第32卷，第8期
2012年8月

光谱学与光谱分析
Spectroscopy and Spectral Analysis

Vol. 32, No. 8, pp2080-2084
August, 2012

温度影响下短波近红外酒精度检测的传递校正

付庆波[1]，索　辉[1]，贺馨平[1,2]，丛玉良[3*]

1. 集成光电子国家重点联合实验室吉林大学实验区，吉林大学电子科学与工程学院，吉林 长春　130012
2. 吉林大学化学学院，吉林 长春　130012
3. 吉林大学通信工程学院，吉林 长春　130012

摘　要　短波近红外光谱结合多元校正方法测量酒精度具有快速、无损的特点，可以现场应用和在线检测。但是温度对预测模型的性能影响很大。作者研究了温度变化对乙醇水溶液的短波近红外光谱的影响，采用四种方法建立了温度变化下的偏最小二乘(PLS)传递校正模型：直接校正、全局校正、正交信号处理(OSC)和广义最小二乘加权法(GLSW)。结果表明，温度变化时直接传递校正存在很大的预测系统偏差，而采用全局校正、OSC和GLSW均能在一定程度上减小系统误差。全局校正需要较多的建模样品数量，得到的模型也更为复杂。而OSC和GLSW方法能得到更为简洁的模型和优良的预测结果。相比之下，GLSW算法得到的预测误差和使用的建模隐含变量数目均小于OSC方法，能够建立更为稳健的传递校正模型。

关键词　短波近红外；温度校正；正交信号处理；广义最小二乘加权法
中图分类号：O657.3　　**文献标识码**：A　　**DOI**：10.3964/j.issn.1000-0593(2012)08-2080-05

引　言

短波近红外(780～1 100 nm)相比长波近红外(1 100～2 526 nm)可以更深地穿透样品并且产生较小的热效应，也可以避免水的强烈吸收干扰[1,2]。低成本又是该技术的一大优势。廉价的钨灯、LED或太阳光都可以作为短波近红外光源。采用价格低廉的硅基CCD构成的多通道短波近红外光谱仪在在线分析中也得到了越来越广泛的应用[3-5]。短波近红外光谱法可用于酒类和饮料中酒精度的定量检测[6-10]。相比国家标准白酒酒精度的比重计测量法，该方法具有无污染、方便快捷、非破坏性、可在线检测等优点。

然而水溶液中羟基的振动易受温度影响：温度的升高会导致伸缩振动的羟基数目减少而自由振动的数目增加，从而产生振动光谱的偏移[11]。因此近红外方法检测酒精度的温度校正是必要的。针对外部变量引起的光谱的偏移，近些年出现了一些光谱标准化算法和正交处理算法(又称为混杂处理，clutter)，包括：分段直接标准化(piecewise direct standardization，PDS)[12,13]，正交信号处理(orthogonal signal correction，OSC)[14,15]，动态正交投影(dynamic orthogonal projection，DOP)[16]，广义最小二乘加权(generalized least squares weighting，GLSW)[17]和模拟退火优选波长法(simulated annealing，SA)[18]等。这些方法可以用于减少由于化学或物理组分的改变引起的背景的变化以及采样的系统误差和仪器漂移所引起的干扰。但是，目前还没有通用的规则来判断何种情况下使用何种方法，而要根据具体问题选择。

本研究目的是在温度变化下建立更为通用的短波近红外酒精度检测校正模型。由于酒厂发酵、蒸馏、灌装车间温度差异较大，建立温度适应性更强的、更为通用的校正模型对短波近红外技术在这方面的实际应用非常有必要。直接传递校正、全局校正，以及OSC和GLSW光谱预处理方法被用于建立传递校正模型，并在建模样品量、实现方式和预测结果方面进行了对比。

1　实验部分

测试样品采用乙醇水溶液。室温下配制20个乙醇水溶液样品，浓度范围30%～70%(体积百分比)，浓度间隔2%。然后均分成两个子集，10个用于作为校正集和10个作为验证集。每个验证样品又被等分成4份，分别保存在15，25，35，40 ℃恒温下。校正集保存在25 ℃恒温。样品收集在带盖玻璃瓶中，防止挥发。

