

贵州烤烟中生物碱的偏最小二乘回归分析

蔡凯, 向章敏, 周淑平*, 耿召良, 葛永辉, 张婕

(贵州省烟草科学研究院, 贵州 贵阳 550081)

摘要:采用多元统计分析方法,对贵州 2010 和 2011 年 124 个烤烟样品中的 8 种生物碱进行偏最小二乘回归分析。结果表明:假木贼碱、二烯烟碱、新烟草碱及可替宁对烟碱含量的影响最大;通过偏最小二乘回归方程对烟碱含量进行预测,训练集 119 个样本的预测值与测定值之间的相对标准偏差都小于 17.04%,其中有 109 个样本小于 10.00%;用检测集 5 个样本进一步对方程进行验证,其预测值与测定值之间的相对标准偏差都小于 6.72%。

关键词:烤烟;生物碱;偏最小二乘回归分析;贵州

中图分类号: S572.01 文献标志码: A 文章编号: 1007-1032(2013)05-0463-03

Partial least squares regression analysis of the alkaloids of flue-cured tobacco in Guizhou

CAI Kai, XIANG Zhang-min, ZHOU Shu-ping*, GENG Zhao-liang, GE Yong-hui, ZHANG Jie

(Guizhou Academy of Tobacco Science, Guiyang 550081, China)

Abstract: Using the method of multivariate statistical analysis, we have conducted the partial least squares regression analysis(PLSRA) of 8 kinds of alkaloids of 124 flue-cured tobacco samples in 2010 and 2011 in Guizhou. The result shows that normicotine, nicotyrine, anatabine and cotinine have the greatest influence on nicotine content. By partial least squares regression equation to predict the content of nicotine. We have found that the relative standard deviation between predicted value of 119 samples of the training set and their measured value is less than 17.04% and that in 109 of them is less than 10.00%. At last, 5 samples of the detecting set have been applied to validate the partial least squares regression equation. The result shows that the relative standard deviation between predicted value and measured value is less than 6.72% and a good fitting has been obtained.

Key words: flue-cured tobacco; alkaloids; partial least squares regression analysis; Guizhou

生物碱是烟草的重要组成成分。生物碱决定烟叶的感官品质及可用性,直接影响烟草制品的评吸质量、烟气特征和安全性等^[1]。烟草中常见的生物碱有 8 种,即烟碱、降烟碱、麦斯明、假木贼碱、二烯烟碱、新烟草碱、2,3'-联吡啶及可替宁^[2],其中烟碱占生物碱总量的 95%左右^[3]。优质烤烟要求烟碱含量适中,烟碱含量过高,刺激性大、味苦、烟味粗糙;含量过低,则烟气没有劲头,吃味平淡。其余几种生物碱含量虽低,但可直接影响烟叶的香气品质^[4-5]。

目前,烟草生物碱之间关系的研究常采用简单相关及组成比例分析方法^[6-8],但这些方法只能同

时对 2 种生物碱进行分析,不能同时阐述多种生物碱之间的关系。笔者采用偏最小二乘回归分析^[9]方法,研究贵州烤烟烟叶中 8 种生物碱含量之间的关系,建立了烟碱与其他 7 种生物碱之间的偏最小二乘回归方程,以期开展生物碱之间的关系研究提供新方法。

1 材料与方法

1.1 材料

2010 和 2011 年于贵州 43 个县的烟区选择有代表性的采样点 62 个,取 C3F 烟叶,共 124 个样本。

收稿日期: 2012-11-09

基金项目: 贵州省烟草专卖局(公司)项目(2009-09; 2009-20; 2011-12)

作者简介: 蔡凯(1986—),男,湖南益阳人,硕士,研究实习员,主要从事烟草化学研究, caiyuenian@163.com; *通信作者, zhoushu_ping@163.com

烤烟品种为当地主栽品种 K326、云烟 87、云 97。2 年的取样点相同或类似。

1.2 烟叶生物碱含量测定

生物碱含量的测定参照文献[10]方法。样本重复测定 3 次后取平均值进行数据分析。

1.3 数据处理

采用 SIMCA-P 11.5, 对 124 个样本(119 个样本作为训练集进行建模, 5 个样本作为检测集进行预测检验)的 8 种生物碱进行偏最小二乘回归分析。以烟碱(1)为因变量 Y(mg/g), 降烟碱(2)、麦斯明(3)、假木贼碱(4)、二烯烟碱(5)、新烟草碱(6)、2,3'-联吡啶(7)和可替宁(8)为自变量 X(μg/g)进行建模。

偏最小二乘回归分析建模: 将数据标准化后, 分别提取自变量组 X 和因变量组 Y 的第 1 对主成分 T₁ 和 U₁, T₁ 和 U₁ 各自尽可能多地提取所在变量组的信息, 且使相关程度达到最大。继续按照相同的要求提取第 2 对主成分 T₂ 和 U₂, 直到回归方程达到一个较满意的精度后停止提取, 最终用提取得到的主成分建立 Y 与 X 的回归方程^[11]。SIMCA-P 软件进行偏最小二乘回归分析建模的基本步骤有模型验证、自变量的重要性指标诊断、因变量 Y 和自变量 X 之间的相关性, 最后得到偏最小二乘回归方程。

