www.sciengine.com/CSD
ISSN 2096-2223
CN 11-6035/N

OPEN ACCESS
CC BY

文献 CSTR：
32001.14.11-6035.ncdc.2024.0020.zh
文献 DOI：
10.11922/11-6035.ncdc.2024.0020.zh
数据 DOI：
10.57760/sciencedb.13892

文献分类：地球科学

收稿日期：2024-06-11
开放同评：2024-08-14
录用日期：2024-11-18
发表日期：2025-02-24

* 论文通信作者
肖建设：xiaojianshe@126.com

专题 青藏高原雪、冰和环境数据集

青藏高原逐月多年平均积雪密度格点数据集

曹晓云[1,2]，史飞飞[1,2]，刘致远[1,2]，张娟[1,2,3]，李弘毅[3]，肖建设[4*]

1. 青海省气象科学研究所，西宁 810001
2. 青海省防灾减灾重点实验室，西宁 810001
3. 中国科学院西北生态环境资源研究院，兰州 730000
4. 成都信息工程大学电子工程学院，成都 610225

摘要：积雪密度是表征积雪特性的一个重要参量，也是将积雪深度转换成雪水当量的重要指标，在山区雪水资源估算、融雪洪水等水资源管理、自然灾害预报以及气候研究等方面具有重要作用。以1960–2020年青藏高原132个逐日国家气象站资料、中国区域地面气象要素驱动数据集、卫星融合雪深数据集为主要数据源，分不同地表类型比较几种机器学习模型在积雪密度模拟中的性能，选取最优模型，综合地面、卫星和再分析资料，制作了青藏高原逐月积雪密度数据集。与132个青藏高原国家气象观测站的逐月多年平均积雪密度数据进行精度检验发现，逐月多年平均积雪密度的平均均方根误差为0.019 g/cm^3，平均相对误差为11.88%，表明数据集具有较高精度。本数据集将为青藏高原水资源评估、水文过程模拟等提供数据支撑。

关键词：青藏高原；积雪密度；气象站点；机器学习

数据库（集）基本信息简介

数据库（集）名称	青藏高原逐月多年平均积雪密度格点数据集
数据作者	曹晓云，史飞飞，刘致远，张娟，李弘毅，肖建设
数据通信作者	肖建设（xiaojianshe@126.com）
数据时间范围	1–12月逐月多年平均值
地理区域	青藏高原，地理范围包括（25°–40°N，73°–105°E）
空间分辨率	0.05°×0.05°
数据量	11.2 MB
数据格式	*. tif
数据服务系统网址	https://doi.org/10.57760/sciencedb.13892 http://www.ncdc.ac.cn/portal/metadata/af7bbd77-93ac-4d77-966d-4104d31d9fee
基金项目	长江流域开放基金项目（CJLY2022Y12）；国家自然科学基金（U22A20556，U21A2021，41761078）；青海省科技厅科技计划项目（2024-ZJ-740）；第二次青藏高原综合科学考察研究项目（2019QZKK0105）；青海省防灾减灾重点实验室项目（QFZ-2021-Z01）。
数据库（集）组成	数据集共包括有12个格式为tif的逐月积雪密度数据。

引　言

积雪密度是表征积雪特性的一个重要参量，影响雪层的热学、力学和光学特性，也是将积雪深度转换成雪水当量的重要指标[1-4]，在山区雪水资源估算、融雪洪水、雪崩、暴风雪等水资源管理、自然灾害预报以及气候研究等方面起着至关重要的作用[3,5-7]。目前，被动微波遥感雪水当量产品换算过程中多使用固定积雪密度参数，普遍认为 0.24 g/cm^3 是合理的全球平均积雪密度[8-11]。随着对积雪深度估算精度的提高，积雪密度估算误差在雪水当量评估中的误差越来越突出[1,3,8-9]。如何改善积雪密度估计成为一个科学问题。然而，由于积雪密度与其影响变量之间的弱相关性导致影响变量对积雪密度估计的解释能力相对有限，实现非常高的积雪密度估计精度具有非常大的挑战性[7]。

