

队伍编号	901957
题号	D

## 脱氧合金化的收得率预测与配料优化

### 摘要

炼钢过程中，脱氧合金化是重要环节，决定了冶炼钢材的质量与成本，对技术水平要求较高。这一过程中需要加入多种合金材料，传统方法为工人根据经验按照规定加入，而随着钢铁产业的发展，我们需要研究一套合理的配料模型，来辅助冶炼过程，降低成本并提高生产效率与质量。

对于问题一，首先利用均值法及碳当量数据填补缺失值，并删去不合理值，从而正常化数据。接着根据定义，利用附件一中的数据计算出历史收得率。对于分析收得率的主要影响因素，本文在无量纲化、归一化数据后，利用PCA得到各成分间关系，结合所查阅资料，提炼出10个一级指标。综合此，建立了合理的收得率影响因素模型。

对于问题二，首先考虑传统方法，采用参考炉次法，利用最近5炉数据预测新一轮钢的收得率。之后采用多元线性回归训练模型。检验结果表明，该模型能预测大体趋势，但误差波动较大。接着我们引入第二个模型，根据第一问中得到的主要指标，建立支持向量机预测的输入向量模型，并利用支持向量机求解。检验结果表面，这一模型预测精度可以符合冶炼标准，误差波动很小。

对于问题三，考虑到实际生产要求质量与安全，建立以钢材质量与安全系数为限制条件，以收得率最高与成本最低为目标的多目标多限制条件优化模型。经过查阅资料，对于炼钢过程仔细考察后，我们着重考虑脱氧剂对于模型的影响，以及采购量对于安全性、成本的影响，因此加入修正系数，进一步提升模型性能。

最后，鲁棒性检验表明，本文建立的模型具有较好的数据波动适应性，能符合实际生产需要，并且在类似的工业生产中有应用价值。

关键词：PCA，支持向量机预测，多目标优化，鲁棒性检验

# 目录

一、 问题重述 . . . . .	1
1.1 背景介绍 . . . . .	1
1.2 问题提出 . . . . .	1
二、 模型假设 . . . . .	1
三、 符号说明 . . . . .	1
四、 问题一的模型建立与求解 . . . . .	2
4.1 数据预处理 . . . . .	2
4.1.1 缺失数据的处理 . . . . .	2
4.1.2 异常数据的修正 . . . . .	3
4.2 历史收得率计算 . . . . .	3
4.3 基于PCA的收得率影响因素确定 . . . . .	4
五、 问题二的模型建立与求解 . . . . .	5
5.1 基于参考炉次法的收得率预测模型 . . . . .	5
5.1.1 模型建立 . . . . .	5
5.1.2 基于多元线性回归的模型训练 . . . . .	6
5.2 基于SVR的收得率预测模型 . . . . .	6
六、 问题三的模型建立与求解 . . . . .	8
七、 建议信 . . . . .	9
八、 灵敏度分析 . . . . .	10
九、 模型评价与推广 . . . . .	11
9.1 模型的优点 . . . . .	11
9.2 模型的缺点 . . . . .	11
十、 参考文献 . . . . .	11
附录1 多元线性回归Matlab代码 . . . . .	12
附录2 SVR的Matlab代码 . . . . .	12

# 一、 问题重述

## 1.1 背景介绍

钢铁的制作过程中，脱氧合金化是至关重要的一步。钢厂需向钢水加入所需合金，从而调整其中氧含量与合金元素含量，使之达到所炼钢种规格的成分范围。为了保证钢水质量的同时降低成本，可在历史合金收得率及当前钢水数据的基础上，建立自动配料模型。国内目前的大多钢铁企业仍依靠操作工人的工作经验、按照冶炼标准选择合金的种类与加入量，造成了钢材质量的限制与成本的居高不下。因此，开发适合脱氧合金化的动态预测与优化模型至关重要。

