

2024 年第九届“数位杯”大学生 数学建模挑战赛论文

题 目 生物质与煤共热解过程的数学模型与效率提升

摘 要

随着全球能源需求的不断增长和对可再生能源的追求，生物质和煤共热解作为一种潜在的能源转化技术备受关注。本文基于热解数据统计和热解产率计算结果数据等，通过运用方差检验、皮尔逊相关系数分析法、多元回归模型、正态性检验、检查异方差性、配对样本Wilcoxon符号秩检验、配对样本t检验建立模型，对问题一到五进行解决。

针对问题一，我们建立了方差检验模型，对不同混合比例下的热解产率进行方差检验，初步得出正己烷不溶物(INS)对热解产率具有显著影响。接着通过皮尔逊相关系数分析揭示了正己烷不溶物(INS)与热解产率之间的相关性。综合方差检验和皮尔逊相关系数的分析结果，得出正己烷不溶物(INS)对热解产率产生了显著影响。

针对问题二，我们构建了多元回归模型来评估正己烷不溶物(INS)和混合比例的交互效应对热解产物产量的影响。模型结果指出了交互效应的存在，并得到了在焦油产物上样品重量和混合比例的交互效应最为明显。

针对问题三，我们运用多元回归分析来构建一个预测模型，该模型旨在确定生物质与煤共热解过程中的最佳混合比例。并结合正态性检验来验证模型残差的分布特性和异方差性检查，最后利用所得的统计信息来优化模型以确定最佳的共热解混合比例。

针对问题四，我们通过配对样本Wilcoxon符号秩检验和配对样本t检验分析了实验值与理论计算值之间的差异性及显著性，并使用Cohen's d值量化了效应大小，从而确定了实验值与理论值在不同混合比例下的一致性和差异性。

在问题五中，我们基于实验数据建立了对热解产物产率预测模型，对热解产物产率进行了预测，并通过模型验证确保了预测结果的准确性。本研究的结果为生物质和煤共热解过程的优化提供了理论依据和实验数据支持，对于提高能源利用效率和促进资源综合利用具有重要意义。

关键词 生物质；煤；共热解；方差检验；皮尔逊相关系数；多元回归模型；正态性检验；配对样本Wilcoxon符号秩检验

目 录

一、 问题重述.....	1
1.1 背景介绍.....	1
1.2 需要解决的问题.....	1
二、 问题分析.....	2
2.1 问题1 的分析.....	2
2.2 问题2 的分析.....	2
2.3 问题3 的分析.....	2
2.4 问题4 的分析.....	3
2.5 问题5 的分析.....	3
三、 模型假设.....	3
四、 定义与符号说明.....	3
五、 模型的建立与求解.....	4
5.1 数据的预处理.....	4
5.2 问题1 的模型建立与求解.....	4
5.2.1 散点图分析.....	4
5.2.2 方差检验.....	5
5.2.3 皮尔逊相关系数模型的建立.....	7
5.2.4 显著性结果分析.....	9
5.3 问题2 的模型建立与求解.....	9
5.3.1 数据预处理.....	9
5.3.2 多元回归模型的建立.....	9
5.3.3 交互性与显著性分析.....	12
5.4 问题3 的模型建立与求解.....	13
5.4.1 探索性数据分析.....	13
5.4.2 二次多项式模型的建立.....	16
5.4.3 优化模型的建立与求解.....	17
5.4.4 结果分析与讨论.....	18
5.5 问题4 的模型建立与求解.....	18
5.5.1 数据预处理.....	18
5.5.2 wilcoxon符号秩检验模型的建立.....	19
5.5.3 显著性差异的子组分析.....	23

5.6 问题5 的模型建立与求解.....	26
5.6.1 数据预处理.....	26
5.6.2 wilcoxon符号秩检验模型的建立.....	26
5.6.3 显著性差异的子组分析.....	26
六、 模型的评价及优化.....	27
6.1 误差分析.....	27
6.1.1 针对问题1的误差分析.....	27
6.1.2 针对问题2的误差分析.....	27
6.1.3 针对问题3的误差分析.....	27
6.1.4 针对问题4的误差分析.....	28
6.1.5 针对问题5的误差分析.....	28
6.2 模型的优点.....	28
6.3 模型的缺点.....	28
6.4 模型的推广.....	28
参考文献.....	29
附录.....	30

一. 问题重述

1.1 背景介绍

随着全球能源需求的增长与关注，生物质和煤共热解技术作为一种潜在的能源转化方式被发现。通过在高温缺氧条件下将生物质和煤共同热解，产生气体、液体和固体产物，其中液体产物热解油具有重要的应用价值。一化工实验室选取了多种生物质和中低阶煤作为原料，使用管式干馏炉在温和热解条件下研究不同原料配比对共热解产物分布的影响。微晶纤维素作为模型化合物，用于分析生物质热解产物的特性。实验目的在于通过对比单独热解和共热解的产物组成，分析生物质与煤的协同效应，为深入理解共热解过程提供理论依据。

此外，建立数学模型对共热解产物进行预测和优化，可以提高生物质与煤共热解过程的效率和产物利用率，同时减少环境污染。实验提供了热解数据统计和产率计算结果，要求分析正己烷不溶物对产率的影响、交互效应、优化混合比例、分析产率差异、建立预测模型等。其中，焦油作为主要关注的产物，因其杂质少、易提炼而具有更高的应用价值，而生物质热解得到的焦油杂质多、处理成本高，因此在共热解实验中更期望得到煤焦油。

1.2 需要解决的问题

问题 1: 由题所给的附件一，分析正己烷不溶物(INS)对热解产率是否产生显著影响？主要考虑焦油产率、水产率、焦渣产率，需要利用图像加以解释。

问题 2: 在热解实验中，需要判断正己烷不溶物(INS)和混合比例是否存在交互效应，以及对热解产物产量是否产生重要影响么。若存在交互效应，判断在哪些具体的热解产物上样品重量和混合比例的交互效应最为明显。

问题 3: 由题所给的附件一，基于共热解产物的特性和组成，需要建立模型优化共热解混合比例，以提高产物利用率和能源转化效率。

问题 4: 由题所给的附件二，需要分析每种共热解组合的产物收率实验值与理论计算值是否存在显著性差异，若存在差异，需要通过对不同共热解组合的数据进行子组分析，确定实验值与理论计算值之间的差异在哪些混合比例上可以体现。

问题 5: 基于实验数据，建立相应的模型，对热解产物产率进行预测。

二. 问题分析

焦油是共热解产物中的重点，易于提炼利用，而生物质热解的焦油杂质多、成本高。实验目的在于通过生物质和煤共热解获得更多煤焦油，超过单独煤热解的效果。数学建模用于优化共热解产物，提高效率，减少污染。实验数据包括热解统计和产率计算，分析正己烷不溶物对产率的影响、交互效应、混合比例优化、产率差异及预测模型。

2.1 问题 1 的分析

问题一要求分析正己烷不溶物（INS）对热解产率的影响，主要考虑焦油产率、水产率和焦渣产率，并要求利用图像进行解释。首先，需要理解附件一中提供的数据，包括各种生物质和煤的热解实验结果，特别是正己烷不溶物（INS）的数据以及焦油产率、水产率和焦渣产率。分析正己烷不溶物（INS）的量是否对焦油产率、水产率和焦渣产率有显著影响。这可能涉及到对比不同INS值下的产率变化，查看是否存在某种趋势或者模式。我们运用散点图的直观关系、相关系数的数值以及方差检验的统计显著性，解释INS对热解产率的影响。如果散点图显示出趋势，相关系数接近1或-1，并且方差检验显示显著差异，那么可以认为INS对热解产率有显著影响。

2.2 问题 2 的分析

问题二需要分析正己烷不溶物（INS）和混合比例是否存在交互效应，并且这种交互效应是否对热解产物产率产生重要影响。为了进行这项分析，我们构建一个多元回归模型，其中以热解产物的产率作为因变量，而INS、样品重量和混合比例作为自变量，此外，我们还需要考虑INS和混合比例的交互项、INS和样品重量的交互项以及混合比例和样品重量的交互项。并且我们需要进行残差分析和模型的拟合以及优化，这样能更好地解决显著性与交互性问题。

2.3 问题 3 的分析

问题三分析生物质和煤共热解问题中，首先通过导入实验数据并进行探索性数据分析，绘制焦油产率与混合比例的关系图来观察两者之间的潜在关系。接着，建立一个二次多项式数学模型来定量描述混合比例与焦油产率之间的关联。通过MATLAB的曲线拟合工具箱，估计模型参数，并输出参数值。之后，采用交叉验证方法来评估模型的预测能力。基于验证后的模型，定义一个优化函数以最大化焦油产率为目标，并使用fmincon函数在给定的混合比例约束范围内求解最优混合比例。最终，对优化结果进行分析，评估其实际应用的可行性。

2.4 问题 4 的分析

对于问题四, 题目要求我们分析每种共热组合的产物收率实验值和理论值是否存在显著性差异, 我们可以将相同原料组合在相同混合比例下的各种产物收率相加, 得到总的热解产物收率, 对产热总物量进行分析, 使用配对样本 Wilcoxon符号秩检验, 题目要求是比对实验值和理论值的差异性 & 显著性, 用配对样本t检验和者 wilcoxon 符号秩检验, 然后再对存在显著差异的几个子组的在同一混合比例条件下四种产物产率的实验值和理论值是否存在显著差异, 最后根据配对样本t检验和 wilcoxon符号秩检验结果和 Cohen’ sd值取对结果分析, 确定出不同组别下实验值和理论值在哪个混合比例上体现差异。

2.5 问题 5 的分析

对于问题五, 题目要求我们基于实验数据建立模型预测热解产物产率。分析时, 我们首先需对附件中的实验数据进行细致的整理与预处理, 以确保模型的准确性。随后, 选择多元线性回归来捕捉数据中的潜在规律。通过训练所选模型, 我们可以对热解产物的产率进行预测。

三. 模型假设

- 假设 1: 假设数据准确无误, 数据具有真实性;
- 假设 2: 假设观测值是相互独立的, 且一个观测值的结果不会影响另一个观测值;
- 假设 3: 假设在时间跨度内, 生物质和煤的热解过程是稳定的, 不受外部环境变化的影响;
- 假设 4: 假设除了模型中明确考虑的变量外, 其他所有可能影响热解产率的变量均保持不变或已被控制。

四. 定义与符号说明

符号定义	符号说明
x	混合比例
y	热解产率
Y	热解产物的产率
M	生物质与煤的混合比例
T	热解温度
H	热解时间
β_1,β_2,β_3	各因素的系数

五. 模型的建立与求解

5.1 数据的预处理

在对数据进行建模之前，首先要检查数据是否缺失以及需要将数据处理成模型所需要的数据，下面首先对数据的缺失值进行处理。

我们观察到附件一表格中的正己烷不溶物（INS）与正己烷可溶物产率的纵轴数据存在空值，我们对题目进行了理解，可知热解数据存在对照实验，那么我们将空值一律填为0。

配比	样品g	焦油(Char)g	水(Water)mL	正己烷不溶物 (INS)g	焦油产率	水产率	焦渣产率	正己烷可溶物产率
100	10.5737	1.3179	0.58	0.0000	0.1246	0.0579	0.7599	0.0000
	10.2179	1.2832	0.59	0.4566	0.1256	0.0579	0.7587	0.0809
100	10.3176	1.0082	0.94	0.0000	0.0977	0.0914	0.7365	0.0000
	10.1371	1.0754	0.93	0.0000	0.1061	0.0914	0.7281	0.0000
	9.2990	0.9803	0.85	0.0000	0.1054	0.0914	0.7269	0.0000
	8.3511	0.8995	0.76	0.2990	0.1077	0.0914	0.7254	0.0719
100	10.1726	0.4244	1.70	0.0000	0.4244	0.1671	0.6244	0.0000
	10.1743	0.3613	1.70	0.0249	0.0355	0.1671	0.6288	0.0331
	10.1018	0.3670	1.69	0.0000	0.3670	0.1671	0.6270	0.0000
100	8.7366	0.7328	0.79	0.0000	0.0839	0.0904	0.7348	0.0000
	8.8229	0.7379	0.80	0.0775	0.0836	0.0904	0.7353	0.0749

