

引用格式:Ma Tengyao, Xiao Pengfeng, Zhang Xueliang, *et al.* Recognition of Snow Cover based on Features Selection in GF-3 Fully Polarimetric Data[J]. Remote Sensing Technology and Application, 2020, 35(6): 1292-1302. [马腾耀, 肖鹏峰, 张学良, 等. 基于特征优选的 GF-3 全极化数据积雪识别[J]. 遥感技术与应用, 2020, 35(6): 1292-1302.]

doi:10.11873/j.issn.1004-0323.2020.6.1292

基于特征优选的 GF-3 全极化数据积雪识别

马腾耀, 肖鹏峰, 张学良, 马威, 郭金金

(南京大学地理与海洋科学学院, 自然资源部国土卫星遥感应用重点实验室, 江苏省地理信息技术重点实验室, 江苏 南京 210023)

摘要:以新疆阿尔泰山南麓克兰河流域典型区为研究区, 利用 GF-3 全极化数据进行积雪探测, 提出了一种基于特征优选的积雪识别方法。首先通过极化分解获取了 GF-3 数据的 22 个极化特征, 并利用随机森林方法计算各特征的重要性, 构建特征优选规则生成最优特征集, 然后基于最优特征集对积雪进行识别。分析特征的重要性发现, 同极化后向散射系数对积雪识别的贡献比交叉极化的贡献大, 面散射和体散射对积雪识别的贡献比二面角散射贡献大。将该方法与最大似然法、支持向量机、BP 神经网络 3 种分类器的对比发现, 使用最优特征集并且利用随机森林方法的积雪识别精度最高(F 指数为 0.86, 总体精度为 0.79)。结果表明: 基于特征优选进行积雪识别, 不仅使得积雪识别效率得到提高, 而且保持精度不变甚至有所增加, 证明了该方法在积雪识别中的有效性。

关键词:积雪识别; GF-3; 极化分解; 特征优选; 随机森林

中图分类号: TP181 **文献标志码:** A **文章编号:** 1004-0323(2020)06-1292-11

1 引言

积雪是冰冻圈的重要组成部分, 影响着全球气候变化和地表能量平衡^[1]。它作为宝贵的融水资源^[2], 是地表、地下径流的重要补充, 与人类活动密切相关^[3]。降雪量过大容易引发雪崩等自然灾害, 影响人类的生产生活^[4]。在全球气候变暖的背景下, 近年来新疆地区的积雪覆盖面积呈缓慢减少的趋势^[5]。而积雪减少会导致河流水资源补给量不足, 使工农业用水紧张。随着遥感技术的发展, 利用遥感图像提取积雪覆盖范围, 使及时应对洪涝和干旱灾害的发生成为了可能。

光学遥感易受天气影响, 难以获得云遮挡下的积雪信息^[6]。被动微波遥感数据的分辨率较低, 难以满足流域尺度的积雪识别要求^[7]。对于流域尺度的积雪识别, 使用对积雪参数比较敏感的合成孔径

雷达 (Synthetic Aperture Radar, SAR) 数据较为合适^[8]。

基于 SAR 数据进行积雪识别的方法较多, 大致可分为基于后向散射系数的方法、基于干涉测量的方法和基于极化分解的方法三类。第一类方法主要基于单频率、单极化数据的变化检测, 计算同一地区湿雪和无雪图像的后向散射系数比值, 再根据阈值识别湿雪^[6,9]。但阈值不具有普适性, 且该方法只针对湿雪制图, 限制了其在积雪识别中的应用。在引入多频率、多极化数据后, 可根据积雪与其它地物在不同极化方式和不同频率的后向散射系数差异识别积雪^[8,10]。第二类方法利用干涉测量技术, 根据降雪前后的相干系数差异识别积雪^[11]。第三类方法利用极化分解技术获取多种散射信息, 因不同地物在这些散射特征中具有不同表征, 可利用

收稿日期: 2019-12-27; 修订日期: 2020-11-15

基金项目: 国家自然科学基金项目 (41671344), 国家科技基础资源调查专项课题 (2017FY100502)。

作者简介: 马腾耀 (1995—), 男, 山西长治人, 硕士研究生, 主要从事积雪遥感和数字图像处理研究。E-mail: tengyaom@gmail.com

通讯作者: 肖鹏峰 (1979—), 男, 湖南宁乡人, 教授, 主要从事积雪遥感和数字图像处理研究。E-mail: xiaopf@nju.edu.cn

积雪与其他地物的表征差异识别积雪^[12];或将它们作为特征向量,利用机器学习方法识别积雪^[13-14]。

目前,若干极化分解方法已经被用于与积雪相关地物的分类^[12-14],但联合多种极化分解技术选择有效特征进行积雪识别仍有待探索。随机森林是一种多分类器集成的算法,可在分类的同时计算每一特征的重要性^[15]。根据特征重要性,结合特征优选规则即可构建最优特征集,进而进行积雪的识别。基于这两个出发点,使用5种极化分解方法提取GF-3数据的极化特征,利用随机森林方法计算特征重要性,从中筛选出最优特征集并将其用于积雪识别。

2 研究区与数据

2.1 研究区

研究区位于新疆阿尔泰山南麓、克兰河流域中部,地理范围为47°39'~47°57' N,87°41'~88°03' E,如图1所示,属于平原与山地的交错地带,区域内高程范围为590~1 625 m。西南部和东北部的地势较高,中部平原相对平坦,零散分布着少量村庄。根据30 m分辨率的土地覆盖数据可知^[16],研究区土地覆盖类型以耕地和戈壁为主,在西南部和北部有部分草地、灌木和少量乔木存在。

克兰河由北向南贯穿阿勒泰市,向西汇入额尔齐斯河。克兰河流域内低山平原区9月上中旬降雪,在翌年2月积雪量达最高值,4月积雪消融^[17]。

径流以积雪融水、降雨和地下水补给为主,其中积雪融水最高可占年径流量的45%^[17-18]。积雪在克兰河流域中扮演重要角色。

2.2 数据

研究采用国产GF-3卫星和美国Landsat 8卫星各一景影像。卫星过境时间分别为2017年11月15日0时19分和11月18日13时01分。此外,还使用了30 m分辨率的航天飞机雷达地形测绘(Shuttle Radar Topography Mission, SRTM) DEM数据,将其用于GF-3数据的地理编码和地形校正。

使用的GF-3数据是L1A级单视复数图像,工作模式为全极化模式。每景数据包括图像数据文件、有理多项式系数(Rational Polynomial Coefficients, RPC)文件、图像元数据等,记录了4种极化方式(HH、HV、VH、VV)的回波信号。图像的像元值由复数构成,表示散射回波的强度和相位信息。Landsat 8数据使用蓝、绿、红和短波红外4个波段,主要利用Landsat 8图像辅助GF-3图像上的训练和验证样本选取。

为确定卫星过境时研究区的积雪干湿状态,使用了中国地面气候资料日值数据集(V3.0)(<http://www.nmic.cn/data>)。该数据集包含了各气象站的0 cm地表温度数据。距离研究区最近的气象站是阿勒泰站。在GF-3卫星过境当日,该气象站平均地温为-2℃,最高地温为0℃,可确定研究区积雪属于干雪。

