



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

使用数据驱动模型优化抗体纯化策略

刘松崧^{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

关键词

抗体纯化
多尺度优化
抗原结合片段
混合整数规划
数据驱动模型
分段线性回归

摘要

本工作致力于抗体片段纯化过程的多尺度优化。优化了生产过程中的色谱决策，包括色谱柱的数量及其大小，每批的循环数以及操作流速。使用基于微型实验数据的制造规模模拟数据集，建立了以负载质量、流速和柱床高度为输入的色谱通量数据驱动模型。与其他方法相比，分段线性回归建模方法具有简单、预测精度高的优点。提出了两种混合整数非线性规划（MINLP）模型，结合数据驱动模型，以最小化每克抗体纯化过程的总成本。然后，使用线性化技术和多参数分解将这些MINLP模型重新构造为混合整数线性规划（MILP）模型。研究了两个具有不同色谱柱尺寸替代品的工业相关案例，以证明所提出模型的适用性。

© 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. 引言

近年来，在传统制造业向智能制造业转型的过程中，全球产业一直在加速变化[1,2]。在转型过程中，产业面临着智能制造带来的诸多挑战，这引起了学术界和实务界的高度关注[3]，尤其是在流程产业中[4]。这项工作的一些挑战包括：

- 数据的使用和分析，特别着重于数据驱动的代理/元模型的开发，以简化复杂的流程并实现制造智能；
- 实施多尺度建模和优化，将战略和规划决策与运营相结合，以支持企业范围内的协调和优化；

• 开发计算效率高的模型、算法和工具，以找到智能制造决策的全局最优解决方案，并实现大规模优化。

在这项工作中，我们旨在基于简单的数据驱动模型为抗体产品的生产过程中的最佳纯化策略开发基于优化的决策模型，以应对生物制药行业中的上述挑战。为了更好地控制过程并提高生产效率，已使用不同的建模和解决方案技术研究了生物制药生产过程的优化问题，如元启发式[5]、动态优化[6]、进化算法[7–9]、马尔可夫决策方法[10]和混合整数规划[11–22]。数据驱动模型（也称为代理模型或元模型）是指基于数据构建的模型，但不依赖于有关过程或系统的理论知识。拥有复杂流程

* Corresponding author.

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

和系统的数据驱动模型使模型更为简单，提高了计算效率[23]，它们与优化的集成需要较少的计算工作量，在工程领域有着广泛的应用[24,25]。这些模型尤其在色谱纯化操作的建模和优化方面具有研究价值[26–29]。但是，将数据驱动模型集成到生物制药纯化过程的优化模型中的尝试很少。Nagrath等[30]开发了一种基于人工神经网络(ANN)的混合模型，用于优化制备色谱过程。Pirrung等[31]从详细的机械模型开发了人工神经网络，并将其整合到生物制药下游工艺的优化中，以获得具有三个不同色谱柱的工艺的最高产量。

抗原结合片段(antigen-binding fragment, Fab)产品被认为是继单克隆抗体(monoclonal antibody, mAb)产品之后的新一代蛋白质生物治疗药物，因其结构简单、体积小而具有诸多优势[32]。在工业实践中开发一种经济高效的Fab生产方法很有必要[33]，但是在这方面的文献很少。本文利用微尺度柱色谱法实验数据，对Fab产品的纯化工艺进行多尺度优化。为了实现最经济有效的过程，除了在设计层面考虑色谱柱分级策略外，还考虑了操作决策，特别是色谱分析步骤的流速。建立了数据驱动模型来估计色谱通量。结合所开发的数据驱动模型，提出了多种混合整数规划模型来寻找最优的Fab净化策略，并通过实例进行了分析。据我们所知，这是文献中首次使用数据驱动模型对Fab净化过程进行多尺度优化。

本文其余部分组织如下：第2节描述了优化问题；第3节建立了色谱操作的数据驱动模型；第4节描述了所提出的数学规划优化模型；第5节介绍了工业相关的案例研究；第6节是计算结果和讨论；结论见第7节。

2. 问题陈述

本文研究了Fab产品的生产工艺优化问题。图1为本文研究的Fab制造工艺流程图。首先，表达Fab的哺乳动物细胞在上游处理(USP)的生物反应器中培养，然

后进入下游处理(DSP)。在DSP中，Fab蛋白产品通过一系列操作进行纯化，包括离心、均质、过滤、超滤/渗滤(UF/DF)和三个柱床色谱分析步骤，包括亲和、阳离子交换和阴离子交换色谱分析步骤。

色谱柱上浆策略对整个纯化过程的效率至关重要。这些策略包括每个步骤的平行的柱数、柱的直径和床层高度以及每批的循环数，这些策略对整个制造过程的成本、时间和产出都有很大的影响。在实际应用中，有不同直径尺寸的标准柱，床身高度设置为典型的整数值范围。因此，在这项工作中，从一组给定的离散替代值中优化了色谱柱的直径和床高。

色谱操作是一个复杂的过程，在对其行为建模方面存在挑战。为了优化色谱操作策略，在这项工作中使用了元建模技术来模拟和预测色谱过程的性能，尤其是色谱通量，这表明了给定时间内一根色谱柱的产物输出速率，这是色谱分析的重要指标。为了开发制造规模的数据驱动模型，需要收集来自微型柱色谱实验室实验的数据，然后进行拟合以获得等温线参数。还建立了具有与比例有关的参数的第一原理色谱模型，以捕获缩放方面的挑战[34]。求解具有等温线参数的第一性原理模型，然后使用COMSOL Multiphysics仿真软件进行仿真运行以生成制造规模数据集[35]。数据集包括在结合和洗脱模式下的两个色谱分析步骤(即亲和色谱和阳离子交换色谱分析步骤)的不同输入条件下的负载质量，以及流速和柱床高度在不同输入条件下的通量输出。然后，将数据集用于导出数据驱动模型，将其合并到建议的优化模型中，以达到最佳的制造规模色谱分析决策。多尺度优化方法的整个过程如图2所示；图中最后三个褐色方框中显示的步骤将在下面的小节中详细描述。请注意，假设考虑的输入条件不影响其他色谱参数，如树脂的产量、结合能力和寿命，这些参数在这个问题中是已知的。

总而言之，本工作中考虑的优化问题描述如下。

考量量：

- Fab产品的工艺流程；

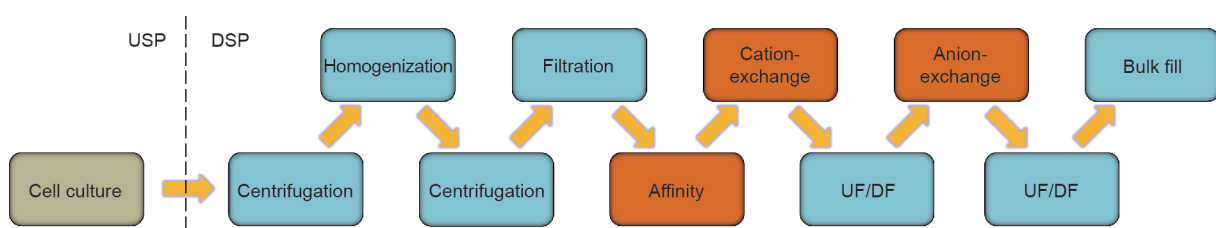


图1. Fab制造过程。橙色框代表色谱操作。USP: 上游处理; DSP: 下游处理; UF/DF: 超滤/渗滤。

- 生物反应器的数量和体积，以及生物反应器的滴定量；

- 色谱操作参数，包括产量、缓冲液和洗脱液用量、动态结合能力、使用寿命等；

- 非色谱操作参数，包括产量、处理速度、缓冲液用量等；

- 时间相关数据，包括处理速率、生物反应时间、年度作业时间等；

- 成本相关数据，包括劳动工资以及树脂、缓冲液、介质、设备等成本；

- 基于第一原理模型和微量柱实验数据的色谱模拟数据；

- 候选色谱柱直径和床层高度，以及最大循环数和色谱柱数。

确定量：

- 色谱柱分级策略，如柱的直径、床层高度和每一步色谱柱的数目；

- 操作策略，如线速度、负载产品质量、亲和色谱分析与阳离子交换色谱分析步骤的循环数；

- 已完成的批次总数；

- 年度总处理时间；

- 年度总产量；

- 年度总费用。

从而使每克Fab产品的商品成本（COG）（即年度

总成本与年度总产出之比）最小化。

3. 数据驱动模型

在本节中,我们开发了基于模拟数据集的色谱性能的简单数据驱动模型。在这项工作中,作为一个关键的性能标准,色谱吞吐率被认为是模型的输出。由于可用的数据有限,只对亲和色谱分析与阳离子交换色谱分析步骤进行了建模。亲和色谱分析与阳离子交换色谱分析步骤的数据集都是基于直径为1 m的单层色谱柱。在数据集中将影响色谱通量的三个关键变量作为数据驱动模型的输入,即负载质量、流速和柱床高度。

为了实现准确而简单的模型,实施了许多广泛使用的方法,包括线性回归、支持向量回归(SVR)[36]、克里金法[37]、步速回归[38]、响应面方法(RSM)[39]和分段线性回归[40]。为了估计这些方法的预测准确性,将执行交叉验证。给定一个数据集, n 次交叉验证将样本随机分成 n 个大小相等的子集。然后,在训练中使用样本的 $(n-1)$ 个子集,其余的集合用于验证所获得的数据驱动模型的预测准确性。在这项工作中,通过生成随机样本拆分执行10轮五折交叉验证,并将所有50个测试集的平均绝对误差(MAE)用作最终误差度量,以比较预测准确性。线性回归、SVR、克里金法和步速回归在WEKA机器学习软件[41]中以默认设置实现,而

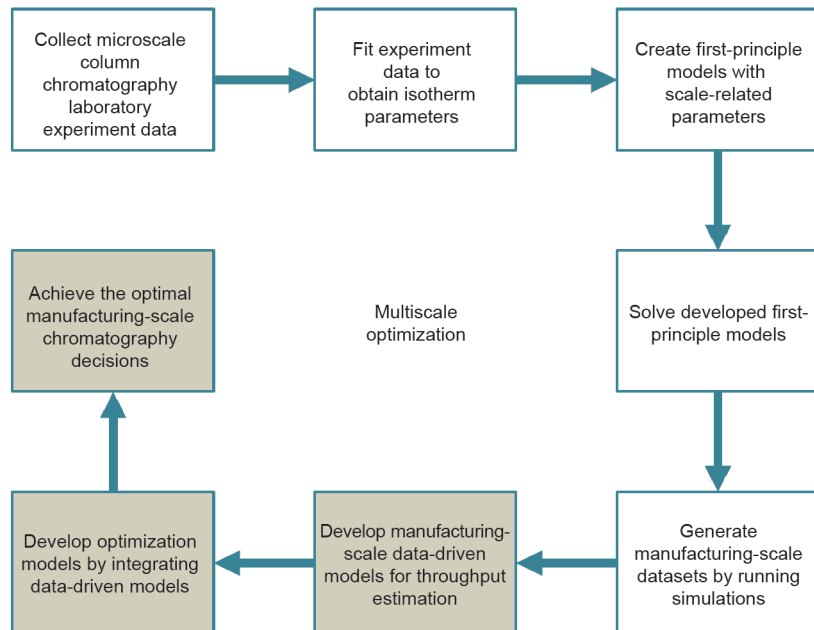


图2. 色谱策略多尺度优化的过程。

RSM和分段线性回归在GAMS [42]中使用CPLEX混合整数线性规划（MILP）求解器运行。表1列出了在对3081个样品的亲和色谱分析和2847个样品的阳离子交换色谱分析的数据集进行交叉验证后获得的预测误差结果。

