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:
Received 31 October 2018
Revised 15 December 2018
Accepted 21 December 2018
Available online 18 October 2019

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.

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

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 (ANN)-based hybrid model for the optimization of preparative chromatographic processes.

⁎ Corresponding authors.
E-mail addresses: s.liu@hit.edu.cn (S. Liu), l.papageorgiou@ucl.ac.uk (L.G. Papageorgiou).
Pirrung et al. [31] developed ANNs from detailed mechanistic models and integrated them into the optimization of biopharmaceutical downstream processes for the maximum yield of a process with three different chromatographic columns.

Antigen-binding fragment (Fab) products are regarded as the next generation of protein-based biotherapeutics after monoclonal antibody (mAb) products, and offer many advantages due to their simpler and smaller structure [32]. There is a need to develop a cost-efficient process for Fab production in industrial practice [33], but few works on this topic exist in the literature. In this work, the multiscale optimization of the purification process of a Fab product is addressed, using microscale column chromatography experiment data for manufacturing-scale chromatography optimization. In order to achieve the most cost-efficient process, besides considering chromatography column-sizing strategies at the design level, operational decisions are taken into account—particularly the flow velocity at chromatography steps. Data-driven models are developed to estimate the chromatography throughput. When incorporating the developed data-driven models, a number of mixed-integer programming models are proposed to find the optimal Fab purification strategies, and are examined through case studies. To the best of our knowledge, this is the first work in the literature on the multiscale optimization of the purification process of Fab using data-driven models.

The rest of this paper is organized as follows: The optimization problem is described in Section 2. The data-driven models of the chromatography operations are developed in Section 3, while the proposed mathematical programming optimization models are described in Section 4. Section 5 presents industrially relevant case studies, followed by the computational results and discussion in Section 6. Finally, concluding remarks are provided in Section 7.

2. Problem statement

This work investigates the optimization of the manufacturing processes of a Fab product. Fig. 1 shows the Fab manufacturing process flowsheet studied in this work. Initially, the mammalian cells expressing the Fab are cultured in bioreactors in the upstream processing (USP) before entering into the downstream processing (DSP). In the DSP, the Fab protein product is purified by a number of operations, including centrifugation, homogenization, filtration, ultrafiltration/diafiltration (UF/DF), and three packed-bed chromatography steps, which include affinity, cation-exchange, and anion-exchange chromatography steps.

Chromatography operations are complex processes with challenges in modeling their behaviors. To optimize the chromatography operational strategies, metamodeling techniques are used in this work to mimic and predict the chromatography process performance, particularly the chromatography throughput, indicating the rate of the product output of one column within a given period, which is an important metric of the chromatography operation. To develop manufacturing-scale data-driven models, data from a microscale column chromatography laboratory experiment are collected and then fitted to obtain isotherm parameters. First-principle chromatography models are also created with scale-related parameters to capture challenges on scaling [34]. First-principle models with isotherm parameters are solved, followed by simulation runs to generate manufacturing-scale datasets using COMSOL Multiphysics simulation software [35]. The datasets include the throughput output under different input conditions of loaded mass, flow velocity, and column bed height at the two chromatography steps in the binding and elution mode—namely, the affinity and cation-exchange chromatography steps. The datasets are then used to derive data-driven models, which are incorporated into the proposed optimization models in order to reach optimal manufacturing-scale chromatography decisions. The whole procedure of the multiscale optimization approach is presented in Fig. 2; the steps presented in the last three tan-colored boxes in the figure will be described in detail in the following sections. Note that it is assumed that the considered input conditions do not affect other chromatography parameters, such as the resin’s yield, binding capacity, and lifetime, which are known parameters in this problem.

In summary, the optimization problem considered in this work is described as follows:

Given:
- The process flowsheet of a Fab product;
- The number of bioreactors and their volumes, along with the bioreactor titer;
- The chromatography operation parameters, including the yield, buffer and eluent usage, dynamic binding capacity, lifetime, etc.;
- The non-chromatography operation parameters, including the yield, processing rate, buffer usage, etc.;
- Time-related data, including the processing rate, bioreaction time, annual operating time, etc.;
- Cost-related data, including the labor wage and the costs of the resin, buffer, media, equipment, etc.;
- Chromatography data from simulations based on first-principle models and microscale column experiment data;
- The candidate column diameter and bed height, and the maximum number of cycles and columns.

Determine:
- Chromatography column-sizing strategies, such as the column diameter, bed height, and number of columns at each chromatography step;
- Operational strategies, such as the liner velocities, loaded product mass, and number of cycles at the affinity and cation-exchange chromatography steps;
The number of total completed batches;  
The annual total processing time;  
The annual total production output;  
The annual total cost.  

So as to minimize the cost of goods (COG) per gram of the Fab product—that is, the ratio of the annual total cost to the annual total production output.

3. Data-driven models

In this section, simple data-driven models are developed for chromatography performance based on simulation datasets. In this work, as a key performance criterion, the chromatography throughput is considered as the output of the models. Only the affinity and cation-exchange chromatography steps are modeled, due to the limited available data. Both the datasets for the affinity and cation-exchange chromatography steps are based on a single chromatography column with a diameter of 1 m. Three key variables influencing the chromatography throughput are considered in the datasets as the inputs of the data-driven models, namely: the loaded mass, flow velocity, and column bed height.

A number of widely used methods are implemented in order to achieve accurate and simple models, including linear regression, support vector regression (SVR) [36], kriging [37], pace regression [38], response surface methodology (RSM) [39], and piecewise linear regression [40]. To estimate the prediction accuracy of these methods, cross-validation is performed. Given a dataset, \( n \)-fold cross-validation randomly splits the samples into \( n \) subsets of equal size. Then \((n - 1)\) subsets of samples are used in the training, and the remaining set is used to validate the prediction accuracy of the obtained data-driven models. In this work, 10 rounds of five-fold cross-validation are performed by generating random sample splits, and the mean absolute error (MAE) over all 50 testing sets is used as the final error metric for comparison of the prediction accuracy. Linear regression, SVR, kriging, and pace regression are implemented in WEKA machine learning software [41] with default settings, while RSM and piecewise linear regression are run in GAMS [42] using the CPLEX mixed-integer linear programming (MILP) solver. Table 1 presents the prediction error results obtained after running the cross-validation on the datasets for affinity chromatography with 3081 samples, and for cation-exchange chromatography with 2847 samples.

Table 1 shows that the piecewise linear regression method gives the best prediction accuracy among all the tested methods. The piecewise linear regression method creates a model to separate samples into multiple complementary intervals on one input variable, with the flexibility of each interval being fitted by its own linear regression function. Considering its ease of modeling and understanding, the piecewise linear regression method is chosen to create the final data-driven model for chromatography throughput estimation, where all samples are used in the training process. In the procedure of the piecewise linear regression [40], each input variable in turn serves once as the partition variable. For each partition variable, an MILP model is solved to determine the breakpoint of the partition variable between only two intervals, and the variable corresponding to the minimum training error is kept as the partition variable. Until the termination criterion is met, the number of intervals is increased and MILP models are solved iteratively with the same partition variable. Following this procedure, the following two models are obtained to estimate throughput, \( TP_i \), at chromatography step \( s \) (af for affinity and ce for cation-exchange):  

\[
TP_i = \begin{cases} 
0.1914 LM_{ia} + 0.3570 V_d - 12.0477 H_d + 230.1318, & \text{if } 654 < LM_{ia} \leq 1643 \\
-0.003294 LM_{ia} + 0.008982 V_d + 43.5598 H_d - 649.3012, & \text{if } 1643 < LM_{ia} \leq 2629 \\
0, & \text{if } 2629 < LM_{ia} \leq 7069 
\end{cases}
\]  

(1)
where the superscript 1 of the variables refers to the column with a diameter of 1 m. In the obtained models, the loaded mass, \( LM^1_s \), at chromatography step \( s \) is determined to separate three intervals by the procedure given in Ref. [40], which provides a smaller prediction error than the other two variables, \( Vs \) and \( Hs \), which are the linear velocity and the column bed height, respectively.

In order to incorporate the above obtained data-driven models into the optimization models for decision-making, they need to be reformulated by introducing a binary variable. The binary variable, \( O_{sr} \), is defined to be equal to 1 if the loaded mass at chromatography step \( s \) lies within interval \( r \); the throughput output is then obtained from the corresponding linear function. As there is only one interval to be selected, the binary variable should satisfy Eq. (3):

\[
\sum_r O_{sr} = 1, \ \forall s \in \{af, ce\}
\]

(3)

The value of the separate input, \( LM^1_s \), in this model should be between the two breakpoints (bp\(_{r,1}\) of the selected interval, which can be formulated as the following linear equation:

\[
\sum_r bp_{r,1} \cdot O_{sr} + \varepsilon \leq LM^1_s \leq \sum_r bp_{r,2} \cdot O_{sr}, \ \forall s \in \{af, ce\}
\]

(4)

where \( \varepsilon \) is a small number to separate two successive intervals at the breakpoints. The general expression of the throughput output is as follows:

\[
TP^1_s = \sum_r \left( O_{sr} \cdot LM^1_s + \rho^b_{s,r} \cdot V_s + \rho^h_{s,r} \cdot H_s + \rho^c_{s,r} \right) \cdot O_{sr},
\]

\( \forall s \in \{af, ce\} \)

(5)

where \( \rho^b_{s,r} \), \( \rho^h_{s,r} \), and \( \rho^c_{s,r} \) are parameters in the function. When interval \( r \) is selected at step \( s \), that is, \( O_{sr} = 1 \), Eq. (4) becomes \( bp_{r,1} < LM^1_s \leq bp_{r,2} \). In this case, the loaded mass lies in the interval \( r \); Eq. (5) then ensures that the throughput is equal to the output of the linear function on the selected interval \( r \):

\[
TP^1_s = \left( O_{sr} \cdot LM^1_s + \rho^b_{s,r} \cdot V_s + \rho^h_{s,r} \cdot H_s + \rho^c_{s,r} \right), \text{if } bp_{r,1} \leq LM^1_s \leq bp_{r,2}
\]

(6)

The above throughput regression model will be incorporated into the optimization models of Fab manufacturing processes in the next section.

