

基于电子鼻的小麦品种鉴别研究

周博, 王俊*

(浙江大学生物系统工程与食品科学学院, 杭州 310029)

摘要: 提出了一种用电子鼻技术快速鉴别小麦品种的新方法。试验以三种小麦品种为研究对象, 首先用主成分分析法、聚类分析法对5种不同温度条件下的麦9023电子鼻数据进行了分析, 确定了试验条件, 保证了试验的可重复性和准确性。用三种小麦第一次测试的结果(每个品种15个样品, 共45个样品)作为训练集来建立模型, 用一周后的测试结果(每个品种5个样品, 共15个样品)作为测试集。分别采用主成分分析法对三种小麦进行了区分, 逐步判别分析法和BP神经网络对三种小麦样品进行了预测。使用主成分分析法不能很好对两次试验的三种小麦样品做出很好的区分, 而使用逐步判别分析法对训练集回判的正确率为100%, 对测试集判别正确率为86.7%。选用三层BP神经网络对训练集回判的正确率为100%, 对测试集判别正确率为93.3%。说明在选取适合的试验条件的情况下, 电子鼻对小麦品种具有很好的鉴别作用, 为小麦品种的鉴别提供了一种新方法。

关键词: 电子鼻; 小麦; 主成分分析; 逐步判别分析; BP神经网络

0 引言

小麦是我国重要的粮食作物之一。小麦品种鉴别是目前农业生产、作物育种和种子检验的重要问题。国内外小麦品种鉴别的常用方法主要有蛋白质电泳鉴定法、DNA分子标记、形态学鉴定、田间鉴别法等^[1]。蛋白质电泳鉴定法、DNA分子标记方法鉴别精度高, 但所需的时间也很长, 而且鉴别过程烦琐, 不适宜对样品进行批量分析和无损在线监测。形态学方法、田间鉴定方法鉴定周期长, 并且容易受环境影响, 鉴别精度不高。

电子鼻是根据仿生学原理, 由选择性的气敏传感器阵列和适当的模式识别系统组成, 是一种新颖的分析、识别和检测复杂气味和大多数挥发性成分的仪器。它与普通的化学分析仪器, 如色谱仪、光谱仪等不同, 得到的不是被测样品中某种或某几种成分的定性与定量结果, 而是给予样品中挥发成分的整体信息, 也称“指纹”数据^[2]。在国内外, 电子鼻在农产品检测中已经有广泛的应用, 如水果新鲜度、肉类、饮料、谷物等^[3-6]。近年来, 许多学者研究了谷物的霉变和微生物污染^[4-7], 以及谷物的储藏条件和时间^[8-9], 但是使用电子鼻对小麦品种的鉴别还没有报道。本文探讨了使用电子鼻进行谷物品种鉴别的条件, 使用逐步判别分析和神经网络建立三种谷物品种鉴别模型, 并对比了模型的鉴别结果。

1 材料与方法

1.1 电子鼻

试验是用德国Airsense公司的PEN2便携式电子鼻

(Portable Electronic nose), 这一电子鼻包含有10个金属氧化物传感器组成的阵列, 传感器对不同的化学成分有不同的响应值。响应信号为传感器接触到样品挥发物后的电导率G与传感器在经过标准活性碳过滤气体的电导率G₀的比值, 用G/G₀来表示。仪器组成主要包含: 传感器通道、采样通道、计算机。该电子鼻具有自动调整、自动校准及系统自动富集的功能。数据分析软件为PEN2电子鼻自带的Winmuster, 通用统计分析软件DPS(Data processing system for practical statistics)和Matlab7.5。

1.2 样品来源及数据采集

试验用的三种小麦种子由杭州东南面粉厂提供, 分别是麦18、麦13、麦9023均适用于江苏淮南地区种植的冬小麦, 小麦种子从外表很难区分。实验前种子被置于冰箱中存放(4℃)。样本从完整、无霉变的种子中挑选。样本盛放在统一容积的烧杯中, 在电子鼻检测实验时, 在4℃下冷藏下取出的样品需先放在室温3-4小时、待升温至室温后再进行实验。试验前样品在恒温状态下用保鲜膜密封, 后进行电子鼻自动顶空取样。采样时间设定为60s, 清洗时间设定为50s。

