Pseudo batch transformation: A novel method to correct for mass removal through sample withdrawal of fed-batch fermentations

Authors: Viktor Hesselberg-Thomsen[1], Teddy Groves[1], Tim McCubbin[2,3], Iván Martínez-Monge[1,4], Igor Marín de Mas[1], Lars Keld Nielsen[1,2,3]

Affiliations

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Lyngby, Denmark.

[2]Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, Brisbane, QLD 4072, Australia

[3]ARC Centre of Excellence in Synthetic Biology, The University of Queensland, Queensland, Australia

[4]AGC Biologics, Upstream Process Development Department. Vandtårnsvej 83 2860 Søborg, Denmark

Abstract

Summary: We present a novel "pseudo batch" transformation algorithm that maps analytical data obtained for fed-batch bioreactor cultivations onto a constant volume batch process, significantly decreasing the complexity of characterizing the fed-batch process.

Availability and implementation: Our method is implemented in both Excel and Python and is available with tutorials and example data from https://github.com/biosustain/pseudobatch. The Python package is also available on PYPI under the name "pseudobatch". A Docker image for reproduction of the results in this study can be found at data.dtu.dk (DOI:10.11583/DTU.23453018).

Contact: Lars Keld Nielsen, e-mail: lars.nielsen@uq.edu.au

Supplementary information: A comprehensive explanation of the simulated fed-batch, parameter estimation procedures, and the Bayesian model can be found in supplementary information (S1-S5).

Introduction

Fed-batch cultivation is commonly used by the biomanufacturing industry to avoid substrate inhibition and/or overflow metabolism in both mammalian and microbial systems [1]. The analysis

of fed-batch data is confounded by the dynamic changes in bioreactor volume caused by sampling and feeding processes. The removal of mass through sampling leads to discontinuous mass time-evolution curves (Figure 1A), which will lead to inaccurate rate, yield, and metabolic flux estimates and inaccurate mass balancing unless carefully considered. Furthermore, this limits the ability of machine learning algorithms to train on fed-batch data.

Traditionally, fed batch processes have been optimized in larger bioreactors where the sample fraction and hence the effect is negligible. However, high-throughput screening for clone selection or feeding strategies is increasingly performed using miniaturized multiplex fermentation systems (e.g., ambr15, BioLector) [2]. Rapid advances in the use of machine learning for optimizing bioprocesses is one of the drivers for high-throughput small volume cultivation [3], [4] and most systems suffer from substantial sampling fraction. Even reactors with larger operating volumes can be subject to substantial errors when significant samples are withdrawn to prevent reactor overflow, extend cultivation time and reduce time between fermentations [5], [6], particularly in the case where reactors are operated as cyclic or repetitive fed-batch modes.

In this study, we present a novel method called pseudo batch transformation that removes the effects of the volume changes from the experimental data. We apply this method to simulated data to illustrate that it estimates the correct rates and yields using standard linear modeling procedures [7].

Methods

Formalization of the Pseudo batch transformation method

The pseudo batch transformation independently projects time-course data for each species (biomass or metabolite) measured across a fed-batch bioreactor culture onto a constant volume batch culture. For convenience, we define an accumulated dilution factor (ADF), defined in equation (1), which scales a concentration at any given measurement time to the original bioreactor volume.

$$\mathrm{AD}F_i = \prod_{1}^{i} \frac{\mathrm{V_i}}{V_{i-1} - S_{i-1}} \tag{1}$$

We employ an integer-valued time index variable i which denotes the order of the measurements. $\mathrm{S_{i-1}}$ is the sample volume at time index $i-1$. This is 0 at time points where no sample was withdrawn. V_i is the volume prior to sampling at time index i. We see that $ADF_i = 1$ when no volume is added to the reactor, i.e. no dilution.

The pseudo batch transformation is defined by equation (2):

$$c^{\star}_{species,\,i} = ADF_i \cdot c_{species,\,i} - \sum_{j=1}^{l} \sum_{k=1}^{k} ADF_k \frac{c_{species,\,feed_j} \cdot (F_{k,j} - F_{k-1,j})}{V_k} \tag{2}$$

$c^{\star}_{species,\,i}$ is the pseudo batch concentration at measurement time index i, which depends on two terms. The first term scales the measured concentration of a species at time index i, $c_{species,\,i}$, by the accumulated dilution factor at time index i, ADF_i. This term essentially transforms a measured concentration to the concentration that would be achieved if the bioreactor was not being diluted by the feed. The second term accumulates species added through feeding. Essentially, this term performs a scaled integration from the beginning until the i'th measurement by iterating k from 1 until i. The concentration of the added species at each time index, k, is scaled by its corresponding ADF at time index k, across all l feeds. $c_{species,\,feed_j}$ is the concentration of the species in the feed j, $F_{k,j}$ is the accumulated volume of feed j added at time k and V_k is the volume at time index k prior to sampling. In the case where the species is not fed, i.e. a product, this second term collapses to 0.

