

MODIS 亚洲高山区积雪面积比例制图

高伟强^{1,2}, 郝晓华², 和栋材¹, 孙兴亮³, 李弘毅², 任鸿瑞¹, 赵琴²

1. 太原理工大学 矿业工程学院, 太原 030000;
2. 中国科学院西北生态环境资源研究院, 兰州 730000;
3. 兰州大学 资源环境学院, 兰州 730000

摘要: 积雪面积比例FSC (Fractional Snow Cover) 能在亚像元尺度上定量描述积雪的覆盖程度, 相比二值积雪更适合反映复杂山区积雪的分布情况, 是山区融雪径流模拟, 气候变化预测的重要输入参数。本研究在亚洲高山区HMA (High Mountain Asia) 基于分地类特征选择的多元自适应回归样条MARS (Multivariate Adaptive Regression Splines) 模型LC-MARS发展了MODIS FSC反演算法, 并制备了亚洲高山区FSC产品。以Landsat 8提取的FSC为参考真值验证LC-MARS模型反演FSC精度, 对比相同训练样本下LC-MARS模型与线性回归模型反演FSC精度, 比较LC-MARS模型制备的FSC与MOD10A1、SnowCCI在亚洲高山区的精度表现。结果表明: (1) LC-MARS模型反演的FSC总Accuracy、Recall分别为93.4%、97.1%, 总体RMSE为0.148, MAE为0.093, 总体精度较高。(2) 相同训练样本下LC-MARS模型在林区、植被和裸地反演FSC精度均高于线性回归模型, 表明LC-MARS模型更适用于山林区FSC反演。(3) MOD10A1总体RMSE为0.178, MAE为0.096; SnowCCI总体RMSE为0.247, MAE为0.131, LC-MARS制备的FSC精度均高于MOD10A1、SnowCCI, 表明由LC-MARS反演的FSC具有一定的应用价值。总体而言, LC-MARS模型可以拟合高维非线性关系, 显著提高山林区FSC的反演精度且模型运算效率高, 适用于制备大尺度长时间序列的FSC产品。本研究基于LC-MARS模型制备了2000年—2021年亚洲高山区逐日MODIS FSC产品, 为亚洲高山区气候变化、水文水资源研究提供重要的数据支撑。

关键词: 遥感, 亚洲高山区, 积雪面积比例, MODIS, MARS, 地形校正

中图分类号: TP79/P2

引用格式: 高伟强, 郝晓华, 和栋材, 孙兴亮, 李弘毅, 任鸿瑞, 赵琴. 2024. MODIS 亚洲高山区积雪面积比例制图. 遥感学报, 28(9): 2223-2239

Gao W Q, Hao X H, He D C, Sun X L, Li H Y, Ren H R and Zhao Q. 2024. Machine learning-based mapping of fraction snow cover in High Mountain Asia by MODIS. National Remote Sensing Bulletin, 28(9): 2223-2239 [DOI: 10.11834/jrs.20242483]

1 引言

积雪是冰冻圈的重要组成部分, 是地球表面最为活跃的自然要素之一 (王建, 1999; 梁顺林等, 2020)。积雪面积比例FSC (Fractional Snow Cover) 描述的是亚像元尺度上积雪的覆盖度, 是区域水资源评估的重要指标和部分气候模式、水文模型的重要输入参数 (Metsämäki 等, 2012; Huang 等, 2011)。积雪面积比例数据可以准确反映地形复杂, 积雪分布不连续的山区积雪的变化特征, 对山区融雪径流模拟和气候变化监测具有

重要的意义 (Notarnicola, 2020)。

亚洲高山区HMA (High Mountain Asia) 是北半球中纬度积雪覆盖面积最大的地区 (Brun 等, 2018)。亚洲高山区积雪分布破碎化严重, 混合像元问题突出, 传统的二值积雪面积数据难以准确估计积雪的时空变化, 而积雪面积比例数据能够在亚像元尺度上定量描述像元内积雪的分布情况, 有效的解决混合像元的问题, 容易反映亚洲高山区积雪的分布特征 (Zhao 等, 2020; 赵宏宇 等, 2018; Hao 等, 2019; Su 等, 2020)。亚洲高山区积雪的时空变化影响全球气候变化, 积雪融水是

收稿日期: 2022-09-30; 预印本: 2024-01-08

基金项目: 国家自然科学基金(编号:U22A20564, 41971325); 国家重点研发计划(编号:2022YFF0711702-05)

第一作者简介: 高伟强, 研究方向为积雪定量遥感。E-mail: gaoweiqiang1055@link.tyut.edu.cn

通信作者简介: 郝晓华, 研究方向为积雪遥感、积雪与气候变化研究。E-mail: haoxh@lzb.ac.cn

周边大型河流的主要补给源,开展亚洲高山区积雪面积比例制图研究可以获取高精度的积雪覆盖数据,对亚洲高山区气候和水文变化研究具有重要的意义(刘时银等,2000;唐志光等,2021)。

随着遥感技术的发展,利用卫星遥感空间覆盖范围大,时间序列长的优势监测积雪覆盖时空变化已被广泛的应用和研究(König等,2001;黄晓东等,2012,Hao等,2022)。其中FSC主要通过光学遥感影像获取,而获取高精度的FSC面临的挑战主要有复杂地表覆盖造成的混合像元问题、地形变化导致山区阴影雪识别难以及林区冠层遮挡对积雪反射率的损失。

目前,基于光学遥感影像提取FSC的方法主要有3种:传统的经验模型、基于混合像元分解模型和机器学习模型(孙兴亮等,2022)。传统的经验模型就是构建FSC与其相关系数之间的线性回归模型,代表性研究主要有Salomonson和Appel构建了FSC与NDSI之间的半经验线性回归模型(Salomonson和Appel,2004),美国NASA发布的MODIS全球积雪覆盖产品就是基于该线性模型制备的。混合像元分解模型主要包含端元选取和丰度估计,先提取混合像元中各类端元,然后利用线性光谱混合分析模型LSMA(Linear Spectral Mixture Analysis)解混获取混合像元的FSC(Li等,2022;Zhang等,2019),主要研究有Rosenthal等通过线性光谱混合像元分解模型制备了Landsat-TM影像的FSC(Rosenthal和Dozier,1996);Painter等结合LSMA和积雪辐射传输模型提出了一种MODIS积雪面积比例和雪粒径反演模型MODSCAG(Painter等,2009)。机器学习在解决非线性拟合问题上具有很大优势(Lu等,2021),凭借高运算率和强大的泛化能力在提取积雪面积比例中被广泛的应用,代表性研究有:Kuter等(2018)分别应用多元自适应回归样条MARS(Multivariate Adaptive Regression Splines)和多层前馈人工神经网络ANNs(Artificial Neural Networks)模型制备了阿尔卑斯山区的MODIS FSC产品,两者皆表现出较高的精度;Czyzowska-Wisniewski等(2015)在复杂山区、林区应用ANNs模型反演了Landsat数据的FSC,与等效参考数据集IKONOS相比具有高度的相关性;Ciftci等(2017)设计比较了ANNs和支持向量机SVM(Support Vector Machines)用于MODIS FSC反演,两种方法都制备了较高精度的

FSC。上述研究表明机器学习方法更多的应用于山区和林区等复杂地形的FSC反演,并且能够取得较高的精度。

对比3类FSC反演模型,其中基于线型回归的经验模型物理意义明确,实现简单,但不适用于解决山林区FSC反演的非线性问题,反演精度低。混合像元分解模型的反演精度很大程度上依赖于端元的选取,在山区等复杂地形,端元选取的不确定性会降低FSC反演精度(Ye等,2021)。对比前两类方法,机器学习方法充分考虑了复杂地形积雪识别的光谱信息,能进一步提高山林区FSC反演精度(Hou和Huang,2014),但ANNs和SVM等机器学习模型复杂度高,可解释性低,可移植性较差,只能线下制备FSC,不适合大尺度长时间序列FSC制图。MARS模型的优势是能够通过简单平滑的连接分段线性多项式来定义自变量和因变量之间潜在的函数关系,为复杂高维的非线性数据建立灵活的回归模型。而获取亚洲高山区准确的积雪覆盖信息需要综合考虑混合像元、地形变化和冠层遮挡等多方面的影响因素,符合MARS模型拟合高维非线性数据的特点,构建好的MARS模型能够以多项式导出,可解释性强且容易移植到Google Earth Engine云平台进行云计算,提高运算效率大大降低了产品制备的时间成本。但在以往的MARS模型反演FSC的研究中(Kuter等,2018),积雪识别特征选择没有考虑到不同地表覆盖类型下积雪分布差异对FSC反演精度的影响。

因此,本研究的主要目的是针对亚洲高山区复杂地形地表,分地类特征选取构建LC-MARS模型,并用于亚洲高山区FSC制图。首先用引入地形校正和分地类特征选取的LC-MARS模型制备亚洲高山区FSC产品,并以Landsat 8 FSC为真值验证LC-MARS模型反演精度,然后对比LC-MARS模型与线性回归模型反演FSC的精度并与SnowCCI和MOD10A1产品进行比较,客观评价LC-MARS模型制备FSC产品的优势。

2 研究区与数据

2.1 研究区概况

亚洲高山区位于亚洲中部(25°N—45°N,67°E—107°E),是地球上除南北极之外的第3大冰冻圈,被称为地球的“第三极”(Pan等,2021)。研究区

范围及山脉分布情况见 [http://www.coldregion.com/2024/01/05/scf_har/\[2022-09-30\]](http://www.coldregion.com/2024/01/05/scf_har/[2022-09-30]), 亚洲高山区以青藏高原为主体, 周边分布着天山、昆仑山、祁连山和喜马拉雅山等众多山系, 平均海拔高达 4000—5000 m (Guo 等, 2021), 拥有除极圈外全球最大的积雪资源。由众多山脉组成亚洲高山区地形地貌复杂, 区域内积雪分布不均匀且变化快, 以 MODIS 空间分辨率来说混合像元现象严重 (Liang 等, 2017; Tang 等, 2017; Huang 等, 2017)。因此, 制备高精度的 FSC 能获取准确的亚洲高山区季节性积雪信息对评估和预测亚洲高山区气候变化具有重要意义。

2.2 数据及预处理

本研究使用 5 种遥感数据。分别是 MOD09GA 地表反射率数据, 哥白尼全球土地服务 CGLS (Copernicus Global Land Service) 分辨率为 100 m 的土地覆盖数据 (CGLS-LC100), 航天飞机雷达地形任务 SRTM (Shuttle Radar Topography Mission) 的 DEM 数据, Landsat 8 地表反射率数据 (Landsat 8 SR), 欧洲航天局雪气候变化倡议 SnowCCI (Snow Climate Change Initiativ) 和 MOD10A1 的 MODIS 全球 FSC 数据。