1.3 数据处理方法

试验得到的每个样品电子鼻数据是10×60(10个传感器, 采样时间60s, 每秒采集1次)的矩阵, 一般取传感器稳定后某一时刻的数据建立数据模型矩阵, 本试验取采样第59s的数据建立初始数据矩阵。由于数据量众多, 选取合适的数据处理方法是十分必要的。主成分分析(PCA)是多元统计中的一种数据挖掘技术。通过对电子鼻传感器数据进行转换, 使数目较少的新变量成为原变量的线性组合, 新变量能最大限度地表征原变量的数据结构特征。判别分析主要是判别未知样本应划归哪一个已知总体, 逐步判别分析是具有筛选能力的判别分析方法, 是在多组判别基础上发展起来的一种方法, 判别准则为贝叶斯判别函数。BP人工神经网络是一种单向传

* 收稿日期:

修订日期:

作者简介: 周博(1971), 男(汉), 湖南, 博士生, 浙江大学生物系统工程与食品科学学院, 邮编310029。Email: zjzhuobo@163.com

通讯作者: 王俊(1965), 男(汉), 浙江, 博士, 教授, 博士生导师, 浙江大学生物系统工程与食品科学学院, 邮编310029。Email: jwang@zju.edu.cn

播多层前馈神经网络，由于其结构简单、可塑性强，故在模式识别、信息分类等领域获得了广泛的应用。

2 试验条件确定

2.1 不同温度电子鼻的响应

图 1 是不同温度下麦9023 的电子鼻的传感器电导率比值随时间变化的曲线图，图 2 是麦9023 在不同温度条件下电子鼻10个传感器响应信号第60s 的雷达图，从图 1中可以看出，电导率比值刚开始接近于 1（与G0相近），随着挥发物在传感器表面的富集，传感器的电导率出现明显变化，随着采样时间的延长在40s 后基本趋于平缓，达到一个稳定的状态。从图 1 中还可以看出，小麦在不

同温度下，电子鼻的响应曲线是不同的，20℃ 时曲线很平缓，随着温度的升高曲线的变化范围也随之扩大。这说明随着温度升高，小麦挥发物的含量也在增多，这从另一方面也说明在常温下小麦挥发物含量比较少，而且挥发物含量受环境影响大，这对用电子鼻来进行小麦品种鉴别是不利的，为了做到对小麦品种准确的鉴别，必须保证试验条件的一致性。从图 2 中可以看出，随着温度的变化每个传感器的变化趋势是一致的。这说明在20℃-50℃范围内，小麦挥发性物的浓度发生了改变但成分没有发生改变，选择合适的条件是可以对小麦品种进行鉴别的。

