

基于分类回归树分析的棉花种植面积提取

——以库、新、沙三县为例

玉苏普江·艾麦提, 玉苏甫·买买提, 阿里木江·卡斯木

(新疆师范大学地理科学与旅游学院, 新疆 乌鲁木齐 830054)

摘要: 以研究区 2012 年的 HJ 卫星 CCD 影像为数据源, 通过物候历和主要农作物的光谱特征分析, 确定棉花识别最佳时相。采用分类回归树分析 (CART) 的决策树方法提取棉花种植面积信息, 并以农田实地调查样点和统计数据为参考对提取的棉花种植面积结果进行评价。结果表明, 基于 HJ-CCD 数据, 使用 CART 算法的决策树可以较好地提取棉花覆盖信息, 最终提取的棉花种植面积总量精度为 94.29%, 位置精度为 88.57%; 本研究采用的决策树方法, 操作方便、容易实现, 分类结果较为实际, 基本满足棉花种植面积遥感监测的需求, 可对棉花种植面积估算和种植结构分析提供一定的参考。

关键词: 棉花; 面积提取; 决策树分类; CART 算法

中图分类号: S29 **文献标志码:** A **文章编号:** 1000-7601(2014)05-0187-05

Cotton planted areas extraction based on the CART analysis

——Taking three Counties (Kuqa, Xinha and Shaya) as examples

Yusupjiang Aimaiti, Yusufu Maimaiti, Alimujiang kasimu

(Institute of Geographical Sciences and Tourism, Xinjiang Normal University, Urumqi, Xinjiang 830054, China)

Abstract: In this paper, the HJ satellite CCD images in research region of 2012 were selected as the source data, through the analysis of phenological calendar and spectral characteristics of major crops, the cotton identifiable best phase was determined. The cotton planted areas information were extracted by using the decision tree method based on CART analysis, also the extracted areas result have been evaluated by the field real survey samples and the statistical data as a reference. The results showed that: Based on the HJ satellite CCD images data, using the decision tree of CART algorithm can be well applied to extract the cotton covering information. The final total accuracy for extracted cotton planted areas was 94.29%, and the position precision was 88.57%. In this research, adopted the decision tree method, has the advantages of convenient operation, easy to implement and more realistic classification results. It can be basically met the needs of remote sensing monitoring for cotton planted areas, and provided a certain reference to estimate the cotton planted areas and analyze the planting structures.

Keywords: cotton; areas extraction; decision tree classification; CART algorithm

棉花作为我国重要的农业经济作物之一, 其产量历来为各级政府所关注。决定棉花产量的关键因素之一是棉花的播种面积, 快速、经济、客观地监测棉花种植面积和长势变化情况, 对于地方生产部门适时掌握棉花生产形势和科学指导棉花生产具有十分重要的意义^[1-3]。

新疆是我国最大的商品棉生产出口基地之一, 与其他产棉区相比, 在棉花单位面积产量、总产量和经济效益等方面有优势, 而南疆的棉花种植面积、总

产量约占到全新疆的 2/3。因此, 棉花是南疆地区的重要支柱产业, 是农民增收的重要保障。南疆地区的棉花播种面积大, 棉田规整且面积大, 统一种植与栽培管理, 再加上南疆地区晴天多, 阴天少, 属于灌溉农业, 具有开展遥感研究的优势^[4]。

目前我国学者用不同的数据和方法来进行棉花种植面积监测方面的研究^[5-7]。其中, 曹卫彬等利用 CBERS-1 与 Landsat TM 影像, 对新疆棉花遥感监测识别区域进行划分^[8]; 韩兰英等利用 SPOT 影

收稿日期: 2013-10-13

基金项目: 国家自然科学基金资助项目(41261056; 41361043; 31260048); 教育部回国人员科研项目; 新疆师范大学研究生创新基金项目(20131229); 新疆维吾尔自治区科技厅优秀科技人才项目(2013721031)