表1显示，分段线性回归方法在所有测试方法中提供了最佳的预测精度。分段线性回归方法建立了一个模型，在一个输入变量上将样本分成多个互补区间，每个区间的灵活性由其自身的线性回归函数拟合。考虑到建模和理解的简便性，选择了分段线性回归方法以创建最终的数据驱动模型进行色谱通量估计，其中所有样本都用于训练过程中。在分段线性回归的过程中[40]，每个输入变量依次作为划分变量。对于每个分区变量，求解MILP模型确定分区变量在两个区间之间的断点，并保留训练误差最小的对应变量作为分区变量。在满足终止标准之前，将增加间隔数，并使用相同的分区变量迭代求解MILP模型。按照此步骤，在色谱分析步骤 s 下（af为亲和力和ce为阳离子交换），获得以下两个模型以估算通量 TP_s^1 ：

$$TP_{af}^1 = \begin{cases} 0.1914 \cdot LM_{af}^1 + 0.3570 \cdot V_{af} - 12.0477 \cdot H_{af} + 230.1318, & \text{if } 654 < LM_{af}^1 \leq 1643 \\ -0.003294 \cdot LM_{af}^1 + 0.008982 \cdot V_{af} + 43.5598 \cdot H_{af} - 649.3012, & \text{if } 1643 < LM_{af}^1 \leq 2629 \\ 0, & \text{if } 2629 < LM_{af}^1 \leq 7069 \end{cases} \quad (1)$$

$$TP_{ce}^1 = \begin{cases} 0.1287 \cdot LM_{ce}^1 + 2.3940 \cdot V_{ce} - 51.4883 \cdot H_{ce} + 895.2814, & \text{if } 3142 < LM_{ce}^1 \leq 12242 \\ -0.002489 \cdot LM_{ce}^1 + 0.05041 \cdot V_{ce} + 260.5745 \cdot H_{ce} - 3890.1058, & \text{if } 12242 < LM_{ce}^1 \leq 19698 \\ 0, & \text{if } 19698 < LM_{ce}^1 \leq 28274 \end{cases} \quad (2)$$

式中，变量的上标1表示直径为1 m的柱。在获得的模型中，色谱分析步骤 s 上的负载质量 LM_s^1 通过参考文献

[40]中给出的程序确定为分离三个区间。其预测误差小于其他两个变量 V_s 和 H_s （分别为线速度和柱床高度）。

为了将以上获得的数据驱动模型合并到用于决策的优化模型中，需要通过引入二进制变量来重新制定它们。如果色谱分析步骤 s 的负载质量在区间 r 内，则将二进制变量 $O_{s,r}$ 定义为等于1；然后从相应的班轮功能获得通量输出。由于只有一个区间可供选择，因此二进制变量应满足等式（3）：

$$\sum_r O_{s,r} = 1, \quad \forall s \in \{af, ce\} \quad (3)$$

在此模型中，单独输入 LM_s^1 的值应位于所选间隔的两个断点($bp_{s,r}$)之间，可以将其表示为以下线性方程式：

$$\sum_r bp_{s,r-1} \cdot O_{s,r} + \varepsilon \leq LM_s^1 \leq \sum_r bp_{s,r} \cdot O_{s,r}, \quad \forall s \in \{af, ce\} \quad (4)$$

式中， ε 是一个很小的数字，用于在断点处分隔两个连续的区间。通量输出的一般表达式如下：

$$TP_s^1 = \sum_r \left(\beta_{s,r}^{LM} \cdot LM_s^1 + \beta_{s,r}^V \cdot V_s + \beta_{s,r}^H \cdot H_s + \beta_{s,r}^0 \right) \cdot O_{s,r}, \quad \forall s \in \{af, ce\} \quad (5)$$

式中， $\beta_{s,r}^{LM}$ ， $\beta_{s,r}^V$ ， $\beta_{s,r}^H$ 和 $\beta_{s,r}^0$ 是函数中的参数。当在步骤 s 选择间隔 r' 时，即 $O_{s,r'} = 1$ ，等式（4）变为 $bp_{s,r'-1} < LM_s^1 \leq bp_{s,r'}$ 。在这种情况下，负载质量在区间 r' 中；然后等式（5）确保通量等于选定间隔 r' 上线性函数的输出：

$$TP_s^1 = \beta_{s,r'}^{LM} \cdot LM_s^1 + \beta_{s,r'}^V \cdot V_s + \beta_{s,r'}^H \cdot H_s + \beta_{s,r'}^0, \quad \text{if } bp_{s,r'-1} \leq LM_s^1 \leq bp_{s,r'} \quad (6)$$

下一部分将把上述通量回归模型合并到Fab制造过程的优化模型中。

表1 MAE不同方法的比较

Method	Affinity chromatography	Cation-exchange chromatography
Linear regression	121.40	777.26
SVR	111.31	700.98
Kriging	115.39	720.67
Pace regression	115.50	712.93
RSM	90.45	527.31
Piecewise linear regression	34.68	233.39

4. 优化模型

在本节中，我们将使用替代柱大小的不同建模方法，使用上述数据驱动模型来开发Fab产品纯化过程的两个混合整数非线性规划（MINLP）模型。然后，使用精确的线性化技术和多参数分解将这些MINLP模型重新构建为MILP模型。

4.1. MINLP 模型 A

MINLP模型A是根据参考资料[16]中提供的模型制定的，用于mAb制造过程。在此模型中，首先根据给定的离散色谱柱直径和床高的组合生成许多可选的色谱柱体积尺寸。在色谱分析步骤 s 上尺寸为 i 的柱体积（ $cv_{s,i}$ ）对应于给定替代方案中的特定直径尺寸（ $dm_{s,i}$ ）和床高（ $h_{s,i}$ ）。

4.1.1. 柱体积

色谱分析步骤 $s \in CS$ （ CS 是色谱分析步骤集合= $\{af, ce, ae\}$ ）的总柱体积（ TCV_s ）由选定大小的柱数（ $CN_{s,i}$ ）乘以相应的柱体积：

$$TCV_s = \sum_i cv_{s,i} \cdot CN_{s,i}, \forall s \in CS \quad (7)$$

通过引入二元变量 $X_{s,i}$ ，在色谱分析步骤 s 中选择柱尺寸 i ，通过以下约束条件可以保证只能选择一个柱尺寸：

$$\sum_i X_{s,i} = 1, \forall s \in CS \quad (8)$$

$$CN_{s,i} \leq \max CN_s \cdot X_{s,i}, \forall s \in CS, i \quad (9)$$

式中， $\max CN_s$ 表示允许的最大柱尺寸。

在每个批次中，色谱分析步骤 s 的可用树脂体积[即总柱体积（ TCV_s ）乘以每批次的循环数（ CYN_s ）]必须足以处理进入该步骤的所有蛋白质质量（ M_{s-1} ），所需的树脂体积（ RV_s ）由动态结合容量（ dbc_s ）和树脂利用率（ μ ）决定。

$$CYN_s \cdot TCV_s \geq RV_s, \forall s \in CS \quad (10)$$

$$RV_s = \frac{M_{s-1}}{dbc_s \cdot \mu}, \forall s \in CS \quad (11)$$

4.1.2. 产品质量

进入DSP过程的初始产品质量（ M_0 ）等于生物反应滴度（ $titer$ ）乘以生物反应器的工作体积，即生物反应器的体积（ brv ）乘以工作体积比（ α ）。

$$M_0 = titer \cdot \alpha \cdot brv \quad (12)$$

从步骤 s 出来的产物蛋白质质量等于上一个步骤 $s-1$ 的质量乘以步骤 s 相应的产量 yd_s 。

$$M_s = yd_s \cdot M_{s-1}, \forall s \quad (13)$$

年产品产量（ AP ）是批量填充步骤（ $s = bf$ ）后的产品质量：

$$AP = \sigma \cdot BN \cdot M_{bf} \quad (14)$$

式中， BN 是已完成的批次数，以最大允许的批次数为上限，而 σ 是批次成功率。

4.1.3. 产品数量

进入DSP（ PV_0 ）的初始产品量等于生物反应器的工作量，公式如下：

$$PV_0 = \alpha \cdot brv \quad (15)$$

对于该过程的前四个步骤，包括第一次离心（ $s = ct_1$ ）、均质化（ $s = ho$ ）、第二次离心（ $s = ct_2$ ）和过滤（ $s = fi$ ）步骤，在步骤 s 之后剩余的产品体积（ PV_s ）等于进入此步骤的产品体积：

$$PV_s = PV_{s-1}, \forall s \in \{ct_1, ho, ct_2, fi\} \quad (16)$$

在亲和力（ $s = af$ ）和阳离子交换色谱分析步骤（ $s = ce$ ）中，产物体积等于洗脱液体积，而在阴离子交换色谱分析步骤（ $s = ae$ ）中，产物体积不变。

$$PV_s = ecv_s \cdot CYN_s \cdot TCV_s|_{s \neq ae} + PV_{s-1}|_{s=ae}, \forall s \in CS \quad (17)$$

式中， ecv_s 是洗脱液体积与色谱柱体积的比率。

在第一个UF/DF步骤（ $s = uf_1$ ），冲洗量被添加到进入步骤的产品量中：

$$PV_{uf_1} = (fvr + 1) \cdot PV_{ce} \quad (18)$$

式中, fvr 是此步骤的冲洗体积比。在第二个UF/DF步骤 ($s = uf_2$), 剩余产品体积等于质量除以填充浓度 $fconc$:

$$PV_{uf_2} = \frac{M_{uf_2}}{fconc} \quad (19)$$

4.1.4. 缓冲液量

在每个步骤中添加的缓冲量 (BV_s) 定义如下:

$$BV_s = bvr_s \cdot PV_{s-1}, \forall s \in \{ct_1, ct_2\} \quad (20)$$

$$BV_s = 0, \forall s \in \{ho, fi\} \quad (21)$$

$$BV_s = bcv_s \cdot CYN_s \cdot TCV_s, \forall s \in CS \quad (22)$$

$$BV_{uf_1} = fvr \cdot PV_{ce} \quad (23)$$

$$BV_{uf_2} = dvr \cdot PV_{uf_2} \quad (24)$$

式中, bvr_s 是缓冲液体积比; bcv_s 是色谱分析步骤 s 的缓冲柱体积比; fvr 和 dvr 分别是第一和第二UF/DF步骤的冲洗体积比和渗滤体积比。

每个批次中所需的总缓冲量 (BBV) 是所有步骤中的缓冲量之和, 年度缓冲量 (ABV) 是所有已完成批次的总缓冲量。

$$BBV = \sum_s BV_s \quad (25)$$

$$ABV = BN \cdot BBV \quad (26)$$

4.1.5. 处理时间

亲和色谱和阳离子交换色谱分析步骤的处理时间 (T_s) 由每根色谱柱的质量输出除以其通量 (TP_s) 确定:

$$T_s = \frac{M_s}{TP_s \cdot \sum_i CN_{s,i}}, \forall s \in \{af, ce\} \quad (27)$$

在阴离子交换色谱分析步骤中, 使用体积流量 (VFR) 分别计算加载产物 (PLT) 和添加缓冲液 (BAT) 的处理时间, 以获得该步骤的处理时间。

$$T_{ae} = PLT + BAT \quad (28)$$

$$PLT = \frac{PV_{uf_1}}{VFR \cdot \sum_i CN_{ae,i}} \quad (29)$$

$$BAT = \frac{CYN_{ae} \cdot bcv_{ae} \cdot \sum_i cv_{ae,i} \cdot X_{ae,i}}{VFR} \quad (30)$$

$$VFR = \frac{1}{1000} \cdot vel \cdot \pi \cdot \sum_i \left(\frac{dm_{ae,i}}{2} \right)^2 \cdot X_{ae,i} \quad (31)$$

式中, vel 是阴离子交换色谱分析步骤的线性流速。假定批量填充步骤的处理时间是恒定的, 而在其他非色谱分析步骤, 处理时间等于相应的产品量除以处理速率 (pr_s):

$$T_s = \frac{PV_s}{pr_s}, \forall s \in \{ct_1, ho, ct_2, fi, uf_1, uf_2\} \quad (32)$$

一批的处理时间 (天) BT , 是所有步骤的总处理时间除以班次持续时间 (sfd) 和每天的班次数量 (sfn):

$$BT = \frac{\sum_s T_s}{sfd \cdot sfn} \quad (33)$$

年度处理时间 (AT) 是所有批次的总处理时间:

$$AT = BN \cdot BT \quad (34)$$

它受年度操作时间 (aot) 减去单批次种子培养链生物反应时间 (st) 和生物反应时间 (bri) ($aot - st - bri$) 的限制。

4.1.6. 数据驱动模型

直径为1 m的色谱柱的通量是根据上一节中获得的分段线性回归模型[包括式 (3) 和式 (4)]计算得出的。由于在色谱分析步骤 s 上选择的床高在该模型中表示为 $\sum_i h_{s,i} \cdot X_{s,i}$, 因此公式 (5) 修改如下:

$$TP_s^1 = \sum_r \left(\beta_{s,r}^{LM} \cdot LM_s^1 + \beta_{s,r}^V \cdot V_s + \beta_{s,r}^H \cdot \sum_i h_{s,i} \cdot X_{s,i} + \beta_{s,r}^0 \right) \cdot O_{s,r}, \forall s \in \{af, ce\} \quad (35)$$