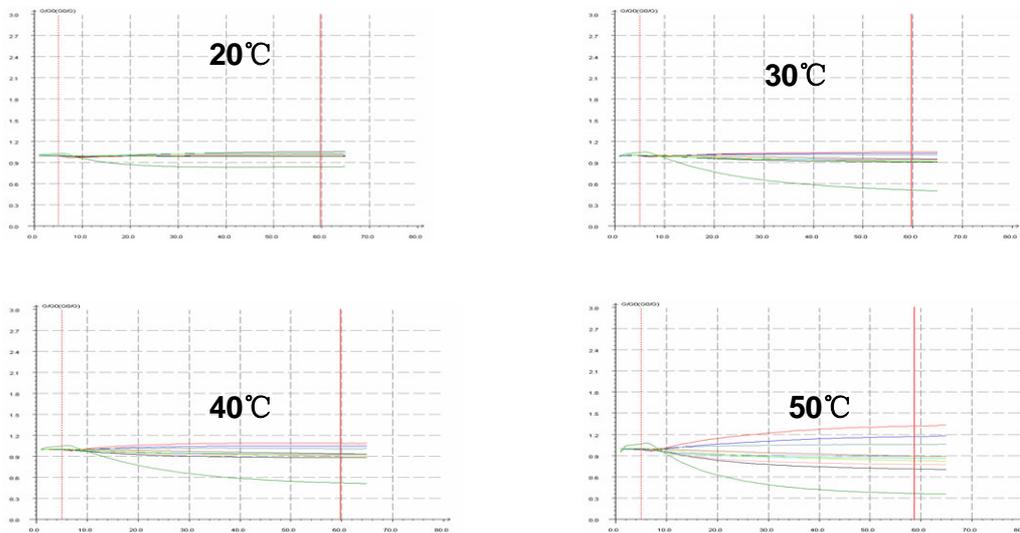


图 1. 电子鼻传感器对不同温度麦 9023 的响应曲线

Fig.1 Responses of the sensors array of the electronic nose to wheat-9023 at different temperature

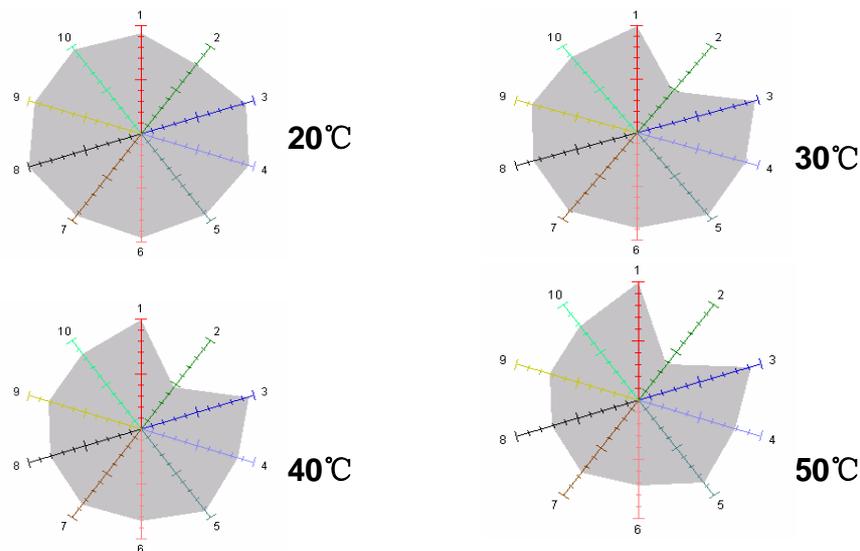


图 2 电子鼻响应 60s 不同温度下麦 9023 的雷达图

Fig.2 Radar charts of electronic nose responses at 60s to wheat-9023 at different temperature

综合以上分析表明，温度对小麦样品影响大，所以试验条件主要确定检测时的温度。将检测温度分别设置为30℃、35℃、40℃、45℃、50℃，共5组麦9023样品，每组5个样品，每样品50g 分别放入500ml 烧杯中，置于恒温箱内。恒温90min 后进行电子鼻顶空取样。

2.2 不同温度的主成分分析

从主成分分析图中可以看出，五种不同温度下的小麦第一主成分的分析值为88.91%（横坐标PC1）与第二主成分的分析值为8.20%（纵坐标PC2）之和已达90%以上，

表明PCA第一主成分、第二主成分的二维图可以代表原始数据的总体信息；随着温度的增加，样品在图中第一主成分的值也随之增大，说明样品电子鼻数据的差异主要是由于温度的不同引起的，第一主成分可以代表样品的温度差异；图中30℃、35℃、40℃、45℃四种样品虽然有差异，但相互之间都有重叠的情况，而50℃时能与其它样品很好的区分并且集中度最好，说明50℃时小麦的试验数据最好。

