

doi: 10.7621/cjarrp.1005-9121.20160804

· 技术方法 ·

# 基于 SAR 遥感的北方旱地秋收作物识别研究\*

东朝霞<sup>1,2</sup>, 王迪<sup>1,2\*</sup>, 周清波<sup>1,2</sup>, 陈仲新<sup>1,2</sup>, 刘佳<sup>1,2</sup>

(1. 中国农业科学院农业资源与农业区划研究所, 北京 100081; 2. 农业部农业信息技术重点实验室, 北京 100081)

**摘要** 在我国北方旱地秋收作物生长关键期, 云雨天气影响较大, 无法及时、有效地获取光学遥感数据, 因此利用雷达遥感进行旱地作物识别研究非常必要。文章以河北省衡水市为研究区, 选择 6 期 RADA-RAST-2 全极化影像作为数据源, 分类方法为随机森林法。首先通过对比不同时相间的组合结果, 优选出了研究区典型秋收作物(玉米、棉花)的最佳识别时相及组合方式。其次, 提取最优识别时相的后向散射信息、纹理信息、极化分解等 3 部分信息, 依据信息间相互组合的结果及随机森林算法对变量的重要性评价, 文中对上述 3 部分信息进行了重要性评估。结果表明: 利用 SAR 识别旱地秋收作物时应着重关注作物生长前期的时相, 其中玉米在 6 月 27 日单一时相下就可获得 90% 以上的高精度; 棉花面积小、地块破碎, 但通过 6 月 3 日与 6 月 27 日两个时相的结合也得到了 70% 以上的精度。在玉米识别中极化信息具有较大的贡献, 极化变量的加入主要增加了玉米和建筑用地的可分离性, 与单纯利用后向散射信息分类相比精度提高了近 7%; 同样, 纹理信息和极化分解信息的加入也使棉花的精度提高了 3%。最后, 利用随机森林算法对变量的重要性评价机制, 优选出对玉米识别最为重要的 5 个变量, 依次为: VH、Alpha、Yamaguchi4-Odd、Freeman-Vol 和 Mean (HV)。该研究利用雷达数据进行旱地作物识别, 验证了雷达影像对旱地秋收作物的识别能力, 不仅保证了数据获取与天气状况的独立性, 还凭借 SAR 独有的数据获取方式, 为光学数据提供了补充。

**关键词** 合成孔径雷达 (SAR) 旱地农作物 分类 随机森林 变量优选

中图分类号: S127 文献标识码: A 文章编号: 1005-9121 [2016] 08-0027-10

## 0 引言

农作物识别是农作物农情监测的初始阶段和关键步骤, 及时准确的监测农作物种类的空间分布是农情信息监测的基础。随着遥感技术的发展和遥感数据的普及, 卫星遥感技术以其宏观、动态、及时、准确等优点成为许多农作物监测的重要手段<sup>[1-3]</sup>, 尤其是应用光学遥感数据进行作物识别已经形成了一套成熟的技术方法, 其作物识别精度已达到较高水平, 理论和技术也已经非常成熟, 然而在实际应用中发现, 在我国北方地区, 多云雨的气候条件限制了光学遥感对作物的监测, 很难在作物生长期内获得完整、连续的观测数据, 致使不能满足农作物类型、面积及其空间分布等监测业务的需求。合成孔径雷达 (synthetic aperture radar, SAR) 的出现, 使农作物监测不受云、雾、雨的影响, 保证了数据获取与当地天气的独立性, 并且微波遥感探测植被信息时, 可获得与光学传感器完全不同的信息, 鉴于以上优点, SAR 被广泛地应用于作物监测与估产<sup>[4-9]</sup>。其中利用 SAR 监测南方水稻的研究最多<sup>[10-13]</sup>, 原因是水稻生长的季节往往为多云雨天气, 很难获得光学遥感数据, 并且由于水稻的介电常数较高, 相对于其他共生植物, 易将水稻和其他地物进行区分。当前国内外的相关研究主要集中在水稻长势监测和水稻制图的研究<sup>[14-16]</sup>, 监测的农作物类型过于单一, 对旱地作物 (小麦、玉米、棉花等) 的识别与面积监测研究相对较少<sup>[17]</sup>。McNa-

收稿日期: 2015-12-23

作者简介: 东朝霞 (1989—), 女, 河北石家庄人, 硕士研究生。研究方向: 农作物遥感监测。\*通信作者: 王迪 (1977—), 男, 辽宁鞍山人, 副研究员。研究方向: 农业遥感基础与应用。Email: wangdicaas@126.com

\* 资助项目: 国家科技重大专项项目“高分农业遥感监测与评价示范系统”(09-Y30B03-9001-13/15)

rin<sup>[18]</sup>使用机载 C 波段极化数据对苜蓿、玉米、大豆和小颗粒谷物(如小麦等)进行了识别研究。Hong Gang 等<sup>[19]</sup>通过融合 MODIS 和 RADARSAT-2 遥感影像对草地和苜蓿进行区分,结果表明光学遥感影像和雷达遥感影像的联合应用可有效地提高作物识别精度。华国强等<sup>[20]</sup>利用 Radarsat-2 全极化影像数据,着重分析了玉米后向散射特征,并利用不同地物后向散射特征的差异提取玉米种植分布情况,精度达到 84%。Jia Kun 等<sup>[21]</sup>利用 ENVISAT-ASAR 和 Terra-SAR 数据,对我国北方旱地作物冬小麦和棉花进行识别和分类,结果表明雷达数据可作为旱地作物识别的有效遥感数据源。这些研究都表明了雷达遥感数据在旱地作物识别方面的可行性和潜力,然而旱地作物在其全生长期内没有水层覆盖,其介电常数与周围共生植被差异不明显,再加上复杂的种植结构,雷达遥感在旱地作物识别中具有一定的挑战性。旱地作物在我国分布极广,及时准确地掌握旱地作物的长势状况和种植面积信息是保障国家粮食安全的基础。文章通过系统的分析雷达对旱地秋收作物(玉米、棉花)的识别能力,掌握了玉米和棉花的最优识别时相及特征变量组合,为数据的获取和信息的利用情况给予了参考,为光学数据提供了补充。

