Supplementary Information for How NOT to Make the Joint Extended Kalman Filter Fail with Unstructured Mechanistic Models

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

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

Contents

S1 Unstructured Mechanistic Models cases in biomanufacturing

S1.1 Case 1

The ODE system 1 represents the dynamics of mammalian cell growth $Xv(t)$, nutrients $N(t)$ and metabolite/production formation $MP(t)$ in a general form[1,2].

$$\begin{aligned}
\frac{dX_V(t)}{dt} &= \mu_{X_v} X_V(t) \\
\frac{dN(t)}{dt} &= -\mu_N X_V(t) \\
\frac{dMP(t)}{dt} &= \mu_{mp} X_V(t)
\end{aligned} \tag{1}$$

S1.2 Case 2

The ODE system 2 is a UMM used for Mab production[3]. This system represents the cell growth, cell dead, uptake of substrates, metabolism, and production process with 35 parameters. More details can be found in[3].

$$\begin{aligned}
\frac{d\,fgr}{dt} &= -k_{11}\frac{fgr}{1+([GLC]/K_{12})} \\
\frac{d\,X_V}{dt} &= \mu_{max}.fgr.X_V.\frac{[GLC]}{K_{21}+[GLC]}.\frac{[GLN]}{K_{22}+[GLN]}.\frac{1}{[LAC]/K_{24}+1}.\frac{1}{[AMM]/K_{23}+1} - k_d.(1-fgr).X_V.(\frac{1}{1+([AMM]/k_{25})^n}+\frac{k_{26}}{[GLC]}) \\
\frac{d\,X_d}{dt} &= k_d.(1-fgr).X_V.(\frac{1}{1+([AMM]/k_{25})^n}+\frac{k_{26}}{[GLC]}) - k_{lys}.X_d \\
\frac{d\,[GLC]}{dt} &= -(\frac{K_{31}.[GLC].[GLN]}{(k_{32}+[GLC])(k_{33}+[GLN])}+\frac{K_{34}.[GLC]}{(k_{35}+[GLC])}).X_V - k_{36}X_V \\
\frac{d\,[GLN]}{dt} &= -k_{41}.X_V.(\frac{[GLC].[GLN]}{(k_{42}+[GLC])(k_{43}+[GLN])}) \\
\frac{d\,[LAC]}{dt} &= -k_{51}.X_V.\frac{d\,[GLC]}{dt} \\
\frac{d\,[ASN]}{dt} &= -k_{61}.X_V.\frac{[ASN]}{(k_{62}+[ASN])} \\
\frac{d\,[ASP]}{dt} &= X_V.(\frac{K_{61}[ASN]}{K_{62}+[ASN]}-\frac{K_{63}[ASP]}{K_{64}+[ASP]}+\frac{K_{65}.[GLC].[GLN]}{(K_{66}+[GLC])(K_{67}+[GLN])}) \\
\frac{d\,[ALA]}{dt} &= X_V.(\frac{K_{65}.[GLC].[GLN]}{(K_{66}+[GLC])(K_{67}+[GLN])}-\frac{K_{68}[ALA]}{K_{69}+[ALA]}) \\
\frac{d\,[AMM]}{dt} &= -k_{71}.\frac{d\,[GLN]}{dt}+k_{72}.X_V.(\frac{K_{61}[ASN]}{K_{62}+[ASN]}+\frac{K_{63}[ASP]}{K_{64}+[ASP]}+\frac{K_{68}.[ALA]}{(K_{69}+[ALA])}) \\
\frac{d\,[mAb]}{dt} &= X_V(k_{81}+k_{82}.[GLC])
\end{aligned} \tag{2}$$

S1.3 Case 3

The ODE system 3 represents the Michaelis-Menten model for enzymatic reactions[4].

$$
\begin{aligned}
\frac{dE(t)}{dt} &= -k_1 E(t)S(t) + k_2 ES(t) + k_3 ES(t) \\
\frac{dS(t)}{dt} &= -k_1 E(t)S(t) + k_2 ES(t) \\
\frac{dES(t)}{dt} &= k_1 E(t)S(t) - k_2 ES(t) - k_3 ES(t) \\
\frac{dP(t)}{dt} &= k_3 ES(t)
\end{aligned}
\tag{3}
$$

S1.4 Case 4

The ODE system 4 is a UMM used for Mab production[5]. This system represents the cell growth, uptake of substrates, metabolism, and production process with 17 parameters as described in the Table S1. More details can be found in[5].

$$
\begin{aligned}
\frac{d\,X_V}{dt} &= (\mu - \mu_d)X_V \\
\mu &= \mu_{max} \cdot \frac{[GLC]}{K_{glc} + [GLC]} \cdot \frac{[GLN]}{K_{gln} + [GLN]} \cdot \frac{K_{Ilac}}{K_{Ilac} + [LAC]} \cdot \frac{K_{Iamm}}{K_{Iamm} + [AMM]} \\
\mu_d &= k_d \cdot \frac{[LAC]}{K_{Dlac} + [LAC]} \cdot \frac{[AMM]}{K_{Damm} + [AMM]} \\
\frac{d\,[GLC]}{dt} &= -\left(\frac{\mu - \mu_d}{Y_{X_V/glc}} + m_{glc}\right).X_V \\
\frac{d\,[LAC]}{dt} &= Y_{lac/glc} \cdot \left(\frac{\mu - \mu_d}{Y_{X_V/glc}}\right).X_V \\
\frac{d\,[GLN]}{dt} &= -\left(\frac{\mu - \mu_d}{Y_{X_V/gln}} + m_{gln}\right).X_V \\
m_{gln} &= \frac{a_1.[GLN]}{a_2 + [GLN]} \\
\frac{d\,[AMM]}{dt} &= Y_{amm/gln} \cdot \left(\frac{\mu - \mu_d}{Y_{X_V/gln}}\right).X_V - r_{amm}.X_V \\
\frac{d\,[mAb]}{dt} &= Q_{mAb}.X_V
\end{aligned}
\tag{4}
$$

S1.5 Case 5

An UMM used to monitoring rAAV production[6] is the following

$$
\begin{aligned}
\frac{dX_V(t)}{dt} &= \mu_{X_v} X_V(t) \\
\frac{dGlc(t)}{dt} &= -\mu_{Glc} X_V(t) \\
\frac{dGln(t)}{dt} &= -\mu_{Gln} X_V(t) \\
\frac{dLac(t)}{dt} &= \mu_{Lac} X_V(t) \\
\frac{dAmm(t)}{dt} &= \mu_{Amm} X_V(t) + k_{deg} Gln(t) \\
\frac{dAAV(t)}{dt} &= \mu_{AAV} X_V(t)
\end{aligned}
\tag{5}
$$

Table S1. Model parameters UMM used for Mab production[5].

Parameter	Description
μ_{max}	Maximum growth rate
k_d	Maximum death rate
$Y_{X/glc}$	Yield coefficient cell concentration/ glucose
$Y_{X/gln}$	Yield coefficient cell concentration/ glutamine
$Y_{lac/glc}$	Yield coefficient lactate/ glucose
$Y_{amm/gln}$	Yield coefficient ammonium/ glutamine
Q_{mAb}	Specific production rate
r_{amm}	Ammonium removal rate
m_{glc}	Glucose maintenance coefficient
a_1	Coefficient for m_{glc}
a_2	Coefficient for m_{gln}
K_{glc}	Monod constant glucose
K_{gln}	Monod constant glutamine
K_{Ilac}	Monod constant lactate for inhibition
K_{Iamm}	Monod constant ammonium for inhibition
K_{Dlac}	Monod constant lactate for death
K_{Damm}	Monod constant glucose ammonium for death

This system represents the cell growth, uptake of substrates, metabolism, and production process with six parameters: the specific cell growth rate (μ_{X_v}), the specific rates of uptake (consumption) of the main nutrients, glucose (μ_{Glc}) and glutamine (μ_{Gln}), the specific rates of production of the metabolite waste, lactate (μ_{Lac}) and ammonium (μ_{Amm}), and specific rate of production of rAAV(μ_{AAV}). In the case of ammonium production, the specific rates must consider the spontaneous glutamine degradation in the medium into ammonium. This process follows first-order rate kinetics concerning glutamine concentration, being k_{deg} the glutamine degradation constant.

S2 Continuous-Discrete Extended Kalman Filter

The EKF can be implemented in Discrete-Discrete, and Continuous-Discrete versions[7–9]. The most common version used in biomanufacturing for modeling nonlinear biochemical dynamic pathways is Continuous-Discrete EKF (CD-EKF)[3,10–12] and we will describe it here. The EKF requires a state–space model to perform estimation on the state variables of a process (nonlinear system) present in a state variable vector $\psi(t)$[4,13,14]. A state-space model consists of process and measurement (observation) models[15]. 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[11,14,16].

Effectively, the nonlinear dynamics are approximated by a time-varying linear system, and the linear Kalman filters equations are applied. Essentially, the mean $\hat{\psi}(t)$ and error covariance matrix $\mathbf{P}(t)$ of the state variables in $\psi(t)$ are recursively corrected. The EKF recursively estimates the (posterior) mean $\hat{\psi}(t_{k/k})$ and error covariance matrix $\mathbf{P}(t_{k/k})$ of the state variables by combining the predicted (a priori) mean $\hat{\psi}(t_{k/k-1})$ and error covariance matrix $\mathbf{P}(t_{k/k-1})$ with the current noisy measurement Z_k[16].

Process Model: An UMM as described in the supplementary information Section S1 can be used as the process model of EKF. The state variables vector to be used by the EKF is composed of the state variables of the UMM (observed and unobserved) and the state variables vector is defined as:

$$\psi(t) = [x_1, x_2, \ldots, x_n]^T. \tag{6}$$

Subsequently, the process model is represented as

$$\frac{d\psi(t)}{dt} = \phi(\psi(t), t) + \omega(t), \tag{7}$$

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

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

$$\mathbf{Z}_k = h(\psi(t_k)) + v. \tag{8}$$

The non-linear function h in the measurement model relates the current state variables to the measurements $\mathbf{Z}_k$. The white measurement noise vector is represented by $v \sim \mathcal{N}(0, \mathbf{R})$, with zero mean and measurement noise variance represented by $\mathbf{R}$. When the some state variables can be measured directly, we have a simple case and h can be a linear model. If h is linear, we have $h(\psi(t_k)) = \mathbf{H}\psi(t_k)$[3,4,11]. Where the matrix $\mathbf{H}$ is a linear operator (row vector) that matches the states variables of $\psi(t_k)$ to the measured variables $\mathbf{Z}_k$ that are obtained at a discrete instance k [3,11]. Consequently, the measurement model (8) can be rewritten as

$$\mathbf{Z}_k = \mathbf{H}\psi(t_k) + v. \tag{9}$$

Probabilistic state-space models: The *probabilistic process model* approximated by EKF is

$$p(\psi(t)|\psi(t-1)) \approx \mathcal{N}(\psi(t)|\phi(\psi(t-1)), \mathbf{Q}), \tag{10}$$

where the non-linear functions ϕ are linearized as follows

$$\begin{aligned} &\phi(\psi(t-1)) \approx \phi(\hat{\psi}(t-1)) + \\ &\quad \mathbf{J}_t^{\phi} \times (\psi(t-1) - \hat{\psi}(t-1)) \end{aligned} \tag{11}$$

and $\mathbf{J}_t^{\phi}$ is the Jacobian matrix of ϕ evaluated at the prior mode[17,18],

$$\mathbf{J}_t^{\phi} = \left. \frac{\partial \phi(\psi(t))}{\partial \psi_i} \right|_{\psi(t) = \hat{\psi}(t-1)}. \tag{12}$$

The *probabilistic measurement model* (measurement likelihood distribution) approximated by EKF is the following

$$p(\mathbf{Z}_k|\psi(t_k)) \approx \mathcal{N}(\mathbf{Z}_k|h(\psi(t_k)), \mathbf{R}) \tag{13}$$

where the non-linear functions h are linearized as follows

$$\begin{aligned} &h(\psi(t_k)) \approx h(\hat{\psi}(t_k)) + \\ &\quad \mathbf{J}_t^{h} \times (\psi(t_k) - \hat{\psi}(t_k)) \end{aligned} \tag{14}$$

and the $\mathbf{J}_{t_k}^{h}$ is the Jacobian matrix of h evaluated at the prior mode[17,18]

$$\mathbf{J}_{t_k}^{h} = \left. \frac{\partial h(\psi(t_k))}{\partial \psi} \right|_{\psi(t) = \hat{\psi}(t_{k/k-1})}. \tag{15}$$

In this work, we will consider the case of h be linear. Then, we have

$$\mathbf{J}_{t_k}^{h} = \left. \frac{\partial (\mathbf{H}\psi(t_k))}{\partial \psi} \right|_{\psi(t) = \hat{\psi}(t_{k/k-1})} = \mathbf{H} \left. \frac{\partial (\psi(t_k))}{\partial \psi} \right|_{\psi(t) = \hat{\psi}(t_{k/k-1})} = \mathbf{H} \tag{16}$$

and consequently

$$p(\mathbf{Z}_k|\psi(t_k)) \approx \mathcal{N}(\mathbf{Z}_k|\mathbf{H}\hat{\psi}(t_k) + \mathbf{H} \times (\psi(t_k) - \hat{\psi}(t_k)), \mathbf{R}) \tag{17}$$

or

$$p(\mathbf{Z}_k|\psi(t_k)) \approx \mathcal{N}(\mathbf{Z}_k|\mathbf{H}\hat{\psi}(t_{k/k-1}) + \mathbf{H} \times (\psi(t_{k/k-1}) - \hat{\psi}(t_{k/k-1})), \mathbf{R}). \tag{18}$$

EKF algorithm: The EKF algorithm is implemented through the initial condition, prediction step (time update) and correction step (measurement update)[3,4,11–13].

Initialization step: The initial condition are composed of the initial mean $\hat{\psi}_0 = E[\psi_0]$, and initial error covariance matrix $\mathbf{P}_0 = \mathbf{P}_{i,i}(t=0) = E[(\psi_0 - \hat{\psi}_0)(\psi_0 - \hat{\psi}_0)^T]$ of state variables vector[16].

