

# 基于机器学习的MODIS亚洲高山区积雪面积比例制图研究

高伟强<sup>1,2</sup>, 郝晓华<sup>2</sup>, 和栋材<sup>1</sup>, 孙兴亮<sup>3</sup>, 李弘毅<sup>2</sup>, 任鸿瑞<sup>1</sup>, 赵琴<sup>2</sup>

1. 太原理工大学 矿业工程学院, 山西 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, 甘肃 兰州 730000

**摘要:** 积雪面积比例 (Fractional Snow Cover, FSC) 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 (High Mountain Asia, HMA) 基于分地类特征选择的多元自适应回归样条 (Multivariate Adaptive Regression Splines, MARS) 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 2000-2021 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:** 亚洲高山区, 积雪面积比例, MODIS, MARS, 地形校正

**中图分类号:** TP79

**引用格式:** 高伟强, 郝晓华, 和栋材, 孙兴亮, 李弘毅, 任鸿瑞, 赵琴. XXXX. 基于机器学习的 MODIS 亚洲高山区积雪面积比例制图研究. 遥感学报, XX(XX): 1-17

GAO Weiqiang, HAO Xiaohua, HE Dongcai, SUN Xingliang, LI Hongyi, REN Hongrui, ZHAO Qin. XXXX. Machine learning-based mapping of fraction snow cover in High Mountain Asia by MODIS. National Remote Sensing Bulletin, 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 (Fractional Snow Cover, FSC) 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 (Metsamaki 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 (High Mountain Asia, HMA) 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

收稿日期: XXXX-XX-XX; 预印本: XXXX-XX-XX

基金项目: 国家自然科学基金项目 U22A20564, 41971325, 国家重点研发计划 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, 研究方向积雪遥感、积雪与气候变化研究。E-mail: haoxh@lzb.ac.cn

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 (Konig 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 的方法主要有三种: 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 (Linear Spectral Mixture Analysis, LSMA) 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 等分别应用多元自适应回归样条 (Multivariate Adaptive Regression Splines, MARS) 和多层前馈人工神经网络 (Artificial Neural Networks, ANNs) 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 (Kuter 等, 2018); Czyzowska-Wisniewski 等在复

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 (Czyzowska-Wisniewski 等, 2015); Ciftci 等设计比较了 ANNs 和支持向量机 (Support Vector Machines, SVM) 用于 MODIS FSC 反演 (Ciftci 等, 2017)。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。

对比三类 FSC 反演模型, 其中基于线性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, 2004), 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°–45° N, 67°–107° E), 是地球上除南北极之外的第三大冰冻圈, 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/](http://www.coldregion.com/2024/01/05/scf_har/) 中图 1, 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 4000~5000m (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 能获得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 (Copernicus Global Land Service, CGLS) 分辨率为 100 米的土地覆盖数据 (CGLS-LC100), 航天飞机雷达地形任务 (Shuttle Radar Topography Mission, SRTM) 的 DEM 数据, Landsat-8 地表反射率数据 (Landsat-8 SR), 欧空局雪气候变化倡议 (Snow Climate Change Initiative, SnowCCI) 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/> [2023-02-02])。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, 其七个单波段反射率用于计算归一化植被指数 (Normalized Difference Vegetation Index, NDVI), 归一化积雪指数 (Normalized Difference Snow Index, NDSI), 归一化林地积雪指数 (Normalized Difference Forest Snow Index, NDFS) 和通用比率积雪指数 (Universal Ratio Snow Index, URSI)。NDVI、NDSI 和 NDFS 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/> [2023-02-02])。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, 为了便于分地类选择模型输入特征, 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 30m, 数据

来源于 (<https://cmr.earthdata.nasa.gov/> [2023-02-02]); SRTM V4 数据, 空间分辨率为 90m, 来源于 (<https://srtm.csi.cgiar.org/> [2023-02-

02] )。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 6500m，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>, [2023-02-02] ) 该反射率数据空间分辨率为 30m，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、2 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% 之间；(3) 影像均匀分布于亚洲高山区主要积雪区。最终在 2013-2018 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 1，36 景作为 LC-MARS 模型反演 FSC 的精度验证，验证影像信息见表 2，训练及验证影像的分布情况见 [http://www.coldregion.com/2024/01/05/scf\\_har/](http://www.coldregion.com/2024/01/05/scf_har/) 中图 2。