## 1.2 问题提出

- 1. 计算C与Mn元素的历史收得率，并分析影响其收得率的主要因素。
- 2. 构建数学模型，预测C与Mn的收得率，并改进模型，使其准确率尽可能高。
- 3. 由于不同材料成本不同，建立数学模型，使钢水在脱氧合金过程中成本优化。
- 4. 根据研究成果，给钢厂领导写一封建议信。

# 二、 模型假设

- 1. 假设题目中所给数据真实可靠。
- 2. 假设忽略实际炼钢环境下的操作失误等不可控变量。
- 3. 不考虑钢炉因使用时间长短等原因部分性质发生改变对转炉冶炼收得率的影响。

# 三、 符号说明

符号	定义
$\eta$	修正系数
$fR$	收得率函数
$fw$	成本函数
$fw$	安全系数函数
$Ceq$	碳当量
$R$	历史收得率
$Con$	连铸正样数据
$End$	转炉终点数据
$M$	钢水净重

表 1 符号说明

## 四、 问题一的模型建立与求解

### 4.1 数据预处理

经过对数据的观察研究并查阅相关资料与历史数据，我们发现原始数据存在部分数据异常，另有些类型的数据缺失，因此首先进行数据正常化处理。

#### 4.1.1 缺失数据的处理

原始数据中，转炉终点Mn与转炉终点P的数据从第7A06618炉开始缺失，连铸正样C、Mn、S、P、Si、Cr的数据从第7A06058炉开始缺失。另存在少量丢失一整行数据的情况，例如7A05410炉数据。这种情况可能是由于数据记录时出现失误。空数据无意义，因此我们直接将其删去。

对于前两种情况，考虑到不同钢号的材料特性相差较大，在具体计算时，我们仅用某一种钢已有数据的均值来填补同一种钢的缺失值，避免混杂各类钢材的数据，导致最终结果出现偏差。

##### (1) 转炉终点数据的处理

对于这一部分缺失数据，我们采用“先剔除，后替换”的策略，先剔除已有数据中的异常值，再采用均值替换法处理，即

$$x_m = \frac{1}{n} \sum_{i=1}^n x_i$$

其中 $n$ 为未缺失且正常的的数据总数， $x_m$ 为我们填充的转炉终点缺失值。

由于转炉终点Mn、转炉终点P值自身总体相差不大，且分布较为稳定，因此用历史炉次的均值来替换是可行的。以7A06618炉为例，填补后的数据为

炉号	钢号	转炉终点Mn
7A06618	HRB400B	0.00103

表 2 7A06618炉转炉终点数据修补

##### (2) 连铸正样数据的处理

虽然元素的连铸正样值大量缺失，但我们注意到数据列“连铸正样Ceq\_val”并未缺失，即钢材的碳当量数据未缺失。国际焊接学会推荐的碳当量公式如下<sup>[1]</sup>

$$Ceq = C + \frac{Mn}{6} + \frac{Cr + V + Mo}{5} + \frac{Cu + Ni}{15}$$

因此元素含量与碳当量的比值应当为一个稳定的值。我们可以计算已有数据的元素含量、碳当量的均值，并采用其比值补充未知的缺失值，即

$$\frac{y_m}{Ceq_m} = \frac{\frac{1}{n} \sum_{i=1}^n y_i}{\frac{1}{n} \sum_{j=1}^n Ceq_j}$$

其中 $y_m$ 为我们填充的连铸正样缺失值， $Ceq_m$ 为缺失值对应的碳当量值， $n$ 为未缺失且正常的数据数量。

以7A06058炉为例，填补后的数据为

炉号	钢号	连铸正样C	连铸正样Mn
7A06058	HRB400B	0.002366043	0.013685299

表 3 7A06058炉连铸正样数据修补

#### 4.1.2 异常数据的修正

原始数据中，有的数据明显过大或过小，显然是不合理的数据。

例如，钢水温度大多在1600℃，因此温度为0℃的少量数据明显应该剔除。再如，可能是由于手工输入的误差，7A06638炉的钢水净重出现明显突变，这种数据异常容易对结果产生不利影响。

