

基于模拟退火波长优化的草莓坚实度近红外光谱检测*

石吉勇¹ 殷晓平¹ 邹小波¹ 赵杰文¹ 鞠时光²

(1. 江苏大学食品与生物工程学院, 镇江 212013; 2. 江苏大学计算机与通信工程学院, 镇江 212013)

【摘要】 为提高近红外光谱技术检测草莓坚实度模型的精度和鲁棒性,研究了一种基于模拟退火算法的波长优选方法,并找到一种与该算法配套的光谱预处理方法。利用光谱仪和物性仪分别采集草莓样品近红外漫反射光谱和坚实度数据,并采用标准正交变换、多元散射校正、一阶导数、二阶导数等方法对原始光谱进行预处理;最后,利用模拟退火算法优选与草莓坚实度高度相关的波数点变量,结合偏最小二乘法建立草莓坚实度预测模型。结果表明:经过标准正交变换预处理后,采用模拟退火算法优选出24个波数点,在主成分数为5时,建立的偏最小二乘模型具有最佳预测效果,模型校正集样本相关系数 r_c 为0.934 2,校正均方根误差为0.665 N/cm²;预测样本相关系数 r_p 为0.919 7,预测均方根误差为0.673 N/cm²。研究表明:模拟退火算法可以提高近红外模型预测草莓坚实度的精度和鲁棒性,并降低预测模型复杂度。

关键词: 草莓 坚实度 近红外光谱 模拟退火算法

中图分类号: O657.33; TS207.3 文献标识码: A 文章编号: 1000-1298(2010)09-0099-05

Detection of Strawberry Firmness by NIR Wavelength Selection Based on Simulated Annealing Algorithm

Shi Jiyong¹ Yin Xiaoping¹ Zou Xiaobo¹ Zhao Jiewen¹ Ju Shiguang²

(1. School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China

2. School of Computer and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract

In order to improve the accuracy and robust of NIR spectroscopy modules in predicting the firmness of strawberry, simulated annealing algorithm (SAA) was used to select the wavenumbers in NIR spectra. A preprocessing method was also selected to adapt the SAA. Firstly, 150 strawberries were selected to collect NIR spectra. Secondly, preprocessing methods, such as SNV, MSC, 1st order derivation, 2nd order derivation, were used to denoise the NIR spectra of strawberry. Thirdly, 24 wavenumbers were selected by simulated annealing algorithm. At last, partial least square was employed to establish the calibration models of firmness. The calibration model was obtained with the correlation coefficient r_c of 0.934 2, the root mean square error of calibration of 0.665 N/cm² and the correlation coefficient r_p of 0.919 7, the root mean square error of prediction of 0.673 N/cm². The results show that SAA can improve the robust and accuracy and simplify NIR spectra models.

Key words Strawberry, Firmness, Near infrared spectroscopy, Simulated annealing algorithm

收稿日期: 2009-11-02 修回日期: 2009-12-17

* 国家“863”高技术研究发展计划资助项目(2008AA10Z208)、国家自然科学基金资助项目(60901079)、江苏省六大人才高峰和青蓝工程资助项目和国家博士后科学基金和优秀博士论文基金资助项目

作者简介: 石吉勇, 博士生, 主要从事农产品、食品品质无损检测研究, E-mail: stoneboy_2007@sohu.com

通讯作者: 邹小波, 副教授, 主要从事农产品、食品品质无损检测研究, E-mail: zou_xiaobo@ujs.edu.cn

引言

坚实度是评价草莓成熟状态和品质的重要指标。目前检测坚实度的常用方法是 M-T 戳穿试验方法 (Magness-Taylor puncture test), M-T 戳穿试验虽然简单易行, 但会破坏检测对象、检测速度慢, 不能对大批量草莓进行快速无损检测。