4. Optimization models

In this section, we will use the above data-driven models to develop two mixed-integer nonlinear programming (MINLP) models of the purification processes of a Fab product, using different modeling methods of alternative column sizes. These MINLP models are then reformulated as MILP models using exact linearization techniques and multiparametric disaggregation.

4.1. MINLP model A

MINLP model A is formulated based on the model provided in Ref. [16] for a mAb manufacturing process. In this model, a number of alternative column volume sizes are first generated from combinations of the given discrete column diameter and bed height. The column volume size \( i \) at chromatography step \( s (c_{si}) \) corresponds to a specific diameter size \( (d_{m,s}) \) and bed height \( (h_{s,i}) \) among the given alternatives.

4.1.1. Column volume

The total column volume \( (TCV_s) \) at chromatography step \( s \in CS \) (CS is the set of chromatography steps, \( CS = \{af, ce, ae\} \), ae stands for anion-exchange) is determined by the number of columns \( (CN_s) \) in the selected size multiplied by the corresponding column volume:

\[
TCV_s = \sum_i c_{si} \cdot CN_s, \ \forall s \in CS
\]

(7)

By introducing a binary variable, \( X_{si} \), for the selection of column size \( i \) at chromatography step \( s \), the following constraints can ensure that only one column size can be selected:

\[
\sum_i X_{si} = 1, \ \forall s \in CS
\]

(8)

\[
CN_s \leq \text{max} CN_i \cdot X_{si}, \ \forall s \in CS, i
\]

(9)

where \( \text{max} CN_i \) refers to the maximum allowed number of columns.

In each batch, the available resin volume at a chromatography step \( s \)—that is, the total column volume \( (TCV_s) \) multiplied by the number of cycles per batch \( (CN_s) \)—must be sufficient to process all protein mass entering into that step \( (M_{s-1}) \), and the required resin volume \( (RV_s) \) is determined by the dynamic binding capacity \( (dbc_s) \) and resin utilization factor \( (\mu_i) \).

\[
CN_s \cdot TCV_s \geq RV_s, \ \forall s \in CS
\]

(10)

\[
RV_s = \frac{M_{s-1}}{dbc_s \cdot \mu}, \ \forall s \in CS
\]

(11)

4.1.2. Product mass

The initial product mass \( (M_0) \) entering into the DSP process is equal to the bioreactor titer \( (titer) \) multiplied by the bioreactor working volume—that is, the bioreactor volume \( (brv) \) times the working volume ratio \( (\rho) \).

\[
M_0 = \text{titer} \cdot \rho \cdot brv
\]

(12)

The product protein mass going out from step \( s \) is equal to the mass from the previous one, step \( s-1 \), multiplied by the corresponding yield of step \( s \), \( yd_s \).

\[
M_s = yd_s \cdot M_{s-1}, \ \forall s
\]

(13)

The annual product output \( (AP) \) is the product mass after the bulk fill step \( (s = bf) \):

\[
AP = \sigma \cdot BN \cdot M_{bf}
\]

(14)

where \( BN \) is the number of completed batches, upper bounded by the maximum allowed batch number, and \( \sigma \) is the batch success rate.

4.1.3. Product volume

The initial product volume entering into the DSP \( (PV_0) \) is equal to the working volume of the bioreactor, formulated as follows:

\[
PV_0 = \rho \cdot brv
\]

(15)

For the first four steps of the process, including the first centrifugation \( (s = ct_1) \), homogenization \( (s = ho) \), second centrifugation \( (s = ct_2) \), and filtration \( (s = fi) \) steps, the product volume remaining after step \( s \) \( (PV_s) \) is equivalent to the product volume entering into this step:

\[
PV_s = PV_{s-1}, \ \forall s \in \{ct_1, ho, ct_2, fi\}
\]

(16)

At the affinity \( (s = af) \) and cation-exchange chromatography \( (s = ce) \) steps, the product volume is equal to the eluent volume, while at the anion-exchange chromatography \( (s = ae) \) step, the product volume does not change.
where $ec \nu_s$ is the eluent volume to column volume ratio.

At the first UF/DF step ($s = uf_1$), the flush volume is added to the product volume entering the step:

$$PV_{uf_1} = (r + 1) \cdot PV_{ce}$$

where $r$ is the flush volume ratio at this step. At the second UF/DF step ($s = uf_2$), the remaining product volume is the mass divided by the filling concentration, $fconc$:

$$PV_{uf_2} = \frac{M_{uf_2}}{fconc}$$

### 4.1.4. Buffer volume

The buffer volume added in each step ($BV_s$) is defined as follows:

$$BV_s = bvf_s \cdot PV_{s-1}, \forall s \in \{ct_1, ct_2\}$$

$$BV_s = 0, \forall s \in \{ho, fi\}$$

$$BV_s = bc \nu_s \cdot CYNs \cdot TCV_s, \forall s \in CS$$

The total required buffer volume in each batch ($BBV_s$) is the sum of the buffer volume at all steps, and the annual buffer volume ($ABV_s$) is the total buffer volume of all the completed batches:

$$BBV = \sum_s BV_s$$

$$ABV = BN \cdot BBV$$

### 4.1.5. Processing time

The processing times ($T_s$) at the affinity and cation-exchange chromatography steps are determined by the mass output of each column divided by its throughput ($TP_s$):

$$T_s = \frac{M_s}{TP_s \cdot \sum CN_{s,i}}, \forall s \in \{af, ce\}$$

At the anion-exchange chromatography step, the processing times for the loading product ($PLT_s$) and adding buffer ($BAT_s$) are calculated separately using the volumetric flow rate ($VFR_s$) to obtain the processing time at the step:

$$T_{ae} = PLT + BAT$$

$$PLT_s = \frac{PV_{uf_1}}{VFR \cdot \sum CN_{ae,i}}$$

$$BAT_s = \frac{CYN_{ae} \cdot bc \nu_{ae} \cdot \sum c \nu_{ae,i} \cdot X_{ae,i} \cdot TCV_s}{VFR}$$

$$VFR_s = \frac{1}{1000} \cdot vel \cdot \pi \cdot \sum_i \left(\frac{dm_{ae,i}}{2}\right)^2 \cdot X_{ae,i}$$

where $vel$ is the linear velocity of flow at the anion-exchange chromatography step. The processing time at the bulk fill step is assumed to be constant, while at the other non-chromatography steps, the process time is equal to the corresponding product volume divided by the processing rate ($pr_s$):

$$T_s = \frac{PV_s}{pr_s}, \forall s \in \{ct_1, ho, ct_2, fi, uf_1, uf_2\}$$

The processing time of one batch, $BT_s$, is the total processing time for all steps, divided by the shift duration ($sf$) and the number of shifts per day ($sfd$):

$$BT = \frac{\sum T_s}{sfd \cdot sfn}$$

The annual processing time ($AT$) is the total processing time of all batches:

$$AT = BN \cdot BT$$

### 4.1.6. Data-driven model

The throughput of the 1 m-diameter column is calculated from the piecewise linear regression model obtained in the previous section, including Eqs. (3) and (4). As the selected bed height at chromatography step $s$ is expressed as $\sum h_{s,i} \cdot X_{s,i}$ in this model, Eq. (5) is modified as follows:

$$TP_s = \frac{\sum \left(\beta_s t^{m} \cdot LM_{s,i} + \beta_s t^{r} \cdot V_s + \beta_s t^{d} \cdot \sum h_{s,i} \cdot X_{s,i} + \beta_s t^{b} \cdot O_{s,i}\right)}{\left(\text{refDM}_s\right)^2}, \forall s \in \{af, ce\}$$

Considering that the throughput of a chromatography column could be regarded as the product of the protein density, linear velocity of flow, and column area, it is assumed that the throughput is proportional to the column area. Thus, it is also proportional to the column diameter squared. In this case, the relationship between the throughput of the selected column and that of the 1 m-diameter column ($TP_{1m}$) is formulated as follows:

$$TP_{1m} = \frac{\sum (dm_{s,i})^2 \cdot X_{s,i}}{(\text{refDM})^2}, \forall s \in \{af, ce\}$$

where $\text{refDM}$ refers to the reference diameter, which is 100 cm in this work.