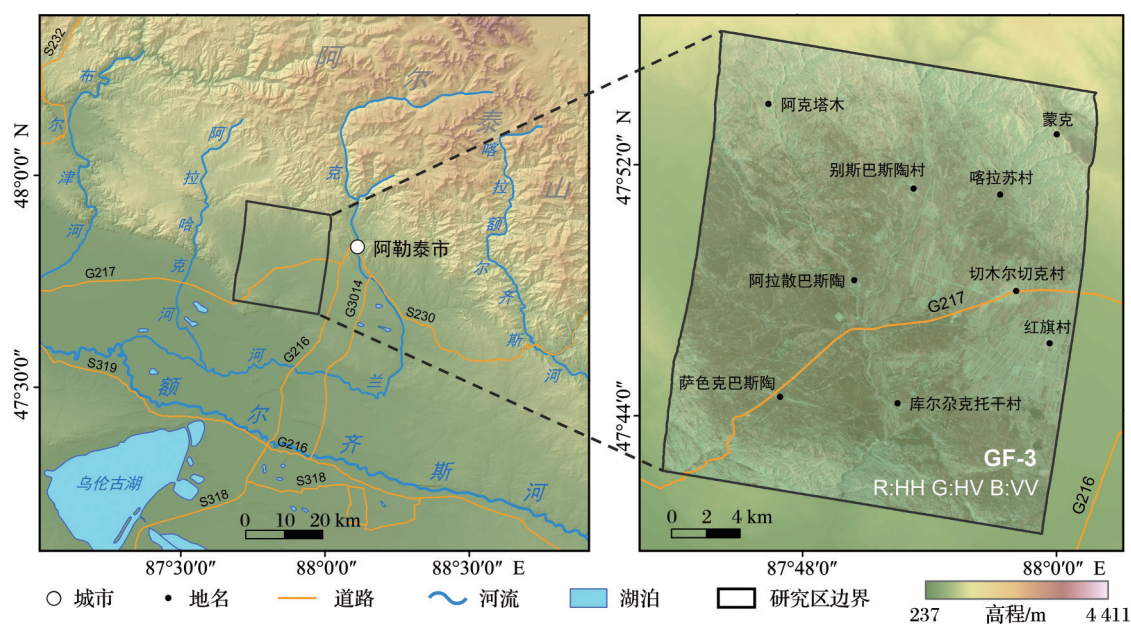


图1 研究区位置图

Fig.1 Location of the study area

2.3 数据预处理

GF-3数据的预处理是将回波信号转化为各种极化方式下的后向散射系数的过程。预处理包括多视、滤波、辐射定标、地理编码以及地形校正,这些步骤在PIE(Pixel Information Expert)软件完成,图1右图所示为GF-3图像预处理结果。

Landsat 8数据预处理是将DN值转换为大气层顶反射率值。预处理包括辐射亮度值转换、大气层顶反射率计算、图像配准、图像裁切,这些步骤在ENVI和ArcGIS软件完成。图2所示为Landsat 8图像预处理的结果,为了突出积雪与非积雪的差异,使用短波红外波段、红波段和绿波段进行RGB

彩色合成,在合成图像中积雪区域显示为青色,非积雪区域显示为深浅不一的红色。

结合Landsat 8和GF-3图像共同确定训练和验证样本。训练样本以感兴趣区域的形式在GF-3图像上选取,由人机交互选择确定,共选取积雪样本52个区域(193像元)、非积雪样本60个区域(200像元)。统计显示,两类训练样本均各自服从正态分布。验证样本则是在GF-3图像上随机产生,生成验证样本347个区域(347像元),再根据光学图像确定验证样本的类别,包括积雪像元262个、非积雪像元85个。训练样本和验证样本的分布如图2所示。

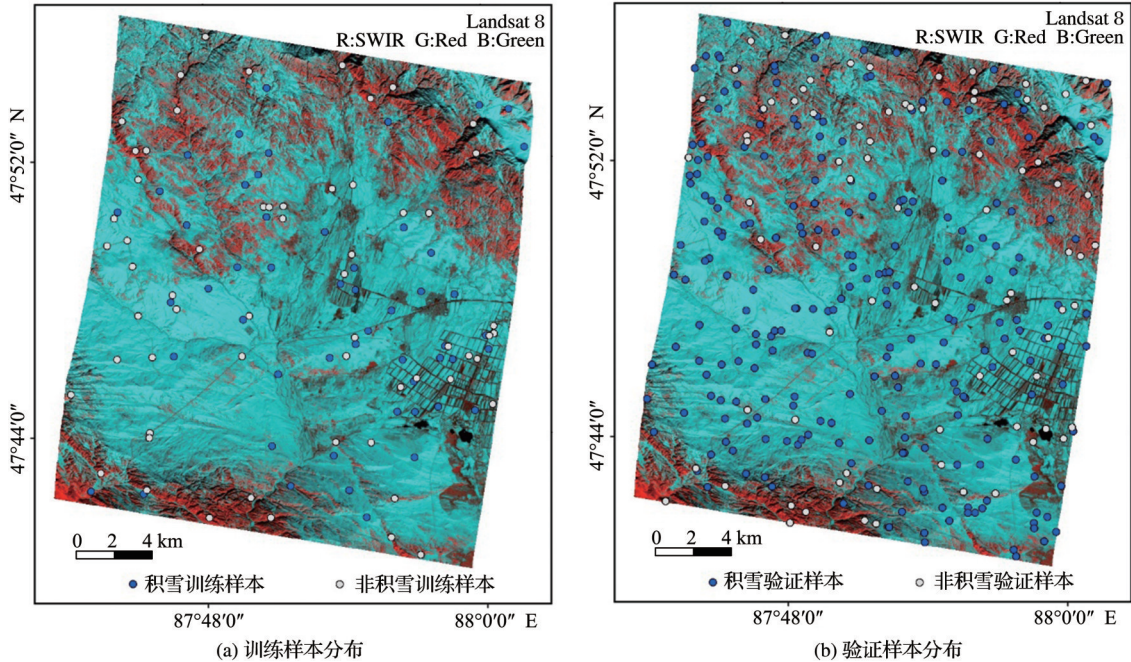


图2 训练样本和验证样本在光学遥感图像上的分布

Fig.2 Distribution of training and validation samples in optical image

3 方法

3.1 散射机制分析

已有研究表明,在干雪且局部入射角低于 60° 条件下,HH、HV、VH、VV极化有积雪覆盖的草地后向散射系数比无积雪覆盖的草地低 $2\sim 3$ dB^[19]。研究区内的积雪主要分布在戈壁和耕地上,为探究识别干雪的可行性,在戈壁中选取了1 132个积雪像元、1 104个非积雪像元,在耕地中选取了829个积雪像元、812个非积雪像元进行表征分析。图3为戈壁和耕地的HH、HV、VV极化后向散射系数的分布,结果显示:各下垫面类型中积雪与非积雪样本均呈分离趋势。3种极化有积雪覆盖的戈壁后向