2.2.1 MOD09GA

MOD09GA 逐日地表反射率数据由 NASA 生产, 空间分辨率为 500 m, 时间分辨率为逐日。数据来源于 ([https://lpdaac.usgs.gov.\[2022-09-30\]](https://lpdaac.usgs.gov.[2022-09-30]))。

MOD09GA 地表反射率数据为反演 FSC 的主要数据源, 使用该数据 QA 质量评估波段提供的云掩膜信息进行去云处理, 其 7 个单波段反射率用于计算归一化植被指数 NDVI (Normalized Difference Vegetation Index), 归一化积雪指数 NDSI (Normalized Difference Snow Index), 归一化林地积雪指数 NDFSII (Normalized Difference Forest Snow Index) 和通用比率积雪指数 URSI (Universal Ratio Snow Index)。NDVI、NDSI 和 NDFSII 等 3 类光谱信息为 MARS 模型反演 FSC 的输入特征。URSI 是可见光绿波段与近红外和短波红外反射率之和的比值 (Wang 等, 2021), 用于构建 FSC 线性回归反演算法。

2.2.2 CGLS-LC100

CGLS-LC100 数据是哥白尼全球陆地服务产品中的一种, 提供了 100 m 空间分辨率的全球土地利用与土地覆盖类型数据, 时间分辨率为 1 年。数据来源于 ([https://land.copernicus.eu.\[2022-09-30\]](https://land.copernicus.eu.[2022-09-30]))。CGLS-LC00 提出了一种连续的土地覆盖分类方案, 相比于 MCD12Q1 能更好的描述异质性较大的土地覆盖区域, 适用于亚洲高山区复杂的地表覆盖 (Zou 等, 2022; García-Álvarez 等, 2022)。

本研究用 CGLS-LC100 数据来获取亚洲高山区土地覆盖信息, 用于地类划分依据。根据不同的土地覆盖区选择不同的模型输入特征来反演 FSC。CGLS-LC100 数据的土地覆盖类型共有 23 种, 如表 1 所示。

表 1 CGLS-LC100 土地重分类方案
Table 1 CGLS-LC100 land reclassification scheme

新类别像元值	土地类型	原像元值	原地类描述
0	水体	80,200	湖泊、河流和海洋等永久性水体
1	林区	20,111—116,121—126	灌木、森林等高大的森林区
2	植被	30,90,100	草地、湿地、苔藓等低矮植被区
3	裸地	40,50,60,70	耕地、裸地、雪冰等无植被区

2.2.3 DEM 数据

本研究使用的 DEM 数据是 SRTM 的 V3 和 V4 数据。SRTM V3 数据, 空间分辨率为 30 m, 数据来源于 ([https://cmr.earthdata.nasa.gov.\[2022-09-30\]](https://cmr.earthdata.nasa.gov.[2022-09-30])); SRTM V4 数据, 空间分辨率为 90 m, 来源于 ([https://srtm.csi.cgiar.org/\[2022-09-30\]](https://srtm.csi.cgiar.org/[2022-09-30]))。SRTM V3 与 Landsat 8 空间分辨率一致, 用于提取高程对

Landsat 8 地表反射率进行地形校正。SRTM V3 空间分辨率较高, 但可描述高程最大值为 6500 m, 与亚洲高山区实际海拔最高值不符, 而 SRTM V4 数据高程区间完整, 相对于 MODIS 数据, 分辨率也较高, 所以 SRTM V4 用于 MOD09GA 地形校正。

2.2.4 Landsat 8 SR 数据

Landsat 8 SR 数据是 NASA 生产的 Landsat 8

OLI/TIRS 传感器的大气校正表面反射率数据产品。数据来源于(<https://earthexplorer.usgs.gov>.[2022-09-30]) 该反射率数据空间分辨率为 30 m, 时间分辨率为 16 日, WGS84-UTM 投影, 包含 5 个可见光和近红外波段、两个短波红外波段以及两个处理为正射校正亮温的热红外波段。本研究使用 Landsat 8 SR 数据来制备 FSC 真值 (L8-FSC), L8-FSC 主要用作模型的训练数据集, 同时也用来对产品进行验证。为了保证所选样本的可靠性, 对 Landsat 8 影像设定了以下的筛选条件: (1) 云覆盖度小于 5%; (2) 积雪覆盖率介于 30%—80%; (3) 影像均匀分布于亚洲高山区主要积雪区。最终在 2013 年—2018 年积雪期 (本年的 11 月 1 日到次年的 3 月 31 日) 共选取了 109 景 Landsat 8 SR 影像, 其中 73 景训练 LC-MARS 模型, 训练影像信息见表 2, 36 景作为 LC-MARS 模型反演 FSC 的精度验证, 验证影像信息见表 3, 训练及验证影像的分布情况见 http://www.coldregion.com/2024/01/05/scf_har/[2022-09-30]。

L8-FSC 提取方法分为两步。第 1 步: 从 Landsat 8 SR 数据中识别积雪像元生成积雪二值数据, 积雪像元识别算法分林区和非林区两种: (1) 非林区采用 Hall 等 (1995) 提出的 SNOMAP 算法 ($\text{NDSI} \geq 0.4$, $\text{B5} < 0.11$, $\text{B3} > 0.1$); (2) 林区采用 Wang 等 (2020) 提出的林地积雪识别算法; 第 2 步: 采用聚合的方法将 30 m 的积雪二值数据升尺度到 500 m 的 FSC 数据。聚合公式如式 (1) 所示。

$$\text{FSC}_i = \frac{s}{n} = \frac{\sum_{k=1}^s 1}{[\frac{500}{30}]^2} \quad (1)$$

式中, $[\]$ 表示取整, n 表示一个 500 m 分辨率像元内所包含的 30 m 分辨率像元的个数, s 表示在一个 500 m 分辨率像元内值为 1 的 30 m 分辨率像元的个数。

2.2.5 SnowCCI 和 MOD10A1

SnowCCI 是欧洲航天局在雪气候变化倡议中制备的全球逐日 MODIS FSC 产品, 该产品空间分辨率为 1 km, 覆盖时间从 2000 年—2020 年, 数据来源于 (<https://climate.esa.int/en/>[2022-09-30])。SnowCCI 参考的 FSC 提取算法为 Metsamaki 等 (2012) 提出的 SCAMod 方法, 这是一种基于光学反射率模型适用于北方森林和苔原的 FSC 提取方法。MOD10A1 是 NASA 发布的 MODIS 逐日积雪产品, 包含 NDSI、

积雪反照率、积雪面积比例和质量评估 (QA) 数据, 空间分辨率 500 m, 数据来源于 (<https://www.earthdata.nasa.gov>.[2022-09-30])。

表 2 Landsat 8 SR 训练影像信息

Table 2 Landsat 8 SR training image information

行列号	日期	云量	行列号	日期	云量
130037	2015-01-11	4.62	139038	2014-12-25	2.21
131038	2019-03-18	0.97	139039	2015-01-10	2.17
131039	2016-03-25	1.35	140035	2019-10-27	1.55
132039	2019-12-06	2.08	140038	2020-01-31	3.26
133038	2018-12-26	2.62	141034	2013-12-20	1.36
133040	2014-01-13	0.97	141035	2013-11-18	1.32
134039	2013-11-17	2.09	141039	2015-01-24	3.44
135039	2016-03-05	2.51	142034	2015-12-17	3.17
136039	2014-02-19	1.23	142036	2020-01-13	2.28
137040	2019-01-07	1.72	142037	2020-01-13	4.64
138040	2019-01-14	2.3	143030	2016-01-25	1.49
140040	2018-12-27	2.33	143036	2021-11-22	0.28
141040	2015-01-24	4.65	143039	2015-02-07	1.89
145039	2019-12-17	1.79	144030	2018-01-05	1.68
147038	2017-02-24	2	144038	2018-12-07	4.18
149036	2015-12-18	1.01	144039	2015-02-14	3.77
150036	2017-02-13	0.68	145030	2016-11-06	1.06
152037	2017-02-11	2.7	145035	2016-01-23	1.76
131038	2018-11-10	1.15	145037	2020-02-03	3.77
132036	2019-12-06	1.69	145038	2015-01-20	3.22
132038	2019-12-06	0.63	147031	2013-11-12	0.59
133035	2018-11-08	1.05	147035	2014-12-17	1.31
133037	2020-11-29	1.54	147037	2015-01-18	0.6
135033	2014-12-29	1.93	148031	2017-01-14	1.23
135034	2019-12-27	4.48	148036	2016-12-13	2
135039	2018-11-22	1.01	149031	2019-02-28	1.85
136033	2014-11-02	1.98	149035	2014-01-29	1.9
136036	2013-12-17	2.84	150031	2017-12-30	1.77
136037	2020-02-04	1.92	150034	2015-01-07	1.89
137033	2019-12-25	3.99	151032	2014-02-28	0.95
137035	2018-11-20	3.9	152031	2017-01-10	0.81
137038	2018-12-22	1.95	152035	2017-12-28	1.06
137039	2018-12-22	0.8	153036	2016-01-15	1.47
138035	2019-10-13	2.72	153037	2019-01-23	1.01
138036	2019-10-13	2.97	154037	2020-02-02	1.89
139034	2019-01-21	4.99	155037	2014-02-08	1.11
139036	2014-12-25	4.72			

MOD10A1 产品使用了 Salomonsona 和 Appel (2004) 提出的 FSC 与 NDSI 的半经验线性模型算法提取 FSC。SnowCCI 与 MOD10A1 作为对比验证数据, 验证 LC-MARS 模型反演 FSC 的精度。

表3 Landsat 8 SR 验证影像信息

Table 3 Landsat 8 SR verification image information

行列号	日期	云量	行列号	日期	云量
130037	2018-11-19	4.24	141036	2013-11-18	4.37
131038	2018-11-10	1.15	142035	2020-11-28	4.69
132038	2016-02-29	0.91	142039	2015-01-31	4.45
133034	2014-11-13	1.55	142040	2020-03-17	3.6
133039	2018-12-26	2.68	143038	2016-01-25	4.17
134036	2018-12-17	1.21	144036	2020-01-11	3.06
134040	2018-02-16	2.22	145036	2014-02-18	3.04
135037	2019-01-25	1.13	146036	2017-02-01	1.51
136035	2019-12-18	4.82	147036	2018-12-28	1.5
136039	2019-01-16	1.86	148035	2013-11-03	1.87
136040	2014-02-19	3.15	148037	2015-02-10	0.94
137036	2017-11-01	1.43	149034	2016-01-19	1.97
137040	2019-01-07	1.72	150032	2020-12-22	0.49
138038	2016-01-22	1.85	151034	2017-12-05	1.75
139037	2015-01-10	3.87	152036	2019-03-05	1.39
139040	2019-01-05	1.42	153031	2014-12-27	0.84
139041	2018-12-20	1.25	154033	2015-02-04	1.38
140039	2018-01-25	4.68	156037	2020-01-15	1.25

