

ARTICLE

A Biomanufacturing Digital Twin for Consistent N-1 Seed Bioreactor Transfers for Commercial Monoclonal Antibody Production

Yu Luo¹  | Sumit Pareek² | Carolyn Zhang³ | Kelly Sarnese⁴ | Diana Ritz⁵ | Sameer Talwar¹ 

¹GSK, Predictive Sciences, Upper Merion, Pennsylvania, USA | ²GSK, Global Supply Chain Digital and Technology, Upper Merion, Pennsylvania, USA | ³GSK, Enterprise AI, Boston, Massachusetts, USA | ⁴GSK, MSAT, Upper Merion, Pennsylvania, USA | ⁵GSK, Biopharmaceutical Drug Substance Development, Upper Merion, Pennsylvania, USA

Correspondence: Yu Luo (yu.8.luo@gsk.com)

Received: 19 March 2026 | **Revised:** 17 August 2026 | **Accepted:** 23 September 2026

ABSTRACT

Process consistency in biomanufacturing, particularly for commercial products, is critical for achieving consistent yield and product quality. The N-1 seed bioreactor stage, where cells grow before the transfer to a production bioreactor, often faces variability in final viable cell density (VCD), which leads to variable seeding densities in the production bioreactor and subsequent variability in total protein output per batch. In this study, we address the challenge of precise batch transfer by developing an upstream, stage-specific, level-2 digital twin (digital shadow). Here we show that a cell growth process model predicts growth accurately and updates forecasts using off-line and in-line measurements, enabling precise determination of the final sample time. Deployment of this seed transfer digital twin led to all N-1 to production transfers falling within acceptable limits and resulted in statistically significant increases in harvest yield relative to historical batches. This improvement is attributed to significantly higher on-target accuracy of final N-1 VCDs and significantly higher working volumes in the production bioreactors. Our findings demonstrate that such digital twins ensure consistent cell density during transfer, improve overall upstream process consistency, and lay a foundation for more advanced process control strategies.

1 | Introduction

In biomanufacturing, it is critical to develop commercial manufacturing processes that result in high process consistency and product quality. Certain sources of variability can be managed directly by controlling process variables via standard process control—for example, pH, temperature, and dissolved oxygen (DO). Other variables (e.g., culture duration) require operation within a specific range but are traditionally managed via scheduling tools instead of process control mechanisms. In a standard biomanufacturing process, cells are scaled up through a series of shake flasks and bioreactors until sufficient biomass is generated to inoculate the production bioreactor. Manufacturing processes are developed with a range of acceptable final viable cell densities (VCDs) across the seed stages. While this is a requirement to enable operational flexibility, it results in batch-to-batch variability,

including variation in culture duration, culture volume, and split ratio (the ratio of fresh medium to spent medium). For the transfer from the N-1 seed bioreactor (the stage prior to the production bioreactor) to the production bioreactor, reducing this variation has the benefit of improving cell health, cell productivity, and culture working volume in the production bioreactor, ultimately resulting in increased productivity from the manufacturing process. In this article, we develop a digital twin that can accurately forecast N-1 cell growth to reduce this variation.

A digital twin represents a “virtual copy” of the physical manufacturing process (Singh et al. 2021; Park et al. 2023, 2021), which can then be used to experiment on, monitor, or control the process without interrupting it. The core component of a digital twin is a predictive model of the process (in the case of this article, a cell

All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

© 2026 Wiley Periodicals LLC.

growth model for the N-1 process). The digital twin updates its forecast throughout the cell culture duration based on the model and new information available from off-line sampling and in-line probe signals. The quality of the forecasts depends on the accuracy of the model and the ability of measurements to represent the cell growth state.

There is an abundance of literature on techniques to monitor and predict cell growth. Process analytical technology (PAT) tools have been developed and deployed in the N-1 stage to measure VCD in real time (Gillespie et al. 2022). Specifically, in-line capacitance probes (Moore et al. 2019; Olin et al. 2024) have demonstrated utility in monitoring VCD in industrial-scale bioreactors (Rittershaus et al. 2022; Pohlscheidt et al. 2013; Xu et al. 2020). Raman spectroscopy, while not a primary tool for VCD monitoring, has the capability to measure VCD, using multivariate chemometric models (Yan et al. 2024; Goldrick et al. 2020). Real-time measurements are useful for monitoring and controlling VCD; however, to determine the time to transfer a batch from one bioreactor to another, predictive cell growth models are necessary to generate that forecast in advance to enable manufacturing operators to ensure the subsequent stage is set up and ready to accept the donating cell culture material.

The mechanism and dynamics of cell growth, especially in the seed train, are well-known. Several cell growth models have been created to describe the dynamics mechanistically (Hernández Rodríguez 2020; Hernández Rodríguez et al. 2022; Frahm 2014; Richelle et al. 2022; Luo et al. 2021; Kontoravdi 2007). In those models, VCD is modeled typically by a combination of cell growth, lysis, and death dynamics. VCD is one of several state variables that describe cell culture processes, including concentrations of metabolites, nutrients, and antibodies. Data-driven and hybrid models have also been developed to use process data efficiently and forecast process trajectories (Park et al. 2023; Bayer et al. 2021; Pinto et al. 2023; Agarwal et al. 2025). Existing cell growth models rarely address the operational need for precise N-1 batch transfer; in contrast, the model presented here was purposely built as a stage-specific digital twin to forecast when a predefined VCD target will be reached and thereby enable precise, Good Manufacturing Practice (GMP)-compatible transfer scheduling. The contribution of this work is not developing a new cell growth model but using the model to answer one specific operational question—when a predefined transfer VCD will be reached. To our knowledge, this is the first reported at-scale commercial deployment with on-target accuracy and positive inoculation volume/harvest yield effects.

In this article, we develop a process model to predict cell growth and update the prediction based on the latest measurements from off-line and in-line analytical instruments (i.e., one-step ahead forecast). The goal of this model and the digital twin is to transfer the N-1 batch when VCD reaches a predetermined target. To maximize the final protein at harvest, we aim for maximum inoculum volume in the production bioreactor when the final N-1 VCD is as close as possible to its predetermined transfer target:

$$V_{\text{inoc}} = \frac{x_v^* V_{\text{pre}}}{x_{v,f} - x_v^*}, \quad (1)$$

where V_{inoc} is the inoculum volume of the production bioreactor; x_v^* is the target initial VCD (i.e., the seeding density) in

the production batch; V_{pre} is the volume of the precharged medium, which contains no cells; $x_{v,f}$ is the final VCD in the N-1 reactor.

