

Article

How Not to Make the Joint Extended Kalman Filter Fail with Unstructured Mechanistic Models

Cristovão Freitas Iglesias, Jr.  and Miodrag Bolic School of Electrical Engineering and Computer Science (EECS), University of Ottawa,
Ottawa, ON K1N 6N5, Canada

* Correspondence: cfrei@uottawa.ca (C.F.I.J.); miodrag.Bolic@uottawa.ca (M.B.)

Abstract: The unstructured mechanistic model (UMM) allows for modeling the macro-scale of a phenomenon without known mechanisms. This is extremely useful in biomanufacturing because using the UMM for the joint estimation of states and parameters with an extended Kalman filter (JEKF) can enable the real-time monitoring of bioprocesses 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 \mathbf{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 equal to zero. In this work, we formally describe this failure case, present the proof of JEKF failure, and propose an approach called SANTO to side-step this failure case. The SANTO approach consists of adding a quantity to the state error covariance between the measured state variable and unshared parameter in the initial $\mathbf{P}(t = 0)$ of the matrix Riccati differential equation to compute the predicted error covariance matrix of the state and prevent the Kalman gain from being zero. Our empirical evaluations using synthetic and real datasets reveal significant improvements: SANTO achieved a reduction in root-mean-square percentage error (RMSPE) of up to approximately 17% compared to the classical JEKF, indicating a substantial enhancement in estimation accuracy.



Citation: Iglesias, C.F., Jr.; Bolic, M. How Not to Make the Joint Extended Kalman Filter Fail with Unstructured Mechanistic Models. *Sensors* **2024**, *24*, 653. <https://doi.org/10.3390/s24020653>

Academic Editors: Maria Gabriella Xibilia and Chamil Abeykoon

Received: 21 November 2023

Revised: 22 December 2023

Accepted: 6 January 2024

Published: 19 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: joint extended Kalman filter; unstructured mechanistic model; bioprocess monitoring

1. 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 (\mathbf{Q}), measurement (\mathbf{R}), and state (\mathbf{P}) [3,4]. The EKF, beyond state estimation, is also used for the parameter estimation (parameter evolution [5]) of nonlinear systems (process models) considering a single joint state variable vector, which includes both the states and parameters of the process model [6–8]. This approach is called the joint estimation of states and parameters with an extended Kalman filter (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 made [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 of the states of the cell culture. Thus, 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 1960s for the estimation of linear systems (in which there is a bilinear relation between the states and parameters) [6–10]. However, the JEKF is still very popular, with several new applications in different areas [5,11–21], and with unsolved problems [22,23]. Furthermore, the 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 microcontrollers (and/or minicomputers and/or microprocessors) requires numerically economical and robust algorithms, such as the JEKF [11,24]. An important area of application of the JEKF is biomanufacturing, that is, the production of biological products from living cells [20,25,26]. The reason for this is that the JEKF with the mechanistic model (MM) as a process model effectively serves as a soft sensor in biomanufacturing. This combination can enable the real-time monitoring of critical process parameters (CPPs) or critical quality attributes (CQAs) that are difficult to measure directly or that can only be measured at low sampling frequencies in a bioprocess [20,27]. There are two types of MM: structured mechanistic models (SMMs) and unstructured mechanistic models (UMMs) [28]. When we have knowledge about a bioprocess, we can use an SMM with the JEKF. On the other hand, when we do not have knowledge about a bioprocess, we can use a UMM with the JEKF because the UMM allows us to model the macro-scale of a phenomenon. It is a mass-balance equation system with few parameters and variables and less complexity than SMMs [29,30].

The UMM used in biomanufacturing typically consists of ODEs with unshared parameters, weak variables, and weak terms. However, these characteristics of UMM in biomanufacturing, together with the use of $\mathbf{P}(t = 0)$ and \mathbf{Q} with uncorrelated elements and the presence of a single measured state variable, represent a failure case that occurs when the JEKF cannot estimate the unshared parameters and the state simultaneously. There are many new bioprocesses for which the literature contains no prior knowledge that the biopharmaceutical industry aims to monitor, such as recombinant adeno-associated virus (rAAV) production [31]. Therefore, enabling the JEKF to side-step the failure case described above may help the industry perform biomanufacturing with the real-time monitoring of bioprocesses with unknown mechanisms. Consequently, this skill can support the biopharmaceutical industry in achieving biomanufacturing 4.0 by becoming more agile and intelligent, thus enhancing product quality, optimizing operations, and reducing costs [25,26,32,33]. 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 the desired productivity and product quality consistently [34].

In this work, we present the common conditions in biomanufacturing that represent a failure case where the JEKF fails to perform the unshared parameter evolution of a UMM, and we propose a solution to side-step this failure case, called SANTO, which consists of a Specific initial condition (SANTO) for the matrix Riccati 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} [35], we only add a quantity to the state error covariance between the measured state variable (MSV) and an unshared parameter (UP) in $\mathbf{P}(t = 0)$ for the MRDE. The proposed approach can avoid JEKF failure by preventing the Kalman gain from being zero throughout the entire process, which is an unrealistic situation that would mean that the predictions of the UMM (used as a process model) are perfect. Our theoretical and empirical results demonstrate the effectiveness of SANTO, which was assessed using synthetic and real datasets. The code and data used in this work are available in the data availability section of this paper to facilitate reproducibility. Our contributions can be summarized as follows:

- We provide proof of JEKF failure when acting as an unshared parameter estimator under specific biomanufacturing conditions that represent a failure case. To our knowledge, this is the first work to formally report this failure case regarding the 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.

2. Related Work

In contrast to JEKF, the dual extended Kalman filter (DEKF) employs two consecutive EKFs, separating the estimation of system states and parameters [36]. This separation can be advantageous in certain scenarios, but JEKF offers three important benefits, particularly in the context of the practical implementation of adaptive controllers using microcontrollers in biomanufacturing that requires numerically economical and robust algorithms such as JEKF [11,24]. First, JEKF avoids the computational overhead associated with running two separate filters, as in DEKF, enhancing computational efficiency [37]. Second, it can provide more accurate and robust estimates in scenarios, such as nonlinear biochemical systems, that commonly occur in biomanufacturing processes [36]. Lastly, the single-filter structure of JEKF is simpler to implement and tune compared to the dual-filter approach of DEKF [8]. The main limitation of JEKF is not guaranteed convergence in some cases, as reported by [6,24,38]. A solution to deal with the convergence problems of JEKF is to use recurrent derivatives [6,38]. 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 be unable to estimate the parameters and the state simultaneously, such as singularity issues [35]. However, until now, 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 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 regard to the UMM parameters, their results showed the unshared parameter evolution with convergence. However, the authors did not describe the problem in detail. They did not present a theoretical justification for the approach used (second linear operator). They clearly stated that the work is an initial study and reported the need for future validation. We named this approach KPH2 because the authors used a second linear operator to enable \mathbf{K} and \mathbf{P} to be updated, and we used this approach in our experimental evaluation for comparison purposes with our proposed approach. A description of KPH2 and a possible interpretation can be found in Section S6 of the Supplementary Material.

3. Background

3.1. Unstructured Mechanistic Model (UMM)

Unstructured Mechanistic Models (or Unstructured Mechanistic Kinetic Models) are models of the temporal evolution of a bioprocess [39]. They are based on first-principle mechanisms that drive the bioprocess under consideration [34]. Examples of bioprocesses are (i) the production of therapeutic monoclonal antibodies (mAbs), which is projected to bring in USD 300 billion by 2025 [34], 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 [19,31]. It is essential to point out that despite UMM being the most suitable option to describe the dynamic behavior of bioprocesses and being considered a crucial foundation for soft sensors in DT development, its industrial use is still in its early stages [28,39,40]. The UMMs 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 [41–43]. Therefore, UMMs are the most suitable option

for explaining observed phenomena, predicting process behavior, and analyzing intrinsic bioprocess characteristics such as controllability [34].

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, the development of SMM for a specific bioprocess requires extensive domain knowledge and substantial effort [34,41]. SMM is unsuitable for the dynamic control of bioprocess in bioreactors used commonly in biomanufacturing because many of the variables used in SMM cannot be manipulated directly [34]. SMM is most suited for cell-line development, in which a cells' genome-level properties are changed to produce the desired process behavior [34].

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 datasets collected under distinct operating circumstances [44]. The Kalman filter approach is commonly implemented with UMM [45] to improve prediction accuracy and generate predictions between sampling instances. Among several data analysis methods, the Kalman filter and its nonlinear 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 Section S1 of Supplementary Material.

3.2. 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 Section S2 of the Supplementary Material. The EKF requires a state-space model to perform an estimation on the state variables of a process (nonlinear system) present in a state variable vector $\boldsymbol{\psi}(t)$ [1,36,44]. 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 [8,20,44].

A 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:

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

Subsequently, the process model is represented as

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

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

The measurement model is treated as a discrete system and defined as

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

The nonlinear 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 $\boldsymbol{v} \sim \mathcal{N}(0, \mathbf{R})$ 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 $h(\boldsymbol{\psi}(t_k)) = \mathbf{H}\boldsymbol{\psi}(t_k)$ [20,36,47] where the matrix \mathbf{H} is a linear operator (row vector) that matches the states variables of $\boldsymbol{\psi}(t_k)$ 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

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

The EKF algorithm is implemented through a state variables vector $\boldsymbol{\psi}(t)$, initial condition, prediction step (time update) and correction step (measurement update) [1,20,21,36,47].

Initialization step: The initial condition is composed of the initial mean $\hat{\boldsymbol{\psi}}_0 = E[\boldsymbol{\psi}_0]$ and initial error covariance matrix $\mathbf{P}_0 = \mathbf{P}(t = 0) = E[(\boldsymbol{\psi}_0 - \hat{\boldsymbol{\psi}}_0)(\boldsymbol{\psi}_0 - \hat{\boldsymbol{\psi}}_0)^T]$ of the 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 $\hat{\boldsymbol{\psi}}(t_{k/k-1})$ and predicted error covariance matrix $\mathbf{P}(t_{k|k-1})$ of state variables vector $\boldsymbol{\psi}(t)$ are obtained. This is completed by numerically integrating $\phi(\boldsymbol{\psi}(t), t, \theta)$ from discrete time t_{k-1} to t_k the following equation

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

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

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

from t_{k-1} to t_k , where a new measurement is obtained at time k [4,49], and \mathbf{J}_t^ϕ is the Jacobian matrix of ϕ evaluated at the prior mode [50,51],

$$\mathbf{J}_t^\phi = \frac{\partial \phi(\boldsymbol{\psi}(t))}{\partial \boldsymbol{\psi}_i} \Big|_{\boldsymbol{\psi}(t) = \hat{\boldsymbol{\psi}}(t-1)}. \quad (7)$$

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

(i) innovation equations

$$e_{Z,k} = \mathbf{Z}_k - \mathbf{H}\hat{\boldsymbol{\psi}}(t_{k/k-1}) \quad (8)$$

$$\mathbf{S}_k = \mathbf{H}\mathbf{P}(t_{k|k-1})\mathbf{H}^T + \mathbf{R} \quad (9)$$

and (ii) update step equations

$$\mathbf{K}_k = \mathbf{P}(t_{k|k-1})\mathbf{H}^T\mathbf{S}_k^{-1} \quad (10)$$

$$\hat{\boldsymbol{\psi}}(t_{k/k}) = \hat{\boldsymbol{\psi}}(t_{k/k-1}) + \mathbf{K}_k e_{Z,k} \quad (11)$$

$$\mathbf{P}(t_{k|k}) = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}(t_{k|k-1}) \quad (12)$$

where $e_{Z,k}$ and \mathbf{S}_k represent, respectively, the innovation error and innovation covariance.

The Kalman gain is a scaling factor (ratio) to estimate the state variables by setting a value between the predicted state and measured state [4,50]. The \mathbf{K}_k chooses a value along the residual range $(\mathbf{Z}_k - \mathbf{H}\hat{\boldsymbol{\psi}}(t_{k/k-1}))$ [8,50]. \mathbf{K}_k enables to set a value for $\hat{\boldsymbol{\psi}}(t_{k/k})$ between the $\hat{\boldsymbol{\psi}}(t_{k/k-1})$ (prediction) and \mathbf{Z}_k (measurement) using Equation (11) and update the belief regarding the state variables based on how certain we are regarding the measurement using Equation (12) [50]. The Kalman gain is computed as a ratio of prior and measurement uncertainty available; see Equation (10). The one-dimensional form of Equation (10) is the following $K = P/(P + R)$ [50]. It is important to point out that linear operator \mathbf{H}

matches the states variables of $\boldsymbol{\psi}(t_k)$ to the measured variables \mathbf{Z}_k that are obtained at a discrete instance.

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

3.3. 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 ($\boldsymbol{\psi}(t) = [x_1, x_2, \dots, x_n]^T$) is considered as extended/augmented as following,

$$\boldsymbol{\psi}(t) = [x_1, x_2, \dots, x_n, \theta_1, \dots, \theta_n]^T. \quad (13)$$

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

$$\theta(t_k) = \theta(t_{k-1}) + \text{noise}, \quad (14)$$

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.

4. Theoretical Analysis

This section presents the theoretical analysis of the JEKF failure to perform unshared parameter evolution with a UMM and SANTO, which is the proposed solution for this problem.