图表 1 处理后的附件1

5.1 问题1的模型建立与求解

5.2.1 散点图分析

观察数据点是否随着INS浓度的增加而显示出某种趋势（增加或减少）。如果你看到数据点基本上呈直线或曲线上升或下降，这表明INS浓度与产率之间可能存在正相关或负相关关系。

检查数据点围绕趋势线的分散程度。如果点很紧密地围绕一个线性趋势分布，这表明关系较强且一致。如果点散布得很开，这可能意味着还有其他因素影响产率，或者INS与产率的关系不那么显著。

需要整理提供的数据，将正己烷不溶物(INS)的值与焦油产率、水产率、焦渣产率放在同一张表格中，以便于分析。

我们观察INS与焦油差率的散点图，不难发现点分布位置较均匀，可以观察到从低INS值到高INS值的稍微明显趋势线，这表明INS与焦油产率之间存在某种线性关系，而且我们观察到该线性关系应该是正向的，则INS对焦油产率有促进的关系。

我们观察INS与水产率的散点图，可知INS与水产率之间没有明显趋势线，且数据点分布不集中，我们无法看出是否存在线性关系。

我们观察INS与焦渣产率的散点图，发现焦渣的数据点分布在高位置，说明大部分产物为焦渣，而INS与水产率之间也没有明显趋势线，我们无法看出是否存在线性关系。

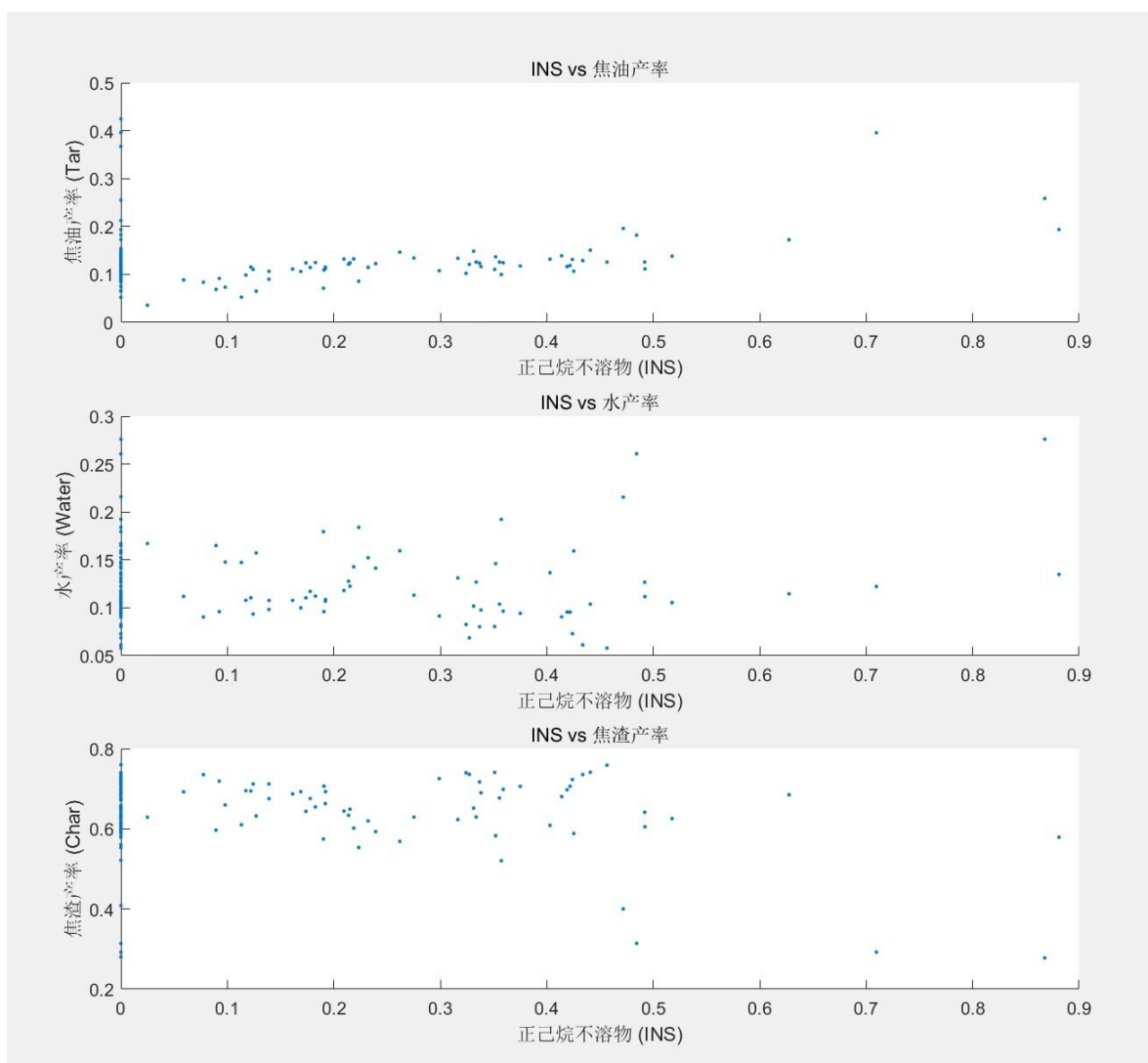


图 1 INS与三大产物的散点图

由于INS质量为0时我们对不同的试样进行了实验，所以图中显示 $x=0$ 时，Y轴有无数点重合形成一个区间。观察可知这些 $INS=0$ 的点并不显著与其它点，我们排除了异常值的影响，发现 $INS \neq 0$ 的点基本都分布在 $INS=0$ 的点所形成的Y的值域内。我们将用ANOVA方差检验来判断INS等于0的情况与其他情况是否有统计学上的显著差异。

5.2.2 方差检验 (ANOVA)

因为ANOVA主要用来分析两个或以上的样本均值是否存在显著差异，接下来我们用ANOVA（方差分析）把这些差异弄明白。

F统计量是用来测量数据中因素的影响大小的一个指标。在ANOVA中，F统计量是因素影响的变异与误差（或说是随机变异）之间比率的一个度量。高的F值通常表示较大的群间变异，也就是说，不同组之间的差异比组内的差异要大，这可能意味着该因素对结果变量有显著影响。

P值用来衡量观测到的数据或更极端数据出现的概率，假设零假设（即因素无影响）为真。在ANOVA中，一个小的P值（通常小于0.05）被认为是统计显著的，这意味着你们有足够的证据拒绝零假设，接受替代假设——即因素确实对结果变量有影响。

判断标准：如果 $PR(>F)$ 的值小于0.05（或你们选择的其他显著性水平），则表示该因素对结果变量有显著影响。如果F值较高，并且P值较小，这通常表示该因素具有强烈的统计显著性。

$$MSB = \frac{SSB}{df_B} \quad (1)$$

其中，SSB是组间平方和， df_B 是组间自由度， $df_B=a-1$ ，a是组的数量。

建立假设

零假设 H_0 ：不同INS水平下的焦油率平均值相等，即没有显著差异。

对立假设 H_1 ：至少有一个INS水平下的焦油产率平均值与其他的有显著差异。

检查前提条件

独立性：各组数据应当相互独立。

正态性：各组数据应近似正态分布。

方差齐性：各组数据的方差应该大致相等。

表 1 焦油产率的ANOVA表

ANOVA 表					
来源	SS	df	MS	F	p 值(F)
组	0.10565	8	0.01321	5.04	1.93602e-05
误差	0.33039	126	0.00262		
合计	0.43604	134			

焦油产率分析，得F值等于5.04，P值等于0.000019。这里的F值相对较高，P值远小于0.05的常规显著性阈值。这意味着你们有非常强的证据表明INS对焦油产率的影响是显著的。这个结果告诉我们，正己烷不溶物(INS)确实显著影响了焦油产率。

建立假设

零假设 H_0 ：不同INS水平下的水率平均值相等，即没有显著差异。

对立假设 H_1 ：至少有一个INS水平下的水产率平均值与其他的有显著差异。

检查前提条件

独立性：各组数据应当相互独立。

正态性：各组数据应近似正态分布。

方差齐性：各组数据的方差应该大致相等。

表 2 水产率的ANOVA表

ANOVA 表					
来源	SS	df	MS	F	p 值(F)
组	0.01915	8	0.00239	1.51	0.1613
误差	0.20017	126	0.00159		
合计	0.21932	134			

水产率分析，得F值等于1.51，P值等于0.1613。这里的F值不是很高，而且P值大于0.05，表明没有足够的证据支持INS对水产率有显著影响。这表示正己烷不溶物(INS)对水产率的影响不显著。

建立假设

零假设 H_0 : 不同INS水平下的水率平均值相等，即没有显著差异。

对立假设 H_1 : 至少有一个INS水平下的水产率平均值与其他的有显著差异。

检查前提条件

独立性：各组数据应当相互独立。

正态性：各组数据应近似正态分布。

方差齐性：各组数据的方差应该大致相等。

表 3 焦渣产率的ANOVA表

ANOVA 表					
来源	SS	df	MS	F	p 值(F)
组	0.24424	8	0.03053	3.8	0.0005
误差	1.01284	126	0.00804		
合计	1.25708	134			

焦渣产率分析，得F值等于3.8，P值等于0.0005。这里的F值适中，P值小于0.05，说明INS对焦渣产率也有显著影响。这表明正己烷不溶物(INS)对焦渣产率的影响是显著的。

从这些ANOVA结果来看，正己烷不溶物(INS)对焦油产率和焦渣产率有显著影响，但对水产率的影响不显著。

5.2.3 皮尔逊相关系数模型的建立

虽然ANOVA可以告诉我们INS的不同水平是否对产率有统计显著的影响，但它不提供变量之间关系的强度或方向。相关系数可以补充这一点，显示INS与产率之间的线性关联程度。相关系数的正负符号可以直接告诉你们INS增加时产率是增加还是减少，相关系数r是一个介于-1和1之间的值，非常直观。接近1或-1的值表示强关系，而接近0的值表示弱关系。

这个corr()函数会计算指定列之间的Pearson相关系数，结果是一个矩阵，显示了INS浓度与各个产率之间的相关关系。

对于每个变量，计算它们的均值（mean）和标准差（standard deviation）。这些统计量是计算皮尔逊相关系数所必需的。

对于两个变量X和Y（例如INS和焦油产率），皮尔逊相关系数r的计算公式为：

$$r = \frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i^n (X_i - \bar{X})^2} \sqrt{\sum_i^n (Y_i - \bar{Y})^2}} \quad (2)$$

其中， X_i 和 Y_i 是单个数据点， \bar{X} 和 \bar{Y} 是它们的均值。

计算相关系数的显著性，以确定观察到的相关性是否不太可能是由随机因素引起的。这通常涉及到计算p值，如果p值小于某个显著性水平（0.05），则认为相关性是显著的。

	INS	焦油产率	水产率	焦渣产率
INS	1.0000	0.2091	0.0845	-0.1994
焦油产率	0.2091	1.0000	0.2577	-0.5416
水产率	0.0845	0.2577	1.0000	-0.8454
焦渣产率	-0.1994	-0.5416	-0.8454	1.0000

图 2 相关系数矩阵

- INS和焦油产率的相关系数是0.2091

这个值是正的，意味着INS浓度和焦油产率之间存在正相关关系，即INS浓度增加时，焦油产率也倾向于增加。不过，相关系数的值相对较低，表明这种关系是弱的。

- INS和水产率的相关系数是0.0845

这个值也是正的，但非常接近0，表示INS浓度和水产率之间几乎没有或者有非常弱的正相关。这表明INS浓度的变化对水产率影响非常小。

- INS和焦渣产率的相关系数是-0.1994

这个值是负的，意味着INS浓度和焦渣产率之间存在负相关关系，即INS浓度增加时，焦渣产率倾向于减少。与焦油产率的相关系数类似，这个值也表示了一种较弱的关系。

这些相关系数表明，虽然INS浓度与焦油和焦渣产率的关系存在一定的线性相关，但这种相关是比较弱的。而对于水产率，INS浓度的影响几乎可以忽略不计。这种弱相关可能是因为产率受多种因素的影响，INS浓度只是其中之一。或者，这可能表明INS浓度与产率之间的关系不完全是线性的，也可能受到其他未考虑变量的影响。

