

乙醇偶合制备C₄烯烃的合理优化

白雨璐

东北大学秦皇岛分校, 数学与统计学院, 河北 秦皇岛

收稿日期: 2022年5月4日; 录用日期: 2022年6月8日; 发布日期: 2022年6月15日

摘要

作为重要的基础化工原料, C₄烯烃被广泛应用于化工产品及医药中间体的生产。乙醇是生产制备C₄烯烃的重要原料, 在制备过程中, C₄烯烃的收率一定程度上取决于反应温度和反应的催化剂组合, 所以如何合理地设计催化剂组合及温度至为关键。本文为探究不同反应条件对乙醇转化率及C₄烯烃选择性的影响, 建立基于OLS + 稳健的标准误的多元线性回归模型, 得到对乙醇转化率及C₄烯烃选择性线性相关显著的影响因素, 并计算标准化回归系数, 对影响因素的重要程度进行排序。为确定最佳反应条件, 建立逐步筛选模型, 并用MATLAB分别对C₄烯烃收率与单变量进行曲线拟合。采用粒子群优化算法, 依次确定C₄烯烃收率最高时, 各变量的最佳取值。

关键词

多元线性回归, WilCoxon符号秩检验, 逐步筛选, 粒子群优化算法

Optimum Preparation of C₄ Olefin by Ethanol Coupling

Yulu Bai

School of Mathematics and Statistics, Northeastern University Qinhuangdao, Qinhuangdao Hebei

Received: May 4th, 2022; accepted: Jun. 8th, 2022; published: Jun. 15th, 2022

Abstract

As an important basic chemical raw material, C₄ olefins are widely used in the production of chemical products and pharmaceutical intermediates. Ethanol is an important raw material for the production and preparation of C₄ olefin. In the preparation process, the yield of C₄ olefin depends on the reaction temperature and catalyst combination, so how to design the catalyst combination and temperature reasonably is the key. In order to explore the effects of different reaction conditions on ethanol conversion and C₄ olefin selectivity, a multiple linear regression model

based on OLS + robust standard error was established to obtain the influential factors with significant linear correlation on ethanol conversion and C₄ olefin selectivity. The standardized regression coefficient is calculated to rank the importance of the influencing factors. In order to determine the optimal reaction conditions, a stepwise screening model was established, and the curve fitting between the yield of C₄ olefin and single variable was carried out by MATLAB. Finally, particle swarm optimization algorithm was used to determine the optimal value of each variable when the yield of C₄ olefin was the highest.

Keywords

Multiple Linear Regression, Wilcoxon Sign Rank Test, Stepwise Screening, Particle Swarm Optimization Algorithm

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

C₄ 烯烃是工业生产及医药生产过程中的重要化工原料, 如何在满足节能、清洁要求的同时高效制备 C₄ 烯烃显得尤为重要。传统的工业生产方式采用化石能源为原料, 但随着经济的快速发展和人们环保意识的提高, 更倾向于采用绿色清洁能源, 乙醇来源广泛, 绿色清洁, 以其为平台化合物生产高附加值产物具有很大的应用前景及经济效益。

在乙醇偶合制备 C₄ 烯烃的过程中, 催化剂组合中 Co 负载量、乙醇浓度、Co/SiO₂ 和 HAP 装料比、装料方式和反应温度都会对乙醇转化率和 C₄ 烯烃的选择性产生影响[1], 从而影响 C₄ 烯烃的收率。传统利用改变实验条件多次重复实验, 来对比实验结果, 从而逐步缩小条件范围, 来确定最佳反应条件的方法会消耗大量时间精力, 同时在一定程度上也会造成资源浪费, 所以利用现有有限的实验数据建立数学模型探究各因素的影响程度, 并且确定可以使 C₄ 烯烃收率达到最大的反应条件, 在实际实验中可以进行针对性的调整, 在一定程度上也可以减少时间、精力、资源的消耗, 从而使生产效率最大化。

