

## OSC-PLS 方法用于渣油裂解装置的软测量建模

成 忠, 诸爱士

(浙江科技学院 生物与化学工程学院, 杭州 310023)

**摘 要:** 渣油裂解反应中, 影响沥青产率的因素多, 反应机理十分复杂, 难以建立准确的机理模型。采用基于正交投影的正交信号校正(OSC)算法对输入变量测量数据进行预处理, 剔除数据中所含的与待测变量如浓度、收率等无关的噪声信息; 再实施 OSC 与偏最小二乘(PLS)回归相结合的 OSC-PLS 方法, 建立渣油裂解装置沥青产率的软测量校正模型。结果显示: 模型精度和稳定性较非线性方法均有显著提高, 而且模型所需 PLS 成分数减少, 模型更简洁。

**关键词:** 正交信号校正; 偏最小二乘回归; 软测量; 裂解装置; 过程建模

中图分类号: TE626.86; TP274

文献标识码: A

文章编号: 1671-8798(2007)01-0010-04

## Soft Sensor Modeling for Cracker of Residue Oil by OSC-PLS

CHENG Zhong, ZHU Ai-shi

(School of Biological and Chemical Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China)

**Abstract:** It is hard to get the satisfied mechanism model in the deasphalting process of residue oil for complicated influential factors and intricate reactive mechanism. A novel partial least squares algorithm that embedded the orthogonal signal correction (OSC) into the regression framework of the partial least squares (PLS) method, termed as OSC-PLS method, is implemented. OSC technique is used to delete the noise signal that is independent of the property variable, and the PLS algorithm is applied to build the cracker soft sensor calibration model. The result indicates that the OSC-PLS approach can not only improve the model accuracy and stability with comparison to some nonlinear modeling methods, but also decreases the PLS factors and the model becomes more concise.

**Key words:** orthogonal signal correction; partial least square regression; soft sensor; cracker; process modeling

原油经常减压精馏后, 会剩余约 40% 的减压渣油。为了提高原油利用率, 需对渣油进行深度加工, 通常采用蒸汽裂解技术, 使渣油发生一系列裂解与缩合反应, 生成裂解气、裂解轻油、裂解重油和沥青。渣油裂解反应不仅受原料理化性质的影响,

还与操作条件有关, 反应机理十分复杂, 很难得到准确的机理模型<sup>[1]</sup>。因此, 建立渣油裂解装置沥青产率的软测量校正模型, 对于品质的及时调控十分关键。

正交信号校正 (Orthogonal signal correction,

收稿日期: 2006-12-06

基金项目: 浙江科技学院科研基金资助项目 (QF200501)

作者简介: 成 忠 (1973—), 男, 江苏盐城人, 讲师, 博士, 主要从事化工系统工程研究。

OSC)由 Wold 等学者于 1998 年首先提出<sup>[2]</sup>,它作为一种数据预处理方法,在对原始数据进行噪音滤除时,可确保去除掉的信息与待测信息(浓度,收率等)无关,具有传统方法所没有的优越性。为此,本文将通过实施 OSC 与偏最小二乘(partial least squares, PLS)回归相结合的 OSC-PLS 方法,由 OSC 先期进行渣油裂解装置输入变量的测量数据降噪处理,而后,针对输入变量间存在的强相关性,选用 PLSR 多变量回归建模方法,负责对高维线性相关系统的数据降维、特征提取和回归建模。实例仿真结果验证了该方法的有效性。

## 1 OSC-PLS 方法

### 1.1 OSC 方法

OSC 方法最初应用于光谱计量学领域,其目标是作为一个数据预处理步骤,用以去除光谱矩阵  $\mathbf{X}^{n \times p}$  中与预测矩阵  $\mathbf{Y}^{n \times q}$  无关的正交部分信息,即正交得分矩阵  $\mathbf{T}_\perp$ ,满足  $\mathbf{T}_\perp^\top \mathbf{Y} = \mathbf{Y}^\top \mathbf{T}_\perp = 0$ ,其中  $n$  为样本容量, $p, q$  分别为光谱矩阵和预测矩阵变量维数。随后,又发展了 OSC 若干改进算法<sup>[3-7]</sup>。现将 Wold 的基于 NIPALS(nonlinear iterative partial least squares)的 OSC 算法<sup>[2]</sup> 归结如下:

第 1 步,将  $\mathbf{X}$  和  $\mathbf{Y}$  矩阵按列,即变量进行标准化预处理;

第 2 步,基于 NIPALS 算法计算  $\mathbf{X}$  阵的第一主成分,作为初始的  $\mathbf{t}_\perp^{(1)}$  开始迭代,它确保初始的  $\mathbf{t}_\perp^{(1)}$  是  $\mathbf{X}$  矩阵中所包含信息的一个较优综合;

第 3 步,计算  $(\mathbf{t}_\perp^{(1)})_{\text{new}} = \mathbf{t}_\perp^{(1)} - \mathbf{Y}(\mathbf{Y}^\top \mathbf{Y})^{-1} \mathbf{Y}^\top \mathbf{t}_\perp^{(1)}$ ,并用一个内置的 PLS 算法,在  $\mathbf{X}$  与  $(\mathbf{t}_\perp^{(1)})_{\text{new}}$  间建立一个多成分 PLS 模型:  $\{\mathbf{X}, (\mathbf{t}_\perp^{(1)})_{\text{new}}\} \xrightarrow{\text{PLS}} \{\mathbf{W}^*, \mathbf{P}^*, \mathbf{q}^*\}$ ,由此得出  $\mathbf{X}$  与  $(\mathbf{t}_\perp^{(1)})_{\text{new}}$  间的回归系数  $\mathbf{b}_\perp^{(1)} = \mathbf{W}^*[(\mathbf{P}^*)^\top \mathbf{W}^*]^{-1} \mathbf{q}^*$ ;

第 4 步,通过  $\mathbf{X}$  和  $\mathbf{b}_\perp^{(1)}$  计算新的得分向量  $\mathbf{t}_\perp^{(1)} = \mathbf{X} \mathbf{b}_\perp^{(1)}$ ,并由此计算  $\|\mathbf{t}_\perp^{(1)} - (\mathbf{t}_\perp^{(1)})_{\text{new}}\| / \|\mathbf{t}_\perp^{(1)}\|$  以判别  $\mathbf{t}_\perp^{(1)}$  的收敛,若满足收敛精度要求,进入下一步,否则返回第 3 步;

第 5 步,计算正交成分的载荷向量  $\mathbf{p}_\perp^{(1)} = \mathbf{X}^\top \mathbf{t}_\perp^{(1)} / ((\mathbf{t}_\perp^{(1)})^\top \mathbf{t}_\perp^{(1)})$ ;

第 6 步,计算修正后的数据  $\mathbf{X}^{\text{OSC}} = \mathbf{X} - \mathbf{t}_\perp^{(1)} (\mathbf{p}_\perp^{(1)})^\top$ 。用  $\mathbf{X}^{\text{OSC}}$  代替  $\mathbf{X}$ ,重复步骤 2~6。最后,所需提取的正交成分数  $r$ ,可通过交叉验证(Cross Validation, CV)<sup>[8]</sup> 确定,且将得到参数矩阵  $\mathbf{P}_\perp =$

$[\mathbf{p}_\perp^{(1)}, \mathbf{p}_\perp^{(2)}, \dots, \mathbf{p}_\perp^{(r)}]$ ,  $\mathbf{B}_\perp = [\mathbf{b}_\perp^{(1)}, \mathbf{b}_\perp^{(2)}, \dots, \mathbf{b}_\perp^{(r)}]$  和正交成分阵  $\mathbf{T}_\perp = [\mathbf{t}_\perp^{(1)}, \mathbf{t}_\perp^{(2)}, \dots, \mathbf{t}_\perp^{(r)}]$ 。