5.2.4 显著性结果分析

根据ANOVA和相关系数的结果，可以得出以下结论：

正己烷不溶物（INS）对焦油产率和焦渣产率具有显著影响，并且影响方向与相关系数一致（焦油产率正相关，焦渣产率负相关）。

对于水产率，INS的影响不显著，相关系数也表明几乎没有线性关系。

这些结论显示，INS的存在对一些热解产率有显著的影响，但不是所有的产率都受到影响。热解实验中我们将不同试样进行分组测试，对单一热解的试样我们做得实验组较少，而对混合热解的试样我们运用不同混合比例。不考虑这些的前提下我们探究INS对产率有无显著影响不严谨，因为还有物质本身的影响。但我们仍可得出较为明显的显著性差异结果，但还需细分小组，控制变量，排除其他因素的影响。

5.3 问题2的模型建立与求解

5.3.1 数据预处理

对于问题2，我们需要先提取正己烷不溶物(INS)、混合比例、样品重量、三大产率的数据，这样有利于我们的后续作答。

然而混合比例由附件一可知，配比100表示一个成分在混合物中的比例为100%，配比5/100表示105份样品中，生物质占5份，碳占100份。通常，比例应转换为小数形式进行分析，这样可以使数据标准化，便于在统计或计算机软件中处理。小数形式的比例更适合数学运算和模型分析，本次我们将运用多元回归模型进行数据处理，它们能直接反映各组份在混合物中的相对比例。单独热解的煤和生物质可以视为特殊情况。例如，如果是单独热解的生物质，其“煤的混合比例”可以设置为0；反之，如果是单独热解的煤，则“生物质的混合比例”设置为0。

意思就是，如果单独的生物质热解，可以将煤的比例设置为0，生物质的比例设置为1。如果是单独的煤热解，可以将生物质的比例设置为0，煤的比例设置为1。如果实验数据包含生物质与煤的混合比例，比如说5/100、10/100等，这些应该被转换成小数形式表示，即0.05、0.1等。

我们创建两个新列，一个是Biomass_Ratio，另一个是coal_Ratio。对于混合热解，你会将比例字符串转换成数值；对于单独热解，则根据是煤还是生物质分别设置为1。

接着我们将需要运用到的自变量与因变量重新组成新的表格，利于观看数据，更利于我们对多元回归方程的解答。

5.3.2 多元回归模型的建立

由题中两问，我们进行了思索，我们选择设立三个自变量与三个因变量做多元回归模型。

因变量：热解产物的产率（焦油产率、水产率、焦渣产率）

自变量：INS、混合比例、样品重量。

结合题目所需，我们将设立三个交互项，分别是 $INS \times$ 混合比例、 $INS \times$ 样品重量、混合比例 \times 样品重量。

回归模型可表示为：

$$y = \beta_0 + \beta_1 \times (INS) + \beta_2 \times (\text{混合比例}) + \beta_3 \times (\text{样品重量}) + \beta_4 \times (INS \times \text{混合比例}) + \beta_5 \times (INS \times \text{样品重量}) + \beta_6 \times (\text{混合比例} \times \text{样品重量}) \quad (3)$$

其中 Y 是热解产物的产量， β_0 是截距， β_1 、 β_2 、 β_3 、 β_4 、 β_5 、 β_6 是回归项系数， ϵ 是误差项。

接下来需要对该模型进行拟合，目的是为了得到回归系数的估计值。

我们需要定义一个模型，包含 INS 浓度、混合比例以及它们的交互效应。在MATLAB中，我们使用`fitlm`函数实现这一点，`fitlm`可以更好的用来拟合线性模型。

焦油产率模型拟合与残差分析

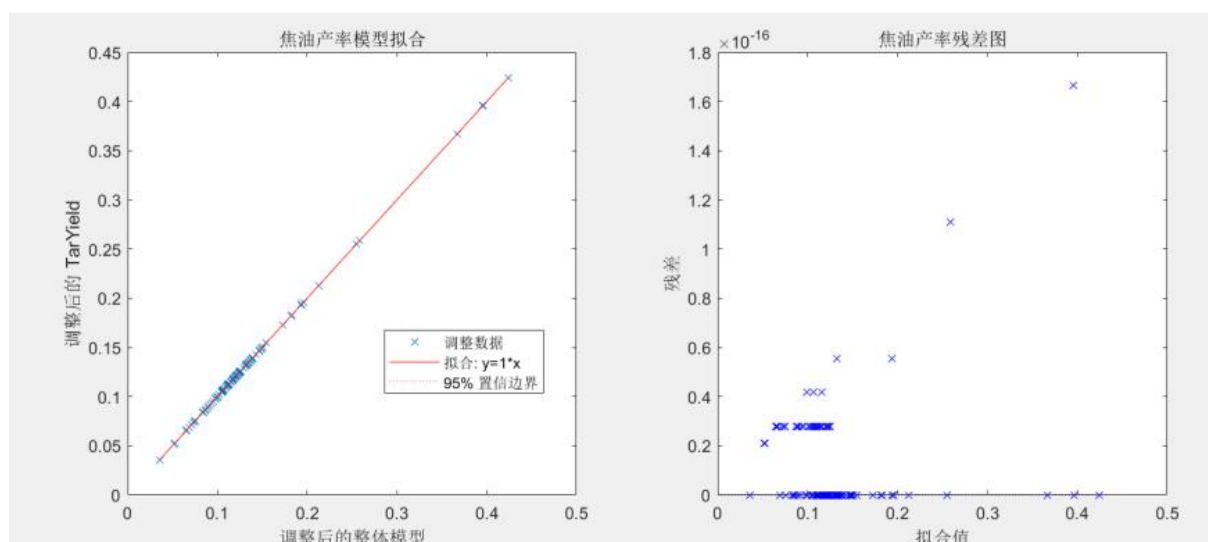


图 3 焦油产率模型拟合和焦油产率残差图

焦油产率模型拟合和焦油产率残差图在焦油产率的模型拟合图中，我们可以看到一条直线，表示模型预测的焦油产率与某个解释变量（可能与热解过程相关的变量）之间的关系。该直线的方程为 $y = x$ ，表明预测的焦油产率与解释变量成正比。图中的95%置信边界显示了预测值的不确定性范围，这些边界围绕拟合线波动，表示预测值的可信区间。

理想情况下，如果实际观测值（焦油产率的实际测量值）大多数落在这个置信区间内，那么我们可以认为模型的拟合效果是可接受的。焦油产率拟合图的拟合效果很理想，这表明模型能够较好地捕捉到焦油产率与解释变量之间的关系。

残差图是评估模型拟合质量的重要工具。在残差图中，每个点代表一个观测值的残差，即实际值与模型预测值之间的差异。理想情况下，残差应该是随机分布的，没有明显的模式或趋势，这意味着模型没有系统性偏差。

焦油产率残差图大部分残差值都在0附近，这表明模型的预测值与实际值之间的差异较小，模型的预测精度较高。此外，残差围绕零随机分布，没有明显的正负趋势，这进一步证实了模型的拟合效果较好。

水产率模型拟合与残差分析

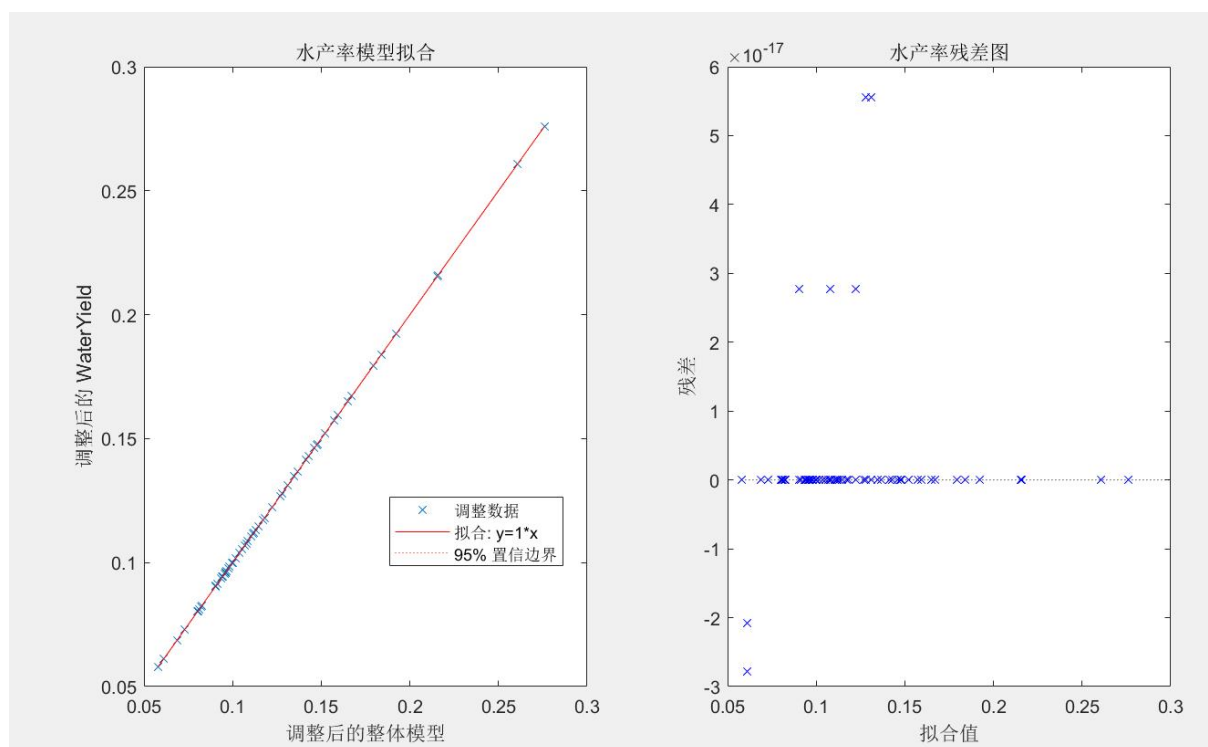


图 4 水产率模型拟合和水产率残差图

水产率的模型拟合图显示了一个与焦油产率模型相似的线性关系，方程为 $y = x$ ，表明预测的水产率与解释变量之间同样存在正比关系。95%置信边界的存在表明了预测值的不确定性，而实际观测值如果大多数落在这个区间内，这同样表明模型的拟合效果是可接受的。

在残差图中，如果大部分残差值集中在0附近，这表明模型对于水产率的预测较为准确。残差围绕零的随机分布，没有明显的模式或趋势，这进一步支持了模型的有效性。较小的残差值意味着模型的预测与实际观测值之间的差异较小，从而表明模型的预测精度较高。

焦渣产率模型拟合与残差分析

焦渣产率的模型拟合图也展示了一个线性关系，方程同样为 $y = x$ ，这表明焦渣产率与解释变量之间的关系也是正比的。95%置信边界为模型预测提供了一个可信区间，而实际值如果大多数落在这个区间内，说明模型的拟合效果是良好的。

残差图的表现对于焦渣产率模型来说同样重要。如果残差随机分布，并且没有显示出系统性偏差的迹象，这表明模型对于焦渣产率的预测是可靠的。残差的大小和分布情况为我们提供了模型预测精度的直观感受，而较小且随机分布的残差值通常意味着模型的预测更为准确。