In the data-driven regression model, the mass loaded to the 1 m-diameter column is used to proportionally calculate the actual loaded mass to the selected column ($LM_s$):

$$LM_s = \frac{\sum (dm_{s,i})^2 \cdot X_{s,i}}{(\text{refDM})^2} \cdot LM_{1m}, \forall s \in \{af, ce\}$$

where the loaded mass, $LM_s$, is defined as the product mass entering each column in each cycle, defined as follows:

$$LM_s = \frac{M_{s-1}}{\sum CN_{s,i}}, \forall s \in \{af, ce\}$$

### 4.1.7. Objective function

In this work, the objective is to minimize COG per gram, which equals the annual total cost ($AC$) divided by the annual total production output ($AP$). All cost terms included in the annual total cost calculation and the related constraints are presented in Supplementary data. The objective function is as follows:

$$COG = \frac{AC}{AP}$$
Thus, the proposed MINLP model A minimizes Eq. (39), subject to the constraints of Eqs. (3), (4), and (7–38), and Eqs. (S1–S19) in Supplementary data.

4.2. MILP model A∗

Next, the obtained MINLP model is reformulated as an MILP model for ease of computation. The new linear constraints are presented below.

4.2.1. Integer variable discretization

Similar to the work in Refs. [17–19], in order to facilitate the linearization of nonlinear terms in other constraints, the integer variables CNs, CYNs, and BN are discretized and expressed by binary variables, as shown in Eqs. (40–44).

\[
\begin{align*}
CN_{i,j} &= \max_{j=1} \sum j \cdot W_{s,i,j}, \forall s \in CS, i \\
\max_{i} \sum W_{s,i,j} &= X_{s,i}, \forall s \in CS, i \\
CYN_{i} &= \max_{k=1} \sum k \cdot Y_{s,i,k}, \forall s \in CS \\
\max_{k=1} Y_{s,i,k} &= 1, \forall s \in CS \\
BN &= \sum_{n=0}^{\log_{2}\max BN} 2^{n-1} \cdot Z_{n}
\end{align*}
\]

where \(W_{s,i,j}, Y_{s,i,k}\), and \(Z_{n}\) are binary variables introduced for discretization of the above integer variables, and \(j\) and \(k\) are the indices of column number and cycle number, respectively.

4.2.2. Column volume linearization

Based on Eqs. (42) and (43), the nonlinear constraint Eq. (10) can be reformulated as follows:

\[
\begin{align*}
\max_{k=1} \sum k \cdot \mathcal{V}_{s,i,k} &\geq RV_{s,i}, \forall s \in CS, k = 1, \ldots, \max CYN_{i} \\
\mathcal{V}_{s,i,k} &\leq \max TCV_{s,i} \cdot Y_{s,i,k}, \forall s \in CS, k = 1, \ldots, \max CYN_{i} \\
\sum_{k=1}^{\max CYN_{i}} \mathcal{V}_{s,i,k} &= TCV_{s,i}, \forall s \in CS
\end{align*}
\]

where an auxiliary variable \(\mathcal{V}_{s,i,k} = Y_{s,i,k} \cdot TCV_{s,i}\) is introduced and defined, and \(\max TCV_{s,i}\) is the maximum total column volume at chromatography step \(s\).

4.2.3. Annual production linearization

By introducing another auxiliary variable, \(\mathcal{Z}_{M,T} = Z_{n} \cdot M_{n}\), to express the bilinear term in Eq. (14), this constraint can be reformulated by the following equations [43]:

\[
\begin{align*}
\mathcal{Z}_{M,T} &\geq \text{titer} \cdot \alpha \cdot \frac{br \cdot v}{Z_{n}}, \forall n = 1, \ldots, \lfloor \log_{2}\max BN \rfloor \\
\mathcal{Z}_{M,T} &\leq M_{\text{bf}}, \forall n = 1, \ldots, \lfloor \log_{2}\max BN \rfloor
\end{align*}
\]

4.2.4. Product and buffer volume linearization

According to Eqs. (46) and (47), Eqs. (17) and (22) can be rewritten as the following two linear equations, respectively:

\[
\begin{align*}
\Pi_{s,i} &= \frac{ec}{v_{s,i}} \cdot \sum_{k=1}^{\max CYN_{i}} k \cdot \mathcal{V}_{s,i,k} + PV_{\text{af}}_{s,i}, \forall s \in CS \\
\mathcal{B}_{s,i} &= \frac{bc}{v_{s,i}} \cdot \sum_{k=1}^{\max CYN_{i}} k \cdot \mathcal{V}_{s,i,k}, \forall s \in CS
\end{align*}
\]

Thus, Eq. (26) can be rewritten as follows:

\[
\begin{align*}
\mathcal{ABV} &= \sum_{n=0}^{\log_{2}\max BN} 2^{n-1} \cdot \mathcal{Z}_{n}
\end{align*}
\]

4.2.5. Processing time linearization

Eq. (27) can be reformulated to include a nonlinear term of the product of one integer variable, \(CN_{s,i}\), and two continuous variables, \(T_{s}\) and \(P_{s}\). Two auxiliary variables are introduced: \(\mathcal{TTP}_{s,i} = T_{s} \cdot P_{s}\) and \(\mathcal{WTTP}_{s,i,j} = W_{s,i,j} \cdot T_{s} \cdot P_{s}\). \(\mathcal{WTTP}_{s,i,j}\) can be determined by \(\mathcal{TTP}_{s,i}\) using the following equations:

\[
\begin{align*}
\sum_{i} \sum_{j=1}^{\max CYN_{i}} j \cdot \mathcal{WTTP}_{s,i,j} &= M_{s}, \forall s \in \{af, ce\} \\
\mathcal{WTTP}_{s,i,j} &\leq \max TP_{s} \cdot \max T_{s} \cdot W_{s,i,j}, \forall s \in \{af, ce\}, i, j = 1, \ldots, \max CYN_{i} \\
\sum_{i} \sum_{j=1}^{\max CYN_{i}} j \cdot \mathcal{WTTP}_{s,i,j} &= \mathcal{TTP}_{s,i}, \forall s \in \{af, ce\}
\end{align*}
\]

Next, \(\mathcal{TTP}_{s}\) is reformulated using multiparametric disaggregation [44–46], where the throughput, \(TP_{s}\), is expressed as a multiparametric sum of active decimal powers determined by binary variable \(\mathcal{BTP}_{s,d}\) and continuous variable \(\mathcal{CTP}_{s,q} \in [0, 1]\), where \(d\) is the digit position ranging from \(p\) to \(\max p\), and \(q\) is the number for power \(d\).

\[
\begin{align*}
TP_{s} &= \max_{d=0}^{p} \sum_{q=d}^{q} 10^{d} \cdot q \cdot \mathcal{BTP}_{s,d,q} + \sum_{q=0}^{q} 10^{p} \cdot q \cdot \mathcal{CTP}_{s,q} \\
\forall s \in \{af, ce\}
\end{align*}
\]

\[
\begin{align*}
\sum_{q=0}^{q} \mathcal{BTP}_{s,d,q} &= 1, \forall s \in \{af, ce\}, d = p, \ldots, \max x \\
\sum_{q=0}^{q} \mathcal{CTP}_{s,q} &= 1, \forall s \in \{af, ce\}
\end{align*}
\]

Based on the above equations, variable \(\mathcal{TTP}_{s}\) is disaggregated into a number of non-negative continuous variables \(\mathcal{BTP}_{s,d,q}\) and \(\mathcal{CTP}_{s,q}\).
\( \prod_{d-p} = \sum_{d-p} \sum_{q-d}^{9} 10^d \cdot q \cdot BTIP_{d-q} + \sum_{q-d}^{9} 10^p \cdot q \cdot CTIP_{q} \),

(64)

\( \forall s \in \{af, ce\} \)

\( BTIP_{d-q} \leq \max(T_d - BTIP_{d-q}) \),

(65)

\( \forall s \in \{af, ce\}, d = p, \ldots, \max(p, q = 0, \ldots, 9 \}

\( CTIP_{d-q} \leq \max(T_d - CTIP_{d-q}) \),

(66)

\( \forall s \in \{af, ce\}, q = 0, 1 \}

\( \sum_{q=0}^{9} BTIP_{d-q} = T_s \),

(67)

\( \forall s \in \{af, ce\} \)

Note that multiparametric disaggregation is a relaxation method, and the above reformulation provides a close approximation of the original equations, and generates a lower bound of the optimum.