针对我国北方旱地作物在关键生长期内受云雨天气影响,进而无法及时、有效获取光学数据的现状,该文使用覆盖作物全生育期的 6 期 Radarsat-2 全极化数据探究雷达数据在北方旱地秋收作物(玉米和棉花)识别中的适用性。该文主要包括 3 部分:(1)通过对研究区内 5 类典型地物后向散射系数进行统计,分析玉米、棉花与其他地类间的混淆程度;(2)通过对单极化、单时相及时相间的不同组合进行分析,探究玉米、棉花的最佳识别时相及组合方式;(3)对比分析后向散射信息、纹理信息和极化分解信息,探究不同信息组合对结果的贡献,并通过变量的优选改善了分类效率。

## 1 研究区和数据

### 1.1 研究区概况

研究区位于河北省衡水市(图 1),地处河北平原中部,介于东经 115°21′~115°50′,北纬 37°42′~38°11′之间,是中国北方典型的旱地作物种植区。地属北温带季风气候,全年大于或等于 0℃的积温为 4 863℃,无霜期 200 d 左右,日照时数 2 563 h,年平均气温 12.8℃,年降水量在 500 mm 左右,降雨主要集中在 6~8 月份。研究区面积为 25 km×25 km,主要土地利用类型为农业用地,作物熟制为一年一熟或一年两熟制,秋收作物主要是玉米和棉花。玉米于 6 月初种植,10 月初收获,棉花为 4 月份种植,10 月下旬收获,玉米占绝大部分,棉花的比例较少。研究区主要覆盖地物类型共 5 类,分别是玉米、棉花、树林、水体和建筑用地。

### 1.2 数据准备

结合玉米和棉花的物候特征,选取了 2014 年 6 月 3 日至 2014 年 10 月 1 日共 6 期精细全极化(Fine-Quad polarization, FQ)模式 Radarsat-2 数据,覆盖了研究区内玉米和棉花的整个生育期。该模式数据标称幅宽为 25 km,重访周期 24 d,分辨率为 5.2 m×7.6 m(距离向×方位向),入射角约为 18°~49°,4 种极化方式即 HH、HV、VH 和 VV。表 1 为雷达数据的具体参数。为全面分析玉米、棉花与研究区其它地类的关系,将研究区的土地覆盖类型归结为 5 类:玉米、棉花、建筑用地、树林和 水体,并于 2014 年 6 月利用差分 GPS 实地采集了玉米、棉花、水体的样方边界和地理位置坐标数据并详细记录了其覆盖类型。与此同时,记录了 30 个地面控制点,用于影像精校正,其均匀分布在研究区内。为减少工作量且考虑到建筑物和树林随时间变动较小,研究获取了 2013 年 5 月份的 GF-1 号光学数据,通过多光谱和全色图像的融合得到了 2 m 分辨率的影像,建筑用地和树林样方的选取利用 GF-1 号光学数据作为辅助。研究共选择了 225 个典型地物样本,其中 164 个为训练样本,61 个为验证样本。

### 1.3 数据处理

#### 1.3.1 数据预处理

GF-1 号光学数据的处理过程包括图像拼接、感兴趣裁剪、投影转换和几何精校正等。其中几何精校正利用地面实测的 30 个控制点进行,保证误差控制在一个像元之内,满足后续研究的需要。

表 1 研究区雷达遥感数据主要参数

获取时间	入射角 (°)	极化	分辨率 (m)	产品级别	玉米生育期	棉花生育期	标注时间
2014-06-03	38.51	全极化	5.2×7.6	SLC	—	苗期	6/03
2014-06-27	38.51	全极化	5.2×7.6	SLC	拔节前期	蕾期前期	6/27
2014-07-21	38.51	全极化	5.2×7.6	SLC	抽穗前期	蕾期后期	7/21
2014-08-14	38.51	全极化	5.2×7.6	SLC	抽穗后期	花铃前期	8/14
2014-09-07	38.51	全极化	5.2×7.6	SLC	乳熟期	花铃后期	9/07
2014-10-01	38.51	全极化	5.2×7.6	SLC	成熟期	吐絮期	10/01

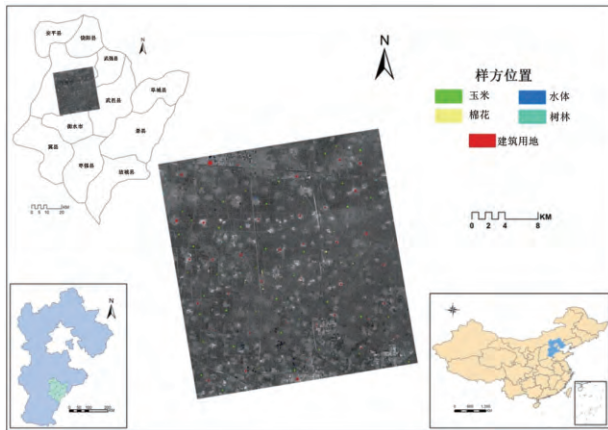


图 1 研究区位置

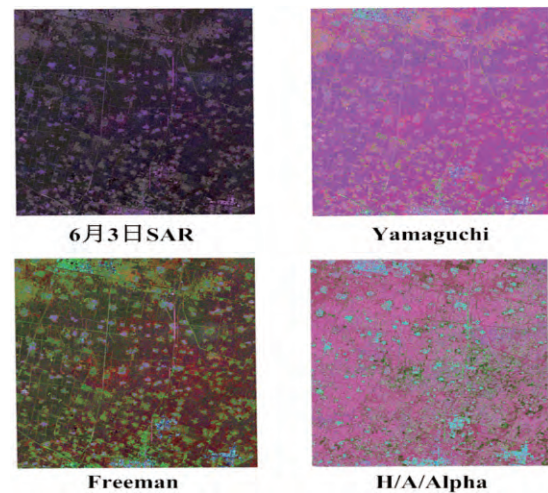


图 2 3种极化分解

SAR 数据预处理过程包括辐射定标、几何校正、影像配准、噪声滤波 (LEE 自适应滤波算法, 窗口大小为  $5 \times 5$ ) 等, 雷达数据的预处理过程主要在 NEST 中进行。影像间相互配准是多时相 SAR 图像应用中必不可少的处理步骤之一, 该研究中影像配准以 GF-1 光学数据作为基准影像, 将 SAR 数据匹配到光学数据中, 因光学影像参与了房屋和树林地面样方数据的选取, 所以需要将其与雷达数据进行配准和坐标系的统一。雷达数据经处理后, 空间分辨率为 8m。

