

学校代码: 10730

分类号: TP79

密级:

兰州大学
硕 士 学 位 论 文
(学 术 学 位)

色林错和纳木错湖区雪深的再分析资料评估

论文题目 (中文) 和基于被动微波数据的反演估计

Evaluation of Reanalysis Snow Depth Datasets

论文题目 (外文) and Retrieval of Snow Depth Based
on Passive Microwave Data in Selin Co

and Nam Co Regions

作 者 姓 名 邬俊飞

学 科 专 业 地理学·自然地理学

研 究 方 向 被动微波遥感积雪参数反演

教 育 类 型 学历教育

指 导 教 师 姚檀栋 院士

论 文 工 作 时 段 2019 年 9 月 至 2022 年 4 月

论 文 答 辩 日 期 2022 年 5 月

校址: 甘肃省兰州市城关区天水南路 222 号

色林错和纳木错湖区雪深的再分析资料评估和

基于被动微波数据的反演估计

中文摘要

积雪是冰冻圈的重要组成要素，对气候变化敏感，能够影响区域能量平衡和水循环过程。青藏高原作为我国中、低纬度的主要积雪分布区，该地区积雪能够影响我国乃至全球气候系统，积雪融水对下游国家与地区的社会经济发展和自然环境稳定也起着重要作用。获取准确的地表雪深信息是开展积雪资源评估、水资源管理以及气候系统模拟研究的基础，具有重要的科学意义与研究价值。当前获取青藏高原地表雪深信息主要通过被动微波反演雪深和再分析雪深数据集，其具有覆盖范围广、时间分辨率高及获取成本低等优势，但由于青藏高原中、西部地区地面雪深资料稀缺，对被动微波雪深反演经验算法和再分析雪深数据集在高原中、西部的区域表现认识不足。此外，当前应用广泛的被动微波雪深反演经验算法多基于被动微波亮温差，较少考虑地形、下垫面等环境要素对雪深反演的影响，导致雪深反演经验算法误差较大。因此，本研究主要基于青藏高原中部色林错和纳木错湖区 2019 年 7 月 ~2021 年 6 月 7 个台站的实测雪深数据和 AMSR2 亮温数据，对 ERA5、GLDAS、JRA55 和 MERRA2 等四种常见的再分析雪深数据集，及 Chang2 算法、SPD 算法、Che 算法、Jiang 算法和 AMSR2 算法等五种被动微波雪深反演算法，在研究区的表现进行了评估。使用 LASSO、SCAD、MCP 和随机森林特征重要度方法，对坡度、坡向和海拔，草地、水体、裸地土地覆被类型，及 AMSR2 亮温数据集中的 10、19、24、37、89 GHz 垂直/水平极化亮温及不同频率亮温差等 61 个变量进行选择，得到了四组雪深反演变量，将四种方法选中的所有变量作为实验对照组（ALL 方法），并基于随机森林回归和支持向量回归，构建了十种机器学习雪深反演组合模型。研究结果显示：

（1）ERA5、GLDAS、JRA55、MERRA2 四种再分析雪深数据集的月平均积雪天数均高于站点观测（6.59 d）。ERA5 再分析雪深数据集在研究区的整体表现最好，与观测雪深的相关系数为 0.62，对研究区雪深整体高估（2.37 cm），其中对深雪高估最大（12.11 cm）。JRA55 数据集雪深与观测雪深的整体相关系数为 0.45，均方根误差在四种数据集中最小（4.06 cm）。GLDAS 和 MERRA2 雪

深数据集在研究区的表现最不理想 (R 值分别为 0.23、0.14)，其中 MERRA2 数据集的整体偏差在四种数据集中最大 (11.71 cm)。四种雪深数据集对地面浅雪的模拟表现均较差 (R 值均小于 0.20)，ERA5 和 JRA55 雪深数据集对地面中雪的模拟表现在四种数据集中最好 (R 值分别为 0.80、0.47)。

(2) Jiang 算法在研究区的整体表现最好，反演雪深与地面观测雪深的相关系数最高 (0.65)，RMSE 为 5.51 cm，对混合土地覆被类型站点的雪深反演表现优于其它四种算法 (R 为 0.56)。Che 算法的均方根误差 (7.36 cm) 仅高于 Jiang 算法。Chang2 算法反演雪深与地面观测雪深的相关性较高 (0.59)，但偏差 (-12.40 cm) 和均方根误差 (17.00 cm) 较大。SPD 算法和 AMSR2 算法在研究区整体表现不理想 (R 分别为 0.42、0.45)。五种算法中，SPD 算法反演雪深对研究区雪深高估 (15.80 cm)，而其余四种算法反演雪深对研究区雪深低估。Chang2 算法、Jiang 算法和 Che 算法对中雪的反演效果较好，反演雪深与观测雪深的相关系数均大于 0.70。

(3) 十种雪深反演模型的决定系数均较高 (0.63~0.82)、偏差分布接近正态分布、平均绝对误差小 (均小于 3.20 cm)，其中基于 LASSO 方法的随机森林回归模型 L-RFR 表现最好 (0.82)。除 I-SVR 模型外 (0.63)，其余九种模型的决定系数均大于 0.67，MAE 均小于 3.05 cm，整体表现优于四种再分析雪深数据集 (R 最大为 0.62) 和五种被动微波雪深反演算法 (R 最大为 0.65)。随机森林回归模型的决定系数 (0.79~0.82) 均高于支持向量回归模型的决定系数 (0.63~0.70)。在随机森林回归或支持向量回归模型中，基于 LASSO、SCAD 方法的变量组合的模型表现最好，基于 MCP 和随机森林特征重要度方法的变量组合的模型表现不佳，基于 ALL 方法变量组合的模型表现介于基于其它四种变量组合的模型表现之间。

关键词：被动微波，积雪深度，色林错，纳木错，机器学习，雪深反演算法，再分析雪深数据集

EVALUATION OF REANALYSIS SNOW DEPTH DATASETS AND RETRIEVAL OF SNOW DEPTH BASED ON PASSIVE MICROWAVE DATA IN SELIN CO AND NAM CO REGIONS

Abstract

Snow cover is an important component of the cryosphere, which is sensitive to climate change and can affect regional energy balance and water cycle process. The Tibetan Plateau is the main snow cover distribution area in the middle and low latitudes of China. The snow cover in this area can affect China and even the global climate system. Snow meltwater also plays an important role in the social and economic development and natural environment stability of downstream countries and regions. Obtaining accurate surface snow depth information is the basis for snow resource assessment, water resource management and climate system simulation, which has important scientific significance and research value. At present, the main ways to obtain surface snow depth information on the Tibetan Plateau are the retrieval of snow depth based on passive microwave and reanalysis snow depth datasets, which have the advantages of wide coverage, high temporal resolution and low acquisition cost. Due to the scarcity of ground snow depth data, there is insufficient understanding of the empirical passive microwave snow-depth retrieval algorithm and the reanalysis snow depth datasets in the central and western Tibetan Plateau regions. In addition, the currently widely used empirical passive microwave snow-depth retrieval algorithms are mostly based on passive microwave brightness temperature difference, and less consideration is given to the influence of environmental factors such as terrain and underlying surface on the retrieval of snow depth, resulting in a large error in the empirical snow-depth retrieval algorithms. Therefore, this study mainly used the measured snow depth data of seven in-situ stations from July 2019 to June 2021 in the Selin Co and Nam Co regions of the central Tibetan Plateau to evaluate the performance

of four common reanalysis snow depths datasets, including ERA5, GLDAS, JRA55 and MERRA2, and five passive microwave snow-depth retrieval algorithms in the study area, including Chang2 algorithm, SPD algorithm, Che algorithm, Jiang algorithm, and AMSR2 algorithm. To recognize the influence of different brightness temperature channels and environment variables such as terrain and land cover types on snow-depth retrieval, and to provide a reference for the study of snow-depth retrieval algorithm based on passive microwave in the Tibetan Plateau, LASSO, SCAD, MCP penalty method, and random forest feature importance method were used to select 61 variables such as slope, aspect, altitude, land cover types, and 10, 18, 23, 37, 89 GHz vertical/horizontal polarization brightness temperature and different frequency brightness temperature difference of AMSR2. Four snow-depth retrieval variables groups were obtained. And all the variables selected by the four methods were used as the experimental control group (ALL method). And then, the five variables groups were combined with random forest regression and support vector regression, respectively, constructing ten models of machine learning for retrieving snow depth. The conclusions of this study are as follows:

(1) The monthly average snow days of the four reanalysis snow depth datasets, ERA5, GLDAS, JRA55, and MERRA2, were all higher than those of the stations' observations (6.59 d). The ERA5 reanalysis snow depth dataset had the best overall performance in the study area, with a correlation coefficient of 0.62. And it overestimated the observed snow depth of the study area overall (2.37 cm), of which the overestimation of deep snow was the largest (12.11 cm). The overall correlation coefficient between the JRA55 reanalysis snow depth dataset and the observed snow depth was 0.45, and the RMSE was the smallest among the four reanalysis snow depth datasets (4.06 cm). The GLDAS and MERRA2 reanalysis snow depth datasets performed the least well in the study area (values of R are 0.23 and 0.14, respectively), and the MERRA2 reanalysis snow depth dataset had the largest overall bias (11.71 cm) among the four analysis snow depth datasets. The correlation coefficient between the four reanalysis snow depth datasets and observed shallow snow was weak (less than 0.20), and the ERA5 and JRA55 reanalysis snow depth datasets had the highest correlation with observed snow depth (values of R are 0.80 and 0.47, respectively).

(2) The Jiang algorithm had the best overall performance among the five algorithms in the study area. The overall correlation coefficient between the snow depth retrieved by the Jiang algorithm and the observed snow depth was the highest (0.65),

and the RMSE was 5.51 cm. The Jiang algorithm outperformed the other four algorithms in snow-depth retrieval for mixed land cover type stations ($R = 0.56$). The RMSE of the Che algorithm is second only to that of the Jiang algorithm (7.36 cm). The correlation coefficient between snow depth retrieved by the Chang2 algorithm and observed snow depth was high (0.59), but the bias (-12.40 cm) and RMSE (17.00 cm) were large. The SPD algorithm and AMSR2 algorithm performed poorly (values of R are 0.42, 0.45, respectively). Among the five algorithms, the snow depth retrieved by the SPD algorithm overestimated the observed snow depth in the study area (15.80 cm), while the snow depth retrieved by the other four algorithms underestimated the observed snow depth in the study area. And the Chang2 algorithm, Jiang algorithm, and Che algorithm had better retrieval effects on medium snow, and the correlation coefficients between the retrieved snow depth and the observed snow depth were all greater than 0.70.

(3) The determination coefficients of the ten snow depth retrieval models were all high (0.63 ~ 0.82). The distribution of their bias was close to the normal distribution, and their MAE was small (all less than 3.2 cm). Among them, the random forest regression model L-RFR based on the LASSO method had the best performance ($R^2 = 0.82$). Except for the I-SVR model ($R^2 = 0.63$), the determination coefficients of the other nine models were all greater than 0.67, their MAE was all less than 3.05 cm, and their overall performances were better than those of the four reanalysis snow depth datasets (the maximum R was 0.62) and the five snow-depth retrieval algorithms based on passive microwave (the maximum R was 0.62). The determination coefficients of the models based on random forest regression (0.79 ~ 0.82) were all higher than those of the models based on support vector regression (0.63 ~ 0.70). Among the random forest regression or support vector regression models, the models based on the variable combination of the LASSO and SCAD methods performed the best, the models based on the variable combination of the MCP and the random forest feature importance method performed poorly, and the performances of models based on the variable combination of the ALL method were between those of the models based on the other four variable combinations.

Keywords: Passive Microwave, Snow Depth, Selin Co, Nam Co, Machine Learning, Snow-depth Retrieval Algorithm, Reanalysis Snow Depth Datasets

目 录

中文摘要.....	I
Abstract.....	III
第一章 绪论.....	1
1.1 选题依据及科学意义	1
1.2 青藏高原雪深研究方法及进展	2
1.2.1 野外实地观测	2
1.2.2 激光雷达	3
1.2.3 GPS 干涉反射法	3
1.2.4 物理模型模拟	4
1.2.5 主被动微波遥感	4
1.3 研究内容和技术路线	9
1.3.1 研究内容	9
1.3.2 技术路线	9
1.4 论文结构	10
第二章 研究区概况和研究方法与数据	11
2.1 研究区概况	11
2.2 数据介绍	12
2.2.1 实测雪深数据	12
2.2.2 DEM 数据	13
2.2.3 土地覆被类型数据	13
2.2.4 AMSR2 亮温数据	13
2.2.5 再分析数据集	14
2.2.6 MODIS 积雪范围产品	15
2.2.7 湖泊范围数据	15

2.3 研究方法	16
2.3.1 雪深数据预处理方法	16
2.3.2 积雪深度分类标准	16
2.3.3 积雪覆盖率数据的处理流程	16
2.3.4 地形数据处理方法	18
2.3.5 土地覆被重分类方法	18
2.3.6 被动微波亮温数据预处理	20
2.3.7 机器学习数据预处理方法	20
2.3.8 机器学习变量选择方法	21
2.3.9 机器学习建模方法	23
2.3.10 机器学习模型的参数优化方法	24
2.3.11 评价指标	24
第三章 再分析雪深数据集与被动微波反演算法评估	26
3.1 站点代表性分析	26
3.1.1 台站雪深与积雪覆盖率分析	26
3.1.2 海拔分析	27
3.1.3 下垫面类型分析	28
3.1.4 站点代表性讨论	30
3.2 再分析雪深数据集评估	30
3.2.1 再分析雪深数据集时间序列特征分析	30
3.2.2 数据集在不同积雪深度的表现	32
3.2.3 结果讨论	34
3.3 被动微波雪深反演算法评估	34
3.3.1 算法反演雪深整体表现	35
3.3.2 算法在不同积雪深度的表现	40
3.3.3 结果与讨论	41

3.4 本章小结	42
第四章 基于机器学习的被动微波雪深反演研究	44
4.1 机器学习模型反演雪深结果与分析	45
4.1.1 模型输入变量选择结果	45
4.1.2 组合模型表现	47
4.1.3 结果讨论	50
4.2 本章小结	52
第五章 结论与展望	54
5.1 主要研究结论	54
5.2 创新点	55
5.3 不足及展望	56
参考文献	57
附录	64
在学期间的研究成果	67
致谢	68

第一章 绪论

1.1 选题依据及科学意义

冰冻圈作为地球气候系统的五大圈层之一，能够直接或间接地影响全球气候和水文循环，青藏高原地区在我国境内平均海拔 4378 m，素有“亚洲水塔”和“世界第三极”之称，是仅次于南北极的全球冰冻圈所在地，孕育了长江、黄河、恒河、印度河等亚洲大江大河^[1, 2]，影响着全世界约 22 %人口的生产生活^[3]。研究显示，过去几十年青藏高原地区升温幅度是全球平均的两倍，青藏高原是全球气候变化的敏感区。而积雪作为冰冻圈的重要组成要素，具有高反照率、低热导率等性质，能够引起气候系统不同时空尺度上的能量平衡和水量循环变化^[4]。

根据积雪存在时间，积雪可被分为永久性积雪和季节性积雪，永久性积雪指的是多年未化的积雪，通常分布于高海拔或高纬度地区，如山地冰川表面、南极大陆等区域，而季节性积雪在冬季积累、春季消融，具有明显的季节性规律^[5]，是许多干旱-半干旱地区地表径流和地下水的重要补给来源，可缓解春旱对于当地农牧业的影响，因此积雪被认为是山区“固态水库”。遥感资料显示，青藏高原地区冬季积雪面积占青藏高原总面积的 65 %^[6]，是中国中、低纬度地区主要的积雪分布区^[7]。季节性积雪融水是山区地表径流的重要来源，Immerzeel et al. 利用分布式水文模型模拟了青藏高原地区积雪融水对山区径流的贡献，积雪融水贡献高达 26 %^[8]，对下游地区的生态系统和社会可持续发展起着重要作用^[9]。

但时空分布不合理的积雪能够导致雪灾^[10]，给人民群众生命财产安全和经济社会发展带来重大威胁。在青藏高原地区雪灾的表现形式多样，常见的有暴风雪、风吹雪、雪崩以及牧区雪灾，能够引起当地气温、能见度骤降，造成道路堵塞、农作物和牲畜受损和死亡等^[11]，如 1996 年 3 月 24 日西藏波密县然乌镇造成 64 人罹难的雪崩灾害，是我国有记录以来伤亡最大的雪崩灾害事件^[12]；1998 年玉树州发生的雪灾，造成近 20 万头牲畜死亡，1552 人冻伤，受灾面积达 1.95 亿亩^[13]。

近年来随着全球气候变化，青藏高原积雪分布表现出较大的时空变化差异^[14]，积雪覆盖范围整体呈下降趋势^[15]，西藏地区的积雪积累时间延后、消融时间提前^[16]，这都给青藏高原山区积雪水文增加了不确定性^[3, 17]，因此受到了越来越多气候变化研究组织和学者的关注^[4]。

积雪深度（Snow Depth, SD）作为积雪场的重要物理属性，是开展山区雪崩

监测、水文预测和气候系统模拟等研究的重要参数,能够影响区域和全球气候系统^[9, 18-23]。研究表明,青藏高原冬春积雪正偏、春季积雪的东多西少与我国夏季南涝北旱格局存在关联^[24],青藏高原的雪深异常(anomaly)能够影响次年中国地区夏季风的登陆时间与降水强度,改变中国夏季降水的时空分布格局^[19]。另外,青藏高原雪深还被发现与北极涛动(Arctic Oscillation)和东亚冬季风间^[25]、El Niño/Southern Oscillation(ENSO)、西风急流和南极涛动等存在关联^[18, 21, 22],因此对青藏高原积雪深度开展研究,能够帮助我们理解山区积雪水文过程及其与气候系统间的作用机制。

1.2 青藏高原雪深研究方法及进展

目前青藏高原雪深研究手段主要有野外实地观测、激光雷达、GPS 干涉反射法、物理模型模拟和主被动微波遥感等方法。

1.2.1 野外实地观测

青藏高原野外雪深观测工作自 20 世纪 50 年代就已经开始了^[26],当时限于社会发展水平,观测工作以人工为主,但随着雪深观测技术发展,目前野外雪深观测以台站自动观察为主,2020 年西藏地区的雪深观测已全面实现自动化^[27]。许多学者基于野外站点对青藏高原雪深开展过相关研究,唐国栋等^[28]对中国西部站点雪深数据进行空间插值,获得了青藏高原雪深的空间分布信息,即青藏高原东部、南部地区属于雪深高值区,年平均雪深约 40 cm,西部属雪深低值区,年平均雪深小于 10 cm。柯长青等^[29]利用青藏高原 55 个野外站点 1963~1992 年的雪深数据进行 EOF(Empirical Orthogonal Functions, 经验正交函数)分析发现,上世纪 60~80 年代积雪的年际波动幅度呈增加趋势。李小兰等^[30]和除多等^[31]的研究显示,上世纪 80 年代至本世纪 00 年代,青藏高原整体平均雪深、最大雪深均呈下降趋势。

青藏高原雪深年际变化研究的结果显示,青藏高原春季雪深自 1980s 出现呈明显下降趋势^[32],冬季雪深的年际变化趋势自 1982 年出现反转,也呈下降趋势,对地表气温的反馈作用也在减弱^[33]。而 You et al.^[34]对青藏高原东部和中部 69 个站点的雪深数据分析发现,高原冬季雪深在 1961~1990 年呈上升趋势,在 1991~2005 年呈下降趋势。研究显示,青藏高原区域间的雪深变化趋势存在差异,西藏地区的年平均雪深在 1979~2010 年呈增加趋势^[35],气候倾向率为 0.26 cm/10 a,而雅江流域的年平均雪深在 1979~2017 年呈显著减少趋势^[36]。

以往基于野外观测台站的青藏高原雪深研究,所使用的雪深观测站点多分布

于青藏高原的东部和南部等低海拔区^[37]，站点在青藏高原中部和西部分布稀疏，对青藏高原中部、西部等偏远山区的雪深及其变化特征认识不足。

1.2.2 激光雷达

激光雷达（Light detection and ranging, Lidar）是通过计算有雪与无雪时期激光发射与传感器接收反射信号的时间差，来获得地表与传感器之间的距离，通过计算两者高程差值进而得到地面雪深数据。激光雷达反演雪深的精度较高，Jacobs et al. ^[38]在美国新罕布什尔州对激光雷达反演雪深精度进行了评估，结果显示激光雷达在小于 20 cm 积雪场的反演偏差仅 0.96 cm；Deems et al. ^[39]在美国科罗拉多利用激光雷达绘制了雪崩形成区的厘米级雪深地图。但激光雷达雪深测量需采集研究区无雪时期的地表高程数据，且对研究区的地形数据要求较高，如激光雷达在 1 km 飞行高度下对地表 45 °坡的高度误差可达 50 cm^[40]。因而激光雷达雪深测量常应用于易到达的当地（local）尺度的雪深反演^[40]。针对交通不便的青藏高原地区的应用成本较高，使得在青藏高原地区未得到大范围应用。

1.2.3 GPS 干涉反射法

GPS 干涉反射法（GPS interferometric reflectometry, GPS-IR）是近二十年出现的新方法，已经得到了许多学者验证^[41, 42]，反演雪深精度可达 4 cm^[43]，其原理是 GPS 卫星信号传输到 GPS 地面接收器时，直接传输信号和经过积雪反射信号在到达接收器时发生相位干涉，通过对接收到的干涉波进行分析处理从而实现雪深反演。监测板块变形、天气状况的 GPS 地面接收器都能被用来反演雪深，并且 GPS 干涉反射计也可用于反演如土壤湿度、冻土变化等其他冰冻圈要素^[44]，该方法在国际上已有较多应用，如 Zhou et al. ^[45]基于三频道（如 L1、L2、L5 波段）GPS 信号提出了一种新的反演算法。Klimánek et al. ^[46]在斯洛伐克山区使用该方法进行了雪深反演，结果表现较好；Zhang et al. ^[47]对格林兰冰盖实测雪深与 GPS 反演雪深比较，发现冰川会导致 GPS 反演雪深与实测雪深存在差异，结果表明 GPS 干涉反射法在极寒地区的适用性不佳。

在青藏高原地区也有学者应该该方法进行过雪深反演，Zhang et al. ^[44]在青藏高原东北部利用 GPS 地面接收器结合气象数据和模型反演了土壤含水量和积雪深度，反演的积雪深度较实测站点偏高，相关系数为 0.72，RMSE 为 7.57 cm，线性关系较好。尽管 GPS 干涉反射计反演雪深成本较低，但其需要固定的地面 GPS 接收器，而相较于其他地区^[48-50]，目前在高原地区的 GPS 地面接收器偏少^[44]，GPS 接收器多架设在国家级的基础观测台站，可用于反演雪深的数据偏少，限制了这一方法在高原的应用。

1.2.4 物理模型模拟

基于模型模拟的再分析雪深数据集是目前青藏高原地区积雪研究中常用的雪深数据源之一，它使用具有实际物理的质量和能量方程，通过微分和积分方程对地球系统中的大气、地表状态进行模拟，得到高时空分辨率的大气、海洋和陆地小时级的模拟数据，并且常常通过数据同化，将大量的地面观测数据融合至模型中以提高模型模拟的精度。再分析数据集是目前在山区开展大范围积雪水文、气候研究使用较多的一类数据集，如王婷曾利用 ERA-Interim 数据集探讨了青藏高原积雪与陆面水热输送过程^[51]；Liu et al.^[52]使用 GLDAS 再分析数据集模拟了青藏高原地区径流，探究了当地径流对亚洲主要大江大河的影响等。

但不同再分析数据集由于核心计算方程如水汽交换和能量收支等的差异^[53]，以及各自物理模型的局限，可能无法对地表复杂过程如区域的风吹雪进行准确模拟，导致模拟的地表雪深数据与实际存在差异，这将影响再分析数据集在青藏高原地区的应用^[53]。而目前由于青藏高原雪深观测资料稀缺，对再分析雪深数据集评估的研究较少，Orsolini et al.^[54]利用遥感数据和 33 个地面台站雪深数据对 ERA5、ERA-Interim、JRA55 和 MERRA-2 再分析雪深数据集在青藏高原的表现进行评估，结果显示 ERA5 雪深数据集对高原雪深存在高估，MERRA2 和 JRA55 再分析数据集表现最好。但该评估结论主要基于青藏高原东部和南部的实测站点，对高原中西部区域的再分析的雪深产品的表现认识不足，给再分析雪深数据集在青藏高原中西部的应用增加较大不确定性，因此对再分析雪深数据集在青藏高原中西部的表现进行评估，能够提升对再分析雪深数据在青藏高原表现的认知。

1.2.5 主被动微波遥感

与传统的光学遥感易受云、可见光影响不同，微波遥感可穿透云层，实现对地面的全天候观测，凭借其低成本、大范围观测等特点，微波遥感成为我们研究山区雪深变化的重要手段。根据是否能发射微波信号，微波遥感被分为主动微波遥感和被动微波遥感两类。

主动微波遥感反演雪深的理论基础是，在一定深度条件下，雪深越大，积雪场产生的后向散射越强^[55]。主动微波遥感卫星发射的微波信号到达地面时，在积雪场发生后向散射，散射信号被卫星传感器接收，而传感器在接收到散射辐射后，会以微波亮度温度（Brightness Temperature，Tb，亮温）形式存储信息，后期经过反演算法得到地表雪深信息，比较常见的主动微波遥感有合成孔径雷达（Synthetic Aperture Radar，SAR）和散射计。1991 年和 1995 年分别发射的 ERS-1、ERS-2 SAR 传感器是最早发射的 SAR 传感器，仅有单一的 VV 极化通道，被用来监测海洋和冰，其后陆续发射了多颗传感器观测地表，多极化通道 SAR 成

为主流传感器类型^[56]。早期主动微波遥感常用来分辨地表湿雪，研究学者通过比较积雪时期与无雪时期的后向散射系数，提出当散射系数低于临界阈值（-3 dB）时，地面积雪为湿雪^[57]。主动微波遥感在积雪深度反演中有较多应用，如 Lievens et al. ^[58]利用哨兵一号（Sentinel-1）合成孔径雷达反演了1 km分辨率的北半球积雪深度，与地面台站雪深数据的相关系数达0.76。Liu et al. ^[59]利用三维变分融合算法结合哨兵一号合成孔径雷达数据反演了新疆巴音布鲁克盆地积雪雪深，提升了主动微波在区域的雪深反演算法的精度。Li et al. ^[60]在新疆北部使用InSAR技术反演了小河流盆地雪深，反演结果与野外实测数据拟合较好。尽管有学者对青藏高原地区也开展过雪深反演工作^[58, 61]，但由于星载主动微波遥感卫星回访周期较长（6~12 d），其中回访周期最短的哨兵一号卫星一般在6 d左右，但在实际应用过程中，其在青藏高原及其周边地区的回访周期一般在12 d左右^[62]，使得主动微波遥感相较被动微波遥感，可用于反演雪深的数据偏少^[63]，无法得到目标研究区连续且稳定的数据，限制了该方法在青藏高原地区的应用。