研究表明积雪密度存在较大的时空差异性，尤其是在高海拔和地形复杂的山区[7,10]，影响积雪密度较大的积雪演化物理过程主要包括新雪的形成、雪的压实、液态水的冻结、蒸汽运动、雪的变质作用等[7,9,12-14]。同时，积雪密度也会受到地形和地表类型等要素的影响[7,9,15]。在海拔较高和辐射较少的斜坡上，雪密度较低，海拔每增加 100 米，雪密度就会下降约 0.006 g/cm^3；平地开阔区平均积雪密度比林区高 8%–13%[1,6]。目前有一些研究通过建立物理模型来模拟积雪密度[16-18]，如能量和质量平衡耦合模型 ISNOBAL，然而物理模型过程复杂，一般仅仅适合小区域积雪密度空间化。还有一些研究利用雪深、雪龄、气象要素和地形变量等不同相关要素与积雪密度建立了统计模型进而估计积雪密度[6,19]，如多元线性回归分析。然而，简单的统计模型没有考虑积雪密度和多个影响要素之间复杂的非线性关系，且这些模型大多根据稀少且短期的观测点数据独立构建，忽略了积雪密度的时空异质性以及与影响要素的关系。近年来，随机森林（Random Forest，RF）、极端梯度提升算法（Extreme Gradient Boosting，XGBoost）、梯度提升回归树模型（Gradient Boosting Regression Tree，GBRT）、轻量级梯度提升机算法（Light Gradient Boosting Machine，LightGBM）、人工神经网络（Artificial Neural Network，ANN）和地理时间加权的神经网络（Geographically and Temporally Weighted Neural Network，GTWNN）等机器学习算法已成功应用于雪密度估算中[6-7,20]，结果不仅具有较高的精度，还提供了每个变量对雪密度估计的详细特征重要性指数，可以增强对各时期雪密度空间分布的主要驱动变量的理解。然而，这些模型都很复杂，需要大量的详细特征信息，且运行需要大量的计算时间。到目前为止，很少有研究综合比较几种机器学习模型在雪密度模拟方面的性能，并分不同地表覆盖类型进行建模。

青藏高原平均海拔约为 4500 m，有“世界屋脊”和“第三极”之称，积雪作为青藏高原冰冻圈的重要组成部分，是重要的淡水资源[11,21-22]。青藏高原积雪具有降雪频繁，局地性强，深度浅，斑块化显著，消融快等特点，稳定积雪和瞬时性积雪同时存在，大部分地区每年存在多次降雪融雪过程，每个雪周期持续几到十几天，并没有表现出典型的致密化。此外，研究发现青藏高原积雪密度的年际变化不明显，但存在独特的季节特征和海拔差异，而且积雪维持时间可以反映每个降雪和融雪周期的开始和状态，在积雪密度估算中表现更优[11,18]。然而，目前没有开发出用于估算青藏高原积雪密度的稳健模型，仅提出了 0.15 g/cm^3 的区域平均值来计算雪水当量[23]，而且，对高原积雪特性的认知不足和地面样本的缺乏，青藏高原积雪密度估计并没有得到本质上的改进。因此，迫切需要开发一种实用的估算方法。

基于此，本数据集以 1960–2020 年青藏高原 132 个逐日国家气象站资料、2000–2018 年的中国区域地面气象要素驱动数据集、卫星融合雪深数据集为主要数据源，将高原分草地、林地和裸地 3

种地表类型，比较 RF、GBRT、XGBoost、LightGBM 四种机器学习模型在积雪密度模拟中的性能，对每种地表类型选取最优模型，综合地面、卫星和再分析资料，实现青藏高原逐月积雪密度制图。本数据集将有效提高对青藏高原积雪密度空间分布的认识，为青藏高原水资源评估、水文过程模拟等提供数据支撑。

1 数据采集和处理方法

1.1 数据获取与预处理

1.1.1 气象站点数据

选取 1960–2020 年青藏高原地区 132 个国家气象站的逐日雪深、雪压、气温和风速数据。其中，积雪密度由逐日雪压和雪深进行计算。青藏高原气象台站分布如图 1 所示，数据来源于中国气象局综合气象信息共享平台（http://10.181.89.55/cimissapiweb/），经过严格的质量控制，准确性及完整性满足科学研究需求。

