

基于对地抽样总量控制下的玉米种植面积提取

王 双, 朱秀芳, 潘耀忠, 徐 超, 李 乐

北京师范大学 环境演变与自然灾害教育部重点实验室, 资源学院, 北京 100875

摘 要: 提出了一种基于统计抽样总量控制下的中高分辨率遥感影像玉米种植面积信息提取方法, 该方法首先利用分层抽样技术对调查目标总体(玉米)进行分层抽样; 然后对抽样小区进行目视解译, 反推区域总量真值; 最后在总量控制下进行区域目标作物的空间分布提取。以河北省三河市中部地区的部分影像为研究区, 以该区 2006-08-21 的 10m 分辨率的 SPOT 5 多光谱影像为基础数据进行了试验研究。结果表明该方法基于群样本检验的总体精度达到 93.8%, Kappa 系数达到 0.88, 均高于最大似然监督分类结果的精度。另外, 所提出的方法在大范围农作物信息提取方面具有一定的实用性和推广性。

关键词: 分层抽样, 总量控制, 群样本检验, 面积监测, 玉米

中图分类号: TP79

文献标识码: A

1 引 言

准确掌握玉米的种植面积、监测其空间分布、长势变化和产量, 不仅能够为政府统计核算提供必要的数据库支持, 还能为国家的粮食安全、战略部署和相关决策提供必要的依据(朱彬, 1998; 陈世军 & 蒋和平, 1998)。

很多学者利用各种方法对玉米面积测量进行了相关的研究, 如 Badhwar(1984)利用多时相 Landsat MSS/TM 影像获取了玉米、大豆、小麦等作物的面积, 并且将该方法从美国推广到阿根廷; 树文和薄立群 (2000)利用生态分类法进行了吉林梨树县玉米种植面积提取的试验, 提取的作物种植面积可用于产量的估算; Jakubauskas 等. (2002)利用 NOAA/AVHRR-NDVI 时间序列数据的谐波函数分析结果, 识别出不同农作物(玉米、大豆、苜蓿), 但分类结果受到影像分辨率低的影响; 钱永兰等(2005)利用 SPOT 5 多光谱遥感图像进行光谱差异增强, 并将增强后的遥感图像与全色波段进行融合, 对融合后的图像进行分类, 得到的分类精度达到 95%以上张林文鹏等(2006)利用多时相 TERRA/MODIS 数据, 采用波谱分析和决策树相结合的方法, 自动提取秋季作物的种植面积, 其结果与统计数据相比, 分类精度达

到 86%以上; 杨桃等(2006)基于 TM 遥感影像数据, 利用多特征空间遥感影像专题信息, 基于 ERDAS 中的 knowledge engineer 设计开发了基于多特征空间的玉米地专题信息库, 利用计算机对北京地区的玉米地进行了自动提取, 分类精度达 92.9%。

总的来说, 目前玉米信息的提取主要利用多时相、多特征信息来实现, 但多时相数据的获取及分类精度受影像分辨率的影响, 普遍认为中高分辨率数据的分类精度要优于低分辨率数据的。但由于同谱异物、同物异谱以及混合像元等原因, 即使是中高分辨率遥感图像在信息提取方面也存在无法避免的不确定性。为此, 本文提出一种基于统计抽样总量控制下的中高分辨率遥感影像夏玉米种植面积信息提取方法。该方法利用抽样小区进行目视解译, 得到区域总量真值; 在总量控制下进行区域目标作物的空间分布提取, 既减少了野外工作量, 提高了工作效率, 又能够有效提取目标作物, 为大范围农作物遥感信息提取提供有效的基础数据。

