Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

A Multi-Level Hybrid EKF–Machine Learning Soft Sensor for Robust Bioprocess Monitoring

Mohammad Reza Boskabadi[a], Rajiv Kailasanathan[ab], Luis Ricardez-Sandoval[c] and Seyed Soheil Mansouri[a*]

[a] Department of Chemical and Biochemical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[b] Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[c] Department of Chemical Engineering, University of Waterloo, Waterloo, N2L 3G1, ON, Canada
* Corresponding Author: seso@kt.dtu.dk

ABSTRACT

Real-time monitoring of bioprocesses is hindered by sparse, heterogeneous measurements of key biological states, such as biomass, substrate, and product concentrations. Extended Kalman Filter (EKF)–based soft sensors offer a physics-grounded solution but are sensitive to limited observability, sensor bias, and process-model mismatch—conditions common in industrial fermentations. This work proposes a Levelized Hybrid Estimation Architecture (LHEA) that systematically enhances physics-based state estimation through increasing robustness and adaptivity while preserving model transparency and regulatory interpretability. The approach is evaluated using the KTB1 benchmark simulation model for continuous lovastatin production under an industrially realistic, cost-constrained sensor configuration combining dissolved oxygen, biomass proxy, volume measurements, and sparse HPLC product assays. Three estimator levels are investigated: (L1) a baseline EKF, (L2) a bias-augmented EKF for sensor-drift robustness, and (L3) a hybrid EKF with physics-constrained residual learning driven by sparse assays. Results show that L1 provides stable state reconstruction under nominal conditions but is sensitive to bias and model mismatch. L2 effectively isolates measurement bias, while L3 adapts to evolving process kinetics and achieves the lowest estimation errors under mismatch. These findings demonstrate that a structured, levelized integration of machine learning can significantly enhance soft-sensor reliability without sacrificing interpretability, providing a practical pathway toward robust digital twins.

Keywords: Biosystems, Process Monitoring, Hybrid Modelling, Soft Sensor

INTRODUCTION

Real-time monitoring and optimization of biomanufacturing processes remains an open challenge due to the limited availability of reliable online sensors for critical variables such as biomass, substrates, and product concentrations [1]. While measurements such as dissolved oxygen, pH, and flow rates are typically available at high frequency, key quality attributes and performance indicators are often accessible only through infrequent and costly offline assays [2]. In continuous biomanufacturing, long operating horizons, tight quality constraints, and economic pressures further exacerbate these challenges. As a result, soft sensors must operate under sparse, asynchronous, and noisy measurements, where sensor drift, calibration errors, and process disturbances can significantly degrade estimation performance.

Model-based state estimation methods, particularly Extended Kalman Filters (EKFs), have been widely applied to infer unmeasured states in bioprocesses using first-principles dynamic models [3], [4]. When accurate models and informative measurements are available, EKFs can provide effective real-time monitoring of biological systems. However, in realistic biomanufacturing environments, model mismatch arising from uncertain kinetics, unmodelled physiological shifts, and slowly varying parameters is unavoidable. Combined with sensor drift and measurement bias, these effects can lead to filter inconsistency, biased estimates, or even divergence in the estimation [5], [6]. Addressing robustness to both model mismatch and measurement imperfections, therefore, remains a central challenge in the development of

reliable soft sensors for bioprocess monitoring.

Machine-learning (ML)–based soft sensors have attracted increasing attention due to their ability to capture complex nonlinear relationships directly from process data [7]. Despite their predictive power, purely data-driven approaches raise concerns related to interpretability, extrapolation beyond training data, sensitivity to operating conditions, and regulatory acceptance [8]. In regulated biomanufacturing contexts, particularly biopharmaceuticals, the lack of explicit physical structure makes it difficult to guarantee mass-balance consistency, diagnose estimation failures, or ensure reliable performance during process regime changes. These limitations motivate hybrid approaches that combine the transparency and structure of first-principles models with the adaptability of data-driven learning [9].

This study presents a *levelized hybrid Extended Kalman Filter–Machine Learning (EKF–ML) approach* for robust soft sensing in bioprocesses, where robustness refers to the preservation of estimator stability and acceptable estimation accuracy in the presence of sensor drift, measurement sparsity, and structural mismatch between assumed and true process kinetics. The methodology is evaluated using the KTB1 MATLAB/Simulink benchmark, a realistic simulator for continuous biomanufacturing that couples an upstream CSTR bioreactor with downstream operations and incorporates sensor noise, delays, and regulatory control [10]. The main contributions of this work are fourfold:

(i) the development of a *robust soft-sensing model* that enhances state estimation accuracy and reliability under realistic measurement limitations;

(ii) the introduction of a *levelized estimation structure*, enabling systematic investigation of increasing model–data hybridization and facilitating future methodological extensions;

(iii) a *comprehensive case study* demonstrating the practical value of the approach under nominal operation, sensor bias, and process-regime changes; and

(iv) the release of an *open-source implementation* to support transparency, reproducibility, and further research in hybrid soft-sensor design (https://github.com/Boskabadi/LHEA).