2. 探究催化剂和温度对反应的影响

为探究不同催化剂组合及温度对乙醇转化率、C₄ 烯烃选择性大小的影响, 且催化剂组合中存在多个影响因素, 因此采用回归方法探究多因素对反应的影响, 首先对催化剂组合中的因素进行拆分, 将温度、乙醇浓度、Co/SiO₂ 和 HAP 装料比、Co 负载量、催化剂总重这五个因素作为自变量($x_1 \sim x_5$), 乙醇转化率和 C₄ 烯烃选择性作为因变量。用部分数据建立模型, 并进行显著性差异检验。具体步骤如下:

Step 1: 随机选取 n 组数据为实验组数据, 建立基于普通最小二乘估计法的回归模型, 其余数据为检验组。

Step 2: 检验是否存在异方差并对模型进行改进。

Step 3: 确定最终模型并对回归结果进行显著性检验, 并验证模型的可行性。

Step 4: 计算标准化回归系数进一步确定自变量对因变量影响的重要程度。

2.1. 模型建立

涉及 5 个自变量的多元线性回归模型[2]可表示为:

$$\begin{cases} y = \omega_0 + \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \omega_4 x_4 + \omega_5 x_5 + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

为了方便推导，通过 n 组实验数据引入矩阵符号分析：

$$Y = \begin{bmatrix} y_1 \\ y_n \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{15} \\ x_{21} & x_{22} & \cdots & x_{25} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{n5} \end{bmatrix}, \quad \omega = \begin{bmatrix} \omega_0 \\ \omega_1 \\ \vdots \\ \omega_5 \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_n \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

其中 X 为模型设计矩阵， X 为常数矩阵， Y 与 ε 是随机向量，

$$Y \sim N_n(X\beta, \sigma^2 I), \quad \varepsilon \sim N_n(0, \sigma^2 I) \quad (I \text{ 为 } n \text{ 阶单位矩阵})$$

ω 为需求回归系数， ε 为不可观测的随机误差。

2.2. 模型求解

以温度、乙醇浓度、Co/SiO₂ 和 HAP 装料比、Co 负载量、催化剂总重这五个因素作为自变量，分别记作 $x_1 \sim x_5$ ，以乙醇转化率、C₄ 烯烃选择性作为因变量，利用实验组数据进行基于普通最小二乘估计法的多元线性回归分析，模型通过了显著性检验，画出残差与拟合值的散点图，如图 1 所示。

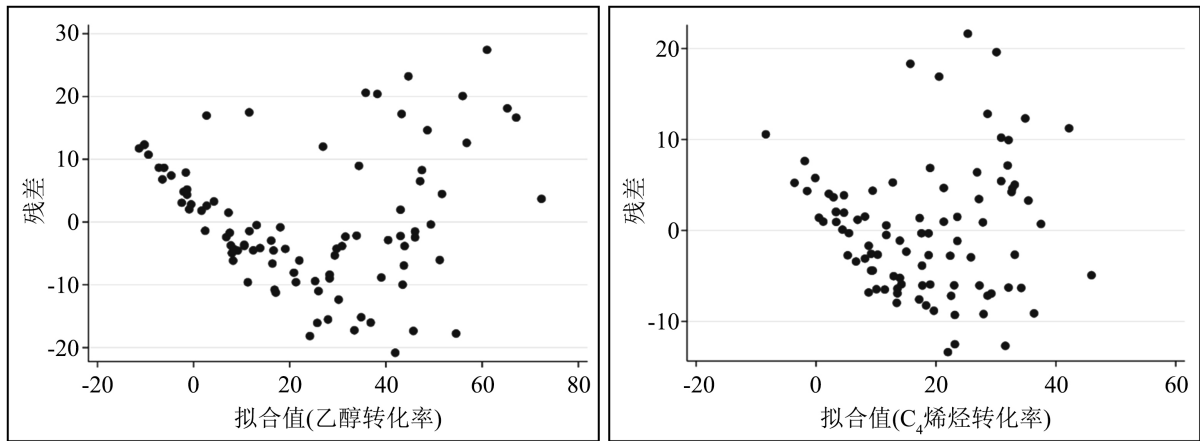


Figure 1. Residual and fitting values scatter plot
图 1. 残差与拟合值散点图

从图 1 中可以看出，扰动项存在异方差。对异方差进行怀特检验，证明了异方差确实存在。因此，在此基础上对模型进行改进，使用 OLS + 稳健的标准误再次进行回归，分别得到温度及催化剂组合对乙醇转化率、C₄ 烯烃选择性的回归方程：