近红外光谱分析技术作为一种快速、无损、简便的检测方法, 被广泛运用于快速检测农产品的糖度、酸度、维生素 C 等品质指标^[1-2], 近红外光谱分析技术预测农产品品质指标的研究主要是光谱预处理方法、波长选择方法和建模方法的研究^[1]。目前国内外已有一些学者利用近红外光谱分析技术对水果坚实度作了一些相关研究^[2-15], 但对草莓坚实度的研究鲜有报道。这些研究中, 光谱预处理方法主要有标准正交变换、多元散射校正、一阶导数、二阶导数等方法, 建模方法主要为偏最小二乘法。近年来研究表明, 采用全光谱偏最小二乘法建立多变量校正模型时, 不仅计算工作量大, 而且校正模型精度和稳定性未必都能达到最优, 其中波长选择对模型的预测精度有直接影响, 因此必须采用一种有效的方法对波长进行优选。目前比较常用的波长选择算法有间隔偏最小二乘法、遗传算法等, 上述波长选择方法均要预先指定区间划分方式 (区间宽度、个数等), 存在一定的主观性; 同时, 遗传算法存在过快收敛, 陷入局部最优解等不足。

模拟退火算法 (simulated annealing algorithm, 简称 SAA) 是 20 世纪 80 年代初发展起来的一种随机组合优化方法^[14]。该算法依据 Metropolis 准则接受新解, 除了接受优化解外, 还在一个限定的范围内接受恶化解, 使其能在较短的时间内求得全局更优近似解^[15], 这是模拟退火算法同其他局部搜索算法的本质区别。SAA 已经在很多领域得到成功应用, 但鲜有将 SAA 应用于近红外光谱检测技术。

本文用近红外漫反射光谱技术预测草莓坚实度。为了使模型具有好的预测能力, 在建模的过程中必须选取合适的波长。为此, 本文设计一种用于优选近红外光谱波长的模拟退火算法, 并将该方法与偏最小二乘法相结合, 寻找草莓漫反射近红外光谱同草莓坚实度的相关关系, 同时比较不同光谱预处理方法对模型预测性能的影响。

1 模拟退火算法优选波长的原理

1.1 模拟退火算法

根据 Metropolis 准则, 粒子在温度 T 时趋于平衡的概率计算公式为

$$P = \exp(-\Delta E / (kT)) \quad (1)$$

式中 E ——温度 T 时粒子的内能

ΔE ——内能改变量

k ——Boltzmann 常数

用固体退火模拟组合优化问题, 先确定初始温度 T_0 , 随机选择一个初始状态并考察该状态的目标函数值 f ; 然后在当前解的领域中, 以一定概率选择一个非局部最优解, 并令这个解重复下去, 从而不会陷入局部最优。算法由控制参数 T 决定, 经过大量迭代变换后, 可求得给定控制参数 T 时优化问题的相对最优解。然后缓慢减小控制参数 T , 重复上述迭代过程。对温度 T 时的所有迭代过程称为一个马尔科夫链, 迭代次数称为马尔可夫链长度 L_k 。当计算完温度 T 对应的马尔科夫链时, 温度 T 按一定冷却率 α 逐渐减小, 重复上述过程直至温度 T 趋于 0°C , 最终得到问题的全局最优近似解。

1.2 控制参数设计

使用模拟退火算法解决实际问题时, 必须合理选择目标函数 f 对应退火过程中粒子的能量 E 。算法收敛速度取决于起始温度 T_0 、冷却率 α 和马尔可夫链长度 L_k , 因此如何合理选择一组控制算法进程的参数, 使算法在有限时间内返回一个近似最优解, 是该算法的关键。这样的一组控制参数通常称为冷却进度表 (cooling schedule), 它主要包括以下参数: 起始温度 T_0 ; 温度衰减函数 $\alpha(T_{k+1} = \alpha T_k)$; 终止温度 T_f ; 马尔科夫链长度 L_k 。

1.2.1 目标函数的选取

在波长优化过程中, 常采用交互验证法来评价模型的预测能力, 即采用交互验证均方根误差、预测残差平方和、待测组分预测值与实测值之间的相关系数等作为目标函数。如采用交互验证均方根误差作为评价指标, 交互验证均方根误差的值越小, 对应校正模型的预测能力越好。将目标函数转换成求最大值问题, 即

$$F(x) = \frac{1}{1 + f(x)} \quad (2)$$

式中 x_k ——优选出来的波长组合

$f(x_k)$ ——用 x_k 中波数点建立的偏最小二乘模型对应的交互验证均方根误差

1.2.2 冷却进度表的设计

冷却进度表的构造是基于算法的准平衡概念, 其定义如下: 设 L_k 是第 k 个马尔科夫链的长度, T_k 是相应的第 k 个温度控制参数值。若第 k 个马尔科夫链的 L_k 次变换后, 解的概率分布充分逼近 $T = T_k$ 时的平稳分布, 则称模拟退火算法达到准平衡。根据上述的准则, 可以得到两个结论: 只要 T 充分大,