2 结果与分析

2.1 模型检验

对烤烟中 8 种生物碱进行 Pearson 相关系数分析, 得到相互之间的相关系数, 都达到了极显著 (P<0.01) 要求。用偏最小二乘回归分析法对其作进一步分析。将数据标准化后, 20 倍交叉验证进行自动拟合, 得到含有 2 个成分模型。模型累积对变量 X 的解释能力 (R²Y) 为 84.3%, 累积预测 Y 的能力 (Q²) 为 88.0%, 表明该模型的模拟效果较好。从图 1 得, 该模型 R²Y-截距和 Q²-截距分别为 -0.012 1 和 -0.129。依文献[12]可知, 可靠的模型 R²Y-截距和

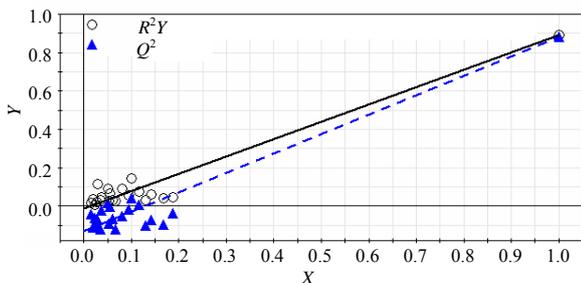


图1 偏最小二乘回归置换检验

Fig. 1 PLS permutation test

Q²-截距应分别小于 0.4 和 0.05, 表明模型可靠, 不存在过拟合现象。

2.2 重要性指标诊断

自变量(X)对因变量(Y)的解释能力常用重要性指标(VIP)来衡量。如果所有 X 对 Y 的解释能力相同, 则所有 X 的 VIP 值为 1; 如果 VIP 大于 1, 则认为该自变量对 Y 具有更强的解释能力。从图 2 可得, 假木贼碱、新烟草碱、二烯烟碱及可替宁的 VIP 值大于 1, 其他稍低于 1, 表明各自变量对 Y 有较强的解释能力, 自变量的选择较合理。

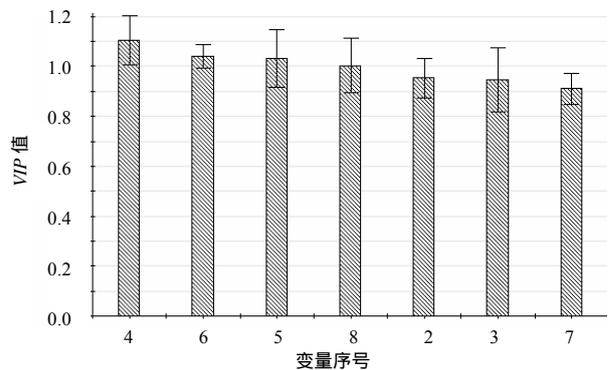


图2 自变量重要性指标

Fig.2 Variable importance in projection

2.3 烟碱与其他生物碱的相关性

从载荷图(图 3)可得, 烟碱与其他 7 种生物碱在第 1 成分上呈高度正相关。其中烟碱与假木贼碱、二烯烟碱、新烟草碱及可替宁在同一区域, 表明这 4 种生物碱对烟碱的含量影响最大。这与 VIP 值得到的结果类似。

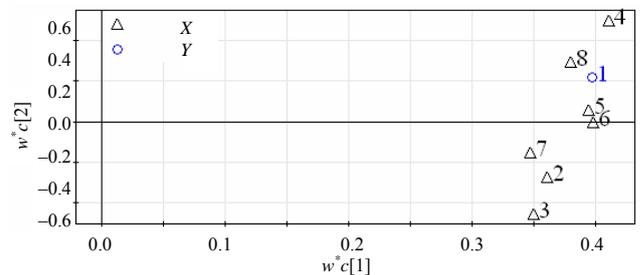


图3 烟碱与其他生物碱的载荷图

Fig. 3 Loading plot between the nicotine and other alkaloids

2.4 回归方程的建立及样本预测

拟合获得标准化方程: Y₁=3.996 4+0.051 38X₂-0.014 63 X₃+0.332 0 X₄+0.175 7 X₅+0.156 8 X₆+0.129 1 X₇-0.012 1 X₈

0.087 10 X₇+0.249 9 X₈。为了能够直接得到预测结果，将系数非标准化后得到方程： $Y_1=5.704 6+0.001 588 X_2-0.040 11 X_3+0.067 88 X_4+0.588 4 X_5+0.002 991 X_6+0.279 2 X_7+0.378 5 X_8$ 。训练集样本的拟合结果如图 4 所示。119 个初烤烟叶中，有 109 个样本的预测值与测定值相对标准偏差小于 10%，7 个样本(编号

