

doi:10.6041/j.issn.1000-1298.2013.01.028

桑蚕鲜茧干壳量的可见/近红外光谱无损检测*

金航峰¹ 黄凌霞¹ 谢琳² 金佩华³ 楼程富¹

(1. 浙江大学动物科学学院, 杭州 310058; 2. 上海大学经济学院, 上海 200444;

3. 浙江农林大学林业与生物技术学院, 杭州 311300)

摘要: 选择 Savitzky-Golay 平滑作为光谱数据的预处理方法, 根据偏最小二乘模型的回归系数进行有效波长的选取, 最终筛选出了桑蚕鲜茧干壳量指标在可见/近红外光谱谱区的 7 个有效波长, 并结合多元线性回归建立干壳量的检测模型。该模型运算简单且检测精度较高, 预测决定系数和剩余预测偏差分别为 0.758 7 和 2.046 4, 是应用可见/近红外光谱检测桑蚕鲜茧干壳量的理想模型。

关键词: 桑蚕鲜茧 干壳量 可见/近红外光谱 有效波长 无损检测

中图分类号: O657.3; S886.3 **文献标识码:** A **文章编号:** 1000-1298(2013)01-0147-05

Nondestructive Detection of Dry Weight of Cocoons Layer of Mulberry Silkmoth Fresh Cocoons Using Visible/Near Infrared Spectroscopy

Jin Hangfeng¹ Huang Lingxia¹ Xie Lin² Jin Peihua³ Lou Chengfu¹

(1. College of Animal Sciences, Zhejiang University, Hangzhou 310058, China 2. School of Economics, Shanghai University, Shanghai 200444, China 3. School of Forestry and Bio-technology, Zhejiang A&F University, Hangzhou 311300, China)

Abstract: Visible/near infrared (Vis-NIR) spectroscopy was investigated to determine the dry weight of the cocoons layer of mulberry silkworm fresh cocoons. Optimal partial least squares (PLS) models were developed with different preprocessing, and the data preprocessed by Savitzky-Golay (SG) smoothing was chosen for the effective wavelengths selection. The selection was operated based on regression coefficients in PLS models, and reduced the original 601 varieties into 7. Then multiple linear regression (MLR) was used for calibration and prediction based on the seven effective wavelengths, compared with the PLS model built on full-spectrum data. The results showed that MLR model was the optimum model for the dry weight of the cocoons layer detection in the process of production and marketing, because of its simple arithmetic and accurate detection. The correlation coefficient and residual predictive deviation were 0.758 7 and 2.046 4.

Key words: Mulberry silkworm fresh cocoon Dry weight of the cocoons layer Visible/near infrared spectroscopy Effective wavelength Nondestructive detection

引言

蚕茧质量的优劣直接关系到缫丝业的经济效益。现行蚕茧质量的主要检测方法是 GB/T 19113—2003《桑蚕鲜茧分级(干壳量法)》中规定的

干壳量法。该方法可以较好地纠正鲜茧茧层含水对计价的负面影响, 使鲜茧收购的质量要求更加规范^[1]。但是切割蚕茧称量干壳量需要破坏样本, 会造成较大的经济浪费。据统计, 我国每年为检测蚕茧而浪费的鲜蚕茧超过千吨, 直接经济损失达到千

收稿日期: 2012-03-22 修回日期: 2012-09-04

* 浙江省自然科学基金资助项目(LY12C17001)、高等学校博士学科点专项科研基金资助项目(20100101120084)、浙江省公益技术研究农业项目(2011C22075)和农业科技成果转化资金项目(2011GB23600008)

作者简介: 金航峰, 博士生, 主要从事蚕茧品质无损检测研究, E-mail: jinhangfeng@zju.edu.cn

通讯作者: 黄凌霞, 讲师, 主要从事数字农业和农产品光谱检测研究, E-mail: lxhuang@zju.edu.cn

万元^[2]。加上该方法操作繁琐,需要耗费较多的时间,所以无法在实际鲜茧产销过程中全面贯彻。目前一些地方仍然以传统的手估目测方法对蚕茧进行评级定价,不可避免的主观性和随机性导致同级蚕茧品质不统一,在一定程度上制约了我国蚕丝行业的产量、质量和效益的提高。因此有必要研究一种能够快速、准确和无损获取鲜茧干壳量信息的检测和分级方法,提高蚕茧生产和收购的信息化水平。