2 | Methods

The cell growth model in this digital twin consists of two components: a linear regression model that predicts the *specific* cell growth rate based on measured process variables and an exponential growth model that predicts future VCD based on the growth rate:

$$\hat{\mu} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p, \quad (2)$$

$$\hat{x}_v(t_1) = x_v(t_0) e^{\hat{\mu}(t_1 - t_0)}, \quad (3)$$

where $\hat{\mu}$ is the predicted specific growth rate based on p measured variables, $x_1, \dots, x_p$; $\beta_0, \dots, \beta_p$ are $p + 1$ model coefficients/parameters (intercept and slopes); $x_v(t_0)$ is the measured VCD at time t_0 ; $\hat{x}_v(t_1)$ is the predicted VCD at a later time t_1 (typically the first off-line sample when VCD is above the acceptable lower limit of 4 million cells/mL). The data in the training sets were cleaned (by retaining observations where VCD and all predictor variables have values), normalized (zero mean and unit variance), and then used to estimate the parameters via ordinary least squares linear regression. Once the specific growth rate model is available, future VCD can be predicted based on exponential growth over a period between t_0 and t_1 . The simplicity of the model structure is intentional and enabled rapid development and deployment to meet the operational objective of transferring an N-1 batch on target.

Leave-one-batch-out cross-validation was used to evaluate how the model generalizes to new data. Specifically, one batch from the training batches in Table 1 was left out each time. The remaining batches were used to estimate the model parameters in (2). This model was then used to predict VCD in the left-out batch based on (3). The root-mean-squared error (RMSE) between measured VCD and predicted VCD was recorded. The cross-validation process ended when every training batch in Table 1 was left out once, and the mean RMSE and its spread were documented in Results and Discussion. The model was finalized by using the entire training set (no batches were left out) and tested against the testing set in Table 1. All batches in Table 1 were at-scale commercial manufacturing batches.

The time to reach a VCD target (i.e., true transfer time) is estimated *retrospectively* by estimating the specific growth rate of two consecutive VCD measurements that are below and above the target:

$$\tilde{\mu} = \frac{\ln x_v(t_1) - \ln x_v(t_0)}{t_1 - t_0}, \quad (4)$$

TABLE 1 | Datasets used in model development.

Dataset	Batch
Training sets	Batch 1, Batch 2, ..., Batch 13
Testing sets	Batch 14, Batch 15, Batch 16

where $\tilde{\mu}$ is the specific growth rate estimated from the two VCD measurements, $x_v(t_0)$ and $x_v(t_1)$. The true transfer time is estimated by inverting (3):

$$t^* = t_0 + \frac{\ln x_v^* - \ln x_v(t_0)}{\tilde{\mu}}, \quad (5)$$

where x_v^* is the transfer target (e.g., 4.3 million cells/mL; a safety factor of 0.3 million cells/mL was added to the process transfer target of 4 million cells/mL; while there was an upper transfer limit of 6 million cells/mL, it was used only in batches without the digital twin to determine process deviations); t^* is the estimated true transfer time. This approach is equivalent to interpolating the VCD measurements exponentially and intersecting the interpolated trend with the transfer target.

The time to reach a VCD target is forecasted by first predicting the specific growth rate based on the model and available measurements and then inverting (3):

$$\hat{t}^* = t_0 + \frac{\ln x_v^* - \ln x_v(t_0)}{\hat{\mu}}, \quad (6)$$

where $\hat{t}^*$ is the transfer time forecast; $\hat{\mu}$ is the predicted specific growth rate based on measured model variables at t_0 . Note that, due to scheduling constraints, only forecasts based on Day 0 or Day 1 measurements are used for this model development, although forecasts based on later measurements may be acceptable for other use cases. Forecasts based on later measurements are potentially more accurate but may not provide enough time for scheduling the final sample, set-up of the accepting bioreactor system, and the subsequent transfer.

The seed transfer digital twin is considered a “level-2 digital twin (digital shadow)” based on the digital twin definitions in Table 2 (Kritzinger et al. 2018). The model output is a datetime value that allows the scheduling team to inform the operations team to take a VCD sample. If the sample result is outside the acceptable range of VCD, the transfer cannot occur; otherwise, the transfer may occur. Use of this twin does not affect existing technical documentation, safety, quality, or process validation, as no reduction or change in sampling is required. The model prediction provides insight to assist with scheduling and does not affect GMP decisions or GMP data; therefore, it is classified as a level-2 digital twin (digital shadow).

Model results are visualized in a Microsoft Power BI dashboard (Figure 1) to guide operators in scheduling the final sampling.

The dashboard contains three components: a timetable, a predicted VCD trend figure, and a model accuracy table. Specifically, the timetable shows the transfer datetime forecast when VCD is projected to pass the transfer criterion based on (6). The operator uses the forecast to plan for the confirmation sample, the transfer to the production bioreactor, and the setting up of the production bioreactor.

The raw data for model training and testing contain eight off-line measurements (VCD, cell culture viability, glucose concentration, lactate concentration, pCO₂, pH, pH adjusted by instrument variability, and pO₂) and eight in-line probe measurements averaged over a 1-h period (bioreactor weight, oxygen micro sparge rate, pH, DO, base tank weight, antifoam tank weight, permittivity, and conductivity). The table was created from an enterprise data lake and a custom data pipeline. All historical and live data are hosted on Microsoft Azure to provide controlled environments, centralized access, and infrastructure for analytical consumption. 13 batches were used as the training set, and another three batches were used as the testing set. Model training, cross-validation, and testing were completed in Python using the scikit-learn and scipy packages. Statistical analysis was conducted in Python using the statsmodels package.