L8-FSC 提取方法分为 2 步。第一步：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 (Hall 等, 1995) 提出的 SNOMAP 算法 ( $NDVI > 0.4$ ,  $B5 < 0.11$ ,  $B3 > 0.1$ )；(2) 林区采用 Wang 等 (Wang 等, 2020) 提出的林地积雪识别算法；第二步：采用聚合的方法将 30m 的积雪二值数据升尺度到 500m 的 FSC 数据。聚合公式如式 (1) 所示：

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

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

500m 分辨率像元内值为 1 的 30m 分辨率像元的个数。

表 2 Landsat-8 SR 训练影像信息

Table2 Landsat-8 SR training image information

行列号	日期	云量	行列号	日期	云量
130037	20150111	4.62	139038	20141225	2.21
131038	20190318	0.97	139039	20150110	2.17
131039	20160325	1.35	140035	20191027	1.55
132039	20191206	2.08	140038	20200131	3.26
133038	20181226	2.62	141034	20131220	1.36
133040	20140113	0.97	141035	20131118	1.32
134039	20131117	2.09	141039	20150124	3.44
135039	20160305	2.51	142034	20151217	3.17
136039	20140219	1.23	142036	20200113	2.28
137040	20190107	1.72	142037	20200113	4.64
138040	20190114	2.3	143030	20160125	1.49
140040	20181227	2.33	143036	20211122	0.28
141040	20150124	4.65	143039	20150207	1.89
145039	20191217	1.79	144030	20180105	1.68
147038	20170224	2	144038	20181207	4.18
149036	20151218	1.01	144039	20150214	3.77
150036	20170213	0.68	145030	20161106	1.06
152037	20170211	2.7	145035	20160123	1.76
131038	20181110	1.15	145037	20200203	3.77
132036	20191206	1.69	145038	20150120	3.22
132038	20191206	0.63	147031	20131112	0.59
133035	20181108	1.05	147035	20141217	1.31
133037	20201129	1.54	147037	20150118	0.6
135033	20141229	1.93	148031	20170114	1.23
135034	20191227	4.48	148036	20161213	2
135039	20181122	1.01	149031	20190228	1.85
136033	20141102	1.98	149035	20140129	1.9
136036	20131217	2.84	150031	20171230	1.77
136037	20200204	1.92	150034	20150107	1.89
137033	20191225	3.99	151032	20140228	0.95
137035	20181120	3.9	152031	20170110	0.81
137038	20181222	1.95	152035	20171228	1.06
137039	20181222	0.8	153036	20160115	1.47
138035	20191013	2.72	153037	20190123	1.01
138036	20191013	2.97	154037	20200202	1.89
139034	20190121	4.99	155037	20140208	1.11
139036	20141225	4.72			

表3 Landsat-8 SR 验证影像信息

Table3 Landsat-8 SR verification image information

行列号	日期	云量	行列号	日期	云量
130037	20181119	4.24	141036	20131118	4.37
131038	20181116	1.15	142035	20201128	4.69
132038	20160229	0.91	142039	20150131	4.45
133034	20141113	1.55	142040	20200317	3.6
133039	20181226	2.68	143038	20160125	4.17
134036	20181217	1.21	144036	20200111	3.06
134040	20180216	2.22	145036	20140218	3.04
135037	20190125	1.13	146036	20170201	1.51
136035	20191218	4.82	147036	20181228	1.5
136039	20190116	1.86	148035	20131103	1.87
136040	20140219	3.15	148037	20150210	0.94
137036	20171101	1.43	149034	20160119	1.97
137040	20190107	1.72	150032	20201222	0.49
138038	20160122	1.85	151034	20171205	1.75
139037	20150110	3.87	152036	20190305	1.39
139040	20190105	1.43	153031	20141227	0.84
139041	20181220	1.25	154033	20150204	1.38
140039	20180125	4.68	156037	20200115	1.25

### 2.2.5 SnowCCI和MOD10A1

SnowCCI是欧空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1km, 覆盖时间从2000年到2020年, 数据来源于(<https://climate.esa.int/en/> [2023-02-02])。SnowCCI参考的FSC提取算法为Metsamaki等提出的SCAmod方法(Metsamaki等, 2012), 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500m, 数据来源于(<https://www.earthdata.nasa.gov/> [2023-02-01])。