3 方 法

基于LC-MARS模型反演亚洲高山区积雪面积比例。首先对MODIS影像进行地形校正，消除地形影响并结合亚洲高山区CGLS-LC100地表分类数据，根据不同地表覆盖类型选取相应的积雪识别特征构建LC-MARS模型。图1是LC-MARS模型

反演FSC的技术路线图。

3.1 地形校正

高亚洲地区主要积雪区大多位于山区，积雪又是各向异性反射体，由于地形引起积雪判别中NDSI等主要参数的变化，因此首先对Landsat 8和MODIS反射率进行地形校正。本研究使用的地形校正方法是C校正，C校正是Teillet等（1982）提出的一种使用像素亮度值与太阳入射角余弦值的经验校正，是基于DEM地形校正的朗伯体反射率模型。该模型既考虑了物理机制，又使用了经验方法。C校正的实现原理如式（2）：

$$L_m = L \left(\frac{\cos \theta + c}{\cos i + c} \right)$$
 (2)

式中， L_m 表示校正后的亮度值， L 表示校正前的像素亮度值， θ 代表太阳天顶角， i 表示太阳水平入射角， c 表示 L 与 $\cos i$ 的回归方程中截距与斜率的比值。图2展示了2018年11月10日亚洲高山区一景Landsat 8 SR某一山区地形校正前后其NDSI单波段影像变化。受地形起伏影响，校正前的影像a阳坡和阴坡接受的太阳辐射能量不平衡，导致NDSI值存在明显的差异，阳坡过度感光NDSI值普遍较高，阴坡受阴影覆盖影响NDSI值较低。经地形校正处理后，影像b阴坡的NDSI值有明显提升，阳坡的NDSI值适当降低，有效的消除了地形起伏引起的地表辐射能量不平衡的影响。

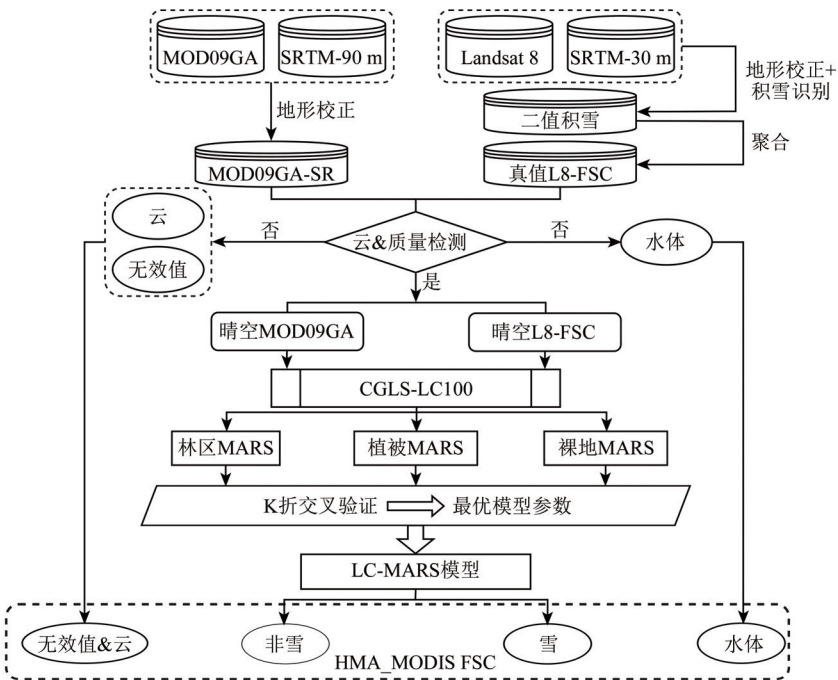


图1 LC-MARS模型反演FSC技术路线

Fig. 1 LC-MARS model inversion FSC technical route

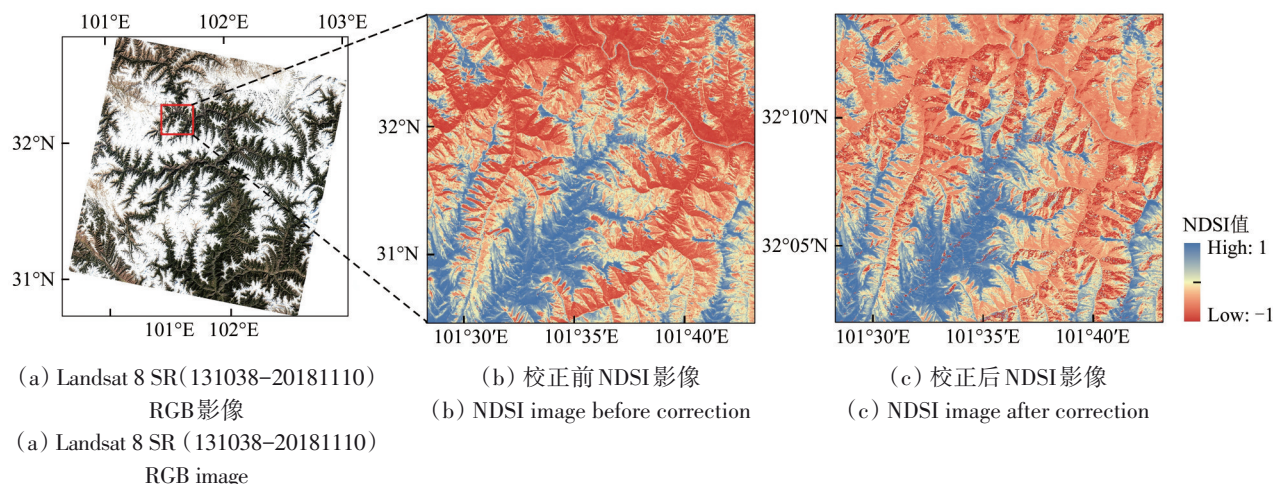


图2 NDSI地形校正前后对比

Fig. 2 Comparison before and after NDSI terrain correction

3.2 分地类特征选取

归一化差值积雪指数 NDSI 是积雪识别中最重要的光谱特征, 是根据积雪在可见光波段和近红外波段具有较高的反射率, 在短波红外波段具有较强的吸收这一光谱特性建立的 (Hall 和 Riggs, 1995)。NDSI 计算公式如式 (3):

$$\text{NDSI} = \frac{G - \text{SWIR}}{G + \text{SWIR}} \quad (3)$$

式中, G 为可见光绿光波段, SWIR 为短波红外波段。NDSI 是 MODIS 积雪信息识别的重要因子, 同时也在 FSC 反演中有广泛的应用 (Abdulkadhim, 2019; Wang 等, 2022; Salomonson 和 Appel, 2004)。但在亚洲高山区的一些林区和植被区, 下垫面类型复杂, 积雪分布破碎化, 混合像元的光谱与 NDSI 呈现非线性的关系, 仅考虑 NDSI 阈值难以准确提取积雪面积 (Hao 等, 2019; 雷华锦 等, 2020)。Klein 等 (1998) 研究表明, 在 NDVI 较高的林区, 积雪判别的 NDSI 阈值较低, 且 NDVI 越大, NDSI 阈值越低, 在林区加入 NDVI 指标可以提高林区积雪的反演精度。NDVI 计算公式如式 (4):

$$\text{NDVI} = \frac{\text{NIR} - R}{\text{NIR} + R} \quad (4)$$

式中, R 为可见光红波段, NIR 为近红外波段。Wang 等 (2020) 使用近红外波段代替 NDSI 计算公式中的可见光波段, 构建了归一化差值林地积雪指数 NDFS, 提出 NDSI 与 NDFS 结合的山区林地积雪制图方法, 在提取林区积雪时具有较高的精度, 表明 NDFS 在林区积雪判别中有一定的作用。NDFS 计算公式如式 (5):

$$\text{NDFS} = \frac{\text{NIR} - \text{SWIR}}{\text{NIR} + \text{SWIR}} \quad (5)$$

式中, NIR 为近红外波段, SWIR 为短波红外波段。考虑到亚洲高山区积雪下垫面复杂, 本研究采用分地类的特征选取提取积雪信息。研究区地表覆盖类型按 CGLS-LC100 重分类标准分为林区、植被、裸地和水体 4 类。在林区, 积雪识别同时受林区冠层遮挡和植被覆盖的影响, 因此选择 NDVI、NDSI 和 NDFS 为该地类积雪识别的输入特征; 在植被覆盖区, 积雪识别信息不受森林冠层的干扰, 只与植被覆盖度有关, 故选择 NDSI 和 NDVI 为草地积雪识别的特征; 在提取裸地积雪信息时, 由于裸地地势平坦且地表类型简单, 积雪信息识别不受混合像元的影响, 积雪覆盖面积与 NDSI 呈现较好的线性关系, 故仅考虑 NDSI 为提取积雪信息的光谱特征。分地类特征选取的方案如图 3 所示。

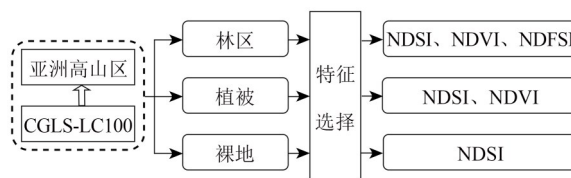


图3 分地类特征选取方案

Fig. 3 Feature selection scheme of land classification

3.3 构建 LC-MARS 模型

多元自适应回归样条 (MARS) 是由统计学家 Jerome Friedman 于 1991 年提出的一种数据分析方法 (Friedman, 1991)。MARS 是专门针对高维数

据的非线性回归算法, MARS 模型原理为式 (6):

$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i B_i(x) \tag{6}$$

式中, $f(x)$ 为 MARS 模型预测值, β_0 为截距, $B_i(x)$ 是第 i 个基函数, α_i 是相应基函数的系数, n 为基函数个数的最大值。基函数为一组线性样条函数, 是组成 MARS 模型的基本单元。基函数的表达形式如式 (7) 所示:

$$\begin{aligned} (x - t)_+ &= \begin{cases} x - t, & x > t \\ 0, & x \leq t \end{cases} \\ (t - x)_+ &= \begin{cases} t - x, & x < t \\ 0, & x \geq t \end{cases} \end{aligned} \tag{7}$$

式 (7) 表达的是一对镜面基函数, t 为结点的位置; $(x - t)_+$ 和 $(t - x)_+$ 为描述结点右侧和左侧的样条函数; 表示对负值取 0, 镜面基函数的图像如图 4 所示。