考虑到色谱柱的通量可以视为蛋白质密度、线速度和柱面积的乘积, 因此可以认为通量与柱面积成正比。因此, 它也与柱直径的平方成正比。在这种情况下, 所选色谱柱的通量和直径为1 m的色谱柱的通量 (TP_s^1) 之间的关系公式如下:

$$TP_s = \frac{\sum_i (dm_{s,i})^2 \cdot X_{s,i}}{(\text{refDM})^2} \cdot TP_s^1, \forall s \in \{af, ce\} \quad (36)$$

其中, refDM 是指参考直径, 在此工作中为100 cm。

在数据驱动的回归模型中, 加载到直径为1 m的柱的质量用于按比例计算所选择柱(LM_s)的实际加载质量。

$$LM_s = \frac{\sum_i (dm_{s,i})^2 \cdot X_{s,i}}{(\text{refDM})^2} \cdot LM_s^1, \forall s \in \{af, ce\} \quad (37)$$

式中, 装载质量 LM_s 定义为每个循环中进入每一个柱的产品质量, 定义如下:

$$LM_s = \frac{M_{s-1}}{CYN_s \cdot \sum_i CN_{s,i}}, \forall s \in \{af, ce\} \quad (38)$$

4.1.7. 目标功能

在这项工作中, 目标是使每克的COG最小化, 这等于每年的总成本(AC)除以每年的总生产量(AP)。年度总成本计算中包括的所有成本条件和相关约束条件均在Appendix A中列出。目标函数如下:

$$\text{COG} = \frac{AC}{AP} \quad (39)$$

因此, 所提出的MINLP模型A使方程式(39)最小。受式(3)、式(4)、式(7)~(38)和Appendix A中式(S1)~(S19)的约束。

4.2. MILP 模型 A*

接下来, 为便于计算, 将所得的MINLP模型重新构造为MILP模型。新的线性约束如下所示。

4.2.1. 整型变量离散化

类似于参考文献[17-19]中的研究, 为了促进在其他约束条件下非线性项的线性化, 将整型变量 $CN_{s,i}$ 、 CYN_s 和 BN 离散化并用二进制变量表示, 如方程(40)~(44)所示。

$$CN_{s,i} = \sum_{j=1}^{\max CN_s} j \cdot W_{s,i,j}, \forall s \in CS, i \quad (40)$$

$$\sum_{j=1}^{\max CN_s} W_{s,i,j} = X_{s,i}, \forall s \in CS, i \quad (41)$$

$$CYN_s = \sum_{k=1}^{\max CYN_s} k \cdot Y_{s,k}, \forall s \in CS \quad (42)$$

$$\sum_{k=1}^{\max CYN_s} Y_{s,k} = 1, \forall s \in CS \quad (43)$$

$$BN = \sum_{n=1}^{\log_2 \max BN} 2^{n-1} \cdot Z_n \quad (44)$$

式中, $W_{s,i,j}$ 、 $Y_{s,k}$ 和 Z_n 是为离散化上述整型变量而引入的二进制变量, 其中, j 和 k 分别是柱数和循环数的指数。

4.2.2. 柱体积线性化

基于方程(42)和(43), 非线性约束方程(10)可以重新表示为:

$$\sum_{k=1}^{\max CYN_s} k \cdot \overline{YV}_{s,k} \geq RV_s, \forall s \in CS \quad (45)$$

$$\overline{YV}_{s,k} \leq \max TCV_s \cdot Y_{s,k}, \forall s \in CS, k = 1, \dots, \max CYN_s \quad (46)$$

$$\sum_{k=1}^{\max CYN_s} \overline{YV}_{s,k} = TCV_s, \forall s \in CS \quad (47)$$

此处引入和定义了辅助变量, $\overline{YV}_{s,k} = Y_{s,k} \cdot TCV_s$, 此外, $\max TCV_s$ 是色谱分析步骤 s 的最大色谱柱体积。

4.2.3. 年产量线性化

通过引入另一个辅助变量, $\overline{ZM}_{s,n} \equiv Z_n \cdot M_s$, 以表示方程(14)中的双线性项。可以通过以下方程[43]来重新表达该约束:

$$AP = \sum_{n=1}^{\lceil \log_2 \max BN \rceil} \sigma \cdot 2^{n-1} \cdot \overline{ZM}_{bf,n} \quad (48)$$

$$\overline{ZM}_{bf,n} \leq \text{titer} \cdot \alpha \cdot brv \cdot Z_n, \forall n = 1, \dots, \lceil \log_2 \max BN \rceil \quad (49)$$

$$\overline{ZM}_{bf,n} \leq M_{bf}, \forall n = 1, \dots, \lceil \log_2 \max BN \rceil \quad (50)$$

$$\overline{ZM}_{bf,n} \geq M_{bf} - \text{titer} \cdot \alpha \cdot brv \cdot (1 - Z_n), \forall n = 1, \dots, \lceil \log_2 \max BN \rceil \quad (51)$$

4.2.4. 产品和缓冲液体积线性化

根据方程(46)和(47), 方程(17)和(22)可以分别重写为以下两个线性方程:

$$PV_s = ecv_s \cdot \sum_{k=1}^{\max CYN_s} k \cdot \overline{YV}_{s,k} \Big|_{s \neq ae} + PV_{uf1} \Big|_{s=ae}, \forall s \in CS \quad (52)$$

$$BV_s = bcv_s \cdot \sum_{k=1}^{\max CYN_s} k \cdot \overline{YV}_{s,k}, \forall s \in CS \quad (53)$$

$\overline{ZV}_n \equiv Z_n \cdot BBV$ 用下列方程进行定义:

$$\overline{ZV}_n \leq \max BBV \cdot Z_n, \forall n = 1, \dots, \lceil \log_2 \max BN \rceil \quad (54)$$

$$\overline{ZV}_n \leq BBV, \forall n = 1, \dots, \lceil \log_2 \max BN \rceil \quad (55)$$

$$\overline{ZV}_n \geq BBV - \max BBV \cdot (1 - Z_n), \quad (56)$$

$$\forall n = 1, \dots, \lceil \log_2 \max BN \rceil$$

因此, 方程 (26) 可以重写如下:

$$ABV = \sum_{n=1}^{\lceil \log_2 \max BN \rceil} 2^{n-1} \cdot \overline{ZV}_n \quad (57)$$

4.2.5. 处理时间线性化

方程 (27) 可以重新构造为包含一个整型变量 $CN_{s,i}$ 及两个连续变量 T_s 和 TP_s 乘积的非线性项。引入了两个辅助变量: $\overline{TTP}_s \equiv T_s \cdot TP_s$ 和 $\overline{WTTTP}_{s,i,j} \equiv W_{s,i,j} \cdot T_s \cdot TP_s$ 。 $\overline{WTTTP}_{s,i,j}$ 可以由 \overline{TTP}_s 通过以下方程确定:

$$\sum_i \sum_{j=1}^{\max CN_s} j \cdot \overline{WTTTP}_{s,i,j} = M_s, \forall s \in \{af, ce\} \quad (58)$$

$$\overline{WTTTP}_{s,i,j} \leq \max TP_s \cdot \max T_s \cdot W_{s,i,j}, \quad (59)$$

$$\forall s \in \{af, ce\}, i, j = 1, \dots, \max CN_s$$

$$\sum_i \sum_{j=1}^{\max CYN_s} j \cdot \overline{WTTTP}_{s,i,j} = \overline{TTP}_s, \forall s \in \{af, ce\} \quad (60)$$

接下来, 使用多参数分解重新定义 \overline{TTP}_s [44–46], 其中, 通量 TP_s 表示为由二进制变量 $BTP_{s,d,q}$ 和连续变量 $CTP_s \in [0,1]$ 确定的有效十进制幂的多参数总和, 其中, d 是从 p 到 $\max p$ 的数位, q 是幂 d 的数字。

$$TP_s = \sum_{d=p}^{\max p} \sum_{q=0}^9 10^d \cdot q \cdot BTP_{s,d,q} + \sum_{q=0}^1 10^p \cdot q \cdot CTP_{s,q}, \quad (61)$$

$\forall s \in \{af, ce\}$

$$\sum_{q=0}^9 BTP_{s,d,q} = 1, \forall s \in \{af, ce\}, d = p, \dots, \max p \quad (62)$$

$$\sum_{q=0}^1 CTP_{s,q} = 1, \forall s \in \{af, ce\} \quad (63)$$

基于以上方程, 变量 \overline{TTP}_s 被分解为多个非负连续变

量 $\overline{BTTP}_{s,d,q}$ 和 $\overline{CTTP}_{s,q}$:

$$\overline{TTP}_s = \sum_{d=p}^{\max p} \sum_{q=0}^9 10^d \cdot q \cdot \overline{BTTP}_{s,d,q} + \sum_{q=0}^1 10^p \cdot q \cdot \overline{CTTP}_{s,q}, \quad (64)$$

$\forall s \in \{af, ce\}$

$$\overline{BTTP}_{s,d,q} \leq \max T_s \cdot BTP_{s,d,q}, \quad (65)$$

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

$$\overline{CTTP}_{s,q} \leq \max T_s \cdot CTP_{s,q}, \forall s \in \{af, ce\}, q = 0, 1 \quad (66)$$

$$\sum_{q=0}^9 \overline{BTTP}_{s,d,q} = T_s, \forall s \in \{af, ce\}, d = p, \dots, \max p \quad (67)$$

$$\sum_{q=0}^1 \overline{CTTP}_{s,q} = T_s, \forall s \in \{af, ce\} \quad (68)$$

需要注意的是, 多参数分解是一种松弛法, 上述重构能给出原始方程式的近似值, 并生成了最优值的下界。

通过使用下述约束条件定义 $\overline{WT}_{s,i,j} \equiv W_{s,i,j} \cdot PLT$:

$$\overline{WT}_{ae,i,j} \leq \max T_{ae} \cdot W_{ae,i,j}, \forall i, j = 1, \dots, \max CN_{ae} \quad (69)$$

$$\sum_i \sum_{j=1}^{\max CN_{ae}} \overline{WT}_{ae,i,j} = PLT \quad (70)$$

以下线性约束条件取代了方程 (29):

$$\frac{1}{1000} \cdot \sum_i \sum_{j=1}^{\max CN_{ae}} j \cdot vel \cdot \pi \cdot \left(\frac{dm_{ae,i}}{2} \right)^2 \cdot \overline{WT}_{ae,i,j} = PV_{uf_1} \quad (71)$$

使用另一个辅助变量, $\overline{XY}_{s,i,k} \equiv X_{s,i} \cdot Y_{s,k}$, 下述约束条件等效于方程 (30):

$$BAT = \sum_i \sum_{k=1}^{\max CYN_{ae}} \frac{bcv_{ae} \cdot CV_{ae,i} \cdot k \cdot \overline{XY}_{ae,i,k}}{(1/1000) \cdot vel \cdot \pi \cdot (dm_{ae,i}/2)^2} \quad (72)$$

$$\sum_i \overline{XY}_{ae,i,k} = Y_{ae,k}, \forall k = 1, \dots, \max CYN_{ae} \quad (73)$$

$$\sum_{k=1}^{\max CYN_{ae}} \overline{XY}_{ae,i,k} = X_{ae,i}, \forall i \quad (74)$$

基于方程 (44), 方程 (34) 可以重写如下:

$$AT = \sum_{n=1}^{\lceil \log_2 \max BN \rceil} 2^{n-1} \cdot \overline{ZT}_n \quad (75)$$