收稿日期：2012-03-12，修订日期：2012-05-30
基金项目：国家自然科学基金面上项目(20773044)资助
作者简介：付庆波，1984年生，吉林大学电子科学与工程学院博士研究生　e-mail：fuqb09@mails.jlu.edu.cn
*通讯联系人　e-mail：ylcong66@yahoo.com.cn

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

采用实验室搭建的 CCD 短波近红外光谱仪系统采集样品的光谱。室温下空气的光谱作为背景。光谱范围在 650～1 200 nm，分辨率 1 nm，积分时间 20 ms，16 次扫描取平均光谱。扫描得到 50 份短波近红外投射光谱：25 ℃下校正集样品的光谱 10 份，验证集样品光谱在 15，25，35，40 ℃下各 10 份。

酒精计(LM79-J10，北京)做为参考方法。精度为 0.1 ℃的水银温度计浸入样品中、样品置于水浴中保持恒定温度。自编软件用于光谱仪控制和数据采集。MATLAB7.9(The Mathworks，Inc.，Natick，MA，USA)和偏最小二乘工具箱 PLS_Toolbox6.2(www.eigenvector.com)用于数据处理。

2　结果讨论

2.1　温度对吸收光谱的影响

30%，50%，60%，70%的样品在 25 ℃下的吸收光谱如图 1 所示。可以清楚地看到酒精水溶液在 900 nm 左右和 950～1 000 nm 之间的吸收强度变化。随着酒精浓度的提高，900 nm 左右的吸收峰得到了增强而 950～1 000 nm 的吸收强度变小。图 2 是 30%的样品在 15，25，35，40 ℃下的吸收光谱。温度升高导致 950～1 000 nm 的吸收出现了明显的增强，并且向短波方向偏移。

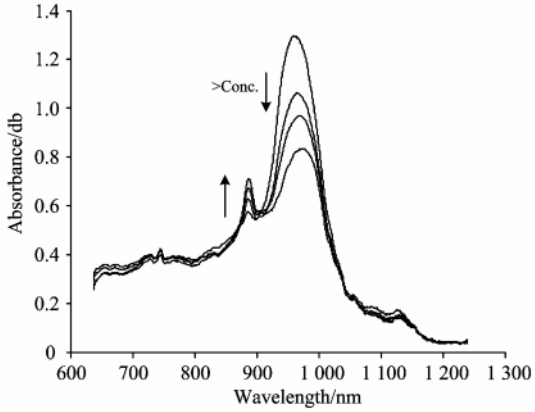

Fig.1　Absorbent spectra of different samples at 25 ℃

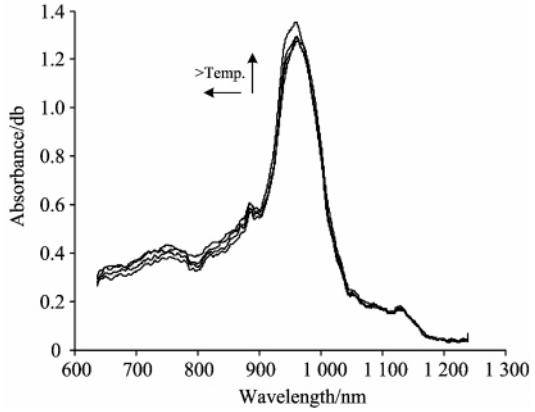

Fig.2　Absorbent spectra of 30% samples at different temperatures

2.2　直接 PLS 校正

在 800～1 100 nm 波段建立 25 ℃下的偏最小二乘校正模型，并预测 15，25，35，40 ℃下的样品。所有光谱进行 11 点均值平滑、均值中心化处理。采用留一交叉检验法。隐含变量的选择基于对均方根误差的考察。最终选择了 2 个隐含变量。

图 3(a)显示了预测结果与参考值的对比。可以看出 25 ℃下预测结果相关性很好，预测系数(R^2)达到了 0.99。其他温度下的预测值虽然和参考方法的线性相关性很好，但是却出现了系统误差，预测偏差(Prediction bias)达到了 5%以上。图 4 中可以看出：预测温度(15 ℃)低于校正温度时出现了较大的正系统误差；当预测温度高于(35 和 45 ℃)校正温度时出现了负系统误差。25 ℃下的误差随机分布，且误差的绝对值小于其他温度下的结果。因此，本地模型(25 ℃)只能在温度不变时得出满意的预测结果。