The N-1 stages were run prior to manufacturing the drug substance of a commercial product in the production facility. It is a Chinese hamster ovary (CHO) process. Cells were revived from a working cell bank in a proprietary medium. Cells were scaled up through a series of shake flasks and seed bioreactors to provide enough biomass. At the end of the production culture, the contents of the production bioreactors were harvested by centrifugation and direct depth filtration; the resulting clarified unprocessed bulk (CUB) was collected for further downstream processing. Off-line cell counts and viability were performed with a ViCell XR cell counter (Beckman, Indianapolis, IN). Glucose and other metabolites were measured using a BioProfile FLEX2 analyzer (Nova Biomedical, Waltham, MA).

3 | Results and Discussion

3.1 | Model Development

The cell growth model in (2) and (3) to support the digital twin was trained using the estimated specific growth rates and the off-line and in-line measurements. Among all 16 variables in the data table, four were selected as the final predictor variables in (2) (see Figure 2 for a histogram of each variable in the training dataset). They were normalized to zero mean and unit variance. The linear

TABLE 2 | Enterprise digital twin definitions.

Classification	Definition
Level-1 Digital twin (Digital model)	A digital model that provides historical or current insights via analytics or visualizations. There is unidirectional flow of data from the physical object to the digital object.
Level-2 Digital twin (Digital shadow)	A digital model that suggests actions via predictive analytics. There is unidirectional flow of data from the physical object to the digital object, as the shadow does not transfer data back to the physical object.
Level-3 Digital twin (Digital twin)	A digital model that prescribes actions for monitoring and control of the process. There is a bidirectional flow of data from the physical object to the digital object.

regression coefficients were estimated via ordinary least squares. The scaling parameters and estimated coefficients are listed in Table 3. The signs of the coefficients indicate the direction of the predictor effects. For instance, higher glucose, VCD, or DO levels are associated with lower specific growth rates.

The training, leave-one-batch-out cross-validation, and testing RMSEs of both the predicted specific growth rate, $\hat{\mu}$, in (2) and the predicted VCD, $\hat{x}_v$, in (3) are listed in Table 4. Note that the cross-validation and testing RMSEs are around 0.3 million cells/mL, the magnitude of which is consistent with the 0.3 million cells/mL safety factor that was applied to the target VCD value to ensure the acceptable range was met.

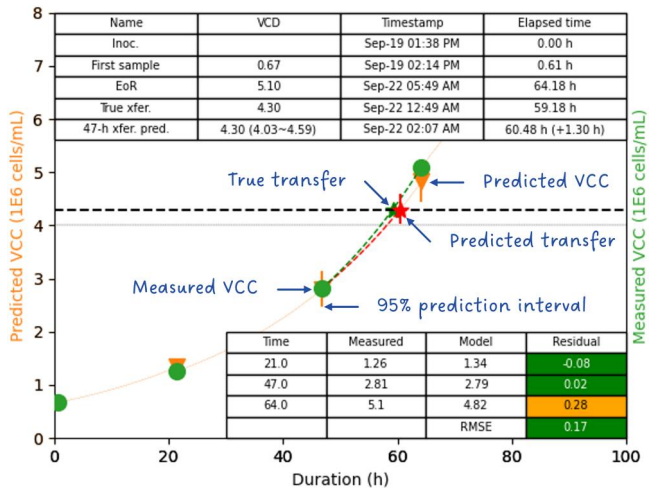


FIGURE 1 | Power BI dashboard.

The time series of VCD measurements and predictions are plotted in Figure 3 for the training and testing batches listed in Table 1. Specifically, the forecasted transfer time and the retrospectively estimated true transfer time show agreement. The worst forecast was 2.48 h later than the true transfer time for Batch 14, corresponding to a batch-wise RMSE of 0.51 million cells/mL. The elevated forecast error observed for Batch 14 likely reflects several contributing factors rather than a single cause, including measurement uncertainty in the ViCell cell counter, the simplified exponential structure of the growth model, and the long forecast horizon from Day 0/1. Because the model treats each ViCell VCD reading as a deterministic input and does not explicitly account for its own measurement error, the specific contribution of that measurement uncertainty was evaluated using a Monte Carlo simulation. The manufacturer’s stated comparability specification, system-to-system agreement within 10% at 95% confidence (Beckman, Indianapolis, IN), corresponding to a per-reading standard deviation of approximately 3.6%, was propagated through the deployed

TABLE 3 | Model parameter estimation results.

Variable name	Scaler (mean)	Scaler (scale)	Estimated coefficient
Intercept			0.0312
Glucose (g/L)	5.34	0.485	−0.000657
VCD (10 ⁶ /mL)	1.57	0.966	−0.00241
Viability (%)	98.9	0.682	0.00166
DO (%)	57.9	29.9	−0.000471