式中, $\overline{ZT}_n \equiv Z_n \cdot BT$, 满足:

$$\overline{ZT}_n \leq (aot - st - brt) \cdot Z_n, \forall n = 1, \dots, \lceil \log_2 \max BN \rceil \quad (76)$$

$$\overline{ZT}_n \leq BT, \forall n = 1, \dots, \lceil \log_2 \max BN \rceil \quad (77)$$

$$\overline{ZT}_n \geq BT - (aot - st - brt) \cdot (1 - Z_n), \quad (78)$$

$$\forall n = 1, \dots, \lceil \log_2 \max BN \rceil$$

4.2.6. 数据驱动的模式重构

方程 (4) 和 (35) 中回归模型的双线性项可以重写为以下线性约束[47]:

$$bp_{s,r-1} + \varepsilon - \max LM_s \cdot (1 - O_{s,r}) \leq LM_s^1 \quad (79)$$

$$\leq bp_{s,r} + \max LM_s \cdot (1 - O_{s,r}), \forall s \in \{af, ce\}, r$$

$$\beta_{s,r}^{LM} \cdot LM_s^1 + \beta_{s,r}^V \cdot V_s + \beta_{s,r}^H \cdot \sum_i h_{s,i} \cdot X_{s,i}$$

$$+ \beta_{s,r}^0 - \max TP_s \cdot (1 - O_{s,r}) \leq TP_s^1 \leq \beta_{s,r}^{LM} \cdot LM_s^1$$

$$+ \beta_{s,r}^V \cdot V_s + \beta_{s,r}^H \cdot \sum_i h_{s,i} \cdot X_{s,i} + \beta_{s,r}^0 \quad (80)$$

$$+ \max TP_s \cdot (1 - O_{s,r}),$$

$$\forall s \in \{af, ce\}, r$$

方程 (36) 可以使用新的辅助变量 $\overline{XTP}_{s,i} \equiv X_{s,i} \cdot TP_s^1$ 和以下约束条件进行线性化:

$$TP_s = \frac{\sum_i (dm_{s,i})^2 \cdot \overline{XTP}_{s,i}}{(\text{refDM})^2}, \forall s \in \{af, ce\} \quad (81)$$

$$\overline{XTP}_{s,i} \leq \max TP_s \cdot X_{s,i}, \forall s \in \{af, ce\}, i \quad (82)$$

$$\sum_i \overline{XTP}_{s,i} = TP_s^1, \forall s \in \{af, ce\} \quad (83)$$

类似地, 使用辅助变量 $\overline{XLM}_{s,i} \equiv X_{s,i} \cdot LM_s^1$, 方程 (37) 等效于:

$$LM_s = \frac{\sum_i (dm_{s,i})^2 \cdot \overline{XLM}_{s,i}}{(\text{refDM})^2}, \forall s \in \{af, ce\} \quad (84)$$

$$\overline{XLM}_{s,i} \leq \max LM_s \cdot X_{s,i}, \forall s \in \{af, ce\}, i \quad (85)$$

$$\sum_i \overline{XLM}_{s,i} = LM_s^1, \forall s \in \{af, ce\} \quad (86)$$

方程 (38) 包含一个非线性项, 它涉及两个整型变量和一个连续变量, 可以将其线性化, 如下所示:

$$\sum_{j=1}^{\max CN_s} \sum_{k=1}^{\max CYN_s} j \cdot k \cdot \overline{WYLM}_{s,i,j,k} = M_{s-1}, \forall s \in \{af, ce\}, i \quad (87)$$

$$\sum_{k=1}^{\max CYN_s} \overline{WY}_{s,i,j,k} = W_{s,i,j}, \forall s \in \{af, ce\}, i, j = 1, \dots, \max CN_s \quad (88)$$

$$\sum_i \sum_{j=1}^{\max CN_s} \overline{WY}_{s,i,j,k} = Y_{s,k}, \forall s \in \{af, ce\}, k = 1, \dots, \max CYN_s \quad (89)$$

$$\overline{WYLM}_{s,i,j,k} \leq \max LM_s \cdot \overline{WY}_{s,i,j,k},$$

$$\forall s \in \{af, ce\}, i, j = 1, \dots, \max CN_s, k = 1, \dots, \max CYN_s \quad (90)$$

$$\sum_i \sum_{j=1}^{\max CN_s} \sum_{k=1}^{\max CYN_s} \overline{WYLM}_{s,i,j,k} = LM_s, \forall s \in \{af, ce\} \quad (91)$$

式中, 有两个辅助变量 $\overline{WY}_{s,i,j,k} \equiv W_{s,i,j} \cdot Y_{s,k}$ 和 $\overline{WYLM}_{s,i,j,k} \equiv W_{s,i,j} \cdot Y_{s,k} \cdot LM_s$.

4.2.7. 目标函数线性化

耗材成本的计算方程 (S9) (见 Appendix A) 涉及非线性。通过引入辅助变量 $\overline{ZYV}_{s,k,n} \equiv Z_n \cdot \overline{YV}_{s,k}$, 方程 (S9) 可以重新编写为:

$$CC = \sum_{s \in CS} \sum_{n=1}^{\log_2 \max BN} \sum_{k=1}^{\max CYN_s} \frac{2^{n-1} \cdot k \cdot of \cdot rpc_s \cdot \overline{ZYV}_{s,k,n}}{l_s} \quad (92)$$

式中, of 为数脂的过填充系数; rpc_s 和 l_s 分别为色谱步骤 s 的树脂价格和树脂寿命。

$$\overline{ZYV}_{s,k,n} \leq \max TCV_s \cdot Z_n, \forall s \in CS, k = 1, \dots, \max CYN_s, \quad (93)$$

$$n = 1, \dots, \lceil \log_2 \max BN \rceil$$

$$\overline{ZYV}_{s,k,n} \leq \overline{YV}_{s,k}, \forall s \in CS, k = 1, \dots, \max CYN_s, \quad (94)$$

$$n = 1, \dots, \lceil \log_2 \max BN \rceil$$

$$\overline{ZYV}_{s,k,n} \geq \overline{YV}_{s,k} - \max TCV_s \cdot (1 - Z_n), \forall s \in CS, \quad (95)$$

$$k = 1, \dots, \max CYN_s, n = 1, \dots, \lceil \log_2 \max BN \rceil$$

在目标函数方程 (39) 中, 年成本可以表示为每克 COG 与年产量的乘积, 可以写成:

$$\overline{COGAP} = AC \quad (96)$$

式中, 辅助变量 $\overline{COGAP} \equiv COG \cdot AP$ 。使用多参数分解并引入新的辅助变量 $\overline{BCOGAP}_{d,q} \equiv COG \cdot \overline{BAP}_{d,q}$ 和 $\overline{CCOGAP}_q \equiv COG \cdot \overline{CAP}_q$, 可以将变量 AP 和 \overline{COGAP} 分解如下:

$$AP = \sum_{d=p}^{\max p} \sum_{q=0}^9 10^d \cdot q \cdot BAP_{d,q} + \sum_{q=0}^1 10^p \cdot q \cdot CAP_q \quad (97)$$

$$\overline{COGAP} = \sum_{d=p}^{\max p} \sum_{q=0}^9 10^d \cdot q \cdot \overline{BCOGAP}_{d,q} + \sum_{q=0}^1 10^p \cdot q \cdot \overline{CCOGAP}_q \quad (98)$$

$$\sum_{q=0}^9 BAP_{d,q} = 1, \forall d = p, \dots, \max p \quad (99)$$

$$\sum_{q=0}^1 CAP_q = 1 \quad (100)$$

$$\overline{BCOGAP}_{d,q} \leq \max COG \cdot BAP_{d,q}, \forall d = p, \dots, \max p, q = 0, \dots, 9 \quad (101)$$

$$\overline{CCOGAP}_q \leq \max COG \cdot CAP_q, \forall q = 0, 1 \quad (102)$$

$$\sum_{q=0}^9 \overline{BCOGAP}_{d,q} = COG, \forall d = p, \dots, \max p \quad (103)$$

$$\sum_{q=0}^1 \overline{CCOGAP}_q = COG \quad (104)$$

因此，重新构建的MILP模型A*包括方程(3)、(7)~(9)、(11)~(13)、(15)、(16)、(18)~(21)、(23)~(25)、(28)、(32)、(33)、(40)~(104)、(S1)~(S8)和(S10)~(S19)中所示的约束条件，并以变量COG为目标。

4.3. MINLP 模型 B

在本节中介绍了可作为选择的MINLP模型B。通过生成一系列柱体积尺寸，模型A和A*得到了大量变量和方程，这有碍于对所提出模型的计算。为了克服这一缺点，在新的MINLP模型B中，我们不使用离散的色谱柱体积大小，并为床高引入了一个整型变量 H_s ，以及一个二进制变量 $E_{s,m}$ 来表示直径 \widetilde{dm}_m 是否被选中。此外，从色谱柱编号变量中删除下标 i ，变量 \widetilde{CN}_s 表示色谱分析步骤 s 的柱数，其上限为 $\max CN_s$ 。因此，模型B中减少了离散变量的数量，从而提高了计算效率。基于上面提出的MINLP模型A，为MINLP模型B的开发提供了许多新变量和约束条件，如下所述。

4.3.1. 柱体积

在模型B中，通过床高(H_s)、直径选择($E_{s,m}$)和柱数(\widetilde{CN}_s)这些变量来计算总柱体积。

$$TCV_s = \frac{1}{1000} \cdot \pi \cdot \sum_m \left(\frac{\widetilde{dm}_m}{2} \right)^2 \cdot E_{s,m} \cdot H_s \cdot \widetilde{CN}_s, \forall s \in CS \quad (105)$$

此外，每个色谱分析步骤只能选择一个直径大小。

$$\sum_m E_{s,m} = 1, \forall s \in CS \quad (106)$$

4.3.2. 处理时间

通过用方程(27)和(29)中的 \widetilde{CN}_s 替换 $\sum_i CN_{s,i}$ ，可以得到以下约束：

$$T_s \cdot TP_s \cdot \widetilde{CN}_s = M_s, \forall s \in \{af, ce\} \quad (107)$$

$$PLT = \frac{PV_{uf_1}}{VFR \cdot \widetilde{CN}_{ae}} \quad (108)$$

同样地，使用新的整型变量 H_s 和二进制变量 $E_{s,m}$ 来表示BAT和VFR：

$$BAT = \frac{CYN_{ae} \cdot bc v_{ae} \cdot (1/1000) \cdot \pi \cdot \sum_m \left(\widetilde{dm}_m / 2 \right)^2 \cdot E_{ae,m} \cdot H_{ae}}{VFR} \quad (109)$$

$$VFR = \frac{1}{1000} \cdot vel \cdot \pi \cdot \sum_m \left(\frac{\widetilde{dm}_m}{2} \right)^2 \cdot E_{ae,m} \quad (110)$$

4.3.3. 数据驱动模型

方程(38)可以使用新的柱数变量编写，如下所示：

$$LM_s = \frac{M_{s-1}}{CYN_s \cdot \widetilde{CN}_s}, \forall s \in \{af, ce\} \quad (111)$$

与方程(36)和(37)类似，可以从1 m直径的柱计算出所选柱的通量和装载质量。

$$TP_s = \frac{\sum_m \left(\widetilde{dm}_m \right)^2 \cdot E_{s,m}}{(\text{refDM})^2} \cdot TP_s^1, \forall s \in \{af, ce\} \quad (112)$$

$$LM_s = \frac{\sum_m \left(\widetilde{dm}_m \right)^2 \cdot E_{s,m}}{(\text{refDM})^2} \cdot LM_s^1, \forall s \in \{af, ce\} \quad (113)$$

4.3.4. 成本

在成本计算中, 方程(S13)对于固定资产投资(FCI)的补充数据(见Appendix A)应替换为以下非线性约束:

$$FCI = lang \cdot (1 + gef) \cdot \left(brc \cdot brn + \sum_{s \in CS} \sum_m \tilde{c}c_{s,m} \cdot \tilde{CN}_s \cdot E_{s,m} + oe\lambda \cdot brc \cdot brn \right) \quad (114)$$

式中, $lang$ 是Lang因子; gef 是一般设备系数; brn 是生物反应器数量; brc 是一种生物反应器的成本; $oe\lambda$ 是其他设备成本与生物反应器成本之比; $\tilde{c}c_{s,m}$ 是色谱分析步骤 s 中直径为 m 的色谱柱成本。