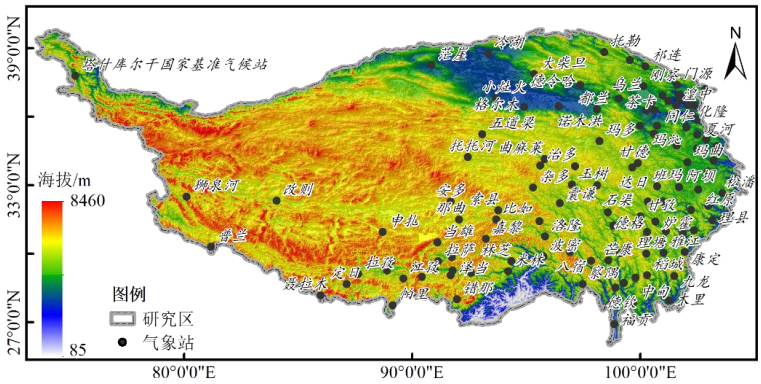


图 1　青藏高原气象站点空间分布

Figure 1　Spatial distribution of meteorological stations on the Tibetan Plateau

1.1.2 基于卫星的多源雪深融合数据

采用 2000–2018 年青藏高原 0.05°逐日积雪深度数据集[24]。该积雪深度降尺度数据集来源于积雪概率数据和中国雪深长时间序列数据集的融合结果，均方根误差由原产品的 2.15 cm 减少到了 1.54 cm。对其计算逐日的积雪维持日数，并处理成逐月多年平均雪深、积雪维持日数数据用于模型输入，该数据可从时空三极环境大数据平台免费获得（http://poles.tpdc.ac.cn/zh-hans/data/0515ce19-5a69-4f86-822b-330aa11e2a28）。

1.1.3 中国区域地面气象要素驱动数据集

采用 1979–2018 年的中国区域地面气象要素驱动数据集[25]。该数据集包括近地面气温、近地面气压、近地面空气比湿、近地面全风速、地面向下短波辐射、地面向下长波辐射、地面降水率共 7 个要素。数据为 NETCDF 格式，时间分辨率为 3 小时，水平空间分辨率为 0.1°，其精度介于气象局观测数据和卫星遥感数据之间，好于国际上已有再分析数据的精度。为与积雪遥感数据进行叠加分析，将其重采样和重投影与雪深一致的空间分辨率和投影，提取气温、风速数据，并将其处理成逐月多年平均值用于模型输入，该数据可从时空三极环境大数据平台免费获得

（http://poles.tpdc.ac.cn/zh-hans/data/8028b944-daaa-4511-8769-965612652c49/）。

1.1.4 基础地理数据

（1）DEM 数据

DEM 数据采用 SRTM（Shuttle Radar Topography Mission）数据集，数据版本为 V003，来源于美国地质勘探局 USGS（United States Geological Survey）（https://lpdaac.usgs.gov/products/srtmgl1v003/），空间分辨率为 90 m。为与积雪遥感数据进行叠加分析，将其重采样和重投影与雪深一致的空间分辨率和投影，并利用 ArcGIS 软件生成海拔和坡向分布图。

（2）地表覆盖类型数据

地表覆盖类型数据选用 2020 年 MODIS IGBP（International Geosphere-Biosphere Program）地表分类数据，来源于美国 LPDAAC（Land Process Distributed Active Archive Center）（http://lpdaacweb.nascom.nasa.gov/data/search.html）。空间分辨率为 500 m，为与积雪遥感数据进行叠加分析，将其重采样和重投影与雪深一致的空间分辨率和投影，并进行重分类，分类成林地、草地、裸地等土地类型，湖泊将其掩膜，不考虑冬季冰面上的积雪（图 2），构建不同土地覆盖类型的积雪密度回归模型。