作者简介: 玉苏普江·艾麦提(1987—), 男, 新疆和田人, 硕士, 研究方向为遥感与环境资源。E-mail: yusup11@163.com。

通信作者: 阿里木江·卡斯木(1976—), 男, 新疆克拉玛依市人, 副教授, 博士, 主要从事环境遥感研究。E-mail: alimkasim@gmail.com。

像,监督分类和非监督分类方法比较进行了新疆阿克苏和喀什棉区棉花面积的估算^[9];刘姣娣等以 Landsat TM 影像为数据源利用混合像元分解技术识别棉花^[10];黄青等利用 MODIS - NDVI 差值模型,对新疆棉花 2009 年的长势进行监测^[11];潘伟等利用 1949—2007 年的棉花调查数据对新疆棉花种植面积的时空变化及适度规模进行分析^[12]。这些研究在一定程度上推进了新疆棉花种植面积监测技术的改善,并且为以后的棉花遥感监测研究提供了技术基础。

HJ 环境卫星作为我国自主研发的民用对地观测卫星,具有重访周期短、幅宽大、监测范围广,且数据处理简单等特点,在棉花遥感监测与估产方面比其他数据源更有优势。刘、欧文浩和王琼等利用 HJ - CCD 影像对农作物覆盖进行监测并获得了较好的效果^[13-15]。研究以库车、新和、沙雅三县为研究区,利用 2012 年 4 月 19 日、6 月 18 日和 9 月 12 日的 HJ - 1 卫星 CCD 影像为数据源,ENVI 下基于 CART 算法的决策树规则提取研究区棉花的种植面

积信息,结合地面实际调查资料,对研究区棉花种植面积进行检测,旨在利用决策树构建 HJ 卫星数据的棉花提取方法,以便为将来在新疆的棉花遥感监测领域中深入地应用国产 HJ 卫星数据提供一定的技术基础。

1 研究区概况

研究区位于新疆维吾尔自治区西部,塔里木盆地的中北部,渭干河库车河流域的下游,中部天山南麓,塔克拉玛干沙漠北缘,介于 39°30' ~ 42°40' N, 81°27' ~ 84°06' E。库车、沙雅和新和三县的土地总面积为 $523.76 \times 10^4 \text{ hm}^2$,其中大部分是沙漠和戈壁,绿洲面积仅有 $56.096 \times 10^4 \text{ hm}^2$,占总面积的 10.7%。从作物种植总面积来看,棉花、冬小麦和夏玉米种植面积居前 3 位,分别占作物种植总面积的 46.89%、24.95% 和 21.09%,三者的种植面积之和占农作物种植总面积的 92.93%。耕作制度包括以秋季为主的一年一熟或两年三熟的旱作轮作制,作物主要生长期在 4—10 月^[16]。