算法会立刻达到准平衡;控制参数 T_k 的衰减量越大,需要的马尔可夫链长度 L_k 越长,才越有利于恢复准平衡,通常选取 T_k 的小衰减量以避免过长的马尔可夫链。同时有效的冷却进度表还要兼顾算法的收敛性和执行效率。综合上面的结论以及参数设置的经验,模拟退火算法使用以下冷却进度表: $T_0 = 200^\circ\text{C}$; $T_k = 0.95T$; $T_f = 0^\circ\text{C}$; $L_k = 50$ 。

模拟退火算法程序在 Matlab 7.4 平台下编程实现。

2 材料和方法

2.1 试验材料

试验用草莓(奶油味)采于镇江市长岗草莓园,共采摘了两批次。第1批主要采摘完全成熟(成熟度依据草莓果的着色率判断)的草莓;第2批采摘成熟度80%左右的草莓。试验当天早上,从草莓园采回试验所需草莓,带回实验室后挑选无碰伤、表面干净的草莓进行编号(共选出150个草莓,100个为校正集,50个为预测集),整个过程保持实验室温度在 25°C ,湿度基本不变。

2.2 近红外光谱数据采集

试验所用近红外数据采集设备为 Antaris II 型傅里叶变换近红外光谱仪(Thermo Fisher,美国),采用 InGaAs 检测器,光谱范围 $10\,000 \sim 4\,000\text{ cm}^{-1}$,扫描次数为32次;分辨率 4 cm^{-1} ,数据采集间隔为 1.928 cm^{-1} 。光谱数据采集时首先将草莓放在检测位置优化增益倍数,并采集背景光谱。整个光谱采集过程保持温度、湿度不变,图1为草莓原始光谱图,图中 R 为近红外光谱反射系数。

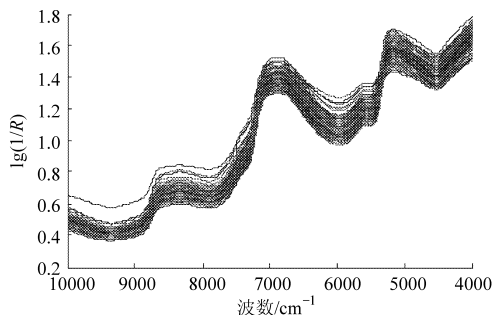


图1 草莓原始光谱图

Fig. 1 Original spectra of strawberry

2.3 坚实度测量

采用 TA-XT2i 型物性仪(英国 SMS 公司)测定草莓坚实度标准值,选用直径 3 mm 的圆平探头(型号 P3),探头插入速度 1.0 mm/s ,插入深度 10 mm,力触发阈值为 0.05 N ,数据采集率为 250 PPS。测量时物性仪探头插入方向由草莓表面指向果心,插入位点与近红外光谱数据采集点对应,本试验位点选

择在草莓赤道圆周上,测定过程中计算机自动记录插入深度和探头受到的压缩力。表1为校正集和预测集样本的坚实度测量值统计结果。

表1 校正集和预测集 M-T 戳穿试验坚实度统计

Tab.1 Statistic of strawberry firmness by Magness-Taylor puncture test

样本数	最大值 $/\text{N}\cdot\text{cm}^{-2}$	最小值 $/\text{N}\cdot\text{cm}^{-2}$	平均值 $/\text{N}\cdot\text{cm}^{-2}$	标准偏差 $/\text{N}\cdot\text{cm}^{-2}$	
校正集	100	8.604 2	2.891 6	5.044 0	1.166 2
预测集	50	8.335 5	3.542 1	5.312 9	1.009 7

3 结果与讨论

3.1 光谱预处理

由于仪器、样品背景、环境条件及其他因素的影响,近红外光谱中常出现噪声、谱图基线漂移和平移等现象,为了消除这些不利因素对模型的影响,应对原始光谱进行预处理。采用标准正交变换、多元散射校正、一阶导数、二阶导数等方法对原始光谱进行预处理,以期望获得好的建模效果。图2为草莓原始光谱经过4种预处理方法处理后的效果图。