METHODOLOGY: LEVELIZED HYBRID ESTIMATION ARCHITECTURE (LHEA)

We consider a nonlinear bioprocess described by a first-principles mass-balance model:

$$\dot{x}(t) = f(x(t), u(t)), \tag{1}$$

where $x(t) \in \mathbb{R}^{n_x}_{\geq 0}$ denotes the process states (e.g., volume, biomass, substrates, product, dissolved oxygen), $u(t) \in \mathbb{R}^{n_x}_{\geq 0}$ represents known non-negative manipulated inputs (e.g., feed flow rates and compositions), and $t \in \mathbb{R}_{\geq 0}$is time. The nonlinear function $f(\cdot)$represents the reactor mass-balance equations derived from conservation laws and kinetic expressions, and is assumed to be continuously differentiable with respect to $x$within the physically admissible operating region. Only a subset of states is measured online at heterogeneous sampling rates, while key quality attributes (e.g., product concentration) are available through sparse laboratory assays, e.g. at the end of the batch. State estimation is performed using an Extended Kalman Filter (EKF) as a baseline estimator. The novelty of this work lies not in the EKF formulation itself, but in a systematic, levelized extension of the estimator to address practical challenges in bioprocess monitoring, i.e. weak observability, sensor drift, and model mismatch.

Level 1 (L1): Physics-based soft sensor

Level 1 implements a standard physics-based EKF using the mechanistic model and available multi-rate measurements. The process dynamics are given by

$$\dot{x}(t) = f(x(t), u(t)), \tag{2}$$

$$y_k = h_k(x) + v_k, \tag{3}$$

where $h_k(\cdot)$represents the subset of sensors available at the time t_k, and v_k is measurement noise. In the multi-rate implementation, each measurement channel is incorporated only when its corresponding sampling time is reached, so the EKF evolves continuously by model prediction between measurement arrivals and applies discrete corrections at the sensor and assay instants. For numerical robustness, the covariance update is implemented using the Joseph form [11], with covariance symmetrization after propagation and update. L1 serves as a baseline physics-consistent soft sensor, providing real-time state estimates under nominal conditions but remaining sensitive to sensor bias and model mismatch.

Level 2 (L2): Bias-robust estimation via state augmentation

Level 2 extends the EKF by explicitly estimating slowly varying sensor biases. This is achieved by augmenting the state vector as:

$$x_a = \begin{bmatrix} x \\ b \end{bmatrix}, \tag{4}$$

where b represents unknown measurement bias terms associated with selected sensors (e.g., volume or dissolved oxygen probes). The bias dynamics are modeled as a random walk,

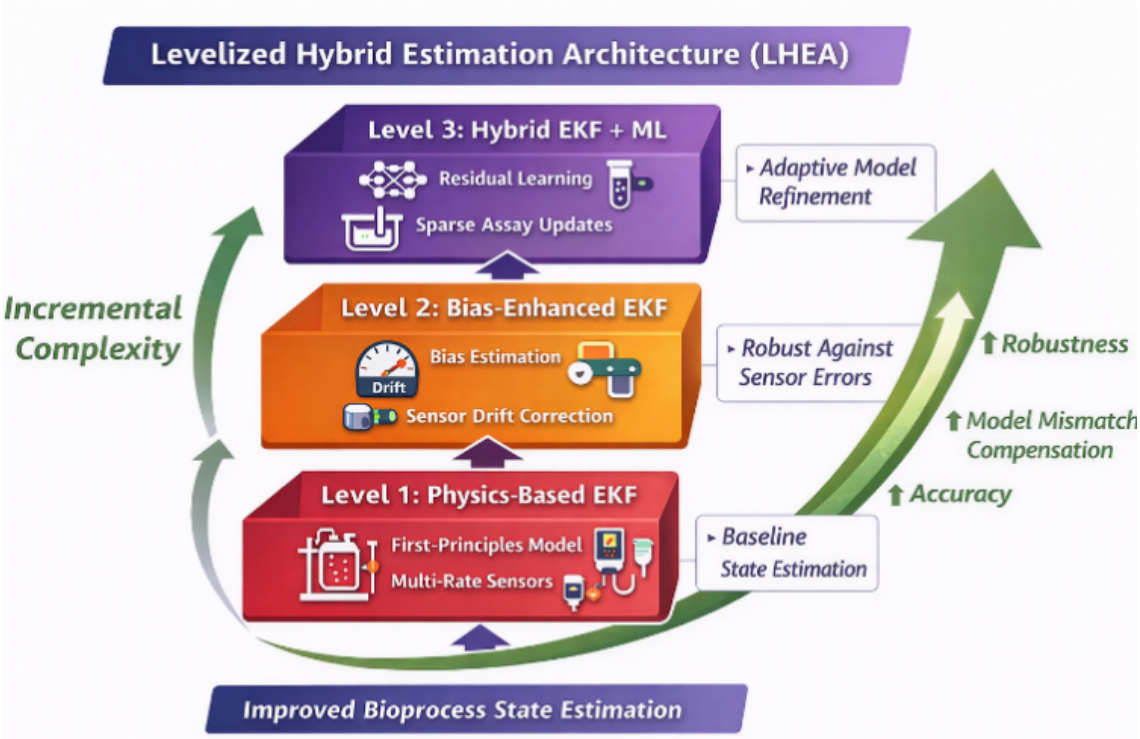

Figure 1. Conceptual illustration of the proposed Levelized Hybrid Estimation Architecture (LHEA). The framework progresses from a physics-based EKF (L1, red), to a bias-robust augmented EKF (L2, orange), and finally to a hybrid EKF–ML estimator (L3, purple). Sparse and multi-rate measurements are incorporated at all levels.