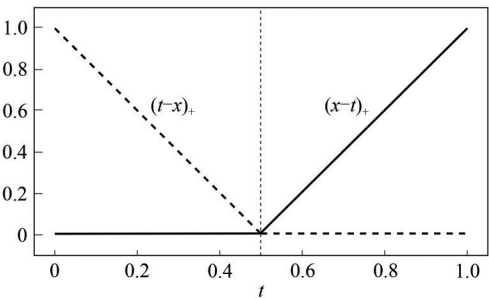


图 4 镜面基函数
Fig. 4 Specular basis function

MARS 模型的构建分“前向逐步”和“后向剪枝”两个过程。第 1 步, 前向逐步增加基函数。每

次选取一对最优的基函数并找到所有可能的结点来提高模型预测性能; 第 2 步, 后向剪枝, 在第一步过程中加入过多的基函数会造成模型的过拟合, 后向剪枝就是去除对模型贡献较小的冗余基函数, 获得最佳模型 (赵忠国 等, 2021)。根据分地类特征选取原则, 将 LC-MARS 模型分为 3 地类子模型, 分别为林区 MARS、植被 MARS 和裸地 MARS。决定 MARS 模型回归性能的参数主要有两个: 基函数的最大数量 (max_term) 和每个基函数中输入最大交互变量数 (max_degree)。通过训练 3 种地类子模型确定各自最优参数进而构建 LC-MARS 模型, 训练过程如下: (1) 将亚洲高山区的林区、植被和裸地的 L8-FSC 作为真值, 林区、植被和裸地所选的特征为样本空间; (2) 把每类地物的真值和特征数据输入到 MARS 模型中并通过 K 折交叉验证训练每类地物的 MARS 模型, 使用平均绝对误差的绝对值 (ABS_MAE) 作为模型性能评估指标, ABS_MAE 值越小, 模型的预测性能越好。图 5 展示了 3 种地类子模型参数训练结果, 图 5 (a) 为林区 MARS 模型参数训练结果, 林区 MARS 模型参数 max_degree=3 和 max_term=15 时, ABS_MAE 值最小, 模型性能最优; 图 5 (b) 为植被 MARS 模型参数训练结果, 植被 MARS 模型参数 max_degree=2 和 max_term=12 时, ABS_MAE 值最小, 模型性能最优; 图 5 (c) 为裸地 MARS 模型参数训练结果, 裸地 MARS 模型参数 max_degree=1 和 max_term=7 时, ABS_MAE 值最小, 模型性能最优。

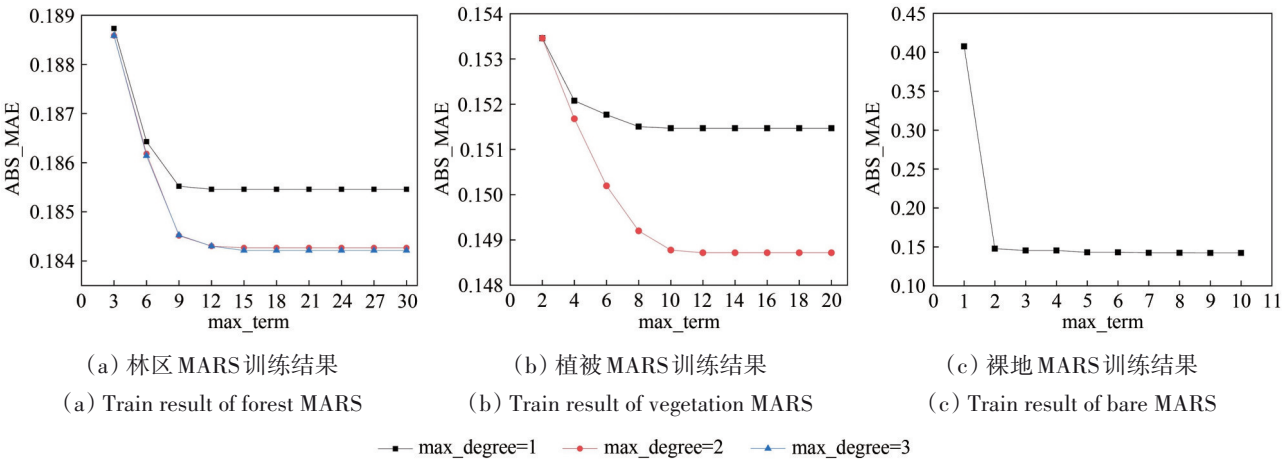


图 5 LC-MARS 模型 max_degree 和 max_term 参数训练变化
Fig. 5 LC-MARS model max_degree and max_term parameter training changes

3.4 精度评估算法

3.4.1 二值验证

二值验证中使用了一组二进制指标，这些二进制指标对应一个MODIS像素是否包含雪的4种可能的结果，分别为：真阳性（TP）、真阴性（TN）、假阳性（FP）和假阴性（FN）。4种可能的结果定义见表4。本研究将MODIS像素有雪和无雪的阈值设置为FSC=15%（Painter等，2009；Rittger等，2013），当MODIS像素的FSC>15%时，该像素被判定为雪，反之，则为无雪。

表4 二值积雪分类表

Table 4 Binary snow classification table		
MODIS-FSC data	L8-FSC(真值FSC)	
	雪	天雪
雪	TP	FP
天雪	FN	TN

FSC 二值验证的评估指标为总体精度(Accuracy)、召回率(Recall)、分类精度(Precision)。总体精度是指一个像元被正确分类的概率，是正确分类的像元数与总像元数的比值；召回率指的是正确分类的积雪像元数与实际所有的积雪像元数的比值；分类精度表示正确分类的积雪像元数与分类出的所有积雪像元数之间的比值。总体精度、召回率和分类精度的计算方法为式(8) — (10)：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (8)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

3.4.2 误差验证

误差验证是用高空间分辨率的L8-FSC作为参考真值，计算MODIS数据反演的FSC与L8-FSC之间的偏差。本研究所用的误差验证的指标为均方根误差RMSE（Root Mean Square Error）和平均绝对误差MAE（Mean Absolute Error）。RMSE和MAE的计算公式如式(11)、(12)。

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{FSC}_i^{\text{MOD}} - \text{FSC}_i^{\text{L8}})^2} \quad (11)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\text{FSC}_i^{\text{MOD}} - \text{FSC}_i^{\text{L8}}| \quad (12)$$

4 结果与精度评估

4.1 结果

4.1.1 LC-MARS模型导出结果

通过K折交叉验证分地类训练LC-MARS模型，确定林区MARS模型、植被MARS模型和裸地MARS模型的最优参数，设定最优参数的MARS模型以基函数和系数的形式导出，表5—7分别为林区MARS模型、植被MARS模型和裸地MARS模型所导出的基函数列表。

表5中林区MARS模型导出的基函数和系数进行线性组合构成林区FSC与林区所选特征NDVI、NDSI和NDFS I之间的非线性函数关系。表达式如式(13)所示：

$$F_{\text{FSC}} = 0.3064 + \sum_{i=1}^{10} \alpha_i \cdot F_i(\text{NDVI}, \text{NDSI}, \text{NDFS I}) \quad (13)$$

式中， F_{FSC} 为林区FSC， $F_i(\text{NDVI}, \text{NDSI}, \text{NDFS I})$ 表示第*i*个林区MARS基函数， α_i 为第*i*个林区基函数的系数。

表5 林区MARS模型基函数

Table 5 Basis functions of MARS model in forest area		
编号(<i>i</i>)	基函数(<i>F</i>)	系数(α)
—	截距	+0.3064
F_1	$\max(0, \text{NDFS I}-0.3446)$	+2.3314
F_2	$\max(0, 0.3446-\text{NDFS I})$	-0.8960
F_3	$\max(0, \text{NDVI}-0.1919)$	-0.7528
F_4	$\max(0, 0.1919-\text{NDVI})$	+0.1115
F_5	$\max(0, \text{NDSI}+0.4499) \cdot \max(0, 0.3446-\text{NDFS I})$	-0.7060
F_6	$\max(0, -0.4499-\text{NDSI}) \cdot \max(0, 0.3446-\text{NDFS I})$	+0.8463
F_7	$\max(0, \text{NDSI}-0.2433)$	-0.6695
F_8	$\max(0, 0.2433-\text{NDSI})$	+0.2429
F_9	$\max(0, \text{NDFS I}-0.6159) \cdot \max(0, \text{NDFS I}-0.3446)$	-1.7362
F_{10}	$\max(0, 0.6159-\text{NDFS I}) \cdot \max(0, \text{NDFS I}-0.3446)$	-0.6795

表6中植被MARS模型导出的基函数和系数进行线性组合为植被区FSC与植被区所选特征NDVI和NDSI之间的非线性函数关系，表达式如式(14)所示：

$$V_{\text{FSC}} = 0.7821 + \sum_{j=1}^{11} \beta_j \cdot V_j(\text{NDVI}, \text{NDSI}) \quad (14)$$

式中， V_{FSC} 为植被FSC， $V_j(\text{NDVI}, \text{NDSI})$ 表示第*j*个植被MARS基函数， β_j 为第*j*个植被基函数的系数。

表6 植被 MARS 模型基函数
Table 6 Basis functions of MARS model in vegetation area

编号(<i>j</i>)	基函数(<i>V</i>)	系数(<i>β</i>)
—	截距	+0.7821
<i>V</i> ₁	max(0, −0.2768−NDSI)	−1.0616
<i>V</i> ₂	max(0, NDSI−0.4478)·max(0, NDSI+0.2768)	−2.0651
<i>V</i> ₃	max(0, 0.4478−NDSI)·max(0, NDSI+0.2768)	+1.4733
<i>V</i> ₄	max(0, NDVI−0.2320)	+0.0658
<i>V</i> ₅	max(0, 0.2320−NDVI)	−1.3970
<i>V</i> ₆	max(0, NDSI−0.4352)	+2.6178
<i>V</i> ₇	max(0, 0.4352−NDSI)	−1.0298
<i>V</i> ₈	max(0, NDSI + 0.1034)·max(0, 0.4352−NDSI)	−2.1499
<i>V</i> ₉	max(0, −0.1034−NDSI)·max(0, 0.4352−NDSI)	+1.2379
<i>V</i> ₁₀	max(0, NDSI+0.1948)·max(0, 0.2320−NDVI)	+1.5105
<i>V</i> ₁₁	max(0, −0.1948−NDSI)·max(0, 0.2320−NDVI)	+2.1519

表7中对裸地 MARS 模型导出的基函数和系数进行线性组合得到裸地 FSC 与裸地积雪识别特征 NDSI 之间的函数关系, 如式 (15) 所示:

$$B_{\text{FSC}} = 0.6025 + \sum_{k=1}^5 \chi_k \cdot B_k(\text{NDSI}) \tag{15}$$

式中, *B*_{FSC} 为裸地 FSC, *B_k*(NDSI) 表示第 *k* 个裸地 MARS 基函数, *χ_k* 为第 *k* 个裸地基函数的系数。上述式 (13)、(14) 和 (15) 分别为亚洲高山林区 FSC、植被 FSC 和裸地 FSC 的反演算法。