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

$$\hat{\psi}(t_{k/k-1}) = \hat{\psi}(t_{k-1}) + \int_{t_{k-1}}^{t_k} \phi(\hat{\psi}(t))dt \Big|_{\hat{\psi}(t_{k-1})} \tag{19}$$

and solving the matrix Riccati Differential equation (MRDE) for predict the state error covariance matrix[7,19]

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{J}_t^{\phi}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{J}_t^{\phi T} + \mathbf{Q} \tag{20}$$

from t_{k-1} to t_k, where a new measurement is obtained at time k[20],[7] and[21]. **The Equation 20 is basically a matrix of ODEs, and the matrix of ODEs solutions obtained from t_{k-1} to t_k represent each state error covariance of the system. See the MRDE represented by Equation 50 and the respective solution represented by Matrix 53 in the supplementary information Section S4.**

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

$$\mathbf{K}_k = \mathbf{P}(t_{k|k-1})\mathbf{H}^T(\mathbf{H}\mathbf{P}(t_{k|k-1})\mathbf{H}^T + \mathbf{R})^{-1} \tag{21}$$

$$\hat{\psi}(t_{k/k}) = \hat{\psi}(t_{k/k-1}) + \mathbf{K}_k(\mathbf{Z}_k - \mathbf{H}\hat{\psi}(t_{k/k-1})) \tag{22}$$

$$\mathbf{P}(t_{k|k}) = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}(t_{k|k-1}) \tag{23}$$

The Kalman gain is a scaling factor (ratio) to estimate the state variables by setting a value between the predicted state and measured state[7,22]. The $\mathbf{K}_k$ chooses a value along the residual range ($\mathbf{Z}_k$ - $\mathbf{H}\hat{\psi}(t_{k/k-1})$)[16,22]. $\mathbf{K}_k$ enables to set a value for $\hat{\psi}(t_{k/k})$ between the $\hat{\psi}(t_{k/k-1})$ (prediction) and $\mathbf{Z}_k$ (measurement) using Equation 22, and update the belief regards the state variables based on how certain we are regards the measurement using the Equation 23,[22] (pages 137 and 209). The Kalman gain is computed as a ratio of prior and measurement uncertainty available; see Equation 21. The one dimensional form Equation 21 is the following $K = P/(P+R)$[22]. It is important to point out that linear operator $\mathbf{H}$ matches the states variables of $\psi(t_k)$ to the measured variables $\mathbf{Z}_k$ that are obtained at a discrete instance. It is linear operator with zeros and ones. Where the elements ones represent elements of the state variable vector that are measured. For example, if the state variables vector has 3 variables and only the first one is measured, we have $\mathbf{H} = [1\ 0\ 0]$.

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

S2.1 Intuitions behind Kalman Gain and unshared parameters

Analyzing the Equation 21 is possible to have the following approximation[22,23]

$$\mathbf{K}_k \approx \frac{Process_{\ Uncertainty}}{System_{\ Uncertainty}}, \tag{24}$$

and extract the following two interpretations[22]: **First**, When the Process uncertainty is large (nominator in Equation 24), K is large and so the corrections (Equation 22) are almost determined by the influence of the measured state variable. Since, $\mathbf{K}$ is multiplied by the residual $(\mathbf{Z}_k - \mathbf{H}\hat{\psi}(t_{k/k-1}))$. So, a large K favors the measurement[24]; and **Second**, On the other hand, if the Process uncertainty is very low, the correction step is almost the estimated without influence of the measurement obtained. Since, $\mathbf{K} \approx 0$ and $\hat{\psi}(t_{k/k}) = \hat{\psi}(t_{k/k-1})$[23].

S3 Analysis of Unstructured Mechanistic Models

S3.1 Unshared and shared parameters

The UMM case S1.1 is the case of an ODE system with only unshared parameters. On another hand, the UMM case S1.3 is a ODE system with only shared parameters. For example, the k_3 is used in different ODEs of this system. The UMM case S1.2 is a case of system with unshared and shared parameters.

S3.2 Weak and strong terms

The three UMM cases (S1.1, S1.2 and S1.3) presented above are examples of ODE systems with weak and strong terms. The UMM case S1.1 is example of ODE system with only terms that could be considered weak, because they have low percentage of variables that compose the state variable vector. For example, let assume the following state variable vector $\psi(t)_{case1} = [Xv, N, MP, \mu_{Xv}, \mu_N, \mu_{mp}]$ with six elements for the UMM of case S1.1. We have that all terms of this UMM has a 1/3 of the state variable vector. On another hand, the UMM case S1.2 is example of ODE system with terms that could be considered weak and strong. For example, let consider the following state variable vector for UMM case S1.2,

$$\psi_{case2} = [X_v, fgr, X_d, GLC, GLN, LAC, ASN, ASP, ALA, AMM, Mab, \mu_{max}, k_{31}, k_d] \tag{25}$$

we have that the first term of equation $\frac{dX_v}{dt}$ is the strongest term in this system, since it has 7/14 of state variable vector, and the first term of the equation $\frac{dMab}{dt}$ as the weakest term, since it has 2/14 of state variable vector. In the context of JEKF, we have that a strong term in an UMM contribute more than weak term to compute of predicted state error covariance $\mathbf{P}(t_{k|k-1})$. Since, many elements of Jacobian $\mathbf{J}_t^{\phi}$, results from the first-order partial derivatives of strong term with respect to the variables of state variable vector $\psi(t)$. For example, let consider the following:

- State variable vector $\psi(t) = [x_1, x_2, x_3, x_4]$
- An UMM composed of an ODE with a strong term by the function S(x1,x2,x3,x4) and three ODEs composed of waek terms represented by the function $W_1(x_1)$, $W_2(x_2)$ and $W_3(x_3)$.
- MRDE (Equation 20).
- $\mathbf{P}$ and $\mathbf{Q}$ uncorrelated for the $\psi(t)$,

$$\mathbf{P} = \begin{bmatrix} P_{1,1} & 0 & 0 & 0 \\ 0 & P_{2,2} & 0 & 0 \\ 0 & 0 & P_{3,3} & 0 \\ 0 & 0 & 0 & P_{4,4} \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} Q_{1,1} & 0 & 0 & 0 \\ 0 & Q_{2,2} & 0 & 0 \\ 0 & 0 & Q_{3,3} & 0 \\ 0 & 0 & 0 & Q_{4,4} \end{bmatrix}. \tag{26}$$

Given this we have the following Jacobian

$$\mathbf{Jacobian}(S, W_1, W_2, W_3) = \begin{bmatrix} \frac{\partial S}{\partial x_1} & \frac{\partial S}{\partial x_2} & \frac{\partial S}{\partial x_3} & \frac{\partial S}{\partial x_4} \\ \frac{\partial W_1}{\partial x_1} & 0 & 0 & 0 \\ 0 & \frac{\partial W_2}{\partial x_2} & 0 & 0 \\ 0 & 0 & \frac{\partial W_3}{\partial x_3} & 0 \end{bmatrix} \tag{27}$$

and the following MRDE to compute the Predicted state error covariance $\mathbf{P}(t_{k/k-1})$ from t_{k-1} to t_k,

$$\frac{d\mathbf{P}(t)}{dt} = \begin{bmatrix} Q_{1,1}+2.P_{1,1}\cdot\frac{\partial S}{\partial x_1} & P_{1,1}\cdot\frac{\partial W_1}{\partial x_1}+P_{2,2}\cdot\frac{\partial S}{\partial x_2} & P_{3,3}\cdot\frac{\partial S}{\partial x_3} & P_{4,4}\cdot\frac{\partial S}{\partial x_4} \\ P_{1,1}\cdot\frac{\partial W_1}{\partial x_1}+P_{2,2}\cdot\frac{\partial S}{\partial x_2} & Q_{2,2} & P_{2,2}\cdot\frac{\partial W_2}{\partial x_2} & 0 \\ P_{3,3}\cdot\frac{\partial S}{\partial x_3} & P_{2,2}\cdot\frac{\partial W_2}{\partial x_2} & Q_{3,3} & P_{3,3}\cdot\frac{\partial W_3}{\partial x_3} \\ P_{4,4}\cdot\frac{\partial S}{\partial x_4} & 0 & P_{3,3}\cdot\frac{\partial W_3}{\partial x_3} & Q_{4,4} \end{bmatrix}. \tag{28}$$

Then, we can see that S contribute with 7 partial derivatives and the others functions with 2 partial derivative only each one. If we consider the MRDE formed with $\mathbf{P}$ correlated, we have S contributing with 32 partial derivatives and the others functions with 8 partial derivative only each one. It is important point out that the element 0 in the MRDE (Equation 28) represents an time invariant ODE $\frac{dP_{x_2,x_4}(t_{k/k-1})}{dt} = 0$ to predicts the covariance between x_4 and x_2 when solved from t_{k-1} to t_k.

S3.3 Weak and strong variables

The three UMM cases (S1.1, S1.2 and S1.3) presented above are examples of the ODE system with weak and strong variables. The variable MP in the UMM case S1.1 and Mab in UMM case S1.2 are examples of weak variable. In these cases, the first-order partial derivatives of all functions with respect to these two variables are equal to zero and this reflects that the variable has a column with zeros in the jacobian $\mathbf{J}_t^{\phi}$. On another hand, in the UMM case S1.1 and S1.2, Xv is an example of strong variable.

S3.4 MRDE to predict the state error covariance $\mathbf{P}(t_{k|k-1})$ based on P and Q with uncorrelated elements

For any UMM, the use of $\mathbf{P}$ and $\mathbf{Q}$ with uncorrelated elements in MRDE, means that the predicted state error covariance $\mathbf{P}(t_{k|k-1})$ will be updated/calculated based only in noise variance of $P_{i,i}$ and $Q_{i,i}$ and elements of Jacobian $\mathbf{J}_t^{\phi}$.

For example, let consider the following conditions:

- The set of functions $\mathbf{f} = [f_1, f_2, f_3, f_4, f_5]$ and state variables vector $\psi(t)_{\mathbf{f}} = [x_1, x_2, x_3, x_4, x_5]$.
- P and Q uncorrelated for the $\psi(t)_{\mathbf{f}}$,

$$\mathbf{P} = \begin{bmatrix} P_{1,1} & 0 & 0 & 0 & 0 \\ 0 & P_{2,2} & 0 & 0 & 0 \\ 0 & 0 & P_{3,3} & 0 & 0 \\ 0 & 0 & 0 & P_{4,4} & 0 \\ 0 & 0 & 0 & 0 & P_{5,5} \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} Q_{1,1} & 0 & 0 & 0 & 0 \\ 0 & Q_{2,2} & 0 & 0 & 0 \\ 0 & 0 & Q_{3,3} & 0 & 0 \\ 0 & 0 & 0 & Q_{4,4} & 0 \\ 0 & 0 & 0 & 0 & Q_{5,5} \end{bmatrix}. \tag{29}$$

- The Jacobian ,

$$\mathbf{Jacobian(f)} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} & \frac{\partial f_1}{\partial x_4} & \frac{\partial f_1}{\partial x_5} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} & \frac{\partial f_2}{\partial x_4} & \frac{\partial f_2}{\partial x_5} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} & \frac{\partial f_3}{\partial x_4} & \frac{\partial f_3}{\partial x_5} \\ \frac{\partial f_4}{\partial x_1} & \frac{\partial f_4}{\partial x_2} & \frac{\partial f_4}{\partial x_3} & \frac{\partial f_4}{\partial x_4} & \frac{\partial f_4}{\partial x_5} \\ \frac{\partial f_5}{\partial x_1} & \frac{\partial f_5}{\partial x_2} & \frac{\partial f_5}{\partial x_3} & \frac{\partial f_5}{\partial x_4} & \frac{\partial f_5}{\partial x_5} \end{bmatrix} \tag{30}$$

Given these conditions and the Equation 20, we have the following MRDE to compute the predicted state error covariance $P(t_{k/k-1})$ from t_{k-1} to t_k,

$$\frac{d\mathbf{P}(t)}{dt} = \begin{bmatrix} Q_{1,1}+2.\frac{\partial f_1}{\partial x_1}.P_{1,1} & P_{1,1}\cdot\frac{\partial f_2}{\partial x_1}+\frac{\partial f_1}{\partial x_2}.P_{2,2} & P_{1,1}\cdot\frac{\partial f_3}{\partial x_1}+\frac{\partial f_1}{\partial x_3}.P_{3,3} & P_{1,1}\cdot\frac{\partial f_4}{\partial x_1}+\frac{\partial f_1}{\partial x_4}.P_{4,4} & P_{1,1}\cdot\frac{\partial f_5}{\partial x_1}+\frac{\partial f_1}{\partial x_5}.P_{5,5} \\ P_{2,2}\cdot\frac{\partial f_1}{\partial x_2}+\frac{\partial f_2}{\partial x_1}.P_{1,1} & Q_{2,2}+2.\frac{\partial f_2}{\partial x_2}.P_{2,2} & P_{2,2}\cdot\frac{\partial f_3}{\partial x_2}+\frac{\partial f_2}{\partial x_3}.P_{3,3} & P_{2,2}\cdot\frac{\partial f_4}{\partial x_2}+\frac{\partial f_2}{\partial x_4}.P_{4,4} & P_{2,2}\cdot\frac{\partial f_5}{\partial x_2}+\frac{\partial f_2}{\partial x_5}.P_{5,5} \\ P_{3,3}\cdot\frac{\partial f_1}{\partial x_3}+\frac{\partial f_3}{\partial x_1}.P_{1,1} & P_{3,3}\cdot\frac{\partial f_2}{\partial x_3}+\frac{\partial f_3}{\partial x_2}.P_{2,2} & Q_{3,3}+2.\frac{\partial f_3}{\partial x_3}.P_{3,3} & P_{3,3}\cdot\frac{\partial f_4}{\partial x_3}+\frac{\partial f_3}{\partial x_4}.P_{4,4} & P_{3,3}\cdot\frac{\partial f_5}{\partial x_3}+\frac{\partial f_3}{\partial x_5}.P_{5,5} \\ P_{4,4}\cdot\frac{\partial f_1}{\partial x_4}+\frac{\partial f_4}{\partial x_1}.P_{1,1} & P_{4,4}\cdot\frac{\partial f_2}{\partial x_4}+\frac{\partial f_4}{\partial x_2}.P_{2,2} & P_{4,4}\cdot\frac{\partial f_3}{\partial x_4}+\frac{\partial f_4}{\partial x_3}.P_{3,3} & Q_{4,4}+2.\frac{\partial f_4}{\partial x_4}.P_{4,4} & P_{4,4}\cdot\frac{\partial f_5}{\partial x_4}+\frac{\partial f_4}{\partial x_5}.P_{5,5} \\ P_{5,5}\cdot\frac{\partial f_1}{\partial x_5}+\frac{\partial f_5}{\partial x_1}.P_{1,1} & P_{5,5}\cdot\frac{\partial f_2}{\partial x_5}+\frac{\partial f_5}{\partial x_2}.P_{2,2} & P_{5,5}\cdot\frac{\partial f_3}{\partial x_5}+\frac{\partial f_5}{\partial x_3}.P_{3,3} & P_{5,5}\cdot\frac{\partial f_4}{\partial x_5}+\frac{\partial f_5}{\partial x_4}.P_{4,4} & Q_{5,5}+2.\frac{\partial f_5}{\partial x_5}.P_{5,5} \end{bmatrix}. \tag{31}$$

Then, we have that each ODE of the $\mathbf{P}(t_{k/k-1})$ is composed only of the elements of $P_{i,i}$ and $Q_{i,i}$ and elements of $\mathbf{Jacobian(f)}$. Furthermore, these ODEs can be time invariant, if the partial derivative are zero. See the equation 28.

S3.5 Only one variable measured

In the UMM case S1.2. For example, let consider the following state variable vector,

$$\psi_{case2} = [X_v, fgr, X_d, GLC, GLN, LAC, ASN, ASP, ALA, AMM, Mab, \mu_{max}, k_{31}, k_d] \tag{32}$$

To estimate the entire state variable vector is need to have the minimum of two measurement, and an option is Xv and GLC. Since, the column regards X_v in MRDE is zero k_{31} ODE, but it is different of zero in GLC column.

