

基于近红外光谱的枸杞化学成分定量分析

汤丽华¹, 刘敦华²

(1. 宁夏食品检测中心, 宁夏银川 750001) (2. 宁夏大学农学院, 宁夏银川 750021)

摘要: 针对目前市场上枸杞掺假现象严重及枸杞品质分析检测繁琐等问题, 本试验采用近红外光谱技术对不同产地的枸杞进行光谱分析, 对枸杞的各种化学成分进行定量分析。通过偏最小二乘回归法 (PLS), 其中 30 个样品为校正集, 10 个样品为预测集, 利用枸杞的近红外光谱数据建立校正模型后, 对枸杞的化学值进行预测。用枸杞样品的主成分在空间的分布对全部样品进行检验, 去除异常样品, 同时采用残余验证方差作为确定主成分数的评价标准。结果表明: 枸杞各成分建立的预测模型, 其校正集相关系数 R_c 均在 0.93 以上, 交叉检验相关系数 R_{cv} 均在 0.83 以上。同时各成分的校正均方差 RMSEC 值均小于交叉检验均方差 RMSEP 值, 且这两个数值没有明显差异性, 预测值与化学值具有良好的相关性。本试验中, 枸杞各成分的化学值与预测值都达到了定量标准, 可以进行定量分析。

关键词: 近红外光谱; 枸杞; 化学成分; 定量分析

文章编号: 1673-9078(2013)9-2306-2310

Quantitative Analysis of *Lycium barbarum* L. Composition Based on Near Infrared Spectrum

TANG Li-hua¹, LIU Dun-hua²

(1. Food Detection Center in Ningxia, Yinchuan 750001, China)

(2. School of Agriculture Ningxia University, Yinchuan 750021, China)

Abstract: To solve the problems in sale and identification of *Lycium barbarum* L., a quantitative analysis method was developed based on near infrared spectrum for quantitative analysis of chemical composition of *Lycium barbarum* L. Partial least-squares regression (PLS) was introduced for establishment of model calibration, in which thirty samples were used for the calibration set and ten samples were used for the prediction set. Then the chemical values of *Lycium barbarum* L. were calculated. The result showed that using this prediction model of *Lycium barbarum* L. composition, the calibrating correlation coefficient R_c were above 0.93 and cross-check correlation coefficients R_{cv} were above 0.83. The mean square error correction RMSEC values of each component were smaller than those of the cross check all variance RMSEP values. The predicted values had a good correlation with the chemistry values. Both of chemical value and predictive value of *Lycium barbarum* L. reached quantitative standard, indicating that this method can be used for quantitative analysis of chemical composition of *Lycium barbarum* L.

Key words: near infrared spectrum; *Lycium barbarum* L.; chemical composition; quantitative analysis

枸杞 (*Lycium barbarum* L.) 为茄科 (Solanaceae) 枸杞属 (*Lycium*) 多分枝灌木, 其果实、皮、叶均可入药^[1]。我国枸杞约有 7 个种 3 个变种, 即宁夏枸杞、枸杞、黑果枸杞、新疆枸杞、截萼枸杞、云南枸杞、柱筒枸杞 7 个种和 3 个变种, 通过人工栽培形成的枸杞林, 主要为宁夏枸杞, 分布在我国川北、华北北部、内蒙古、陕西、甘肃、宁夏、青海和新疆等地^[2]。

枸杞具有抗肿瘤、抗衰老、降血脂、抗氧化等多

收稿日期: 2013-06-12

基金项目: 国家科技支撑项目 (2009BAI72B04)

作者简介: 汤丽华 (1986-), 女, 硕士研究生, 主要从事食品检测工作

种功效^[3], 人们对其所含的各种化学成分进行了深入的研究。枸杞中的化学成分主要有糖类、蛋白质与氨基酸、脂肪类、生物碱、胡萝卜素及色素、黄酮类成分、维生素、矿质元素等。目前, 枸杞化学成分的测定普遍采用国标方法, 虽然可以保证测量的标准统一, 便于对照, 但传统的化学测定方法所需的化学试剂多、测试仪器操作复杂、测定步骤较多、测定结果误差大、检验人员多等使测定过程费时、费力。需要选择一种简单、方便、易操作的成分测定方法, 以提高枸杞生产企业效率。

