

The Research and Application of Raohe Honey Tracing Based on Inductively Coupled Plasma Mass Spectrometer

Haihua Zhang*, Zhanfeng Ma, Zhiyong Liu

Harbin Product Quality Supervision and Inspection Institute, Harbin Heilongjiang
Email: [*zhanghaihua523@126.com](mailto:zhanghaihua523@126.com)

Received: Jun. 5th, 2015; accepted: Jun. 23rd, 2015; published: Jun. 30th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The study detected Boron isotope and Strontium isotope of 102 Raohe origin honeys and 31 non-local honeys with ICP-MS. We used principal component analysis (pca) to dimensionally reduce 7 isotope abundance ratio variables to four principal components, and selected three of the four principal components to build models. By Agilent MPP data analysis software, we built five models with the three new principal components. The models are: Decision Tree, Naive Bayes, Neural Network, Partial Least Square Discriminate and Support Vector Machine. Finally, we optimized Decision Tree model as Raohe origin honey traceability model, the resolution accuracy rate is 93.75%.

Keywords

Raohe Original Honey, ICP-MS, Principal Component Analysis, Forecasting Model

基于电感耦合等离子体质谱仪在饶河蜂蜜溯源中的应用

张海华*, 马占峰, 刘志勇

哈尔滨市产品质量监督检验院, 黑龙江 哈尔滨
Email: [*zhanghaihua523@126.com](mailto:zhanghaihua523@126.com)

*通讯作者。

收稿日期：2015年6月5日；录用日期：2015年6月23日；发布日期：2015年6月30日

摘要

本研究对102个饶河蜂蜜和31个外地蜂蜜中的硼同位素和锶同位素通过电感耦合等离子体质谱(ICP-MS)进行检测。采用主成分分析法将7个同位素丰度比值变量降维至4个主成分，并选取了主成分1、主成分2和主成分3进行模型的建立。用三个主成分新变量通过Agilent MPP数据分析软件建立Decision Tree、Naive Bayes、Neural Network、Partial Least Square Discriminate和Support Vector Machine五种模型，最终优选出以Component1、Component2和Component3交互图建立的Decision Tree模型作为饶河蜂蜜的产地溯源模型，判别准确率为93.75%。

关键词

饶河蜂蜜，ICP-MS，主成分分析，预判模型

1. 引言

饶河椴树蜜是20世纪初从乌苏里江东引入的东北黑蜂从东北特有糠椴树和紫椴树花中采集而来。饶河县从1997年设立了东北黑蜂国家级自然保护区，蜂种纯正。糠椴蜜和紫椴蜜具有高波美度。东北黑蜂抗病力强，几乎不需要使用抗生素免疫，从根本上解决了蜂产品常见的兽药残留问题[1]。

近年来，由于饶河椴树蜜以其优良的品质特性在国内外博览会多次得奖，饶河椴树蜜受到广大消费者的青睐。但与此同时，在利益的驱动下，有些商户及厂家对非饶河椴树蜜进行假冒贴牌包装，冒充饶河椴树蜜。这些冒牌饶河椴树蜜以其包装精美，价格便宜堂而皇之地进入一些大型超市和商场，这不仅欺骗了广大消费者也使饶河当地养蜂业的收入受到影响，更严重的，有些质量不好的假椴树蜜还会使消费者对饶河椴树蜜的品质产生质疑，极大的损害了饶河椴树蜜的声誉。对饶河蜂蜜原产地进行溯源和辨别技术亟待开发[2]。

食品原产地追溯和辨别技术是食品安全领域的重要技术手段，其中，同位素分析技术是公认的最有效技术之一[3] [4]。目前，食品追溯主要集中在两方面研究：1) 探索区分不同地域来源食品的有效溯源参数，为建立同位素“矿物元素溯源数据库或同位素”矿物元素地图提供理论与方法依据；2) 结合地形“气候”地质等因素探讨同位素矿物元素的变化规律[5]。

自然界中锶是第五周期IIA族元素，它具有4个同位素，即 84Sr ， 86Sr ， 87Sr ， 88Sr 。尽管动植物的吸收与代谢过程会象改变S、C、H、O和N同位素一样改变锶的同位素比率，但由于放射衰变产生的一定量的 87Sr 可作为地域溯源的指标[5]-[7]。动植物体中的 $87\text{Sr}/86\text{Sr}$ 与岩床中能被生物体利用的含锶矿物有关。锶同位素比值是判断动植物产地来源、鉴别真假的一种有效指标。当生物体中D18O和DD相同时，即在气候差异比较小的地区，锶同位素比率的判别效果比较好。