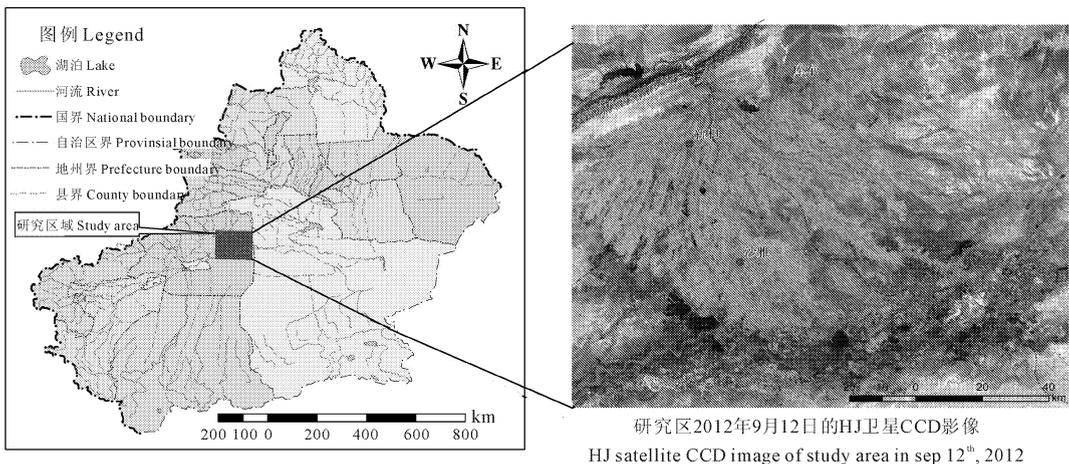


图 1 研究区地理位置

Fig. 1 The geographical position of research region

2 数据与研究方法

2.1 遥感数据的选择

环境与灾害监测预报小卫星星座 A、B 星(简称环境小卫星,简写 HJ - 1A/1B)是我国继气象、海洋、国土资源卫星之后一个全新的民用卫星。于 2009 年 3 月 30 日开始正式交付使用,HJ - 1 - A 星搭载了 CCD 相机和超光谱成像仪(HSI),HJ - 1 - B 星搭载了 CCD 相机和红外相机(IRS),CCD 相机重访周期为 96 h(HJ - 1/A 和 HJ - 1/B 组网后 48 h),空间分辨率为 30 m,共有蓝、绿、红和近红外 4 个波段。其卫星系统建设的主要任务是利用我国自主小卫星

星座,形成对我国生态环境和灾害遥感监测的能力,为我国环境保护与防灾减灾提供遥感信息与技术支撑,全面提高我国环境和灾害信息的获取、处理和应用的水平。

2.2 非遥感数据的选择

研究区收集的非遥感数据有:(1) 研究区 1:10 万地形图、行政区划矢量图和 2000—2012 年的统计资料;(2) 根据研究区作物类型和作物分布情况,沿一定的路线,用手持式全球定位系统(GPS)获取调查点的经纬度、高度和方向,采样了 179 个调查点的土壤剖面并记录了调查点的作物覆盖类型及健康状况等信息,分别用于室内试验、影像分类的训练样本

和位置精度验证。

2.3 最佳时相选择

最佳时相的选择是根据研究区不同作物的物候规律、季相节律的差异性和时相变化等信息来选取的。已有研究成果表明新疆棉花遥感较为容易识别期为6月中旬前后,9月上、中旬期间是一个最佳识别棉花的时相期^[8]。本文以棉花遥感识别的最佳时相为基准,还考虑遥感影像的可获取性、天气状况和数据质量等因素,选用2012年6月18日和9月12日的HJ卫星CCD遥感数据来提取棉花种植面积信息。

2.4 数据的预处理

卫星遥感数据获取后需要对数据进行定标、大气纠正和几何精校正等预处理。对于HJ-1号卫星CCD影像,利用ENVI 5.0软件FLAASH大气校正模块对影像进行了大气校正,将辐射亮度值转为地表反射率。以有精确地理坐标的LANDSAT TM卫星遥感数据为参考数据,对采集的2012年6月18日的HJ-1号卫星数据进行几何精校正,误差控制在半个像元以内。以纠正后的2012年6月18日的遥感影像为基准,分别对4月19日和9月12日的HJ卫星CCD遥感影像进行图像间的配准,误差控制在半个像元内。为增强图像的目视效果和提高解译精度,对影像进行线性拉伸和滤波等增强处理。包含研究区的最大矩形为感兴趣区,分别对3景CCD影像进行研究区的裁剪,得到了研究区HJ数据图像。

2.5 分类回归树(CART)算法

分类回归树 CART(Classification and Regression Tree)是Breiman等提出的一种决策树构建算法,并不断进行了改进^[17]。基本原理是通过由测试变量和目标变量构成的训练数据集的循环二分形成二叉树形式的决策树结构。其特点是充分利用二叉树的结构(Binary Tree-structured),即根节点包含所有样本,在一定的分割规则下根节点被分割成两个子节点,这个过程又在子节点上重复进行,成为一个回归过程直至不可再分成子节点为止。在土地利用/覆被分类中,目标变量是土地利用/覆被的类型值,测试变量为所利用的分类特征^[18-19]。基本技术流程包括:数据预处理,构建多元数据,获取规则,应用规则和分类后处理等。