3.2 坚实度定标模型的建立

根据1.2节设定的控制参数,采用模拟退火算法对预处理后的光谱优选波长,并建立了偏最小二乘模型。为了考察窗口宽度对模型精度的影响,在 $10 \sim 100$ 波数点范围内,对标准正交变换预处理后的光谱优化了间隔偏最小二乘模型窗口宽度。结果如图3所示,其中横坐标为波数点数即窗口宽度,纵坐标为目标函数值,目标函数值越大,说明对应的窗口宽度越好。从图3中可以看出,当窗口宽度为24个波数点时,目标函数值取得最大值。

为了研究模拟退火算法的收敛情况,在窗口宽度为50时,跟踪了目标函数值随退火温度降低的变化情况。结果如图4所示,其中横坐标表示退火温度,退火初始温度 T_0 为 200°C ,按照一定冷却率逐渐冷却,直到达到结束温度 0°C 。纵坐标表示目标函数值,函数值越大,表明对应的解越好。图4中圆圈所在曲线上的点代表当前退火温度前算法计算出来的最优解,它保证了在整个优化过程中,最优解不会发生退化。图4中方框所在曲线上的点代表当前退火温度所得到的最优解,从图4中可以看出,方框所在曲线上的点并不是随着退火温度的变化而持续递增,而是在某些温度下目标函数值变小了。这是由于模拟退火算法根据 Metropolis 准则(式(1))接受新解时,不但会接受优化解,还会按一定概率接受恶化解,所以在图中会出现目标函数值随着退火温