针对目前枸杞市场掺假现象严重, 枸杞成分测定繁琐等问题下, 利用近红外光谱技术快速、无损、无

污染、操作简单等特点,建立不同枸杞产地溯源模型,为枸杞产地鉴定提供一种新的技术和依据。同时,建立枸杞成分的定量分析模型,为枸杞品质的判断提供依据,为企业生产提供帮近红外光谱技术能够实现样品化学成分的定量检测,在不损伤样品的情况下,进行高效、快速测定,目前利用近红外光谱结合 PLS 对物质成分做定量分析的研究还是较少的,主要集中于药品、粮食和茶叶等方面,对于枸杞成分的近红外定量分析还未见有报道,因此本试验利用近红外光谱技术为枸杞定量分析提供了一种新的思路。

1 材料与方法

1.1 试验材料

以中宁、南梁、同心、惠农、甘肃、青海、内蒙、河北等 8 个产地的枸杞共计 40 个样品作为训练集,建立枸杞个成分的定量模型。

1.2 试验仪器

本试验使用瑞典波通公司生产的 DA7200 近红外成分分析仪进行光谱收集与分析,该仪器的波长范围:950~1650 nm;扫描次数:100 次/秒;扫描间隔:2.00 nm;噪音水平:小于 20 μ A;利用软件 Simplicity 进行光谱数据采集。采用 Unscrambler MVA 光谱分析软件对原始光谱进行数据处理分析。

1.3 试验方法

1.3.1 样品制备

将采集后的枸杞样品去除杂质后,放入烘箱中 55 $^{\circ}$ C 烘一周,除去初水分,用粉碎机粉碎后过 60 目筛,将枸杞样品粉末充分混合均匀,装入自封袋中,贴上标签,注明样品的采集地点及样本号,放入干燥器中保存,用来测定化学参考值和光谱数据。

1.3.2 枸杞各成分的测定方法

水分:GB 5009.3-2010;灰分:GB 5009.4-2010;蛋白质:GB 5009.5-2010;脂肪:GB 5009.6-2003;总糖:GB/T 5009.7-2008;还原糖:GB/T 5009.7-2008;总酸:GB/T 12456-2008;维生素 C:2,4-二硝基苯肼法^[4];氨基酸态氮:双指示剂法^[5];多糖:超声波辅助测定法^[5];黄酮:超声波辅助测定法^[5];甜菜碱:分光光度法^[6];类胡萝卜素:有机溶剂测定法^[6]。

1.3.3 近红外光谱的定量分析方法

本文采用偏最小二乘回归法(PLS)作为枸杞近红外光谱的定量分析方法。偏最小二乘回归(PLS)^[7-8]是近红外光谱分析中应用最多的一种建模回归方法,

采用全光谱分析将光谱变量压缩为少量的独立变量建立回归方程,通过交叉验证防止过拟合的现象,保证主成分一定与某化学组分浓度相关,消除数据之间的冗余信息,以达到降维的目的。

1.4 衡量校正模型的参数

本试验对枸杞样品采用全波段光谱扫描,建立校正模型后,必须通过对验证集样品的测量来判断校正模型的稳定性和准确性,校正模型验证常用的评估参数包括^[9,10]。

相关系数(Correlation coefficient, R):是一个标量,其值介于 0~1 之间。相关系数用于考察样本预测值与实测值之间的相关程度,R 越接近于 1,说明预测值与实测值间的相关程度越好。

校正集预测误差均方根(Root mean square error of calibration, RMSEC):是衡量模型好坏的一个重要指标。

预测集预测误差均方根(Root mean square error of prediction, RMSEP):是外部验证中的一个重要参数,用于衡量预测集样品预测结果的准确性。RMSEP 越小,表明模型对外部样本的预测能力越高;反之 RMSEP 越大,其预测能力越低。

交叉验证均方差(Root Mean Square Error of Cross Validation, RMSECV):是内部交叉验证中的一个重要参数,主要用于评价所建模型对内部样本的预测能力。RMSECV 越小,表明模型对内部样本的预测能力越高,校正集样品的预测结果越接近理论值;RMSECV 越大,误差越大。