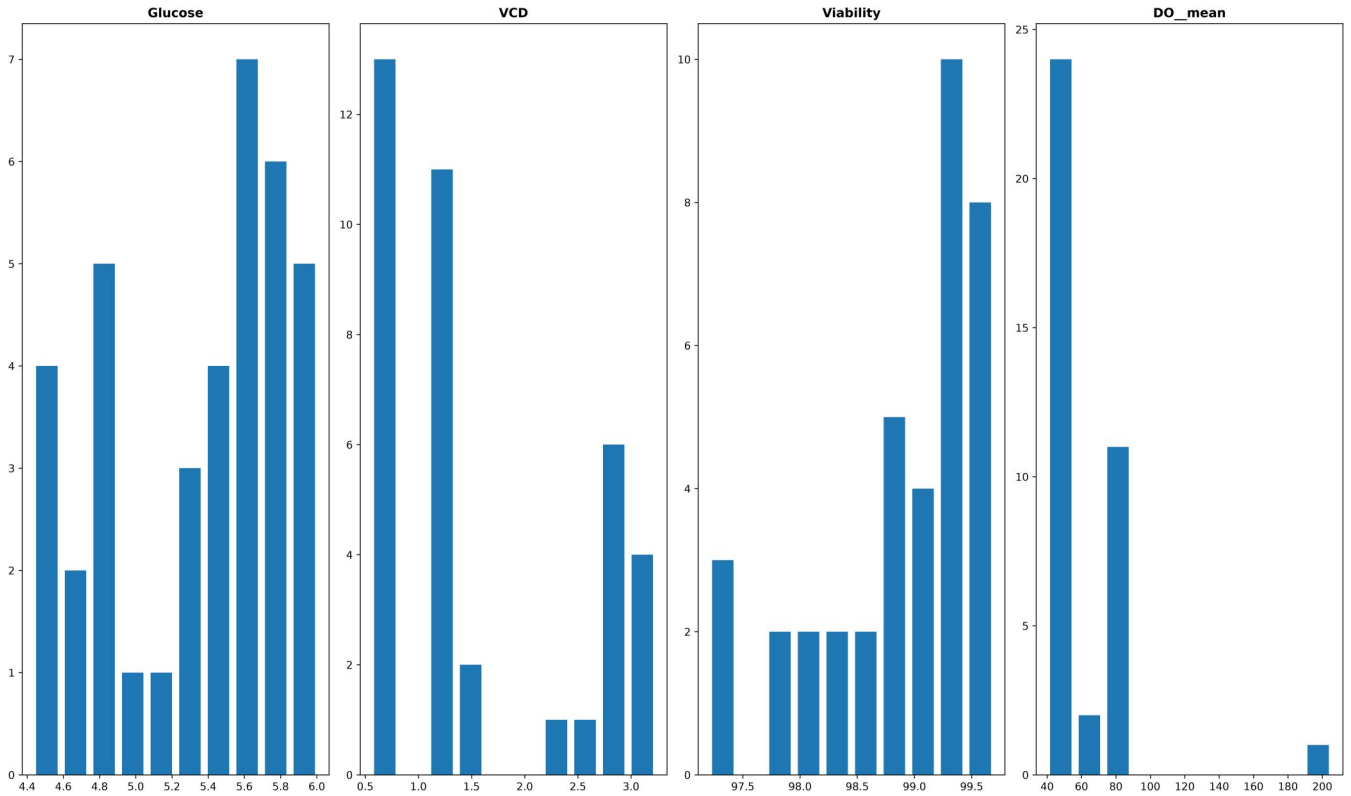


FIGURE 2 | Histograms of measured variables in the training dataset.

model over 20,000 draws per prediction point, yielding a standard deviation in transfer time of approximately 1.2 h, or 95% confidence interval (CI) of approximately ± 2.4 h, across the held-out test batches, and approximately 1.1 h (± 2.2 h) across all batches. The 2.48-h deviation observed for Batch 14 falls at the edge of this interval, indicating that ViCell measurement uncertainty alone is a quantitatively sufficient, though not the only plausible, explanation. The direction of the deviation is also operationally favorable: a later forecast corresponds to a higher VCD, so the confirmation sample would still be expected to fall above the transfer target and would not trigger a re-sample action, whereas an early forecast could place VCD below the acceptable lower limit. This source of uncertainty is distinct from the 95% prediction interval already reported in Figure 3, which reflects only the model's coefficient uncertainty and not the additional variability introduced by the VCD measurement itself. Reducing this input-side uncertainty in future iterations will require reconciling the ViCell reading with additional signals, such as in-line capacitance, through a state estimator, an approach already implemented in the successor version of this tool developed for a different process (see Limitations and Future Work).

TABLE 4 | Model performance.

Set	RMSE ($\hat{\mu}$, h^{-1})	RMSE ($\hat{x}_p$, $10^6/\text{mL}$)
Training	0.0023	0.24
Cross-validation	0.0025 ± 0.001	0.23 ± 0.06
Testing	0.0027	0.31

3.2 | Digital Twin Deployment and Results

The digital twin was deployed for three N-1 batches in the manufacturing campaign of the same product. In these batches, the final sample times were suggested by the twin and adopted in manufacturing. The batches without using the twin were also included in the analysis as a reference. All batches were contemporaneous commercial batches of the same product. There were no concurrent changes to the process besides the usage of the twin. The analysis in this section is framed according to a controlled operational comparison, not a powered causal study typically in the form of a design of experiments.

As a result, all N-1 transfers (with and without using the twin) were above the acceptable lower limit of 4 million cells/mL (Figure 4). Notably, the N-1 batches using the digital twin ended close to the predefined target of 4.3 million cells/mL with an RMSE of 0.094 million cells/mL, which is much lower than an RMSE of 0.788 million cells/mL observed for the batches without using the twin.

The twin batches also showed a smaller systematic bias from target (-0.06 million cells/mL vs. $+0.57$ million cells/mL for the no-twin batches, indicating a consistent overshoot when the twin was not used), and all three twin batches (100%) landed within the ± 0.3 million cells/mL acceptance band around the target, compared with 44% of the no-twin batches. Because the twin is intended to improve accuracy specifically (not simply to shift it in either direction), we tested this as a one-sided, directional hypothesis: a permutation test on the mean absolute deviation from target and a Mann-Whitney test both showed the twin batches were significantly closer to target ($p = 0.020$).