By defining \( \mathcal{WTT}_{sij} \equiv W_{sij} \cdot PLT \) using the following constraints:

\( \mathcal{WTT}_{sij} \leq \max(T) \cdot W_{sij} \),

(69)

\( \forall i = 1, \ldots, \maxCN_{w} \)

\( \sum_{i} \sum_{j=1}^{\maxCN_{w}} \mathcal{WTT}_{sij} = PLT \)

(70)

The following linear constraint replaces Eq. (29):

\( \frac{1}{1000} \sum_{i} \sum_{j=1}^{\maxCN_{w}} \frac{i \cdot \text{vel} \cdot \pi \cdot (\text{dem}_{ej}/2)^2}{\mathcal{WTT}_{sij} = \text{PV}_{ufi}} \)

(71)

Using another auxiliary variable, \( \mathcal{XY}_{sij} \equiv X_{sij}, Y_{sij}, \) and the following constraints are equivalent to Eq. (30):

\( \mathcal{BAT}_{sij} = \sum_{k=1}^{\maxCN_{y}} \frac{bc}{\text{CV}_{ajt} \cdot k \cdot \mathcal{XY}_{ajt}} \cdot \mathcal{Y}_{ajt} = K_{ajt} \cdot \mathcal{XY}_{ajt} \quad \forall k = 1, \ldots, \maxCN_{y} \)

(72)

\( \maxCN_{y} \sum_{k=1}^{\maxCN_{y}} \mathcal{XY}_{ajt} = X_{ajt} \),\( \forall i \)

(74)

Based on Eq. (44), Eq. (34) can be rewritten as follows:

\( AT = \sum_{n=1}^{\log_{10}(\maxBN)} 2^{n-1} \cdot \mathcal{ZT}_n \)

(75)

where \( \mathcal{ZT}_n = Z_n \cdot BT \), satisfying:

\( \mathcal{ZT}_n \leq (aot - st - btr) \cdot Z_n, \forall n = 1, \ldots, \log_{10}(\maxBN) \)

(76)

\( \mathcal{ZT}_n \leq BT, \forall n = 1, \ldots, \log_{10}(\maxBN) \)

(77)

\( \mathcal{ZT}_n \geq BT - (aot - st - btr) \cdot (1 - Z_n), \forall n = 1, \ldots, \log_{10}(\maxBN) \)

(78)

4.2.6. Data-driven model reformulation

The bilinear terms of the regression models in Eqs. (4) and (35) can be rewritten as the following linear constraints [47]:

\( bp_{sij} = b - \max(LM)_{i} \cdot (1 - O_{s}) \leq LM_{i} \)

\( \leq bp_{sij} + \max(LM)_{i} \cdot (1 - O_{s}), \forall s \in \{af, ce\}, r \)

(79)

\( \beta_{j}^{L} \cdot LM_{j} \leq \beta_{j}^{L} \cdot V_{j} + \beta_{j}^{H} \cdot \sum_{i} h_{ij} \cdot X_{ij} \)

+ \beta_{j}^{H} \cdot \max(TP_{j}) \cdot (1 - O_{ij}) \leq \beta_{j}^{L} \cdot LM_{j} \)

\( + \beta_{j}^{H} \cdot V_{j} + \beta_{j}^{H} \cdot \sum_{i} h_{ij} \cdot X_{ij} + \beta_{j}^{H} \cdot \max(TP_{j}) \cdot (1 - O_{ij}) \),

\( \forall s \in \{af, ce\}, r \)

(80)

Eq. (36) can be linearized using a new auxiliary variable, \( \mathcal{XTP}_{ij} \equiv X_{ij} \cdot TP_{ij} \), and the following constraints:

\( TP_{ij} = \sum_{i=1}^{(dem}_{ij})^2 \cdot \mathcal{XTP}_{ij} \), \( \forall s \in \{af, ce\} \)

(81)

\( \mathcal{XTP}_{ij} \leq \max(TP_{ij} \cdot X_{ij} \), \( \forall s \in \{af, ce\}, i \)

(82)

\( \sum_{i} \mathcal{XTP}_{ij} = TP_{ij}, \forall s \in \{af, ce\} \)

(83)

Similarly, using the auxiliary variable \( \mathcal{XM}_{ij} \equiv X_{ij} \cdot LM_{ij} \), Eq. (37) is equivalent to:

\( LM_{ij} = \sum_{i=1}^{(dem}_{ij})^2 \cdot \mathcal{XM}_{ij} / \text{refDM} \)

(84)

\( \mathcal{XM}_{ij} \leq \max(LM)_{ij} \cdot X_{ij}, \forall s \in \{af, ce\}, i \)

(85)

\( \sum_{i} \mathcal{XM}_{ij} = LM_{ij}, \forall s \in \{af, ce\} \)

(86)

(87)

Eq. (38) includes a nonlinear term involving two integer variables and one continuous variable, which can be linearized as follows:

\( \sum_{i=1}^{maxCN_{y}} \sum_{k=1}^{maxCN_{y}} j \cdot k \cdot \mathcal{XY}_{sijk} = M_{i-1}, \forall s \in \{af, ce\}, i \)

(88)

\( \sum_{i=1}^{maxCN_{y}} \sum_{k=1}^{maxCN_{y}} \mathcal{XY}_{sijk} = Y_{sijk}, \forall s \in \{af, ce\}, i, j = 1, \ldots, \maxCN_{i} \)

(89)

\( \sum_{i=1}^{maxCN_{y}} \sum_{k=1}^{maxCN_{y}} \mathcal{XY}_{sijk} = Y_{sijk}, \forall s \in \{af, ce\}, k = 1, \ldots, \maxCN_{i} \)

(90)

\( \sum_{i=1}^{maxCN_{y}} \sum_{k=1}^{maxCN_{y}} \mathcal{XY}_{sijk} = LM_{s}, \forall s \in \{af, ce\} \)

(91)

where there are two auxiliary variables, \( \mathcal{WXY}_{sijk} \equiv W_{sijk}, \) and \( \mathcal{WXY}_{sijk} \equiv W_{sijk} \cdot Y_{sijk} \cdot LM_{s} \).

4.2.7. Objective function linearization

The calculation of consumables cost in Eq. (9) in Supplementary data involves nonlinearities. By introducing auxiliary variable \( \mathcal{ZV}_{sijk} \equiv Z_{sijk}, \) Eq. (9) can be reformulated as follows:

\( CC = \sum_{i=1}^{log_{10}(\maxBN)} \sum_{k=1}^{maxCN_{y}} \sum_{i=1}^{maxCN_{y}} k \cdot \text{rp}_{c} \cdot \mathcal{ZV}_{sijk} / \text{refVC} \)

(92)

where of refers to the resin overpacking factor.

\( \mathcal{ZV}_{sijk} \leq \mathcal{ZV}_{sijk}, \forall s \in \mathcal{CS}, k = 1, \ldots, \maxCN_{i}, n = 1, \ldots, \log_{10}(\maxBN) \)

(93)

\( \mathcal{ZV}_{sijk} \leq \mathcal{ZV}_{sijk}, \forall s \in \mathcal{CS}, k = 1, \ldots, \maxCN_{i}, n = 1, \ldots, \log_{10}(\maxBN) \)
where $r_{pc}$ and $l_i$ are the resin price and resin lifetime at chromatography step $s$, respectively.

In the objective function Eq. (39), the annual cost can be expressed as the product of COG per gram and the annual production, which can be written as follows:

$$\text{COGAP} = AC$$

where auxiliary variable $\text{COGAP} = \text{COG} \cdot AP$. Using multiparametric disaggregation and introducing new auxiliary variables $\text{BCOGAP}_{dp} = \text{COG} \cdot \text{BAP}_{dp}$ and $\text{CCOGAP}_{pq} = \text{COG} \cdot \text{CAP}_{pq}$, the variables $AP$ and $\text{COGAP}$ can be disaggregated as follows:

$$AP = \max_{d-p} \left[ \sum_{q=0}^{9} 10^q \cdot \text{BAP}_{dp} + \sum_{q=0}^{9} 10^q \cdot \text{CAP}_{pq} \right]$$

$$\text{COGAP} = \max_{d-p} \left[ \sum_{q=0}^{9} 10^q \cdot \text{BCOGAP}_{dp} + \sum_{q=0}^{9} 10^q \cdot \text{CCOGAP}_{pq} \right]$$

$$\sum_{q=0}^{9} \text{BAP}_{dp} = 1, \forall d = p, \ldots, \max_{p}$$

$$\sum_{q=0}^{1} \text{CAP}_{pq} = 1$$

$$\text{BCOGAP}_{dp} \leq \max_{d-p} \text{COG} \cdot \text{BAP}_{dp}, \forall d = p, \ldots, \max_{p}, q = 0, \ldots, 9$$

$$\text{CCOGAP}_{pq} \leq \max_{d-p} \text{COG} \cdot \text{CAP}_{pq}, \forall q = 0, 1$$

Thus, the reformulated MILP model $A^*$ includes the constraints shown in Eqs. (3), (7–9), (11–13), (15), (16), (18–21), (23–25), (28), (32), (33), and (40–104); and Eqs. (S1–S8) and (S10–S19) in Supplementary data, with the variable COG as the objective.

4.3. MINLP model B

In this section, an alternative MINLP model B is introduced. By generating a set of column volume sizes, the models A and $A^*$ result in a large number of variables and equations, which hinder the computation of the proposed models. To overcome this shortcoming, in the new MINLP model B, we get rid of the discrete column volume sizes, and introduce an integer variable, $H_s$, for the bed height, and a binary variable, $E_{m,s}$, to indicate whether diameter size, $d_m$, is selected. In addition, the subscript $i$ is removed from the chromatography column number variable, and variable $CN_i$ expresses the number of columns at chromatography step $s$, which is upper bounded by $\max CN_s$. Thus, the number of discrete variables in model B is reduced with improved computational efficiency. Based on the proposed MINLP model A presented above, a number of new variables and constraints are developed for the MINLP model B, as introduced below.