2 结果与分析

2.1 枸杞近红外光谱的采集

中宁、南梁、同心、惠农、甘肃、青海、内蒙、河北等 8 个产地的枸杞分别进行近红外扫描,每个样品扫描 2 次,采集不同产地枸杞的近红外原始光谱。如图 1 所示。

由图 1 可以看出,光谱图中由上至下依次为:河北、中宁、惠农、内蒙、甘肃、青海、同心、南梁 8 个产地 40 个枸杞样品光谱图,其变化趋势相似,说明采集到的原始光谱信息量较大,光谱具有一定的代表性,达到了近红外光谱预测的要求。由于光谱信息重叠情况严重,样品因种间差异和种植环境因素等的不同,主要化合物的含量也会有差别。因此,必须对原始光谱进行处理,才能利用近红外光谱贵枸杞化学成分进行定量分析。本文采用一阶导数+五点平滑+矢量

归一化 (SNV) 的方法对光谱进行处理, 处理后的光谱如图 2 所示, 从图 2 中可以看出全光谱信息特征更加突出, 因此以该处理后的光谱信息作为定量分析的预测集。

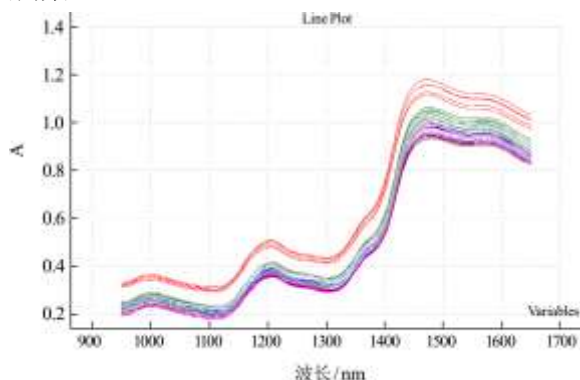


图 1 8 个产地枸杞的近红外光谱图

Fig.1 NIR spectra of *Lycium barbarum* L from eight regions

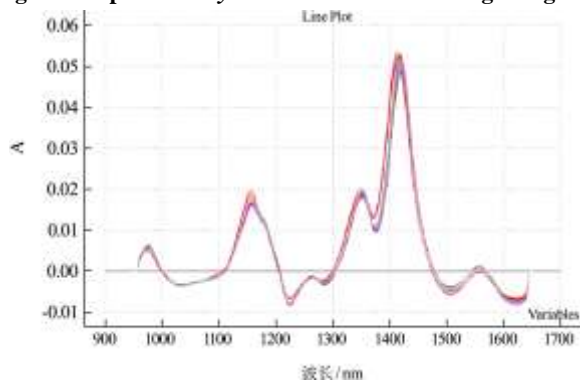


图 2 一阶导数 (5 点平滑) +SNV 全波段光谱图

Fig.2 First-order derivative (five point smoothing) and Full-band SNV spectra

2.2 枸杞中各化学值的采集

枸杞 40 个样品化学成分分析结果见表 1, 由表中可以看出, 枸杞中水分、灰分、脂肪、蛋白质、总糖、还原糖、总酸、多糖、甜菜碱、黄酮、类胡萝卜素、

维生素 C 和氨基酸态氮的含量均有明显差异, 其含量变化范围基本上覆盖了当前的枸杞品种所具有的化学成分变化范围, 可以建立枸杞样品定量分析模型。

表 1 枸杞样品化学成分的分析结果

Table 1 The chemical value of *Lycium barbarum* L

成分	样品数 N	最大值 Max.	最小值 Min.	平均值 Mean
水分/%	40	19.08	10.84	14.10
灰分/%	40	5.43	3.89	4.55
脂肪/%	40	8.36	2.61	4.45
蛋白质/%	40	15.03	9.68	11.82
总糖/%	40	53.78	36.70	48.35
还原糖/%	40	50.84	33.38	42.79
总酸/%	40	1.53	0.42	1.11
多糖/%	40	6.45	3.05	4.48
甜菜碱/%	40	0.65	0.29	0.47
黄酮/(mg/g)	40	1.18	0.65	0.82
类胡萝卜素/(mg/g)	40	1.62	0.97	1.20
维生素 C/(10 ⁻² mg/g)	40	32.32	13.64	22.19
氨基酸态氮/%	40	0.27	0.14	0.18

