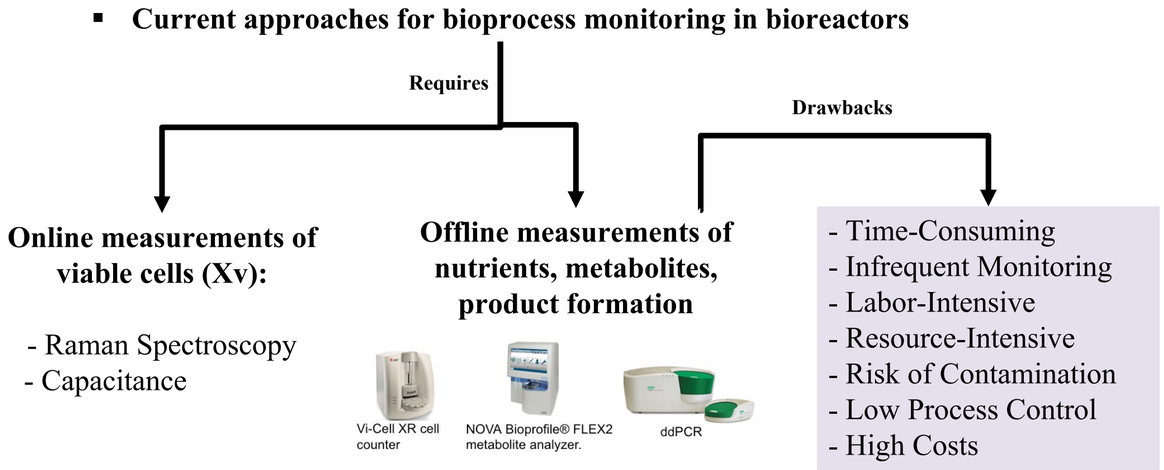
Limitations of Joint and Dual Nonlinear Kalman Estimators in Low-Cost Bioprocess Monitoring

Cristovão F. Iglesias Jr[1], Luis Pessoa[2], Claudio Miceli de Farias[2], Miodrag Bolic[1]
[1]EECS, University of Ottawa, Ottawa, Canada
[2]Federal University of Rio de Janeiro, RJ, Brazil.

Overview

- **Biopharmaceutical industry requires fast and low-cost bioprocess monitoring to improve efficiency, reduce costs, and ensure high-quality bioprocess productions.**

- **Current approaches for bioprocess monitoring in bioreactors**



- **Biomanufacturing conditions for fast and low-cost bioprocess monitoring:**

1. A state variables vector defined by $\mathbf{x} = [x_{msv}, x_2, \ldots, x_n]$.
2. x_{msv} as the unique measured state variable (MSV).
3. A system of nonlinear differential equations representing a UMM of the form:

$$\mathbf{g}(\mathbf{x}, \boldsymbol{\theta}) = \begin{array}{ll} \frac{dx_{msv}}{dt} & = f_1(x_{msv}, x_2, \ldots, x_{n-1}, \theta_1, \theta_2, \ldots, \theta_m), \\ \frac{dx_2}{dt} & = f_2(x_{msv}, x_2, \ldots, x_{n-1}, \theta_1, \theta_2, \ldots, \theta_m), \\ & \vdots \\ \frac{dx_n}{dt} & = f_n(x_{msv}, \theta_{up}) \end{array} \tag{2.31}$$

where x_{msv} and $x_2, \ldots, x_n$ are the variables of the system, $f_1, f_2, \ldots, f_n$ are the functions defining the system represented by $\mathbf{g}(\cdot)$, and $\theta_1, \theta_2, \ldots, \theta_m$ are the parameters of the system, and θ_{up} an unshared parameter.

4. $\mathbf{R}$ as measurement noise variance of x_{msv}.
5. θ_{up} as the unshared parameter (UP) to be evolved (estimated) in real-time. It is important to note that UP is part of a weak term related to a weak variable (x_n).
6. $\mathbf{P}(0)$ and $\mathbf{Q}$ with uncorrelated elements. They are defined as

$$\mathbf{P}(0) = Diag([P_{x_{msv},x_{msv}}, P_{x_2,x_2}, \cdots, P_{n,n}, P_{\theta_{up},\theta_{up}}])$$
$$\mathbf{Q} = Diag([Q_{x_{msv},x_{msv}}, Q_{x_2,x_2}, \cdots, Q_{n,n}, Q_{\theta_{up},\theta_{up}}]).$$

Motivation

Soft-sensor based on Data-Driven models (deep learning models) [1,2] → - Lack of interpretability - Large data requirements

Soft-sensor based on Nonlinear Kalman Estimators (NKE) and Unstructured Mechanistic Models (UMM) → - Joint EKF fails with biomanufacturing conditions described above

Possible solution

Theorem 1 (**S**pecific Initi**A**l co**N**di**T**i**O**n - SANTO): The addition of a quantity (λ) to the $P_{MSV,UP}(0)$ and $P_{UP,MSV}(0)$ (where $P_{MSV,UP} = P_{UP,MSV}$) in P(0) to initialize the matrix Ricatti differential equation with a specific initial condition can prevent the Kalman gain being zero in the entire execution of JEKF and prevent the JEKF failure [3].

