Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Smart Process Manufacturing: Deep Integration of AI and Process Manufacturing—Article

Optimal Antibody Purification Strategies Using Data-Driven Models

Songsong Liu [a,b,*], Lazaros G. Papageorgiou [c,*]

[a] School of Management, Harbin Institute of Technology, Harbin 150001, China
[b] School of Management, Swansea University, Swansea SA1 8EN, UK
[c] Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, London WC1E 7JE, UK

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Antibody purification
Multiscale optimization
Antigen-binding fragment
Mixed-integer programming
Data-driven model
Piecewise linear regression

ABSTRACT

This work addresses the multiscale optimization of the purification processes of antibody fragments. Chromatography decisions in the manufacturing processes are optimized, including the number of chromatography columns and their sizes, the number of cycles per batch, and the operational flow velocities. Data-driven models of chromatography throughput are developed considering loaded mass, flow velocity, and column bed height as the inputs, using manufacturing-scale simulated datasets based on microscale experimental data. The piecewise linear regression modeling method is adapted due to its simplicity and better prediction accuracy in comparison with other methods. Two alternative mixed-integer nonlinear programming (MINLP) models are proposed to minimize the total cost of goods per gram of the antibody purification process, incorporating the data-driven models. These MINLP models are then reformulated as mixed-integer linear programming (MILP) models using linearization techniques and multiparametric disaggregation. Two industrially relevant cases with different chromatography column size alternatives are investigated to demonstrate the applicability of the proposed models.

1. Introduction

The global industry has been experiencing accelerating changes during the recent transformation of traditional manufacturing into smart manufacturing [1,2]. During the conversion process, industries face a number of challenges posed by smart manufacturing, which have attracted great attention in both academic and practitioner communities [3], particularly in the process industry [4]. Some of the challenges to be covered in this work include:

- The use and analysis of data, with a particular focus on the development of data-driven surrogate/metamodels to simplify complex processes and to enable manufacturing intelligence;
- The implementation of multiscale modeling and optimization to integrate strategic and planning decisions with operations in order to support enterprise-wide coordination and optimization;
- The development of computationally efficient models, algorithms, and tools in order to find global optimal solutions for smart manufacturing decision-making and to enable large-scale optimization.

In this work, we aim to develop optimization-based decision-making models for optimal purification strategies in the manufacturing process of an antibody product based on simple data-driven models, in an attempt to cope with the above challenges in the biopharmaceutical industry. In order to achieve better control of the processes and improve production efficiency, biopharmaceutical manufacturing process optimization problems have been investigated using different modeling and solution techniques, such as metaheuristic [5], dynamic optimization [6], evolutionary algorithm [7–9], Markov decision method [10], and mixed-integer programming [11–22]. Data-driven models—also known as surrogate models or metamodels—refer to models that are built on the basis of data, but are not dependent on theoretical knowledge of the concerned processes or systems. Data-driven models of complex processes and systems provide model simplicity and computational efficiency [23], and their integration with optimization requires less computational effort and has a broad application in the engineering field [24,25]. In particular, such models have demonstrated research benefits in the modeling and optimization of chromatography purification operations [26–29]. However, only a few attempts have been made to integrate data-driven models into optimization models for biopharmaceutical purification processes. Nagrath et al. [30] developed an artificial neural network

* Corresponding authors.
E-mail addresses: s.liu@hit.edu.cn (S. Liu), l.papageorgiou@ucl.ac.uk (L.G. Papageorgiou).

https://doi.org/10.1016/j.eng.2019.10.011

Nomenclature

Indices

d	Position in multiparametric disaggregation = p, . . ., $maxp$
i	Column volume size
j	Column number = 1, . . .,$maxCYN_s$
k	Cycle number = 1, . . .,$maxCN_s$
m	Diameter size
n	Digit of the binary representation = 1, . . .,$\lceil \log_2 maxBN \rceil$
q	Integer number in multiparametric disaggregation
r	Interval in piecewise regression function
s	Downstream step = ct_1 (centrifugation 1), *ho* (homogenization), ct_2 (centrifugation 2), *fi* (filtration), *af* (affinity chromatography), *ce* (cation-exchange chromatography), uf_1 (UF/DF 1), *ce* (anion-exchange chromatography), uf_2 (UF/DF 2), *bf* (bulk fill)