相较于上述其他方法，被动微波遥感具有采样频率高、易获取等特点，被认为是反演大范围积雪深度最有效的手段^[64, 65]，目前在青藏高原遥感雪深反演领域应用最多、最广。被动微波遥感反演雪深的理论基础：在低于积雪深度反演阈值时，积雪深度越大，对下层土壤的散射作用越强，地表辐射经积雪场散射后，卫星传感器接收到的亮温也越高^[66]，通过对亮温数据分析进而得到雪深。国内外学者基于辐射传输（RT）理论^[67]，提出了多个雪深反演模型，我国金亚秋院士在1994年提出了矢量辐射传输理论（VRT），并用于模拟积雪亮温^[68]。Wiesmann et al. ^[69]在1999年提出的MEMLS模型，将积雪场分为多个雪层，基于六流近似理论计算辐射在雪层间的反射、散射和吸收，实现了对积雪场亮温的模拟。Tsang et al. ^[70]通过假设积雪场为密集分布的中等尺度粒子，考虑积雪粒子内部及相互之间的影响，使用密集介质辐射传输模型（Dense Media Radiative Transfer, DMRT）模拟了积雪场的亮温，对1 m雪深的模拟结果较好。尽管这些辐射传输模型对积雪场亮温的模拟较为准确，但在实际应用中，模拟要求的输入参数包括积雪含水率、密度、粒径等指标，往往需要实地测量。由于对偏远山区的实地测量难度较大，这使得国内外学者对模型进行了简化，得到了被动微波雪深反演经验性算法。

最初的经验算法由Chang et al. ^[66]提出，其利用辐射传输模型模拟了不同属性积雪场的多频率辐射亮温，揭示了18/37 GHz频率微波水平亮温差数据用于反演雪深的理论可行性，并基于亚欧、加拿大地区的野外实测积雪数据，建立了全球的线性雪深反演算法（Chang1算法），该算法假定积雪颗粒是半径为0.3 mm的理想球体，积雪密度为0.3 g/cm³，公式如下：

$$SD = 1.59 \times (Tb_{18H} - Tb_{37H}) \quad (1-1)$$

式中: SD 为雪深 (下同), Tb_{18H} 、 Tb_{37H} 分别代表 18、37 GHz 水平极化亮温 (下同)。

其后 Aschbacher^[71]基于扫描多通道微波辐射计 (Scanning Multichannel Microwave Radiometer, SMMR) 的 18 GHz 和 37 GHz 的水平和垂直极化下的亮温作为雪深反演参数, 提出了 SPD (Spectral Polarization Difference) 算法, 该算法降低了冰壳、雪粒粒径等对雪深反演的影响, SPD 算法公式如下:

$$SD = 0.68 \times SPD + 0.67 \quad (1-2)$$

$$SPD = (Tb_{18V} - Tb_{37V}) + (Tb_{18V} - Tb_{18H}) \quad (1-3)$$

式中: SPD 代表极化光谱差, Tb_{18V} 、 Tb_{37V} 分别代表 18 GHz 垂直极化亮温、37 GHz 垂直极化亮温。

考虑到森林覆被对地表散射辐射传输的影响, Foster et al. ^[72]通过引入森林覆盖率参数并依据当地积雪粒径数据对 Chang1 算法进行了修正, 增加了 Chang1 算法在北美和欧亚大陆的适用性, 修正后的算法 (Foster 算法) 如下所示:

$$SD = 0.78 \times \frac{Tb_{18H} - Tb_{37H}}{1 - f} \quad (1-4)$$

式中: f 表示森林覆盖率 (下同)。

Kelly^[73]利用地球观测系统的高级微波扫描辐射计 (The Advanced Microwave Scanning Radiometer for Earth Observing System, AMSR-E) 亮温数据提出了针对森林区和非森林区的反演算法, 利用 89 GHz 通道亮温数据提高了算法对浅雪 (小于 5 cm) 的检测能力和森林稀疏区的雪深反演能力。在这个基础上, 基于二代高级微波扫描辐射计 (Advanced Microwave Scanning Radiometer 2, AMSR2) 传感器的改进雪深反演算法被提出, 其被作为 AMSR2 雪深产品的标准业务化算法^[74], 其具体的算法公式 (AMSR2 算法) 如下:

$$SD = f * (SD_f) + (1 - f) * (SD_o) \quad (1-5)$$

$$SD_f = \frac{1}{\log_{10}(Pol_{36}) \times (Tb_{19V} - Tb_{36V}) / (1 - 0.6 * f)} \quad (1-6)$$

$$SD_o = \frac{1}{\log_{10}(Pol_{36}) \times (Tb_{10V} - Tb_{36V})} + \frac{1}{\log_{10}(Pol_{19}) \times (Tb_{10V} - Tb_{19V})} \quad (1-7)$$

$$Pol_{36} = Tb_{36V} - Tb_{36H} \quad (1-8)$$

$$Pol_{19} = Tb_{19V} - Tb_{19H} \quad (1-9)$$

式中: SD_f 代表遥感卫星瞬时视场中森林覆被地表区域的反演雪深, SD_o 代表遥感卫星瞬时视场中非森林覆被地表区域的反演雪深, 下标中的 H 和 V 分别代表不同频率亮温通道的水平和垂直极化方向, 下标的 10、19、36 分别代表 10、

19、36 GHz（下同）。

青藏高原地区被动微波雪深反演研究自 20 世纪 90 年代就已开始，早期的 Chang1 算法是基于美国科罗拉多地区积雪属性建立的，而中国西部积雪密度普遍偏小、深度偏浅^[75]，因此 Chang et al.^[76]针对中国西部地区对 Chang1 算法进行了修正，修正后的 Chang2 算法公式如下：

$$SD = 2.0 \times (Tb_{18H} - Tb_{37H}) - 8 \quad (1-10)$$

曹梅盛和李培基^[75]利用 SMMR 传感器亮温数据，针对青藏高原不同地貌单元对 Chang1 算法提出修正，得到了对高原雪深的初步认识，修正后的算法（Cao 算法），Cao 算法公式如下：

$$SD = 1.59 \times (Tb_{18H} - Tb_{37H}) - 8 \quad (1-11)$$

柏延臣等^[77]利用特殊传感器微波/图像（Special Sensor Microwave/Image, SSM/I）19/37 GHz 水平极化亮温差对 Cao 算法的评估结果显示，Cao 算法对高原雪深整体存在高估。Che et al.^[78]基于中国地面站点数据，分别针对 SMMR 和 SSM/I 传感器，修正了 Chang1 算法，并提出了适合中国地区的动态雪深反演算法（简称 Che 算法），两传感器反演雪深的 RMSE 分别为 6.22 和 5.22 cm，并基于该算法制作了中国雪深长时间序列数据集（WESTDC, Long-term snow depth dataset of China）雪深产品，Che 算法公式如下：

$$SD = 0.66 \times (Tb_{18H} - Tb_{37H}) + b \quad (1-12)$$

式中：b 代表冬季 Che 算法的雪深补偿值，具体值见参考文献[78]。

Jiang et al.^[79]利用风云三号 B 星（FY3B）亮温数据及站点数据，建立了地表覆被像元线性混合的反演算法（简称 Jiang 算法），Jiang 算法公式如下所示：

$$SD = f_{farmland} \times SD_{farmland} + f_{grass} \times SD_{grass} + \quad (1-13)$$

$$f_{baresoil} \times SD_{baresoil} + f_{forest} \times SD_{forest}$$

$$SD_{farmland} = -4.235 + 0.432 \times (Tb_{18H} - Tb_{36H}) \quad (1-14)$$

$$+ 1.074 \times (Tb_{89V} - Tb_{89H})$$

$$SD_{grass} = 4.320 + 0.506 \times (Tb_{18H} - Tb_{36H}) - 0.131 \times (Tb_{18V} - Tb_{18H}) \quad (1-15)$$

$$+ 0.183 \times (Tb_{10V} - Tb_{89H}) - 0.123 \times (Tb_{18V} - Tb_{89H})$$

$$SD_{baresoil} = 3.143 + 0.532 \times (Tb_{36H} - Tb_{89H}) - 1.424 \times (Tb_{10V} - \quad (1-16)$$

$$Tb_{89V}) + 0.183 \times (Tb_{18V} - Tb_{89V}) - 0.238 \times (Tb_{36V} - Tb_{89V})$$

$$SD_{forest} = 11.128 - 0.474 \times (Tb_{18H} - Tb_{36V}) - 1.441 \times (Tb_{18V} \quad (1-17)$$

$$- Tb_{18H}) + 0.678 \times (Tb_{10V} - Tb_{89H}) - 0.649 \times (Tb_{36V} - Tb_{89H})$$

式中： $f_{farmland}$ 、 f_{grass} 、 $f_{baresoil}$ 、 f_{forest} 分别表示 FY3B 卫星瞬时视场中耕地、草

地、裸地及森林土地覆被类型的面积百分比,而 $SD_{farmland}$ 、 SD_{grass} 、 $SD_{baresoil}$ 、 SD_{forest} 代表相应土地覆被类型反演的雪深,当被动微波遥感瞬时视场的土地覆被类型大于 85 %时,就认为当前瞬时视场为该土地覆被类型的“纯”像元。

Dai et al. [65] 基于被动微波 AMSR-E 亮温数据,反演了青藏高原地区的积雪深度,反演雪深与野外实测雪深的相关系数达到了 0.574。Dai et al. [80] 将地表辐射纳入雪深反演算法中,提高了雪深反演精度。Xiao et al. [81] 提出了支持向量回归方法 (Support vector regression, SVR),与以往的四种算法进行了对比,降低了雪深的不确定性。Wang et al. [23] 提出利用降尺度方法对被动微波遥感 AMSR2 进行处理,反演雪深的 RMSE 为 0.58 cm,误差小于 3 cm。

尽管目前已有激光雷达、GPS 干涉反射法、物理模型模拟和主被动微波遥感等多种雪深监测方法,但由于青藏高原地形复杂和交通不便,再分析模拟数据和被动微波遥感反演雪深是目前获取青藏高原雪深信息的经济、高效的研究手段,在青藏高原地区应用较多。

以往学者对再分析雪深资料和被动微波雪深产品开展过评估工作,如肖林等 [82] 对中国区域多源雪深产品评估发现,AMSR-E 雪深产品在青藏高原东南部存在高估。李小兰等 [30] 通过对比中国地区实际雪深与被动微波遥感反演雪深资料,发现被动微波反演雪深在青藏高原整体存在低估,在东南部存在高估。乔德京 [53] 评估结果显示,AMSR-E 雪深产品在青藏高原地区整体呈现高估,WESTDC 雪深产品在青藏高原的综合表现较好。但目前评估工作中使用的青藏高原地面台站资料多集中分布于青藏高原的东部和南部,部分地面台站数据分布于低海拔地区,高原中西部雪深评估资料稀疏,导致对青藏高原中西部雪深特征认识不足。同时,针对多种雪深反演算法的评估多以被动微波雪深反演算法的标准化雪深产品为对象,但产品所使用的被动微波遥感数据源和数据处理流程等也不同 [73, 78, 83]。此外,当前应用广泛的被动微波雪深反演经验算法多基于被动微波亮温差,较少考虑地形、下垫面等环境要素对雪深反演的影响,导致雪深反演经验算法误差较大。

因此本文基于青藏高原中部色林错、纳木错湖区连续两年的雪深观测资料和 AMSR2 亮温数据,对雪深再分析数据集和被动微波雪深反演算法在高原中部区域表现进行评估,分析了当前雪深产品和被动微波雪深反演在研究区的适用性,并利用被动微波亮温和环境要素等变量建立了机器学习模型,探究不同频率亮温、下垫面和地形要素对被动微波雪深反演的影响,以期结果能为青藏高原被动微波雪深反演研究提供参考。

1.3 研究内容和技术路线

1.3.1 研究内容

（1）再分析雪深数据集评估

以地面台站数据作为验证数据，对 ERA5、JRA55、GLDAS、MERRA2 四种常见的再分析雪深数据集在青藏高原中部色林错、纳木错湖区的适用性进行评估，评估结果能够降低再分析雪深资料在青藏高原中部应用的不确定性。

（2）被动微波雪深反演算法评估

基于观测雪深数据与 AMSR2 亮温数据，对 Chang2 算法、SPD 算法、Che 算法、Jiang 算法和 AMSR2 算法等五种被动微波雪深反演算法在研究区表现进行评估，结果有助于为青藏高原中部区域的被动微波雪深反演业务化提供参考。

（3）基于机器学习的被动微波雪深反演模型

使用 LASSO、SCAD、MCP 和随机森林特征重要度四种方法对草地、裸地和水体下垫面变量，坡度、坡向和海拔等地形变量，5 个不同频率（10、19、24、37、89 GHz）的垂直/水平亮温变量及不同亮温间的亮温差变量等 61 个变量进行，基于筛选得到的变量组合，使用随机森林回归和支持向量回归，构建了十种机器学习雪深反演模型，期望实验结果为青藏高原被动微波雪深反演算法研究提供新的思路。

1.3.2 技术路线

为了实现上述研究内容，先对雪深数据和亮温数据分别进行预处理，再基于再分析雪深数据集和 AMSR2 亮温数据集对应空间范围的土地覆被类型、海拔、积雪覆盖率数据，对研究区七个台站的代表性进行了评估，然后基于代表性较好站点的观测雪深数据，对 ERA5、JRA55、GLDAS、MERRA2 再分析雪深数据集和 Chang2 算法、SPD 算法、Che 算法、Jiang 算法和 AMSR2 算法进行评估。其后，使用 LASSO、SCAD、MCP 和随机森林特征重要度四种方法，对坡度、坡向、海拔，草地、水体、裸地土地覆被类型占比，和 AMSR2 亮温数据中的 10、19、24、37、89 GHz 垂直/水平极化亮温及不同频率亮温差等 61 个变量进行选择。根据变量筛选结果，使用随机森林回归和支持向量回归建立机器学习雪深反演模型。

具体的技术路线图见图 1-1：

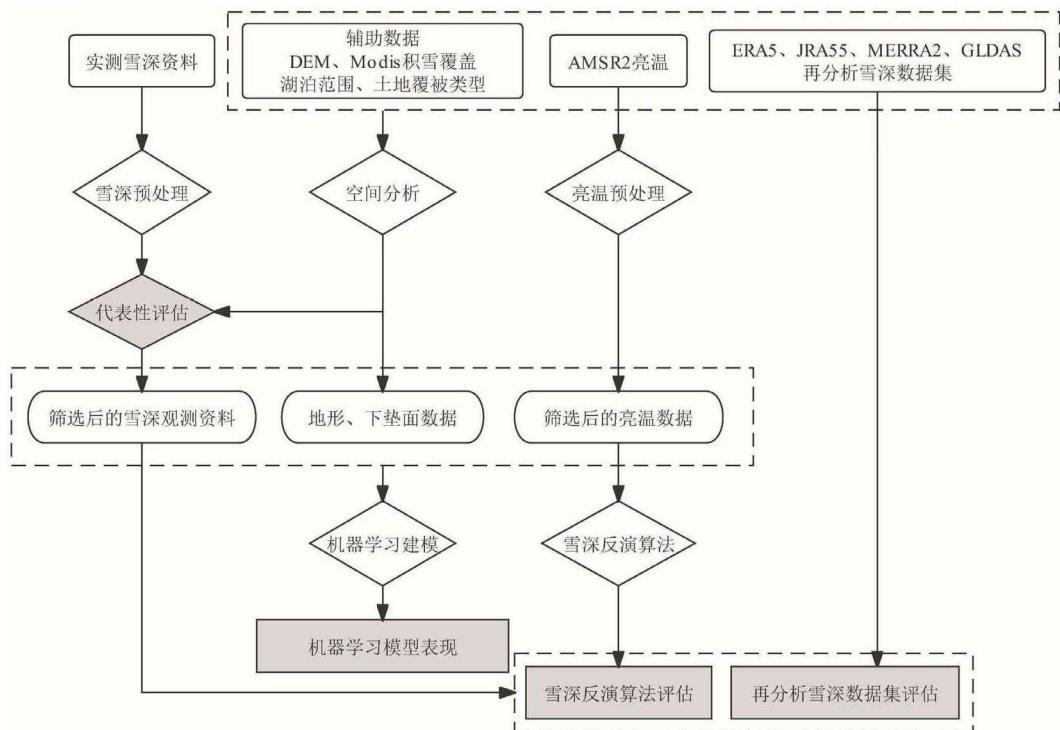


图 1-1 技术路线图

Figure 1-1 The scheme of technology roadmap

1.4 论文结构

本文共五章，结构如下：

第一章为绪论，介绍选题依据、研究意义、研究现状、研究内容和论文结构。

第二章为研究区概况和研究方法与数据，介绍了本文研究区概况、实验数据来源与研究方法。

第三章为再分析雪深产品与被动微波反演算法评估，对站点代表性进行评价，基于代表性较好的实验数据，评估了各再分析数据集雪深资料和被动微波雪深反演算法在研究区的表现。

第四章为基于机器学习的被动微波雪深反演，对机器学习模型初始变量的选择结果与机器学习雪深反演模型表现进行分析与讨论。

第五章为结论与展望，概括了本文的主要结论和创新点，对本研究的不足之处提出对应的解决思路。

第二章 研究区概况和研究方法与数据

2.1 研究区概况

青藏高原分布着约 1200 个面积大于 1 km^2 的湖泊^[84]，总面积约占全国湖泊面积一半，按地貌、水系和湖泊分布特点，湖泊群可被分为青海湖—柴达木盆地湖区、羌塘湖区、高原东部湖区、高原南部湖区，分别位于阿尔金山—祁连山和东昆仑山—布尔汗布达山之间、昆仑山与冈底斯山—念青唐古拉山之间、青海湖—柴达木盆地湖区以南和羌塘湖区以东、冈底斯山—念青唐古拉山和喜马拉雅山之间^[85]。

本文研究区色林错、纳木错湖区位于青藏高原中部（见图 2-1），属羌塘湖区，其中色林错湖区的多年平均气温为 -1.8°C ，降水量达 389.4 mm ，风速为 $3.7 \text{ m}\cdot\text{s}^{-1}$ ^[86]。色林错和纳木错两地湖区气候类型均为高原亚寒带，以半干旱气候为主^[87]，当地植被类型以高寒草甸为主，湖泊海拔高度分别为 4539 m 、 4723 m ^[88]，湖泊面积分别是 2178 km^2 、 2018 km^2 ，分别是西藏自治区面积第一、第二大湖泊，其中色林错自上世纪八十年代以来湖泊面积扩张较明显，面积增长率约 40% ，是我国当前面积第二大的咸水湖。两地积雪类型均属于非周期性不稳定积雪区，多年平均连续积雪天数小于 10 天，多年平均累计积雪天数在 $10 \sim 60$ 天之间^[5]。纳木错湖区年均雪深为 6.84 cm ，年均雪水当量达 13.5 mm ，积雪密度平均为 $0.18 \text{ g}\cdot\text{cm}^{-3}$ ^[89]。

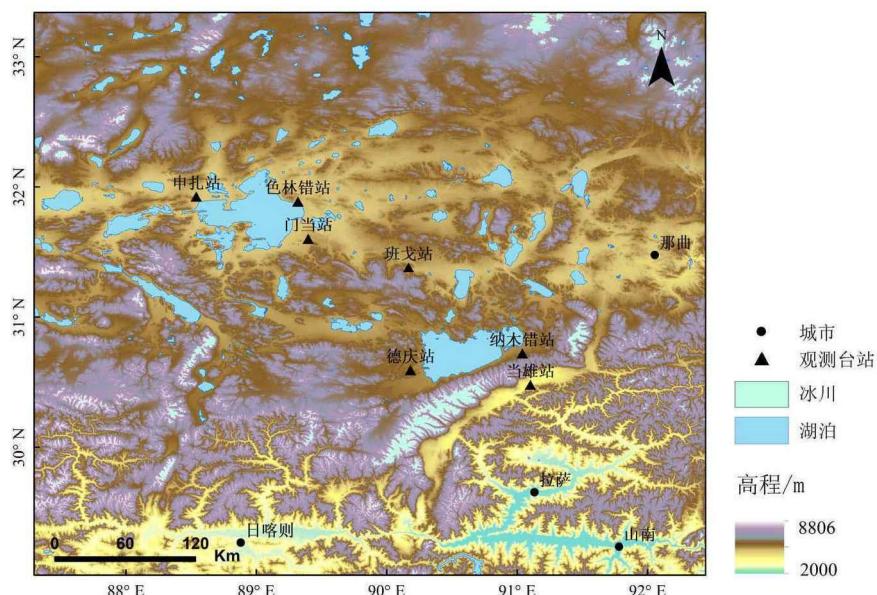


图 2-1 研究区观测站点分布

Figure 2-1 The distribution of meteorological stations in the study area

2.2 数据介绍

2.2.1 实测雪深数据

本文雪深观测数据来自于 5 个野外自动观测气象站以及 2 个国家气象局 (<http://data.cma.cn/>) 气象站 (见表 2-1)，其中纳木错站、德庆站、色林错站、门当站和申扎站等五个站点使用 SR50A 超声波雪深观测仪监测地表雪深，该仪器在青藏高原地区的表现经过检验^[90]，监测雪深与人工测量雪深的相关系数能达到 0.97^[91]，五个台站收集了 2019 年 5 月至 2021 年 8 月期间的雪深数据 (申扎站由于仪器故障，仅收集到 2019 年 7 月 ~ 2020 年 10 月间的雪深数据，且冬季雪深存在缺失)，仪器声波速度经过温度校正，测量精度为 ± 0.6 cm，时间分辨率为 0.5 h。

SR50A 传感器均架设距地面 1.4 m 左右高度，其通过记录超声波脉冲发射和到达目标后返回的时间，计算得到传感器与目标之间的高度 (见图 2-2)。当地面无雪时，传感器测量高度为仪器架设高度；当地面存在积雪时，传感器测量距离为传感器与积雪场上层之间的距离，通过计算两时期传感器距离差值，可得到地表雪深数据。

当雄站和班戈站为国家气象局观测站 (见图 2-1)，雪深数据是气象站工作人员于每日上午进行人工观测，观测雪深结果取三次测量平均值，可靠性高。

为了减少积雪季节变化对本文结果的影响，本文将前一年的 7 月 31 日至第二年的 6 月 30 日作为一个积雪年，雪深评估以积雪年为单位，因此本文评估所使用的雪深数据时间范围为 2019 年 7 月 31 日 ~ 2020 年 6 月 30 日、2020 年 7 月 31 日 ~ 2021 年 6 月 30 日。

表 2-1 雪深观测台站基本信息

Table 2-1 Basic information of meteorological stations

台站名称	经度	纬度	海拔/m
纳木错	91°02'	30°43'	4925
德庆	90°10'	30°35'	4864
色林错	89°18'	31°53'	4685
门当	89°23'	31°36'	4747
申扎	88°32'	31°55'	4706
当雄	91°06'	30°29'	4200
班戈	90°01'	31°23'	4700

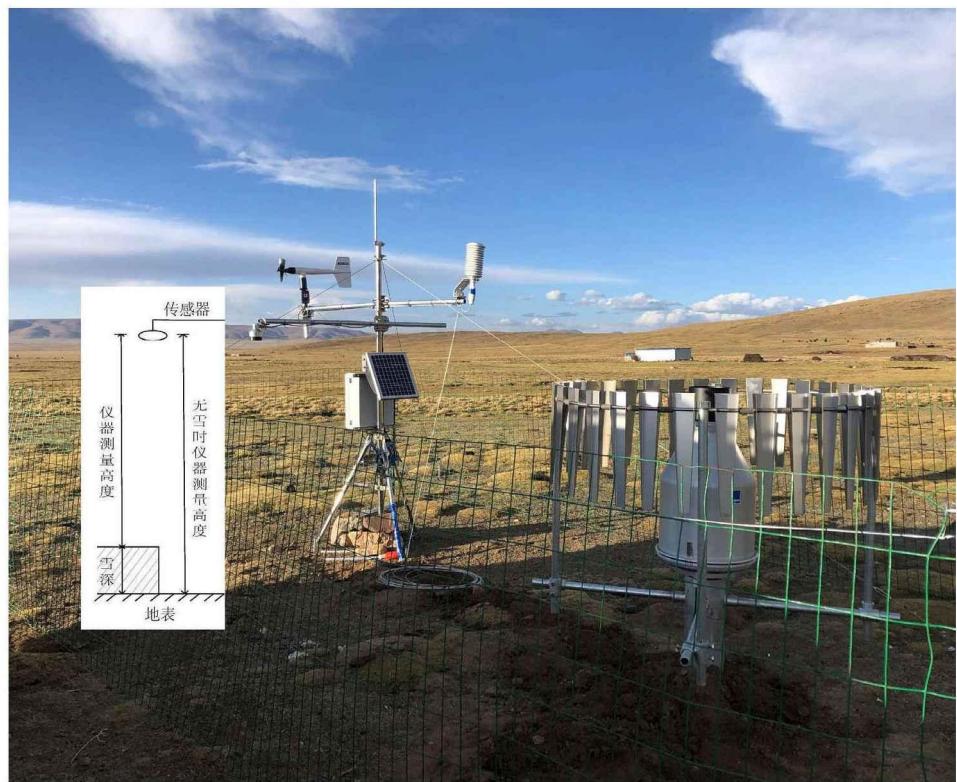


图 2-2 SR50A 工作原理示意图
Figure 2-2 Schematic diagram of the working principle of SR50A

2.2.2 DEM 数据

本文所使用的研究区数字高程模型 (DEM) 是第四版的 SRTM (Shuttle Radar Topography Mission) 数据, 由国际热带农业中心 (International Center for Tropical Agriculture, CIAT) 制作得到, 数据集的空间分辨率为 90 m, 数据下载于国家青藏高原科学数据中心 (<https://data.tpdc.ac.cn/>) [92]。

2.2.3 土地覆被类型数据

本文所使用的土地覆被 (Land cover) 类型数据来源于美国宇航局 (NASA) 生产的中分辨率成像光谱仪 (Moderate Resolution Imaging Spectroradiometer, MODIS) 地表覆被全球产品 (MCD12Q1), 该数据集使用 MODIS 传感器每年收集到的数据处理得到, 产品空间分辨率为 500 m。本文选取了 2019 年地表覆被产品, 下载于美国地质勘探局 (The United States Geological Survey, USGS) (<https://earthexplorer.usgs.gov/>), 该产品使用了国际地圈—生物圈计划 (the International Geosphere-Biosphere Program, IGBP) 土地分类规则。

2.2.4 AMSR2 亮温数据

AMSR2 传感器搭载于全球变化观测计划中的水循环变动观测卫星 (Global

Change Observation Mission 1st-Water, GCOM-W1) 上, 性能较以往微波辐射计得到了较大提升, 传感器参数见表 2-2。

本文使用的 AMSR2 被动微波亮温数据通过 NASA 戈达德地球科学数据和信息服务中心 (Goddard Earth Sciences Data and Information Services Center, GES DISC) 下载 (<https://disc.gsfc.nasa.gov/>), 该产品对日本航空航天局 (Japan Aerospace Exploration Agency, JAXA) 的 AMSR2 一级亮温数据进行了重采样, 重采样后的产品空间分辨率为 $25 \times 25 \text{ km}$, 数据经过校准和质量控制。