由此可见, 所提出的MINLP模型B使方程(39)最简化, 但要满足方程(3)~(5)、(10)~(26)、(28)、(32)~(34)、(105)~(114)、(S1)~(S12)和(S14)~(S19)中所示的约束条件。

4.4. MILP 模型 B*

在MILP模型B*中, 将MINLP模型B的所有非线性约束线性化。除了在MILP模型A*中显示的线性约束外, 下面还给出了新开发的约束条件。

4.4.1. 整型变量离散化

首先, 使用二进制变量 $F_{s,j}$ 将整型变量 \tilde{CN}_s 离散化, $F_{s,j}$ 表示在色谱分析步骤 s 是否使用 j 柱, 如下所示:

$$\tilde{CN}_s = \sum_{j=1}^{\max CN_s} j \cdot F_{s,j}, \quad \forall s \in \{af, ce\} \quad (115)$$

$$\sum_{j=1}^{\max CN_s} F_{s,j} = 1, \quad \forall s \in \{af, ce\} \quad (116)$$

4.4.2. 柱体积线性化

为了线性化方程(105), 引入了一个辅助变量 $\overline{EFH}_{s,m,j} \equiv E_{s,m} \cdot F_{s,j} \cdot H_s$, 且有以下约束条件:

$$\overline{EFH}_{s,m,j} \leq \max H_s \cdot \overline{EF}_{s,m,j}, \quad \forall s \in CS, m, j = 1, \dots, \max CN_s \quad (117)$$

$$\sum_m \sum_{j=1}^{\max CN_s} \overline{EFH}_{s,m,j} = H_s, \quad \forall s \in CS \quad (118)$$

式中, $\overline{EF}_{s,m,j} \equiv E_{s,m} \cdot F_{s,j}$ 定义如下:

$$\sum_m \overline{EF}_{s,m,j} = F_{s,j}, \quad \forall s \in \{af, ce\}, j = 1, \dots, \max CN_s \quad (119)$$

$$\sum_{j=1}^{\max CN_s} \overline{EF}_{s,m,j} = E_{s,m}, \quad \forall s \in \{af, ce\}, m \quad (120)$$

因此, 将方程(105)重新构造为以下线性约束:

$$TCV_s = \sum_m \sum_{j=1}^{\max CN_s} j \cdot \pi \cdot \left(\frac{\tilde{d}m_m}{2} \right)^2 \cdot \overline{EFH}_{s,m,j}, \quad \forall s \in \{af, ce\} \quad (121)$$

4.4.3. 处理时间线性化

根据方程(61)~(68), $T_s \cdot TP_s$ 这一项可以由 \overline{TTP}_s 表示。因而, 引入 $\overline{FTTP}_{s,j}$ 来表达 $F_{s,j} \cdot \overline{TTP}_s$, 但具有以下约束:

$$\overline{FTTP}_{s,j} \leq \max TP_s \cdot \max T_s \cdot F_{s,j}, \quad \forall s \in \{af, ce\}, j = 1, \dots, \max CN_s \quad (122)$$

$$\sum_{j=1}^{\max CN_s} j \cdot \overline{FTTP}_{s,j} = \overline{TTP}_s, \quad \forall s \in \{af, ce\} \quad (123)$$

因此, 方程(107)重写如下:

$$\sum_{j=1}^{\max CN_s} j \cdot \overline{FTTP}_{s,j} = M_s, \quad \forall s \in \{af, ce\} \quad (124)$$

基于引入的辅助变量 $\overline{EFH}_{s,m,j} \equiv \overline{EF}_{s,m,j} \cdot PLT$ 及方程(119)和(120), 方程(108)可以如下线性化:

$$\frac{1}{1000} \cdot \sum_m \sum_{j=1}^{\max CN_{ae}} j \cdot vel \cdot \pi \cdot \left(\frac{\tilde{d}m_m}{2} \right)^2 \cdot \overline{EFT}_{ae,m,j} = PV_{uf_1} \quad (125)$$

$$\overline{EFT}_{ae,m,j} \leq \max T_{ae} \cdot \overline{EF}_{ae,m,j}, \quad \forall m, j = 1, \dots, \max CN_{ae} \quad (126)$$

$$\sum_m \sum_{j=1}^{\max CN_{ae}} \overline{EFT}_{ae,m,j} = PLT \quad (127)$$

此外, 通过 $\overline{YH}_{s,k} \equiv Y_{s,k} \cdot H_s$, 以下约束条件可以在MILP模型中代替方程(109):

$$BAT = \sum_{k=1}^{\max CYN_{ae}} \frac{bc v_{ae} \cdot k \cdot \overline{YH}_{ae,k}}{vel} \quad (128)$$

$$\overline{YH}_{ae,k} \leq \max H_{ae} \cdot Y_{ae,k}, \quad \forall k = 1, \dots, \max CYN_s \quad (129)$$

$$\sum_{k=1}^{\max CYN_s} \overline{YH}_{ae,k} = H_{ae} \quad (130)$$

4.4.4. 数据驱动模型的重构

类似于方程 (80), 方程 (5) 中给出的数据驱动模型约束可以改写为线性约束[47], 如下所示:

$$\begin{aligned} & \beta_{s,r}^{LM} \cdot LM_s^1 + \beta_{s,r}^V \cdot V_s + \beta_{s,r}^H \cdot H_s + \beta_{s,r}^0 - \max TP_s \cdot (1 - O_{s,r}) \\ & \leq TP_s^1 \leq \beta_{s,r}^{LM} \cdot LM_s^1 + \beta_{s,r}^V \cdot V_s + \beta_{s,r}^H \cdot H_s + \beta_{s,r}^0 + \\ & \max TP_s \cdot (1 - O_{s,r}), \forall s \in \{af, ce\}, r \end{aligned} \quad (131)$$

通过引入辅助变量 $\overline{FYLM}_{s,j,k} \equiv F_{s,j} \cdot Y_{s,k} \cdot LM_s$ 和 $\overline{FY}_{s,j,k} \equiv F_{s,j} \cdot Y_{s,k}$, 方程 (111) 可线性化为方程 (132)。

$$\sum_{j=1}^{\max CN_s} \sum_{k=1}^{\max CYN_s} j \cdot k \cdot \overline{FYLM}_{s,j,k} = M_{s-1}, \forall s \in \{af, ce\} \quad (132)$$

$$\sum_{k=1}^{\max CYN_s} \overline{FY}_{s,j,k} = F_{s,j}, \forall s \in \{af, ce\}, j = 1, \dots, \max CN_s \quad (133)$$

$$\sum_{j=1}^{\max CN_s} \overline{FY}_{s,j,k} = Y_{s,k}, \forall s \in \{af, ce\}, k = 1, \dots, \max CYN_s \quad (134)$$

$$\begin{aligned} & \overline{FYLM}_{s,j,k} = \max LM_s \cdot \overline{FY}_{s,j,k}, \forall s \in \{af, ce\}, \\ & j = 1, \dots, \max CN_s, k = 1, \dots, \max CYN_s \end{aligned} \quad (135)$$

$$\sum_{j=1}^{\max CN_s} \sum_{k=1}^{\max CYN_s} \overline{FYLM}_{s,j,k} = LM_s, \forall s \in \{af, ce\} \quad (136)$$

在方程 (112) 和 (113) 中, 非线性涉及 $E_{s,m}$ 与连续变量的乘积。通过引入辅助变量 $\overline{ETP}_{s,m}$ 和 $\overline{ELM}_{s,m}$, 以下线性方程式可用作线性约束:

$$TP_s = \frac{\sum_m (\widetilde{dm}_m)^2 \cdot \overline{ETP}_{s,m}}{(\text{refDM})^2}, \forall s \in \{af, ce\} \quad (137)$$

$$\overline{ETP}_{s,m} \leq \max TP_s \cdot E_{s,m}, \forall s \in \{af, ce\}, m \quad (138)$$

$$\sum_m \overline{ETP}_{s,m} = TP_s^1, \forall s \in \{af, ce\} \quad (139)$$

$$LM_s = \frac{\sum_m (\widetilde{dm}_m)^2 \cdot \overline{ELM}_{s,m}}{(\text{refDM})^2}, \forall s \in \{af, ce\} \quad (140)$$

$$\overline{ELM}_{s,m} \leq \max CLM_s \cdot E_{s,m}, \forall s \in \{af, ce\}, m \quad (141)$$

$$\sum_m \overline{ELM}_{s,m} = LM_s^1, \forall s \in \{af, ce\} \quad (142)$$

4.4.5. 成本线性化

最后, 为了重新定义固定资产投资公式, 可以使用 $\overline{EF}_{s,m,j}$ 来线性化方程 (114), 如下所示:

$$\begin{aligned} FCI = & lang \cdot (1 + gef) \cdot \\ & \left(brc \cdot brn + \sum_{s \in CS} \sum_m \sum_{j=1}^{\max CN_s} \widetilde{cc}_{s,m} \cdot k \cdot \overline{EF}_{s,m,j} \right. \\ & \left. + oe\lambda \cdot brc \cdot brn \right) \end{aligned} \quad (143)$$

总而言之, MILP模型B*包括方程(3)、(11)~(13)、(15)、(16)、(18)~(21)、(23)~(25)、(28)、(32)、(33)、(42)~(57)、(61)~(68)、(75)~(79)、(92)~(104)、(106)、(115)~(143)、(S1)~(S8)、(S10)~(S12)和(S14)~(S19)中显示的约束条件。

总之, 表2总结了所有四个提出的模型所用的方程。

5. 案例研究

在本节中会将提出的四个优化模型应用于与行业相关的案例研究中, 以检验其性能。此示例的过程流程图如图1所示, 其中包含一个生物反应器和三个色谱分析步骤: 亲和色谱、阳离子交换色谱和阴离子交换色谱分析步骤。为了确定色谱柱的大小, 每个色谱分析步骤使用的柱数限制为四个, 而每批最多允许10个循环。考虑了两种不同的色谱柱直径和床高选择方案。案例1包括10个从50 cm到200 cm不等的离散直径和11个在15~25 cm之间的离散床高, 而案例2有26个在50~300 cm之间的离散直径和21个在10~30 cm范围内的整数的离散床高。在使用离散柱体积尺寸的模型A和A*中, 案例1有110种可选方案, 案例2有546种可选方案。表3给出了详细的可选色谱柱直径和床高。

色谱树脂利用率 μ 为0.95。亲和色谱和阳离子交换色谱分析步骤中的流动线速度限制在200~600 $\text{cm} \cdot \text{h}^{-1}$ 之间, 而阴离子交换色谱分析步骤中的流动线速度 (vel) 固定为300 $\text{cm} \cdot \text{h}^{-1}$ 。表4给出了三个色谱分析步骤的其他参数。

该过程 aot 的最后阶段为340天, 批处理成功率 σ 为90%。表5中提供了非色谱分析步骤的工艺参数。与成本相关的数据可以在Appendix A (Table S1) 中找到。

表2 模型总结

Constraints	MINLP A	MILP A*	MINLP B	MILP B*
Integer variable discretization	—	Eqs. (40–44)	—	Eqs. (42–44), (115), (116)
Column volume	Eqs. (7–11)	Eqs. (7–9), (11), (45–47)	Eqs. (10), (11), (105), (106)	Eqs. (11), (45–47), (106), (117–121)
Product mass	Eqs. (12–14)	Eqs. (12), (13), (48–51)	Eqs. (12–14)	Eqs. (12), (13), (48–51)
Product volume	Eqs. (15–19)	Eqs. (15), (16), (18), (19), (52)	Eqs. (15–19)	Eqs. (15), (16), (18), (19), (52)
Buffer volume	Eqs. (20–26)	Eqs. (20), (21), (23–25), (53–57)	Eqs. (20–26)	Eqs. (20), (21), (23–25), (53–57)
Processing time	Eqs. (27–34)	Eqs. (28), (32), (33), (58–78)	Eqs. (28), (32–34), (107–110)	Eqs. (28), (32), (33), (61–68), (75–78), (122–130)
Data-driven model	Eqs. (3), (4), (35–38)	Eqs. (3), (79–91)	Eqs. (3–5), (111–113)	Eqs. (3), (79), (131–142)
Cost and objective	Eqs. (39), (S1–S19)	Eqs. (92–104), (S1–S8), (S10–S19)	Eqs. (114), (S1–S12), (S14–S19)	Eqs. (92–104), (143), (S1–S8), (S10–S12), (S14–S19)