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

## 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

高山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(Teillet等, 1982)。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):

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

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,  $\theta$ 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存在明显的差异, 阳坡过度感光NDSI普遍较高, 阴坡受阴影覆盖影响NDSI较低。经地形校正处理后, 影像b阴坡NDSI有明显提升, 阳坡NDSI适当降低, 有效地消除了地形起伏引起的地表辐射能量不平衡的影响。

### 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 and Riggs 2010)。NDSI计算公式如式(3):

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

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等, 2019; Wang等, 2022; Salomonson and 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(Klein等，1998)。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)：

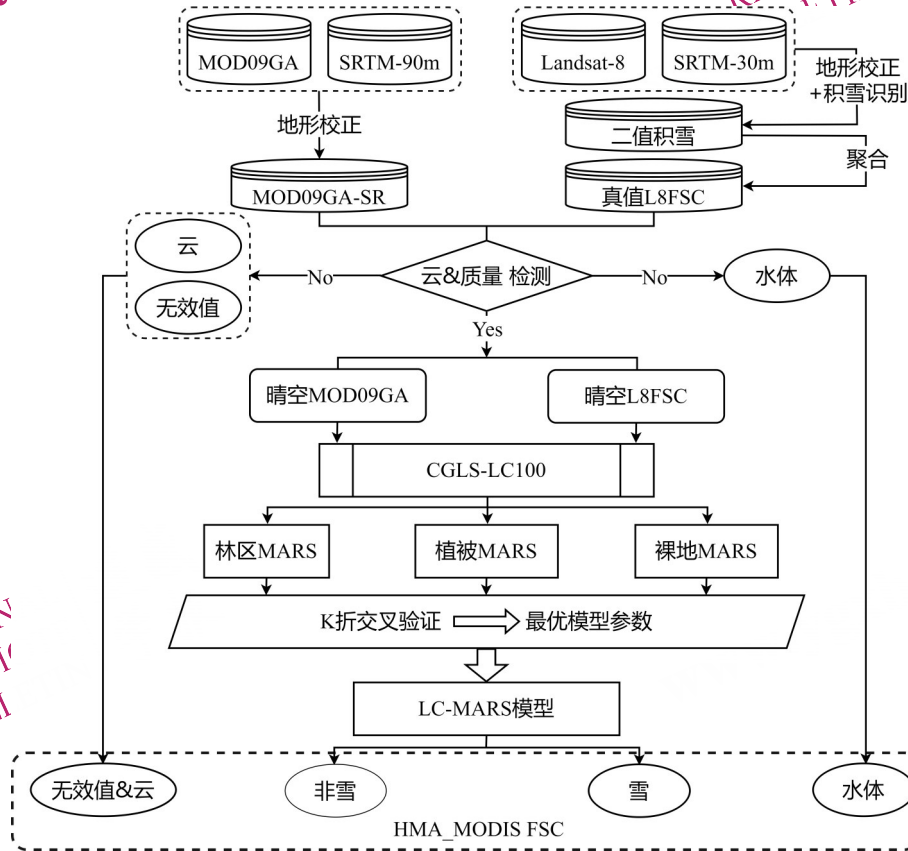


图1 LC-MARS模型反演FSC技术路线

Fig. 1 LC-MARS model inversion FSC technical route

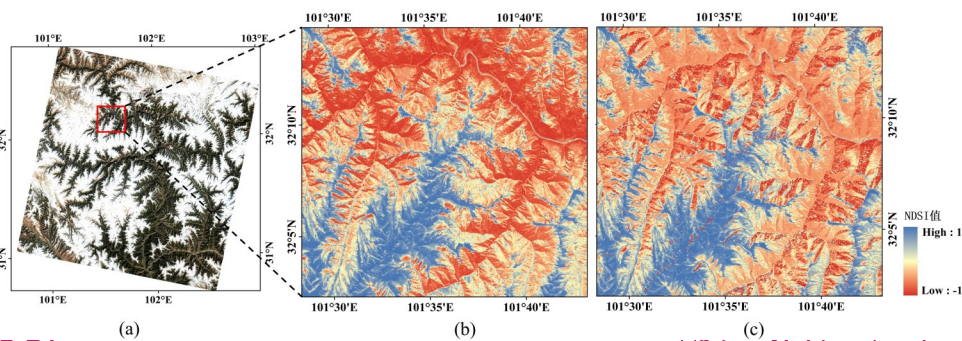


图2 NDSI地形校正前后对比

Fig. 2 Comparison before and after NDSI terrain correction

(a) Landsat-8 SR (131038-20181110) RGB image (b) NDSI image before correction (c) NDSI image after correction

