

文章编号: 1007-4619(2004)06-0551-19

基于两个独立抽样框架的农作物种植面积遥感估算方法

吴炳方, 李强子

(中国科学院 遥感应用研究所, 北京 100101)

摘要: 通过分析遥感技术在中国农作物种植面积估算中所遇到的难点, 针对运行化的农作物遥感估产系统对主要农作物种植面积估算的需求, 提出在农作物种植结构区划的基础上, 采用整群抽样和样条采样技术相结合的方法, 进行农作物种植面积估算。整群抽样技术利用遥感影像估算农作物总种植成数, 样条采样是一种适合中国农作物种植结构特征的采样技术, 用于调查不同农作物类别在所有播种作物中的分类成数。在中国现有的耕地数据库基础上, 根据两次抽样获得的成数, 计算得到具体某一种农作物类别的种植面积。最后给出了2003年早稻种植面积估算的实例。

关键词: 种植成数; 分类成数; 整群采样; 样条采样; 遥感; 种植面积

中图分类号: TP79 **文献标识码:** A

1 引言

采用遥感技术进行农作物产量估算, 历来受到各国政府的重视, 发展到今已经取得了很大成功。如美国是一个农产品出口大国, 为了在世界粮食市场占据主动地位, 应用遥感技术不仅对自己国内的农作物估产, 还专门在农业部设立机构监测全球主要产粮区和粮食消费国的粮食产量, 取得了巨大成绩^[1,2]。欧共体、俄罗斯、法国、日本和印度等国为了本国的利益也都应用遥感技术进行农作物长势监测和产量测算, 均取得了不同程度的效益^[1]。中国从1983—1998年的15年间, 通过国家科技攻关计划和一些部门科研项目研究, 监测作物种类从冬小麦单一作物估产发展到小麦、水稻和玉米等多种农作物遥感估产, 监测范围从小区域到横跨11省市的遥感估产, 在农作物遥感估产的各方面取得了长足的进步^[3]。

在农作物遥感估产中, 农作物种植面积的遥感估算是作物产量预测的基础和主要内容, 长期以来, 一直是大面积农作物遥感估产的重要技术研究课

题^[4-6]。

开展全国范围的农作物面积遥感估算必须解决3个问题: (1) 精度问题。一些方法在小范围内进行实验可达到理想的高精度, 但在大范围应用时精度难以保证。(2) 速度与效率问题。全国范围多种农作物的耕种与收获季节集中在4—10月, 时间十分紧凑。为达到预报的目的, 估产任务时间很紧。估算方法必须满足速度快、效率高的要求, 一些适合在小区域应用的精确处理方法不适应大范围内运行的要求。(3) 成本问题。一些在小范围试验的成功方法推广到全国时因成本太高而不实用。因此, 全国性大范围的农作物面积估算技术, 必须同时满足4个条件: (1) 准确度高能够满足决策的需要; (2) 速度快能够满足应用部门的要求; (3) 经费省使应用部门足以支付得起; (4) 有统一规范且可操作性强等特点^[7]。

中国农情遥感速报系统在1:10万全国资源环境数据库中的耕地层数据的支持下, 将抽样技术与遥感估算相结合, 通过两次抽样, 提出适合中国特点的农作物种植面积遥感估算技术体系, 实现了对全国农作物种植面积和种植结构的快速准确的估算。

收稿日期: 2003-09-30; 修订日期: 2003-12-12

基金项目: 中国科学院95重大项目(KZ951-A1-302-02)和特别支持项目(KZ95F-03-02); 中国科学院知识创新重要方向项目(KZCX2-313); 科技部十五攻关项目(2001BA513B02); 中国科学院遥感所领域前沿项目: 样条采样框架的理论基础与精度检验方法研究。

作者简介: 吴炳方(1962—), 男, 工学博士, 研究员, 博士生导师, 江西省玉山人, 从事农业与生态环境遥感研究, 发表论文40余篇。E-mail: wubf@isa.ac.cn

2 农作物播种面积估算技术思路

采用遥感技术进行大面积农作物种植面积运行化监测,直接采用遥感信息源进行种植面积估算,面临着 3 个难题^[8]:

(1) 遥感数据无法在作物生长季内实现全国范围内的全覆盖。不但数据费用昂贵,难以负担,而且即便是能够保证数据费用支出,也无法在短期内处理完毕,因此,全覆盖方式不能满足遥感估产的运行性需求。

(2) 由于天气因素的影响,使得无法获取能够满足遥感估产需要的具有合适时相的高分辨率遥感数据。中国南方地区,多阴雨天气,作物生长期内的遥感数据获取率很低。影响面积提取精度在时间上取决于合适的时相,南方水稻的最佳时相是移栽后的 2—5 周;北方冬小麦的最佳时相是每年 4 月;东北玉米的最佳时相是 8 月中旬和 9 月上旬;东北大豆的最佳时相是 8 月上旬。这些时相要求很难得到满足。

(3) 在中国现行土地使用制度下,土地使用权力分散,农业生产地块破碎,插花种植现象明显,间作套作普遍,如南方地区的早、中、晚稻,东北地区的

春小麦、春玉米和大豆,北方地区的夏玉米、棉花等交错种植,因此,从遥感监测作物种植面积的能力方面分析,即使利用合适时相的 Landsat TM 或 CBERS CCD 等光学遥感数据,甚至使用 QuickBird 数据也很难高精度地直接提取某一种农作物的种植面积,如在开封地区的实验表明,对于夏收作物和秋收作物的识别,其正确率不足 90%^[9]。图 1 是山西太谷地区的 QuickBird 影像窗口,在一个 TM 像元范围内,多种作物类型插花严重,据调查,周边 5km×5km 的范围,农作物类型达 50 多种。在某些地区,利用合适时相的影像可以实现对某些作物种植面积的高精度估算,但往往局限于某些特定地区和某些特殊作物,如用 Landsat TM 的差值植被指数进行冬小麦面积监测,局部精度可达 90% 以上^[4],这种方法取得较高精度的主要原因是北方冬小麦种植具有空间上的一致性,连片种植,且在返青早期没有其他绿色植被,对于南方地区,冬小麦种植成数偏低,且与油菜插花现象比较严重,该方法就极不理想。同时差值分割法中阈值的选取直接制约着该方法的大范围推广,因为不同地区、不同时相的阈值差异也较大,且不同年份间的阈值也有所不同。同样道理,如果进行北方的油菜、春小麦、春玉米或南方土豆、玉米、冬小麦等低成数作物的种植面积监测,将面临同样困难。

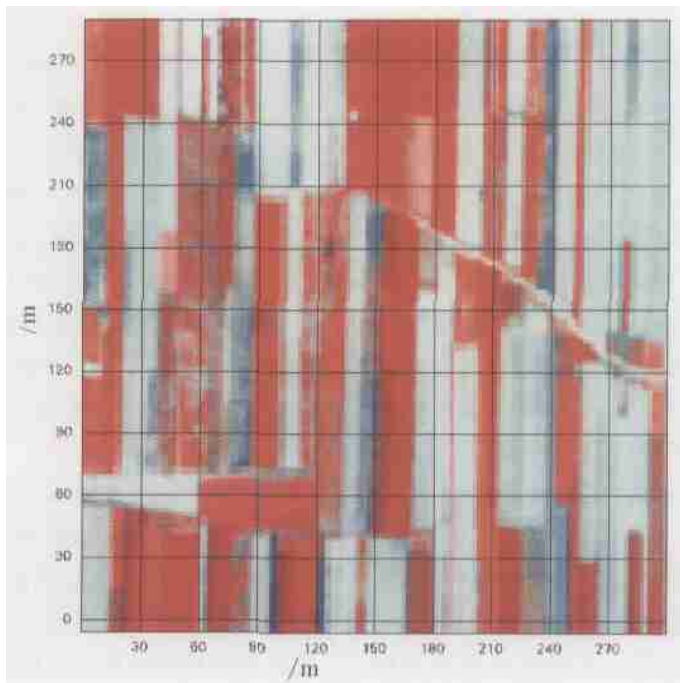


图 1 山西太谷地区 QuickBird 影像
(时相: 2003-07-15, 波段组合: 4, 3, 2)

Fig. 1 QuickBird imagery of Taigu region in Shanxi province
(Acquisite date: Jul. 15, 2002, Band combination: 4, 3, 2)

因此, 实现全国运行化分类型作物种植面积监测, 必须采取遥感技术与抽样技术相结合的方式, 这不但解决了全覆盖方式的影像获取困难与资金节约, 而且可以用抽样技术进行质量控制与误差分析。抽样技术在全球性大型农业调查项目中得到广泛使

用, 如欧盟 MARS 计划采用了分层面积抽样方法^[10]、美国大面积农作物估产计划(LACIE) 以及农业和资源空间遥感调查计划(AGRISTARS) 等都使用了面积抽样框架^[1]。中国农情遥感速报系统采用了两个独立的分层抽样方法(图 2)。