对于新样本个体  $\mathbf{x}_{\text{new}}$ ,首先利用第 1 步中得到的标准化参数进行预处理,然后可按下述过程进行 OSC 修正:取  $i = 1$ ,从  $\mathbf{B}_\perp$  中取出第  $i$  列作为  $\mathbf{b}_\perp^{(i)}$ ,计算  $\mathbf{t}_{\perp, \text{new}}^{(i)} = \mathbf{x}_{\text{new}}^\top \mathbf{b}_\perp^{(i)}$ ,再从  $\mathbf{P}_\perp$  中取出第  $i$  列作为  $\mathbf{p}_\perp^{(i)}$ ,计算  $\mathbf{x}_{\text{new}}^{\text{OSC}} = \mathbf{x}_{\text{new}} - \mathbf{t}_{\perp, \text{new}}^{(i)} \mathbf{p}_\perp^{(i)}$ 。令  $i = i + 1$ ,  $\mathbf{x}_{\text{new}} = \mathbf{x}_{\text{new}}^{\text{OSC}}$ ,重复上述过程,直至去除  $r$  个正交成分。

### 1.2 OSC-PLS 方法

现将光谱矩阵经过正交信号校正后的  $\mathbf{X}^{\text{OSC}}$  与预测矩阵  $\mathbf{Y}$  组合为新的样本阵  $\{\mathbf{X}^{\text{OSC}}, \mathbf{Y}\}$ ,代入线性 PLS 算法<sup>[9]</sup> 以建立回归校正模型。在确定所需提取  $h$  个 PLS 成分数后,最终得自变量成分矩阵  $\mathbf{T} = [\mathbf{t}^{(1)}, \mathbf{t}^{(2)}, \dots, \mathbf{t}^{(h)}]$ 、转换权矩阵  $\mathbf{W} = [\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \dots, \mathbf{w}^{(h)}]$ 、因变量成分矩阵  $\mathbf{U} = [\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(h)}]$ 、载荷向量矩阵  $\mathbf{Q} = [\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(h)}]$ ,成分对  $\{\mathbf{t}^{(j)}, \mathbf{u}^{(j)}\} (j = 1, 2, \dots, h)$  间的回归系数矩阵  $\mathbf{B} = \text{diag}(\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots, \mathbf{b}^{(h)})$ ,以及  $\mathbf{X}^{\text{OSC}}$  和  $\mathbf{Y}$  间的回归系数矩阵  $\mathbf{C} = (\mathbf{W}(\mathbf{P}^\top \mathbf{W})^{-1} \mathbf{B}^\top)$ 。

至此,线性 PLS 算法所构建的模型,可实现自变量  $\mathbf{x}_{\text{new}}^{\text{OSC}} \in \mathcal{R}^p$  预报因变量  $\mathbf{y}_{\text{new}}$ 。先由  $\mathbf{W}$  计算  $\mathbf{W}^\top \mathbf{x}_{\text{new}}^{\text{OSC}} = [\mathbf{t}_{\text{new}}^{(1)}, \mathbf{t}_{\text{new}}^{(2)}, \dots, \mathbf{t}_{\text{new}}^{(h)}]^\top$ ,然后应用相应成分对间的线性回归模型:  $\mathbf{u}^{(j)} = \mathbf{b}^{(j)} \mathbf{t}^{(j)} + \mathbf{e}^{(j)}$ ,  $\mathbf{e}^{(j)}$  为回归误差,由  $\mathbf{t}_{\text{new}}^{(j)} (j = 1, 2, \dots, h)$  预报相应的  $\mathbf{u}_{\text{new}}^{(j)} (j = 1, 2, \dots, h)$

值,最后算出  $\mathbf{y}_{\text{new}}$  的预报值  $\hat{\mathbf{y}}_{\text{new}} = \sum_{j=1}^h \mathbf{u}_{\text{new}}^{(j)} (\mathbf{q}^{(j)})^\top$ 。