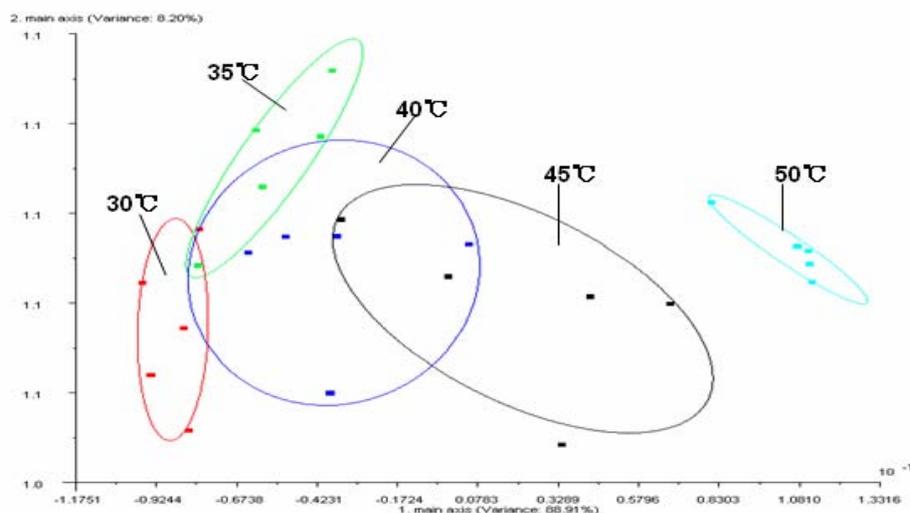


图3 五种不同温度麦9023的主成分分析图

Figure.3 Principal component analysis of wheat-9023 at five different temperature

2.3 聚类分析

聚类分析是研究样品分类问题的一种多元统计分析方法。聚类分析是在事前没有分类的情况下对未知数据结构进行分类的方法，是将分类对象置于一个多维空间中，按照它们空间关系的亲疏程度进行分类。选用分层聚类分析方法，分类原则采用最短距离法，分类数量设定为5类。图4和表1是5组样品电子鼻取样59s 时的数据所做的聚类分析结果（a、b、c、d、e分别代表30℃、35℃、40℃、45℃、50℃时的样品）。从聚类结果中可以看出，13个样品被分为1类，2个样品被分为2类，4个样品被分为3类，5个样品被分为4类，1个样品被分为5类。多数样品被分为1类，说明样品间的差异不是很显著，这与主成分分析中前四组样品有相互重叠的情况相符合，50℃时的样品有4个被分为4类，与其它样品区分度最为明显，说明50℃的样品与其它组样品的差异较大，集中度较好，这也与主成分分析的结果相符合。

综合分析结果，样品在50℃时与其它组差异大，并且集中度好，试验结果稳定。所以进行三种小麦品种试验时选择试验条件为：试验温度50℃、恒温时间90min、

每样品重量50g、烧杯容积500ml。

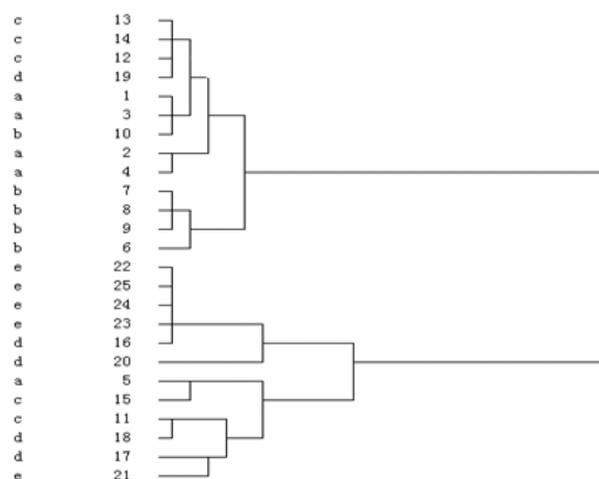


图4 五种不同温度麦9023的聚类分析图

Figure.4 Cluster analysis of wheat-9023 at five different temperature

表 1 五种不同温度麦 9023 的聚类分析结果

Table 1 Results of cluster analysis for wheat-9023 at five different temperature

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
初始类	a	a	a	a	a	b	b	b	b	b	c	c	c	c	c	d	d	d	d	d	e	e	e	e	e
分类	1	1	1	1	2	1	1	1	1	1	3	1	1	1	2	4	3	3	1	5	3	4	4	4	4