S4 Proof of Lemma: Inability to Update Kalman Gain for Unshared parameters

The Kalman gain cannot be updated/calculated (by Eq 21) for an unshared parameter that is part of a state variable vector and part of a weak term in an UMM, if the initial state error covariance **P**(t=0) and **Q** are formed by uncorrelated elements and there is only one state variable measured.

Proof. Let's consider the following:

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

$$\frac{dx_{msv}}{dt} = f_1(x_{msv}, x_2, \dots, x_{n-1}, p_1, p_2, \dots, p_m) \tag{33}$$
$$\frac{dx_2}{dt} = f_2(x_{msv}, x_2, \dots, x_{n-1}, p_1, p_2, \dots, p_m) \tag{34}$$
$$\vdots \tag{35}$$
$$\frac{dx_n}{dt} = f_n(x_{msv}, \theta_{up}) \tag{36}$$

  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 $p_1, p_2, \dots, p_m$ are the parameters of the system and θ_{up} an unshared parameter.

- A joint state variables vector defined as

$$\psi(t)_{general} = [x_{msv}, x_2, \dots, x_n, \theta_{up}]. \tag{37}$$

- A process model defined as

$$\frac{d\psi(t)_{general}}{dt} = \phi(\psi(t)_{general}, t) + \omega(t) = \frac{d}{dt}\begin{bmatrix} x_{msv} \\ x_2 \\ \vdots \\ x_n \\ \theta_{up} \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \\ 0 \end{bmatrix} + \omega(t). \tag{38}$$

- 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.
- P and Q uncorrelated elements for the $\psi(t)$ (Equation 37),

$$\mathbf{P} = \begin{bmatrix} P_{x_{msv},x_{msv}} & 0 & \dots & 0 & 0 \\ 0 & P_{x_2,x_2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & P_{n,n} & 0 \\ 0 & 0 & \dots & 0 & P_{\theta_{up},\theta_{up}} \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} Q_{x_{msv},x_{msv}} & 0 & \dots & 0 & 0 \\ 0 & Q_{x_2,x_2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & Q_{n,n} & 0 \\ 0 & 0 & \dots & 0 & Q_{\theta_{up},\theta_{up}} \end{bmatrix}. \tag{39}$$

- The Jacobian $\mathbf{J}_t^{\phi}$, (Equation 12) with the $\psi(t)_{general}$ (Equation 37),

$$\mathbf{J}_t^{\phi}(\phi(\psi(t)_{general},t)) = \begin{bmatrix} \frac{\partial f_1}{\partial x_{msv}} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} & 0 \\ \frac{\partial f_2}{\partial x_{msv}} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_{msv}} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}. \tag{40}$$

Given these conditions and the Equation 20, we have the following MRDE (based on **P** uncorrelated)

$$\frac{d\mathbf{P}(t)}{dt} = \begin{bmatrix} \frac{dP_{x_{msv},x_{msv}}(t)}{dt} = Q_{1,1} + 2P_{1,1}\frac{\partial f_1}{\partial x_{msv}} & \frac{dP_{x_2,x_{msv}}(t)}{dt} = (P_{1,1}+P_{2,2})\frac{\partial f_1}{\partial x_2} & \dots & \frac{dP_{x_n,x_{msv}}(t)}{dt} = (P_{1,1}+P_{3,3})\frac{\partial f_1}{\partial x_n} & \frac{dP_{\theta_{up},x_{msv}}(t)}{dt} 0 \\ \frac{dP_{x_{msv},x_2}(t)}{dt} = (P_{1,1}+P_{2,2})\frac{\partial f_2}{\partial x_{msv}} & \frac{dP_{x_2,x_2}(t)}{dt} = Q_{2,2} + 2P_{2,2}\frac{\partial f_2}{\partial x_2} & \dots & \frac{dP_{x_n,x_2}(t)}{dt} = (P_{3,3}+P_{2,2})\frac{\partial f_2}{\partial x_n} & \frac{dP_{\theta_{up},x_2}(t)}{dt} = 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{dP_{x_{msv},x_n}(t)}{dt} = (P_{1,1}+P_{n,n})\frac{\partial f_n}{\partial x_{msv}} & \frac{dP_{x_2,x_n}(t)}{dt} = (P_{2,2}+P_{n,n})\frac{\partial f_n}{\partial x_2} & \dots & \frac{dP_{x_n,x_n}(t)}{dt} = Q_{n,n} + 2P_{n,n}\frac{\partial f_n}{\partial x_n} & \frac{dP_{\theta_{up},x_n}(t)}{dt} = 0 \\ \frac{dP_{x_{msv},\theta_{up}}(t)}{dt} = 0 & \frac{dP_{x_2,\theta_{up}}(t)}{dt} = 0 & \dots & \frac{dP_{x_n,\theta_{up}}(t)}{dt} = 0 & \frac{dP_{\theta_{up},\theta_{up}}(t)}{dt} = 0 \end{bmatrix}. \tag{41}$$

Now, using this Equation 41 to compute the predicted state error covariance matrix $\mathbf{P}(t_{k/k-1})$ from t_{k-1} to t_k with a initial predicted state error covariance matrix $\mathbf{P}(t_{k-1}) = \mathbf{P}_0 = \mathbf{P}_{init}(t=0)$ with uncorrelated elements as following

$$\mathbf{P}_{init}(t=0) = \begin{bmatrix} P_{x_{msv},x_{msv}}(t=0) & 0 & \dots & 0 & 0 \\ 0 & P_{x_2,x_2}(t=0) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & P_{n,n}(t=0) & 0 \\ 0 & 0 & \dots & 0 & P_{\theta_{up},\theta_{up}}(t=0) \end{bmatrix}, \tag{42}$$

we have

$$\mathbf{P}(t_{k/k-1}) = \begin{bmatrix} P_{x_{msv},x_{msv}}(t_{k/k-1}) & P_{x_2,x_{msv}}(t_{k/k-1}) & \dots & P_{n,x_{msv}}(t_{k/k-1}) & P_{\theta_{up},x_{msv}}(t_{k/k-1}) \\ P_{x_{msv},x_2}(t_{k/k-1}) & P_{x_2,x_2}(t_{k/k-1}) & \dots & P_{n,x_2}(t_{k/k-1}) & P_{\theta_{up},x_2}(t_{k/k-1}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ P_{x_{msv},n}(t_{k/k-1}) & P_{x_2,n}(t_{k/k-1}) & \dots & P_{n,n}(t_{k/k-1}) & P_{\theta_{up},n}(t_{k/k-1}) \\ P_{x_{msv},\theta_{up}}(t_{k/k-1}) = 0 & P_{x_2,\theta_{up}}(t_{k/k-1}) & \dots & P_{n,\theta_{up}}(t_{k/k-1}) & P_{\theta_{up},\theta_{up}}(t_{k/k-1}) \end{bmatrix}. \tag{43}$$

Now, using $\mathbf{P}(t_{k/k-1})$, **H** and R to compute the Kalman gain for all variables in the state variable vector $\psi(t)_{general}$ (Equation 37), we have

$$\mathbf{K}_k = \mathbf{P}(t_{k|k-1})\mathbf{H}^T(\mathbf{H}\mathbf{P}(t_{k|k-1})\mathbf{H}^T + \mathbf{R})^{-1} = \begin{bmatrix} K_{x_{msv}} \\ K_{x_2} \\ \vdots \\ K_{x_n} \\ K_{\theta_{up}} \end{bmatrix} = \begin{bmatrix} \frac{P_{x_{msv},x_{msv}}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \frac{P_{x_{msv},x_2}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \vdots \\ \frac{P_{x_{msv},n}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \frac{P_{x_{msv},\theta_{up}}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \end{bmatrix} = \begin{bmatrix} \frac{P_{x_{msv},x_{msv}}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \frac{P_{x_{msv},x_2}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \vdots \\ \frac{P_{x_{msv},n}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \frac{0}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \end{bmatrix} = \begin{bmatrix} \frac{P_{x_{msv},x_{msv}}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \frac{P_{x_{msv},x_2}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ \vdots \\ \frac{P_{x_{msv},n}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \\ 0 \end{bmatrix}. \tag{44}$$

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, $P_{x_{msv},\theta_{up}}(t_{k/k-1}) = Cov(x_{msv}, \theta_{up}) = 0$. Since 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 **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, we have 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 Eq 23). Since

$$\mathbf{P}(t_{k|k}) = (\mathbf{I}-\mathbf{K}_k\mathbf{H})\mathbf{P}(t_{k|k-1}) = \begin{bmatrix} \vdots & \iddots \\ P_{x_{msv},\theta_{up}}(t_{k/k-1}) - K_{\theta_{up}}.P_{x_{msv},x_{msv}}(t_{k/k-1}) & \dots \end{bmatrix} = \begin{bmatrix} \vdots & \iddots \\ 0-0.P_{x_{msv},x_{msv}}(t_{k/k-1}) & \dots \end{bmatrix}. \tag{45}$$

Therefore, we have that $P_{x_{msv},\theta_{up}}(t_{k/k}) = P_{x_{msv},\theta_{up}}(t_{k/k-1}) = 0$, and as $P_{x_{msv},\theta_{up}}(t_{k/k}) = 0$ have to be used as a new initial condition for MRDE (Equation 41), 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 41 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. ■

S4.1 Example

To illustrate this Lemma, we show that the Kalman gain value cannot be updated for the unshared parameter μ_{mp} of the UMM with weak terms presented in Case S1.1. Let's consider the following:

- The state variables vector

$$\psi(t)_{case1} = [Xv, N, MP, \mu_{Xv}, \mu_N, \mu_{mp}]. \tag{46}$$

- Xv is the unique measured variable and $\mathbf{H} = [1\ 0\ 0\ 0\ 0\ 0]$.
- R as measurement noise variance of Xv.
- μ_{mp} the unshared parameter to be evolved (estimated) and that is related to a weak term.
- **P** and **Q** with uncorrelated elements for the $\psi(t)_{case1}$ (Equation 46),

$$\mathbf{P} = \begin{bmatrix} P_{1,1} & 0 & 0 & 0 & 0 & 0 \\ 0 & P_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & P_{3,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & P_{4,4} & 0 & 0 \\ 0 & 0 & 0 & 0 & P_{5,5} & 0 \\ 0 & 0 & 0 & 0 & 0 & P_{6,6} \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} Q_{1,1} & 0 & 0 & 0 & 0 & 0 \\ 0 & Q_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & Q_{3,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & Q_{4,4} & 0 & 0 \\ 0 & 0 & 0 & 0 & Q_{5,5} & 0 \\ 0 & 0 & 0 & 0 & 0 & Q_{6,6} \end{bmatrix}. \tag{47}$$

- The Jacobian $\mathbf{J}_t^{\phi}$, (Equation 12) with the $\psi(t)_{case1}$ (Equation 46),

$$\mathbf{J}_t^{\phi} = \begin{bmatrix} \mu_{Xv} & 0 & 0 & Xv & 0 & 0 \\ \mu_N & 0 & 0 & 0 & Xv & 0 \\ \mu_{mp} & 0 & 0 & 0 & 0 & Xv \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{48}$$

Given these conditions and the Equation 20, we have the following MRDE (based on **P** with uncorrelated elements)

$$\frac{d\mathbf{P}(t)}{dt} = \begin{bmatrix} Q_{1,1}+2.P_{1,1}.\mu_{Xv} & P_{1,1}.\mu_N & P_{1,1}.\mu_{mp} & P_{4,4}.Xv & 0 & 0 \\ P_{1,1}.\mu_N & Q_{2,2} & 0 & 0 & P_{5,5}.Xv & 0 \\ P_{1,1}.\mu_{mp} & 0 & Q_{3,3} & 0 & 0 & P_{6,6}.Xv \\ P_{4,4}.Xv & 0 & 0 & Q_{4,4} & 0 & 0 \\ 0 & P_{5,5}.Xv & 0 & 0 & Q_{5,5} & 0 \\ 0 & 0 & P_{6,6}.Xv & 0 & 0 & Q_{6,6} \end{bmatrix} = \tag{49}$$

$$\frac{d\mathbf{P}(t)}{dt} = \begin{bmatrix} \frac{P_{X_v,X_v}(t)}{dt} = Q_{1,1} + 2.P_{1,1}.\mu_{Xv} & \frac{P_{N,X_v}(t)}{dt} = P_{1,1}.\mu_N & \dots & \frac{P_{\mu_N,X_v}(t)}{dt} = 0 & \frac{P_{\mu_{mp},X_v}(t)}{dt} = 0 \\ \frac{P_{X_v,N}(t)}{dt} = P_{1,1}.\mu_N & \frac{P_{N,N}(t)}{dt} = Q_{2,2} & \dots & \frac{P_{\mu_N,N}(t)}{dt} = P_{5,5}.Xv & \frac{P_{\mu_{mp},N}(t)}{dt} = 0 \\ \frac{P_{X_v,MP}(t)}{dt} = P_{1,1}.\mu_{mp} & \frac{P_{N,MP}(t)}{dt} = 0 & \dots & \frac{P_{\mu_N,MP}(t)}{dt} = 0 & \frac{P_{\mu_{mp},MP}(t)}{dt} = P_{6,6}.Xv \\ \frac{P_{X_v,\mu_{X_v}}(t)}{dt} = P_{4,4}.Xv & \frac{P_{N,\mu_{X_v}}(t)}{dt} = 0 & \dots & \frac{P_{\mu_N,\mu_{X_v}}(t)}{dt} = 0 & \frac{P_{\mu_{mp},\mu_{X_v}}(t)}{dt} = 0 \\ \frac{P_{X_v,\mu_N}(t)}{dt} = 0 & \frac{P_{N,\mu_N}(t)}{dt} = P_{5,5}.Xv & \dots & \frac{P_{\mu_N,\mu_N}(t)}{dt} = Q_{5,5} & \frac{P_{\mu_{mp},\mu_N}(t)}{dt} = 0 \\ \frac{P_{X_v,\mu_{mp}}(t)}{dt} = 0 & \frac{P_{N,\mu_{mp}}(t)}{dt} = 0 & \dots & \frac{P_{\mu_N,\mu_{mp}}(t)}{dt} = 0 & \frac{P_{\mu_{mp},\mu_{mp}}(t)}{dt} = Q_{6,6} \end{bmatrix}. \tag{50}$$