(a) Landsat-8 SR (131038-20181110) RGB image (b) NDSI image before correction (c) NDSI image after correction

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

式中， $R$ 为可见光红波段， $NIR$ 为近红外波段。Wang等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(Wang等，2020)，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。NDFSIS计算公式如式(5):

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

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NDFSIS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:

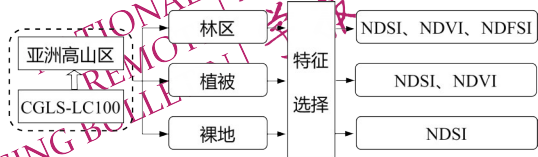


图3 分地类特征选取方案

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。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):

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

式中,  $f(x)$ 为MARS模型预测值,  $\beta_0$ 为截距,  $B_i(x)$ 是第*i*个基函数,  $\alpha_i$ 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:

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

上式表达的是一对镜面基函数, 式中*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:

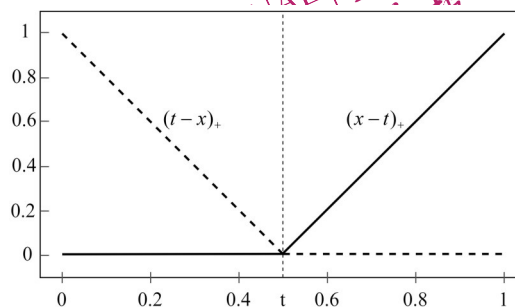


图4 镜面基函数

Fig. 4 Specular basis function

MARS模型的构建分“前向逐步”和“后向剪枝”2个过程。第一步, 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; 第二步, 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(赵忠阔等, 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)。通过训练三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展示了三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。

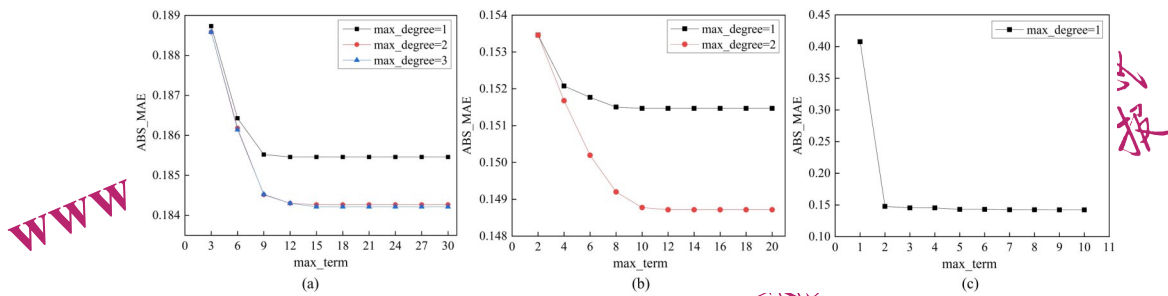


图5 LC-MARS模型max\_degree和max\_term参数训练变化

Fig. 5 LC-MARS model max\_degree and max\_term parameter training changes

(a) 林区MARS训练结果 (b) 植被MARS训练结果 (c) 裸地MARS训练结果

(a) train result of forest MARS (b) train result of vegetation MARS (c) train result of bare MARS

### 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像素是否包含雪的四种可能的结果，分别为：真阳性（TP）、真阴性（TN）、假阳性（FP）和假阴性（FN）。四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

Table4 Binary snow classification table

MODIS-FSC data	L8-FSC(真值FSC)	
	Snow	Not Snow
Snow	TP	FP
Not Snow	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如下：

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

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

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

#### 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（Root Mean Square Error, RMSE）和平均绝对误差（Mean Absolute Error, MAE）。RMSE和MAE的计算公式如下：