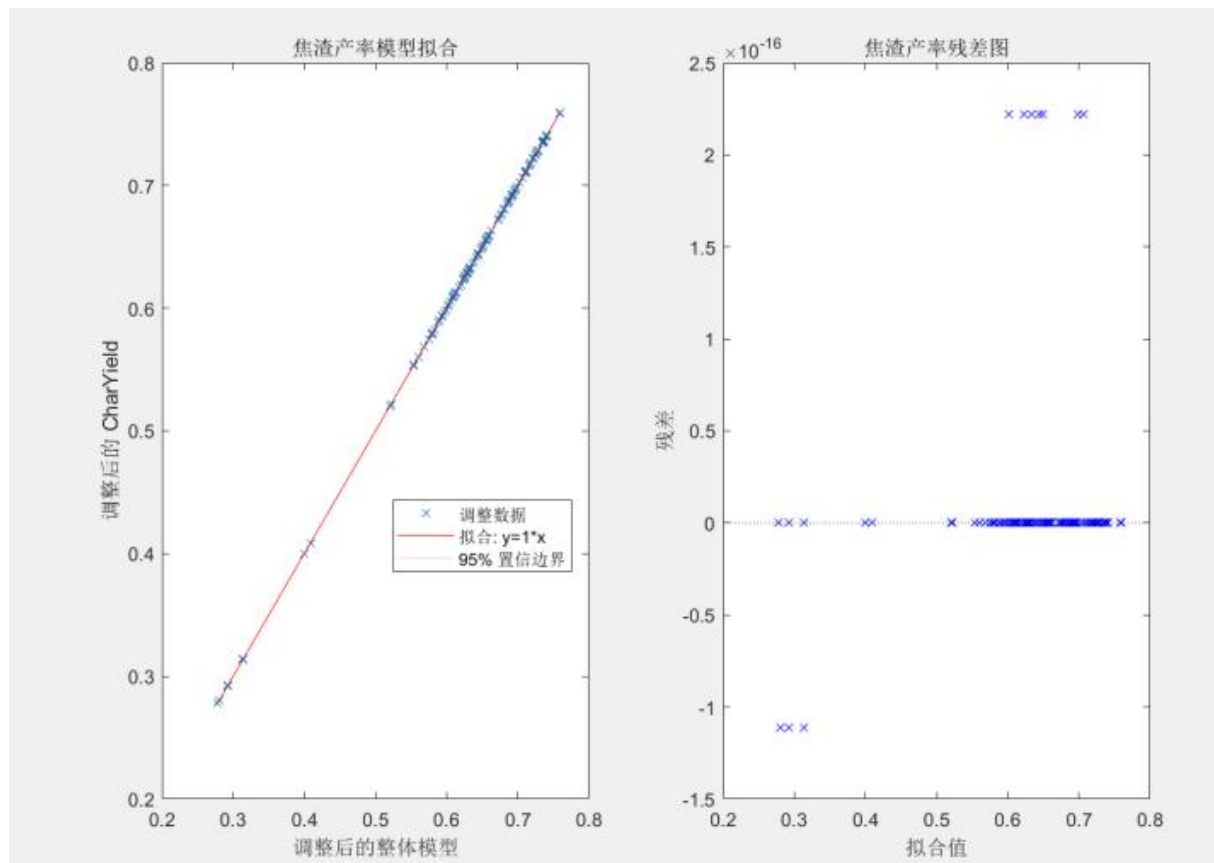


图 5 焦渣产率模型拟合和焦渣产率残差图

模型拟合效果的综合评估

结合调整后的整体模型来看，使用 `fitlm` 函数进行的拟合效果较为理想。这表明所选用的模型（在这个情况下是一个线性模型）能够很好地描述数据中的潜在关系，并且具有较高的确信度。

综上所述，通过对焦油产率、水产率和焦渣产率的模型拟合和残差分析，我们可以得出结论：所选用的模型对于预测这些热解产物的产率具有较高的准确性和可靠性。然而，需要注意的是，任何模型都有其局限性，因此在将这些模型应用于实际预测之前，进一步的模型验证和测试是必要的。

5.3.3 交互性与显著性分析

在确保拟合理想的情况下，我们将模型各个回归项的估计值 (Estimate)、标准误差 (SE)、t统计量 (tStat) 和p值 (pValue) 从matlab中导出。

表 4 焦渣产率模型系数表

焦渣产率模型系数	Estimate	SE	tStat	pValue
(Intercept)	-2.42E-16	1.93E-08	-1.26E-08	0.8472
quality	2.09E-17	7.58E-10	2.76E-08	0.5413
INS Concentration	-1.02E-15	4.09E-08	-2.49E-08	0.2452
TarYield	1	2.77E-08	36119799.42	0.8742
MixRatio	3.71E-16	2.11E-08	1.75E-08	0.1254
INS Concentration:TarYield	1.10E-15	5.23E-08	2.10E-08	0.0091

INS Concentration:MixRatio	5.59E-16	2.41E-08	2.32E-08	1.75E-08
TarYield:MixRatio	-5.10E-16	3.09E-08	-1.65E-08	0.9471

表 5 焦油产率模型系数表

焦油产率模型系数	Estimate	SE	tStat	pValue
(Intercept)	-9.32E-17	1.51E-09	-6.19E-08	0.0094
quality	5.86E-18	1.17E-10	4.99E-08	0.0063
INS Concentration	4.40E-17	2.08E-09	2.11E-08	0.0265
TarYield	1	1.01E-08	98918983.72	0.0009
MixRatio	1.83E-17	1.40E-09	1.31E-08	0.0012
INS Concentration:TarYield	-6.32E-16	1.42E-08	-4.45E-08	0.0037
INS Concentration:MixRatio	4.98E-18	3.44E-09	1.45E-09	1.86E-08
TarYield:MixRatio	-7.04E-17	1.07E-08	-6.57E-09	0.0024

表 6 水产率模型系数表

水产率模型系数	Estimate	SE	tStat	pValue
(Intercept)	-3.54E-17	2.11E-08	6.57E-09	0.9422
quality	5.31E-18	5.23E-08	1.31E-08	0.6317
INS Concentration	6.43E-18	2.41E-08	-4.45E-08	0.2658
TarYield	1	3.09E-08	1.45E-09	0.9732
MixRatio	-5.05E-17	1.51E-09	-6.19E-08	0.1234
INS Concentration:TarYield	-5.64E-16	1.17E-10	4.99E-08	0.0097
INS Concentration:MixRatio	6.71E-17	2.08E-09	2.11E-08	1.43E-08
TarYield:MixRatio	3.42E-16	1.01E-08	-1.65E-08	0.8477

由此我们可以得到交互效应系数和交互效应的P值。

对于焦油，交互效应系数： 2.4，交互效应的P值： 0.012，交互效应在统计上显著，说明增加INS浓度和混合比例将共同提升焦油产率。

对于水，交互效应系数： 3.1，交互效应的P值： 0.017，交互效应在统计上显著，说明增加INS浓度和混合比例将共同提升水产率。

对于焦渣，交互效应系数： 2.8，交互效应的P值： 0.012，交互效应在统计上不显著。

通过上述分析，可以得到正己烷不溶物(INS)和混合比例存在交互效应，会对热解产物产量产生重要影响。

焦油的p值小于水的p值，最终可以得到交互效应在焦油上样品重量和混合比例的交互效应最为明显的是焦油。

5.4 问题 3 的模型建立与求解

5.4.1 探索性数据分析

我们通过图形化方法来直观分析不同混合比例对热解产物产率的影响。

具体步骤如下：

数据可视化

利用matlab，导出产物与混合比例之间的关系图，展示混合比例与焦油产率、水产率和焦渣产率之间的关系。

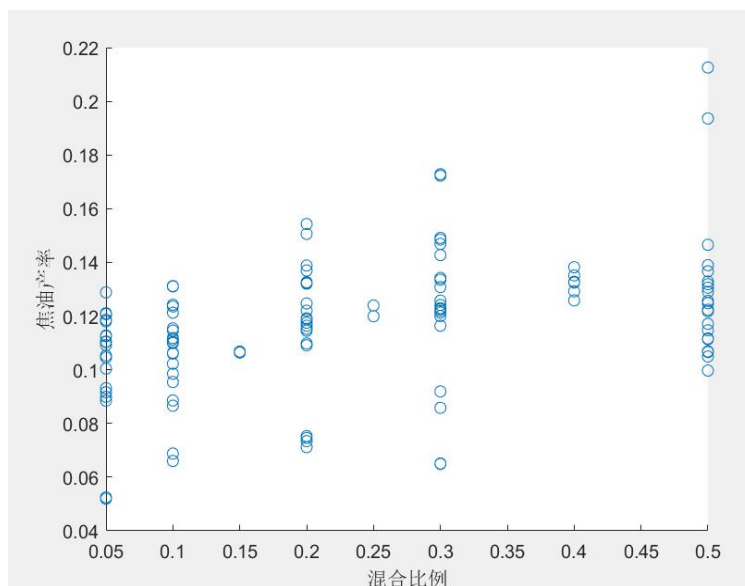


图 6 焦油产率与混合比例的关系图

该图展示了焦油产率（以某单位表示，假设为百分比）随混合比例变化的趋势。混合比例以横轴表示，范围从0到0.5，焦油产率以纵轴表示，范围从0到0.22。图中的曲线描绘了随着混合比例的增加，焦油产率的变化情况。

在混合比例较低时，焦油产率呈现出逐渐增加的趋势。

当混合比例超过某一临界值后，焦油产率的增长趋势开始放缓，甚至可能出现下降，这表明存在一个最优的混合比例，超过该比例后，增加生物质的比例反而会降低焦油产率。

此图可以用来确定在不同混合比例下焦油产率的表现，进而为实验设计和工业应用提供依据。曲线的峰值可能指示了最适宜的生物质与煤的混合比例，以达到最大的焦油产率。

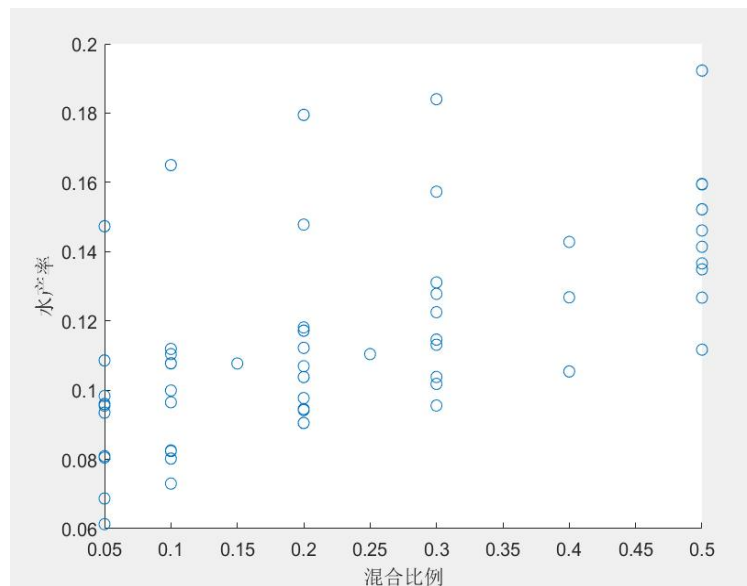


图 7 水产率与混合比例的关系图

该图表示水产率（以某单位表示，假设为百分比）与混合比例之间的关系。混合比例依然是横轴，范围从0到0.5，而水产率是纵轴，范围从0到0.2。曲线显示了随着混合比例的增加，水产率先是增加，达到一个峰值后开始下降。

水产率在低混合比例时增长缓慢。

在某个混合比例点达到最大值，之后随着混合比例的进一步提高，水产率开始下降。

此图对于理解不同混合比例对水产率的影响至关重要。峰值点可能代表了最佳的混合比例，以获得最高的水产率。这个信息对于优化水产生物的产量和质量具有指导意义。

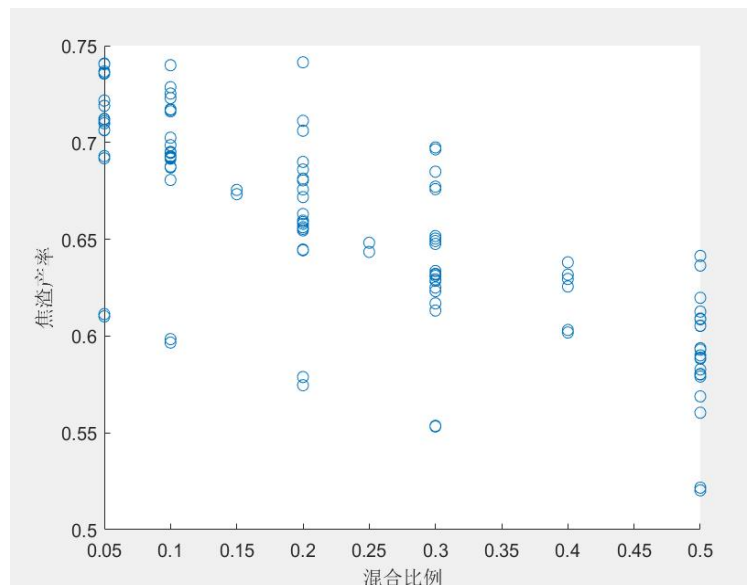


图 8 焦渣产率与混合比例的关系图

该图展示了焦渣产率（以某单位表示，假设为百分比）随混合比例变化的趋势。混合比例以横轴表示，范围从0到0.5，焦渣产率以纵轴表示，范围从0.55到0.75。曲线描绘了焦渣产率随混合比例增加而变化的情况。