3 棉花种植面积信息的提取

3.1 时序遥感影像 NDVI 特征变化

归一化植被指数(NDVI)是目前比较广泛应用的一种植被指数,可以作为植被总量、结构和状况的光谱度量,是植被生长状态及植被覆盖度的最佳指示因子。NDVI计算公式如下: $NDVI = (NIR - RED) / (NIR + RED)$ 其中,NIR为HJ-1号卫星近红外波段反射率,RED为红光波段反射率。利用ENVI 5.0下分别计算三期HJ-CCD影像的NDVI值并制作NDVI图(图2所示)。

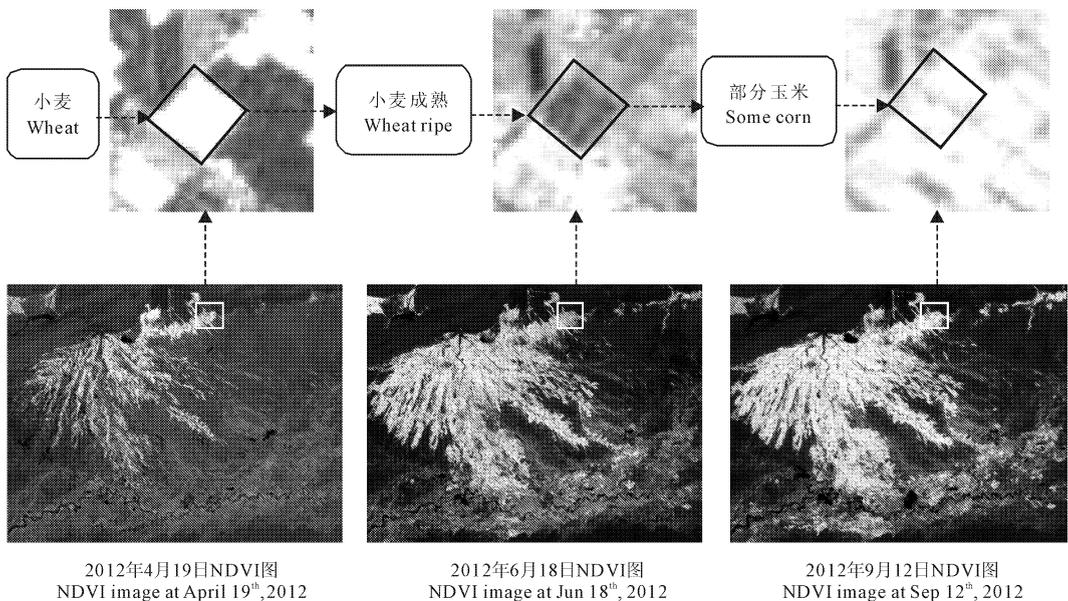


图2 研究区3个不同时期的NDVI变化图

Fig.2 NDVI change map of study area in three different periods

从图2中可以看出,作物在不同物候期的植被指数有相应的变化,4月中下旬是小麦处于生长旺

盛期NDVI值较高,到6月中下旬开始成熟NDVI值明显下降;9月上中旬,大部分春玉米已处于成熟

期,此时棉花则处于吐絮期,棉花的 NDVI 值与其差异较大,可以有效地提取棉花种植面积,但是免不了麦后复播的一些(夏玉米等)作物对棉花面积提取的干扰(见图 2)。

3.2 棉花种植信息提取

本文使用 CART 算法获取规则,基于规则提取棉花覆盖信息。首先准备每一期的 HJ 卫星影像的 NDVI 图,对每一期影像进行非监督分类,最大分类数量为 10,然后分别对每一期的 HJ 卫星影像构建多源数据集。数据集由蓝、绿、红、近红外 4 个波段、NDVI、ISODATA 非监督分类结果组合成一个 6 个波段的文件。其次以野外调查样点为基准,在实验区影像中选择一定数量的训练样本,利用 ENVI5.0 下的 CART 扩展工具,使用 CART 算法获取专家知识分类规则。本文提取棉花覆盖信息的总体技术流程如图 3 所示。