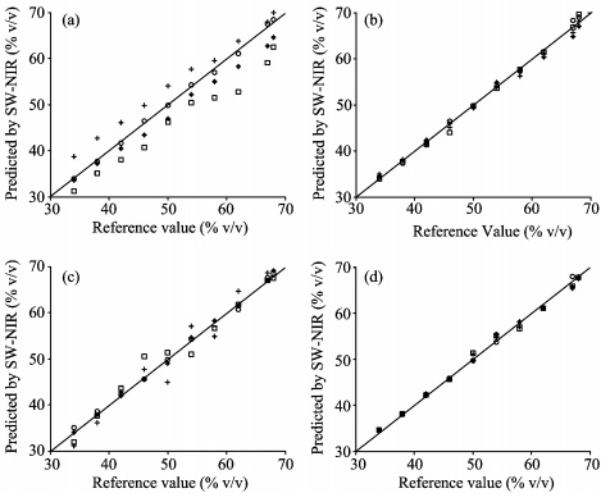

Fig.3　Predication values versus reference vales

(a)：Direct calculation；(b)：Glbal calibration；(c)：OSC method；(d)：GLSW method. Validation set at 25 ℃(circle)，15 ℃(plus)，35 ℃(star) and 40 ℃(sguare)

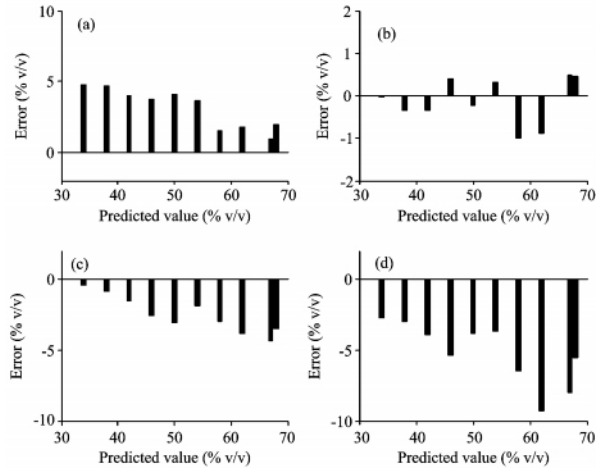

Fig.4　Predication errors for validation set at 25 ℃(a)，15 ℃(b)，35 ℃(c) and 40 ℃(d)

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

2.3 全局校正

全局校正即在每个温度下都选择一定数量的具有代表性的样品来建立校正模型。代表性样品的选择有多种方法，可以按照光谱间距离、建模样品的杠杆值等。我们采用欧式距离作为判断样品光谱差异的手段，从 15，35 和 40 ℃验证集中分别选择 5 个样品。25 ℃下的校正集直接加入校正样品集中。得到总数为 25 的全局校正样品。留一法交叉验证，选择了 4 个隐含变量。

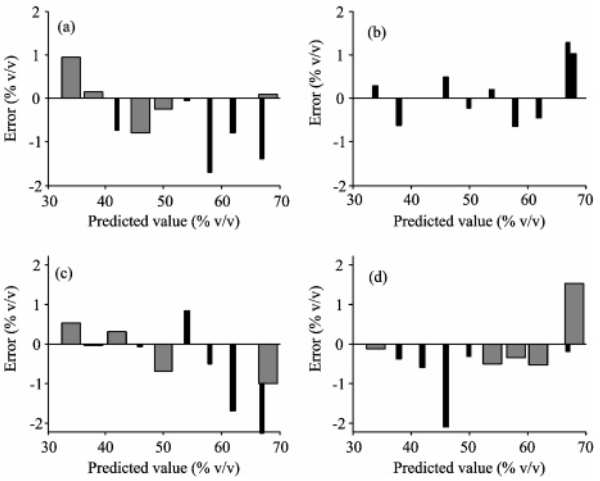

Fig. 5 Predication errors of the global model validation set at 25 ℃(a)，15 ℃(b)，35 ℃(c) and 40 ℃(d)

