Supplementary Material: Comparison of Joint State and Parameter Estimation with NKEs for Bioprocess Monitoring

Annonymus

CONTENTS

I. BACKGROUND

TABLE I
GLOSSARY

JEKF	Joint estimation of states and parameters with Extended Kalman Filter
JUKF	Joint estimation of states and parameters with Unscented Kalman Filter
JCKF	Joint estimation of states and parameters with Cubature Kalman Filter
NKE	Nonlinear Kalman Estimator
UMM	Unstructured Mechanistic Model
MRDE	Matrix Ricatti Differential Equation
MSV	Measured State Variable
UP	Unshared parameter
SANTO	Specific initiAl coNdiTiOn
CD-EKF	Continuous-Discrete EKF
SD	Synthetic dataset
mAb	Monoclonal Antibody

A. Unstructured Mechanistic Model (UMM)

UMMs also known as Unstructured Mechanistic Kinetic Models, are pivotal in modeling the temporal progression of bioprocesses like the production of therapeutic monoclonal antibodies (mAbs), projected to generate USD 300 billion by 2025, and rAAV production, a leading viral vector technology for gene therapy [6, 10, 14, 17]. These models, grounded in fundamental principles, are key to understanding and simulating bioprocess dynamics at the macro-scale, such as cell density, viability, and nutrient/metabolite concentrations. Despite their critical role in digital twin (DT) development and soft sensors, the industrial application of UMMs is still nascent [9, 11, 5, 16]. In contrast to Structured Mechanistic Models (SMMs), which delve into the intracellular details of a homogeneous cell population and are more complex, requiring extensive expertise for development, UMMs are less detailed but more practical for dynamic control in common biomanufacturing bioreactors [10, 13]. SMMs are better suited for cell-line development, focusing on genomic-level alterations for desired process behaviors. However, the predictive capability of simple UMMs is limited, often failing to accurately estimate process states across different operating conditions [8]. To enhance their predictive accuracy, UMMs are frequently integrated with the Kalman filter and its nonlinear variants like the extended Kalman filter, effectively predicting unobserved states.

B. Nonlinear Kalman Estimators

The two-step recursive algorithmic process used by Nonlinear Kalman Estimators (EKF, UKF, RKF, and CKF) is summarized as follows [7, 15]:

- **Prediction step**: This step is where the state and error are propagated forward in time. In this step the predicted mean of state, $\hat{\mathbf{x}}_{k/k-1}$, and predicted error covariance matrix of state $\mathbf{P}_{k/k-1}$ are obtained using a process model (nonlinear system dynamics), a initial condition ($\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$) and $\mathbf{Q}$. See the Figure 1. Here is a description of the prediction step performed by each of the Nonlinear Kalman Estimators:
  - EKF: Linearizes the system's dynamics around the current state estimate to predict the next state [1].
  - RKF: Similar to EKF but incorporates mechanisms to handle model uncertainties and outliers. It adapts to model uncertainties and outliers, enhancing resilience to deviations from nominal assumptions [12].
  - UKF: Uses a set of deterministically chosen sample points (sigma points) to capture the mean and covariance of the state estimate and propagates these through the nonlinear system dynamics. UKF uses a deterministic sampling to approximate nonlinear transformations without linearization [2].
  - CKF: Employs cubature rules to compute the integral of the state transition function over the state distribution, effectively predicting the next state without linearization [3].
- **Update step**: This step is the same for all Nonlinear Kalman Estimators (EKF, UKF, RKF, and CKF). In this step the Predictions ($\hat{\mathbf{x}}_{k/k-1}$ and $\mathbf{P}_{k/k-1}$) are combined with the measured values ($\mathbf{y}_k$) to provide updated states and errors ($\hat{\mathbf{x}}_{k/k}$ and $\mathbf{P}_{k/k}$) using $\mathbf{R}$. This involves calculating the Kalman gain, determining how much the state prediction should be corrected based on the new measurement, and updating the error covariance to reflect the reduced uncertainty after incorporating the measurement. In addition, the updated state becomes the initial condition for the next prediction performed by the prediction step, $\hat{\mathbf{x}}_0 = \hat{\mathbf{x}}_{k/k}$ and $\mathbf{P}_0 = \mathbf{P}_{k/k}$. See the Figure 1.

C. JEKF, JUKF, and JCKF Overview

1) JEKF (Joint Extended Kalman Filter): JEKF is a Bayesian filter-based approach for joint estimation in nonlinear dynamical systems. It concatenates states x_i and parameters θ of a process model into a single joint state vector. The state variables vector $\boldsymbol{\psi}(t)$ in JEKF is extended as:

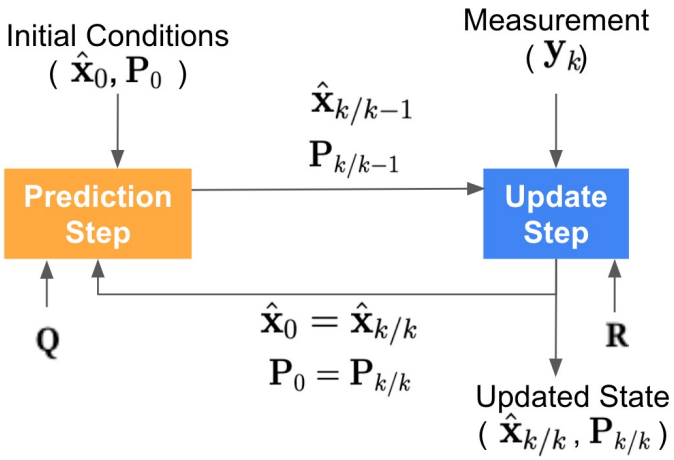

Fig. 1. Two-step recursive algorithmic process used by Nonlinear Kalman Estimators such as EKF, UKF, RKF and CKF. The state predictions are updated with new measurements, and this updated state becomes the basis for the next prediction.