4.1. 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 Section S3.1 of the Supplementary Material.
- **Weak and Strong term of an ODE:** A *weak term* is a term of an ODE with a low percentage of variables of the state variable vector, and a *strong term* is one with a high percentage of variables of the state variable vector. See the example in Section S3.2 of the Supplementary Material.
- **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 Section S3.3 of the Supplementary Material.

4.1.1. Failure Case: Biomanufacturing Conditions

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

- **ODEs of UMM with unshared parameters.** This parameter type is commonly used in ODE to model the dynamic of product formation in biomanufacturing [53–55]. See the example in Section S3.1 of the Supplementary Material.
- **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 $\mathbf{P}(t = 0)$).

This assumption raises two scenarios:

1. The use of **P** with uncorrelated elements to build the MRDE (Equation (6)) and $\mathbf{P}(t = 0)$ with uncorrelated elements as the 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^ϕ . See the example in Section S3.4 of the Supplementary Material. It is important to point out that depending on the 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 Section S3.2 of the Supplementary Material.
 2. The use of **P** with correlated elements to build the MRDE (Equation (6)) and $\mathbf{P}(t = 0)$ with uncorrelated elements as the 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^ϕ result from the partial derivation of a strong term. See the example in Section S3.2 of the Supplementary Material.
 - **ODEs of UMM with weak variables.** In the Jacobian \mathbf{J}_t^ϕ , 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 Section S3.3 of Supplementary Material.
 - **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 the 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 Equation (10). 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 Section S3.5 of the Supplementary Material.

4.1.2. Lemma: Inability to Update Kalman Gain for Unshared Parameters based $\mathbf{P}(t = 0)$ and **Q** with Uncorrelated Elements

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

Lemma 1. *The Kalman gain cannot be updated (by Equation (10)) for an unshared parameter that is part of a state variable vector and part of a weak term in a UMM if the initial state error covariance matrix $\mathbf{P}(t = 0)$ and **Q** are formed by uncorrelated elements and there is only one state variable measured.*

The proof of this lemma is in the following, and an example can be found in Section S4 of the Supplementary Material.

Proof of Lemma 1. Let us consider the following:

- A general UMM with an unshared parameter in a weak term represented by a system of nonlinear differential equations of the form:

$$\frac{dx_{msv}}{dt} = f_1(x_{msv}, x_2, \dots, x_{n-1}, \theta_1, \theta_2, \dots, \theta_m) \quad (15)$$

$$\frac{dx_2}{dt} = f_2(x_{msv}, x_2, \dots, x_{n-1}, \theta_1, \theta_2, \dots, \theta_m) \quad (16)$$

$$\vdots \quad (17)$$

$$\frac{dx_n}{dt} = f_n(x_{msv}, \theta_{up}) \quad (18)$$

where x_{msv} and x_2, \dots, x_n are the variables of the system, f_1, f_2, \dots, f_n are the functions defining the system, and $\theta_1, \theta_2, \dots, \theta_m$ are the parameters of the system, and θ_{up} is an unshared parameter.

- A joint state variables vector defined as

$$\boldsymbol{\psi}(t)_{general} = [x_{msv}, x_2, \dots, x_n, \theta_{up}]. \quad (19)$$

- A process model defined as

$$\frac{d\boldsymbol{\psi}(t)_{general}}{dt} = \boldsymbol{\phi}(\boldsymbol{\psi}(t)_{general}, t) + \boldsymbol{\omega}(t) = \frac{d}{dt} \begin{bmatrix} x_{msv} \\ x_2 \\ \vdots \\ x_n \\ \theta_{up} \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \\ 0 \end{bmatrix} + \boldsymbol{\omega}(t). \quad (20)$$

- x_{msv} as the unique measured state variable (MSV) and $\mathbf{H} = [1 \ 0 \ \dots \ 0 \ 0]$.
- \mathbf{R} as measurement noise variance of x_{msv} .
- θ_{up} as the unshared parameter (UP) to be evolved (estimated) and presented in only one weak term.
- \mathbf{P} and \mathbf{Q} with uncorrelated elements for the $\boldsymbol{\psi}(t)_{general}$ (Equation (19)),

$$\mathbf{P} = \begin{bmatrix} P_{x_{msv}, x_{msv}} & 0 & \dots & 0 & 0 \\ 0 & P_{x_2, x_2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & P_{n, n} & 0 \\ 0 & 0 & \dots & 0 & P_{\theta_{up}, \theta_{up}} \end{bmatrix}, \quad (21)$$

$$\mathbf{Q} = \begin{bmatrix} Q_{x_{msv}, x_{msv}} & 0 & \dots & 0 & 0 \\ 0 & Q_{x_2, x_2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & Q_{n, n} & 0 \\ 0 & 0 & \dots & 0 & Q_{\theta_{up}, \theta_{up}} \end{bmatrix}. \quad (22)$$

- The Jacobian \mathbf{J}_t^ϕ (Equation (7)), with the $\boldsymbol{\psi}(t)_{general}$ (Equation (19)),

$$\mathbf{J}_t^\phi(\boldsymbol{\psi}(t)_{general}, t) = \begin{bmatrix} \frac{\partial f_1}{\partial x_{msv}} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} & 0 \\ \frac{\partial f_2}{\partial x_{msv}} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_{msv}} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} & \frac{\partial f_n}{\partial \theta_{up}} \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \quad (23)$$

Given these conditions and Equation (6), we have the following MRDE (based on \mathbf{P} uncorrelated)

$$\frac{d\mathbf{P}(t)}{dt} = \begin{bmatrix} \frac{dP_{x_{msv},x_{msv}}(t)}{dt} = Q_{1,1} + 2P_{1,1} \frac{\partial f_1}{\partial x_{msv}} & \frac{dP_{x_2,x_{msv}}(t)}{dt} = (P_{1,1} + P_{2,2}) \frac{\partial f_1}{\partial x_2} & \cdots & \frac{dP_{x_n,x_{msv}}(t)}{dt} = (P_{1,1} + P_{3,3}) \frac{\partial f_1}{\partial x_n} & \frac{dP_{\theta_{up},x_{msv}}(t)}{dt} = 0 \\ \frac{dP_{x_{msv},x_2}(t)}{dt} = (P_{1,1} + P_{2,2}) \frac{\partial f_2}{\partial x_{msv}} & \frac{dP_{x_2,x_2}(t)}{dt} = Q_{2,2} + 2P_{2,2} \frac{\partial f_2}{\partial x_2} & \cdots & \frac{dP_{x_n,x_2}(t)}{dt} = (P_{3,3} + P_{2,2}) \frac{\partial f_2}{\partial x_n} & \frac{dP_{\theta_{up},x_2}(t)}{dt} = 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{dP_{x_{msv},x_n}(t)}{dt} = (P_{1,1} + P_{n,n}) \frac{\partial f_n}{\partial x_{msv}} & \frac{dP_{x_2,x_n}(t)}{dt} = (P_{2,2} + P_{n,n}) \frac{\partial f_n}{\partial x_2} & \cdots & \frac{dP_{x_n,x_n}(t)}{dt} = Q_{n,n} + 2P_{n,n} \frac{\partial f_n}{\partial x_n} & \frac{dP_{\theta_{up},x_n}(t)}{dt} = 0 \\ \frac{dP_{x_{msv},\theta_{up}}(t)}{dt} = 0 & \frac{dP_{x_2,\theta_{up}}(t)}{dt} = 0 & \cdots & \frac{dP_{x_n,\theta_{up}}(t)}{dt} = 0 & \frac{dP_{\theta_{up},\theta_{up}}(t)}{dt} = 0 \end{bmatrix}. \quad (24)$$

Now, using this Equation (24) to compute the predicted state error covariance matrix $\mathbf{P}(t_{k/k-1})$ from t_{k-1} to t_k with an initial predicted state error covariance matrix $\mathbf{P}(t_{k-1}) = \mathbf{P}_0 = \mathbf{P}_{init}(t=0)$ with uncorrelated elements as the following

$$\mathbf{P}_{init}(t=0) = \begin{bmatrix} P_{x_{msv},x_{msv}}(t=0) & 0 & \cdots & 0 & 0 \\ 0 & P_{x_2,x_2}(t=0) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & P_{n,n}(t=0) & 0 \\ 0 & 0 & \cdots & 0 & P_{\theta_{up},\theta_{up}}(t=0) \end{bmatrix}, \quad (25)$$

we have

$$\mathbf{P}(t_{k/k-1}) = \begin{bmatrix} P_{x_{msv},x_{msv}}(t_{k/k-1}) & P_{x_2,x_{msv}}(t_{k/k-1}) & \cdots & P_{n,x_{msv}}(t_{k/k-1}) & P_{\theta_{up},x_{msv}}(t_{k/k-1}) \\ P_{x_{msv},x_2}(t_{k/k-1}) & P_{x_2,x_2}(t_{k/k-1}) & \cdots & P_{n,x_2}(t_{k/k-1}) & P_{\theta_{up},x_2}(t_{k/k-1}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ P_{x_{msv},n}(t_{k/k-1}) & P_{x_2,n}(t_{k/k-1}) & \cdots & P_{n,n}(t_{k/k-1}) & P_{\theta_{up},n}(t_{k/k-1}) \\ P_{x_{msv},\theta_{up}}(t_{k/k-1}) = 0 & P_{x_2,\theta_{up}}(t_{k/k-1}) & \cdots & P_{n,\theta_{up}}(t_{k/k-1}) & P_{\theta_{up},\theta_{up}}(t_{k/k-1}) \end{bmatrix}. \quad (26)$$

Now, using $\mathbf{P}(t_{k/k-1})$, \mathbf{H} and \mathbf{R} to compute the Kalman gain for all variables in the state variable vector $\boldsymbol{\psi}(t)_{general}$ (Equation (19)), we have

$$\mathbf{K}_k = \mathbf{P}(t_{k/k-1})\mathbf{H}^T(\mathbf{H}\mathbf{P}(t_{k/k-1})\mathbf{H}^T + \mathbf{R})^{-1} = \begin{bmatrix} K_{x_{msv}} \\ K_{x_2} \\ \vdots \\ K_{x_n} \\ K_{\theta_{up}} \end{bmatrix} = \begin{bmatrix} \frac{P_{x_{msv},x_{msv}}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \frac{P_{x_{msv},x_2}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \vdots \\ \frac{P_{x_{msv},n}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \frac{P_{x_{msv},\theta_{up}}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \end{bmatrix} = \begin{bmatrix} \frac{P_{x_{msv},x_{msv}}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \frac{P_{x_{msv},x_2}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \vdots \\ \frac{P_{x_{msv},n}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{P_{x_{msv},x_{msv}}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \frac{P_{x_{msv},x_2}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \vdots \\ \frac{P_{x_{msv},n}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ 0 \end{bmatrix}. \quad (27)$$

\mathbf{H} selected the first column of $\mathbf{P}(t_{k/k-1})$, since it is related to the measured value x_{msv} . However, in this column, we have that the predicted state error covariance between x_{msv} and θ_{up} is zero, $P_{x_{msv},\theta_{up}}(t_{k/k-1}) = Cov(x_{msv}, \theta_{up}) = 0$. The solution of $\frac{dP_{x_{msv},\theta_{up}}(t)}{dt} = 0$ obtained from t_{k-1} to t_k is equal to the initial condition that is zero due to $\mathbf{P}(t=0)$ with uncorrelated elements, and we have $Cov(x_{msv}, \theta_{up}) = P_{x_{msv},\theta_{up}}(t_{k-1}) = P_{x_{msv},\theta_{up}}(t=0) = 0$.

Then, the Kalman gain value for the unshared parameter is zero, $K_{\theta_{up}} = 0$, and consequently, the predicted state error covariance $P_{x_{msv}, \theta_{up}}(t_{k/k-1})$ cannot be updated (by Equation (12)). Since

$$\mathbf{P}(t_{k|k}) = (\mathbf{I} - \mathbf{K}_k \mathbf{H}) \mathbf{P}(t_{k|k-1}) = \begin{bmatrix} \vdots & \vdots & \ddots \\ P_{x_{msv}, \theta_{up}}(t_{k/k-1}) - K_{\theta_{up}} \cdot P_{x_{msv}, x_{msv}}(t_{k/k-1}) & \dots & \dots \\ \vdots & \vdots & \ddots \\ 0 - 0 \cdot P_{x_{msv}, x_{msv}}(t_{k/k-1}) & \dots & \dots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \ddots \\ 0 & \dots & \dots \\ \vdots & \vdots & \ddots \\ 0 & \dots & \dots \end{bmatrix}. \quad (28)$$

Therefore, we have that $P_{x_{msv}, \theta_{up}}(t_{k/k}) = P_{x_{msv}, \theta_{up}}(t_{k/k-1}) = 0$, and as $P_{x_{msv}, \theta_{up}}(t_{k/k}) = 0$ has to be used as a new initial condition for MRDE (Equation (24)), we have $K_{\theta_{up}} = 0$ for all $P_{x_{msv}, \theta_{up}}(t_{k/k-1})$ obtained from t_{k-1} to t_k using Equation (24) and consequently $K_{\theta_{up}}$ and $P_{x_{msv}, \theta_{up}}(t_{k/k}) = P_{x_{msv}, \theta_{up}}(t_{k/k-1}) = 0$ are always zero and cannot be updated. \square

4.1.3. Theorem: JEKF Failure

The consequence of Lemma 1 (Section 4.1.2) is the following theorem:

Theorem 1. *The JEKF (Section 3.3) 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 $\mathbf{P}(t=0)$ and \mathbf{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 proof of Theorem 1 is in the following, and an example of this theorem can be found in Section S5 of the Supplementary Material.