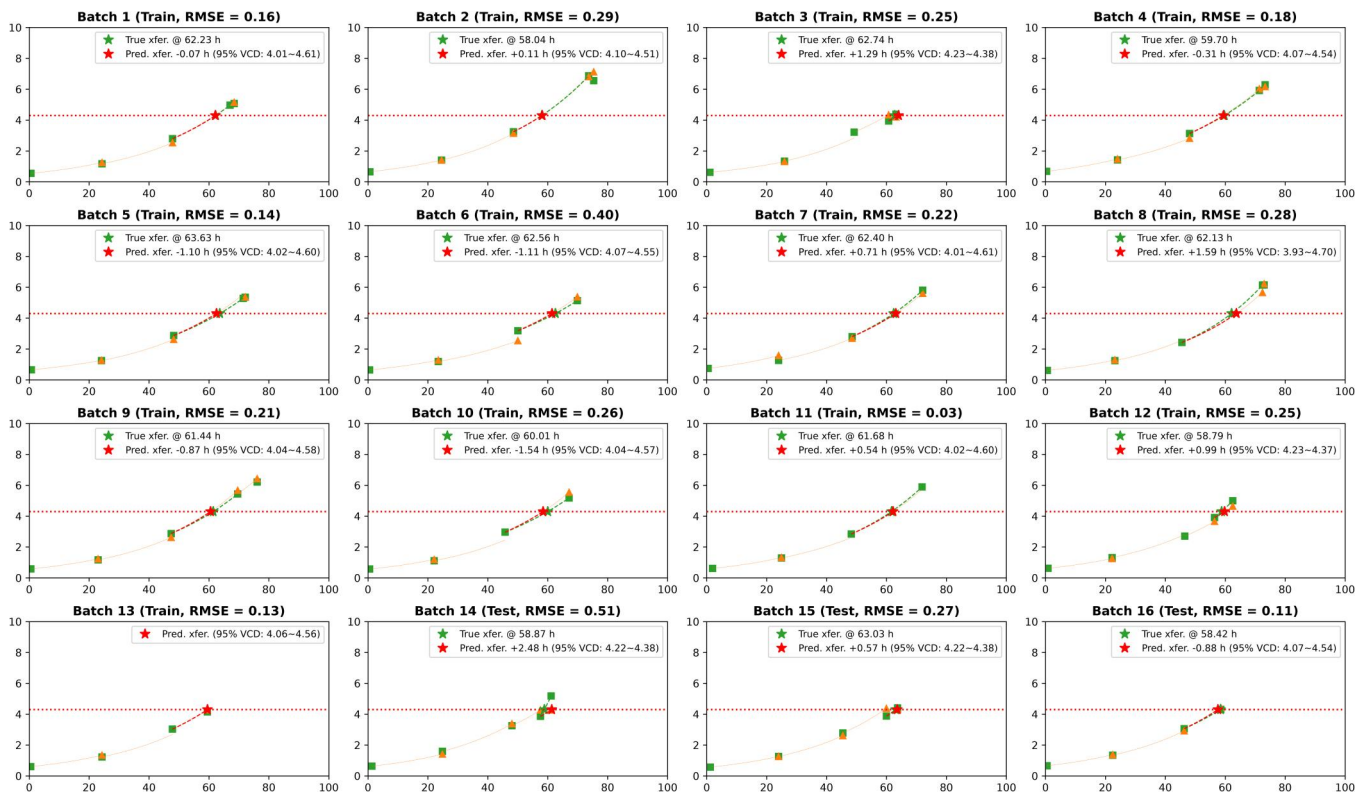


FIGURE 3 | Measured VCD (green squares), predicted VCD (orange triangles), true transfer time (green stars), and predicted transfer time (red stars) of the training and testing batches. In addition, the 95% prediction interval of VCD at the predicted transfer time is included as text.

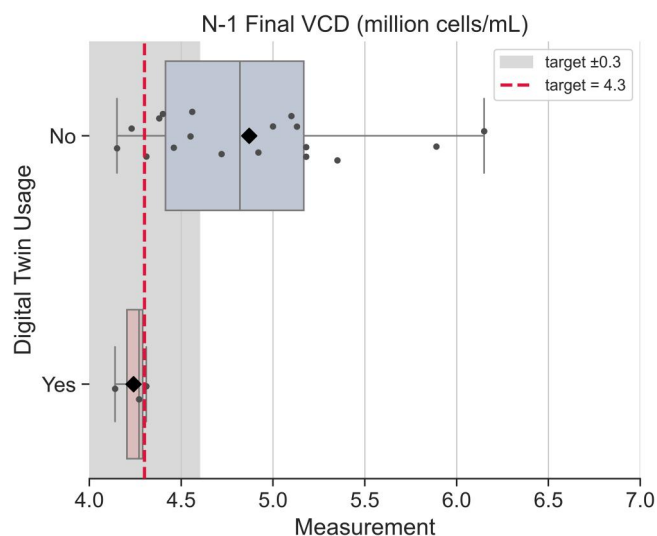


FIGURE 4 | Final VCDs of N-1 bioreactors between using and not using the transfer digital twin.

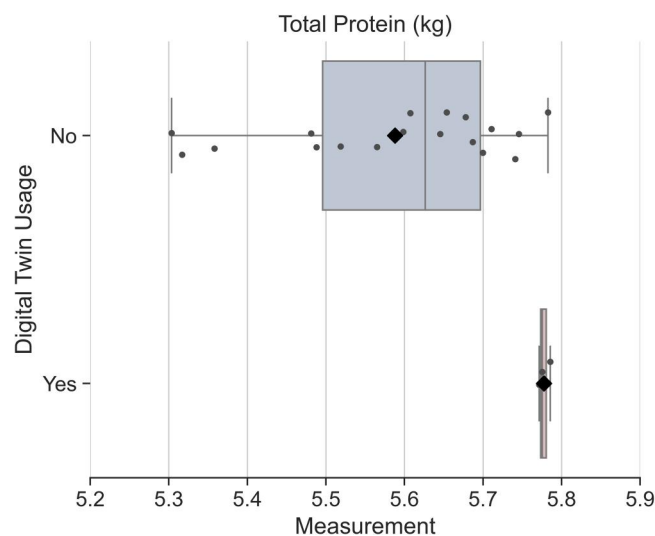


FIGURE 6 | Total protein masses between using and not using the transfer digital twin.

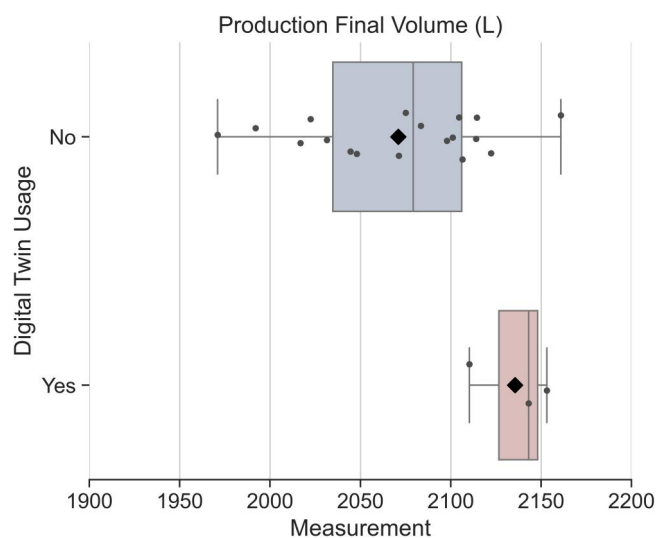


FIGURE 5 | Final volumes of production bioreactors between using and not using the transfer digital twin.