关于蚕茧干壳量简易、快速检测技术的研究,早期研究方法主要有动量分离法、弹力测定法和扣减水分法等。前2种方法经试验证实可行性较差,而扣减水分法需要预先测得鲜茧壳的质量,同样会对样本造成破坏,所以均没有在实际生产中得到应用^[3]。近几年随着无损检测技术的快速发展,一些新技术被用于蚕茧干壳量的无损检测研究^[4-5]。但这些研究都没有对干壳量指标进行直接判别,在一定程度上影响了检测精度,所以没有得到广泛的推广应用。

近红外(Near infrared, NIR)光谱技术是一种低成本、无污染的分析技术,操作简便,且测定速度快。它可以充分利用全谱段或多波长下的光谱数据对物质的品质、种类、化学成分等进行定性和定量分析,已被广泛应用于食品、农业、制药等领域^[6-9]。在蚕茧品质检测方面也有相关的应用,如对雌雄鲜茧、死笼茧的非破坏性识别和蚕茧茧层量的无损检测^[10-13]。

本文研究目的是验证可见/近红外光谱技术用于检测桑蚕鲜茧干壳量的可行性,并选取干壳量指标在可见/近红外光谱谱区的有效波长,建立干壳量快速无损检测模型,探讨蚕茧产销过程中快速、无损检测干壳量指标的可行方法。

1 试验材料与方法

1.1 样品准备

试验选取2011年秋季(9~10月)浙江省湖州市农业科学研究院蚕桑科学研究所提供的秋丰×白玉、白玉×秋丰、菁松×皓月、皓月×菁松4个品种的鲜茧,在上簇后第5天采鲜茧。光谱测量前,在实验室25℃恒温条件下先避光贮藏3h。

1.2 光谱采集

试验使用美国ASD(Analytical Spectral Device)公司的Handheld FieldSpec光谱仪,其波长范围为325~1075 nm,探头视场角45°,光源为14.5 V卤素灯。分析软件为Unscramble V9.6。

将鲜茧剥去茧衣后,采用漫反射模式采集光茧

的近红外光谱。连接光谱仪的光纤探头垂直放置于距离样本表面10 cm处。光源与水平位置成45°角。光谱扫描稳定后进行光谱数据的采集,每个样本翻转扫描2次,取平均值。

1.3 干壳量测定

光谱扫描后,先用电子天平测得新鲜蚕茧的全茧质量,记为 M_1 。然后将蚕茧削开一个小口,倒出蚕蛹,观察蚕蛹是否发育良好。如发现毛脚茧、死笼茧、烂茧等下车茧,须从样本集中剔除。每个品种挑选100个上车茧的茧壳,共400个样本放入120℃干燥箱内,2 h后取出,迅速用电子天平测得无水茧壳的质量,记为 M_2 。干壳量定义为50 g鲜上车茧的光茧壳干燥后的质量。

将样本随机分成建模集和预测集,建模集包含300个样本,每个品种75个;预测集包含100个样本,每个品种25个。蚕茧干壳量的统计描述参数如表1所示。建模集、预测集和全部样本集所对应的平均值和标准差相近,说明样品抽取的随机性较好,所建模型的普适性较广。

表1 蚕茧干壳量的统计描述参数

Tab.1 Descriptive statistic data of dry weight of the cocoons layer

数据集	样本数	平均值/g	测量范围/g	标准差/g
建模集	300	11.802 2	9.427 8 ~ 13.938 6	1.057 1
预测集	100	11.825 1	9.656 0 ~ 13.850 8	1.017 5
全集	400	11.807 8	9.427 8 ~ 13.938 6	1.046 1

1.4 光谱数据预处理

由于光谱信号在采集范围的首端与末端存在较大的波动,所以取相对稳定的400~1000 nm波段共601个波长的吸光度作为数据处理对象。同时为了去除光谱信号中由于随机误差叠加的噪声,提高光谱的信噪比,需要对光谱数据进行平滑去噪。Savitzky-Golay卷积平滑是一种常用的信号平滑方法,通过多项式来对移动窗口内的数据进行多项式最小二乘拟合,实现去噪。本文选择经Savitzky-Golay平滑处理的光谱数据作为后续处理的对象。