AMSR2 卫星的过境频率为一天两次, AMSR2 亮温数据存在升轨和降轨两类, 升轨产品在研究区的过境时间在北京时间下午三点左右, 降轨产品的过境时间在北京时间凌晨两点左右。由于凌晨研究区温度较低, 不利于积雪融化, 因此本文选择了 2019 年 7 月至 2021 年 6 月两个积雪年期间的 AMSR2 降轨产品。

表 2-2 AMSR2 传感器参数信息
Table 2-2 Sensor parameters of AMSR2

中心频率/ GHz	极化方向	波束宽度/MHz	灵敏度/K	瞬时视场 (km × km)
6.93	V, H	350	0.3	35×62
7.3	V, H	350	0.3	35×62
10.65	V, H	100	0.6	24×42
18.7	V, H	200	0.6	14×22
23.8	V, H	400	0.6	15×26
36.5	V, H	1000	0.6	7×12
89.0	V, H	3000	1.1	3×5

2.2.5 再分析数据集

本文选择进行评估的再分析数据集为 ERA5、JRA55、GLDAS、MERRA2, 它们是目前青藏高原地区雪深研究中常用到的数据集。

ERA5 雪深数据集是由欧洲中尺度天气预报中心 (European Centre for Medium-Range Weather Forecasts, ECMWF) 发布的第五代再分析数据集, 数据集的时间跨度从 1979 年至今, 时间分辨率为 1 h, 空间分辨率为 $0.1^\circ \times 0.1^\circ$ ^[93]。

JRA55 数据集是第一个应用四维变分分析的再分析数据集, 与以往版本相比, 其提高了模型分辨率与温度分析的时间一致性, 并增加了新的同化数据源。本文所使用的 JRA55 数据集下载于美国国家大气研究中心 (National Center for

Atmospheric Research, NCAR) (<https://rda.ucar.edu/>)^[94], 数据集空间分辨率为 $0.56^\circ \times 0.56^\circ$, 时间分辨率为 6 h。

GLDAS (Global Land Data Assimilation System) 数据集包括四种地表模型, 分别为 Noah、Catchment (CLSM)、社区土地模型 (Community Land Model, CLM) 和可变渗透能力 (Variable Infiltration Capacity, VIC), 本文使用的是 Noah 模式下的雪深数据集^[95], 其空间分辨率为 $0.25^\circ \times 0.25^\circ$, 时间分辨率为 3 h, 数据下载于 GES DISC (<https://disc.gsfc.nasa.gov/>)。

第二次现代时代研究和应用回顾分析数据集 (The second Modern-Era Retrospective analysis for Research and Applications, MERRA2) 使用了新的戈达德地球观测系统 (Goddard Earth Observing System Model Version 5, GEOS-5) 和全球统计插值 (GSI) 分析方案, 能够同化最新的微波探测仪和高光谱红外辐射仪数据, 该数据集时间范围为 1980 年至今, 时间分辨率为 1 h, 空间分辨率为 $0.5^\circ \times 0.625^\circ$, 数据集下载于 GES DISC (<https://disc.gsfc.nasa.gov/>)。

表 2-3 再分析数据集的基本信息
Table 2-3 Introduction to reanalysis datasets

数据名称	发布机构	时间范围	时间分辨率	空间分辨率
ERA5	ECMWF	1979 年至今	1 h	$0.1^\circ \times 0.1^\circ$
JRA55	NCAR	1958 年至今	6 h	$0.56^\circ \times 0.56^\circ$
GLDAS	GES DISC	2000 年至今	3 h	$0.2^\circ \times 0.25^\circ$
MERRA2	GES DISC	1980 年至今	1 h	$0.5^\circ \times 0.625^\circ$

2.2.6 MODIS 积雪范围产品

本文使用的积雪范围数据是 MODIS 积雪范围 (Snow Cover) 产品 (MOD10_L2), 该产品使用了地球观测系统 (Earth Observation System, EOS) 中 Terra 卫星的 MODIS 传感器数据, 利用卫星波段 4 和波段 6 计算得到的归一化积雪指数 (Normalized Difference Snow Index, NDSI), Terra 卫星的过境时间在当地上午 10: 30 左右, 数据下载自美国国家冰雪数据中心 (Nation Snow and Ice Data Center, NSIDC) (<https://nsidc.org/>)^[96], 数据格式为 HDF-EOS2, 空间分辨率为 500 m, 数据时间范围为 2019 年 7 月至 2021 年 6 月两个积雪年。

2.2.7 湖泊范围数据

本文所使用的纳木错、色林错湖区的湖泊范围数据来源于青藏高原大于 1 平方公里湖泊数据集^[97], 该数据集发布于国家青藏高原科学数据中心

(<http://data.tpcdc.ac.cn/>)，该数据集基于 Landsat 遥感影像数据处理获得，时间范围涵盖 1970 ~ 2018 年，为了减小湖泊范围变化对本文评估结果的影响，本文选取了 2018 年最新的研究区湖泊范围数据。

2.3 研究方法

2.3.1 雪深数据预处理方法

为了剔除杂草对雪深数据影响，在评估前需要对雪深数据进行预处理，处理流程如下：首先结合站点同步观测的降水、气温数据，进行交叉验证判断降雪事件是否发生，并剔除异常、错误数据。降雪事件判断标准为：有降水且日测量高度与前一日变化超过 1.2 cm 且温度低于零摄氏度。

然后依据研究区地表草地生长周期，设置地面初始高度。由于研究区降雪事件多发生于当年 9 月至次年 5 月^[35]，对应草地的枯黄期与返青期^[36]，故将前一年枯黄期稳定无雪的测量高度作为积雪场积累期、稳定期的地面高度，将第二年积雪完全融化后的草地返青期的平均高度作为积雪消融期的地面高度，其中地面无雪判断标准为：连续三日无降水，日积温高于 0 °C，且测量高度变化幅度在 1.2 cm 范围内。

为了保证评估数据的一致性，所选取的雪深数据均与研究资料的时间尺度保持一致，在对再分析雪深资料评估时，使用台站的日平均雪深数据，在对被动微波雪深反演算法评估及机器学习反演雪深中使用 AMSR2 过境时刻的地表雪深数据，计算过程通过 NCL (The NCAR Command Language) 和 Python 实现。

2.3.2 积雪深度分类标准

在评估再分析数据集和被动微波雪深反演算法在不同深度的表现时，本文选择 5 cm、20 cm 深度阈值划分浅雪、中雪和深雪，其中 5 cm 雪深标准与以往学者的划定标准一致^[98]，而为了保证中、深雪阈值设定的客观性和合理性以及评估结果的可比性，统计了台站观测雪深中大于 5 cm 的雪深的中位值 (19.73 cm, ~ 20 cm)，将其作为中、深雪的划分阈值标准。

2.3.3 积雪覆盖率数据的处理流程

本文将再分析数据集、AMSR2 亮温数据在各站点的最小栅格范围作为积雪覆盖率的计算范围，为了避免数据处理对评估结果的影响，数据未进行空间插值处理，积雪覆盖率的处理流程如下：

首先通过空间分析，获得 ERA5、JRA55、GLDAS、MERRA2 再分析数据集

和 AMSR2 亮温数据在各站点的栅格数据的矢量范围数据，其中纳木错站和当雄站同时位于 MERRA2、JRA55 数据集的同一矢量范围内（见图 2-3a），色林错站和门当站位于 JRA55 数据集的同一矢量范围内。

李晓锋^[99]的研究表明，高原湖冰的平均 NDSI 为 0.8，而青藏高原草地覆被类型的雪深阈值仅为 0.33^[100]，表明矢量范围内湖泊下垫面的出现能够导致区域的平均 NDSI 值偏高。遥感实测数据也显示，冬季色林错、纳木错湖泊的 NDSI 值较周围陆面明显偏高（见图 2-4），整体也高于念青唐古拉山冰川积雪区的 NDSI 值。因此为了消除湖泊对 NDSI 计算结果的影响，本文基于青藏高原大于 1 平方公里湖泊数据集，将数据集对应区域内的湖泊区域剔除（见图 2-3b），通过 ArcGIS 中 Erase 功能实现。然后基于获得的 MODIS 每日积雪范围产品（MOD10_L2），计算得到各数据集在 7 个站点的日均 NDSI 值。

研究表明^[101]，积雪覆盖率可以通过 NDSI 的经验公式表示，如下：

$$SCF = a + b * NDSI \quad (2-1)$$

其中 SCF 代表积雪覆盖率，a、b 为固定参数，针对 Terra/Aqua 卫星有不同值。Terra 卫星过境时间为当地上午 10: 30，而 Aqua 卫星过境时间为当地下午 1: 30，本文选择的是 Terra 卫星数据，a、b 值分别为-0.01、1.45，得到的 SCF 数据经过质量检验。将 SCF 的雪深阈值设为 0.5，即当 SCF 大于 0.5，认为地面存在积雪。最后将各站点不同数据集所对应的 SCF 数据与各站点地面是否存在积雪进行相关分析，雪深数据使用日平均值，若同一栅格区域内存在两个台站，取两台站的平均值。

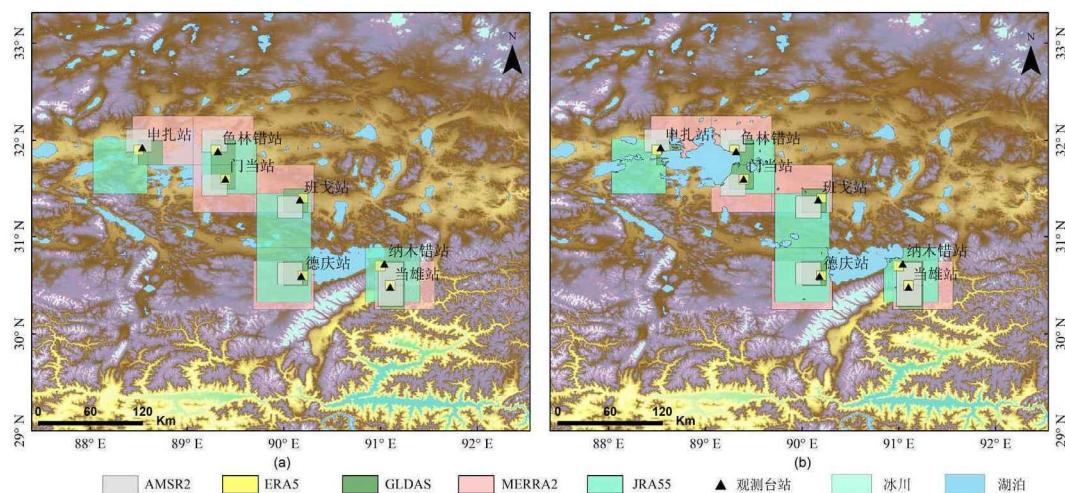


图 2-3 各数据集栅格区域空间分布，(a) 为未剔除湖泊范围，(b) 为剔除湖泊范围，数据来源见 2.2.4 AMSR2 亮温数据、2.2.5 再分析数据集与 2.2.7 湖泊范围数据

Figure 2-3 Spatial distribution of raster regions for each dataset. (a) The spatial extent of the unexcluded lakes. (b) The spatial extent of the lake is excluded. For data sources, see 2.2.4 AMSR2 Brightness Temperature Dataset, 2.2.5 Reanalysis Datasets and 2.2.7 Lake Cover Dataset.

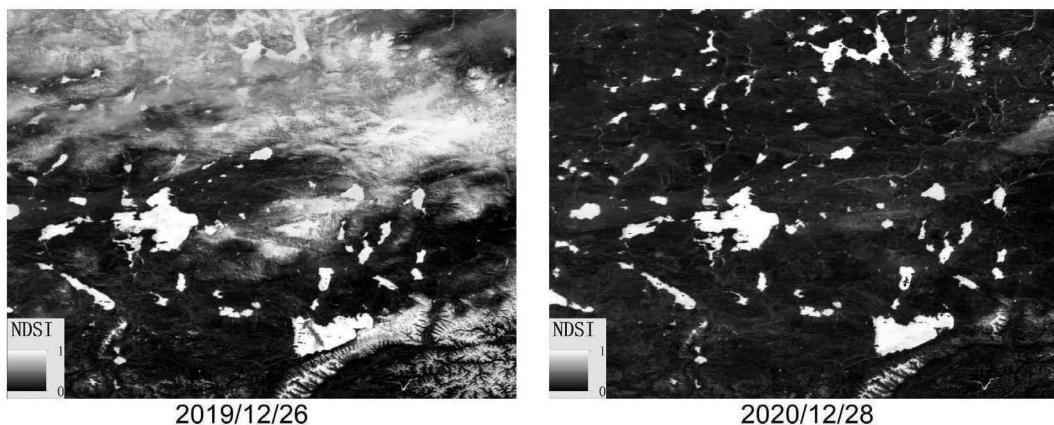


图 2-4 研究区冬季的 NDSI 分布, 数据来源见 2.2.6 MODIS 积雪范围产品

Figure 2-4 Spatial distribution of NDSI in winter in the study area. For data sources, see 2.2.6 MODIS Snow Cover Product.

2.3.4 地形数据处理方法

本文所使用到的坡度、坡向和海拔地形数据, 为各数据集对应栅格范围 (见 2.3.3 积雪覆盖率的数据处理方法, 图 2-3a) 的平均值, 其获取流程如下:

使用 2.2.2 DEM 数据, 计算不同各数据集对应区域的平均高程、平均坡向和平均坡度数据, 借助 ArcGIS 中 Spatial Analyst 工具模块完成。

2.3.5 土地覆被重分类方法

在评估的五种被动微波雪深反演算法中, Jiang 算法使用了地表农田、草地、裸地和森林四种地表覆被占比数据, AMSR2 算法使用了森林覆被率数据。为了满足被动微波雪深反演算法评估的需求, 本文对已有的 IGBP 地表覆被产品进行重分类, 分类规则见表 2-4, 将地表覆被分为森林、灌木、草地、耕地、裸地和水体六类。地表覆被重分类结果显示 (见图 2-5), 研究区土地覆被类型以草地、裸地和水体为主, 灌木类型占比较少, 重分类通过 ArcGIS 中 Spatial Analyst 工具模块中的重分类功能完成。

表 2-4 土地覆被重分类规则
Table 2-4 Land cover reclassification rules

编码	IGBP 分类规则	重分类
0	水体	水体
1	常绿针叶林	
2	常绿阔叶林	
3	落叶针叶林	森林
4	落叶阔叶林	

表 2-4 (续)
Table 2-4 (Continued)

编码	IGBP 分类规则	重分类
5	混合林	森林
6	密集灌木丛	灌木
7	稀疏灌木丛	
8	多树热带草原	
9	热带稀树草原	
10	草原	草地
11	永久湿地	
12	耕地	耕地
14	镶嵌型耕地/自然植被	
13	城市和建筑用地	
15	积雪和冰	裸地
16	裸地或稀疏植被	
254	未分类	未分类

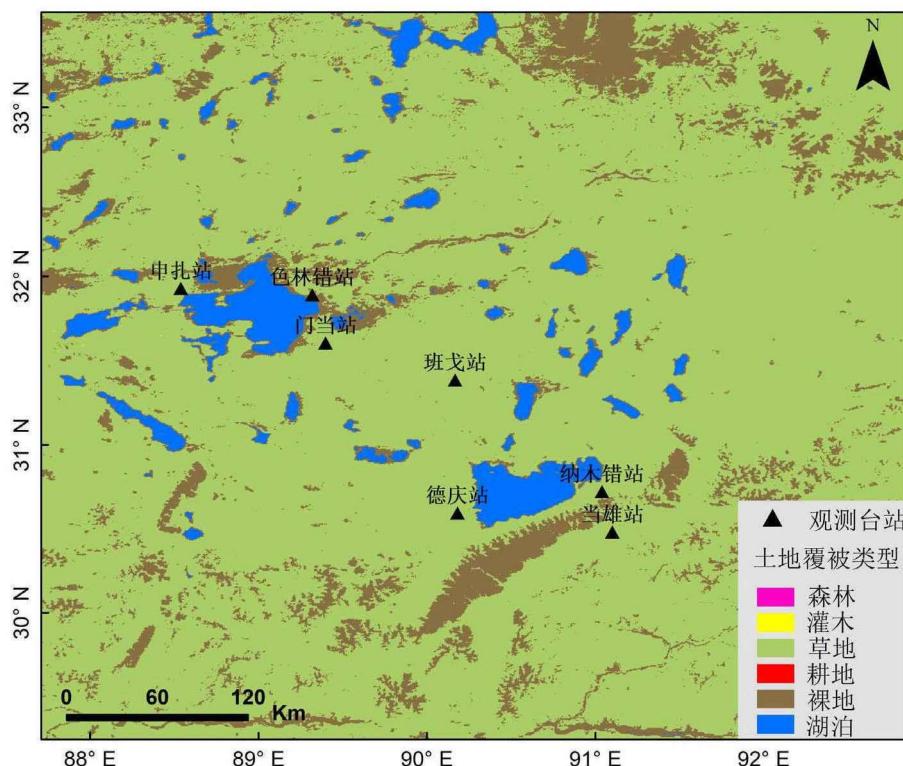


图 2-5 研究区重分类后的土地覆被类型, 数据来源见 2.2.3 土地覆被类型数据
Figure 2-5 Land cover types after reclassification in the study area. For data sources, see 2.2.3
Land Cover Types Dataset.

2.3.6 被动微波亮温数据预处理

降雨、寒漠和冻土具有与积雪十分相似的微波辐射特性，且当积雪场含水量增加时，积雪场对微波的散射能力减弱、吸收能力增强。本文仅考虑干雪雪深反演过程，为了避免非积雪散射体及湿雪对实验结果的影响，在利用被动微波亮温反演雪深时，需剔除非积雪散射体和湿雪亮温数据。本文采用 Kelly 等^[73]方法剔除降雨、寒漠和冻土等散射体，采用 Walker^[102]提出的方法剔除湿雪，具体积雪判别流程见图 2-6。

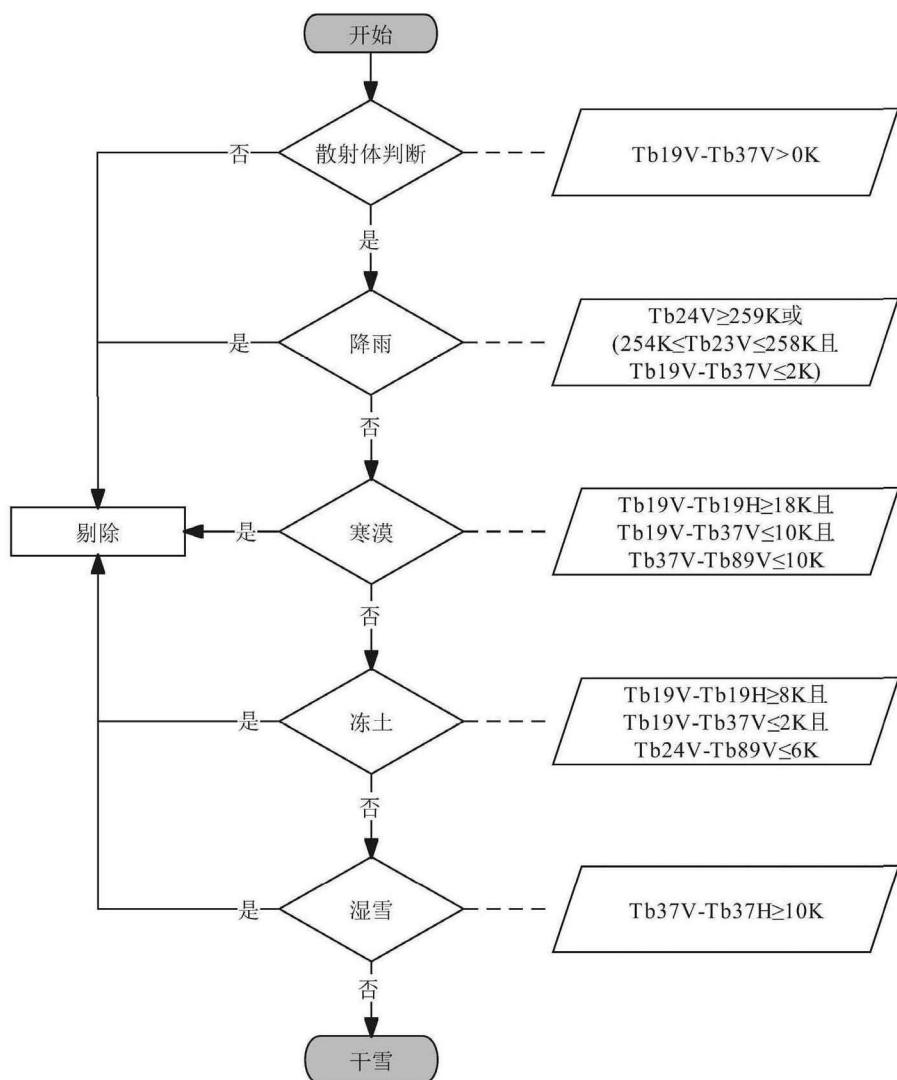


图 2-6 干雪积雪场识别流程

Figure 2-6 Discrimination process for dry snow

2.3.7 机器学习数据预处理方法

机器学习往往涉及较多变量，当变量间单位不统一时，可能会导致模型拟合精度下降，因此在数据被用于机器学习建模前，会对数据进行标准化处理，将变

量整体均值变为 0、方差变为 1，这一过程不改变原有数据分布特征，该操作借助 Python 语言中的“Sklearn”包完成。

2.3.8 机器学习变量选择方法

在统计分析建模中，如果将很多的变量引入模型，虽然可以提高模型的可解释性，但会增加模型的复杂度，而且有些变量对模型没有影响或者影响很小，可以忽略不计。为此，我们需要进行变量选择，将无关变量剔除掉，这不仅可以降低数据维数，而且可以提高预测准确性、鲁棒性。因此在使用机器学习的方法对数据进行建模前，需对数据进行变量选择，以剔除无关、冗余变量。本文分别使用基于惩罚的 LASSO、SCAD 和 MCP 三种方法和随机森林特征重要度的方法进行变量选择。

(1) 惩罚方法：

惩罚的思想是通过将无关、冗余变量的系数缩小到零来剔除无关变量。本文使用了 LASSO、SCAD 和 MCP 三种方法，其中 LASSO (Least absolute shrinkage and selection operator，最小绝对收缩和选择算法) 方法通过引入 L1 范数进行惩罚，限制回归系数绝对值之和小于某个固定正值，将无关变量系数压缩为 0，从而剔除无关、冗余变量，LASSO 方法会导致一些显著变量的系数被压缩，使得到的惩罚项系数估计结果是有偏的全局最优解，不具有 Oracle 性质（“Oracle”表示在数值模拟的过程中真实模型已知）。

SCAD (Smoothly Clipped Absolute Deviation，平滑削边绝对偏离) 方法的惩罚力度能够随变量系数的估计量变化而变化，通过迭代的方式将较小系数压缩为 0，最终得到无偏结果，具有 Oracle 性质，但该方法得到的结果可能是局部最优解，且得到的系数结果不稳定。

MCP (Minimax Concave Penalty，极大极小凹点惩罚) 方法通过对 SCAD 函数增加了凹度限制条件，优化了 SCAD 方法系数结果不稳定和陷入局部的问题，同时该方法还可以应用于特征数大于样本数的场景。

惩罚方法变量选择的表达式如下所示：

$$\underset{\theta}{\operatorname{argmin}}\{l(\theta) + P_{\lambda}(|\theta|)\} \quad (2-2)$$

其中 $l(\theta)$ 为损失函数， $P_{\lambda}(|\theta|)$ 为惩罚函数， $\lambda > 0$ 是截断参数，本文选用平方损失函数，形式如下：

$$l(\theta) = \sum_{i=1}^n (Y_i - X_i^T \theta)^2 \quad (2-3)$$

其中 Y_i 为一维随机变量, $X_i = (X_{i1}, X_{i2}, X_{i3} \dots X_{ip})^T$ 是 p 维随机变量, $\theta = (\theta_1, \theta_2, \theta_3 \dots \theta_p)^T$ 变量系数。

LASSO、SCAD、MCP 方法的差异主要体现在惩罚函数的不同, 三种方法的惩罚函数表达式如下所示^[103]:

LASSO:

$$p_\lambda(\theta) = \lambda|\theta| \quad (2-4)$$

SCAD:

$$p_{\lambda,\gamma}(\theta) = \begin{cases} \frac{\lambda\theta}{\gamma\lambda\theta-0.5(\theta^2+\lambda^2)} & \theta \leq \lambda \\ \frac{\gamma-1}{\lambda} & \lambda < \theta \leq \gamma\lambda \\ \frac{\lambda^2(\gamma^2-1)}{2(\gamma-1)} & \theta > \gamma\lambda \end{cases} \quad (2-5)$$

MCP:

$$p_{\lambda,\gamma}(\theta) = \begin{cases} \lambda\theta - \frac{\theta^2}{2\gamma} & \theta \leq \gamma\lambda \\ \frac{1}{2}\gamma\lambda^2 & \theta > \gamma\lambda \end{cases} \quad (2-6)$$

其中 SCAD 要求 $\gamma > 2$, MCP 要求 $\gamma > 1$ 。

可通过 AIC、BIC 准则或交叉核实等方法来选取截断参数(Tuning Parameter) $\lambda > 0$, 本文选用 10 折交叉验证方法(见 2.3.10 机器学习模型的参数优化方法)选取截断参数, 惩罚变量筛选使用 R 3.6 “ncvreg” 数据包完成。

(2) 随机森林特征重要度 (Random Forest Feature Importance, RFFI) :

随机森林特征重要度方法能够忽略重要性低的特征, 并且使用特征变量的重要性进行排序, 将对反演结果影响较大的变量筛选出来, 它是通过计算特征变量在决策树中的基尼指数 (Gini) 平均减少值得到的, 基尼指数是指随机选中的样本被分错的概率, 其函数表达式为:

$$GI_m = \sum_{k=1}^{|K|} p_k(1-p_k) = 1 - \sum_{k=1}^{|K|} p_k^2 \quad (2-7)$$

其中 GI_m 表示基尼指数, K 表示一共有 K 个类别, p_k 表示样本属于 k 类的概率。

2.3.9 机器学习建模方法

本文选用支持向量回归和随机森林回归方法^[104]对雪深进行反演回归建模。

(1) 支持向量回归 (Support Vector Regression, SVR) :

支持向量机回归是在支持向量机的基础上发展而来的，它的基本思想是在n维空间中存在一个超平面能够将输入的所有数据划分为两类，当数据在原有空间无法进行划分时，通过分线性映射函数，将原有样本映射到高维平面，然后通过最大化间隔，找到一个分割超平面，使尽可能多的样本点落入间隔内。具体形式为：

$$\min_{\omega} \frac{1}{2} \|\omega\|^2, \quad (2-8)$$

$$s.t. |y_i - \omega^T x_i - b| \leq \epsilon. \quad (2-9)$$

但是在实际问题中，往往很难直接确定准确的 ϵ ，为了确保大部分数据都能在间隔带内，为每个样本点引入松弛变量 ξ_i 或 $\hat{\xi}_i$ ，从而使函数的间隔要求变的放松，也就是允许一些样本可以不在间隔带内。令 $f(x_i) = \omega^T x_i + b$ ，支持向量回归函数形式为：

$$\min_{\omega} \left[\min_{\omega} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \hat{\xi}_i) \right] \quad (2-10)$$

$$s.t. \begin{cases} f(x_i) - y_i \leq \epsilon + \xi_i, \\ y_i - f(x_i) \leq \epsilon + \hat{\xi}_i, \\ \xi_i, \hat{\xi}_i \geq 0 \end{cases} \quad (2-11)$$