$$\psi(t) = [x_1, x_2, ..., x_n, \theta_1, ..., \theta_n]^T. \tag{1}$$

In JEKF, the learning involves both states x_i and parameters θ_i of a discrete-time nonlinear system (e.g., UMM). It corrects system states and model parameters simultaneously based on observed noisy signals Z_k. JEKF is recognized for parameter evolution, where parameters are treated as random variables with noise added at each timestep:

$$\theta(t_k) = \theta(t_{k-1}) + \text{noise}, \tag{2}$$

This approach is efficient for updating process model parameters, especially when near optimal parameters for specific conditions. In JEKF, parameter estimation refers to this ongoing evolution of parameters.

2) JUKF (Joint Unscented Kalman Filter): JUKF extends the UKF to joint estimation scenarios. Unlike JEKF, JUKF does not linearize the process and measurement models but instead uses a deterministic sampling technique (the unscented transform) to capture the mean and covariance estimates. This makes JUKF more accurate in capturing the true state of a nonlinear system. JUKF also concatenates states and parameters into a single state vector and simultaneously estimates them using the unscented transform and Kalman filter equations.

3) JCKF (Joint Cubature Kalman Filter): Similar to JUKF, JCKF is designed for joint estimation of states and parameters in nonlinear systems. JCKF employs the cubature Kalman filter, which uses cubature rules to approximate the integrals in the state and covariance propagation. This approach avoids linearization errors and is computationally more efficient than JUKF. JCKF, like the other joint estimation methods, concatenates states and parameters into a single state vector for simultaneous estimation.

Each of these filters - JEKF, JUKF, and JCKF - has unique characteristics and is suitable for different nonlinear estimation scenarios in processes such as biomanufacturing.

D. 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:** Parameters unique to one term of an ODE and not shared with other ODEs in the UMM are typical for modeling product formation dynamics in biomanufacturing.
- **P and Q with Uncorrelated Elements:** Often, limited data leads to assuming error covariance matrices **P** (process error covariance) and **Q** (measurement error covariance) with uncorrelated elements, meaning they are diagonal with nonzero diagonal elements (noise variances) and zero off-diagonal elements. This results in two scenarios: 1) *Using **P** with uncorrelated elements for MRDE construction and as **P**(t=0) initial condition:* Here, the MRDE ODEs depend only on noise variances of $\mathrm{P}_{i,i}$ and $\mathrm{Q}_{i,i}$, and elements of Jacobian $\mathbf{J}_t^{\phi}$. 2) *Using **P** with correlated elements for MRDE construction and **P**(t=0) with uncorrelated elements as initial condition:* This can reduce time-invariant ODEs predicting state error covariance between two state variables.
- **ODEs of UMM with Weak Terms:** Weak terms in an ODE have less impact on the predicted state error covariance $\mathbf{P}(t_{k|k-1})$ compared to strong terms. The Jacobian $\mathbf{J}_t^{\phi}$ is more influenced by strong terms.

- **ODEs of UMM with Weak Variables:** Weak variables, present only in the first member of an ODE, do not contribute to the computation of predicted error covariance $\mathbf{P}(t_{k|k-1})$, as their first-order partial derivatives in Jacobian $\mathbf{J}_t^{\phi}$ are zero. In contrast, strong variables significantly influence $\mathbf{P}(t_{k|k-1})$ computation.
- **Only One Measured State Variable:** In some JEKF applications, only a single state variable is measured. This variable determines the column of predicted state error covariance $\mathbf{P}(t_{k|k-1})$ used for Kalman gain computation. If a row in this column is zero (no covariance between the measured and state variable), the Kalman gain for the state variable represented by that row cannot be computed.

These conditions emphasize the complexities and limitations in applying classical JEKF in biomanufacturing, where unique parameter characteristics and measurement constraints can impact its effectiveness [18].

E. SANTO approach

The SANTO approach is designed to address the failure case of the classical Joint Extended Kalman Filter (JEKF) in biomanufacturing scenarios. This approach specifically targets the initial condition of the Matrix Riccati Differential Equation (MRDE), which is the initial state error covariance matrix $\mathbf{P}_0 = \mathbf{P}(t = 0)$. In typical situations where $\mathbf{P}_0$ is composed of uncorrelated elements ($\mathrm{P}_{i,j} = 0$), certain initial conditions of time-invariant ODEs in the MRDE are zero, leading to zero solutions for these ODEs from t_{k-1} to t_k. In the presence of biomanufacturing conditions, the Kalman gain for the unshared parameter (K_{UP}) and the predicted state error covariance between the measured state variable and the unshared parameter ($P_{MSV,UP}(t_{k|k-1})$) are also zero, resulting in $K_{UP} = 0$ and $P_{MSV,UP}(t_{k|k-1}) = 0$. This indicates an unrealistic scenario where the prediction regarding the unshared parameter is considered perfect, without the need for measurement influence in the JEKF correction step.