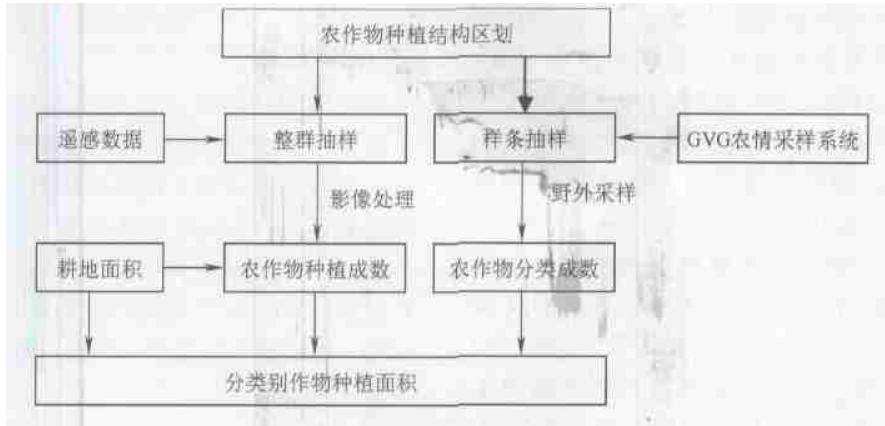


图 2 农作物种植面积估算技术路线

Fig. 2 Flowchart of crop acreage estimation

首先, 把全国分成种植面积相对稳定、种植结构相对一致、自然条件相近的区域构成抽样层, 即农作物种植结构区划。区划共分 3 级, 第一级全国共包括 11 个区划单元, 第二级区划包括 44 个二级单元, 第三级区划包括 102 个三级单元。然后, 在区划的基础上, 利用两个抽样估算作物种植面积。第一个抽样以二级区划单元为基础, 以 1: 10 万地形图国际标准分幅为抽样框, 利用遥感数据进行分层整群抽样。利用遥感能够高精度提取作物总的种植面积的特点, 在每个层内抽取能够满足估算精度要求的一定数量的面积样方, 通过影像处理与分类解译提取作物总种植面积, 并借助土地资源数据库估算出每个层的农作物总种植成数(即所有作物的种植面积占耕地面积的比例)。第二个抽样以三级区划单元为基础, 采用样条采样框架技术, 借助 GVC 农情采样系统, 通过野外调查, 估算区划单元内每种作物类型的分类成数(即某类农作物种植面积占所有作物种植面积的比例)。最后, 用作物总种植成数与每种作物类别的分类成数相乘, 再乘以各区划单元的耕地面积就得到每种作物的面积。

采用该方法体系, 一方面是因为我国已经建成了 1: 10 万比例尺的土地资源数据库, 可以将耕地层提取出来作为作物种植面积估算的数据基础, 另一方面, 虽然利用 Landsat TM 数据难以精确地提取作物种类的种植面积, 但是却可以精确地提取作物总的种植面积。

另外, 在进行水稻种植面积估算时, 同时采用了雷达数据和光学遥感数据(TM, Cbers CCD 等), 由于通过两个时相雷达数据的后向散射系数差异, 可以高精度地将水稻区分出来^[11], 满足运行化要求, 而光学数据的提取精度相对较低, 因此, 水稻种植面积估算是混合性的方法, 对于有雷达数据的区划单元采用面积抽样框架方法, 对于有光学数据的区划单元采用两个分层抽样方法。

3 农作物种植结构区划分层

由于中国农业自然条件的地区差异, 形成了不同地区不同的作物种植制度, 加上长期的农业开发历史, 最终导致形成了耕地地块小而分散的农作物种植格局, 其主要特点是: (1) 除东北、华北、长江中下游等小部分地区外, 大多为丘陵与山地, 地势起伏明显, 水田旱田插花严重, 导致农作物种植插花现象明显; (2) 作物种植间作套作现象严重。中国现行的农业土地使用制度下, 农业用地耕作由农民自由支配, 插花种植现象普遍, 即使自然条件相同的同一田块内也存在复杂的农作物插花种植, 并有套作种植现象; (3) 农作物种植成数偏低。中国农业发展很早, 农业用地的开发利用程度也较高, 但是由于地形崎岖以及过量人口活动, 导致中国的耕作面积所占整个土地总面积的比例并不高, 农作物种植成数处于较低状态。

农作物种植结构区划是提高农作物遥感监测精度的有效途径之一。区划工作之前将全国划分为农业区和非农业区,然后在农业区进行。同时,为了保持与统计信息获取与发布的一致性,区划时保持县级行政单元的完整性。

农作物种植结构区划共分 3 级,一级区划直接采用中国农作物种植制度气象区划的第一级,主要考虑温度、降水、太阳辐射、土壤、地貌和作物类型等因素,将全国共划分为 11 个一级区划单元^[12]。其分区指标见表 1。二级区划采用农作物种植结构信息作为主导因子,包括小麦、水稻、玉米和大豆 4 种作物的播种面积比例以及农业集约化程度等(表 1)。通过多指标聚类形成 44 个单元(表 2)^[13]。每个区划单元内部,种植结构相似,种植制度基本一致。如东南丘陵平原的水稻一元结构区、黄淮海平原丘陵的小麦玉米二元

表 1 农作物种植结构区划因子表

Table 1 Factors considered in crop proportion zoning

区划等级	分区指标
一级区	温度、降水、太阳辐射、土壤、地貌
二级区	小麦播种面积比例、水稻播种面积比例、玉米播种面积比例、大豆播种面积比例、农业集约化程度

结构区、东北平原丘陵的大豆小麦二元结构区兼玉米区等。三级区划在农业区划中考虑垦殖水平,参考全国 1: 10 万比例尺的耕地数据库将全国划分为 5 种类型区,反映不同垦殖水平(耕地密度 > 80%, 50%—80%, 15%—50%, 0%—15%, 0),与二级区划叠置。同时为了计算方便,将区划层与省界叠加,使每个区划单元只属于一个省。整理后形成 102 个单元(图 3)。

表 2 中国农作物种植结构区划

Table 2 Zones used for crop proportion survey

一级区划	二级区划	一级区划	二级区划
青藏高原喜凉作物区(I)	小麦为主,无稻区(I 1)	黄淮海平原丘陵半湿润旱作物为主区(VI)	小麦为主,兼玉米大豆区(VI5)
	小麦一元结构,无稻区(I 2)		玉米为主,兼水稻大豆区(VI6)
	经济作物为主,兼粮食作物(I 3)		玉米为主,兼小麦大豆水稻等构区(VII)
内蒙古东南部黄土高原西部半干旱喜凉作物区(II)	小麦一元结构区(II 1)	西南东部高原山地湿润水旱兼作区(VII)	水稻为主,兼小麦水稻等构区(VII2)
	春麦为主区(II 2)		小麦玉米为主,兼水稻区(VII3)
	春麦玉米大豆三元结构区(II 3)		水稻为主,兼玉米区(VII4)
内蒙古陕晋高原山地易旱喜温作物区(III)	小麦为主,兼玉米区(III1)	长江中下游平原丘陵湿润水田为主区(VIII)	小麦玉米二元结构,兼大豆区(VI5)
	大豆玉米小麦三元结构区(III2)		水稻一元结构,兼大豆区(VIII1)
	玉米为主,兼大豆区(III3)		水稻一元结构区(VIII2)
	小麦玉米二元结构区(III4)		水稻为主,兼小麦区(VIII3)
东北平原丘陵半湿润湿润喜凉作物区(IV)	玉米为主,兼水稻大豆区(IV1)	四川盆地平原丘陵山地湿润水旱兼作区(IX)	水稻小麦二元结构区(VIII4)
	大豆小麦二元结构,兼玉米区(IV2)		四元结构区(VI5)
	四元结构区(IV3)		水稻为主,小麦玉米大豆等构区(IX1)
	小麦一元结构区(IV4)		小麦水稻玉米三元结构区(IX2)
新疆河西走廊及河套干旱灌溉区(V)	小麦为主,兼玉米大豆区(V 1)	东南丘陵山地湿润双季单季水稻兼作区(X)	水稻为主,兼小麦玉米区(IX3)
	四元结构区(V 2)		水稻一元结构区(X 1)
	小麦为主,兼玉米区(V 3)		水稻玉米二元结构,兼大豆区(X 2)
黄淮海平原丘陵半湿润旱作物为主区(VI)	小麦一元结构区(V 4)	东南丘陵平原湿润双季稻热作区(XI)	水稻一元结构,兼大豆区(X 3)
	小麦玉米为主,兼大豆区(VI1)		水稻为主,兼小麦玉米区(X 4)
	水稻为主,兼玉米区(VI2)		水稻一元结构区(XI1)
	小麦玉米二元结构区(VI4)		小麦为主区(XI2)
			水稻为主,兼玉米大豆区(XI3)

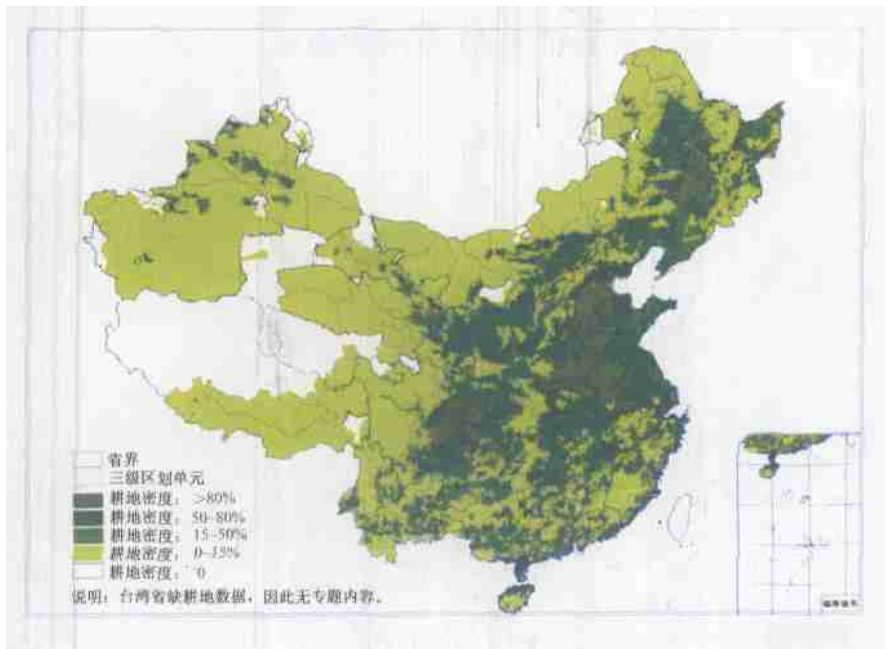


图3 中国农作物种植结构区划

Fig. 3 Zones used for crop proportion estimation

4 农作物种植成数的遥感估算

在中国现行的土地使用制度下, 寄希望于利用遥感数据直接提取各种作物的种植面积是不现实的, 然而遥感图像(Landsat TM, CBERS CCD)却能较好地反映出植被信息, 因此可以精确地提取耕地上所有作物的总种植面积, 而不管是什么类型的作物, 非常容易提取, 效率高, 且能在很短的时间内完成大量的数据处理工作, 能满足运行要求, 这是农作物种植成数遥感估算的基础。

