RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

RETROSPECTIVE SOFT-SENSOR ASSESSMENT OF CHINESE HAMSTER OVARY CULTURES USING CAMPAIGN-GROUPED CROSS-VALIDATION

AVALIAÇÃO RETROSPECTIVA DE SENSORES VIRTUAIS EM CULTIVOS DE CÉLULAS DE OVÁRIO DE HAMSTER CHINÊS POR VALIDAÇÃO CRUZADA AGRUPADA POR CAMPANHA

EVALUACIÓN RETROSPECTIVA DE SENSORES VIRTUALES EN CULTIVOS DE CÉLULAS DE OVARIO DE HÁMSTER CHINO MEDIANTE VALIDACIÓN CRUZADA AGRUPADA POR CAMPAÑA

Rubens Barbosa Rezende
Postgraduate in Biotechnology, Federal University of São Paulo, Brazil
E-mail: rubensrezende420@gmail.com

Abstract

Background: Retrospective estimation of viable cell density (VCD) from bioreactor signals requires careful time alignment and campaign-held-out testing. **Objective**: The study aimed to evaluate the association of seven concurrent inline sensor signals with CHO-K1 VCD estimation and to assess a separate, reduced growth model on external aggregate trajectories. **Methods**: I reanalyzed 378 offline observations and 274,781 inline records from 24 CHO-K1 cultivations in 12 campaigns. A seven-signal estimator underwent post-hoc nested leave-one-campaign-out assessment in nine batch campaigns. Separately, CHO-K1-trained early-growth coefficients were applied without refitting to three NISTCHO configuration means from one independent study, with observed day-zero VCD used for initialization. **Results**: Complete seven-signal matches were available for 350 offline observations. Across the nine batch campaigns, campaign-averaged mean absolute error (MAE) decreased from 1.346 x 10^6 cells/mL with time-only splines to 0.956 x 10^6 cells/mL with time plus seven signals (28.9% exploratory reduction), but early-period gains were less consistent. In the separate cross-dataset transfer assessment, mean MAE was 0.382 *versus* 0.800 x 10^6 cells/mL for the constant baseline across days 1-3; the transferred model performed slightly worse than the constant baseline for pDs. **Conclusions**: Benefits were retrospective and context dependent. The original seven-sensor estimator was not validated externally, and neither analysis demonstrates prospective forecasting or industrial control.

Keywords: Chinese hamster ovary cells; soft sensors; viable cell density; process analytical technology; cross-validation.

Resumo

Introdução: A estimativa retrospectiva da densidade de células viáveis (DCV) a partir de sinais de biorreatores exige alinhamento temporal rigoroso e avaliação independente por campanha experimental. **Objetivo**: Avaliar a associação de sete sinais simultâneos de sensores em linha com a estimativa de DCV de células CHO-K1 e examinar, separadamente, a transferência de um modelo reduzido de crescimento para trajetórias médias externas. **Métodos**: Foram reanalisadas 378

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

observações laboratoriais e 274.781 registros de sensores em linha provenientes de 24 cultivos CHO-K1 distribuídos em 12 campanhas. Um estimador com sete sinais foi submetido, retrospectivamente, à validação cruzada aninhada com exclusão de uma campanha por vez em nove campanhas batch. Separadamente, coeficientes de crescimento inicial ajustados em CHO-K1 foram aplicados, sem novo ajuste, às médias de três configurações NISTCHO de um único estudo independente, com inicialização pela DCV observada no dia zero. **Resultados**: Houve correspondência completa dos sete sinais para 350 observações laboratoriais. O erro absoluto médio (MAE) das campanhas batch diminuiu de 1,346 x 10^6 células/mL com *splines* baseados apenas no tempo para 0,956 x 10^6 células/mL com tempo e sete sinais (redução exploratória de 28,9%); no período inicial, os ganhos foram menos consistentes. No teste externo distinto, o MAE médio foi de 0,382 contra 0,800 x 10^6 células/mL para a referência constante nos dias 1-3, com desempenho inferior na configuração pDs. **Conclusões**: Os benefícios observados foram retrospectivos e dependentes do contexto. O estimador original com sete sensores não foi validado externamente, e nenhuma análise demonstra previsão prospectiva ou controle industrial.

Palavras-chave: Células de ovário de hamster chinês; sensores virtuais; densidade de células viáveis; tecnologia analítica de processos; validação cruzada.

Resumen

Introducción: La estimación retrospectiva de la densidad de células viables (DCV) a partir de señales de biorreactores requiere una alineación temporal rigurosa y una evaluación independiente por campaña experimental. **Objetivo**: Evaluar la asociación de siete señales simultáneas de sensores en línea con la estimación de DCV de células CHO-K1 y examinar, por separado, la transferencia de un modelo reducido de crecimiento a trayectorias medias externas. **Métodos**: Se reanalizaron 378 observaciones de laboratorio y 274.781 registros de sensores en línea procedentes de 24 cultivos CHO-K1 distribuidos en 12 campañas. Un estimador de siete señales se sometió retrospectivamente a una validación cruzada anidada, dejando fuera una campaña por vez, en nueve campañas de cultivo discontinuo. Por separado, los coeficientes de crecimiento inicial ajustados con CHO-K1 se aplicaron, sin reajuste, a las medias de tres configuraciones NISTCHO de un único estudio independiente, utilizando la DCV observada en el día cero para la inicialización. **Resultados**: Se dispuso de correspondencias completas de las siete señales para 350 observaciones de laboratorio. El error absoluto medio (MAE) de las campañas discontinuas disminuyó de 1,346 x 10^6 células/mL con splines basados únicamente en el tiempo a 0,956 x 10^6 células/mL con el tiempo y las siete señales (reducción exploratoria del 28,9%), aunque las mejoras en el período inicial fueron menos consistentes. En la prueba externa diferenciada, el MAE medio fue de 0,382 frente a 0,800 x 10^6 células/mL para la referencia constante durante los días 1-3, con un rendimiento inferior en la configuración pDs. **Conclusiones**: Los beneficios fueron retrospectivos y dependientes del contexto. El estimador original de siete sensores no se validó externamente y ninguno de los análisis demuestra pronóstico prospectivo ni control industrial.

Palabras clave: Células de ovario de hámster chino; sensores virtuales; densidad de células viables; tecnología analítica de procesos; validación cruzada.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

1. Introduction

Bioreactor performance emerges from coupled biological and engineering phenomena, including cell physiology, mass and heat transfer, mixing, and nutrient-feeding strategies. Stirred-tank, pneumatically mixed, perfusion, and single-use platforms address different operating requirements, particularly oxygen delivery, hydrodynamic stress, and scale-up (Nienow, 2006; Sharma; Harrison; Tai, 2022). Growing instrumentation capabilities nevertheless raise a complementary question: are the resulting measurements sufficiently complete and representative to support trustworthy inference?

In biopharmaceutical manufacturing, continuous monitoring can provide finer temporal resolution than intermittent sampling, but it does not eliminate calibration error, sensor drift, synchronization problems, or ambiguity in signal interpretation. Capacitance, optical signals, dissolved oxygen and carbon dioxide, pH, and conductivity respond to different physical and chemical phenomena; none should automatically be treated as a direct measurement of viable cell density or product quality. Sensor selection must therefore be tied to the target variable, operating context, and calibration evidence (Pineda Rincon *et al*., 2025; Reardon, 2021; Reyes *et al*., 2022; Wasalathanthri *et al*., 2020).

Soft sensors integrate instrument readings to estimate states that cannot be measured continuously. Hybrid models can additionally combine mechanistic balances with statistical learning. Neither approach is inherently a digital twin or an autonomous controller: the latter require operational connectivity, updating strategies, uncertainty management, and evaluation under actual process constraints (Satwekar *et al*., 2026). This distinction matters when interpreting retrospective reductions in prediction error, as emphasized in practical assessments of machine learning for bioprocesses (Helleckes *et al*., 2023).

Chinese hamster ovary (CHO) cells provide an important case because of their widespread use in recombinant therapeutic-protein manufacturing. A recent integrative review has also addressed CHO cells and the yeast *Komagataella phaffii* as production platforms for biotherapeutics (Rezende, 2026). Experiments with the

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

NISTCHO platform demonstrate that feeding conditions can influence growth, metabolism, and product attributes simultaneously; viable cell density (VCD) should not be used as a surrogate for antibody titer or quality (Schäpertöns *et al*., 2026). Reference protocols also emphasize consistent operational documentation and comparability between campaigns (Karageorgos *et al*., 2026).

Open bioreactor time series raise issues of time-base alignment, units, instrumentation gaps, and correlated repeated measurements. Repeated measurements from a single trajectory cannot be treated as independent cultivations. The CHO-K1 dataset released by Uhlendorff *et al*. (2026) contains both high-frequency inline signals and intermittent laboratory data, allowing the consequences of these design choices to be evaluated in an open and traceable setting.