散射系数比无积雪覆盖的戈壁低5 dB左右,两类具有很好的区分度;HH、VV极化有积雪覆盖的耕地后向散射系数比无积雪覆盖的耕地低 $2\sim 3$ dB左右,HV极化低 $3\sim 4$ dB左右。造成这些差异的原因很可能是电磁波穿透积雪介质时自身的衰减和干雪中微量的水分,这使得积雪覆盖地表的后向散射系数比无积雪覆盖的低,可为积雪识别提供依据。

积雪的后向散射由空气—雪界面面散射、雪层体散射、雪—地界面面散射和雪层体散射和雪—地界面面散射作用项组成。在干雪条件下,电磁波将穿透雪层,散射类型以雪—地界面的面散射为主,

并包含少量的雪层体散射。当雪层中湿度较大时,将导致电磁波的穿透深度降低,此时散射包括空气—雪界面的面散射和雪层体散射。随着湿度的增加,穿透深度进一步降低,包含的雪层体散射会

不断减少,空气—雪界面的面散射占主导地位。极化分解技术可将地物的后向散射系数分解为不同的散射类型,能提供更丰富的散射信息,可为积雪识别提供一种行之有效的方法。

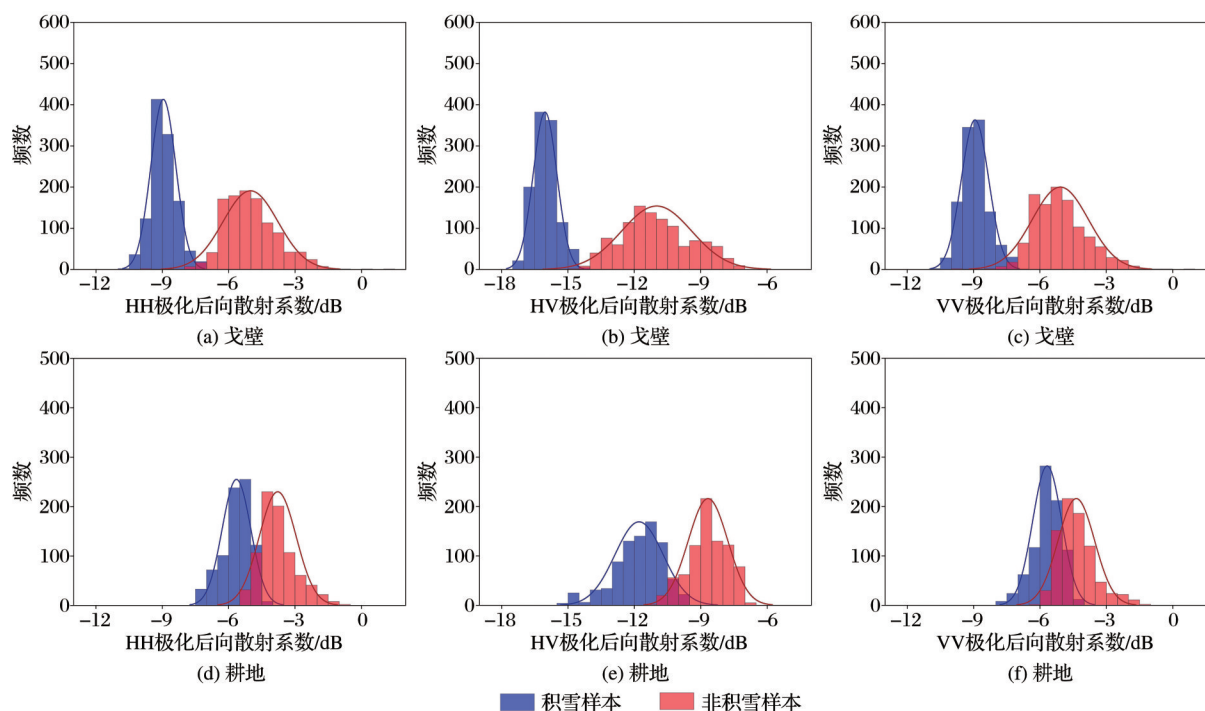


图3 不同下垫面类型积雪、无积雪覆盖地表后向散射系数分布

Fig.3 Distribution of backscattering coefficient of snow-covered and snow-free pixels under different surface conditions

3.2 极化特征获取

利用极化散射矩阵进行相干目标分解,利用极化相干矩阵或极化协方差矩阵进行非相干目标分解,以此得到积雪识别的特征。Pauli分解是常见的相干目标分解方法,H-A- $\bar{\alpha}$ 分解、Freeman分解、Yamaguchi分解和Anyang分解是常见的非相干目标分解方法。

Pauli分解将极化散射矩阵转化为3种散射基的加权和。满足天线单站互易时,可得到Pauli分解的3种散射功率。从物理意义上可以理解为面散射功率、二面角散射功率和对应于非相干目标分解中的体散射功率^[20]。

H-A- $\bar{\alpha}$ 分解,也称为Cloude分解^[21],是从极化相干矩阵的特征值和特征向量入手,求算目标散射回波中占主导地位的散射分量。该分解可获得4种分量:极化散射熵 H 描述不同散射类型表现出的混乱程度;散射各向异性度 A 用于确定第二、三特征值之间的关系,理论上 H 大于0.7时, A 才用于散射机制的识别;平均散射角 $\bar{\alpha}$,范围在0~90度,对应的散射方式从面散射到体散射,再从体散射到二面角散

射;第4个分量是这3个特征值的加权均值。

Freeman分解对体散射、二面角散射和表面散射3种散射机制分别建模,将极化协方差矩阵分解为这3种散射的线性和^[22]。该方法基于物理散射模型,对自然散射体的描述效果较好。求解极化协方差矩阵可得到二面角散射、体散射和面散射3种分量的散射功率。

为了对包含复杂几何散射结构的目标进行极化分解,克服Freeman分解在地形复杂区域适用性较差的问题,Yamaguchi等^[23]向模型中加入第四种散射分量,该散射相当于螺旋体散射,弥补了Freeman分解无法应用于不对称散射的缺陷。并根据 $10 \log \left(\frac{\langle |S_{vv}|^2 \rangle}{\langle |S_{hh}|^2 \rangle} \right)$ 的大小使用不同的体散射分量,求解极化协方差矩阵得到二面角散射、体散射、面散射和螺旋体散射的功率。

Yamaguchi等^[24]基于极化相干矩阵分解时,面散射和二面角散射的贡献会出现负值,这与散射功率的结果不符。针对这一问题,安文韬等^[25-26]提出Anyang极化分解方法。在Yamaguchi分解前使用

定向角补偿的方法,令相干矩阵变换,可解决散射分量中贡献值为负值的问题,而且还可以增强二面角散射功率。结合能量约束和定向角补偿过程,得到 4 种散射条件下的散射功率。