2.3 近红外光谱对枸杞各化学成分的定量分析

2.3.1 异常值的去除和最佳主成分数的确定

对近红外光谱定量分析来说, 其数据异常包括光谱或化学值的测量值与真实值的显著性异常和样本的光谱或化学值与建模集中样本的平均光谱或化学值范围的显著性差异, 可分为光谱异常和化学值异常, 为提高模型的质量, 需去除样品中的异常值^[1]。各成分去除异常值个数如表 2 所示。

表 2 各成分去除异常值个数

Table 2 The number of the components with removed outliers

	水分	灰分	脂肪	蛋白质	总糖	还原糖	总酸	多糖	黄酮	维生素 C	甜菜碱	类胡萝卜素	氨基酸态氮
异常个数	0	0	3	2	0	1	2	3	8	1	2	3	1

表 3 各成分的最佳主成分数

Table 3 The best principal components number of various ingredients

	水分	灰分	脂肪	蛋白质	总糖	还原糖	总酸	多糖	黄酮	维生素 C	甜菜碱	类胡萝卜素	氨基酸态氮
最佳主成分数	2	4	6	2	2	7	6	9	8	4	9	8	8

本试验采用偏最小二乘法 (PLS) 建立校正模型, 确定的主成分中第一个主成份最重要, 随着主成份数的增加, 重要程度逐渐降低, 直至后面的许多主成份反应的是噪音信息^[2]。主成分数的选择影响模型的实际预测能力。本试验采用残余验证方差 (Residual

validation variance) 作为确定主成分数的评价标准。通过残余方差验证, 得到各成分的最佳主成分数如表 3 所示。

2.3.2 近红外光谱对枸杞各化学成分的定量分析结果

试验中随机挑选 40 个样品中的 30 个样本组成校正集, 根据化学方法测得的数据, 采用枸杞样品的近红外光谱和枸杞化学值之间建立数学模型, 在全光谱范围内, 确定合理的主成分数, 选用 PLS 进行计算, 利用该数学模型预测 10 个预测集枸杞样品中化学成分。本试验随机取 30 个枸杞样品为校正集, 剩下 10 个作为预测集, 在全光谱 950~1650 nm 波长范围内, 对原始光谱进行一阶导数+五点平滑+矢量归一化 (SNV) 预处理后, 采用 PLS 方法建立枸杞各化学成分的定量分析模型。表 4 是枸杞 13 种化学成分的定标模型参数, 以相关系数(Correlation coefficient, R)、校正集预测误差均方根(Root mean square error of calibration, RMSEC)、预测集预测误差均方根(Root mean square error of prediction, RMSEP)来评价建立模型的优劣。

表 4 13 种枸杞化学成分的定标模型参数

Table 4 Calibration model parameters for 13 chemical constituents of *Lycium barbarum* L

	去除异常点	主成分数	Rc	Rcv	RMSEC	RMSEP	SLOP
水分	0	2	0.94	0.95	0.73	0.79	0.89
灰分	0	4	0.93	0.90	0.14	0.18	0.87
脂肪	3	6	0.96	0.93	0.41	0.61	0.93
蛋白质	2	2	0.94	0.93	0.52	0.57	0.88
总糖	0	2	0.95	0.95	1.31	1.41	0.91
还原糖	1	7	0.96	0.95	1.26	2.22	0.92
总酸	2	6	0.96	0.91	0.09	0.13	0.92
多糖	3	9	0.97	0.83	0.26	0.73	0.94
黄酮	8	8	0.99	0.96	0.018	0.045	0.99
维生素 C	1	4	0.93	0.91	2.21	3.32	0.87
甜菜碱	2	9	0.98	0.89	0.017	0.045	0.97
类胡萝卜素	3	8	0.98	0.94	0.028	0.058	0.97
氨基酸态氮	1	8	0.97	0.84	0.0091	0.022	0.95

用 PLS 法建立各化学成分的数学模型, 表 4 中 Slope 值是化学测定值与预测值组成曲线的斜率, 当值越接近于 1, 说明预测值与化学分析值越接近, 由此可知各成分的校正曲线和交叉验证曲线的 Slope 值是水分: 0.89, 0.88; 灰分: 0.87, 0.83; 脂肪: 0.93, 0.89; 蛋白质: 0.88, 0.87; 总糖: 0.91, 0.91; 还原糖: 0.92, 0.90; 总酸: 0.92, 0.87; 多糖: 0.94, 0.73; 黄酮: 0.99, 0.90; 维生素 C: 0.87, 0.85; 甜菜碱: 0.97, 0.86; 类胡萝卜素: 0.97, 0.90; 氨基酸态氮: 0.95, 0.89。其中脂肪、总糖、还原糖、总酸、多糖、黄酮、甜菜碱、类胡萝卜素、氨基酸态氮的校正曲线 Slope 值大部分在 0.90, 而且各点大多分布在曲线附