This study involved an original retrospective quantitative analysis of CHO-K1 cultivation measurements, supported by relevant scientific literature. The objectives were to audit synchronized sensor availability; compare time-only and sensor-enhanced VCD estimators using experimental campaigns held out from model training; examine sensitivity to cultivation phase and missing data; and assess the transfer of a separately fitted, reduced early-growth model to three NISTCHO configuration-level mean trajectories from an independent experimental study. The original seven-signal estimator was not evaluated using an independent instrumentation platform, and no real-time closed-loop control was implemented.

2. Materials and Methods

2.1 Study design

I conducted an original retrospective quantitative study using secondary data from publicly deposited experimental datasets. I analyzed CHO-K1 cultivation records to assess sensor-time alignment, missingness, concurrent VCD estimation, and campaign-grouped model performance. Separately, I evaluated a reduced early-growth model against three NISTCHO configuration-level mean trajectories from one independent study. No new cell-culture experiments or real-time control interventions

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

were performed. Selected experimental and methodological publications were used to contextualize the results; no systematic literature search or formal study-level risk-of-bias assessment was conducted.

2.2. Data sources and provenance

The publicly deposited dataset from Uhlendorff *et al*. (2026) was reanalyzed (DOI: 10.5281/zenodo.20829178). It contains 24 CHO-K1 cultivations in two parallel reactors across 12 campaigns: nine batch and three fed-batch experiments. The archive reports 378 offline laboratory observations and time-aligned inline files (one inline and one offline CSV per reactor). Experimental time was calculated from inoculation as defined in the repository.

An independent dataset was reported by Barczyk *et al*. (2026) and deposited in DaRUS (DOI: 10.18419/DARUS-6466). All Data.xlsx and Figure3.csv report three NISTCHO scale-down bioreactor configurations (SMCB, pDs and Reference), with at-line VCD and other measures over 12 days. VCD values in the source file Figure3.csv were converted from units of 10^5 to 10^6 viable cells/mL. Numerical agreement between the rounded CSV and the more precise Excel data was checked, and both files matched their repository MD5 records. These are three system-level curves derived from one study, not three independently replicated external studies. Because the identical seven-channel CHO-K1 sensor panel was unavailable, I performed only a distinct, reduced time-only transfer assessment for days 0-3.

Regarding cell-line provenance, the CHO-K1 lineage comprises Chinese hamster (*Cricetulus griseus*), female, ovary-derived cells (RRID:CVCL_0214; Cellosaurus). This lineage-level identification does not prove the identity, supplier, passage history or mycoplasma-negative status of the particular stocks in the source experiments. Stock-specific authentication and testing records were not independently verified from the repository material considered here; no new cell lines were handled in this secondary-data study.

2.3. Variables and analytical procedures

For each cultivation, I extracted the numbers of offline and inline observations, peak VCD, initial and final glucose, peak lactate, final ammonium, feeding indicators,

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

and data-quality flags. Means and ranges were summarized separately by operation mode. Measurements at different times within one cultivation were treated as repeated observations, not independent biological replicates. Mode-specific descriptive statistics were not adjusted for differences in feeding conditions or other operating parameters, and no test of batch versus fed-batch superiority was performed.

The experimental campaign (EXP), rather than the individual time point, defined the data partition. In each validation fold, both reactors from the held-out campaign were excluded from model fitting. Instrumental covariates were summarized only over windows ending at the corresponding laboratory sampling time; future information was never used. Experimental indicators were retained with their original units, and transformations were documented in the reproducible scripts.

2.4. Modeling approaches and analytical scope

Three different modeling tasks were conducted. First, an internal early batch-growth benchmark used the median of reactor-specific log-growth coefficients from the training campaigns in each leave-one-experiment-out (LOEO) fold. Second, concurrent VCD was estimated from cultivation time and matched inline signals using Ridge regression, with an additional post-hoc nested LOEO comparison. Third, a separate time-only model was fitted to all eligible CHO-K1 batch training records by weighted regression and evaluated on early NISTCHO system means without coefficient refitting. The median-coefficient internal benchmark and the weighted-regression external model are not the same estimator, so their mean absolute errors (MAEs) should not be interpreted as performance changes for one unchanged model. None is a first-principles mechanistic model. Equations 1-2 are theoretical identities; Equations 3-14 describe the applied calculations and estimators.

2.5. Data-integrity criteria

Recognized limitations include much sparser offline than inline sampling, only three independent fed-batch experiments, two reactors per campaign, and glucose readings of exactly zero that might reflect analytical detection or quantification limits. The origin of these zero values requires confirmation against the source protocol.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

Biological explanations for anomalous cell trajectories were not inferred without independent evidence.

2.6. Temporal indicators from offline measurements

Within each reactor, observations were ordered by elapsed time since inoculation. I calculated peak VCD, time to peak, trapezoidal area under the VCD-time curve, duration-normalized area, initial and final glucose, maximum lactate, and final ammonium. Values of exactly zero were retained rather than automatically labeled missing. The apparent early specific growth rate, μ (h^{-1}), was estimated as the slope of *ln* (VCD) against time for positive observations collected between 0 and 72 h. It is a local descriptive growth-rate estimate, not a physiological constant validated over the entire cultivation.

Mathematical definitions and numerical evaluation. For positive VCD, the instantaneous apparent growth rate is defined as the time derivative of log-density (Equation 1). This derivative provides kinetic interpretation; it was not estimated by finite-differencing noisy samples.

$$\mu_{\mathrm{app}}(t) = \frac{1}{X_v(t)} \frac{\mathrm{d}X_v(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \ln X_v(t) \quad (1)$$

Integrating the continuous-time growth-rate identity gives Equation 2. It is shown to explain the exponential benchmark: the study did not numerically integrate a time-varying growth-rate function.

$$\ln \frac{X_v(t)}{X_v(0)} = \int_0^t \mu_{\mathrm{app}}\,(u)\,\mathrm{d}u \quad (2)$$

In the actual early-growth analysis, the reactor-level coefficient was estimated from the least-squares slope of observed ln(VCD) against elapsed time (Equation 3), using positive observations in the specified 0-72 h window.

$$\hat{\mu}_r = \frac{\sum_{i=1}^{n_r} (t_i - \bar{t}_r)\left(\ln X_{v,i} - \overline{\ln X_v}_r\right)}{\sum_{i=1}^{n_r} (t_i - \bar{t}_r)^2} \quad (3)$$

Integrated viable-cell exposure was calculated numerically from the irregularly sampled offline trajectory by the composite trapezoidal rule (Equation 4), approximating the time integral of VCD. The duration-normalized area was obtained using Equation 5.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

$$\mathrm{AUC}_{X_v} = \int_{t_1}^{t_n} X_v\,(t)\,\mathrm{d}t \approx \sum_{i=1}^{n-1} \frac{X_{v,i}+X_{v,i+1}}{2}\,(t_{i+1}-t_i) \quad (4)$$

$$\overline{X}_{v,\mathrm{time}} = \frac{\mathrm{AUC}_{X_v}}{t_n - t_1} \quad (5)$$

Here, X_v denotes viable cell density (10⁶ cells/mL), t is elapsed time (h), μ is in h⁻¹, and AUC has units of 10⁶ cells·h/mL. These are descriptive growth measures; no fed-batch mass-balance differential equation was calibrated.

2.7. Held-out-campaign kinetic benchmark and sensitivity analyses

For the internal early-growth benchmark, nine batch campaigns (EXP001-EXP009) were evaluated. Both reactors of each held-out campaign were excluded together. In every LOEO fold, reactor-specific apparent growth coefficients were calculated from the remaining eight campaigns and their median was transferred to the held-out pair. Predictions between inoculation and 72 h were initialized with each test reactor's observed initial VCD; a constant-initial-VCD prediction served as the baseline. Hence initial density was provided for calibration, and only the growth coefficient was estimated from other campaigns. This procedure is distinct from the weighted pooled regression used for NISTCHO transfer in Section 2.10.

The held-out-campaign prediction and its constant baseline are defined in Equation 6. The training coefficient μ̃-g is the median of the reactor-level estimates from the other eight batch campaigns; X_v(0) is available only from the held-out reactor's initial laboratory sample.

$$\hat{X}_v(t) = X_v(0)\exp\left(\tilde{\mu}_{-g}t\right), \qquad \hat{X}_{v,0}(t) = X_v(0) \quad (6)$$

Absolute prediction errors were expressed in 10⁶ cells/mL and summarized per observation and per campaign (the paired reactors combined). A descriptive 95% bootstrap percentile interval for the average percentage reduction in campaign-level error used 10,000 resamples of nine campaigns (fixed seed: 185). A paired Wilcoxon test was exploratory because only nine campaigns were available. Analyses at 48 and 96 h re-estimated the growth coefficient within each respective training window. This internal LOEO kinetic benchmark is distinct from the later NISTCHO transfer experiment in Section 2.10.