硼同位素由于不同的地球化学过程会引起其分馏效应，从而导致岩石、海洋沉积物和自然水中 $11\text{B}/10\text{B}$ 比率变化较大。硼同位素发生自然分馏的另一个重要机制是硼酸、 $\text{B}(\text{OH})_3$ 与硼酸盐离子、 $\text{B}(\text{OH})_4^-$ 之间随着pH的改变会发生交换作用，这种交换会导致硼酸中富集 11B 。以上这些自然过程会导致D11B值高达90%。硼同位素组成除受自然因素影响外，农业生产中施加含硼的化肥也会影响 $11\text{B}/10\text{B}$ 比率，这就导致不同土壤中硼同位素组成有较大差异。

针对以上两种元素同位素在地源性方面上所存在的特殊针对性, 本研究通过对饶河蜂蜜进行产地溯源来鉴别饶河蜂蜜的真假。利用ICP-MS对来自于饶河多个地区的102个蜂蜜样品和其他地域的31个蜂蜜样品中B和Sr及其同位素的含量进行分析。所得数据与化学计量学方法结合并探寻这些元素和蜂蜜样品产地之间的关系。采用主成分分析的方法, 对所有的数据采用MPP化学计量软件进行统计学分析, 并构建了Decision Tree、Naive Bayes、Neural Network、Partial Least Square Discriminate和Support Vector Machine五种判别模型。随机抽取饶河蜂蜜11个和外地蜂蜜5个, 对五个模型进行验证。

2. 实验材料与方法

2.1. 实验材料

2.1.1. 实验仪器

电感耦合等离子体质谱仪 7700 (ICP-MS, 美国 Agilent 公司), 配置玻璃同心雾化器, 镍采样锥, Ar 和 He 作为碰撞气, MARS Xpress 微波消解仪(美国 CEM 公司), MilliQ 超纯水仪。

2.1.2. 实验试剂与样品

1) 实验试剂

锶同位素标准物质: NIST987; 硼同位素标准物质 NIST951 (美国 National Institute of Science and Technology (NIST)研制)。硝酸; 双氧水。

2) 样品

样品分别采集于蜂农和五家加工企业。蜂蜜产地分别来自于大佳河镇、小佳河镇、红旗岭镇、五林洞林场、永幸林场等多个区域, 并选取其它地域有代表性的蜂蜜样品 31 个。

2.2. 实验方法

称取蜂蜜样品 0.1 g, 并加入 HNO_3 和 H_2O_2 将其微波消解至透明, 用 ICP-MS 检测样品中的 Sr84, Sr86, Sr87, Sr88, B10, B11, He 碰撞气去同量异位素干扰, 并用标准物质进行质量校正。对所测结果进行同位素丰度比并运用 Agilent MPP 软件对各同位素丰度比值进行数据分析和建立模型。

3. 结果与讨论

3.1. 主成分分析

本研究检测所得数据饶河和外地蜂蜜的同位素丰度比为 84/86 Sr, 84/87 Sr, 84/88 Sr, 86/87 Sr, 86/88 Sr, 87/88 Sr, 10/11B, 一共 7 组数据, 根据需要对 7 组数据进行降维处理, 将原来的 7 个变量重新组合成一组新的互相无关的 4 个综合变量, 同时根据实际需要从 4 个综合变量中取出 2 个尽可能多地反映原来变量的信息的综合变量进行分析。

3.1.1. 数据的降维处理

利用 MPP 计量软件对以上 7 组数据进行降维处理, 降维后的 4 个主成分的贡献率分别如下表所示:

由表 1 可知, 4 个主成分累计贡献率达到 96.32%, 数据表明降维后的 4 个主成分能代表样品的绝大部分信息, 可应用建模处理。

3.1.2. 对四个主成分分别进行两交互处理, 选择最优的两个主成分建模

1) Component 1 和 Component 2 交互图

如图 1 所示, Component 1 为 52.92%, Component 2 为 23.98%, 两个主成分贡献率之和为 76.9%, 大

Table 1. The cumulative contribution rate of four principal components

表 1. 4 个主成分累计贡献率

主成分	F1	F2	F3	F4
贡献率	52.92%	23.98%	12.52%	6.9%

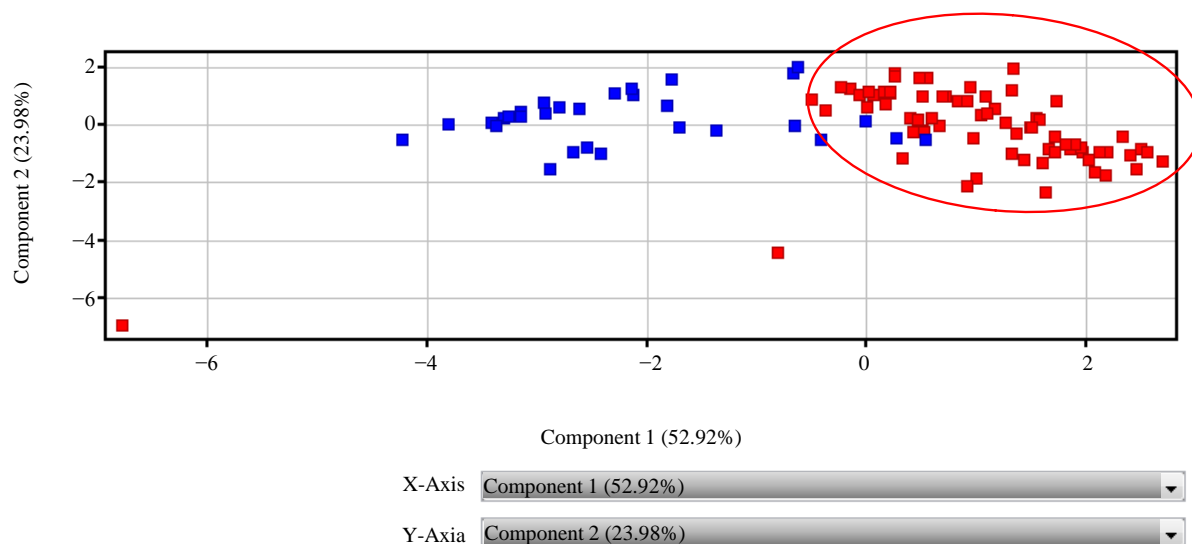


Figure 1. Interaction diagram of Component 1 and Component 2

图 1. Component 1 和 Component 2 交互图

于 65%，符合统计学的建模要求。红色代表饶河蜂蜜，深蓝色代表外地蜂蜜，饶河蜂蜜与外地蜂蜜区分明显。

2) Component 1 和 Component 3 交互图

如图 2 所示，Component 1 为 52.92%，Component 3 为 12.52%，两个主成分贡献率之和为 68.44%，大于 65%，符合统计学的建模要求。红色代表饶河蜂蜜，深蓝色代表外地蜂蜜，饶河蜂蜜与外地蜂蜜区分明显，但和图 1 相比，交叉点的个数多于图 1，并且点的分布相对松散，交互建模效果不如图 1，但仍可以满足统计学的建模需要。

3) Component 1 和 Component 4 交互图

如图 3 所示，Component 1 为 52.92%，Component 4 为 6.9%，两个主成分贡献率之和为 59.82%，小于 65%，不符合统计学的建模要求。红色代表饶河蜂蜜，深蓝色代表外地蜂蜜，图 3 可以看出，数据点分布较为松散，不能明显的对饶河蜂蜜和其他地域蜂蜜进行很好的区分，所以不能采用 Component 1 和 Component 4 交互建模。

4) Component 2 和 Component 3 交互图

如图 4 所示，Component 2 为 23.98%，Component 3 为 12.52%，两个主成分贡献率之和为 36.5%，小于 65%，不符合统计学的建模要求。红色代表饶河蜂蜜，深蓝色代表外地蜂蜜，图 4 可以看出，数据点交叉在一起，差异性不显著，所以不能采用 Component 2 和 Component 3 交互建模。

5) Component 2 和 Component 4 交互图

如图 5 所示，Component 2 为 23.98%，Component 4 为 6.9%，两个主成分贡献率之和为 30.88%，小于 65%，不符合统计学的建模要求。红色代表饶河蜂蜜，深蓝色代表外地蜂蜜，图 5 可以看出，数据点不集中，差异性不显著，所以不能采用 Component 2 和 Component 4 交互建模。