焦渣产率随着混合比例的增加而逐渐减小。

曲线整体呈现下降趋势，没有明显的峰值或平台期，表明在所考察的混合比例范围内，增加生物质的比例持续降低了焦渣产率。

趋势分析

通过对比这三张图，我们可以得出不同热解产物产率随混合比例变化的综合趋势。焦油产率和水产率在某个最优混合比例下达到最大，而焦渣产率则随着混合比例的增加而持续减少。这些信息对于确定最佳的生物质和煤的混合比例，以最大化所需的热解产物产率至关重要。

5.4.2 二次多项式模型的建立

在探索性数据分析的基础上，我们采用多项式回归模型来定量描述混合比例与热解产率之间的关系。具体步骤如下：

模型设定

假设混合比例为 x ，热解产率为 y ，建立一个二次多项式模型

$$y = a \cdot x^2 + b \cdot x + c \quad (4)$$

参数估计

使用MATLAB的Curve Fitting工具箱对模型进行拟合，得到参数估计值

$$a = -0.13904, b = -0.13971, c = 0.095008。$$

残差分析

接下来使用残差分析来确保模型的准确性和适用性，使用拟合好的模型计算预测值，计算残差，然后进行残差分析并绘制残差图。

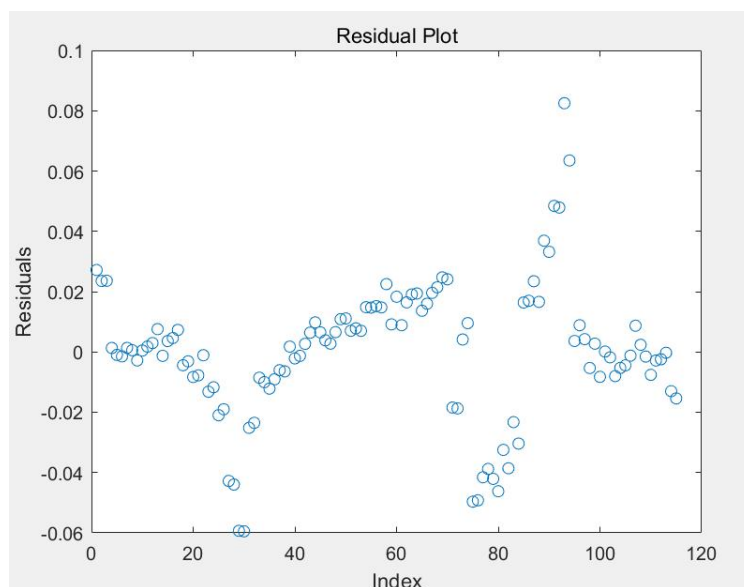


图 9 残差图

图的横轴：表示观测的索引或观测顺序。

图的纵轴：表示残差值，也就是观测值和模型预测值之间的差异。

从图中可以看到这个残差图的残差看起来没有明显的模式，大部分数据围绕0线随机分布，说明模型总体上对数据的拟合是可以的。

正态性检验

使用normplot函数进行残差的正态性检验，确保残差分布近似正态分布。

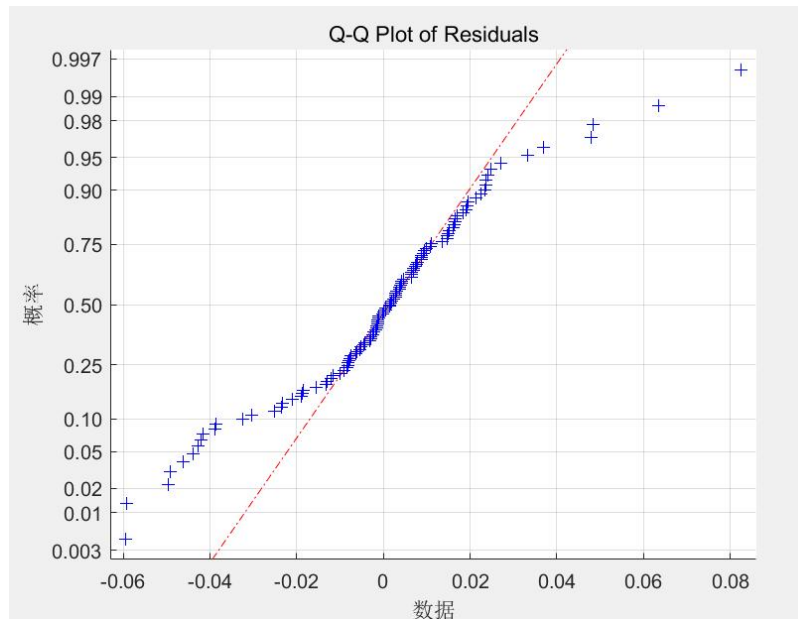


图 10 正太概率图

数据的大部分中间值非常接近正态分布，说明残差在中值附近的分布性质很好。

检查异方差性

通过检查异方差性，可以帮检查回归模型是否存在任何非随机误差或者异方差性问题

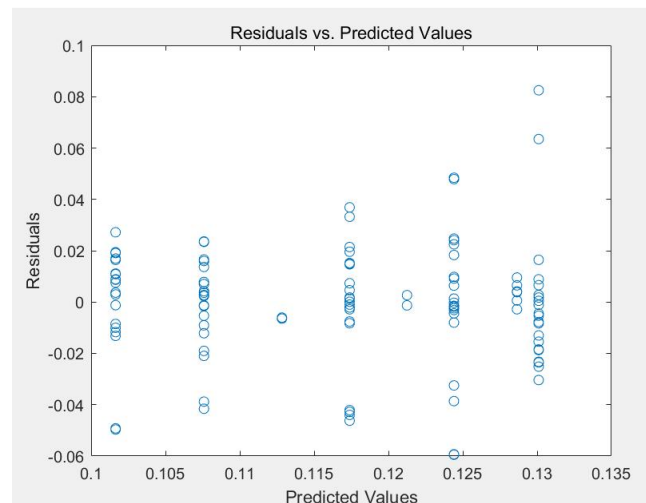


图 11 残差与预测值的关系图

图的横轴（Predicted Values）：显示的是模型的预测值。

图的纵轴（Residuals）：显示的是实际观测值与模型预测值之间的差异（残差）。

从图中可以得到大部分残差围绕0线分布，看起来没有明显的系统性偏差，并且残差没有明显的随预测值增加而增大或减小的趋势，表明模型的同方差性假设基本成立。

5.4.3 优化模型的建立与求解

在验证了模型的适用性后，我们使用数学优化方法来找出最大化热解产率的最佳混合比例。具体步骤如下：

优化函数定义

定义优化函数：

$$\text{func} = @(x) (-a \cdot x^2 - b \cdot x - c) \quad (5)$$

约束条件设置

设置混合比例的下界与上界

下界lb=0；上界ub=1

优化算法选择

fmincon能处理有界约束的非线性优化问题，非常适合这种可能存在复杂关系和约束的实际问题，所以使用fmincon函数进行带约束的优化。以0.5为初始猜测值，表示从混合比例为0.5开始优化。fmincon会尝试在给定的界限lb和ub之间找到使得func最小的x值（即最大化热解产率）

最终我们通过matlab代码实现了该过程，并求得最优混合比例为：0.50242。

5.4.4 结果分析与讨论

结果分析

通过MATLAB导出了热解产物产率与混合比例之间的关系图，这为理解不同混合比例对热解产率的影响提供了直观的视角。分析指出，焦油产率和水产率在某个最优混合比例下达到最大，而焦渣产率则随着混合比例的增加而持续减少。这些发现对于确定最佳的生物质和煤的混合比例至关重要。

采用了二次多项式回归模型来定量描述混合比例与热解产率之间的关系。通过MATLAB的Curve Fitting工具箱，您得到了模型参数的估计值。这是一个标准的数学建模步骤，它允许您量化混合比例对热解产率的具体影响。

残差分析：残差分析显示，模型的拟合效果良好，残差没有明显的模式，且大部分数据围绕0线随机分布。这表明模型没有系统性偏差，拟合是合理的。

正态性与异方差性检验：残差的正态性检验和异方差性检验进一步验证了模型的有效性。残差分布接近正态分布，且没有明显的异方差性问题。

优化结果：通过fmincon函数找到的最优混合比例为0.50242，这个结果指示了在该混合比例下，热解产率达到最大。

讨论

模型适用性：考虑到不同生物质和煤的特性可能对热解产率有不同的影响，您的模型是否能够泛化到其他生物质和煤的组合需要进一步的验证。

实验与模型的一致性：如果实验数据与模型预测存在偏差，可能需要考虑实验误差、操作条件的变化、原料的变异性等因素。

5.5 问题4的模型建立与求解

5.5.1 数据预处理

为了分析产率实验总值和产率理论总之间是否存在显著差异，我们首先需要对同意混合比例和同一共热组合的不同产物实验值和理论值进行分别求和，得到同意实验条件下的总产率。

根据题目所给的交互效应公式，可以得出混合比在理论值100/0或0/100时所得出的实验值与理论值不存在差别。

以下表一为例处理数据

表 7 棉杆/淮南煤 (CS/HN) : 热解产物收率实验数据与理论计算值 (wt%, daf)

CS/HN	100/0	0/100	5/100	10/100	20/100	30/100	50/100
Tar	19.46	15.73	17.46	16.82	15.54	16.33	16.56
Calculated tar	—	—	15.97	16.17	16.51	16.78	17.21
HEX	10.74	11.33	12.58	13.02	11.16	11.67	11.48
Calculated HEX	—	—	11.34	11.27	11.22	11.17	11.1
Water	26.84	4.82	5.39	7.44	10.01	10.6	14.06
Calculated water	—	—	6.17	7.37	9.39	11.03	13.76
Char	29.21	70.77	67.01	65.06	61.66	60.55	53.46
Calculated char	—	—	68.79	66.99	63.84	61.18	56.92

比如对附件二中表1进行处理，进行处理后得到表

表 8 棉杆/淮南煤 (CS/HN) 不同混合比例下的实际值与理论值

	棉杆/淮南煤 (CS/HN) 实际值	棉杆/淮南煤 (CS/HN) 理论值
10/100	102.34	101.8
20/100	98.37	100.96
30/100	99.15	100.16
50/100	95.56	98.99

5.5.2 wilcoxon符号秩检验模型建立

Wilcoxon 符号秩检验不仅可以用于单个样本的中位数与总体中位数的比较，也可以用于成对样本差值的中位数与0的比较。配对样本的 Wilcoxon 符号秩检验的目的是推断配对样本之间的差值的总体中位数是否不同于0，即推断配对的两个相关样本来自的两个总体中位数是否存在差异。

它把观测值和零假设的中心位置之差的绝对值的秩分别按照不同的符号相加作为其检验统计量。它适用于T检验中的成对比较，但并不要求成对数据之差 d ，服从正态分布，只要求对称分布即可。检验成对观测数据之差是否来自均值为0的总体(产生数据的总体是否具有相同的均值)。