### 1.3.2 纹理信息

纹理是遥感图像上的重要信息和基本特征, 是进行图像分析和图像理解的重要信息源<sup>[22]</sup>。分类过程中纹理信息的加入已经成为一种重要的提高遥感影像分类精度的手段, 因为纹理信息可以帮助抑制异物同谱和同物异谱现象的发生。许多研究也表明了纹理信息可以提高图像识别的精确性<sup>[23-24]</sup>。文中纹理信息的提取利用了 Haralick 提出的灰度共生矩阵 (GLCM), 它是一种最常见和广泛应用的纹理统计分析方法<sup>[25]</sup>。纹理提取过程中窗口大小设置为  $3 \times 3$ , 灰度量级级别为 64, 共有 8 个基于二阶矩阵的纹理滤波, 这些滤波包括: 均值 (Mean)、方差 (Variance)、协同性 (Homogeneity)、对比度 (Contrast)、相异性 (Dissimilarity)、信息熵 (Entropy)、二阶矩 (Second Moment) 和相关性 (Correlation)。

### 1.3.3 极化信息

全极化合成孔径雷达可测量观测目标的全散射矩阵, 全散射矩阵包含着所观测目标的重要信息。利用目标分解理论对全极化数据进行目标散射机制的解译, 可进一步分析物体的物理和几何特征<sup>[26-28]</sup>。首先将地物回波的复杂散射过程分解为几种互不相关的单一的散射分量, 由于这些单一的散射分量都对应于具有不同物理和几何特征以及分布特征的地物, 从而提供了更加丰富的地表覆盖信息, 这样就很大程度地改善了地物信息的分类精度<sup>[29]</sup>。

将散射矢量矩阵  $k$  与其共轭转置矢量矩阵  $k^*$  求外积, 可得到三维极化相干矩阵  $T_3$ , 如式 (1) 所示:

$$T_3 = \langle k \cdot k^* \rangle = \frac{1}{2} \begin{bmatrix} \langle |S_{HH} + S_{VV}|^2 \rangle & \langle (S_{HH} + S_{VV})(S_{HH} - S_{VV})^* \rangle & 2 \langle (S_{HH} + S_{VV})(S_X^*) \rangle \\ \langle (S_{HH} - S_{VV})(S_{HH} + S_{VV})^* \rangle & \langle |S_{HH} - S_{VV}|^2 \rangle & 2 \langle (S_{HH} - S_{VV})S_X^* \rangle \\ 2 \langle S_X(S_{HH} + S_{VV})^* \rangle & 2 \langle (S_X(S_{HH} - S_{VV})^0) \rangle & 4 \langle |S_X|^2 \rangle \end{bmatrix} \quad (1)$$

式中,  $S_X = S_{HV} = S_{VH}$ ;  $S_{HH}$ 、 $S_{VV}$ 为同极化分量;  $S_{HV}$ 、 $S_{VH}$ 为交叉极化分量。

在 PolSARPro - v4. 2. 0 软件中利用 T3 矩阵进行分解, 分解方法选择了 Cloude-Pottier 分解<sup>[30]</sup>, Freeman 分解 (FM)<sup>[31]</sup>以及 Yamaguchi 分解 (YG)<sup>[32]</sup>。图 2 给出了不同分解方式下的图像表达。

### 1. 3. 4 分类方法

随机森林法 (Random Forest, RF) 是由美国科学家 Leo Breiman Ho<sup>[33]</sup>于 2001 年联合发表的一种新的机器学习算法。该算法以 K 个决策树为基本分类器, 随机森林输出的分类结果由每个决策树的分类结果简单投票决定<sup>[34]</sup>。其已被成功的应用到土地覆盖制图中<sup>[35]</sup>。随机森林算法有很多优点: 1) 对于很多种资料, 它是可以产生高准确度的分类器; 2) 它可以处理大量的输入变量; 3) 它可以在决定类别时, 评估变量的重要性; 4) 在大的、高维数据训练时, 不容易出现过拟合而且速度较快。RF 分类过程在 EnMAP Box 软件<sup>[36]</sup>中实现, 该软件包含多个内置分类算法, 其中就包括随机森林算法。