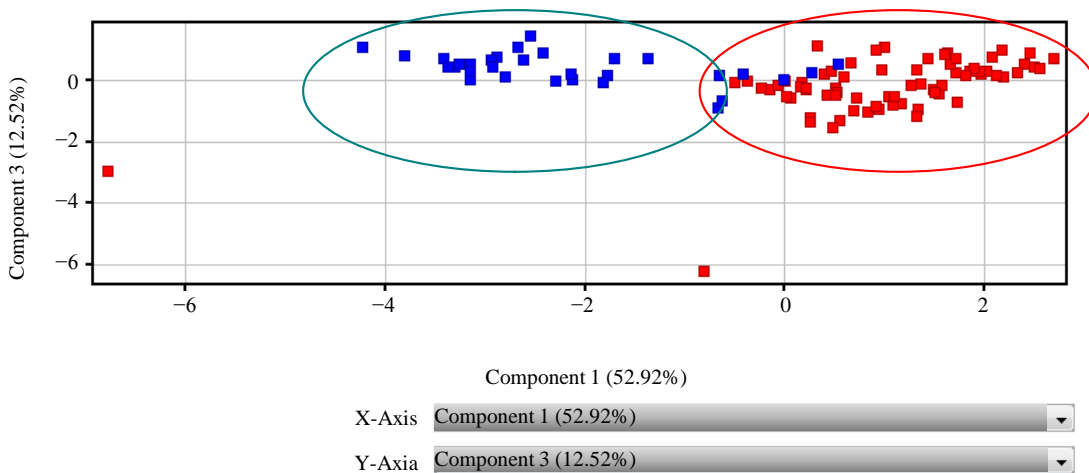


Figure 2. Interaction diagram of Component 1 and Component 3

图 2. Component 1 和 Component 3 交互图

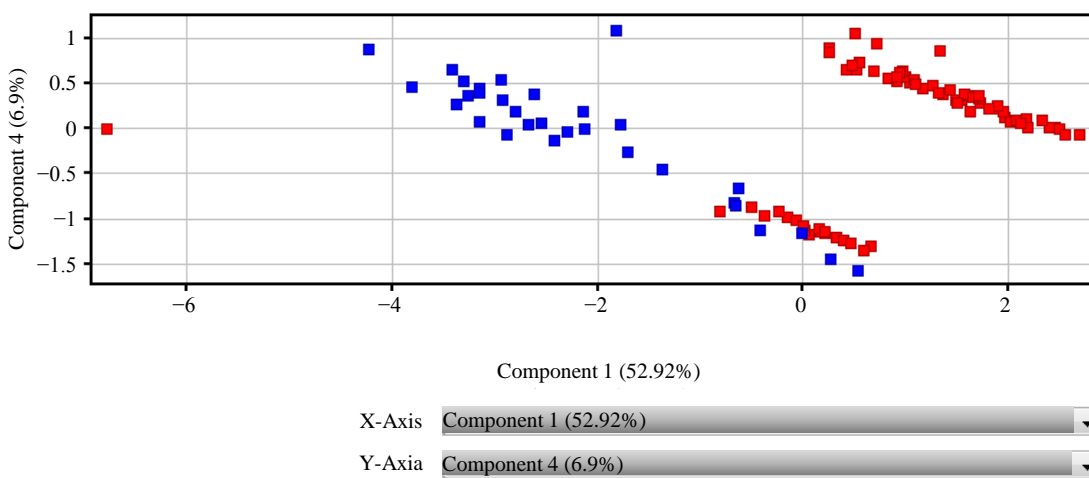


Figure 3. Interaction diagram of Component 1 and Component 4

图 3. Component 1 和 Component 4 交互图

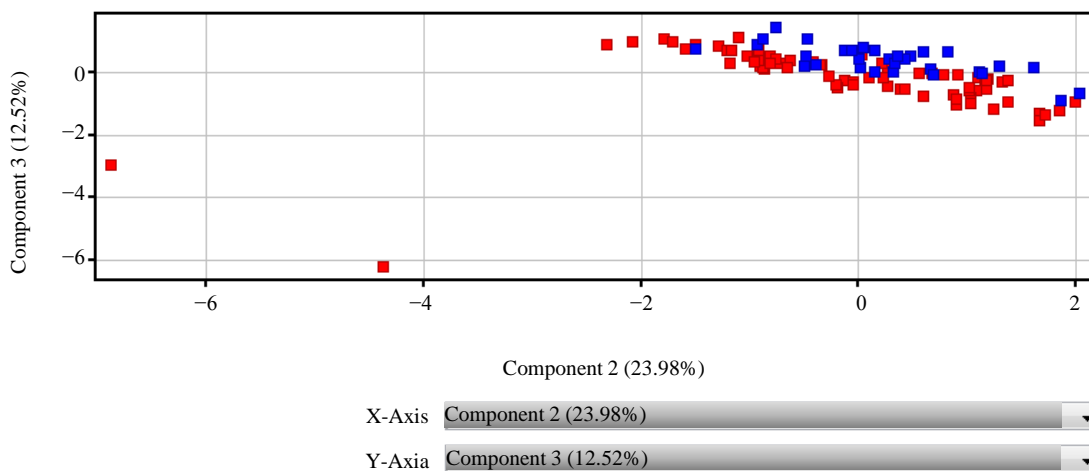


Figure 4. Interaction diagram of Component 2 and Component 3

图 4. Component 2 和 Component 3 交互图

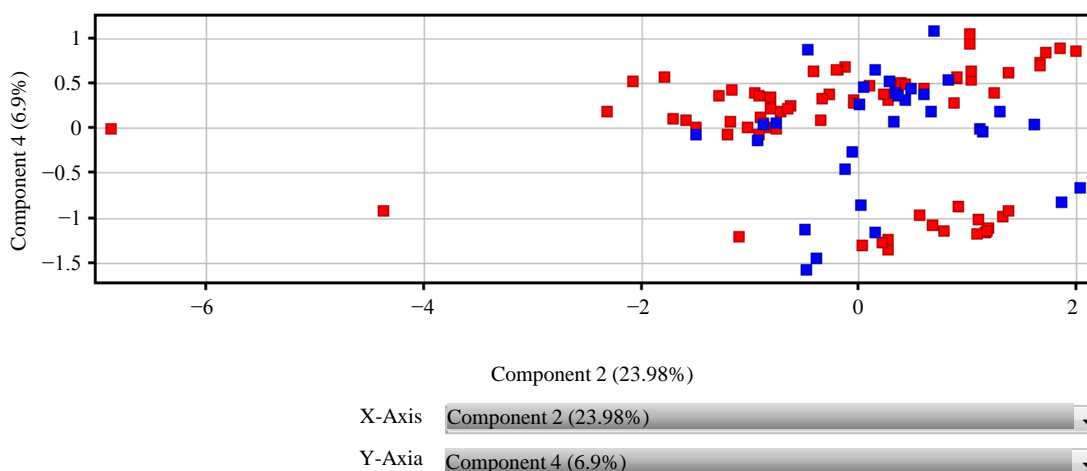


Figure 5. Interaction diagram of Component 2 and Component 4
图 5. Component 2 和 Component 4 交互图

6) Component 3 和 Component 4 交互图

如图 6 所示, Component 3 为 12.52%, Component 4 为 6.9%, 两个主成分贡献率之和为 19.42%, 小于 65%, 不符合统计学的建模要求。红色代表饶河蜂蜜, 深蓝色代表外地蜂蜜, 图 6 可以看出, 数据点不集中, 差异性不显著, 所以不能采用 Component 3 和 Component 4 交互建模。

3.2. 模型建立