综上所述,OSC-PLS 方法的框架思路为,首先将实测光谱矩阵  $\mathbf{X}$  通过 OSC 方法分割成两部分:

$$\mathbf{X} = \mathbf{X}^{\text{OSC}} + \mathbf{X}_\perp \quad (1)$$

式(1)中:  $\mathbf{X}_\perp = \sum_{i=1}^r \mathbf{t}_\perp^{(i)} (\mathbf{p}_\perp^{(i)})^\top$ ,为过滤掉的与  $\mathbf{Y}$  正交的无关信息。而后,将  $\mathbf{X}^{\text{OSC}}$  与  $\mathbf{Y}$  组合实施 PLSR 算法,它可以更少的 PLS 成分数  $h$  得到回归校正模型。在实施 PLSR 算法后,  $\mathbf{X}^{\text{OSC}}$  和  $\mathbf{Y}$  可表示为:

$$\mathbf{X}^{\text{OSC}} = \sum_{j=1}^h \mathbf{t}^{(j)} (\mathbf{p}^{(j)})^\top + \mathbf{E} \quad (2)$$

$$\mathbf{Y} = \sum_{j=1}^h \mathbf{u}^{(j)} (\mathbf{q}^{(j)})^\top + \mathbf{F} \quad (3)$$

式(2)(3)中:  $\mathbf{E}$  和  $\mathbf{F}$  分别为  $\mathbf{X}^{\text{OSC}}$  和  $\mathbf{Y}$  可忽略的残差阵。由此可看出,OSC-PLS 方法将 OSC 作为数据预处理的过滤器,去除  $\mathbf{X}$  中跟  $\mathbf{Y}$  无关的正交信息  $\mathbf{X}_\perp$ ,以期让后继的  $\mathbf{X}^{\text{OSC}}$  与  $\mathbf{Y}$  在用 PLSR 方法建模时所需提取的 PLS 成分数  $h$  减少,模型得以简化,且易于解

释和理解<sup>[7,10]</sup>。

## 2 OSC-PLS 方法为渣油裂解装置软测量建模

由于渣油热裂解反应机理复杂,很多细节之处尚不清楚,所以,纯机理模型难以准确地反映裂解装置中的真实情况,从而影响模型的性能和精度。考虑到稳态生产过程,尽管从机理上解析指标与影响因素间为严格非线性关系,但生产过程中描述它们的数据一般呈现线性相关,同时因 OSC 在每一步中都去除  $X$  中与  $Y$  无关的成分,所以,即使原来  $X$  与  $Y$  的相关性不大,经过去除足够多的正交成分后,  $X^{\text{OSC}}$  与  $Y$  的相关性也将变得很大。由此,基于样本数据建立它们间的线性经验模型,可获得较高的学习和泛化性能<sup>[11-13]</sup>。为此,本文拟用 OSC-PLS 方法为裂解装置建模,以期在实际采集的样本数据的基础上建立一个能预测沥青产率的非机理模型。

### 2.1 样本数据说明

依据机理分析与生产经验,本文拟用 8 个因素作为模型的自变量,它们为:渣油的理化特性,包括黏度、密度和残碳;裂解反应的工艺参数,包括装置的出口温度、反应温度、反应时间、蒸汽流量和循环比等。即  $p = 8$ 。模型的因变量取为沥青的收率,即  $q = 1$ 。样本容量  $n = 66$ 。对建模样本,用线性最小二乘回归方法,其回归方差比为 985.06,大于临界值 2.97(显著性水平为 0.01),故可确认模型有效。而样本数据相关矩阵  $X^T X$  以及经  $r = 1$  的 OSC 校正后  $(X^{\text{OSC}})^T X^{\text{OSC}}$  的最小特征根分别为  $8.2 \times 10^{-6}$  和  $4.9 \times 10^{-6}$ ,它们均非常接近于 0,表明自变量间存在复共线性,宜选用 PLSR 方法建模。

### 2.2 试验方式与性能评价指标

以“留一法”的交叉验证方式<sup>[8]</sup>,从样本中依次取出 1 个体(序号为  $k$ )用于预测,其余  $n-1$  个个体用于训练,自检相对误差  $\text{SREr}$  和预报相对误差  $\text{PREr}$  如(4)、(5)式:

$$\text{SREr}(k) = \frac{1}{n-1} \sum_{l=1, l \neq k}^n |[\hat{y}(x_l) - y_l]/y_l| \quad (4)$$

$$\text{PREr}(k) = |[\hat{y}(x_k) - y_k]/y_k| \quad (5)$$

式中:  $l$  为训练样本序号,  $\hat{y}$  为估计值。再计算所有  $\text{SREr}(k)$  的均值,记为 MSRE,反映模型的拟合精度;计算所有  $\text{PREr}$  的均值和标准差,记为 MPRE 与 PSv,反映模型的预报精度和稳定性。按(6)、(7)式计算提取的  $h$  个成分在自变量数据总量中所占的比