表3 两个案例下色谱柱尺寸的替代方法

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

表4 色谱操作参数

Step	Affinity chromatography	Cation-exchange chromatography	Anion-exchange chromatography
Yield, yd_s	95%	95%	98%
Dynamic binding capacity, dbc_s ($\text{g}\cdot\text{L}^{-1}$)	15	50	100
Eluate volume ratio, ecv_s	2	3	0
Buffer volume ratio, bcv_s	15	22	18
Resin lifetime, l_s (cycle)	100	100	100
Resin price, rpc_s ($\text{GBP}\cdot\text{L}^{-1}$)	60 000	1 200	1 400

GAMS 24.7 [42]在具有Intel Core i5-3330 3.00 GHz处理器和8.0 GB RAM的64位基于Windows 7的计算机上,使用BARON作为MINLP解决器,使用CPLEX作为MILP解决器,在GAMS 24.7 [42]中实现了提出的优化模型。每个模型的中央处理器时间(CPU)限制为1 h。

6. 结果和讨论

提出的模型已应用于上述两种案例,其中色谱柱大小不同。本节展示并讨论了计算结果。

6.1. 案例 1

表6列出了将四个提出的模型应用于案例1的模型统计和计算结果,其中所有四个模型均取得其全局最优解。需要注意的是,尽管所报道的四个模型最优目标的小数点后第一位相同,但是MILP模型的解实际上非常接近相应MINLP模型的全局最优,并且具有相同

的柱大小和关键操作决策。它们比MINLP模型的最优目标值约低 1×10^{-4} ,这是多参数分解理论所显示的下界。MINLP模型A能够在494 s内得到201.7英镑 $\cdot\text{g}^{-1}$ 的最优值。线性化后,尽管MILP模型A*中方程和变量的数量增加,但找到最优解花费的时间稍短(381 s)。同时,如表6所示,MINLP模型B具有与MINLP模型A相同的连续变量,但是方程和离散变量的数量大大减少了一个数量级。因此,该模型能够在47 s内找到最优解。MINLP模型B*是通过将MINLP模型B线性化而得到的,与MILP模型A*相比,其方程和变量要少得多,并且仅需5 s即可找到最优解,从而为CPU节省了两个数量级运算量。通过比较可得,在所提出的四个模型中,在计算方面,模型B和B*明显优于其他两个模型。尤其是MILP模型B*,它需要的计算工作量最少,并且在较大规模的问题处理上具有最大的潜力,这在下一部分讨论的较大示例(案例2)中也很明显。需要指出的是,所得分段线性回归模型的预测误差对优化模型的最优目标

值影响很小。例如，估计的输出通量有17%的变化只会导致最优目标值相差0.1%。对于案例2也有类似的发现。

接着，将详细讨论案例1的最优解。图3中给出了最

表5 非色谱单元操作的参数

Unit operation parameter	Value
Cell culture	
Titer, $titter$ ($\text{g}\cdot\text{L}^{-1}$)	2
Bioreaction time, bri (d)	4
Seed train bioreaction time, st (d)	2
Number of bioreactors, brn	1
Bioreactor working volume ratio, a	75%
Bioreactor volume, brv (L)	20 000
Centrifugation 1	
Yield, yd_s	95%
Buffer volume ratio, bvr_s	0.9
Processing rate, pr_s ($\text{L}\cdot\text{h}^{-1}$)	250
Homogenization	
Yield, yd_s	100%
Processing rate, pr_s ($\text{L}\cdot\text{h}^{-1}$)	143
Centrifugation 2	
Yield, yd_s	95%
Buffer volume ratio, bvr_s	0.1
Processing rate, pr_s ($\text{L}\cdot\text{h}^{-1}$)	250
Filtration	
Yield, yd_s	98%
Processing rate, pr_s ($\text{L}\cdot\text{h}^{-1}$)	125
UF/DF 1	
Yield, yd_s	98%
Flush volume ratio, fvr	6
Processing rate, pr_s ($\text{L}\cdot\text{h}^{-1}$)	40
UF/DF 2	
Yield, yd_s	98%
Defiltration volume ratio, dvr	7
Processing rate, pr_s ($\text{L}\cdot\text{h}^{-1}$)	40
Bulk fill	
Yield, yd_s	98%
Filling time, T_s (h)	6
Final concentration, $fconc$ ($\text{g}\cdot\text{L}^{-1}$)	10

表6 提出模型在案例1中的统计和计算性能

Model	No. of equations	No. of continuous variables	No. of discrete variables	Optimal objective ($\text{GBP}\cdot\text{g}^{-1}$)	CPU (s)
MINLP A	423	85	670	201.7	494
MILPA*	13 646	11 092	12 097	201.7	381
MINLP B	93	85	46	201.7	47
MILP B*	1 508	700	595	201.7	5

佳的色谱柱尺寸调整策略，其中色谱柱的直径与绘制的形状的宽度成比例，而床高与形状的高度成比例。虚线形状代表每批所需的循环。在每个色谱分析步骤中，仅使用一个色谱柱。亲和色谱分析步骤的色谱柱直径为180 cm，床高为15 cm，而阳离子交换色谱分析步骤则使用较小的色谱柱，直径为90 cm，床高为21 cm。阴离子交换色谱分析步骤中的色谱柱的最小直径为80 cm，但使用最大床高为25 cm。

在此，检查了通量回归模型和操作决策的性能。在最优解中，元模型的输出通量在亲和色谱分析步骤中为 $1869.8 \text{ g}\cdot\text{h}^{-1}$ ，在阳离子交换色谱分析步骤中为 $1823.9 \text{ g}\cdot\text{h}^{-1}$ 。如图3所示，在这两个色谱分析步骤中，每批分别需要五个和四个循环。在每个循环中加载到每个柱上的所得产物质量分别为5306.7 g和6301.7 g。转换为直径为1 m的柱之后，在亲和色谱分析步骤中，装载的质量落入分段线性回归模型的第一个区间，并且使用相应的函数来估算通量，如下所示：

$$TP_{af} = 0.1914 \cdot LM_{af} + 1.8^2 \cdot (0.3570 \cdot V_{af} - 12.0477 \cdot H_{af} + 230.131) \quad (144)$$

式中，根据方程(37)，添加 1.8^2 是用来将1 m直径的柱的性能转换为所选的1.8 m直径的柱的性能， $LM_{af} = 1.8^2 \cdot LM_{af0}$ 。在阳离子交换色谱分析步骤中，还选择了回归模型中的第一个间隔。同样，使用的回归模型如下：

$$TP_{ce} = 0.1287 \cdot LM_{ce} + 0.9^2 \cdot (2.3940 \cdot V_{ce} - 51.4883 \cdot H_{ce} + 895.2814) \quad (145)$$

在以上两个函数中，线速度变量均对吞吐率有积极影响，因此，两个步骤的流动线速度都达到其上限 $600 \text{ cm}\cdot\text{h}^{-1}$ 。

最后，从图4所示的最优成本分布来看，由于所使用的亲和树脂价格高昂，耗材成本（即此问题中的树脂成本）占了最大比例，超过30%。此外，设备的资本成本、化学试剂（缓冲液和介质）成本以及人工成本都在总成本中占很大比例。

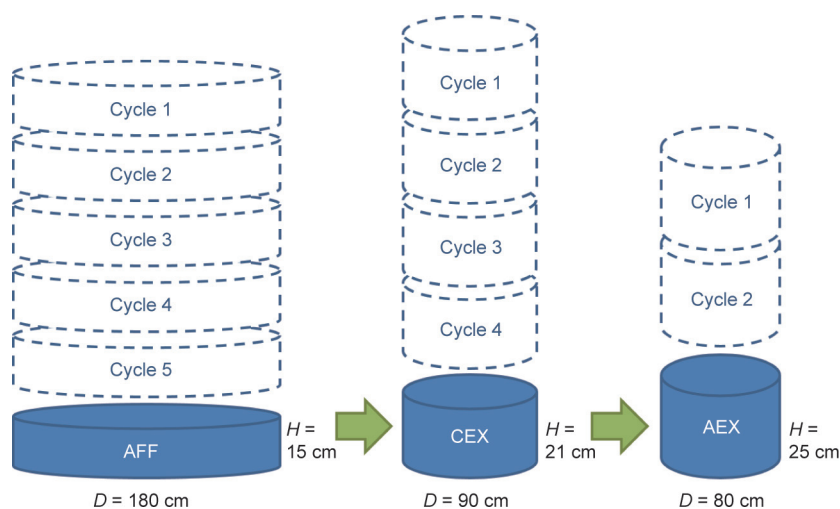


图3. 案例1的最佳色谱柱大小调整策略。CEX: 阳离子交换色谱法, AEX: 阴离子交换色谱法, D : 直径, H : 床高。

6.2. 案例 2

在案例2中, 考虑了更多的可供选择的直径和床高, 从而生成了较大规模的优化模型, 如表7所示。与案例1相比, MINLP模型B的离散变量较多, 而其他模型涉及的方程和变量也同时增加了。需要注意的是, 模型A和A*未能在计算时间限制3600 s之内终止, 尽管与最优解的差距分别仅为0.6%和0.4%。根据图5中所示的两个模型的求解过程, MILP模型A*在220 s左右找到了一个很好的可行解, 实际上在10 min内就算出了最优解。但是, 在分支定界过程中, 溶液的下界收敛得十分缓慢, 以至于在给定的时限内无法证明所获得目标的最优值是200.3英镑·g⁻¹。同时, MINLP模型A计算相对较慢, 大约在1000 s后才算出第一个可行解, 并且在近30 min时获得了良好的可行解。与模型A和A*相比, 模型B和B*表现出明显改善的计算性能, 并在4 min内得到了最优解。MINLP模型B大约需要1 min得到一个接近的可行解, 需要192 s得到最优解。MILP模型B*实现了一个数量级的CPU节省, 算得最优解仅需24 s, 这也是MINLP模型B*由于多参数分解而产生的一个非常接近的下界。对于在最优值1%范围内的良好可行解, 模型B*仅需6 s即可得到。由此可证明MILP模型B*的计算优势。

关于决定柱大小, 案例2有更多可能的柱大小选项。从图6可以看出, 与案例1的最优决策相比, 在色谱前两步安装了一个直径更大但床高更小的柱。选择的床高(11 cm和14 cm)超出案例1允许的床高范围(15~25 cm)。另外, 案例1中没有案例2选择的直径(190 cm和110 cm)。由于可选参数较多, 案例1的最优解仅为案例2的可行解, 案例2的最优每克COG提高1.4英镑。同时,

阴离子交换色谱分析步骤选择相同的柱(直径为80 cm, 床高25 cm)。选择较大直径的柱会导致设备投资成本升高, 而床高越小, 树脂用量越少, 相应的成本也越低。这些成本的差别相对很小, 因此, 成本分布与案例1非常相似, 这里不再进一步讨论。

在亲和色谱和阳离子色谱两个步骤中, 最高允许流速为600 cm·h⁻¹。由于所选直径尺寸的差异, 这两个步骤的实际通量回归模型与案例1略有不同:

$$TP_{af} = 0.1914 \cdot LM_{af} + 1.9^2 \cdot (0.3570 \cdot V_{af} - 12.0477 \cdot H_{af} + 230.131) \quad (146)$$

$$TP_{ce} = 0.1287 \cdot LM_{ce} + 1.1^2 \cdot (2.3940 \cdot V_{ce} - 51.4883 \cdot H_{ce} + 895.2814) \quad (147)$$

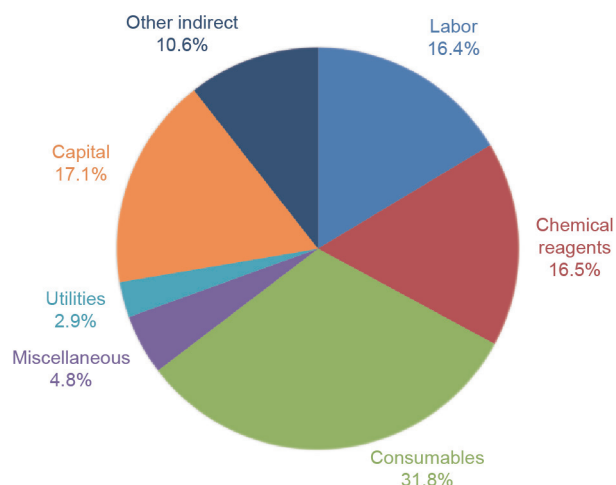


图4. 案例1的最优成本分配。

如图6所示,与案例1相比,亲和色谱分析步骤多使用了一个循环,因此在每个循环中加载的质量更少(4422.3 g)。但由于床高较小,直径较大,通量增加到 $1972.0 \text{ g}\cdot\text{h}^{-1}$ 。对于阳离子交换色谱分析步骤,虽然采用了与案例1相同的循环次数和加载质量,但由于选择了更大的直径和更小的床层高度,得到了更高的通量 $2760.2 \text{ g}\cdot\text{h}^{-1}$ 。