估算全国范围内农作物总的种植成数, 采用遥感技术与抽样技术相结合的方法。考虑到遥感数据的特点, 采用整群抽样技术估算农作物种植成数。

4.1 整群抽样设计

整群抽样技术首先将总体基本单元划分为若干的群, 从群中抽样一部分样本群, 然后对样本群中的所有基本单元进行统计分析。抽样误差主要取决于群间方差的大小。群与群之间的差异越小, 抽样效率就越高。中国农情遥感速报系统通过整群抽样, 抽取满足精度要求的抽样群, 以选中的抽样群的范围订购合适时相的遥感数据, 用来监测和估算作物种植成数。

4.1.1 抽样框设计

中国农情遥感速报系统采用农作物种植结构区划的二级单元作为抽样总体, 在每个单元内划分抽样群, 然后按精度要求抽取一定数量的群。

整群抽样过程中, 以耕地区的1:10万地形图国际标准分幅作为抽样群, 每个抽样群的面积在1270—1950km²之间, 相当于一景Landsat TM影像的1/16(平均面积约等于1977km²)。参照1:10万比例尺土地利用图剔除没有耕地的群, 以所有包含耕地的标准分幅格网作为整群抽样框(图4), 全国整群抽样的总群数 N 为3095(没有包括西藏和台湾), 每个区划单元内的群数见表3。然后通过空间位置建立抽样群与遥感影像之间的对应关系, 进行遥感影像的选择。

4.1.2 样本量

整群抽样的样本量需要根据群内方差和估计误差限计算, 或者根据群设计效应进行计算, 但是在全国农作物种植成数的估计时, 由于作物种植结构调整等各方面原因, 每年作物种植成数变化较大, 群内方差计算较难, 因此, 对于农作物种植成数的抽样估计采用简单随机抽样的样本量计算公式, 同时考虑到整群抽样的样本量要大于简单随机抽样, 因此对根据简单随机抽样样本量计算出来的样本量增加了5%的修正量, 以期满足整群抽样的要求:

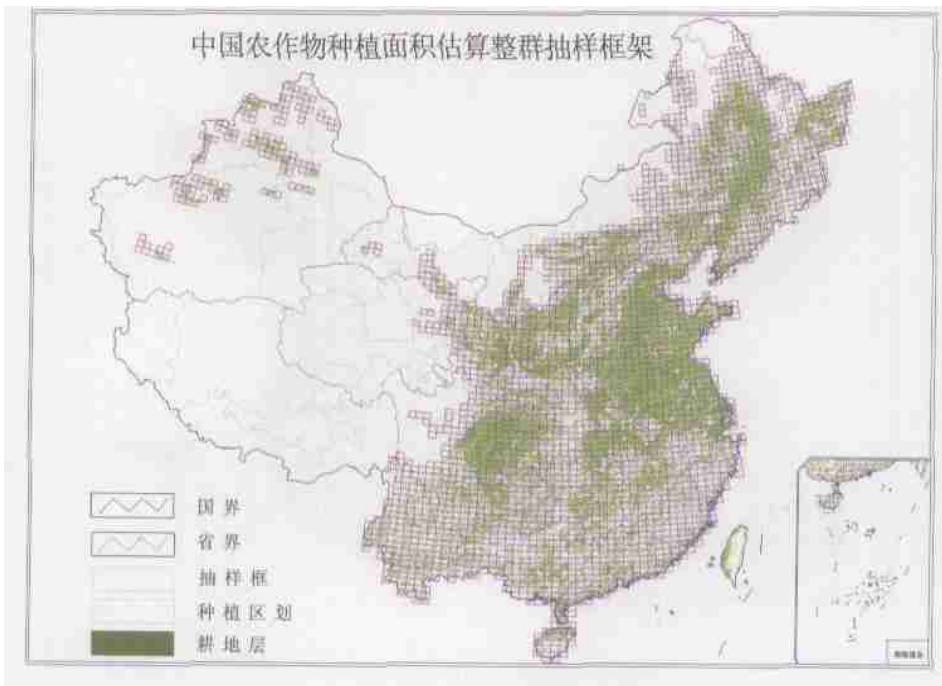


图 4 全国农作物种植面积估算整群抽样框架

Fig. 4 Sampling frame used for the estimation of the crop proportion

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}} \quad (1)$$

$$n_0 = \frac{t_{\alpha p}^2 (1 - p)}{d^2} \quad (2)$$

式中, n 为最终的样本量, n_0 为先期计算的样本量, N 为样本总体, d 为预期的计算误差, t 为正态离

差, p 为作物种植成数。根据估产要求, 估算精度应达 95% 以上, 则 d 取 0.05; 在 95% 置信度下的 t_{α} 为 1.96, 结合前期的预抽样分析结果, 确定 p 取保守值 0.75, 依此求得 $n = 264$, 增加 5% 的修正量后, 抽样群数 $n_{\text{总}} = 264 \times (1 + 5\%) = 278$ 。最终求得各区划单元的样本量(表 3)。最后全国的平均抽样率 $f \approx 9\%$ 。

表 3 整群抽样样本总体及样本量

Table 3 Population and sampling size in cluster sampling

分层				层群数				分层				层群数			
一级	二级	总群	样群	一级	二级	总群	样群	一级	二级	总群	样群	一级	二级	总群	样群
I	I 1	7	1	V	V 1	44	4	VIII	VII2	158	14	IX	IX1	28	3
	I 2	3	1		V 2	14	1		VII3	64	6		IX2	46	4
	I 3	43	4		V 3	64	6		VII4	77	7		IX3	41	4
II	II 1	100	9		V 4	100	9		VII5	14	1		X	X 1	85
	II 2	49	4	VI	VI 1	166	15		IX	IX1	28			3	X 2
	II 3	87	8		VI 2	5	1	IX2	46	4	X 3	68		6	
III	III 1	61	5		VI 4	73	7	XI	XI 1	202	18	XI	XI 1	202	18
	III 2	69	6	VI 5	28	3	XI 2		8	1	XI 2		8	1	
	III 3	166	15	VI 6	84	8	X 2		64	6	XI 3		19	2	
	III 4	40	4	VII	VII 1	79	7		X 3	68	6	全国总计	44	3095	282
IV	IV 1	203	18		VI 2	78	7	X 4	51	5	全国总计		44	3095	282
	IV 2	133	12		VI 3	154	14	XI	XI 1	202			18	全国总计	44
	IV 3	131	12	VI 4	38	3	XI 2		8	1		全国总计	44		3095
	IV 4	36	3	VI 5	35	3	XI 3		19	2	全国总计		44		3095
	IV 5	23	2	VIII	VIII 1	57	5	全国总计	44	3095			282		

4.1.3 农作物种植成数的抽样估计

由于组成抽样群的 1:10 万地形图幅在不同纬度下存在面积差异, 中国农情遥感速报系统采用不等群随机抽样方式, 抽取满足精度估算要求数量的样本群, 用于订购遥感数据, 经过图像处理与分类进行种植面积提取, 最后采用整群抽样的比例计算方法来估算作物播种面积占耕地面积的百分比, 即种植成数。

种植成数

$$p = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n m_i} \quad (3)$$

其中, p 为种植成数, 即有作物耕地像元和耕地总像元数之比; n 为区划单元内整群抽样抽取的群数, 即欲订购的影像数; a_i 为影像中作物像元数总和; m_i 为影像对应区域中耕地像元数总和, 即抽样群中的耕地像元数。

估计方差

$$v(p) = \frac{1-f}{nm^2} \cdot \frac{\sum_{i=1}^n (a_i - pm_i)^2}{n-1} \quad (4)$$

式中 $m = \frac{1}{n} \sum_{i=1}^n m_i$ 为每景影像中平均耕地像元数总和, $f = n/N$ 为抽样率, 即抽取群数 n 占总群数 N 的比例。

估计精度

$$A(p) = 1 - t_\alpha \cdot \sqrt{v(p)}/p \quad (5)$$

式中 t_α 为一定置信度下标准正态分布的双侧 α 分位数。

4.2 遥感数据选取及时相要求

中国农情遥感速报系统的作物种植面积监测目标包括 4 大类 7 种作物的面积估算, 即早稻、中稻(一季稻)、晚稻、冬小麦、春小麦、玉米和大豆。根据其物候特征和空间分布, 欲使监测精度有所保证, 需要明确遥感数据类型及合适的时相。

1998 年以前, 利用遥感数据估计作物的种植面积主要采用 MSS, TM 和 NOAA AVHRR 数据, 也有采用 TM 快视图像的^[14]。中国农情遥感速报系统主要依靠 Landsat TM 数据, 也使用少量的 Cbers CCD 数据。同时, 由于监测业务运行化对时相的要求较高, 而可见光遥感数据的获取由于受到天气因素的影响, 很难在短时期内获取满足要求的数据, 尤其是对南方水稻面积提取时, 情况更为严重。为此在 Landsat TM 等光学遥感数据获取率比较低的地区, 选用全天候、全天时、不受阴雨天气影响的雷达数据作为信息源提取水稻播种面积。

在确定影像的时相选择时, 主要考虑主要农作物的空间布局、生育期、轮作规律和田块分布特征以及同类作物的时间跨度, 中国农作物种植除冬小麦生长期内作物类型较单一外, 其他作物大多混种、插花, 如南方地区的早、中、晚稻, 东北地区的春小麦、春玉米和大豆; 华北地区的夏玉米、花生、棉花等交错种植。随着种植结构的调整力度加大, 作物熟制减少, 使得作物种植时间的灵活性加大, 因此在选择遥感图像时, 影像的时相既要能保证能反映作物的空间分布, 又能满足运行的时间要求。图 5 是不同类型作物的时相要求和监测结果发布时间。