1.5 基于回归系数的有效波长选择

全波段建模存在数据冗余、计算处理时间较长等缺点,为缩短数据处理时间以及充分挖掘原始光谱对预测结果的有效信息,需要对全波段数据进行有效波长(Effective wavelength, EW)提取。常用的有效波长选取方法有连续投影算法(Successive projections algorithm, SPA)、回归系数法(Regression coefficient, RC)和独立组分分析法(Independent component analysis, ICA)等。

回归系数是回归分析中度量因变量对自变量相依程度的指标,回归系数的绝对值越大表示自变量对因变量的影响越大。本文中波长对应的回归系数绝对值越大,表示该波长对桑蚕鲜茧干壳量的检测效果的影响就越大。该方法运算简单且适用性广,所以可以用于有效波长的选择。

1.6 建模效果评价

在对模型的建模效果进行评价时,常用的评价指标主要有建模集的决定系数(R_c^2)和均方根误差(RMSEC),以及建模集交互验证的决定系数(R_{cv}^2)和均方根误差(RMSECV)。建模效果较好的模型需要有较高的 R_c^2 和 R_{cv}^2 ,同时为了防止模型出现过拟合,RMSEC和RMSECV之间的差值应较小。所以提出了一个新的模型建模效果评价指标

$$V = \frac{R_c^2 + R_{cv}^2}{1 + \text{abs}(R_{\text{MSEC}} - R_{\text{MSECV}})} \quad (1)$$

当 V 取到较大值时,该模型的建模效果较好,且没有出现拟合现象。对于模型预测效果的评价,常用指标主要有预测集的决定系数(R_p^2)和均方根误差(RMSEP),以及剩余预测偏差(RPD)。RPD是预测集样品干壳量指标的方差与模型预测标准误差(SEP)的比值,RPD值越大,对应模型的预测效果越好。

表 2 基于不同波长数和主成分数的模型预测结果

Tab.2 Prediction results of models with different wavelengths and principal component numbers

阈值	波长数	主成分数	R_c^2	RMSEC	V	R_p^2	RMSEP	RPD
0	601	8	0.850 4	0.408 2	1.630 2	0.811 0	0.440 0	2.312 5
16	374	7	0.850 5	0.408 0	1.647 0	0.814 5	0.435 9	2.334 2
30	157	8	0.849 9	0.408 8	1.622 0	0.809 9	0.441 3	2.305 7
16/10	92	7	0.841 3	0.420 3		0.806 4	0.447 0	2.276 3
30/20	43	7	0.828 1	0.437 4		0.802 0	0.450 3	2.259 6
多元线性回归	7		0.789 2	0.484 4		0.758 7	0.497 2	2.046 4

2.2 回归系数阈值确定

从主成分数为 8 的 PLS 模型中获得各波长对应的回归系数绝对值后,通过设定不同的阈值,将回归系数绝对值低于所设阈值对应数值的波长剔除,剩余波长对应的光谱数据进行重新组合并建立 PLS 模型。将阈值设定为 0~99,对应数值为回归系数绝对值中最大值的不同百分比。601 个波长在各阈值区间内的分布如图 2 所示,当阈值较小时,区间内的波长分布较多,且对应的回归系数绝对值较小,提高了有效波长筛选的效率;而当阈值较大时,波长对模型检测效果的影响较大,区间内波长数几乎均小于 5,确保了有效信号最大程度上的保留。

根据设定的阈值 0~99,当阈值为 0 时,全部波长的光谱数据均用于建模;当阈值为 50 时,回归系

2 结果与分析

2.1 回归系数获取

将 Savitzky - Golay 平滑后的波长数据作为模型输入变量,将蚕茧的干壳量作为输出变量,通过建模集样本建立偏最小二乘(Partial least squares, PLS)模型,并用完全交互验证对模型进行验证。通过比较基于不同主成分数的模型对预测集样本的预测效果,发现当主成分数取 8 时,模型的 V 值最高(见图 1),为 1.630 2,说明建模效果较好。且 R_p^2 为 0.811 0,RPD 为 2.312 5(表 2),说明预测效果较好。从该模型中获取各波长对应回归系数的绝对值,其中波长 400 nm 的回归系数绝对值最大,为 22.103。