7. 结语

本文研究了抗体制备过程的多尺度优化问题。在操

作层面,为了模拟色谱过程的复杂性能,使用基于微尺度实验数据的制造规模的模拟数据集,开发了数据驱动模型来估计色谱的通量。通过对多种元模型生成方法的比较,建立了基于模拟数据集的分段线性回归模型。

在工艺设计层面,为了确定最佳色谱柱尺寸策略,提出了两种可选择的最小二乘法模型以最小化每克COG。采用线性化技术,建立了两个MILP模型。将通量回归模型合并到优化模型中,以确定最优的操作决策,即流速和每个批加工的循环数。

通过研究两种不同柱尺寸方案的工业相关案例,比较了四种优化模型的计算性能。综上所述,模型B和B*

表7 案例2的统计和计算性能

Model	No. of equations	No. of continuous variables	No. of discrete variables	Optimal objective ($\text{GBP}\cdot\text{g}^{-1}$)	CPU (s)
MINLPA	1731	85	3286	200.5 ^a	3600 ^a
MILPA*	63 350	52 948	59 185	200.3 ^b	3600 ^b
MINLP B	93	85	94	200.3	192
MILP B*	1972	828	1027	200.3	24

^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.

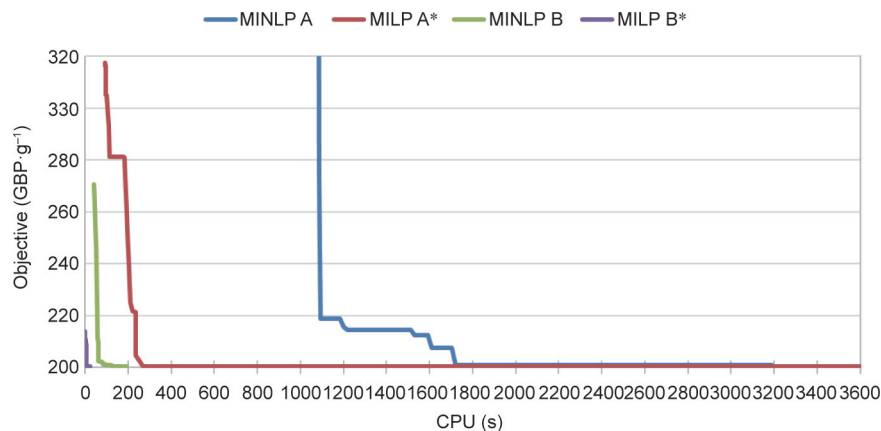


图5. 案例2中四种模型的求解过程。

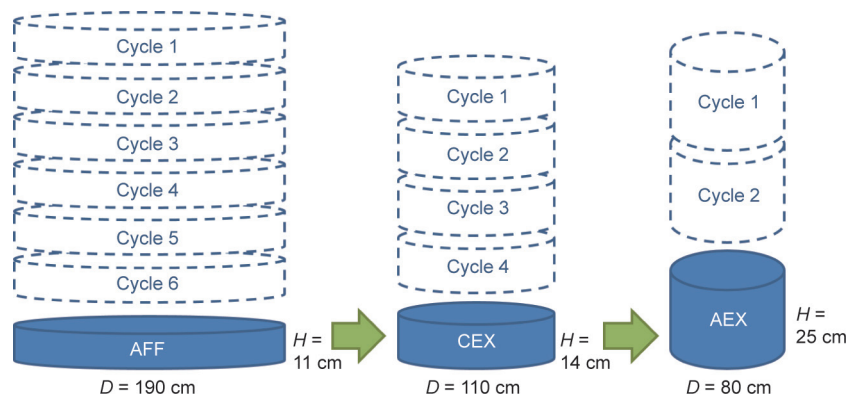


图6. 案例2的最佳色谱柱尺寸策略。

表现出更高效的计算性能。尤其是第二个MILP模型被证明是计算效率最高的，因此可以推荐用于大规模优化研究。此外，还讨论了最优解的细节，并证明了数据驱动模型能够很好地实现最优通量。

本研究的未来研究方向可能是开发阴离子交换色谱的数据驱动元模型，以及更多的色谱参数，如产量和结合能力，以及更多的输入变量，如pH值和温度。此外，净化过程的其他性能指标，如除杂能力，可以考虑用于多目标优化[9,21,22]。也可以考虑产量、滴度等参数的不确定性[19,22]。最后，在纯化过程中，一次性色谱法可被认为是智能生物制药[48]的一个重要方向。

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 = $p, \dots, \max p$
i	Column volume size
j	Column number = $1, \dots, \max CN_s$
k	Cycle number = $1, \dots, \max CYN_s$
m	Diameter size
n	Digit of the binary representation = $1, \dots, [\log_2 \max BN]$
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), ae (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 to column 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, $GBP \cdot L^{-1}$
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 , $g \cdot L^{-1}$
$dm_{s,i}$	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, $g \cdot L^{-1}$
fvr	Flush volume ratio of the first UF/DF step
gef	General equipment factor
gu	General utility unit cost, $GBP \cdot L^{-1}$

$h_{s,i}$	Bed height of column size i at chromatography step s , cm	sfd	Duration per shift, h
ir	Interest rate	sfn	Number of shifts per day, d^{-1}
$i\lambda$	Ratio of insurance cost to fixed capital investment	st	Seed train bioreaction time, d
l_s	Lifetime of resin at chromatography step s , cycle	$s\lambda$	Supervisors cost ratio to direct labor cost
$lang$	Lang factor	$titer$	Upstream product titer, $g \cdot L^{-1}$
$maxBBV$	Maximum buffer volume per batch, L	$t\lambda$	Tax cost ratio to the fixed capital investment
$maxBN$	Maximum number of batches	uon	Number of operators per bioreactor in USP
$masCN_s$	Maximum number of columns at chromatography step s	uot	USP operating time per day
$maxCOG$	Maximum COG per gram, $GBP \cdot L^{-1}$	vel	Linear velocity of flow at the anion-exchange chromatography step, $cm \cdot h^{-1}$
$maxCYN_s$	Maximum number of cycles at chromatography step s	w	Wage of an operator, $GBP \cdot L^{-1}$
$maxH_s$	Maximum column bed height size at chromatography step s , cm	yd_s	Product yield at step s
$maxLM_s$	Maximum product mass loaded at chromatography step s , g	α	Bioreactor working volume ratio
$maxp$	Maximum position in multiparametric disaggregation	$\beta_{s,r}^0$	Constant coefficient in interval r at chromatography step s
$maxT_s$	Maximum processing time per batch at chromatography step s , h	$\beta_{s,r}^H$	Coefficient for bed height in interval r at chromatography step s
$maxTCV_s$	Maximum total column volume at chromatography step s , L	$\beta_{s,r}^{HM}$	Coefficient for loaded mass in interval r at chromatography step s
$maxTP_s$	Maximum throughput at chromatography step s , $g \cdot L^{-1}$	$\beta_{s,r}^V$	Coefficient for velocity in interval r at chromatography step s
$ma\lambda$	Maintenance cost ratio to the fixed capital investment	ε	A small number
$mepc$	Media price, $GBP \cdot L^{-1}$	θ	Media overfill allowance
$mi\lambda$	Miscellaneous material cost ratio to chemical reagent and consumable costs	μ	Chromatography resin utilization factor
$m\lambda$	Management cost ratio to direct labor cost	σ	Batch success rate
$oe\lambda$	Other equipment cost ratio to the bioreactor cost	<i>Continuous variables</i>	
of	Resin overpacking factor	ABV	Annual buffer volume, L
pr_s	Processing rate of step s , $L \cdot h^{-1}$	AP	Annual product output, g
$q\lambda$	Ratio of QCQA cost to direct labor cost	AT	Annual downstream operating time, d
rpc_s	Resin price at chromatography step s , $GBP \cdot L^{-1}$	BAT	Time for adding buffer per batch at the anion-exchange chromatography step, h
refCC	Reference cost of a chromatography column, GBP	BBV	Buffer volume added per batch, L
refDM	Reference diameter of a chromatography column, cm	BC	Buffer cost, GBP
		BRC	Bioreactor cost, GBP
		BT	Downstream processing time per batch, d
		BV_s	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

CAP_q	Continuous variable for annual production in multiple disaggregation at digit q	VFR	Volumetric flow rate at anion-exchange chromatography step, $L \cdot h^{-1}$
$CTP_{s,q}$	Continuous variable for throughput in multiple disaggregation at digit q , step s	<i>Binary variables</i>	
DLC	Direct labor cost, GBP	$BAP_{d,q}$	1 if digit q for power d is selected for annual production output; 0 otherwise
FCI	Fixed capital investment, GBP	$BTP_{s,d,q}$	1 if digit q for power d is selected for throughput at chromatography step s ; 0 otherwise
GUC	General utility cost, GBP	$E_{s,m}$	1 if diameter size m is selected at chromatography step s ; 0 otherwise
IC	Insurance cost, GBP	$F_{s,j}$	1 if there are j columns at chromatography step s ; 0 otherwise
LC	Labor cost, GBP	$O_{s,r}$	1 if the function at interval r is selected at chromatography step s ; 0 otherwise
LM_s	Mass loaded to single column at chromatography step s , g	$W_{s,i,j}$	1 if there are j columns of size i at chromatography step s ; 0 otherwise
LM_s^1	Mass loaded to single 1-m-diameter column at chromatography step s , g	$X_{s,i}$	1 if column size i is selected at chromatography step s ; 0 otherwise
M_0	Initial product mass entering downstream processes per batch, g	$Y_{s,k}$	1 if there are k cycles at chromatography step s ; 0 otherwise
M_s	Product mass per batch after step s , g	Z_n	1 if the n th digit of the binary representation of variable BN is equal to 1; 0 otherwise
MAC	Maintenance cost, GBP	<i>Integer variables</i>	
MC	Management cost, GBP	BN	Number of completed batches
MEC	Media cost, GBP	CN_s	Number of columns of size i at chromatography step s
MIC	Miscellaneous material cost, GBP	\widetilde{CN}_s	Number of columns at chromatography step s
OIC	Other indirect costs, GBP	CYN_s	Number of cycles per batch at chromatography step s
PLT	Time for loading product per batch at anion-exchange chromatography step, h	H_s	Bed height of column at chromatography step s , cm
PV_0	Initial product volume entering downstream processes per batch, L	<i>Auxiliary variables</i>	
PV_s	Product volume per batch after step s , L	$\overline{BCOGAP}_{d,q}$	$\equiv COG \cdot BAP_{d,q}$
QC	QCQA cost, GBP	$\overline{BTTP}_{s,d,q}$	$\equiv T_s \cdot BTP_{s,d,q}$
RV_s	Resin volume required at chromatography step s , L	\overline{CCOGAP}_q	$\equiv COG \cdot CAP_q$
SC	Supervisors cost, GBP	\overline{COGAP}	$\equiv COG \cdot AP$
T_s	Processing time per batch of step s , h	$\overline{CTTP}_{s,q}$	$\equiv T_s \cdot CTP_{s,q}$
TC	Tax cost, GBP	$\overline{EF}_{s,m,j}$	$\equiv E_{s,m} \cdot F_{s,j}$
TCV_s	Total column volume at chromatography step s , L	$\overline{EFH}_{s,m,j}$	$\equiv E_{s,m} \cdot F_{s,j} \cdot H_s$
TP_s	Throughput of single column at chromatography step s , $g \cdot h^{-1}$		
TP_s^1	Throughput of single 1-m-diameter column at chromatography step s , $g \cdot h^{-1}$		
UC	Utilities cost, GBP		
V_s	Linear velocity of flow at chromatography step s , $cm \cdot h^{-1}$		