利用上述 5 种极化分解方法对 GF-3 全极化数据进行极化分解,得到 18 种极化特征,加上 4 种极化后向散射系数,共计获取 22 种特征(表 1)。根据相关文献和训练样本在各特征中的分布,大部分特征对积雪比较敏感,除散射各向异性度 A 特征外,各特征在积雪区域的值比非积雪区域的值低。

表 1 特征列表
Table 1 List of features

类型	特征			
后向散射系数	HH	HV	VH	VV
Pauli 分解	P_{odd}	P_{dbl}	P_{vol}	
H-A- α 分解	H	A	α	λ
Freeman 分解	F_{dbl}	F_{vol}	F_{odd}	
Yamaguchi 分解	Y_{dbl}	Y_{vol}	Y_{odd}	Y_{hix}
Anyang 分解	A_{dbl}	A_{vol}	A_{odd}	A_{hix}

3.3 特征优选与积雪识别

基于特征优选的积雪识别方法包括两部分。首先,使用全部特征利用随机森林方法进行积雪的初步识别,同时计算出各特征在积雪识别中的重要性;其次,根据特征的重要性进行特征优选,利用最优特征集识别积雪。

随机森林是一种集成学习算法,其作用机制是从所有特征中随机选出适量的特征构建指定数量的决策树。所有决策树参与总体决策,按照集成规则确定最终类别^[27-28],在 SAR 图像分类中常被使用^[29-30]。本文先使用全部特征利用随机森林方法对积雪进行初步识别,使用的随机森林来自 Matlab 中的 TreeBagger 方法^[27]。训练每棵决策树的特征数量设定在 5~22 个,决策树使用 1~100 棵。使用 100 棵决策树识别积雪时,特征数量对袋外误差的影响在 0.02 范围内,说明特征数量对识别结果影响不大。因此决定使用 100 棵决策树,每棵树随机抽选 5 个特征用于积雪的初步识别。

计算各特征的重要性采用的方法是向所有特征中逐一加入噪声,如果相应的特征对识别的贡献较强,分类器识别袋外数据时加入噪声前后的准确率会发生明显变化。利用识别准确率作为衡量标准反映特征的重要程度。利用样本集 D_i 训练决策树 h_i ,训练过程中的袋外数据记为 D_i^{oob} ;再利用 D_i^{oob}

计算 h_i 的袋外误分率,记为 O_i ;对特征 X^j 做轻微改动,袋外数据更新为 D_{ic}^{oob} ,利用改动后的样本集计算 h_i 的误分率为 O_{ic}^j ;特征 X^j 的重要程度计算公式为:

$$I(X^j) = \frac{1}{k} \sum_{i=1}^k |O_{ic}^j - O_i| \quad (1)$$

其中: k 为特征 X^j 参与训练的决策树数目, I 表示特征的重要程度。

通过上述过程,可以计算出每一特征在积雪识别中的重要性。特征优选是按照特征重要性大小向随机森林方法中逐个添加特征。每加入一个特征进行一次积雪识别,记录每次加入特征后积雪识别的总体精度,根据总体精度的变化确定最优特征集。利用最优特征集进行积雪识别,以期获得最优的识别结果。

3.4 积雪识别精度评价

本文基于随机产生的 347 个验证样本对积雪识别结果的精度进行评价。通过积雪识别结果与真值的比较,计算 F 测度和总体精度指标,以此评价积雪识别的精度。 F 测度包括准确率(Precision)、召回率(Recall)、 F 指数(F -score),其中准确率指积雪识别结果中积雪被正确识别的比例,记为 P ;召回率指积雪样本中积雪被正确识别的比例,记为 R ; F 指数则是两者的协调平均值,记为 F ;总体精度指所有验证样本被正确识别的比例,记为 A 。计算公式如下:

$$P = \frac{TP}{TP+FP} \quad R = \frac{TP}{TP+FN} \quad F = 2 \times \frac{P \times R}{P+R} \quad (2)$$

$$A = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

其中:TP 指积雪样本被正确识别的个数,FP 指非积雪样本被识别为积雪的个数,FN 指积雪样本被识别为非积雪的个数,TN 指非积雪样本被正确识别的个数。

4 结果与讨论

4.1 特征优选前的积雪识别

特征优选前使用全部特征进行积雪识别。为了检验本文方法的性能,使用最大似然法、支持向量机和 BP 神经网络 4 种分类器作比较。基于选取的训练样本,利用十折交叉验证法对支持向量机和 BP 神经网络进行参数优化。使用径向基函数(Radial Basis Function, RBF)作为支持向量机的核函数,优化的参数为惩罚因子 C 和支持向量因子 γ ^[31]。最终确定 C 的最优参数为 163, γ 的最优参数为

0.01。BP 神经网络优化的参数有激励函数、学习率和误差容忍度。最优参数确定为使用 Logistic 激励函数,初始学习率为 0.1,误差容忍度为 0.001,此参数下的最优精度为 0.931。统计显示,选取的训练样本基本服从正态分布,满足最大似然法的参数要求。

图 4 为特征优选前 4 种分类器的积雪识别结果,其中图 4(a)、(b)、(c)、(d)分别为随机森林、最大似然法、支持向量机和 BP 神经网络的识别结果。

结果表明,4 种分类器都能正确地识别主要积雪范围。其中,随机森林和最大似然法的识别结果与光学影像表征比较相近。两种识别结果的区别在东北部的山区,随机森林的识别结果中误判积雪像元的数量比最大似然法的少。支持向量机的识别结果中积雪范围被低估,尤其在中部农田和北部山区,大量积雪像元被识别为非积雪像元。BP 神经网络的识别结果比较破碎,识别效果最差。