$$\dot{b}(t) = w_b(t), \tag{5}$$

reflecting slow drift due to fouling, calibration loss, or aging. Measurements affected by bias are written as

$$y_k = h(x_k) + b + v_k. \tag{6}$$

Importantly, neither the presence, magnitude, nor onset of bias is assumed to be known a priori. Only the possibility of bias in specific measurement channels is assumed based on engineering knowledge. The EKF jointly estimates process states and bias terms using innovation statistics, allowing sensor drift to be absorbed without corrupting the underlying state estimates. At this level, the process model remains unchanged, and no data-driven components are introduced.

Level 3 (L3): Physics-constrained hybrid EKF–ML

Even with bias handling, first-principles models inevitably exhibit a mismatch due to uncertain kinetics, physiological adaptation, or operating-regime changes. Level 3 introduces a physics-constrained residual learning mechanism within the EKF to compensate for such a mismatch while preserving interpretability. The process model is augmented as

$$\dot{x}(t) = f\big(x(t), u(t)\big) + r_\theta\big(\phi(x(t))\big), \tag{7}$$

where $r_\theta(\cdot)$ is a learned residual correction and $\phi(x)$ is a low-dimensional, interpretable feature vector (e.g., biomass and dissolved oxygen). The residual is restricted to selected dynamics (e.g., product formation) and is bounded to ensure physical plausibility, meaning that it (i) does not violate mass-balance structure, (ii) preserves state positivity, and (iii) remains bounded within a small fraction of the nominal kinetic rates. These constraints prevent violations of positivity, mass conservation, or known stoichiometric relationships.

The residual parameters θ are updated online only when sparse laboratory assays become available (e.g., HPLC product measurements every 24 h), using a lightweight adaptive law (e.g., normalized gradient or RLS-type update), triggered only when sparse, high-fidelity laboratory assays become available. Learning therefore, occurs within the same run, without offline retraining. The feasibility of this adaptation is ensured by (i) fast online measurements that constrain the state trajectory, (ii) sparse but informative assays that anchor slow dynamics, and (iii) the low dimensionality of the residual parameterization. The preceding observability analysis confirms that inclusion of sparse product measurements renders the system observable over the estimation horizon,