15、67、86、93、96、103、106)小于 15%，3 个样本(编号 18、29、33)分别是 17.04%、16.00%和 16.26%。用检测集对方程进一步验证，预测结果与测定结果差别较小(表 1)，表明该模型能较好地解释 8 种生物碱之间的关系。

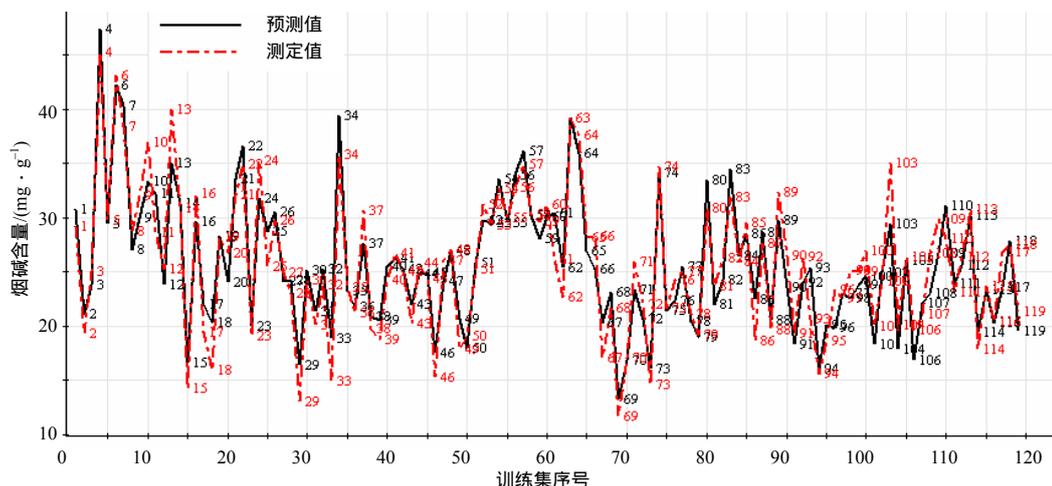


图4 训练集预测结果

Fig. 4 Prediction of training set

表1 检测集预测结果

Table 1 Prediction of detecting set				
地点	品种	生物碱检测结果/ (mg·g ⁻¹)	生物碱预测结果/ (mg·g ⁻¹)	RSD/%
贵阳	云 87	29.490	26.816	6.72
黔东南	云 87	30.653	29.148	3.56
遵义	云 97	17.535	19.055	5.87
黔西南	云 87	19.651	20.754	3.86
遵义	K326	24.446	22.822	3.56

参考文献:

[1] 张楠, 焦芳婵, 许自成, 等. 烤烟生物碱含量的杂种优势、亲子相关及与感官质量的关系分析[J]. 西南农业学报, 2011, 24(2): 490-495.

[2] 许燕娟, 白长敏, 钟科军, 等. 气相色谱/质谱分析烟草中的主要生物碱[J]. 分析化学, 2006, 34(3): 382-384.

[3] Cai J B, Liu B Z, Lin P, et al. Fast analysis of nicotine related alkaloids in tobacco and cigarette smoke by megabore capillary gas chromatography[J]. J Chromatogr A, 2003, 1017(1/2): 187-193.

[4] 史宏志, 黄元炯, 刘国顺, 等. 我国烟草和卷烟生物碱含量和组成比例分析[J]. 中国烟草学报, 2001, 7(2): 8-12.

[5] 丁丽, 盛良全, 董红武, 等. 溶剂萃取-毛细管气相色

谱法测定烟草中主要生物碱[J]. 分析化学, 2004, 32(9): 1161-1164.

[6] 廉芸芸, 王允白, 邱军, 等. 不同产区烤烟中主要生物碱含量和组成比例分析[J]. 中国烟草科学, 2008, 29(4): 6-9.

[7] 高志强, 邓小华, 周清明, 等. 湖南烤烟生物碱含量及其相关性分析[J]. 湖南农业大学学报: 自然科学版, 2008, 34(3): 288-292.

[8] 程森, 杜咏梅, 张骏, 等. 烤烟不同生物碱含量特征及其与烟叶内在质量关系研究初报[J]. 中国烟草科学, 2009, 30(6): 1-4.

[9] 薛丹. 黔东南自治州生态足迹预测研究[D]. 成都: 西南交通大学, 2009: 13-18.

[10] Cai K, Xiang Z M, Zhang J, et al. Determination of eight tobacco alkaloids in flue-cured tobacco samples by gas chromatography with nitrogen chemiluminescence detection (NCD)[J]. Anal Methods, 2012, 4(7): 2095-2100.

[11] 秦浩, 林志娟, 陈景武. 偏小二乘回归原理、分析步骤及程序[J]. 数理医药学杂志, 2007, 20(4): 450-451.

[12] Eriksson L, Johansson E, Kettaneh-Wold N, et al. Multi-and Megavariate Data Analysis [M]. Umea: Umetrics AB, 2006: 374-375.

责任编辑: 罗慧敏
英文编辑: 张健