表 2 为精度评价结果。随机森林的 F 指数和总

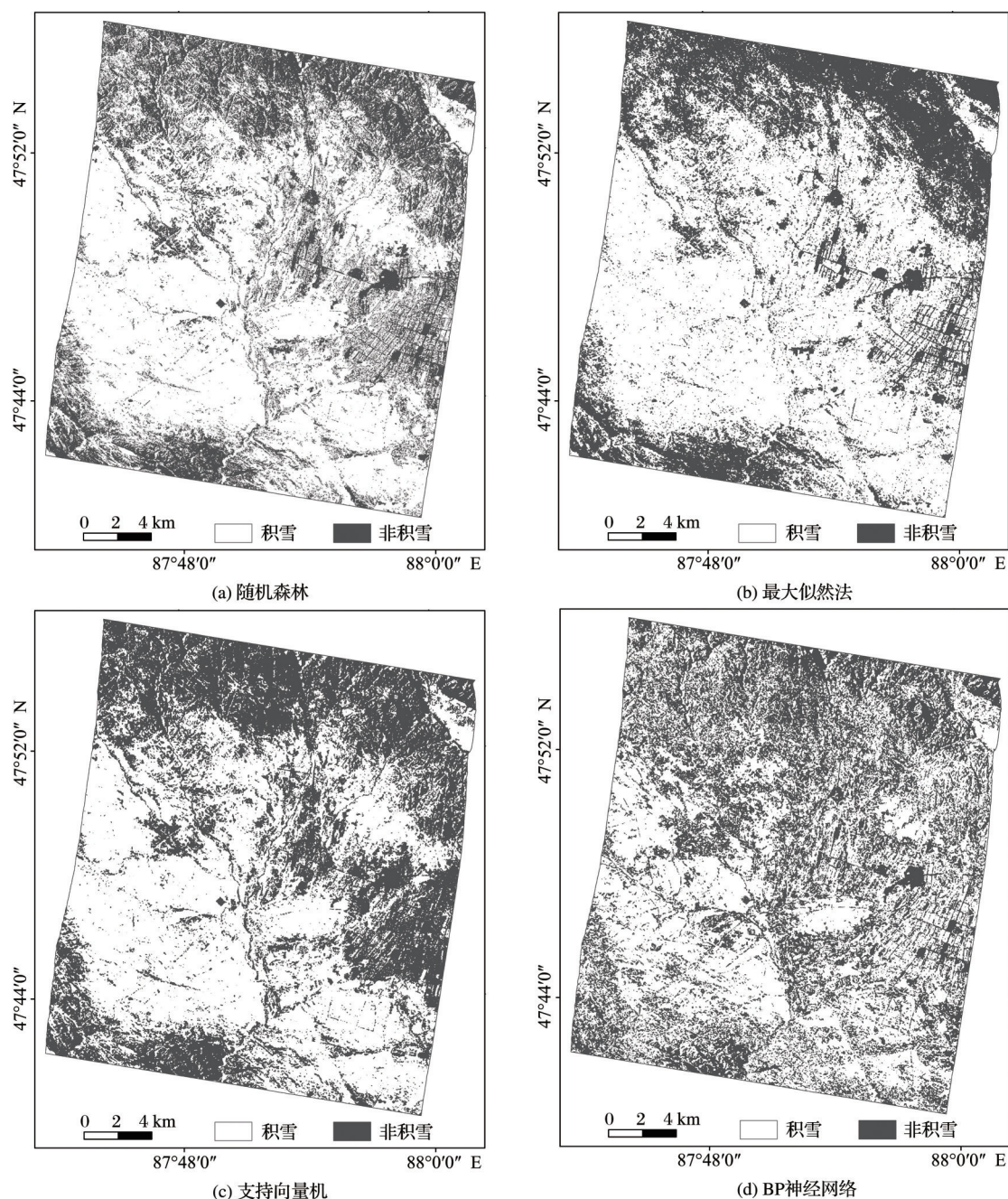


图 4 特征优选前的积雪识别结果

Fig.4 Results of snow cover recognition before features selection

体精度均是最高,达到了0.86和0.79。最大似然法的 F 指数和总体精度为0.83和0.75,识别效果次于随机森林。支持向量机的准确率最高,达到了0.90,但召回率仅有0.74,说明支持向量机的识别结果中虚警率低、漏检率高,导致识别结果中积雪范围有所低估。BP神经网络的识别结果最差,各指标均比其他分类器低。从积雪的初步识别结果来看,随机森林的识别精度最高。

4.2 特征优选

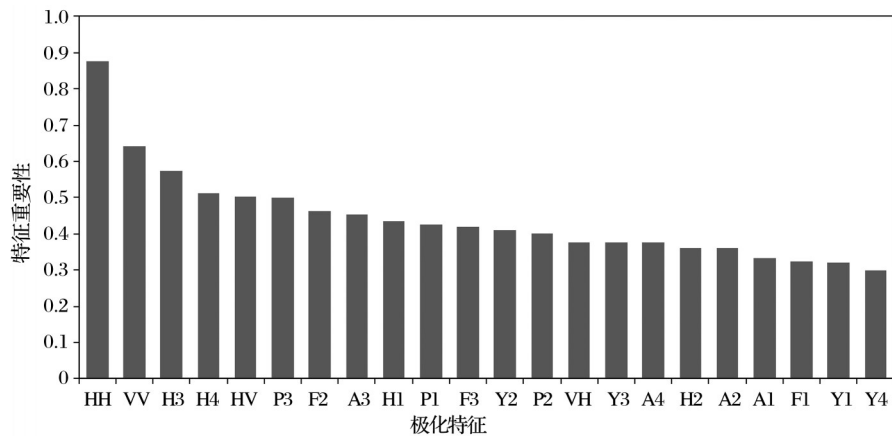
特征优选是一个根据特定规则筛选特征的过程,包括特征的重要性评估和特征的筛选两个步骤。首先,对参与积雪识别的22个特征进行重要性评估,各特征的重要性在特征优选前进行积雪识别时产生,排序结果如图5所示。HH和VV极化后向

表2 特征优选前不同分类器的识别精度比较

Table 2 Accuracy comparison of the different classifiers before features selection

分类器	准确率(P)	召回率(R)	F 指数(F)	总体精度(A)
随机森林	0.87	0.84	0.86	0.79
最大似然法	0.88	0.77	0.83	0.75
支持向量机	0.90	0.74	0.81	0.74
BP神经网络	0.82	0.66	0.73	0.63

散射系数的重要性最高,分别为0.88和0.64。其次,H-A- $\bar{\alpha}$ 极化分解中第3分量的重要性为0.58,剩余特征的重要性在0.3~0.5之间。说明同极化后向散射系数对积雪识别的贡献比交叉极化后向散射系数的大。面散射分量(A3、F3、Y3)和体散射分量(A2、F2、Y2)对积雪识别的贡献比对应极化分解中二面角散射分量(A1、F1、Y1)的大。



(H3表示H-A- $\bar{\alpha}$ 极化分解的第3个分量,P3表示Pauli分解的第3个分量,其他缩写类似)

图5 特征重要性排序

Fig.5 Ranking of importance of features

进行特征重要性评估后,根据特征的重要性大小筛选最优特征集。随着特征数量的增加,总体精

度的变化如图6所示。使用HH一个特征,总体精度仅为0.73;使用HH和VV两个特征,总体精度为

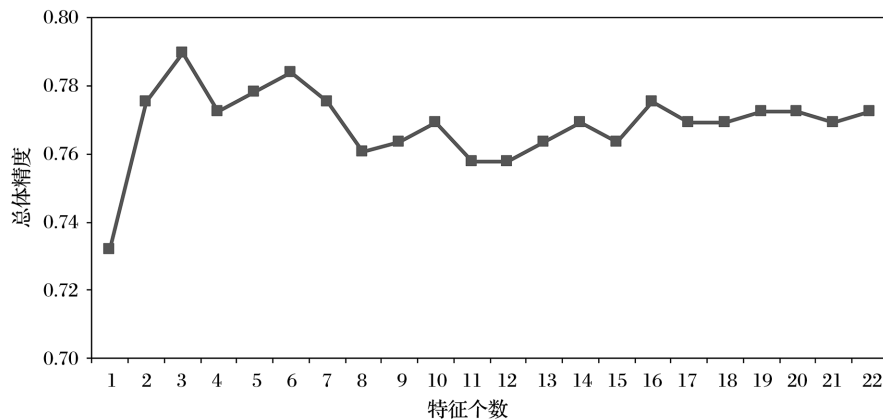


图6 特征个数及积雪识别总体精度图

Fig.6 The number of features and overall accuracy of snow cover recognition

0.77;使用重要性最高的3个特征时,总体精度达到最高值0.79。此后,总体精度在0.76~0.79之间变化。因此确定最优特征集为同极化后向散射系数(HH、VV)和H-A- \bar{a} 分解的平均散射角分量(H3)组成的集合。