近, 说明预测值和化学分析值符合良好。而水分、灰分、蛋白质、维生素 C 的校正曲线 Slope 值在 0.83 以上, 且各点在曲线附近比较分散, 说明由于人为或环境等影响, 预测值和化学分析值有一定偏差。

如表 4 所示, 枸杞中水分、灰分、脂肪、蛋白质、总糖、还原糖、总酸、多糖、黄酮、维生素 C、甜菜碱、类胡萝卜素、氨基酸态氮的校正集相关系数 Rc 分别为: 0.95、0.93、0.96、0.94、0.95、0.96、0.96、0.99、0.97、0.97、0.98、0.98、0.93, 交叉检验相关系数 Rcv 分别为: 0.94、0.90、0.93、0.93、0.95、0.95、0.91、0.83、0.84、0.91、0.89、0.94、0.96。其中水分、脂肪、蛋白质、总糖、还原糖、总酸、黄酮、维生素 C、类胡萝卜素校正集相关系数 Rc 及交叉检验相关系数 Rcv 均告高于 0.90, 说明建立的模型与化学值有很好的相关性。而多糖、甜菜碱、氨基酸态氮的交叉检验相关系数在 0.83 以上, 建立的预测模型, 存在一定误差, 是由于多糖、甜菜碱、氨基酸态氮等含量低, 近红外光谱信息含量复杂, 化学成分含量低导致在建模过程中信息提取难度增大, 造成模型的预测效果不理想。

对于一个模型在最佳主成分数的情况下, 校正预测均方差 RMSEC 一定小于交叉验证均方差 RMSEP, 但二者不应该有显著的差异, 否则存在建模信息提取不充分、建模代表性差、和试验分析数据准确度不高等原因^[13]。表 1~4 表明水分、灰分、脂肪、蛋白质、总糖、还原糖、总酸、多糖、黄酮、维生素 C、甜菜碱、类胡萝卜素、氨基酸态氮的校正均方差 RMSEC 分别为: 0.73、0.14、0.41、0.52、1.31、1.26、0.09、0.26、0.018、2.21、0.017、0.0091、0.028, 交叉检验均方差 RMSEP 分别为: 0.79、0.18、0.61、0.57、1.41、2.22、0.13、0.73、0.045、3.32、0.045、0.022、0.058。可以看出各成分的校正均方差 RMSEC 值均小于交叉检验均方差 RMSEP 值, 而且这两个数值没有明显差异性。由以上这些数据一致表明对水分、灰分、脂肪、蛋白质、总糖、还原糖、总酸、多糖、黄酮、维生素 C、甜菜碱、类胡萝卜素、氨基酸态氮等 13 个枸杞化学成分建立了稳定可靠的校正模型。

3 结论

3.1 在全光谱 950~1650 nm 波长范围内, 结合偏小二乘法 (PLS), 对原始光谱进行一阶导数+五点平滑+矢量归一化 (SNV) 预处理, 用枸杞样品的主成分在空间的分布对全部样品进行检验, 去除异常样品, 同时采用残余验证方差作为确定主成分数的评价标准。结果表明, 水分、灰分、脂肪、蛋白质、总糖、还原

糖、总酸、多糖、黄酮、维生素 C、甜菜碱、类胡萝卜素、氨基酸态氮的测定中样品异常值去除的个数分别为: 0、0、3、2、0、1、2、3、8、1、2、1、3; 主成分数分别是: 2、4、6、2、2、7、6、9、8、4、9、8、8。

3.2 对枸杞各化学成分采用偏最小二乘法(PLS)建立模型, 其中 30 个样品为校正集, 10 个样品为预测集, 建立校正模型后, 对预测集进行预测。结果表明: 水分、脂肪、蛋白质、总糖、还原糖、总酸、黄酮、维生素 C、类胡萝卜素等成分建立准确的预测模型, 其校正集相关系数 R_c 及交叉检验相关系数 R_{cv} 均在 0.90 以上, 可以对其进行快速准确的预测工作。而灰分、多糖、甜菜碱、氨基酸态氮的交叉检验相关系数在 0.83 以上, 建立的预测模型, 存在一定误差, 可在以后的试验中进一步完善模型。