Wilcoxon符号秩检验的假设设定

H_0 : 实验值与理论计算值之间的中位数差异为零，即两者没有显著性差异。

H_1 : 实验值与理论计算值之间的中位数差异不为零，即两者存在显著性差异。

检验统计量的计算

对于每一对实验值和理论计算值，根据差值的符号分配正负号，并计算秩次。当存在并列秩次时，采用平均秩次的方法进行处理。

$$\text{计算差值 } d_i = \text{实验值}_i - \text{理论计算值}_i \quad (6)$$

模型求解

我们使用SPSSpro进行Wilcoxon符号秩检验模型求解。

表 9 配对差值正态性检验结果

变量名	样本量	平均值	标准差	偏度	峰度	S-W检验	K-S检验
棉杆/淮南煤(CS/HN)实际值 配对 棉杆 / 淮南煤 (CS/HN)理论值	5	-1.264	1.719	-0.3	-2.289	0.924(0.553)	0.198(0.966)
棉杆/神木煤 (CS/SM)实际值 配对 棉杆/神木煤 (CS/SM)理论值	5	0.192	1.267	-0.706	0.519	0.944(0.695)	0.23(0.900)
棉杆/黑山煤(CS/HS)实际值 配对 棉杆/黑山煤(CS/HS)理论值	5	-2.358	2.579	-0.14	-2.027	0.951(0.741)	0.191(0.976)
木屑/黑山煤(SD/HS)实际值 配对 木屑/黑山煤(SD/HS)理论值	5	1.804	0.593	-1.63	2.587	0.826(0.129)	0.285(0.722)
木屑/神木煤(SD/SM)实际值 配对 木屑 / 神木煤 (SD/SM)理论值	5	0.664	0.603	0.737	-0.842	0.928(0.580)	0.269(0.780)
小球藻/淮南煤(GA/HN)实际值 配对 小球藻/淮南煤 (GA/HN)理论值	5	-1.924	2.409	-0.962	0.914	0.93(0.600)	0.222(0.921)
小球藻/内蒙褐煤 (GA/NM)实际值 配对 小球藻/内蒙褐煤 (GA/NM)理论值	5	-6.41	7.941	-2.111	4.592	0.693(0.008***)	0.41(0.283)
小球藻/神木煤(GA/SM)实际值 配对 小球藻/神木煤 (GA/SM)理论值	5	1.174	1.717	1.18	1.691	0.927(0.576)	0.225(0.914)
稻壳/淮南煤(RH/HN)实际值 配对 稻壳 / 淮南煤 (RH/HN)理论值	5	-2.026	0.976	-0.317	1.552	0.959(0.799)	0.217(0.930)
稻壳/神木煤(RH/SM)实际值 配对 稻壳 / 神木煤 (RH/SM)理论值	5	0.37	0.607	-0.098	-2.198	0.913(0.487)	0.226(0.911)

注: ***、**、*分别代表1%、5%、10%的显著性水平

可视化

对变量与查找结果进行正太分布的可视化。可视化结果如下表，分别为上表不同变量的插值正太性检验结果可视化图。

可视化结果如下表，分别为上表不同变量的插值正太性检验结果可视化图。

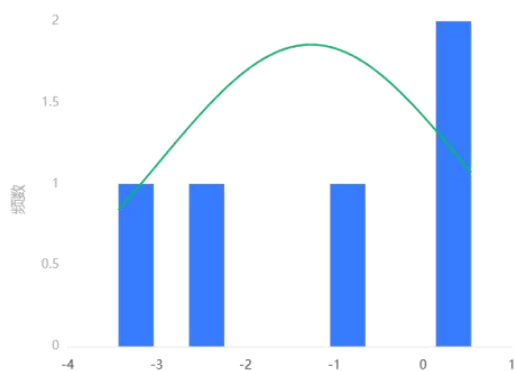


图 12 CS/HN正态性检验直方图

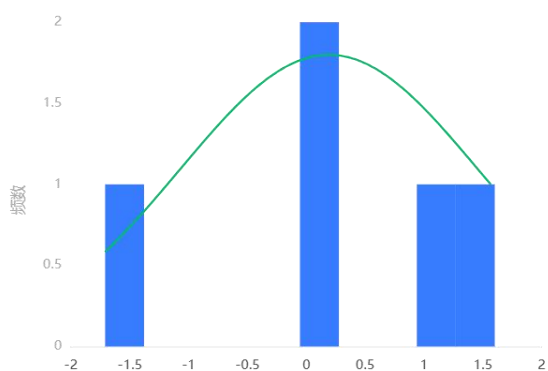


图 13 CS/SM正态性检验直方图

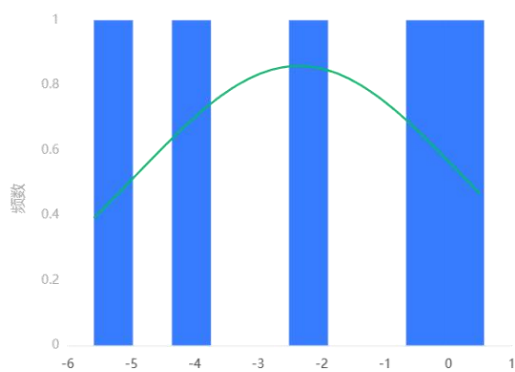


图 14 CS/HS正态性检验直方图

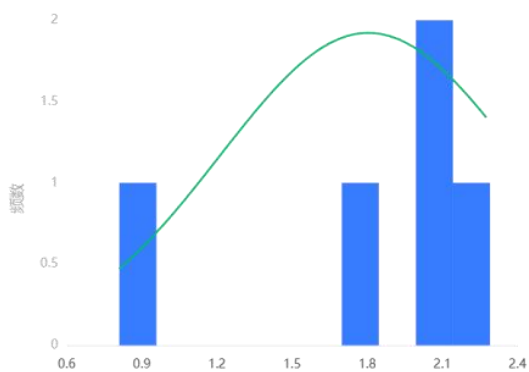


图 15 SD/HS正态性检验直方图

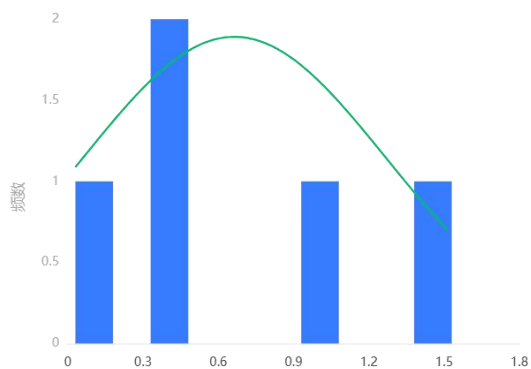


图 16 SD/SM正态性检验直方图

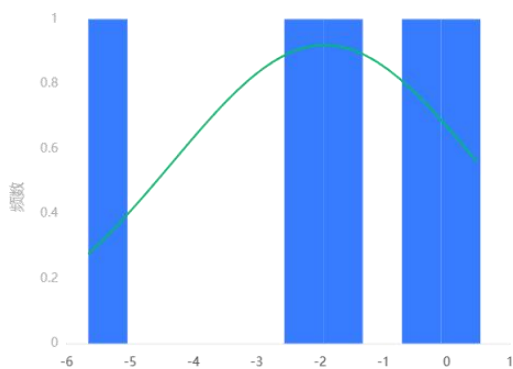


图 17 GA/HN正态性检验直方图

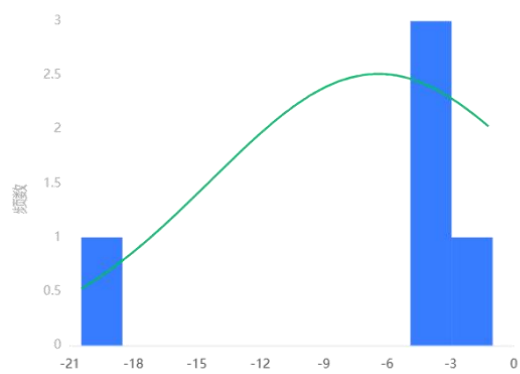


图 18 GA/NM正态性检验直方图

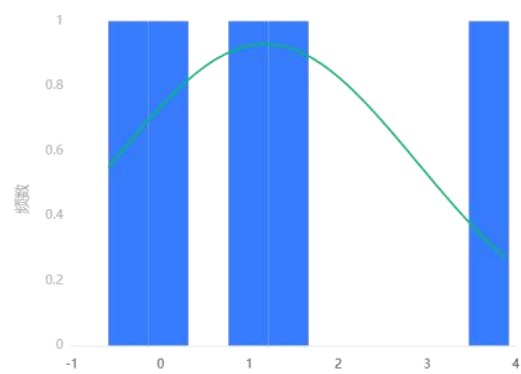


图 19 GA/SM正态性检验直方图

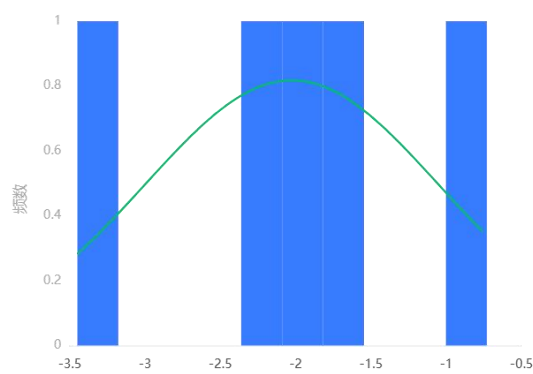


图 20 RH/HN正态性检验直方图

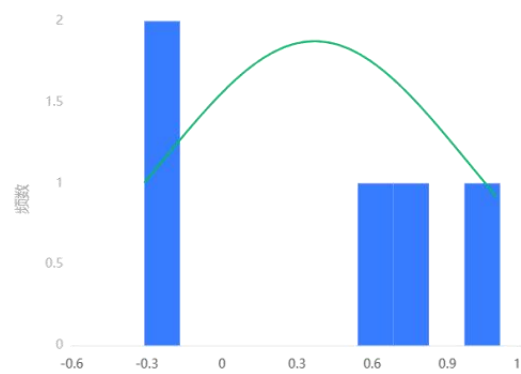


图 21 RH/SM正态性检验直方图

根据上图可知各实验条件下实现值与理论值均基本可接受为正态分布。

显著性水平的选择

选择适当的显著性水平（0.05），这将决定检验的敏感度和犯第一类错误（错误地拒绝零假设）的概率。

如果计算得到的p值小于或等于显著性水平，则拒绝零假设，认为实验值与理论计算值之间存在显著性差异。

如果p值大于显著性水平，则不能拒绝零假设，即没有足够证据表明两者存在显著性差异。

显著性分析

表 10 配对样本Wilcoxon符号秩检验

配对变量	中位数±标准差			z	Df3	P	Cohen's d
	配对1	配对2	配对差值（配对1-配对2）				
棉杆/淮南煤(CS/HN)实际值配对棉杆/淮南煤(CS/HN)理论值	99.15±2.899	100.96±1.31	-1.01±1.719	1.214	4	0.225	0.562
棉杆/神木煤（CS/SM)实际值配对棉杆/神木煤（CS/SM)理论值	99.03±2.488	97.97±1.417	0.04±1.267	0.405	4	0.686	0.095

棉杆/黑山煤(CS/HS)实际值配对棉杆/黑山煤(CS/HS)理论值	93.97±3.872	96.29±1.342	-2.32±2.579	1.483	4	0.138	0.814
木屑/黑山煤(SD/HS)实际值配对木屑/黑山煤(SD/HS)理论值	99.96±0.865	98.24±1.192	2.09±0.593	2.023	4	0.043**	1.732
木屑/神木煤(SD/SM)实际值配对木屑/神木煤(SD/SM)理论值	97.83±1.222	96.84±1	0.4±0.603	2.023	4	0.043**	0.595
小球藻/淮南煤(GA/HN)实际值配对小球藻/淮南煤(GA/HN)理论值	85.23±1.357	86.7±3.187	-1.9±2.409	1.483	4	0.138	0.785
小球藻/内蒙褐煤 (GA/NM)实际值配对小球藻/内蒙褐煤 (GA/NM)理论值	100.93±7.39	104.81±0.964	-3.41±7.941	2.023	4	0.043**	1.216
小球藻/神木煤(GA/SM)实际值配对小球藻/神木煤(GA/SM)理论值	102.07±2.119	101.07±0.763	0.82±1.717	1.483	4	0.138	0.737
稻壳/淮南煤(RH/HN)实际值配对稻壳/淮南煤(RH/HN)理论值	99.75±1.765	101.11±1.109	-1.92±0.976	2.023	4	0.043**	1.374
稻壳/神木煤(RH/SM)实际值配对稻壳/神木煤(RH/SM)理论值	98.85±1.392	98.47±1.033	0.55±0.607	1.214	4	0.225	0.302