Proof of Theorem 1. This proof can be completed using the conditions and results described previously in the proof of Lemma 1 (Section 4.1.2).

Then, let us consider the following:

- $\mathbf{H} = [1 \ 0 \ \dots \ 0 \ 0]$ and $\mathbf{K}_k = [K_{x_{msv}}, K_{x_2}, \dots, K_{x_n}, K_{\theta_{up}}]^T$ as obtained in the proof of Lemma 1 in Section 4.1.2, where $K_{\theta_{up}} = 0$.
- \mathbf{Z}_k as a measured value of x_{msv} .
- Predicted mean of the state variable vector $\hat{\boldsymbol{\psi}}(t_{k/k-1})_{general} = [\hat{x}_{msv}, \hat{x}_2, \dots, \hat{x}_n, \hat{\theta}_{up}]^T$ with regard to the general UMM used in the proof of Lemma 1 in Section 4.1.2.

Now, using Equation (11) to compute the estimated mean of the state variable vector $\hat{\boldsymbol{\psi}}(t_{k/k})_{general}$, we have

$$\hat{\boldsymbol{\psi}}(t_{k/k})_{general} = \hat{\boldsymbol{\psi}}(t_{k/k-1})_{general} + \mathbf{K}_k (\mathbf{Z}_k - \mathbf{H} \hat{\boldsymbol{\psi}}(t_{k/k-1})_{general}) \quad (29)$$

$$\hat{\boldsymbol{\psi}}(t_{k/k})_{general} = \begin{bmatrix} \hat{x}_{msv} \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \\ \hat{\theta}_{up} \end{bmatrix} + \begin{bmatrix} K_{x_{msv}} \\ K_{x_2} \\ \vdots \\ K_{x_n} \\ K_{\theta_{up}} \end{bmatrix} \cdot (\mathbf{Z}_k - \hat{x}_{msv}) = \begin{bmatrix} \hat{x}_{msv} + K_{x_{msv}} \cdot (\mathbf{Z}_k - \hat{x}_{msv}) \\ \hat{x}_2 + K_{x_2} \cdot (\mathbf{Z}_k - \hat{x}_{msv}) \\ \vdots \\ \hat{x}_n + K_{x_n} \cdot (\mathbf{Z}_k - \hat{x}_{msv}) \\ \hat{\theta}_{up} + 0 \end{bmatrix} \quad (30)$$

Then, we have that the estimated mean of the unshared parameter $\hat{\theta}_{up}(t_{k/k})$ (composing the $\hat{\boldsymbol{\psi}}(t_{k/k})_{general}$) is equal to the predicted mean of unshared parameter $\hat{\theta}_{up}(t_{k/k-1})$ (composing the $\hat{\boldsymbol{\psi}}(t_{k/k-1})_{general}$) for all steps from t_{k-1} to t_k . In other words, the JEKF fails to perform the parameter evolution, since it does not have a noise component to evolve the parameter as described in the $\theta(t_k) = \theta(t_{k-1}) + noise$ (Equation (14)); then, $\hat{\theta}_{up}(t_{k/k}) = \hat{\theta}_{up}(t_{k/k-1})$ for all steps from t_{k-1} to t_k . \square

4.2. 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 JEFK failure described in Theorem 1. The initial condition of MRDE is the initial state error covariance matrix $\mathbf{P}_0 = \mathbf{P}(t = 0)$. When it is composed of uncorrelated elements ($P_{i,j} = 0$), some initial conditions of time-invariant ODEs ($\frac{dP_{i,j}(t_{k|k-1})}{dt} = 0$) in the MRDE are zero, and consequently, the obtained solutions from t_{k-1} to t_k for some of these time-invariant ODEs are zero, too. Furthermore, in the presence of the biomanufacturing conditions (failure case presented in Section 4.1.1), 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 with regard to the unshared parameter (see Lemma 1), and they are constant and equal to zero during the entire process execution of JEFK. 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, that $K_{UP} = 0$ during the entire JEFK execution reflects 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 JEFK since there is no uncertainty in the prediction regarding the unshared parameter. This reflects the second intuition behind Kalman gain described in Section S2 of the Supplementary Material. 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 ($P_{i,j} = 0$) 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 ($P_{i,j} = 0$), we can consider only the key off-diagonal element (that is $P_{MSV,UP}(t = 0)$) with an initial value different of zero ($P_{MSV,UP}(t = 0) \neq 0$) to avoid the failure case. This value could be a positive quantity, λ , since the off-diagonal elements of $\mathbf{P}(t = 0)$ can show a positive covariance between two variables, indicating that they tend to increase or decrease together. Furthermore, the value of λ 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 ($P_{MSV,UP}(t = 0)$). If we add it to the initial state error covariance matrix $\mathbf{P}(t = 0)$ with the other uncorrelated elements, we have a specific initial condition for MRDE that enables us 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 JEFK failure.

Theorem 2 (SANTO—Proposed approach to avoid the JEFK failure). *The addition of a positive quantity (λ) to the $P_{MSV,UP}(t = 0)$ in $\mathbf{P}(t = 0)$ to initialize the MRDE with a specific initial condition can prevent the Kalman gain being zero in the entire execution of JEFK and prevent the JEFK failure (Section 4.1).*

Proof. The proof of the SANTO approach can be completed using the conditions described previously in the proof of Lemma 1 (Section 4.1.2) and Theorem 1 (Section 4.1.3).

Then, let us consider the following:

- A **positive** quantity λ .
- x_{msv} as the unique measured state variable (MSV) and $\mathbf{H} = [1 \ 0 \ \dots \ 0 \ 0]$.
- R as measurement noise variance of x_{msv} .
- θ_{up} as the unshared parameter (UP) to be evolved (estimated) and presented in only one weak term.
- A specific initial predicted state error covariance matrix $\mathbf{P}(t_{k-1}) = \mathbf{P}_0 = \mathbf{P}_{santo}(t = 0)$ with uncorrelated elements and $P_{MSV,UP}(t = 0) = P_{x_{msv},\theta_{up}}(t = 0) = \lambda$ as following

$$\mathbf{P}_{santo}(t=0) = \begin{bmatrix} P_{x_{msv},x_{msv}}(t=0) & 0 & \dots & 0 & 0 \\ 0 & P_{x_2,x_2}(t=0) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & P_{n,n}(t=0) & 0 \\ P_{x_{msv},\theta_{up}}(t=0) = \lambda & 0 & \dots & 0 & P_{\theta_{up},\theta_{up}}(t=0) \end{bmatrix}, \quad (31)$$

Now, using this Equation (24) to compute the predicted state error covariance matrix $\mathbf{P}(t_{k/k-1})$ from t_{k-1} to t_k with the specific initial predicted state error covariance matrix $\mathbf{P}(t_{k-1}) = \mathbf{P}_0 = \mathbf{P}_{santo}(t=0)$, we have

$$\mathbf{P}(t_{k/k-1}) = \begin{bmatrix} \vdots & \vdots & \ddots \\ P_{x_{msv},\theta_{up}}(t_{k/k-1}) = \lambda & P_{x_2,\theta_{up}}(t_{k/k-1}) & \dots \end{bmatrix}. \quad (32)$$

where $P_{x_{msv},\theta_{up}}(t_{k/k-1}) = \lambda$ because the solution of $\frac{dP_{x_{msv},\theta_{up}}(t)}{dt} = 0$ obtained from t_{k-1} to t_k is equal to the initial condition that is λ in \mathbf{P}_0 . Now, using $\mathbf{P}(t_{k/k-1})$, \mathbf{H} and \mathbf{R} to compute the Kalman gain for all variables in the state variable vector $\psi(t)$ (Equation (19)), we have

$$\mathbf{K}_k = \mathbf{P}(t_{k/k-1})\mathbf{H}^T(\mathbf{H}\mathbf{P}(t_{k/k-1})\mathbf{H}^T + \mathbf{R})^{-1} = \begin{bmatrix} \vdots \\ K_{\theta_{up}} \end{bmatrix} = \begin{bmatrix} \vdots \\ \frac{P_{x_{msv},\theta_{up}}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1}) + \mathbf{R}} \end{bmatrix} = \begin{bmatrix} \vdots \\ \frac{\lambda}{P_{x_{msv},x_{msv}}(t_{k/k-1}) + \mathbf{R}} \end{bmatrix}. \quad (33)$$

Then, we have the Kalman gain value for the unshared parameter as

$$K_{\theta_{up}} = \lambda(P_{x_{msv},x_{msv}}(t_{k/k-1}) + \mathbf{R})^{-1} \neq 0, \quad (34)$$

and consequently, the predicted state error covariance $P_{x_{msv},\theta_{up}}(t_{k/k-1})$ can be updated by Equation (12) and predicted mean of the state variable vector with regard to UP, $\hat{\theta}_{up}(t_{k/k-1})$ can be updated as Equation (11). Therefore, we have $P_{x_{msv},\theta_{up}}(t_{k/k}) \neq P_{x_{msv},\theta_{up}}(t_{k/k-1})$ and $\hat{\theta}_{up}(t_{k/k}) \neq \hat{\theta}_{up}(t_{k/k-1})$ during the entire execution of JEFK. \square

It is essential to point out that the SANTO is inspired by the idea of a regularization technique used to avoid the singularity problem in the state error covariance matrix [35,56]. However, instead of adding a small quantity to the diagonal elements of the state error covariance matrix \mathbf{P} , such as the perturbed-P algorithm [35], we only add a positive quantity (λ) to the $P_{MSV,UP}(t=0)$ in $\mathbf{P}(t=0)$ to initialize the MRDE. Furthermore, a positive quantity to the $P_{MSV,UP}(t=0)$ can be defined by empirical tuning. One of the most common ways to define a 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].

Figure 1 shows the steps to develop a soft sensor for bioprocess monitoring based on JEFK-SANTO.

- Step 1: Data Collection and Preprocessing. The first step in developing a soft sensor for bioprocess monitoring using the JEFK-SANTO approach involves comprehensive data collection and preprocessing. Once collected, these data must be meticulously cleaned and preprocessed to remove outliers and address any missing values. This preprocessing is crucial to ensure the quality and reliability of the data, which forms the foundation for accurate modeling and estimation in subsequent steps.
- Step 2: Analyze the Biomanufacturing Conditions. This step involves a comprehensive analysis of the biomanufacturing conditions where JEFK fails to estimate an unshared parameter that is part of a state variable vector and part of a weak term in a UMM if the initial state error covariance matrix $\mathbf{P}(t=0)$ and \mathbf{Q} are composed of uncorrelated elements, and there is only one state variable measured.
- Step 3: Implement JEFK with the SANTO approach. Implement the JEFK algorithm, defining the process model and the measurement model. Modify the initial state error covariance matrix $\mathbf{P}(t=0)$ as per the SANTO approach, adding a specific positive

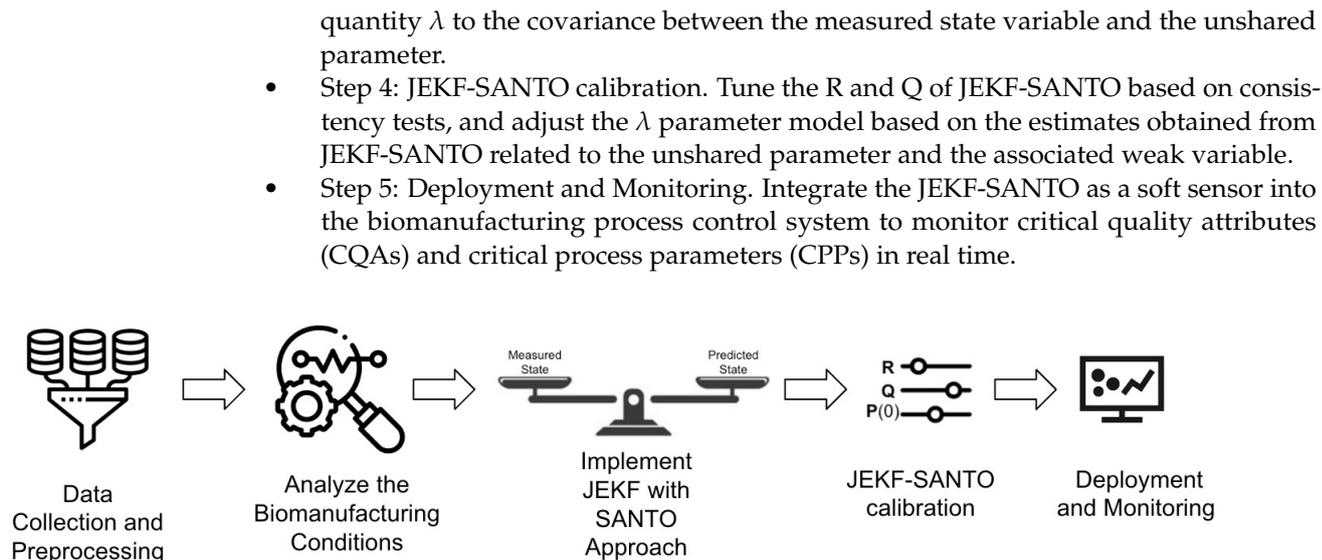


Figure 1. The basic steps to develop a soft sensor for bioprocess monitoring based on JEFK-SANTO.

