges, Buildings and Other Infrastructures*, 223–252 (World Scientific, 2022).
6. Ljung, L. Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Autom. Control.* **24**, 36–50 (1979).
7. Kopp, R. E. & Orford, R. J. Linear regression applied to system identification for adaptive control systems. *Aiaa J.* **1**, 2300–2306 (1963).
8. Haykin, S. S. & Haykin, S. S. *Kalman filtering and neural networks*, vol. 284 (Wiley Online Library, 2001).
9. Cox, H. On the estimation of state variables and parameters for noisy dynamic systems. *IEEE Transactions on automatic control* **9**, 5–12 (1964).
10. Urrea, C. & Agramonte, R. Kalman filter: historical overview and review of its use in robotics 60 years after its creation. *J. Sensors* **2021** (2021).
11. Aswal, N., Sen, S. & Mevel, L. Switching kalman filter for damage estimation in the presence of sensor faults. *Mech. Syst. Signal Process.* **175**, 109116 (2022).
12. Stojanovic, V., He, S. & Zhang, B. State and parameter joint estimation of linear stochastic systems in presence of faults and non-gaussian noises. *Int. J. Robust Nonlinear Control.* **30**, 6683–6700 (2020).

13. Beelen, H., Bergveld, H. J. & Donkers, M. Joint estimation of battery parameters and state of charge using an extended kalman filter: a single-parameter tuning approach. *IEEE Transactions on Control. Syst. Technol.* **29**, 1087–1101 (2020).

14. Dhanalakshmi, R. *et al.* Onboard pointing error detection and estimation of observation satellite data using extended kalman filter. *Comput. Intell. Neurosci.* **2022** (2022).

15. Huang, K., Yuen, K.-V. & Wang, L. Real-time simultaneous input-state-parameter estimation with modulated colored noise excitation. *Mech. Syst. Signal Process.* **165**, 108378 (2022).

16. Huang, K. & Yuen, K.-V. Online dual-rate decentralized structural identification for wireless sensor networks. *Struct. Control. Heal. Monit.* **26**, e2453 (2019).

17. Yuen, K.-V. & Huang, K. Real-time substructural identification by boundary force modeling. *Struct. Control. Heal. Monit.* **25**, e2151 (2018).

18. Kleyman, V. *et al.* State and parameter estimation for retinal laser treatment. *arXiv preprint arXiv:2203.12452* (2022).

19. Iglesias Jr, C. F. *et al.* Monitoring the recombinant adeno-associated virus production using extended kalman filter. *Processes* **10**, 2180 (2022).

20. Yousefi-Darani, A., Paquet-Durand, O. & Hitzmann, B. The kalman filter for the supervision of cultivation processes. *Digit. Twins* 95–125 (2020).

21. Paquet-Durand, O., Zettel, V., Yousefi-Darani, A. & Hitzmann, B. The supervision of dough fermentation using image analysis complemented by a continuous discrete extended kalman filter. *Processes* **8**, 1669 (2020).

22. Song, H. & Hu, S. Open problems in applications of the kalman filtering algorithm. In *2019 International Conference on Mathematics, Big Data Analysis and Simulation and Modelling (MBDASM 2019)*, 185–190 (Atlantis Press, 2019).

23. Khodarahmi, M. & Maihami, V. A review on kalman filter models. *Arch. Comput. Methods Eng.* 1–21 (2022).

24. Nelson, L. & Stear, E. The simultaneous on-line estimation of parameters and states in linear systems. *IEEE Transactions on automatic Control.* **21**, 94–98 (1976).

25. Yousefi-Darani, A., Paquet-Durand, O. & Hitzmann, B. *The Kalman Filter for the Supervision of Cultivation Processes*, 95–125 (Springer International Publishing, Cham, 2021).

26. Herwig, C., Pörtner, R. & Möller, J. *Digital Twins: tools and concepts for smart biomanufacturing*, vol. 176 (Springer Nature, 2021).

27. Herwig, C., Pörtner, R. & Möller, J. *Digital Twins: Applications to the Design and Optimization of Bioprocesses*, vol. 177 (Springer Nature, 2021).

28. Sinner, P., Daume, S., Herwig, C. & Kager, J. *Usage of Digital Twins Along a Typical Process Development Cycle*, 71–96 (Springer International Publishing, Cham, 2021).

29. Moser, A., Appl, C., Brüning, S. & Hass, V. C. Mechanistic mathematical models as a basis for digital twins. *Digit. Twins* 133–180 (2020).

30. Narayanan, H., Sokolov, M., Morbidelli, M. & Butté, A. A new generation of predictive models: the added value of hybrid models for manufacturing processes of therapeutic proteins. *Biotechnol. Bioeng.* **116**, 2540–2549 (2019).

31. Fernandes-Platzgummer, A., Badenes, S. M., da Silva, C. L. & Cabral, J. M. Bioreactors for stem cell and mammalian cell cultivation. *Bioprocess. Technol. for Prod. Biopharm. Bioprod.* 131–173 (2018).

32. Iglesias Jr, C. F., Ristovski, M., Bolic, M. & Cuperlovic-Culf, M. raav manufacturing: The challenges of soft sensing during upstream processing. *Bioengineering* **10**, 229 (2023).

33. Gargalo, C. L. *et al.* Towards smart biomanufacturing: a perspective on recent developments in industrial measurement and monitoring technologies for bio-based production processes. *J. Ind. Microbiol. & Biotechnol. Off. J. Soc. for Ind. Microbiol. Biotechnol.* **47**, 947–964 (2020).