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

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

## 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、表6和表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。

由表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：

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

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

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

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

表5 林区MARS模型基函数  
Table5 Basis functions of MARS model in forest area

编号(i)	基函数(F)	系数( $\alpha$ )
	截距	+ 0.3064
$F_1$	$\max(0, NDFS1 - 0.3446)$	+ 2.3314
$F_2$	$\max(0, 0.3446 - NDFS1)$	- 0.8960
$F_3$	$\max(0, NDVI - 0.1919)$	- 0.7528
$F_4$	$\max(0, 0.1919 - NDVI)$	+ 0.1115
$F_5$	$\max(0, NDSI + 0.4499) * \max(0, 0.3446 - NDFS1)$	- 0.7060
$F_6$	$\max(0, -0.4499 - NDSI) * \max(0, 0.3446 - NDFS1)$	+ 0.8463
$F_7$	$\max(0, NDSI - 0.2433)$	- 0.6695
$F_8$	$\max(0, 0.2433 - NDSI)$	+ 0.2429
$F_9$	$\max(0, NDFS1 - 0.6159) * \max(0, NDFS1 - 0.3446)$	- 1.7362
$F_{10}$	$\max(0, 0.6159 - NDFS1) * \max(0, NDFS1 - 0.3446)$	- 0.6795

表6 植被MARS模型基函数  
Table6 Basis functions of MARS model in vegetation area

编号(j)	基函数(V)	系数( $\beta$ )
	截距	+ 0.7821
$V_1$	$\max(0, -0.2768 - NDSI)$	- 1.0616
$V_2$	$\max(0, NDSI - 0.4478) * \max(0, NDSI + 0.2768)$	- 2.0651
$V_3$	$\max(0, 0.4478 - NDSI) * \max(0, NDSI + 0.2768)$	+ 1.4733
$V_4$	$\max(0, NDVI - 0.2320)$	+ 0.0658
$V_5$	$\max(0, 0.2320 - NDVI)$	- 1.3970
$V_6$	$\max(0, NDSI - 0.4352)$	+ 2.6178
$V_7$	$\max(0, 0.4352 - NDSI)$	- 1.0298
$V_8$	$\max(0, NDSI + 0.1034) * \max(0, 0.4352 - NDSI)$	- 2.1499
$V_9$	$\max(0, -0.1034 - NDSI) * \max(0, 0.4352 - NDSI)$	+ 1.2379
$V_{10}$	$\max(0, NDSI + 0.1948) * \max(0, 0.2320 - NDVI)$	+ 1.5105
$V_{11}$	$\max(0, -0.1948 - NDSI) * \max(0, 0.2320 - NDVI)$	+ 1.5519

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:

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

式中,  $B_{FSC}$ 为裸地FSC,  $B_k(NDSI)$ 表示第k个裸地MARS基函数,  $\chi_k$ 为第k个裸地基函数的系数。上述的式(13)、式(14)和式(15)分别为亚洲高山林区FSC、植被FSC和裸地FSC的反演算法。

#### 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的三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 $LST \geq 275K$ 且 $DEM \leq 1300m$ 或者 $LST \geq 281K$ 且 $DEM \geq 1300m$ (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/](http://www.coldregion.com/2024/01/05/scf_har/) 中的图 3, 展示了 2018 年 11 月 10 日的亚洲高山区 FSC。

表 7 裸地 MARS 模型基函数

Table 7 Basis functions of MARS model in vegetation area

编号( $k$ )	基函数( $B$ )	系数( $\chi$ )
	截距	+ 0.6025
$B_1$	$\max(0, -0.183687 - \text{NDSI})$	+ 0.0288
$B_2$	$\max(0, 0.596954 - \text{NDSI})$	- 1.1126
$B_3$	$\max(0, 0.223459 - \text{NDSI})$	+ 0.7618
$B_4$	$\max(0, \text{NDSI} + 0.277521)$	+ 0.3568
$B_5$	$\max(0, -0.277521 - \text{NDSI})$	+ 0.3162

## 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, 其中裸地三项指标均高于植被, 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。从表 9 可以看出, 林区、植被和裸地的 RMSE 均小于 0.2, 三种地表的 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。三种地表的 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。

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

### 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500m以下精度最高, 海拔2500到3500m精度最低。海拔低于500m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到3500m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3500m逐渐上升至5500m, 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

Table10 Variation of error in different altitudes

海拔区间/m	样本数	RMSE	MAE
<500	2356	0.034	0.018
[500,1500]	96137	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、积雪和融雪三个过程, 三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。