5. 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: JEFK-Classic, JEFK-SANTO and JEFK-KPH2. First, the goals are the following:

- (**G1**) Experimentally test Theorem 1 (JEFK Failure) in Section 4.1.3;
- (**G2**) Test whether SANTO can avoid the JEFK 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 completed by JEFK-Classic 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 completed 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?

5.1. Experimental Setup

5.1.1. 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 comprises two runs (A-SD and B-SD) with different cell expansions and maximums of the mAb (titer) production. The runs of SD can be seen in Figure 2, and the runs have a sample rate of 7.5 minutes during 103 hours of the process. The runs were generated using the UMM proposed by [59] with small variations in parameters μ_{max} (maximum growth rate) and Q_{mAb} (mAb specific production rate) (see Table S1 of the Supplementary Material) but with the same initial concentrations of states variables (viable cell density (X_v), glucose (GLC), glutamine (GLN), lactate (LAC), ammonium (AMM) and mAb) and with different conditions of pH and temperature as completed in the synthetic dataset of [55]. The run A-SD (red lines in plots of Figure 2) was generated using the original parameters proposed by [59], which are the parameters $\mu_{max} = 5.8 \times 10^{-9} (\text{h}^{-1})$ and $Q_{mAb} = 7.21 (\times 10^{-9} \text{ mg cells}^{-1} \text{ h}^{-1})$. Run B-SD (blue lines in plots of Figure 2) has the maximum cell expansions and a maximum of mAb (titer) production of SD, and they were obtained with the parameters $\mu_{max} = 7.5 \times 10^{-9} (\text{h}^{-1})$ and $Q_{mAb} =$

$9.21 (\times 10^{-9} \text{ mg cells}^{-1}\text{h}^{-1})$. Furthermore, the run B-SD has samples regarding X_V (cell/L) with Gaussian white noise, and they were created by adding the Gaussian white noise with a standard deviation of 20×10^7 to the data represented in blue and green lines. The X_v of B-SD with noise is highlighted in light blue in the first plot. It is essential to point out that X_V samples with Gaussian white noise represent a possible online measurement with a sensor that includes noises. This noise is used to evaluate the performance of the NSEs (JEKF-Classic, JEKF-SANTO, and JEKF-KPH2) to estimate mAb and QmAb.

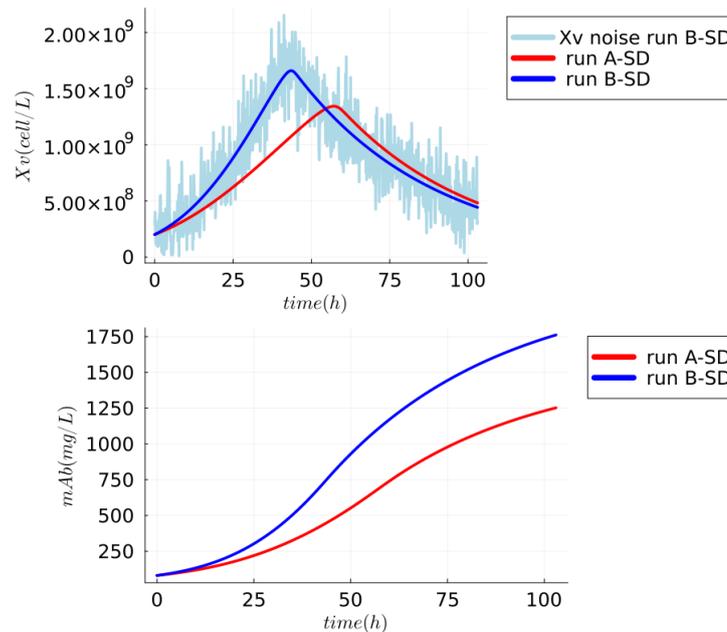


Figure 2. Synthetic dataset regarding mAb production. The run A-SD (red lines) was generated using the original parameters proposed by [59]. Run B-SD (blue lines) has the maximum cell expansions and the maximum mAb (titer) production of SD. The X_v of B-SD with noise is highlighted in light blue in the first plot. This noise is used to evaluate the performance of the NSE to estimate mAb and QmAb.

5.1.2. Real Dataset: AAV Production

The real dataset (RD) contains data regarding rAAV productions, which are 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 [60]. The RD has two runs with online and offline measurements of the state variables viable cell density (X_v), 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 X_v 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 [19].

5.1.3. 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 Section S1.4 of the Supplementary Material as a process model and the same initial concentration regarding the state variables; see Table S2 of the Supplementary Material. The NSEs were used to correct (estimate) the predictions regarding state variables (X_v and mAb) and to evolve the unshared parameter (QmAb) of the process model. This was accomplished using the X_v 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 of the process model (see Tables S1 and S2 of Supplementary Material). 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 X_v with noise. For example, the initial value used for Q_{mAb} is the value of run A-SD ($Q_{mAb} = 7.21 \times 10^{-9} \text{ mg cells}^{-1}\text{h}^{-1}$), and it should be evolved to the value of run B-SD ($9.21 \times 10^{-9} \text{ mg cells}^{-1}\text{h}^{-1}$) based on X_v with the noise of run B-SD. Furthermore, the X_v (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 made with MRDE formed by \mathbf{P} with correlated elements (MRDE-PC) and uncorrelated elements (MRDE-PU). In addition, MRDE-PC and MRDE-PU were combined with standard and specific $\mathbf{P}(t = 0)$ to check the sensitivity of SANTO (with regard to $P_{MSV,UP}(t = 0)$) and KPH2 (with regard to $P_{UP,UP}(t = 0)$). The standard $\mathbf{P}(t = 0)$ means that all NSEs used the same $\mathbf{P}(t = 0)$. On the other hand, the specific $\mathbf{P}(t = 0)$ means that each NSE used a different $\mathbf{P}(t = 0)$ that enables its best performance. For example, the specific $\mathbf{P}(t = 0)$ for SANTO contains a specific value of $P_{MSV,UP}(t = 0)$, and the specific $\mathbf{P}(t = 0)$ for KPH2 includes specific value of $P_{UP,UP}(t = 0)$. The specific $\mathbf{P}(t = 0)$ was obtained by trial and error, and a standard \mathbf{Q} was used for all NSEs. For example, the specific $\mathbf{P}(t = 0)$ for SANTO contains a specific value of $P_{MSV,UP}(t = 0) = \lambda$, and the specific $\mathbf{P}(t = 0)$ for KPH2 includes a specific value of $P_{UP,UP}(t = 0)$. The values of specific $\mathbf{P}(t = 0)$ (including λ) were obtained by trial and error. Furthermore, a standard and specific \mathbf{Q} were also used for all NSEs. In addition, the root-mean-square percentage error (RMSPE) 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 Section S7.3 of the Supplementary Material.

5.1.4. NSEs Assessment with Real Dataset to Address RQ3-G2

The NSEs (JEKF-Classic, JEKF-SANTO, and JEKF-KPH2) used the UMM described in Section S1.5 of the Supplementary Material as a process model and the same initial concentration regarding the state variables; see Table S8 of the Supplementary Material. These three NSEs were used to correct (estimate) the predictions regarding X_v , GLC, LAC, and rAAV (titer) and to evolve the unshared parameters (μ_{Lac} , μ_{GLC} and μ_{rAAV}) of the process model. This was accomplished using the X_v 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 Table S9 of the Supplementary Material). 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 X_v with noise. However, in this case, the NSEs have to correct three different unshared parameters simultaneously based on X_v with the noise of run B-RD. Furthermore, the RMSPE 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 essential to point out that the estimations were also completed with MRDE-PC and with specific $\mathbf{P}(t = 0)$. The details about the design of NSEs with RD can be found in Section S7.4 of the Supplementary Material.

5.1.5. Checking Consistency and Efficiency

The calibration of standard and specific \mathbf{Q} were based on consistency tests, specifically the innovation magnitude bound (IMB) test and the normalized innovations squared (NIS) Chi-square test [61]. These two tests are used to check that the NSEs are performing correctly with \mathbf{Q} and \mathbf{R} selected [50,62].

IMB Test. It checks that the innovation is consistent with its covariance by verifying that the magnitude of the innovation is bounded by $\pm 2\sqrt{\mathbf{S}_k}$. A positive result in this test occurs when at least 95% of the values of the innovation lie within the $\pm 2\sqrt{\mathbf{S}_k}$. Figure 3 presents the innovation error sequence for the NSEs configured with MRDE-PC, utilizing specific \mathbf{Q} and $\mathbf{P}(0)$ settings as detailed in Tables S4–S6 of the Supplementary Material using run B of the synthetic dataset. This figure demonstrates that the innovation errors

are unbiased with approximately 95.14% of the values falling within the $\pm 2\sqrt{\mathbf{S}_k}$ bounds as required. Similar outcomes were observed for NSEs configured with both MRDE-PC and MRDE-PU, and irrespective of whether standard or specific \mathbf{Q} and $\mathbf{P}(0)$ settings were employed, as shown in Figures S1–S3 using \mathbf{Q} and $\mathbf{P}(0)$ as detailed in Tables S4, S5, and S7 of the Supplementary Material. Each of these configurations yielded similar innovation error characteristics, underscoring the robustness of the NSEs under varying conditions. Furthermore, similar results were also obtained with run B of the real dataset. Figure 4 presents the innovation error sequence for the NSEs configured with MRDE-PC, utilizing specific \mathbf{Q} and $\mathbf{P}(0)$ settings as detailed in Tables S10 and S11 of the Supplementary Material. This figure demonstrates that the innovation errors are unbiased with approximately 95.9% of the values falling within the $\pm 2\sqrt{\mathbf{S}_k}$ bounds as required.

Complementarily, standard error (SE) plots, based on the \mathbf{P} matrix's diagonal, demonstrate the changing uncertainty in state estimates. Filter stability and consistency are indicated by SEs, related to the measured state variable, converging to a stable value. This convergence signifies the adaptability and equilibrium of a filter in making accurate predictions. The alignment of positive innovation test results with this convergent SE (of measured state variable) trend substantiates the overall stability and consistency of a filter. Figure 5 depicts the SE over time of X_V (measured state variable) estimated by NSEs with a synthetic dataset using MRDE-PC and specific $\mathbf{P}(0)$. Initially, these errors exhibited an increase, reflecting a period of adaptation as the filter assimilated the initial data. However, after this initial phase, the standard errors converged around a stable value. This convergence signifies the increasing reliability of the filter in estimating the state of X_V as it processed more data. The initial increase followed by a steady convergence of the standard errors, in tandem with the favorable innovation test results, compellingly demonstrates the robustness of the NSEs. Similar results were obtained with NSEs with the synthetic dataset using MRDE-PU and specific $\mathbf{P}(0)$ (Figure S4 of the Supplementary Material) and with NSEs with the real dataset using MRDE-PC and specific $\mathbf{P}(0)$; see Figure 6. It is important to point out that Figures S5–S8 in the Supplementary Material show the normal behavior of standard errors for the state variables (QmAb and mAb) estimated by JEFK-SANTO and JEFK-KPH2 with the synthetic dataset. Similarly, Figures S10–S14 in the Supplementary Material show the standard errors for the state variables (GLC, LAC, rAAV, μ_{GLC} , μ_{LAC} , and μ_{rAAV}) estimated by JEFK-SANTO and JEFK-KPH2 with the synthetic dataset.

NIS Chi-square Test. It verifies that the innovation is unbiased and white by using hypothesis testing (χ^2 test) [50,62]. The NIS is defined as $NIS_k = e_{Z,k} \mathbf{S}_k^- e_{Z,k}$, and the mean of NIS is defined as $\mu(NIS) = \frac{1}{N} \sum_{k=1}^N e_{Z,k} \mathbf{S}_k^- e_{Z,k}$ from a single run of a JEFK. Therefore, the NIS test involves verifying that $\mu(NIS)$ lies in the confidence interval $[r1, r2]$ defined by the hypothesis H_0 that $N \times \mu(NIS)$ is χ_{Nm}^2 distributed with probability $1-\alpha$, such that $P(N \times \mu(NIS) \in [r1, r2] | H_0) = 1 - \alpha$ where m is the number of measured state variables and N is the number of samples from the measured state variables. In our case, $m=1$ because we have only one measure state variable, and $N = 824$ for SD and $N = 2901$ for RD. Furthermore, for the case of a two-sided 95% confidence region, we have $[r1, r2] = [\chi_{Nm}^2(0.025), \chi_{Nm}^2(0.975)]$. Therefore, the NIS test of NSEs with the synthetic dataset is concerned with answering the following question: Is $N \times \mu(NIS)$ inside of $[\chi_{824}^2(0.025), \chi_{824}^2(0.975)] = [745.39, 904.39]$ where $N=824$, such that $P(N \times \mu(NIS) \in [745.39, 904.39] | H_0) = 1 - \alpha$? All the NSEs designed with the synthetic dataset using the \mathbf{Q} , \mathbf{R} and $\mathbf{P}(0)$ defined in Tables S3–S7 of the Supplementary Material presented the $N \times \mu(NIS)$ falling inside of the confidence bound defined by the χ^2 test. The NSEs with MRDE-PU had an $N \times \mu(NIS) = 835.29$ and NSEs with MRDE-PC presented a $N \times \mu(NIS) = 830.35$. Furthermore, the NSEs with the real dataset had a positive result in the χ^2 test with the following question: Is $N \times \mu(NIS)$ inside of $[\chi_{2901}^2(0.025), \chi_{2901}^2(0.975)] = [2752.63, 3051.15]$ where $N=2901$? NSEs with MRDE-PC and, \mathbf{Q} , \mathbf{R} and $\mathbf{P}(0)$ defined in Tables S10 and S11 of the Supplementary Material presented an $N \times \mu(NIS) = 2840.55$.