4.3.1. Column volume

In model B, the total column volume is calculated using the variables of bed height ($H_s$), diameter selection ($E_{m,s}$), and number of columns ($CN_i$).

$$\text{TCV}_s = \sum_{m} \left( \frac{d_m}{2} \right)^2 \cdot E_{m,s} \cdot H_s \cdot CN_i, \forall s \in CS$$

In addition, only one diameter size can be selected at each chromatography step.

$$\sum_{m} E_{m,s} = 1, \forall s \in CS$$

4.3.2. Processing time

By replacing $\sum_i CN_{i,s}$ by $CN_i$ in Eqs. (27) and (29), the following constraint can be obtained:

$$T_s \cdot TP_s \cdot CN_i = M_i, \forall s \in \{af, ce\}$$

$$PLT = \frac{PV_{af}}{\text{VFR} \cdot CN_{af}}$$

Similarly, $\text{BAT}$ and $\text{VFR}$ are formulated using the new integer variable, $H_s$, and the binary variable, $E_{m,s}$:

$$\text{BAT} = \sum_{m} \left( \frac{d_m}{2} \right)^2 \cdot E_{m,s} \cdot H_s$$

$$\text{VFR} = \frac{1}{1000} \cdot \text{vel} \cdot \pi \cdot \sum_m \left( \frac{d_m}{2} \right)^2 \cdot E_{m,s}$$

4.3.3. Data-driven model

Eq. (38) can be written using the new column number variable as follows:

$$\text{LM}_i = \frac{M_{s,1}}{CN_i \cdot CN_i}, \forall s \in \{af, ce\}$$

Similar to Eqs. (36) and (37), the throughput and loaded mass of the selected columns can be calculated from those of the 1 m-diameter columns.

$$TP_i = \sum_m \left( \frac{d_m}{2} \right)^2 \cdot E_{m,s} \cdot \text{TP}_{s}^1, \forall s \in \{af, ce\}$$

$$\text{LM}_i = \sum_m \left( \frac{d_m}{2} \right)^2 \cdot E_{m,s} \cdot \text{LM}_{s}^1, \forall s \in \{af, ce\}$$

4.3.4. Cost

In the cost calculation, Eq. (S13) for fixed capital investment (FCI) in Supplementary data should be replaced by the following nonlinear constraint:

$$\text{FCI} = \text{lang} \cdot (1 + \text{gef}) \cdot \left( \text{brc} \cdot \text{brn} + \sum_{i=CS} \sum_m \tilde{c}_{i,m} \cdot CN_i \cdot E_{m,s} + \text{oel} \cdot \text{brc} \cdot \text{brn} \right)$$

where $\text{lang}$ is the Lang factor; $\text{gef}$ is the general equipment factor; $\text{brn}$ is the number of bioreactors; $\text{brc}$ is the cost of one bioreactor; $\text{oel}$ is the ratio of other equipment cost to the bioreactor cost; $\tilde{c}_{i,m}$ is the chromatography column cost of diameter size $m$ at chromatography step $s$. 
Thus, the proposed MINLP model B minimizes Eq. (39), subject to the constraints shown in Eqs. (3–5), (10–26), (28), (32–34), and (105–114); and Eqs. (S1–S12) and (S14–S19) in Supplementary data.

4.4. MILP model $B^*$

In MILP model $B^*$, all nonlinear constraints of MINLP model B are linearized. Besides the linear constraints presented in MILP model $A^*$, the newly developed ones are given below.

4.4.1. Integer variable discretization

At first, integer variable $CN_i^*$ is discretized using binary variable $F_{sj}$, which represents whether or not $j$ columns are used at chromatography step $s$, as follows:

$$CN_i^* = \sum_{j=1}^{maxCN_i} F_{sj}, \forall s \in \{af, ce\}$$

$$\max_{j=1}^{maxCN_i} F_{sj} = 1, \forall s \in \{af, ce\}$$

4.4.2. Column volume linearization

In order to linearize Eq. (105), an auxiliary variable, $EF_{s,m,j} = E_s \cdot F_{sj} \cdot H_s$, is introduced, with the following constraints:

$$EF_{s,m,j} \leq maxH_s \cdot EF, \forall s \in CS, m, j = 1, \ldots, maxCN_i$$

$$\sum_{m} \sum_{j=1}^{maxCN_i} EF_{s,m,j} = H_s, \forall s \in CS$$

where $EF_{s,m,j} = E_s \cdot F_{sj}$ is defined as follows:

$$\sum_{m} EF_{s,m} = F_{sj}, \forall s \in \{af, ce\}, j = 1, \ldots, maxCN_i$$

$$\max_{m} EF_{s,m} = E_s, \forall s \in \{af, ce\}, m$$

Therefore, Eq. (105) is reformulated to the following linear constraint:

$$TCV_s = \sum_{m} \sum_{j=1}^{maxCN_i} j \cdot \pi \cdot \left( \frac{dm_m}{2} \right)^2 \cdot EF_{s,m,j}, \forall s \in \{af, ce\}$$

4.4.3. Processing time linearization

According to Eqs. (61–68), the term $T_s \cdot TP_s$ can be expressed by $TTP_s$. Therefore, $TTP_{sj}$ is introduced to express $F_{sj} \cdot TTP_s$, with the following constraints:

$$TTP_{sj} \leq maxT_s \cdot maxF_{sj}, \forall s \in \{af, ce\}, j = 1, \ldots, maxCN_i$$

$$\sum_{j=1}^{maxCN_i} j \cdot TTP_{sj} = TTP_s, \forall s \in \{af, ce\}$$

Thus, Eq. (107) is rewritten as follows:

$$\sum_{j=1}^{maxCN_i} j \cdot TTP_{sj} = M_s, \forall s \in \{af, ce\}$$

Based on the introduction of auxiliary variable $EFT_{s,m,j} = EF_{s,m,j} \cdot PLT$ and Eqs. (119) and (120), Eq. (108) can be linearized as follows:

$$\frac{1}{1000} \sum_{m} \sum_{j=1}^{maxCN_i} j \cdot vel \cdot \pi \cdot \left( \frac{dm_m}{2} \right)^2 \cdot EFT_{s,m,j} = PV_{sd}$$

$$\max_{m} EFT_{s,m,j} = maxT_{ae} \cdot EF_{s,m,j}, \forall m, j = 1, \ldots, maxCN_{ae}$$

$$\sum_{m} \sum_{j=1}^{maxCN_{ae}} EFT_{s,m,j} = PLT$$

$$\max_{j=1}^{maxCN_i} \sum_{m} H_{ae,j} \cdot \gamma_{ae,j}, \forall k = 1, \ldots, maxCN_{ae}$$

4.4.4. Data-driven model reformulation

Similar to Eq. (80), the data-driven model constraint given in Eq. (5) can be re-formulated to linear constraints [47], as follows:

$$PLT^1 \leq \beta_{LM}^1 \cdot LM^1 + \beta_{re}^1 \cdot V_s + \beta_{PLT} H_s + \beta_{PLT} - \max TP, (1 - O_s)$$

$$\max TP \leq \beta_{PLT} \cdot LM^1 + \beta_{re}^1 \cdot V_s + \beta_{PLT} H_s + \beta_{PLT}$$

$$\max TP \cdot (1 - O_s), \forall s \in \{af, ce\}, r$$

Eq. (111) is linearized to Eq. (132) by introducing auxiliary variables $FYLM_{s,j,k} = F_{sj} \cdot Y_{s,k} \cdot LM_s$ and $FY_{s,j,k} = F_{sj} \cdot Y_{s,k}$.

$$\sum_{j=1}^{maxCN_i} \sum_{k=1}^{maxCN_{ae}} j \cdot k \cdot FYLM_{s,j,k} = M_{s,k-1}, \forall s \in \{af, ce\}$$

$$\sum_{k=1}^{maxCN_{ae}} FY_{s,j,k} = F_{sj}, \forall s \in \{af, ce\}, j = 1, \ldots, maxCN_i$$

$$\sum_{k=1}^{maxCN_{ae}} FY_{s,j,k} = Y_{s,k}, \forall s \in \{af, ce\}, k = 1, \ldots, maxCN_i$$

$$\max FYLM_{s,j,k} = maxLM_s \cdot FY_{s,j,k}, \forall s \in \{af, ce\}, j = 1, \ldots, maxCN_i, k = 1, \ldots, maxCN_i$$

$$\max FYLM_{s,j,k} = LM_s, \forall s \in \{af, ce\}$$

In Eqs. (112) and (113), the nonlinearities involve the product of $E_{s,m}$ and a continuous variable. With the introduction of the auxiliary variables, $EFP_{s,m}$ and $ELM_{s,m}$, the following linear equations are used instead as linear constraints:

$$TP_s = \frac{\sum_{m} \left( \frac{dm_m}{2} \right)^2 \cdot EFP_{s,m}}{(refDM)^2}, \forall s \in \{af, ce\}$$

$$EFP_{s,m} \leq \max TP_s \cdot E_{s,m}, \forall s \in \{af, ce\}, m$$

$$\sum_{m} EFP_{s,m} = TP_s, \forall s \in \{af, ce\}$$

$$LM_s = \frac{\sum_{m} \left( \frac{dm_m}{2} \right)^2 \cdot ELM_{s,m}}{(refDM)^2}, \forall s \in \{af, ce\}$$
\[
\begin{align*}
\sum_m \mathcal{FM}_{im} & \leq \max M \cdot E, \quad \forall s \in \{af, ce\}, m  \\
\sum_m \mathcal{FM}_{im} = LM_i, \quad \forall s \in \{af, ce\}
\end{align*}
\]

(141) \hspace{1cm} (142)

4.4.5. Cost linearization

At last, in order to reformulate the fixed capital investment formula, \( \mathcal{EF}_{i,m} \) can be used to linearize Eq. (114) as follows:

\[
FCI = \text{lang} \cdot (1 + \text{gef}) \cdot \left( brc - brn + \sum_{s \in \{af, ce\}} \sum_{m} \max \{c_{x,m}, k \cdot \mathcal{EF}_{i,m}\} + \alpha e \cdot brc - brn \right)
\]

(143)

Overall, MILP model B* includes constraints shown in Eqs. (3), (11–13), (15), (16), (18–21), (23–25), (28), (32), (33), (42–57), (61–68), (75–79), (92–104), (106), and (115–143); and Eqs. (51–58), (510–512), and (514–519) in Supplementary data.