Now, using this Equation 50 to compute the predicted state error covariance matrix $\mathbf{P}(t_{k/k-1})$ (for the Case S1.1) from t_{k-1} to t_k with a initial predicted state error covariance matrix $\mathbf{P}(t_{k-1}) = \mathbf{P}(t=0)$ with uncorrelated elements as following

$$\mathbf{P}(t=0) = \begin{bmatrix} P_{X_v,X_v}(t=0) & 0 & 0 & 0 & 0 & 0 \\ 0 & P_{N,N}(t=0) & 0 & 0 & 0 & 0 \\ 0 & 0 & P_{MP,MP}(t=0) & 0 & 0 & 0 \\ 0 & 0 & 0 & P_{\mu_{X_v},\mu_{X_v}}(t=0) & 0 & 0 \\ 0 & 0 & 0 & 0 & P_{\mu_N,\mu_N}(t=0) & 0 \\ 0 & 0 & 0 & 0 & 0 & P_{\mu_{mp},\mu_{mp}}(t=0) \end{bmatrix}, \tag{51}$$

we have

$$\mathbf{P}(t_{k/k-1}) = \begin{bmatrix} P_{X_v,X_v}(t_{k/k-1}) & P_{N,X_v}(t_{k/k-1}) & P_{MP,X_v}(t_{k/k-1}) & P_{\mu_{X_v},X_v}(t_{k/k-1}) & P_{\mu_N,X_v}(t_{k/k-1}) & P_{\mu_{mp},X_v}(t_{k/k-1}) \\ P_{X_v,N}(t_{k/k-1}) & P_{N,N}(t_{k/k-1}) & P_{MP,N}(t_{k/k-1}) & P_{\mu_{X_v},N}(t_{k/k-1}) & P_{\mu_N,N}(t_{k/k-1}) & P_{\mu_{mp},N}(t_{k/k-1}) \\ P_{X_v,MP}(t_{k/k-1}) & P_{N,MP}(t_{k/k-1}) & P_{MP,MP}(t_{k/k-1}) & P_{\mu_{X_v},MP}(t_{k/k-1}) & P_{\mu_N,MP}(t_{k/k-1}) & P_{\mu_{mp},MP}(t_{k/k-1}) \\ P_{X_v,\mu_{X_v}}(t_{k/k-1}) & P_{N,\mu_{X_v}}(t_{k/k-1}) & P_{MP,\mu_{X_v}}(t_{k/k-1}) & P_{\mu_{X_v},\mu_{X_v}}(t_{k/k-1}) & P_{\mu_N,\mu_{X_v}}(t_{k/k-1}) & P_{\mu_{mp},\mu_{X_v}}(t_{k/k-1}) \\ P_{X_v,\mu_N}(t_{k/k-1}) & P_{N,\mu_N}(t_{k/k-1}) & P_{MP,\mu_N}(t_{k/k-1}) & P_{\mu_{X_v},\mu_N}(t_{k/k-1}) & P_{\mu_N,\mu_N}(t_{k/k-1}) & P_{\mu_{mp},\mu_N}(t_{k/k-1}) \\ P_{X_v,\mu_{mp}}(t_{k/k-1}) & P_{N,\mu_{mp}}(t_{k/k-1}) & P_{MP,\mu_{mp}}(t_{k/k-1}) & P_{\mu_{X_v},\mu_{mp}}(t_{k/k-1}) & P_{\mu_N,\mu_{mp}}(t_{k/k-1}) & P_{\mu_{mp},\mu_{mp}}(t_{k/k-1}) \end{bmatrix} = \tag{52}$$

$$\mathbf{P}(t_{k/k-1}) = \begin{bmatrix} Cov(X_v,X_v) & Cov(N,X_v) & Cov(MP,X_v) & Cov(\mu_{X_v},X_v) & Cov(\mu_N,X_v) & Cov(\mu_{mp},X_v) \\ Cov(X_v,N) & Cov(N,N) & Cov(MP,N) & Cov(\mu_{X_v},N) & Cov(\mu_N,N) & Cov(\mu_{mp},N) \\ Cov(X_v,MP) & Cov(N,MP) & Cov(MP,MP) & Cov(\mu_{X_v},MP) & Cov(\mu_N,MP) & Cov(\mu_{mp},MP) \\ Cov(X_v,\mu_{X_v}) & Cov(N,\mu_{X_v}) & Cov(MP,\mu_{X_v}) & Cov(\mu_{X_v},\mu_{X_v}) & Cov(\mu_N,\mu_{X_v}) & Cov(\mu_{mp},\mu_{X_v}) \\ Cov(X_v,\mu_N) & Cov(N,\mu_N) & Cov(MP,\mu_N) & Cov(\mu_{X_v},\mu_N) & Cov(\mu_N,\mu_N) & Cov(\mu_{mp},\mu_N) \\ Cov(X_v,\mu_{mp}) & Cov(N,\mu_{mp}) & Cov(MP,\mu_{mp}) & Cov(\mu_{X_v},\mu_{mp}) & Cov(\mu_N,\mu_{mp}) & Cov(\mu_{mp},\mu_{mp}) \end{bmatrix}. \tag{53}$$

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

$$\mathbf{K}_k = \mathbf{P}(t_{k|k-1})\mathbf{H}^T(\mathbf{H}\mathbf{P}(t_{k|k-1})\mathbf{H}^T + \mathbf{R})^{-1} = \begin{bmatrix} K_{X_v} \\ K_N \\ K_{MP} \\ K_{\mu_{X_v}} \\ K_{\mu_N} \\ K_{\mu_{mp}} \end{bmatrix} = \begin{bmatrix} \frac{P_{X_v,X_v}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,N}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,MP}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,\mu_{X_v}}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,\mu_N}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,\mu_{mp}}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \end{bmatrix} = \begin{bmatrix} \frac{P_{X_v,X_v}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,N}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,MP}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,\mu_{X_v}}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,\mu_N}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{0}{P_{X_v,X_v}(t_{k/k-1})+R} \end{bmatrix} = \begin{bmatrix} \frac{P_{X_v,X_v}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,N}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,MP}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,\mu_{X_v}}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ \frac{P_{X_v,\mu_N}(t_{k/k-1})}{P_{X_v,X_v}(t_{k/k-1})+R} \\ 0 \end{bmatrix}. \tag{54}$$

$\mathbf{H}$ selected the first column of $\mathbf{P}(t_{k/k-1})$ since it is related to the measured value X_v. However, in this column, we have that the predicted error covariance between X_v and μ_{mp} is zero, $Cov(X_v, \mu_{mp}) = 0$. Since the solution of $\frac{P_{X_v,\mu_{mp}}(t)}{dt} = 0$ obtained from t_{k-1} to t_k is zero, and we have $P_{X_v,\mu_{mp}}(t_{k/k-1}) = P_{\mu_{X_v},\mu_{mp}}(t_{k-1}) = P_{\mu_{X_v},\mu_{mp}}(t=0) = 0$. This means that due to $\mathbf{P}$(t=0) with uncorrelated elements the obtained solution is equal to the initial condition. Then, we have the kalman gain value for the unshared parameter is zero, $K_{\mu_{mp}}=0$, and consequently the predicted state error covariance $P_{X_v,\mu_{mp}}(t_{k/k-1})$ cannot be updated in $\mathbf{P}(t_{k|k-1})$ by Eq 23. Since

$$\mathbf{P}(t_{k|k})=(\mathbf{I}-\mathbf{K}_k\mathbf{H})\mathbf{P}(t_{k|k-1})=\begin{bmatrix} P_{X_v,X_v}(t_{k/k-1})-K_{Xv}.P_{X_v,X_v}(t_{k/k-1}) & \dots \\ P_{X_v,N}(t_{k/k-1})-K_{N}.P_{X_v,X_v}(t_{k/k-1}) & \dots \\ P_{X_v,MP}(t_{k/k-1})-K_{mp}.P_{X_v,X_v}(t_{k/k-1}) & \dots \\ P_{X_v,\mu_{Xv}}(t_{k/k-1})-K_{\mu_{Xv}}.P_{X_v,X_v}(t_{k/k-1}) & \dots \\ P_{X_v,\mu_N}(t_{k/k-1})-K_{\mu_N}.P_{X_v,X_v}(t_{k/k-1}) & \dots \\ P_{X_v,\mu_{mp}}(t_{k/k-1})-K_{\mu_{mp}}.P_{X_v,X_v}(t_{k/k-1}) & \dots \end{bmatrix}= \tag{55}$$

$$\mathbf{P}(t_{k|k})=(\mathbf{I}-\mathbf{K}_k\mathbf{H})\mathbf{P}(t_{k|k-1})=\begin{bmatrix} P_{X_v,X_v}(t_{k/k-1})-K_{Xv}.P_{X_v,X_v}(t_{k/k-1}) & \dots \\ P_{X_v,N}(t_{k/k-1})-K_{N}.P_{X_v,X_v}(t_{k/k-1}) & \dots \\ P_{X_v,MP}(t_{k/k-1})-K_{mp}.P_{X_v,X_v}(t_{k/k-1}) & \dots \\ P_{X_v,\mu_{Xv}}(t_{k/k-1})-K_{\mu_{Xv}}.P_{X_v,X_v}(t_{k/k-1}) & \dots \\ P_{X_v,\mu_N}(t_{k/k-1})-K_{\mu_N}.P_{X_v,X_v}(t_{k/k-1}) & \dots \\ 0-0.P_{X_v,X_v}(t_{k/k-1}) & \dots \end{bmatrix}. \tag{56}$$

We have that $P_{X_v,\mu_{mp}}(t_{k/k})=P_{X_v,\mu_{mp}}(t_{k/k-1})=0$, and as the $P_{X_v,\mu_{mp}}(t_{k/k})=0$ have to be used as a new initial condition for MRDE (Equation 50), we have $K_{\mu_{mp}}=0$ for all $P_{X_v,\mu_{mp}}(t_{k/k-1})$ in $\mathbf{P}(t_{k/k-1})$ that are obtained from t_{k-1} to t_k using Equation 50 and consequently no updates for $K_{\mu_{mp}}$ and $P_{X_v,\mu_{mp}}(t_{k/k-1})$.

S5 Proof of Theorem (JEKF failure)

The JEKF 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 **P**(t=0) and **Q** are composed of uncorrelated elements, and there is only one state variable measured. Since the Kalman gain value for the unshared parameter is equal to zero for all steps of execution of JEKF algorithm.

Proof. This proof can be done using the conditions and results described previously in the proof of the Lemma S4.

Then, let consider the following:

- **H**=[1 0 ... 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 S4, where $K_{\theta_{up}}=0$.
- $\mathbf{Z}_k$ as measured value of x_{msv} .
- Predicted mean of the state variable vector $\hat{\psi}(t_{k/k-1})_{general}=[\hat{x_{msv}},\hat{x_2},\dots,\hat{x_n},\hat{\theta_{up}}]^T$ regards to the general UMM used in the proof S4.

Now, using the Equation 22 to compute the estimated mean of the state variable vector $\hat{\psi}(t_{k/k})_{general}$, we have

$$\hat{\psi}(t_{k/k})_{general}=\hat{\psi}(t_{k/k-1})_{general}+\mathbf{K}_k(\mathbf{Z}_k-\mathbf{H}\hat{\psi}(t_{k/k-1})_{general}) \tag{57}$$

$$\hat{\psi}(t_{k/k})_{general}=\begin{bmatrix}\hat{x_{msv}}\\ \hat{x_2}\\ \vdots\\ \hat{x_n}\\ \hat{\theta_{up}}\end{bmatrix}+\begin{bmatrix}K_{x_{msv}}\\ K_{x_2}\\ \vdots\\ K_{x_n}\\ K_{\theta_{up}}\end{bmatrix}.(\mathbf{Z}_k-\hat{x_{msv}})=\begin{bmatrix}\hat{x_{msv}}+K_{x_{msv}}.(\mathbf{Z}_k-\hat{x_{msv}})\\ \hat{x_2}+K_{x_2}.(\mathbf{Z}_k-\hat{x_{msv}})\\ \vdots\\ \hat{x_n}+K_{x_n}.(\mathbf{Z}_k-\hat{x_{msv}})\\ \hat{\theta_{up}}+0\end{bmatrix} \tag{58}$$

Then, we have that the estimated mean of the unshared parameter $\hat{\theta_{up}}(t_{k/k})$ (composing the $\hat{\psi}(t_{k/k})_{general}$) is equal to the predicted mean of unshared parameter $\hat{\theta_{up}}(t_{k/k-1})$ (composing the $\hat{\psi}(t_{k/k-1})_{general}$) for all step from t_{k-1} to t_k. In another words, the JEKF failure 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 11 in Background Section of the main text), then $\hat{\theta_{up}}(t_{k/k})=\hat{\theta_{up}}(t_{k/k-1})$ for all step from t_{k-1} to t_k. ∎

S5.1 Example

To illustrate this Theorem, we show the JEKF failure with the condition and results used in the example S4.1 of the Lemma S4.

Then, let's consider the following:

- the UMM of Case S1.1.
- **H**=[1 0 0 0 0 0] and $\mathbf{K}_k = [\mathbf{K}_{X_v}, \mathbf{K}_N, \mathbf{K}_{MP}, \mathbf{K}_{\mu_{X_v}}, \mathbf{K}_{\mu_N}, \mathbf{K}_{\mu_{mp}}]^T$ as obtained in the proof S4, where $\mathbf{K}_{\mu_{mp}} = 0$.
- $\mathbf{Z}_k$ as measured value of X_v .
- $\hat{\psi}(t_{k/k-1})_{case1} = [\hat{X}_v, \hat{N}, \hat{MP}, \hat{\mu_{X_v}}, \hat{\mu_N}, \hat{\mu_{mp}}]^T$.

Now, using the Equation 22 to provide the estimated mean of the state variable vector $\hat{\psi}(t_{k/k})_{case1}$, we have

$$\hat{\psi}(t_{k/k})_{case1} = \hat{\psi}(t_{k/k-1})_{case1} + \mathbf{K}_k(\mathbf{Z}_k - \mathbf{H}\hat{\psi}(t_{k/k-1})_{case1}) \tag{59}$$

$$\hat{\psi}(t_{k/k})_{case1} = \begin{bmatrix} \hat{X}_v \\ \hat{N} \\ \hat{MP} \\ \hat{\mu_{X_v}} \\ \hat{\mu_N} \\ \hat{\mu_{mp}} \end{bmatrix} + \begin{bmatrix} \mathbf{K}_{X_v} \\ \mathbf{K}_N \\ \mathbf{K}_{MP} \\ \mathbf{K}_{\mu_{X_v}} \\ \mathbf{K}_{\mu_N} \\ \mathbf{K}_{\mu_{mp}} \end{bmatrix} .(\mathbf{Z}_k - X_v) = \begin{bmatrix} \hat{X}_v + \mathbf{K}_{X_v}.(\mathbf{Z}_k - \hat{X}_v) \\ \hat{N} + \mathbf{K}_N.(\mathbf{Z}_k - \hat{X}_v) \\ \hat{MP} + \mathbf{K}_{MP}.(\mathbf{Z}_k - \hat{X}_v) \\ \hat{\mu_{X_v}} + \mathbf{K}_{\mu_{X_v}}.(\mathbf{Z}_k - \hat{X}_v) \\ \hat{\mu_N} + \mathbf{K}_{\mu_{mp}}.(\mathbf{Z}_k - \hat{X}_v) \\ \hat{\mu_{mp}} + 0 \end{bmatrix} \tag{60}$$

Then, we have the estimated mean $\hat{\psi}(t_{k/k})_{case1}$ of the unshared parameter is equal to the predicted mean $\hat{\psi}(t_{k/k-1})_{case1}$ of unshared parameter for all step from t_{k-1} to t_k. In another words, the JEKF failure to perform the parameter evolution, since $\hat{\mu_{mp}}(t_{k/k}) = \hat{\mu_{mp}}(t_{k/k-1})$ all step from t_{k-1} to t_k.