Normalized estimation error squared (NEES) test. It is the metric used to evaluate the efficiency of the JEFK-SANTO as an estimator. This involves verifying that the actual

estimation errors ($e_{x,k}$) appropriately match the predictions made by the $\mathbf{P}(t_{k/k})$ [62]. Essentially, if the $\mathbf{P}(t_{k/k})$ predicts a certain degree of uncertainty, it is expected for the real-world errors $e_{x,k}$ to match this prediction. This match is crucial for the estimator to be considered accurate and reliable. NEES is calculated as $NEES(k) = e_{x,k}^T \mathbf{P}(t_{k/k})^{-1} e_{x,k}$ where $e_{x,k}$ is the estimation error at time step k , defined as $e_{x,k} = \mathbf{x}(t_k) - \hat{\mathbf{x}}(t_{k/k})$, with $\mathbf{x}(t_k)$ being the true state and $\hat{\mathbf{x}}(t_{k/k})$ being the estimated state. Then, for the case of a single run, the $NEES(k)$ is Chi-square distributed with n_x degrees of freedom. In our case, we have $n_x = 3$ because we are concerned with evaluating the performance of JEFK-SANTO to estimate the states X_V , QmAb and mAb of the synthetic dataset. Therefore, we consider a one-sided 95% probability region as seen in Bar-Shalom et al. in [62] for single-run simulation tests with small degrees of freedom. We have the hypothesis H_0 that JEFK-SANTO's efficiency ($e_{x,k}$ matches $\mathbf{P}(t_{k/k})$), and H_0 is accepted if $P(NEES(k) \leq \chi_3^2(0.95) = 7.815 | H_0) = 1 - \alpha$. Figure 7 depicts the result of NEES for the JEFK-SANTO with the synthetic dataset using MRDE-PC and MRDE-PU. The designated upper threshold for the acceptance region is set at 7.815. The majority of the $NEES(k)$ values are observed to fall within the defined confidence interval $[0, \chi_3^2(0.95) = 7.815]$, which means the estimation error and the covariance are compatible with each other, and the estimation of the JEFK-SANTO is reliable and credible. Moreover, these findings are in alignment with those reported by Bar-Shalom et al. in [62], particularly in the context of single-run simulation tests with a small number of degrees of freedom.

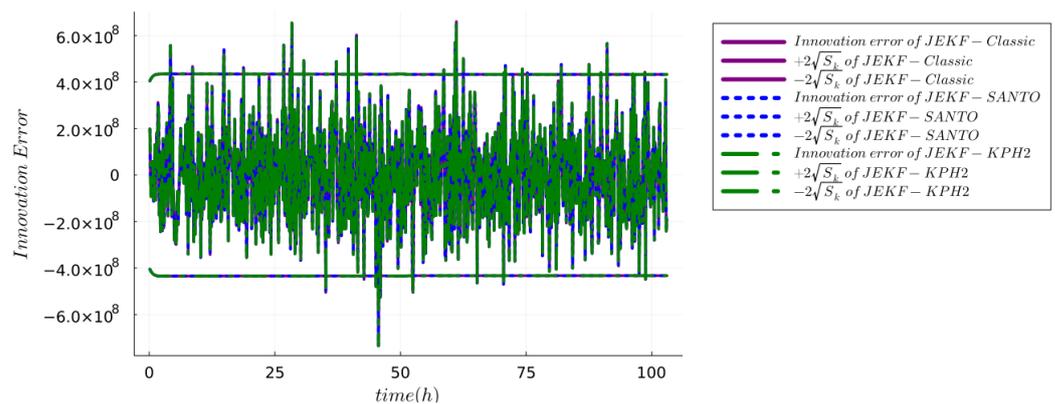


Figure 3. Innovation magnitude bound test using the run B of synthetic dataset for the NSEs with MRDE-PC and specific \mathbf{Q} and $\mathbf{P}(0)$.

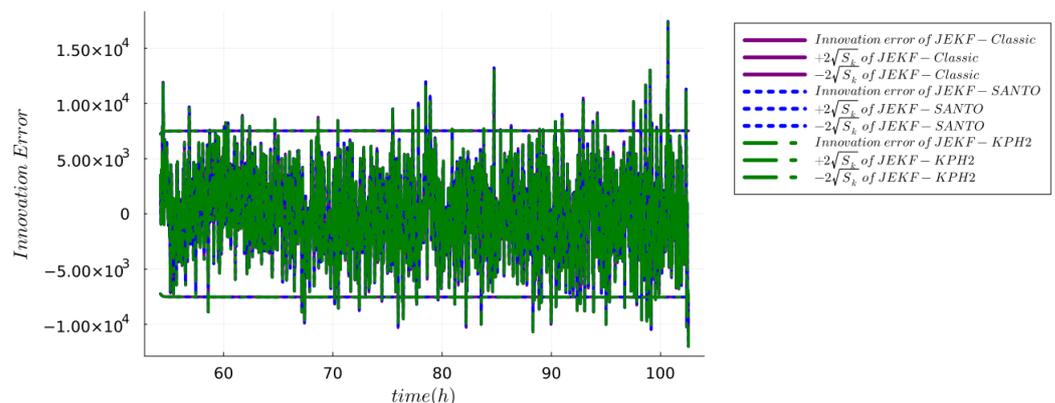


Figure 4. Innovation magnitude bound test using the run B of real dataset for the NSEs with MRDE-PC and specific \mathbf{Q} and $\mathbf{P}(0)$.

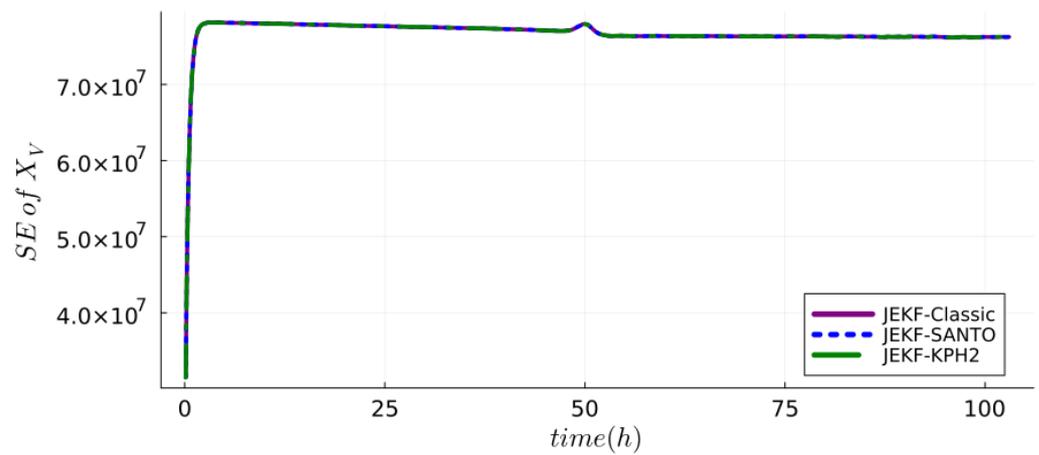


Figure 5. Standard error of X_V at each k estimated by NSEs with synthetic dataset (run B) using MRDE-PC and specific $P(0)$.

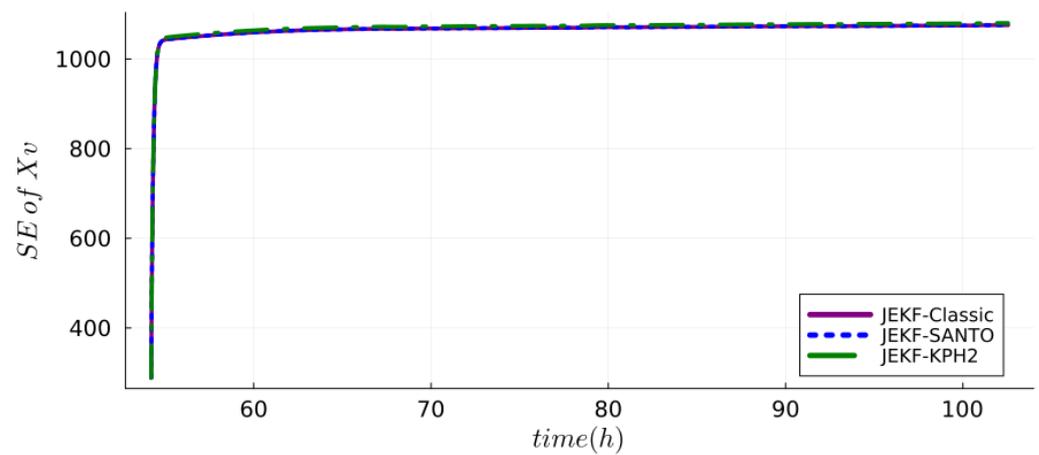


Figure 6. Standard error of X_V at each k estimated by NSEs with real dataset using MRDE-PC and specific $P(0)$.

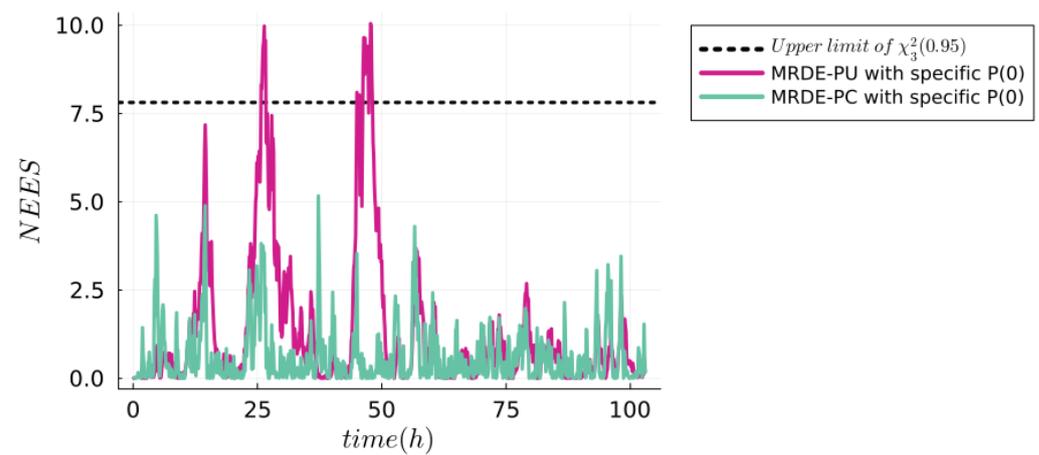


Figure 7. NEES test of JEFK-SANTO with synthetic dataset. In case of JEFK-SANTO with MRDE-PC with specific $P(0)$, we have that 100% of all NEES computed are found inside the one-sided 95% probability region where the 5% tail is $\chi_3^2(0.95) = 7.815$ (upper limit). In case of JEFK-SANTO with MRDE-PU with specific $P(0)$, we have 98.4% of the NEES inside of confident interval $[0, \chi_3^2(0.95) = 7.815]$.

6. Results

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

Answer to RQ1-G1. The results of the experimental test of Theorem 1 (JEKF failure) can be seen in Figures 8 and 9. We also reported the estimations made using JEKF-SANTO and JEKF-KPH2 in regard to X_v , mAb , and $QmAb$ of mAb production (run B-SD) using MRDE-PC and MRDE-PU with specific $P(t=0)$. In plot A of Figures 8 and 9, we can see that all NSEs estimated the X_v close to the ground truth. However, the JEKF-Classic (purple line) was not able to evolve (update) the unshared parameter $QmAb$, because the estimations about $QmAb$ were constant and equal to the initial value of $7.21 \times 10^{-9} \text{ mg cells}^{-1}\text{h}^{-1}$ during the entire process. Consequently, the JEKF-Classic estimation regarding mAb was far from the ground truth (red dash line) of run B-SD (see plots B and C in Figure 8), and it had a high RMSPE value of 18.65%; see Table 1. The same results regarding the JEKF-Classic were obtained using MRDE-PU; see Figure 9. It is important to point out that the Kalman gain over time obtained by JEKF-Classic with SD is constant and equal to zero using MRDE-PU or MRDE-PC (see Figure 10). Furthermore, the Kalman gain values obtained by JEKF-SANTO with MRDE-PC were more stable than those obtained by JEKF-KPH2.