To avoid this failure, the SANTO approach modifies the initial condition of MRDE. Instead of considering all off-diagonal elements of $\mathbf{P}(t = 0)$ as zero, a key off-diagonal element, specifically $P_{MSV,UP}(t = 0)$, is assigned an initial value different from zero ($P_{MSV,UP}(t = 0) \neq 0$). This value can be a positive or negative quantity, λ, reflecting the covariance between two variables. The choice of λ is crucial: it must be small enough to not significantly affect the filter's estimates but large enough to prevent the failure case.

Theorem of SANTO: The introduction of a positive quantity λ to the $P_{MSV,UP}(t = 0)$ in $\mathbf{P}(t = 0)$, initializing the MRDE with a specific initial condition, can prevent the Kalman gain from being zero throughout the JEKF execution, thereby averting the failure case [18].

II. UMM for Empirical Evaluation

A. UMM for monoclonal antibody (mAb) production

The ODE system 3 is a UMM used for Mab production [4]. This system represents the cell growth, uptake of substrates, metabolism, and production process with 16 parameters described in the Table II. It is important to point out that $QmAb$ denotes the specific mAb production rate, and is an example of unshared parameter. More details can be found in [4].

$$
\begin{aligned}
\frac{d\ X_V}{dt} &= (\mu - \mu_d)X_V \\
\frac{d\ X_t}{dt} &= \mu X_V - k_{lysis}(X_t - X_V) \\
\mu &= \mu_{max} . \frac{[GLC]}{K_{glc} + [GLC]} . \frac{[GLN]}{K_{gln} + [GLN]} . \frac{K_{Ilac}}{K_{Ilac} + [LAC]} . \frac{K_{Iamm}}{K_{Iamm} + [AMM]} \\
\mu_d &= \frac{\mu_{d,max}}{1 + (K_{d,amm} + [AMM])^2} \\
\frac{d\ [GLC]}{dt} &= -Q_{glc} X v \\
\frac{d\ [GLN]}{dt} &= -Q_{gln} X v - K_{d,gln}[GLN] \\
\frac{d\ [LAC]}{dt} &= Q_{lac} X v \\
\frac{d\ [AMM]}{dt} &= Q_{amm} X v + K_{d,gln}[GLN] \\
Q_{glc} X v &= \frac{\mu}{Y_{x,glc}} + m_{glc} \\
Q_{gln} X v &= \frac{\mu}{Y_{x,gln}} + m_{gln} = \frac{\mu}{Y_{x,gln}} + \frac{\alpha_2 [GLN]}{\alpha_2 + [GLN]} \\
Q_{lac} X v &= Y_{lac,glc} Q_{glc} \\
Q_{amm} X v &= Y_{amm,gln} Q_{gln} \\
\frac{d\ [mAb]}{dt} &= (2 - \gamma\mu) Q_{mAb} . X_V
\end{aligned}
\tag{3}
$$

Let's break down the components of this ODE system:

1) **Cell Growth and Death Dynamics:**
   - $\frac{dX_V}{dt} = (\mu - \mu_d)X_V$: This equation models the rate of change of viable cell density (X_V) over time. The growth rate (μ) minus the death rate (μ_d) is multiplied by the current viable cell density.
   - $\frac{dX_t}{dt} = \mu X_V - k_{lysis}(X_t - X_V)$: This equation describes the total cell density (X_t), considering both viable and non-viable cells. The rate of total cell density change is determined by the growth of viable cells and the lysis (breakdown) of cells, where k_{lysis} is the lysis rate constant.
2) Growth Rate (μ) and Death Rate (μ_d):
   - μ: Defined as a function of substrate concentrations ([GLC] for glucose and [GLN] for glutamine) and inhibitors ([LAC] for lactate and [AMM] for ammonium). This function reflects how cell growth rate is influenced by the availability of nutrients and the presence of metabolic byproducts.
   - μ_d: The death rate, modeled as a function of the ammonium concentration, with $\mu_{d,max}$ representing the maximum death rate and $K_{d,amm}$ as a constant.
3) Substrate Consumption and Metabolite Production:
   - The following set of equations ($\frac{d[GLC]}{dt}$, $\frac{d[GLN]}{dt}$, $\frac{d[LAC]}{dt}$, $\frac{d[AMM]}{dt}$) represent the rates of change in concentrations of glucose, glutamine, lactate, and ammonium, respectively. These are key substrates and metabolites in the cell culture. The terms $Q_{glc}, Q_{gln}, Q_{lac}, Q_{amm}$ denote specific consumption/production rates of these components, and $K_{d,gln}$ is the degradation constant for glutamine.
4) Balancing Equations for Substrate Consumption and Product Formation:
   - The equations relating $Q_{glc}Xv, Q_{gln}Xv, Q_{lac}Xv, Q_{amm}Xv$ establish relationships between growth rate, substrate consumption, and metabolite production rates. These are based on yield coefficients ($Y_{x,glc}, Y_{x,gln}, Y_{lac,glc}, Y_{amm,gln}$) and maintenance coefficients ($m_{glc}, m_{gln}, \alpha_2$).
5) Monoclonal Antibody (mAb) Production:
   - $\frac{d[mAb]}{dt} = (2 - \gamma\mu)Q_{mAb}.X_V$: This equation models the rate of mAb production. The specific mAb production rate (Q_{mAb}) is multiplied by the viable cell density and a factor considering the growth rate, where γ is a constant.