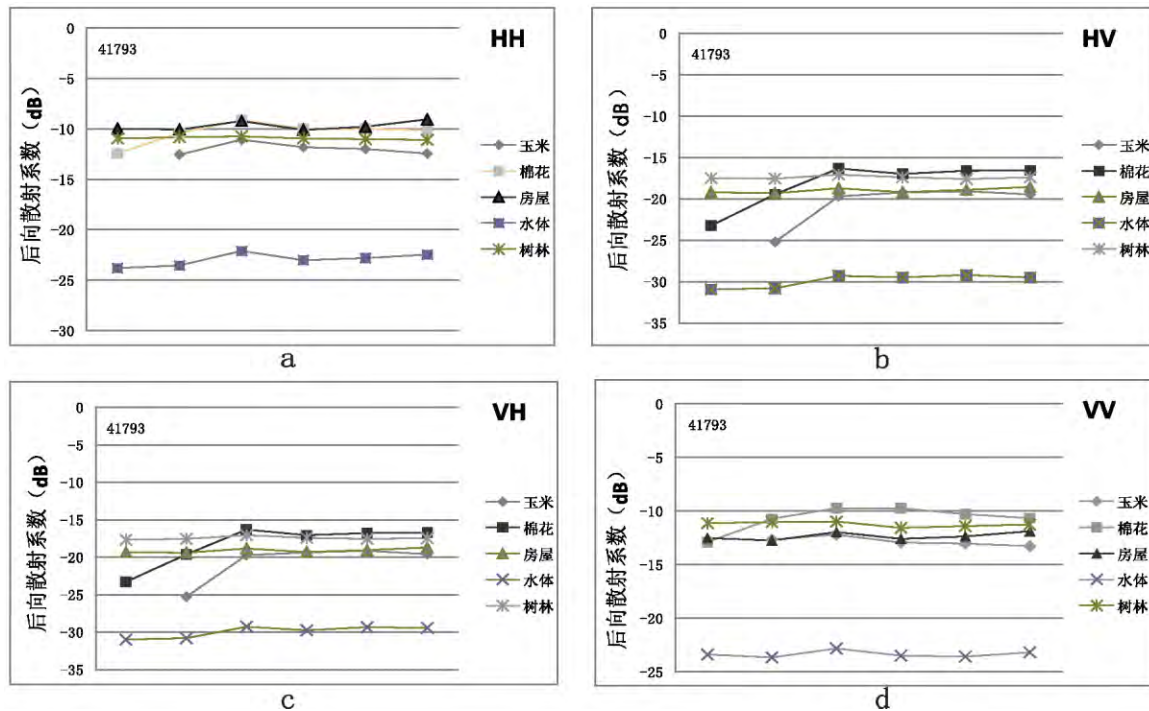


图 3 研究区典型地物不同极化雷达后向散射特征 (a: HH b: HV c: VH d: VV)

表 2 单时相条件下玉米的识别结果

J-M 距离	6/27	7/21	8/14	9/07	10/01
	玉米	玉米	玉米	玉米	玉米
棉花	1.44	0.85	1.19	1.04	1.30
树林	1.84	0.60	0.52	0.50	0.85
建筑物	1.59	0.53	0.51	0.53	0.79
水体	1.92	1.82	1.88	1.80	1.81
制图精度 (%)	87.11	46.26	45.84	50.01	63.39
用户精度 (%)	95.64	51.99	48.44	47.72	58.48

表 3 单时相条件下棉花的识别结果

J-M 距离	6/3	6/27	7/21	8/14	9/07	10/01
	棉花	棉花	棉花	棉花	棉花	棉花
玉米	0.99	1.44	0.85	1.19	1.05	1.30
树林	1.66	0.51	0.37	0.41	0.26	0.21
房屋	1.17	0.79	0.96	1.05	0.83	0.87
水体	1.99	2.0	1.99	1.96	1.92	1.95
制图精度 (%)	35.14	16.89	9.46	22.64	10.14	3.72
用户精度 (%)	45.02	16.89	11.43	15.99	9.97	5.26

## 2 结果与讨论

### 2.1 玉米、棉花与其他典型地物间的混淆性分析

利用获取的训练样方数据, 从多时相全极化 SAR 数据中提取了玉米、棉花、水体、建筑用地和树林的后向散射系数, 如图 3 所示。6 月 3 日至 10 月 1 日分别对应棉花的苗期至吐絮期, 棉花的后向散射系数先后经历了先上升后稳定的过程, 6 月 3 日棉花处于出苗期, 被地膜覆盖, 裸露地表所占比重较大, 表面散射强, 此时其后向散射强度最小, 与其他地物差异最大, 随着棉花进入生长盛期, 冠层体散射占据主要的后向散射部分, 导致后向散射系数增加。玉米在 6 月 27 日交叉极化方式下与其他地物的差异最明显, 除水体外, 玉米与其他地物的差异达到 5.7dB, 易与其它地物进行区分, 经过拔节期茎叶的迅速生长, 7 月 21 日玉米

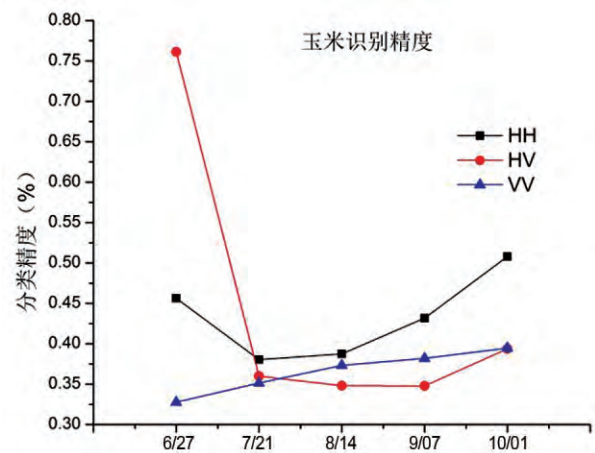


图 4 单极化对玉米识别精度的影响

的后向散射强度基本达到最大。从图 4 可以看出, 玉米和棉花的后向散射系数均在 7 月 21 日左右达到最大, 其后, 除水体外的四类地物后向散射系数集中在 3.4dB 范围之内, 互相混淆。研究区的城区高大建筑物因其独特的建筑结构和坚硬的水泥表面会形成二面角反射形成高亮区域, 易与作物区分, 但在农村地区情况却不同, 部分村庄建筑物分布稀疏且较多为砖土结构的房屋, 二面角反射不明显, 加上周围常会有蔬菜或树林的影响, 降低了整体建筑物后向散射系数的均值, 对玉米的识别形成阻碍。树林的组成包括树苗、果树和树林, 成分复杂, 其与棉花的后向散射系数接近, 在棉花的识别中会产生影响。