S6 Proof of SANTO (Proposed approach to avoid the JEKF failure)

The SANTO approach prevents the JEKF failure by only adding a small positive quantity (λ) to the $P_{MSV,UP}(t=0)$ in $\mathbf{P}(t=0)$ to initialize the MRDE with a specific initial condition and prevent the Kalman gain being zero in the entire execution of JEKF.

Proof. The proof of SANTO approach can be done using the conditions described previously in the proof of the Lemma S4 and the Theorem S5.

Then, let consider the following:

- A small positive quantity λ.
- x_{msv} as the unique measured state variable (MSV) and **H** = [1 0 ... 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.
- An 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

$$\mathbf{P}_{santo}(t=0) = \begin{bmatrix} P_{x_{msv},x_{msv}}(t=0) & 0 & \dots & 0 & 0 \\ 0 & P_{x_2,x_2}(t=0) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & P_{n,n}(t=0) & 0 \\ P_{x_{msv},\theta_{up}}(t=0) = \lambda & 0 & \dots & 0 & P_{\theta_{up},\theta_{up}}(t=0) \end{bmatrix}, \tag{61}$$

Now, using this Equation 41 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

$$\mathbf{P}(t_{k/k-1}) = \begin{bmatrix} \vdots & \vdots & \iddots \\ P_{x_{msv},\theta_{up}}(t_{k/k-1}) = \lambda & P_{x_2,\theta_{up}}(t_{k/k-1}) & \dots \end{bmatrix}. \tag{62}$$

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 R to compute the Kalman gain for all variables in the state variable vector $\psi(t)$ (Equation 37), we have

$$\mathbf{K}_k = \mathbf{P}(t_{k|k-1})\mathbf{H}^T(\mathbf{H}\mathbf{P}(t_{k|k-1})\mathbf{H}^T + \mathbf{R})^{-1} = \begin{bmatrix} \vdots \\ K_{\theta_{up}} \end{bmatrix} = \begin{bmatrix} \vdots \\ \frac{P_{x_{msv},\theta_{up}}(t_{k/k-1})}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \end{bmatrix} = \begin{bmatrix} \vdots \\ \frac{\lambda}{P_{x_{msv},x_{msv}}(t_{k/k-1})+R} \end{bmatrix}. \tag{63}$$

Then, we have the kalman gain value for the unshared parameter as $K_{\theta_{up}} = \lambda(P_{x_{msv},x_{msv}}(t_{k/k-1}) + R)^{-} \neq 0$, and consequently the predicted state error covariance $P_{x_{msv},\theta_{up}}(t_{k/k-1})$ can be updated by Equation 23 and predicted mean of the state variable vector regards to UP $\hat{\theta_{up}}(t_{k/k-1})$ can be updated Equation 22. 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 JEKF. ■

S7 Related work: Approach KPH2

In this section, we describe the approach KPH2[6] that can to side-step JEFK failure. The authors did not give details about the approach, because the focus was to report application developed to monitoring a rAAV production that is a new bioprocess. Basically, the KPH2 tries to prevent the Kalman gain value regards to an unshared parameter from being zero. Because Kalman gain value equal to zero resulted from an low process uncertainty would mean that the prediction regarding the unshared parameter is perfect and does not need the influence of the measurement in the correction step of JEKF since there is no uncertainty in the prediction regarding the unshared parameter. This is an unrealistic situation and therefore, there is the need to increase the Process uncertainty, $\mathbf{K}_k \approx \frac{Process\ _{Uncertainty}}{System\ _{Uncertainty}}$ to obtain $|K| > 0$ and enable the predicted unshared parameter to be corrected by the influence of residual error in JEKF algorithm. The KPH2 approach tries to fulfill this need by adding more information about the prior error covariances regard to an unshared parameter in two update steps: i) in the Kalman gain computation (Equation 21) and ii) in the update of predicted state error covariance matrix (Equation 23) of EKF algorithm. In the following, we describe the details of this.

The Process uncertainty (Equation 24) is only composed of prior error covariance related to the measured state variable and this information is "incomplete" with regards to unshared parameter, $Cov(MSV, UP) = 0$ in the initial condition $\mathbf{P}$(t=0) and in the predictions during the process of execution of JEKF algorithm. However, the prior error covariances related to unshared parameter $\mathbf{P}_{i,UP}(t_{k|k-1})$ are informations that are already available in the $\mathbf{P}(t_{k|k-1})$ and can be easily extracted from it. Then, an approach to increase process uncertainty (in Equation 24) is to add the Prior error covariances of unshared parameter $\mathbf{P}_{i,UP}(t_{k|k-1})$ to $\mathbf{P}(t_{k|k-1})\mathbf{H}^T$ in Equation 21 as following. Given that $\mathbf{P}(t_{k|k-1})\mathbf{H}^T$ is a vector with prior error covariances of each state variable (SV) with measured state variable (MSV)

$$\mathbf{P}(t_{k|k-1})\mathbf{H}^T = \begin{bmatrix} Cov(SV_i, MSV) \\ \vdots \\ Cov(SV_n, MSV) \end{bmatrix}, \tag{64}$$

and $\mathbf{P}_{i,UP}(t_{k|k-1})$ is also a vector with prior error covariances of each state variable (SV) with all unshared parameter to be estimated

$$\mathbf{P}_{i,UP}(t_{k|k-1}) = \begin{bmatrix} Cov(SV_i, \sum_i^j UP_i) \\ \vdots \\ Cov(SV_n, \sum_i^j UP_i) \end{bmatrix}, \tag{65}$$

we have that the sum of $\mathbf{P}_{i,UP}(t_{k|k-1})$ and $\mathbf{P}(t_{k|k-1})\mathbf{H}^T$ in the equation 21 is

$$\mathbf{K}_k = (\mathbf{P}(t_{k|k-1})\mathbf{H}^T + \mathbf{P}_{i,UP}(t_{k|k-1}))(\mathbf{HP}(t_{k|k-1})\mathbf{H}^T + \mathbf{R})^{-1}. \tag{66}$$

Since, all information that we need to perform $\mathbf{P}(t_{k|k-1})\mathbf{H}^T + \mathbf{P}_{i,UP}(t_{k|k-1})$ are available in $\mathbf{P}(t_{k|k-1})$, we can apply a specific linear operator $\mathbf{H}_2$ (row vector) to $\mathbf{P}(t_{k|k-1})$ to extract all information easily. We need only to defining $\mathbf{H}_2$ with the state variable vector "position" of MSV and UP. For example, if the position of MSV and UP in the state variable vector, $\psi(t) = [MSV, x_2, x_3, x_4, UP]$, is 1st and 5th, we have $\mathbf{H}_2$ =[1 0 0 0 1]. Then, the final version of the Equation 66 is

$$\mathbf{K}_k = \mathbf{P}(t_{k|k-1})\mathbf{H}_2^T(\mathbf{HP}(t_{k|k-1})\mathbf{H}^T + \mathbf{R})^{-1}. \tag{67}$$

In theory, this Equation 67 could prevent $\mathbf{K} = 0$. However, in the case of too many unshared parameters to be estimated, it can unbalance the ratio in the Equation 24, since the Process uncertainty (PU) can become too large in relation to the System uncertainty (SU) in the entire process of execution of JEKF algorithm. Then, to try to preserve a more realistic ratio between PU and SU, the PU that compose the SU can be increased. Since $SU = PU + Measurement_{uncertainty} = \mathbf{HP}(t_{k|k-1})\mathbf{H}^T + \mathbf{R}$[22]. Given this, an approach is add the total sum of all prior error covariances of unshared parameters to be estimated $\mathbf{P}_{UP_{total}}(t_{k|k-1})$ to $\mathbf{HP}(t_{k|k-1})\mathbf{H}^T$ that is part of the system uncertainty (Equation 67) as following. Given that $\mathbf{HP}(t_{k|k-1})\mathbf{H}^T$ results in a scalar representing prior error variance of MSV

$$\mathbf{HP}(t_{k|k-1})\mathbf{H}^T = Cov(MSV, MSV), \tag{68}$$

and $\mathbf{P}_{UP_{total}}(t_{k|k-1})$ is also a scalar, but representing the total sum of all prior error covariances related with the unshared parameters to be estimated

$$\mathbf{P}_{UP_{total}}(t_{k|k-1}) = Cov(MSV, \sum_i^n UP_i) + \sum_j^n Cov(UP_j, MSV + \sum_i^n UP_i), \tag{69}$$

we have that the sum of $\mathbf{P}_{UP_{total}}(t_{k|k-1})$ and $\mathbf{HP}(t_{k|k-1})\mathbf{H}^T$ in the equation 67 is

$$\mathbf{K}_k = (\mathbf{P}(t_{k|k-1})\mathbf{H}^T + \mathbf{P}_{i,UP}(t_{k|k-1}))(\mathbf{HP}(t_{k|k-1})\mathbf{H}^T + \mathbf{P}_{UP_{total}}(t_{k|k-1}) + \mathbf{R})^{-1}. \tag{70}$$

Since all information that we need to perform the sum of $\mathbf{P}_{UP_{total}}(t_{k|k-1})$ and $\mathbf{HP}(t_{k|k-1})\mathbf{H}^T$ are available in $\mathbf{P}(t_{k|k-1})$. We can apply a specific linear operator $\mathbf{H}_2$ to $\mathbf{P}(t_{k|k-1})$ to extract all information easily. Then, the final version of the Equation 70 is

$$\mathbf{K}_k = \mathbf{P}(t_{k|k-1})\mathbf{H}_2^T(\mathbf{H}_2\mathbf{P}(t_{k|k-1})\mathbf{H}_2^T + \mathbf{R})^{-1}. \tag{71}$$

The Equation 71 tries to preserve a more realistic ratio between PU and SU by increasing the PU that compose the SU. However, if the kalman gain continue being small, the $\mathbf{P}(t_{k|k-1})$ could be updated slowly by the Equation 23. Because $\mathbf{K}_k\mathbf{HP}(t_{k|k-1})$ is the factor that update $\mathbf{P}(t_{k|k-1})$ and it is totally dependent of $\mathbf{K}$, as we can see in

$$\mathbf{K}_k\mathbf{HP}(t_{k|k-1}) = \begin{bmatrix} Cov(MSV, SV_1).K_1, \dots, Cov(MSV, SV_n).K_1, \\ \vdots \\ Cov(MSV, SV_1).K_n, \dots, Cov(MSV, SV_n).K_n, \end{bmatrix}. \tag{72}$$

Them, an artefact to avoid a possible slow update of $\mathbf{P}(t_{k|k-1})$ can be to add the Prior error covariance related to unshared parameter $\mathbf{P}_{i,UP}(t_{k|k-1})$ to the $\mathbf{K}_k\mathbf{HP}(t_{k|k-1})$ in the Equation 23 as the following way

$$\mathbf{K}_k\mathbf{H}_2\mathbf{P}(t_{k|k-1}) = \begin{bmatrix} (Cov(MSV, SV_1) + Cov(\sum_i^j UP_i, SV_1)).K_1, \dots, (Cov(MSV, SV_n) + Cov(\sum_i^j UP_i, SV_n)).K_1, \\ \vdots \\ (Cov(MSV, SV_1) + Cov(\sum_i^j UP_i, SV_1)).K_n, \dots, (Cov(MSV, SV_n) + Cov(\sum_i^j UP_i, SV_n)).K_n, \end{bmatrix} \tag{73}$$

$$\mathbf{P}(t_{k|k}) = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_2)\mathbf{P}(t_{k|k-1}), \tag{74}$$

where $\mathbf{H}_2$ is the same linear operator used in Equation 71 .

S8 Empirical Evaluation - Extension

S8.1 Synthetic dataset development - mAb production

The Synthetic dataset (SD) is composed of two runs (A-SD and B-SD) with different conditions. The runs have different samples regarding the state variables Xv, GLC, GLN, LAC, AMM, and mAb and were generated using the UMM case S1.4 with two set of different parameters. These parameters are presented in the Table S2 and were obtained with the the "in silico" runs 41 and 91 presented in[2] and Bayesian inference in the same way done in[6]. A-SD and B-SD were generated with the same initial concentration regards the state variables (see Table S3) and with a sample rate of 3 minutes during 336 hours by simulation. Run A-SD has a maximum of mAb (Titer) equal to 671.0 mg/L with pH varying from 7.1 to 6.7 and temperature constant and equal to 36º C. On the other hand, run B-SD has a maximum of mAb (Titer) equal to 930.9 mg/L with pH and temperature varying from 7.1 to 7 and from 37.5º C to 35º C, respectively. In the Figure S1 we can view the runs A-SD and B-SD that were generated.

Table S2. Parameters used in UMM case S1.4 to generate the runs A-SD and B-SD of Synthetic Dataset (SD).

Parameter	Name	Values of run B-SD	Values of run A-SD
$\mu_{max}(h^{-})$	Maximum growth rate	0.35	0.0777599
$k_{glc}(mM)$	Monod constant glucose	0.2	0.01
$k_{gln}(mM)$	Monod constant glutamine	0.02	0.01
$k_{Ilac}(mM)$	Monod constant lactate for inhibition	3.0	6.0
$k_{Iamm}(mM)$	Monod constant ammonium for inhibition	11.5	15.0
$k_{d}(h^{-})$	Maximum death rate	0.07	0.07
$k_{Dlac}(mM)$	Monod constant lactate for death	45.0	45.0
$k_{Damm}(mM)$	Monod constant ammonium for death	13.0	13.0
$a1(10^{-12} mmols\ cells^{-}\ h^{-})$	Coefficient for m_{gln}	0.01	0.0221436
$a2(mM)$	Coefficient for m_{gln}	5.0	6.0
$Y_{X/glc}(10^{9} cells\ mmol^{-})$	Yield coefficient cell conc./glucose	4.0	3.0
$m_{glc}(10^{-12} mmols\ cells^{-}\ h^{-})$	Glucose maintenance coefficient	0.037	0.05
$Y_{X/gln}(10^{9} cells\ mmol^{-})$	Yield coefficient cell conc./glutamine	1.47864	2.07621
$Y_{lac/glc}(mmol\ mmol^{-})$	Yield coefficient lactate/glucose	9.0	8.0
$Y_{amm/gln}(mmol\ mmol^{-})$	Yield coefficient ammonium/glutamine	40.1	30.5
$r_{amm}(10^{-12} mmol\ cells^{-}\ h^{-})$	Ammonium removal rate	1.22174e-13	7.18176e-13
$Q_{mAb}(10^{-12} g\ cells^{-}\ h^{-})$	Specific production rate	1.00487	0.79

Table S3. Initial conditions of state variables of UMM case S1.4 for the JEKF test with Synthetic Dataset.