图 3(b)显示了全局校正的预测结果与参考值对比。四个温度下预测系数(R^2)均达到了 0.99。由图 5 中可以看出：参与校正的样品获得了满意的误差值(随机分布在 0 点的上下，如宽柱所示)，而没有参与校正的样品的预测结果依然存在一定的系统误差。全局校正在一定程度上消除了温度对预测的影响，预测系统误差在 0.5%以内。但是考虑到有一部分验证样品同时参与了校正，如果从预测样品集中剔除这一部分样品，只能得到更大的预测系统误差。

2.4 正交信号处理

正交信号处理算法可用来消除光谱矩阵 X 中与浓度矩阵 Y 中的不相干信息。使用 OSC 校正之前，要求 X 和 Y 数据必须进行标度，一般做均值中心化处理[14]。

采用欧式距离的方法从 25℃下验证集样品中选择 5 个代表性样品，然后从其他温度下的样品集中选择同样浓度值的样品组成一个新的样品集用于实现正交滤波模型的建立。建立好正交滤波模型之后，所有温度下的经过中心化处理后的光谱需要经过该滤波模型的处理，然后使用重新获得的光谱数据建立 PLS 预测模型。

校正集和验证集样品均经过正交信号处理后重新建立 PLS 校正模型。采用正交信号处理需要调节正交模型的因子(components，C)，决定有多少个主成分参与信号的滤波。以 PLS 模型的隐含变量数(LV)和不同温度下的 RMSEP 作为评价标准，来优选合适的因子数。从表 1 列出的结果中可以看出 C=1 是最佳的因子选择。

Table 1 Figures of merit of PLS model with different numbers of components

C	15 ℃		25 ℃		35 ℃		40 ℃	
	LV	RMSEP*	LV	RMSEP*	LV	RMSEP*	LV	RMSEP*
1	3	0.24	3	0.08	3	0.05	3	0.17
2	2	3.17	2	0.35	2	0.41	2	2.65
3	3	2.20	3	0.52	3	1.20	3	1.33

* 数据经过中心化处理，不可与其他方法中的数据纵向对比，只在此横向对比用于选择 C

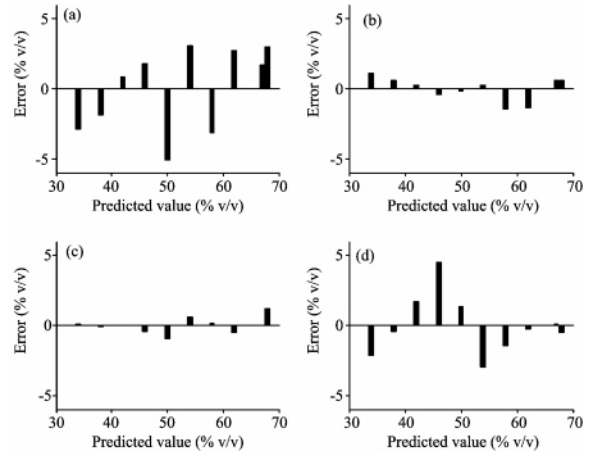

Fig. 6 Predication errors of the OSC model validation set at 25 ℃(a)，15 ℃(b)，35 ℃(c) and 40 ℃(d)

模型在四个温度下预测结果与参考方法的对比如图 3(c)所示。四个温度下的预测值与参考值有很好的相关性。由图 6 中可以看出：25 和 35 ℃下的预测误差小于 2%，优于 15 和 40 ℃下的结果；与全局校正对比发现，25 和 35 ℃下的预测结果相当，而 15 和 40 ℃下存在相对大的预测误差。正交信号处理得到的预测系统偏差小于 1%。

2.5 广义最小二乘加权法处理

广义最小二乘加权算法(GLSW)通过消除相似浓度样品对应的光谱的差异建立一个用于滤波的矩阵，来消除外部变量引起的 X 矩阵的变化。在外部变量的作用下，本该相似的同一样品的光谱可能出现某种程度的偏移。GLSW 的目的就是较少外部变量引发的光谱改变相对于 y 改变引发的光谱改变的权重，从而使在存在外部作用下的同一样品的光谱更为接近[17]。

以欧式距离法，从 25 ℃下验证集样品中选择距离最大的前 5 个样品，然后从其他温度下的样品集中选择同样编号的样品组成一个新的样品集用于建立 GLSW 滤波模型。所有温度下光谱需要经过该滤波模型的前处理，然后使用获得的光谱重新建立 PLS 预测模型。