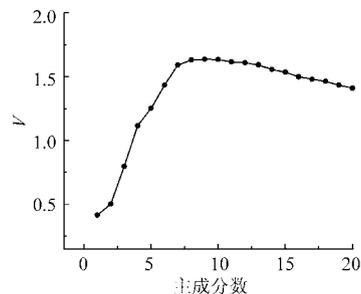


图 1 基于不同主成分数的 PLS 模型评价价值

Fig.1 V value of PLS models with different principal components

数绝对值小于最大值 50% 的波长数据将会被剔除;而当阈值为 100 时,所有波长的数据均会被剔除,模型无法建立。

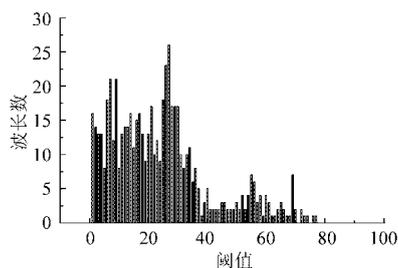


图 2 各阈值区间内波长数分布图

Fig.2 Numbers of wavelength in each threshold intervals

2.3 PLS 模型建立

以最大值 22.103 的不同百分比的值作为挑选

波长的标准,当阈值取 50 时,回归系数绝对值小于最大值的 50% (11.051 5) 的 534 个波长将会被剔除,剩余 67 个波长的数据用于重新建模。同时选取 1~20 不同的主成分数,共建立了 2 000 个不同阈值和主成分数组合的 PLS 模型。最后计算各模型的 V 值,如图 3 所示。

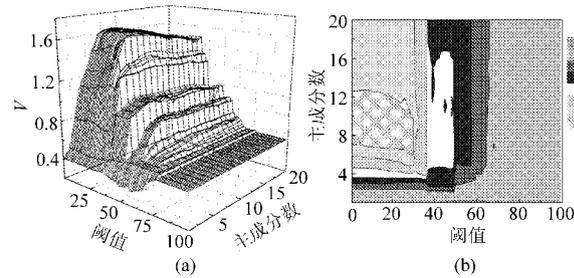


图 3 各 PLS 模型的评价值

Fig. 3 Plot of V value of PLS models

(a) 三维分布图 (b) 二维分布图

从图 3 中可以看出,当阈值大于 65 时,对应模型的 V 值均较低,这是因为用于建模的波长已少于 20 个,过多的信息被剔除,导致建模效果不佳。当阈值小于 30,且主成分数在 7~13 之间时,多数模型的 V 值大于 1.6,说明该区域内的阈值和主成分数的组合建模效果较好。而当主成分数大于 13 时, V 值开始降低,这是因为由于主成分数选择过多,模型出现了过拟合现象。

当阈值取 16,主成分数取 7 时,对应模型的 V 值最大,为 1.647 0,374 个波长的数据被用于建模。16-7-PLS 模型的 R_p^2 为 0.814 5,RPD 为 2.334 2。与全波段光谱相比,超过 1/3 的波长数据被剔除,但 R_p^2 和 RPD 分别增大了 0.43% 和 0.94%。

同时在 V 大于 1.6 的 164 个模型中,取阈值的最大值 30,用于建模的波长数为 157。当主成分数取 8 时,对应模型的 V 最大,为 1.622 0。30-8-PLS 模型的 R_p^2 为 0.809 9,RPD 为 2.305 7。与全波段光谱相比,仅剩 1/4 的波长数据被用于建模,但 R_p^2 和 RPD 仅分别减小了 0.14% 和 0.29%。说明根据回归系数对无用波长数据进行剔除的方法是可行的。

2.4 有效波长选择

为了进一步选取与桑蚕鲜茧干壳量相关的有效波长信息,根据 16-7-PLS 和 30-8-PLS 模型中各波长对应的回归系数进行二次筛选。如图 4 所示,将 16-7-PLS 模型中回归系数绝对值小于 10 的波长剔除,剩余 92 个波长数据重新建立 PLS 模型。该模型的 R_p^2 为 0.806 4,RPD 为 2.276 3。同时将 30-8-PLS 模型中回归系数绝对值小于 20 的波长剔除,剩余 43 个波长数据重新建立 PLS 模型。该