表7 裸地 MARS 模型基函数
Table 7 Basis functions of MARS model in vegetation area

编号(<i>k</i>)	基函数(<i>B</i>)	系数(<i>χ</i>)
—	截距	+0.6025
<i>B</i> ₁	max(0, −0.183687−NDSI)	+0.0288
<i>B</i> ₂	max(0, 0.596954−NDSI)	−1.1126
<i>B</i> ₃	max(0, 0.223459−NDSI)	+0.7618
<i>B</i> ₄	max(0, NDSI+0.277521)	+0.3568
<i>B</i> ₅	max(0, −0.277521−NDSI)	+0.3162

4.1.2 基于 LC-MARS 模型的 HMA FSC 产品

基于 Google Earth Engine 云平台, 针对 LC-MARS 的 3 种地类子模型导出结果, 分地类计算 *F_{FSC}*, *V_{FSC}*, *B_{FSC}* 然后镶嵌为 L1 级 FSC 产品。L1 级 FSC 低于 10% 的误差较大, 需要进行温度掩膜消除误差, 温度掩膜通过设定 DEM 和地表温度 (LST) 阈值判断雪和非雪, 非雪条件为 LST≥275 K 且 DEM≤1300 m 或者 LST≥281 K 且 DEM≥1300 m (Hao 等, 2021)。本研究基于 LC-MARS 制备了 MODIS 亚洲

高山区逐日 FSC 初级产品, 见 http://www.coldregion.com/2024/01/05/scf_har/[2022-09-30] 中的图 3, 展示了 2018 年 11 月 10 日的亚洲高山区 FSC。

4.2 精度评估

本研究用亚洲高山区 36 景 L8-FSC 为真值验证 LC-MARS 模型反演 FSC 的精度, 从二值验证和误差验证两方面来评估 LC-MARS 模型反演亚洲高山区 FSC 的精度。

4.2.1 LC-MARS 二值验证

二值验证能够定量评估 LC-MARS 模型识别雪和非雪的准确度。亚洲高山区地表类型复杂, 不同下垫面的积雪识别难易程度不同, 为了全面评估 LC-MARS 模型在不同地表下分类雪和非雪的能力, 结合地表分类数据, 本研究统计了 LC-MARS 模型在总体和不同地表类型下二值验证的结果, 如表 8 所示: 可以看出 LC-MARS 模型反演的 FSC 在全地表覆盖类型下识别雪和非雪具有较高的准确度。总体 Accuracy=0.934, 表示全地表覆盖类型下, 正确分类的像元数占总的像元数的 93.4%; 总体 Recall=0.971, 表示全地表覆盖类型下, 正确分类的雪像元数占实际总的雪像元数的 97.1%; 总体 Precision=0.929, 表示全地表覆盖类型下, 分类为雪的所有像元中, 正确分类雪像元占 92.9%。对比不同地表分类下的二值验证评估指标 Accuracy、Recall 和 Precision, 其中裸地 3 项指标均高于植被, 植被三项指标均高于林区; 这表明 LC-MARS 模型在裸地识别雪和非雪的准确度最高, 其次是在植被区, 在林区识别雪和非雪的准确度相对最低, 这与积雪混合像元问题在林区最严重, 在植被区次之, 在裸地几乎不受影响现象相一致, 进一步表明二值验证结果的可靠性。

表8 总体及分地类二值验证
Table 8 Overall and Landcover binary verification

地表分类	样本数	Accuracy	Recall	Precision
林区	320169	0.880	0.892	0.783
植被	790984	0.937	0.974	0.939
裸地	646529	0.956	0.986	0.954
总体	1757682	0.934	0.971	0.929

4.2.2 LC-MARS 误差验证

误差验证可以获取 LC-MARS 模型反演的 FSC

与真值 L8-FSC 之间更具体的误差值，误差验证结果如表 9 所示，由 36 景验证影像验证的 LC-MARS 模型总体精度为 RMSE=0.148，MAE=0.093，总体精度较高。由于不同下垫面的积雪分布存在明显差异，为进一步验证 LC-MARS 模型在不同地表下反演 FSC 的精度，结合地表分类数据，表 9 也统计了 3 种下垫面 FSC 的验证结果。

表 9 总体及分地类误差验证

Table 9 Overall and landcover error validation

地表分类	样本数	RMSE	MAE
林区	320169	0.181	0.112
植被	790984	0.145	0.094
裸地	646529	0.133	0.082
总体	1757682	0.148	0.093

从表 9 可以看出，林区、植被和裸地的 RMSE 均小于 0.2，3 种地表的 FSC 精度都较高；分析 RMSE 和 MAE 两类评估指标，其中林区 RMSE 和 MAE 分别为 0.181、0.112，植被 RMSE 和 MAE 分别为 0.145、0.094，裸地 RMSE 和 MAE 分别为 0.133、0.082。两类精度评估指标均表明，裸地 FSC 精度最高，植被 FSC 次之，林区 FSC 精度较低。3 种地表的 FSC 精度差异是由于积雪在不同下垫面分布特征差异所引起的，林区地形复杂，混合像元较多，积雪呈斑状分布，积雪识别受森林冠层阴影的影响，所以精度较差；植被地形较为平缓，且没有高大树木遮挡积雪信息，但仍存在混合像元问题，所以精度略低；裸地地形平坦，积雪期地表几乎无作物生长，积雪分布连续且范围广，几乎无混合像元问题，所以精度较高。

4.2.3 误差来源分析

亚洲高山区地形地貌复杂，地貌分布受海拔影响较大，不同地貌的积雪分布特征不同致使误差存在差异，因此本文统计了不同海拔区间 LC-MARS 模型反演的 FSC 精度变化，结果如表 10 所示。可以看出，随着海拔逐渐上升，RMSE 值先增后减，其中海拔 500 m 以下精度最高，海拔 2500—3500 m 精度最低。海拔低于 500 m 时，地貌多表现为平原，地形平坦，积雪识别受地形阴影影响较小，所以 FSC 精度最高。海拔在 2500—3500 m 范围时，地形复杂度较高且地表覆盖类型多样，积雪分布破碎化，积雪信息识别同时受到混合像元

和地形的影响，所以精度最低。随着海拔从 3500 m 逐渐上升至 5500 m，尽管地形依旧复杂，但植物量逐渐减少，地表覆盖类型趋于单一，积雪混合像元问题随之消失，所以 FSC 精度逐渐提高。

表 10 不同海拔区间误差变化

Table 10 Variation of error in different altitudes

海拔区间/m	样本数	RMSE	MAE
<500	2356	0.034	0.018
[500,1500]	96127	0.095	0.046
[1500,2500]	188096	0.148	0.081
[2500,3500]	236329	0.173	0.104
[3500,4500]	478267	0.156	0.098
[4500,5500]	637794	0.143	0.094
>5500	118713	0.123	0.085

本研究针对亚洲高山区积雪期积雪开展 FSC 制图。积雪期的积雪是不断变化的，包含降雪、积雪和融雪 3 个过程，3 个过程又分别对应不同的季节，地表林区、植被等光谱信息会随季节不断变化，从而导致积雪反演过程的不确定性。为此本文深入评价了 LC-MARS 模型在不同积雪期的表现，为避免高程变化影响，所选的验证影像都是同一行列号不同积雪期的，验证结果如表 11 所示，积雪期 RMSE 最小，FSC 精度最高；融雪期次之；降雪期 FSC 反演精度最低。降雪期和融雪期，对应积雪从无到有和逐渐消融，期间积雪动态变化明显，LC_MARS 模型难以准确监测，所以降雪期和融雪期 FSC 精度较差。

表 11 不同积雪期误差变化

Table 11 Variation of error in different snow cover periods

积雪期	样本数	月份	RMSE
降雪	199275	9,10,11	0.186
积雪	233195	12,1,2	0.161
融雪	205772	3,4,5	0.179

亚洲高山区地形崎岖起伏，地表同一地物接受太阳辐射能量不均衡，处于阳坡的积雪像元具有较高的辐亮度，而阴坡的积雪多数被阴影覆盖，辐亮度较低，这种差异影响山区积雪信息的提取精度。本文通过地形校正来消除这种影响，为验证地形校正对山区积雪反演精度的影响，本研究在青藏高原东南地区、横断山脉和祁连山脉分别选取一处复杂地形，并对 3 处地形校正前后 LC-

MARS 模型反演 FSC 精度进行对比分析, 3 处复杂地形分布见 http://www.coldregion.com/2024/01/05/scf_har/[2022-09-30]中的图 4, 表 12 为地形校正前后精度验证结果。可以看出在藏东南、横断山脉、祁连山脉所选的 3 处复杂地形 FSC 反演精度普遍较低, 但经过地形校正后, FSC 精度均有提升, 其中藏东南山区 RMSE 提高了 0.022; 横断山脉 RMSE 提高了 0.011; 祁连山脉 RMSE 提高了 0.012。结果表明, 地形校正能有效的消除山区复杂地形对积雪信息识别的影响, 进而提高 FSC 反演精度。

表 12 地形校正前后 FSC 精度对比

Table 12 Comparison of FSC accuracy before and after mountainous terrain correction

山区	地形校正	RMSE	MAE
藏东南	校正前	0.202	0.154
	校正后	0.180	0.134
横断山脉	校正前	0.221	0.123
	校正后	0.210	0.119
祁连山脉	校正前	0.197	0.121
	校正后	0.185	0.114

4.2.4 LC-MARS 模型与线性回归模型的比较

本研究通过对目标区域样本选取构建 LC-MARS 模型并提取亚洲高山区 FSC。考虑到样本选取存在区域限制因素, 为了客观评估 LC-MARS 模型自身的算法优势, 本研究用与训练 LC_MARS 模型同一区域相同的训练样本拟合线性回归算法, 并用 36 景验证样本比较拟合的线性回归模型与 LC-MARS 模型反演 FSC 精度。建立了两种线性回归模型, 分别为 NDSI、NDFS I 与 FSC 的线性回归模型 (NN-Linear)、URSI、NDFS I 与 FSC 的线性回归模型 (UN-Linear)。线性回归模型拟合结果如式 (16)、(17) 所示:

$$FSC_{NN-Linear} = 0.421 \times NDSI + 0.581 \times NDFS I + 0.233$$

(16)

$$FSC_{UN-Linear} = 0.634 \times URSI + 0.671 \times NDFS I - 0.072$$

(17)

图 6 为 36 景验证影像下 LC-MARS 模型与 NN-Linear、UN-Linear 反演 FSC 总体精度和分地类精度的雷达分布图。4 个方向轴线分别表示总体、林区、植被和裸地的 RMSE 值, 可以看出 LC-MARS 模型雷达图面积远小于 NN-Linear 和 UN-Linear,