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, 并对三处地形校正前后LC-MARS模型反演FSC精度进行对比分析, 三处复杂地形分布见 [http://www.coldregion.com/2024/01/05/scf\\_har/](http://www.coldregion.com/2024/01/05/scf_har/)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、祁连山脉所选的三处复杂地形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11 不同积雪期误差变化

Table11 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

表12 地形校正前后FSC精度对比

Table12 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:

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

$$0.233 \quad (16)$$

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

$$0.072 \quad (17)$$

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。图中三种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。

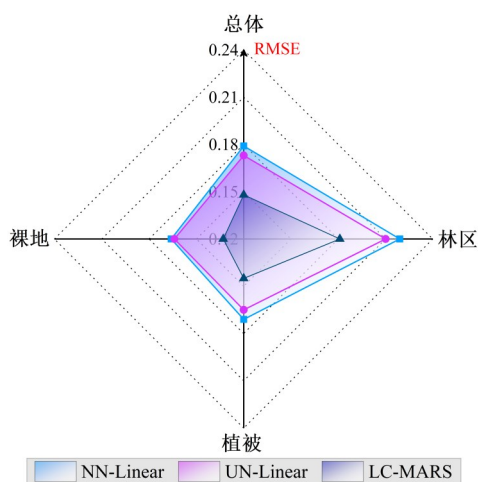


图6 LC-MARS与NN-Linear、UN-Linear模型精度对比  
Fig. 6 LC-MARS and NN-Linear, UN-Linear model accuracy comparison

## 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，三种FSC产品精度验证的参考真值还是亚洲高山区36景L8-FSC。

二值验证评估三种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三类指标，LC-MARS模型识别积雪的准确率显著高于MOD10A1和SnowCCI。

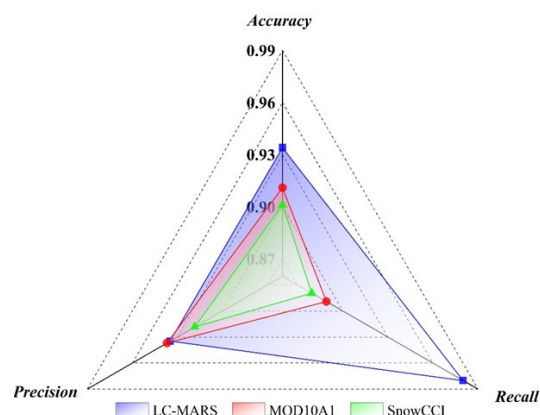


图7 LC-MARS与MOD10A1、SnowCCI二值验证对比图  
Fig. 7 Comparison of LC-MARS and MOD10A1, SnowCCI binary verification

36景验证影像均匀分布于亚洲高山区, 三种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择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, 三类地类下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。