炉号	钢号	钢水净重
7A06638	HRB500D	243250

表 4 7A06638炉钢水净重异常

对于类似的异常数据，我们利用其本身或者临近数据进行修改替换。

## 4.2 历史收得率计算

由题意，合金的历史收得率为脱氧合金化过程中，被钢水吸收的合金元素重量与加入的该元素总重量之比，结合所给数据，以C的收得率计算为例，我们可以得到某一炉钢的计算公式。

$$R_C = \frac{Con_C \times M - End_C \times M}{\sum_{i=1}^k (M_{Ai} \times W_{Ci})} \quad (1)$$

其中 $R_C$ 代表该炉钢的C收得率； $Con_C$ 代表连铸正样C数据，即脱氧合金化后钢水中碳元素的含量； $End_C$ 为转炉终点C数据，即脱氧合金化前钢水中碳元素含量； $M$ 为该炉钢水的净重； $M_{Ai}$ 为加入钢水的第 $i$ 种合金的质量； $W_{Ci}$ 是这种合金中碳元素的含量。

因此，某一钢号的材料的碳收得率为该种钢的碳收得率的平均值。最终结果如下表。

钢号	Mn历史收得率	C历史收得率
20MnKA	0.9594	0.9049
20MnKB	0.9378	0.9420
HRB400B	0.8925	0.9310
HRB400D	0.8991	0.9423
HRB500B	0.9870	0.8977
HRB500D	0.9482	0.9693
Q345B	0.9241	0.9011

表 5 C和Mn的历史收得率

### 4.3 基于PCA的收得率影响因素确定

#### (1) 数据处理

由于不同类型数据绝对数值相差很大，因此我们首先对数据进行无量纲化与归一化处理。我们按如下方式进行标准化处理：

$$\widetilde{M}_i = \frac{M - \overline{M}}{S}$$

其中 $\overline{M}$ 与 $S$ 分别为均值和标准差，这样可以在不改变数据相关性的前提下，消除不同变量量纲的影响。

#### (2) 确定主成分

根据以上步骤，我们利用Matlab，我们将38种元素整和为10个一级指标，其中BPLOT反应了元素之间的相关度，LOADING PLOT反应了PCA后新组成的几个影响因子内部的权重。