In summary, the equations of all four proposed models are summarized in Table 2.

5. Case study

In this section, the four proposed optimization models are applied to industrially relevant case studies to examine their performances. The process flowsheet of this example is shown in Fig. 1, where there are one bioreactor and three chromatography steps. The process flowsheet of this example is shown in Fig. 1, where there are one bioreactor and three chromatography steps. The process flowsheet of this example is shown in Fig. 1, where there are one bioreactor and three chromatography steps. The process flowsheet of this example is shown in Fig. 1, where there are one bioreactor and three chromatography steps.

Table 2

<table>
<thead>
<tr>
<th>Constraints</th>
<th>MINLP A</th>
<th>MILP A*</th>
<th>MINLP B</th>
<th>MILP B*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integer variable discretization</td>
<td>–</td>
<td>Eqs. (40–44)</td>
<td>–</td>
<td>Eqs. (42–44), (115), (116)</td>
</tr>
<tr>
<td>Column volume</td>
<td>Eqs. (7–11)</td>
<td>Eqs. (7–9), (11), (45–47)</td>
<td>Eqs. (10), (11), (105), (106)</td>
<td>Eqs. (11), (45–47), (106), (117–121)</td>
</tr>
<tr>
<td>Product mass</td>
<td>Eqs. (12–14)</td>
<td>Eqs. (12), (13), (48–51)</td>
<td>Eqs. (12–14)</td>
<td>Eqs. (12), (13), (48–51)</td>
</tr>
<tr>
<td>Product volume</td>
<td>Eqs. (15–19)</td>
<td>Eqs. (15), (16), (18), (19), (52)</td>
<td>Eqs. (15–19)</td>
<td>Eqs. (15), (16), (18), (19), (52)</td>
</tr>
<tr>
<td>Buffer volume</td>
<td>Eqs. (20–26)</td>
<td>Eqs. (20), (21), (23–25), (53–57)</td>
<td>Eqs. (20–26)</td>
<td>Eqs. (20), (21), (23–25), (53–57)</td>
</tr>
</tbody>
</table>

Table 3

| Alternatives of chromatography column sizes in two cases. |
|-----------------|-----------------|-----------------|
| Step            | Case 1           | Case 2           |
| Diameter (cm)   | 50, 60, 70, 80, 90, 100, 120, 160, 180, 200 | 50 to 300 with a step size of 10 |
| Bed height (cm) | 15 to 25 with a step size of 1 | 10 to 30 by with a step size of 1 |

Table 4

| Parameters of chromatographic operations. |
|-----------------|-----------------|-----------------|
| Step            | Affinity chromatography | Cation-exchange chromatography | Anion-exchange chromatography |
| Yield, yd,      | 95%              | 95%              | 98%              |
| Dynamic binding capacity, dbc, (g L\(^{-1}\)) | 15              | 50              | 100              |
| Eluate volume to column volume ratio, ecv, | 2              | 3              | 0               |
| Buffer volume to column volume ratio, bcv, | 15              | 22              | 18              |
| Resin lifetime, t, (cycle) | 100             | 100             | 100             |
| Resin price, rpc, (GBP L\(^{-1}\)) | 60 000          | 1 200           | 1 400           |
ing MINLP models with the same column sizing and key operational decisions. They are about $1 \times 10^{-4}$ lower than the optimal objective values of the MINLP models, being lower bounds as indicated by the theory of multiparametric disaggregation. The MINLP model A is able to achieve the optimum of 201.7 GBP $\cdot$ h$^{-1}$ within 494 s. After linearization, the MILP model A$^*$ takes a slightly shorter time (381 s) to find the optimal solution, despite having an increased number of equations and variables in the model. Meanwhile, as shown in Table 6, the MINLP model B has the same continuous variables as the MINLP model A, but significantly reduces the number of equations and discrete variables by one order of magnitude. As a result, the model is able to find the optimal solution within 47 s. The MINLP model B$^*$, which results from linearizing the MINLP model B, involves much fewer equations and variables than the MILP model A$^*$, and takes only 5 s to find the optimal solution, saving CPU by two orders of magnitude. From the comparison, it can be concluded that among the four proposed models, the models B and B$^*$ show obvious computational advantages over the other two models. In particular, the MILP model B$^*$ requires the smallest computational effort and has the most potential for larger scale problems, which also becomes evident from the larger example (Case 2) discussed in the next subsection. Note that the prediction errors of the resulting piecewise linear regression models have very minor effects on the optimal objective values of the optimization models; for example, a 17% shift of the estimated throughput output results in only a 0.1% difference in the optimal objective value. A similar observation is made for Case 2.

Next, the detailed optimal solutions of Case 1 are discussed. The optimal chromatography column-sizing strategies are given in Fig. 3, where the diameters of the chromatography columns are proportional to the widths of the shapes plotted, while the bed heights are proportional to the shapes’ heights. The dashed-line shapes represent the cycles needed in each batch. At each chromatography step, only one column is utilized. The column at the affinity chromatography step has a diameter of 180 cm and a bed height of 15 cm, while the cation-exchange chromatography step uses a smaller column with a diameter of 90 cm and a bed height of 21 cm. The chromatography column at the anion-exchange chromatography step has the smallest diameter, 80 cm, but the largest bed height, 25 cm.

Here, the performance of the throughput regression models and operational decisions are examined. In the optimal solution, the throughput output of the metamodel is 1869.8 g $\cdot$ h$^{-1}$ at the affinity chromatography step and 1823.9 g $\cdot$ h$^{-1}$ at the cation-exchange chromatography step. As shown in Fig. 3, five and four cycles per batch are required at these two chromatography steps, respectively. The resulting product masses loaded to each column in each cycle are 5306.7 and 6301.7 g, receptively. After converting to the values for a 1 m-diameter column, the loaded mass falls into the first interval in the piecewise linear regression model at the affinity chromatography step, and the corresponding function is used to estimate the throughput, as given below:

$$TP_{af} = 0.1914 \cdot LM_{af} + 1.8^2 \cdot (0.3570 \cdot V_{af} - 12.0477 \cdot H_{af} + 230.131)$$

where $1.8^2$ is added to convert the performance of the 1 m-diameter column to that of the selected 1.8 m-diameter column, and $LM_{af} = 1.8^2 \cdot LM_{af}^L$, according to Eq. (37). At the cation-exchange chromatography step, the first interval in the regression model is also selected. Similarly, the regression model used is as follows:

$$TP_{ca} = 0.1287 \cdot LM_{ca} + 0.9^2 \cdot (2.3940 \cdot V_{ca} - 51.4883 \cdot H_{ca} + 895.2814)$$

In both of the above functions, the linear velocity variable contributes to the throughput positively; therefore, the linear velocity of the flows at both steps reaches its upper bound, 600 cm $\cdot$ h$^{-1}$.

At last, from the optimal cost distribution shown in Fig. 4, the consumables cost—that is, the resin cost in this problem—represents the largest portion, over 30%, due to the high price of the affinity resin used. Also, the capital cost of the equipment, chemical reagents (buffer and media) cost, and labor cost all make a significant contribution to the total cost.

6.2. Case 2

In Case 2, a larger number of alternative diameter and bed height are considered, resulting in larger scale optimization models, as indicated in Table 7. Compared with Case 1, MINLP model B has slightly more discrete variables, while the other models involve an increased number of both equations and variables. It is worth noting that models A and A$^*$ fail to terminate before the computational time limit of 3600 s, although the gaps to the best possible