使用二次规划算法求解上述优化问题，其拉格朗日函数为：

$$\begin{aligned} L(\omega, b, \epsilon, \hat{\epsilon}, \mu, \hat{\mu}, \alpha, \hat{\alpha}) = & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\xi_i + \hat{\xi}_i) - C \sum_{i=1}^N (\mu_i \xi_i + \hat{\mu}_i \hat{\xi}_i) \quad (2-12) \\ & + \sum_{i=1}^N \alpha_i (y_i - \omega^T x_i - b - \epsilon - \xi_i) + \sum_{i=1}^N \hat{\alpha}_i (\omega^T x_i + b - y_i - \epsilon - \hat{\xi}_i) \end{aligned}$$

其中 $\mu, \hat{\mu}, \alpha, \hat{\alpha} \geq 0$ 。引入核函数 $K(\cdot, \cdot)$ ，最终求得函数表达式：

$$f(x) = (\alpha_i - \hat{\alpha}_i) K(x_i, x) + b \quad (2-13)$$

式中： $\alpha_i, \hat{\alpha}_i$ 为拉格朗日乘子。

该方法拟合结果为全局最优解，同时支持向量机算法对积雪属性变化敏感^[105]，本实验选用的核函数为径向基核函数 (Radial Basis Function Kernel)，支持向量回归借助 Python 的 Sklearn 包完成。

(2) 随机森林回归 (Random Forest Regression, RFR) :

随机森林 (Random Forest, RF) 算法是基于多个决策树的监督学习算法, 通过 Bootstrap 重采样方法抽取训练集建立决策树, 在每个决策树上随机选择多个特征建模, 模型回归值取所有决策树的平均值, 因此随机森林具有不易出现过拟合、受噪声数据干扰小等特点^[53], 其函数为:

$$\bar{h}(x) = \frac{1}{T} \sum_{t=1}^T \{h(x, \theta_t)\} \quad (2-14)$$

其中 $\bar{h}(x)$ 为模型预测结果, $h(x, \theta_t)$ 为基于 x 和 θ_t 的输出, x 为自变量, θ_t 为独立同分布随机向量, T 为回归决策树的个数。

2.3.10 机器学习模型的参数优化方法

K 折交叉验证是机器学习中算法参数择优、模型评估的常见方法, 其思想是将原始数据分成 K 组, 不重复的抽取其中一组作为测试组, 将其余 K-1 组作为训练组来拟合模型, 将测试组数据作为真值, 对模型表现进行评分, 此过程重复 K 次, 最终输出结果取 K 次平均, 使用 K 折交叉验证能够减少评估结果的随机性, 增强结果的说服力, 本文折数 K 选为 10。

2.3.11 评价指标

为了更好描述台站代表性、再分析雪深数据集、被动微波雪深反演算法和机器学习雪深反演模型的表现, 使用了相关系数 (R)、偏差 (Bias)、均方根误差 (RMSE)、平均绝对误差 (MAE) 和决定系数 (R^2) 等指标进行评价, 各指标介绍如下:

偏差 (Bias) 指反演雪深与观测雪深差值的平均值, 用来评价雪深数据整体的准确性, 偏差的绝对值越小, 表明反演雪深整体与实际雪深越接近, 反演算法表现越好, 偏差为正, 表示反演雪深整体高估实际雪深; 反之, 表示低估。表达式为:

$$Bias = \frac{1}{n} \sum_{i=1}^n (x_i - y_i) \quad (2-15)$$

均方根误差 (Root Mean Square Error, RMSE), 指雪深数据与站点雪深观测值之间差值平方的平均值的平方根, RMSE 越小, 表示反演雪深越接近实际站点雪深, 且 RMSE 对偏差较大数据更敏感, 其公式为:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (2-16)$$

相关系数 (Correlation coefficient, R) 用来评估两变量之间的相关性, R 值范围为 [-1, 1], 正值表示两变量正相关, 负值表示负相关, 0 值表示不相关, 其公式为:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2-17)$$

上述公式中 x_i 、 y_i 分别代表两变量 X、Y 集合中的第 i 个数据, \bar{x} 、 \bar{y} 分别代表两变量集合的平均值, n 表示数据样本数。

平均绝对误差 (Mean Absolute Error, MAE) 指标是指数据集合中模拟雪深雪深与观测雪深偏差绝对值的平均值, MAE 大于 0, 越小表示, 模型反演雪深的表现越好, 其函数计算公式为:

$$MAE = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n} \quad (2-18)$$

决定系数 (Determination coefficient, R^2) 可用于评估模型拟合效果, 其中决定系数越高表示自变量对因变量的解释程度越高, 其取值范围为 [0 - 1] 之间, 其函数表达式为:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2-19)$$

式中 \hat{y}_i 为第 i 个预测值, y_i 为观测雪深数据中第 i 个值, \bar{y} 为观测雪深的平均值。

第三章 再分析雪深数据集与被动微波反演算法评估

3.1 站点代表性分析

地面台站实测雪深数据是检验被动微波遥感反演雪深与再分析雪深数据集的重要参考资料,由于本文所使用数据集的空间分辨率存在较大差异,本文在开展相关实验前,对站点代表性展开讨论,以选择对区域积雪场有较好代表性的台站雪深数据进行实验,减少本文结果的不确定性。

本文通过分析站点与数据集对应区域的积雪覆盖率(Snow Cover Fraction, SCF)、海拔和土地覆被类型特征来探究台站的代表性。

3.1.1 台站雪深与积雪覆盖率分析

通过对台站观测雪深数据与四种再分析数据集和 AMSR2 亮温数据集对应区域的积雪覆盖率进行相关分析,得到了各数据集 SCF 与台站实测雪深间的相关系数表(见表 3-1)。

在四种再分析雪深数据集中, JRA55 数据集的积雪覆盖率与台站观测雪深数据整体的相关性最好(R 值为 0.54), ERA5 数据集次之(R 值为 0.52), GLDAS、MERRA2 数据集的相关性最弱, R 值分别为 0.49、0.42。AMSR2 亮温数据集的 SCF 与站点观测雪深的整体相关系数为 0.56, 高于四种再分析数据集, 其中与纳木错、色林错站观测雪深的相关系数最高, 均大于 0.6, 与申扎站相关系数最低, 为 0.06。

表 3-1 数据集积雪覆盖率与站点观测雪深的相关系数

Table 3-1 Correlation coefficient between SCF of datasets and observed snow depth

数据集	台站名称							
	纳木错	德庆	色林错	门当	申扎	当雄	班戈	所有站点
ERA5	0.65**	0.62**	0.36*	0.38**	0.02	0.44*	0.51**	0.52**
GLDAS	0.61**	0.38**	0.57**	0.43**	0.01	0.32**	0.37**	0.49**
JRA55	0.72**	0.35**	0.55**	0.55**	0.08*	0.72**	0.30**	0.54**
MERRA2	0.69**	0.32**	0.44**	0.35**	0.02	0.69**	0.32**	0.42**
AMSR2	0.68**	0.47**	0.63**	0.38**	0.06*	0.29*	0.41**	0.56**

**表示 $P < 0.01$, *表示 $P < 0.05$ 。

在七个台站中, 五种数据集 SCF 与纳木错站点观测雪深的相关系数高于其

余六个站点，均大于 0.6 且通过 0.01 显著性检验，其中 JRA55 数据集 SCF 的相关系数最大，为 0.72。除申扎站外的六个站点雪深数据与四种再分析数据集 SCF 的相关系数均超过了 0.3。申扎站与五种数据集 SCF 的相关系数均低于 0.1，且仅 JRA55、AMSR2 通过 0.05 显著性检验，表明申扎站对数据集对应区域的代表性较弱，这主要由于申扎站观测仪器出现故障，使得该站点观测到的冬季雪深数据偏少，导致观测雪深与各数据集积雪覆盖率的相关系数均较低。

3.1.2 海拔分析

海拔对于积雪的形成和分布具有重要影响，通常随着海拔升高，雪深呈现增加趋势，但雪深最大处并非为海拔最高处^[106]，这是由于随着海拔增加，气温降低，上升水汽更易形成降雪，而当海拔超过某一高度后继续上升，此时大气水汽含量较低，不易产生降雪。在研究区七个雪深观测台站中，当雄站海拔最低，为 4200 m，其余台站海拔在 4600 ~ 5000 m 之间。各数据集对应平均海拔与站点海拔的海拔差在当雄站最大（见图 3-1），除 ERA5 数据集外，其余四个数据集海拔差均大于 200 m，其中 JRA55 数据集的海拔差最大，超过 600 m。

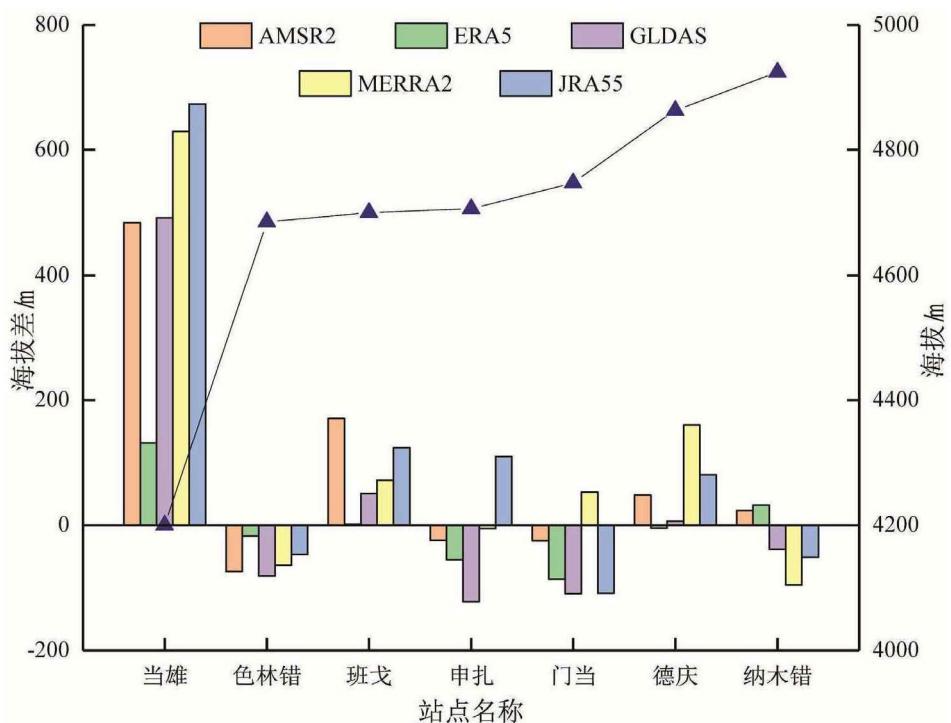


图 3-1 数据集平均海拔与台站海拔
Figure 3-1 Elevation of datasets and meteorological stations

这主要由于当雄站位于山谷城镇地区，与周围山体距离较近，数据集栅格范围中山体占较大，导致数据集对应区域的平均海拔比台站海拔偏大。纳木错站、德庆站、色林错站、门当站、申扎站、班戈站海拔差均较小，部分数据集对应区

域的平均海拔略低于台站海拔。在五种数据集中, ERA5 数据集的海拔差均较小, 这与 ERA5 数据集空间分辨率最高有关, 其空间分辨率约为 $10 \text{ km} \times 10 \text{ km}$, 受区域地形影响相对较小。本文选择将海拔差超过 200 m 的对应雪深观测数据剔除。

3.1.3 下垫面类型分析

研究区下垫面类型以草地、裸地和水体为主, 其中草地覆被类型占比最高, 七个台站架设点的土地覆被类型均为草地。研究显示^[100, 107], 裸地与草地覆被类型对雪深影响较灌木、森林等土地覆被类型小, 而由于水体与陆地的理化性质存在较大差异, 积雪在两种环境下产生、积累和消融过程有较大不同^[100, 108], 通常再分析数据集针对水体和陆面过程使用的模型策略也不一样^[109], 并且再分析数据集雪深数据为栅格区域的平均值, 当再分析数据集范围内水体占比过高时, 数据集雪深数据的准确性将显著下降。因此本文主要考虑了水体下垫面对数据集数据代表性的影响, 将水体占比是否超过 15% 作为判断阈值, 当水体超过该阈值时, 认为再分析数据集数据受水体影响较大, 不适用于与台站观测雪深数据进行对比分析。

对 4 种再分析数据集在各站点对应区域的土地覆被类型的计算结果显示, 四种再分析数据集在色林错站的水体占比均高于 15% (见表 3-2), GLDAS、MERRA2 数据集在申扎站的水体占比超过 15% , JRA55 数据集在门当站的水体占比超过 15% 。

当前学者在利用被动微波遥感进行雪深反演时常剔除湖泊周围数据^[65, 81], 这是由于水体介电常数较大, 能够引起被动微波亮温偏低, 导致算法反演雪深出现高估。当冬季湖泊处于结冰期时^[110], 不同于陆面下垫面, 湖冰上的被动微波雪深反演还需要考虑湖面气温、风力等气候因子^[108]。通过对比前人研究, 发现模型模拟水体结冰导致降低的 19 GHz 极化亮温差^[111]、观测到的冰层导致 $17/38 \text{ GHz}$ 亮温变化^[112]、以及模拟的积雪体散射所引起的亮温变化值均为同一数量级变化^[98], 而 AMSR2 亮温数据集在各站点的水体占比均小于 10% (见表 3-2), 因此可以认为水体引起的被动微波反演雪深高估影响在可接受范围内, 对本文评估结果影响较小。

表 3-2 五种数据集在各站点对应区域的土地覆被类型

Table 3-2 The land cover types of the five datasets in the meteorological stations

站点	草地/ %					裸地/ %					水体/ %					
	ERA5	GLDAS MERRA2	JRA55	AMSR2	ERA5	GLDAS MERRA2	JRA55	AMSR2	ERA5	GLDAS MERRA2	JRA55	AMSR2	ERA5	GLDAS MERRA2	JRA55	AMSR2
纳木错站	91.75	85.76	89.11	80.38	81.96	8.25	14.24	10.89	13.01	18.04	0	0	0	0	6.61	0
德庆站	99.74	99.96	93.21	98.51	100	0.26	0.04	5.67	1.32	0	0	0	0	1.12	0.17	0
色林错站	54.25	43.51	54.31	60.4	54.07	22.25	36.9	29.8	24.16	42.53	23.5	19.59	15.89	15.45	3.4	
门当站	100	79.92	78.09	60.4	81.87	0	11.84	8.91	24.16	10.5	0	8.24	13	15.45	7.63	
申扎站	97.75	35.8	46.14	83.18	88.21	2.25	20.16	21.3	8.92	10.36	0	44.04	32.53	7.93	1.43	
当雄站	99	98.69	93.21	80.38	98.13	1	1.31	5.67	13.01	1.87	0	0	1.12	6.61	0	
班戈站	99.75	99.51	99.54	94.01	99.83	0.25	0.49	0.46	2.97	0.17	0	0	0	3.02	0	

3.1.4 站点代表性讨论

对各数据集在不同站点对应区域的积雪覆盖率、平均海拔和下垫面类型特征的分析结果表明, 申扎站与五种数据集对应区域 SCF 的相关性均较低, 说明申扎站对五种数据集对应区域的积雪场的代表性较弱。五种数据集对应区域的平均海拔与台站海拔的差值在当雄站最大, 表明当雄站对区域代表性较弱。四种再分析数据集对应区域在色林错站的水体占比均超过 15 %, GLDAS、MERRA2 数据集对应区域在申扎站和 JRA55 数据集对应区域在门当站的水体占比也超过 15 %。

为了减少评估结果的不确定性, 在对再分析数据集和被动微波雪深反演算法进行评估时, 需将五种数据集中积雪覆盖率、平均海拔和下垫面类型表现不佳的站点数据和数据集数据进行剔除, 最终数据筛选结果为: 在对四种再分析雪深数据集评估时均不使用色林错站、当雄站数据, 对 GLDAS、MERRA2、JRA55 数据集评估时, 分别不使用当雄站、门当站和当雄站、当雄站数据, 使用 AMSR2 被动微波亮温数据时不使用申扎站、当雄站数据。

3.2 再分析雪深数据集评估

3.2.1 再分析雪深数据集时间序列特征分析

选用纳木错、德庆、门当、当雄和班戈五个站点两个积雪年的雪深数据, 对 ERA5、JRA55、GLDAS、MERRA2 再分析雪深数据集在研究区的表现进行评估, 结果显示(见图 3-2), ERA5、GLDAS、JRA55 数据集雪深在积雪积累期、消融期的变化趋势与地面站点观测雪深接近, 四种再分析数据在连续积雪年均表现出相似的年际波动。

ERA5 数据集雪深峰值在纳木错、当雄、班戈站大于观测雪深, 在德庆、门当站与观测雪深接近。ERA5 数据集雪深峰值出现在 3 ~ 5 月, 而台站观测雪深峰值出现在 1 ~ 3 月, ERA5 再分析数据集雪深峰值存在滞后, 其中在纳木错站 2019 ~ 2020 年积雪年滞后时间最久, 实际雪深峰值出现在 2019 年 12 月至 2020 年 1 月, 而 ERA5 数据集雪深峰值出现在 2020 年 4 月。GLDAS 数据集雪深冬季在纳木错站的积雪深度范围为 5 ~ 15 cm, 低于实际 15 ~ 30 cm 积雪深度, GLDAS 数据集的雪深峰值在德庆、门当、班戈站与观测峰值接近。

JRA55 数据集雪深均较小, 整体为低于 5 cm 的浅雪, 对各站点雪深表现出低估, 其中在纳木错站的低估最大。同时, 相对 ERA5、GLDAS 数据集雪深, JRA55 数据集雪深表现出较大雪深波动。MERRA2 数据集雪深在冬季雪深固定为 17 cm, 表现出平峰特征, 对冬季积雪的积累、消融过程表现不佳, 而夏季雪深波动频率和范围在四种再分析数据集中最大。

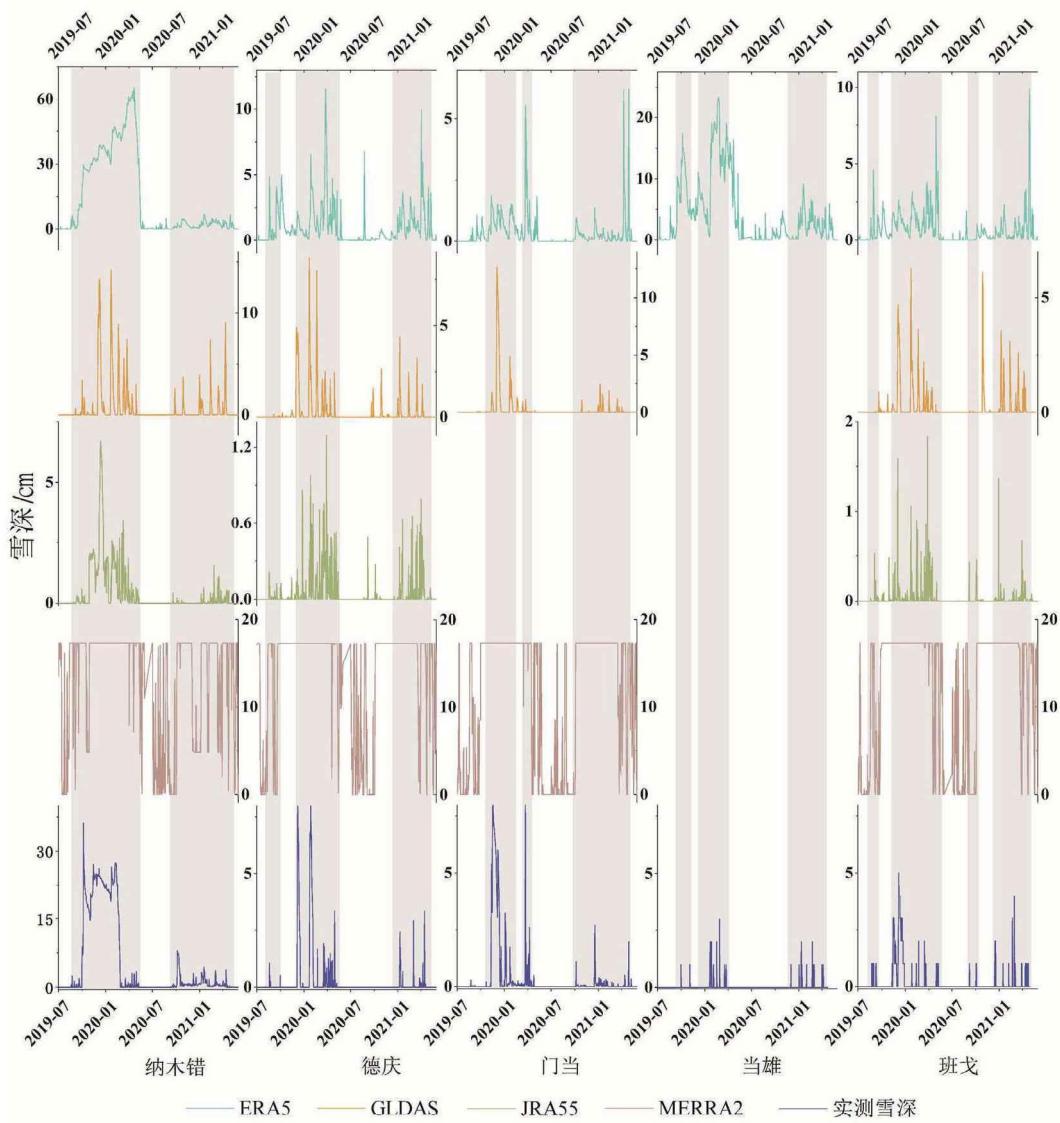


图 3-2 再分析数据集与站点观测的时间序列雪深

Figure 3-2 Time series snow depth of reanalysis datasets and meteorological stations

对站点及再分析数据集的月平均积雪天数统计发现（见表 3-3），研究区气象站点观测的月平均积雪天数为 6.59 d，均小于四个再分析数据集的月平均积雪天数。JRA55 的月平均月积雪天数为 10.33 d，GLDAS 和 ERA5 月平均积雪天数较为接近，分别为 9.43、9.66 d，MERRA2 再分析数据集的月平均积雪天数最大为 25.31 d，与地面观测相差最大。

表 3-3 再分析雪深资料与台站观测的月平均积雪天数

Table 3-3 Monthly average snow cover days for stations and reanalysis datasets

	野外站点	ERA5	GLDAS	JRA55	MERRA2
天数/d	6.59	9.66	9.43	10.33	25.31

基于 R、Bias、RMSE 评价指标，对四种再分析雪深数据集进行评估，结果显示（见表 3-4），四种再分析雪深数据集中，ERA5 数据集雪深与观测雪深的相关系数最高，为 0.62，MERRA2 数据集雪深与观测雪深的相关系数最小，R 值仅 0.14。MERRA2 雪深数据集的均方根误差最大（13.92 cm），GLDAS 和 JRA55 雪深数据集的均方根误差最小，分别为 4.19、4.06 cm。

ERA5 和 MERRA2 数据集雪深对研究区雪深整体高估，其中 MERRA2 数据集的高估值最大（11.71 cm），而 GLDAS 和 JRA55 数据集雪深对研究区雪深整体低估，两数据集的偏差在四种数据集中最小，分别为 -0.68、-0.91 cm。

表 3-4 再分析雪深资料表现

Table 3-4 The performance of reanalysis datasets

指标	ERA5	GLDAS	JRA55	MERRA2
R	0.62**	0.23**	0.41**	0.14**
Bias/cm	2.37	-0.68	-0.91	11.71
RMSE/cm	7.78	4.19	4.06	13.92

**表示 $P < 0.01$ ，*表示 $P < 0.05$ 。

3.2.2 数据集在不同积雪深度的表现

从再分析雪深数据集的雪深散点图可以看出（见图 3-3），四种再分析数据集的散点分布呈现出不同特征，ERA5、GLDAS 和 JRA55 数据集雪深散点在图像原点附近分布最集中。GLDAS、JRA55 和 MERRA2 数据集的雪深散点呈现出平行横向坐标轴的分布规律，GLDAS 和 JRA55 数据集雪深散点多分布于“再分析数据集雪深=0 cm”，MERRA2 数据集雪深散点在“再分析数据集雪深=17 cm”处分布最集中。这表明 GLDAS 和 JRA55 再分析雪深数据集对地面积雪的识别能力较弱，易出现“无雪”误判，而 MERRA2 数据集对地面积雪出现“中雪”误判频率较高。

ERA5 雪深数据集散点分布具有一定的线性趋势，而其余三种再分析数据集的雪深散点的线性分布特征不显著。在 0~5 cm 的观测雪深区间，ERA5 数据集雪深散点距 1:1 趋势线的距离在四种数据集中最大，最大雪深在 60 cm 左右。

以 R、Bias 和 RMSE 为评价指标，对四种再分析数据集在浅、中和深雪的表现进行评估。结果显示（见表 3-5），四个再分析数据集在不同观测雪深的表现存在明显差异，在 0~5 cm 的浅雪观测雪深区间，四种雪深数据集与台站雪深的相关系数均小于 0.20，与观测雪深的相关性均较弱，其中 GLDAS 数据集雪深与台站观测雪深的相关系数最大，为 0.18。

在 5 ~ 20 cm 的中雪观测雪深区间, ERA5 数据集雪深与台站观测雪深的相关系数最高 (0.80), MERRA2 数据集雪深与台站观测雪深的相关系数为-0.61, 是唯一与实测雪深表现出负相关的数据集, GLDAS 数据集雪深与实测雪深的偏差在四种雪深再分析数据集中最小。

在大于 20 cm 的深雪观测雪深区间, ERA5、JRA55、GLDAS 雪深数据集的偏差绝对值和均方根误差均大于各数据集在浅、中雪观测雪深区间的偏差绝对值和均方根误差, 而 MERRA2 雪深数据集的偏差绝对值和均方根误差表现与之相反, 与其 17 cm 雪深数据占比较多有关。