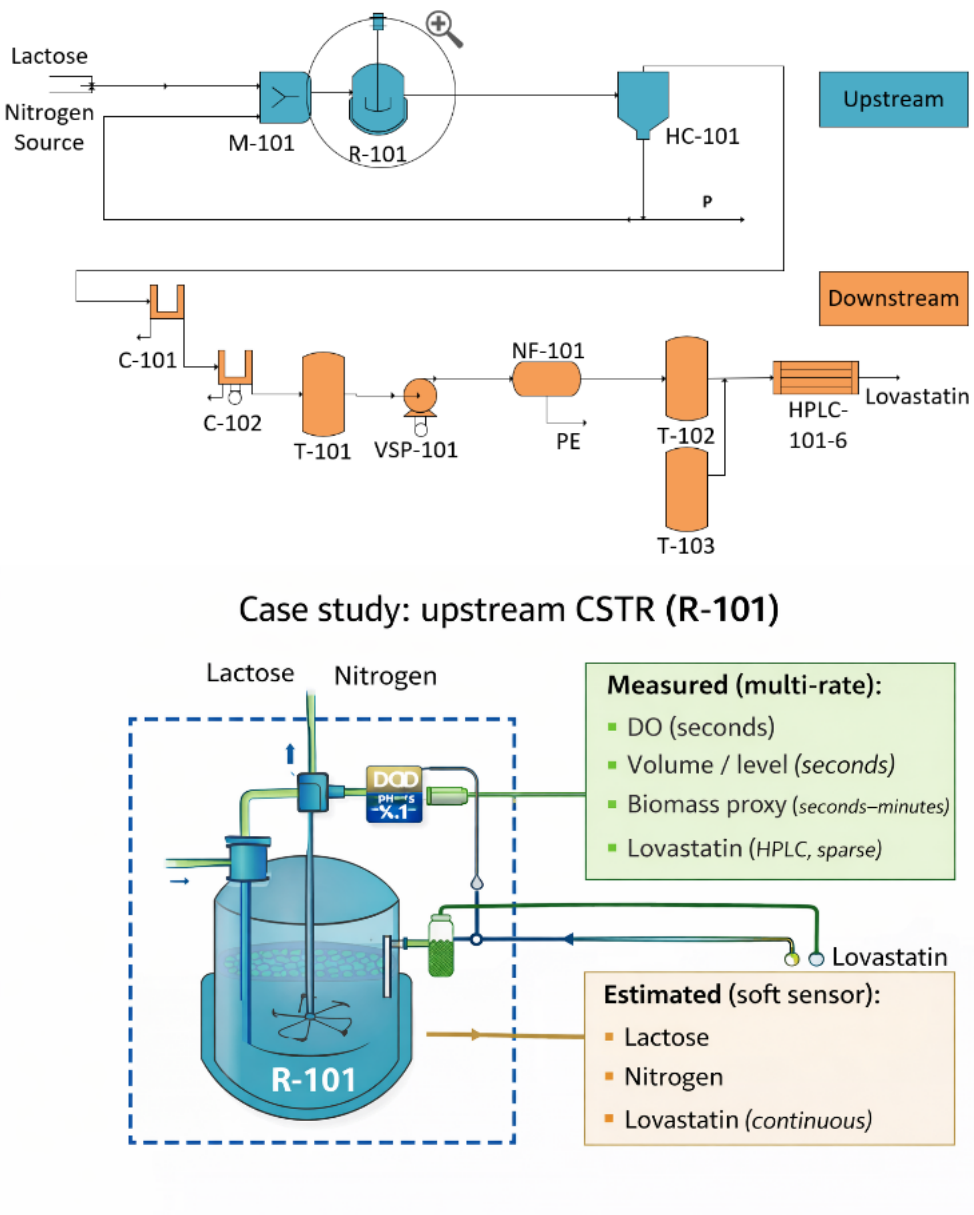

Figure 2. (a) Overall process flow diagram (PFD) of the KTB1 continuous biomanufacturing benchmark, including upstream fermentation and downstream separation units for lovastatin production. (b) Zoomed-in schematic of the CSTR reactor (R-101) highlighting available online sensors, sparse laboratory assays, and the hybrid EKF-based soft-sensor role for real-time state estimation.

enabling meaningful online learning.

In addition, slowly varying kinetic effects can be represented through a scalar scaling parameter:

$$q_{\text{eff}}(t) = q_{\text{nom}}\, q_{\text{scale}}(t), \tag{8}$$

which is updated using assay information and constrained within physically meaningful bounds.

Figure 1 summarizes the proposed *Levelized Hybrid Estimation Architecture (LHEA)* and illustrates how estimation capability is progressively enhanced through structured hybridization. Starting from a purely physics-based estimator (L1), each subsequent level introduces additional mechanisms to address practical limitations encountered in bioprocess monitoring, including sensor bias, model mismatch, and sparse measurements. The color coding highlights the incremental nature of the

framework: L1 (red) provides baseline real-time state reconstruction using first-principles models; L2 (orange) augments the estimator with bias states to improve robustness against sensor drift without altering the underlying process model; and L3 (purple) incorporates physics-constrained, sparsely trained machine-learning residuals to compensate for unmodelled kinetics or slowly varying parameters. Multi-rate and sparse measurements (green) are consistently integrated across all levels, reflecting realistic PAT availability. As illustrated by the upward trend in the figure, each level systematically increases estimation robustness, tolerance to model mismatch, and predictive accuracy while preserving physical interpretability.

CASE STUDY: CONTINUOUS LOVASTATIN PRODUCTION IN THE KTB1 SIMULATOR

Process Description

This study uses the KTB1 simulation platform, implemented in MATLAB/Simulink, as a representative case study for continuous biomanufacturing (Figure 2 a). KTB1 models the continuous production of the Activate Pharmaceutical Ingredient (API) lovastatin and includes realistic upstream and downstream unit operations. The upstream section consists of a mixer, a continuous stirred-tank bioreactor (CSTR), pumps, and a hydrocyclone, while the downstream section includes centrifuges, buffer tanks, a variable-speed pump, nanofiltration, and HPLC columns for product isolation. The simulator is able to account for sensor noise, actuator delays, and regulatory control, making it a suitable benchmark for evaluating advanced monitoring and soft-sensing strategies under industrially relevant conditions [10].

The focus of this work is the upstream CSTR reactor (R-101), where lovastatin is produced by *Aspergillus terreus*. Continuous monitoring of key states inside the reactor—particularly substrate and product concentrations—is essential for maintaining process stability, ensuring consistent product quality, optimizing yield, and enabling early fault detection. However, continuous monitoring in bioprocesses remains challenging, and this difficulty is especially pronounced for secondary metabolites such as lovastatin [12].

Continuous monitoring of lovastatin inside the reactor is challenging due to several practical limitations. Online sensors are affected by noise, drift, and delays, which can significantly degrade estimation performance during long continuous runs. Moreover, lovastatin is produced at low concentrations in dilute fermentation broths, resulting in weak analytical signals. Direct in-line quantification remains limited, and product measurements typically rely on sparse at-line or offline HPLC assays. In addition, slow biological kinetics and strong upstream–downstream coupling in continuous operation amplify the impact of measurement delays, making adequate feedback control difficult.