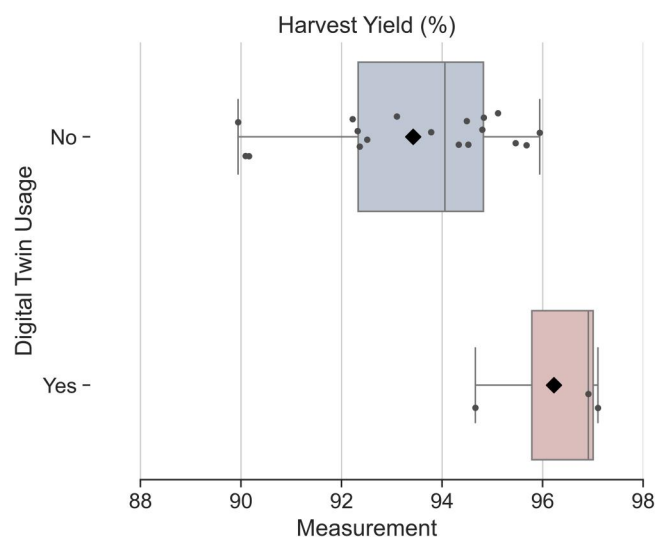


FIGURE 7 | Harvest yields between using and not using the transfer digital twin.

and $p = 0.025$, respectively; the two-sided permutation p -value was 0.082 for reference).

As a secondary observation, the N-1 final VCDs were also less variable when the twin was used. A standard analysis of variance (ANOVA) tests differences between group means, not differences in spread. To evaluate this variance claim directly, we used tests designed for that purpose, specifically Levene's test and the Brown-Forsythe test (both compare within-group variability around a central point) and a bootstrap comparison of the variance ratio (which resamples the data to build a confidence interval directly on how much tighter one group's spread is than the other's). All three agreed that the twin group was narrower ($p = 0.050$, $p = 0.048$, and $p = 0.025$ with a 95% CI of [0.0012, 0.064] that excludes 1, respectively). Since a variance-ratio CI containing 1 would indicate no difference, this bootstrap result likewise points to a statistically significant difference in variance. Given the small deployment sample ($n = 3$), these results are best interpreted as a controlled operational comparison rather than a fully powered statistical test.

The on-target N-1 final VCDs led to higher inoculation volumes in the subsequent production batches, which in turn increased the working volumes in the production bioreactors according to (1). As shown in Figure 5, the final production bioreactor volumes were significantly higher when the twin was used (Welch $p = 0.005$, Mann-Whitney $p = 0.017$). The volume benefit via (1) is specific to processes that maximize inoculation volume against a fixed seeding density. For processes with a fixed split ratio or target initial working volume, the same twin can still provide monitoring and scheduling benefits based on its high on-target VCD accuracy. The forecasting capability generalizes even though the specific benefit depends on the inoculation strategy.

The total protein masses (Figure 6; Welch $p < 0.001$, Mann-Whitney $p = 0.003$) and harvest yields (Figure 7; Welch $p = 0.022$, Mann-Whitney $p = 0.017$) of the subsequent production batches were significantly higher than those of the batches that did not use the digital twin. Notably, among the

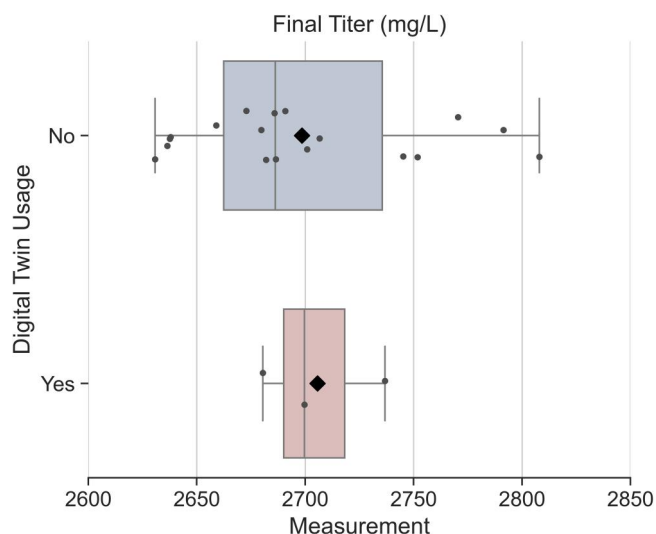


FIGURE 8 | Final titers of production bioreactors between using and not using the transfer digital twin.

three deployment batches, two reached the highest harvest yields to date for the product. Harvest yield is defined as follows:

$$\eta = \frac{V_c C_c}{\left(1 - \frac{w_h - w_t}{w_g - w_t}\right) V_b C_b} \times 100\%, \quad (7)$$

where η is harvest yield; V_c is the volume of the CUB; C_c is the concentration of the CUB; w_h is the weight of the pellet plus the centrifuge bottle; w_t is the tare weight of the bottle; w_g is the weight of the cell culture broth plus the centrifuge bottle; V_b is the total cell culture broth volume; C_b is the average cell culture broth protein concentration.

Given similar productivity (no significant titer difference; Figure 8; Welch $p = 0.748$, Mann-Whitney $p = 0.740$) between batches using and not using the digital twin, transferring the N-1 batches at the lower end of the transfer range was the primary driver of the higher overall total protein outputs and subsequently higher harvest yields due to the higher inoculation volumes.

Using the twin did not affect product quality either. As shown in Figure 9, the charge variant main species was indistinguishable between the twin and no-twin groups (Welch $p = 0.597$, Mann-Whitney $p = 0.339$).