Sets

CS	Set of chromatography steps ={*af*, *ce*, *ae*}

Parameters

a, b, c	Utilities cost coefficients
aot	Annual operating time, d
bcv_s	Buffer volume ratio at chromatography step s
$bp_{s,r}$	Breakpoint of loaded mass between intervals r and $r+1$ at chromatography step s, g
bpc	Buffer price, GBP·L^{-1}
brc	Bioreactor cost, GBP
brn	Number of bioreactors
brt	Bioreaction time, d
brv	Bioreactor volume, L
bvr_s	Buffer volume ratio at centrifugation step s
$cc_{s,i}$	Column cost of size i at chromatography step s, GBP
$\widetilde{cc}_{s,m}$	Column cost of diameter size m at chromatography step s, GBP
$cv_{s,i}$	Volume of column size i at chromatography step s, L
dbc_s	Dynamic binding capacity at chromatography step s, g·L^{-1}
$dm_{s,i}$	Diameter of column size i at chromatography step s, cm
$\widetilde{dm}_m$	Diameter of size m at chromatography step s, cm
don	Number of operators for downstream processing
dvr	Diafiltration volume ratio at the second UF/DF step
ecv_s	Elute volume ratio at chromatography step s
el	Equipment lifetime, year
$fconc$	Final concentration of product, g·L^{-1}
fvr	Flush volume ratio of the first UF/DF step
gef	General equipment factor
gu	General utility unit cost, GBP·L^{-1}
$h_{s,i}$	Bed height of column size i at chromatography step s, cm
ir	Interest rate
$i\lambda$	Ratio of insurance cost to fixed capital investment
l_s	Lifetime of resin at chromatography step s, cycle
$lang$	Lang factor
$maxBBV$	Maximum buffer volume per batch, L
$maxBN$	Maximum number of batches
$maxCN_s$	Maximum number of columns at chromatography step s
$maxCOG$	Maximum COG per gram, GBP·L^{-1}
$maxCYN_s$	Maximum number of cycles at chromatography step s
$maxH_s$	Maximum column bed height size at chromatography step s, cm
$maxLM_s$	Maximum product mass loaded at chromatography step s, g
$maxp$	Maximum position in multiparametric disaggregation
$maxT_s$	Maximum processing time per batch at chromatography step s, h
$maxTCV_s$	Maximum total column volume at chromatography step s, L
$maxTP_s$	Maximum throughput at chromatography step s, g·L^{-1}
$ma\lambda$	Maintenance cost ratio to the fixed capital investment
$mepc$	Media price, GBP·L^{-1}
$mi\lambda$	Miscellaneous material cost ratio to chemical reagent and consumable costs
$m\lambda$	Management cost ratio to direct labor cost
$oe\lambda$	Other equipment cost ratio to the bioreactor cost
of	Resin overpacking factor
pr_s	Processing rate of step s, L·h^{-1}
$q\lambda$	Ratio of QCQA cost to direct labor cost
rpc_s	Resin price at chromatography step s, GBP·L^{-1}
$refCC$	Reference cost of a chromatography column, GBP
$refDM$	Reference diameter of a chromatography column, cm
sfd	Duration per shift, h
sfn	Number of shifts per day, d^{-1}
st	Seed train bioreaction time, d
$s\lambda$	Supervisors cost ratio to direct labor cost
$titer$	Upstream product titer, g·L^{-1}
$t\lambda$	Tax cost ratio to the fixed capital investment
uon	Number of operators per bioreactor in USP
uot	USP operating time per day, h·d^{-1}
vel	Linear velocity of flow at the anion-exchange chromatography step, cm·h^{-1}
w	Wage of an operator, GBP·L^{-1}
yd_s	Product yield at step s
α	Bioreactor working volume ratio
$\beta^0_{s,r}$	Constant coefficient in interval r at chromatography step s
$\beta^H_{s,r}$	Coefficient for bed height in interval r at chromatography step s
$\beta^{LM}_{s,r}$	Coefficient for loaded mass in interval r at chromatography step s
$\beta^V_{s,r}$	Coefficient for velocity in interval r at chromatography step s
ε	A small number
θ	Media overfill allowance
μ	Chromatography resin utilization factor
σ	Batch success rate

Continuous variables

ABV	Annual buffer volume, L
AC	Annual total cost, GBP
AP	Annual product output, g
AT	Annual downstream operating time, d
BAT	Time for adding buffer per batch at the anion-exchange chromatography step, h
BBV	Buffer volume added per batch, L
BC	Buffer cost, GBP
BT	Downstream processing time per batch, d
BV_s	Buffer volume per batch in step s, L
CAC	Capital cost, GBP
CC	Consumables cost, GBP
COG	Cost of goods per gram, GBP·g^{-1}
CRC	Chemical reagents cost, GBP
CAP_q	Continuous variable for annual production in multiple disaggregation at digit q
$CTP_{s,q}$	Continuous variable for throughput in multiple disaggregation at digit q, step s
DLC	Direct labor cost, GBP
FCI	Fixed capital investment, GBP
GUC	General utility cost, GBP
IC	Insurance cost, GBP
LC	Labor cost, GBP
LM_s	Mass loaded to single column at chromatography step s, g

Please cite this article as: S. Liu and L. G. Papageorgiou, Optimal Antibody Purification Strategies Using Data-Driven Models, Engineering, https://doi.org/10.1016/j.eng.2019.10.011