### Table 5
Parameters of non-chromatographic unit operations.

<table>
<thead>
<tr>
<th>Unit operation parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titer, titer (g $\cdot$ L$^{-1}$)</td>
<td>2</td>
</tr>
<tr>
<td>Bioreaction time, btr (d)</td>
<td>4</td>
</tr>
<tr>
<td>Seed train bioreaction time, st (d)</td>
<td>2</td>
</tr>
<tr>
<td>Number of bioreactors, bnv</td>
<td>1</td>
</tr>
<tr>
<td>Bioreactor working volume ratio, x</td>
<td>75%</td>
</tr>
<tr>
<td>Bioreactor volume, brv (L)</td>
<td>20 000</td>
</tr>
<tr>
<td>Centrifugation 1</td>
<td></td>
</tr>
<tr>
<td>Yield, yd,</td>
<td>95%</td>
</tr>
<tr>
<td>Buffer volume ratio, bvr,</td>
<td>0.9</td>
</tr>
<tr>
<td>Processing rate, pr, (L $\cdot$ h$^{-1}$)</td>
<td>250</td>
</tr>
<tr>
<td>Homogenization</td>
<td></td>
</tr>
<tr>
<td>Yield, yd,</td>
<td>100%</td>
</tr>
<tr>
<td>Processing rate, pr, (L $\cdot$ h$^{-1}$)</td>
<td>143</td>
</tr>
<tr>
<td>Centrifugation 2</td>
<td></td>
</tr>
<tr>
<td>Yield, yd,</td>
<td>95%</td>
</tr>
<tr>
<td>Buffer volume ratio, bvr,</td>
<td>0.1</td>
</tr>
<tr>
<td>Processing rate, pr, (L $\cdot$ h$^{-1}$)</td>
<td>250</td>
</tr>
<tr>
<td>Filtration</td>
<td></td>
</tr>
<tr>
<td>Yield, yd,</td>
<td>98%</td>
</tr>
<tr>
<td>Processing rate, pr, (L $\cdot$ h$^{-1}$)</td>
<td>125</td>
</tr>
<tr>
<td>UF/DF 1</td>
<td></td>
</tr>
<tr>
<td>Yield, yd,</td>
<td>98%</td>
</tr>
<tr>
<td>Flush volume ratio, fvr</td>
<td>6</td>
</tr>
<tr>
<td>Processing rate, pr, (L $\cdot$ h$^{-1}$)</td>
<td>40</td>
</tr>
<tr>
<td>UF/DF 2</td>
<td></td>
</tr>
<tr>
<td>Yield, yd,</td>
<td>98%</td>
</tr>
<tr>
<td>Defiltration volume ratio, dvf</td>
<td>7</td>
</tr>
<tr>
<td>Processing rate, pr, (L $\cdot$ h$^{-1}$)</td>
<td>40</td>
</tr>
<tr>
<td>Bulk fill</td>
<td></td>
</tr>
<tr>
<td>Yield, yd,</td>
<td>98%</td>
</tr>
<tr>
<td>Filling time, Tf (h)</td>
<td>6</td>
</tr>
<tr>
<td>Final concentration, fconc (g $\cdot$ L$^{-1}$)</td>
<td>10</td>
</tr>
</tbody>
</table>

### Table 6
Statistics and computational performance of the proposed models for Case 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of equations</th>
<th>No. of continuous variables</th>
<th>No. of discrete variables</th>
<th>Optimal objective (GBP $\cdot$ g$^{-1}$)</th>
<th>CPU (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MINLP A</td>
<td>423</td>
<td>85</td>
<td>670</td>
<td>201.7</td>
<td>494</td>
</tr>
<tr>
<td>MILP A$^*$</td>
<td>13 646</td>
<td>11 092</td>
<td>12 097</td>
<td>201.7</td>
<td>381</td>
</tr>
<tr>
<td>MINLP B</td>
<td>93</td>
<td>85</td>
<td>46</td>
<td>201.7</td>
<td>47</td>
</tr>
<tr>
<td>MILP B$^*$</td>
<td>1 508</td>
<td>700</td>
<td>595</td>
<td>201.7</td>
<td>5</td>
</tr>
</tbody>
</table>
solutions are only 0.6% and 0.4%, respectively. According to the solution process of the two models shown in Fig. 5, the MILP model A* finds a good feasible solution at around 220 s, and actually achieves the optimum within 10 min. However, the lower bound of the solution in the branch and bound process converges so slowly that the optimum of the obtained objective, 200.3 GBP$^{−1}$ cannot be proven within the given time limit. Meanwhile, the MINLP model B takes around 1 min for a close feasible solution and 192 s for the optimum. The MILP model B* achieves a CPU saving of one order of magnitude, with only 24 s for its optimum, which is also a very close lower bound of the MINLP model B* due to multiparametric disaggregation. For a good feasible solution within 1% of the optimum, only 6 s is needed for the model B*. Thus, the computational advantage of the MILP model B* is demonstrated.

Regarding the column-sizing decisions, Case 2 has more possible column size options. Fig. 6 shows that in comparison with the optimal decisions of Case 1, a column with a larger diameter but smaller bed height is installed at the first two chromatography steps. The selected bed heights (11 and 14 cm) are beyond the bed height range allowed in Case 1 (15–25 cm). In addition, the selected diameters (190 and 110 cm) are not available in Case 1. With more possible alternatives, the optimal solution of Case 1 is only a feasible solution of Case 2, and there is an improvement of 1.4 GBP$^{−1}$ in the optimal COG per gram of Case 2. Meanwhile, the same column (80 cm diameter and 25 cm bed height) is selected at the anion-exchange chromatography step. The selected larger diameter columns result in higher cost in equipment investment, while the smaller bed heights lead to less resin and relevant cost. The differences in these costs are relatively very small; therefore, the cost distribution is very similar to that of Case 1, which is not discussed further here.

At both the affinity and cation chromatography steps, the highest allowed flow velocity, 600 cm$^3$/C1$^{−1}$h$^{−1}$, is implemented. The actual throughput regression models at the two steps are slightly different from those in Case 1, due to the differences in the selected diameter sizes:

\[
TP_{af} = 0.1914 \cdot LM_{af} + 1.92 \cdot (0.3570 \cdot V_{af} \cdot 12.0477 \cdot H_{af} + 230.131)
\]

Table 7

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of equations</th>
<th>No. of continuous variables</th>
<th>No. of discrete variables</th>
<th>Optimal objective (GBP$^{−1}$)</th>
<th>CPU (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MINLP A</td>
<td>1 731</td>
<td>85</td>
<td>3 286</td>
<td>200.5$^a$</td>
<td>3600$^a$</td>
</tr>
<tr>
<td>MILP A*</td>
<td>63 350</td>
<td>52 948</td>
<td>59 185</td>
<td>200.3$^b$</td>
<td>3600$^b$</td>
</tr>
<tr>
<td>MINLP B</td>
<td>93</td>
<td>85</td>
<td>94</td>
<td>200.3</td>
<td>192</td>
</tr>
<tr>
<td>MILP B*</td>
<td>1 972</td>
<td>828</td>
<td>1 027</td>
<td>200.3</td>
<td>24</td>
</tr>
</tbody>
</table>

$^a$ Obtained solution has an optimality gap of 0.6% when the CPU limit is reached.

$^b$ Obtained solution has an optimality gap of 0.4% when the CPU limit is reached.
\[ TP_{ce} = 0.1287 \cdot LM_{ce} + 1.1^2 \]
\[ (2.3940 \cdot V_{ce} - 51.4883 \cdot H_{ce} + 895.2814) \quad (149) \]

As shown in Fig. 6, one more cycle is used at the affinity chromatography step than in Case 1, and therefore less mass (4422.3 g) is loaded in each cycle. However, due to the smaller bed height and larger diameter, the throughput increases to 1972.0 g h\(^{-1}\). For the cation-exchange chromatography step, although the same number of cycles and loaded mass as Case 1 are obtained, a higher throughput of 2760.2 g h\(^{-1}\) is achieved, due to the chosen larger diameter and smaller bed height.

7. Concluding remarks

In this work, the multiscale optimization of an antibody manufacturing process has been investigated. At the operational level, to mimic the complex behavior of the chromatography process, data-driven models were developed to estimate the chromatography throughput, using manufacturing-scale simulated datasets based on microscale experimental data. Through a comparison of a number of methods for metamodeling, piecewise linear regression models were developed based on the simulated datasets.

At the process design level, in order to determine the optimal chromatography column-sizing strategies, two alternative MINLP models were proposed to minimize the COG per gram. Adopting linearization techniques, two MILP models were developed. The throughput regression models were incorporated into the optimization models to determine the optimal operational decisions—that is, the flow velocity and the number of cycles per batch.

By studying two industrially relevant cases with different column size alternatives, the computational performance of the four proposed optimization models were compared. In conclusion, models B and B\(^*\) demonstrated more efficient computational performances. In particular, the second MILP model was shown to be the most computationally efficient, and can thus be recommended for large-scale optimization studies. In addition, optimal solution details were discussed, and the data-driven models were shown to work well to achieve optimal throughput.

A future research direction of this work could be the development of data-driven metamodels for anion-exchange chromatography, and for more chromatography parameters, such as yield and binding capacity, with more input variables, such as pH value and temperature. In addition, other performance criteria of the purification process, such as impurity removal capacity, could be taken into account for multi-objective optimization [9,21,22]. The uncertainty of the parameters, such as yield and titer, could also be considered [19,22]. Finally, single-use chromatography could be considered in the purification process as an important direction in smart biopharmaceutical manufacturing [48].

Acknowledgements

Funding from the UK Engineering & Physical Sciences Research Council (EP/I033270/1 and EP/M027856/1) is gratefully acknowledged. The authors would like to thank Dr. Spyridon Gerontas for providing data and useful discussions, and Dr. Lingjian Yang for implementing a preliminary analysis of regression models.
Compliance with ethics guidelines

Songsong Liu and Lazaros G. Papageorgiou declare that they have no conflicts of interest or financial conflicts to disclose.