且在每个轴线方向上 LC-MARS 模型 RMSE 值均远小于 NN-Linear 和 UN-Linear, 这表明相比于线性回归模型, LC-MARS 模型在任一地类上反演 FSC 均具有较大的优势。NN-Linear 雷达图面积最大, 精度最低; 其次是 UN-Linear; LC-MARS 面积最小, 精度最高。UN-Linear 总体和分地类精度略高于 NN-Linear, 这是由于 URSI 对土地覆盖类型的敏感性较弱, 相比 NDSI 是一种较为稳定的积雪指数 (Wang 等, 2021), 所以由 URSI、NDFS I 构建的 UN-Linear 模型能减少复杂地表类型对 FSC 反演精度的影响。图 6 中 3 种 FSC 提取算法均表现为林区精度最低, 裸地最高; 这表明 LC-MARS 模型在反演林区积雪时, 还会受林区复杂地形地貌特征制约的影响, 仅通过地形校正并未完全消除地形因素对积雪反演的影响, 山区阴影下的积雪识别还需要进一步探究, 其次 LC-MARS 模型由于是区域样本训练, 故仅适用于亚洲高山区范围的 FSC 反演。

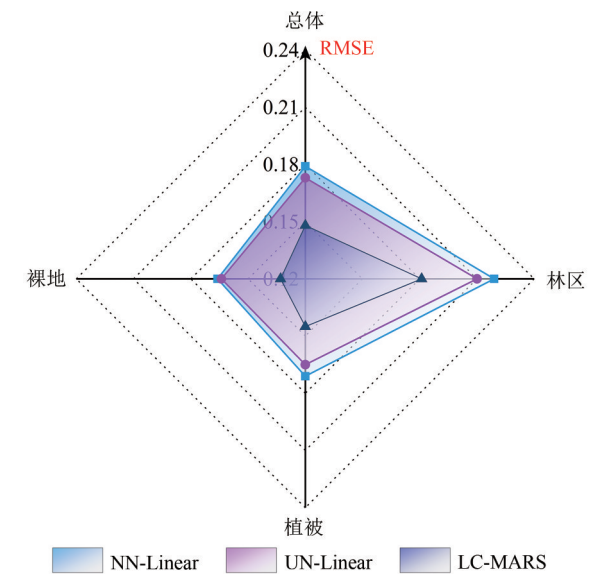


图 6 LC-MARS 与 NN-Linear、UN-Linear 模型精度对比
Fig. 6 LC-MARS and NN-Linear, UN-Linear model accuracy comparison

5 讨 论

5.1 与其他 FSC 产品的比较

本研究的主要目的是制备亚洲高山区长时间序列的 FSC 产品, 为了从产品精度层面评估 LC-MARS 模型反演亚洲高山区 FSC 的有效性, 本研究

用LC-MARS模型反演的FSC与现有的MOD10A1、SnowCCI两种知名的MODIS FSC全球产品进行比较, 3种FSC产品精度验证的参考真值还是亚洲高山区36景L8-FSC。

二值验证评估3种FSC产品在积雪定性上的精度, 图7展示了LC-MARS模型反演的FSC与MOD10A1、SnowCCI的二值验证雷达图, 可以看出LC-MARS模型制备的FSC与MOD10A1、SnowCCI的二值验证精度存在明显差异。雷达图中面积越大表示积雪产品越准确, 从图7可以看出LC-MARS制备的FSC最精确, 其次是MOD10A1, SnowCCI精确度最差。与MOD10A1、SnowCCI相比, LC-MARS模型反演的FSC的Accuracy值最高, 表明LC-MARS模型在雪和非雪的分类中, 正确分类的比率高于MOD10A1和SnowCCI; LC-MARS模型的Recall值最高, 表明LC-MARS模型正确识别的雪像元数比MOD10A1和SnowCCI多; 针对Precision指标, LC-MARS模型略低于MOD10A1, 这表明相较于MOD10A1, LC-MARS模型在识别雪像元时, 将少数非雪识别为雪, 但综合考虑Accuracy、Recall和Precision等3类指标, LC-MARS模型识别积雪的准确率显著高于MOD10A1和SnowCCI。

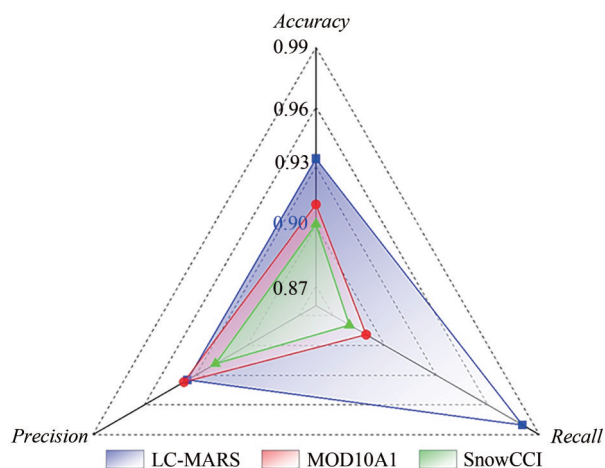


图7 LC-MARS与MOD10A1、SnowCCI二值验证对比图
Fig. 7 Comparison of LC-MARS and MOD10A1, SnowCCI binary verification

36景验证影像均匀分布于亚洲高山区, 3种FSC产品在每一景验证影像上的验证结果如图8所示, 可以明显看出与MOD10A1和SnowCCI相比, LC-MARS模型的RMSE均最低, 这表明在亚洲

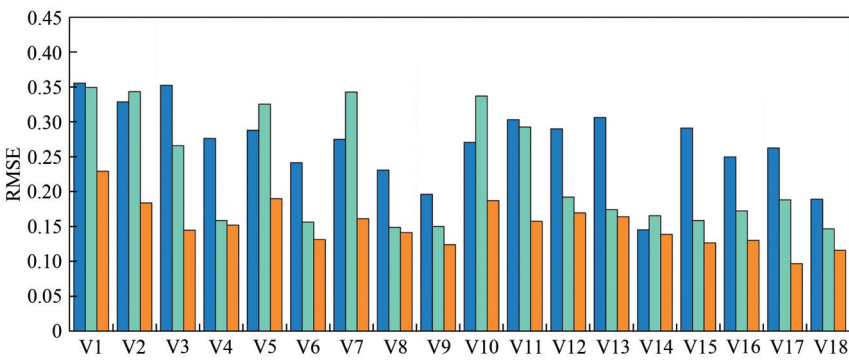
高山区, 由LC-MARS反演的FSC精度普遍高于MOD10A1和SnowCCI。MOD10A1和SnowCCI均为时空大尺度的全球FSC产品, 而本研究提出的LC-MARS是在亚洲高山区分地类训练得到, 为突出小尺度分地类选样训练对精度提高的效果, 表13展示了LC-MARS模型与MOD10A1、SnowCCI分地类对比验证结果。仅分析RMSE可以看出, 3类地表下LC-MARS模型的RMSE均最小, FSC精度最高; LC-MARS模型与MOD10A1、SnowCCI相比, 林区RMSE分别提高了0.055、0.136; 植被区RMSE分别提高了0.039、0.089; 裸地RMSE分别提高了0.02、0.077, 林区精度提高最多, 植被区次之, 裸地较少, 这表明分地类特征选取构建的LC-MARS制备的FSC和其他FSC产品相比, 能更精确的反映森林、灌木和草地等下垫面积雪的分布。

5.2 NDVI和NDFSI对LC-MARS精度的影响

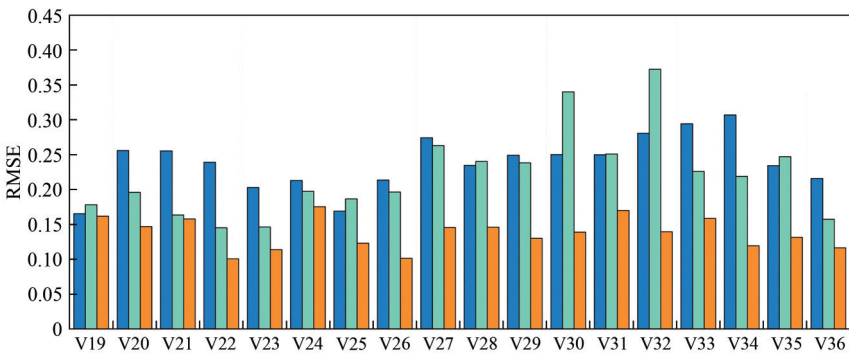
本研究通过分析不同地表覆盖下积雪分布的差异, 进而选取相应的积雪识别特征来构建LC-MARS模型, 在林区选取NDVI、NDSI、NDFSI, 在植被区选NDVI和NDSI, 在裸地选取NDSI。林区、植被和裸地均包含NDSI。为证实引入NDVI、NDFSI是否提高林区FSC反演精度以及NDFSI对林区精度提升的贡献, 引入NDVI是否对植被区FSC精度提高起作用, 本研究引入了两种MARS模型, 一是全地表覆盖类型下仅考虑NDSI特征构建的NDSI-MARS模型; 二是林区输入NDSI和NDFSI构建的NDFSI-MARS模型。以36景L8-FSC为参考真值, 比较NDSI-MARS模型、NDFSI-MARS和LC-MARS反演FSC的精度, 验证总体精度和分地类精度结果如表14所示。可以看出, LC-MARS模型总体RMSE、MAE为0.148、0.093, NDSI-MARS模型总体RMSE、MAE为0.154、0.099, 表明引入NDFSI、NDVI的LC-MARS模型总体精度高于单一特征的NDSI-MARS模型。在林区, LC-MARS模型RMSE、MAE分别为0.181、0.112; NDFSI-MARS模型RMSE、MAE分别为0.185、0.114; NDSI-MARS模型RMSE、MAE分别为0.192、0.119。结果表明引入NDFSI、NDVI后林区FSC精度有所提高, 其中RMSE提高了0.011, MAE提高了0.007; 在林区仅引入NDFSI, RMSE提高了0.007, MAE提高了0.005。植被区LC-MARS模型和NDSI-

MARS 模型 RMSE 分别为 0.145、0.153, MAE 分别为 0.094、0.104, 植被区 RMSE 提高了 0.008, MAE 提高了 0.01; 结果表明, 引入 NDVI 能够提高植被区 FSC 的反演精度。在裸地, LC-MARS 模型和

NDSI-MARS 模型的 MAE 相同, RMSE 仅相差 0.001, 这表明在裸地, 两种模型都仅考虑 NDSI 单一特征时, 最终反演的 FSC 精度基本一致, 进一步说明两模型对比验证结果的可靠性。