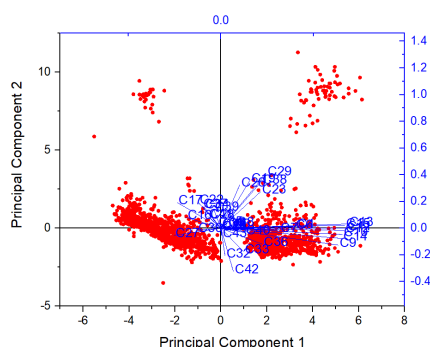


图 1 BPLOT

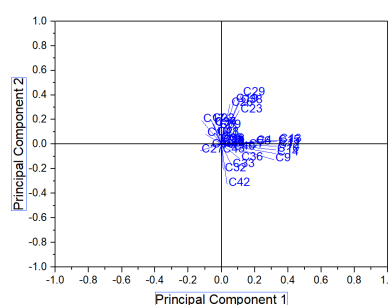


图 2 LOADING PLOT

我们保留了权重总和95%的元素，剔除数据不足元素的同时，额外补充实际工业冶炼中有明确影响，但题给数据不足的部分，完成收得率影响因素变量的整个体系架构。

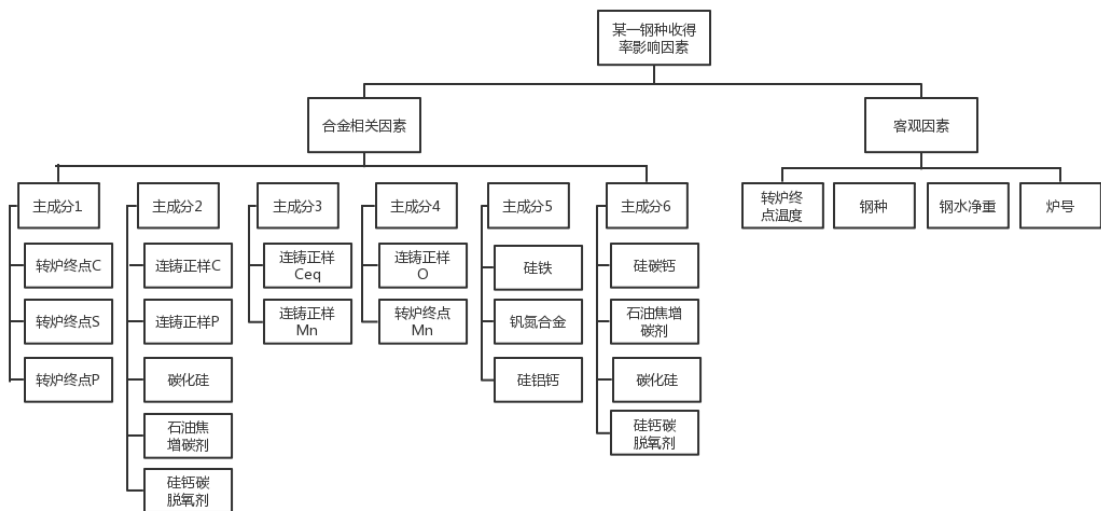


图 3 收得率影响因素体系构架表

## 五、 问题二的模型建立与求解

由于不同钢种的模型原理相同，以下模型皆以HRB400B号钢的C收得率为例。

### 5.1 基于参考炉次法的收得率预测模型

#### 5.1.1 模型建立

考虑到以下方面的原因，我们借鉴以往研究经验<sup>[2]</sup>，采用参考炉次法计算合金的收得率：

1. 实际冶炼环境复杂，条件多变，用统一的模型衡量较为困难。
2. 参考炉次法具有自我更新的特点，可自动学习历史数据的特征。
3. 该方法可以跟踪冶炼过程随炉次增加的变化情况。

在上一问计算出的收得率数据基础上，参考炉次法计算合金收得率的公式为<sup>[3]</sup>：

$$R_E = \sum_{i=1}^5 \alpha_i R_{Ei} + \beta R_{benchmark} + \gamma \quad (2)$$

其中 $R_E$ 为本炉次合金收得率； $R_i$ 为参考炉次的收得率； $R_{benchmark}$ 代表某一钢种的基准收得率； $\alpha$ 、 $\beta$ 、 $\gamma$ 为待定系数。

对于参考炉次，我们选择生产相同钢种，且距离当前炉次最近的五个冶炼炉次。当前炉生产完成后，若也满足条件，则也加入参考炉次中。

对于基准收得率，选取相同钢种正常冶炼的历史数据，对其收得率取平均值。该值随生产的进行不断更新。

5.1.2 基于多元线性回归的模型训练

对于待定系数，我们利用多元线性回归法，抽取第一问所得数据的70%来训练该模型。训练模型所得系数如下：

$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$
0.2475	0.1269	0.0582	0.0758
$\alpha_5$	$\beta$	$\gamma$	
0.0678	0.4181	0.00089	

表 6 历史数据训练得到的待定系数值

得到模型后，我们代入被划分出来的30%数据，并计算其与真实值的差异，得到以下结果：



图 4 参考法碳收得率模型检验

可以看出，这一模型大致满足预测的需要，预测趋势随炉次变化调整较为灵敏，总体趋势符合实际。但该模型仅仅参考过往炉次，误差略大且误差波动剧烈，不能更精准地预测，不能够完全满足实际生产时的需要。因此我们建立第二个预测模型。

5.2 基于SVR的收得率预测模型

传统的回归方法要求当回归函数 $f(x)$ 完全等于 $y$ 时才认为正确，而支持向量回归SVR则认为只要 $f(x)$ 与 $y$ 的偏离程度在规定阈值 $\epsilon$ 内，都可认为误差为0，因此只计算 $|f(x) - y| > \epsilon$ 的数据点的误差。