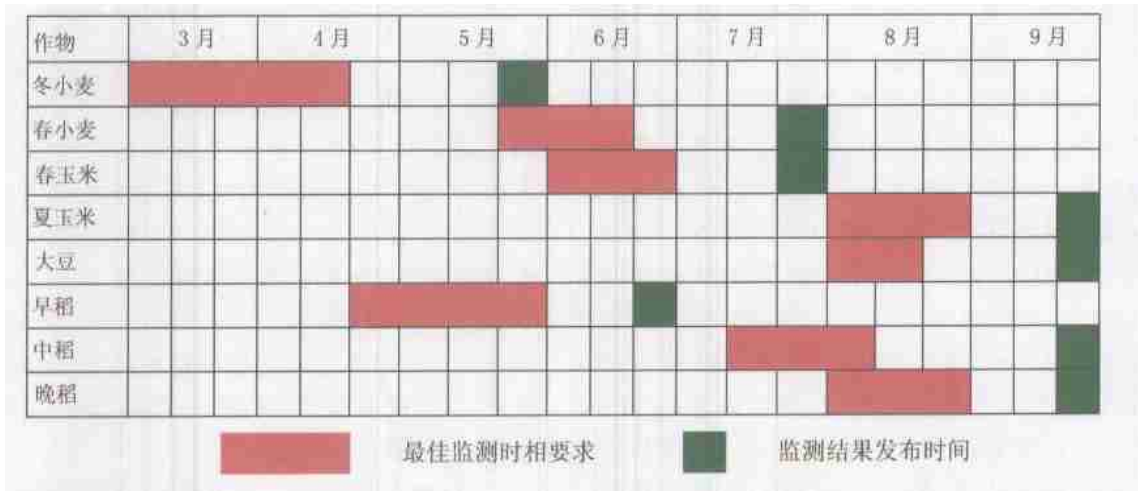


图 5 农作物遥感监测最佳时相及最佳监测结果发布时间

Fig. 5 Optimal date for image acquisition and results release for crop acreage monitoring

4.3 遥感影像处理与分类

根据农情遥感监测系统快速、高效、高精度的运

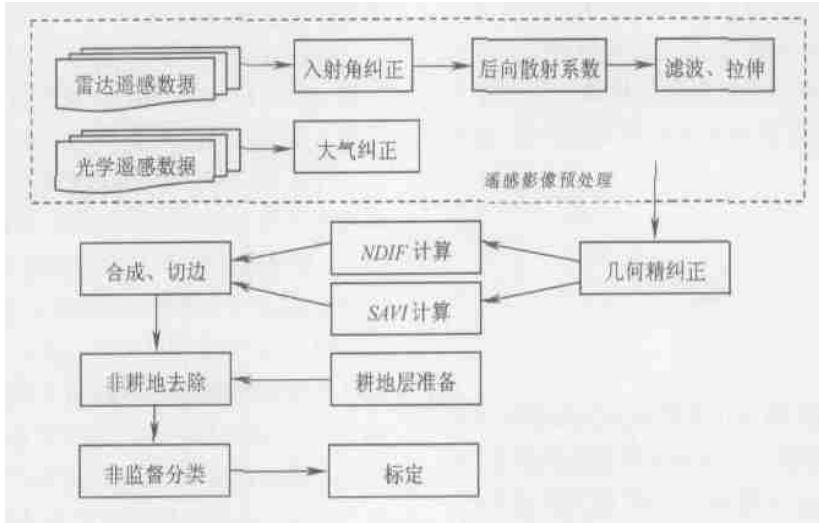


图 6 农作物种植成数监测中的遥感影像处理流程

Fig. 6 Image processing flow in the crop proportion monitoring

遥感影像两种不同的处理方式。光学遥感影像的处理主要是作大气纠正; 雷达遥感影像处理比较复杂, 包括入射角纠正、后向散射系数计算、滤波、拉伸和 multispectral 合成等。其中, 入射角纠正是为了校正不同扫描入射角所造成的差异。雷达影像的滤波是为了消除斑点噪声的影响, 同时雷达影像因地形起伏以及地物类型的差异, 主要依据滤波效果和信息量的损失大小来选用滤波算子, 多数情况采用 Lee 滤波算法, 窗口大小采用 3×3 的滑动窗口。另外, 用线性拉伸的方法将后向散射系数进行拉伸处理。

(2) 遥感影像几何精纠正。采用二阶多项式进行遥感影像的几何精校正。首先对照地形图或参考影像建立地面控制点集, 最后采用最邻近距离法重采样生成几何精校正后影像。为了保证几何精纠正的质量, 建立了标准的参考影像库和控制点集。在没有参考影像的情况下, 用地形图作为参考进行纠正, 并保证控制点的空间分布接近均匀分布。对于 Landsat TM 等光学遥感数据, 每个地面控制点 RMS 误差与 RMS 累积误差均要求不超过一个像元。雷达数据可以适当放松控制点精度, 每个地面控制点 RMS 误差与 RMS 累积误差允许放宽到两个像元。同时, 从控制点个数上进行精度约束, 要求每帧 CBERS CCD 影像或全景的 TM 影像, 要求其控制点个数不低于 20 个, 对于 Radarsat SCAN SAR 和 Landsat TM 整景影

像要求控制点个数不低于 25 个。

(1) 遥感影像预处理。分光学遥感影像和雷达

像要求控制点个数不低于 25 个。

(3) 植被指数计算。植被指数是反映作物播种信息的重要指标, 植被指数参与分类和解译可以明显提高精度。其中光学遥感影像的植被指数采用 SAVI 公式:

$$SAVI = (NIR - Red) / (NIR + Red + 0.5) \times 100 \quad (4)$$

对于雷达影像, 根据水稻的后向散射系数在二期雷达影像上的较大差异, 采用两期影像的后向散射系数进行归一化计算:

$$NDIF = (BS2 - BS1) / (BS2 + BS1) \times 100 \quad (5)$$

其中, $BS1$ 为第一时相的后向散射系数, $BS2$ 为第二时相的后向散射系数。

(4) 影像合成。对于光学影像, 将计算求得的 SAVI 数据层叠加到多光谱数据之后, 形成包括多光谱数据与 SAVI 信息的合成数据集。对于雷达影像, 用二期雷达影像计算求得的后向散射系数以及 NDIF 数据层进行叠加合成, 形成包括两期后向散射系数与 NDIF 的合成数据。同时, 切除不同波段或不同期影像之间的数据交错边缘。

(5) 非耕地去除。从 1:10 万土地资源数据中提取出遥感图像对应区域的耕地数据, 去除非耕地像元, 保留耕地范围内的像元, 使耕地内的作物信息突出, 有利于后期分类。

(6) 非监督分类。采用 ISO DATA 算法, 对去除非耕地后的图像进行非监督分类, 生成分类影像。为保证分类质量, 每帧影像的分类类别不低于 20

类,并要求最大迭代次数不低于15次。

(7) 标定。标定是非监督分类工作的后期处理,其目的是从非监督分类所生成的各个类中提取出农作物种植区,这个过程因数据的不同而有差异。由于雷达数据与光学遥感数据不同的早稻监测机理,因此解译的方法也不尽相同。雷达数据通过两期对比,对南方水稻具有精确的识别能力,因此可以直接解译识别出具体的作物种植区,因此其解译对象是水稻种植区。而对于光学遥感数据,因为不同作物类别之间的光谱差异不太明显,标定工作的目标是作物种植区。在解译过程中,综合运用各个分类类别的统计量数据、空间分布情况、各种先验知识(主要是物候历、长势监测成果、气象信息)以用地面调查信息(如有关种植结构调整的政策等),解译出各个类别是否种植作物。标定工作由经验丰富并对影像区域的农作物种植制度比较熟悉的工作人员来完成,并实行专人负责制。区分有农作物覆盖区和无覆盖区,除了工作人员的专业知识外,还必须参考不同作物的物候历及实时的气象信息,并通过相关资料查询有关农作物种植变化的情况。另外,从历史档案中调出遥感数据进行对比分析,也有助于精确识别。为了保证标定的质量,采用了3类自校验方法。根据遥感数据的时相和目标作物的类型,设定SAVI和NDIF的阈值范围,超出阈值范围的像元定为非作物像元,阈值的设定随空间和时间变化;根据初步标定结果,进行分县或分区统计,将统计值与耕地面积和统计数据中的种植面积相比较,当相差大于某一特定阈值时,表明误差太大,需要重新标定;在标定时,大部分类别都能得到准确的识别,少量的几个类别可能是造成误差的主要因素,需要地面经验或电话咨询当地人员。

4.4 种植成数估算

遥感数据经过纠正、分类、标定等预处理过程后,使用之前根据需要计算区划单元和省级行政单元的农作物种植成数。各个区划单元的农作物种植成数由区划单元内抽取的各帧遥感影像的处理情况,根据公式(3)计算求得,即各帧影像中的总种植面积除以总耕地面积。省级行政单元的农作物种植成数由省内各个区划单元的农作物种植成数按耕地面积加权平均计算。

5 农作物分类成数调查

农作物分类成数监测最直接的方法是使用Quickbird或IKONOS等高分辨率遥感影像,考虑到高分辨率遥感影像的价格因素,采用抽样方式是最合适的,但是从全国性作物分类成数监测的运行性来看,如果采用QuickBird或IKONOS等影像进行抽样,由于QuickBird等高空间分辨率影像的获得性较差,往往无法满足抽样要求的影像数据,而且成本太高。因此,最实用的方法是采用野外调查的方式,既可以保证精度,也可以控制成本。

中国农情遥感速报系统在对现有的农情抽样技术分析基础上,结合中国农业种植结构特点,提出了农作物分类成数的样条采样框架方案。即在成数抽样理论的指导下,在中国农作物种植结构区划的基础上,在每个区划单元内设计用于调查农作物分类成数的调查线路,然后安排采样调查队伍到沿各个采样线进行采样,最后统计汇总形成各个区划单元的农作物分类成数。

5.1 样条采样框架

目前,国内外用于农作物分类成数调查的抽样技术主要是面积抽样技术,这些面积抽样技术可以总结为两个范式,即TPS(Tessellated plane sampling)范式和CPS(Continuous plane sampling)范式。TPS范式中,调查区域被分割为连续的等面积的抽样块(block),由它们组成抽样框架。这些抽样块可以由卫星影像的像素组成。由此范式得出的推论,是根据面积框架中有限数量的调查估计得出的。欧盟的MARS计划、美国的森林调查与分析(Forest Inventory and Analysis, FIA)项目和自然资源调查项目(Natural Resource Inventory)都是在此范式的指导之下建立起来的抽样计划^[15]。CPS范式则采用的是随机放置抽样地块(plot)的方式,根据这种范式进行抽样不需要建立面积抽样框架,而是直接在抽样空间中随机布样,有文献证明,这种范式比TPS范式有更可靠的估计能力^[16]。CPS范式的介绍与应用例子也很多^[17-22]。TPS范式的最大潜在优点是它比CPS可用更多的层,从而可以更大限度地降低估计误差^[23],而TPS范式的不足主要是面积框架的质量问题,即当具有周期性抽样目标的空间分布周期与面积框架的格网密度相似时,可能会造成严重的偏差。CPS范式要求各个层的界限清晰,野外容易找到。