度降低而减小的情况。同时,根据式(1)可知,随着退火温度减小,出现目标函数值减小频率和减小幅

度会降低,而图中曲线的变化趋势也印证了这一规律。

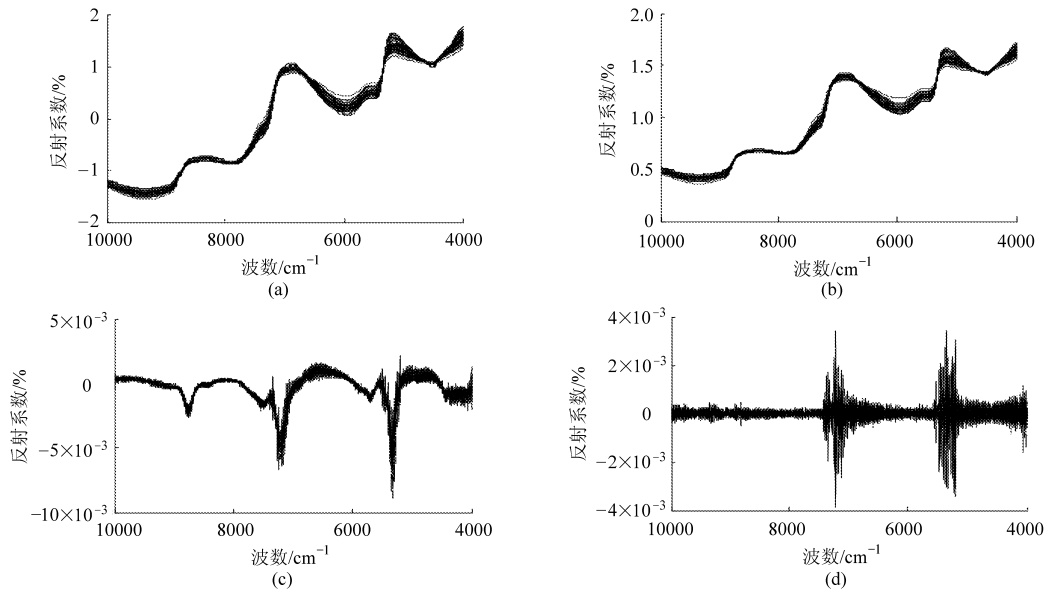


图 2 光谱预处理后效果图

Fig. 2 Spectra after different pre-method

(a) 标准正交变换 (b) 多元散射校正 (c) 一阶导数 (d) 二阶导数

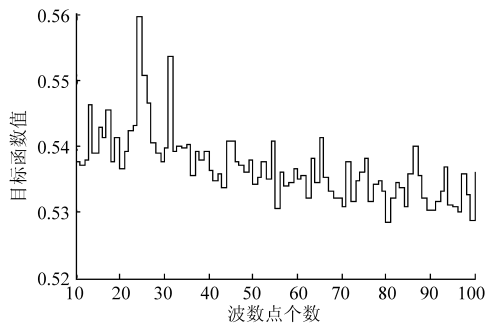


图 3 波数点优化结果

Fig. 3 Number of selected wavenumbers after SAA

预处理后的数据经模拟退火算法优选波长后建立了定标模型。在 10 ~ 100 范围内对建模波数点数

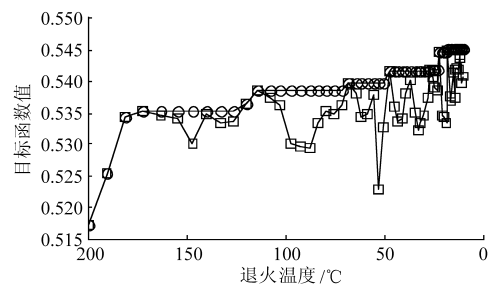


图 4 模拟退火算法优选波长效果

Fig. 4 Best f value of SAA wave select method

(窗口宽度)进行了优化,在建模过程中用交互验证法同时优化窗口宽度和主成分数,结果如表 2 所示。

从表 2 中可以看出,不同预处理方法会对模型

表 2 不同预处理方法偏最小二乘建模效果比较

Tab. 2 Comparison of PLS model results after spectra being treated by different pre-methods

预处理方法	窗口宽度	主成分数	校正均方根误差 $\sqrt{N \cdot \text{cm}^{-2}}$	r_c	预测均方根误差 $\sqrt{N \cdot \text{cm}^{-2}}$	r_p
标准正交变换	24	5	0.665	0.934 2	0.673	0.919 7
多元散射校正	28	7	0.669	0.927 7	0.678	0.905 2
一阶导数	35	7	0.674	0.901 1	0.679	0.895 4
二阶导数	40	8	0.671	0.902 4	0.682	0.889 4
无预处理	42	8	0.679	0.900 2	0.687	0.875 3

产生一定的影响。其中标准正交变换预处理方法建模效果要比其他方法所得的效果略好些。这是由于草莓表面凹凸不平,以及不同成熟度的草莓密度有差异,这些因素会影响光的漫反射过程,而标准正交变换可以消除由于颗粒大小不均匀和密度不均而产生的散射影响,所以标准正交变换能得到相对较好的建模结果。

4 结束语

利用近红外光谱漫反射技术结合模拟退火算法建立草莓坚实度预测模型。对近红外光谱数据进行标准正交变换预处理后,采用模拟退火算法优选出了 24 个波数点,并在主成分数为 5 时采用偏最小二乘建立了最佳定标模型。模型对校正集样本相关系

数 r_c 为 0.934 2,校正均方根误差为 0.665 N/cm²; 误差为 0.673 N/cm²。
模型对预测样本相关系数 r_p 为 0.919 7,预测均方根