3.3 | Limitations and Future Work

There are limitations of the digital twin. First, the model was trained on data from repeated N-1 batches of the same product. By design, the variation (and hence the information content) was lower than that in data from experiments designed to evaluate process variability and effects of perturbations. As a result, the model is appropriate for forecasting but cannot be used for control because the information about how manipulated inputs affect cell growth does not exist in the data. While this is a limitation, the simplicity of the model structure and ease of training made the twin easy to develop and fast to deploy. Second, the model is

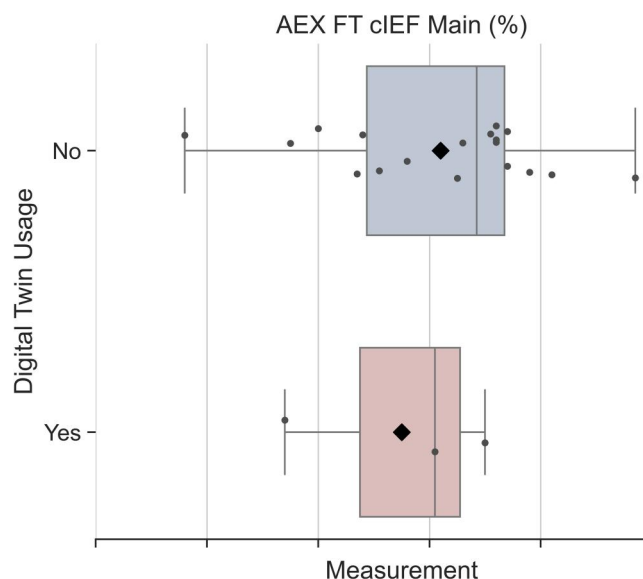


FIGURE 9 | Main species of production bioreactors between using and not using the transfer digital twin. Actual values are masked.

process- and product-specific and not readily usable in another manufacturing process. However, techniques such as transfer learning can be used to adapt the model to a new process with minimal training data. Third, applications like this must be designed within the constraints imposed by a GMP environment. Depending on the level of decision-making outlined in Table 2, a digital twin may require more validation than in this case before deployment. Fourth, the small sample size of digital twin deployment batches ($n = 3$) weakens the conclusions of statistical significance and cannot effectively exclude unobserved confounders. Nonetheless, the strongest argument comes from the deterministic evaluation of on-target final VCD, which maximizes inoculation volume and final product amount. Lastly, a more robust, regularized model could be developed in retrospect with fewer, more stable model variables to strengthen the model performance further.

Potential improvements to the digital twin include better estimation of growth rates. Specific growth rates were estimated using ViCell, which has lower accuracy and precision than other off-line analytical instruments such as FLEX2. As a result, the estimated growth rates could be inaccurate, which in turn affects the accuracy of the optimal transfer time. A state estimator can mitigate measurement errors by reconciling ViCell measurements, model predictions, and in-line capacitance measurements to generate more accurate and precise VCD estimates for the twin.

The digital twin can be further strengthened by incorporating additional in-line PAT tools, such as Raman spectroscopy and capacitance. These technologies are increasingly used across the biomanufacturing industry to provide real-time measurements, which could enable more timely and automated forecasting of transfer readiness. As in-line sensing capabilities continue to advance, the opportunity to reduce reliance on off-line sampling and improve overall forecasting efficiency will also grow. A capacitance probe has been used as the sole VCD instrument in later deployments of the digital twin in other campaigns in the same commercial manufacturing facility.

More broadly, digital twins like the one in this article can serve as foundational components within larger predictive frameworks that span the entire upstream process from thawing through production bioreactor operation. Emerging industry trends point toward integrating multiple models to support end-to-end process understanding, improve consistency, and enhance decision-making. In this context, the seed transfer twin contributes by providing reliable forecasts that help maintain consistent inoculum conditions and complement other predictive technologies, such as model predictive control (MPC) (Luo et al. 2023; Nikolakopoulou et al. 2025), which are designed to support stable and efficient production-scale performance.

4 | Conclusion

The developed seed transfer digital twin, employing a cell growth process model, successfully addresses the challenge of ensuring consistent inoculum volume during N-1 to production bioreactor transfers in monoclonal antibody manufacturing. By accurately forecasting VCD, the twin provides a reliable estimate for the optimal transfer time, thereby mitigating batch-to-batch variability. The practical application of this digital twin delivered business value, leading to more consistent transfer timings, higher on-target accuracy and reduced variability of final VCDs, and higher production bioreactor working volumes.

While the current model is process-specific and primarily functions as a forecasting tool rather than for direct process control like MPC, its simplicity facilitated rapid development and deployment. Future enhancements could include exploring PAT for real-time forecasts, using transfer learning techniques to adapt the model to new processes with minimal data, and integrating state estimators to mitigate measurement inaccuracies from analytical instruments. This seed transfer digital twin represents a valuable step toward a more robust and consistent biomanufacturing process, aligning with broader ambitions for advanced process control and digitalization in the industry.

Author Contributions

Yu Luo: conceptualization, formal analysis, writing – original draft. **Sumit Pareek:** investigation, writing – review and editing. **Carolyn Zhang:** project administration, writing – review and editing. **Kelly Sarnese:** supervision, writing – review and editing. **Diana Ritz:** supervision, writing – review and editing. **Sameer Talwar:** supervision, writing – review and editing.

Acknowledgments

The authors thank Sreetama Das, Ghifari Pradana, Jeff Simkins, and Matthieu Duvinage for productionizing the digital twin; Paul Renaut for creating and maintaining the data pipeline for model development; Lucas Kimerer for curating the data used in the deployment analysis; Fay Chios, Evan Hernick, Nathaniel Berendson, and David Aguirre for business case and value definition, evaluation of training and testing data for data quality, user interface development, industrialization and generation of data, and consultation for future improvement; Theodore Vu, Nathaniel Berendson, and Annabelle Cram for contributing to the development and testing of a previous iteration of the digital twin; Neil Walker for suggesting the statistical methodology used to compare the deployment outcomes; and Zach Hatzenbeller and Toma Ramonaite for

reviewing the data used in this article. This work was funded by GSK. No external or third-party funding was received.

Conflicts of Interest

All authors are employees of GSK. This work was funded by GSK and was carried out using data from a GSK process. One or more authors hold GSK stock as part of their employment compensation. The authors have no other interests or relationships to declare.

Data Availability Statement

The authors have nothing to report.