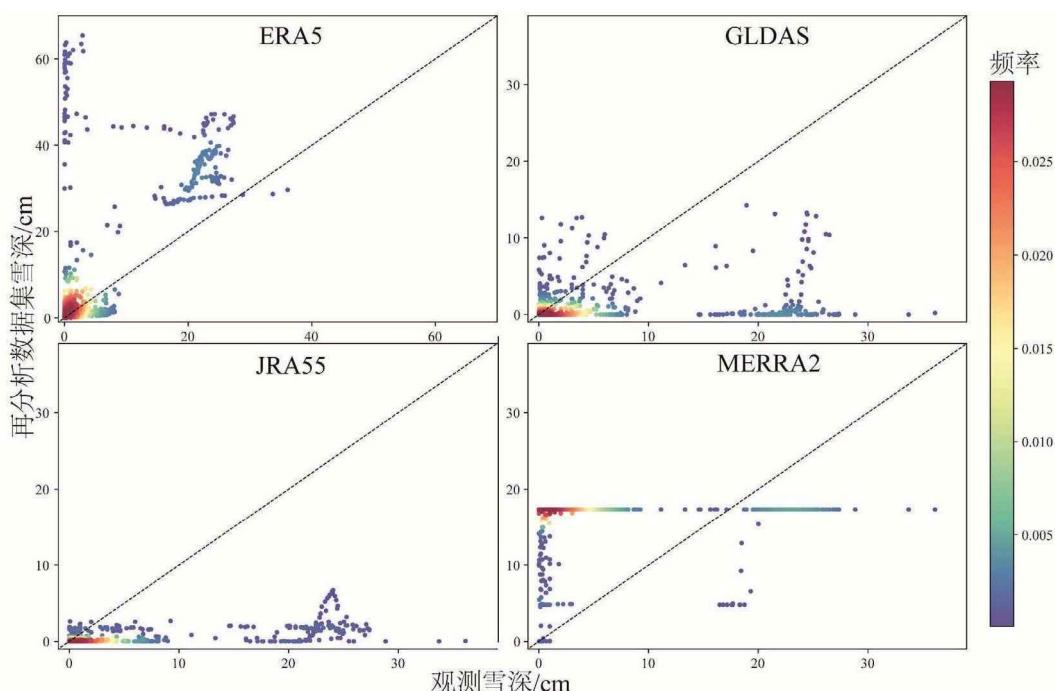


图 3-3 再分析数据集的雪深散点图

Figure 3-3 Snow depth scatterplot for the reanalysis datasets

表 3-5 再分析雪深数据在不同深度表现

Table 3-5 The performance of the reanalysis datasets at different snow depths

雪深/cm	R	Bias/cm	RMSE/cm	R	Bias/cm	RMSE/cm
		ERA5			JRA55	
> 20	0.15	12.11	13.21	0.08	-20.90	21.15
5 ~ 20	0.80**	2.54	8.32	0.47*	-0.92	4.39
< 5	0.09**	2.32	7.81	0.16**	-0.81	4.07
		GLDAS			MERRA2	
> 20	0.16	-20.84	21.28	0.13	-6.17	6.63
5 ~ 20	0.01	-0.75	4.53	-0.61**	11.46	13.80
< 5	0.18**	-0.63	4.21	0.15**	11.68	13.92

**表示 $P < 0.01$, *表示 $P < 0.05$ 。

3.2.3 结果讨论

在四种再分析雪深数据集中, ERA5 雪深数据集的整体表现最好, 与台站观测雪深的整体相关性最好, 雪深变化趋势与地面观测雪深相似, 说明 ERA5 再分析雪深数据集在研究区具有较好的应用潜力。但 ERA5 再分析数据集雪深对研究区地面观测雪深整体高估, 其中对于浅雪观测雪深的偏差值能达到 60 cm, 这与其在纳木错站冬季雪深峰值表现出的滞后现象有关, 因此在实际雪深应用中, 需要考虑 ERA5 再分析数据集高值雪深数据可能存在的较大偏差。

JRA55 和 GLDAS 再分析雪深数据集在研究区表现不佳, 数据集雪深散点呈现出贴近横向坐标轴的分布趋势, 在不同观测雪深区间的表现在均不佳。两再分析雪深数据集对研究区地面雪深存在低估, 其中雪深值为 0 cm 的占比较多, 说明两再分析数据集所使用的物理模型对研究区地表积雪的识别能力较弱, 在实际雪深应用中将给研究结果增加较大不确定性。

MERRA2 雪深数据集在研究区的表现最不理想, 这是因为该数据集雪深数据多为 17 cm 的固定值, 且在夏季地面无雪时期又出现多次降雪。而 MERRA2 数据集雪深为 17 cm 固定值与其使用的模型参数有关^[113, 114], 其使用的物理模型将每个栅格分为多个瓦片 (Tiles), 每个瓦片的雪压默认值为 26 kg/m², 单个瓦片的积雪量是固定值, 只有当单次降雪事件的降雪量超过单个瓦片的积雪量时, 此次降雪事件才会被模型记录, 而当栅格中所有瓦片状态均为有雪时, 此时单个栅格的积雪覆盖率为 100 %, 雪深为 17 cm, 而当积雪继续增加时, 栅格雪深并不会增加, 从而表现出雪深“饱和”现象。这种处理方法虽然可以减少薄雪对模型结果的影响并降低模型复杂度, 但导致了 MERRA2 雪深数据集在研究区的适用性较差。

3.3 被动微波雪深反演算法评估

本文选择了 Chang2 算法、SPD 算法、Che 算法、Jiang 算法和 AMSR2 算法等五种典型被动微波雪深反演算法进行评估, 算法公式见表 3-6(详细参数见 1.2.5 主被动微波遥感)。AMSR2 算法将瞬时视场中森林区和非森林区分别计算, 而本文研究区土地覆被以草地为主, 无森林覆被类型, 因此本文 AMSR2 算法选用其非森林区算法, 五种算法中的 18 (19) GHz 的亮温数据在本文中统一使用 AMSR2 的 18.7 GHz 亮温数据。

表 3-6 被动微波雪深反演算法公式
Table 3-6 Formula of passive microwave snow depth retrieval algorithms

算法名称	算法公式	公式序号	参考文献
Chang2 算法	$SD = 2.0 \times (Tb_{18H} - Tb_{37H}) - 8$	(1 - 10)	[76]
SPD 算法	$SPD = (Tb_{18V} - Tb_{37V}) + (Tb_{18V} - Tb_{18H})$	(1 - 2)	[71]
	$SD = 0.68 \times SPD + 0.67$		
Che 算法	$SD = 0.66 \times (Tb_{19H} - Tb_{37H}) + b$	(1 - 12)	[78]
Jiang 算法	$SD = f_{farmland} \times SD_{farmland} + f_{grass} \times D_{grass} + f_{baresoil} \times SD_{baresoil} + f_{forest} \times D_{forest}$	(1 - 13)	[79]
AMSR2 算法	$SD = f * (SD_f) + (1 - f) * (SD_0)$	(1 - 5)	[74]

3.3.1 算法反演雪深整体表现

基于 AMSR2 亮温数据及土地覆被类型数据, 使用 Chang2 算法、SPD 算法、Che 算法、Jiang 算法和 AMSR2 算法等五种典型算法对研究区雪深进行反演, 得到了五种算法的反演雪深。从反演雪深散点图可以看出(见图 3-4), 深雪雪深均发生于纳木错站, 德庆、色林错、门当、班戈站的雪深类型为浅雪和中雪。

五种算法反演雪深散点在德庆、色林错、门当、班戈站未呈现出显著线性关系, 散点分布趋势平行于纵向坐标轴, 其中 Chang2 算法反演雪深散点多分布于 1:1 趋势线下部, SPD 算法反演雪深散点全部位于 1:1 趋势线上部, Che 算法、AMSR2 算法和 Jiang 算法反演雪深散点在 1:1 趋势线上下部均有分布。五种算法反演雪深散点在纳木错站多分布于 1:1 趋势线附近, 表现出一定线性特征。

五种算法反演的时间序列雪深数据显示(见图 3-5), 五种算法反演的雪深在两个积雪年内均表现出相同变化趋势, 即呈先上升后下降趋势, 五种算法反演雪深的峰值均出现在地面存在积雪时期(图 3-5 灰色区域)。五种算法反演的雪深与站点观测雪深的年际变化趋势也较为相似, 五种算法反演雪深与地面观测雪深的年际变化在纳木错站均呈下降趋势, 而在其余四个站点, 五种算法反演雪深与观测雪深的年际变化均较小。

SPD 算法反演的雪深最小值均大于 5 cm, Che 算法反演的雪深在 -20 ~ 20 cm 范围内波动, 波动幅度在五种算法中最小。Jiang 算法反演雪深峰值与观测雪深峰值具有较好对应关系, Jiang 算法反演雪深在色林错站 2020 ~ 2021 积雪年的波动幅度较小。AMSR2 算法反演雪深与 Chang2 算法反演雪深表现出相似变化趋势, 但 AMSR2 算法反演的雪深值多高于 Chang2 算法反演的雪深值。

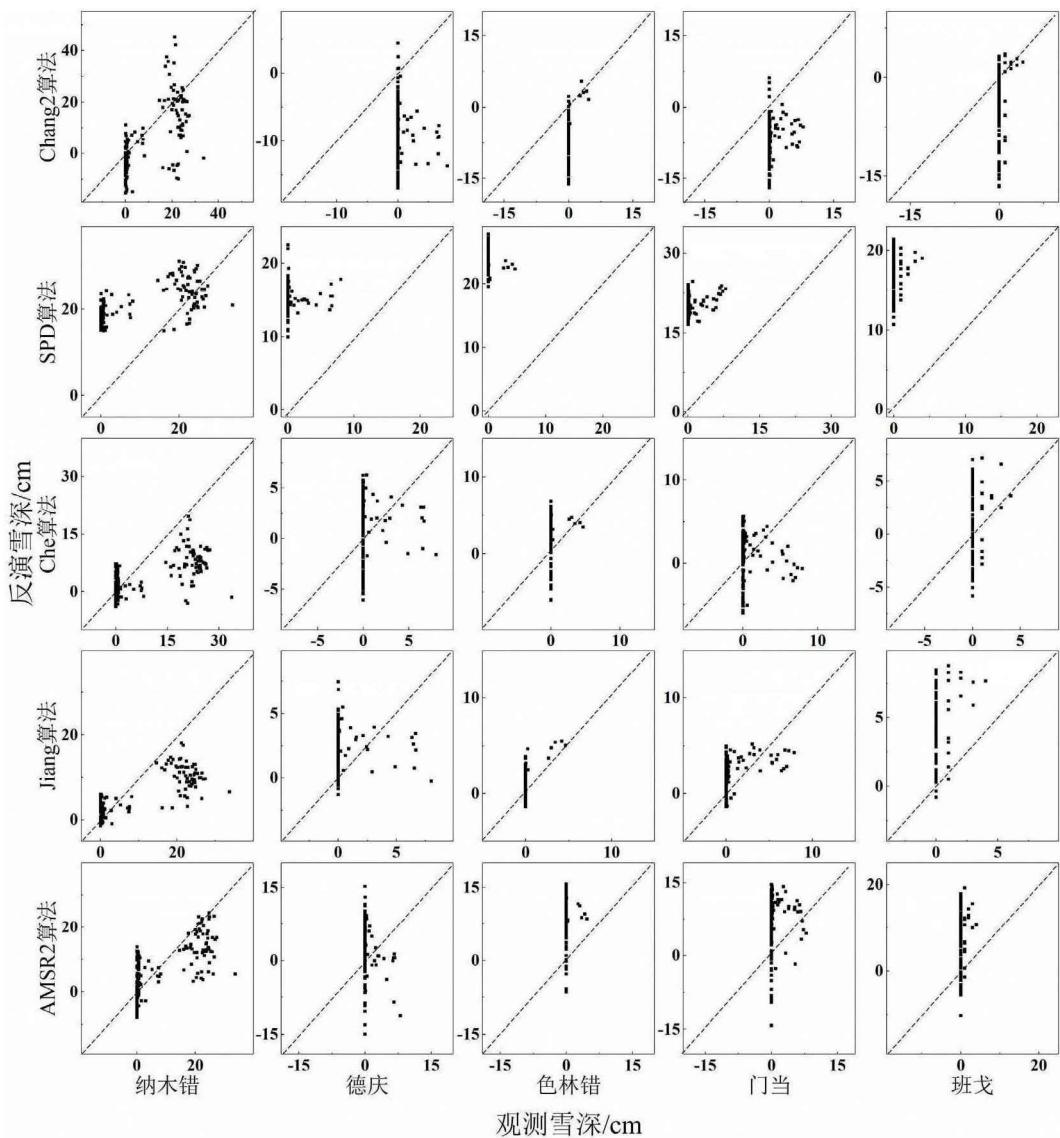


图 3-4 算法反演雪深散点图

Figure 3-4 Retrieved snow depth scatterplot for the algorithms

对五种被动微波雪深反演算法的评估结果显示, Jiang 算法反演雪深与实测雪深的整体相关性最高(R 值为 0.65, 见图 3-6), Chang2 算法和 Che 算法次之, 相关系数分别为 0.59、0.48, AMSR2 算法的相关系数为 0.45, SPD 算法的相关系数最小 (0.42), 五种算法相关系数均通过 0.01 显著性检验。

五种算法反演雪深与观测雪深的相关性存在区域差异, 在纳木错、班戈站五种算法反演雪深与实测雪深的相关性最高, 相关系数在 0.58 ~ 0.81 之间, 其中 Jiang 算法在纳木错站的相关系数值最大, Che 算法在纳木错站点的相关系数最小。五种算法在德庆站点反演的雪深与观测雪深的相关系数均小于 0.1, 且均未通过 0.05 显著性检验, 表明五种算法在德庆站点的雪深反演效果均不佳。五种

算法中仅 Jiang 算法和 Chang2 算法在色林错、门当站点同时通过 0.01 显著性检验，其中 Jiang 算法在两台站的相关系数最高，分别为 0.56、0.30。

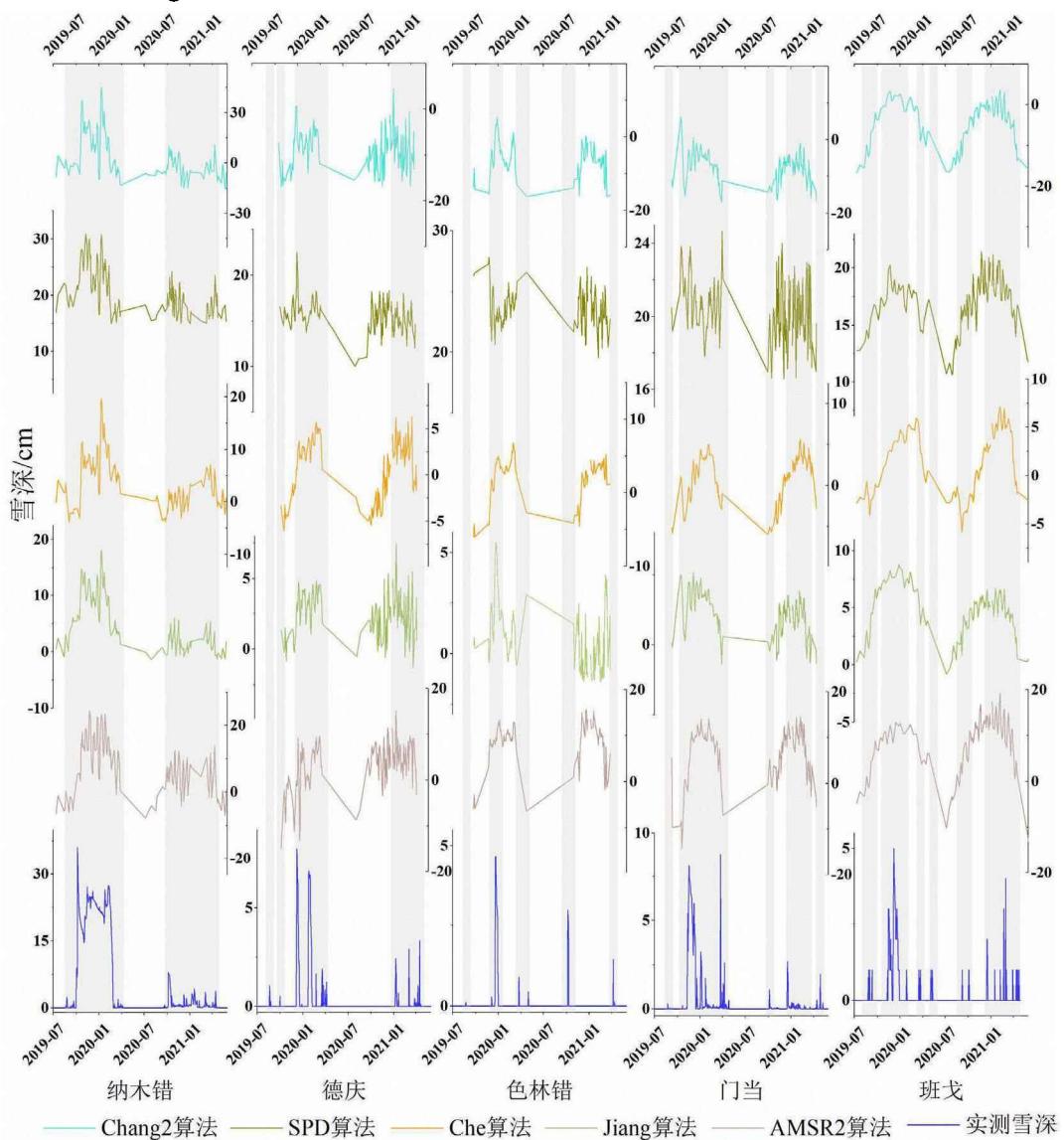


图 3-5 算法反演的时间序列雪深

Figure 3-5 Time series snow depth retrieved by five algorithms

在五种算法的均方根误差整体表现中，Che 算法和 Jiang 算法的 RMSE 值最小（见图 3-7），分别为 7.36、5.51 cm，Chang2 算法和 SPD 算法的 RMSE 值最大，分别为 17.00、17.10 cm。Jiang 算法、Che 算法和 AMSR2 算法在各站点的 RMSE 均较小 (< 10 cm)。Jiang 算法反演雪深的 RMSE 在色林错站、门当站均小于 3 cm，在纳木错站最大 (7.68 cm)。Che 算法反演雪深的 RMSE 值在德庆、门当、班戈站点要低于纳木错、色林错站，AMSR2 算法的 RMSE 值变化在五种

算法中最小。Chang2 算法和 SPD 算法的 RMSE 值在色林错站最大 (24.40、20.40 cm) , 在纳木错站最小 (12.10、13.20 cm) 。

五种算法反演雪深的偏差结果显示 (见图 3-8) , Che 算法、Jiang 算法和 A MSR2 算法的整体 Bias 均为负值, 在 -5 ~ 0 cm 范围内。AMSR2 算法反演雪深的整体偏差的绝对值最小 (-0.50 cm) ; Che 算法和 Jiang 算法反演雪深与实测雪深的整体偏差的绝对值仅次于 AMSR2 算法, Bias 分别为 -3.43、-1.49 cm; SPD 算法的偏差最大 (15.80 cm) , 同时是五种算法中唯一对观测雪深整体高估的算法。

Jiang 算法反演雪深在门当站的偏差绝对值最小 (0.41 cm) , SPD 算法反演雪深在色林错站点的偏差绝对值最大 (20.30 cm) 。在纳木错、色林错、门当、班戈站, SPD 算法偏差均最大且为正值, 在 11 ~ 20.30 cm 区间, Chang2 算法反演雪深的偏差均为负, 且偏差绝对值仅次于 SPD 算法, Bias 在 -16.00 ~ -7.42 cm 间。在德庆站点, Chang2 算法的偏差 (-14.1 cm) 绝对值高于 SPD 算法 (1 3.1 cm) 。

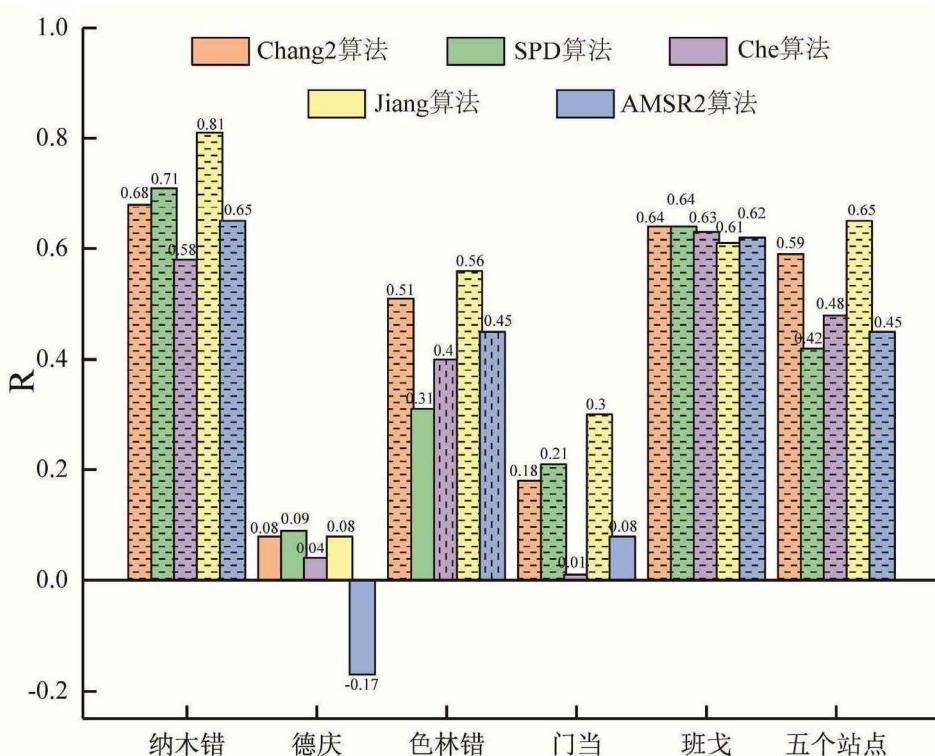


图 3-6 算法反演雪深与站点雪深的相关系数, 其中横向虚线表示 $P < 0.01$, 纵向虚线表示 $P < 0.05$, 无图案表示未通过显著性检验

Figure 3-6 Correlation coefficient between snow depth retrieved by algorithms and observed snow depth. The horizontal dashed line represents $P < 0.01$, the vertical dashed line represents $P < 0.05$, and no pattern means failed the significance test.

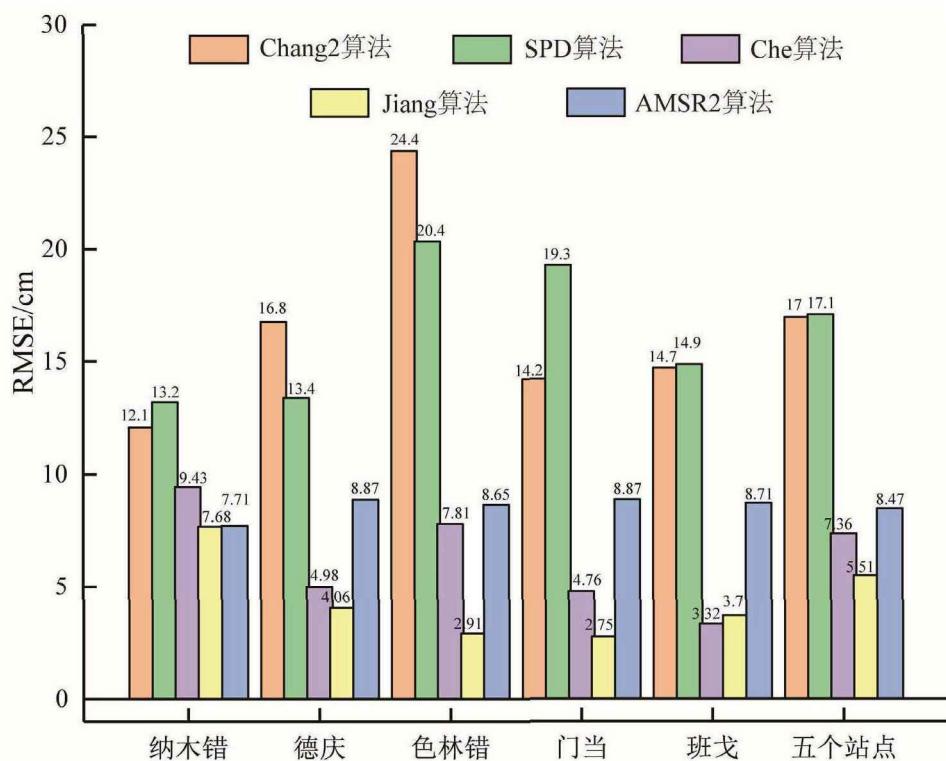


图 3-7 算法反演雪深与观测雪深的均方根误差

Figure 3-7 RMSE between snow depth retrieved by algorithms and observed snow depth

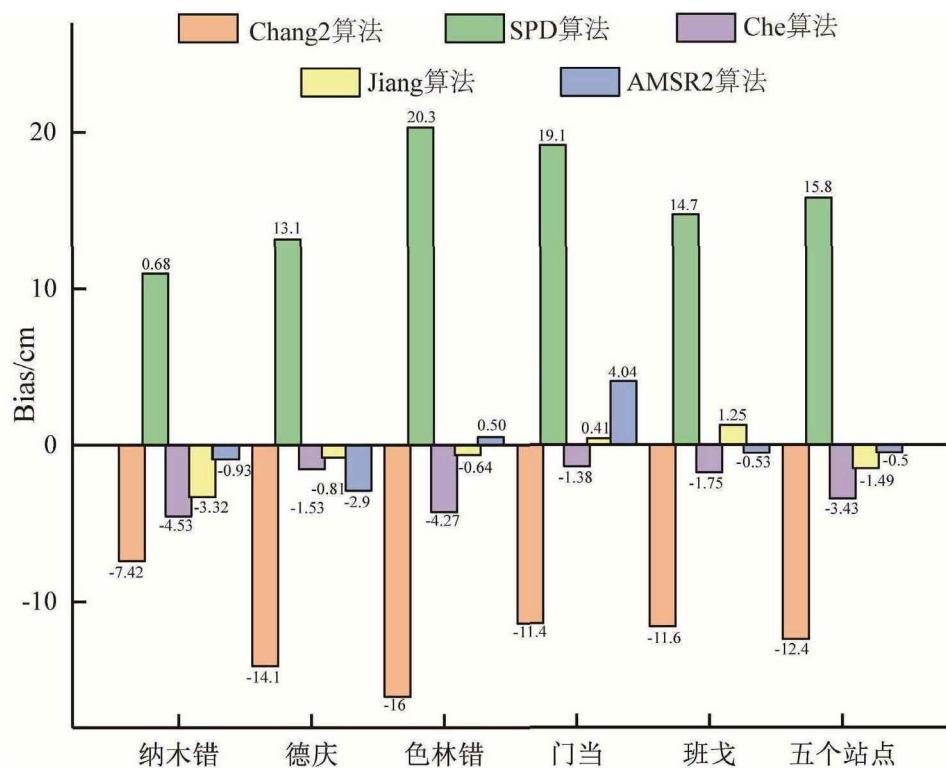


图 3-8 算法反演雪深与观测雪深的偏差

Figure 3-8 Bias between snow depth retrieved by algorithms and observed snow depth

3.3.2 算法在不同积雪深度的表现

算法反演雪深散点的分布结果显示（见图 3-4），算法对不同积雪深度的敏感度存在差异。当观测雪深为浅雪时，SPD 算法反演雪深的散点均分布于 1:1 趋势线上方，Chang2 算法反演雪深散点多分布于 1:1 趋势线下部，而其余三种算法反演雪深散点在 1:1 趋势线上下部均有分布，表明 SPD 算法对浅雪存在高估，Chang2 算法对浅雪易低估。当观测雪深为中、深雪时，Chang2 算法、SPD 算法反演雪深的散点在 1:1 趋势线上下部均有分布，而 Che 算法、Jiang 算法和 AMSR2 算法反演雪深的散点多分布于 1:1 趋势线下方，这说明 Che 算法、Jiang 算法和 AMSR2 算法对中、深雪反演易出现低估。

对五种算法在不同深度表现的评价结果显示（见表 3-7），五种算法对深雪的反演雪深与观测深雪的相关系数均为负值，且仅 Che 算法和 Jiang 算法的相关性经过了 0.05 的显著性检验，表明五种算法反演雪深与观测雪深负相关，算法对深雪的反演表现均较差。