4.3 特征优选后的积雪识别

基于特征优选获得的最优特征集,利用随机森

林方法进行积雪识别。同时,使用最大似然法、支持向量机和BP神经网络3种分类器与之对比。结果如图7所示,积雪范围总体上与图4保持一致。不过,由于特征数量的减少,特征优选前后4种分类器在北部山区的积雪识别结果有所变化。另外,最大似然法、支持向量机、BP神经网络3种分类器在中部农田区域均低估了积雪范围。

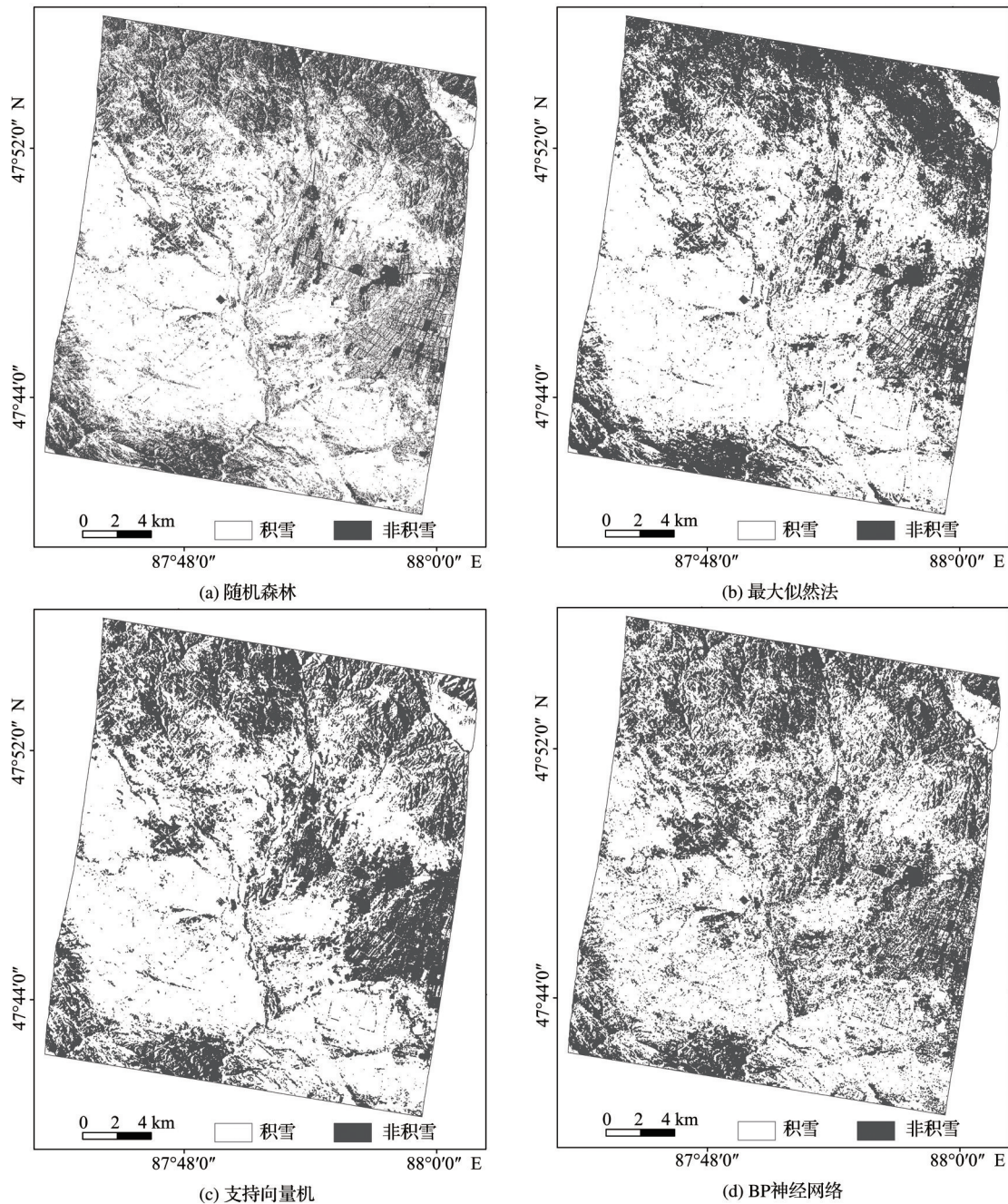


图7 特征优选后的积雪识别结果

Fig.7 Result of snow cover recognition after features selection

为了更准确地分析和对比4种分类器使用最优特征集时的识别结果,在图像中选取了3处有代表

性的区域开展进一步分析(图8)。黄色框表示所选区域在研究区中的位置。

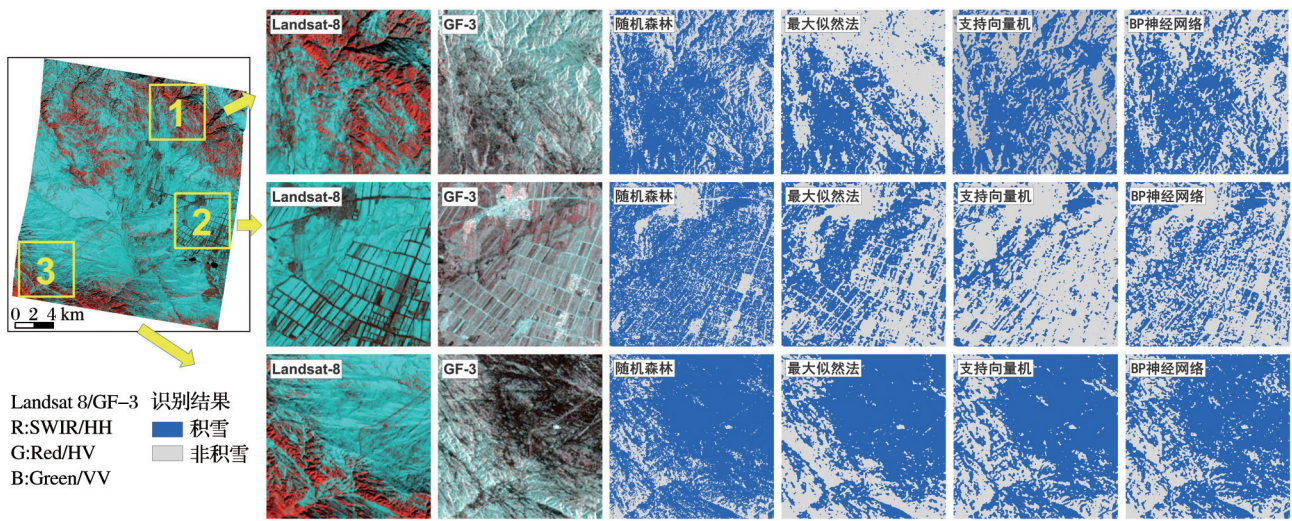


图 8 Landsat 8 和 GF-3 图像及积雪识别结果细节图

Fig.8 Details of Landsat-8, GF-3 and results of snow cover recognition

区域 1 位于研究区东北部,属于地形起伏较大的山区,土地覆盖主要以草地、林地为主,因此积雪信号受地形和下垫面影响比较严重。比较 4 种分类器的识别结果发现,效果最好的是随机森林和支持向量机,可识别出主要的积雪范围;最大似然法效果最差,只有少部分积雪像元被正确识别。