The dynamic CSTR model in KTB1 explicitly represents six key reactor states: biomass concentration, nitrogen (adenine), lactose (substrate), lovastatin (product), dissolved oxygen, and liquid volume/level. Table 1 summarizes these states, their symbols and units, typical measurement availability, preferred sensing technologies, and representative sampling cadences reported in the literature. While dissolved oxygen and volume/level are routinely available at fast sampling rates, substrate and product concentrations are typically measured only sparsely or inferred via soft sensors.

Table 1. Reactor states considered in the CSTR model, including symbols, units, measurability, typical sensing technologies, and realistic sampling cadences.

State	Symbol & unit	Preferred sensors & cadence	Key refs
Biomass	C_X (g/L)	Capacitance (5–15 s); OD/Turbidity (1–5 s)	[13]
Nitrogen	C_N (g/L)	HPLC-UV 260 nm (10–30 min); UV/Vis or Raman + PLS (30–120 s)	[14]
Lactose	C_LAC (g/L)	HPLC/enzymatic (5–30 min); Raman/NIR + PLS (30–120 s)	[15]
Lovastatin	C_MEV (g/L)	HPLC (30–120 min); LC-MS for intermediates; Raman soft-sensor (1–3 min)	[16]
Dissolved oxygen	C_DO (g/L)	Optical DO probe (1–5 s); polarographic (legacy)	[17]
Liquid volume/level	V (L)	Level sensor 1–5 s + geometry	[17]

Observability Analysis and Sensor Scenarios

To systematically assess how realistic sensor configurations affect state reconstruction, a multi-rate observability analysis was performed on the CSTR model. Nine sensor scenarios were defined, reflecting progressively richer combinations of online, at-line, and sparse measurements commonly encountered in industrial practice. These scenarios are summarized in Table 2.

Table 2. Sensor configuration scenarios used for the observability analysis, showing included measurements, and sampling cadences

Scenario	Sensors included	Sampling cadence
1	DO	5 s
2	DO, Volume/Level	DO: 5 s, Level: 5 s
3	DO, Biomass proxy (capacitance)	DO: 5 s, Biomass: 10 s
4	DO, Biomass proxy, Volume/Level	DO: 5 s, Biomass: 10 s, Level: 5 s
5	DO, Lactose (Raman)	DO: 5 s, Lactose: 60 s
6	DO, Biomass proxy, Lactose	DO: 5 s, Biomass: 10 s, Lactose: 60 s
7	DO, Biomass proxy, Lactose, Volume/Level	DO: 5 s, Biomass: 10 s, Lactose: 60 s, Level: 5 s
8	Scenario 7 + Product (MEV, HPLC)	DO: 5 s, Biomass: 10 s, Lactose: 60 s, Level: 5 s, MEV: 30 min
9	Scenario 4 + Product (MEV, HPLC)	DO: 5 s, Biomass: 10 s, Level: 5 s, MEV: 24 h

The observability rank for each scenario was evaluated over a 24-hour horizon using a base integration step of 1 s. The results are summarized graphically in Figure 3. As shown, DO-only sensing (Scenario 1) yields limited observability (rank 4), as oxygen dynamics alone are insufficient to uniquely reconstruct all material balances. Adding either volume/level measurements (Scenario 2) or a biomass proxy (Scenario 3) increases the observability rank to 5, demonstrating the importance of dilution effects and growth-related information. However, neither configuration alone achieves full state observability.

Scenario 4, which combines DO, biomass proxy, and volume/level measurements at realistic online cadences, consistently achieves an observability rank of 5 out of 6. This indicates that all states except the product concentration are practically observable, a result that aligns well with industrial experience: product formation dynamics are only weakly coupled to fast online sensors. The inclusion of substrate measurements (Scenarios 5–7) does not fundamentally improve observability unless combined with complementary sensors, highlighting that additional measurements must contribute independent dynamic information to be effective.

Full observability (rank 6) is achieved only in Scenarios 8 and 9, where sparse product (lovastatin) measurements are included, because the product state cannot be uniquely inferred from fast online measurements alone. Occasional product assays provide independent information that anchors the slow product dynamics and closes the observability gap when fused with faster measurements. Notably, Figure 3 shows that even very slow product assays—on the order of tens of minutes to several hours—are sufficient to close the observability gap when fused with faster online measurements over an extended horizon. Based on these results, Scenario 9 was selected as the primary sensor configuration for the subsequent estimation studies. This scenario reflects a cost-constrained yet industrially realistic sensing setup and serves as a stringent test case for the proposed levelized hybrid estimation architecture.

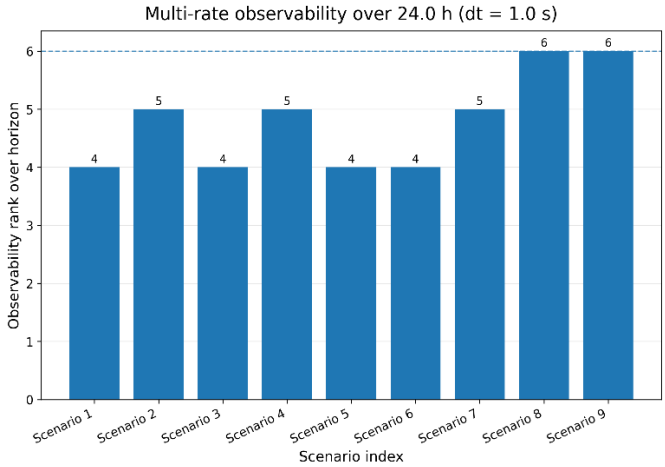

Figure 3. Multi-rate observability rank of the CSTR reactor states over a 24 h horizon for different sensor configurations (Scenarios 1–9). The dashed line indicates full observability (rank 6). Scenarios 8 and 9 have full observability.