五种算法对于中雪雪深的反演结果均较好，相关系数在 0.52~0.78 之间，且均通过了 0.01 显著性检验，其中 Jiang 算法、Che 算法和 Chang2 算法的相关系数均大于 0.7。SPD 算法反演雪深的偏差绝对值最大（10.63 cm），AMSR2 算法和 Jiang 算法反演雪深的偏差绝对值和均方根误差均最小，分别为 4.16 cm 和 7.06 cm、4.91 cm 和 5.90 cm，Chang2 算法反演雪深的 RMSE 值最大（14.47 cm）。

表 3-7 算法在不同雪深下的表现

Table 3-7 Performance of algorithms at different snow depths

雪深/cm	R	Bias/cm	RMSE/cm	R	Bias/cm	RMSE/cm
Chang2 算法						
>20	-0.19	-10.40	16.71	-0.19	0.61	4.76
5~20	0.73**	-7.82	14.47	0.52**	10.63	11.65
<5	0.11**	-12.15	16.87	-0.08*	15.86	17.20
Jiang 算法						
>20	-0.23*	-13.65	14.32	-0.14	-9.14	11.45
5~20	0.78**	-4.91	5.90	0.57**	-4.16	7.06
<5	0.31**	-1.48	5.59	0.16*	-0.20	8.45
Che 算法						
>20	-0.20*	-16.12	17.19			
5~20	0.77**	-7.73	8.42			
<5	0.26**	-3.42	7.45			

**表示 $P < 0.01$ ，*表示 $P < 0.05$ 。

五种雪深反演算法对浅雪的反演表现较中雪明显欠佳，算法反演雪深与观测雪深间的相关性较弱，Jiang 算法的相关系数最大（0.31），Che 算法（0.26）和

AMSR2 算法 (0.16) 次之, 而 SPD 算法的相关系数为负 (-0.08)。

Che 算法和 Jiang 算法的偏差绝对值和 RMSE 在浅雪雪深较中雪雪深偏小, 而 Chang2 算法和 SPD 算法与之相反。AMSR2 算法反演雪深的偏差绝对值在浅雪雪深小于对应中雪雪深, 但 RMSE 值表现相反。

3.3.3 结果与讨论

在各频率被动微波亮温中, 37 GHz 频率亮温被认为是地表温度的最佳反演波段^[103, 115], 受地表温度影响较大, 是地面无雪时期算法反演雪深变化的重要原因, 而其作为雪深反演算法的重要输入参数, 对算法反演雪深具有一定影响。以往研究结果显示被动微波雪深产品在青藏高原地区存在低估^[30], 而本文除 SPD 算法外, 其余四种算法整体均对研究区观测雪深低估, 这与以往研究结论一致。

Chang2 算法在研究区整体表现出低估, 偏差绝对值和均方根误差仅次于 SPD 算法, 这主要由于不同频率被动微波亮温对不同深度积雪场的敏感性存在差异, 理论上频率越高的亮温, 反演雪深的“饱和深度”越低, 而 Chang2 算法中使用的 19、37 GHz 频率亮温数据, 理论上只能反演 0.05~1 m 范围内的雪深^[98], 导致对浅雪的低估较大, 使得算法整体偏差较大。这也解释了 Chang2 算法在中、深雪记录较多的纳木错站点的表现要优于其余浅雪站点的现象, 目前普遍认为 89 GHz 频率亮温通道对浅雪的敏感性最好^[116], 但其受大气云中液态水含量影响较大^[117]。

算法评估结果显示, Che 算法对浅雪的反演效果较 Chang2 算法有较大提升, Che 算法的相关系数更高, 偏差绝对值和均方根误差更小, 这是因为 Che 算法考虑到了积雪场属性的动态变化, 针对不同时期进行了雪深补偿, 其补偿值基于中国区域雪深观测资料确定, 而中国绝大部分地区的积雪深度在 5 cm 以下^[30], 与研究区多个台站观测雪深接近, 使得该算法对浅雪反演效果较好, 但 Che 算法对中、深雪的反演表现不如 Chang2 算法, 表明 Che 算法可能更适合应用于浅雪地区的雪深反演。

SPD 算法在研究区五个站点均表现出高估, 对浅雪高估尤为突出, 在五种算法中其 R、Bias、RMSE 指标评价结果均表现最差。SPD 算法通过引入多个 19、37 GHz 垂直/水平极化亮温变量降低了地表冰壳、雪粒粒径等对雪深反演的干扰, 但算法中使用的亮温变量较多且变量系数较大, 导致反演雪深结果高估最大。另外, 相比 Chang2 算法, SPD 算法反演雪深表现出更大波动幅度, 这也是 SPD 算法表现不理想的原因之一, 这可能是由于引入的多个垂直极化亮温变量对雪深变化敏感导致的^[93]。

周胜男等^[118]研究表明, AMSR-E 雪深产品在草地覆被类型存在高估, 而本

文研究区土地覆被类型以草地为主, 使用了 AMSR2 亮温数据进行算法评估, 但得出了低估结论, 出现这一现象的可能原因有: 当地 (local) 尺度上存在比土地覆被类型权重更高的影响雪深反演的要素, 或由于被动微波数据源及处理流程差异导致评估结果不同。

Jiang 算法是基于混合像元思想建立的被动微波雪深反演算法, 使用了 10、19、37、89 GHz 频率亮温数据, 及不同土地覆被类型数据。在本研究区中纳木错、色林错、门当站是混合土地覆被类型的站点, 包含水体、草地和裸地类型, Jiang 算法在纳木错、色林错、门当站的相关系数、偏差和均方根误差指标均表现最好, 说明了 Jiang 算法在土地覆被类型复杂地区应用的巨大潜力。另外, Jiang 算法对浅雪的雪深反演表现最好, 可能与其使用了 89 GHz 频率亮温数据有关。

五种算法对深雪的反演表现较差, 相关系数均为负值且均未通过 0.01 显著性检验, 其中 AMSR2 算法及 Jiang 算法使用的 10 GHz 频率亮温数据具有监测深雪的能力^[119], 这表明现有算法对深雪的反演效果存在较大局限, 需要未来进一步深入研究。

3.4 本章小结

本章基于积雪覆盖率、土地覆被类型和海拔特征, 对台站的代表性进行讨论, 最终在对四种再分析雪深数据集评估时均剔除色林错、当雄站数据, 对 GLDAS、MERRA2、JRA55 数据集评估时, 分别剔除当雄、门当和当雄、当雄站点数据, 使用被动微波雪深反演雪深时不使用申扎、当雄站点数据。在剔除相应台站数据后, 分别对 ERA5、GLDAS、JRA55、MERRA2 四种再分析雪深数据集及 Chang2 算法、SPD 算法、Jiang 算法、Che 算法和 AMSR2 算法在色林错和纳木错湖区表现进行了评估。

评估结果显示, 四种数据集对研究区积雪天数均存在高估, 对浅雪和深雪表现均不佳。ERA5 数据集雪深与台站观测雪深的整体相关性最好, 对研究区雪深存在高估, 而其它三种再分析雪深数据集均表现不佳, JRA55 和 GLDAS 数据集对地面积雪出现“无雪”误判频率较高, MERRA2 雪深数据集存在“饱和”雪深现象, 在研究区表现最不理想。

五种被动微波雪深反演算法评估结果表明, 五种算法均对中雪表现最好, 对浅雪、深雪表现不理想。Jiang 算法在五种算法中整体表现最好, 对混合土地覆被类型站点的雪深反演效果也最好。SPD 算法表现最不理想, 同时是五种算法中唯一对研究区雪深高估的算法。AMSR2 算法表现仅优于 SPD 算法, 对研究区雪深低估。Chang2 算法反演雪深与台站观测雪深的相关性较好, 但偏差和均方根

误差较大,而 Che 算法反演的雪深的偏差和均方根误差整体较小,反演雪深散点较 Chang2 算法反演雪深散点分布范围更小。

第四章 基于机器学习的被动微波雪深反演研究

前文对五种被动微波雪深反演算法的评估结果表明, 算法反演雪深与研究区地面观测雪深间的线性关系较弱且精度不够理想, 这是由于五种被动微波雪深反演算法本质是对被动微波亮温数据集合与实际地面雪深数据集合间的线性回归, 常应用于雪深观测资料稀缺地区, 而在自然环境复杂的青藏高原地区, 被动微波辐射传输受地表冻融过程、下垫面特性、积雪场属性和短期大气状态等多种环境要素影响^[79], 使得基于线性经验算法的被动微波雪深反演在青藏高原地区应用仍存在诸多挑战^[64, 65], 因此本章针对被动微波雪深反演这一典型非线性多变量问题, 尝试使用机器学习模型来探究该问题, 期望实验结果为青藏高原被动微波雪深反演算法研究提供新的思路。

机器学习近年来在被动微波雪深反演研究领域应用较多^[53, 104, 105], 其具有高效数据挖掘和建模能力, 能够处理多维数据并充分挖掘数据间的信息, 减少样本噪声影响, 对维数较多的被动微波雪深反演具有较好的应用效果, 本章基于观测雪深、被动微波亮温变量、土地覆被类型和地形等数据, 建立机器学习雪深反演模型, 建模流程见图 4-1。

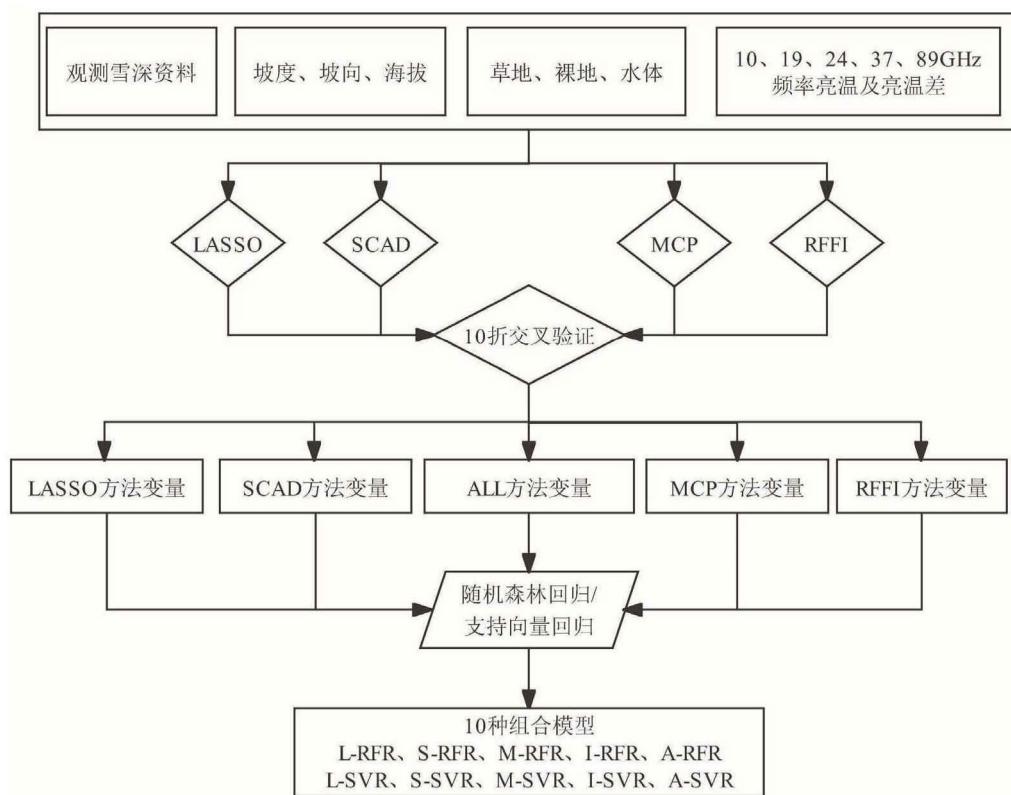


图 4-1 机器学习雪深反演模型建模流程

Figure 4-1 The flowchart of machine learning modeling for retrieval of snow depth

本文选择的机器学习方法为支持向量回归 (Support Vector Regression, SVR) 和随机森林回归 (Random Forest Regression, RFR) , 实验流程可简单分为模型输入变量选择、模型参数调优和模型模拟三个步骤(详细方法见 2.3 研究方法), 将获取到的 61 个下垫面、地形和亮温及亮温差变量进行筛选, 筛选出具有代表性的特征变量, 将部分数据作为训练样本输入模型, 优化模型参数, 最后利用剩余数据作为测试样本, 评估模型表现。

4.1 机器学习模型反演雪深结果与分析

4.1.1 模型输入变量选择结果

本次实验使用了 AMSR2 亮温数据集瞬时视场区域中草地、裸地和水体土地覆被类型占比数据、坡度、坡向、海拔、5 个不同频率 (10、19、24、37、89 GHz) 的垂直、水平亮温及其不同频率亮温间差值数据等 61 个变量数据 (见附录附表 1) , 对 61 个变量分别使用 LASSO、SCAD、MCP 和随机森林特征重要度方法进行筛选。

筛选结果显示 (见附录附表 2、附表 3、附表 4 和附表 5) , LASSO 方法共筛选出了 16 个变量, SCAD 方法共筛选出 15 个变量, MCP 方法筛选出 15 个变量, 随机森林特征重要度的筛选阈值设为 0.015, 最终筛选出 23 个变量 (见图 4-2) , 四种方法共筛选出 34 个变量 (见表 4-1) , 其中被四种方法同时选中的变量有 6 个, 被三种方法同时选中的变量有 6 个。仅被一种方法选择的变量有 17 个, 其中 12 个为随机森林特征重要度筛选出的变量。

为了分析不同方法筛选出的变量组合对模型表现的影响, 并增强模型反演雪深的解释性, 将四种方法筛选出来的 34 个变量作为实验对照组 (ALL 方法) , 结合 LASSO、SCAD、MCP 和随机森林特征重要度方法筛选出的四组变量组合, 得到了五组不同的变量组合。将五组变量组合分别作为支持向量回归和随机森林回归方法的实验变量, 得到了 10 种不同的组合模型 (见表 4-2) 。

表 4-1 机器学习模型初始变量的筛选结果

Table 4-1 Screening variables for initial variables of machine learning modeling

变量类型	变量名称	筛选结果
地形变量	坡度	LSMI
	坡向	LSMI
	海拔	LSMI

表中 L、S、M、I 分别代表变量被 LASSO、SCAD、MCP 和随机森林特征重要度方法筛选出来, “/”表示变量同时未被四种方法选择。

表 4-1 (续)
Table 4-1 (Continued)

变量类型	变量名称	筛选结果			
下垫面类型变量	水体	LSM			
	草地	LM			
	裸地	SI			
亮温变量	10V	/			
	10H	/			
	19V	M			
	19H	/			
	24V	/			
	24H	/			
	37V	LSI			
	37H	I			
	89V	LSMI			
	89H	LI			
变量名称	筛选结果	变量名称	筛选结果	变量名称	筛选结果
10V10H	LSMI	10H89V	/	24V24H	I
10V19V	/	10H89H	/	24V37V	/
10V19H	L	19V19H	I	24V37H	/
10V24V	SM	19V24V	LSM	24V89V	I
10V24H	/	19V24H	SMI	24V89H	I
10V37V	L	19V37V	LSI	24H37V	I
10V37H	/	19V37H	/	24H37H	M
10V89V	/	19V89V	/	24H89V	/
10V89H	I	19V89H	I	24H89H	/
10H19V	I	19H24V	I	37V37H	LSMI
10H19H	/	19H24H	LSM	37V89V	/
10H24V	I	19H37V	I	37V89H	/
10H24H	/	19H37H	/	37H89V	/
10H37V	/	19H89V	/	37H89H	L
10H37H	/	19H89H	/	89V89H	SM

表中 L、S、M、I 分别代表变量被 LASSO、SCAD、MCP 和随机森林特征重要度方法筛选出来，“/”表示变量同时未被四种方法选择。

表 4-2 10 种机器学习模型组合类型
Table 4-2 Ten types of machine learning models

回归方法	变量组合方法				
	LASSO	SCAD	MCP	RFFI	ALL
随机森林回归 (RFR)	L-RFR	S-RFR	M-RFR	I-RFR	A-RFR
支持向量回归 (SVR)	L-SVR	S-SVR	M-SVR	I-SVR	A-SVR

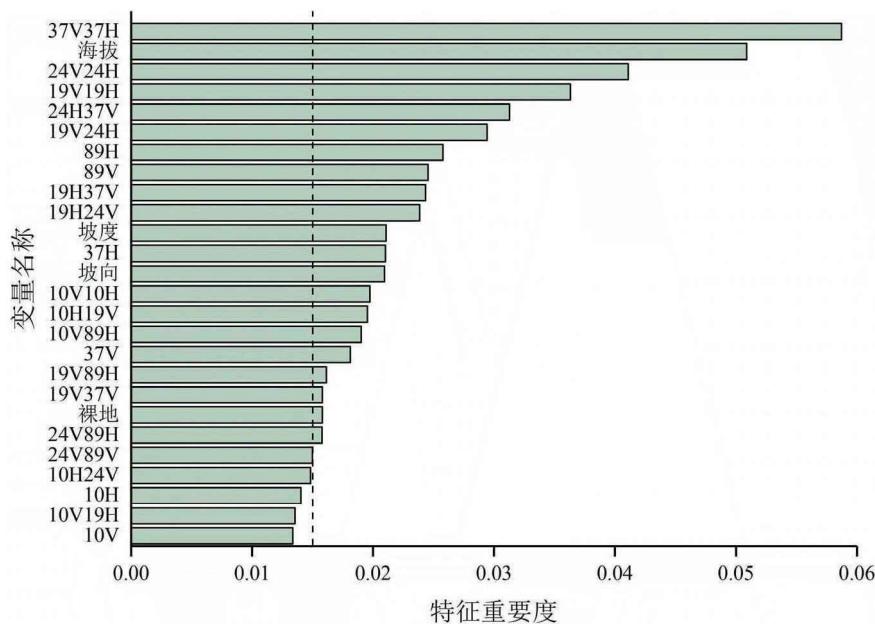


图 4-2 机器学习模型初始变量的特征重要度

Figure 4-2 Feature importance of initial variables for machine learning modeling

4.1.2 组合模型表现

对十种组合模型决定系数分析发现，L-RFR 模型决定系数最高（见图 4-3），为 0.82，I-SVR 模型的决定系数最低（为 0.63），随机森林回归模型（L-RFR、S-RFR、M-RFR、I-RFR、A-RFR）的决定系数均接近 0.8，而支持向量回归模型（L-SVR、S-SVR、M-SVR、I-SVR、A-SVR）的决定系数在 0.6~0.7 区间，同一变量组合与随机森林回归的组合模型的决定系数要优于对应变量组合与支持向量回归的组合模型。

在相同回归方法（支持向量回归或随机森林回归）的组合模型中，基于 LASSO、SCAD 和 MCP 三种方法的组合模型的决定系数依次减小，其中基于 LASSO 方法的组合模型（L-SVR、L-RFR）的决定系数最高，而基于 ALL 方法的组合模型的决定系数介于基于其它四种方法的组合模型的决定系数之间。

十种组合模型的 MAE 指标与决定系数表现出协同变化趋势（见图 4-3 和图 4-4），即决定系数高的模型，其 MAE 值相对较低。十种组合模型的 MAE 值主要分布在 2~3 cm 间，其中 L-RFR 的 MAE 值最小（2.04 cm），I-SVR 模型的 MAE 值最大（3.18 cm）。五种随机森林回归组合模型的 MAE 为 2 cm 左右，而五种支持向量回归组合模型的 MAE 值为 3 cm 左右，支持向量回归组合模型表现出更大的平均绝对误差。在相同回归方法（支持向量回归或随机森林回归）的组合模型中，基于 LASSO、SCAD 和 MCP 三种惩罚方法的组合模型的 MAE 值依次增加，基于 ALL 方法的组合模型（A-SVR 和 A-RFR）的 MAE 值仍介于基于其它四种方法的组合模型的 MAE 值之间。

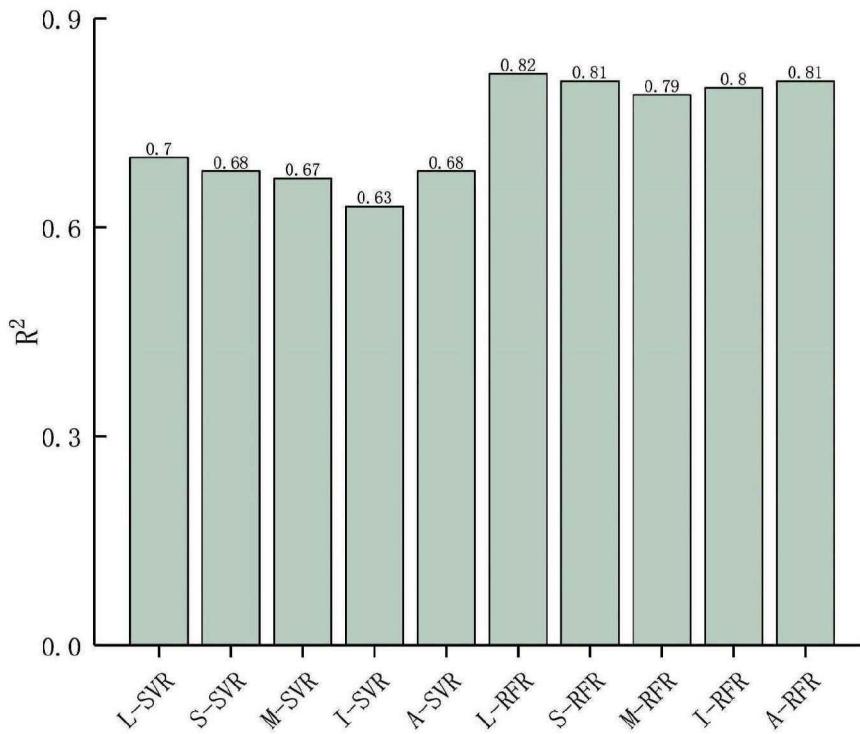


图 4-3 十种机器学习模型的决定系数

Figure 4-3 Determination coefficient of ten machine learning models

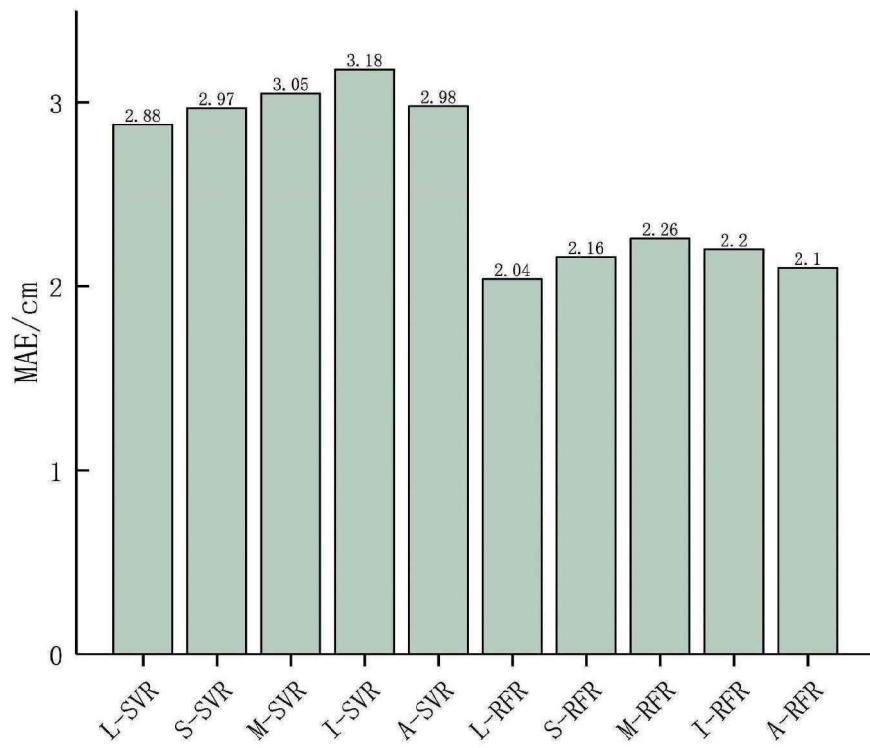


图 4-4 十种机器学习模型的平均绝对误差

Figure 4-4 Mean absolute error of ten machine learning models

十种组合模型的偏差分布结果显示，十种组合模型的偏差分布均接近正态分

布（见图 4-5），其中五种基于随机森林回归的组合模型的偏差峰值占比均超过 30%，而五种基于支持向量回归的组合模型的偏差分布峰值占比均低于 30%。五种变量组合与 RFR 的组合模型的偏差分布范围均较对应 SVR 的组合模型的偏差分布范围小，五种基于支持向量回归的组合模型的偏差在 -20 ~ -10 cm 范围内的分布频率较对应基于随机森林回归的组合模型的偏差的分布频率高。

不同变量组合对组合模型的偏差分布也存在影响，在随机森林回归的五种组合模型中，S-RFR 模型的偏差分布峰值最高，而其余四种组合模型的偏差分布峰值较为接近，其中 I-RFR 的偏差值最大（在 -35 cm 附近）。

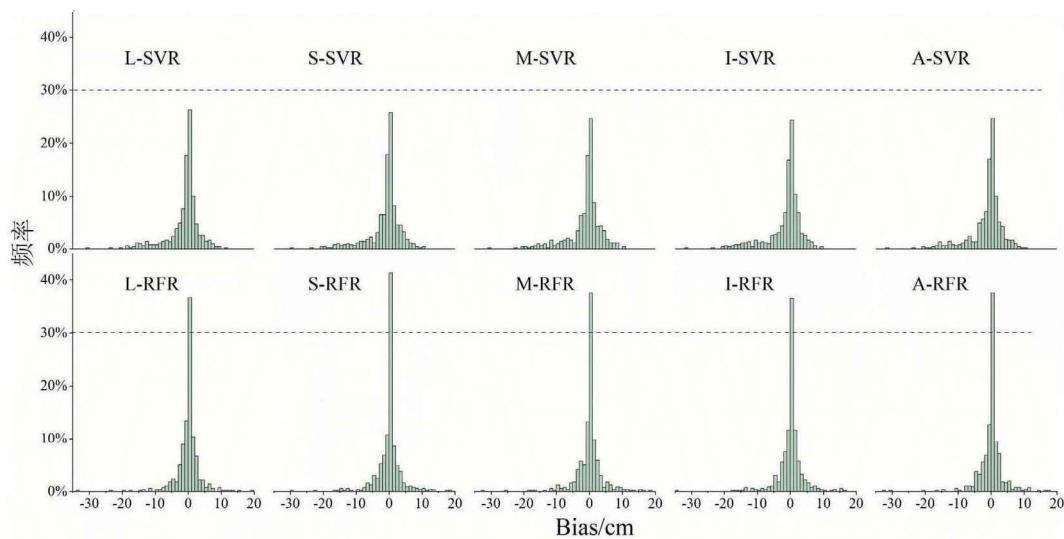


图 4-5 十种机器学习模型的偏差分布频率直方图

Figure 4-5 Frequency histogram of bias for ten machine learning models

十种组合模型反演雪深散点分布结果显示（见图 4-6），十种组合模型反演雪深均为正值，散点在浅雪区间的分布频率较高，与使用的研究区雪深数据多为浅雪有关，各组合模型对不同深度雪深的反演表现存在差异。