1) 乙醇转化率：

$$y = 0.555x_1 - 7.183x_2 + 0.0514x_3 - 87.892$$

2) C₄ 烯烃选择性：

$$y = 0.192x_1 + 2.914x_2 + 3.248x_3 - 293.508x_4 + 0.0453x_5 - 55.499$$

2.3. 显著性检验

在进行结果分析之前，对两组回归结果分别进行检验[3]：

a) 模型 F 检验

Table 1. Test results

表 1. 检验结果

乙醇转化率		C ₄ 烯烃选择性	
F 检验统计量	P 值	F 检验统计量	P 值
52.99	0.0000	51.85	0.0000

由表 1 可知, $P = 0.0000 < 0.0001$, 上述模型通过了联合显著性检验, 且置信度达到 99% 以上。

b) 回归系数显著性检验

Table 2. Significance test of regression coefficient of ethanol conversion

表 2. 乙醇转化率回归系数显著性检验

	系数	t 检验	t 检验 p 值
温度	0.355 ^{***}	15.02	0.000
乙醇浓度	-7.183 ^{**}	-2.41	0.018
装料比	-2.103	-0.78	0.436
Co 负载量	93.939	0.91	0.368
催化剂总重	0.0514 ^{***}	4.90	0.000
常数	-87.892 ^{***}	-8.51	0.000

注: ^{***} $p < 0.01$, ^{**} $p < 0.05$, ^{*} $p < 0.1$ 。

如表 2, 其中装料比与 Co 负载量的偏相关系数没有通过显著性检验, 说明装料比和 Co 负载量对乙醇转化率线性相关不显著, 其余自变量均线性相关显著。

Table 3. Significance test of C₄ olefins selective regression coefficient表 3. C₄ 烯烃选择性回归系数显著性检验

	系数	t 检验	t 检验 p 值
温度	0.192 ^{***}	13.13	0.000
乙醇浓度	2.914 [*]	1.81	0.074
装料比	3.248 [*]	1.68	0.096
Co 负载量	-293.508 ^{***}	-3.46	0.001
催化剂总重	0.0453 ^{***}	5.60	0.000
常数	-55.499 ^{***}	-9.89	0.000

注: ^{***} $p < 0.01$, ^{**} $p < 0.05$, ^{*} $p < 0.1$ 。

如表 3，其中乙醇浓度和装料比与烯烃选择性在 90% 的置信水平上线性相关显著，其余在 99% 的置信水平上线性相关显著。

c) 多重共线性检验

考虑自变量之间是否有相关性，此时将对回归系数估计不准确，引入方差膨胀因子 VIF ，假设有 k 个自变量，那么对第 m 个自变量的方差膨胀因子定义为：

$$VIF_m = \frac{1}{1 - R_{1-k/m}^2}$$

其中 $R_{1-k/m}^2$ 是将第 m 个自变量作为因变量对剩下的 $k-1$ 个自变量进行回归得到的拟合优度。 VIF_m 越大，说明第 m 个变量和其他变量的相关性越大，利用该规则进行多重共线性检验：