Answer to RQ2-G2. The results of JEKF-SANTO avoiding the JEKF failure (using runs B-SD) can be seen in the plots B and C of Figures 8 and 9. In these plots, we can see that JEKF-SANTO (blue line) evolved the $QmAb$ from the initial value to the ground truth (red dash line) and consequently estimated the mAb close to the ground truth of run B-SD (red dash line) with MRDE-PU and MRDE-PC. These results are the opposite of the ones obtained with JEKF-Classic. In addition, JEKF-SANTO had the smallest RMSPE values between the NSEs in all cases; see Table 1. On the other hand, the JEKF-KPH2 did not perform similarly to JEKF-SANTO. The unique case where JEKF-KPH2 (green line) had a good performance was in run B-SD with MRDE-PU with specific $P(t=0)$. In that case, JEKF-KPH2 estimations were near to the ground truth (red dash line); see plots B and C of Figure 9. However, JEKF-KPH2 did not present stability, and the estimation converged to values far from the ground truth in run B-SD (with MRDE-PC with specific $P(t=0)$). The best performances of JEKF-SANTO and JEKF-KPH2 were obtained by the use of specific $P(t=0)$ because when we used a standard $P(t=0)$ for JEKF-SANTO and JEKF-KPH2, their estimations are worse with runs B-SD. The results using standard $P(t=0)$ with runs B-SD can be found in Figures S15 and S16 and Table S12 of the Supplementary Material. These results (with standard and specific $P(t=0)$) show that JEKF-KPH2 is sensitive to the initial $P_{QmAb, QmAb}(t=0)$, and JEKF-SANTO is sensitive to $P_{X_v, QmAb}(t=0)$, since their better results were obtained with their specific $P(t=0)$. Table S4 of the Supplementary Material shows the specific $P(t=0)$ used in JEKF-KPH2, and Table S5 of the Supplementary Material shows the specific $P(t=0)$ used in JEKF-SANTO. It is important to point out that the best results of JEKF-SANTO were obtained with $P_{X_v, QmAb}(t=0)$ with positive values in case of run B-SD; see Table S5 of the Supplementary Material.

Answer to RQ3-G2. In Figure 11, we show the estimations made by JEKF-SANTO and JEKF-KPH2 with regard to X_v , GLC, LAC, $rAAV$ and the three 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 plot A of Figure 11, we can see that JEKF-SANTO and JEKF-KPH2 estimated the X_v inside of the noise range of the real online measurement of X_v by the capacitance probe. The following plots, B, C, and D, show the estimation obtained for the variables GLC, LAC, and $rAAV$. JEKF-SANTO (blue line) and JEKF-KPH2 (green line) were able to evolve the three unshared parameters simultaneously converging to values that enabled the estimation of GLC, LAC, and $rAAV$ near the ground truth (red points in plots B, C and D). In these plots of Figure 11, and the RMSPE in Table 2, we can see that JEKF-SANTO and JEKF-KPH2 made similar estimations. Nevertheless, JEKF-SANTO had a slightly better performance than JEKF-KPH2 estimating GLC, LAC, and $rAAV$ (titer); see the RMSPE Table 2. It is important to point out that the Kalman gain over time obtained by JEKF-Classic with RD is constant and equal to zero. See Figure 12. Consequently, JEKF-

Classic had the worst performance and could not evolve the three unshared parameters simultaneously; see plots of Figure 11, and the RMSPE Table 2.

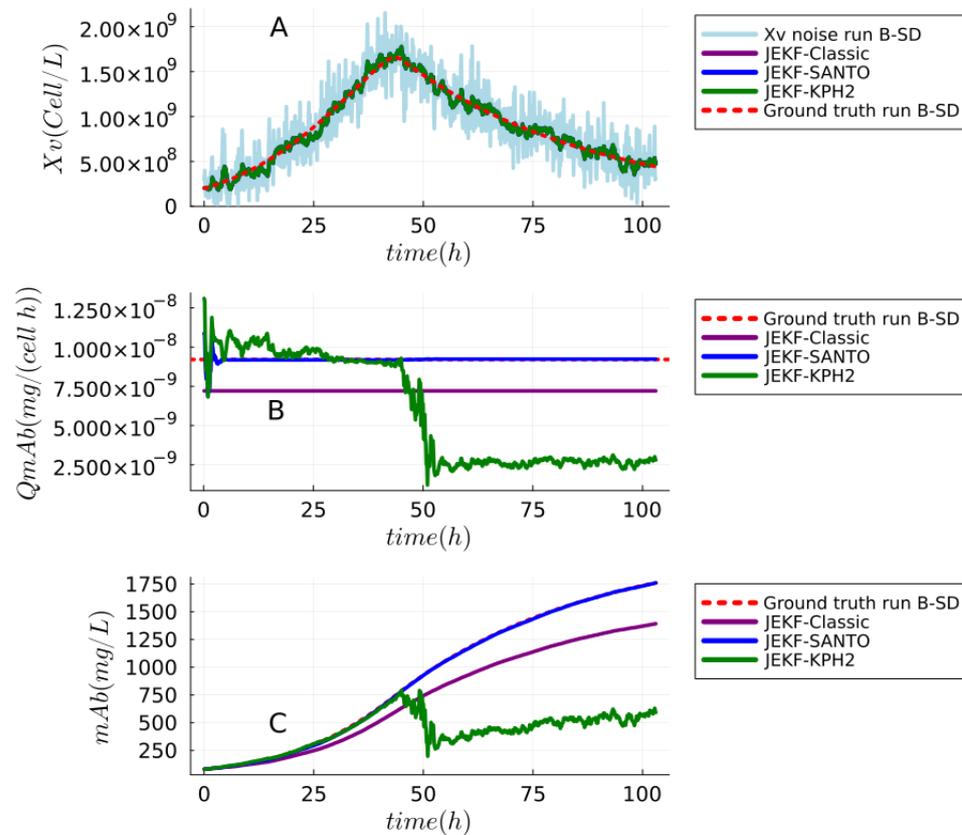


Figure 8. Experimental test of the theorem (JEKF failure) and the JEKF-SANTO to avoid the JEKF failure with the biomanufacturing conditions (failure case). This experiment used run B of the synthetic dataset, and plot (A) shows that all estimations with regard to X_v were close to the ground truth. Plots (B,C) show the estimations with regard to the unshared parameter Q_{mAb} and m_{Ab} (titer), respectively. The JEKF-SANTO was able to evolve Q_{mAb} with convergence to the ground truth value, but JEKF-KPH2 and JEKF-Classic failed. They were not able to evolve the m_{Ab} . All NSEs were executed with MRDE-PC and specific $P(t = 0)$.

Table 1. RMSPE between NSEs estimations about m_{Ab} and ground truth of run B in synthetic dataset with specific $P(t = 0)$.

NSE	RMSPE (MRDE-PU)	RMSPE (MRDE-PC)
JEKF-SANTO	2.06%	1.30%
JEKF-KPH2	10.11%	48.44%
JEKF-Classic	18.80%	18.65%

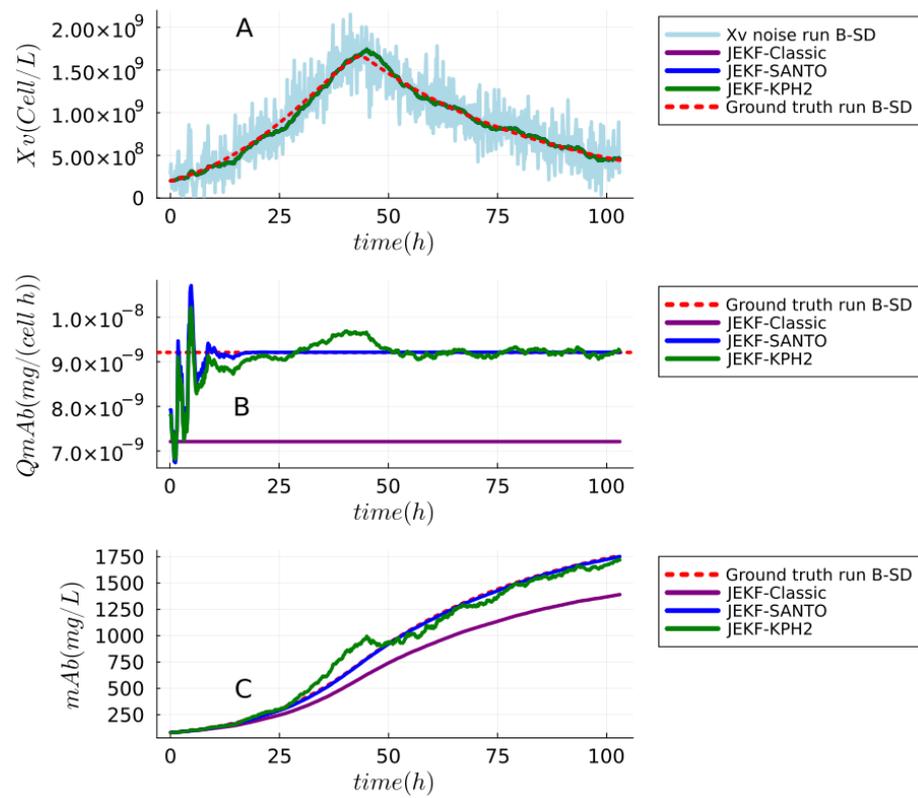


Figure 9. Experimental test that JEFK-Classic cannot avoid the JEFK failure with run B of the synthetic dataset. First, plot (A) shows the estimations regarding X_v , and all estimations were close to the ground truth. The plots (B,C) show the estimations regarding the unshared parameter Q_mAb and mAb (titer), respectively. All NSEs evolved Q_mAb with convergence to the ground truth value except JEFK-Classic. All NSEs were executed with MRDE-PU and specific $P_{UP,UP}(t = 0)$.

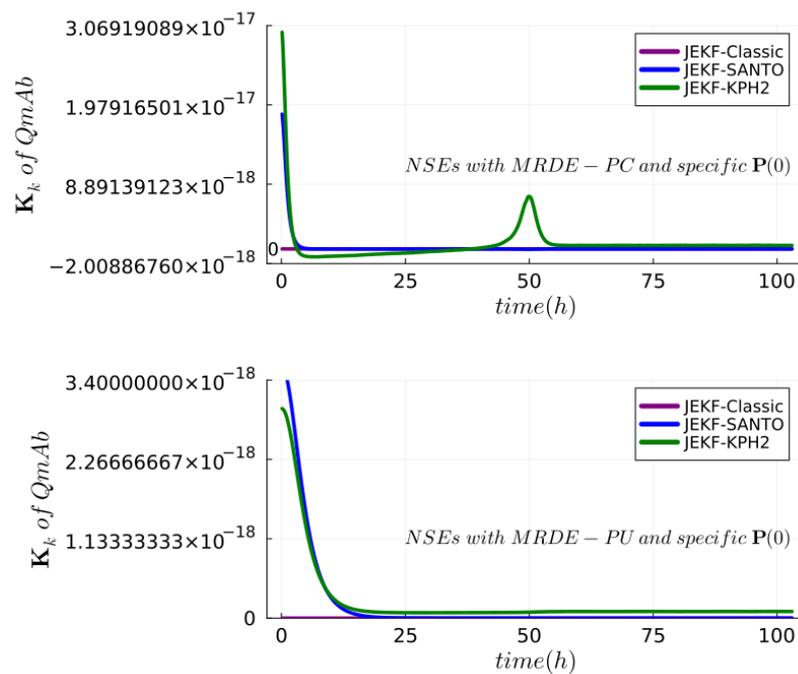


Figure 10. Kalman gain over time for the NSEs with run B of synthetic dataset. In all cases, JEFK-Classic is constant and equal to zero.

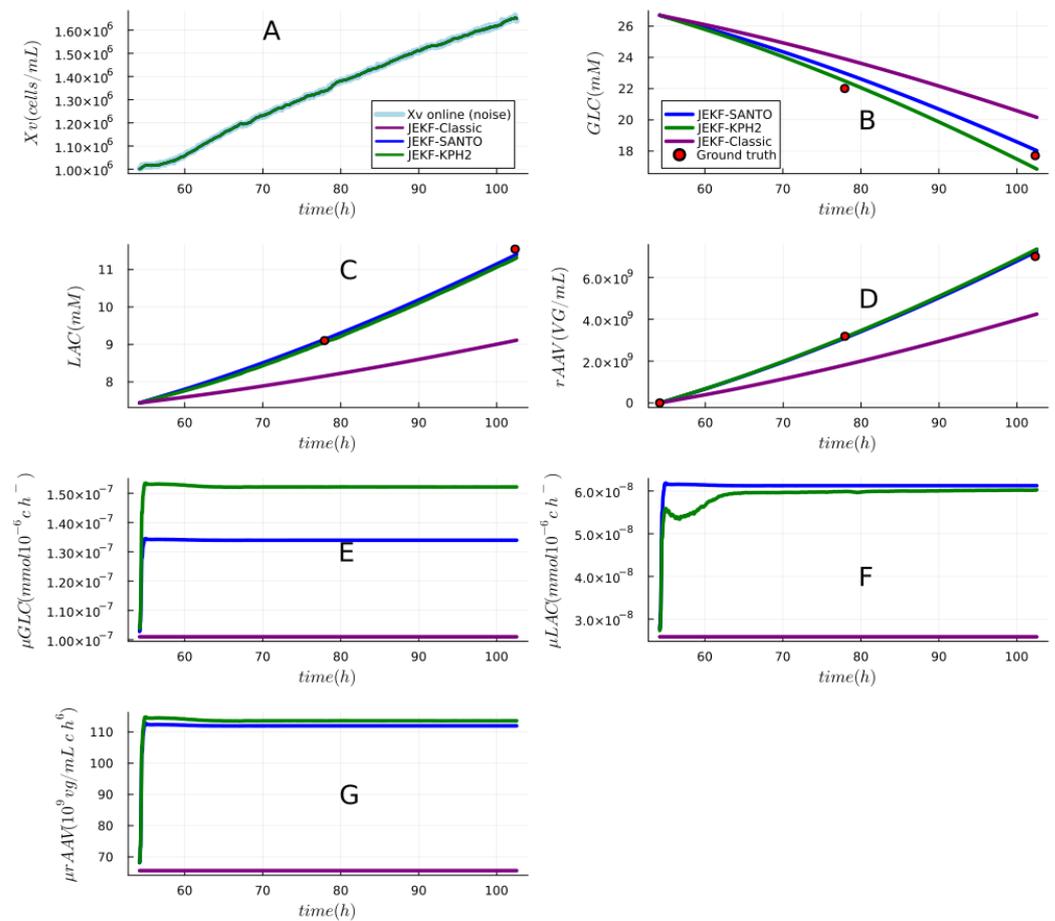


Figure 11. Simultaneous unshared parameters estimation by JEFK-SANTO and JEFK-KPH2 with real dataset (rAAV production). Plot (A) shows the estimations regarding X_v , and all estimations were inside of the noise range of the real online measurement of X_v by the capacitance probe. Plots (B–D) show the estimation obtained regarding the variables GLC, LAC, and rAAV, respectively. In these plots, we can see that JEFK-SANTO and JEFK-KPH2 had similar estimations. They evolved the μ_{GLC} , μ_{LAC} , and μ_{rAAV} (unshared parameters) with convergence, and their estimations related to GLC, LAC, and rAAV were close to the ground truth (red points); see plots (E–G). All NSEs were executed with MRDE-PC and specific $\mathbf{P}(t = 0)$.