支持向量回归中，核函数是决定其性能的关键点。通过查阅相关资料，钢水中还原剂的总量对于C的收得率影响很大，因此我们将钢水中还原剂总含量、钢水总量与还



原剂比值其列入支持向量机的输入向量。再结合第一问PCA的结果，我们首先建立核函数的输入向量模型如下。

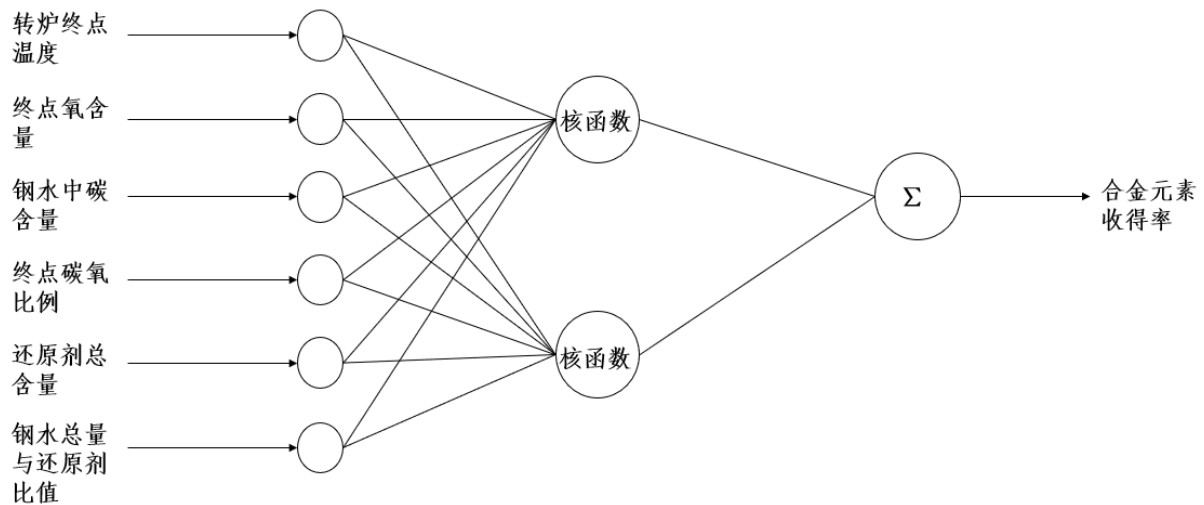


图 5 SVR输入向量模型

为了在传统的参考炉次法的基础上获得性能提升，我们只要求当前模型的误差小于参考炉次法的误差。因此阈值 $\epsilon$ 规定如下：

$$\epsilon = \frac{\sum_{i=1}^n (\sum_{i=1}^5 \alpha_i R_{Ei} + \beta R_i + \gamma)}{n} \quad (3)$$

模型训练时，落在阈值范围内的点，我们不计算其误差，即认为其预测准确。而对于落在阈值外的点，我们定义不敏感的损失函数如下：

$$L_{\epsilon}(z) = \begin{cases} 0, & \text{if } |z| \leq \epsilon \\ |z| - \epsilon, & \text{else} \end{cases} \quad (4)$$

因此，在模型训练中，SVR模型的求解可以转化为

$$\min_{w_1 b} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^m L_{\epsilon}(f(x_i) - y_i) \quad (5)$$

对于该式最小值的求解。

我们依旧随机抽取70%的数据作为训练样本，30%的数据作为检验数据，得到模型效果如下图。可以明显看出，提升后的SVR模型效果明显好于传统的参考炉次法，误差很小且预测趋势准确。二者的对比图如下右图。