2.8. Inline-data integration and retrospective soft-sensor benchmarking

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

The aligned CSV files for all 24 reactors (24 inline and 24 offline) were extracted from Uhlendorff *et al*. (2026). At each offline sampling time *t*, the median of instrumental readings in [*t* - 0.25 h, *t*] was calculated without using observations after *t*. Seven signals were retained: permittivity, transmission, reflection, conductivity, dissolved O_2, dissolved CO_2, and pH. Of 378 offline observations, 350 had a complete seven-signal window; the remaining 28 were excluded from model comparisons but retained in the audit trail. The target was VCD (10^6 cells/mL).

To prevent future-data leakage, each sensor feature was the median of its valid observations during the 0.25 h immediately preceding laboratory sampling, as expressed in Equation 7. If any of the seven necessary features lacked an eligible value, the pair was excluded from model fitting and retained in the missingness audit.

$$\tilde{s}_j(t_i) = \text{median}\{s_j(u): t_i - 0.25 \le u \le t_i\}, \quad j = 1, \dots, 7 \quad (7)$$

Within each cultivation mode, LOEO evaluation held both reactors from the test campaign outside model fitting. The initial time-only comparator used a cubic-spline basis of time/100 with five knots defined from training data. The sensor-enhanced model added seven contemporaneous covariates. Feature scaling was fitted exclusively on training data; Ridge regression used α = 10, and negative VCD predictions were truncated to zero. Both 0-72 h and full-observation horizons were investigated. Because alternative model families and penalties had been explored during development, this fixed-setting benchmark was treated as exploratory rather than as an independent or nested hyperparameter-validation result.

The design vector combined a cubic B-spline expansion of scaled culture time (t/100) with J contemporary instrumental features; J = 0 for the temporal baseline, J = 1 for the permittivity comparison, and J = 7 for the complete sensor panel (Equation 8). For all model specifications, feature standardization was fitted on training observations only.

$$\mathbf{z}_i = \left[B_1(t_i/100), \dots, B_K(t_i/100), \tilde{s}_1(t_i), \dots, \tilde{s}_J(t_i)\right] \quad (8)$$

Ridge regression estimated its intercept and coefficients by minimizing the penalized least-squares criterion in Equation 9. The intercept was not regularized; α controlled the L_2 penalty. As in the implementation, negative regression outputs were

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

truncated to zero (Equation 10). These equations describe retrospective estimation of present-time VCD, neither mechanistic simulation nor forecasting.

$$(\hat{\beta}_0, \hat{\boldsymbol{\beta}}) = \underset{\beta_0, \boldsymbol{\beta}}{\operatorname{argmin}} \left[\sum_{i \in \mathcal{T}} \left(y_i - \beta_0 - \mathbf{z}_i^{\top} \boldsymbol{\beta} \right)^2 + \alpha \parallel \boldsymbol{\beta} \parallel_2^2 \right] \quad (9)$$

$$\hat{y}_i = \max\left(0, \hat{\beta}_0 + \mathbf{z}_i^{\top} \hat{\boldsymbol{\beta}}\right) \quad (10)$$

The primary metric was campaign-specific MAE, followed by an equally weighted mean across campaigns. Between-model differences were resampled at the campaign level with 20,000 bootstrap draws (seed: 20261022) to obtain descriptive 95% bootstrap percentile intervals. The seven-sensor model improved eight of nine full-horizon batch experiments, but overlapping training sets across LOEO folds do not amount to nine external validations. Neither causation by sensors nor operational process control was demonstrated (Lemke *et al*., 2023; Metcalfe *et al*., 2025; Siegl *et al*., 2024; Suman *et al*., 2025; Sun *et al*., 2025).

2.8.1. Temporal matching audit and missingness analysis

Offline sampling times were compared with the actual temporal coverage of the 24 original inline CSVs. The primary eligibility criterion remained a median computed from readings in the preceding 15 minutes, including time t, without interpolation or access to future readings. For each sample, I recorded the latest preceding inline time stamp, its age, eligibility under alternative windows (30 min, 1, 2, 4, 8, and 24 h), and whether the sample followed termination of instrumental recording. The principal window was not changed after examining the results. Missing values were not imputed; coverage analyses were descriptive.

2.8.2. Grouped nested cross-validation and sensitivity analysis

As a post-hoc extension, three batch-mode specifications were compared: time-only cubic splines, time plus permittivity, and time plus all seven instrumental covariates. In each outer LOEO iteration, both reactors from the held-out EXP were reserved for testing. For each specification, spline knot counts (4, 5, or 7) and Ridge penalties (α = 1, 10, or 100) were selected using LOEO cross-validation within the remaining eight training campaigns, minimizing the equally weighted campaign-level MAE. The model was then refitted on those eight campaigns and applied to the outer held-out campaign. Standardization and spline fitting occurred within training data

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

only. Negative outputs were truncated at zero. Predictor sets and hyperparameter grids were chosen during manuscript development and were not preregistered.

For the outer held-out campaign g, Equation 11 summarized selection of the tuning pair θ = (number of spline knots, Ridge penalty) by the mean inner LOEO MAE across the eight other campaigns. Each inner score was computed from a model trained without the corresponding inner campaign. The hyperparameter set θ and the three predictor families were chosen post hoc.

$$\hat{\theta}_{-g} = \operatorname{argmin}_{\theta\in\Theta} \frac{1}{8}\sum_{h\neq g} \mathrm{MAE}_h^{(\mathrm{inner})}(\theta), \quad \Theta = \{4,5,7\}\times\{1,10,100\} \quad (11)$$

Results were summarized as unweighted means of the nine batch-campaign MAEs. Between-model differences were evaluated using 30,000 campaign-level bootstrap resamples and an exploratory paired sign-flip distribution with 2^9 configurations. These analyses were treated as conditional on the exploratory design and model choices, not as confirmatory tests. Both the full observed period and the initial 72 h were assessed, with a separate inner selection for each horizon. Nested evaluation was not applied to fed-batch cultures because only three independent campaigns were available.

The primary error metric was calculated at the campaign level before aggregation, as in Equation 12: n_g is the number of matched samples from both reactors in campaign g, and G = 9 for batch experiments. Thus campaigns, not time points, have equal weight.

$$\mathrm{MAE}_g = \frac{1}{n_g}\sum_{i\in g} |y_i - \hat{y}_i|, \qquad \overline{\mathrm{MAE}} = \frac{1}{G}\sum_{g=1}^{G} \mathrm{MAE}_g \quad (12)$$

Absolute and percentage improvements relative to the time-only comparator followed Equation 13; positive Δ indicates a lower MAE with sensors.

$$\Delta_{\mathrm{MAE}} = \overline{\mathrm{MAE}}_{\mathrm{time}} - \overline{\mathrm{MAE}}_{\mathrm{sensors}}, \quad R = 100\frac{\Delta_{\mathrm{MAE}}}{\overline{\mathrm{MAE}}_{\mathrm{time}}} \quad (13)$$

For descriptive uncertainty, entire campaigns were resampled with replacement, and the 2.5th and 97.5th percentiles of the resampled paired difference were reported (Equation 14; 30,000 draws for nested analysis). The intervals remain exploratory because model-family selection was not preregistered.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

$$\mathrm{CI}_{95\%}(\Delta) = [Q_{0.025}(\Delta^*),\ Q_{0.975}(\Delta^*)] \quad (14)$$

Computational verification and software. The numerical evaluation and primary dataset integrity checks were independently re-executed using *Python 3.13.5*, *NumPy 2.3.5* and *scikit-learn 1.8.0*. The original 227.6-MB Zenodo archive matched its published MD5, and all 378 source offline measurements and recorded sensor-window medians were checked against the corresponding aligned *CSV* files. The accompanying reproducibility package contains executable scripts, intermediate matrices, held-out campaign predictions, descriptive metrics and checksum manifests. This audit verifies numerical reproduction; it does not transform post-hoc model development into prospective validation.

2.9. Contextual literature and evidence boundaries

Scientific interpretations were separated into three classes: results calculated from CHO-K1 data (Uhlendorff *et al*., 2026), independent experimental findings in other CHO or bioreactor settings, and methodological recommendations. For the independent studies, interpretation considered the cell line, cultivation mode, sensing modality, measured outcome and validation design. Results from another system were not used as independent validation of the present seven-signal estimator. This contextual comparison did not involve complete screening or full-text assessment of the wider literature.