RESULTS AND DISCUSSION

Considering the observability analysis in Figure 3, prior experience with similar industrial fermentations, and economic constraints associated with advanced sensing, *Scenario 9* was selected for subsequent estimation studies. This scenario combines volume/level measurement, a capacitance-based biomass proxy, dissolved oxygen, and sparse HPLC product assays, representing an industrially realistic and cost-constrained sensor configuration.

EXP1. Nominal estimation performance

Figure 4 compares the lovastatin concentration estimates obtained under nominal operating conditions. The normal reactor trajectory, shown as the blue line, remains nearly constant at approximately 1.2 g/L over the operating horizon. The Level 1 estimate, based solely on the first-principles model together with the available online measurements and sparse 24 h product assays, remains numerically stable and exhibits small discrete corrections at the assay instants. These update events confirm that the baseline EKF does assimilate the sparse

laboratory measurements; however, between assay times, a mild downward drift is still observed, indicating sensitivity to residual process-model mismatch. The Level 2 bias-augmented EKF estimate follows the Level 1 trajectory closely and shows the same assay-driven corrections, with only marginal deviation from Level 1 under nominal conditions. This behavior indicates that, in the absence of systematic measurement bias, the additional bias states remain weakly excited and therefore provide limited improvement in estimation accuracy**.** In contrast, the Level 3 hybrid estimator (dashed red), in which the physics-based EKF is augmented with residual learning driven by sparse HPLC product assays, remains tightly bound around the normal reactor trajectory, with small corrective adjustments occurring at assay times. These results demonstrate that, under nominal conditions, even when all estimators receive the sparse assays, the primary performance improvement arises from adaptive compensation of process-model mismatch rather than sensor bias handling.

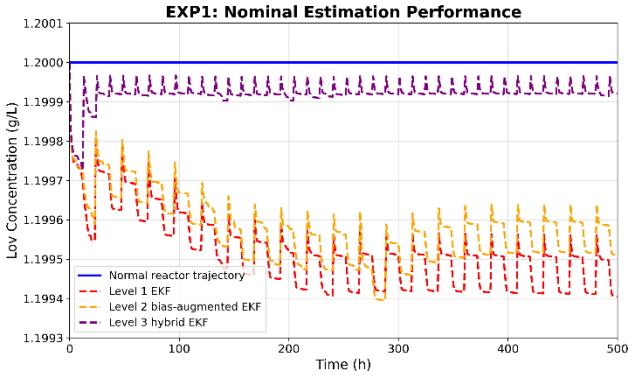

Figure 4. Lovastatin concentration estimates under nominal operation. The blue curve represents the true lovastatin concentration. Level 1 (EKF), Level 2 (bias-augmented EKF), and Level 3 hybrid are overlaid for direct comparison

EXP2. Robustness to measurement bias

Figure 5 evaluates estimator robustness to sensor bias by introducing a step change only in the measured reactor volume (5000 → 5050 L at t = 50 h), while the true process remains unchanged. The normal reactor trajectory, shown as the blue line, remains consistent with the nominal operating case. The Level 1 estimate (dashed red) directly follows the biased volume signal; once the bias is introduced, its lovastatin estimate progressively deviates upward from the normal reactor trajectory, indicating that the standard EKF propagates the biased measurement through the process model and mass balances. The Level 2 bias-augmented EKF estimate (dashed orange) effectively absorbs the artificial measurement shift through its bias states, maintaining a lovastatin estimate consistent with the nominal case. The Level 3 estimator (dashed purple) exhibits similar stability under this scenario, as no kinetic mismatch is present. These results demonstrate that the primary contribution of Level 2 is robustness to sensor bias and drift rather than improved nominal accuracy.

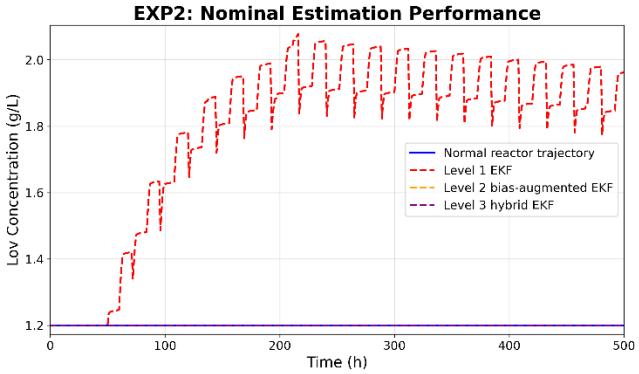

Figure 5. Lovastatin concentration estimates under a measurement bias scenario. The normal reactor trajectory (blue) is shown together with Level 1 EKF (dashed red), Level 2 bias-augmented EKF (dashed orange), and Level 3 hybrid EKF (dashed purple).