模型的 R_p^2 为 0.802 0,RPD 为 2.259 6。与全波段光谱相比,仅剩 7.15% 的波长数据被用于建模,但 R_p^2 和 RPD 仅分别减小了 1.11% 和 2.13%,说明剩余的 43 个波长的数据包含了大部分与桑蚕鲜茧干壳量相关的有效信息。

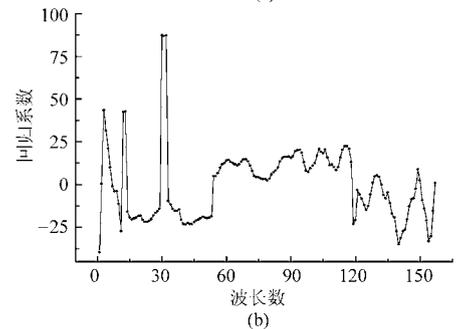
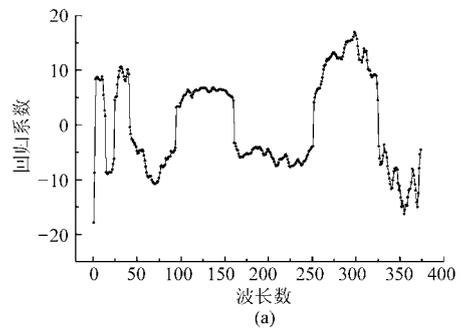


图 4 模型的回归系数图

Fig. 4 Regression coefficients of PLS models

(a) 16-7-PLS (b) 30-8-PLS

从 30-8-PLS 模型中筛选出的 43 个有效波长在原始光谱曲线上的分布如图 5 所示。从图中可以看出,有效波长主要分布在 402~415、458~459、514~523、704~706、823~831、916~947 和 976~992 nm 这 7 个光谱波段范围内。根据 43 个有效波长建立的 PLS 模型中各波长的回归系数,从每个波段中挑选出最大绝对值对应的波长,分别为 402、458、514、705、829、916 和 992 nm 共 7 个有效波长。

2.5 建模方法选择

为了进一步简化模型,使得更适用于实际产销过程中仪器开发和推广应用,对筛选出的 7 个有效波长进行多元线性回归 (Multiple linear regression, MLR) 建模,所建立的多元线性回归方程为: $y = 8.784 + 19.365\lambda_{402} + 85.644\lambda_{458} - 290.19\lambda_{514} + 237.665\lambda_{705} - 271.843\lambda_{829} + 477.733\lambda_{916} - 253.539\lambda_{992}$, 方程中 y 为桑蚕鲜茧的干壳量, λ 为各波长对应的蚕茧样本的吸光度,下角标数字表示对应的波长。该模型的 R_p^2 为 0.758 7,RPD 为 2.046 4。与全波段光谱相比,仅剩 1.16% 的波长数据被用于建模, R_p^2 和 RPD 分别减少了 6.45% 和 11.51%。预测值与真实值的相关关系见图 6。