State Variable	Name	Value
Xv	Viable cells	0.2×10^{6} c/mL
GLC	Glucose	60 mM
GLN	Glutamine	8 mM
LAC	Lactate	0.1 mM
AMM	Ammonium	0.1 mM
mAb	Monoclonal Antibody (titer)	0 mg/L
QmAb	Specific production rate of mAb	$0.79 \times 10^{12}\ g\ cells^{-1} h^{-1}$

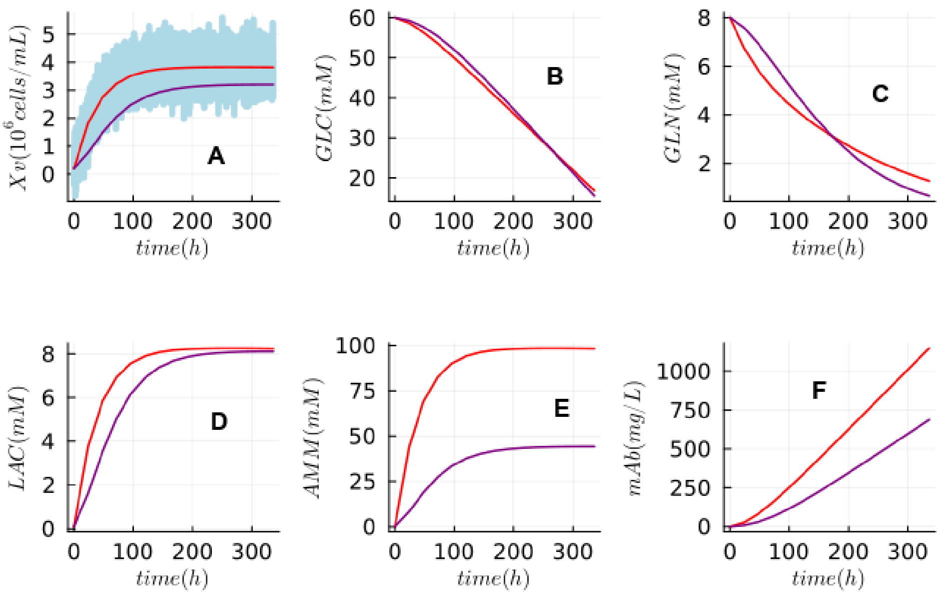

Figure S1. Synthetic dataset regards to mAb production. The curves in red represent the run B-SD and the curves in purple represent the run A-SD. The X_v with noise is highlighted in light blue in the first plot.

S8.2 Real dataset development - rAAV production

The details of Real dataset (RD) can be found in[6].

S8.3 NSEs (JEKF-classic, JEKF-SANTO and JEKF-KPH2) design to address RQ1-G1 and RQ2-G2

The process model (based on UMM case S1.4) and joint state variable vector used by JEKF-Classic, JEKF-SANTO and JEKF-KPH2 are the following:

$$\psi(t)_{case4} = [X_V, GLC, GLN, LAC, AMM, mAb, QmAb]^T, \tag{75}$$

and

$$\frac{d}{dt}\begin{bmatrix} X_V \\ GLC \\ GLN \\ LAC \\ AMM \\ mAb \\ QmAb \end{bmatrix} = \begin{bmatrix} f_{X_V} \\ f_{GLC} \\ f_{GLN} \\ f_{LAC} \\ f_{AMM} \\ f_{mAb} \\ 0 \end{bmatrix} + \omega(t). \tag{76}$$

The standard and specific **P**(t=0) that were used to address the RQ1-G1 and RQ2-G2 are in Tables S4, S5 and S6. Standard **P**(t=0) for the NSEs were obtained following $\mathbf{P}(t=0) = diag((\psi_{case4}(t=0) - \hat{\psi}_{case4}(t=0))(\psi_{case4}(t=0) - \hat{\psi}_{case4}(t=0))^T)$ as done in[4,6]. Then, we have $\mathrm{P_{QmAb,QmAb}} = (1.00487 - 0.79)^2 = 0.0462(g\ cells^{-1}h^{-1})^2$, see Table S4. Since $\mathrm{P_{X_v,QmAb}}$ is a off-diagonal element, we defined it as 1/4 of $\mathrm{P_{QmAb,QmAb}}$ $(\mathrm{g\ cells^{-1}h^{-1}})^2$.

Then, $\mathrm{P_{X_v,QmAb}} = 0.0462 * 1/4 = 0.01155\ (\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$. On the other hand, the specific **P**(t=0) for the NSEs were obtained by trial and error. Furthermore, the **R** and **Q** used by the NSEs are presented in Table S7. It is important point out that all NSEs used a standard **Q**. This means that they used same **Q** perform the estimations to address the RQ1-G1 and and RQ2-G2.

Table S4. Standard initial state error covariance matrix (standard **P**(t=0)) for JEKF-Classic, JEKF-KPH2 and JEKF-SANTO with Synthetic Dataset.

Parameter	Name	$\mathbf{P}_{i,i}$ for JEKF-Classic and JEKF-KPH2 in MRDE-PC and MRDE-PU	$\mathbf{P}_{i,i}$ for JEKF-SANTO in MRDE-PC and MRDE-PU
$\mathrm{P_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	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	0.0462	0.0462
$\mathrm{P_{X_v,QmAb}}$ $(\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$	Initial $Cov(X_v, QmAb)$	0.0	0.01155

Table S5. Specific initial state error covariance matrix (specific **P**(t=0)) for JEKF-KPH2 with Synthetic Dataset.

Parameter	Name	Value in MRDE-PU	Value in MRDE-PC
$\mathrm{P_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	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	0.0212	0.0252

Table S6. Specific initial state error covariance matrix (specific **P**(t=0)) for JEKF-SANTO with Synthetic Dataset

Parameter	Name	Value in MRDE-PC	Value in MRDE-PU
$\mathrm{P_{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	$Cov(x_v, x_v)$	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	0.0462	0.0462
$\mathrm{P_{X_v,QmAb}}$ $(\mathrm{c^2/mL^2})(\mathrm{g\ cells^{-1}h^{-1}})$	Initial $Cov(X_v, QmAb)$	0.01058	0.0118

Table S7. Measurement noise variance **R** and error covariance matrix of process model (standard **Q**) for the JEKF-Classic, JEKF-SANTO and JEKF-KPH2 with SD.

Parameter	Name	Value in MRDE-PU and MRDE-PC
R (c^2/mL^2)	Viable cells MNV [1]	0.5
Q_{X_v,X_v} (c^2/mL^2)	Viable cells PNV [2]	0.005
$Q_{GLC,GLC}$ (mM^2)	Glucose PNV	0.005
$Q_{GLN,GLN}$ mM^2	Glutamine PNV	0.005
$Q_{LAC,LAC}$ (mM^2)	Lactate PNV	0.005
$Q_{AMM,AMM}$ (mM^2)	Ammonium PNV	0.005
$Q_{mAb,mAb}$ (VG^2/mL^2)	Monoclonal Antibody (titer) PNV	0.005
$Q_{QmAb,QmAb}$ (h^{-2})	Specific production rate of mAb	0.0462×10^{-4}

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

S8.4 NSEs (JEKF-SANTO and JEKF-KPH2) design to address RQ3-G2

The process model (based on UMM of Section S1.5) and joint state variable vector used by JEKF-SANTO and JEKF-KPH2 to address the RQ3-G2 are the following:

$$\psi(t) = [X_v, GLC, GLN, LAC, AMM, AAV, \mu_{X_v}, \mu_{Glc}, \mu_{Gln}, \mu_{Lac}, \mu_{Amm}, k_{deg}, \mu_{AAV}]^T. \tag{77}$$

and

$$\frac{d\psi(t)}{dt} = \phi(\psi(t), t) + \omega(t), \tag{78}$$

$$\frac{d}{dt}\begin{bmatrix} X_V \\ Glc \\ Gln \\ Lac \\ Amm \\ AAV \\ \mu_{X_v} \\ \mu_{Glc} \\ \mu_{Gln} \\ \mu_{Lac} \\ \mu_{Amm} \\ k_{deg} \\ \mu_{AAV} \end{bmatrix} = \begin{bmatrix} \mu_{X_v} X_V \\ -\mu_{Glc} X_V \\ -\mu_{Gln} X_V \\ \mu_{Lac} X_V \\ \mu_{Amm} X_V + k_{deg} Gln \\ \mu_{AAV} X_V \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \omega(t). \tag{79}$$

The standard and specific **P**(t=0) that were used by the NSes to address the RQ3-G2 are in Tables S10, S11 and S12. The specific **P**(t=0) for JEKF-KPH2 with MRDE-PC and specific **Q** come from Table 5 of article of Iglesias et al[6]. Furthermore, the **R**, and standard and specific **Q** used by the NSEs are presented in Table S13. The specific **Q** come from Table 6 of article of Iglesias et al[6].

Table S8. Initial conditions of state variables of UMM case S1.5 for the JEKF-SANTO and JEKF-KPH2 test with run B-RD (Source[6]).

State Variable	Name	run B-RD
Xv	Viable cells	1.0011×10^6 c/mL
GLC	Glucose	26.7219 mM
GLN	Glutamine	4.0299 mM
LAC	Lactate	7.2925 mM
AMM	Ammonium	1.5469 mM
rAAV	rAAV viral titer	0 VG/mL

Table S9. Initial parameters obtained with A-RD for the JEKF-SANTO and JEKF-KPH2 test with run B-RD (Source[6]).

State Variable	run B-RD
μ_{Xv} (h^{-1})	0.0065
μ_{GLC} (mmol 10^{-6}c h^{-1})	0.0973
μ_{GLN} (mmol 10^{-6}c h^{-1})	0.0213
μ_{LAC} (mmol 10^{-6}c h^{-1})	0.0214
μ_{AMM} (mmol 10^{-6}c h^{-1})	0.0001
k_{deg} (h^{-1})	0.0020
μ_{AAV} (10^9 vg/mL h 10^6c)	0.0644

Table S10. Standard initial state error covariance matrix (standard **P**(t=0)) for JEKF-SANTO and JEKF-KPH2 with real dataset.

Parameter unit	Initial error covariance	JEKF-KPH2	JEKF-SANTO
$\mathrm{P}_{\mathrm{X_v,X_v}}$ $(\mathrm{c^2/mL^2})$	$Cov(X_v, X_v)$	0.00	0.00
$\mathrm{P}_{\mathrm{GLC,GLC}}$ $(\mathrm{mM^2})$	$Cov(GLC, GLC)$	0.00	0.00
$\mathrm{P}_{\mathrm{GLN,GLN}}$ $(\mathrm{mM^2})$	$Cov(GLN, GLN)$	0.00	0.00
$\mathrm{P}_{\mathrm{LAC,LAC}}$ $(\mathrm{mM^2})$	$Cov(LAC, LAC)$	0.00	0.00
$\mathrm{P}_{\mathrm{AMM,AMM}}$ $(\mathrm{mM^2})$	$Cov(AMM, AMM)$	0.00	0.00
$\mathrm{P}_{\mathrm{rAAV,rAAV}}$ $(\mathrm{VG^2/mL^2})$	$Cov(rAAV, rAAV)$	0.00	0.00
$\mathrm{P}_{\mu_{\mathrm{Xv}},\mu_{\mathrm{Xv}}}$ $(\mathrm{h^{-2}})$	$Cov(\mu_{Xv}, \mu_{Xv})$	9.0×10^{-8}	9.0×10^{-8}
$\mathrm{P}_{\mu_{\mathrm{GLC}},\mu_{\mathrm{GLC}}}$ (mmol 10^{-12}c $\mathrm{h^{-2}}$)	$Cov(\mu_{GLC}, \mu_{GLC})$	2.5×10^{-5}	2.5×10^{-5}
$\mathrm{P}_{\mu_{\mathrm{GLN}},\mu_{\mathrm{GLN}}}$ (mmol 10^{-12}c $\mathrm{h^{-2}}$)	$Cov(\mu_{GLN}, \mu_{GLN})$	9.61×10^{-6}	9.61×10^{-6}
$\mathrm{P}_{\mu_{\mathrm{LAC}},\mu_{\mathrm{LAC}}}$ (mmol 10^{-12}c $\mathrm{h^{-2}}$)	$Cov(\mu_{LAC}, \mu_{LAC})$	9.61×10^{-6}	9.61×10^{-6}
$\mathrm{P}_{\mu_{\mathrm{AMM}},\mu_{\mathrm{AMM}}}$ (mmol 10^{-12}c $\mathrm{h^{-2}}$)	$Cov(\mu_{AMM}, \mu_{AMM})$	8.41×10^{-10}	8.41×10^{-10}
$\mathrm{P}_{\mathrm{k_{deg},k_{deg}}}$ $(\mathrm{h^{-2}})$	$Cov(k_{deg}, k_{deg})$	9.0×10^{-8}	9.0×10^{-8}
$\mathrm{P}_{\mu_{\mathrm{rAAV}},\mu_{\mathrm{rAAV}}}$ ($\mathrm{vg^2/mL^2\ h^2\ 10^{12}c}$)	$Cov(\mu_{rAAV}, \mu_{rAAV})$	7.29×10^{-6}	7.29×10^{-6}
$\mathrm{P}_{\mathrm{X_v},\mu_{\mathrm{GLC}}}$ (c/mL)(mM)	$Cov(X_v, \mu_{GLC})$	0.00	1.5×10^{-7}
$\mathrm{P}_{\mathrm{X_v},\mu_{\mathrm{LAC}}}$ (c/mL)(mM)	$Cov(X_v, \mu_{LAC})$	0.00	5.766×10^{-8}
$\mathrm{P}_{\mathrm{X_v},\mu_{\mathrm{rAAV}}}$ (c/mL)(VG/mL)	$Cov(X_v, \mu_{rAAV})$	0.00	4.374×10^{-8}

Table S11. Specific initial state error covariance matrix (specific **P**(t=0)) for JEKF-KPH2 with MRDE-PU and MRDE-PC.