通过以上分析, 可以看出主成分 1、主成分 2 和主成分 3 的贡献率较大, 相互交互图差异性较明显, 应用三个主成分建立三维立体模型, 采用三个主成分两两交互建立二维模型, 运用 Agilent MPP 软件建立 Decision Tree、Naive Bayes、Neural Network、Partial Least Square Discriminate 和 Support Vector Machine 五种模型。并随机筛选饶河蜂蜜 11 个和外地蜂蜜 5 个, 对五个模型进行验证。

如图 7 所示, 红色点代表饶河蜂蜜的数据点, 蓝色点代表其他地域蜂蜜样品的数据点, 从图中可以看出两种样品差异性显著。

3.2.1. Decision Tree 模型及预判结果

如表 2 和图 8 所示, Confident Meatur 数值都接近 1, 证明该模型可信度很高, 对应的结果显示只有一个为误判, 即 Sample 为 waidifengmi1 (表中带菱形标注的样品) 误判为饶河。总体预判准确率为 93.75%, 判定准确率 > 90%, 符合模式识别判别模型的统计学要求。

3.2.2. Naive Bayes 模型及预判结果

如表 3 所示, Naive Bayes 模型中 Confident Measurer 值有几个较低, 不如 Decision Tree 高, 该模型结果显示有 5 个样品(表中带菱形标注的样品)为误判, 总体预判准确率为 68.75%, 不符合统计学模式识别模型判定的要求(判别正确率 $\geq 90\%$)。

3.2.3. Neural Network 模型及预判结果

如表 4 所示, Neural Network 模型有 3 个样品(表中带菱形标注的样品)进行了错误预判, 总体预判准确率为 81.25%。不符合统计学模式识别模型判定的要求(判别正确率 $\geq 90\%$)。