2 研究区及其数据

2.1 研究区

研究区(图 1)位于河北省三河市中部, 范围为

收稿日期: 2008-02-26; 修订日期: 2008-06-02

基金项目: 国家高技术研究发展计划(课题编号: 2006AA120101)。

第一作者简介: 王双(1981—), 女, 硕士研究生, 从事土地覆盖遥感监测研究。

通讯作者: 朱秀芳, E-mail: xiufang.zhu@gmail.com。

10km×10km。三河位于华北平原(海河流域, 处于东经 116°45′—117°15′, 北纬 39°48′—40°05′之间), 地处京津唐金三角地带, 西襟北京, 南连天津, 东临唐山, 北依燕山山脉, 平均气温 12℃, 年平均最高气温 18.3℃。年平均最低气温 6.8℃。全年无霜期 203d, 年日照时数 2561.5h, 年总降水量 369.8mm, 属于暖湿半湿润地区, 满足玉米生长需水较多、喜温、喜光的特点, 适宜种植玉米, 是河北省的玉米种植大市。该市玉米种植以夏玉米为主, 播期通常在 6—9 月份, 同期作物还有大豆和棉花等。

2.2 数据准备

本研究采用数据包括 SPOT5 多光谱影像以及野外样方数据。其中 SPOT5 多光谱影像时相为 2006-08-21(此时, 玉米正处抽雄期, 大豆和棉花分别处于结荚和开花期, 在图像上表现处不同的光谱色调, 易于识别, 图 1), 像元大小为 10m, 经过辐射校正将 DN 值转换为反射率、几何纠正误差小于一个像元, 投影类型为 Albers 等积投影, 椭球体为 WGS84。野外调查时间为 2006 年 9 月, 采用简单随机抽样方法, 结合交通的可达性, 在研究区布设采样点。野外样方数据主要为点数据, 包括纯的玉米、裸地、树林、菜地等, 要求该种地块的面积较大, 可以在影像上清楚的识别, 为后续室内玉米的识别与提取提供分类模板和参照, 同时为精度验证提供实测数据集, 另外还有一些地面控制点, 辅助几何纠正和点位漂移的控制。

3 技术路线

技术流程主要包括如下 3 个步骤(图 2)。

(1) 抽样入样总体的获得。首先对夏玉米及其他植被类型进行光谱分析, 提取植被指数并对图像进行光谱差异增强, 为玉米信息提取进行数据预处理, 然后利用最大似然分类得到研究区的玉米分布的图像, 将得到的图像和玉米种植面积作为入样总体。

(2) 抽样反推得到研究区的真实的玉米种植总面积: 对第一步得到的结果进行分层抽样, 对抽中的样本进行目视解译得到样区内玉米真实种植面积, 进而反推整个研究区的玉米种植面积, 并将其视为真实值。

(3) 在真实总量的控制下进行玉米信息提取, 得到最终的玉米空间分布图。

3.1 入样总体的获得

研究区中地物类型有玉米、大豆、棉花、荷塘和林地等, 在正式分类前为了增大各种地物间的光谱差异, 提高分类精度, 借鉴已有相关研究(钱永兰等, 2005; 邓劲松等, 2004), 对原始图像进行了加和比植被指数 PRVI 的计算和光谱差异增强处理。然后利用最大似然分类将研究区分为玉米、大豆、棉花、水体(荷塘)、林地和建筑用地。最后经过合并得到玉米分布的二值图(图 3(a), 0 为图中黑色区域表示非玉米, 1 为图中白色区域表示玉米), 将得到的图像和玉米种植面积作为入样总。值得一提的是, 对原始图像进行了加和比植被指数 PRVI 的计算和光谱差异增强处理后, 通过 N 维可视化(图 4(a))我们可以发现, 在 SPOT4、1、2 三波段特征空间中, 玉米和棉花、大豆已经能够有效的分开, 但是玉米和林地的存在一定混淆, 我们进一步作出了玉米、树木、棉花和大豆 4 种植被样本的 PRVI 均值和方差(见图 4(b)),