而且它的估计在样本较少时可能会导致估计偏差,因此要求较多的样本量。

中国学者也认识到了遥感技术的局限性,很早就将抽样技术与遥感技术相结合,进行农作物种植面积、耕地面积变化等监测,有局部实验,也有全国性实验。王延颐等利用 TM 影像进行成数抽样对江苏省兴化县的水稻种植面积进行监测^[24]。周华茂则借鉴 MARS 计划的面积抽样框架在成都平原进行了水稻播种面积遥感调查实验^[25]。陈仲新等用分层抽样方法实现了全国冬小麦面积变化遥感监测的外推^[26]。刘海启用分层随机抽样方法在全国选取了 8 个调查区,进行大尺度的耕地变化监测^[7]。焦险峰等使用 1:2.5 万比例尺地形图标准幅建立抽样框架,完成了新疆维吾尔自治区的棉花种植面积监测^[27]。

从抽样的角度看,中国农作物分类成数的调查属于低成数、超大范围、高空间异质性的空间抽样问题。采用 CPS 范式对于样本量的要求比较高,否则可能会造成较大的偏差,而且由于 CPS 自身的特点,在抽样分层数量比较多时,工作量巨大,因此可行性不高,不适合分类成数的调查。TPS 范式,则要求大量的野外样地测量工作,不可能在短时间内完成野外调查工作,无法满足运行化监测要求,此外,面积抽样框架主要借助遥感影像建立,容易导致样本数量不足或样本在空间上过于集中等现象。

样条抽样框架是典型的分层二阶段抽样。它以农作物种植结构区划的三级区划为基础,进行二阶段抽样。在第一阶段的抽样中,继承 TPS 范式的面积抽样框架,完成初级抽样单元的布设,面积抽样框架采用耕地区范围内的 4km × 4km 格网。

在第二阶段的抽样中,为了提高抽样的执行效率,节省抽样时间,在一级抽样选取的样本格网中,以道路为基础,按一定抽样率抽取满足精度要求的路段(样线),以样线另一边的缓冲区(目视可辨别作物类别的有效范围)作为样区(即样条),进行农作物分类成数调查。

5.2 基于面积抽样框架的第一阶段抽样

第一阶段抽样采用面积抽样框架技术,其中最基础的工作是面积抽样框架的建设。农业调查中面积抽样框架中样方的大小因项目不同而不同。如欧盟 MARS 计划采用 10km × 10km 面积

抽样框架^[28-31]、美国 NASS JES 项目调查的 1mi × 1mi (1mi = 1.61km) 面积抽样框架^[32],法国 TERUTI 项目调查 6km × 6km 面积抽样框架^[33],FAO 采用多框架抽样方法^[34,35]。我们结合中国农作物种植结构的特点和遥感估产系统运行化的需要,在第一阶段抽样的抽样框架采用 4km × 4km 格网。去除无耕地区域,全国样本总数为 196170 个(表 4,图 7)。

第一阶段抽样的样本量计算采用简单随机抽样的样本量计算公式,在给定置信水平与允许误差的情况下,影响抽样样本量大小的因子主要是作物成数。但是,中国农作物种植成数因地因时差异巨大。华北地区夏收作物简单,主要为冬小麦,还有小部分油菜等,而秋收作物种植结构复杂,包括玉米、水稻、大豆、花生、棉花等。南方各省区则夏收作物相对复杂,包括冬小麦、油菜、红花草等,夏季则主要是水稻和薯类,其他作物很少。西北地区与华北具有相似性,东北地区只有一季作物,但结构相当复杂,包括玉米、大豆、水稻、土豆等。由于样条抽样框架的采样目标是获取不同作物的分类成数,考虑到不同地区不同季节的作物成数的较大变化,因此,在随机抽样样本量计算中其比率取为 0.5,以满足作物分类成数的复杂性特点,同时也能保证足够的样本。样本的选择一般有简单随机取样和系统取样两种方法,从使用角度看,这两种方法都非常容易,而且可以使样本在空间上呈分散分布,足以反映抽样变量的空间变化特征。但是,Womer 研究证明^[36],当抽样变量在空间上呈随机分布时,两种抽样方法基本上可以取得比较好的效果。然而,当抽样对象在空间呈镶嵌分布格局的情况下,随机抽样可以达到非常接近理论概率的程度,而系统布样就要差得多。由于中国农作物种植结构从南到北、从东到西差异明显,距离随机分布的程度还很远,因此,在第一阶段抽样中采用简单随机抽样方式。

5.3 样条抽样框布设

在样条抽样框布设过程中,由于样条抽样框架以初级抽样单元内公路一侧的缓冲区作为样本。假定第一阶段抽样格网大小为 $K \text{ km} \times K \text{ km}$,则抽样率可表达成:

$$n = \frac{LD}{K^2} \times 100\% \quad (6)$$

表 4 中国农作物分类成数调查第一阶段抽样样本量

Table 4 Sample size used in crop type proportion survey

一级	二级	抽样域	无耕地区	初级抽样总体	初级抽样样本量	初级抽样率/ %	样条长度/ km
I	I 1	8970	8052	918	46	5	147.2
	I 2	12788	12625	163	8	5	25.6
	I 3	58225	55065	3160	96	3	307.2
II	II 1	13707	7533	6174	96	2	307.2
	II 2	5171	1516	3655	96	3	307.2
	II 3	9198	3113	6085	96	2	307.2
III	III1	7663	3321	4342	96	2	307.2
	III2	6743	932	5811	96	2	307.2
	III3	17460	5676	11784	96	1	307.2
	III4	3512	422	3090	96	3	307.2
IV	IV1	17497	1482	16015	96	1	307.2
	IV2	12490	5487	7003	96	1	307.2
	IV3	12838	5160	7678	96	1	307.2
	IV4	9225	8097	1128	56	5	179.2
	IV5	1972	370	1602	80	5	256
V	V 1	22704	20901	1803	90	5	288
	V 2	1252	814	438	22	5	70.4
	V 3	59307	53353	5954	96	2	307.2
	V 4	38292	32910	5382	96	2	307.2
VI	VI1	14813	2176	12637	96	1	307.2
	VI2	323	100	223	11	5	35.2
	VI4	6493	2345	4148	96	2	307.2
	VI5	2550	121	2429	96	4	307.2
	VI6	7978	720	7258	96	1	307.2
VII	VII1	8936	6803	2133	96	5	307.2
	VII2	8481	6421	2060	96	5	307.2
	VII3	16257	12069	4188	96	2	307.2
	VII4	3916	3225	691	35	5	112
	VII5	3544	2773	771	39	5	124.8
VIII	VIII1	5748	3531	2217	96	4	307.2
	VIII2	15713	7249	8464	96	1	307.2
	VIII3	6410	2485	3925	96	2	307.2
	VIII4	7875	1971	5904	96	2	307.2
	VIII5	1128	433	695	35	5	112
IX	IX1	3311	972	2339	96	4	307.2
	IX2	4481	414	4067	96	2	307.2
	IX3	3722	656	3066	96	3	307.2
X	X 1	9619	2917	6702	96	1	307.2
	X 2	7052	1805	5247	96	2	307.2
	X 3	6682	1654	5028	96	2	307.2
	X 4	4704	966	3738	96	3	307.2
XI	XI1	17594	3251	14343	96	1	307.2
	XI2	223	19	204	10	5	32
	XI3	2187	679	1508	75	5	240
全国		488754	292584	196170	3579	1.82	11452.8

注: 初级抽样单元样本量计算采用简单随机抽样的样本量计算公式。样条采样的抽样率为 2%。



图 7 中国农作物分类成数调查第一阶段抽样样本分布图

Fig. 7 Primary sampling unit used for crop type proportion survey

其中, L 为样条长度, D 为缓冲区宽度, K 为初级抽样单元边长。在具体抽样过程中, K 与 D 均已知, 那么样条的长度 L 实际反映了第二阶段的抽样水平, 因此抽样水平由样条长度反映出来,

$$L = \frac{K^2}{D} n_s \quad (7)$$

理论上, 第二阶段的抽样水平应该根据每个初级抽样单元的实际情况确定, 但在实际抽样过程中, 由于每个初级抽样单元的自然条件不同, 农作物种植结构也有所不同, 如果根据农作物种植结构和农作物分布特征计算样本量, 目前还没有有效的方法, 因此, 中国农情遥感速报系统以最低抽样率的形式指定最小抽样线路长度, 以保证抽样精度。

借鉴相关研究的经验和中国种植结构的特点^[25], 确定农作物分类成数抽样的第二阶段抽样采用 2% 的抽样率。初级抽样单元尺寸 $K = 4\text{km}$, 缓冲区宽度 $D = 0.1\text{km}$, 则最短抽样线路长度 $L = 3.2\text{km}$ 。按照这个标准和第一阶段抽样计算求得的初级抽样单元数目, 求得全国样条总里程为 1.15 万 km (表 4)。

尽管第二阶段抽样仅从抽样率上进行了限定, 但在具体选择采样线路时, 还要根据实际情况, 考虑

以下因素: (1) 线路的连通性, 以便于进行连续采样和减少线路行程; (2) 尽量选取居民地少和视野比较开阔的地段, 以便于能够充分观察道路沿线的农作物种植结构; (3) 尽量选取低等级公路和等外路甚至土路, 以避免公路两边的种植结构与其他地区的差异; (4) 线路一经选定就固定下来不变, 每年对选定的线路进行采样, 如果遇到道路施工或其他情况可以采用替代线路, 但下一次采样时如能恢复到原线路, 仍采用原线路。