区域 2 位于研究区东部,包含了农田及少量村庄,地势平坦。比较 4 种分类器的识别结果发现,最大似然法的识别结果中,规则农田的轮廓识别的比较完整,但积雪范围被低估;随机森林可识别出主要的积雪范围,但农田的轮廓不完整;支持向量机的识别结果精度最差,出现了大量的误判现象。

区域 3 位于研究区的西南部,平均海拔比周边地区高百米左右,主要以戈壁、草地和少量灌木组成。比较识别结果发现,随机森林的结果最佳,积雪与非积雪区域的边界比较清晰;其他分类器的识别结果中,积雪与非积雪的边界区域误判现象较多,其中 BP 神经网络的识别结果最差。

4 种分类器识别结果和真值的比较结果如表 3 所示,粗体数值表示特征优选后的精度比特征优选前的高,正体数值表示特征优选前后精度不变,正体数值加下划线表示特征优选后精度降低。随机森林方法使用最优特征集与使用全部特征时精度基本保持不变。此外,4 种分类器的召回率均有所提高,即漏检率减少,说明特征优选后对积雪识别有用的特征增多,从而提高了识别精度。从结果来看,基于特征优选和随机森林方法进行积雪识别时效果最佳。

表 3 特征优选后不同分类器的识别精度比较

Table 3 Accuracy comparison of the different classifiers after features selection

分类器	准确率(P)	召回率(R)	F 指数(F)	总体精度(A)
随机森林	<u>0.85</u>	0.87	0.86	0.79
最大似然法	0.88	0.79	0.83	0.76
支持向量机	<u>0.89</u>	0.77	0.82	0.75
BP 神经网络	0.87	0.76	0.81	0.73

5 结 论

本文利用国产 GF-3 数据提出了一种基于特征优选的积雪识别方法。首先,通过极化分解技术提取积雪识别的候选特征;然后,利用随机森林方法计算各候选特征的重要性;再根据特征的重要性选择对识别贡献较大的特征,构建最优特征集进行积雪识别。

分析特征的重要性发现,在使用最重要的 3 个特征时积雪识别精度达到最高,因此确定最优特征集为 HH、VV 和 $H-A-\bar{\alpha}$ 分解的平均散射角分量(H3)组成的集合。本文方法与最大似然法、支持向量机和 BP 神经网络 4 种分类器的对比发现,使用最优特征集和随机森林方法的识别精度最高。特征优选后积雪识别精度基本保持不变,甚至有所增加。从分类器的选择和特征优选前后各指标值的变化来看,本文提出的方法具有一定的优势,可有效用于积雪识别。

虽然在实验中,特征优选前后积雪识别的精度只有较小幅度的提高,但是因为随机森林的训练速度快,可利用特征优选进行数据的降维处理,所以

本文方法在进行大尺度、大样本的积雪识别时具有很大的潜力。

参考文献(References):

- [1] Cai Dihua, Guo Ni, Wang Xing, *et al.* The Spatial and Temporal Variations of Snow Cover over the Qilian Mountains based on MODIS Data[J]. Journal of Glaciology and Geocryology, 2009, 31(6): 1028-1036.[蔡迪花, 郭妮, 王兴, 等. 基于MODIS的祁连山区积雪时空变化特征[J]. 冰川冻土, 2009, 31(6): 1028-1036.]
- [2] Wang Jian. Comparison and Analysis on Methods of Snow Cover Mapping by Using Satellite Remote Sensing Data[J]. Remote Sensing Technology and Application, 1999, 14(4): 29-36.[王建. 卫星遥感雪盖制图方法对比与分析[J]. 遥感技术与应用, 1999, 14(4): 29-36.]
- [3] Cui Caixia, Wei Rongqing, Li Yang. Long-term Change of Seasonal Snowcover and Its Effects on Runoff Volume in the Upper Reaches of the Tarim River[J]. Arid Land Geography, 2005, 28(5): 569-573.[崔彩霞, 魏荣庆, 李杨. 塔里木河上游地区积雪长期变化趋势及其对径流量的影响[J]. 干旱区地理, 2005, 28(5): 569-573.]
- [4] Wang Shijin, Ren Jiawen. A Review of the Progresses of Avalanche Hazards Research[J]. Progress in Geography, 2012, 31(11): 1529-1536.[王世金, 任贾文. 国内外雪崩灾害研究综述[J]. 地理科学进展, 2012, 31(11): 1529-1536.]
- [5] Lou Mengyun, Liu Zhihong, Lou Shaoming, *et al.* Temporal and Spatial Distribution of Snow Cover in Xinjiang from 2002 to 2011[J]. Journal of Glaciology and Geocryology, 2013, 35(5): 1095-1102.[娄梦筠, 刘志红, 娄少明, 等. 2002~2011年新疆积雪时空分布特征研究[J]. 冰川冻土, 2013, 35(5): 1095-1102.]
- [6] Nagler T, Rott H. Retrieval of Wet Snow by Means of Multitemporal SAR Data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2000, 38(2): 754-765.
- [7] Xiao Pengfeng, Feng Xuezhi, Xie Shunping, *et al.* Research Progresses of High-resolution Remote Sensing of Snow in Manasi River Basin in Tianshan Mountains, Xinjiang Province[J]. Journal of Nanjing University (Natural Sciences Edition), 2015, 51(5): 909-920.[肖鹏峰, 冯学智, 谢顺平, 等. 新疆天山玛纳斯河流域高分辨率积雪遥感研究进展[J]. 南京大学学报(自然科学版), 2015, 51(5): 909-920.]
- [8] Huang L, Li Z, Tian B S, *et al.* Recognition of Supraglacial Debris in the Tianshan Mountains on Polarimetric SAR Images[J]. Remote Sensing of Environment, 2014, 145(8): 47-54.
- [9] Schellenberger T, Ventura B, Zebisch M, *et al.* Wet Snow Cover Mapping Algorithm based on Multitemporal COSMO-SkyMed X-Band SAR Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2012, 5(3): 1-9.
- [10] Rott H, Davis R E. Multifrequency and Polarimetric SAR Observations on Alpine Glaciers[J]. Annals of Glaciology, 1993, 17: 98-104.
- [11] Li Zhen, Guo Huadong, Li Xinwu, *et al.* SAR Interferometry Coherence Analysis and Snow Mapping[J]. Journal of Remote Sensing, 2002, 6(5): 334-338.[李震, 郭华东, 李新武, 等. SAR干涉测量的相干性特征分析及积雪划分[J]. 遥感学报, 2002, 6(5): 334-338.]
- [12] Singh G, Venkataraman G. Application of Incoherent Target Decomposition Theorems to Classify Snow Cover over the Himalayan Region[J]. International Journal of Remote Sensing, 2012, 33(13): 4161-4177.
- [13] Huang L, Li Z, Tian B S, *et al.* Classification and Snow Line Detection for Glacial Areas Using the Polarimetric SAR Image[J]. Remote Sensing of Environment, 2011, 115(7): 1721-1732.
- [14] Tsai Y S, Dietz A, Oppelt N, *et al.* Wet and Dry Snow Detection Using Sentinel-1 SAR Data for Mountainous Areas with a Machine Learning Technique[J]. Remote Sensing, 2019, 11(8): 895. doi:10.3390/rs11080895.
- [15] Pal M. Random Forest Classifier for Remote Sensing Classification[J]. International Journal of Remote Sensing, 2005, 26(1): 217-222.
- [16] Chen J, Chen J, Liao A P, *et al.* Global Land Cover Mapping at 30 m Resolution: A POK-based Operational Approach[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 103: 7-27.
- [17] Shen Yongping, Wang Guoya, Su Hongchao, *et al.* Hydrological Processes Responding to Climate Warming in the Upper Reaches of Kelan River Basin with Snow-dominated of the Altay Mountains Region, Xinjiang, China[J]. Journal of Glaciology and Geocryology, 2007, 29(6): 845-854.[沈永平, 王国亚, 苏宏超, 等. 新疆阿尔泰山区克兰河上游水文过程对气候变暖的响应[J]. 冰川冻土, 2007, 29(6): 845-854.]
- [18] Zhou Bochong. An Analysis on the Relationship between Streamflow and Precipitation in Altay Mountains Region[J]. Journal of Glaciology and Cryopenology, 1983, 5(4): 49-56.[周伯诚. 我国阿尔泰山的降水及河流径流分析[J]. 冰川冻土, 1983, 5(4): 49-56.]
- [19] He Guangjun, Feng Xuezhi, Xiao Pengfeng, *et al.* Characterization of C band SAR Image for Snow in Mountainous Areas of Manasi River Basin[J]. Journal of Nanjing University (Natural Sciences Edition), 2015, 51(5): 955-965.[何广军, 冯学智, 肖鹏峰, 等. 玛纳斯河流域山区积雪的C波段SAR图像表征[J]. 南京大学学报(自然科学), 2015, 51(5): 955-965.]
- [20] Wang Chao, Zhang Hong, Chen Xi, *et al.* Image Processing of Fully Polarimetric SAR[M]. Beijing: Science Press, 2008.[王超, 张红, 陈曦, 等. 全极化合成孔径雷达图像处理[M]. 北京: 科学出版社, 2008.]
- [21] Cloude S R, Pottier E. An Entropy based Classification Scheme for Land Application of Polarimetric SAR[J]. IEEE Transactions on Geoscience and Remote Sensing, 1997, 35(1): 68-78.
- [22] Freeman A, Durden S L. A Three-Component Scattering