The model's strength lies in its ability to capture the interplay between cell growth, nutrient consumption, metabolite accumulation, and product formation, which are crucial for optimizing and monitoring biomanufacturing processes. The parameter Q_{mAb}, representing the specific mAb production rate, is particularly notable as it's an unshared parameter, meaning its value is unique to this process and not shared with other models or components within this system.

III. Empirical Evaluation - Extension

A. Synthetic dataset development - mAb production

The Synthetic dataset (SD) is composed of three runs (A-SD, B-SD, and C-SD). The runs have different samples regarding the state variables Xv, GLC, GLN, LAC, AMM, and mAb and were generated using the UMM II-A with three set of different parameters (Table II), and the same initial condition (Table III).

TABLE II

INITIAL PARAMETERS USED IN UMM CASE II-A TO GENERATE THE RUNS A-SD, B-SD AND C-SD OF SYNTHETIC DATASET (SD).

Parameter	**Name**	**run A-SD**	**run B-SD**	**run C-SD**
$\mu_{max}(h^-)$	Maximum growth rate	5.8×10^{-2}	7.5×10^{-2}	5×10^{-2}
$k_{glc}(mM)$	Monod constant glucose	7.5×10^{-1}	7.5×10^{-1}	7.5×10^{-1}
$k_{gln}(mM)$	Monod constant glutamine	7.5×10^{-2}	7.5×10^{-2}	7.5×10^{-2}
$k_{Ilac}(mM)$	Monod constant lactate for inhibition	1.72×10^{2}	1.72×10^{2}	1.72×10^{2}
$k_{Iamm}(mM)$	Monod constant ammonium for inhibition	2.85×10^{1}	2.85×10^{1}	2.85×10^{1}
$\mu_{d,max}(h^-)$	Maximum death rate	3.0×10^{-2}	3.0×10^{-2}	3.0×10^{-2}
$K_{d,amm}(mM)$	Monod constant ammonium for death	1.76	1.76	1.76
$K_{lysis}(h^-)$	Breakdown of cell membranes	5.51×10^{-2}	5.51×10^{-2}	5.51×10^{-2}
$Y_{X,glc}(cells\ mmol^-)$	Yield coefficient cell conc./glucose	1.06×10^{8}	1.06×10^{8}	1.06×10^{8}
$m_{glc}(mmol/cells\ h)$	Glucose maintenance coefficient	4.85×10^{-14}	4.85×10^{-14}	4.85×10^{-14}
$Y_{X,gln}(cells/mmmol)$	Yield coefficient cell conc./glutamine	5.57×10^{8}	5.57×10^{8}	5.57×10^{8}
$\alpha_1(mmols\ cells^-\ h^-)$	Coefficient for m_{gln}	3.40×10^{-13}	3.40×10^{-13}	3.40×10^{-13}
$\alpha_2(mM)$	Coefficient for m_{gln}	4.0	4.0	4.0
$k_{d,gln}(h^-)$	Monod constant glutamine for death	9.6×10^{-3}	9.6×10^{-3}	9.6×10^{-3}
$Y_{lac/glc}(1)$	Yield coefficient lactate/glucose	1.4	1.4	1.4
$Y_{amm/gln}(1)$	Yield coefficient ammonium/glutamine	4.27×10^{-1}	4.27×10^{-1}	4.27×10^{-1}
γ	constant parameter	4.27×10^{-1}	4.27×10^{-1}	4.27×10^{-1}
$Q_{mAb}(mg\ cells^-\ h^-)$	mAb specific production rate	7.21×10^{-9}	9.21×10^{-9}	4.21×10^{-9}

TABLE III

INITIAL CONDITIONS OF STATE VARIABLES OF UMM CASE II-A.

State Variable	**Name**	**Value**
Xv	Viable cells density	2×10^8 c/mL
Xt	total cells density	2×10^8 c/mL
GLC	Glucose	29.1 mM
GLN	Glutamine	4.9 mM
LAC	Lactate	0 mM
AMM	Ammonium	0.31 mM
mAb	Monoclonal Antibody (titer)	80.6 mg/L
QmAb	Specific production rate of mAb	$7.21 \times 10^{-9}\ mg\ cells^{-1}h^{-1}$

B. NKEs design to address RQ1 and RQ2

The process model (based on UMM case II-A) and joint state variable vector used by NKEs (JEKF-Classic, JUKF-Classic, JCKF-Classic, JEKF-SANTO, JUKF-SANTO and JCKF-SANTO) are the following:

$$\boldsymbol{\psi}(t)_{case4} = [X_V, X_t, GLC, GLN, LAC, AMM, mAb, QmAb]^T, \tag{4}$$

and

$$\frac{d}{dt}\begin{bmatrix} X_V \\ X_t \\ GLC \\ GLN \\ LAC \\ AMM \\ mAb \\ QmAb \end{bmatrix} = \begin{bmatrix} f_{X_V} \\ f_{GLC} \\ f_{GLN} \\ f_{LAC} \\ f_{AMM} \\ f_{mAb} \\ 0 \end{bmatrix} + \boldsymbol{\omega}(t). \tag{5}$$

The **P**(t=0) that were used by the NKEs with run B of Synthetic Dataset are in Tables IV, V, and VI. Furthermore, the **P**(t=0) that were used by the NKEs with run C of Synthetic Dataset are in Tables IX, X, and XI. It is important to point out that in case of run B-SD, we applied the SANTO approach by adding a small positive quantity to $\mathrm{P_{X_v,QmAb}}$ $(\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$, and in case of run C-SD we added a small negative positive quantity to $\mathrm{P_{X_v,QmAb}}$ $(\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$, see Tables IV, V, VI, IX, X, and XI.