$$\overline{EFT}_{s,m,j} \equiv E_{s,m} \cdot F_{s,j} \cdot PLT$$

$$\overline{ELM}_{s,m} \equiv E_{s,m} \cdot LM_s^1$$

$$\overline{ETP}_{s,m} \equiv E_{s,m} \cdot TP_s^1$$

$$\overline{FTTP}_{s,j} \equiv F_{s,j} \cdot T_s \cdot TP_s$$

$$\overline{FY}_{s,j,k} \equiv F_{s,j} \cdot Y_{s,k}$$

$$\overline{FYL M}_{s,j,k} \equiv F_{s,j} \cdot Y_{s,k} \cdot LM_s$$

$$\overline{TTP}_s \equiv T_s \cdot TP_s$$

$$\overline{WT}_{s,i,j} \equiv W_{s,i,j} \cdot PLT$$

$$\overline{W TTP}_{s,i,j} \equiv W_{s,i,j} \cdot T_s \cdot TP_s$$

$$\overline{WY}_{s,i,j,k} \equiv W_{s,i,j} \cdot Y_{s,k}$$

$$\overline{WYLM}_{s,i,j,k} \equiv W_{s,i,j} \cdot Y_{s,k} \cdot LM_s$$

$$\overline{XLM}_{s,i} \equiv X_{s,i} \cdot LM_s^1$$

$$\overline{XLM}_{s,i} \equiv X_{s,i} \cdot LM_s^1$$

$$\overline{XY}_{s,i,k} \equiv X_{s,i} \cdot Y_{s,k}$$

$$\overline{YH}_{s,k} \equiv Y_{s,k} \cdot H_s$$

$$\overline{YV}_{s,k} \equiv Y_{s,k} \cdot TCV_s$$

$$\overline{ZM}_{s,n} \equiv Z_n \cdot M_s$$

$$\overline{ZT}_n \equiv Z_n \cdot BT$$

$$\overline{ZV}_n \equiv Z_n \cdot BBV$$

$$\overline{ZYV}_{s,k,n} \equiv Z_n \cdot Y_{s,k} \cdot TCV_s$$

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

- [1] Davis J, Edgar T, Porter J, Bernaden J, Sarli M. Smart manufacturing, manufacturing intelligence and demand-dynamic performance. *Comput Chem Eng* 2012;47:145–56.
- [2] Thoben KD, Wiesner S, Wuest T. "Industrie 4.0" and smart manufacturing—a review of research issues and application examples. *Int J Automat Technol* 2017;11(1):4–16.
- [3] Kang HS, Lee JY, Choi SS, Kim H, Park JH, Son JY, et al. Smart manufacturing: past research, present findings, and future directions. *Int J Pr Eng Man-GT* 2016;3(1):111–28.
- [4] Bogle IDL. A perspective on smart process manufacturing research challenges for process systems engineers. *Engineering* 2017;3(2):161–5.
- [5] Simaria AS, Turner R, Farid SS. A multi-level meta-heuristic algorithm for the optimisation of antibody purification processes. *Biochem Eng J* 2012;69:144–54.
- [6] Brunet R, Guillén-Gosálbez G, Pérez-Corraea JR, Caballero JA, Jiménez L. Hybrid simulation-optimization based approach for the optimal design of single-product biotechnological processes. *Comput Chem Eng* 2012;37:125–35.
- [7] Allmendinger R, Simaria AS, Farid SS. Efficient discovery of chromatography equipment sizing strategies for antibody purification processes using evolutionary computing. In: *Proceedings of the 12th International Conference on Parallel Problem Solving from Nature-Volume Part II*; 2012 Sep 1–5; Taormina, Italy. Berlin: Springer; 2012. p. 468–77.
- [8] Allmendinger R, Simaria AS, Turner R, Farid SS. Closed-loop optimization of chromatography column sizing strategies in biopharmaceutical manufacture. *J Chem Technol Biotechnol* 2014;89(10):1481–90.
- [9] Allmendinger R, Simaria AS, Farid SS. Multiobjective evolutionary optimization in antibody purification process design. *Biochem Eng J* 2014;91:250–64.
- [10] Martagan T, Krishnamurthy A, Leland PA, Maravelias CT. Performance guarantees and optimal purification decisions for engineered proteins. *Oper Res* 2018;66(1):18–41.
- [11] Asenjo JA, Montagna JM, Vecchiotti AR, Iribarren OA, Pinto JM. Strategies for the simultaneous optimization of the structure and the process variables of a protein production plant. *Comput Chem Eng* 2000;24(9–10):2277–90.
- [12] Montagna JM, Vecchiotti AR, Iribarren OA, Pinto JM, Asenjo JA. Optimal design of protein production plants with time and size factor process models. *Biotechnol Prog* 2000;16(2):228–37.
- [13] Pinto JM, Montagna JM, Vecchiotti AR, Iribarren OA, Asenjo JA. Process performance models in the optimization of multiproduct protein production plants. *Biotechnol Bioeng* 2001;74(6):451–65.
- [14] Simeonidis E, Pinto JM, Lienqueo ME, Tsoka S, Papageorgiou LG. MINLP models for the synthesis of optimal peptide tags and downstream protein processing. *Biotechnol Prog* 2005;21(3):875–84.
- [15] Liu S, Simaria AS, Farid SS, Papageorgiou LG. Mixed integer optimisation of antibody purification processes. In: *Kraslawski A, Turunen I, editors. Computer aided chemical engineering. Proceedings of the 23rd European Symposium on Computer Aided Process Engineering*; 2013 June 9–12; Lappeenranta, Finland. Amsterdam: Elsevier; 2013. p. 157–62.
- [16] Liu S, Simaria AS, Farid SS, Papageorgiou LG. Designing cost-effective biopharmaceutical facilities using mixed-integer optimization. *Biotechnol Prog* 2013;29(6):1472–83.
- [17] Liu S, Simaria AS, Farid SS, Papageorgiou LG. Optimising chromatography strategies of antibody purification processes by mixed integer fractional programming techniques. *Comput Chem Eng* 2014;68:151–64.
- [18] Liu S, Simaria AS, Farid SS, Papageorgiou LG. Mathematical programming approaches for downstream processing optimisation of biopharmaceuticals. *Chem Eng Res Des* 2015;94:18–31.
- [19] Liu S, Farid SS, Papageorgiou LG. Integrated optimization of upstream and downstream processing in biopharmaceutical manufacturing under uncertainty: a chance constrained programming approach. *Ind Eng Chem Res* 2016;55(16):4599–612.
- [20] Liu S, Gerontas S, Gruber D, Turner R, Titchener-Hooker NJ, Papageorgiou LG. Optimization-based framework for resin selection strategies in biopharmaceutical purification process development. *Biotechnol Prog* 2017;33(4):1116–26.
- [21] Liu S, Papageorgiou LG. Optimal production of biopharmaceutical manufacturing. In: *Yuan Z, Singh R, editors. Process systems engineering for pharmaceutical manufacturing. Computer aided chemical engineering. Amsterdam: Elsevier; 2018. p. 569–95.*
- [22] Liu S, Papageorgiou LG. Multi-objective optimisation for biopharmaceutical manufacturing under uncertainty. *Comput Chem Eng* 2018;119:383–93.
- [23] Bhosekar A, Ierapetritou M. Advances in surrogate based modeling, feasibility analysis, and optimization: a review. *Comput Chem Eng* 2018;108:250–67.
- [24] Wang GG, Shan S. Review of metamodelling techniques in support of engineering design optimization. *J Mech Des* 2007;129(4):370–80.
- [25] Forrester A, Sobester A, Keane A. *Engineering Design via Surrogate Modelling: A Practical Guide*. Chichester: John Wiley & Sons; 2008.
- [26] Song M, Breneman CM, Bi J, Sukumar N, Bennett KP, Cramer S, et al. Prediction of protein retention times in anion-exchange chromatography systems using support vector regression. *J Chem Inf Comput Sci* 2002;42(6):1347–57.
- [27] Mandenius CF, Brundin A. Bioprocess optimization using design-of-experiments methodology. *Biotechnol Prog* 2008;24(6):1191–203.
- [28] Ghose S, Zhang J, Conley L, Caple R, Williams KP, Cecchini D. Maximizing binding capacity for protein A chromatography. *Biotechnol Prog* 2014;30(6):1335–40.
- [29] Wang G, Briskot T, Hahn T, Baumann P, Hubbuch J. Root cause investigation of deviations in protein chromatography based on mechanistic models and artificial neural networks. *J Chromatogr A* 2017;1515:146–53.
- [30] Nagrath D, Messac A, Bequette BW, Cramer SM. A hybrid model framework for the optimization of preparative chromatographic processes. *Biotechnol Prog* 2004;20(1):162–78.
- [31] Pirrung SM, van der Wielen LAM, van Beckhoven RFWC, van de Sandt EJAX, Eppink MHM, Ottens M. Optimization of biopharmaceutical downstream processes supported by mechanistic models and artificial neural networks. *Biotechnol Prog* 2017;33(3):696–707.
- [32] Nascimento A, Pinto IF, Chu V, Aires-Barros MR, Conde JP, Azevedo AM. Studies on the purification of antibody fragments. *Sep Purif Technol* 2018;195:388–97.
- [33] Scanlan C, Shumway J, Castano J, Wagner M, Waghmare R. Challenges and strategies for the downstream purification and formulation of Fab antibody fragments. *Biopharm Int* 2014;27(1):42–4.
- [34] Gerontas S, Asplund M, Hjorth R, Bracewell DG. Integration of scaledown experimentation and general rate modelling to predict manufacturing scale chromatographic separations. *J Chromatogr A* 2010;1217(44):6917–26.
- [35] Boushaba R, Baldascini H, Gerontas S, Titchener-Hooker NJ, Bracewell DG. Demonstration of the use of windows of operation to visualize the effects of fouling on the performance of a chromatographic step. *Biotechnol Prog* 2011;27(4):1009–17.
- [36] Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KK. Improvements to

- the SMO algorithm for SVM regression. *IEEE Trans Neural Netw* 2000;11(5):1188–93.
- [37] MacKay DJC. Introduction to Gaussian processes. In: Bishop CM, editor. *Neural networks and machine learning*, NATO ASI series: Ser. F: computer and systems science. Berlin: Springer; 1998. p. 133–66.
- [38] Wang Y, Witten IH. Modeling for optimal probability prediction. In: *Proceedings of the Nineteenth International Conference on Machine Learning*; 2002 July 8–12; Sydney, NSW, Australia; 2002. p. 650–7.
- [39] Box GEP, Draper NR. *Response surfaces, mixtures, and ridge analyses*. 2nd ed. Hoboken: John Wiley & Sons; 2007.
- [40] Yang L, Liu S, Tsoka S, Papageorgiou LG. Mathematical programming for piecewise linear regression analysis. *Expert Syst Appl* 2016;44:156–67.
- [41] Frank E, Hall MA, Witten IH. *Data mining: practical machine learning tools and techniques*. 4th ed. San Francisco: Morgan Kaufmann; 2016.
- [42] GAMS Development Corporation. *GAMS: a user's guide*. Washington, DC: GAMS Development Corporation; 2016.
- [43] Glover F. Improved linear integer programming formulations of nonlinear integer problems. *Manage Sci* 1975;22(4):455–60.
- [44] Kolodziej S, Castro PM, Grossmann IE. Global optimization of bilinear programs with a multiparametric disaggregation technique. *J Glob Optim* 2013;57(4):1039–63.
- [45] Castro PM. Normalized multiparametric disaggregation: an efficient relaxation for mixed-integer bilinear problems. *J Glob Optim* 2016;64(4):765–84.
- [46] Koleva MN, Styan CA, Papageorgiou LG. Optimisation approaches for the synthesis of water treatment plants. *Comput Chem Eng* 2017;106:849–71.
- [47] Lin MH, Carlsson JG, Ge D, Shi J, Tsai JF. A review of piecewise linearization methods. *Math Probl Eng* 2013;2013:101376.
- [48] Langer ES, Rader RA. Single-use technologies in biopharmaceutical manufacturing: a 10-year review of trends and the future. *Eng Life Sci* 2014;14(3):238–43.