EXP3. Adaptation to process-model mismatch

Figure 6 examines estimator performance under a true process-model mismatch, introduced by increasing the intrinsic product formation rate $q_{\max,\mathrm{MEV}}$by 20% at $t =$ 50h. The resulting reactor lovastatin concentration, shown as the blue line, exhibits a gradual but persistent deviation from the nominal trajectory due to the altered kinetics. Under this mismatch, the Level 1 estimate (dashed red) and the Level 2 bias-augmented EKF estimate (dashed orange) are repeatedly corrected by the sparse assay updates; however, both remain limited by the fixed process model and therefore do not fully recover the transient lovastatin evolution. In contrast, the Level 3 estimator (dashed purple) tracks the changes in the trajectory much more closely by combining physics-based prediction with a bounded residual correction learned from the assay information. Notably, the Level 3 estimate is not simply a sample-and-hold reconstruction: it evolves between assay times and provides a smoother and more accurate transient estimate than a zero-order-hold (ZOH) baseline. The ZOH baseline yielded an RMSE of 3.24×10^{-4}g/L and an MAE of 2.10×10^{-4}g/L, whereas Level 3 achieved lower errors, corresponding to a 41.3% reduction in RMSE and a 38.0% reduction in MAE relative to ZOH. These results demonstrate that residual learning provides a capability fundamentally distinct from bias handling, enabling accurate estimation under evolving bioprocess kinetics.

Table 3 summarizes the quantitative estimation performance of the three estimator levels across the three

evaluation scenarios: nominal operation, measurement bias, and kinetic mismatch. Under nominal conditions, all estimators achieve low prediction errors, with Level 1 and Level 2 showing comparable RMSE and MAE values, reflecting the adequacy of the baseline physics-based EKF when model assumptions hold. When a measurement bias is introduced, the error metrics of Level 1 increase substantially, whereas Level 2 maintains nearly unchanged performance due to its explicit bias-state formulation.

Under kinetic parameter uncertainty or change, Level 3 provided the lowest estimation error, while Levels 1 and 2 remain more strongly affected by the structural model discrepancy. Overall, the results confirm that each estimator level contributes a distinct and complementary capability: baseline state reconstruction (L1), robustness to sensor bias (L2), and adaptation to process-model mismatch (L3), supporting the proposed levelized estimation strategy.

Table 3. Quantitative comparison of lovastatin concentration estimation accuracy (RMSE, MAE) for Level 1–3 estimators under nominal operation, measurement bias, and kinetic mismatch scenarios.

Level	Index	EXP1	EXP2	EXP3
Level 1	RSME	0.00049	0.64920	0.00121
	MAE	0.00048	0.59454	0.00099
Level 2	RSME	0.00043	0.00043	0.00102
	MAE	0.00042	0.00042	0.00082
Level 3	RSME	8.5e-5	8.3e-5	0.00019
	MAE	7.9e-5	7.9e-5	0.00013

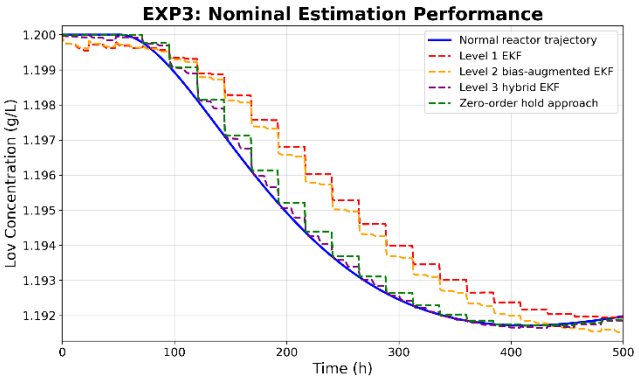

Figure 6. Lovastatin concentration under a true process–model mismatch. The reactor trajectory (blue) is shown together with Level 1 EKF (dashed red), Level 2 bias-augmented EKF (dashed orange), Level 3 hybrid EKF (dashed purple), and zero-order hold baseline (dashed green).

Perspective and Future Work

This work demonstrates the first three layers of the proposed Levelized Hybrid Estimation Architecture (LHEA), showing how physics-based EKF estimation can be systematically enhanced to improve robustness and accuracy under realistic sensing constraints. While this study focuses on state reconstruction, bias robustness, and physics-constrained residual learning, several extensions remain for future work. These include learned measurement models that convert auxiliary PAT signals (e.g., Raman or capacitance) into soft measurements, innovation-adaptive noise tuning for automated robustness, and physics-regularized surrogate models for operation under severe regime changes. Integrating the levelized estimators with closed-loop control and validating the framework on experimental or pilot-scale data are also important next steps. Together, these directions complete the envisioned framework and position LHEA as a scalable, interpretable pathway for incorporating machine learning into high-reliability soft sensing for continuous biomanufacturing

DIGITAL SUPPLEMENTARY MATERIAL

Digital supplementary material and codes associated with this paper are available at the following link: https://github.com/Boskabadi/LHEA

ACKNOWLEDGEMENTS

Luis Ricardez-Sandoval acknowledges the support from the Canada Research Chairs (CRC) program.