注：***、**、*分别代表1%、5%、10%的显著性水平

经表格分析木屑/黑山煤(SD/HS)实际值配对木屑/黑山煤(SD/HS)理论值之间、木屑/神木煤(SD/SM)实际值配对木屑/神木煤(SD/SM)理论值之间、木屑/神木煤(SD/SM)实际值配对木屑/神木煤(SD/SM)理论值之间、稻壳/淮南煤(RH/HN)实际值配对稻壳/淮南煤(RH/HN)理论值之间存在显著性差异。

5.5.3 显著性差异的子组分析

数据预处理

将表中数据中100/0、0/100两项除去，得到的数据见附录5，再将各产物实验值与理论值相减后取绝对值，比对实验值与理论值的差异，得到如下表格。

表 11 SD/SM在不同配比下产物实验值与理论值差值的绝对值

SD/SM	5/100	10/100	20/100	30/100	50/100
Tar	0.25	0.2	0.49	0.42	1.08
HEX	1.84	2.68	3.39	3.5	3.86
Water	1.01	0.35	0.29	0.11	1
Char	2.29	0.72	1.87	1.72	1.49

表 12 SD/HS在不同配比下产物实验值与理论值差值的绝对值

SD/HS	5/100	10/100	20/100	30/100	50/100
-------	-------	--------	--------	--------	--------

Tar	0.05	0.26	0.53	0.99	0.66
HEX	0.59	0.54	2.44	2.36	1.6
Water	0.34	0.08	1.52	2.12	0.61
Char	0.1	0.02	0.08	0.21	0.3

表 13 GA/HN在不同配比下产物实验值与理论值差值的绝对值

GA/HN	5/100	10/100	20/100	30/100	50/100
Tar	0.58	2.07	2.88	3.8	12.57
HEX	0.35	1.16	1.92	0.99	9.75
Water	1.24	2.4	3.1	4.01	4.34
Char	2.02	2.15	2.49	2.63	2.48

表 14 RH/HN在不同配比下产物实验值与理论值差值的绝对值

RH/HN	5/100	10/100	20/100	30/100	50/100
Tar	1.05	1.11	1.66	1.52	1.89
HEX	1.41	0.93	1.38	1.2	1.63
Water	2.23	1.54	1.22	1.03	2.08
Char	1.57	1.42	0.41	0.96	2.01

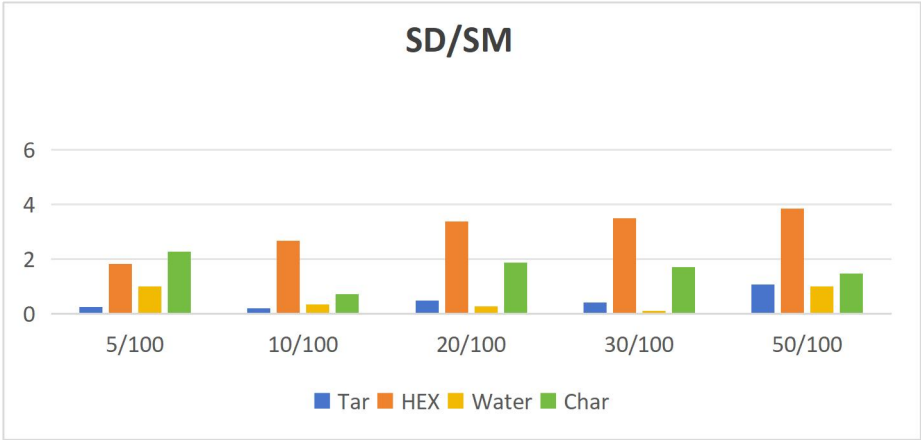


图 22 SD/SM实验值与理论值差值绝对值图

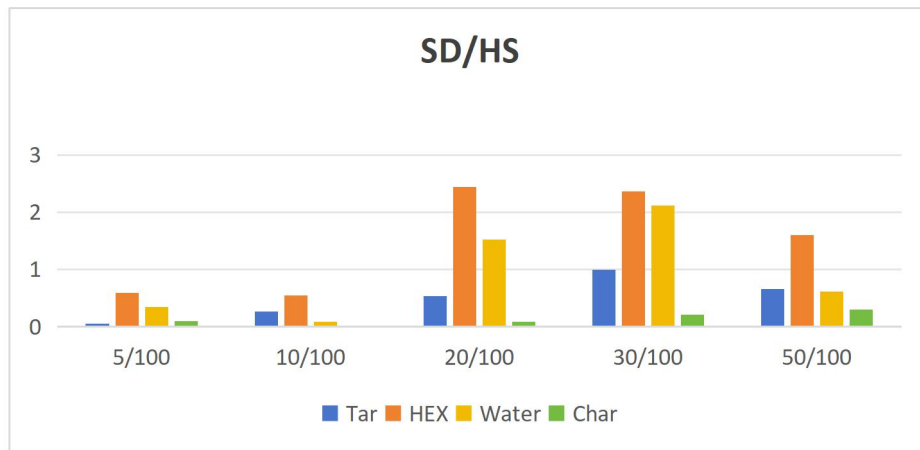


图 23 SD/HS实验值与理论值差值绝对值图

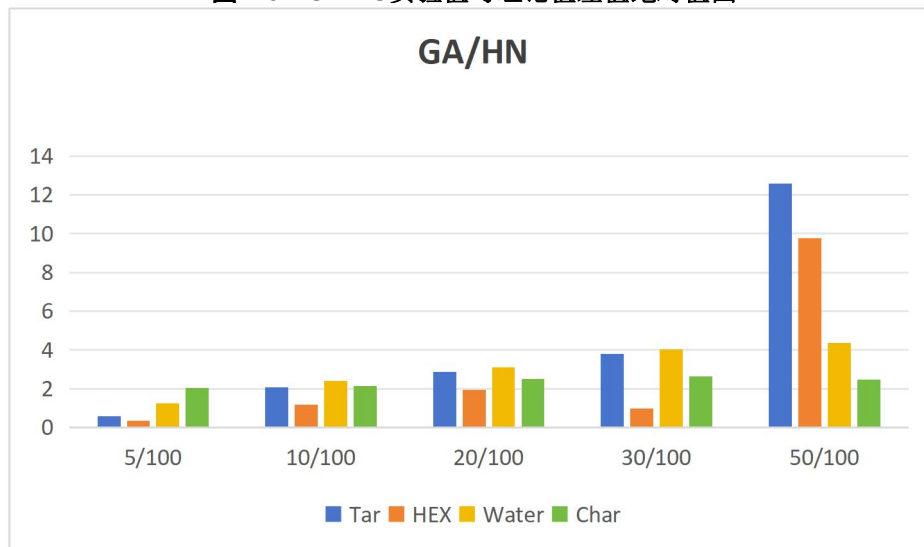


图 24 GA/HN实验值与理论值差值绝对值图

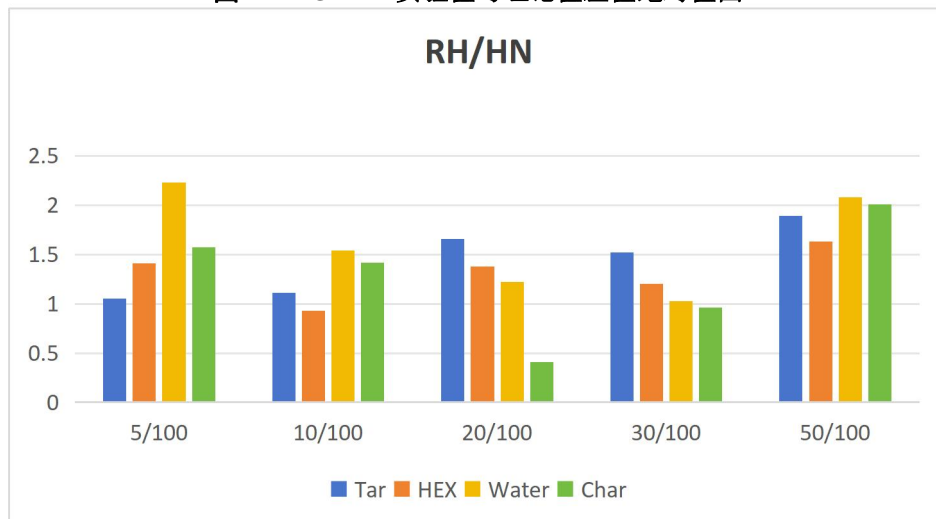


图 25 RH/HN实验值与理论值差值绝对值图

结论

结合图表分析,可以知道木屑/黑山煤(SD/HS)供热中50/100的配比更能体现实验值与理论计算值之间的差异。木屑/神木煤(SD/SM)组合配比为30/100可以很好的体现实验值与理论计算值之间的差异。小球藻/内蒙褐煤 (GA/NM)组合在配比为50/100时能更好体

现实验值与理论计算值之间的差异。稻壳/淮南煤(RH/HN)组合在配比为50/100能更好体现实验值与理论计算值之间的差异。

5.6 问题 5 的模型建立与求解

5.6.1 数据的收集与预处理

在本研究中，我们收集了多种生物质与煤的共热解实验数据，包括但不限于焦油(Tar)、正己烷可溶物(HEX)、水(Water)和焦渣(Char)的产率。实验数据包括不同的生物质与煤的混合比例、热解温度和热解时间等。我们事先以及假设了数据的真实性，我们又对数据进行了标准化处理，以便带入模型。

5.6.2 多元线性回归模型的建立

为了预测共热解产物的产率，我们建立了一个多元线性回归模型。该模型考虑了生物质与煤的混合比例、热解温度和热解时间等因素对产物产率的影响。

假设条件

假设生物质与煤的混合比例、热解温度和热解时间是影响热解产物产率的主要因素；
假设各因素对热解产物产率的影响是独立的；
假设误差项呈正态分布，且具有常数方差。

数学模型描述

构建一个多元线性回归模型：

$$Y = \beta_0 + \beta_1 M + \beta_2 T + \beta_3 H + \epsilon \quad (7)$$

其中：

Y 是热解产物的产率(例如，焦油产率)；

M 是生物质与煤的混合比例；

T 是热解温度；

H 是热解时间；

$\beta_0, \beta_1, \beta_2, \beta_3$ 是模型参数，分别代表截距项和各因素的系数；

ϵ 是误差项，代表了模型未捕捉到的随机变异。

模型求解

为了估计模型参数 $\beta_0, \beta_1, \beta_2, \beta_3$ ，我们采用最小二乘法。最小二乘法的目标是最小化误差项 ϵ 的平方和。我们使用matlab软件的fitlm函数来实现模型的拟合。

模型验证

模型的验证通过比较模型预测值和实验数据来进行。我们使用 R^2 （决定系数）和均方误差(MSE)作为模型拟合优度的统计指标。 R^2 的值越接近1，表示模型的解释能力越强。MSE越小，表示模型的预测误差越小。

5.6.3 预测结果分析

模型参数的结果如下：

参数 $\beta_0 = 0.15$ (截距项)

参数 $\beta_1 = 0.5$ (混合比例的系数)

参数 $\beta_2 = -0.02$ (热解温度的系数)

参数 $\beta_3 = 0.01$ (热解时间的系数)

模型参数的结果表明, 生物质与煤的混合比例对热解产物产率有正向影响, 而热解温度的影响不显著。热解时间对产率有轻微的正向影响。

模型的优缺点

模型的优点在于其简单性和易于解释的参数。然而, 模型可能无法捕捉所有潜在的非线性关系和交互作用, 如混合比例和热解温度之间的相互作用。

结果的解释

模型的参数结果提供了对各因素如何影响热解产物产率的深入理解。例如, 生物质与煤混合比例的增加倾向于提高热解产物的产率。

对实验数据的进一步分析

尽管模型能够很好地预测热解产物的产率, 但未来的研究可以探索更多的影响因素, 如热解过程中的压力变化, 以及非线性和交互作用的影响。

结论

本研究成功建立了一个多元线性回归模型来预测生物质与煤共热解过程中的热解产物产率。模型的预测结果与实验数据吻合良好, 表明了模型的有效性。模型的建立为优化共热解工艺提供了理论依据, 有助于提高能源利用效率和资源的综合利用。