对 2012-09-12 日的合成影像,初步利用 CART 算法获取的分类规则进行分类,提取棉花的种植面积信息(此时提取的棉花信息中包含着部分错分类的其它农作物);其次对 2012-06-18 日的合成影像使用分类规则进行分类,提取棉花种植面积信息(此时小麦已成熟,而春玉米处在拔节期,夏玉米还没种植或没长出,所以这份地块的 NDVI 值较低(图 2)显得较暗)。因此在 ArcGIS10.0 软件的支持下对两期影像进行相交(Intersect)处理,剔除棉花种植面积中的玉米和其它差异较小的作物种植面积。由于分类结果图斑较为细碎,需要对栅格分类图进行分类后处理,因此进行了聚类统计、过滤和去除分析来提高分类精度。最终获取研究区棉花的种植面积总量与空间分布图(图 4)。

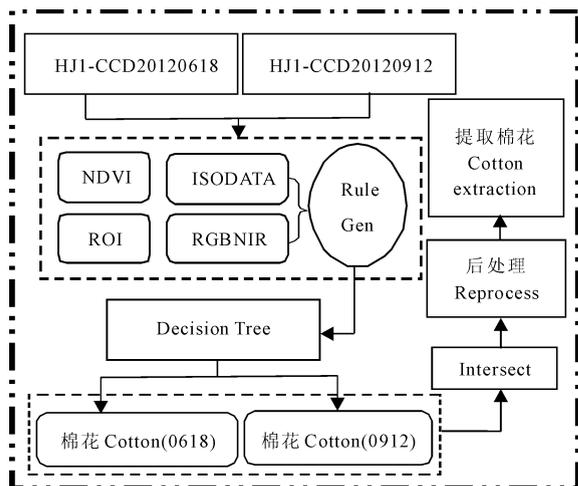


图 3 基于 HJ 卫星的棉花种植面积提取流程图

Fig.3 Flowchart of extracted cotton areas based on HJ satellite

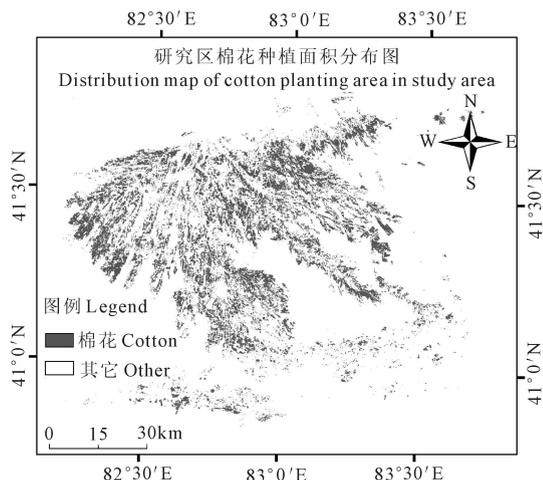


图 4 基于 HJ 卫星的棉花提取结果

Fig.4 Extracted result of cotton area based on HJ satellite

3.3 精度验证

为评价棉花种植面积提取的结果,本文从总量精度和面积位置精度两个方面进行精度评价^[20-21]。总量精度是指一定范围的自然或行政单元内棉花种植面积的总量精度,该研究以农业普查统计的农作物种植面积总量为参考值,对提取的棉花面积与已有的统计数据进行比较,其基于 HJ 卫星提取的棉花种植面积为 155 781.9 hm²,与政府统计数据 147 360.7 hm² 相比,准确率达 94.29%;位置精度主要是通过棉田野外调查样点与遥感实测棉田进行对比验证的。首先从棉田野外调查样点中选取未参与光谱训练的地面调查地块 105 个,一个一个与遥感图像解释结果进行对比,准确分类的棉田地块为 93 个,位置准确率达 88.57%。

4 结论与讨论

根据棉花的物候历特征,研究区关键时期多时相 HJ-CCD 卫星数据结合棉田实地调查数据进行处理得到了以下几方面的结论:

1) 充分利用 HJ 卫星 CCD 影像具有高时空分辨率、成本低和获取方便的优势,选取研究区 2012 年的棉花不同生育时期的 2 期 HJ 卫星影像,根据农作物不同物候期的影像特征,较准确地提取了研究区棉花的种植面积,总体面积精度达到了 94.29%,位置精度为 88.57%。说明基于 HJ 卫星的棉花种植面积提取在南疆地区有较高的精度和普适性,其在实际农业生产遥感监测方面有一定的潜力。

2) 研究过程中,首先使用基于 CART 算法的决策树规则提取了研究区不同时相的棉花的种植面积信息,并分析不同时期的农作物影像的特征进行图

像间的叠加处理,剔除棉花种植面积中的玉米和其它差异较小的作物种植面积,成功地提取了棉花种植面积。表明该方法与单纯进行监督分类相比,操作方便,分类精度较高,可基本满足棉花种植面积遥感监测的需求。

3) 本研究使用 CART 算法获取规则来提取棉花覆盖信息,其训练样本的正确选取,是一定程度上决定了影像分类精度,野外调查时如能利用高光谱仪收集典型农作物的高光谱数据,则有助于更准确地获取各样点农作物的信息,因而更精确地选择训练样区。此外,采用 HJ-CCD 卫星遥感影像仍存在混合像元的问题,若今后能结合我国 CBERS2B-HR 影像这类高分辨率遥感影像数据,进行图像之间的融合技术来提取棉花信息,可有效地提高农作物判读与分类精度,从而提高棉花面积监测的精确性。

参考文献:

- [1] 马众模, 娄 径, 裴志远, 等. 长江流域棉花种植面积变化遥感监测——以安庆市和池州市为例[J]. 合肥工业大学学报(自然科学版), 2005, 28(7): 738-742.
- [2] LI Min, ZHAO Gengxing, CAI Mingqing, et al. Extraction of cotton information using remote sensing and precision management zoning at county scale[J]. Journal of Natural Resources, 2012, 27(11): 1971-1980.
- [3] Zhao Dehua, Huang Liangmei, Li Jianlong, et al. A comparative analysis of broadband and narrowband derived vegetation indices in predicting LAI and CCD of a cotton canopy[J]. ISPRS Journal of Photogrammetry & Remote Sensing, 2007, 62: 25-33.
- [4] 伍维模, 王家强, 牛建龙, 等. 基于 CBERS-02 卫星遥感图像的棉花光谱识别模型的建立[J]. 塔里木大学学报, 2009, 21(2): 41-50.
- [5] 张小平, 曹卫彬, 刘姣娣. 基于遥感影像的棉花种植面积提取方法研究[J]. 安徽农业科学, 2011, 39(7): 4226-4228.
- [6] 曹卫彬, 杨邦杰, 宋金鹏. 基于 Landsat TM 图像棉花面积提取中线状地物的扣除方法研究[J]. 农业工程学报, 2004, 20(2): 164-167.
- [7] 柏军华, 李少昆, 吴洪永, 等. 基于 LANDSAT-5 像元尺度的棉田生长密度遥感监测初步研究[J]. 中国农业科学, 2009, 42(4): 1197-1206.
- [8] 曹卫彬, 刘姣娣, 马 蓉. 新疆棉花遥感监测识别区域的划分[J]. 农业工程学报, 2008, 24(4): 172-176.
- [9] 韩兰英, 陈全功, 韩 涛, 等. 基于 3S 技术的棉花面积估测方法研究[J]. 干旱区研究, 2008, 25(2): 207-211.
- [10] 刘姣娣, 曹卫彬, 刘 学, 等. 棉花遥感识别的混合像元分解[J]. 农业工程学报, 2011, 27(6): 182-186.
- [11] 黄 青, 王利民, 滕 飞. 利用 MODIS-NDVI 数据提取新疆棉花播种面积信息及长势监测方法研究[J]. 干旱地区农业研究, 2011, 29(2): 213-217.
- [12] 潘 伟, 杨德刚, 杨 莉, 等. 新疆棉花种植面积的时空变化及适度规模研究[J]. 中国生态农业学报, 2011, 19(2): 415-420.
- [13] 刘 璠, 田庆久, 黄 彦, 等. 利用多时相 HJ 卫星 CCD 遥感影像提取嘉祥县秋收作物[J]. 遥感信息, 2012, (2): 67-70.
- [14] 欧文浩, 苏 伟, 薛文振, 等. 基于 HJ-1 卫星影像的三大农作物估产最佳时相选择[J]. 农业工程学报, 2010, 26(11): 176-182.
- [15] 王 琼, 王克如, 李少昆, 等. HJ 卫星数据在棉花种植面积提取中的应用研究[J]. 棉花学报, 2012, 24(6): 503-510.
- [16] 满苏尔·沙比提, 买买提·阿扎提. 渭干河-库车河三角洲绿洲主要作物不同生育期需水量估算[J]. 干旱区研究, 2009, 26(1): 102-108.
- [17] Breiman L, Friedman J H, Olshen R A, et al. Classification and Regression trees[M]. Monterey, California, USA: Wadsworth International Group, 1984: 1-358.
- [18] 姜丽华, 杨晓蓉. 基于决策树分类技术的遥感影像分类方法研究[J]. 农业网络信息, 2009, (10): 34-42.
- [19] 赵 萍, 傅云飞, 郑刘根, 等. 基于分类回归树分析的遥感影像土地利用/覆被分类研究[J]. 遥感学报, 2005, 9(6): 708-716.
- [20] 马 丽, 顾晓鹤, 徐新刚, 等. 地块数据支持下的玉米种植面积遥感测量方法[J]. 农业工程学报, 2009, 25(8): 147-151.
- [21] 赵丽花, 李卫国, 杜培军. 基于多时相 HJ 卫星的冬小麦面积提取[J]. 遥感信息, 2011, (2): 41-50.