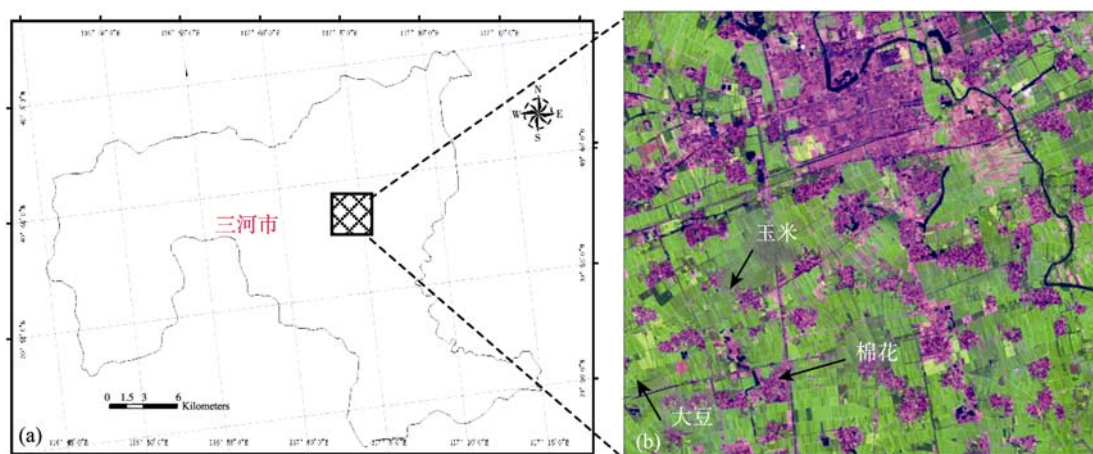


图 1 研究区

(a) 三河市区研究区地理位置; (b) 研究区的 SPOT 5 遥感影像(RGB412)

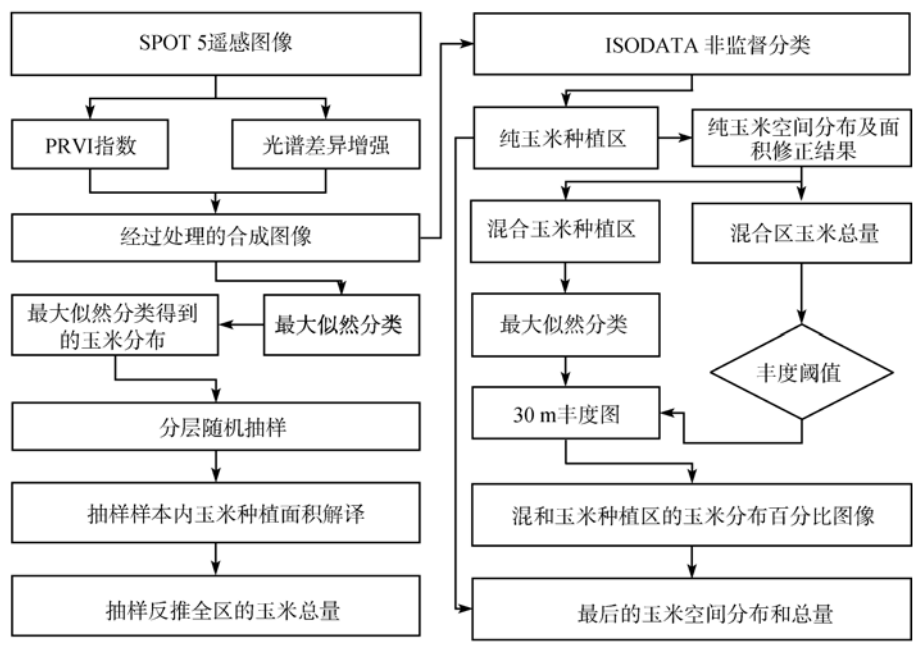


图 2 技术路线图

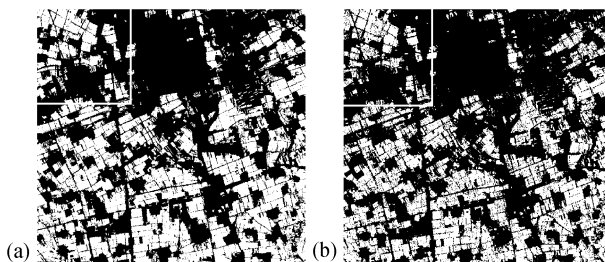


图 3 玉米分布图

(a) 最大似然法得到的玉米分布图; (b) 总量控制法得到的玉米分布图

可以明显的看到玉米样本的 PRVI 上限低于棉花和大豆 PRVI 值的下限, 因此, 玉米可以和棉花、大豆有效的分离。但是玉米和树木的 PRVI 取值范围有部分重合, 因此, 玉米和树木容易混淆。这也是我们后续处理的重点所在。