References

- Agarwal, P., C. McCready, S. K. Ng, et al. 2025. “Hybrid Modeling for in Silico Optimization of a Dynamic Perfusion Cell Culture Process.” *Biotechnology Progress* 41, no. 1: e3503.
- Bayer, B., M. Duerkop, G. Striedner, and B. Sissolak. 2021. “Model Transferability and Reduced Experimental Burden in Cell Culture Process Development Facilitated by Hybrid Modeling and Intensified Design of Experiments.” *Frontiers in Bioengineering and Biotechnology* 9: 740215.
- Frahm, B. 2014. *Seed Train Optimization for Cell Culture*, in *Methods in Molecular Biology*, 355–367. Humana Press.
- Gillespie, C., D. P. Wasalathanthri, D. B. Ritz, et al. 2022. “Systematic Assessment of Process Analytical Technologies for Biologics.” *Biotechnology and Bioengineering* 119, no. 2: 423–434.
- Goldrick, S., A. Umprecht, A. Tang, et al. 2020. “High-Throughput Raman Spectroscopy Combined With Innovate Data Analysis Workflow to Enhance Biopharmaceutical Process Development.” *Processes* 8, no. 9: 1179.
- Hernández Rodríguez, T., C. Posch, and R. Pörtner. 2020. “Dynamic Parameter Estimation and Prediction over Consecutive Scales, Based on Moving Horizon Estimation - Applied to An Industrial Cell Culture Seed Train.” *Bioprocess and Biosystems Engineering* 44: 793–808. <https://doi.org/10.1007/s00449-020-02488-1>.
- Hernández Rodríguez, T., A. Sekulic, M. Lange-Hegermann, and B. Frahm. 2022. “Designing Robust Biotechnological Processes Regarding Variabilities Using Multi-Objective Optimization Applied to a Biopharmaceutical Seed Train Design.” *Processes* 10, no. 5: 883.
- Kontoravdi, C., P. A. Steven, E. N. Pistikopoulos, and A. Mantalaris. 2007. “Development of a Dynamic Model of Monoclonal Antibody Production and Glycosylation for Product Quality Monitoring.” *Computers and Chemical Engineering* 31, no. 5: 392–400.
- Kritzing, W., M. Karner, G. Traar, J. Henjes, and W. Sihn. 2018. “Digital Twin in Manufacturing: A Categorical Literature Review and Classification.” *IFAC-PapersOnLine* 51, no. 11: 1016–1022.
- Luo, Y., V. Kurian, L. Song, E. A. Wells, A. S. Robinson, and B. A. Ogunnaike. 2023. “Model-Based Control of Titer and Glycosylation in Fed-Batch mAb Production: Modeling and Control System Development.” *AIChE Journal* 69, no. 4: e18075.
- Luo, Y., R. J. Lovelett, J. V. Price, et al. 2021. “Modeling the Effect of Amino Acids and Copper on Monoclonal Antibody Productivity and Glycosylation: A Modular Approach.” *Biotechnology Journal* 16, no. 2: 2000261.
- Moore, B., R. Sanford, and A. Zhang. 2019. “Case Study: The Characterization and Implementation of Dielectric Spectroscopy (Biocapacitance) for Process Control in a Commercial GMP CHO Manufacturing Process.” *Biotechnology Progress* 35, no. 3: e2782.
- Nikolakopoulou, A., Y. Luo, and Z. Hatzenbeller. Towards “Tighter” Titer: Model Predictive Control in Fed-batch Bioreactors. in 2025 American Control Conference (ACC). 2025.
- Olin, M., N. Wolnick, H. Crittenden, et al. 2024. “An Automated High Inoculation Density Fed-Batch Bioreactor, Enabled Through N-1

Perfusion, Accommodates Clonal Diversity and Doubles Titters.” *Biotechnology Progress* 40, no. 2: e3410. <https://doi.org/10.1002/btpr.3410>.

Park, S. Y., S. J. Kim, C. H. Park, J. Kim, and D. Y. Lee. 2023. “Data-Driven Prediction Models for Forecasting Multistep Ahead Profiles of Mammalian Cell Culture Toward Bioprocess Digital Twins.” *Biotechnology and Bioengineering* 120, no. 9: 2494–2508.

Park, S.-Y., C. H. Park, D. H. Choi, J. K. Hong, and D. Y. Lee. 2021. “Bioprocess Digital Twins of Mammalian Cell Culture for Advanced Biomanufacturing.” *Current Opinion in Chemical Engineering* 33: 100702.

Pinto, J., J. R. C. Ramos, R. S. Costa, S. Rossell, P. Dumas, and R. Oliveira. 2023. “Hybrid Deep Modeling of a CHO-K1 Fed-Batch Process: Combining First-Principles With Deep Neural Networks.” *Frontiers in Bioengineering and Biotechnology* 11: 1237963. <https://doi.org/10.3389/fbioe.2023.1237963>.

Pohlscheidt, M., M. Jacobs, S. Wolf, et al. 2013. “Optimizing Capacity Utilization by Large Scale 3000 L Perfusion in Seed Train Bioreactors.” *Biotechnology Progress* 29, no. 1: 222–229.

Richelle, A., B. Corbett, P. Agarwal, A. Verneris, J. Trygg, and C. McCready. 2022. “Model-Based Intensification of CHO Cell Cultures: One-Step Strategy From Fed-Batch to Perfusion.” *Frontiers in Bioengineering and Biotechnology* 10: 948905. <https://doi.org/10.3389/fbioe.2022.948905>.

Rittershaus, E. S. C., M. S. Rehmann, J. Xu, et al. 2022. “N-1 Perfusion Platform Development Using a Capacitance Probe for Biomanufacturing.” *Bioengineering* 9, no. 4: 128.

Singh, M., E. Fuenmayor, E. Hinchey, Y. Qiao, N. Murray, and D. Devine. 2021. “Digital Twin: Origin to Future.” *Applied System Innovation* 4, no. 2: 36.

Xu, J., X. Xu, C. Huang, et al. 2020. “Biomanufacturing Evolution From Conventional to Intensified Processes for Productivity Improvement: A Case Study.” *mAbs* 12, no. 1: 1770669.

Yan, X., X. Dong, Y. Wan, et al. 2024. “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.” *Biotechnology Journal* 19, no. 1: e2300395. <https://doi.org/10.1002/biot.202300395>.