参考文献

- [1] 杨薇,谭自力.枸杞的价值及栽培技术[J].河北农业科技, 2008,6:35-36
Yang Wei, Tan Zi li. The value of Lycium barbarum L and cultivation techniques [J]. HeiBei Agricultural Science and Technology, 2008, 6: 35-36
- [2] 郑国琦,胡正海.宁夏枸杞的生物学和化学成分的研究进展[J].中草药,2008,5:796-780
Zheng Guo qi, Hu Zheng hai. The research of Ningxia wolfberry biological and chemical composition [J]. Chinese Traditional and Herbal Drugs, 2008, 5: 796-780
- [3] 李小亭,李瑞盈,相海恩,等.基于 HPLC 指纹图谱及聚类分析对不同产地枸杞质量评价研究[J].现代食品科技,2012, 28(9):1251-1252
Li Xiao ting, Li Rui ying, Xiang Haien, et al. Study on Quality Evaluation of Lycium Chinese from Different Areas by HPLC Fingerprint and Cluster Analysis [J]. Modern Food Science and Technology, 2012, 28(9):1251-1252
- [4] 侯曼玲.食品分析[M].化学工业出版社,2004
Hou Manling. Food Analysis [M]. Chemical Industry Press, 2004
- [5] 张自萍,黄文波.枸杞总黄酮和多糖的超声提取及含量的测定[J].农业科学研究,2006,27(1):22-24
Zhang Ziping, Huang Haibo. Ultrasonic extraction and determination of total flavonoids and polysaccharides from Lycium barbarum L [J]. Journal of Agricultural Sciences, 2006, 27(1): 22-24
- [6] 白寿宁.宁夏枸杞研究(上下)[M].宁夏人民出版社,2000
Bai Shouning. The Research of Lycium [M]. The People's Press of NingXia, 2000
- [7] N Ganze Turana, Basak Mesci, Okan Ozgonenel. The use of artificial neural networks (ANN) for modeling of adsorption of Cu(II) from industrial leachate by pumice [J]. Chemical Engineering Journal, 2011, 171: 1091- 1097
- [8] Xu H, Liu Z C, Cai W S, et al. A wavelength selection method based on randomization test for near-infrared spectral analysis [J]. Chemometrics and Intelligent Laboratory Systems, 2009, 97(2): 189-193
- [9] 姜伟.基于近红外技术的苎麻化学成分快速定量分析研究[D].青岛大学硕士论文,2009
Jiang Wei. Based on near-infrared technology ramine rapid quantitative analysis of chemical composition [D]. Master Thesis in Qingdao University, 2009
- [10] 张华秀.近红外光谱法快速检测牛奶中蛋白质与脂肪含量[D].中南大学,2010
Zhang Huaxiu. Using Near-infrared Spectroscopy to detect the content of protein and fat in milk [D]. Centeal South University, 2010
- [11] 高文骥,袁洪福,邱藤,等.原位实时近红外光谱研究核壳乳液聚合过程[J].高等学校化学学报,2009,30(7):1293-1299
Gao Wen ji, Yuan Hongfu, Qiu Teng, et al. Research of Core-shell Emulsion Polymerization Using in-situ Near Infrared Spectroscopy [J]. Chemical Journal of Chinese University, 2009, 30(7): 1293-1299
- [12] 穆怀彬.近红外光谱技术在玉米营养品质和青贮玉米品质评定中的研究[D].内蒙古农业大学博士论文,2008
Mu Huai bing. Application of near-infrared Reflectance Spectroscopy (NIRS) in the Analysis of Maize Nutritional Qualities and Silage Qualities [D]. NeiMenggu Agricultural University PhD thesis, 2008
- [13] 史永刚,冯新沪,孙萍.水中有机污染物的近红外光谱快速鉴别[J].光谱实验室,2005,22(3):575-577
Shi Yong gang, Feng Xin lu, Sun Ping. Use Near Infrared Spectroscopy for Rapid Identification of Organic Contaminants in Waste Water [J]. Spectroscopy Laboratory, 2005, 22(3): 575-577