2.10. Independent NISTCHO evaluation of reduced time-only kinetics

External evaluation of the seven-signal estimator would require a matched independent dataset with reference VCD, the same seven concurrent sensor types, compatible measurement units and the preceding-15-minute feature window. NISTCHO (Barczyk *et al*., 2026) does not supply this panel. I therefore examined a different, reduced transfer question using nine CHO-K1 batch campaigns (18 reactors; 98 offline records during days 0-3). Weighted least squares was applied to $\ln[VCD(t)/VCD(0)]$, with t in days, fitting either $\ln[VCD(t)/VCD(0)] = \beta_1 t$ (exponential) or $\ln[VCD(t)/VCD(0)] = \beta_1 t + \beta_2 t^2$ (log-quadratic). Contributions were balanced by campaign and reactor. Constant initial VCD served as the baseline. Coefficients were fitted exclusively on CHO-K1, with no NISTCHO refitting or selection. Unlike the

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

internal LOEO median-coefficient model, this external test used coefficients obtained from pooled, weighted CHO-K1 records.

The descriptive test covered days 0-3 of the SMCB, pDs and Reference NISTCHO system-level mean trajectories, corresponding to the early cultivation period (Barczyk *et al*., 2026). For each configuration, the observed day-zero VCD was supplied as the model's initialization value and excluded from error calculations. MAE and RMSE were computed for days 1-3 in 10^6 cells/mL. Consequently, there were nine scored daily system-level means (three per configuration), not nine independent cultivations. The underlying published means summarize biological duplicates; raw duplicate-level trajectories were not reconstructed from these means. Results were reported by system without inferential confidence intervals, significance tests or claims of multi-site generalization. The day-0 initialization and retrospective choice of evaluation window restricted the scope of this transfer assessment; the analysis did not validate the seven-signal estimator.

3. Results

3.1. Dataset structure and measurement completeness

The dataset comprised 24 cultivations conducted in 12 independent experimental campaigns: nine batch and three fed-batch. There were 378 offline laboratory measurements and 274,781 inline instrumental records. Offline variables included cell- and metabolite-related measurements; inline channels captured dielectric, optical, and physicochemical signals. Table 1 summarizes cultivation modes and descriptive outcomes. These unadjusted results cannot establish causal differences between operating modes.

Table 1. Descriptive characteristics of CHO-K1 cultivation campaigns derived from the secondary data audit.

Indicator	Batch	Fed-batch
Cultivations (n)	18	6
Independent campaigns (n)	9	3
Mean peak VCD (10^6 cells/mL)	9.42	9.01
Range of peak VCD (10^6 cells/mL)	3.79-14.00	1.83-16.17

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

Indicator	Batch	Fed-batch
Mean peak lactate (g/L)	2.71	6.74
Mean final ammonium (mM)	7.33	18.59
Mean offline measurements per cultivation	14.78	18.67

Source: calculations from Uhlendorff *et al*. (2026); unadjusted for campaign operating conditions.

The descriptive differences in Table 1 reflect cultivation records that were not adjusted for campaign operating conditions; they do not establish that one cultivation mode caused the observed outcomes.

3.2. Between-reactor heterogeneity within campaigns

Paired reactors sometimes differed substantially. Peak VCD in EXP001 was 14.00 versus 8.49 x 10^6 cells/mL, whereas EXP010 reactors reached 16.07 and 16.17 x 10^6 cells/mL. These contrasts justify retaining the campaign, not the measurement or reactor, as the held-out unit.

3.3. Data-quality flags and influential observations

Among 378 offline observations, 135 glucose readings were exactly zero (35.7%); depletion, quantification limits and export artifacts could not be distinguished. In EXP011/R02, inline acquisition ended at approximately 67.1 h, although offline sampling continued to 195 h. This reactor also contained five zero-VCD measurements and a final ammonium value near 48.93 mM. Without instrument logs, these anomalies cannot be interpreted causally.

3.4. Early growth and between-campaign variation

Early batch growth varied across the 18 reactors: median apparent specific growth rate was 0.0261 h^{-1} (range 0.0203-0.0366 h^{-1}) and peak VCD ranged from 3.79 to 14.00 x 10^6 cells/mL. This heterogeneity motivates campaign-held-out evaluation.

3.5. Performance on campaigns excluded from training

The early (0-72 h) LOEO benchmark evaluated 80 post-inoculation observations across nine batch campaigns. Equally weighted campaign MAE was 0.819 x 10^6 cells/mL for constant initial VCD and 0.471 x 10^6 cells/mL for the fold-wise median-coefficient exponential model (42.6% descriptive reduction; campaign-bootstrap interval for the relative reduction, 25.5%-54.9%). Eight campaigns

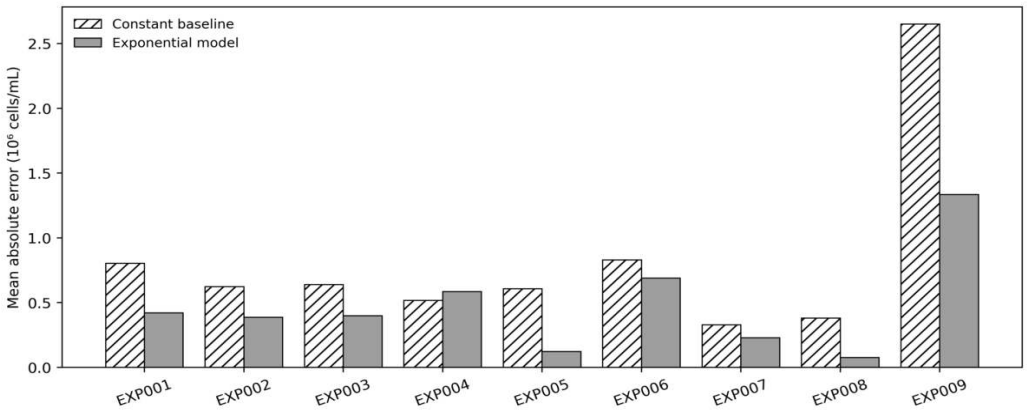
CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

improved; EXP004 worsened. Figure 1 shows this internal benchmark. Its growth-rate estimator differs from the pooled weighted estimator used in Section 3.10, so the two error summaries are not directly interchangeable.

Figure 1. Campaign-level mean absolute error of the constant and exponential models. Bars denote model errors, not biological replicates or confidence intervals.



Source: author's analysis of the CHO-K1 dataset described in the manuscript (Uhlendorff *et al.*, 2026).

Figure 1 illustrates the heterogeneous errors: EXP009 remained difficult for both models. Repeated observations from one cultivation do not constitute independent biological replicates.

3.6. Sensitivity to the time horizon and comparator choice

At 48, 72 and 96 h, observation-weighted MAE values were 0.319 *versus* 0.219, 0.822 *versus* 0.477 and 1.754 *versus* 0.791 x 10^6 cells/mL (constant *versus* exponential), respectively. The exponential model improved seven of nine campaigns at 48 h and eight of nine at 72 and 96 h. These post-hoc comparisons show gains over a simple baseline, not superiority over mechanistic, recursive-filtering or hybrid models.

3.7. Temporal matching and inline soft-sensor benchmark

Complete seven-signal windows were available for 350 of 378 offline samples (92.6%; 254 batch and 96 fed-batch) using the preceding 15 minutes. Table 2 reports concurrent LOEO VCD estimates; Figure 2 compares observations and estimates. The 274,781 inline records were not counted as independent cultures.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

Table 2. Retrospective LOEO performance with and without inline signals (MAE in 10^6 cells/mL).

Mode / time window	Campaigns (n)	Matched samples (n)	Time-only MAE	Time + sensors MAE	Reduction	95% interval for difference
Batch; 0–72 h	9	92	0.293	0.223	23.8%	−0.018 to 0.179
Batch; full	9	254	1.363	1.035	24.1%	0.149 to 0.512
Fed-batch; 0–72 h	3	38	0.280	0.234	16.2%	−0.133 to 0.203
Fed-batch; full	3	96	3.737	2.394	35.9%	0.772 to 1.659

Note: The 95% bootstrap percentile intervals describe between-campaign MAE differences; estimates for the three fed-batch campaigns are particularly uncertain and are not confirmatory.
Source: retrospective reanalysis of Uhlendorff *et al*. (2026).