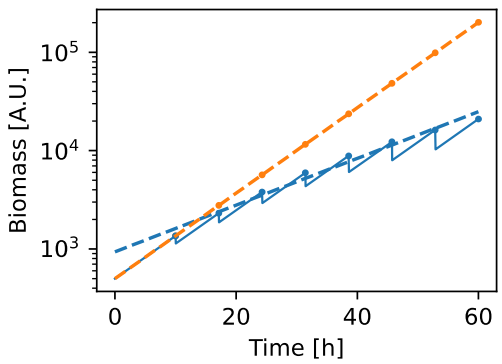
The pseudo batch concentrations represent the effective concentration of catabolized substrates or generated products and can be directly utilized within yield calculations. The method can be expanded to capture substrates or products that are not solubilized in the media, such as gas data, by converting these species into a pseudo concentration (see S2 for details). Furthermore, pseudo concentrations are allowed to be negative. Importantly, they should not be interpreted as an actual prediction of the theoretical fed-batch process without volume changes. This would likely result in a different environment in the bioreactor, e.g., due to larger cell densities. Rather the method scales the observed phenotypic behavior to be more interpretable both visually and through statistical models.

Results and discussion

Assessing the validity of the pseudo batch method with simulated data

To illustrate the need and utility of pseudo batch transformation, we simulated a microbial culture under fed-batch conditions. In this simulation glucose is consumed, and biomass, product, and CO_2 are produced. The feed addition follows an exponential feeding profile designed to keep the substrate concentration constant (further detailed in S3). When samples are withdrawn from this culture, mass is removed, and the true mass time course becomes discontinuous (Figure 1A - blue line and figure S1). Thus, if we estimate the growth rate as the slope of a log-linear model, we obtain a biased estimate which is 34% lower than the simulated growth rate (Figure 1A – dashed blue line and B). However, if the pseudo batch transformation is applied to the data before estimating the growth rate, the correct values are obtained (Figure 1A – dashed orange line). The same issue applies to the yield calculations without the pseudo batch transformation (Figure 1B), except where a yield is calculated between two products which cancels out the majority of the error. Often the production or consumption rates are of interest, the calculation of which will compound the error from the yields and growth rate estimates.

A



B

	Unit	True	Uncorrected	Corrected
Y_{xs}	$\frac{gGlucose}{gBiomass}$	1.85	3.36 (-82%)	1.85 (0 %)
Y_{xp}	$\frac{gProduct}{gBiomass}$	0.84	0.83 (-2%)	0.82 (0 %)
Y_{xCO_2}	$\frac{gCO_2}{gBiomass}$	0.05	0.08 (-81%)	0.05 (0 %)
μ	$\frac{1}{h}$	0.10	0.05 (45%)	0.10 (0 %)

Figure 1 **A)** Simulated exponential fed-batch with sampling (blue line), sample measurements (blue dots) and the pseudo batch transform measurements (orange dots). Fitted growth rate model for raw measurements (blue dashed line) and pseudo batch transformed measurements (orange dashed line). **B)** Uncorrected and corrected estimated values of the yields and growth rate (see S4 for estimation approach). The parenthesis holds the percent error of the estimate compared to the true value.

The pseudo batch transformation introduces negligible additional error

The pseudo batch concentrations are calculated from several measured quantities: concentration, bioreactor volume, feeding volume, and concentrations in the feed medium. Uncertainty in these quantities will combine to produce uncertainty in the pseudo batch concentration. We therefore created a Bayesian statistical model that propagates known uncertainties and sources of measurement error into uncertainties for pseudo batch transformed concentrations. We fit this model for a typical case under a plausible range of error assumptions, assuming independent 5% errors on reactor volume, sample volume measurements and a consistent roughly +/- 10% error in feed volume delivered.

In our case study, the logarithmic-scale uncertainty, as captured by the marginal coefficients of variation of posterior samples, was approximately the same for transformed and untransformed concentrations under a range of plausible measurement error assumptions (Figure S4). This indicates that, at least in such a typical case, the pseudo batch transformation introduces negligible additional error. We also used these posterior samples to recover known growth rates and yield coefficients using standard methods for batch fermentation data. This demonstrates, again for this typical case, that analyses using this data are appropriately accurate (Figure S5 and Table S3).