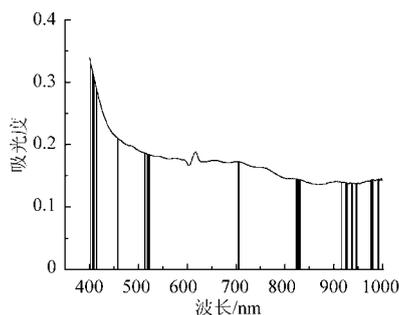


图5 43个有效波长在原始光谱曲线上的分布

Fig.5 Distribution of 43 effective wavelengths on Vis/NIR spectra

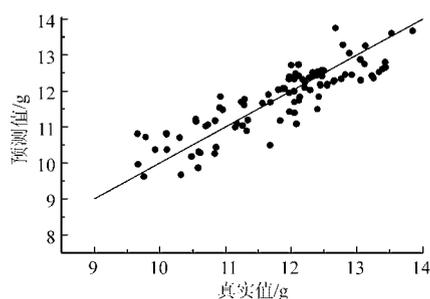


图6 基于7个有效波长的MLR模型预测值与真实值的相关关系

Fig.6 Correlativity between predicted and reference values in MLR model

3 结论

(1) 可见/近红外光谱技术实现了桑蚕鲜茧干壳量指标的无损检测,验证了该方法的可行性。

(2) 与声波振动信号处理技术和数字图像处理技术相比,可见/近红外光谱技术不需要其他指标的参考,可直接对桑蚕鲜茧的干壳量进行检测,检测精

度有了较大的提升。

(3) 可见/近红外光谱技术可对单个样本进行检测,实现了蚕茧品质的精确评价,有利于蚕茧等级的正确划分,提高了茧丝绸产品的产量和质量。

(4) 最终建立的检测模型相对简单,且检测精度较高,可实现蚕茧干壳量实时、在线检测的目标。

参 考 文 献

- 1 许建华. 解读 GB/T 19113—2003《桑蚕鲜茧分级(干壳量法)》[J]. 中国纤检, 2005(6):23.
- 2 顾国达, 姜丽花. 我国桑蚕茧评级标准和收购价格的演变[J]. 丝绸, 2002(11):44~47.
- 3 胡祚忠, 文勇. 桑蚕茧干壳量快速检测研究[J]. 纤维标准与检验, 1998(11):1~3.
- 4 胡兴明, 吴恢, 叶楚华, 等. 基于神经网络集成的蚕茧干壳量无损检测方法研究[J]. 蚕业科学, 2008, 34(4):781~785. Hu Xingming, Wu Hui, Ye Chuhua, et al. Research on the non-destructive testing method for dried shell weight of cocoon based on neural network ensemble[J]. Science of Sericulture, 2008, 34(4):781~785. (in Chinese)
- 5 甘勇, 孔庆华, 韦荔甫. 基于图像处理技术的蚕茧干壳量无损检测方法及其影响因素分析[J]. 丝绸, 2008(12):39~41.
- 6 Sivertsen A H, Kimiya T, Heia K. Automatic freshness assessment of cod (*Gadus morhua*) fillets by Vis/Nir spectroscopy[J]. Journal of Food Engineering, 2011, 103(3):317~323.
- 7 Liu Y D, Sun X D, Dong X L, et al. Non-destructive measurement of soluble solids content and vitamin C in gannan navel oranges by Vis-NIR spectroscopy[J]. Sensor Letters, 2011, 9(3):1133~1139.
- 8 Vohland M, Michel K, Ludwig B. Use of near-infrared spectroscopy to distinguish carbon and nitrogen originating from char and forest-floor material in soils; usefulness of a genetic algorithm[J]. Journal of Plant Nutrition and Soil Science, 2011, 174(5):695~701.
- 9 Said M M, Gibbons S, Moffat A C, et al. Near-infrared spectroscopy (NIRS) and chemometric analysis of Malaysian and UK paracetamol tablets; a spectral database study[J]. International Journal of Pharmaceutics, 2011, 415(1~2):102~109.
- 10 潘沈元, 金同铭, 陆国权, 等. 蚕茧近红外反射(NIR)光谱的模式识别——I. 对雌雄鲜茧、死茧的非破坏性识别[J]. 生物物理学报, 1995, 11(1):53~59.
- 11 颜辉, 陈斌, 金尚忠. 基于NIR的主成分结合支持向量机鉴别蚕茧雌雄的研究[J]. 计算机与应用化学, 2008, 25(10):1261~1264. Yan Hui, Chen Bin, Jin Shangzhong. Study on the method of determining cocoon sexuality by PCA plus SVM based on NIR[J]. Computers and Applied Chemistry, 2008, 25(10):1261~1264. (in Chinese)
- 12 李玉品, 陈斌, 颜辉, 等. 近红外光谱分析法在蚕茧性别鉴定上的应用[J]. 江苏农业科学, 2009(1):237~239.
- 13 黄凌霞, 吴迪, 金航峰, 等. 基于变量选择的蚕茧茧层量可见-近红外光谱无损检测[J]. 农业工程学报, 2010, 26(2):231~236. Huang Lingxia, Wu Di, Jin Hangfeng, et al. Non-destructive detection of cocoon shell weight based on variable selection by visible and near infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering, 2010, 26(2):231~236. (in Chinese)
- 14 王娟, 张荣芳, 王相友. 双孢蘑菇硬度的近红外漫反射光谱无损检测[J]. 农业机械学报, 2012, 43(11):163~168. Wang Juan, Zhang Rongfang, Wang Xiangyou. Non-destructive detection of *Agaricus bisporus* firmness based on near-infrared diffused spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(11):163~168. (in Chinese)