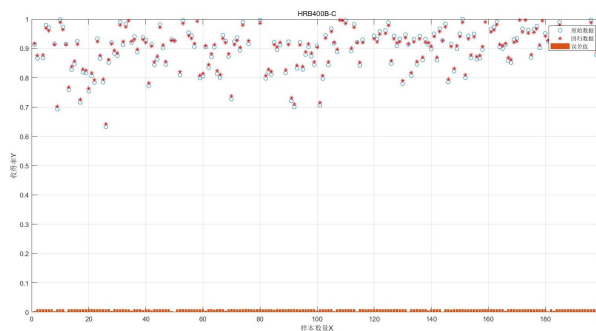


图 6 SVR模型检验

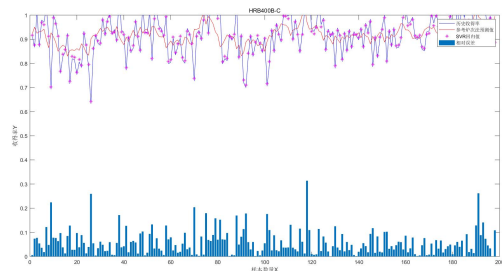


图 7 对比图

## 六、 问题三的模型建立与求解

问题三要求我们根据第二问的预测结果，优化脱氧合金化过程中加入的合金种类及比例。因此我们建立优化模型。

在考虑自变量范围时，我们需要明确各个变量之间的关系。考虑到碳元素的含量及收得率与钢水中脱氧剂的含量关联密切，为了在保证钢材质量的情况下，尽可能降低成本、提高收得率，我们主要研究脱氧剂的相关因素，以此研究更优的合金配比。综合以上分析考察，我们可以建立框图模型如下：

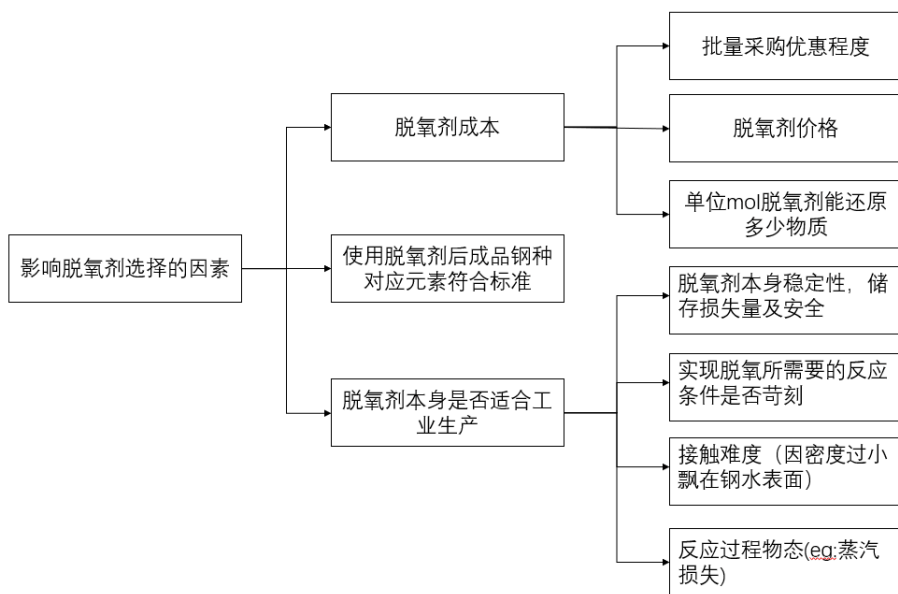


图 8 优化模型

在优化模型建立之前，我们先明确限制条件与优化目标。生产的目的是得到合格

的钢材，由题目中给的条件可知钢材HBR400B的标准。因此模型的限制条件1为：

$$\begin{cases} \sum_{i=1}^n m_i \times \rho_{Ci} \geq 0.23 \\ \sum_{i=1}^n m_i \times \rho_{Si} \geq 0.55 \\ \sum_{i=1}^n m_i \times \rho_{Mni} \geq 1.45 \\ \sum_{i=1}^n m_i \times \rho_{Pi} \geq 0.04 \\ \sum_{i=1}^n m_i \times \rho_{Si} \geq 0.04 \end{cases} \quad (6)$$