Further details on the model and results can be found in Section S5. Additionally, we exposed our model in the package through a function. This allow users to quantify their own pseudo batch transformation uncertainties (Further detailed in https://biosustain.github.io/pseudobatch/Tutorials/7%20-%20Pseudobatch%20transformation%20with%20uncertainties.html).

The pseudo batch transformation enables the utilization of an extensive toolkit

The pseudo batch transformed data can be used in the standard linear model workflows in Excel, R, Python or other software, to estimate rates and yields from experimental measurements.

Furthermore, the continuous nature of the pseudo batch transformed data enables a big suite of more advanced statistical and machine learning methods to work straight out of the box, which eases data integration. For example the pseudo batch transformation enables estimation of rates through gaussian process regression [8], spline fitting [9], and dynamic flux balance analysis [9]. Additionally, our method enables training of machine learning algorithms on data from sampled microbioreactor cultivations. Such algorithms could include physics-informed neural networks for bioprocess modelling [10] or strain optimization through a probabilistic ensemble model [4]. Machine learning methods are crucial techniques for high throughput data processing pipelines and automated decision making [3]. The pseudo batch transformation enables seamless use of fed-batch data from microbioreactor systems in such pipelines.

Conclusion

The pseudo batch transformation method presented in this study effectively transforms measurements from sampled fed-batch cultivations into a space where it is possible to calculate rates and yields using standard methods. The transformation does not increase the uncertainty of the measurements or estimated values. The pseudo batch transformation can facilitate data-driven decision making and strain characterization based on fed-batch fermentations, which can enhance bioprocess optimization, clone selection, and metabolic studies and enables integration of fed-batch data with machine learning models in the biotechnology industry and academia.

Acknowledgements

None

Funding

This work was supported by the Novo Nordisk Foundation [grant numbers NNF20CC0035580 and NNF14OC0009473] and by the Australian Government through the Australian Research Council Centre of Excellence in Synthetic Biology [project number CE200100029].

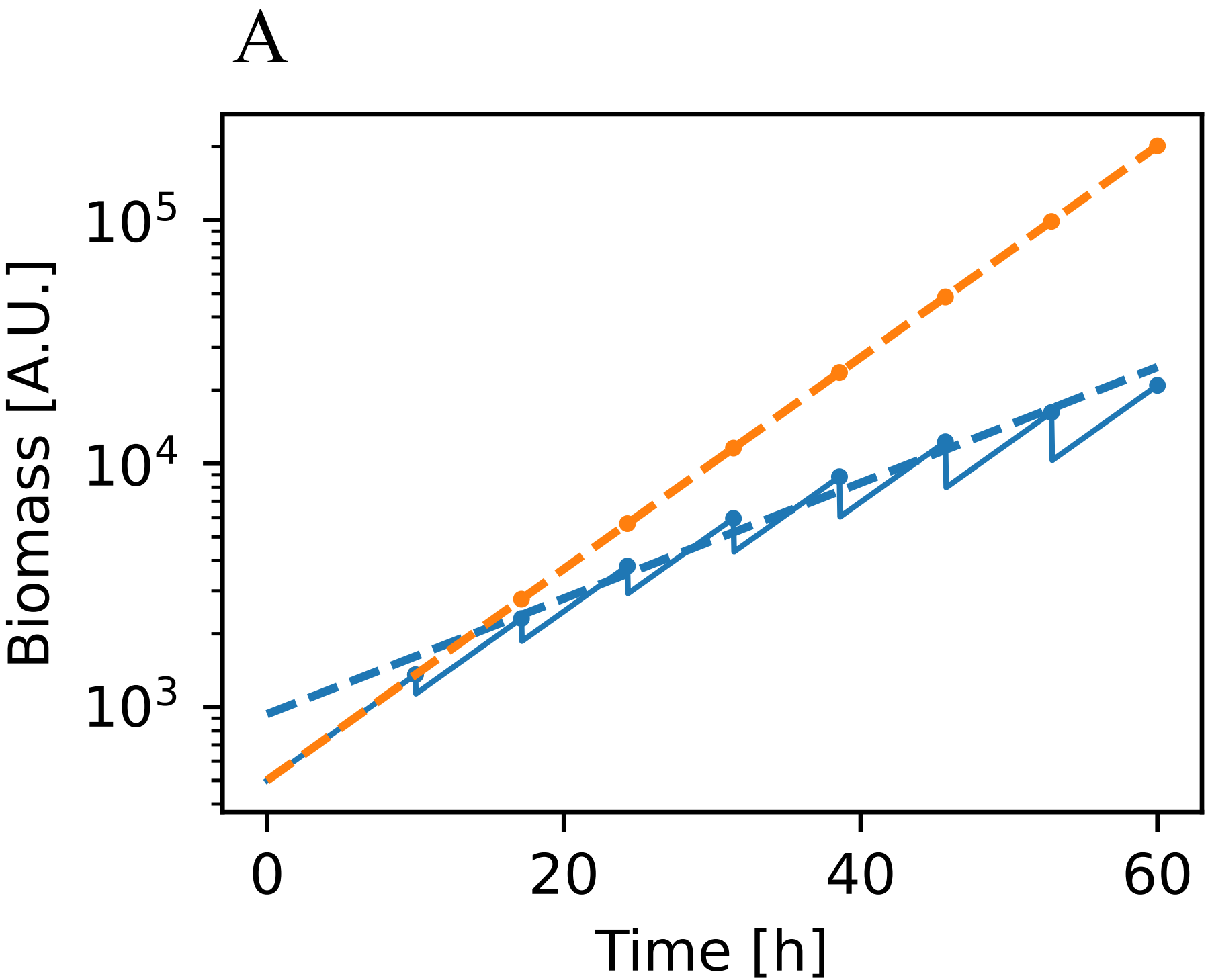
Reference:

[1] J. Villadsen, J. Nielsen, and G. Lidén, *Bioreaction Engineering Principles*. Boston, MA: Springer US, 2011. doi: 10.1007/978-1-4419-9688-6.

[2] J. Hemmerich, S. Noack, W. Wiechert, and M. Oldiges, “Microbioreactor Systems for Accelerated Bioprocess Development,” *Biotechnology Journal*, vol. 13, no. 4, p. 1700141, Apr. 2018, doi: 10.1002/biot.201700141.

[3] L. M. Helleckes, J. Hemmerich, W. Wiechert, E. Von Lieres, and A. Grünberger, “Machine learning in bioprocess development: from promise to practice,” *Trends in Biotechnology*, vol. 41, no. 6, pp. 817–835, Jun. 2023, doi: 10.1016/j.tibtech.2022.10.010.

[4] J. Zhang *et al.*, "Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism," *Nat Commun*, vol. 11, no. 1, p. 4880, Sep. 2020, doi: 10.1038/s41467-020-17910-1.
[5] S. Singh, B. Sithole, P. Lekha, K. Permaul, and R. Govinden, "Optimization of cultivation medium and cyclic fed-batch fermentation strategy for enhanced polyhydroxyalkanoate production by Bacillus thuringiensis using a glucose-rich hydrolyzate," *Bioresour. Bioprocess.*, vol. 8, no. 1, p. 11, Dec. 2021, doi: 10.1186/s40643-021-00361-x.
[6] J. Kopp, S. Kittler, C. Slouka, C. Herwig, O. Spadiut, and D. J. Wurm, "Repetitive Fed-Batch: A Promising Process Mode for Biomanufacturing With E. coli," *Front. Bioeng. Biotechnol.*, vol. 8, p. 573607, Nov. 2020, doi: 10.3389/fbioe.2020.573607.
[7] C. P. Long and M. R. Antoniewicz, "High-resolution 13C metabolic flux analysis," *Nat Protoc*, vol. 14, no. 10, pp. 2856–2877, Oct. 2019, doi: 10.1038/s41596-019-0204-0.
[8] P. S. Swain *et al.*, "Inferring time derivatives including cell growth rates using Gaussian processes," *Nat Commun*, vol. 7, no. 1, p. 13766, Dec. 2016, doi: 10.1038/ncomms13766.
[9] V. S. Martínez, M. Buchsteiner, P. Gray, L. K. Nielsen, and L.-E. Quek, "Dynamic metabolic flux analysis using B-splines to study the effects of temperature shift on CHO cell metabolism," *Metabolic Engineering Communications*, vol. 2, pp. 46–57, Dec. 2015, doi: 10.1016/j.meteno.2015.06.001.
[10] M. S. F. Bangi, K. Kao, and J. S.-I. Kwon, "Physics-informed neural networks for hybrid modeling of lab-scale batch fermentation for β-carotene production using Saccharomyces cerevisiae," *Chemical Engineering Research and Design*, vol. 179, pp. 415–423, Mar. 2022, doi: 10.1016/j.cherd.2022.01.041.


A
Biomass [A.U.]
10^5
10^4
10^3
0
20
40
60
Time [h]