### 2.2 玉米、棉花最佳识别时相的优选

首先利用单一时相进行玉米和棉花的识别, 目的是探究单时相雷达数据对旱地秋收作物的识别能力及对特定地物的识别时相进行优劣排序。统计了玉米、棉花与研究区典型地物间的 J-M 距离, Jeffries-Matusita 距离<sup>[37]</sup> (J-M) 用于确定 2 个类别间的差异性程度, 这是一种定量的方式来衡量训练样本 (ROI) 的可分离性, 它不需要假定地物的正态分布, 具有较好的通用性<sup>[38]</sup>。这个参数的值在 0~2.0 之间, 越接近 1.8 说明样本之间可分离性越好。结果见表 2 和表 3。如表 2 所示, 在玉米的整个生育期中 6 月 27 日获取了最高的识别精度, 并且在该时相条件下玉米与其它四类地物的 J-M 距离均为最大, 说明在玉米的整个生长期中 6 月 27 日即玉米的拔节前期为识别的最优时相。原因是在该时期玉米处于苗期和拔节期之间, 植株普遍偏小, 且玉米播种时行距与株距本身就较大, 土壤裸露比例高, 地表结构简单, 在这种情况下, 造成交叉极化方式表现出明显区别于其它地物的特性, 从图 3 可以明显地发现, 在 6 月 27 日交叉极化条件下, 除水体外, 玉米与其它地物的差异达到了 5.7dB。7 月 21 日后玉米识别精度大幅降低, 原因是树林、建筑物对玉米的影响增大, 其 J-M 距离在 0.5 左右, 其中 8 月 14 日玉米识别效果最差, 该时期玉米与树林、建筑物的可分离度降到最低。因此, 以制图精度作为标准, 识别玉米的优势时相依次为: 6 月 27 日、10 月 1 日、9 月 7 日、7 月 21 日、8 月 14 日。从表 3 可以看出, 在覆盖棉花全生育期的 6 个时相中, 6 月 3 日即棉花的苗期取得了最高的识别精度, 该时期玉米尚未播种, 取代玉米的为处于收获期的小麦, 此时棉花植株小、覆盖率低, 大部分土壤被地膜覆盖, 由统计的 J-M 距离可以看出, 该时期棉花与其它地物的差异性基本达到最大, 说明棉花的苗期为识别的最佳时相。以分类后的制图精度作为标准, 识别棉花的优势时相依次为: 6 月 3 日、8 月 14 日、6 月 27 日、9 月 7 日、7 月 21 日、10 月 1 日。通过表 2、表 3 的对比可以发现, 利用单时相进行玉米和棉花识别时, 玉米的精度要优于棉花, 原因可能是由于玉米为研究区的主要作物类型, 棉花所占的比例很小, 给识别带来了困难。

接着, 选取玉米识别的前 3 个优势时相, 即 6 月 27 日、10 月 1 日和 9 月 7 日; 棉花识别的前 3 个优

表 4 时相组合对玉米识别的影响

J-M 距离	6/27	6/27 + 10/01	6/27 + 9/07	10/01 + 9/07	6/27 + 10/01 + 9/07
	玉米	玉米	玉米	玉米	玉米
棉花	1.44	1.76	1.67	1.55	1.84
树林	1.84	1.89	1.86	1.05	1.90
房屋	1.59	1.77	1.75	0.97	1.83
水体	1.92	1.97	1.97	1.88	1.98
制图精度 (%)	87.11	89.87	88.82	67.38	89.46
用户精度 (%)	95.64	95.71	95.97	62.33	96.09

表 5 时相组合对棉花识别的影响

J-M 距离	6/03	6/03 + 8/14	6/03 + 6/27	8/14 + 6/27	6/03 + 8/14 + 6/27
	棉花	棉花	棉花	棉花	棉花
玉米	0.99	1.52	1.73	1.72	1.85
树林	1.66	1.77	1.74	0.86	1.82
房屋	1.17	1.77	1.60	1.38	1.845
水体	1.19	2.0	2.0	2.0	2.0
制图精度 (%)	35.14	47.97	68.92	27.03	63.85
用户精度 (%)	45.02	45.08	64.76	27.49	77.78

势时相,即6月3日、8月14日和6月27日。将时相间进行组合,以优选出合适的时相组合。结果见表4、表5。表4为针对玉米识别进行的组合分析,从表中可以看出6月27日单一时相条件下玉米的识别精度达到了87.11%,识别精度最好的为6月27日与10月1日两个时相的结合,识别精度提高了约2.7%,发现通过6月27日与其他时相的结合精度并没有太大的提升,考虑到该识别精度只利用了后向散射信息,还未充分的利用极化信息和纹理信息,因此后期只选择了6月27日单一时相进行玉米的识别。从表5可以看出,6月3日与6月27日两个时相的组合使棉花获取了最优的识别精度,与最优时相6月3日的分类结果相比精度提高了近34%,在该条件下棉花与其他地物的J-M距离均在1.6以上,说明6月3日与6月27日为棉花识别的最优组合方式。

最后,为比较极化贡献的差异性,对所有时相的单极化进行了分类,由于研究区棉花面积小,地块破碎,对结果的影响因素多,该过程只对玉米进行。结果见图4。从图中可以看出在玉米整个生育期中,交叉极化对精度的贡献浮动最大,同极化变化平稳。在玉米播种前期交叉极化对精度的贡献要远优于同极化,随着生育期的推进,交叉极化的优势迅速减小,7月21日后同极化略高于交叉极化。同极化间进行比较发现,HH极化优于VV极化,且在生育期的初始期和结束期优势明显。

表 6 不同信息组合对玉米、棉花识别精度的影响

玉米	不同组合 (6/27)				棉花	不同组合 (6/03 + 6/27)			
	1	2	3	4		1	2	3	4
后向散射系数	*	*	*	*	后向散射系数	*	*	*	*
纹理信息		*		*	纹理信息		*		*
极化信息			*	*	极化信息			*	*
制图精度 (%)	87.11	89.18	93.98	93.68	制图精度 (%)	68.92	69.93	67.23	71.96
用户精度 (%)	95.64	98.29	99.06	99.19	用户精度 (%)	64.76	84.84	82.23	85.89

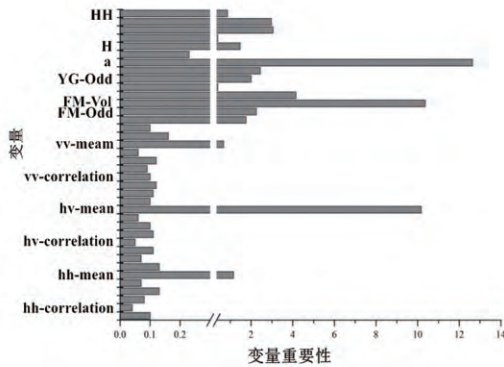


图 5 玉米变量重要性评价

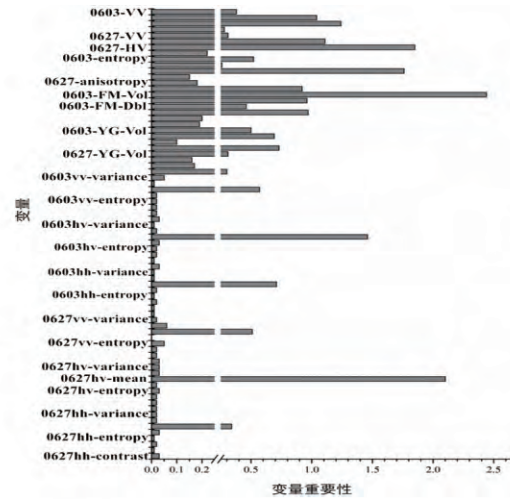


图 6 棉花变量重要性评价 (0603 为 6 月 3 日; 0627 为 6 月 27 日)