(a) 验证影像 V1-V18 精度对比
(a) Validation image V1-V18 accuracy comparison



(b) 验证影像 V19-V36 精度对比
(b) Validation image V19-V36 accuracy comparison

MOD10A1 SnowCCI LC-MARS

图 8 LC-MARS 与 MOD10A1、SnowCCI 验证影像精度对比
Fig. 8 Comparison of LC-MARS and MOD10A1, SnowCCI verification image accuracy

表 13 LC-MARS 模型与 MOD10A1、SnowCCI 分地类精度对比

Table 13 Comparison of LC-MARS model and MOD10A1, SnowCCI land classification accuracy			
地表分类	FSC 产品	RMSE	MAE
林区	MOD10A1	0.236	0.120
	SnowCCI	0.317	0.185
	LC-MARS	0.181	0.112
植被	MOD10A1	0.184	0.104
	SnowCCI	0.234	0.138
	LC-MARS	0.145	0.094
裸地	MOD10A1	0.153	0.082
	SnowCCI	0.210	0.086
	LC-MARS	0.133	0.082

表 14 NDSI-MARS 与 LC-MARS 精度验证

Table 14 NDSI-MARS and LC-MARS accuracy verification			
地表类型	模型	RMSE	MAE
林区	NDSI-MARS	0.192	0.119
	NDFS-MARS	0.185	0.114
	LC-MARS	0.181	0.112
植被	NDSI-MARS	0.153	0.104
	LC-MARS	0.145	0.094
裸地	NDSI-MARS	0.134	0.082
	LC-MARS	0.133	0.082
总体	NDSI-MARS	0.154	0.099
	LC-MARS	0.148	0.093

6 结 论

本研究基于MARS机器学习模型,考虑了亚洲高山区复杂地形地貌下积雪分布破碎化和混合像元问题对FSC估计的影响,分地类选取积雪识别特征构建LC-MARS模型制备亚洲高山区FSC。利用L8-FSC数据作为参考真值验证LC-MARS模型反演FSC精度,同时与MOD10A1、SnowCCI等全球FSC产品进行比较;探究了NDVI和NDSFI在LC-MARS模型中对山林区FSC反演精度提高的影响,并将LC-MARS模型移植在Google Earth Engine云平台,制备亚洲高山区FSC产品。研究结论如下:

(1) LC-MARS模型反演亚洲高山区FSC精度较高,其中总体RMSE、MAE分别为0.148、0.093;总体Accuracy、Recall和Precision分别为93.4%、97.1%、92.9%。LC-MARS模型提取山林区FSC的表现优于线性回归模型且LC-MARS反演的FSC产品在二值验证和误差验证上精度均高于MOD10A1和SnowCCI。

(2) 引入NDVI、NDSFI特征构建的LC-MARS与仅NDSI构建的NDSI-MARS相比;LC-MARS总体精度高于NDSI-MARS模型,分地类精度提高主要体现在林区、植被区, RMSE分别降低了0.011、0.008,有效的解决了林区、植被区积雪识别精度不足的问题。

(3) LC-MARS移植到Google Earth Engine云平台制备了2000年—2020年MODIS亚洲高山区逐日FSC产品。

总体而言,LC-MARS模型反演FSC精度较高,运算效率也较高,基于该模型制备的高精度的亚洲高山区FSC产品,优于国际同类产品,该产品可用于分析亚洲高山区积雪时空变化,同时也为研究亚洲高山区水文、气候和融雪径流模拟提供了重要的输入数据。

参考文献(References)

Abdulkadhim A H. 2019. Estimating snow cover area in south of Turkey using the Normalized Difference Snow Index (NDSI) form MODIS satellite images. *Journal of Physics: Conference Series*, 1279: 012047 [DOI: 10.1088/1742-6596/1279/1/012047]

Brun F, Berthier E, Wagnon P, Käb A and Treichler D. 2018. A spatially resolved estimate of High Mountain Asia glacier mass balances

from 2000 to 2016. *Nature Geoscience*, 11(7): 543 [DOI: 10.1038/s41561-018-0171-z]

Çiftçi B B, Kuter S, Akyürek Z and Weber G W. 2017. Fractional snow cover mapping by artificial neural networks and support vector machines. 4th International Geoadvances Workshop. Safranbolu, Karabuk: ISPRS: 179-187 [DOI: 10.5194/isprs-annals-IV-4-W4-179-2017]

Czyzowska-Wisniewski E H, Van Leeuwen W J D, Hirschboeck K K, Marsh S E and Wisniewski W T. 2015. Fractional snow cover estimation in complex alpine-forested environments using an artificial neural network. *Remote Sensing of Environment*, 156: 403-417 [DOI: 10.1016/j.rse.2014.09.026]

Friedman J H. 1991. Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1): 1-67 [DOI: 10.1214/AOS/1176347963]

García-Álvarez D, Hinojosa J L, Pérez J J and Villaraso J Q. 2022. Global general land use cover datasets with a time series of maps// García-Álvarez D, Olmedo M T C, Paegelow M and Mas J F, eds. *Land Use Cover Datasets and Validation Tools: Validation Practices with QGIS*. Cham: Springer: 287-311 [DOI: 10.1007/978-3-030-90998-7_15]

Guo Z M, Geng L, Shen B S, Wu Y W, Chen A A and Wang N L. 2021. Spatiotemporal variability in the glacier snowline altitude across high mountain Asia and potential driving factors. *Remote Sensing*, 13(3): 425 [DOI: 10.3390/rs13030425]

Hall D K, Riggs G A and Salomonson V V. 1995. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, 54(2): 127-140 [DOI: 10.1016/0034-4257(95)00137-P]

Hao S R, Jiang L M, Shi J C, Wang G X and Liu X J. 2019. Assessment of MODIS-based fractional snow cover products over the Tibetan Plateau. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(2): 533-548 [DOI: 10.1109/Jstars.2018.2879666]

Hao X H, Huang G H, Che T, Ji W Z, Sun X L, Zhao Q, Zhao H Y, Wang J, Li H Y and Yang Q. 2021. The NIEER AVHRR snow cover extent product over China-a long-term daily snow record for regional climate research. *Earth System Science Data*, 13(10): 4711-4726 [DOI: 10.5194/essd-13-4711-2021]

Hao X H, Huang G H, Zheng Z J, Sun X L, Ji W Z, Zhao H Y, Wang J, Li H Y and Wang X Y. 2022. Development and validation of a new MODIS snow-cover-extent product over China. *Hydrology and Earth System Sciences*, 26(8): 1937-1952 [DOI: 10.5194/hess-26-1937-2022]

Hou J L and Huang C L. 2014. Improving mountainous snow cover fraction mapping via artificial neural networks combined with MODIS and ancillary topographic data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9): 5601-5611 [DOI: 10.1109/Tgrs.2013.2290996]

Huang X D, Deng J, Wang W, Feng Q S and Liang T G. 2017. Impact of climate and elevation on snow cover using integrated remote sensing snow products in Tibetan Plateau. *Remote Sensing of En-*

- vironment, 190: 274-288 [DOI: 10.1016/j.rse.2016.12.028]
- Huang X D, Hao X H, Yang Y S, Wang W and Liang T G. 2012. Advances in snow-cover monitoring using optical remote sensing. *Pratacultural Science*, 29(1): 35-43 (黄晓东, 郝晓华, 杨永顺, 王伟, 梁天刚. 2012. 光学积雪遥感研究进展. *草业科学*, 29(1): 35-43)
- Huang X D, Liang T G, Zhang X T and Guo Z G. 2011. Validation of MODIS snow cover products using Landsat and ground measurements during the 2001-2005 snow seasons over northern Xinjiang, China. *International Journal of Remote Sensing*, 32(1): 133-152 [DOI: 10.1080/01431160903439924]
- Klein A G, Hall D K and Riggs G A. 1998. Improving snow cover mapping in forests through the use of a canopy reflectance model. *Hydrological Processes*, 12(10-11): 1723-1744 [DOI: 10.1002/(SICI)1099-1085(199808/09)12:10/11<1723::AID-HYP691>3.0.CO;2-2]
- König M, Winther J G and Isaksson E. 2001. Measuring snow and glacier ice properties from satellite. *Reviews of Geophysics*, 39(1): 1-27 [DOI: 10.1029/1999RG000076]
- Kuter S, Akyurek Z and Weber G W. 2018. Retrieval of fractional snow covered area from MODIS data by multivariate adaptive regression splines. *Remote Sensing of Environment*, 205: 236-252 [DOI: 10.1016/j.rse.2017.11.021]
- Lei H J, Li H Y, Wang J, Hao X H, Zhao H Y and Zhang J. 2020. MODIS fractional snow cover products preparing on Tibetan Plateau based on environmental information and regression model. *Remote Sensing Technology and Application*, 35(6): 1303-1311 (雷华锦, 李弘毅, 王建, 郝晓华, 赵宏宇, 张娟. 2020. 基于环境信息和回归模型的青藏高原MODIS积雪面积比例产品制备. *遥感技术与应用*, 35(6): 1303-1311) [DOI: 10.11873/j.issn.1004-0323.2020.6.1303]
- Li H S, Wu K and Xu Y. 2022. An integrated change detection method based on spectral unmixing and the CNN for hyperspectral imagery. *Remote Sensing*, 14(11): 2523 [DOI: 10.3390/rs14112523]
- Liang H, Huang X D, Sun Y H, Wang Y L and Liang T G. 2017. Fractional snow-cover mapping based on MODIS and UAV data over the Tibetan Plateau. *Remote Sensing*, 9(12): 1332 [DOI: 10.3390/rs9121332]
- Liang S L, Bai R, Chen X N, Cheng J, Fan W J, He T, Jia K, Jiang B, Jiang L M, Jiao Z T, Liu Y B, Ni W J, Qiu F, Song L L, Sun L, Tang B H, Wen J G, Wu G P, Xie D H, Yao Y J, Yuan W P, Zhang Y G, Zhang Y T, Zhang Y T, Zhang X T, Zhao T J and Zhao X. 2020. Review of China's land surface quantitative remote sensing development in 2019. *Journal of Remote Sensing (in Chinese)*, 24(6): 618-671 (梁顺林, 白瑞, 陈晓娜, 程洁, 范闻捷, 何涛, 贾坤, 江波, 蒋玲梅, 焦子锦, 刘元波, 倪文俭, 邱风, 宋柳霖, 孙林, 唐伯惠, 闻建光, 吴桂平, 谢东辉, 姚云军, 袁文平, 张永光, 张玉珍, 张云腾, 张晓通, 赵天杰, 赵祥. 2020. 2019年中国陆表定量遥感发展综述. *遥感学报*, 24(6): 618-671) [DOI: 10.11834/jrs.20209476]
- Liu S Y, Ding Y J, Ye B S, Wang N L and Xie Z C. 2000. Regional characteristics of glacier mass balance variations in high Asia. *Journal of Glaciology and Geocryology*, 22(2): 97-105 (刘时银, 丁永建, 叶佰生, 王宁练, 谢自楚. 2000. 高亚洲地区冰川物质平衡变化特征研究. *冰川冻土*, 22(2): 97-105) [DOI: 10.3969/j.issn.1000-0240.2000.02.001]
- Lu R X, Qiu G H, Zhang Z J, Deng X Z, Yang H, Zhu Z M and Zhu J Y. 2021. A mixture varying-gain dynamic learning network for solving nonlinear and nonconvex constrained optimization problems. *Neurocomputing*, 456: 232-242 [DOI: 10.1016/j.neucom.2021.05.037]
- Metsämäki S, Mattila O P, Pulliainen J, Niemi K, Luojus K and Böttcher K. 2012. An optical reflectance model-based method for fractional snow cover mapping applicable to continental scale. *Remote Sensing of Environment*, 123: 508-521 [DOI: 10.1016/j.rse.2012.04.010]
- Notarnicola C. 2020. Hotspots of snow cover changes in global mountain regions over 2000-2018. *Remote Sensing of Environment*, 243: 111781 [DOI: 10.1016/j.rse.2020.111781]
- Painter T H, Rittger K, McKenzie C, Slaughter P, Davis R E and Dozier J. 2009. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, 113(4): 868-879 [DOI: 10.1016/j.rse.2009.01.001]
- Pan X D, Guo X J, Li X, Niu X L, Yang X J, Feng M, Che T, Jin R, Ran Y H, Guo J W, Hu X L and Wu A D. 2021. National Tibetan Plateau data center: promoting earth system science on the third pole. *Bulletin of the American Meteorological Society*, 102(11): E2062-E2078 [DOI: 10.1175/Bams-D-21-0004.1]
- Rittger K, Painter T H and Dozier J. 2013. Assessment of methods for mapping snow cover from MODIS. *Advances in Water Resources*, 51: 367-380 [DOI: 10.1016/j.advwatres.2012.03.002]
- Rosenthal W and Dozier J. 1996. Automated mapping of montane snow cover at subpixel resolution from the Landsat Thematic Mapper. *Water Resources Research*, 32(1): 115-130 [DOI: 10.1029/95WR02718]
- Salomonson V V and Appel I. 2004. Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote Sensing of Environment*, 89(3): 351-360 [DOI: 10.1016/j.rse.2003.10.016]
- Su X, Jiang L M, Wang G X and Wang J. 2020. The validation of snow cover product over high mountain Asia//*Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. Waikoloa: IEEE: 2946-2949 [DOI: 10.1109/Igarss39084.2020.9324316]
- Sun X L, Hao X H, Wang J, Zhao H Y and Ji W Z. 2022. Research on retrieval of MODIS fraction snow cover based on spectral environmental random forest regression model. *Journal of Glaciology and Geocryology*, 44(1): 147-158 (孙兴亮, 郝晓华, 王建, 赵宏宇, 纪文政. 2022. 基于光谱—环境随机森林回归模型的MODIS积雪面积比例反演研究. *冰川冻土*, 44(1): 147-158) [DOI: 10.7522/j.issn.1000-0240.2022.0026]
- Tang Z G, Deng G, Hu G J, Wang X, Jiang Z L and Sang G Q. 2021.