The **R** and **Q** used by the NSEs (for runs B of Synthetic Dataset) are presented in Tables VIII and VIII. Furthermore, The **R** and **Q** used by the NSEs (for runs C of Synthetic Dataset) are presented in Tables XIII and XIII.

It is important point out that all NSEs used a **R**, **P**(t=0), and **Q** that were obtained by by trial and error until achieve positive results in the Normalized Innovations Squared Chi-square Test.

TABLE IV
STANDARD INITIAL STATE ERROR COVARIANCE MATRIX (STANDARD **P**(T=0)) FOR JEKF-CLASSIC, AND JEKF-SANTO WITH RUN B OF SYNTHETIC DATASET.

Parameter	**Name**	$\mathbf{P}_{i,i}$ for JEKF-Classic	$\mathbf{P}_{i,i}$ for JEKF-SANTO
$\mathrm{P_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00	0.00
$\mathrm{P_{X_t,X_t}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00	0.00
$\mathrm{P_{GLC,GLC}}$ $(\mathrm{mM^2})$	Glucose	0.00	0.00
$\mathrm{P_{GLN,GLN}}$ $(\mathrm{mM^2})$	Glutamine	0.00	0.00
$\mathrm{P_{LAC,LAC}}$ $(\mathrm{mM^2})$	Lactate	0.00	0.00
$\mathrm{P_{AMM,AMM}}$ $(\mathrm{mM^2})$	Ammonium	0.00	0.00
$\mathrm{P_{mAb,mAb}}$ $(\mathrm{mg/L})^2$	Monoclonal Antibody (titer)	0.00	0.00
$\mathrm{P_{QmAb,QmAb}}$ $(\mathrm{g\ cells^{-1}h^{-1}})^2$	Specific production rate of mAb	3.9e-18	3.9e-18
$\mathrm{P_{X_v,QmAb}}$ $(\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$	Initial $Cov(X_v, QmAb)$	0.0	0.8404

TABLE V
STANDARD INITIAL STATE ERROR COVARIANCE MATRIX (STANDARD **P**(T=0)) FOR JUKF-CLASSIC, AND JUKF-SANTO WITH RUN B OF SYNTHETIC DATASET.

Parameter	**Name**	$\mathbf{P}_{i,i}$ for JCKF-Classic	$\mathbf{P}_{i,i}$ for JCKF-SANTO
$\mathrm{P_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00001	0.00001
$\mathrm{P_{X_t,X_t}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00001	0.00001
$\mathrm{P_{GLC,GLC}}$ $(\mathrm{mM^2})$	Glucose	0.00001	0.00001
$\mathrm{P_{GLN,GLN}}$ $(\mathrm{mM^2})$	Glutamine	0.00001	0.00001
$\mathrm{P_{LAC,LAC}}$ $(\mathrm{mM^2})$	Lactate	0.00001	0.00001
$\mathrm{P_{AMM,AMM}}$ $(\mathrm{mM^2})$	Ammonium	0.00001	0.00001
$\mathrm{P_{mAb,mAb}}$ $(\mathrm{mg/L})^2$	Monoclonal Antibody (titer)	0.00001	0.00001
$\mathrm{P_{QmAb,QmAb}}$ $(\mathrm{g\ cells^{-1}h^{-1}})^2$	Specific production rate of mAb	0.01	10000.01
$\mathrm{P_{X_v,QmAb}}$ $(\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$	Initial $Cov(X_v, QmAb)$	0.0	3.1e-1

TABLE VI
STANDARD INITIAL STATE ERROR COVARIANCE MATRIX (STANDARD **P**(T=0)) FOR JCKF-CLASSIC, AND JCKF-SANTO WITH RUN B OF SYNTHETIC DATASET.

Parameter	**Name**	$\mathbf{P}_{i,i}$ for JCKF-Classic	$\mathbf{P}_{i,i}$ for JCKF-SANTO
$\mathrm{P_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00001	0.00001
$\mathrm{P_{X_t,X_t}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00001	0.00001
$\mathrm{P_{GLC,GLC}}$ $(\mathrm{mM^2})$	Glucose	0.00001	0.00001
$\mathrm{P_{GLN,GLN}}$ $(\mathrm{mM^2})$	Glutamine	0.00001	0.00001
$\mathrm{P_{LAC,LAC}}$ $(\mathrm{mM^2})$	Lactate	0.00001	0.00001
$\mathrm{P_{AMM,AMM}}$ $(\mathrm{mM^2})$	Ammonium	0.00001	0.00001
$\mathrm{P_{mAb,mAb}}$ $(\mathrm{mg/L})^2$	Monoclonal Antibody (titer)	0.00001	0.00001
$\mathrm{P_{QmAb,QmAb}}$ $(\mathrm{g\ cells^{-1}h^{-1}})^2$	Specific production rate of mAb	0.01	10000.01
$\mathrm{P_{X_v,QmAb}}$ $(\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$	Initial $Cov(X_v, QmAb)$	0.0	2.9e-1

TABLE VII
MEASUREMENT NOISE VARIANCE **R** AND ERROR COVARIANCE MATRIX OF PROCESS MODEL (**Q**) FOR THE JEKF-CLASSIC, JUKF-CLASSIC AND JCKF-CLASSIC WITH RUN B OF SYNTHETIC DATASET.