Parameter	**Initial error Covariance**	**MRDE-PU with standard Q**	**MRDE-PC with standard Q**	**MRDE-PC with specific Q**
P_{X_v,X_v} (c^2/mL^2)	$Cov(X_v,X_v)$	0.00	0.00	0.00
$P_{GLC,GLC}$ (mM^2)	$Cov(GLC,GLC)$	0.00	0.00	0.00
$P_{GLN,GLN}$ (mM^2)	$Cov(GLN,GLN)$	0.00	0.00	0.00
$P_{LAC,LAC}$ (mM^2)	$Cov(LAC,LAC)$	0.00	0.00	0.00
$P_{AMM,AMM}$ (mM^2)	$Cov(AMM,AMM)$	0.00	0.00	0.00
$P_{rAAV,rAAV}$ (VG^2/mL^2)	$Cov(rAAV,rAAV)$	0.00	0.00	0.00
$P_{\mu_{Xv},\mu_{Xv}}$ (h^{-2})	$Cov(\mu_{Xv},\mu_{Xv})$	9.0×10^{-8}	9.0×10^{-8}	7.92×10^{-7}
$P_{\mu_{GLC},\mu_{GLC}}$ $(mmol\ 10^{-12}c\ h^{-2})$	$Cov(\mu_{GLC},\mu_{GLC})$	9.3×10^{-3}	10.5×10^{-1}	2.56×10^{-5}
$P_{\mu_{GLN},\mu_{GLN}}$ $(mmol\ 10^{-12}c\ h^{-2})$	$Cov(\mu_{GLN},\mu_{GLN})$	9.61×10^{-6}	9.61×10^{-6}	1.05×10^{-5}
$P_{\mu_{LAC},\mu_{LAC}}$ $(mmol\ 10^{-12}c\ h^{-2})$	$Cov(\mu_{LAC},\mu_{LAC})$	8.0×10^{-3}	21.91×10^{-1}	9.59×10^{-6}
$P_{\mu_{AMM},\mu_{AMM}}$ $(mmol\ 10^{-12}c\ h^{-2})$	$Cov(\mu_{AMM},\mu_{AMM})$	8.41×10^{-10}	8.41×10^{-10}	6.71×10^{-10}
$P_{k_{deg},k_{deg}}$ (h^{-2})	$Cov(k_{deg},k_{deg})$	9.0×10^{-8}	9.0×10^{-8}	8.71×10^{-8}
$P_{\mu_{rAAV},\mu_{rAAV}}$ $(vg^2/mL^2\ h^2\ 10^{12}c)$	$Cov(\mu_{rAAV},\mu_{rAAV})$	7.89×10^{-3}	10.29×10^{-1}	4.30×10^{-6}
$P_{X_v,\mu_{GLC}}$ $(c/mL)(mM)$	$Cov(X_v,\mu_{GLC})$	0.00	0.00	0.00
$P_{X_v,\mu_{LAC}}$ $(c/mL)(mM)$	$Cov(X_v,\mu_{LAC})$	0.00	0.00	0.00
$P_{X_v,\mu_{rAAV}}$ $(c/mL)(VG/mL)$	$Cov(X_v,\mu_{rAAV})$	0.00	0.00	0.00

Table S12. Specific initial state error covariance matrix (specific **P**(t=0)) for JEKF-SANTO with MRDE-PU and MRDE-PC.

Parameter	**Initial error covariance**	**Value in MRDE-PU and MRDE-PC**
P_{X_v,X_v} (c^2/mL^2)	$Cov(X_v,X_v)$	0.00
$P_{GLC,GLC}$ (mM^2)	$Cov(GLC,GLC)$	0.00
$P_{GLN,GLN}$ (mM^2)	$Cov(GLN,GLN)$	0.00
$P_{LAC,LAC}$ (mM^2)	$Cov(LAC,LAC)$	0.00
$P_{AMM,AMM}$ (mM^2)	$Cov(AMM,AMM)$	0.00
$P_{rAAV,rAAV}$ (VG^2/mL^2)	$Cov(rAAV,rAAV)$	0.00
$P_{\mu_{Xv},\mu_{Xv}}$ (h^{-2})	$Cov(\mu_{Xv},\mu_{Xv})$	9.0×10^{-8}
$P_{\mu_{GLC},\mu_{GLC}}$ $(mmol\ 10^{-12}c\ h^{-2})$	$Cov(\mu_{GLC},\mu_{GLC})$	2.5×10^{-5}
$P_{\mu_{GLN},\mu_{GLN}}$ $(mmol\ 10^{-12}c\ h^{-2})$	$Cov(\mu_{GLN},\mu_{GLN})$	9.61×10^{-6}
$P_{\mu_{LAC},\mu_{LAC}}$ $(mmol\ 10^{-12}c\ h^{-2})$	$Cov(\mu_{LAC},\mu_{LAC})$	9.61×10^{-6}
$P_{\mu_{AMM},\mu_{AMM}}$ $(mmol\ 10^{-12}c\ h^{-2})$	$Cov(\mu_{AMM},\mu_{AMM})$	8.41×10^{-10}
$P_{k_{deg},k_{deg}}$ (h^{-2})	$Cov(k_{deg},k_{deg})$	9.0×10^{-8}
$P_{\mu_{rAAV},\mu_{rAAV}}$ $(vg^2/mL^2\ h^2\ 10^{12}c)$	$Cov(\mu_{rAAV},\mu_{rAAV})$	7.29×10^{-6}
$P_{X_v,\mu_{GLC}}$ $(c/mL)(mM)$	$Cov(X_v,\mu_{GLC})$	0.002
$P_{X_v,\mu_{LAC}}$ $(c/mL)(mM)$	$Cov(X_v,\mu_{LAC})$	0.0017
$P_{X_v,\mu_{rAAV}}$ $(c/mL)(VG/mL)$	$Cov(X_v,\mu_{rAAV})$	0.0016

Table S13. Measurement noise variance **R**, and standard and specific error covariance matrix of process model $\mathbf{Q}_{i,i}$ for the JEKF-SANTO and JEKF-KPH2.

Parameter	Name	Standard Q for KPH2 and SANTO	Specific Q for KPH2
R (c^2/mL^2)	Viable cells MNV [1]	0.006	0.006
Q_{X_v,X_v} (c^2/mL^2)	Viable cells PNV [2]	0.0006	0.000006
$Q_{GLC,GLC}$ (mM^2)	Glucose PNV	0.0006	0.0006
$Q_{GLN,GLN}$ mM^2	Glutamine PNV	0.0006	0.0006
$Q_{LAC,LAC}$ (mM^2)	Lactate PNV	0.0006	0.0006
$Q_{AMM,AMM}$ (mM^2)	Ammonium PNV	0.0006	0.0006
$Q_{rAAV,rAAV}$ (VG^2/mL^2)	AAV viral titer PNV	0.0006	0.0006
$Q_{\mu_{Xv},\mu_{Xv}}$ (h^{-2})	μ_{Xv} PNV	7.92×10^{-8}	7.92×10^{-8}
$Q_{\mu_{GLC},\mu_{GLC}}$ (mmol 10^{-12}c h^{-2})	μ_{GLC} PNV	1.56×10^{-5}	1.56×10^{-5}
$Q_{\mu_{GLN},\mu_{GLN}}$ (mmol 10^{-12}c h^{-2})	μ_{GLN} PNV	1.05×10^{-5}	1.05×10^{-5}
$Q_{\mu_{LAC},\mu_{LAC}}$ (mmol 10^{-12}c h^{-2})	μ_{LAC} PNV	15.59×10^{-6}	15.59×10^{-6}
$Q_{\mu_{AMM},\mu_{AMM}}$ (mmol 10^{-12}c h^{-2})	μ_{AMM} PNV	0.11×10^{-8}	0.11×10^{-8}
$Q_{k_{deg},k_{deg}}$ (h^{-2})	k_{deg} PNV	0.71×10^{-8}	0.71×10^{-8}
$Q_{\mu_{rAAV},\mu_{rAAV}}$ (vg^2/mL^2 h^2 10^{12}c)	μ_{rAAV} PNV	15.30×10^{-6}	15.30×10^{-6}

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

S8.5 Results with synthetic dataset

The experimental test of the JEKF failure theorem can be seen in the Figures S2, S3 and S4. In all cases the JEKF-Classic could not evolve the QmAb (unshared parameter). It was constant in the entire execution. Consequently, the JEKF-Classic estimation regards to mAb were far from the ground truth, see Table S14. Futhermore, by comparing the results from Figures S2, S3 and S4, we can see that the JEKF-SANTO and JEKF-KPH2 are sensible to **P**(t=0) because the best results were obtained with specific **P**(t=0), see Figure S2.

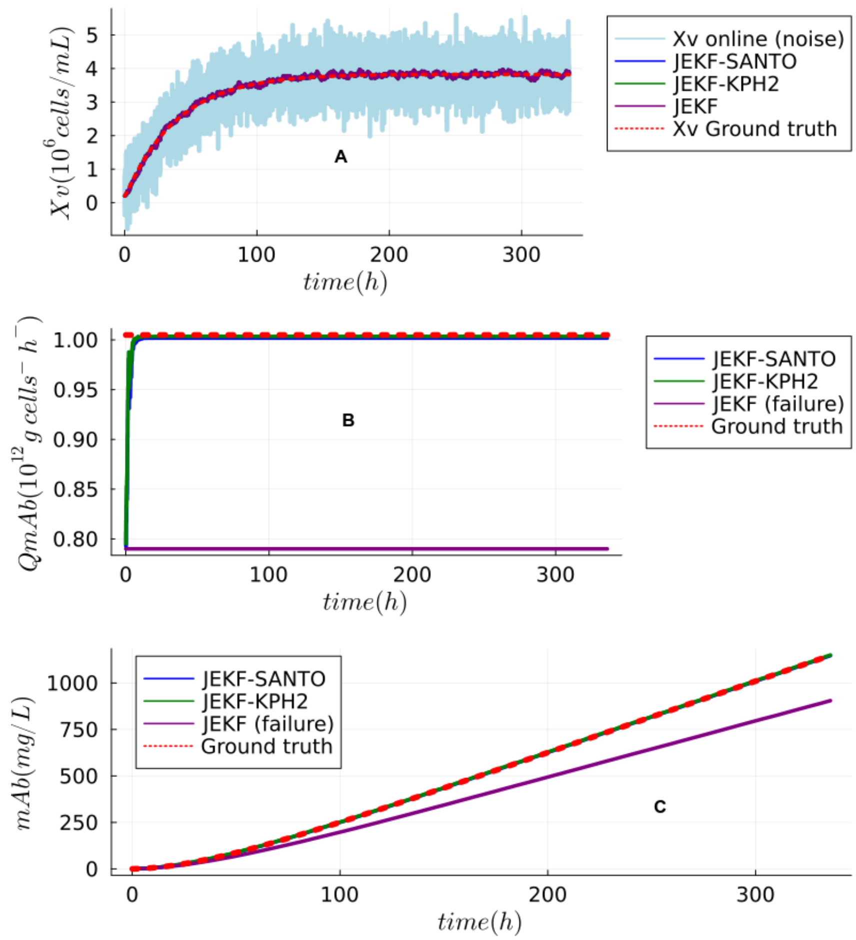

Figure S2. Experimental test that the JEKF-classic cannot avoid the JEKF failure with Synthetic dataset. First, plot A shows the estimations regards Xv, and all estimations were close the ground truth. The plots B and C show the estimations regards the unshared parameter QmAb and mAb (titer), respectively. All NSEs were able to evolve QmAb with convergence to the ground truth value with exception to JEKF-Classic. All NSEs were executed with **MRDE-PU** and **specific** $P_{UP,UP}(t=0)$.

Table S14. RMSE between NSEs estimations about mAb and ground truth of synthetic dataset with standard P(t=0).

NSE	RMSE (MRDE-PU)	RMSE (MRDE-PC)
JEKF-SANTO	3.02	13.47
JEKF-KPH2	68.65	52.11
JEKF-Classic	132.4	132.4

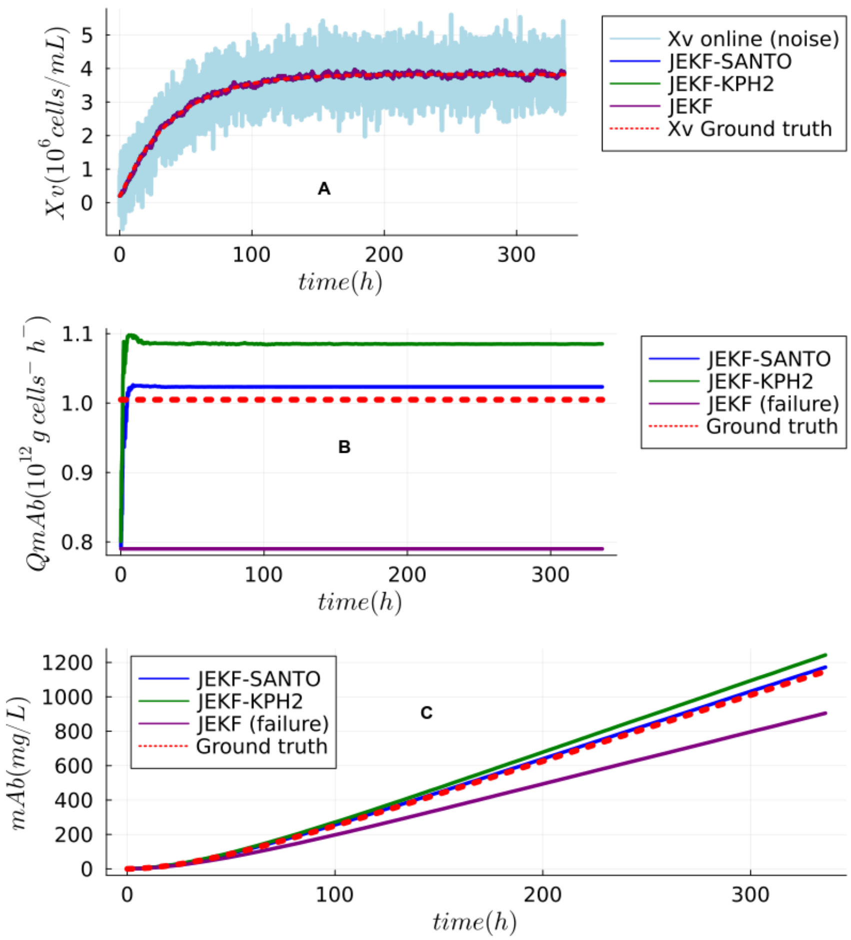

Figure S3. The JEKF-SANTO and JEKF-KPH2 avoid the JEKF failure in SD, but they need an specific $P(t=0)$. First, plot A shows the estimations regards Xv, and all estimations were close the ground truth. The plots B and C show the estimations regards the unshared parameter QmAb and mAb (titer), respectively. The NSEs were executed with **MRDE-PC** and **standard** $P(t=0)$.