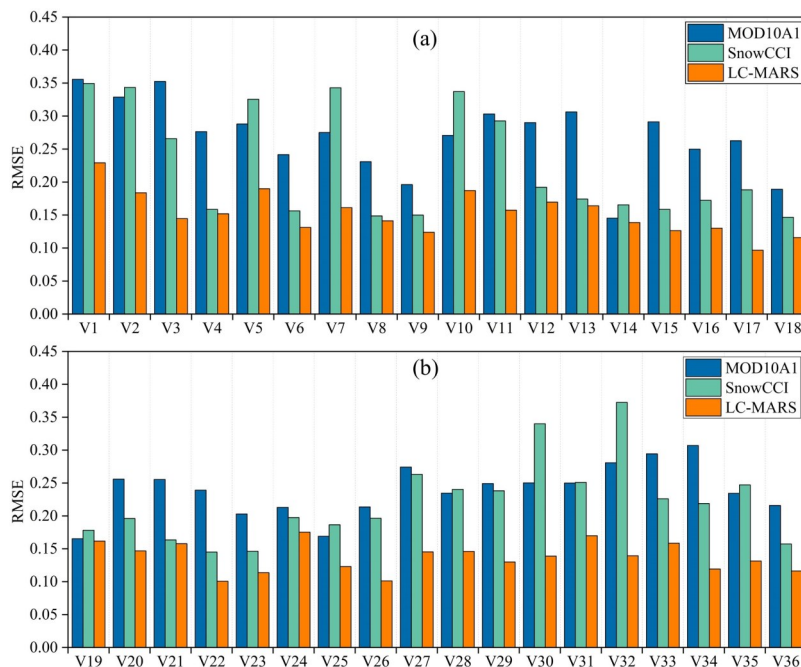


图8 LC-MARS与MOD10A1、SnowCCI验证影像精度对比

Fig. 8 Comparison of LC-MARS and MOD10A1, SnowCCI verification image accuracy

(a) 验证影像V1-V18精度对比 (b) 验证影像V19-V36精度对比

(a) validation image V1-V18 accuracy comparison (b) validation image V19-V36 accuracy comparison

## 5.2 NDVI和NDFS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、NDFS, 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、NDFS是否提高林区FSC反演精度以及NDFS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NDFS构建的NDFS-MARS模型。以36景L8-FSC为参考

真值, 比较NDSI-MARS模型、NDFS-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NDFS、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; NDFS-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NDFS、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NDFS I，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。

表13 LC-MARS模型与MOD10A1、SnowCCI分地类精度对比

**Table13 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 I-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NDFS I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、NDFS I特征构建的LC-MARS与仅由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，有效地解决了林区、植被区积雪识别精度不足的问题。

(3) LC-MARS模型移植到Google Earth Engine云平台制备了2000-2020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, Adnan H. 2019. Estimating snow cover area in south of Turkey using the Normalized Difference Snow Index (NDSI) from MODIS Satellite Images. Journal of Physics: Conference Series[DOI: 10.1088/1742-6596/1279/1/012047]Brun F, Berthier E, Wagnon P, Kaab A and Treichler D.2018.A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016 (vol 10, pg 668, 2017). Nature Geoscience 11 (7):543-[DOI: 10.1038/s41561-018-0171-z]
- Brun F, Berthier E, Wagnon P, Kaab A, and Treichler D. 2018. "A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016 (vol 10, pg 668, 2017)." Nature Geoscience 11 (7):543-[doi: 10.1038/s41561-018-0171-z]