Nomenclature

Indices
- \( d \): Position in multiparametric disaggregation = 1, ..., \( \text{maxp} \)
- \( i \): Column volume size
- \( j \): Column number = 1, ..., \( \text{maxCN}_i \)
- \( k \): Cycle number = 1, ..., \( \text{maxCYN}_s \)
- \( m \): Diameter size
- \( n \): Digit of the binary representation = 1, ..., \( \log_2(\text{maxBN}) \)
- \( q \): Integer number in multiparametric disaggregation
- \( r \): Interval in piecewise regression function
- \( s \): Downstream step = \( \text{ct}_1 \) (centrifugation 1), \( \text{ho} \) (homogenization), \( \text{ct}_2 \) (centrifugation 2), \( \text{fi} \) (filtration), \( \text{af} \) (affinity chromatography), \( \text{ce} \) (cation-exchange chromatography), \( \text{uf}_1 \) (UF/DF 1), \( \text{ae} \) (anion-exchange chromatography), \( \text{uf}_2 \) (UF/DF 2), \( \text{bf} \) (bulk fill)

Sets
- \( \text{CS} \): Set of chromatography steps = \{\( \text{af}, \text{ce}, \text{ae} \)\}

Parameters
- \( a, b, c \): Utilities cost coefficients
- \( \text{aot} \): Annual operating time, d
- \( \text{bcv}_s \): Buffer volume to column volume ratio at chromatography step \( s \)
- \( \text{bp}_{r,s} \): Breakpoint of loaded mass between intervals \( r \) and \( (r + 1) \) at chromatography step \( s \), g
- \( \text{bpc} \): Buffer price, GBP.L\(^{-1}\)
- \( \text{brc} \): Bioreactor cost, GBP
- \( \text{brn} \): Number of bioreactors
- \( \text{bri} \): Bioreactor time, d
- \( \text{brv} \): Bioreactor volume, L
- \( \text{brvr}_s \): Buffer volume ratio at centrifugation step \( s \)
- \( \text{cc}_{s,i} \): Column cost of size \( i \) at chromatography step \( s \), GBP
- \( \text{cc}_{s,m} \): Column cost of diameter size \( m \) at chromatography step \( s \), GBP
- \( \text{cv}_{s,i} \): Volume of column size \( i \) at chromatography step \( s \), L
- \( \text{dbc}_{s} \): Dynamic binding capacity at chromatography step \( s \), g.L\(^{-1}\)
- \( \text{dm}_{s,i} \): Diameter of column size \( i \) at chromatography step \( s \), cm
- \( \text{dm}_{s,m} \): Diameter of column size \( m \) at chromatography step \( s \), cm
- \( \text{don} \): Number of operators for downstream processing
- \( \text{dvr} \): Dialfiltration volume to column volume ratio at the second UF/DF step
- \( \text{ecv}_s \): Elute volume to column volume ratio at chromatography step \( s \)
- \( \text{el} \): Equipment lifetime, yeara
- \( \text{fconc} \): Final concentration of product, g.L\(^{-1}\)
- \( \text{fvr} \): Flush volume ratio at the first UF/DF step
- \( \text{gef} \): General equipment factor
- \( \text{gu} \): General utility unit cost, GBP.L\(^{-1}\)
- \( \text{hs}_{s,i} \): Bed height of column size \( i \) at chromatography step \( s \), cm
- \( \text{ir} \): Interest rate
- \( \text{i\lambda} \): Ratio of insurance cost to fixed capital investment
- \( \lambda \): Lifetime of resin at chromatography step \( s \), cycle
- \( \text{lang} \): Lang factor
- \( \text{maxBBV} \): Maximum buffer volume per batch, L
- \( \text{maxBN} \): Maximum number of batches
- \( \text{maxCN}_s \): Maximum number of columns at chromatography step \( s \)
- \( \text{maxCOG} \): Maximum COG per gram, GBP.L\(^{-1}\)
- \( \text{maxCYN}_s \): Maximum number of cycles at chromatography step \( s \)
- \( \text{maxH}_s \): Maximum column bed height size at chromatography step \( s \), cm
- \( \text{maxLM}_s \): Maximum product mass loaded at chromatography step \( s \), g
- \( \text{maxp} \): Maximum position in multiparametric disaggregation
- \( \text{maxT}_s \): Maximum processing time per batch at chromatography step \( s \), h
- \( \text{maxTCV}_s \): Maximum total column volume at chromatography step \( s \), L
- \( \text{maxTP}_{s} \): Maintenance throughput at chromatography step \( s \), g.L\(^{-1}\)
- \( \text{mae} \): Maintenance cost ratio to the fixed capital investment
- \( \text{mepc} \): Media price, GBP.L\(^{-1}\)
- \( \text{ml} \): Miscellaneous material cost ratio to chemical reagent and consumable costs
- \( \text{mcl} \): Management cost ratio to direct labor cost
- \( \text{oecl} \): Other equipment cost ratio to the bioreactor cost
- \( \text{of} \): Resin overpacking factor
- \( \text{pr}_{s} \): Processing rate of step \( s \), L.h\(^{-1}\)
- \( \text{q} \): Ratio of QCQA cost to direct labor cost
- \( \text{rpc}_{s} \): Resin price at chromatography step \( s \), GBP.L\(^{-1}\)
- \( \text{refCC} \): Reference cost of a chromatography column, GBP
- \( \text{refDM} \): Reference diameter of a chromatography column, cm
- \( \text{sfd} \): Duration per shift, h
- \( \text{sfh} \): Number of shifts per day, d\(^{-1}\)
- \( \text{std} \): Seed train bioreaction time, d
- \( \text{st} \): Supervisors cost ratio to direct labor cost
- \( \text{titer} \): Upstream product titer, g.L\(^{-1}\)
- \( \text{tx} \): Tax cost ratio to the fixed capital investment
- \( \text{uon} \): Number of operators per bioreactor in USP
- \( \text{uot} \): USP operating time per day
- \( \text{vel} \): Linear velocity of flow at the anion-exchange chromatography step, cm.h\(^{-1}\)
- \( \text{w} \): Wage of an operator, GBP.L\(^{-1}\)
- \( \text{yd}_{s} \): Product yield at step \( s \)
- \( \alpha \): Bioreactor working volume ratio
- \( \beta_{s}^{0} \): Constant coefficient in interval \( r \) at chromatography step \( s \)
- \( \beta_{s}^{H} \): Coefficient for bed height in interval \( r \) at chromatography step \( s \)
- \( \beta_{s}^{L} \): Coefficient for loaded mass in interval \( r \) at chromatography step \( s \)
- \( \beta_{s}^{V} \): Coefficient for velocity in interval \( r \) at chromatography step \( s \)
- \( \epsilon \): A small number to separate two successive intervals at the breakpoint \( s \)
- \( \theta \): Media overfill allowance
- \( \mu \): Chromatography resin utilization factor
- \( \sigma \): Batch success rate

Continuous variables
- \( \text{ABV} \): Annual buffer volume, L
- \( \text{AP} \): Annual product output, g
AT  Annual downstream processing time, d
BAT  Time for adding buffer per batch at anion-exchange chromatography step, h
BBV  Buffer volume added per batch, L
BC  Buffer cost, GBP
BRC  Bioreactor cost, GBP
BT  Downstream processing time per batch, d
BV  Buffer volume per batch in chromatography step s, L
CAC  Capital cost, GBP
CC  Consumables cost, GBP
COG  Annual cost of goods, GBP
CRC  Chemical reagents cost, GBP
CAPq  Continuous variable for annual production in multiple disaggregation at digit q
CTP  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  Mass loaded to single column at chromatography step s, g
LM1  Mass loaded to single 1 m-diameter column at chromatography step s, g
M0  Initial product mass entering downstream processes per batch, g
M1  Product mass per batch after step s, g
MAC  Maintenance cost, GBP
MC  Management cost, GBP
MEC  Media cost, GBP
MIC  Miscellaneous material cost, GBP
OIC  Other indirect costs, GBP
PLT  Time for loading product per batch at anion-exchange chromatography step, h
PV  Initial product volume entering downstream processes per batch, L
PV1  Product volume per batch after step s, L
QC  QCQA cost, GBP
RV  resin volume required at chromatography step s, L
SC  Supervisors cost, GBP
T  Processing time per batch of step s, h
TC  Tax cost, GBP
TCV  Total column volume at chromatography step s, L
TP  Throughput of single column at chromatography step s, g h⁻¹
TP1  Throughput of single 1 m-diameter column at chromatography step s, g h⁻¹
UC  Utilities cost, GBP
V  Linear velocity of flow at chromatography step s, cm h⁻¹
VFR  Volumetric flow rate at anion-exchange chromatography step, L h⁻¹

O_r  1 if the function at interval r is selected at chromatography step s; 0 otherwise
W  1 if there are j columns of size i at chromatography step s; 0 otherwise
X  1 if column size i is selected at chromatography step s; 0 otherwise
Y  1 if there are k cycles at chromatography step s; 0 otherwise
Z  1 if the nth digit of the binary representation of variable BN is equal to 1; 0 otherwise

Integer variables
BN  Number of completed batches
CN  Number of columns of size i at chromatography step s
ZN  Number of columns at chromatography step s
CYN  Number of cycles per batch at chromatography step s
H  Bed height of column at chromatography step s, cm

Auxiliary variables
CCOG  COG · BAPd
BTTP  T · BTP
CCOGAP  COG · CAP
CTTP  T · CTP
EF  Every cycle to process
EM  Every cycle to process
ET  Extra time
FTTP  T · TP
FTY  Fsj · Ysk
FTLM  Fsj · Ysk · LM
TP  T · TP
WT  Wsij · PLT
WTTP  Wsij · T · TP
WT  Wsij · Ysk
WTLM  Wsij · Ysk · LM
XLM  Xsj · LM
XTP  Xsj · TP
XT  Xsj · Ysk
XV  Xsk · TV
ZLM  Z · M
ZT  Z · BT
ZTV  Z · BBV
ZT  Z · TV

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2019.10.011.

References