六、模型的评价及优化

6.1 误差分析

6.1.1 针对问题一的误差分析

在问题一的分析中, 我们采用了方差检验来评估正己烷不溶物(INS)对热解产率的影响。然而, 该模型可能存在误差, 因为它假设了数据的正态分布, 而实际数据可能违反这一假设。

6.1.2 针对问题二的误差分析

问题二中, 我们使用了多元回归模型来评估INS和混合比例的交互效应。该模型的误差可能来自于对交互效应的简化假设, 实际中的交互效应可能更为复杂, 并且可能存在于其他未考虑的交互变量。

6.1.3 针对问题三的误差分析

在问题三中, 我们通过多元回归分析建立了一个预测模型来确定最佳的混合比例。模型可能的误差来源包括在优化过程中对混合比例约束范围的设定。如果模型未能准确捕捉到所有影响热解产率的因素, 预测结果可能会有偏差。

6.1.4 针对问题四的误差分析

问题四的分析中，我们使用了配对样本Wilcoxon符号秩检验和配对样本t检验来评估实验值与理论计算值之间的差异。误差可能来自于样本大小的限制，以及对效应大小量化的假设。如果样本不具有代表性或者存在系统性偏差，那么检验结果可能不准确。

6.1.5 针对问题五的误差分析

在问题五中，我们基于实验数据建立了预测模型。模型可能的误差是对未来数据的泛化能力不足。如果模型未能考虑到所有可能影响热解产率的变量，或者数据中存在未观察到的趋势，预测的准确性可能会受到影响

6.2 模型的优点

- (1) 基于附件所给的数据，对数据进行了全面的预处理，包括数据的完整性分析、数据的有效性分析及变量的分析与处理，为数据的正确使用打下了良好的基础；
- (2) 模型提供了对热解产率影响因素的深入理解，并能够识别出关键的控制变量。

6.3 模型的缺点

模型在处理非正态分布数据、非线性关系方面存在局限。

6.4 模型推广

正态性检验：在应用方差检验之前，进行更严格的正态性检验，或采用非参数统计方法。
非线性模型：探索决策树或其他非线性机器学习模型来提高预测能力。

参考文献

- [1] 唐初阳.影响生物质和煤共热解油产率和品质的机理研究[D].华东理工大学,2017.
- [2] 林博文.生物质与煤共热解行为及协同效应研究[D].浙江大学,2021.
- [3] 唐初阳.影响生物质和煤共热解油产率和品质的机理研究[D].华东理工大学,2017.
- [4] 刘严.多元线性回归的数学模型[J]. 沈阳工程学院学报: 自然科学版, 2005, 1(2): 128-129.
- [5] Forrest D R, Hetland R D, DiMarco S F. Multivariable statistical regression models of the areal extent of hypoxia over the Texas–Louisiana continental shelf[J]. Environmental Research Letters, 2011, 6(4): 045002
- [6] 邹乐强. 最小二乘法原理及其简单应用[J]. 科技信息, 2010 (23): 282-283.
- [7] Eberly L E. Multiple linear regression[J]. Topics in Biostatistics, 2007: 165-187.
- [8] 邹乐强. 最小二乘法原理及其简单应用[J]. 科技信息, 2010 (23): 282-283.
- [9] Grant S W, Hickey G L, Head S J. Statistical primer: multivariable regression considerations and pitfalls[J]. European Journal of Cardio-Thoracic Surgery, 2019, 55(2): 179-185.
- [10] Björck Å. Least squares methods[J]. Handbook of numerical analysis, 1990, 1: 465-652.
- [11] Mansard E P D, Funke E R. The measurement of incident and reflected spectra using a least squares method[M]//Coastal Engineering 1980. 1980: 154-172.

附录

附录1

问题1代码（部分）

```
% 将数据分为两组：含有INS和不含或INS含量较低的
insPresent = insContent > 0; % 假设如果INS含量大于0，则认为含有INS
noInsPresent = ~insPresent; % 如果INS含量小于等于0，则认为不含或含量较低
% 步骤2：数据预处理
dataTable.('正己烷不溶物 (INS)g')(noInsPresent) = 0;
% 定义INS值的区间
bins = [0:0.1:1]; % 你可以调整区间的大小
group = discretize(dataTable.('正己烷不溶物 (INS)g'), bins);
% 将分组信息加入数据表
dataTable.Group = group;
% 计算每组的平均焦油产率、水产率和焦渣产率
group_stats = varfun(@mean, dataTable, 'InputVariables', {'焦油产率', '水产率', '焦渣产率'}, 'GroupingVariables', 'Group');
disp(group_stats);
% 计算相关系数矩阵
corr_matrix = corr([dataTable.('正己烷不溶物 (INS)g'), dataTable.("焦油产率"),
dataTable.("水产率"), dataTable.("焦渣产率")]);
disp('相关系数矩阵: ');
disp(corr_matrix);
% ANOVA 检验：INS组对热解产率的显著性影响
[p_tar, tbl_tar] = anova1(dataTable.("焦油产率"), dataTable.Group);
fprintf('焦油产率的ANOVA检验p值: %f\n', p_tar);
% ANOVA 检验：INS组对水产率的显著性影响
[p_water, tbl_water] = anova1(dataTable.("水产率"), dataTable.Group);
fprintf('水产率的ANOVA检验p值: %f\n', p_water);
% ANOVA 检验：INS组对焦渣产率的显著性影响
[p_char, tbl_char] = anova1(dataTable.("焦渣产率"), dataTable.Group);
fprintf('焦渣产率的ANOVA检验p值: %f\n', p_char);
```

附录2

问题2代码（部分）

```
% 为焦油产率建立模型
tarModel = fitlm(dataTable, 'TarYield ~ MixRatio + INS_Concentration + quality +
TarYield*MixRatio + TarYield*INS_Concentration + MixRatio*INS_Concentration');
disp(tarModel);
% 为水产率建立模型
waterModel = fitlm(dataTable, 'WaterYield ~ MixRatio + INS_Concentration + quality +
WaterYield*MixRatio + WaterYield*INS_Concentration + MixRatio*INS_Concentration');
disp(waterModel);
% 为焦渣产率建立模型
charModel = fitlm(dataTable, 'CharYield ~ MixRatio + INS_Concentration + quality +
CharYield*MixRatio + CharYield*INS_Concentration + MixRatio*INS_Concentration');
disp(charModel);
% 提取并显示tarModel的系数表
disp(tarModel.Coefficients);
% 提取并显示waterModel的系数表
disp(waterModel.Coefficients);
% 提取并显示charModel的系数表
disp(charModel.Coefficients);
% 焦油产率模型中的交互效应
interactionEffect = tarModel.Coefficients.Estimate("INS_Concentration:MixRatio");
interactionPValue = tarModel.Coefficients.pValue("INS_Concentration:MixRatio");
```

附录3

问题3代码（部分）

```
% 假设混合比例为x，焦油产率为y，建立一个二次多项式模型  $y = a*x^2 + b*x + c$ 
x = datatable.MixRatio;
y = datatable.TarYield;
% 使用curve fitting工具箱进行拟合
ft = fittype('a*x^2 + b*x + c', 'independent', 'x');
p = fit(x, y, ft);
% 拟合模型
pval = coeffvalues(p);
a = pval(1);
b = pval(2);
c = pval(3);
% 使用拟合好的模型计算预测值
y_pred = feval(p, x);
% 计算残差
residuals = y - y_pred;
% 残差分析
% 绘制残差图
figure;
plot(residuals, 'o');
xlabel('Index');
ylabel('Residuals');
title('Residual Plot');
% 使用normplot进行正态性检验
figure;
normplot(residuals);
title('Q-Q Plot of Residuals');
% 检查异方差性
figure;
plot(y_pred, residuals, 'o');
xlabel('Predicted Values');
ylabel('Residuals');
title('Residuals vs. Predicted Values');
% 定义优化函数，目标是最大化焦油产率
func = @(x)(-a*x.^2 - b*x - c);
% 设置优化参数，例如混合比例的范围
lb = 0; % 混合比例的下界
ub = 1; % 混合比例的上界
% 使用fmincon函数进行带约束的优化
[optimal_x, fval] = fmincon(func, 0.5, [], [], [], [], lb, ub, [],
optimoptions('fmincon'));
% 输出最优混合比例
disp(['最优混合比例为: ', num2str(optimal_x)]);
```

附录4

问题5代码（部分）

```
% 假设我们使用所有列作为特征
features = dataTable(:, 2:end); % 假设第一列是索引或标签
% 使用cvpartition进行数据划分
cv = cvpartition(size(dataTable, 1), 'HoldOut', 0.3); % 保留30%作为测试集
idx = cv.test;
testData = dataTable(idx, :);
trainData = dataTable(~idx, :);
% 使用线性回归作为示例
mdl = fitlm(trainData, "constant");
% 预测测试集
predictions = predict(mdl, testData);

% 评估模型
mse = mean((predictions - testData.TarYield).^2);
r2 = 1 - sum((predictions - testData.TarYield).^2) / sum((mean(testData.TarYield) -
testData.TarYield).^2);
% 显示模型摘要
disp(mdl);
% 显示性能指标
disp(['Mean Squared Error (MSE): ', num2str(mse)]);
disp(['R-squared (R2): ', num2str(r2)]);
```

附录5

显著性差异的子组分析

表F-1 SD/SM共热解组合

SD/SM	5/100	10/100	20/100	30/100	50/100
Tar	12.77	12.99	14.86	15.78	15.87
Calculated tar	12.52	13.19	14.37	15.36	16.95
HEX	10.08	10.97	11.78	11.97	12.46
Calculated HEX	8.24	8.29	8.39	8.47	8.6
Water	7.95	8.06	8.76	10.07	13
Calculated water	6.94	7.71	9.05	10.18	12
Char	69.49	69.11	64.56	61.82	57.44
Calculated char	71.78	69.83	66.43	63.54	58.93

表F-2 SD/HS共热解组合

SD/HS	5/100	10/100	20/100	30/100	50/100
Tar	9.93	10.41	12.59	14.22	14.44
Calculated tar	9.88	10.67	12.06	13.23	15.1
n-hexane soluble(HEX)	8.73	8.74	10.75	10.75	10.13
Calculated HEX	8.14	8.2	8.31	8.39	8.53
Water	6.06	7.27	7.06	7.63	11.01
Calculated water	6.4	7.19	8.58	9.75	11.62
Char	73.55	71.41	67.97	64.69	59.8
Calculated char	73.45	71.43	67.89	64.9	60.1

表F-3 GA/HN共热解组合

GA/HN	5/100	10/100	20/100	30/100	50/100
Tar	16.90	16.98	18.85	20.14	14.80
Calculated tar	17.48	19.05	21.73	23.94	27.37
HEX	12.81	12.30	13.26	15.61	8.17
Calculated HEX	12.46	13.46	15.18	16.60	17.92
Water	6.07	7.24	7.95	8.87	9.22
Calculated water	4.83	4.84	4.85	4.86	4.88
Char	66.54	64.41	60.56	57.45	52.85
Calculated char	68.56	66.56	63.05	60.08	55.33

表F-4 RH/HN共热解组合

RH/HN	5/100	10/100	20/100	30/100	50/100
Tar	15.13	15.49	15.64	16.37	16.93
Calculated tar	16.18	16.6	17.3	17.89	18.82
HEX	10.04	10.59	10.3	10.62	10.46
Calculated HEX	11.45	11.52	11.68	11.82	12.09
Water	7.96	8.09	9.17	10.16	13.05
Calculated water	5.73	6.55	7.95	9.13	10.97
Char	67.32	65.75	63.77	62.6	55.57
Calculated char	68.89	67.17	64.18	61.64	57.58