Table 2. RMSPE between NSEs estimations and ground truth of real dataset with MRDE-PC, specific $\mathbf{P}(t = 0)$ and (standard \mathbf{Q}).

Ground Truth	JEFK-Classical	JEFK-SANTO	JEFK-KPH2
GLC	11.6%	3.48%	3.58%
LAC	16.66%	1.01%	1.61%
rAAV (titer)	41.41%	2.87%	3.28%

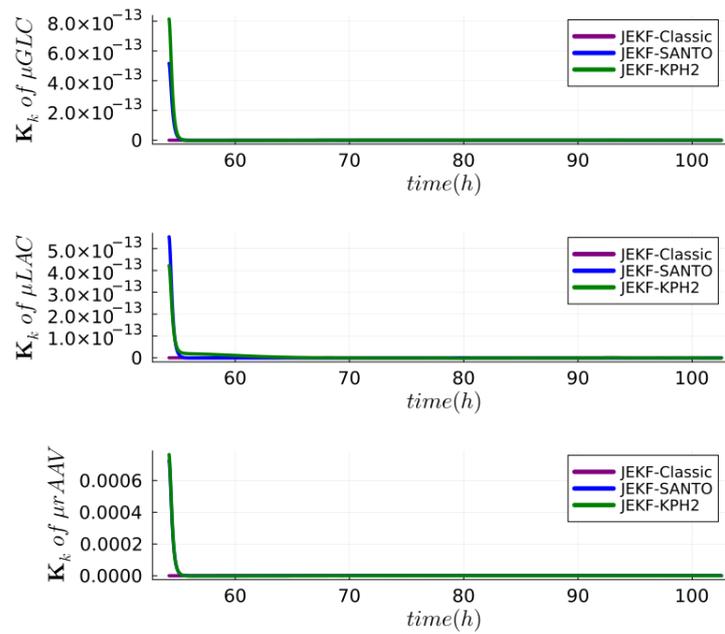


Figure 12. Kalman gain over time for the NSEs with run B of real dataset. In all cases, the JEFK-Classic is constant and equal to zero.

7. Discussion

Our theoretical and empirical results showed the JEFK failure with biomanufacturing conditions. These results showed that JEFK-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 from the beginning to the end of the processes tested. On the other hand, the results showed that the JEFK-SANTO and JEFK-KPH2 approaches can avoid the JEFK failure. However, the JEFK-SANTO had a more accurate estimation than JEFK-KPH2 while having faster and stable unshared parameters evolution to values that allowed the best performance of the process model tested. It is essential to point out that JEFK-SANTO performed best in two different situations, which were represented by run B-SD with MRDE-PC and MRDE-PC. The best performance of JEFK-KPH2 was only with run B-SD. Furthermore, the results showed that both approaches are sensitive to $\mathbf{P}(t=0)$. JEFK-KPH2 is sensible to the $P_{UP,UP}(t=0)$, and JEFK-SANTO is sensible to $P_{MSV,UP}(t=0)$. It is essential to point out that the JEFK-SANTO approach did not change the probabilistic view of JEFK, and the minimization cost function in JEFK remained the same. Therefore, the JEFK-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 JEFK-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. Beyond the SANTO approach, several methods have been established to tackle singularities and convergence issues in EKF. Rank reduction techniques address ill-conditioned covariance matrices by reducing their dimensionality, thus preventing singularities [63]. Time-correlated noise analysis allows for a more accurate state estimation by adjusting noise covariance matrices based on observed temporal correlations in system noise, providing a more realistic noise model [64]. These methods, used in conjunction with JEFK-SANTO, offer a comprehensive approach to EKF optimization in challenging scenarios like biomanufacturing. It is important to note that our analysis did not explicitly consider observability and stability conditions. However, this omission does not invalidate our study. The focus of our research was on addressing a specific failure case of JEFK under certain biomanufacturing conditions. Our proposed solution, SANTO, was developed to specifically address this issue based on experiments with JEFK that are consistent. Therefore, in our study, while we did not explicitly detail the

observability and stability analysis in the traditional sense, we implicitly addressed these aspects through empirical evaluation methods. Regarding observability, our approach primarily focused on the empirical performance of the JEFK in the given case study rather than a formal observability analysis.

8. Conclusions

In this work, firstly, we presented the common conditions in biomanufacturing that represent a failure case for the classical JEFK. Secondly, we proved that the classical JEFK, with these conditions, cannot 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 JEFK failure case by adding a positive quantity (λ) regarding the initial state error covariance between a measured state variable and an unshared parameter ($P_{MSV,UP}(t = 0)$) in $P(t = 0)$. Our empirical evaluation demonstrated that the SANTO approach effectively estimates unshared parameters and states simultaneously, aligning closely with ground truth values in the tested datasets. SANTO notably outperformed both JEFK-Classic and JEFK-KPH2 in accuracy. In a rigorously controlled test using a synthetic dataset, JEFK-SANTO, whether paired with MRDE-PC or MRDE-PU, exhibited a substantial improvement in RMSPE, achieving up to approximately 17% enhancement compared to JEFK-Classic. Meanwhile, JEFK-KPH2 showed an improvement of around 8.7% in RMSPE, but this was limited to its execution with MRDE-PU. This highlights the effectiveness of SANTO in overcoming the limitations of classical JEFK in biomanufacturing applications. Our future works will focus on the development of an auto-tuning mechanism based on an objective function to systematically calibrate Q , R and λ , as seen in [65], but also investigate the potential of the Unscented Kalman Filter (UKF) to estimate the unshared parameters and the state simultaneously with the biomanufacturing conditions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/s24020653/s1>, Figure S1: Innovation Magnitude Bound Test using the run B of Synthetic dataset for the NSEs with MRDE-PU and specific Q and $P(0)$; Figure S2: Innovation Magnitude Bound Test using the run B of Synthetic dataset for the NSEs with MRDE-PC and standard Q and $P(0)$; Figure S3: Innovation Magnitude Bound Test using the run B of Synthetic dataset for the NSEs with MRDE-PU and standard Q and $P(0)$; Figure S4: Standard Error of X_V at each k estimated by NSEs with Synthetic Dataset using MRDE-PU and specific $P(0)$; Figure S5: Standard Error of $QmAb$ at each k estimated by NSEs with Synthetic Dataset using MRDE-PC and specific $P(0)$; Figure S6: Standard Error of $QmAb$ at each k estimated by NSEs with Synthetic Dataset using MRDE-PU and specific $P(0)$; Figure S7: Standard Error of mAb at each k estimated by NSEs with Synthetic Dataset using MRDE-PC and specific $P(0)$; Figure S8: Standard Error of mAb at each k estimated by NSEs with Synthetic Dataset using MRDE-PU and specific $P(0)$; Figure S9: Standard Error of GLC at each k estimated by NSEs with Real Dataset using MRDE-PC and specific $P(0)$; Figure S10: Standard Error of LAC at each k estimated by NSEs with Real Dataset using MRDE-PC and specific $P(0)$; Figure S11: Standard Error of $rAAV$ at each k estimated by NSEs with Real Dataset using MRDE-PC and specific $P(0)$; Figure S12: Standard Error of μGLC at each k estimated by NSEs with Real Dataset using MRDE-PC and specific $P(0)$; Figure S13: Standard Error of μLAC at each k estimated by NSEs with Real Dataset using MRDE-PC and specific $P(0)$; Figure S14: Standard Error of $\mu rAAV$ at each k estimated by NSEs with Real Dataset using MRDE-PC and specific $P(0)$; Figure S15: The JEFK-SANTO and JEFK-KPH2 avoid the JEFK failure in B-SD, but they need a specific $P(t = 0)$. First, plot A shows the estimations regards X_v , and all estimations were close the ground truth. The plots B and C show the estimations regards the unshared parameter $QmAb$ and mAb (titer) far from the ground truth, respectively. The NSEs were executed with **MRDE-PC** and **standard** $P(t = 0)$; Figure S16: The JEFK-SANTO and JEFK-KPH2 avoid the JEFK failure in B-SD, but they need a specific $P(t = 0)$. First, plot A shows the estimations regards X_v , and all estimations were close the ground truth. The plots B and C show the estimations regards the unshared parameter $QmAb$ and mAb (titer) far from the ground truth, respectively. The NSEs were executed with **MRDE-PU** and **standard** $P(t = 0)$.; Table S1: Parameters used in UMM case 1.4 to generate the runs A-SD, and B-SD

of Synthetic Dataset (SD); Table S2: Initial conditions of state variables of UMM case 4 for the JEKF test with Synthetic Dataset; Table S3: Standard initial state error covariance matrix (standard $\mathbf{P}(t=0)$) for JEKF-Classic, JEKF-KPH2 and JEKF-SANTO with run B of Synthetic Dataset; Table S4: Specific initial state error covariance matrix (specific $\mathbf{P}(t=0)$) for JEKF-KPH2 with run B of Synthetic Dataset; Table S5: Specific initial state error covariance matrix (specific $\mathbf{P}(t=0)$) for JEKF-SANTO with run B of Synthetic Dataset; Table S6: Measurement noise variance \mathbf{R} and error covariance matrix of process model (\mathbf{Q}) for the JEKF-Classic, JEKF-SANTO and JEKF-KPH2 with run B of Synthetic Dataset using MRDE-PC; Table S7: Measurement noise variance \mathbf{R} and error covariance matrix of process model (\mathbf{Q}) for the JEKF-Classic, JEKF-SANTO and JEKF-KPH2 with run B of Synthetic Dataset using MRDE-PU; Table S8: Initial conditions of state variables of UMM case 5 for the JEKF-SANTO and JEKF-KPH2 test with run B-RD (Source [19]); Table S9: Initial parameters obtained with A-RD for the JEKF-SANTO and JEKF-KPH2 test with run B-RD (Source [19]); Table S10: Specific initial state error covariance matrix (specific $\mathbf{P}(t=0)$) for for the JEKF-Classic, JEKF-SANTO and JEKF-KPH2 with Real Dataset (run B) using MRDE-PC; Table S11: Measurement noise variance \mathbf{R} , and error covariance matrix of process model $\mathbf{Q}_{i,j}$ for the JEKF-Classic, JEKF-SANTO and JEKF-KPH2 with Real Dataset (run B) using MRDE-PC; Table S12: RMSPE between NSEs estimations about mAb and ground truth of run B in synthetic dataset with standard $\mathbf{P}(t=0)$.

Author Contributions: Conceptualization, C.F.I.J.; Implemented algorithms and conducted the experiments, C.F.I.J.; Performed analysis on experimental results and wrote the manuscript, C.F.I.J.; Provided insightful discussions, reviewed the results and revised the manuscript, M.B.; Supervision, M.B.; Project administration, M.B.; funding acquisition, M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research Council through AI for Design Challenge Program (operating grants AI4D-103-1).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and code used in this study are available in github <https://github.com/cristovaoiglesias/JEKF-SANTO> (accessed on 20 November 2023).