Table 4. Multicollinearity test
表 4. 多重共线性检验

变量	VIF
温度	1.01
乙醇浓度	1.12
装料比	1.00
Co 负载量	1.10
催化剂总重	1.18

由表 4 可以看出，各自变量的 $VIF < 10$ ，通过多重共线性检验。

d) 回归结果的检验

利用检验组数据对模型的可行性进行验证，样本实验数据与拟合值如下图 2、图 3：

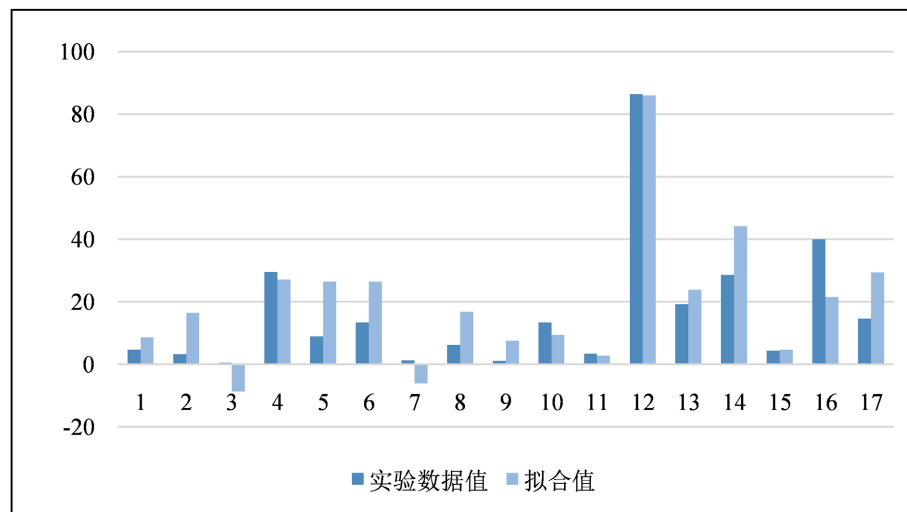


Figure 2. Ethanol conversion rate data contrast
图 2. 乙醇转化率数据对比

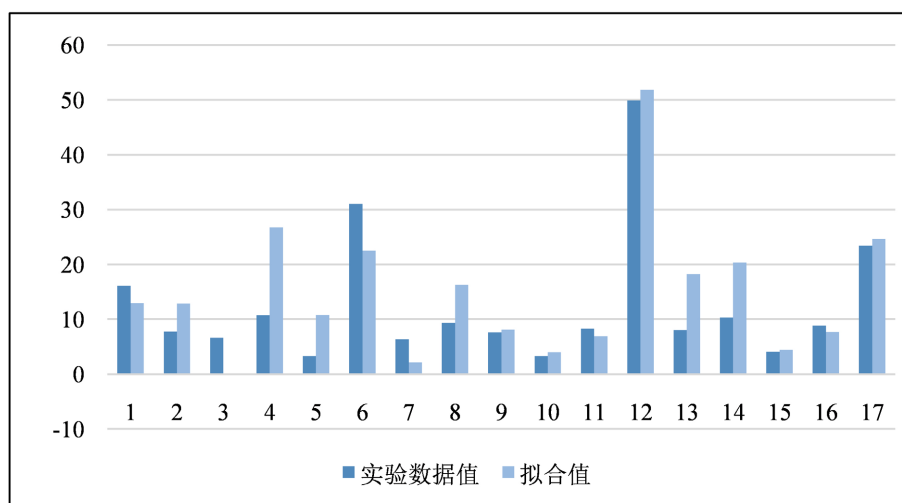


Figure 3. C₄ olefins selectivity data contrast
图 3. C₄ 烯烃选择性数据对比

利用 SPSS 进行 Wilcoxon 符号秩检验，由于实验数据较少且无法对分布形态做简单假定，因此使用非参数检验来确定多元线性回归模型的计算值与实际的实验数据是否有显著性差异。其中， H_0 ：两组数据值无显著性差异， H_1 ：两组数据值有显著性差异。

Table 5. Wilcoxon sign rank test results
表 5. Wilcoxon 符号秩检验结果

	乙醇转化率	C ₄ 烯烃选择性
P 值	0.227	0.246

由表 5 可知， $p > 0.05$ ，故接受原假设，认为通过模型计算的数据和原始实验数据没有显著性差异，得到的线性回归方程可信。

2.4. 结果分析

由上述回归结果可以剔除对 C₄ 烯烃收率影响不显著的自变量，为了更精准的研究显著自变量对因变量影响程度，计算出标准化回归系数，标准化系数的绝对值越大，说明对因变量的影响程度越大，根据计算结果，对乙醇转化率影响由高到低依次为温度 > 催化剂总重 > 乙醇浓度，对 C₄ 烯烃选择性影响由高到低依次为温度 > 催化剂总重 > Co 负载量 > 乙醇浓度。