34. Udugama, A. *et al.* Towards digitalization in bio-manufacturing operations: A survey on application of big data and digital twin concepts in denmark. *Front. Chem. Eng* **3**, 727152 (2021).

35. Luo, Y., Kurian, V. & Ogunnaike, B. A. Bioprocess systems analysis, modeling, estimation, and control. *Curr. Opin. Chem. Eng.* **33**, 100705 (2021).

36. Wang, K., Li, Y. & Rizos, C. Practical approaches to kalman filtering with time-correlated measurement errors. *IEEE Transactions on Aerosp. Electron. Syst.* **48**, 1669–1681 (2012).

37. Ljung, L. & Söderström, T. *Theory and practice of recursive identification* (MIT press, 1983).

38. Kyriakopoulos, S. *et al.* Kinetic modeling of mammalian cell culture bioprocessing: the quest to advance biomanufacturing. *Biotechnol. J.* **13**, 1700229 (2018).

39. Park, S.-Y., Park, C.-H., Choi, D.-H., Hong, J. K. & Lee, D.-Y. Bioprocess digital twins of mammalian cell culture for advanced biomanufacturing. *Curr. Opin. Chem. Eng.* **33**, 100702 (2021).

40. Tsopanoglou, A. & del Val, I. J. Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses. *Curr. Opin. Chem. Eng.* **32**, 100691 (2021).

41. Mears, L., Stocks, S. M., Albaek, M. O., Sin, G. & Gernaey, K. V. Mechanistic fermentation models for process design, monitoring, and control. *Trends biotechnology* **35**, 914–924 (2017).

42. Reyes, S. J., Durocher, Y., Pham, P. L. & Henry, O. Modern sensor tools and techniques for monitoring, controlling, and improving cell culture processes. *Processes* **10**, 189 (2022).

43. Zhang, D., Del Rio-Chanona, E. A., Petsagkourakis, P. & Wagner, J. Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization. *Biotechnol. bioengineering* **116**, 2919–2930 (2019).

44. Kourti, T. Multivariate statistical process control and process control, using latent variables. (2020).

45. Ji, Z. & Brown, M. Joint state and parameter estimation for biochemical dynamic pathways with iterative extended kalman filter: comparison with dual state and parameter estimation. *The Open Autom. Control. Syst. J.* **2** (2009).

46. Brockwell, P. Time series analysis. *Encycl. Stat. Behav. Sci.* (2005).

47. Ohadi, K., Legge, R. L. & Budman, H. M. Development of a soft-sensor based on multi-wavelength fluorescence spectroscopy and a dynamic metabolic model for monitoring mammalian cell cultures. *Biotechnol. bioengineering* **112**, 197–208 (2015).

48. Assimakis, N. & Adam, M. Kalman filter riccati equation for the prediction, estimation, and smoothing error covariance matrices. *Int. Sch. Res. Notices* **2013** (2013).

49. Kulikova, M. V. & Kulikov, G. Y. Adaptive ode solvers in extended kalman filtering algorithms. *J. Comput. Appl. Math.* **262**, 205–216 (2014).

50. Fraser, C. T. *Adaptive extended Kalman filtering strategies for autonomous relative navigation of formation flying spacecraft*. Ph.D. thesis, Carleton University (2019).

51. Labbe Jr, R. R. Kalman and bayesian filters in python. (2020).

52. Haldar, A. & Al-hussein, A. A. A. Recent developments in structural health monitoring and assessment-opportunities and challenges: Bridges, buildings and other infrastructures. (2022).

53. Goudar, C. T. Computer programs for modeling mammalian cell batch and fed-batch cultures using logistic equations. *Cytotechnology* **64**, 465–475 (2012).

54. Kornecki, M. & Strube, J. Accelerating biologics manufacturing by upstream process modelling. *Processes* **7**, 166 (2019).

55. Narayanan, H. *et al.* Hybrid-ekf: Hybrid model coupled with extended kalman filter for real-time monitoring and control of mammalian cell culture. *Biotechnol. Bioeng.* **117**, 2703–2714 (2020).

56. Ueno, G. & Nakamura, N. Bayesian estimation of the observation-error covariance matrix in ensemble-based filters. *Q. J. Royal Meteorol. Soc.* **142**, 2055–2080 (2016).

57. Michel, V., Gramfort, A., Varoquaux, G., Eger, E. & Thirion, B. Total variation regularization for fmri-based prediction of behavior. *IEEE transactions on medical imaging* **30**, 1328–1340 (2011).

58. Jyothilekshmi, I. & Jayaprakash, N. Trends in monoclonal antibody production using various bioreactor systems. (2021).

59. Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W. & Gao, G. Viral vector platforms within the gene therapy landscape. *Signal transduction targeted therapy* **6**, 53 (2021).

Acknowledgements

We want to acknowledge the support provided by National Research Council Canada through AI4D grant. We are grateful to Dr. Amine A. Kamen and his group at McGill University for providing the data and for providing expert domain knowledge during our meetings and discussion. We would also like to thank Dr. Nabil Belacel for his constructive comments and discussions.

Author contributions statement

All authors reviewed the manuscript. C.F.I.J conceptualization, C.F.I.J implemented algorithms and conducted the experiments, C.F.I.J performed analysis on experimental results and wrote the manuscript, M.B. provided insightful discussions, reviewed the results and revised the manuscript, M.B. supervision, M.B. project administration, and M.B. funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Data Availability The data and code used in this study are available in GitHub: https://github.com/cristovaoiglesias/JEKF-SANTO;

Supplementary Information The online version contains supplementary material available GitHub: https://github.com/cristovaoiglesias/JEKF-SANTO/blob/main/article/Supplementary_Information.pdf .