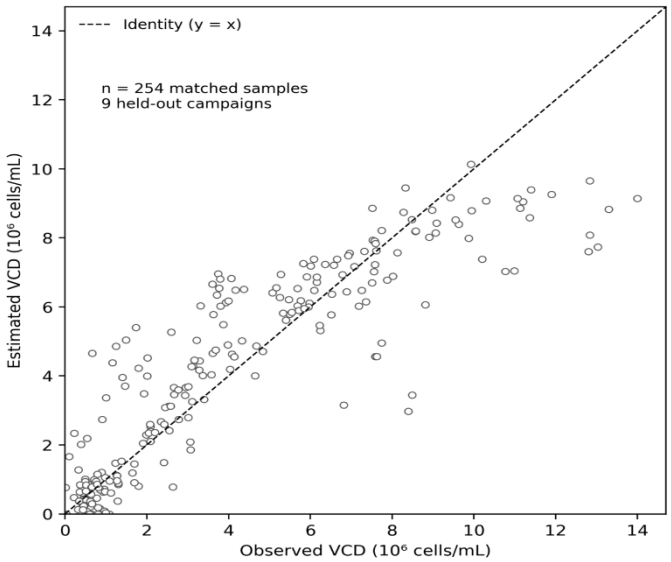
With fixed hyperparameters, seven sensors lowered full-period batch MAE by 0.329 x 10^6 cells/mL (95% campaign-bootstrap interval, 0.149-0.512), improving eight of nine campaigns. At 0-72 h, only five of nine improved and the interval included zero, showing horizon dependence. Figure 2 presents sample-level agreement; observations nested within campaigns are not independent validation units.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

Figure 2. Observed *versus* sensor-estimated viable cell density under batch-campaign LOEO evaluation. Each point is one matched offline sample; the dashed line indicates equality.



Source: Reanalysis of data from Uhlendorff et al. (2026).

The point-level association in Figure 2 is within one CHO-K1 repository and does not establish accuracy at an independently instrumented site.

3.8. Fed-batch observations and limits of extrapolation

The sensor-enhanced estimator had a lower full-period MAE in the fed-batch subset, but only three independent campaigns were available, precluding robust inference or extrapolation. Batch and fed-batch data were evaluated separately, without interpreting mode differences as causal feeding effects. All 28 offline observations lacking eligible sensor matches were retained in the audit trail, without imputation or artificial reconstruction of the signals.

3.9. Missing-signal audit and grouped nested cross-validation

Of 28 unmatched offline samples, 13 followed termination of inline acquisition and 15 occurred within internal signal gaps. Widening the matching window to 60 minutes would recover only one additional sample. In EXP011/R02, inline acquisition ended at 67.13 h, although offline sampling continued until 195 h. Missingness was not assumed to be random.

An additional nested evaluation used the same 254 matched samples from nine batch campaigns while selecting hyperparameters without consulting the outer

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

held-out campaign. Experiment-weighted mean MAE was 1.346 x 10^6 cells/mL with a time-only spline, 1.045 x 10^6 cells/mL with time plus permittivity, and 0.956 x 10^6 cells/mL with time plus seven signals. The last model achieved a 28.9% exploratory reduction versus the time-only comparator and lower error in eight campaigns. The mean paired difference in campaign-level MAE was 0.389 x 10^6 cells/mL (descriptive 95% campaign-bootstrap percentile interval: 0.198-0.582). Table 3 and Figure 3 provide the detailed comparison.

Table 3. Post-hoc nested cross-validation by batch campaign; MAE in 10^6 cells/mL, with nine campaigns equally weighted.

Window / model	Time-only MAE	Model MAE	Reduction	Campaigns improved
Full: time + permittivity	1.346	1.045	22.3%	8/9
Full: time + seven sensors	1.346	0.956	28.9%	8/9
0–72 h: time + seven sensors	0.301	0.226	24.8%	5/9

Source: retrospective reanalysis of Uhlendorff *et al*. (2026).

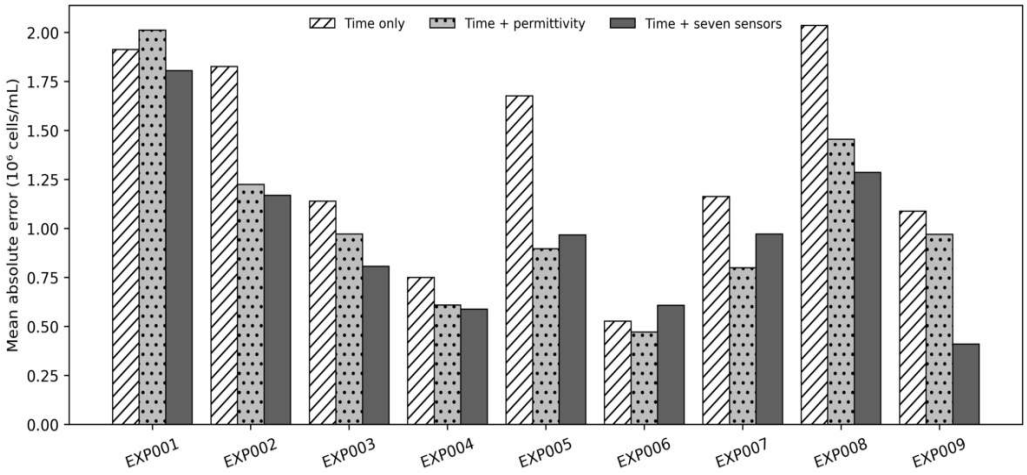
During the first 72 h, MAE declined from 0.301 to 0.226 x 10^6 cells/mL, but only five of nine campaigns improved. The campaign-bootstrap 95% interval for the difference (-0.012 to 0.177 x 10^6 cells/mL) included zero. Although inner-fold tuning excluded the held-out campaign from parameter selection, the predictor families and horizons were chosen post hoc; these results are retrospective, not prospective.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

Figure 3. Nested comparison of three models, with hyperparameters selected exclusively within training campaigns. Bars indicate MAE; unsupported dispersion bars were not added.



Source: Reanalysis of data from Uhlendorff et al. (2026).

Figure 3 displays the campaign-level variability among the three retrospective estimators. The seven-signal model reduced full-period mean error in eight of nine held-out batch campaigns, but early-period improvements were not uniform.

3.10. External transfer assessment of time-only growth kinetics

The independent NISTCHO source provided 39 published system-level VCD means (13 days x 3 configurations), of which 9 post-baseline means from days 1-3 were scored in this reduced transfer assessment. The 3 curves arise from 1 independent study and summarize biological duplicates; they are not 9 separate external cultivation experiments. Day-zero observations calibrated the initial condition. The Excel and rounded CSV files agreed to the published CSV precision and matched the repository MD5 values (Barczyk *et al*., 2026).

Using 98 observations from 18 CHO-K1 reactors in the 0-72 h training window, the exponential growth coefficient was fixed at 0.53728 d^{-1}; log-quadratic coefficients were 0.12262 d^{-1} and 0.18295 d^{-2}. No NISTCHO outcomes were used to refit these parameters. Table 4 reports the transfer errors.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

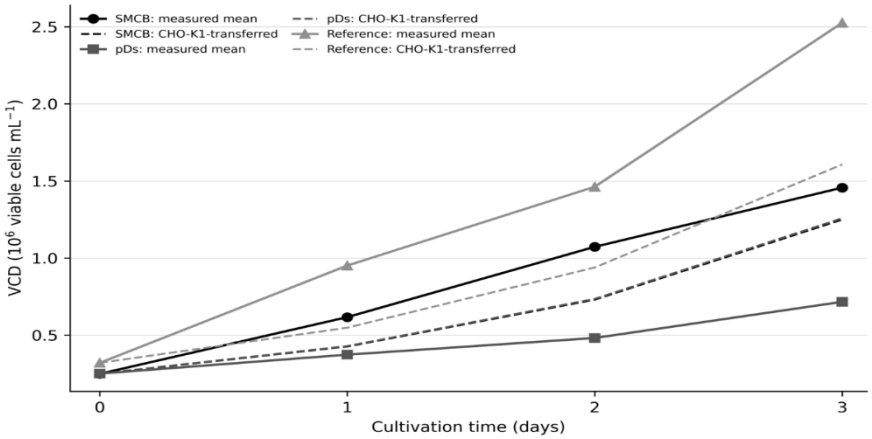
Table 4. Descriptive errors for reduced CHO-K1-trained growth estimators transferred without refitting to the three NISTCHO configuration-mean trajectories from a single external study (days 1-3).

NISTCHO system	Scored days (n)	Constant MAE	Exponential MAE	Log-quadratic MAE	Exponential RMSE
SMCB	3	0.800	0.247	0.367	0.256
pDs	3	0.273	0.284	0.461	0.347
Reference	3	1.326	0.615	0.416	0.653
Unweighted mean	—	0.800	0.382	0.415	0.419

Note: MAE and RMSE are in 10^6 viable cells/mL. Each configuration contributes three scored daily means; underlying curves summarize biological duplicates, not individual replicate-level trajectories. The across-configuration arithmetic mean is descriptive, not a replicate-level estimate; the mean row aggregates three configurations, not three scored days. Day-zero VCD was provided for initialization. No external confidence interval or significance test was computed. Source: reanalysis of Uhlendorff *et al*. (2026) and Barczyk *et al*. (2026).