Parameter	**Name**	JEKF-Classic	JUKF-Classic	JCKF-Classic
$\mathrm{R^2}$ $(\mathrm{c^2/mL^2})$	Viable cells MNV [1]	$(20\times\ 10^7)^2$	$(20\times\ 10^7)^2$	$(20\times\ 10^7)^2$
$\mathrm{Q_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	Viable cells PNV [2]	$(20\times\ 10^6)^2$	$(20\times\ 10^6)^2$	$(20\times\ 10^6)^2$
$\mathrm{Q_{X_t,X_t}}$ $(\mathrm{c^2/mL^2})$	Viable cells PNV [2]	0.001	0.00001	0.00001
$\mathrm{Q_{GLC,GLC}}$ $(\mathrm{mM^2})$	Glucose PNV	0.001	0.00001	0.00001
$\mathrm{Q_{GLN,GLN}}$ $\mathrm{mM^2}$	Glutamine PNV	0.001	0.00001	0.00001
$\mathrm{Q_{LAC,LAC}}$ $(\mathrm{mM^2})$	Lactate PNV	0.001	0.00001	0.00001
$\mathrm{Q_{AMM,AMM}}$ $(\mathrm{mM^2})$	Ammonium PNV	0.001	0.00001	0.00001
$\mathrm{Q_{mAb,mAb}}$ $(\mathrm{VG^2/mL^2})$	Monoclonal Antibody (titer) PNV	0.001	00001	0.001
$\mathrm{Q_{QmAb,QmAb}}$ $(\mathrm{h^{-2}})$	Specific production rate of mAb	$1\times\ 10^{-18}$	0.01	0.1

[1] MNV—measurement noise value; [2] PNV—process noise value.

TABLE VIII
MEASUREMENT NOISE VARIANCE **R** AND ERROR COVARIANCE MATRIX OF PROCESS MODEL (**Q**) FOR THE JEKF-SANTO, JUKF-SANTO AND JCKF-SANTO WITH RUN B OF SYNTHETIC DATASET.

Parameter	**Name**	JEKF-SANTO	JUKF-SANTO	JCKF-SANTO
$\mathrm{R^2}$ $(\mathrm{c^2/mL^2})$	Viable cells MNV [1]	$(20\times\ 10^7)^2$	$(20\times\ 10^7)^2$	$(20\times\ 10^7)^2$
$\mathrm{Q_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	Viable cells PNV [2]	$(80\times\ 10^6)^2$	$(20\times\ 10^6)^2$	$(20\times\ 10^6)^2$
$\mathrm{Q_{X_t,X_t}}$ $(\mathrm{c^2/mL^2})$	Viable cells PNV [2]	0.001	0.00001	0.00001
$\mathrm{Q_{GLC,GLC}}$ $(\mathrm{mM^2})$	Glucose PNV	0.001	0.00001	0.00001
$\mathrm{Q_{GLN,GLN}}$ $\mathrm{mM^2}$	Glutamine PNV	0.001	0.00001	0.00001
$\mathrm{Q_{LAC,LAC}}$ $(\mathrm{mM^2})$	Lactate PNV	0.001	0.00001	0.00001
$\mathrm{Q_{AMM,AMM}}$ $(\mathrm{mM^2})$	Ammonium PNV	0.001	0.00001	0.00001
$\mathrm{Q_{mAb,mAb}}$ $(\mathrm{VG^2/mL^2})$	Monoclonal Antibody (titer) PNV	0.001	0.001	0.001
$\mathrm{Q_{QmAb,QmAb}}$ $(\mathrm{h^{-2}})$	Specific production rate of mAb	0.001	0.001	0.001

[1] MNV—measurement noise value; [2] PNV—process noise value.

TABLE IX
STANDARD INITIAL STATE ERROR COVARIANCE MATRIX (STANDARD **P**(T=0)) FOR JEKF-CLASSIC, AND JEKF-SANTO WITH RUN C OF SYNTHETIC DATASET.

Parameter	**Name**	$\mathbf{P}_{i,i}$ for JEKF-Classic	$\mathbf{P}_{i,i}$ for JEKF-SANTO
$\mathrm{P_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00	0.00
$\mathrm{P_{X_t,X_t}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00	0.00
$\mathrm{P_{GLC,GLC}}$ $(\mathrm{mM^2})$	Glucose	0.00	0.00
$\mathrm{P_{GLN,GLN}}$ $(\mathrm{mM^2})$	Glutamine	0.00	0.00
$\mathrm{P_{LAC,LAC}}$ $(\mathrm{mM^2})$	Lactate	0.00	0.00
$\mathrm{P_{AMM,AMM}}$ $(\mathrm{mM^2})$	Ammonium	0.00	0.00
$\mathrm{P_{mAb,mAb}}$ $(\mathrm{mg/L})^2$	Monoclonal Antibody (titer)	0.00	0.00
$\mathrm{P_{QmAb,QmAb}}$ $(\mathrm{g\ cells^{-1}h^{-1}})^2$	Specific production rate of mAb	8.9e-18	8.9e-18
$\mathrm{P_{X_v,QmAb}}$ $(\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$	Initial $Cov(X_v, QmAb)$	0.0	-0.1805

TABLE X
STANDARD INITIAL STATE ERROR COVARIANCE MATRIX (STANDARD **P**(T=0)) FOR JUKF-CLASSIC, AND JUKF-SANTO WITH RUN C OF SYNTHETIC DATASET.