3. 确定最佳反应条件

在上述模型结果的基础上，采用基于控制变量思想的逐步筛选法，将催化剂组合中的 5 个自变量对 C₄ 烯烃收率的影响分别分析。其次使用 MATLAB 工具箱对 C₄ 烯烃收率与单变量进行曲线拟合，并进行检验。最后，选用粒子群优化算法[4] [5]对逐步筛选出的数据求解最佳取值。

3.1. 模型建立

建立基于控制变量思想的逐步筛选法使 C₄ 烯烃收率最大化的模型。

控制变量法的思想是每一次只改变其中的某一个因素，而控制其余几个因素不变，从而研究被改变的这个因素对事物的影响，分别研究各因素，最后再综合解决，它是科学探究中的重要思想方法。

逐次筛选法的思想是按照自变量的重要程度，依次控制单变量对数据进行分析，从而逐步求解 C₄ 烯烃的最大收率。

模型过程如下图 4:

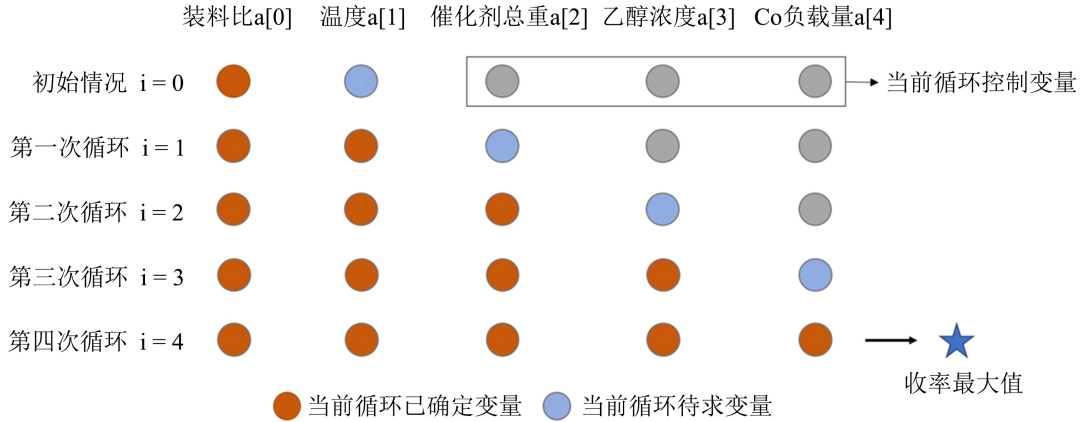


Figure 4. Model process diagram
图 4. 模型过程图

模型伪代码描述

```

1. Function<粒子群最优化算法> (a[i],ymax);
2. var I,y,ymax 数组 a[4]={装料比,温度,催化剂总重,乙醇浓度,Co 的负载量};
初始化 数组 a[0] = {装料比 1:1,0,0,0,0}; y = <粒子最优算法>(a[0]);
for i: =1 to 4
begin
if<a[i-1] > 0 then
begin
通过粒子群最优算法确定 使得 C4 烯烃收率最大的最优值 a[i];
y = <粒子最优算法> (a[i]); //当前使得 C4 烯烃的收率最大
end
else 则重新进行上一步骤的赋值;
ymax = <粒子最优算法> (a[4]);
end
输出装料比, 温度, 催化剂总重, 乙醇浓度, Co 的负载量均取到最优解时 C4 烯烃的最大值 ymax;
    
```

3.2. 模型求解及分析

依据上述思想确定最佳反应条件，阅读相关文献[1]了解到 Co/SiO₂ 和 HAP 装料比为 1:1 时，C₄ 烯烃的收率达到最高。因此，将其作为假设，并从实验数据中，挑选出仅装料比不同的数据进行绘图，并对假设进行检验。其 C₄ 烯烃收率的柱状图如图 5 所示。

从图 5 中可以看出，在其他 4 个因素保持不变的情况下，无论处于哪种温度条件下，Co/SiO₂ 和 HAP 装料比为 1:1 时，C₄ 烯烃收率均达到最大值，所以将 Co/SiO₂ 和 HAP 装料比设定为 1:1 作为初始情况。