(1) 光谱差异增强计算公式

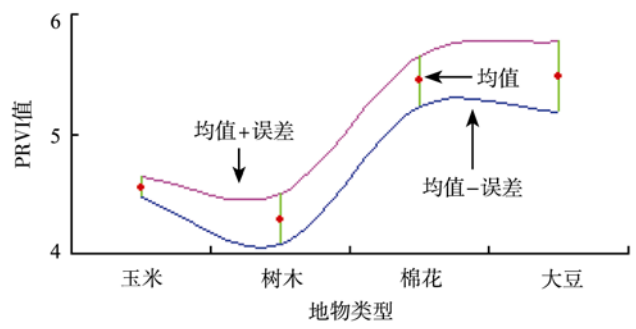
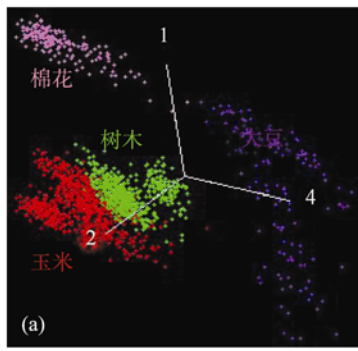


图 4 植被光谱分析图

(a) 各植被类型的 N 维特征空间; (b) 各种地物的 PRVI 均值和方差

$$B_N P'_{ij} = B_N P_{ij} + \left(B_N P_{ij} - \frac{\sum_{M=1}^n B_M P_{ij}}{n} \right) \quad (1)$$

式中, $B_N P_{ij}$ 表示原始图像第 N 波段位于 (i, j) 位置的像元值, N 为对应的波段, n 为波段数, $B_N P'_{ij}$ 表示第 N 波段图像位于 (i, j) 位置的新的像元值。由公式可知, 新生成的图像体现了地物在个体波段以及全部波段差异的和, 考虑了所有波段量化值的影响, 加大了地物之间的光谱差异。

(2) PRVI 计算公式

$$PRVI = (\rho_{sw} + \rho_{nir}) / \rho_R \quad (2)$$

式中, ρ_{sw} , ρ_R 和 ρ_{nir} 分别是远红外, 红光, 红外波段的反射率。

3.2 抽样反推玉米真实总量

一般来说分层数目越多, 抽样统计越准确, 同

时抽样费用和工作量也会大大增加,大量研究表明层数为6时比较合适,如陈仲新等(2000年)在全国冬小麦面积变化遥感监测中,将全国1411个冬小麦生产县划分为6层,建立全国冬小麦面积变化遥感监测抽样外推模型。刘海启(2001)在进行大尺度耕地变化监测的遥感抽样方法研究时,通过6、8、10、12层分层抽样试验后,最终确定分层数为6。焦险峰等(2002)利用分层抽样进行全国棉花种植面积的遥感监测时,也确定分层层数为6层。为此,本研究以遥感图像的最大似然预分类结果为入样总体,以 $100\text{m} \times 100\text{m}$ 的大小为抽样框,采用分层随机抽样,在置信度为95%、允许误差为5%条件下,对分层数为6—10层的不同抽样结果(表1)进行了分析,最后选取总量精度最高、抽样总量最小的一组抽样框图像(即层数为8时的图像)作为抽样样本库。

表1 分层抽样精度表

分层数	抽样图像个数	抽样精度
6	13	0.951
7	14	0.993
8	8	0.999
9	9	0.992
10	10	0.997

在抽取抽样框后,对抽样框图像进行目视解译屏幕数字化,获取样框内玉米的种植面积,然后利用分层抽样反推公式反推出研究区总的玉米种植面积为 46305012m^2 (表2)。