注: a、b、c、d、e分别代表30℃、35℃、40℃、45℃、50℃时的样品

3 试验结果与分析

为了验证试验的可重复性, 试验分两次进行。第一次, 三种小麦各取15个样品, 在设定试验条件下进行, 得到的试验数据作为初始的训练样本; 第二次为间隔1周后, 每组5个样品, 作为预测样本。以W18、W13、W9023分别代表麦18、麦13、麦9023三类小麦品种。采用主成分分析对试验数据进行处理, 判断区分效果, 用逐步判别分析和BP神经网络建立模型对样品进行预测。

3.1 主成分定性分析

对第一次试验数据进行主成分分析得到前两个主成分的贡献率分别为: 83.8%、11.59%, 累积贡献率为: 95.39%。分类结果如图5, 可以看出3个品种的小麦可以被的区分。对第二次试验数据进行主成分分析得到前两个主成分的贡献率分别为: 87.37%、7.8%, 累积贡献率为: 95.17%。分类结果如图6, 可以看3种小麦在主成分图中的位置与图6相似, 但麦18与麦13有交叉重叠。将两次实验数据同时进行主成分分析得到前两个主成分的贡献率分别为: 82.23%、10.73%, 累积贡献率为: 92.96%。分类结果如图7, 可以看出3种小麦第一次与第二次试验均有重叠, 但重叠程度不同, 麦9023的重合度最好, 麦13重合度一般, 麦18重合度最差。

比较图5和图6可以看出, 两次试验结果主成分分析有差异, 说明随着时间的延长, 小麦挥发性成分会产生轻微的改变, 对试验的重复性有影响。从图7可以看出, 麦18与麦13有重叠, 说明仅用主成分分析方法不能对三种小麦进行很好的区分。

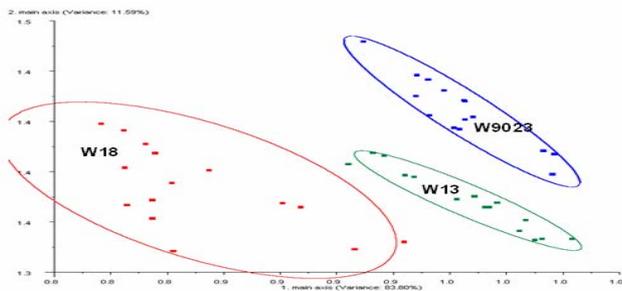


图 5 第一次试验 45 个样品主成分 (主成分 1、主成分 2) 得分图
Fig.5 Scores plot obtained from the PCA (PC1 vs PC2) of the first 45 samples

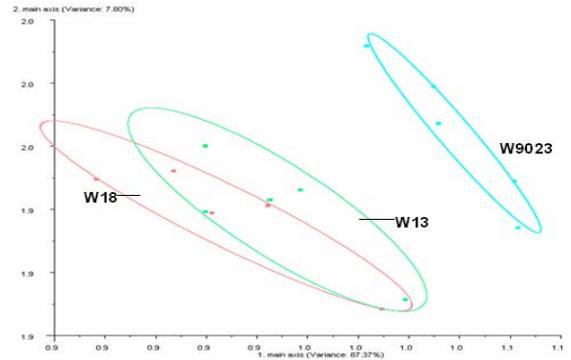


图 6 第二次试验 15 个样品主成分 (主成分 1、主成分 2) 得分图
Fig.6 Scores plot obtained from the PCA (PC1 vs PC2) of the second 15 samples