装料比为 1:1 即为定值，控制催化剂总量为 400 mg、乙醇浓度为 1.68 ml/min、Co 负载量为 1 wt% 不变，将温度作为变量，筛选出新的数据，记为数据 a，为拟合 C₄ 烯烃收率与温度的函数提供数据。

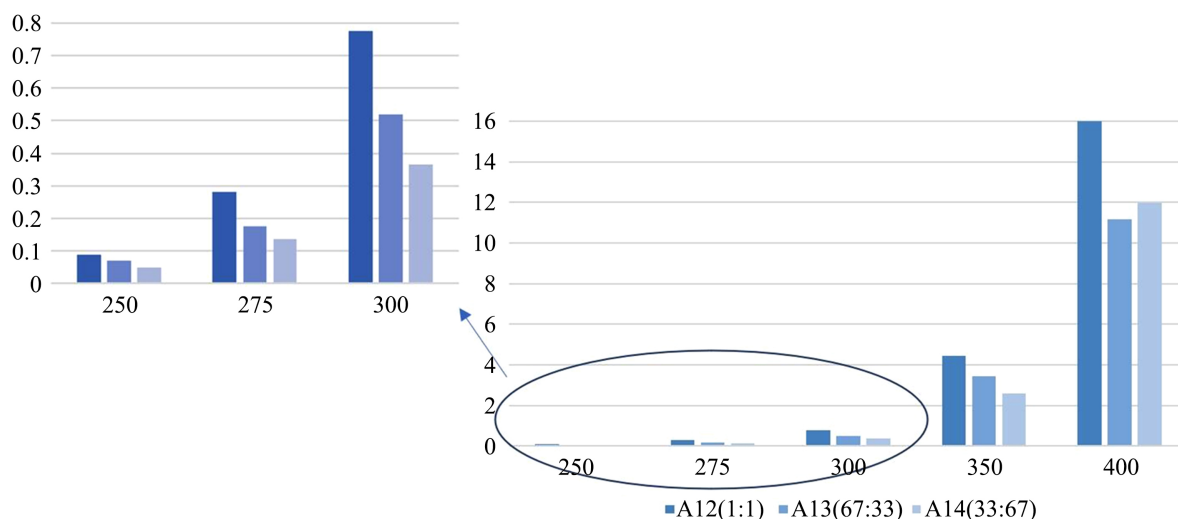


Figure 5. C₄ olefin yield of three charge ratios at different temperatures

图 5. 不同温度下三种装料比的 C₄ 烯烃收率

进而，在数据 a 的基础上，将计算得到的最优装料比 1:1 和温度 400 °C 作为定值，并保持乙醇浓度为 1.68 ml/min、Co 负载量为 1 wt% 不变，催化剂总重为变量，进一步筛选出符合的数据，记为数据 b，为拟合 C₄ 烯烃收率与催化剂总量的函数提供数据。

最后，由于满足条件的实验数据较少，仍将已经计算得到的最优装料比 1:1 和温度 400 °C 作为定值，并保持 Co 负载量为 1 wt%、催化剂总重为 100 mg 不变，乙醇浓度为变量，筛选出符合的数据，记为数据 c，为拟合 C₄ 烯烃收率与乙醇浓度的函数提供数据。

3.3. C₄ 烯烃收率与单变量的曲线拟合

采用 MATLAB Curve Fitting 工具箱对已处理的数据进行曲线拟合，并计算已调整后的 R 方进行曲线拟合效果的检验，为求解 C₄ 烯烃收率的最优值做准备。

- C₄ 烯烃的收率与温度的拟合函数：

$$Y = 17.04 * \exp\left(-\left(\frac{x-400}{8.389}\right)^2\right) + 4.724 * \exp\left(-\left(\frac{x-350}{7.481}\right)^2\right) + 0.8007 * \exp\left(-\left(\frac{x-300}{5.761}\right)^2\right), \quad R^2 = 0.9958$$

- C₄ 烯烃的收率与催化剂总重的拟合函数：

$$Y = 19.41 * \exp\left(-\left(\frac{x-176.1}{72.22}\right)^2\right) + 11.42 * \exp\left(-\left(\frac{x-86.41}{49.96}\right)^2\right), \quad R^2 = 0.9771$$