其中n为附件二中给出的合金材料数量。这一限制条件保证了生产的钢材符合质量标准，即保证了最低要求。

同时，考虑到钢铁炼制时安全性很重要，因此我们的限制条件2是需要保证该种方案的安全系数在警戒值以上，即

$$f(s) \geq safety\_standard$$

下面是优化目标。在生产合格钢材的基础上，我们模型的目标应使收得率 $f(R)$ 最高、成本 $f(w)$ 最低，即

$$max : f(R) \quad min : f(w)$$

基于以上分析，我们在以上三个目标的基础上建立一个综合的定价方案。

在成本的考量过程中，最基础的成本计算即为购入量与价格的乘积。考虑到材料有损耗的可能性，我们增加参数 $p$ 作为调整。同时，当同一种材料大量购入时，批量购买可以使得团购成本下降，但其存储风险上升，即安全性下降。因此我们添加修正系数 $\eta$ ，额外考虑了安全成本。综合以上分析，我们得到成本公式 $f(w)$ 如下：

$$f(w) = \sum_{i=1}^n \left( \frac{m_i}{p} \times w_i + \eta \right) \quad (7)$$

其中 $n$ 为加入材料的总种类数， $m_i$ 为理想条件下需要的某种材料质量， $p$ 为材料损失的可能性系数， $w_i$ 为对应材料的单价。

而修正系数经过我们的试验，规定如下：

$$\eta = \frac{1}{1 + 1.31 \ln m_i} \quad (8)$$

而对于安全度的计算，我们引入工业生产中每种材料的安全系数 $s$ 。修正系数原理同上。因此得到安全系数公式 $f(s)$ 如下：

$$f(s) = \frac{\sum_{i=1}^n s}{n} \times \frac{1}{1 + \frac{\ln m}{1.31}} \quad (9)$$

我们以生产1吨合格的HRB400B钢材为例，优化的配料比为：

## 七、 建议信

尊敬的企业老板，你好！

材料	所需质量
硅锰面	218
钒铁(FeV50-B)	45
碳化钙	63
硅钙碳脱氧剂	127

表 7 1吨HRB400B钢材优化后配料比

随着钢铁行业竞争的激烈，钢材产品的利润空间逐渐变小，在保证质量的同时提高转化率成了钢厂进一步发展所必须要考虑的问题。我们主张，在挖掘已有设备的功能潜力下，合理调整产品配比、反应环境，通过工艺技术进步，合理控制成本，在适应大生产的条件下，降低炼钢过程的生产消耗，有效降低成本，提高毛利率，为钢厂的进一步发展做好准备。

首先，在精炼过程中，有很多影响转炉冶炼收得率的因素，经过我们分析，温度，碳氧比，还原剂投放量是主要的影响因素。其次，我们发现，合金元素收得率的预测对于原料配比至关重要，保证钢水质量的同时最大限度降低合金钢的生产成本，很大程度上依赖于合金元素收得率的计算。因此我们针对性的建立了关于收得率预测的在线预测模型，新的模型符合实际生产环境中对实时性的要求，有效提高了收得率的预测准确度，让进一步降低成本成为可能。

目前，国外已借助以合金收得率预测及成本优化算为主题的自动配料模型实现计算机自动配料。为实现钢厂的持续发展，可以引进国外技术，学习并持续优化相关模型，提升自己的核心竞争力，结合历史数据多次实践以后投入生产创造收益；

进一步地，对于合金成分控制模型，为实现最低成本，企业可从原料的价格和数量两方面着手。第一，企业应多关注新化合物在脱氧合金化中的作用，分出关键原料和次要原料，并分别结合相关原料的价格，并分析原料市场价格的波动，以及自身生产条件，确定合适的购买方案。在采购过程中根据成本，利润，产钢中的合金含量动态模型并结合自身情况来选择一个平衡点，降低成本的同时提高质量。

此外，可以将传统的冶炼技术进行改良，更有效的提高接触面积，增大利用率。

希望钢厂的发展越来越好。

祝好！

## 八、 灵敏度分析

我们修改原始数据，使其波动 $\pm 2.5\%$ 及 $\pm 5\%$ ，观察预测结果的波动情况如下。可以看出，我们的模型所得结果在数据波动的情况下变化并不剧烈，因此具有较好的鲁棒性。