率,以及对因变量的解释能力<sup>[14]</sup>

$$\xi_h = 1 - \text{tr}((X^{(h)})^T X^{(h)}) / \text{tr}(X^T X) \quad (6)$$

$$\zeta_h = 1 - \text{tr}((Y^{(h)})^T Y^{(h)}) / \text{tr}(Y^T Y) \quad (7)$$

式中  $\text{tr}(\cdot)$  为矩阵的迹。

### 2.3 结果与分析

2.3.1 模型精度与稳定性分析 OSC-PLS 方法的试验结果列于表 1,其中正交信号校正成分数(OSC factors)  $r$  取为 0~7,第二列  $h$  的取值为各行对应  $r$  下最优的 PLS 成分数。

表 1 自检误差、预报误差及预报标准差比较

$r$	$h$	MSRE/%	MPRE/%	PSv/%
0	6	0.418 4	0.473 3	0.339 8
1	5	0.417 7	0.472 2	0.339 3
2	4	0.416 1	0.472 1	0.337 6
3	3	0.416 0	0.471 9	0.335 9
4	2	0.414 2	0.471 1	0.337 6
5	1	0.379 3	0.767 9	1.164 6
6	1	0.018 7	5.275 7	3.945 3
7	1	0.000 9	5.284 2	3.960 3

先分析  $r$  对模型结构的影响,随着  $r$  的逐次增多,  $h$  呈现逐次减少,即回归模型所需提取的成分个数减少。再看  $r$  对模型性能指标的影响,随着  $r$  的逐次增多,MSRE 值开始呈相对平稳减小,但于  $r = 6$  处后急剧变小,即模型出现过拟合;而 MPRE 和 PSv 值开始也呈相对平稳减小,但于  $r = 6$  处后急剧变大,即模型泛化能力下降。同时,由表 1 看到,同最优 PLSR 模型( $r = 0, h = 6$ )相比,OSC-PLS 方法所构建的校正模型,在经过 OSC 算法一定数目  $r$  的预处理后,其拟合能力、预报能力和模型稳定性都得以保持或差别不大。从图 1 对沥青收率的显示结果也可以看出,PLS 和 OSC-PLS( $r = 4$ )两种方法所构建的模型预测能力均很好。另外,文献[1]中三种非线性方法 MEP-GRNN、TPG-GRNN 和 RBF-PLS,它们的建模结果均表现为对训练样本的过拟合,而预测能力均不如本文的 PLS 和 OSC-PLS 线性方法。

### 2.3.2 成分解释能力和模型结构分析比较

OSC-PLS 和 PLS 两种方法的  $\xi_h$  和  $\zeta_h$  列于表 2,其中 OSC-PLS 方法的正交成分数取为 4,即  $r = 4$ 。从表 2 结果可以看到,在  $h = 1$  和  $h = 2$  下,OSC-PLS 方法的  $\xi_h$  值均小于 PLS 的  $\xi_h$  值,但其  $\zeta_h$  值却均大于 PLS 方法。由此说明,OSC-PLS 所提取的成分在自变量数据的总量中所占的份额虽然不多,但它们对因变量的解释能力却明显提高。同时,从表 1 的结果可以看到,PLS 方法的最优成分数  $h = 6$ ,而

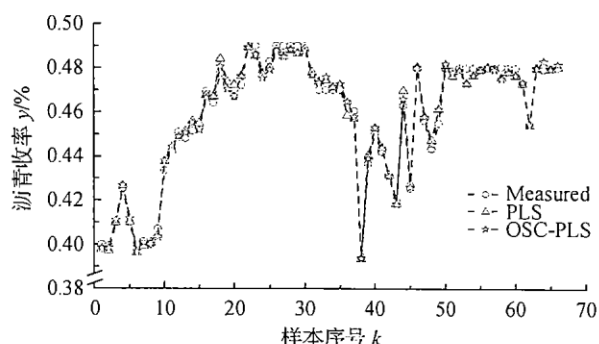


图1 沥青收率预测结果比较

OSC-PLS方法在 $r=4$ 下最优成分数 $h=2$ 。由此说明,经过OSC步骤后,模型所需提取的成分个数下降,模型结构变得更加简洁。另外,两种方法在预报精度达到最高时, $\xi_h$ 值分别为99.27%和99.23%,而 $\zeta_h$ 值均较低,分别为85.81%和72.31%,这说明在自变量中,仍有一部分数据量并不能解释因变量的变化。