表 7 组合 1 (玉米) 精度表

	建筑物	棉花	树林	水体	玉米
建筑物	7 131	200	1 467	1	592
棉花	84	56	100	0	42
树林	98	33	238	0	5
水体	0	0	0	910	23
玉米	192	7	10	7	4 829

玉米制图精度 87.11% 用户精度 95.64%

表 8 组合 3 (玉米) 精度

	建筑物	棉花	树林	水体	玉米
建筑物	7 302	73	885	0	213
棉花	69	137	134	0	116
树林	94	80	793	0	0
水体	0	0	0	918	7
玉米	40	6	3	0	5 155

玉米制图精度 93.98% 用户精度 99.19%

### 2.3 纹理和极化信息的重要性分析

根据上述分析,选择 6 月 27 日单一时相进行玉米的识别,棉花的识别时相确定为 6 月 3 日与 6 月 27 日 2 个时相的结合。以上分类过程只利用了后向散射信息,下面将在后向散射信息的基础上添加纹理和极化信息,探究二者的加入对分类精度的影响,结果见表 6。从表 6 可以看出,纹理信息与极化信息的加入使玉米的识别精度有了较明显的提升,其中纹理信息的加入使精度由原来的 87.11% 增加到 89.18%,增幅为 2.07%,极化信息的加入使精度由 87.11% 增加到 93.98%,增幅达 6.87%,说明纹理信息和极化信息都在一定程度上提高了玉米的识别精度,相比之下,在玉米识别过程中极化信息的重要性要优于纹理信息。表 7、表 8 分别为组合 1 和组合 3 的具体精度,通过二者的比较发现,在玉米识别过程中建筑用地对其影响最大,而极化信息的加入主要减少了玉米和建筑用地的混淆,进而提升了玉米的识别精度。组合 4 将后向散射系数、纹理信息和极化信息进行结合后分类,其分类效果与组合 3 类似,表明在玉米识别过程中后向散射信息与极化信息结合就能获得理想的结果。对于棉花,从表中可以看出纹理信息和极化信息的加入均起到了积极的作用,制图精度没有明显改善,用户精度提高了近 20%。通过组合 2 和组合 3 的比较可知在棉花的识别过程中纹理信息的作用略优于极化信息,组合 2 相较于组合 3 而言制图精度和用户精度均有 2.6% 以上的提升,组合 4 为 3 部分数据的结合,其获取了最高了识别精度,说明在棉花的识别中,纹理信息和极化信息对于精度的提升都起到了积极的促进作用。图 5、图 6 分别为对玉米、棉花识别过程中所有变量的重要性评价结果,二者趋势一致,从图中可以看出,后向散射信息和极化信息的重要性都要远优于纹理信息,除均值信息外所有的纹理信息对结果的贡献均较小,在玉米识别过程中贡献度最大的 3 个变量依次为  $\alpha$ 、FM - Vol、HV - Mean; 棉花识别过程中最优的 3 个变量依次为 0627-HV、0603-FM

- Vol、0627-HV - Mean。将玉米识别的组合 3 与棉花识别的组合 4 相互结合，得到最终分类结果，见图 7。

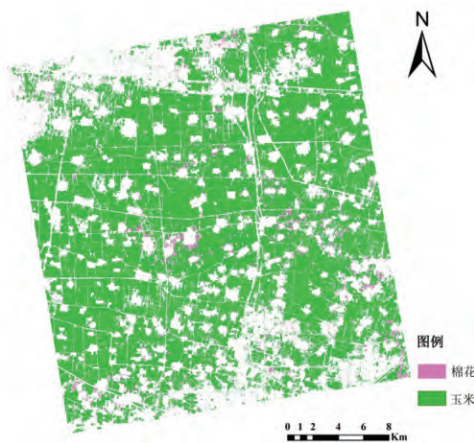


图 7 研究区玉米、棉花分布

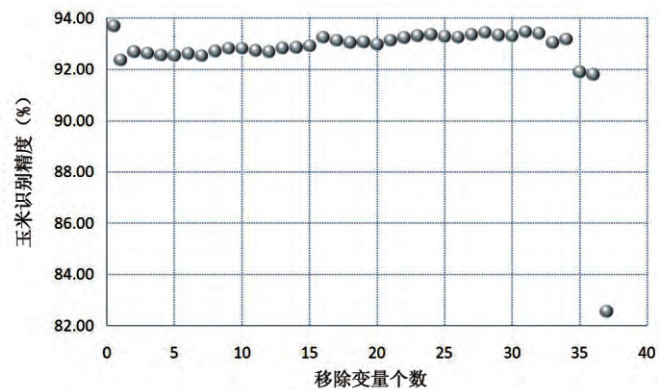


图 8 玉米变量优选

玉米的识别利用了 6 月 27 日单一时相，加上纹理信息和极化信息，特征数为 38 个（图 5 是对不同变量重要性的评价结果）；棉花的识别利用了两个时相的结合，其特征数共 76 个。理想情况下，每个变量都可提供额外的信息以提高分类精度，然而在机器学习的实际应用中，特征数量往往较多，其中可能存在不相关的特征，特征之间也可能存在相互依赖，特征个数的增加，会导致分析特征、训练模型所需的时间长，且容易引起“维度灾难”；除此之外，多源数据的加入必然会引起因配准误差而产生的噪声。因此，该文将探究在玉米识别的多特征基础上，通过特征选择剔除不相关和冗余信息，以提高计算效率。由于研究区棉花面积小，地块破碎，受配准误差的影响要远大于玉米，特征的优选只针对玉米进行。为进行变量的优选，研究主要利用随机森林算法对变量的重要性评价机制，通过迭代依次除去重要性最小的变量，再对剩余的变量进行分类并进行精度评价，以此类推。结果见图 8。在移除 33 个变量之前，玉米的识别精度浮动微小，剩余 5 个变量时精度迅速降低，这 5 个变量分别为：VH、Alpha、Yamaguchi4-Odd、Freeman-Vol、Mean (HV)。