5.4 样条采样框架的可行性分析

样条采样框架的可行性主要集结于两个方面的分析。

首先是道路长度能否满足分类成数采样的需求。采样的执行以道路为基础, 在给定最短样线长度的前提下, 分类成数采样的可行性首先取决于道路网络的密度。根据国家统计局公布的公路里程资料及各省区面积计算出来的公路密度(表 5), 除台湾、香港、澳门因没有资料不对其进行作物抽样, 其余的 31 个省区中, 大部分省区的公路密度大于 $0.2\text{km}/\text{km}^2$, 对于一个 $4\text{km} \times 4\text{km}$ 的初级抽样单元,

平均公路线路长度相当于 $0.2 \times 16 = 3.2\text{km}$, 可以满足 2% 的抽样率。从山东省的资料看(表 6), 平均公路密度 47.24% ($0.47\text{km}/\text{km}^2$), 除德州市外各市公路密度均在 $0.4\text{km}/\text{km}^2$ 以上, 全部可以满足 2% 的抽样要求。不可否认, 全国部分地区, 甚至东部的一些地区, 公路密度小于 $0.2\text{km}/\text{km}^2$, 但这些地区非农区所占比例相对较大(如内蒙古、新疆和西藏等), 如果

考虑到仅对农区的公路密度进行计算, 一般都会大于 $0.2\text{km}/\text{km}^2$, 基本满足抽样的需求。另外, 如果将非统计在内的乡间道路计算在内, 公路密度将会远远超过表 5 中的数字。更重要的是, 从抽样执行角度讲, 由于公路多接近于居民点, 接近于耕作区, 如果采样时有意避开的非农作物种植区, 那么实际抽样率将会高于表 5 中的数字。

表 5 2000 年全国各省区公路运输线长度及其密度表

Table 5 Road length and road density of provinces in 2000

省区	公路里程/km	面积/ km^2	密度/(km/km^2)	省区	公路里程/km	面积/ km^2	密度/(km/km^2)
北京	12825	16380.14	0.783	江苏	27749	100209.2	0.277
天津	8844	11519.3	0.768	云南	102405	381271.5	0.269
上海	4231	6107.344	0.693	江西	36918	167387.8	0.221
广东	95610	177234.9	0.539	广西	51378	236262.5	0.217
海南	17124	33852.25	0.506	陕西	43212	206034.4	0.21
山东	67847	153551.5	0.442	宁夏	10015	51562.65	0.194
福建	50202	120986.6	0.415	贵州	33973	176168.7	0.193
浙江	40266	101461	0.397	四川	89318	481920.4	0.185
河南	60330	165644.3	0.364	吉林	34516	190335.9	0.181
重庆	28086	82759.41	0.339	黑龙江	49928	452366	0.11
山西	52807	156434.1	0.338	甘肃	36212	393892.5	0.092
河北	58162	186853.7	0.311	内蒙古	63824	1150730	0.055
辽宁	45020	145993.2	0.308	青海省	18268	723317	0.025
湖北	55389	185399.5	0.299	新疆	33484	1640480	0.02
安徽	40857	139782.9	0.292	西藏	22475	1202440	0.019
湖南	60416	211758.6	0.285	全国	1351691	9600000	0.141

注: 公路里程数据来自统计局公布数字(<http://www.stats.gov.cn/ndsj/zgnj/2000/003c.htm>), 行政单元面积没有查到国家公布的相关数字, 根据 1:100 万行政界线数据计算。

样条采样框架可行性分析的另一个焦点, 集中在农作物种植结构的分布差异方面。道路两边, 农作物种植结构的分布与其他地区可能存在较大差异, 因此, 沿道路采样可能存在偏差。如公路两边, 尤其是高等公路往往较多种植蔬菜、瓜果等作物的现象。为了克服公路两边种植结构的异常, 在实际样线选取过程中, 确定了以下原则:

(1) 优先选择乡道、等外路, 避开等级路的选样, 这样既可以避免道路两边建筑对采样工作的影响, 也提高了样线的代表性;

(2) 避开城镇选择样线, 因为在城市外围往往有蔬菜、瓜果等的经济农业圈层结构, 这在很大程度上也避免了样线两边的种植结构异常;

(3) 培养专业的地面调查队伍, 选择有经验专业人员执行样线选择及调查工作。他们除了对各个

初级抽样单内的种植结构熟悉之外, 对于种植结构的变化及其影响因素也很熟悉, 这些知识对于他们选择合适的样线有很大的帮助。

5.5 样条采样框架优点分析

从样条采样框架的实用性方面分析, 样条采样框架具有以下优点:

(1) 继承了面积采样框架适合于作物分布零散、插花、成数低等面积采样框架的优点。

(2) 保证了样区的面状特征, 保证了更大信息量的获取。

(3) 样区呈线状分布, 即样区由长而窄的样条组成, 适合于借助交通工具和摄影工具, 由专业的采样队伍在较短的时间内完成采样, 提高了采样效率。

表 6 2002 年山东省公路里程及公路密度
Table 6 Road length and density of Shandong province in 2002

名称	公路里程/km	等级公路/km	高级与次高级/km	高速公路/km	公路密度/%
济南市	4450.5	4373.9	4330.8	189.9	54.27
青岛市	5591	5591	3666.4	287	52.25
淄博市	3529	3529	3109.9	147.9	58.82
枣庄市	1925.1	1920.8	1723.6	115.2	41.85
东营市	4507.7	4507.7	4507.7	132.8	57.06
烟台市	6161	6161	3812.3	200.1	44.64
潍坊市	7476.2	7466.9	4645.8	204.5	47.02
济宁市	5109.1	5109.1	4868.5	145.6	47.75
泰安市	4350.5	4350.5	3535.1	208.4	55.78
威海市	2405	2350.2	1714.8	13.6	43.73
日照市	2317.1	2317.1	1127.2	78.5	43.72
莱芜市	1088.7	1088.7	801.7	97.2	47.33
临沂市	4029.8	4029.8	4029.8	25.4	42.42
德州市	3822.4	3822.4	3814	142.7	36.75
聊城市	3868.5	3868.5	3868.3	111.8	44.47
滨州市	7753.2	7753.2	5080.2	205.6	45.08
菏泽市	5644.5	5644.5	5559.1	104.5	46.27
全省	74029.3	73884.3	60195.2	2410.7	47.24

注: 数字引自山东省统计局资料, 网址: <http://www.stats-sd.gov.cn/disp/tjsj.asp?id=1000000587>。

5.6 GVG 农情采样系统

中国农情遥感速报系统以样条采样框架技术为基础, 建立了全国性农情采样网络, 每年定期派出农情采样队伍进行农作物分类成数的现场调查^[37]。但是, 要完成总里程达 1.15 万 km 的作物分类成数调查, 工作量巨大且成本极高, 时间需要长。因此, 研制了 GVG 农情采样系统^[38], 配合样条采样框架的野外调查。

GVG 农情采样系统是 GPS, Video 和 GIS 技术的集成。通过视频技术大量采集反映作物种植情况的现场照片, 同时用 GPS 的定位功能确定照片的空间位置, 然后通过室内解译确定每张照片的作物分类成数, 并利用 GIS 技术统计计算每条样线、每个抽样单元、县级行政单元、区划单元和省级行政单元等各个空间尺度的作物分类成数。

GVG 农情采样系统是样条采样技术的执行系统, 为了保证野外采样质量, 制订了标准的野外采样运行规程, 明确规定采样时间(前后不能相差 1 周)

和标准的仪器操作规范, 如车速限制、拍摄方位、拍摄距离以及软件操作等内容。

5.7 分类成数计算

分类成数通过解译 GVG 系统所拍摄的野外照片, 确定每张照片内不同农作物类型的分类成数, 然后汇总到不同的统计单元, 得到指定统计单元内的作物分类成数。采用公式:

$$P_j = \frac{\sum_{i=1}^N a_{ij}}{\sum_{i=1}^N A_i} \quad j = 1, 2, \dots, M \quad (8)$$

式中 P_j 是第 j 种作物的分类成数, a_{ij} 是第 j 种作物在第 i 张 GVG 照片中的比例, A_i 是第 i 张 GVG 照片中作物的百分比。 N 表示统计单元内有 N 个有效 GVG 照片记录, M 表示区划单元内有 M 种作物类型。

分类成数的统计过程中, 统计单元的范围是关键。考虑到分类成数的应用及采样在空间上的异质性特点, 汇总时分 4 类统计单元进行汇总。

第 1 类统计单元即初级抽样单元, 面积是 4km × 4km。该类统计的目的是为了提取微观尺度上的分类成数信息, 需要对最多有 90 条左右的记录 (即照片判读记录) 进行汇总。

第 2 类统计单元是县级行政单元。按县级区划单元对 GVG 影像记录进行汇总。

第 3 类统计单元是种植结构区划单元。将各个初级抽样单元的分类成数汇总起来, 即可得到种植结构区划单元内的作物分类成数。种植结构区划单元内, 有几个甚至几十个县, 采样时也许仅对其中几个县进行了采样, 区划单元内的作物分类成数代表着单元内所有县的平均水平。

第 4 类统计单元是省级行政单元。当将县级行政单元或种植结构区划单元的作物分类成数汇总到省级行政单元的时候, 由于省级行政单元范围内的种植结构和物候期差异很大, 简单地采用上述统计方法会造成严重的误差。以河南省夏季作物采样为例, 在河南省南部的的主要作物有小麦和油菜, 且油菜占有很大的比例, 而在北部主要是冬小麦, 油菜的比例很低。如果简单统计, 会夸大冬小麦的成数, 而减少油菜的成数, 造成很大的偏差, 因此在省级或全国尺度的统计时, 采用面积加权方法:

$$P_j = \frac{\sum_{i=1}^N P_{ij} B_{ij}}{\sum_{i=1}^N B_{ij}} \quad j = 1, 2, \dots, M \quad (9)$$

式中 P_j 是第 j 种作物的分类成数, P_{ij} 是第 i 个统计单元 (县或区划单元) 中第 j 种作物的成数, B_{ij} 是第 i 个统计单元第 j 种作物上年的种植面积, N 是统计单元数, 即省内所有的县级行政单元或区划单元的