- Spatiotemporal dynamics of snow phenology in the high mountain Asia and its response to climate change. *Journal of Glaciology and Geocryology*, 43(5): 1400-1411 (唐志光, 邓刚, 胡国杰, 王欣, 蒋宗立, 桑国庆. 2021. 亚洲高山区积雪物候时空动态及其对气候变化的响应. *冰川冻土*, 43(5): 1400-1411) [DOI: 10.7522/j.issn.1000-0240.2021.0092]
- Tang Z G, Wang X R, Wang J, Wang X, Li H Y and Jiang Z L. 2017. Spatiotemporal variation of snow cover in Tianshan Mountains, Central Asia, based on cloud-free MODIS fractional snow cover product, 2001-2015. *Remote Sensing*, 9(10): 1045 [DOI: 10.3390/rs9101045]
- Teillet P M, Guindon B and Goodenough D G. 1982. On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, 8(2): 84-106 [DOI: 10.1080/07038992.1982.10855028]
- Wang G X, Jiang L M, Shi J C and Su X. 2021. A universal ratio snow index for fractional snow cover estimation. *IEEE Geoscience and Remote Sensing Letters*, 18(4): 721-725 [DOI: 10.1109/lgrs.2020.2982053]
- Wang G X, Jiang L M, Xiong C and Zhang Y S. 2022. Characterization of NDSI variation: implications for snow cover mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 4304318 [DOI: 10.1109/Tgrs.2022.3165986]
- Wang J. 1999. Comparison and analysis on methods of snow cover mapping by using satellite remote sensing data. *Remote Sensing Technology and Application*, 14(4): 29-36 (王建. 1999. 卫星遥感雪盖制图方法对比与分析. *遥感技术与应用*, 14(4): 29-36) [DOI: 10.3969/j.issn.1004-0323.1999.04.005]
- Wang X Y, Chen S Y and Wang J. 2020. An adaptive snow identification algorithm in the forests of Northeast China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 5211-5222 [DOI: 10.1109/Jstars.2020.3020168]
- Ye C L, Liu S W, Xu M M, Du B, Wan J H and Sheng H. 2021. An endmember bundle extraction method based on multiscale sampling to address spectral variability for hyperspectral unmixing. *Remote Sensing*, 13(19): 3941 [DOI: 10.3390/rs13193941]
- Zhang C S, Ma L, Chen J, Rao Y H, Zhou Y and Chen X H. 2019. Assessing the impact of endmember variability on linear Spectral Mixture Analysis (LSMA): a theoretical and simulation analysis. *Remote Sensing of Environment*, 235: 111471 [DOI: 10.1016/j.rse.2019.111471]
- Zhao H Y, Hao X H, Wang J A, Li H Y, Huang G H, Shao D H, Su B, Lei H J and Hu X J. 2020. The Spatial-Spectral-environmental extraction endmember algorithm and application in the MODIS fractional snow cover retrieval. *Remote Sensing*, 12(22): 3693 [DOI: 10.3390/rs12223693]
- Zhao H Y, Hao X H, Zheng Z J, Wang J, Li H Y, Huang G H, Shao D H, Wang X, Gao Y and Lei H J. 2018. A new algorithm of fractional snow cover basing on FY-3D/MERSI-II. *Remote Sensing Technology and Application*, 33(6): 1004-1016 (赵宏宇, 郝晓华, 郑照军, 王建, 李弘毅, 黄广辉, 邵东航, 王轩, 高扬, 雷华锦. 2018. 基于FY-3D/MERSI-II的积雪面积比例提取算法. *遥感技术与应用*, 33(6): 1004-1016) [DOI: 10.11873/j.issn.1004-0323.2018.6.1004]
- Zhao Z G, Zhang F and Zheng J H. 2021. Evaluation of landslide susceptibility by multiple adaptive regression spline method. *Geomatics and Information Science of Wuhan University*, 46(3): 442-450 (赵忠国, 张峰, 郑江华. 2021. 多元自适应回归样条法的滑坡敏感性评价. *武汉大学学报(信息科学版)*, 46(3): 442-450) [DOI: 10.13203/j.whugis20190136]
- Zou F L, Hu Q W, Li H D, Lin J, Liu Y C and Sun F L. 2022. Dynamic disturbance analysis of grasslands using neural networks and spatiotemporal indices fusion on the Qinghai-Tibet Plateau. *Frontiers in Plant Science*, 12: 760551 [DOI: 10.3389/fpls.2021.760551]

Machine learning-based mapping of fraction snow cover in High Mountain Asia by MODIS

GAO Weiqiang^{1,2}, HAO Xiaohua², HE Dongcai¹, SUN Xingliang³, LI Hongyi², REN Hongrui¹, ZHAO Qin²

1. College of Mining Engineering, Taiyuan University of Technology, Taiyuan 030000, China;

2. Northwest Institute of Eco-environment and Resources Chinese Academy of Sciences, Lanzhou 730000, China;

3. College of Resources and Environment, Lanzhou University, Lanzhou 730000, China

Abstract: High Mountain Asia (HMA) is the richest high altitude region in the world except for the poles in terms of glacier and snow resources, The accurate monitoring of HMA snowpack distribution is important for HMA snowmelt runoff simulation, climate change prediction and ecosystem evolution. Fractional Snow Cover (FSC) can quantitatively describe the extent of snow cover at the sub-image scale, and is more suitable for reflecting the distribution of snow in complex mountainous areas than binary snow. The objective of this study is to develop a new HMA snow area ratio inversion algorithm and integrate the algorithm into Google Earth Engine to prepare a set of long time series HMA snow area ratio products.

Considering the influence of HMA topography and sub-bedding type on the accuracy of snow accumulation information extraction, this

paper proposes a Multivariate Adaptive Regression Splines (MARS) model LC-MARS to invert the proportion of snow accumulation area in Asia by integrating topography correction and subland class feature extraction. The FSC extracted by Landsat 8 is used as the true value, and the LC-MARS model is tested for inversion FSC accuracy using binary and error validation methods, and the performance of linear regression models trained with the same training samples and the LC-MARS model for inversion HMAFSC accuracy is compared, and the accuracy of the FSC inversion of the LC-MARS model with SnowCCI and MOD10A1 is also compared.

(1) The overall accuracy of FSC binary validation of LC-MARS model inversion showed that Accuracy and Recall were 93.4% and 97.1%, respectively, and the overall accuracy of error validation showed that RMSE was 0.148 and MAE was 0.093, both binary validation and error validation indicated that the FSC accuracy of LC-MARS model inversion was higher. (2) The LC-MARS model trained based on the same training samples has higher FSC accuracy than the linear regression model in forest area, vegetation and bare land inversions, indicating that the LC-MARS model is more suitable for FSC inversions in mountain and forest areas. (3) The overall RMSE of MOD10A1 is 0.178 and MAE is 0.096; the overall RMSE of SnowCCI is 0.247 and MAE is 0.131. The accuracy of FSC prepared by LC-MARS is higher than that of MOD10A1 and SnowCCI, indicating that FSC inversion by LC-MARS has some application value.

The LC-MARS model can fit high-dimensional nonlinear relationships and significantly improve the inversion accuracy of FSC in mountain and forest areas. The computational efficiency of the LC-MARS model based on Google Earth Engine is high, and it is suitable for preparing FSC products of large scale long time series. In this study, the day-by-day MODIS FSC products of HMA from 2000 to 2021 were prepared based on the LC-MARS model, which provides important data support for the study of climate change, hydrological and water resources in HMA.

Key words: remote sensing, High Mountain Asia (HMA), fractional snow cover, MODIS, MARS, terrain correction

Supported by National Natural Science Foundation of China (No. U22A20564, 41971325); National Key Research and Development Program of China (No. 2022YFF0711702-05)