- C₄ 烯烃收率与乙醇浓度的拟合函数：

$$Y = 27.94 * \exp\left(-\left(\frac{x-0.3372}{0.7401}\right)^2\right) + 16.02 * \exp\left(-\left(\frac{x-1.877}{0.6187}\right)^2\right), \quad R^2 = 0.9207$$

- C₄ 烯烃收率与 Co 负载量关系的函数：

$$Y = 9.493 * \exp\left(-\left(\frac{x-0.02}{0.001747}\right)^2\right) + 5.33 * \exp\left(-\left(\frac{x-0.01}{0.002555}\right)^2\right) + 5.151 * \exp\left(-\left(\frac{x-0.005}{0.004494}\right)^2\right), \quad R^2 = 0.9283$$

3.4. 求解最佳温度与催化剂组合

利用粒子群优化算法，对 C₄ 烯烃收率与单变量拟合出的函数进行最优值求解，探讨出在相同实验条

件下能够使得 C₄ 烯烃收率尽可能高的催化剂组合和温度。

算法的步骤大致分为以下五步：

- ① 设置粒子速度的区间，对粒子群进行初始化。
- ② 计算粒子的适应值。
- ③ 全局搜索粒子个体的极值，并对当前最优解和历史最优解进行比较，选择出当前历史的最优解。

粒子更新迭代公式为如下：

$$v_i = v_i + c_1 \times rand() \times (pbest_i - x_i) + c_2 \times rand() \times (gbest_i - x_i)$$

$$x_i = x_i + v_i$$

- ④ 寻找全局的最优解。据图公式如下：

$$v_i = \omega \times v_i + c_1 \times rand() \times (pbest_i - x_i) + c_2 \times rand() \times (gbest_i - x_i)$$

$$\omega^{(t)} = (\omega_{ini} - \omega_{end}) \times (G_k - g) / G_k + \omega_{end}$$