- Ciftci B B, Kuter S, Akyurek Z and Weber G W. 2017. Fractional Snow Cover Mapping by Artificial Neural Networks and Support Vector Machines. 4th International Geoadvances Workshop - Geoadvances 2017: Isprs Workshop on Multi-Dimensional & Multi-Scale Spatial Data Modeling 4-4 (W4): 179-87[DOI: 10.5194/isprs-annals-IV-4-W4-179-2017]
- Czyżowska-Wisniewski, Elzbieta H, Willem J D, van Leeuwen, Katherine K, Hirschboeck, Stuart E, Marsh, and Wit T. Wisniewski. 2015. Fractional snow cover estimation in complex alpine-forested environments using an artificial neural network. Remote Sensing of Environment[DOI: 10.1016/j.rse.2014.09.026]
- Friedman J. 1991. Multivariate Adaptive Regression Splines [DOI: 10.1214/AOS/1176347963]
- García-Álvarez, David, Javier Lara Hinojosa, Francisco José Jurado Pérez and Jaime Quintero Villaraso. 2022. Global General Land Use Cover Datasets with a Time Series of Maps. In Land Use Cover Datasets and Validation Tools: Validation Practices with QGIS, edited by David García-Álvarez, María Teresa Camacho Olmedo, Martin Paegelow and Jean François Mas, 287-311. Cham: Springer International Publishing.
- 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)[DOI:ARTN42510.3390/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):127-40[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): 533-48[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): 4711-26[DOI: 10.5194/essd-13-4711-2021]
- Hao X H, Huang G H, Zheng Z J, Sun X L, Ji 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): 1937-52[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): 5601-11[DOI: 10.1109/Tgrs.2013.2290996]
- 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(01):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 (01): 35-43[DOI: http://ir. casnw. net/handle/362004/11956])
- 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 Environment 190:274-88[DOI: 10.1016/j.rse.2016.12.028]
- 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):133-52[DOI: 10.1080/01431160903439924]
- Klein, Andrew G, Dorothy K. Hall, and George A. Riggs. 1998. Improving snow cover mapping in forests through the use of a canopy reflectance model. Hydrological Processes 12 (10-11):1723-44[DOI: 10.1002/(SICI)10991085(199808/09)12: 10/11<1723:: AID-HYP691>3.0.CO;2-2]
- Konig 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: 236-52 [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(06):1303-11 (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 (06): 1303-11. [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) [DOI: ARTN 252310.3390/rs14112523]
- Liang S L, Bai R, Chen X N, Chen J, Fan W J, He T, Jia Ket al. 2020. Review of China's land surface quantitative remote sensing development in 2019. Journal of Remote Sensing(Chinese)24(06)618-671 (梁顺林, 白瑞, 陈晓娜, 程洁, 范闻捷, 何涛, 贾坤. 2020. 2019年中国陆表定量遥感发展综述. 遥感学报 24 (06):618-71. [DOI:10.11834/jrs.20209476])
- 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) [ DOI: ARTN 133210.3390/rs9121332]
- Liu S Y, Ding Y J, Ye B S, Wang N L and Xie Z C. 2002. Regional Characteristics of Glacier Mass Balance Variations in High Asia. Journal of Glaciology and Geocryology(02):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 (02):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* 456:232-42[DOI: 10.1016/j.neucom.2021.05.037]
- Metsamaki S, Mattila O P, Pulliainen J, Niemi K, Luojus K and Bottcher K. 2012. An optical reflectance model-based method for fractional snow cover mapping applicable to continental scale. *Remote Sensing of Environment* 123: 508-21[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 [DOI:ARTN11178110.1016/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):868-79[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 Tet al. 2021. National Tibetan Plateau Data Center Promoting Earth System Science on the Third Pole. *Bulletin of the American Meteorological Society*102(11): E2062E78[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* 51:367-80[DOI: 10.1016/j.advwatres.2012.03.002]
- Rosenthal, Walter and Jeff Dozier.1996. Automated Mapping of Mountain Snow Cover at Subpixel Resolution from the Landsat Thematic Mapper. *Water Resources Research* 32 (1):115-30[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):351-60[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. *Igarss 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*:2946-9 [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(01):147-58 (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 (01):147-58[ 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(05):1400-11 (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 (05):1400-11.[ 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) [ DOI: ARTN 104510.3390/rs9101045]
- Teillet P m, Guindon B, and D.g.Goodenough.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 J. 1999. Comparison and analysis on methods of snow cover mapping by using satellite remote sensing data. *Remote Sensing Technology and Application*, (04):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* (04):29-36. [DOI: 10.3969/j.issn.1004-0323.1999.04.005])
- 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[DOI: Artn 430431810.1109/Tgrs.2022.3165986]
- Wang G X, Jiang L M, Shi J C and Xu S. 2021. A Universal Ratio Snow Index for Fractional Snow Cover Estimation. *IEEE Geoscience and Remote Sensing Letters* 18 (4):721-5. [DOI: 10.1109/lgrs.2020.2982053]
- Wang G X, Jiang L M, Shi J C, Liu X J, Yang J W and Cui H Z.2019. Snow-Covered Area Retrieval from Himawari-8 AHI Imagery of the Tibetan Plateau. *Remote Sensing* 11 (20) [ DOI: ARTN 239110.3390/rs11202391]
- 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* 13:5211-22[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) [DOI:ARTN394110.3390/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[DOI:ARTN11147110.1016/j.rse.2019.111471]
- 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(06):1004-16 (赵宏宇, 郝晓华, 郑照军, 王建, 李弘毅, 黄广辉, 郝东航, 王轩, 高扬, 霍华锦. 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 (06):1004-16[DOI:CNKI:SN:YGJS.0.2018-06-003].)
- 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(03): 442-50 (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 (03):442-50[ DOI: 10.13203/j.whugis20190136])
- 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. *RemoteSensing* 12(22) [DOI: ARTN 369310.3390/rs12223693]

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[DOI: ARTN 76055110.3389/fpls.2021.760551]

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

GAO Weiqiang<sup>1,2</sup>,HAO Xiaohua<sup>2</sup>,HE Dongcai<sup>1</sup>,SUN Xingliang<sup>3</sup>,LI Hongyi<sup>2</sup>,REN Hongrui<sup>1</sup>,ZHAO Qin<sup>2</sup>

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:** The abstract of this study contains four sections objective, method,result and conclusion 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. Method 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.Result (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.Conclusion 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:** High Mountain Asia, Fractional snow cover, MODIS, MARS, Terrain correction

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