反推公式定义如下:

$$T\hat{X} = A\bar{x}_x = A \frac{\sum_{i=1}^n A_i x_i}{\sum_{i=1}^n A_i} \quad (3)$$

式中, A 为总研究区的面积, A_i 为抽样的面积, x_i 为抽样样本中目标地物的比例, n 为抽样单元的数目。

表2 分层抽样总量反推结果表

抽样框	目视解译夏玉米数目/像元	夏玉米面积比例	各抽样框夏玉米面积总量/ m^2	反推总量/ m^2
1	198	0.4950	19800	46305012
2	205	0.5125	20500	
3	75	0.1875	7500	
4	13	0.0325	1300	
5	379	0.9475	37900	
6	163	0.4075	16300	
7	275	0.6875	27500	
8	151	0.3775	15100	

3.3 总量控制下的玉米提取

对光谱差异增强图像进行 isodata 非监督分类,结合野外测量数据,对非监督分类得到的每一个类别进行属性定义,然后利用 ERDAS 中的 recode 和 mask 功能,将整个研究区归并为纯玉米区域、混和玉米种植区和非玉米区域。其中提取的纯玉米种植区由于异物同谱的原因,还包括了部分城镇内部的细碎绿色植被,本研究中利用城镇斑块与周边绿色植被的邻接特征,对该地区进行修正,最后得到纯玉米像元 342745 个。

混合区包括玉米、林地和城镇内部及边缘细碎的植被区的处理包括如下几个步骤:

(1) 用抽样反推的研究区玉米总种植面积(46305012m^2)减去纯夏玉米区种植面积(34274500m^2)得到混合区玉米种植面积为 12030512m^2 。

(2) 对混和区单独进行最大似然分类,将分类结果合并成玉米和非玉米图,然后将该图降维生成 $30\text{m} \times 30\text{m}$ 的玉米种植百分比图,通过反复调整阈值,使得混和区内的玉米种植面积尽可能的接近第一步得到的总量,然后将丰度值大于等于最佳丰度阈值(本研究中的最佳丰度阈值为0.4)以上的像元定义为玉米,其他为非玉米。

(3) 将纯玉米区面积提取结果与混合玉米区提取结果拼接,得到全区的玉米分布图,见图3(b)。

3.4 精度评价

本研究是将抽样反推的结果视为真实值,在该总量控制下进行的分类,因此,精度评价的重点在于玉米空间位置的合理性。在评价空间位置精度时本文用的是群检验的方法。

3.4.1 群样本检验最小样本数的确定

群样本检验需要确定一个最小样本数,考虑到样本量需要根据群内方差和估计误差限计算,群内的二级样本点限定在所选的基础样本内,空间距离接近,不适合用传统的统计抽样计算方法。同时群内每一个像元有一个精度值,所有像元群的精度值是连续的(刘旭拢等,2006),采用了连续尺度的样本量公式计算群样本的最小样本数更适合本研究数据特征,因此,文中采用了连续尺度的样本量计算公式(式(4))。参考吴炳芳和李强子(2004)进行全国粮食种植区抽样时,利用连续抽样样本量公式计算得到群样本量后又添加了一定比例的修正量,得到最终的样本量的方法,本试验中也加入了5%的修正量,以保证足够的样本。

$$n = \left(\frac{s \times u_{1-\alpha/2}}{d \times \bar{x}} \right)^2 \quad (4)$$

式中, s 是试探性样本的标准差, $\bar{x}$ 是试探性样本精度值的均值, u 是对应于一定置信水平从正态分布的概率表上所查的值, d 为误差允许范围。

3.4.2 群样本获取

试探中, 本文采用了系统抽样的方法, 以 1% 的抽样比例, 获得了 100 个样本(大小为 100m × 100m), 对抽样框进行目视解译, 计算群样本统计量。得到试探性样本的标准差 $s=0.1791$, 试探性样本精度值