表2 不同成分数下自变量方差份额和因变量解释能力比较

$h$	PLS		OSC-PLS	
	$\xi_h/\%$	$\zeta_h/\%$	$\xi_h/\%$	$\zeta_h/\%$
1	24.87	79.59	23.26	91.39
2	48.03	94.28	37.30	99.15
3	56.97	99.18	57.25	99.22
4	67.84	99.26	72.31	99.23
5	80.11	99.27	80.79	99.23
6	85.51	99.27	90.27	99.23
7	88.78	99.28	100.00	99.23
8	100.00	99.28	100.00	99.23

### 3 结 语

为建立准确的渣油裂解装置经验模型,本文采用了正交信号校正技术和偏最小二乘回归相结合的OSC-PLS方法,它先由OSC实施对输入变量测量数据的预处理,剔除数据中所含的与待测变量沥青收率无关的噪声信息,然后将其交由PLSR多元回归建模方法,进行众多沥青收率影响因素的数据降维、特征提取和回归建模。由此所构建的渣油裂解装置沥青产率的软测量校正模型,不仅在模型精度和稳定性较多种非线性方法有显著提高,而且该方法提取的PLS成分对因变量的解释能力明显提高,模型所需成分个数减少,模型结构更加简洁。

### 参考文献:

- [1] 张兵,陈德钊,俞欢军. 通用回归神经网络及其用于渣油裂解建模[J]. 浙江大学学报:工学版,2004,38(6): 653-657.
- [2] WOLD S, ANTTI H, LINDGREN F, et al. Orthogonal signal correction of near-infrared spectra [J]. Chemometrics and intelligent laboratory systems,1998,44: 175-185.
- [3] SJÖBLOM J, SVENSSON O, JOSEFSON M, et al. An evaluation of orthogonal signal correction applied to calibration transfer of near-infrared spectra[J]. Chemometrics and intelligent laboratory systems,1998,44: 229-244.
- [4] ANDERSSON C A. Direct orthogonalization[J]. Chemometrics and intelligent laboratory systems,1999,47: 51-63.
- [5] FEAN T. On orthogonal signal correction[J]. Chemometrics and intelligent laboratory systems,2000,50: 47-52.
- [6] WESTERHUIS J A, DEJONG S, SMILDE A K. Direct orthogonal signal correction [J]. Chemometrics and intelligent laboratory systems,2001,56:13-25.
- [7] TRYGG J, WOLD S. Orthogonal projections to latent structures (O-PLS) [J]. Journal of Chemometrics, 2002, 16:119-128.
- [8] GELADI P, KOWALSKI B R. Partial least-squares regression: a tutorial [J]. Analytica Chimica Acta, 1986, 185(1):1-17.
- [9] 王惠文. 偏最小二乘回归方法及其应用[M]. 北京:国防工业出版社,1999.
- [10] TRYGG J, WOLD S. O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter [J]. Journal of Chemometrics, 2003,17:53-64.
- [11] 杨民助,席西民,汪应洛. 时变系数线性模型的加权易适应最小二乘方法[J]. 西安交通大学学报,1997,31(S0):67-73.
- [12] 钱宇,姚志湘,李秀喜. 化工过程系统的状态空间分析[J]. 华南理工大学学报:自然科学版,2003,31(11):14-19.
- [13] 李春富,王桂增,叶昊. 正交信号修正法及其在软测量中的应用[J]. 信息与控制,2004,33(4):500-504.
- [14] 陈德钊. 多元数据处理[M]. 北京:化学工业出版社,1998.