### 3 结论

该文以河北省衡水市为研究区，获取了覆盖旱地秋收作物（玉米和棉花）全生育期的 6 期影像，系统分析了玉米、棉花在各个生长时期的后向散射系数，并结合随机森林算法对玉米和棉花进行了识别，验证了雷达数据在旱地作物识别中的可行性，并为光学数据的获取提供了补充。结论如下：

(1) 研究利用全生育期的单极化数据对玉米进行识别，结果发现在不同生育期 4 种极化方式对结果的影响存在差异。拔节期之前，交叉极化方式在玉米的识别中贡献最为突出；全生育期内 HH 极化优于 VV 极化，且在生育期的初始和结束期优势明显。

(2) 研究利用单一时相及时相间的组合对玉米和棉花进行了识别，结果表明 6 月 27 日即玉米的拔节前期为玉米识别的最优时相，6 月 3 日与 6 月 27 日为棉花识别的最优时相组合。由于作物本身物候期的差异，导致在作物耕作的交替期往往会表现出极大的差异性，因此利用雷达数据识别旱地秋收作物时应重点关注作物生长前期的时相。

(3) 研究对后向散射信息、纹理信息、极化分解信息等 3 部分信息进行了不同方式的组合，并针对玉米进行了变量优选，结果表明在北方地区玉米识别过程中，极化信息贡献度较大，其在后向散射信息分类结果的基础上精度提高了近 7%，极化信息的加入主要增加了玉米和建筑用地的可分离性。纹理信息和极化信息的加入也使棉花的精度提高了 3%。通过针对玉米的变量优选，其中 5 个变量的组合可使玉米获



得较高的精度, 然而所有信息结合获取的结果仍为最优。

该文利用雷达数据对华北平原地区的旱地作物进行识别研究可能还存在一定不足: 研究区内旱地作物品种单一(大多数农田农作物类型仅包括玉米、棉花两类), 且棉花面积小、地块破碎, 今后的研究将在作物类型复杂的地区进一步深入。

## 参考文献

- [1] Gonzalez-Sanpedro M C, Le Toan T, Moreno J, et al. Seasonal variations of leaf area index of agricultural fields retrieved from Landsat data. *Remote Sensing of Environment*, 2008, 112: 810 ~ 824
- [2] Wang Dan Jin Hiu, Chen Jinsong, et al. Application of multi-temporal ENVISAT ASAR data to agricultural area mapping in the Pearl River Delta. *International Journal of Remote Sensing*, 2010, 31: 1555 ~ 1572
- [3] 李志鹏, 李正国, 刘珍环, 等. 基于中分辨 TM 数据的水稻提取方法对比研究. *中国农业资源与区划*, 2014, 35 (1): 27 ~ 33
- [4] Liu Chen Shang Jiali, Vachon P W, et al. Multiyear crop monitoring using Polarimetric RADARSAT-2 Data. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 51 (4): 2227 ~ 2240
- [5] 邵芸, 郭华东, 范湘涛, 等. 水稻时域散射特征分析及其应用研究. *遥感学报*, 2001, 5 (5): 340 ~ 345
- [6] Le Toan T Laur H, Mougin E, et al. Multitemporal and dual-polarization observations of agricultural vegetation covers by X-band SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 1989, 27 (6): 709 ~ 718
- [7] 谭炳香, 李增元. SAR 数据在我国南方水稻分布图快速更新中的应用研究. *国土资源遥感*, 2000, 43 (20): 24 ~ 27
- [8] 李坤, 邵芸, 张风丽. 基于 RadarSat-2 全极化数据的水稻识别. *遥感技术与应用*, 2012, 27 (1): 86 ~ 93
- [9] 丁娅萍, 陈仲新. 基于最小距离法的 RADARSAT-2 遥感数据旱地作物识别. *中国农业资源与区划*, 2014, 35 (6): 79 ~ 84
- [10] 董彦芳, 孙国清, 庞勇. 基于 ENVISAT ASAR 数据的水稻监测. *中国科学 D 辑*, 2005, 35 (7): 682 ~ 689
- [11] 凌飞龙, 汪小钦, 史晓明. 多时相 SAR 图像水稻分布信息提取方法研究. *福建师范大学学报(自然科学版)*, 2007, 23 (3): 15 ~ 19
- [12] 汪小钦, 王钦敏, 史晓明, 等. 基于主成分变换的 ASAR 数据水稻种植面积提取. *农业工程学报*, 2008, 24 (10): 122 ~ 126
- [13] 杨沈斌, 李秉柏, 申双和, 等. 基于多时相多极化差值图的稻田识别研究. *遥感学报*, 2008, 12 (4): 614 ~ 618
- [14] 张萍萍, 申双和, 李秉柏, 等. 水稻极化散射特征分析及稻田分类方法研究. *江苏农业科学*, 2006, 1: 148 ~ 151
- [15] 杨沈斌, 李秉柏, 申双和, 等. 基于 ENVISAT ASAR 数据的水稻遥感监测. *江苏农业学报*, 2008, 24 (1): 33 ~ 38
- [16] Haldar D Patnaik C. Synergistic use of Multi-temporal Radarsat Sar and AWiFS data for Rabi rice identification. *J. Indian Soc. Remote Sens*, 2010, 38: 153 ~ 160
- [17] 王迪, 周清波, 陈仲新, 等. 基于合成孔径雷达的农作物识别研究进展. *农业工程学报*, 2014, 30 (16): 203 ~ 212
- [18] McNairn H Van der Sanden J J, Brown R J, et al. The potential of RADARSAT-2 for crop mapping and assessing crop condition // Proceedings of the Second International Conference on Geospatial Information in Agriculture and Forestry. Lake Buena Vista, FL, 2000, 2: 81 ~ 88
- [19] Gang Hong Zhang Aiming, Zhou Fuqun, et al. Integration of optical and synthetic aperture radar (SAR) images to differentiate grassland and alfalfa in prairie area. *International Journal of Applied Earth Observation and Geoinformation*, 2014, 28: 12 ~ 19
- [20] 化国强, 肖靖, 黄晓军, 等. 基于全极化 SAR 数据的玉米后向散射特征分析. *江苏农业科学*, 2011, 39 (3): 562 ~ 565
- [21] Jia Kun Li Qiangzi, Tian Yichen, et al. Crop classification using multi-configuration SAR data in the North China Plain. *International Journal of Remote Sensing*, 2012, 33 (1/2): 170 ~ 183
- [22] 舒宁. 关于多光谱和高光谱影像的纹理问题. *武汉大学学报(信息科学版)*, 2004, 29 (4): 292 ~ 295
- [23] 张顺谦, 杨秀荣. 神经网络和分形纹理在夜间云雾分离中的应用. *遥感学报*, 2006, 10 (4): 497 ~ 502
- [24] Shaban M A Dikshit O. Improvement of classification in urban area by the use of textural features: The case study of lucknow city, Uttar Pradesh. *International Journal of Remote Sensing*. 2001, 22: 565 ~ 593
- [25] 刘龙飞, 陈云浩, 李京. 遥感影像纹理分析方法综述与展望. *遥感技术与应用*, 2003, 18 (6): 441 ~ 448
- [26] Armando Marino Irena Hajnssek. A change detector based on an optimization with polarimetric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52 (8): 4781 ~ 4798
- [27] V Alberga. A study of land cover classification using polarimetric SAR parameters. *International Journal of Remote Sensing*, 2007, 28 (17): 3815 ~ 3870
- [28] Stefan Uhlmann Serkan Kiranyaz. Integration color features in polarimetric SAR image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52 (4): 2197 ~ 2216
- [29] 陈劲松, 邵芸, 李震. 基于目标分解理论的全极化 SAR 图像神经网络分类方法. *中国图象图形学报*, 2004, 9 (5): 552 ~ 557
- [30] Cloude S R Pottier E. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 1997, 35 (1): 68 ~ 78