参 考 文 献

- 1 Butz P, Hofmann C, Tauscher B. Recent developments in noninvasive techniques for fresh fruit and vegetable internal quality analysis[J]. *Journal of Food Science*, 2005, 70(9): 131 ~ 141.
- 2 Camps C, Guillermin P, Mauget J C, et al. Discrimination of storage duration of apples stored in a cooled room and shelf-life by visible-near infrared spectroscopy[J]. *Journal of Near Infrared Spectroscopy*, 2007, 15(3): 169 ~ 177.
- 3 Lu R, Ariana D. A near-infrared sensing technique for measuring internal quality of apple fruit[J]. *Applied Engineering in Agriculture*, 2002, 18(5): 585 ~ 590.
- 4 Guyer D, Yang X K. Use of genetic artificial neural networks and spectral imaging for defect detection on cherries[J]. *Computers and Electronics in Agriculture*, 2000, 29(3): 179 ~ 194.
- 5 Teerachaichayut S, Kil K Y, Terdwongworakul A, et al. Non-destructive prediction of translucent flesh disorder in intact mangosteen by short wavelength near infrared spectroscopy[J]. *Postharvest Biology and Technology*, 2007, 43(2): 202 ~ 206.
- 6 Peiris K H S, Dull G G, Leffler R G, et al. Spatial variability of soluble solids or dry-matter content within individual fruits, bulbs, or tubers: implications for the development and use of NIR spectrometric techniques[J]. *Hort. Science*, 2003, 34(1): 114 ~ 118.
- 7 Huang H B, Yu H Y, Xu H R, et al. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages [J]. *Journal of Food Engineering*, 2008, 87(3): 303 ~ 313.
- 8 Sirisomboon P, Tanaka M, Fujita S, et al. Evaluation of pectin constituents of Japanese pear by near infrared spectroscopy [J]. *Journal of Food Engineering*, 2007, 78(2): 701 ~ 707.
- 9 Tsai Chao-Yin, Chen Hsi-Jien, Hsieh Junn-Fu, et al. Fabrication of a near infrared online inspection system for pear fruit [J]. *International Agricultural Engineering Journal*, 2007, 16(1 ~ 2): 57 ~ 70.
- 10 Walsh K B, Long R L, Middleton S G. Use of near infrared spectroscopy in evaluation of source-sink manipulation to increase the soluble sugar content of stonefruit[J]. *Journal of Horticultural Science & Biotechnology*, 2007, 82(2): 316 ~ 322.
- 11 Golic M, Walsh K B. Robustness of calibration models based on near infrared spectroscopy for the in-line grading of stonefruit for total soluble solids content[J]. *Analytica Chimica Acta*, 2006, 555(2): 286 ~ 291.
- 12 Lu R F, Peng Y K. Hyperspectral scattering for assessing peach fruit firmness[J]. *Biosystems Engineering*, 2006, 93(2): 161 ~ 171.
- 13 Li X L, Hu X Y, He Y. New approach of discrimination of varieties of juicy peach by near infrared spectra based on PCA and MDA model[J]. *Journal of Infrared and Millimeter Waves*, 2006, 25(6): 417 ~ 420.
- 14 John H, Kalivas. Chemometrics with undergraduates[J]. *Chemometrics and Intelligent Laboratory Systems*, 1992, 15(2 ~ 3): 127 ~ 135.
- 15 Kirkpatrick S, Gelatt Jr C D, Vecchi M P. Optimization by simulated annealing[J]. *Science*, 1983, 220: 671 ~ 680.
- 16 邹小波, 赵杰文. 农产品无损检测技术与数据分析方法[M]. 北京:中国轻工业出版社, 2007:336 ~ 337.
- 17 李桂峰, 赵国建, 刘兴华, 等. 苹果硬度的傅里叶变换近红外光谱无损检测[J]. *农业机械学报*, 2009, 40(1): 120 ~ 123.
Li Guifeng, Zhao Guojian, Liu Xinghua, et al. Using FT-NIR spectra in non-invasive measurements of apple firmness[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2009, 40(1): 120 ~ 123. (in Chinese)
- 18 陈香维, 杨公明. 测试部位对猕猴桃近红外光谱响应特性的影响[J]. *农业机械学报*, 2009, 40(1): 124 ~ 128.
Chen Xiangwei, Yang Gongming. Influences of test locations on sugar content prediction of kiwifruit with near-infrared diffuse reflectance technique [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2009, 40(1): 124 ~ 128. (in Chinese)
- 19 夏俊芳, 李小昱. 脐橙 VC 含量近红外光谱消噪小波优化[J]. *农业机械学报*, 2009, 40(4): 143 ~ 146.
Xia Junfang, Li Xiaoyu. Wavelet optimization for near-infrared spectra denoising of vitamin C content of umbilical orange [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2009, 40(4): 143 ~ 146. (in Chinese)
- 20 孙旭东, 章海亮, 欧阳爱国, 等. 南丰蜜桔可溶性固形物近红外特征波段选择[J]. *农业机械学报*, 2009, 40(7): 129 ~ 132.
Sun Xudong, Zhang Hailiang, Ouyang Aiguo, et al. Selection of NIR characteristic wavelength bands for soluble solids content in Nanfeng mandarin fruit[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2009, 40(7): 129 ~ 132. (in Chinese)