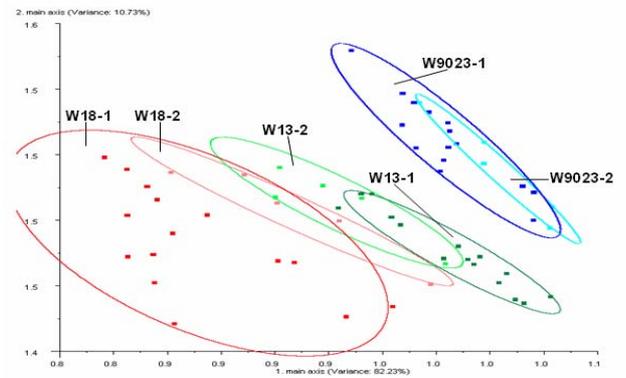


图 7 两次试验的主成分 (主成分 1、主成分 2) 得分图
Fig.7 Scores plot obtained from the PCA (PC1 vs PC2) of the two experiments

3.2 逐步判别分析模型预测

以符号1、2、3分别代表麦18、麦13、麦9023三种小麦品种, 用三种样品第一次测试的结果 (每个品种15个样品, 共45个样品) 作为训练集来建立模型, 用一周后的测试结果 (每个品种5个样品, 共15个样品) 作为测试集, 进行逐步判别分析。训练集样本的回判结果见表2, 正确率均为100%, 说明所建立的判别模型是可靠的。

对未参与建立模型的15个测试集样本进行验证。15个样本依次是麦18、麦13、麦9023各5个。其中类1代

表麦18, 类2代表麦13, 类3代表麦9023, 得到表3. 从表3中我们可以看到, 除了麦13(2类)中的(9)和(10)号样本被误判为麦18(1类)其它样品均判别正确, 模型对测试集的预测正确率达到86.7%。

表2 逐步判别分析对训练集回判结果

Table 2 Discrimination rate of training set by SDA

来自 \ 判为	麦18	麦13	麦9023	正确率%
麦18	15	0	0	100
麦13	0	15	0	100
麦9023	0	0	15	100

表3 逐步判别分析对测试集的判别结果

Table 3 Discrimination of testing set by SDA

样本序号	真实类	预测类	后验概率	样本序号	真实类	预测类	后验概率	样本序号	真实类	预测类	后验概率
(1)	1	1	1	(6)	2	2	0.76	(11)	3	3	1
(2)	1	1	1	(7)	2	2	0.90	(12)	3	3	1
(3)	1	1	1	(8)	2	2	0.74	(13)	3	3	1
(4)	1	1	1	(9)	2	1	0.99	(14)	3	3	1
(5)	1	1	1	(10)	2	1	1	(15)	3	3	1

3.3 BP神经网络模型预测

试验建立一个三层的BP神经网络结构, 用10个传感器的响应值作为神经网络的输入变量建立预测模型, 小麦类别作为网络的输出。通过反复测试来调整隐含层神经元数来优化网络结构, 得到最佳网络结构为10(输入) - 18(隐含) - 1(输出)的三层BP神经网络。设定目标误差为0.01, 网络学习速率为0.05, 训练迭代次数为1000

次。设定预测误差的阈值为0.2, 即预测误差大于0.2就算分类错误。模型对于训练集样品回判的正确率为100%, 对预测集样本的预测结果如表4所示, 除了麦13中的(10)号样本的预测值为1.5740, 误差超过了设定的阈值, 其它样本的预测值均在误差范围之内, 模型对于测试集的预测正确率为93.3%。

表4 BP神经网络对测试集的判别结果

Table 4 Discrimination of testing set by BPNN

样本序号	真实值	预测值	样本序号	真实值	预测值	样本序号	真实值	预测值
(1)	1	1.1536	(6)	2	2.0126	(11)	3	3.0532
(2)	1	0.9913	(7)	2	1.9688	(12)	3	2.9988
(3)	1	1.0008	(8)	2	1.8711	(13)	3	3.0246
(4)	1	1.1211	(9)	2	1.8891	(14)	3	3.0044
(5)	1	1.1249	(10)	2	1.5740	(15)	3	3.0067