- ⑤ 粒子之间进行比较并修改粒子的速度和位置，并设置使得粒子停止迭代的条件。

将 MATLAB 拟合并通过检验的函数代入粒子群优化算法中，可以计算得到最佳催化剂组合和温度，结果如图 6 所示：

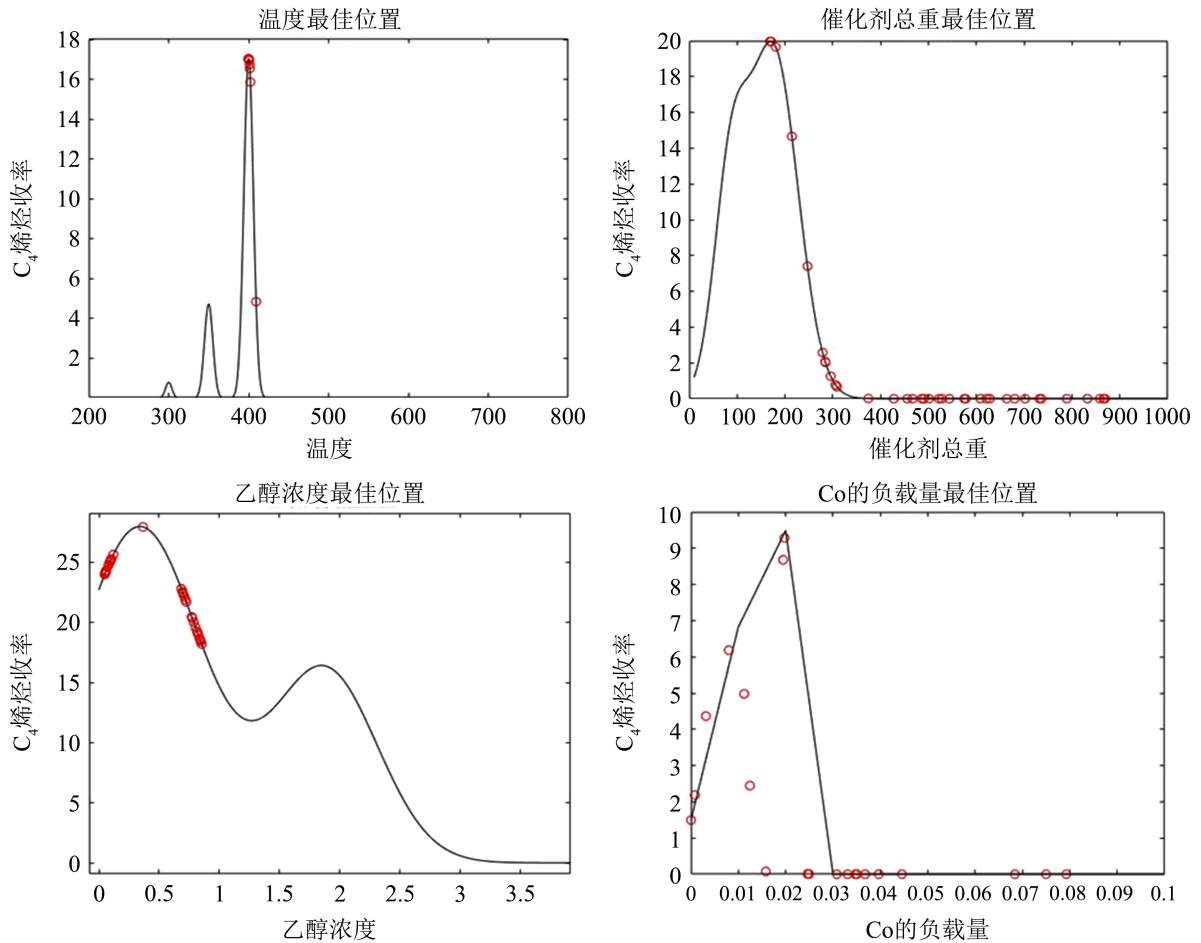


Figure 6. Solving results

图 6. 求解结果

结果显示, 当温度为 399.9999℃, 催化剂总重 169.6884 mg, 乙醇浓度为 0.33982 ml/min, Co 负载量为 0.020001, 即 2.0001 wt% 时时, C₄ 烯烃收率达到最高。综合以上, 最佳温度及催化剂组合的选择如表 6 所示。

Table 6. Optimal reaction conditions
表 6. 最佳反应条件

装料比	温度	催化剂总量	乙醇浓度	Co 负载量
1:1	400℃	169.6884 mg	0.33982 ml/min	2 wt%

4. 总结

在本文中, 为探究各因素对反应的影响, 将数据分为两组, 利用检验组对所求结果进行 WilCoxon 符号秩检验, 相比于用所有数据用于回归模型, 增设模型检验, 因而模型更具可信度。在求解最佳反应条件时, 提出基于控制变量思想的逐步筛选法, 改善了数据处理的方式, 对解决多变量问题进行简化。

本文所建立的基于 OLS + 稳健标准误的多元线性回归模型以及对于该模型效果的检验方法, 可适用于探究其他化学反应中反应条件对产物收率的影响, 并确定影响产物收率的关键因素。基于控制变量思想的逐步筛选模型可适用于实验数据较少时, 求解产物收率最优化问题。上述模型可推广至其他工业生产中, 具有一定的参考价值与经济意义。

参考文献

- [1] 吕绍沛. 乙醇偶合制备丁醇及 C₄ 烯烃[D]: [硕士学位论文]. 大连: 大连理工大学, 2018.
- [2] 赵明容, 刘平清, 杨凯, 海雪, 卢珊, 朱泓宇. 六盘水市旅游收入主要影响因素的实证分析——基于多元线性回归模型[J]. 山西农经, 2022(6): 23-25+71. <https://doi.org/10.16675/j.cnki.cn14-1065/f.2022.06.007>
- [3] 翟维. 基于多元线性回归的环境监测数据分析与研究[J]. 微型电脑应用, 2022, 38(3): 186-188+192.
- [4] 崔乃丹. 基于粒子群优化算法与 BP 神经网络的高铁客运量预测算法[J]. 自动化技术与应用, 2022, 41(4): 148-150.
- [5] 李兆北, 王印松, 苏杰. 基于非支配排序遗传算法和多目标粒子群算法的脱硫系统运行策略优化[J/OL]. 热力发电: 1-7, 2022-05-12. <https://doi.org/10.19666/j.rlfid.202112258>