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

采用 GLSW 处理需要调节参数 α，决定加权处理的强弱。α 一般选择 0.000 1～1 之间的数值，较小的值提供较强的滤波作用。从表 2 中可以看出 0.01 是最佳的因子值。

Table 2　Figures of merit of PLS model with various α

α	15 ℃		25 ℃		35 ℃		40 ℃	
	LV	RMSEP	LV	RMSEP	LV	RMSEP	LV	RMSEP
0.001	1	0.81	1	0.55	1	0.81	1	0.95
0.01	1	0.74	1	0.55	1	0.82	1	0.89
0.1	1	0.80	1	0.70	1	1.24	1	1.19

Table 3　Figures of merit of comparison for four calibation methods

Methods	LV	RMSEP%				Pred. Bias%			
		15 ℃	25 ℃	35 ℃	40 ℃	15 ℃	25 ℃	35 ℃	40 ℃
直接校正	2	3.39	0.52	2.77	5.57	3.11	−0.12	−2.48	−5.17
全局校正	4	0.87	0.64	1.04	0.89	−0.46	0.13	−0.45	−0.35
OSC	3	2.84	0.80	0.56	2.02	−1.05	−0.30	−0.25	−0.89
GLSW	1	0.74	0.55	0.82	0.89	−0.09	−0.09	−0.11	−0.07

模型在四个温度下的预测结果与参考方法的对比如图 3(d)所示。四个温度下的预测值与参考值均有很好的相关性。系统误差明显减小。由图 7 中可以看出：四个温度下的误差都随机分布在 0 点的上下且四个温度下误差小于 1.5%。对比发现，四个温度下的预测结果均优于 OSC 信号处理的结果。四种校正方法的最终对比结果列在了表 3 中。可以看出，GLSW 滤波消除了温度对预测的影响，预测系统偏差低于 0.1%。模型的预测性能优于 OSC 信号校正模型，也优于全局模型下的结果，且需要最少的隐含变量。

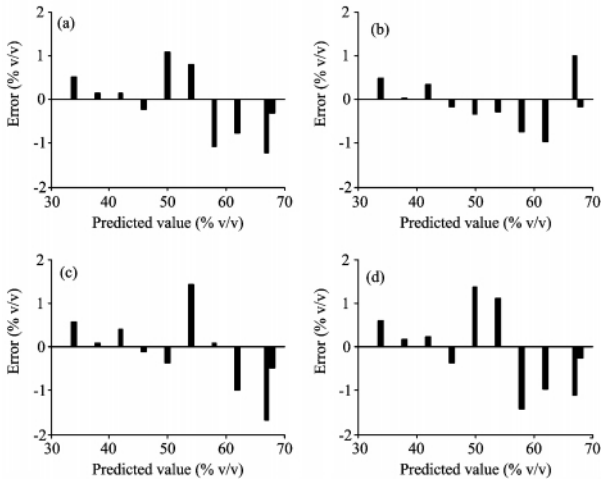

Fig. 7　Predication errors of the GLSW model validation set at 25 ℃ (a), 15 ℃ (b), 35 ℃ (c) and 40 ℃ (d)

3　结　论

温度的变化会对酒精度检测的短波近红外 PLS 校正模型带来很大的预测系统偏差，必须采取传递校正方法来消除。全局校正需要在每个温度下选择校正样品，需校正样品数目最多，所建立的模型也较为复杂；OSC 滤波和全局建模有相当的预测能力，但是由于可以作为一种光谱前处理方法使用，减少了建模样品的个数；相比前两者，GLSW 方法给出了最小的系统偏差和测量误差，也可以作为一种光谱前处理方法使用，并且所建模型也更为简洁。结果表明 GLSW 方法要优于全局校正和 OSC 方法，能够建立更为稳健的传递校正模型。

References

[1] Reeves J B J. Ⅲ. Journal of Near Infrared Spectroscopy, 1994, 2: 199.

[2] YAN Yan-lu(严衍禄). Near Infrared Spectroscopy: Principle and Application(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京：中国轻工业出版社), 2005. 5.