Research Questions

RQ1) Can Dual EKF estimate the unshared parameters and the state simultaneously under the biomanufacturing conditions for fast and low-cost bioprocess monitoring?

RQ2) How are the performances of Joint UKF and Joint CKF under the same biomanufacturing conditions?

RQ3) Can the SANTO approach improve the performance of Joint UKF and Joint CKF?

Results

Theorem 2 : The DEKF cannot estimate an unshared parameter (parameter evolution) that is part of a weak term in a UMM if the unshared parameter is not part of the nonlinear function that models the unique state variable measured.

UMM for mAb production

$$\frac{dX_V}{dt} = (\mu - \mu_d)X_V$$
$$\frac{dX_t}{dt} = \mu X_V - k_{lysis}(X_t - X_V)$$
$$\mu = \mu_{max} \cdot \frac{[GLC]}{K_{glc} + [GLC]} \cdot \frac{[GLN]}{K_{gln} + [GLN]} \cdot \frac{K_{Ilac}}{K_{Ilac} + [LAC]} \cdot \frac{K_{Iamm}}{K_{Iamm} + [AMM]}$$
$$\mu_d = \frac{\mu_{d,max}}{1 + (K_{d,amm} + [AMM])^2}$$
$$\frac{d[GLC]}{dt} = -Q_{glc}Xv$$
$$\frac{d[GLN]}{dt} = -Q_{gln}Xv - K_{d,gln}[GLN]$$
$$\frac{d[LAC]}{dt} = Q_{lac}Xv$$
$$\frac{d[AMM]}{dt} = Q_{amm}Xv + K_{d,gln}[GLN]$$
$$Q_{glc}Xv = \frac{\mu}{Y_{x,glc}} + m_{glc}$$
$$Q_{gln}Xv = \frac{\mu}{Y_{x,gln}} + m_{gln} = \frac{\mu}{Y_{x,gln}} + \frac{\alpha_2[GLN]}{\alpha_2 + [GLN]}$$
$$Q_{lac}Xv = Y_{lac,glc}Q_{glc}$$
$$Q_{amm}Xv = Y_{amm,gln}Q_{gln}$$
$$\frac{d[mAb]}{dt} = (2 - \gamma\mu)Q_{mAb}.X_V$$

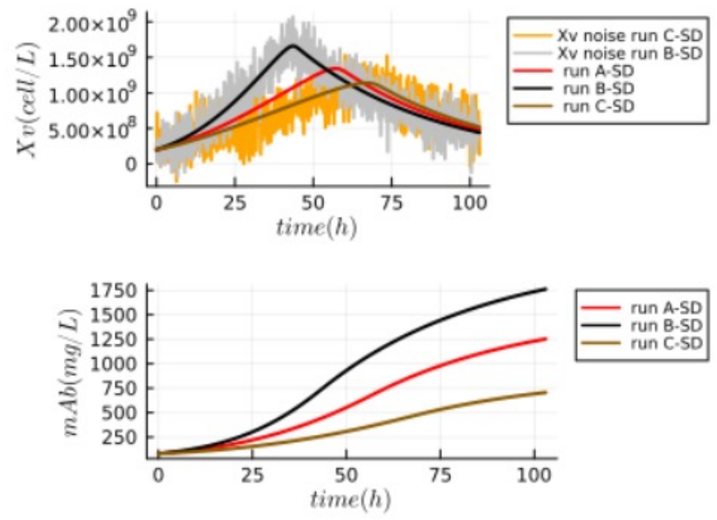


Figure 1. Synthetic dataset regarding the mAb production.