3.2.4. Partial Least Squares Discrimination 模型及预判结果

如表 5 所示, Neural Network 模型有 4 个样品(表中带菱形标注的样品)进行了错误预判, 总体预判准

确率为 75%。不符合统计学模式识别模型判定的要求(判别正确率 $\geq 90\%$)。

3.2.5. Support Vector Machine 模型及预判结果

如表 6 和图 9 所示, Support Vector Machin Confident Measurer 值较低, 认为模型可信度不够高。模型中存在 1 个样品(表中带菱形标注的样品)被误判, 总体预判准确率为 93.75%。判定准确率 $> 90\%$, 符合模式识别判别模型的统计学要求。

以上分析中省略了部分不合格样品二维主成分相互交互图。

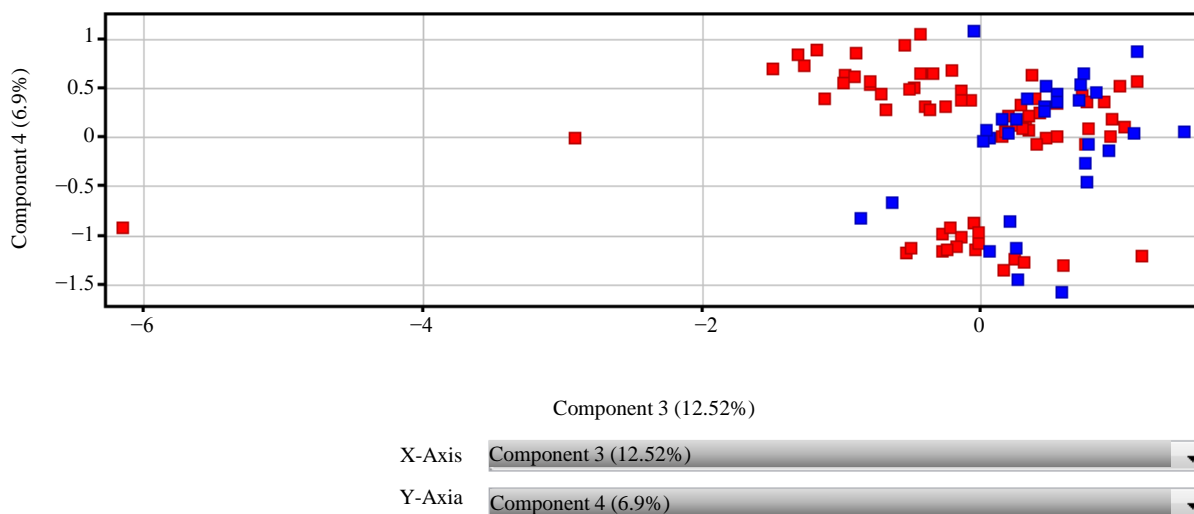


Figure 6. Interaction diagram of Component 3 and Component 4

图 6. Component 3 和 Component 4 交互图

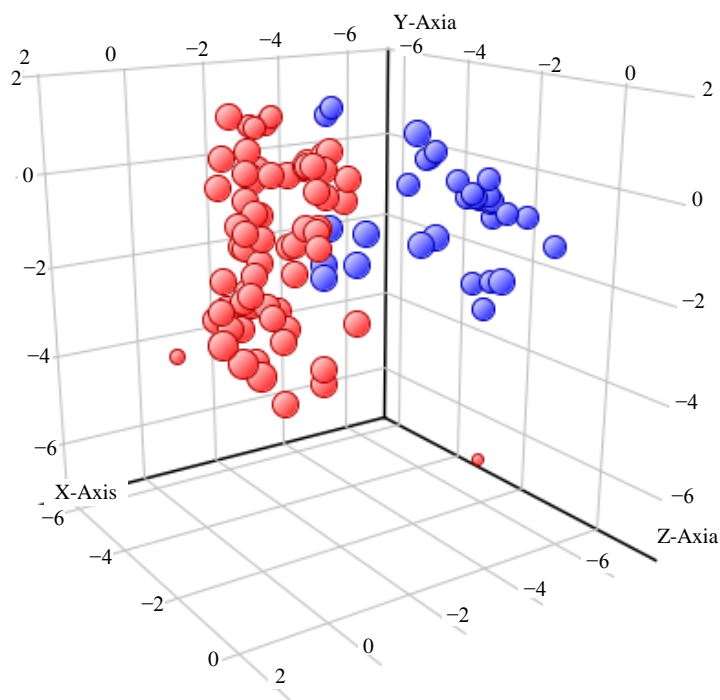


Figure 7. 3D ditribution diagram

图 7. 3D 分布图

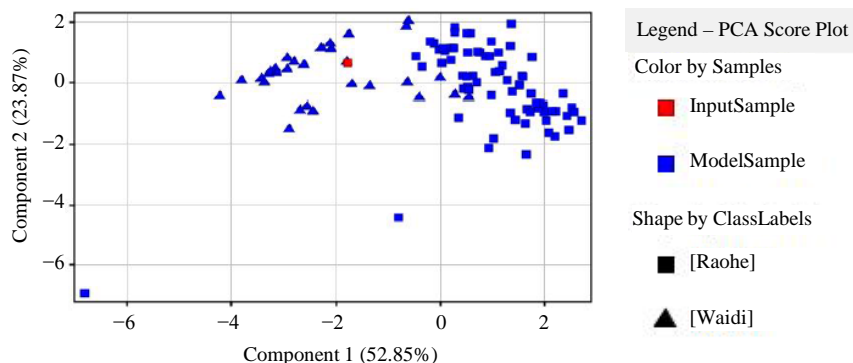


Figure 8. Component 1 and Component 2 interaction diagram of waidifengmi 1 in Decision Tree model

图 8. Decision Tree 模型中 waidifengmi 1 的 Component 1 和 Component 2 交互图

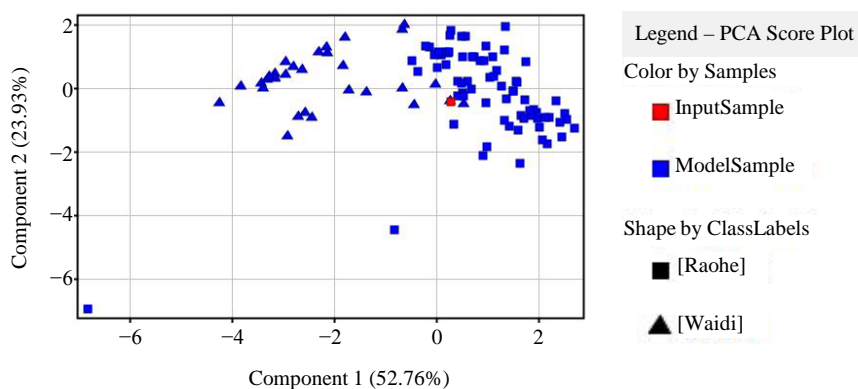


Figure 9. Sample discrimination diagram of the Sample is 5 in Support Vector Machine model

图 9. Support Vector Machine 模型中的 Sample 为 5 的样品判别图