The mean descriptive MAE across three NISTCHO configurations was 0.382 × 10^6 cells/mL for the CHO-K1-trained exponential model versus 0.800 × 10^6 cells/mL for constant initial VCD (a 52.3% relative difference). Performance was better for SMCB and Reference but slightly worse for pDs (0.284 *vs* 0.273). The log-quadratic model yielded mean MAE 0.415. Table 4 and Figure 4 show this system dependence. These errors refer to aggregate daily means from one external study, not independent biological-replicate validation or a seven-sensor soft-sensor test.

Figure 4. Early (days 0-3) NISTCHO viable-cell-density means and transferred CHO-K1 exponential predictions for SMCB, pDs and Reference. Day-zero means calibrate the initial condition; only days 1-3 are scored. Solid lines and symbols show observed biological-duplicate means; dashed lines show predictions with CHO-K1-only fitted growth rate (0.53728 d^{-1}). The plotted lines connect observations and do not imply more frequent sampling.



Source: reanalysis of Uhlendorff *et al*. (2026) and Barczyk *et al*. (2026).

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

With day-zero VCD supplied, CHO-K1 growth kinetics transferred unevenly to NISTCHO: pDs did not improve. Three system-level mean curves from one study cannot establish replicate-level uncertainty or cross-site and hardware transfer. The original seven-sensor model was not externally tested.

4. Discussion

4.1. Bioreactor engineering and the information carried by sensors

Oxygen transfer, mixing, control interventions and dissolved-gas dynamics influence inline measurements. A dissolved-oxygen signal is not equivalent to cellular oxygen uptake without gas-transfer and control-action data, and mass-transfer parameters depend on reactor geometry and operating regime (Nienow, 2006; Pineda Rincon *et al*., 2025). No mechanistic oxygen-transfer model was calibrated in this study. Oxygen-based estimates of CHO glucose demand reported elsewhere address a different outcome (Jossen *et al*., 2026).

4.2. Predictive estimation is not autonomous control

Machine-learning predictions alone do not constitute robust real-time control. A deployable soft sensor needs an independent reference target, defined operating and evaluation domains, characterized uncertainty, drift detection, and documented performance on unseen campaigns (Helleckes *et al*., 2023; Wasalathanthri *et al*., 2020). An operational digital twin additionally requires synchronization with the physical process, structured communication, and control actions bounded by safety constraints. The framework discussed by Satwekar *et al*. (2026) supports distinguishing modeling capability from connected operation. Hybrid CHO models can be informative but depend on training-set size, predictor collinearity and operating context (Doyle *et al*., 2026; Ghodba *et al*., 2025; Richelle *et al*., 2026; Robitaille; Chen; Jolicoeur, 2015). Grouping both reactors in each campaign reduces leakage; it does not remove unmeasured process differences or post-hoc model-selection bias.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

4.3. Biological interpretation: glucose, lactate, and product quality

The primary CHO-K1 reanalysis describes viable biomass and extracellular metabolites, not monoclonal-antibody titer or glycosylation. Peak VCD or low lactate must therefore not be described as evidence of superior therapeutic-protein output. In the separate NISTCHO system, Schäpertöns *et al*. (2026) compared controlled feeding conditions and found differences in culture performance and glycan profiles, with reduced glucose availability associated with impaired titer and altered N-glycosylation.

Walsh *et al*. (2025) used a different CHO fed-batch dataset to model both titer and glycosylation-related responses. Dynamic CHO metabolic modeling and metabolite reviews likewise distinguish nutrient states, growth and product-related endpoints (Nolan; Lee, 2011; Pereira; Kildegaard; Andersen, 2018). These findings justify considering product-specific endpoints in future work; they are not outcomes measured in the reanalyzed CHO-K1 dataset and cannot be imported into the present model.

4.4. Campaign-level independence and inferential strength

Although 80 time-point predictions were examined, only nine independent batch campaigns were available to evaluate early-period transfer between experimental runs. Treating all 80 predictions as separate biological replicates would produce spuriously narrow uncertainty intervals. LOEO grouping limits leakage between trajectories but does not remove common-parameter bias, irregular sampling, or undocumented process differences. The inferential problem is particularly acute for the three fed-batch campaigns, which were therefore used primarily for descriptive characterization. Recent CHO hybrid-model studies underscore the importance of combining physiological knowledge, operating conditions, and genuinely independent assessment (Peng *et al*., 2026; Richelle *et al*., 2020; Richelle *et al*., 2026; Wang; Harcum; Xie, 2026).

4.5. Sensor physics and limits of concurrent estimation

Dielectric and spectroscopic measurements provide potentially informative signals, but responses depend on cellular and instrumental properties, including cell

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

size, membrane-related electrical properties, and medium composition (Lemke *et al*., 2023; Opel; Li; Amanullah, 2010; Salimi *et al*., 2025). I evaluated VCD at the same time as an offline sample, not future states. Prospective application would require characterization of latency, drift, recalibration, and performance in newly acquired campaigns. Estimating present-time VCD does not imply prediction of future critical quality attributes.

4.6. Comparison with independent CHO monitoring studies

Experimental literature supports physical plausibility of sensor-enhanced estimation but not numerical equivalence across bioprocesses. Lemke *et al*. (2023) assessed dielectric spectroscopy with online O-PLS in CHO perfusion, whereas Salimi *et al*. (2025) showed that impedance responses vary with cell concentration, size and dielectric characteristics. Metze *et al*. (2020) described capacitance-assisted biomass monitoring during CHO fed-batch scale-up. Suman *et al*. (2025) combined optical and dielectric information in a separate industrial case study.

In contrast, Pinto *et al*. (2023) used a hybrid mass-balance/deep-learning approach on a distinct CHO-K1 fed-batch dataset; those 24 cultivations should not be confused with the 24 cultivations reanalyzed here. Commercial-scale Raman monitoring and generic Raman calibrations have distinct measurement and calibration requirements (Mehdizadeh *et al*., 2015; Yan *et al*., 2024); a separate Raman-dielectric data-fusion study targeted arginine, not the present VCD outcome (Hajdú *et al*., 2026). Because endpoints, operating modes and validation designs differ, no cross-study pooling of prediction errors is justified.

The present study uses seven contemporaneous signals and regularized regression against nonlinear time-only baselines. Full-period batch errors decreased in most held-out campaigns, but improvements were inconsistent during the first 72 h and the same sensor panel was not independently tested. In contrast, previously reported Raman-guided feeding studies implemented operational interventions (Domján *et al*., 2022; Gibbons *et al*., 2023; Webster *et al*., 2021; Webster *et al*., 2023). Real-world application requires drift monitoring, missingness management,

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

independent hardware calibration and prospectively held-out data (Metcalfe *et al*., 2025; Satwekar *et al*., 2026; Siegl *et al*., 2024).

These limitations also apply to scale transfer. Controlled scale-down CHO experiments show that mixing heterogeneity can alter cell physiology and process signals (Gaugler *et al*., 2024). Thus, apparent sensor gain within one instrument platform cannot establish generalization across scales or independent sites.

4.7. Internal validity and limited cross-dataset transfer

Holding out both reactors in an entire campaign prevented that campaign from contributing to spline fitting, feature standardization and Ridge hyperparameter selection for the specified nested comparison. Nevertheless, the predictor panel, alternative algorithms and analysis horizons had been explored during development. The resulting intervals and paired comparisons are conditional on these exploratory choices and nine batch campaigns; they should not be interpreted as confirmatory estimates of cross-site generalization. Moreover, because reference VCD and instrument features were indexed to the same time point, the task was concurrent estimation, not forecasting. Reviews of cross-domain soft sensors and bioprocess transfer learning similarly warn against untested transportability (Barón Díaz *et al*., 2026; Liu *et al*., 2026).

Missing sensor matches may also select which biological phases are represented. Among the three fed-batch campaigns, the median VCD was approximately 0.265×10^6 cells/mL in 16 unmatched samples, compared with 3.315×10^6 cells/mL among 96 matched samples. This unadjusted contrast cannot identify the cause of missingness, but it suggests that complete-case analysis may underrepresent declining cultures. The extended interruption in EXP011/R02 further illustrates that sensor availability must be monitored alongside measurement precision.

Barczyk *et al*. (2026) supplied independent NISTCHO measurements for a restricted time-only transfer assessment. The exponential model improved over a constant-VCD baseline for SMCB and Reference but was slightly worse for pDs (Table 4; Figure 4). All three series came from a single external study, were available

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

only as means of biological duplicates, and were initialized with observed day-zero VCD. Mixing and cell-line differences could influence transfer, but this small descriptive comparison cannot establish causal mechanisms, statistically reliable transportability or prospective forecasting. The seven-signal CHO-K1 regression remains unvalidated on a matching independent instrument panel. Independent reactor-level runs with harmonized measurements, prespecified evaluation rules and uncertainty quantification are required.