数目。

5.8 样条采样框架用于农作物分类成数调查的精度检验

从 2000 年开始建立地面实验样区进行分类成数调查的精度检验, 2000 年与 2001 年布设了江苏江宁与吉林德惠 2 个实验样区, 2002 年重新布设了吉林双阳、河北栾城和湖北新洲 3 个实验样区, 共 5 个实验样区, 每个实验样区大小为 5km × 5km。

5.8.1 精度检验方法

一方面, 在样区内选取采样路线, 利用 GVG 系统采样计算样区内各种农作物的分类成数。同时用高分辨率遥感影像进行作物分类, 并借助大比例尺 (1:1 万) 土地利用图进行外业调查, 生成样区农作物种植分布图, 并计算各种农作物的分类成数。最后将两种方法得到的结果进行对比分析, 以检验样条采样的精度。

5.8.2 精度检验结果及分析

各年的作物分类成数调查精度通过表 7 可以反映出来。水稻的精度最高 99.9%, 最低为 89.1%, 大部分在 94% 以上。冬小麦的验证精度也很高, 可以达到 98.9%。玉米精度除 2000 年和 2001 年由于样区内的小面积种植, 所抽取的样线内出现的面积小导致精度不足外, 2002 年在两个样区的检验精度均在 90% 以上。棉花由于同样的原因, 检验精度仅 70%, 该地区棉花的种植成数较低, 仅 20% 左右。表明样条采样框架对于大宗作物具有较高的抽样调查精度, 能够满足中国农情遥感速报系统的分类成数调查要求。

表 7 样条采样框架各样区历年检验精度

Table 7 Transaction sampling precise validated from 2000 to 2002

样区	2000 年		2001 年		2002 年			
	水稻	玉米	水稻	玉米	水稻	玉米	冬小麦	棉花
德惠	94.3	70.0	94.4	80.3	-	-	-	-
双阳	-	-	-	-	95.9	92.5	-	-
栾城	-	-	-	-	-	100.0	98.9	-
新洲	-	-	-	-	89.1	-	-	70.0
江宁	96.0	-	99.9	-	92.4	-	-	-

6 农作物种植面积估算

中国农情遥感速报系统在耕地数据库的支持

下, 利用作物种植成数和分类成数计算作物种植面积。农作物种植面积估算可以按不同单元进行计算。

作物种植面积根据耕地面积 (早中晚稻为水田

面积) 与种植成数和分类成数连乘后得到。但是由于雷达数据可以直接从影像中解译出水稻种植面积, 因此作特殊处理。

(1) 雷达影像覆盖区域, 水稻种植面积用耕地面积乘以水稻种植成数, 此时水稻种植成数即直接从雷达影像解译出种植成数:

$$\text{水稻种植面积} = \text{耕地面积} \times \text{整群抽样获取的水稻种植成数}$$

(2) 光学影像覆盖区域, 作物种植面积即耕地面积与作物种植成数和分类成数的乘积, 即:

$$\text{作物种植面积} = \text{耕地面积} \times \text{种植成数} \times \text{分类成数}$$

按照这种计算方法求得不同单元的农作物种植面积, 最后汇总到省级行政单元, 得到最终发布的各省作物种植面积。汇总方法采用加权求和方式, 权重是省内各单元的耕地面积或前一年的农作物种植面积, 两种权重比较起来, 前一年的农作物种植面积要比耕地面积更准确, 应优先采用。即:

$$A_p = \sum_{i=1}^n A_i W_i \setminus \sum_{i=1}^n A_i \quad (10)$$

其中, A_p 是省级行政单元的作物种植面积, A_i 为省内下属单元的作物种植面积, W_i 为 i 单元的耕地面

积或前一年作物种植面积, n 是单元数。单元可以是县或区划单元。

7 应用实例 ——2003 年早稻种植面积估算

中国农情遥感速报系统 2003 年早稻种植面积的估算工作开始于 4 月中旬, 派出农情采样队伍对广东、广西、海南、云南、江西、福建等 6 省区的分类成数进行调查, 5 月中派出农情采样队伍对浙江、安徽、湖北、湖南 4 省的分类成数进行调查, 6 月上旬各采样队伍陆续回到驻地, 在室内进行分类成数统计, 6 月 20 日收到各采样队伍上报的早稻主产 10 省区的分类成数。同时, 于 6 月上旬订购遥感卫星数据 16 帧(其中包括 Landsat ETM/TM 数据 14 景, Radarsat SCAN SAR 数据 2 景, 图 8) 用于种植成数遥感监测。遥感图像经过处理后, 解译汇总生成早稻种植成数。

早稻种植面积估算在区划单元水平上估算, 然后汇总到省级行政单元, 估算表明 2003 年度全国早稻种植面积为 5928km²(表 8)^[39]。



图 8 2003 年早稻种植面积估算所用的 TM/ETM 和 Radarsat 数据分布图
Fig. 8 Landsat TM and Radarsat images used for early rice acreage estimation

8 结 论

中国农情遥感速报系统结合中国农作物种植制度特点, 提出在农作物种植结构区划的基础上, 利用

中国现有的 1: 10 万比例尺耕地数据库, 通过两次抽样分别获得农作物总种植成数和具体某一种农作物的分类成数, 最后将耕地面积与这两个成数相乘得到某一种作物的种植面积, 这样从总体思路保证了农作物种植面积的监测质量。

表8 2003年早稻种植面积估算结果

Table 8 Acreage estimated of early rice

省 区	水田面积/ hm ²	种植成数	分类成数	早稻面积/hm ²
浙江省	1 958 220.46	0.1758	—	344 255.16
安徽省	1 679 472.58	0.2256	—	378 889.01
福建省	1 035 037.82	0.7322	0.5315	402 799.77
江西省	2 336 754.53	0.8527	0.5354	1 066 811.58
湖北省	1 411 892.50	0.6861	0.4145	401 525.92
湖南省	3 111 681.94	0.8197	0.5450	1 390 101.90
广东省	2 103 162.20	0.6929	0.5691	829 338.67
广西壮族自治区	2 305 494.17	0.7878	0.5404	968 117.20
海南省	304 860.62	0.6394	0.5508	107 366.28
云南省	669 415.20	0.2487	0.1941	32 314.46
10省合计	16 915 992.02	—	—	5 921 519.95
全国预计	296 937 519.39	—	—	5 928 159.10

注: 浙江、安徽两省的早稻种植面积根据雷达数据推算出来, 无分类成数; 全国早稻面积根据10省早稻面积推算出来, 推算方法是用10省区早稻面积除以10省区早稻面积占全国早稻面积的历史平均比例。

农作物种植结构区划把全国分成种植面积相对稳定、种植结构相对一致、自然条件相近的区域构成抽样层, 从而针对不同的区划单元进行抽样, 以少量的样本量, 精确地估算农作物遥感种植面积和种植结构, 达到降低成本的目的, 是解决中国农作物种植结构复杂现状, 提高农作物遥感监测精度的有效途径之一。

采用整群抽样技术, 克服了对大面积监测区域遥感监测的技术瓶颈, 在整群抽样理论的指导下, 避免了全覆盖的遥感数据订购方法, 而是通过选择合适的遥感数据, 经分类解译得到最终的农作物总种植面积, 结合耕地数据, 最终得到每个区划单元的作物总种植成数, 既节约了成本, 又保证了精度。

为了实现对中国农作物的分类成数进行抽样调查, 并能适应中国农作物种植的高度插花现象, 结合中国土地使用制度和农作物种植制度的特点, 提出的样条采样框架技术, 具有效率高、费用低、灵活性强、简单易用和多用途的特点, 可以对大宗作物成数进行调查, 精度证明可达到95%以上。经过1998—2003年共6年的运行和不断完善, 已经形成一套比较成熟、可靠的作物面积遥感估算技术体系, 逐步实现标准化、流程化和系统化。

参 考 文 献 (References)

[1] Liu H Q, Jin M Y. Applications of Remote Sensing in Agriculture in the United States [J]. *China Agricultural Resources and Regional Planning*, 1999, 20 (2): 56—60. [刘海启, 金敏毓. 美国农业遥

感技术应用状况概述[J]. *中国农业资源与区划*, 1999, 20 (2): 56—60.]

- [2] Liu H Q. The Introduction of Mars Plan of European Commission and Chinese Agriculture Remote Sensing Application [J]. *China Agricultural Resources and Regional Planning*, 1999, 20 (3): 55—57. [刘海启. 欧盟MARS计划简介与我国农业遥感应用思路[J]. *中国农业资源与区划*, 1999, 20 (3): 55—57.]
- [3] Yang B J, Pei Z Y, Zhou Q B, et al. Key Technologies of Crop Monitoring Using Remote Sensing at a National Scale: Progress and Problems [J]. *Transactions of Case*, 2002, 18 (3): 191—194. [杨邦杰, 裴志远, 周清波等. 我国农情遥感监测关键技术研究进展[J]. *农业工程学报*, 2002, 18 (3): 191—194.]
- [4] Xu X R. Bulletin on Environment Monitoring and Crop Yield Estimation with Remote Sensing [C]. Beijing: Beijing University Press, 1991. [徐希孺. 环境监测与作物估产的遥感研究论文集[C]. 北京: 北京大学出版社, 1991.]
- [5] Winter Wheat Remote Sensing Production Estimation Cooperation Team. Winter Wheat Dynamic Monitoring Using meteorological Satellite [M]. Beijing: Meteorological Press, 1993. [全国冬小麦遥感综合测产协作组. 冬小麦气象卫星遥感动态监测与估产[M]. 北京: 气象出版社, 1993.]
- [6] Sun J L. Pandect on Dynamic Monitoring and Yield Estimation for Crop in China [M]. Beijing: China Science and Technology Press, 1996. [孙九林. 中国农作物遥感动态监测与估产总论[M]. 北京: 中国科学技术出版社, 1996.]
- [7] Liu H Q. Sampling Method with Remote Sensing for Monitoring of Cultivated Land Changes on Large Scale [J]. *Transactions of CSAE*, 2001, 17 (2): 168—171. [刘海启. 大尺度耕变化监测的遥感抽样方法研究[J]. *农业工程学报*, 2001, 17 (2): 168—171.]
- [8] Wu B F. Operational Remote Sensing Methods for Agricultural Statistics [J]. *Acta Geographica Sinica*, 2000, 55 (1): 23—35 [吴炳方. 全国农情监测与估产的运行化遥感方法[J]. *地理学报*, 2000,