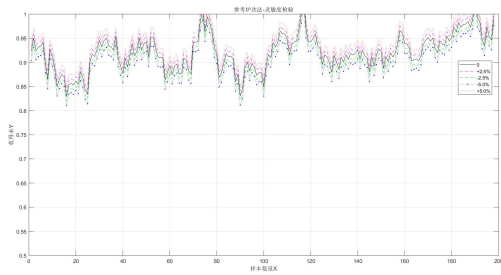


图 9 参考炉次法灵敏度检验

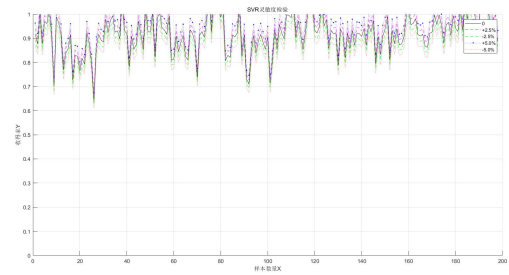


图 10 SVR模型灵敏度检验

## 九、 模型评价与推广

### 9.1 模型的优点

1. 本文模型预测精度高，鲁棒性好，符合实际生产需要。
2. 成本优化模型综合考虑实际生产情况，考虑到了安全性、批量采购等，实际应用价值明显。

### 9.2 模型的缺点

1. 由于缺乏实际数据，成本模型不够精确。

## 十、 参考文献

- [1] 斯重遥等. 焊接手册(第二卷). 北京: 机械工业出版社, 1992: 59-65.
- [2] Notman GK. Development in BOS process control at british steel[J]. Steel Times, 1991(7): 360.
- [3] Watson DR. Computerized basic oxygen furnace operation with substance dynamic control[A]. Steelmaking Conference Proceedings. 1984(121)

## 附录1 多元线性回归Matlab代码

```
clc;clear;

x1 = xlsread('400B.xlsx','BA3:BA200');
x2 = xlsread('400B.xlsx','BA4:BA201');
x3 = xlsread('400B.xlsx','BA5:BA202');
x4 = xlsread('400B.xlsx','BA6:BA203');
x5 = xlsread('400B.xlsx','BA7:BA204');
y = xlsread('400B.xlsx','BA2:BA199');

RE = repmat(0.9310,[198 1]);
%X=[ones(size(y)) x1 x2 x3 x4 x5 RE];
X=[RE x1 x2 x3 x4 x5];

[b,bint,r,rint,stats] = regress(y,X);

YY = b(1,1)*RE+b(2,1).*x1+b(3,1).*x2+b(4,1).*x3+b(5,1).*x4+b(6,1).*x5;
XX = 1:1:198;
plot(XX,y,'b');
hold on
plot(XX,YY,'k');
axis([0 200 0 1]);
CHA = abs(YY-y);
hold on
bar(CHA);
```

## 附录2 SVR的Matlab代码

```
tic;
close all;
clear;
clc;
format compact;

x = (1:1:198)';
y = xlsread('400B.xlsx','BA2:BA199');
model = svmtrain(y,x,'-s 3 -t 2 -c 2.2 -g 2.8 -p 0.01');
```

```

[py,mse,prob] = svmpredict(y,x,model,'-b_0');
figure;
plot(x,y,'o');
hold on;
plot(x,py,'r*');
legend('ORIGIN_DATA','POCESSED_DATA');
CHA = abs(py-y);
hold on
bar(CHA);
grid on;

testx = [199;200;201];
display('REAL_DATA');
testy = xlsread('400B.xlsx','BA200:BA202')
[ptesty,tmse,prob2] = svmpredict(testy,testx,model,'-b_0');
display('PREDICTED_DATA');
ptesty
axis([0 200 0.5 1]);
toc

plot(x,py,'k');
py1 = py*(1+0.025);
py2 = py*(1-0.025);
py3 = py*(1+0.05);
py4 = py*(1-0.05);
plot(x,py1,'m—');
hold on;
plot(x,py2,'g—');
hold on;
plot(x,py3,'b:.');
hold on;
plot(x,py4,'r:');
hold on;

```