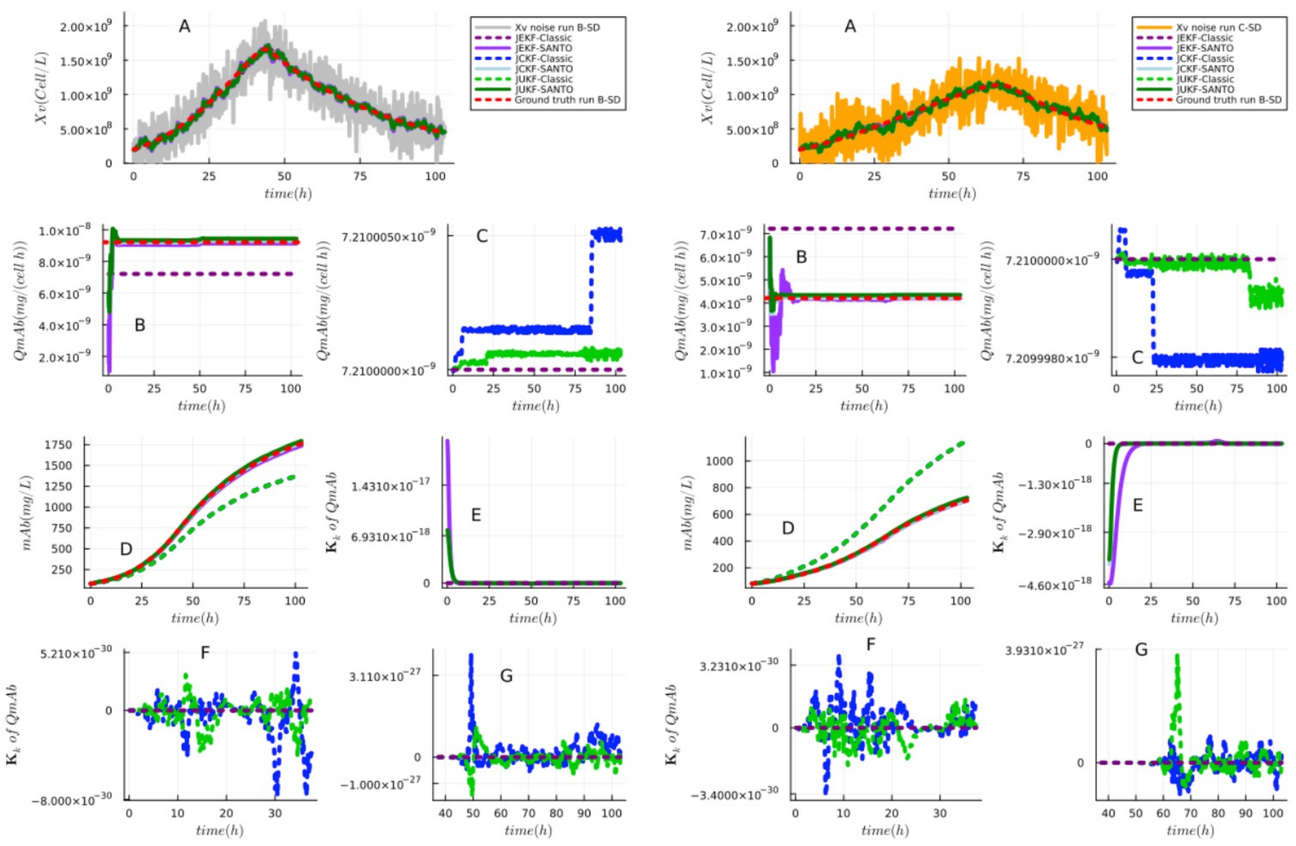


Figure 2. Empirical results with run B-SD.

Figure 3. Empirical results with run C-SD.

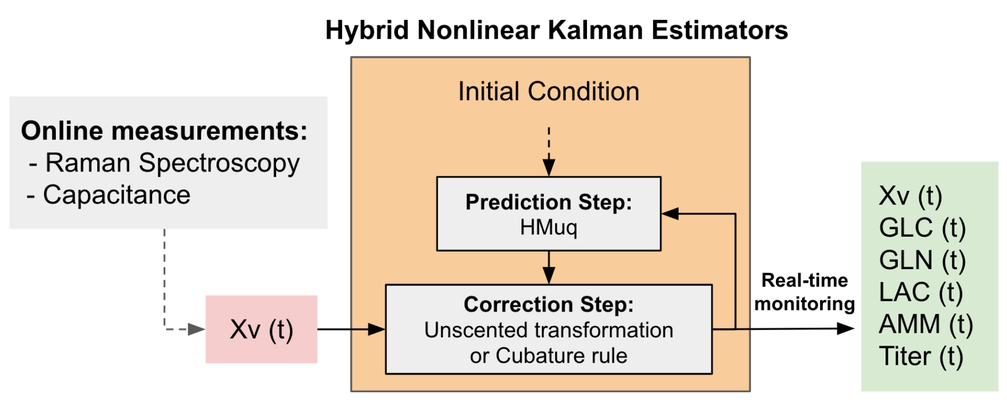
Conclusion

Answering RQ1: No, based on theorem 2**.**

Answering RQ2: Joint and dual NKE are inefficient for fast and low-cost bioprocess monitoring within the studied biomanufacturing conditions.

Answering RQ3: The SANTO approach can improve performance, but in some cases, it leads to unconventional Kalman gain values.

Next Steps



Advantages

- Fast monitoring: estimations in high frequency
- Low-cost monitoring: only one device required
- Reduced tuning requirements
- Auto-initialization of P(0)
- Reduced large data requirements

References

1 - Khuat, Thanh Tung, et al. "Applications of machine learning in antibody discovery, process development, manufacturing and formulation: Current trends, challenges, and opportunities." *Computers & Chemical Engineering* (2024): 108585.
2- Narayanan, Harini, et al. "Hybrid models based on machine learning and an increasing degree of process knowledge: Application to capture chromatographic step." *Industrial & Engineering Chemistry Research* 60.29 (2021): 10466-10478.
3 - Iglesias Jr, Cristovão Freitas, and Miodrag Bolic. "How Not to Make the Joint Extended Kalman Filter Fail with Unstructured Mechanistic Models." *Sensors* 24.2 (2024).