Table 2. The prediction result of Decision Tree model

表 2. Decision Tree 模型预判结果

Sample	Predicted	Confident Measure
RF11	Raohe	0.9861111
hfy19	Raohe	0.9861111
17	Waidi	0.9677419
◆ waidifengmi1	Raohe	0.9861111
RF15	Raohe	0.9861111
RF1	Raohe	0.9861111
laoyingou5	Raohe	0.9861111
HFY2	Raohe	0.9861111
30	Waidi	0.9677419
5	Waidi	0.9677419
31	Waidi	0.9677419
HFY20	Raohe	0.9861111
HFY21	Raohe	0.9861111
yongxing15	Raohe	0.9861111
HFY3	Raohe	0.9861111
HFY4	Raohe	0.9861111

Table 3. The prediction result of Naive Bayes model
表 3. Naive Bayes 模型预判结果

Sample	Predicted	Confident Measurer
RF11	Raohe	1.0
hfy19	Raohe	1.0
◆ 17	Raohe	0.95757383
◆ waidifengmi1	Raohe	1.0
RF15	Raohe	1.0
RF1	Raohe	1.0
laoyinggou5	Raohe	1.0
HFY2	Raohe	0.8869561
30	Waidi	0.9999964
◆ 5	Raohe	0.79181457
31	Waidi	0.9999988
◆ HFY20	Waidi	0.7664125
HFY21	Raohe	0.7178487
yongxing15	Raohe	1.0
HFY3	Raohe	1.0
◆ HFY4	Waidi	1.0

Table 4. The prediction result of Neural Network model
表 4. Neural Network 模型预判结果