- [31] Freeman A ,Durde S L. A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing* , 1998 ,36 ( 3) : 963 ~973
- [32] Yamaguchi Y ,Moriyama T , Ishido M , et al. Four-component scattering model for polarimetric SAR image decomposition. *IEEE Transactions on Geoscience and Remote Sensing* ,2005 ,43 ( 8) : 1699 ~1706
- [33] Breiman L. Random forests. *Machine Learning* ,2001 ,45 ,5 ~32
- [34] 董师师,黄哲学. 随机森林理论浅析. *集成技术* ,2013 ,2 ( 1) : 1 ~7
- [35] Gislason P O ,Benediktsson J A , Sveinsson J R. Random Forests for land cover classification. *Pattern Recognition Letters* ,2006 ,27 ,294 ~300
- [36] Jakimow B ,Oldenburg C , Rabe A , et al. Manual for Application: image RF ( 1.1) ,2012
- [37] Richards J. A. *Remote Sensing Digital Image Analysis*. Berlin: Springer-Verlag , 1999
- [38] 凌飞龙,汪小钦,史晓明. 多时相 SAR 图像水稻分布信息提取方法研究. *福建师范大学学报 ( 自然科学版)* ,2007 ,23 ( 3) : 15 ~19

## DRYLAND CROP IDENTIFICATION BASED ON SYNTHETIC APERTURE RADAR IN THE NORTH CHINA PLAIN\*

Dong Zhaoxia<sup>1,2</sup> ,Wang Di<sup>1,2\*</sup> ,Zhou Qingbo<sup>1,2</sup> ,Chen Zhongxin<sup>1,2</sup> ,Liu Jia<sup>1,2</sup>

( 1. Key Laboratory of Agri-informatics ,Ministry of Agriculture ,Beijing 100081 ,China;

2. Institute of Agricultural Resources and Regional Planning ,Chinese Academy of Agricultural Sciences ,Beijing 100081 ,China)

**Abstract** Crop identification is the initial phase and key step of crops condition monitoring. Timely and accurately monitoring the spatial distribution of crop types is the basis of monitoring the information of farming condition. With the development of remote sensing technology and the popularization of remote sensing data , remote sensing has become an important means in agricultural monitoring. In particular , a complete set of technological method using optical remote sensing data for crop identification has already formed. However , during the key grow period of dryland crops in North China , since rainy or cloudy it is difficult to obtain timely and effective optical remote sensing data , it's impossible to ensure the acquisition of the data during the key growth period of crops by using optical remote sensing technology , thus the continuous monitoring and research of crops are hindered. Therefore , it is necessary to conduct the study of dryland crops identification by the means of Synthetic aperture radar ( SAR) , which overcomes the above restrictions due to its all - weather , all - day imaging capability. In addition , when detecting vegetation information with microwave remote sensing technology , information different from that acquired through optical remote sensing can be acquired. Based on the SAR data , this paper chose Shenzhou , Hengshui city of Hebei Province as the study area and used the RADARSAT - 2 images covering the area from June 3 to October 1 , 2014 as the study data source. Classification method was the Random Forest ( RF) . First , by contrasting the combination of different time , the optimal phase and phase combination was selected for typical dryland crop ( corn/cotton) in the study area. Second , we extracted backscattering coefficients , texture information and polarimetric decomposition from optimal phase , and combine the information in different ways. The contribution of different combination of input data to a random forest classifier was evaluated with map accuracy statistics and the evaluation of variables importance. The results showed that when identifying dryland crops , radar remote sensing should pay attention to the early phases during the growth period of crops. The accuracy can reach more than 90% for corn on condition of single phase , cotton areas were usually small and distributed sporadically , but it also reached more than 70% through the combination of June 3 and June 27. While using texture information and polarimetric decomposition it also increased 3% . Finally , variable importance was estimated following by VH、Alpha、Yamaguchi4 - Odd、Freeman - Vol and Mean(HV) . The results demonstrated the capacity of SAR can identify the dryland crops and the independence of data. SAR with unique acquisition mode provided a complement to optical data.

**Keywords** synthetic aperture radar ( SAR) ; dryland crop; classification; random forest ( RF) ; variable selection