的均值 $\bar{x}=0.9382$ 。取 $\alpha=0.1$, 即置信度水平为 95%, 查表得 $u=1.96$, 误差允许范围为 $d=\pm 5\%$, 根据式(4)计算得到 $n=54$ 。加入 5% 的修正量, 得到样本量 n 值应为 56 个。系统抽样的样本框为 100 个, 超过了检验需求量, 满足检验要求。

3.4.3 基于对地抽样的群样本检验精度结果

利用上面的 100 个群样本目视解译的结果为真实值, 分别对最大似然分类结果和抽样总量控制下的提取结果进行群样本检验, 检验结果以误差矩阵形式给出(表 3)。

表 3 最大似然法分类和总量结果的群样本检验精度

参考图 被评价图		参考值：目视解译的真实值			参考图 被评价图		参考值：目视解译的真实值		
		玉米	非玉米	合计			玉米	非玉米	合计
最大似然分类结果	玉米	4327	962	5289	总量控制下的结果	玉米	4216	359	4575
	非玉米	157	4714	4871		非玉米	268	5317	5585
	合计	4484	5676	10160		合计	4484	5676	10160
总体精度=88.9%, Kappa=0.78					总体精度=93.8%, Kappa=0.88				

4 结论与讨论

本文利用 SPOT 5 多光谱影像, 采用基于对地抽样总量控制下的玉米种植面积遥感信息提取方法, 获取研究区的玉米种植面积和空间分布图。研究的主要结论如下:

(1) 本研究基于玉米光谱特性, 将其分为纯玉米种植区和混和区 2 类, 分别逐级进行信息提取。并结合各类地物的邻接特征, 对纯玉米种植区提取结果进行修正; 对混和区玉米, 根据其空间分布特征, 采用丰度值对其进行信息提取, 可以较好的提取边界地物和破碎地块的信息。

(2) 本文提出的基于对地抽样总量控制下的中高分辨率遥感图像夏玉米种植面积提取方法, 该方法能够较好的提取目标作物, 野外工作量小, 精度水平较高。对研究区的试验结果表明, 基于群样本检验的总体精度为 93.8%, Kappa 系数为 0.88, 均高于最大似然监督分类。

同时本文的研究中还存在一些值得继续深入的问题, 比如本研究采用分层抽样方法进行抽样图像提取时, 将预分类结果作为入样总体会对抽样结果有一定影响, 这种影响究竟有多大, 或者预分类精度到达多少情况下才能满足抽样要求等。但是无论如何, 本文提出的方法在大范围农作物信息提取方面, 具有很好的实用性。因为, 在大范围农作物中高

分辨率遥感图像很难全覆盖获取, 基于一定抽样地区的中高分辨率遥感图像进行高精度的信息提取后, 抽样反推得到区域总量; 然后在总量控制下进行低分辨率的遥感图像信息提取, 不仅可以获得较高的总体精度, 还可以得到较好的像元精度, 既克服依靠抽样方法仅能得到总量数值而无法得到作物空间分布的缺点, 又改善了低分辨率全覆盖遥感测量的精度。本文的后续工作将会选择合适的研究区, 同时收集研究区内的中高分辨率和低分辨率的遥感数据, 进行该方法的验证和可行性研究。

REFERENCES

- Badhwar G D. 1984. Classification of corn and soybeans using multitemporal thematic mapper data. *Remote Sensing of Environment*, **16**: 175—182
- Chen S J and Jiang H P. 1998. The supply and demands of corn in China. *The World of Survey and Research*, **10**: 17—22
- Chen Z X, Liu H Q, Zhou Q B, Yang G X and Liu J. 2000. Sampling and scaling scheme for monitoring the change of winter wheat acreage in China. *Transactions of the CSAE*, **16**(5): 126—129
- Deng J S, Wang K, Shen Z Q and Xu H W. 2004. Design tree algorithm of automatically extracting farmland information from SPOT-5 image based on characteristic bands. *Transactions of the CSAE*, **20**(6): 145—148
- Jakubauskas M E, Legates D R and Kastens J H. 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI

- data. *Computers and Electronics in Agriculture*, **37**: 127—139
- Jiao X F, Yang B J and Pei Z Y. 2002. Design of sampling method for cotton field area estimation using remote sensing at a national level. *Transactions of the CSAE*, **18**(4): 159—162
- Lin W P, Wang C Y, Chu D P, Niu Z and Qian Y L. 2006. Extraction of fall crop types based on spectral analysis. *Transactions of the CSAE*, **22**(9): 128—132
- Liu H Q. 2001. Sampling method with remote sensing for monitoring of cultivated land changes on large scale. *Transactions of the CSAE*, **17**(2): 168—1701
- Liu X L, He C Y, Pan Y Z, Yang M C and Zhang J S. 2006. Accuracy assessment of thematic classification based on point and cluster sample. *Journal of Remote Sensing*, **10**(3): 366—372
- Qian Y L, Yang B J, Pei Z Y, Jiao X F, Zhang S L, Wu Q, Wang Q F and Wang F. 2005. Image spectral difference enhance and crop identification. Proceedings of Conference on Chinese Society of Agricultural Engineering
- Wu B F and Li Q Z. 2004. Crop acreage estimation using two individual sampling frameworks with stratification. *Journal of Remote Sensing*, **8**(6): 551—569
- Yang G, Zhang B, Bian H F, Wang Z M and Wang Z Q. 2006. Automatically extracting information from maize field based on TM remote sensing image. *Resources Science*, **28**(4): 91—96
- Zhang S W and BO L Q. 2000. Study on ecoland classification of remote sensing image in crop estimation. *Scientia Geographica Sinica*, **20**(6): 415—425
- Zhu B. 2000. Analysis of Some Grain Breeds in China under the Goal of Grain Security. Thesis Defense Committee for Master Degree of Nanjing Agricultural University
- ### 附中文参考文献
- 陈世军, 蒋和平. 1998. 中国玉米供给与需求. 调研世界, **10**: 17—22
- 陈仲新, 刘海启, 周清波, 杨桂霞, 刘佳. 2000. 全国冬小麦面积变化遥感监测抽样外套方法的研究. 农业工程学报, **16**(5): 126—129
- 邓劲松, 王珂, 沈掌泉, 徐红卫. 2004. 基于特征波段的 SPOT 5 卫星影像耕地信息自动提取的方法研究. 农业工程学报, **20**(6): 145—148
- 焦险峰, 杨邦杰, 裴志远. 2002. 全国棉花种植面积遥感监测抽样方法设计. 农业工程学报, **18**(4): 159—162
- 林文鹏, 王长耀, 储德平, 牛铮, 钱永兰. 2006. 基于光谱特征分析的主要秋季作物类型提取研究. 农业工程学报, **22**(9): 128—132
- 刘海启. 2001. 大尺度耕地变化监测的遥感抽样方法研究. 农业工程学报, **17**(2): 168—1701
- 刘旭拢, 何春阳, 潘耀忠, 杨明川, 张锦水. 2006. 遥感图像分类精度的点、群样本检验与评估. 遥感学报, **10**(3): 366—372
- 钱永兰, 杨邦杰, 裴志远, 焦险峰, 张松岭, 吴全, 汪庆发, 王飞. 2005. 遥感图像光谱差异增强与农作物识别. 中国农业工程学会 2005 年学术年会论文集
- 吴炳方, 李强子. 2004. 基于两个独立抽样框架的农作物种植面积遥感估算方法. 遥感学报, **8**(6): 551—569
- 杨桃, 张柏, 边红枫, 王宗明, 王志强. 2006. 基于 TM 遥感影像的玉米地专题信息自动提取. 资源科学, **28**(4): 91—96
- 张树文, 薄立群. 2000. 遥感图像生态土地分类法在农作物种植面积提取中的应用. 地理科学, **20**(6): 415—425
- 朱彬. 2000. 粮食安全目标下我国主要粮食品种结构分析. 南京农业大学硕士学位论文