Parameter	**Name**	$\mathbf{P}_{i,i}$ for JUKF-Classic	$\mathbf{P}_{i,i}$ for JUKF-SANTO
$\mathrm{P_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00001	0.00001
$\mathrm{P_{X_t,X_t}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00001	0.00001
$\mathrm{P_{GLC,GLC}}$ $(\mathrm{mM^2})$	Glucose	0.00001	0.00001
$\mathrm{P_{GLN,GLN}}$ $(\mathrm{mM^2})$	Glutamine	0.00001	0.00001
$\mathrm{P_{LAC,LAC}}$ $(\mathrm{mM^2})$	Lactate	0.00001	0.00001
$\mathrm{P_{AMM,AMM}}$ $(\mathrm{mM^2})$	Ammonium	0.00001	0.00001
$\mathrm{P_{mAb,mAb}}$ $(\mathrm{mg/L})^2$	Monoclonal Antibody (titer)	0.00001	0.00001
$\mathrm{P_{QmAb,QmAb}}$ $(\mathrm{g\ cells^{-1}h^{-1}})^2$	Specific production rate of mAb	0.01	10000.01
$\mathrm{P_{X_v,QmAb}}$ $(\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$	Initial $Cov(X_v, QmAb)$	0.0	-1.518e-1

TABLE XI
STANDARD INITIAL STATE ERROR COVARIANCE MATRIX (STANDARD **P**(T=0)) FOR JCKF-CLASSIC, AND JCKF-SANTO WITH RUN C OF SYNTHETIC DATASET.

Parameter	**Name**	$\mathbf{P}_{i,i}$ for JCKF-Classic	$\mathbf{P}_{i,i}$ for JCKF-SANTO
$\mathrm{P_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00001	0.00001
$\mathrm{P_{X_t,X_t}}$ $(\mathrm{c^2/mL^2})$	Viable cells	0.00001	0.00001
$\mathrm{P_{GLC,GLC}}$ $(\mathrm{mM^2})$	Glucose	0.00001	0.00001
$\mathrm{P_{GLN,GLN}}$ $(\mathrm{mM^2})$	Glutamine	0.00001	0.00001
$\mathrm{P_{LAC,LAC}}$ $(\mathrm{mM^2})$	Lactate	0.00001	0.00001
$\mathrm{P_{AMM,AMM}}$ $(\mathrm{mM^2})$	Ammonium	0.00001	0.00001
$\mathrm{P_{mAb,mAb}}$ $(\mathrm{mg/L})^2$	Monoclonal Antibody (titer)	0.00001	0.00001
$\mathrm{P_{QmAb,QmAb}}$ $(\mathrm{g\ cells^{-1}h^{-1}})^2$	Specific production rate of mAb	0.01	10000.01
$\mathrm{P_{X_v,QmAb}}$ $(\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$	Initial $Cov(X_v, QmAb)$	0.0	-1.581e-1

TABLE XII
MEASUREMENT NOISE VARIANCE **R** AND ERROR COVARIANCE MATRIX OF PROCESS MODEL (**Q**) FOR THE JEKF-CLASSIC, JUKF-CLASSIC AND JCKF-CLASSIC WITH RUN C OF SYNTHETIC DATASET.

Parameter	**Name**	JEKF-Classic	JUKF-Classic	JCKF-Classic
$\mathrm{R^2}$ $(\mathrm{c^2/mL^2})$	Viable cells MNV [1]	$(20\times\ 10^7)^2$	$(20\times\ 10^7)^2$	$(20\times\ 10^7)^2$
$\mathrm{Q_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	Viable cells PNV [2]	$(20\times\ 10^6)^2$	$(20\times\ 10^6)^2$	$(20\times\ 10^6)^2$
$\mathrm{Q_{X_t,X_t}}$ $(\mathrm{c^2/mL^2})$	Viable cells PNV [2]	0.001	0.00001	0.00001
$\mathrm{Q_{GLC,GLC}}$ $(\mathrm{mM^2})$	Glucose PNV	0.001	0.00001	0.00001
$\mathrm{Q_{GLN,GLN}}$ $\mathrm{mM^2}$	Glutamine PNV	0.001	0.00001	0.00001
$\mathrm{Q_{LAC,LAC}}$ $(\mathrm{mM^2})$	Lactate PNV	0.001	0.00001	0.00001
$\mathrm{Q_{AMM,AMM}}$ $(\mathrm{mM^2})$	Ammonium PNV	0.001	0.00001	0.00001
$\mathrm{Q_{mAb,mAb}}$ $(\mathrm{VG^2/mL^2})$	Monoclonal Antibody (titer) PNV	0.001	001	0.001
$\mathrm{Q_{QmAb,QmAb}}$ $(\mathrm{h^{-2}})$	Specific production rate of mAb	0.001	0.01	0.1

[1] MNV—measurement noise value; [2] PNV—process noise value.