4 结论

小麦品种在检测中受温度影响最大, 在温度为50℃的试验条件下小麦的试验重复性最好。由于时间差异造成样品品质有轻微改变, 用主成分分析方法不能对三种不同时间检测的小麦进行很好的区分。用逐步判别分析法和BP神经网络对三种小麦样品进行了预测, 训练集的判别正确率均为100%, 但对测试集判别用逐步判别分析法正确率为86.7%, 用BP神经网络的正确率为93.3%, BP神经网络的判断正确率高于逐步判别分析方法。

[参 考 文 献]

- [1] 李汝玉 李群 等. 利用 SSR 标记进行中国小麦品种鉴定的研究[J]. 种子, 2008, (02) :91~96.
- [2] 张红梅 王俊. 电子鼻传感器阵列优化及其在小麦储藏年限检测中的应用[J]. 农业工程学报, 2006, (12)
- [3] M. Peris, L. Escuder-Gilabert, A 21st century technique for food control: electronic noses[J]. Analytica Chimica Acta (2009) : 1921-1926
- [4] Roberto Paolesse, Adriano Alimelli Detection of

fungal contamination of cereal grain samples by an electronic nose[J]. *Sensors and Actuators B* 119 (2006) 425 - 430

- [5] S. Balasubramaniana, S. Panigrahi Evaluation of an artificial olfactory system for grain quality discrimination[J]. *LWT* 40 (2007) 1815 - 1825.
- [6] D.S. Presicce, A. Forleo, A.M. Taurino Response evaluation of an E-nose towards contaminated wheat by *Fusarium poae* fungi[J]. *Sensors and Actuators B* 118 (2006) 433 - 438
- [7] D. Abramson, R. Hulasare Mycotoxins, ergosterol, and odor volatiles in durum wheat during granary storage at 16% and 20% moisture content[J]. *Journal of Stored*

Products Research, Volume 41, Issue 1, 2005, Pages 67-76

- [8] Hongmei Zhang, Jun Wang Optimization of sensor array and detection of stored duration of wheat by electronic nose[J]. *Journal of Food Engineering* 82 (2007) 403 - 408
- [9] Hongmei Zhang, Jun Wang Detection of age and insect damage incurred by wheat, with an electronic nose[J]. *Journal of Stored Products Research* 43 (2007) 489 - 495

Discrimination of Varieties of Wheat seeds Based on Electronic Nose

Bo Zhou, Jun Wang*

(Department of Agricultural Engineering, Zhejiang University, 268 Kaixuan Road, Hangzhou 310029, PR China)

Abstract: A new method for discrimination of varieties of wheat seeds by means of electronic nose (E-nose) was developed. The wheat seeds used for this study involved three varieties: wheat-18, wheat-13, wheat-9023. Firstly, the principal component analysis (PCA) and cluster analysis (CA) were used to classify five groups of wheat 9023 at different temperature. The experimental conditions were fixed. Secondly, the data model was set up by two experiments. In this model, 45 samples of the first experiment were used as the training data set and 15 samples of the second experiment which performed a week later after the first experiment were used as the testing data set. Thirdly, the PCA was used to distinguish three varieties of wheat seeds. The analysis results showed that the PCA could not distinguish the three varieties of wheat seeds properly. The stepwise discriminant analysis (SDA) and a three-layer back-propagation neural network (BPNN) were developed for pattern recognition models. Calculation results showed that both models for training data set the discrimination rate of the three wheat seed varieties was 100 %, for testing data set the SDA discrimination rate was 86.7% and the BPNN was 93.3%. So this study can offer a new approach to the fast discrimination of varieties of wheat seeds.

Keywords: E-nose; Wheat; Principal component analysis (PCA); Stepwise discriminant analysis (SDA); Back-propagation neural network (BPNN)