- 55 (1): 23—35.]
- [9] Li Q Z, Wu B F. Discrimination of Seed Rape and Winter Wheat using two date ETM Data[A]. International Conference on Computer Graphics and Spatial Information System[C]. Beijing: China Meteorological Press, 2002.
- [10] Liew S C, Kam S, Tuong T, *et al.* Application of Multitemporal ERS-2 Synthetic Aperture Radar in Delineating Rice Cropping Systems in the Mekong River Delta[A]. Vietnam, IEEE Transactions on Geosciences and Remote Sensing[C]. 1998, **36** (5): 1412—1420.
- [11] Sushil Pradhan. Crop Area Estimation Using GIS, Remote Sensing and Area Frame Sampling[J]. *JAG*, 2001, **3** (1): 86—92.
- [12] Li S K. A General Introduction to the Complete Set of Research Achievements “A Study of National Agroclimatic Resources and Agroclimatic Regionalization in China” [J]. *Journal of Arid Land Resources and Environment*, 1994, **4** (2): 116—122. [李世奎.《全国农业气候资源和农业气候区划研究》系列成果综述[J]. 干旱区资源与环境, 1994, **4** (2): 116—122.]
- [13] Zhou Y, Tian G L. Chinese Crop Planting structure Stratification [R]. Institute of Remote Sensing Applications, Chinese Academy of Sciences, 1999. [周艺, 田国良. 中国农作物种植结构区划[R], 中国科学院遥感应用研究所, 1999.]
- [14] Kang X F, Wang N B, Yang X H. Process of Information Extraction Methods for Crop Planting Area[J]. *Resources Sciences*, 2002, **24** (5): 8—12 [康晓风, 王乃斌, 杨小映. 粮食种植面积提取方法的发展与现状[J]. 资源科学, 2002, **24** (5): 8—12.]
- [15] Michael S Williams, Marian Eriksson. Comparing the Two Paradigms for Fixed-area Sampling in Large-scale Inventories[J]. *Forest Ecology and Management*, 2002, **168**: 135—148.
- [16] Bush J, House C. The Area Frame: a Sampling Base for Establishment Surveys[A]. Cox B G, Binder D A, Chinnappa B N, *et al.* Business Survey Methods[C]. Wiley, New York, 1995.
- [17] Ripely B D. Spatial Statistics[M]. Wiley, New York, 1981.
- [18] Diggle P J. Statistical Analysis of Spatial Point Patterns[M]. Academic Press, London, 1983.
- [19] Stephen V Stehman, W Scott Overton. Spatial Sampling[M]. Sandra lach Arlinghaus, Practical Handbook of Spatial Statistics, CRC Press, 1996.
- [20] Kaiser L. Unbiased Estimation in Line-intersect Sampling [J]. *Biometrics*, 1983, **39**: 965—976.
- [21] Barlett R F. Estimating the Total of a Continuous Population[J]. *Stat. Plan. Infer.* 1986, **13**: 51—66.
- [22] Stevens Jr, D L. Variable Density Grid-based Sampling Designs for Continuous Spatial Populations [J]. *Environmetrics*, 1997, **8**: 167—195
- [23] Michael S Williams, Marian Eriksson. Comparing the Two Paradigms for Fixed-area Sampling in Large-scale Inventories[J]. *Forest Ecology and Management*, 2002, **168**: 135—148.
- [24] Wang Y Y, Ma J F, Chen Y Q. Percentage Sampling on TM Image for Rice Area Inventory [J]. *Jiangsu Journal of Agricultural Sciences*, 1989, **5** (3): 34—39. [王延颐, 马金福, 陈玉泉. 用卫片成数抽样法进行稻田面积调查[J]. 江苏农业学报, 1989, **5** (3): 34—39.]
- [25] Zhou H M. A Procedure for Sampling Investigation of Paddy Area [J]. *Southwest China Journal of Agricultural Sciences*, 1996, **9** (3): 100—105. [周华茂. 水稻播面遥感抽样调查技术探讨[J]. 西南农业学报, 1996, **9** (3): 100—105.]
- [26] Chen Z X, Liu H Q, Zhou Q B, *et al.* Sampling and Scale Scheme for Monitoring the Change of Winter Wheat Acreage in China[J]. *Transaction of CSAE*, 2000, **16** (5): 126—129. [陈仲新, 刘海启, 周清波等. 全国冬小麦面积变化遥感监测抽样外推方法的研究[J]. 农业工程学报, 2000, **16** (5): 126—129.]
- [27] Jiao X F, Yang B J. Design of Sampling Method for Cotton Field Area Estimation Using Remote Sensing at a National Level [J]. *Transaction of CSAE*, 2002, **18** (4): 126—129. [焦险峰, 杨邦杰. 全国棉花种植面积遥感监测抽样方法设计[J]. 农业工程学报, 2002, **18** (4): 126—129.]
- [28] Gallego F J. Agriculture Sampling Frames of Square Segments[M]. Official Publications of the European Communities, 1995.
- [29] Gallego F J. Double Sampling for Area Estimates and Map Accuracy Assessment in Quantifying Spatial Uncertainty in Natural Resources [A]. Mower, H T, Congalton R G. Theory and Application for GIS and Remote Sensing[C]. Ann arbor Press, Chelsea, Michian, USA, 2000: 65—77.
- [30] Tsiligrades T A. Remote Sensing as a Tool for Agricultural Statistics: a Case Study of Area Frame Sampling Methodology in Hellas [J]. *Computers and Electronics in Agriculture*, 1998, **20**: 45—77.
- [31] Sushil Pradhan. Crop Area Estimation Using GIS, Remote Sensing and Area Frame Sampling[J]. *JAG*, 2001, **3** (1): 86—92.
- [32] David J. Megill. Recommendations on Sampling and Estimation Methodology for Rwanda Agricultural Surveys[M]. Food Security Research Project (FSRP), Rwanda Ministry of Agriculture, Livestock and Forestry, U. S. Agency for International Development, 1999.
- [33] Deneufchatel D, Pordier J C. Area Frame Surveys and Remote Sensing: Guide to Surveying by Segments and Points[A]. Study Group on Food and Agricultural Statistics in Europe [C]. Geneva, 5—9 July 1993.
- [34] FAO. Multiple Frame Agricultural Surveys[C]. Volume 1: Current Surveys Based on Area and List Sampling Methods, Rome, 1996.
- [35] FAO. Multiple Frame Agricultural Surveys[C]. Volume 2: Agricultural Survey Programmes Based on Area Frame or Dual Frame (Area and List) Sample Designs, Rome, 1998.
- [36] Womer S P, Shah R, Chapman R B. Systematic Versus Simple Random Sampling in Plant Protection[A]. Proc. 52nd N. Z. Plant Protection Conf [C]. 1999.
- [37] Zhang L, Wu B F, Li Q Z. Agro-sampling Network and Its Implementation in China [J]. *Journal of Remote Sensing*, 2004, **8** (6): 593—601. [张磊, 吴炳方, 李强子. 中国农情采样网络的组织与实施[J]. 遥感学报, 2004, **8** (6): 593—601.]
- [38] Wu B F, Tian Y C, Li Q Z. GVG, a Crop Type Process Monitoring instrument[J]. *Journal of Remote Sensing*, 2004, **8** (6): 570—580. [吴炳方, 田亦陈, 李强子, GVG 农情采样系统及其应用[J]. 遥感学报, 2004, **8** (6): 570—580.]
- [39] Wu B F. China Crop Watch Bulletin[R]. 2003. [吴炳方. 中国农情遥感速报[R].]

Crop Acreage Estimation Using Two Individual Sampling Frameworks with Stratification

WU Bing-fang, LI Qiang-zi

(*Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, 100101 China*)

Abstract: Crop plots are very small in China due to special farm-use rules. Image classification techniques are limited in crop acreage survey using remote sensing. In this paper, we analyze this problem and provide a substitute methodology to estimate crop acreage.

In this methodology, crop stratification is fundamental. Proportions of main crop types as well as physical factors of temperature, precipitation, soil type and sun radiation are considered. There are about 11 strata in China at the first level based on physical factors, 44 strata at the second level based on crop proportion and 102 strata at the third level based on arable land intensity.

Two individual sampling frameworks are used. The cluster sampling is used to estimate the proportion of planted area on arable land with remote sensing data, mostly Landsat and Radarsat data, currently also ENVISAT ASAR data are used. The clusters are defined as a map sheet at a scale of 1: 100 000 (about 1/16 Landsat TM scene). And images are selected based on cluster sample randomly for each crop season. After atmospheric correction, geometric correction non-arable land masking, remotely sensed images are classified by ISODATA unsupervised classification, and then the planted areas are labeled by considering *NDVI* value. The planted area proportions are calculated for each stratum.

The transect sampling framework is used to estimate the proportions of different crop type within planted area. To identify the crop proportion of a small parcel, field works should be used since it is impossible to make crop classification with remote sensing data cost-effectively. The transect sampling actually is a two-stage sampling. In the first stage, PSUs are selected randomly on a 4km × 4km area frame. In the second stage, the selected PSUs are sampled only along the road within PSUs, called transect line. The sampling works in the field are to take pictures along the road within PSUs with 100 m buffer, and a GVG system is designed for this purpose. Proportions of every crop type are calculated for each stratum.

Crop acreages are calculated under the support of current arable database. For every crop type, the planted acreage is the arable area multiply by planted proportion and crop type proportion of stratum.

The estimation of early rice acreage in 2003 in China is presented as a case study. Results show this methodology is feasible. This methodology has adopted in China since 1998. And the experience shows that the stratification schema is efficiency, and the two individual sample frameworks can generate accurate estimation of crop acreage.

Key words: crop proportion; crop type proportion; cluster sampling; transect sampling; remote sensing; crop acreage