(上接第 126 页)

- [7] 白秀梅, 卫正新, 郭汉清. 起垄覆膜微集雨技术对玉米生长发育及产量的影响[J]. 山西水土保持科技, 2007, 6(2): 12-15.
- [8] 张 蔚, 蒋志荣, 陈 锋, 等. 地膜覆盖法调节水热状况的研究进展[J]. 安徽农业科学, 2008, 36(19): 8184-8186.
- [9] Cook H F, Valdes Gerardo S B, Lee H C. Mulch effects on rainfall interception, soil physical characteristics and temperature under *Zea mays* L. [J]. Soil and Tillage Res., 2006, 91: 227-235.
- [10] Unger P W. Role of mulchs in dryland agriculture[C]//Gupta U S. Crop physiology. New Delhi: Oxford and IBH, 1975: 237-260.
- [11] Gelmond H. Problems in crop seed germination[C]//Gupta U S. Crop physiology. New Delhi: Oxford and IBH, 1978: 7-15.
- [12] Boatwright G O, Ferguson H, Sims J R. Soil temperature around the crown node influences early growth, nutrient uptake, and nutrient translocation of spring wheat[J]. Agron. J., 1976, 68: 227-281.
- [13] Halse J N, Weir R N. Effect of temperature on spikelet number in wheat[J]. Aust. J. Agric. Res., 1974, 25: 687-695.
- [14] Rawson H M, Bagga A K. Influence of temperature between floral initiation and flag-leaf emergence on grain number in wheat[J]. Aust. J. Plant Physiol., 1979, 6: 391-400.
- [15] Chaudhary T N, Chhpra U K. Effect of soil covers on growth and yield of irrigated wheat planted at two dates[J]. Field Crops Res, 1983, 6: 293-304.
- [16] Zaongo C G L, Wendt C W, Lascano R J, et al. Interactions of water, mulch and nitrogen on Sorghum in Niger[J]. Plant and Soil, 1997, 197: 119-126.