Acknowledgments: We want to acknowledge the support provided by National Research Council Canada through the AI4D grant. We are grateful to 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 Nabil Belacel from National Research Council Canada for his constructive comments and discussions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

JEKF	Joint estimation of states and parameters with Extended Kalman Filter
NSE	Nonlinear State Estimator
UMM	Unstructured Mechanistic Model
SMM	Structured Mechanistic Model
MRDE	Matrix Ricatti Differential Equation
MSV	Measured State Variable
UP	Unshared parameter
SANTO	Specific initial coNdiTiOn
CD-EKF	Continuous-Discrete EKF
RD	Real dataset
SD	Synthetic dataset
rAAV	Recombinant Adeno-Associated Virus
mAb	Monoclonal Antibody

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* **2021**, *21*, 2085. [CrossRef] [PubMed]
2. Alexander, R.; Campani, G.; Dinh, S.; Lima, F.V. Challenges and opportunities on nonlinear state estimation of chemical and biochemical processes. *Processes* **2020**, *8*, 1462. [CrossRef]
3. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng. Mar* **1960**, *82*, 35–45. [CrossRef]
4. Jazwinski, A. *Stochastic Processes and Filtering Theory*; Mathematics in Science and Engineering; Academic Press: Cambridge, MA, USA, 1970. Available online: <https://books.google.ch/books?id=nGISNvKyY2MC> (accessed on 20 November 2023).
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*; World Scientific: Singapore, 2022; pp. 223–252.
6. Ljung, L. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Trans. Autom. Control* **1979**, *24*, 36–50. [CrossRef]
7. Kopp, R.E.; Orford, R.J. Linear regression applied to system identification for adaptive control systems. *Aiaa J.* **1963**, *1*, 2300–2306. [CrossRef]
8. Haykin, S.S.; Haykin, S.S. *Kalman Filtering and Neural Networks*; John Wiley & Sons, Ltd: Hoboken, NJ, USA, 2001; Volume 284. [CrossRef]
9. Cox, H. On the estimation of state variables and parameters for noisy dynamic systems. *IEEE Trans. Autom. Control* **1964**, *9*, 5–12. [CrossRef]
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*, 1–21. [CrossRef]
11. Aswal, N.; Sen, S.; Mevel, L. Switching Kalman filter for damage estimation in the presence of sensor faults. *Mech. Syst. Signal Process.* **2022**, *175*, 109116. [CrossRef]
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* **2020**, *30*, 6683–6700. [CrossRef]
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 Trans. Control Syst. Technol.* **2020**, *29*, 1087–1101. [CrossRef]
14. Dhanalakshmi, R.; Bhavani, N.; Raju, S.S.; Shaker Reddy, P.C.; Marvaluru, D.; Singh, D.P.; Batu, A. Onboard Pointing Error Detection and Estimation of Observation Satellite Data Using Extended Kalman Filter. *Comput. Intell. Neurosci.* **2022**, *2022*, 4340897. [CrossRef] [PubMed]
15. Huang, K.; Yuen, K.V.; Wang, L. Real-time simultaneous input-state-parameter estimation with modulated colored noise excitation. *Mech. Syst. Signal Process.* **2022**, *165*, 108378. [CrossRef]
16. Huang, K.; Yuen, K.V. Online dual-rate decentralized structural identification for wireless sensor networks. *Struct. Control Health Monit.* **2019**, *26*, e2453. [CrossRef]
17. Yuen, K.V.; Huang, K. Real-time substructural identification by boundary force modeling. *Struct. Control Health Monit.* **2018**, *25*, e2151. [CrossRef]
18. Kleyman, V.; Schaller, M.; Mordmuller, M.; Wilson, M.; Brinkmann, R.; Worthmann, K.; Muller, M.A. State and parameter estimation for retinal laser treatment. *arXiv* **2022**, arXiv:2203.12452.
19. Iglesias, C.F., Jr.; Xu, X.; Mehta, V.; Akassou, M.; Venereo-Sanchez, A.; Belacel, N.; Kamen, A.; Bolic, M. Monitoring the Recombinant Adeno-Associated Virus Production using Extended Kalman Filter. *Processes* **2022**, *10*, 2180. [CrossRef]
20. Yousefi-Darani, A.; Paquet-Durand, O.; Hitzmann, B. The Kalman filter for the supervision of cultivation processes. *Digital Twins* **2020**, *177*, 95–125.
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* **2020**, *8*, 1669. [CrossRef]
22. Song, H.; Hu, S. Open Problems in Applications of the Kalman Filtering Algorithm. In *Proceedings of the 2019 International Conference on Mathematics, Big Data Analysis and Simulation and Modelling (MBDASM 2019)*, Changsha, China, 30–31 August 2019; Atlantis Press: Amsterdam, The Netherlands, 2019; pp. 185–190.
23. Khodarahmi, M.; Maihami, V. A Review on Kalman Filter Models. *Arch. Comput. Methods Eng.* **2022**, *30*, 727–747. [CrossRef]
24. Nelson, L.; Stear, E. The simultaneous on-line estimation of parameters and states in linear systems. *IEEE Trans. Autom. Control* **1976**, *21*, 94–98.
25. Herwig, C.; Pörtner, R.; Möller, J. *Digital Twins: Tools and Concepts for Smart Biomanufacturing*; Springer Nature: Berlin/Heidelberg, Germany, 2021; Volume 176. [CrossRef]
26. Herwig, C.; Pörtner, R.; Möller, J. *Digital Twins: Applications to the Design and Optimization of Bioprocesses*; Springer Nature: Berlin/Heidelberg, Germany, 2021; Volume 177. [CrossRef]
27. Sinner, P.; Daume, S.; Herwig, C.; Kager, J. Usage of Digital Twins Along a Typical Process Development Cycle. In *Digital Twins: Tools and Concepts for Smart Biomanufacturing*; Herwig, C., Pörtner, R., Möller, J., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 71–96. [CrossRef]
28. Moser, A.; Appl, C.; Brüning, S.; Hass, V.C. Mechanistic mathematical models as a basis for digital twins. *Digit. Twins* **2020**, *176*, 133–180.

29. 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.* **2019**, *116*, 2540–2549. [[CrossRef](#)] [[PubMed](#)]
30. Fernandes-Platzgummer, A.; Badenes, S.M.; da Silva, C.L.; Cabral, J.M. Bioreactors for Stem Cell and Mammalian Cell Cultivation. *Bioprocess. Technol. Prod. Biopharm. Bioprod.* **2018**, *4*, 131–173.
31. Iglesias, C.F., Jr.; Ristovski, M.; Bolic, M.; Cuperlovic-Culf, M. rAAV Manufacturing: The Challenges of Soft Sensing during Upstream Processing. *Bioengineering* **2023**, *10*, 229. [[CrossRef](#)] [[PubMed](#)]
32. Gargalo, C.L.; Udugama, I.; Pontius, K.; Lopez, P.C.; Nielsen, R.F.; Hasanzadeh, A.; Mansouri, S.S.; Bayer, C.; Junicke, H.; Gernaey, K.V. 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. Ind. Microbiol. Biotechnol.* **2020**, *47*, 947–964. [[CrossRef](#)] [[PubMed](#)]
33. Udugama, A.; Öner, M.; Lopez, P.C.; Beenfeldt, C.; Bayer, C.; Huusom, J.K.; Gernaey, K.V.; Sin, G. Towards Digitalization in Bio-Manufacturing Operations: A Survey on Application of Big Data and Digital Twin Concepts in Denmark. *Front. Chem. Eng.* **2021**, *3*, 727152. [[CrossRef](#)]
34. Luo, Y.; Kurian, V.; Ogunnaike, B.A. Bioprocess systems analysis, modeling, estimation, and control. *Curr. Opin. Chem. Eng.* **2021**, *33*, 100705. [[CrossRef](#)]
35. Wang, K.; Li, Y.; Rizos, C. Practical approaches to Kalman filtering with time-correlated measurement errors. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 1669–1681. [[CrossRef](#)]
36. 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. *Open Autom. Control Syst. J.* **2009**, *2*, 69–77. [[CrossRef](#)]
37. Mariani, S.; Corigliano, A. Impact induced composite delamination: State and parameter identification via joint and dual extended Kalman filters. *Comput. Methods Appl. Mech. Eng.* **2005**, *194*, 5242–5272. [[CrossRef](#)]
38. Ljung, L.; Söderström, T. *Theory and Practice of Recursive Identification*; MIT Press: Cambridge, MA, USA, 1983; Volume 4, pp. 136–249.
39. Kyriakopoulos, S.; Ang, K.S.; Lakshmanan, M.; Huang, Z.; Yoon, S.; Gunawan, R.; Lee, D.Y. Kinetic modeling of mammalian cell culture bioprocessing: The quest to advance biomanufacturing. *Biotechnol. J.* **2018**, *13*, 1700229. [[CrossRef](#)] [[PubMed](#)]
40. 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.* **2021**, *33*, 100702. [[CrossRef](#)]
41. 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.* **2021**, *32*, 100691. [[CrossRef](#)]
42. Mears, L.; Stocks, S.M.; Albaek, M.O.; Sin, G.; Gernaey, K.V. Mechanistic fermentation models for process design, monitoring, and control. *Trends Biotechnol.* **2017**, *35*, 914–924. [[CrossRef](#)] [[PubMed](#)]
43. Reyes, S.J.; Durocher, Y.; Pham, P.L.; Henry, O. Modern Sensor Tools and Techniques for Monitoring, Controlling, and Improving Cell Culture Processes. *Processes* **2022**, *10*, 189. [[CrossRef](#)]
44. 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. Bioeng.* **2019**, *116*, 2919–2930. [[CrossRef](#)]
45. Kourtí, T. *Multivariate Statistical Process Control and Process Control, Using Latent Variables*; Elsevier: Amsterdam, The Netherlands, 2020. [[CrossRef](#)]
46. Brockwell, P. Time series analysis. *Encycl. Stat. Behav. Sci.* **2005**. [[CrossRef](#)]
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. Bioeng.* **2015**, *112*, 197–208. [[CrossRef](#)]
48. Assimakis, N.; Adam, M. Kalman filter Riccati equation for the prediction, estimation, and smoothing error covariance matrices. *Int. Sch. Res. Not.* **2013**, *2013*, 249594. [[CrossRef](#)]
49. Kulikova, M.V.; Kulikov, G.Y. Adaptive ODE solvers in extended Kalman filtering algorithms. *J. Comput. Appl. Math.* **2014**, *262*, 205–216. [[CrossRef](#)]
50. Särkkä, S.; Svensson, L. *Bayesian Filtering and Smoothing*; Cambridge University Press: Cambridge, UK, 2023; Volume 17, [[CrossRef](#)]
51. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012; Volume 1, pp. 631–660.
52. Haldar, A.; Al-hussein, A.A.A. *Recent Developments in Structural Health Monitoring and Assessment-Opportunities and Challenges: Bridges, Buildings and Other Infrastructures*; World Scientific: Singapore, 2022. [[CrossRef](#)]
53. Goudar, C.T. Computer programs for modeling mammalian cell batch and fed-batch cultures using logistic equations. *Cytotechnology* **2012**, *64*, 465–475. [[CrossRef](#)]
54. Kornecki, M.; Strube, J. Accelerating biologics manufacturing by upstream process modelling. *Processes* **2019**, *7*, 166. [[CrossRef](#)]
55. Narayanan, H.; Behle, L.; Luna, M.F.; Sokolov, M.; Guillén-Gosálbez, G.; Morbidelli, M.; Butté, A. Hybrid-EKF: Hybrid Model coupled with Extended Kalman Filter for real-time monitoring and control of mammalian cell culture. *Biotechnol. Bioeng.* **2020**, *117*, 2703–2714. [[CrossRef](#)] [[PubMed](#)]
56. Ueno, G.; Nakamura, N. Bayesian estimation of the observation-error covariance matrix in ensemble-based filters. *Q. J. R. Meteorol. Soc.* **2016**, *142*, 2055–2080. [[CrossRef](#)]
57. Michel, V.; Gramfort, A.; Varoquaux, G.; Eger, E.; Thirion, B. Total variation regularization for fMRI-based prediction of behavior. *IEEE Trans. Med Imaging* **2011**, *30*, 1328–1340. [[CrossRef](#)] [[PubMed](#)]

58. Jyothilekshmi, I.; Jayaprakash, N. Trends in monoclonal antibody production using various bioreactor systems. *J. Microbiol. Biotechnol.* **2021**, *31*, 349–357. [[CrossRef](#)]
59. Liu, Y.; Gunawan, R. Bioprocess optimization under uncertainty using ensemble modeling. *J. Biotechnol.* **2017**, *244*, 34–44. [[CrossRef](#)]
60. Bulcha, J.T.; Wang, Y.; Ma, H.; Tai, P.W.; Gao, G. Viral vector platforms within the gene therapy landscape. *Signal Transduct. Target. Ther.* **2021**, *6*, 53. [[CrossRef](#)]
61. Evangelidis, A.; Parker, D. Quantitative verification of Kalman filters. *Form. Asp. Comput.* **2021**, *33*, 669–693. [[CrossRef](#)]
62. Bar-Shalom, Y.; Li, X.R.; Kirubarajan, T. *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*; John Wiley & Sons: Hoboken, NJ, USA, 2001; Volume 5, pp. 199–265.
63. Simon, D. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*; John Wiley & Sons: Hoboken, NJ, USA, 2006. [[CrossRef](#)]
64. Maybeck, P.S. *Stochastic Models, Estimation, and Control*; Academic Press:Cambridge, MA, USA, 1982; Volume 3, pp. 126–132.
65. Boulkroune, B.; Geebelen, K.; Wan, J.; van Nunen, E. Auto-tuning extended Kalman filters to improve state estimation. In Proceedings of the 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, AK, USA, 4–7 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.