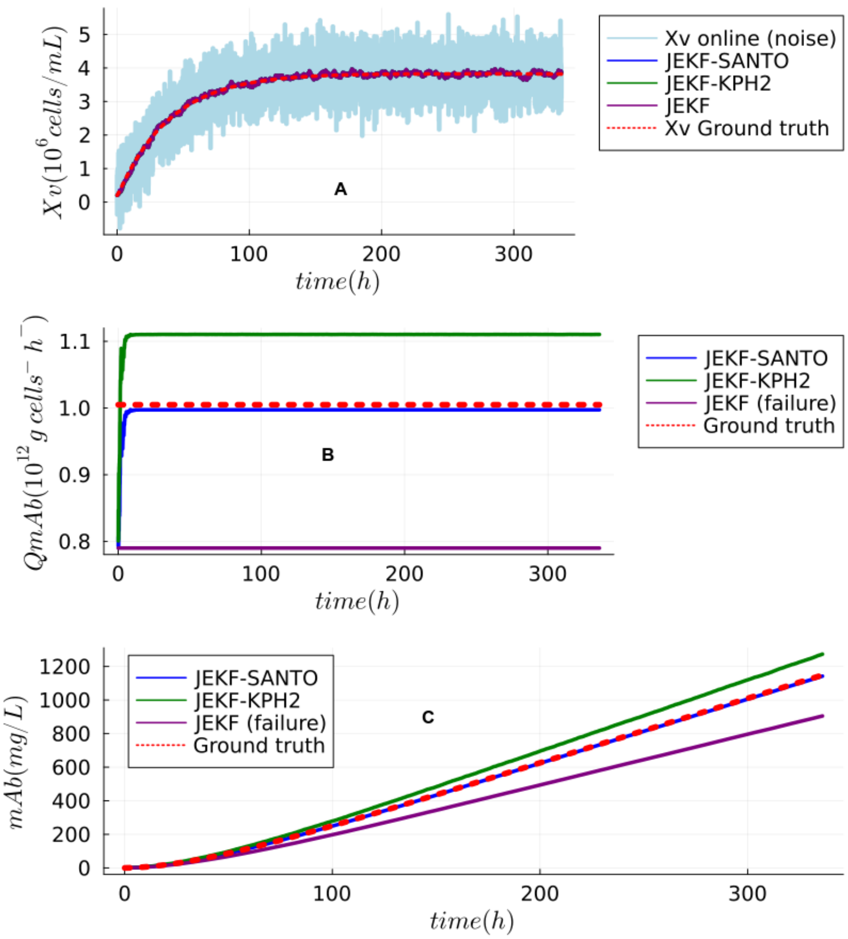

Figure S4. The JEKF-SANTO and JEKF-KPH2 avoid the JEKF failure in SD, but they need an specific $P(t=0)$. First, plot A shows the estimations regards Xv, and all estimations were close the ground truth. The plots B and C show the estimations regards the unshared parameter QmAb and mAb (titer), respectively. The NSEs were executed with **MRDE-PU** and **standard** $P(t=0)$**.**

S8.6 Results with real dataset

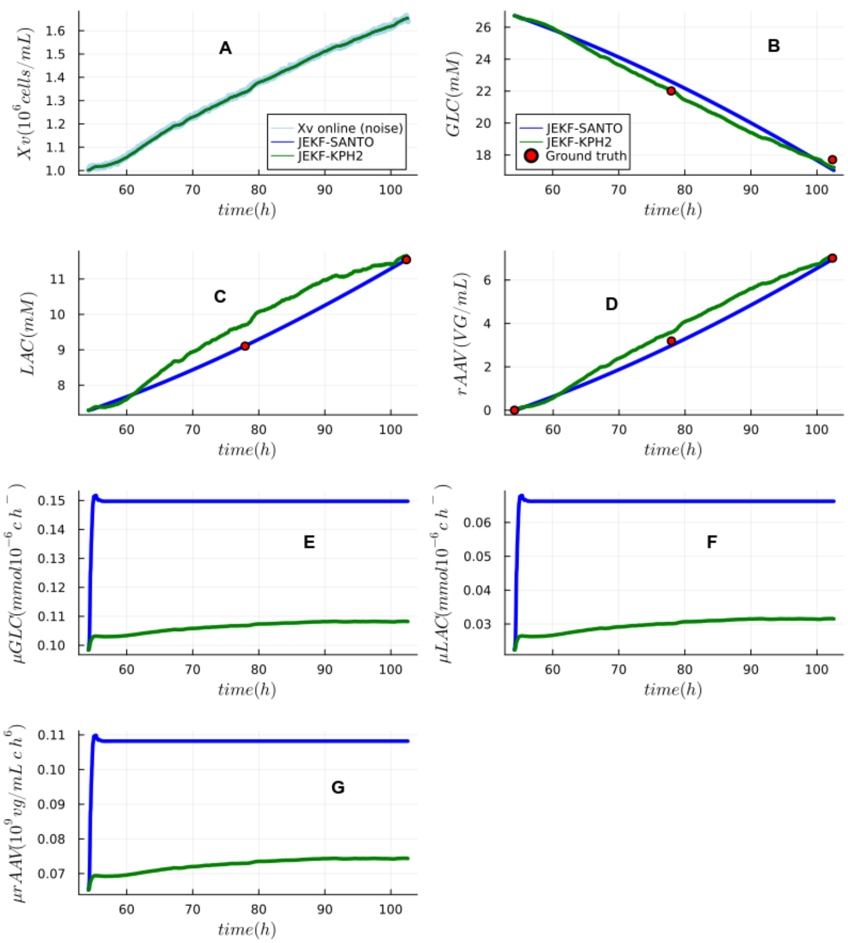

Figure S5. Simultaneous unshared parameters estimation by the JEKF-SANTO and JEKF-KPH2 with real dataset (rAAV production) with MRDE-PU and specific **P**(t=0). Plots A, B, C and D show the estimations for Xv, GLC, LAC and rAAV, respectively. The evolution of the unshared parameters (μGLC, μLAC and $\mu rAAV$) are presented in the plots E, F and G.

Table S15. RMSE between NSEs estimations and ground truth of real dataset with MRDE-PU and specific $\mathbf{P}$(t=0).

Ground truth	JEKF-SANTO	JEKF-KPH2
GLC	0.601	0.354
LAC	0.006	0.430
rAAV (titer)	0.124	0.241

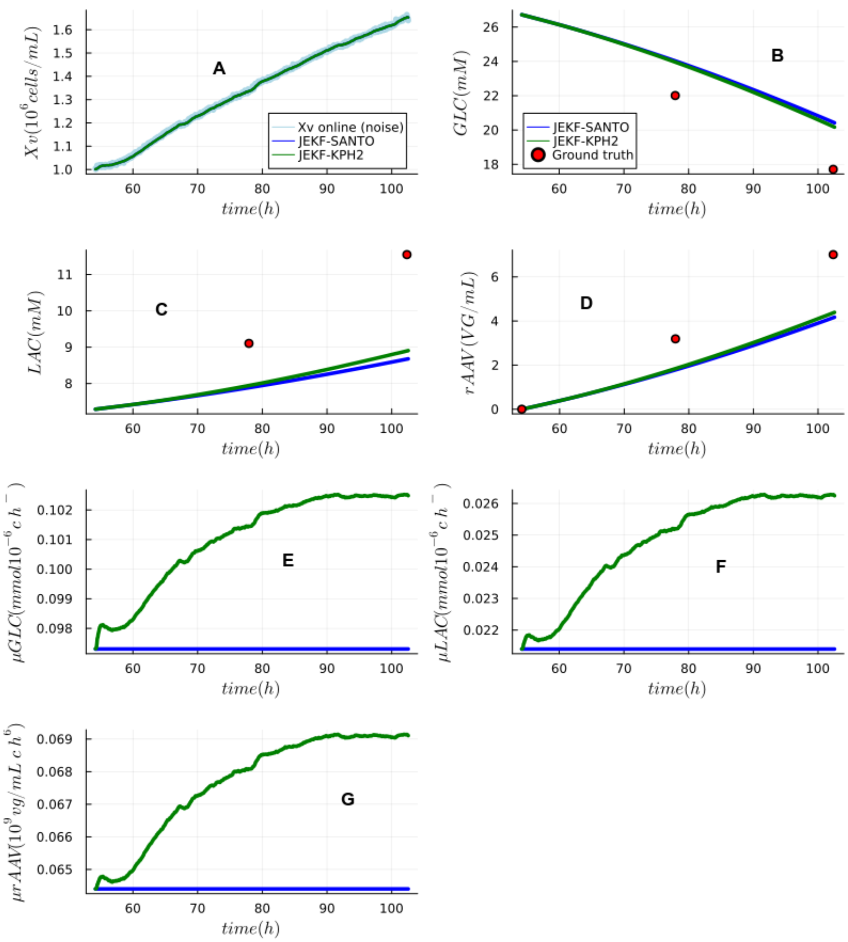

Figure S6. Simultaneous unshared parameters estimation by the JEKF-SANTO and JEKF-KPH2 with real dataset (rAAV production) with MRDE-PC and standard **P**(t=0). Plots A, B, C and D show the estimations for Xv, GLC, LAC and rAAV, respectively. The evolution of the unshared parameters (μGLC, μLAC and $\mu rAAV$) are presented in the plots E, F and G.

Table S16. RMSE between NSEs estimations and ground truth of real dataset with MRDE-PC and standard $\mathbf{P}(t{=}0)$.

Ground truth	JEKF-SANTO	JEKF-KPH2
GLC	2.4086	2.239
LAC	2.207	2.0453
rAAV (titer)	1.840	1.712

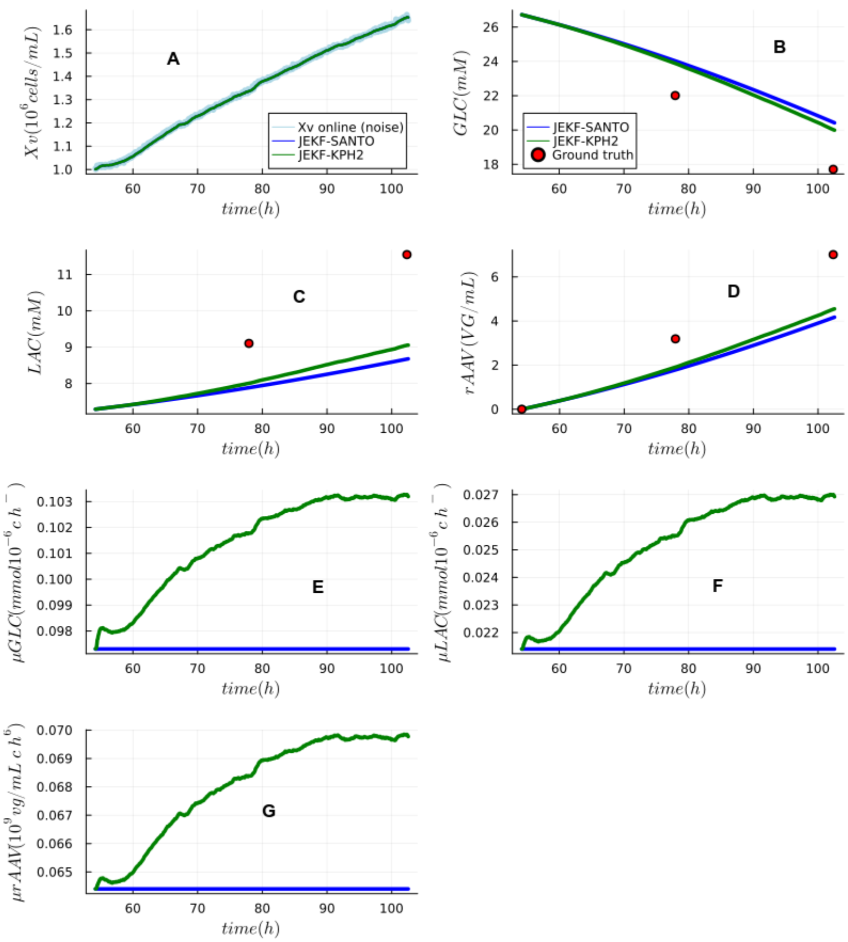

Figure S7. Simultaneous unshared parameters evolution by the JEKF-SANTO and JEKF-KPH2 with real dataset (rAAV production) with MRDE-PU and standard **P**(t=0). Plots A, B, C and D show the estimations for Xv, GLC, LAC and rAAV, respectively. The evolution of the unshared parameters (μGLC, μLAC and $\mu rAAV$) are presented in the plots E, F and G.

Table S17. RMSE between NSEs estimations and ground truth of real dataset with MRDE-PU and standard $\mathbf{P}(t=0)$.

Ground truth	JEKF-SANTO	JEKF-KPH2
GLC	2.408	2.103
LAC	2.207	1.925
rAAV (titer)	1.840	1.610

S8.7 Codes and datasets availability

All source codes and datasets used in this work can be found in GitHub: https://github.com/cristovaoiglesias/JEKF-SANTO.

References

1. Goudar, C. T. Computer programs for modeling mammalian cell batch and fed-batch cultures using logistic equations. *Cytotechnology* **64**, 465–475 (2012).
2. Narayanan, H. *et al.* Hybrid-ekf: Hybrid model coupled with extended kalman filter for real-time monitoring and control of mammalian cell culture. *Biotechnol. Bioeng.* **117**, 2703–2714 (2020).
3. Ohadi, K., Legge, R. L. & Budman, H. M. Development of a soft-sensor based on multi-wavelength fluorescence spectroscopy and a dynamic metabolic model for monitoring mammalian cell cultures. *Biotechnol. bioengineering* **112**, 197–208 (2015).
4. Ji, Z. & Brown, M. Joint state and parameter estimation for biochemical dynamic pathways with iterative extended kalman filter: comparison with dual state and parameter estimation. *The Open Autom. Control. Syst. J.* **2** (2009).
5. Kornecki, M. & Strube, J. Accelerating biologics manufacturing by upstream process modelling. *Processes* **7**, 166 (2019).
6. Iglesias Jr, C. F. *et al.* Monitoring the recombinant adeno-associated virus production using extended kalman filter. *Processes* **10**, 2180 (2022).
7. Jazwinski, A. Stochastic processes and filtering theory. *Math. science engineering (* (1970).
8. Frogerais, P., Bellanger, J.-J. & Senhadji, L. Various ways to compute the continuous-discrete extended kalman filter. *IEEE Transactions on Autom. Control.* **57**, 1000–1004 (2011).
9. Kulikov, G. Y. & Kulikova, M. V. Accurate numerical implementation of the continuous-discrete extended kalman filter. *IEEE Transactions on Autom. Control.* **59**, 273–279 (2013).
10. Albiol, J., Robusté, J., Casas, C. & Poch, M. Biomass estimation in plant cell cultures using an extended kalman filter. *Biotechnol. progress* **9**, 174–178 (1993).
11. Yousefi-Darani, A., Paquet-Durand, O. & Hitzmann, B. The kalman filter for the supervision of cultivation processes. *Digit. Twins* 95–125 (2020).
12. Paquet-Durand, O., Zettel, V., Yousefi-Darani, A. & Hitzmann, B. The supervision of dough fermentation using image analysis complemented by a continuous discrete extended kalman filter. *Processes* **8**, 1669 (2020).
13. Jin, X.-B., Robert Jeremiah, R. J., Su, T.-L., Bai, Y.-T. & Kong, J.-L. The new trend of state estimation: from model-driven to hybrid-driven methods. *Sensors* **21**, 2085 (2021).
14. Zhang, D., Del Rio-Chanona, E. A., Petsagkourakis, P. & Wagner, J. Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization. *Biotechnol. bioengineering* **116**, 2919–2930 (2019).
15. Brockwell, P. Time series analysis. *Encycl. Stat. Behav. Sci.* (2005).
16. Haykin, S. S. & Haykin, S. S. *Kalman filtering and neural networks*, vol. 284 (Wiley Online Library, 2001).
17. Särkkä, S. Bayesian estimation of time-varying systems: Discrete-time systems. *Writ. material for course held Spring* **2012** (2012).
18. Murphy, K. P. *Machine learning: a probabilistic perspective* (MIT press, 2012).

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

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

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

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

23. Benhamou, E. Kalman filter demystified: from intuition to probabilistic graphical model to real case in financial markets. *arXiv preprint arXiv:1811.11618* (2018).

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