Sample	Predicted	Confident Measurer
RF11	Raohe	0.9611365
hfy19	Raohe	0.9925651
◆ 17	Raohe	0.79242414
Waidifengmi1	Waidi	0.9553002
RF15	Raohe	0.99281746
RF1	Raohe	0.9895421
Laoyinggou5	Raohe	0.9927961
HFY2	Raohe	0.9827965
30	Waidi	0.9793519
◆ 5	Raohe	0.79376185
31	Waidi	0.97935563
HFY20	Raohe	0.7371074
HFY21	Raohe	0.97476923
Yongxing15	Raohe	0.9930735
HFY3	Raohe	0.99306345
◆ HFY4	Waidi	0.9422084

Table 5. The prediction result of Partial Least Squares Discrimination model
表 5. Partial Least Squares Discrimination 模型预判结果

Sample	Predicted	Confident Measurer
RF11	Raohe	1.0
hfy19	Raohe	1.0
17	Waidi	1.0
waidifengmi1	Waidi	1.0
RF15	Raohe	1.0
RF1	Raohe	1.0
laoyinggou5	Raohe	1.0
◆ HFY2	Waidi	1.0
30	Waidi	1.0
5	Waidi	1.0
31	Waidi	1.0
◆ HFY20	Waidi	1.0
◆ HFY21	Waidi	1.0
yongxing15	Raohe	1.0
HFY3	Raohe	1.0
◆ HFY4	Waidi	1.0

Table 6. The prediction result of Support Vector Machine model
表 6. Support Vector Machine 模型预判结果

Sample	Predicted	Confident Measurer
RF11	Raohe	1.0
hfy19	Raohe	0.3352127
17	Waidi	0.021802539
waidifengmi1	Waidi	0.055281788
RF15	Raohe	0.65562
RF1	Raohe	0.13493183
laoyinggou5	Raohe	0.29780075
HFY2	Raohe	0.19093436
30	Waidi	0.64581287
◆ 5	Raohe	0.076684095
31	Waidi	0.46378288
HFY20	Raohe	0.0061945207
HFY21	Raohe	0.23153588
yongxing15	Raohe	0.63791883
HFY3	Raohe	0.5314536
HFY4	Raohe	1.0

3.3. 讨论

分析图 1 到图 6 的 6 个交互判别图可以发现, Component 1 和 Component 2 交互图和 Component 1 和 Component 3 交互图差异性显著, 能够满足统计学建模的要求(累计贡献率 $\geq 65\%$), 而其他四个图不满足该条件, 可作为参考。对五个模型所给出的预判结果进行观察, 发现 Decision Tree、Neural Network、Partial Least Square Discriminate 和 Support Vector Machine 都仅有 1 个样品被误判, 判定准确率大于 90%, 符合统计学建模的要求, 其余 3 个模型则不符合该条件, 故舍弃。而 Decision Tree 模型的 Confident Measurer 的值明显高于 Support Vector Machine 模型, 说明 Decision Tree 模型溯源更为可信, 故本研究最终采用了以 Component 1、Component 2 和 Component 3 交互图建立的 Decision Tree 模型作为饶河蜂蜜的产地溯源模型。

4. 结论

运用电感耦合等离子体质谱(ICP-MS)结合主成分分析和建立预判模型对饶河蜂蜜进行溯源是可行的。Decision Tree 和 Support Vector Machine 模型能够准确的辨别饶河和外地蜂蜜。对于 Naive Bayes 和 Neural Network 模型, 在以后可以通过调整主成分个数, 参数范围来进行完善, 以提高其预判准确率。

基金项目

国家质检总局科技计划项目, 项目编号: 2013QK221。

参考文献 (References)

- [1] 别梅 (2012) 蜂产品中多种农兽药残留同步检测技术研究. 硕士论文, 天津商业大学, 天津.
- [2] 张龙 (2012) 植源性农产品溯源以及鉴别技术研究. 硕士论文, 浙江大学, 杭州.
- [3] 李玉梅 (2002) 碳、氧稳定同位素分析在古环境研究中的应用. 博士学位论文, 中国科学院地质与地球物理研究所, 北京.
- [4] 王国安 (2001) 中国北方草本植物及表土有机质碳同位素组成. 博士学位论文, 中国科学院地质与地球物理研究所, 北京.
- [5] 马英军, 刘丛强 (2001) 花岗岩化学风化过程中的 Sr 同位素演化——矿物相对风化速率的影响. *中国科学(D 辑)*, **8**, 634-640.
- [6] Åberg, G. (1995) The use of natural strontium isotopes as tracers in environmental studies. *Water, Air and Soil Pollution*, **79**, 309-322.
- [7] Capo, R.C., Stewart, B.W. and Chadwick, O.A. (1998) Strontium isotopes as tracers of ecosystem processes: Theory and methods. *Geoderma*, **82**, 197-225.