- Model for Polarimetric SAR Data[J]. IEEE Transactions on Geoscience and Remote Sensing, 1998, 36(3):963-973.
- [23] Yamaguchi Y, Moriyama T, Ishido M, *et al.* Four-Component Scattering Model for Polarimetric SAR Image Decomposition [J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 104(8):1699-1706.
- [24] Yajima Y, Yamaguchi Y, Yamada H, *et al.* A Four-Component Decomposition of POLSAR Images based on the Coherency Matrix [J]. IEEE Geoscience and Remote Sensing Letters, 2006, 3(3):292-296.
- [25] An W, Yi C, Jian Y. Three-Component Model-based Decomposition for Polarimetric SAR Data [J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(6):2732-2739.
- [26] An W, Xie C, Yuan X, *et al.* Four-Component Decomposition of Polarimetric SAR Images with Deorientation [J]. IEEE Geoscience and Remote Sensing Letters, 2011, 8(6):1090-1094.
- [27] Breiman L. Random Forests [J]. Machine Learning, 2001, 45(1):5-32.
- [28] Xu Qiao, Zhang Xiao, Yu Shaohuai, *et al.* Multi-feature-based Classification Method Using Random Forest and Superpixels for Polarimetric SAR Images [J]. Journal of Remote Sensing, 2019, 23(4):685-694. [徐乔, 张霄, 余绍淮, 等. 综合多特征的极化 SAR 图像随机森林分类算法 [J]. 遥感学报, 2019, 23(4):685-694.]
- [29] Du P J, Samat A, Waske B, *et al.* Random Forest and Rotation Forest for Fully Polarized SAR Image Classification Using Polarimetric and Spatial Features [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 105:38-53. doi: 10.1016/j.isprsjprs.2015.03.002.
- [30] Tang Tingyuan, Fu Bolin, He Suyun, *et al.* Identification of Typical Land Features in the Lijiang River Basin with Fusion Optics and Radar [J]. Remote Sensing Technology and Application, 2020, 35(2):448-457. [唐廷元, 付波霖, 何素云, 等. 基于 GF-1 和 Sentinel-1A 的漓江流域典型地物信息提取 [J]. 遥感技术与应用, 2020, 35(2):448-457.]
- [31] Cherkassky V, Ma Y. Practical Selection of SVM Parameters and Noise Estimation for SVM Regression [J]. Neural Networks, 2004, 17(1):113-126.

Recognition of Snow Cover based on Features Selection in GF-3 Fully Polarimetric Data

Ma Tengyao, Xiao Pengfeng, Zhang Xueliang, Ma Wei, Guo Jinjin

(Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China)

Abstract: This study proposed a recognition method for snow cover based on feature selection using GF-3 fully polarimetric data. The study area was selected from the typical area of the Kelan River Basin in the southern piedmont of the Altai Mountains, Xinjiang Province. First, we obtained 22 polarization features of GF-3 data by polarization decomposition. The importance of each feature was calculated by using Random Forest (RF) method. Then, we designed the rules of feature selection to generate the optimal feature sets, which were used to recognize snow cover with RF method. Analyzing the importance of the features, we can find that, for snow cover recognition, the contribution of the same polarization backscattering coefficient is greater than that of the cross polarization backscattering coefficient, and the contribution of the surface scattering or volume scattering is greater than that of the dihedral angle scattering. Finally, a comparison with the Maximum Likelihood, Support Vector Machine, and BP neural network was made for testing the performance of the proposed method. It is found that the optimal feature sets using RF method to recognize snow cover have the highest accuracy (F -score is 0.86, overall accuracy is 0.79). From the selection of classifiers and the results of features selection, the proposed method is very effective in recognition of snow cover.

Key words: Snow cover recognition; GF-3; Polarization decomposition; Features selection; Random Forest