比较不同回归方法对组合模型结果影响发现，五种 RFR 的模型对浅雪的回归表现不佳，偏离 1:1 趋势线的回归雪深散点占比较高，其中较多模型回归雪深在 20 cm 附近的占比较高，对浅雪存在高估，而 RFR 模型对深雪的回归雪深散点在 1:1 趋势线附近分布频率较高。SVR 模型对深雪的回归雪深散点在 1:1 趋势线下方分布占比较高，反演雪深散点在 0 ~ 15 cm 范围的分布频率较高，表明 SVR 模型对深雪易低估。

不同变量组合对组合模型的表现也有影响，在五种基于支持向量回归的组合模型中，I-SVR 模拟雪深散点在 20 cm 深雪附近偏离 1:1 趋势线距离最大，是造成其平均绝对误差最大的重要原因。在五种 RFR 组合模型中，基于 LASSO 方法

的 L-RFR 组合模型对浅雪和深雪的反演雪深散点在 1:1 趋势线上的占比最高，表现出较好的线性趋势，在五种变量选择方法中表现最好。

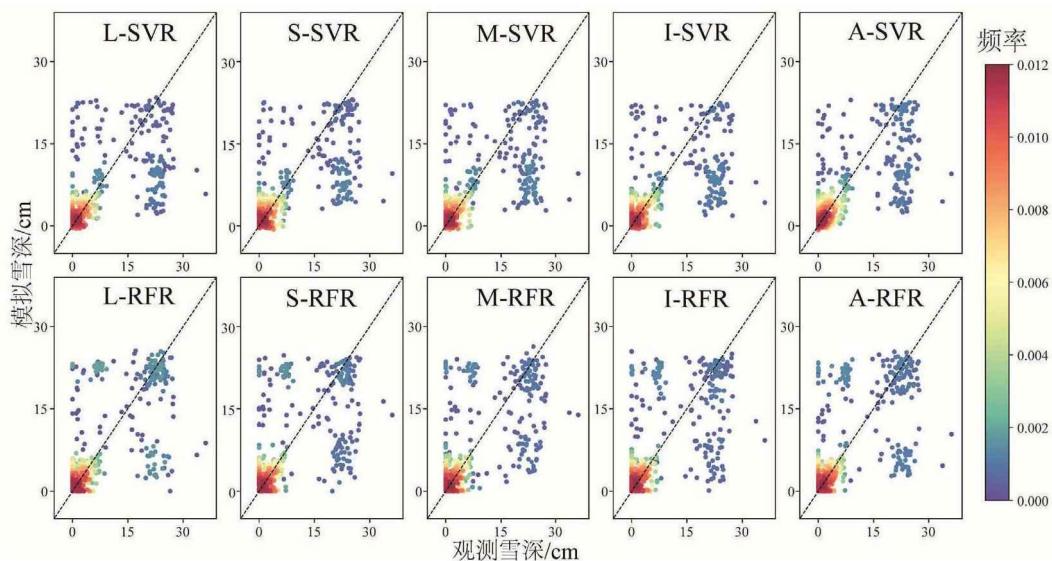


图 4-6 十种机器学习模型的模拟雪深散点分布

Figure 4-6 Snow depth scatterplot of ten machine learning models

4.1.3 结果讨论

在三种惩罚方法中，若变量的回归系数等于 0，表示该变量的变化不会引起雪深的变化，而当变量的回归系数大于（小于）0 时，则表示该变量越大，雪深值也越大（越小）。三种惩罚方法在地形、下垫面变量中仅同时选择了坡度、坡向、海拔和水体变量，其中水体、坡度和海拔变量的回归系数均为正，表明水体、坡度和海拔变量与研究区雪深具有正相关关系；而坡向变量的回归系数均为负，研究区的坡向在 160 ~ 190 度间，表明研究区的东南坡向较西南坡向更有利于雪深的积累。在亮温变量中，三种惩罚方法得到的变量回归系数均为负，表明亮温越大，雪深越小，这与以往研究结果一致^[98]，即雪深越大，积雪场对微波辐射的散射作用越强，传感器所接收到的亮温越小。而在亮温差变量中，仅 19H24H、37V37H 两变量的回归系数为负，其它被选中变量的回归系数均为正值。

地形、下垫面类型变量中，仅坡度、坡向和海拔等地形变量被四种方法同时选中，这说明坡度、坡向和海拔变量对研究区雪深影响较大，研究结果显示不同坡度、坡向和海拔组合能够导致积雪产生较大空间差异^[120]，这主要由于坡度、坡向能够影响地面接收的太阳辐射和区域气候系统的水汽传输，海拔能够引起气温的梯度变化，适宜的坡度、坡向和海拔地形要素组合能够为积雪的累积、保存提供良好的自然环境，这也是青藏高原高山区常年积雪的重要原因之一。

在亮温变量中，仅有 89V 变量被四种方法同时选中，这可能由于研究区雪

深观测数据多为 0~5 cm 浅雪, 而 89 GHz 垂直极化亮温对浅雪较为敏感, 使得该频率通道亮温对研究区雪深变化的解释性强。而在亮温差变量中, 10V10H 和 37V37H 亮温差变量同时被四种算法选中, 这可能与 10、37 GHz 频率被动微波在雪深反演中的特性有关, 10 GHz 频率亮温通常被认为适用于厚雪监测^[73], 而前人利用辐射传输模型的模拟结果则揭示了 37 GHz 频率亮温对雪深的敏感性^[66], 37 GHz 频率亮温是当前被动微波雪深反演算法中最常用的亮温变量^[71]。在所有被选中的被动微波亮温变量中, 垂直极化亮温被选中的频率要高于水平极化亮温, 这可能与垂直极化亮温对雪深变化更敏感有关^[93]。

在十种组合模型中, 除 I-SVR 模型外 (0.63), 其余九种模型的决定系数均大于 0.67, 而四种再分析雪深数据集与地面观测雪深的相关系数最大为 0.62, 五种被动微波算法反演雪深与地面雪深的相关系数最大为 0.65。十种组合模型的 MAE 均小于 3.05 cm, 反演雪深的散点多分布于 1:1 趋势线附近, 而五种被动微波算法的整体偏差绝对值均大于 5 cm, 尽管四种再分析雪深数据集中 ERA5、GLDAS、JRA55 再分析雪深数据集偏差较小, 但其雪深散点分布趋势较十种组合模型的反演雪深散点分布趋势差。这表明十种组合模型表明机器学习模型反演雪深在研究区的整体表现要优于四种再分析雪深数据集与五种被动微波算法,

通过与前人在青藏高原地区开展的被动微波雪深反演研究对比发现, 以往机器学习雪深反演模型未考虑坡度、坡向变量, 模型决定系数低于 0.7^[119], 而本文在机器学习模型引入坡度、坡向变量后, 实验结果的决定系数超过 0.8, 也从侧面反映了坡度、坡向对基于被动微波的雪深反演较为重要。同时本文构建的机器学习模型表现要优于融合了积雪覆盖影像数据的降尺度雪深反演方法(相关系数为 0.56)^[121]、积雪过程模型与微波辐射传输模型的耦合结果(相关系数为 0.78)^[122], 这说明本文构建的机器学习雪深反演模型在青藏高原具有较好的应用潜力。

基于不同方法的变量组合对回归模型的表现影响较大, 基于 LASSO 方法的组合模型模拟结果优于基于 SCAD 和 MCP 方法的组合模型, 这是因为基于 LASSO 方法选择的重要变量比其它两个方法多, 在拟合时会得到较大的 R^2 和较小的 MAE。随机森林特征重要度方法筛选的变量组合在支持向量回归模型中表现最差, 而在随机森林回归模型中表现优于 MCP 惩罚方法得到的变量组合, 这可能与随机森林特征重要度变量筛选阈值设置有关, 选中的 23 个变量中仍存在无关、冗余变量, 而随机森林回归方法对过拟合表现较好, 使得基于随机森林特征重要度方法的变量组合在随机森林回归方法中表现优于支持向量回归方法。SCAD 和随机森林特征重要度方法筛选得到的变量组合表现劣于 ALL 方法的变量组合表现, 说明 SCAD 和随机森林特征重要度方法对研究区被动微波雪深反演变量的筛选效果不佳。同时不难发现, 基于 ALL 方法的组合模型在基于五种

方法的组合模型中表现并非最优,表明此时 ALL 方法的变量组合中可能存在过多无关变量,这也表明将过多的变量引入模型,并不一定会增加模型拟合的精度与准确性。

比较基于五种方法的变量组合发现,基于 SCAD 方法与基于 MCP 方法的变量组合中的不同变量数最少,对两组合变量去重后可知,SCAD 方法组合变量的剩余变量为裸地、37V 和 19V37V,而 MCP 方法组合变量的剩余变量为草地、19V 和 24H37H,说明裸地、37V 和 19V37V 变量对研究区雪深变化的解释性要强于草地、19V 和 24H37H 变量。

随机森林特征重要度方法得到的变量重要度结果显示,裸地、37V 和 19V37V 变量的重要度均大于 0.015,而草地、19V 和 24H37H 变量均小于 0.015,这侧面表明裸地、37V 和 19V37V 的组合变量对研究区雪深变化具有重要作用,但模型模拟过程通常涉及多个变量,当变量间存在多重共线性时,不同变量间可能会相互影响,例如下垫面类型也会影响被动微波亮温数据^[123],将给特征变量得分带来较大影响,同时变量重要度得分是针对当前所选中的决策树模型进行计算,训练样本的选取对模型结果具有较大影响,因此裸地、37V 和 19V37V 三个变量对雪深的敏感性强弱还不清楚,还需要更多实验数据的支持。

评估结果表明,相较支持向量回归的组合模型,随机森林回归的组合模型决定系数高、偏差分布集中且平均绝对误差小,模型反演雪深的散点具有较好线性趋势,模拟雪深散点在 1:1 趋势线附件分布频率较高,综合表现优于支持向量回归模型,实验结果与以往结果一致^[119],这可能是因为(1)随机森林回归使用的决策树算法无需先验假设,在解决模型过拟合问题上表现较好^[124]。(2)随机森林回归在每个决策树上使用多个模型进行模拟,最终对多个模型择优,所以随机森林受噪声数据干扰小,拟合较好。(3)由于算法采用了随机采样,训练出的模型的方差小,泛化能力强。(4)支持向量机回归结果的优劣受核函数影响较大,对一个实际问题而言,如何根据实际的数据模型选择合适的核函数是至关重要的。目前比较成熟的核函数及其参数的选择都是人为的,根据经验来选取的,带有一定的不确定性。最终导致在雪深反演模型中,基于支持向量回归的模型表现不是最优的。

4.2 本章小结

针对被动微波雪深反演的非线性多变量特征,基于草地、裸地、水体、坡度、坡向、海拔、5 个不同频率(10、19、24、37、89 GHz)的垂直、水平亮温及亮温差数据等 61 个变量数据和雪深观测数据,使用 LASSO、SCAD、MCP 和随机

森林特征重要度四种方法筛选重要变量，并将四种方法选中的所有变量作为实验对照组（ALL 方法），得到了五种不同的变量组合。将不同变量组合分别与随机森林回归（RFR）和支持向量回归（SVR）组合，得到了十种机器学习模型，并使用了决定系数、偏差和平均绝对误差指标对模型表现进行评价。

变量筛选结果显示，在地形、下垫面要素中坡度、坡向和海拔变量被四种方法同时选中，10、37、89 GHz 频率亮温和垂直频率亮温被选中频率较高，表明这些变量对区域雪深变化的解释性好。

模型评价结果显示，十种组合模型的决定系数在 0.63~0.82 间，平均绝对误差在 2~3 cm 间，偏差均接近正态分布，决定系数与平均绝对误差表现出协同变化趋势。五种变量组合中基于 LASSO、SCAD 方法的变量组合表现最好，基于 MCP 和随机森林特征重要度方法的变量组合表现相对不佳，而基于 ALL 方法的变量组合表现介于四种方法之间。随机森林回归对区域雪深反演整体表现优于支持向量回归。RFR 模型和 SVR 模型对不同深度雪深反演表现存在差异，SVR 组合模型对深雪低估的频率较高，而 RFR 模型对浅雪易出现高估，反演雪深散点在 1:1 趋势线附近的整体分布频率较高。机器学习模型的雪深反演表现整体优于四种再分析雪深数据集和五种被动微波雪深反演算法。

第五章 结论与展望

5.1 主要研究结论

本文主要基于青藏高原中部色林错、纳木错湖区 2019 年 7 月至 2021 年 6 月的地表雪深观测资料和 AMSR2 亮温数据，使用海拔、土地覆被类型和积雪覆盖率特征探讨了研究区雪深观测站点对区域积雪场的代表性，在剔除代表性较弱的站点雪深数据后，对 ERA5、GLDAS、JRA55、MERRA2 四种再分析雪深数据集和 Chang2 算法、SPD 算法、Che 算法、Jiang 算法、AMSR2 算法在研究区的表现分别进行了评估。其后，为探究不同亮温通道与地形、下垫面等环境要素对雪深反演的影响，使用 LASSO、SCAD、MCP 和随机森林特征重要度方法对获取的 61 个环境变量进行筛选，建立了基于随机森林回归和支持向量回归的十种机器学习雪深反演模型，使用了决定系数、偏差和平均绝对误差指标对模型表现进行评估，得到的主要研究结论有：

(1) 对 ERA5、GLDAS、JRA55、MERRA2 四种再分析雪深数据集的评估结果显示，四种再分析雪深数据集的月平均积雪天数均高于站点观测数据 (6.59 d)，其中 MERRA2 数据集的高估最大 (25.31 d)。ERA5 再分析雪深数据集在研究区的整体表现最好，与观测雪深的相关系数为 0.62，对研究区雪深整体高估 (2.37 cm)，其中对深雪高估最大 (12.11 cm)。JRA55 数据集雪深与观测雪深的整体相关系数为 0.41，均方根误差在四种数据集中最小 (4.06 cm)。GLDAS 和 MERRA2 雪深数据集在研究区表现最不理想 (R 值分别为 0.23、0.14)，其中 GLDAS 数据集的整体偏差最小 (-0.68 cm)，而 MERRA2 数据集的整体偏差最大 (11.71 cm)，在四种数据集中表现最不理想。四种数据集雪深与地面浅雪的相关性均较弱 (小于 0.20)，ERA5 和 JRA55 数据集雪深与地面中雪的相关性最高，R 值分别为 0.80、0.47。

(2) 五种被动微波雪深反演算法的评估结果显示，Jiang 算法在研究区的整体表现最好，反演雪深与地面观测雪深的相关系数最高 (0.65)，RMSE 为 5.51 cm，对混合土地覆被类型的色林错站点的雪深反演表现最好 (R 为 0.56)。SPD 算法和 AMSR2 算法表现不理想，R 值分别为 0.42、0.45，其中 SPD 算法的偏差 (15.80 cm) 和均方根误差 (17.10 cm) 最大。Chang2 算法反演雪深与地面观测雪深的相关性较高 (0.59)，但偏差 (-12.4 cm) 和均方根误差 (17 cm) 仅次于 SPD 算法。Che 算法的偏差 (-3.43 cm) 和均方根误差 (7.36 cm) 整体较小。五种算法中 SPD 算法对研究区雪深高估，而其余四种算法对研究区雪深低估。

Chang2 算法、Jiang 算法和 Che 算法对中雪的反演效果较好, 反演雪深与观测雪深的相关系数均大于 0.7。

(3) 十种雪深反演组合模型的决定系数高 (0.63~0.82)、偏差分布接近正态分布、平均绝对误差小(均小于 3.2 cm), 其中 L-RFR 模型的表现最好 (0.82)。除 I-SVR 组合模型外 (0.63), 其余九种模型的决定系数大于 0.67, MAE 小于 3.05 cm, 九种组合模型的表现均优于四种再分析雪深数据集 (R 为 0.62) 和五种被动微波雪深反演算法 (R 为 0.65)。基于随机森林回归的组合模型的决定系数 (0.79~0.82) 均高于基于支持向量回归的组合模型的决定系数 (0.63~0.7), 五种变量组合中, LASSO、SCAD 方法变量组合表现最好, MCP 方法和随机森林特征重要度方法变量组合表现不佳, ALL 方法变量组合表现介于其它四组变量组合之间。

(4) 惩罚方法的变量筛选结果表明, 水体、坡度和海拔变量与研究区雪深具有正相关关系, 东南坡向更有利于雪深的积累, 筛选出的亮温变量与研究区雪深均具有负相关关系, 筛选出的亮温差变量仅 19H24H、37V37H 两变量与研究区雪深表现出负相关, 其它被筛选变量均表现出正相关。坡度、坡向、海拔和 89V、10V10H 和 37V37H 变量被四种变量选择方法同时选中, 对研究区雪深变化的解释性强, 在所有被选中的亮温变量中, 垂直极化亮温变量被选中的频率要高于水平极化亮温变量。裸地、37V 和 19V37V 的组合变量对研究区雪深反演的表现要优于草地、19V 和 24H37H 的组合变量。

5.2 创新点

本研究对四种再分析雪深资料和五种被动微波雪深反演算法在研究区的表现进行了评估, 结果显示 ERA5 再分析雪深资料与 Jiang 算法在研究区表现最好, 评估结果提升了对再分析雪深资料与被动微波雪深反演算法的区域认知。首次使用 LASSO、SCAD、MCP 和随机森林特征重要度方法, 探索了地形、下垫面、亮温及亮温差变量对雪深反演的影响, 结果显示坡度、坡向和海拔等地形要素对研究区雪深反演具有重要影响。基于土地覆被类型、地形、被动微波亮温及亮温差变量建立的机器学习雪深反演模型, 整体表现要优于四种再分析雪深资料与五种被动微波雪深反演算法, 表明了基于机器学习的被动微波雪深反演模型在青藏高原地区具有较大应用潜力。

5.3 不足及展望

本文对 ERA5、GLDAS、JRA55、MERRA2 四种再分析雪深数据集和 Chang2 算法、SPD 算法、Che 算法、Jiang 算法、AMSR2 算法在研究区的应用表现进行了评估, 填补了四种再分析雪深数据集与五种被动微波雪深反演算法在研究区区域表现的空白。基于 LASSO、SCAD、MCP 和随机森林特征重要度变量筛选方法与随机森林回归和支持向量回归建立了雪深反演的机器学习模型, 模型表现优于四种再分析雪深数据集和五种被动微波雪深反演算法, 探究了地形、下垫面和被动微波亮温及亮温差变量对研究区雪深反演的影响, 研究结果能为青藏高原地区的被动微波雪深反演算法研究提供参考, 但本文也存在一些不足之处, 有待未来进一步完善:

(1) 地表积雪场的积雪密度、粒径大小、积雪场结构等属性变化均会影响积雪场的微波辐射特性, 由于实测数据获取难度较高, 本文在进行被动微波雪深反演算法评估和建立机器学习模型反演雪深时, 未考虑到积雪场属性动态变化对被动微波亮温的影响。同时, 评估中所使用的 89 GHz 频率亮温对大气状态较为敏感, 本文在使用时也未进行大气校正, 可能会给实验结果带来偏差, 这需要未来收集研究区的积雪场属性与环境信息, 并结合辐射传输模型以深入研究。

(2) 由于本文所使用的是台站雪深数据仅为两个连续积雪年, 雪深数据类型多为浅雪, 对于中雪、深雪的观测数据偏少, 且部分台站由于对区域积雪场的代表性较弱在实验中被剔除, 导致最终可用于再分析雪深资料和被动微波雪深反演算法评估及机器学习建模的数据样本量偏少, 给研究结果增加了不确定性, 未来可考虑增加青藏高原中、西部地区的雪深观测台站, 以收集更多中雪、深雪观测数据。

参考文献

- [1] 姚檀栋, 陈发虎, 崔鹏, 等. 从青藏高原到第三极和泛第三极[J]. 中国科学院院刊, 2017, 32(09): 924-931.
- [2] 姚檀栋, 邬光剑, 徐柏青, 等. “亚洲水塔”变化与影响[J]. 中国科学院院刊, 2019, 34(11): 1203-1209.
- [3] Immerzeel, W.W., Lutz, A., Andrade, M., et al. Importance and vulnerability of the world's water towers[J]. Nature, 2020, 577(7790): 364-369.
- [4] You, Q., Wu, T., Shen, L., et al. Review of snow cover variation over the Tibetan Plateau and its influence on the broad climate system [J]. Earth-Science Reviews, 2020, 201: 103043.
- [5] 张廷军, 钟歆玥. 欧亚大陆积雪分布及其类型划分[J]. 冰川冻土, 2014, 36(03): 481-490.
- [6] 王建, 车涛, 李震, 等. 中国积雪特性及分布调查[J]. 地球科学进展, 2018, 33(01): 12-26.
- [7] 张欢, 邱玉宝, 郑照军, 等. 基于 MODIS 的青藏高原季节性积雪去云方法可行性比较研究[J]. 冰川冻土, 2016, 38(03): 714-724.
- [8] Lutz, A., Immerzeel, W., Shrestha, A., et al. Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation[J]. Nature Climate Change, 2014, 4(7): 587-592.
- [9] Sexstone, G.A., Driscoll, J.M., Hay, L.E., et al. Runoff sensitivity to snow depletion curve representation within a continental scale hydrologic model[J]. Hydrological Processes, 2020, 34(11): 2365-2380.
- [10] Mcclung, D.M. Avalanche character and fatalities in the high mountains of Asia[J]. Annals of Glaciology, 2016, 57(71): 114-118.
- [11] 秦大河. 冰冻圈科学辞典[M]. 北京: 气象出版社, 2014.
- [12] 汶林科, 向灵芝, 蔡毅, 等. 雪崩的形成机理研究[J]. 山地学报, 2016, 34(01): 1-11.
- [13] 郭晓宁. 青海高原近 50a 来雪灾特征研究[D]. 兰州: 兰州大学, 2010.
- [14] Wang, Z., Wu, R., Huang, G. Low-frequency snow changes over the Tibetan Plateau[J]. International Journal of Climatology, 2018, 38(2): 949-963.
- [15] Li, C., Su, F., Yang, D., et al. Spatiotemporal variation of snow cover over the Tibetan Plateau based on MODIS snow product, 2001–2014[J]. International Journal of Climatology, 2018, 38(2): 708-728.
- [16] Basang, D., Barthel, K., Olseth, J.A. Satellite and ground observations of snow cover in Tibet during 2001–2015[J]. Remote Sensing, 2017, 9(11): 1201.
- [17] Immerzeel, W.W., Beek, L.P.H.V., Bierkens, M.F.P. Climate Change Will Affect the Asian Water Towers[J]. Science, 2010, 328(5984): 1382-1385.
- [18] Shaman, J., Tziperman, E. The Effect of ENSO on Tibetan Plateau Snow Depth: A Stationary Wave Teleconnection Mechanism and Implications for the South Asian Monsoons[J]. Journal of Climate, 2005, 18(12): 2067-2079.
- [19] Qian, Y.F., Zheng, Y.Q., Zhang, Y., et al. Responses of China's summer monsoon climate to snow anomaly over the Tibetan Plateau[J]. International Journal of Climatology, 2010, 23(6): 593-613.

- [20] Flanner, Mark, G. Snowpack radiative heating: Influence on Tibetan Plateau climate[J]. *Geophysical Research Letters*, 2005, 32(6): L06501.
- [21] Bao, Y., You, Q. How do westerly jet streams regulate the winter snow depth over the Tibetan Plateau?[J]. *Climate Dynamics*, 2019, 53(1): 353-370.
- [22] Lü, J.M., Ju, J.H., Kim, S.J., et al. Arctic Oscillation and the autumn/winter snow depth over the Tibetan Plateau[J]. *Journal of Geophysical Research: Atmospheres*, 2008, 113(D14): D14117.
- [23] Wang, Y., Huang, X., Wang, J., et al. AMSR2 snow depth downscaling algorithm based on a multifactor approach over the Tibetan Plateau, China[J]. *Remote Sensing of Environment*, 2019, 231: 111268.
- [24] 丁锋, 孙照渤, 刘敬乐. 青藏高原冬春雪深分布与中国夏季降水的关系[J]. *大气科学学报*, 2009, 32(06): 783-791.
- [25] Li, Z., Song, Y., Zhang, W., et al. Interdecadal correlation of solar activity with Tibetan Plateau snow depth and winter atmospheric circulation in East Asia[J]. *Sciences in Cold and Arid Regions*, 2016, 8(06): 524-535.
- [26] 朱崑崙, 徐淑英, 段月薇, 等. 气象记录在经济建设中的应用(I)——全国风压、雪深及气温区域的划分[J]. *地理学报*, 1954(01): 45-58.
- [27] 中国西藏新闻网. 西藏地面气象观测迈入全面自动化时代[EB/OL]. (2020-04-02) [2021-11-22]. http://www.xzxw.com/xw/202004/t20200402_2999618.html.
- [28] 唐国栋, 柯长青. 中国西部主要台站积雪深度的空间插值研究[J]. *气候与环境研究*, 2007(04): 566-572.
- [29] 柯长青, 李培基. 用 EOF 方法研究青藏高原积雪深度分布与变化[J]. *冰川冻土*, 1998(01): 65-68.
- [30] 李小兰, 张飞民, 王澄海. 中国地区地面观测积雪深度和遥感雪深资料的对比分析[J]. *冰川冻土*, 2012, 34(04): 755-764.
- [31] 除多, 洛桑曲珍, 林志强, 等. 近 30 年青藏高原雪深时空变化特征分析[J]. *气象*, 2018, 44(02): 233-243.
- [32] Xu, W., Ma, L., Ma, M., et al. Spatial-temporal variability of snow cover and depth in Qinghai-Tibetan Plateau[J]. *Journal of Climate*, 2017, 30(4): 1521-1533.
- [33] Chen, X., Yang, Y. Spring snow depth changes and feedback to surface air temperature across the Tibetan plateau from 1961 to 2013[J]. *International Journal of Climatology*, 2022, 42(1): 32-47.
- [34] You, Kang, S.C., Ren, Y.G., et al. Observed changes in snow depth and number of snow days in the eastern and central Tibetan Plateau[J]. *Climate Research*, 2011, 46(2): 171-183.
- [35] 白淑英, 史建桥, 沈渭寿, 等. 近 30 年西藏雪深时空变化及其对气候变化的响应[J]. *国土资源遥感*, 2014, 26(01): 144-151.
- [36] 谭秋阳, 程磊, 徐宗学, 等. 1979—2017 年雅鲁藏布江流域雪深时空分布特征及其影响因素分析[J]. *冰川冻土*, 2021, 43(04): 1049-1059.
- [37] Frei, A., Tedesco, M., Lee, S., et al. A review of global satellite-derived snow products[J]. *Advances in Space Research*, 2012, 50(8): 1007-1029.