5. Limitations, Bias, and External Validity

Inference is limited to nine batch and three fed-batch CHO-K1 experimental campaigns. Only 350 of 378 offline samples satisfied the seven-signal matching rule; missingness may bias complete-case results. Sensor sets, model families and analysis horizons were explored during development, so cross-validation results and bootstrap intervals are exploratory rather than prospective or preregistered confirmation. The sensor model estimates contemporaneous VCD, not future antibody titer, glycosylation or control actions.

The reduced external analysis used only three NISTCHO configuration means from one study, each derived from biological duplicates, and three post-baseline scored days per configuration, with measured day-zero VCD supplied for initialization. It is a descriptive transfer stress test, not a replicate-level external validation of CHO-K1 kinetics or any validation of the seven-sensor model. Finally, the supporting literature was used for contextual discussion without a systematic search or formal critical-appraisal procedures.

6. Conclusions

Valid soft-sensor evaluation requires both synchronized measurements and independent test campaigns. Of 378 CHO-K1 laboratory samples, 350 had usable seven-signal matches; the other 28 remained documented rather than imputed. Under post-hoc nested batch-campaign assessment, the mean MAE decreased from

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

1.346 to 0.956 x 10^6 cells/mL with seven concurrent signals added to cultivation time (28.9% exploratory relative difference).

Eight of nine batch campaigns improved over their full observed trajectories, but the effect was inconsistent during the first 72 h. In a distinct reduced transfer test, CHO-K1-trained exponential growth kinetics gave a descriptive NISTCHO mean MAE of 0.382 *versus* 0.800 × 10^6 cells/mL for a constant-initial-VCD baseline across days 1-3. That improvement was not uniform across the three configurations and required measured day-zero external VCD. Neither test demonstrates transportability of the original seven-sensor model.

The principal contribution is an auditable retrospective workflow for sensor time alignment, campaign-grouped model assessment, missing-signal reporting and a deliberately restricted cross-dataset growth stress test. All reported principal numerical estimates were recomputed from the available source-derived matrices, and original matching summaries were compared against the primary aligned files (see supporting reproducibility package). Independent reactor-level validation with matching instrumentation remains necessary before seven-signal estimation can support operational decisions.

Declaration of generative AI and AI-assisted technologies

OpenAI ChatGPT was used to assist with English-language editing, mathematical exposition, *Python* code development, and figure/table presentation. The analysis code and source-derived outputs were computationally checked against the deposited primary data, but the author retains responsibility for independent scientific judgment, interpretation, final verification and submission.

References

BARCZYK, J.; LOPEZ MEDINA, B.; TELEKI, A.; *et al*. Data for: Scale-down bioreactors reveal that prolonged mixing times deteriorate the performance of antibody-producing Chinese hamster ovary cells. *DaRUS*, 2026. DOI: https://doi.org/10.18419/DARUS-6466.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

BARÓN DÍAZ, D.; DROMMERSHAUSEN, A. L.; GRÜNBERGER, A.; HOLTMANN, D. Transfer Learning Approaches in Bioprocess Engineering: Opportunities and Challenges. *Biotechnol Bioeng.*, v. 123, p. 1417-1431, 2026. DOI: https://doi.org/10.1002/bit.70186.

DOMJÁN, J.; PANTEA, E.; GYÜRKÉS, M.; *et al*. Real-time amino acid and glucose monitoring system for the automatic control of nutrient feeding in CHO cell culture using Raman spectroscopy. *Biotechnol J.*, v. 17, p. e2100395, 2022. DOI: https://doi.org/10.1002/biot.202100395.

DOYLE, K.; YANG, O.; COLARUSSO, T.; BANO, G.; GLENNON, B.; JIMÉNEZ DEL VAL, I. Hybrid Dynamic Modelling With Gaussian Process Regression for Intensified Fed-Batch CHO Cell Culture Processes. *Biotechnol Bioeng.*, 2026. DOI: https://doi.org/10.1002/bit.70339.

GAUGLER, L.; HOFMANN, S.; SCHLÜTER, M.; TAKORS, R. Mimicking CHO large-scale effects in the single multicompartment bioreactor: A new approach to access scale-up behavior. *Biotechnology and Bioengineering*, v. 121, p. 1243–1255, 2024. DOI: https://doi.org/10.1002/bit.28647.

GHODBA, A.; RICHELLE, A.; MCCREADY, C.; RICARDEZ-SANDOVAL, L.; BUDMAN, H. A novel dynamic flux balance analysis for modeling CHO cell fed-batch cultures with pH and temperature shifts. *J Biotechnol*., v. 408, p. 61-71, 2025. DOI: https://doi.org/10.1016/j.jbiotec.2025.08.010.

GIBBONS, L.; MASLANKA, F.; LE, N.; *et al*. An assessment of the impact of Raman based glucose feedback control on CHO cell bioreactor process development. *Biotechnol Prog.*, v. 39, p. e3371, 2023. DOI: https://doi.org/10.1002/btpr.3371.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

HAJDÚ, D. K.; MAROSI, G. J.; NAGY, Z. K.; *et al*. Advancing bioprocess monitoring: data fusion and ANN-based prediction of arginine concentration in monoclonal antibody-producing CHO cell cultures. *New Biotechnol.*, v. 93, p. 185–193, 2026. DOI: https://doi.org/10.1016/j.nbt.2026.03.003.

HELLECKES, L. M.; HEMMERICH, J.; WIECHERT, W.; VON LIERES, E.; GRÜNBERGER, A. Machine learning in bioprocess development: From promise to practice. *Trends Biotechnol*., v. 41, p. 817-835, 2023. DOI: https://doi.org/10.1016/j.tibtech.2022.10.010.

JOSSEN, V.; JOSEPH, A.; MESADO, G.; LIM, S.; BROCCARD, G. Oxygen-based soft-sensing of glucose demand in Chinese hamster ovary fed-batch cultures across scales. *Appl Microbiol Biotechnol*., v. 110, p. 257, 2026. DOI: https://doi.org/10.1007/s00253-026-13882-2.

KARAGEORGOS, I.; O'DELL, W.; BOLLA, V.; *et al*. NISTCHO Cell Culture and Fed-Batch Bioreactor Protocols. *NIST Special Publication*. 1200-32, 2026. DOI: https://doi.org/10.6028/NIST.SP.1200-32.

LEMKE, J.; SÖLDNER, R.; AUSTERJOST, J.; *et al*. Online deployment of an O-PLS model for dielectric spectroscopy-based inline monitoring of viable cell concentrations in Chinese hamster ovary cell perfusion cultivations. *Engineering in Life Sciences*, v. 23, p. e2200053, 2023. DOI: https://doi.org/10.1002/elsc.202200053.

LIU, Y.; ZHU, J.; YANG, C.; CHEN, T.; WONG, D. S. H.; YAO, Y. Transfer Learning for Soft Sensors in Process Industries: A Review and Future Perspectives. *Ind Eng Chem Res.*, v. 65, p. 8103-8125, 2026. DOI: https://doi.org/10.1021/acs.iecr.5c05144.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

MEHDIZADEH, H.; LAURI, D.; KARRY, K. M.; MOSHGBAR, M.; PROCOPIO-MELINO, R.; DRAPEAU, D. Generic Raman-based calibration models enabling real-time monitoring of cell culture bioreactors. *Biotechnol Prog.*, v. 31, p. 1004-1013, 2015. DOI: https://doi.org/10.1002/btpr.2079.

METCALFE, B.; ACOSTA-PAVAS, J. C.; ROBLES-RODRIGUEZ, C. E.; *et al*. Towards a machine learning operations (MLOps) soft sensor for real-time predictions in industrial-scale fed-batch fermentation. *Computers & Chemical Engineering*, v. 194, p. 108991, 2025. DOI: https://doi.org/10.1016/j.compchemeng.2024.108991.

METZE, S.; RUHL, S.; GRELLER, G.; GRIMM, C.; SCHOLZ, J. Monitoring online biomass with a capacitance sensor during scale-up of industrially relevant CHO cell culture fed-batch processes in single-use bioreactors. *Bioprocess Biosyst Eng.*, v. 43, p. 193–205, 2020. DOI: https://doi.org/10.1007/s00449-019-02216-4.

NIENOW, A. W. Reactor Engineering in Large Scale Animal Cell Culture. *Cytotechnology*, v. 50, p. 9–33, 2006. DOI: https://doi.org/10.1007/s10616-006-9005-8.

NOLAN, R. P.; LEE, K. Dynamic model of CHO cell metabolism. *Metab Eng.*, v. 13, p. 108-124, 2011. DOI: https://doi.org/10.1016/j.ymben.2010.09.003.