AUTHOR IDENTIFIERS

Author ORCIDs:
Mohammad Reza Boskabadi: 0009-0009-9619-3735
Rajiv Kailasanathan: 00009-0009-8915-7111
Luis Ricardez-Sandoval: 0000-0001-9867-6778
Seyed Soheil Mansouri: 0000-0002-0924-4947

REFERENCES

1. Kailasanathan R, Clements WR, Boskabadi MR, Gibford S, Papadakis E, Savoie CJ, Mansouri SS. Quantum enhanced ensemble gans for anomaly detection in continuous biomanufacturing. Ind. Eng. Chem. Res. 65:5046-5057 (2026). https://doi.org/10.1021/acs.iecr.5c04183
2. M. Drobnjakovic, R. Hart, B. Kulvatunyou, N. Ivezic, and V. Srinivasan, "Current challenges and recent advances on the path towards continuous biomanufacturing, " *Biotechnol. Prog.*, vol. 39, no. 6, p. e3378, 2023.
3. Liñán DA, McCormick-Mantilla JA, Ricardez-Sandoval LA. Economic nonlinear model predictive control and scheduling of multiple fed-batch fermenters in a lignocellulosic biorefinery.

Computers & Chemical Engineering 201:109223 (2025). https://doi.org/10.1016/j.compchemeng.2025.109223
4. Ghodba A, Richelle A, McCready C, Ricardez-Sandoval L, Budman H. A robust batch-to-batch optimization framework for pharmaceutical applications. Computers & Chemical Engineering 193:108935 (2025). https://doi.org/10.1016/j.compchemeng.2024.108935
5. Valipour M, Ricardez-Sandoval LA. Abridged gaussian sum extended kalman filter for nonlinear state estimation under non-gaussian process uncertainties. Computers & Chemical Engineering 155:107534 (2021). https://doi.org/10.1016/j.compchemeng.2021.107534
6. Valipour M, Ricardez-Sandoval LA. Assessing the impact of EKF as the arrival cost in the moving horizon estimation under nonlinear model predictive control. Ind. Eng. Chem. Res. 60:2994-3012 (2021). https://doi.org/10.1021/acs.iecr.0c06095
7. Boskabadi MR, Murugaiah M, Nielsen TR, Sivaram A, Sin G, Mansouri SS. Virtual sensor for sustainable large-scale process monitoring. Ind. Eng. Chem. Res. 64:3902-3917 (2025). https://doi.org/10.1021/acs.iecr.4c03342
8. Strielkowski W, Vlasov A, Selivanov K, Muraviev K, Shakhnov V. Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: a review. Energies 16:4025 (2023). https://doi.org/10.3390/en16104025
9. R. Gani *et al.*, "Challenges and opportunities for process systems engineering in a changed world, " *Comput. Aided Chem. Eng.*, vol. 49, pp. 7–20, 2022.
10. Boskabadi MR, Ramin P, Kager J, Sin G, Mansouri SS. Kt-biologics i (KTB1): a dynamic simulation model for continuous biologics manufacturing. Computers & Chemical Engineering 188:108770 (2024). https://doi.org/10.1016/j.compchemeng.2024.108770
11. Bar-Shalom Y, Li XR, Kirubarajan T. Estimation with applications to tracking and navigation. Wiley (2002). https://doi.org/10.1002/0471221279
12. Kroll P, Sagmeister P, Reichelt W, Neutsch L, Klein T, Herwig C. ex situonline monitoring: application, challenges and opportunities for biopharmaceuticals processes. Pharmaceutical Bioprocessing 2:285-300 (2014). https://doi.org/10.4155/pbp.14.22
13. Bergin A, Carvell J, Butler M. Applications of bio-capacitance to cell culture manufacturing. Biotechnology Advances 61:108048 (2022). https://doi.org/10.1016/j.biotechadv.2022.108048
14. Popović I, Dončević L, Biba R, Košpić K, Barbalić M, Marinković M, Cindrić M. Advancements in adenine nucleotides extraction and quantification from a single drop of human blood. Molecules 29:5630 (2024). https://doi.org/10.3390/molecules29235630
15. Gerzon G, Sheng Y, Kirkitadze M. Process analytical technologies – advances in bioprocess integration and future perspectives. Journal of Pharmaceutical and Biomedical Analysis 207:114379 (2022). https://doi.org/10.1016/j.jpba.2021.114379
16. Garavaglia M, McGregor C, Bommareddy RR, Irorere V, Arenas C, Robazza A, Minton NP, Kovacs K. Stable platform for mevalonate bioproduction from co₂. ACS Sustainable Chem. Eng. 12:13486-13499 (2024). https://doi.org/10.1021/acssuschemeng.4c03561
17. Gibford SM, Boskabadi MR, Savoie CJ, and Mansouri SS. Quantum Synthetic Data Generation for Industrial Bioprocess Monitoring. arXiv preprint (2025) https://arxiv.org/pdf/2510.17688