- [38] Jacobs, J.M., Hunsaker, A.G., Sullivan, F.B., et al. Snow depth mapping with unpiloted aerial system lidar observations: a case study in Durham, New Hampshire, United States[J]. *The Cryosphere*, 2021, 15(3): 1485-1500.
- [39] Deems, J.S., Gadomski, P.J., Vellone, D., et al. Mapping starting zone snow depth with a ground-based lidar to assist avalanche control and forecasting[J]. *Cold Regions Science Technology*, 2015, 120: 197-204.
- [40] Deems, J.S., Painter, T.H., Finnegan, D.C. Lidar measurement of snow depth: a review[J]. *Journal of Glaciology*, 2013, 59(215): 467-479.
- [41] Nievinski, F.G., Larson, K.M. Inverse Modeling of GPS Multipath for Snow Depth Estimation—Part II: Application and Validation[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2014, 52(10): 6564-6573.
- [42] Larson, K.M., Gutmann, E.D., Zavorotny, V.U., et al. Can we measure snow depth with GPS receivers?[J]. *Geophysical Research Letters*, 2009, 36(17): L17502.
- [43] Larson, K.M. GPS interferometric reflectometry: applications to surface soil moisture, snow depth, and vegetation water content in the western United States[J]. *Wiley Interdisciplinary Reviews Water*, 2016, 3(6): 775-787.
- [44] Zhang, J., Liu, L., Su, L., et al. Three-in-one GPS-IR measurements of ground surface elevation changes, soil moisture, and snow depth at a permafrost site in the northeastern Qinghai-Tibet Plateau[J]. *The Cryosphere Discussions*, 2021, 15(6): 3021-3033.
- [45] Zhou, W., Liu, L., Huang, L., et al. A new GPS SNR-based combination approach for land surface snow depth monitoring[J]. *Scientific Reports*, 2019, 9(1): 1-20.
- [46] Klimánek, M., Mikita, T., Lizuch, M., et al. Using GPS for snow depth and volume measurement of centennial avalanche field in High Tatras[J]. *Cold Regions Science Technology*, 2011, 65(3): 392-400.
- [47] Zhang, S., Meiling, Z., Yajie, W., et al. Ground-based GPS Used in the Snow Depth Survey of Greenland[J]. *Journal of Geodesy and Geoinformation Science*, 2021, 4(02): 47-55.
- [48] Boniface, K., Braun, J.J., McCreight, J.L., et al. Comparison of Snow Data Assimilation System with GPS reflectometry snow depth in the Western United States[J]. *Hydrological Processes*, 2015, 29(10): 2425-2437.
- [49] Zhang, S., Wang, X., Zhang, Q. Avoiding errors attributable to topography in GPS-IR snow depth retrievals[J]. *Advances in Space Research*, 2016, 59(6): 1663-1669.
- [50] Yu, K., Ban, W., Zhang, X., et al. Snow Depth Estimation Based on Multipath Phase Combination of GPS Triple-Frequency Signals[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(9): 5100-5109.
- [51] 王婷. 青藏高原积雪对陆面过程水热输送的影响研究[D]. 成都: 成都信息工程大学, 2019.
- [52] Liu, Z., Yao, Z., Wang, R., et al. Estimation of the Qinghai-Tibetan Plateau runoff and its contribution to large Asian rivers[J]. *Science of the Total Environment*, 2020, 749: 141570.
- [53] 乔德京. 多源雪深数据不确定性分析及其融合研究[D]. 西安: 西安科技大学, 2020.
- [54] Orsolini, Y., Wegmann, M., Dutra, E., et al. Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalyses using in situ and satellite remote sensing observations[J]. *The Cryosphere*, 2019, 13(08): 2221-2239.

- [55] Tedesco, M. Remote sensing of snow depth and snow water equivalent[J]. *Remote Sensing of the Cryosphere*, 2015, 12: 73-98.
- [56] 崔玉荣. 基于主动微波的积雪参数反演研究[D]. 北京: 中国科学院大学(中国科学院遥感与数字地球研究所), 2018.
- [57] Baghdadi, N., Gauthier, Y., Bernier, M. Capability of multitemporal ERS-1 SAR data for wet-snow mapping[J]. *Remote Sensing of Environment*, 1997, 60(2): 174-186.
- [58] Lievens, H., Demuzere, M., Marshall, H.P., et al. Snow depth variability in the Northern Hemisphere mountains observed from space[J]. *Nature Communications*, 2019, 10(1): 1-12.
- [59] Yang, L., Lanhai, L., Jinming, Y., et al. Estimating Snow Depth Using Multi-Source Data Fusion Based on the D-InSAR Method and 3DVAR Fusion Algorithm[J]. *Remote Sensing*, 2017, 9(11): 1195.
- [60] Li, H., Wang, Z., He, G., et al. Estimating Snow Depth and Snow Water Equivalence Using Repeat-Pass Interferometric SAR in the Northern Piedmont Region of the Tianshan Mountains[J/OL]. *Journal of Sensors*, 2017, DOI: <https://doi.org/10.1155/2017/8739598>.
- [61] Majumdar, S., Thakur, P.K., Chang, L., et al. Snow Depth and Snow Water Equivalent Estimation in the Northwestern Himalayan Watershed using Spaceborne Polarimetric SAR Interferometry[EB/OL]. (2020-05-24) [2022-1-7]. <https://eartharxiv.org/>. DOI: <https://doi.org/10.31223/osf.io/6v4h3>.
- [62] 蒋玲梅, 崔慧珍, 王功雪, 等. 积雪、土壤冻融与土壤水分遥感监测研究进展[J]. 遥感技术与应用, 2020, 35(06): 1237-1262.
- [63] 施建成, 熊川, 蒋玲梅. 雪水当量主被动微波遥感研究进展[J]. 中国科学: 地球科学, 2016, 46(04): 529-543.
- [64] 黄晓东, 李旭冰, 刘畅宇, 等. 青藏高原积雪范围和雪深/雪水当量遥感反演研究进展及挑战[J]. 冰川冻土, 2019, 41(05): 1138-1149.
- [65] Liyun, D., Tao, C., Yongjian, D., et al. Evaluation of snow cover and snow depth on the Qinghai-Tibetan Plateau derived from passive microwave remote sensing[J]. *Cryosphere*, 2017, 11(4): 1933-1948.
- [66] Chang, A.T.C. Nimbus-7 SMMR derived global snow cover parameters[J]. *Annals of Glaciology* 1987, 9: 39-44.
- [67] Schuster, A. Radiation through a foggy atmosphere[J]. *The Astrophysical Journal*, 1905, 21: 1.
- [68] 金亚秋. 矢量辐射传输理论和参数反演[M]. 郑州: 河南科学技术出版社, 1994.
- [69] Wiesmann, A., Mätzler, C. Microwave emission model of layered snowpacks[J]. *Remote Sensing of Environment*, 1999, 70(3): 307-316.
- [70] Tsang, L., Chen, C.-T., Chang, A.T., et al. Dense media radiative transfer theory based on quasicrystalline approximation with applications to passive microwave remote sensing of snow[J]. *Radio Science*, 2000, 35(3): 731-749.
- [71] Aschbacher, J. Land surface studies and atmospheric effects by satellite microwave radiometry[D]. Austria: University of Innsbruck, 1993.
- [72] Foster, J. Comparison of snow mass estimates from a prototype passive microwave snow algorithm, a revised algorithm and a snow depth climatology[J]. *Remote Sensing of Environment*, 1997, 62(2): 132-142.

- [73] Kelly, R. The AMSR-E snow depth algorithm: Description and initial results[J]. *Journal of the Remote Sensing Society of Japan*, 2009, 29(1): 307-317.
- [74] Kelly, R.E., Chang, A.T., Tsang, L., et al. A prototype AMSR-E global snow area and snow depth algorithm[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2): 230-242.
- [75] 曹梅盛, 李培基. 中国西部积雪微波遥感监测[J]. *山地学报*, 1994(04): 230-234.
- [76] Chang, A., Foster, J.L., Hall, D.K., et al. The use of microwave radiometer data for characterizing snow storage in western China[J]. *Annals of Glaciology*, 1992, 16: 215-219.
- [77] 柏延臣, 冯学智, 李新, 等. 基于被动微波遥感的青藏高原雪深反演及其结果评价[J]. *遥感学报*, 2001(03): 161-165.
- [78] Che, T., Xin, L., Jin, R., et al. Snow depth derived from passive microwave remote-sensing data in China[J]. *Annals of Glaciology*, 2008, 49(1): 145-154.
- [79] Lingmei, J., Pei, W., Lixin, Z., et al. Improvement of snow depth retrieval for FY3B-MWRI in China[J]. *Science China Earth Science*, 2014, 57(6): 1278-1292.
- [80] Dai, L., Che, T., Xie, H., et al. Estimation of Snow Depth over the Qinghai-Tibetan plateau based on AMSR-E and MODIS data[J]. *Remote Sensing*, 2018, 10(12): 1989.
- [81] Xiao, X., Zhang, T., Zhong, X., et al. Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data[J]. *Remote Sensing of Environment*, 2018, 210: 48-64.
- [82] 肖林, 车涛, 戴礼云. 多源雪深数据在中国的空间特征评估[J]. *遥感技术与应用*, 2019, 34(06): 1133-1145.
- [83] Solberg, R., Wangensteen, B., Metsämäki, S., et al. GlobSnow snow extent product guide product version 1.0[EB/OL]. (2011-12-19) [2022-01-09]. https://www.globsnow.info/se/archive_v1.2/GlobSnow_SE_product_readme_v1.2.pdf.
- [84] Zhang, G., Luo, W., Chen, W., et al. A robust but variable lake expansion on the Tibetan Plateau[J]. *Science Bulletin*, 2019,
- [85] 叶红, 易桂花, 张廷斌, 等. 2000—2019年青藏高原积雪时空变化[J]. *资源科学*, 2020, 42(12): 2434-2450.
- [86] 王坤鑫, 张寅生, 张腾, 等. 1979—2017年青藏高原色林错流域气候变化分析[J]. *干旱区研究*, 2020, 37(03): 652-662.
- [87] 郑景云, 卞娟娟, 葛全胜, 等. 1981~2010年中国气候区划[J]. *科学通报*, 2013, 58(30): 3088-3099.
- [88] Zhang, G.Q., Xie, H.J., Yao, T.D., et al. Water balance estimates of ten greatest lakes in China using ICESat and Landsat data[J]. *Chinese Science Bulletin*, 2013, 58(31): 3815-3829.
- [89] 万欣, 康世昌, 李延峰, 等. 2007-2011年西藏纳木错流域积雪时空变化及其影响因素分析[J]. *冰川冻土*, 2013, 35(06): 1400-1409.
- [90] 肖瑶, 赵林, 李韧, 等. 唐古拉地区超声雪深传感器 SR-50 监测研究[J]. *应用气象学报*, 2013, 24(03): 342-348.
- [91] 除多, 边巴次仁, 扎珠, 等. SR-50A 超声雪深仪在西藏高原的适用性研究[J]. *高原气象*, 2018, 37(02): 382-393.
- [92] 国家青藏高原科学数据中心. 青藏高原 SRTM DEM 数据集 (2012) [DB/OL]. (2021-08-02) [2021-12-22]. <https://data.tpdc.ac.cn/>. DOI: 10.11888/Geogra.tpdc.270486.

- [93] 沙依然.外力. 基于新一代先进卫星遥感 AMSR2、VIIRS 数据融合积雪监测模型及应用研究[D]. 南京: 南京信息工程大学, 2017.
- [94] Kobayashi, S., Y. Ota, Y. Harada, et al. The JRA-55 Reanalysis: General Specifications and Basic Characteristics[J]. Journal of the Meteorological Society of Japan, 2015, 93(1): 5-48.
- [95] Beaudoin, H. and M. Rodell. GLDAS Noah Land Surface Model L4 3 hourly 0.25 x 0.25 degree V2.1[EB/OL]. (2021-6-18) [2022-02-18]. <https://disc.gsfc.nasa.gov/>. DOI: <https://doi.org/10.5067/E7TYRXPJKWOQ>.
- [96] Hall, D. K. and G. A. Riggs.. MODIS/Terra Snow Cover 5-Min L2 Swath 500m, Version 61[EB/OL]. (2021-12) [2022-02-18]. <https://nsidc.org/>. DOI: https://doi.org/10.5067/MODIS/MOD10_L2.061.
- [97] 张国庆. 青藏高原大于 1 平方公里湖泊数据集 (V3.0) (1970s-2021) [EB/OL]. (2022-04-15) [2022-04-18]. <https://data.tpdc.ac.cn/>. DOI: 10.11888/Hydro.tpdc.270303.
- [98] Chang, A., Gloersen, P., Schmugge, T.J., et al. Microwave emission from snow and glacier ice[J]. Journal of Glaciology, 1976, 16(74): 23-39.
- [99] 李晓锋. 基于 MODIS 数据的高原湖泊冰情遥感监测方法研究[D]. 兰州: 西北师范大学, 2018.
- [100] 高扬, 郝晓华, 和栋材, 等. 基于不同土地覆盖类型 NDSI 阈值优化下的青藏高原积雪判别[J]. 冰川冻土, 2019, 41(05): 1162-1172.
- [101] Salomonson, V.V., Appel, I. Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results[J]. IEEE Transactions on Geoscience Remote Sensing of Environment, 2006, 44(7): 1747-1756.
- [102] Walker, A.E., Goodison, B.E.J.a.O.G. Discrimination of a wet snow cover using passive microwave satellite data[J]. Annals of Glaciology, 1993, 17: 307-311.
- [103] Huang, J., Breheny, P., Ma, S.J.S.S.a.R.J.O.T.I.O.M.S. A selective review of group selection in high-dimensional models[J]. Statistical Science: A Review Journal of the Institute of Mathematical Statistics, 2012, 27(4).
- [104] 王泽坤, 甘甫平, 闫柏琨, 等. 雪深和雪水当量被动微波反演及应用进展[J/OL]. 自然资源遥感: 1-9[2022-05-16]. <http://kns.cnki.net/kcms/detail/10.1759.P.20211214.1718.012.html>.
- [105] 肖雄新, 张廷军. 基于被动微波遥感的积雪深度和雪水当量反演研究进展[J]. 地球科学进展, 2018, 33(06): 590-605.
- [106] 缪文飞, 刘时银, 朱钰, 等. 梅里雪山地区气温和降水的时空分异及海拔效应[J]. 气候变化研究进展: 1-18[2022-05-16]. <http://kns.cnki.net/kcms/detail/11.5368.P.20220223.1323.002.html>.
- [107] 仲桂新. 土地覆盖类型对 MODIS 积雪识别的精度影响分析[D]. 长春: 东北师范大学, 2010.
- [108] 汪关信. 青海湖湖冰特征及其变化[D]. 兰州: 兰州大学, 2020.
- [109] Hersbach, H., Bell, B., Berrisford, P., et al. ERA5 hourly data on single levels from 1979 to present[DB/OL]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), (2018-06-14) [2021-12-11]. <https://cds.climate.copernicus.eu/>. DOI: 10.24381/cds.adbb2d47.

- [110] Cai, Y., Ke, C.Q., Li, X., et al. Variations of lake ice phenology on the Tibetan Plateau from 2001 to 2017 based on MODIS data[J]. *Journal of Geophysical Research: Atmospheres*, 2019, 124(2): 825-843.
- [111] 金亚秋. 星载 SSM/I微波遥感渤海海冰的辐射特征分析[J]. *海洋学报(中文版)*, 1998(03): 40-46.
- [112] 陶安琪, 柯长青, 谢红接, 等. 北极地区夏季太阳辐照度变化研究[J]. *光谱学与光谱分析*, 2012, 32(08): 2037-2042.
- [113] Toure, A.M., Reichle, R.H., Forman, B.A., et al. Assimilation of MODIS snow cover fraction observations into the NASA catchment land surface model[J]. *Remote Sensing*, 2018, 10(2): 316.
- [114] Reichle, R.H., Draper, C.S., Liu, Q., et al. Assessment of MERRA-2 land surface hydrology estimates[J]. *Journal of Climate*, 2017, 30(8): 2937-2960.
- [115] 李永康. 被动微波地表温度降尺度对比研究及空间结构相似性评价[D]. 乌鲁木齐: 新疆农业大学, 2021.
- [116] 王云龙. 中国典型区积雪遥感监测及其时空变化特征研究[D]. 兰州: 兰州大学, 2020.
- [117] 石利娟. 被动微波大气影响分析与校正研究[D]. 北京: 中国科学院大学(中国科学院遥感与数字地球研究所), 2018.
- [118] 周胜男, 车涛, 戴礼云. 基于地面站点类型代表性的积雪遥感产品精度评价[J]. *遥感技术与应用*, 2017, 32(02): 228-237.
- [119] 王健顺. 基于 AMSR2 被动微波数据的青藏高原雪深反演算法研究[D]. 兰州: 兰州大学, 2020.
- [120] 曹晓云, 肖建设, 郝晓华, 等. 2001—2020 年三江源地区积雪日数变化及地形分异 [J/OL]. 干旱区地理: 1-13[2022-05-16]. <http://kns.cnki.net/kcms/detail/65.1103.X.20220214.1143.001.html>.
- [121] 胡晓静. 基于多源遥感数据的雪深反演研究[D]. 兰州: 兰州交通大学, 2021.
- [122] 陈涛. 基于微波遥感的积雪模型算法优化及在新疆和青藏地区的验证[D]. 乌鲁木齐: 新疆大学, 2020.
- [123] 刘晓敬. 被动微波遥感雪深反演混合像元问题研究[J]. *测绘学报*, 2022, 51(02): 313.
- [124] 史翔宇. 基于机器学习回归算法的地震预测研究及其在中国地震科学实验场的应用 [D]. 北京: 中国地震局地震预测研究所, 2021.

附录

附表 1 机器学习模型建模的初始变量

Supplementary table 1 Initial variables for modeling machine learning models

变量类型	变量名称	含义				
地形变量	坡度	平均坡度, 0~90 度间				
	坡向	平均坡向, 0~359.9 度间, 以正北为基准方向顺时针测量				
	海拔	平均海拔				
下垫面类型变量	水体	水体土地覆被类型占比				
	草地	草地土地覆被类型占比				
	裸地	裸地土地覆被类型占比				
亮温变量	10V	10.65 GHz 垂直极化亮温				
	10H	10.65 GHz 水平极化亮温				
	19V	18.7 GHz 垂直极化亮温				
	19H	18.7 GHz 水平极化亮温				
	24V	23.8 GHz 垂直极化亮温				
	24H	23.8 GHz 水平极化亮温				
	37V	36.5 GHz 垂直极化亮温				
	37H	36.5 GHz 水平极化亮温				
	89V	89.0 GHz 垂直极化亮温				
	89H	89.0 GHz 水平极化亮温				
亮温差变量	变量名称	含义	变量名称	含义	变量名称	含义
	10V10H	10V-10H	10H89V	10H-89V	24V24H	24V-24H
	10V19V	10V-19V	10H89H	10H-89H	24V37V	24V-37V
	10V19H	10V-19H	19V19H	19V-19H	24V37H	24V-37H
	10V24V	10V-24V	19V24V	19V-24V	24V89V	24V-89V
	10V24H	10V-24H	19V24H	19V-24H	24V89H	24V-89H
	10V37V	10V-37V	19V37V	19V-37V	24H37V	24H-37V
	10V37H	10V-37H	19V37H	19V-37H	24H37H	24H-37H
	10V89V	10V-89V	19V89V	19V-89V	24H89V	24H-89V
	10V89H	10V-89H	19V89H	19V-89H	24H89H	24H-89H
	10H19V	10H-19V	19H24V	19H-24V	37V37H	37V-37H
	10H19H	10H-19H	19H24H	19H-24H	37V89V	37V-89V
	10H24V	10H-24V	19H37V	19H-37V	37V89H	37V-89H
	10H24H	10H-24H	19H37H	19H-37H	37H89V	37H-89V
	10H37V	10H-37V	19H89V	19H-89V	37H89H	37H-89H
	10H37H	10H-37H	19H89H	19H-89H	89V89H	89V-89H

附表 2 基于 LASSO 惩罚方法得到的初始变量系数
 Supplementary table 2 Coefficient of initial variables based on LASSO penalty method

变量	系数	变量	系数	变量	系数	变量	系数
坡度	0.509	10V10H	0.169	10H89V	0	24V24H	0
坡向	-0.199	10V19V	0	10H89H	0	24V37V	0
水体	0.276	10V19H	0.096	19V19H	0	24V37H	0
草地	0.044	10V24V	0	19V24V	0.174	24V89V	0
裸地	0	10V24H	0	19V24H	0	24V89H	0
海拔	0.237	10V37V	0.060	19V37V	0.006	24H37V	0
10V	0	10V37H	0	19V37H	0	24H37H	0
10H	0	10V89V	0	19V89V	0	24H89V	0
19V	0	10V89H	0	19V89H	0	24H89H	0
19H	0	10H19V	0	19H24V	0	37V37H	-0.158
24V	0	10H19H	0	19H24H	-0.145	37V89V	0
24H	0	10H24V	0	19H37V	0	37V89H	0
37V	-0.146	10H24H	0	19H37H	0	37H89V	0
37H	0	10H37V	0	19H89V	0	37H89H	0.162
89V	-0.178	10H37H	0	19H89H	0	89V89H	0
89H	-0.017						

附表 3 基于 SCAD 惩罚方法得到的初始变量系数
 Supplementary table 3 Coefficient of initial variables based on SCAD penalty method

变量	系数	变量	系数	变量	系数	变量	系数
坡度	0.522	10V10H	0.157	10H89V	0	24V24H	0
坡向	-0.239	10V19V	0	10H89H	0	24V37V	0
水体	0.332	10V19H	0	19V19H	0	24V37H	0
草地	0	10V24V	0.101	19V24V	0.132	24V89V	0
裸地	0.034	10V24H	0	19V24H	0.091	24V89H	0
海拔	0.228	10V37V	0	19V37V	0.050	24H37V	0
10V	0	10V37H	0	19V37H	0	24H37H	0
10H	0	10V89V	0	19V89V	0	24H89V	0
19V	0	10V89H	0	19V89H	0	24H89H	0
19H	0	10H19V	0	19H24V	0	37V37H	-0.218
24V	0	10H19H	0	19H24H	-0.238	37V89V	0
24H	0	10H24V	0	19H37V	0	37V89H	0
37V	-0.083	10H24H	0	19H37H	0	37H89V	0
37H	0	10H37V	0	19H89V	0	37H89H	0
89V	-0.361	10H37H	0	19H89H	0	89V89H	0.049
89H	0						

附表 4 基于 MCP 惩罚方法得到的初始变量系数
 Supplementary table 4 Coefficient of initial variables based on MCP penalty method

变量	系数	变量	系数	变量	系数	变量	系数
坡度	0.523	10V10H	0.157	10H89V	0	24V24H	0
坡向	-0.241	10V19V	0	10H89H	0	24V37V	0
水体	0.326	10V19H	0	19V19H	0	24V37H	0
草地	-0.036	10V24V	0.101	19V24V	0.132	24V89V	0
裸地	0	10V24H	0	19V24H	0.192	24V89H	0
海拔	0.228	10V37V	0	19V37V	0	24H37V	0
10V	0	10V37H	0	19V37H	0	24H37H	0.119
10H	0	10V89V	0	19V89V	0	24H89V	0
19V	-0.061	10V89H	0	19V89H	0	24H89H	0
19H	0	10H19V	0	19H24V	0	37V37H	-0.339
24V	0	10H19H	0	19H24H	-0.238	37V89V	0
24H	0	10H24V	0	19H37V	0	37V89H	0
37V	0	10H24H	0	19H37H	0	37H89V	0
37H	0	10H37V	0	19H89V	0	37H89H	0
89V	-0.361	10H37H	0	19H89H	0	89V89H	0.049
89H	0						

附表 5 基于随机森林特征重要度方法得到的初始变量特征重要度
 Supplementary table 5 Feature importance of initial variables based on random forest feature importance method

变量	系数	变量	系数	变量	系数	变量	系数
坡度	0.021	10V10H	0.020	10H89V	0.012	24V24H	0.041
坡向	0.021	10V19V	0.010	10H89H	0.012	24V37V	0.013
水体	0.003	10V19H	0.014	19V19H	0.036	24V37H	0.011
草地	0.012	10V24V	0.011	19V24V	0.012	24V89V	0.015
裸地	0.016	10V24H	0.012	19V24H	0.029	24V89H	0.016
海拔	0.051	10V37V	0.010	19V37V	0.019	24H37V	0.031
10V	0.013	10V37H	0.009	19V37H	0.013	24H37H	0.008
10H	0.014	10V89V	0.012	19V89V	0.013	24H89V	0.012
19V	0.012	10V89H	0.019	19V89H	0.016	24H89H	0.010
19H	0.013	10H19V	0.20	19H24V	0.024	37V37H	0.059
24V	0.013	10H19H	0.009	19H24H	0.013	37V89V	0.008
24H	0.012	10H24V	0.015	19H37V	0.024	37V89H	0.010
37V	0.018	10H24H	0.007	19H37H	0.012	37H89V	0.008
37H	0.021	10H37V	0.013	19H89V	0.011	37H89H	0.010
89V	0.025	10H37H	0.009	19H89H	0.013	89V89H	0.011
89H	0.026						

在学期间的研究成果

一、发表论文

1. 邬俊飞, 姚檀栋, 戴玉凤, 陈文锋. 色林错、纳木错湖区被动微波雪深反演算法评估[J]. 遥感技术与应用. (待刊, 2022年第37卷6期)

二、参与课题

1. 国家自然科学基金委青年科学基金项目“青藏高原典型湖泊的湖泊效应降水及其对环境的影响”(项目号: 41801049)
2. 科技部“第二次青藏高原综合科学考察研究”, “亚洲水塔动态变化与影响”任务, “冰川、积雪、冻土变化与影响及应对”专题(专题号: 2019QZKK0201)

致 谢

日月既往，不可追复，行文至此，落笔为终。山色横侵遮不住，萃英山下好读书，离家千余里至金城求学，三年榆中，四年城关，兰大见证了的成长。

饮水思其源，学成念吾师。三生有幸，能够师从姚檀栋院士，本文从选题到撰写，姚老师都给予了悉心指导。姚老师渊博的学识与严谨、包容的学术态度以及儒雅随和的个人魅力都让我印象深刻、高山仰止。在今后的工作和生活中，我将牢记您的教诲，不负韶华。

学贵为师，亦贵为友。感谢师门的戴玉凤师姐和陈文锋师兄，像自己的哥哥姐姐，在我迷茫无措时，给我指导，同时在我做错事时，严厉批评。“世事洞明皆学问，人情练达即文章”，师兄师姐教导的人生道理，让自己少走了许多弯路。还要感谢朱飞师兄对我野外和生活的帮助，巨细无遗地为我讲解野外事项，帮我解决北京交流期间的住宿问题等，十分感谢！

感谢青藏所王伟财和汶林科老师对我的指导与帮助，感谢旺扎老师对我野外工作的支持！

特别感谢杜悦同学对我硕士论文的帮助，特别感谢舍友兼同门雷鹏嗣同学在研究生期间给予的支持与帮助！

感谢观云楼 999 自习室的丁国强、姬盼盼师兄、陈曼丽师姐、尚雪雪和唐倩师妹、逸凡和子恒师弟，舍友李垚和徐绍源，球友云博、向奕、兰兰、启航和培研等，以及我的好兄弟陈凯文、胡耀辉、童志、张彧圣、黄河、吴淼、万梓涵、程禹瑞、谢作鹏、涂建新、付雨晖和饶毅等，一路来给予的关心与鼓励，祝愿大家未来一帆风顺！

感谢家人多年来对我学业的支持，希望他们身体健康！

感谢兰大和青藏所，给了我不一样的诗和远方！