TABLE XIII
MEASUREMENT NOISE VARIANCE **R** AND ERROR COVARIANCE MATRIX OF PROCESS MODEL (**Q**) FOR THE JEKF-SANTO, JUKF-SANTO AND JCKF-SANTO WITH RUN C OF SYNTHETIC DATASET.

Parameter	**Name**	JEKF-SANTO	JUKF-SANTO	JCKF-SANTO
$\mathrm{R^2}$ $(\mathrm{c^2/mL^2})$	Viable cells MNV [1]	$(20\times\ 10^7)^2$	$(20\times\ 10^7)^2$	$(20\times\ 10^7)^2$
$\mathrm{Q_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	Viable cells PNV [2]	$(20\times\ 10^6)^2$	$(20\times\ 10^6)^2$	$(20\times\ 10^6)^2$
$\mathrm{Q_{X_t,X_t}}$ $(\mathrm{c^2/mL^2})$	Viable cells PNV [2]	0.001	0.00001	0.00001
$\mathrm{Q_{GLC,GLC}}$ $(\mathrm{mM^2})$	Glucose PNV	0.001	0.00001	0.00001
$\mathrm{Q_{GLN,GLN}}$ $\mathrm{mM^2}$	Glutamine PNV	0.001	0.00001	0.00001
$\mathrm{Q_{LAC,LAC}}$ $(\mathrm{mM^2})$	Lactate PNV	0.001	0.00001	0.00001
$\mathrm{Q_{AMM,AMM}}$ $(\mathrm{mM^2})$	Ammonium PNV	0.001	0.00001	0.00001
$\mathrm{Q_{mAb,mAb}}$ $(\mathrm{VG^2/mL^2})$	Monoclonal Antibody (titer) PNV	0.001	0.001	0.001
$\mathrm{Q_{QmAb,QmAb}}$ $(\mathrm{h^{-2}})$	Specific production rate of mAb	0.001	0.001	0.001

[1] MNV—measurement noise value; [2] PNV—process noise value.

REFERENCES

[1] Simon J Julier and Jeffrey K Uhlmann. "New extension of the Kalman filter to nonlinear systems". In: *Signal processing, sensor fusion, and target recognition VI*. Vol. 3068. Spie. 1997, pp. 182–193.

[2] Eric A Wan and Rudolph Van Der Merwe. "The unscented Kalman filter for nonlinear estimation". In: *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*. Ieee. 2000, pp. 153–158.

[3] Ienkaran Arasaratnam and Simon Haykin. "Cubature kalman filters". In: *IEEE Transactions on automatic control* 54.6 (2009), pp. 1254–1269.

[4] Yang Liu and Rudiyanto Gunawan. "Bioprocess optimization under uncertainty using ensemble modeling". In: *Journal of biotechnology* 244 (2017), pp. 34–44.

[5] Lisa Mears et al. "Mechanistic fermentation models for process design, monitoring, and control". In: *Trends in biotechnology* 35.10 (2017), pp. 914–924.

[6] Sarantos Kyriakopoulos et al. "Kinetic modeling of mammalian cell culture bioprocessing: the quest to advance biomanufacturing". In: *Biotechnology Journal* 13.3 (2018), p. 1700229.

[7] Alper Akca and M Önder Efe. "Multiple model Kalman and Particle filters and applications: A survey". In: *IFAC-PapersOnLine* 52.3 (2019), pp. 73–78.

[8] Dongda Zhang et al. "Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization". In: *Biotechnology and bioengineering* 116.11 (2019), pp. 2919–2930.

[9] André Moser et al. "Mechanistic mathematical models as a basis for digital twins". In: *Digital Twins* (2020), pp. 133–180.

[10] Yu Luo, Varghese Kurian, and Babatunde A Ogunnaike. "Bioprocess systems analysis, modeling, estimation, and control". In: *Current Opinion in Chemical Engineering* 33 (2021), p. 100705.

[11] Seo-Young Park et al. "Bioprocess digital twins of mammalian cell culture for advanced biomanufacturing". In: *Current Opinion in Chemical Engineering* 33 (2021), p. 100702.

[12] Kaio DT Rocha and Marco H Terra. "Robust Kalman filter for systems subject to parametric uncertainties". In: *Systems & Control Letters* 157 (2021), p. 105034.

[13] Apostolos Tsopanoglou and Ioscani Jiménez del Val. "Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses". In: *Current Opinion in Chemical Engineering* 32 (2021), p. 100691.

[14] Cristovão Freitas Iglesias Jr et al. "Monitoring the Recombinant Adeno-Associated Virus Production using Extended Kalman Filter". In: *Processes* 10.11 (2022), p. 2180.

[15] Masoud Khodarahmi and Vafa Maihami. "A Review on Kalman Filter Models". In: *Archives of Computational Methods in Engineering* (2022), pp. 1–21.

[16] Sebastian Juan Reyes et al. "Modern Sensor Tools and Techniques for Monitoring, Controlling, and Improving Cell Culture Processes". In: *Processes* 10.2 (2022), p. 189.

[17] Cristovão Freitas Iglesias Jr et al. "rAAV Manufacturing: The Challenges of Soft Sensing during Upstream Processing". In: *Bioengineering* 10.2 (2023), p. 229.

[18] Cristovao Iglesias Jr. *How NOT To Make the Joint Extended Kalman Filter Fail with Unstructured Mechanistic Models*. Sensors - MDPI, 2024.