[3] YUAN Hong-fu, ZHU Xiao-li, LU Wan-zhen, et al(袁洪福，褚小立，陆婉珍，等). Chinese Journal of Analytical Chemistry(分析化学), 2004, 32(2): 255.

[4] LU Wan-zhen(陆婉珍). Modern Near Infrared Spectroscopy Analytical Technology(现代近红外光谱分析技术). Beijing: China Petrochemical Press Co., Ltd.(北京：中国石化出版社有限公司), 2005. 92..

[5] LI Jun-hui, ZHAO Long-lian, LAO Cai-lian, et al(李军会，赵龙莲，劳彩莲，等). Modern Scientific Instruments(现代科学仪器), 2005, 1: 17.

[6] LU Jia-hui, TENG Li-rong, JIANG Fu-ming, et al(逯家辉，腾利荣，蒋富明，等). Journal of Jilin University(Science Edition)(吉林大学学报·理学版), 2002, 41(2): 245.

[7] LI Ji-guang, ZHANG Gen-sheng, MAO Di-rui(李继光，张根生，毛迪锐). Food and Machinery(食品与机械), 2004, 20(1): 22.

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

[8] WANG You-bing(王友兵). Journal of Baicheng Teachers College(白城师范学院学报), 2004, 18(4): 44.
[9] Yu H Y, Niu X Y, Lin H J, et al. Food Chemistry, 2009, 113: 291.
[10] Maretto D A, Mello C, Poppi R J. Journal of Near Infrared Spectroscopy, 2008, 16(3): 249.
[11] Czarnecki M A, Maeda H, Ozaki Y, et al. J. Phys. Chem., 1998, 102: 9117.
[12] Bouveresse E, Massart D L. Vibrational Spectroscopy, 1996, 11(1): 3.
[13] Wülfert F, Kok W, De Noord O E, et al. Analytical Chemistry, 2000, 72(7): 1639.
[14] Wold S, Antti H, Lindgren F, et al. Chemometrics and Intelligent Laboratory Systems, 1998, 44: 175.
[15] Barboza F D, Poppi R J. Analytical and Bioanalytical Chemistry, 2003, 377(4): 695.
[16] Zeaiter M, Roger J M, Bellon-Maurel V. Chemometrics and Intelligent Laboratory Systems, 2006, 80(2): 227.
[17] Martens H, Hoy M, Wise B M, et al. Journal of Chemometrics, 2003, 17(3): 153.
[18] Swierenga H, Wülfert F, De Noord O E, et al. Analytica Chimica Acta, 2000, 411(1-2): 121.

Transfer Calibration for Alcohol Determination Using Temperature-Induced Shortwave Near Infrared Spectra

FU Qing-bo[1], SUO Hui[1], HE Xin-ping[1,2], CONG Yu-liang[3]*

1. State Key Laboratory on Integrated Optoelectronics, College of Electronic Science and Engineering, Jilin University, Changchun 130012, China
2. College of Chemistry, Jilin University, Changchun 130012, China
3. College of Communication Engineering, Jilin University, Changchun 130012, China

Abstract The authors studied the temperature influence on short-wave near-infrared spectra of ethanol aquatic solution and utilized four methods to establish the transfer partial least squares (PLS) calibration model: direct transfer calibration, global calibration, orthogonal signal correction (OSC) and generalized least squares weighting (GLSW). The PLS models were built at four temperatures: 15, 25, 35 and 40 ℃. The results showed that direct calibration provided high prediction bias: significantly high positive prediction bias for a temperature lower than calibration temperature and negative bias for higher temperatures. By using the global correction, OSC and GLSW, the systematic errors could be reduced. However, the global correction needed more calibration samples and built a more complex model. The OSC and GLSW methods provided better predictions using fewer latent variables. By using the GLSW method, prediction bias less than 0.1% and RMSEP less than 0.9% were obtained. The absolute prediction error of GLSW method was less than 1.5%. Additionally, the GLSW provided smaller prediction error at every researched temperature using fewer latent variables than OSC. Thus, GLSW was superior to OSC and could establish more robust transfer calibration model.

Keywords Short-wave near infrared; Temperature correction; Orthogonal signal correction; Generalized least squares weighting

(Received Mar. 12, 2012; accepted May 30, 2012)

* Corresponding author

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net