OPEL, C. F.; LI, J.; AMANULLAH, A. Quantitative modeling of viable cell density, cell size, intracellular conductivity, and membrane capacitance in batch and fed-batch CHO processes using dielectric spectroscopy. *Biotechnol Prog.*, v. 26, p. 1187-1199, 2010. DOI: https://doi.org/10.1002/btpr.425.

PENG, J.; KHUAT, T. T.; OTTE, E.; MUSIAL, K.; GABRYS, B. Learning from limited data and feedback for cell culture process monitoring: A comparative study. *Chem Eng J.*, v. 535, p. 175264, 2026. DOI: https://doi.org/10.1016/j.cej.2026.175264.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

PEREIRA, S.; KILDEGAARD, H. F.; ANDERSEN, M. R. Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients. *Biotechnol J.,* v. 13, p. 1700499, 2018. DOI: https://doi.org/10.1002/biot.201700499.

PINEDA RINCON, D. A.; CORRALES, D. C.; DABOUSSI, F.; CESCUT, J.; FILLAUDEAU, L.; ACEVES-LARA, C. A. Benchmarking and Trends Around In-situ and Ex-situ Sensors in Bioreactors. *Chemical Engineering Transactions*, v. 117, p. 103–108, 2025. DOI: https://doi.org/10.3303/CET25117018.

PINTO, J.; RAMOS, J. R. C.; COSTA, R. S.; ROSSELL, S.; DUMAS, P.; OLIVEIRA, R. Hybrid deep modeling of a CHO-K1 fed-batch process: Combining first-principles with deep neural networks. *Front Bioeng Biotechnol.*, v. 11, p. 1237963, 2023. DOI: https://doi.org/10.3389/fbioe.2023.1237963.

REARDON, K. F. Practical monitoring technologies for cells and substrates in biomanufacturing. *Curr Opin Biotechnol*., v. 71, p. 225-230, 2021. DOI: https://doi.org/10.1016/j.copbio.2021.08.006.

REYES, S. J.; DUROCHER, Y.; PHAM, P. L.; HENRY, O. Modern Sensor Tools and Techniques for Monitoring, Controlling, and Improving Cell Culture Processes. *Processes*, v. 10, p. 189, 2022. DOI: https://doi.org/10.3390/pr10020189.

REZENDE, R. B. Cultivo de células CHO e da levedura Komagataella phaffii na produção de bioterapêuticos: revisão integrativa. *RSV*, v. 8, n. 10, p. 1–22, 2026. DOI: https://doi.org/10.66104/7rgz3791.

RICHELLE, A.; LEE, B. W.; PORTELA, R. M. C.; RALEY, J.; VON STOSCH, M. Analysis of Transformed Upstream Bioprocess Data Provides Insights into Biological

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

System Variation. *Biotechnol J.*, v. 15, p. 2000113, 2020. DOI: https://doi.org/10.1002/biot.202000113.

RICHELLE, A.; ANDERSSON, D.; ANTONAKOUDIS, A.; JAKOBSSON, J.; PIJEAUD, S.; *et al*. A Hybrid Modeling Framework for Predictive Digital Twins of CHO Cell Culture. *Computational and Structural Biotechnology Journal*, v. 35, p. 0078, 2026. DOI: https://doi.org/10.34133/csbj.0078.

ROBITAILLE, J.; CHEN, J.; JOLICOEUR, M. A Single Dynamic Metabolic Model Can Describe mAb Producing CHO Cell Batch and Fed-Batch Cultures on Different Culture Media. *PLoS One*, v. 10, p. e0136815, 2015. DOI: https://doi.org/10.1371/journal.pone.0136815.

SALIMI, E.; ABSALAN, S.; ROBITAILLE, J.; MONTES, J.; BUTLER, M.; THOMSON, D.; BRIDGES, G. Sensitivity of bulk electrical impedance spectroscopy (bio-capacitance) probes to cell and culture properties: Study on CHO cell cultures. *Biotechnology Progress*, v. 41, p. e3519, 2025. DOI: https://doi.org/10.1002/btpr.3519.

SATWEKAR, A.; GRACIEUX-SINGLETON, M. C.; CUMMINGS, C.; *et al*. Operationalizing digital twins in biomanufacturing through interoperable process analytical technology. *Bioprocess and Biosystems Engineering*, v. 49, p. 2061–2083, 2026. DOI: https://doi.org/10.1007/s00449-026-03369-9.

SCHÄPERTÖNS, V.; HOFER, L.; BERGER, T.; *et al*. Effects of feeding strategies on culture performance and product quality in NISTCHO. *npj Systems Biology and Applications*, v. 12, p. 67, 2026. DOI: https://doi.org/10.1038/s41540-026-00686-3.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

SHARMA, R.; HARRISON, S. T. L.; TAI, S. L. Advances in Bioreactor Systems for the Production of Biologicals in Mammalian Cells. *ChemBioEng Rev.*, v. 9, p. 42–62, 2022. DOI: https://doi.org/10.1002/cben.202100022.

SIEGL, M.; GEIER, D.; ANDREESSEN, B.; *et al*. Data synchronization techniques and their impact on the prediction performance of automated recalibrated soft sensors in bioprocesses. *Biotechnology and Bioprocess Engineering*, v. 29, p. 929–941, 2024. DOI: https://doi.org/10.1007/s12257-024-00120-7.

SUMAN, S.; MURR, M.; CROWE, J.; *et al*. In-line prediction of viability and viable cell density through machine learning-based soft sensor modeling and an integrated systems approach: An industrially relevant PAT case study. *Biotechnology Progress*, v. 41, n. 3, p. e3520, 2025. DOI: https://doi.org/10.1002/btpr.3520.

SUN, Y.; ZHANG, Q.; HE, Y.; *et al*. Real-Time Auto Controlling of Viable Cell Density in Perfusion Cultivation Aided by In-Line Dielectric Spectroscopy With Segmented Adaptive PLS Model. *Biotechnology and Bioengineering*, v. 122, n. 4, p. 858–869, 2025. DOI: https://doi.org/10.1002/bit.28930.

UHLENDORFF, S.; FULEK, R.; EIMLER, J.; PEIN-HACKELBUSCH, M.; FRAHM, B. Dataset Based on Chinese Hamster Ovary (CHO) Cultivations including Turbidity, Permittivity, O2 and CO2 Measurements. *Zenodo*, 2026. DOI: https://doi.org/10.5281/zenodo.20829178.

WALSH, I.; SHOZUI, F.; SATO, A.; PARK, U.; CHO, H.; CHIA, C. Toward Machine Learning-Guided CHO Bioprocess and Media Optimization for Improved Titer and Glycosylation. *Biotechnol J.*, v. 20, p. e70149, 2025. DOI: https://doi.org/10.1002/biot.70149.

CC BY

RSV
ISSN 2674-8584

Received: 10/10/2026 - **Accepted:** 08/10/2026
Vol: 08.11
DOI: 10.66104/actz3g62
Pages: 1-33

WANG, K.; HARCUM, S. W.; XIE, W. Multi-Scale Hybrid Modeling to Predict Cell Culture Process With Metabolic Phase Transitions. *Biotechnology and Bioengineering*, v. 123, n. 7, p. 1745–1770, 2026. DOI: https://doi.org/10.1002/bit.70205.

WASALATHANTHRI, D. P.; REHMANN, M. S.; SONG, Y.; *et al*. Technology outlook for real-time quality attribute and process parameter monitoring in biopharmaceutical development: A review. *Biotechnol Bioeng*., v. 117, p. 3182-3198, 2020. DOI: https://doi.org/10.1002/bit.27461.

WEBSTER, T. A.; HADLEY, B. C.; DICKSON, M.; BUSA, J. K.; JAQUES, C.; MASON, C. Feedback control of two supplemental feeds during fed-batch culture on a platform process using inline Raman models for glucose and phenylalanine concentration. *Bioprocess Biosyst Eng.*, v. 44, p. 127-140, 2021. DOI: https://doi.org/10.1007/s00449-020-02429-y.

WEBSTER, T. A.; HADLEY, B. C.; DICKSON, M.; *et al*. Automated Raman feedback control of multiple supplemental feeds to enable an intensified high inoculation density fed-batch platform process. *Bioprocess Biosyst Eng.*, v. 46, p. 1457–1470, 2023. DOI: https://doi.org/10.1007/s00449-023-02912-2.

YAN, X.; DONG, X.; WAN, Y.; *et al*. Development of an in-line Raman analytical method for commercial-scale CHO cell culture process monitoring: Influence of measurement channels and batch number on model performance. *Biotechnol J.*, v. 19, p. e2300395, 2024. DOI: https://doi.org/10.1002/biot.202300395.

CC BY