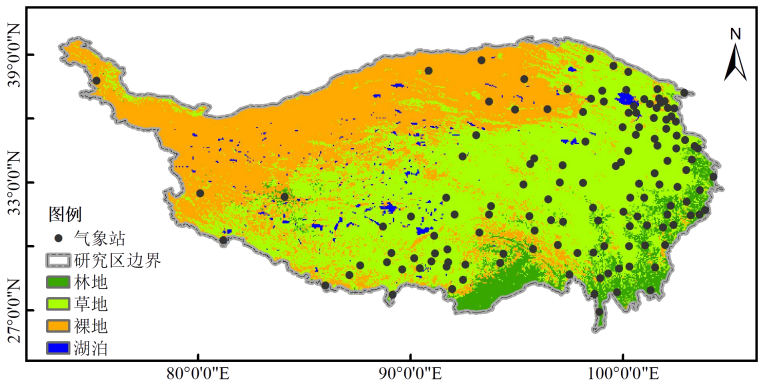


图 2 青藏高原土地覆盖类型空间分布

Figure 2 Spatial distribution of land cover types on the Tibetan Plateau

1.2 数据处理方法

1.2.1 数据处理流程

本数据集的制作主要分为以下几个步骤：数据获取及预处理、站点的积雪密度反演模型构建、格点的积雪密度估算和精度验证，具体流程如图 3 所示。首先，基于气象站点的逐日雪深、雪压数据计算得到每个站点逐日的积雪密度。之后，选取基于站点的逐日雪深、积雪维持日数、平均气温、风速、积雪密度、海拔、坡度、坡向、月份和经纬度为特征要素，按照 2:8 将样本集随机划分为测试集和训练集。选用 RF、XGBoost、GBRT 和 LightGBM 为基础模型，采用网格搜索结合 5 折交叉验证的方法进行超参数优化，并采用十折交叉验证算法来比较各模型对林地、草地和裸地三种地表类型积雪密度的估算性能，选取每种地表类型积雪密度反演的最优模型。随后，选取基于遥感的逐月多年平均雪深和积雪维持日数，气象驱动数据中的逐月多年平均气温、风速，海拔、坡度、坡向、月份和经纬度为特征要素，分不同地表类型用最优模型进行逐网格的积雪密度反演。最后，选取 1980–2010 年青藏高原 132 个国家气象观测站的逐月多年平均积雪密度数据为“真值”进行精度验

证，最终输出青藏高原逐月多年平均积雪密度格点数据集。

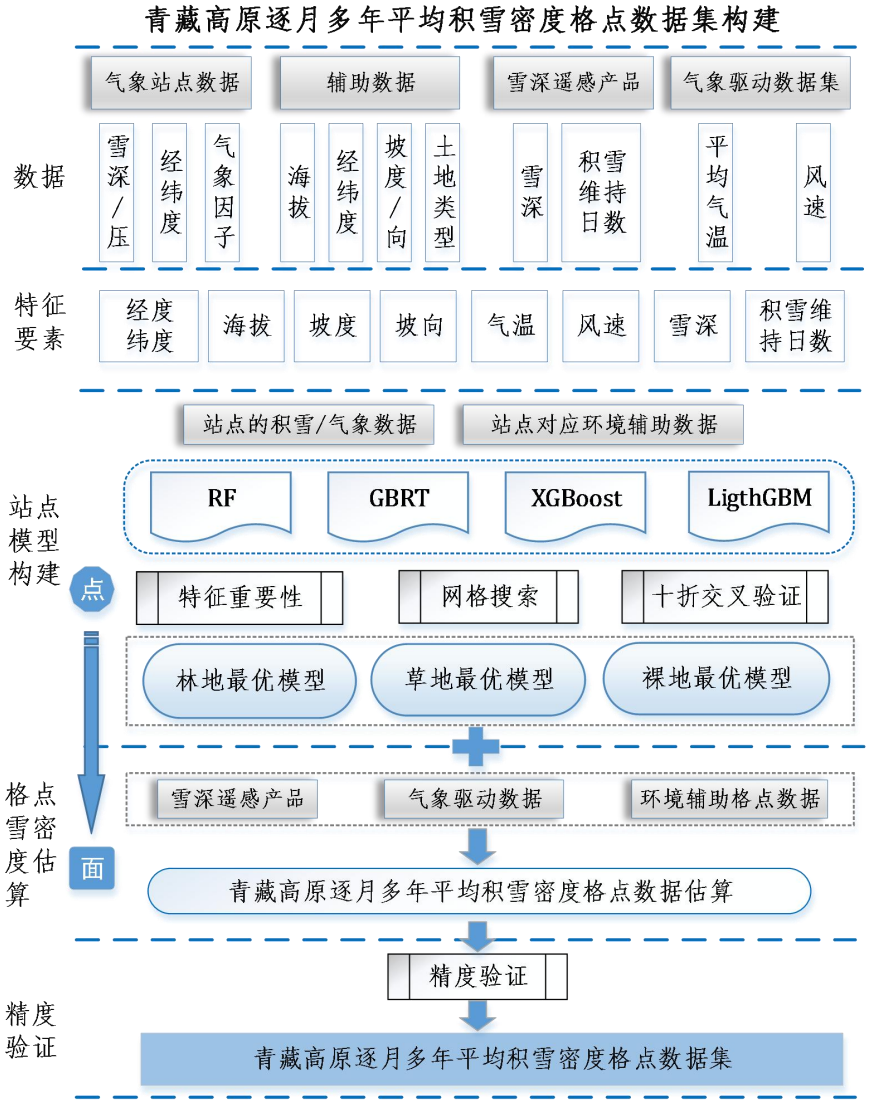


图 3　数据处理流程

Figure 3　Data processing flowchart

1.2.2　气象站点积雪密度计算

气象站点的积雪密度可以使用雪深和雪压数据进行计算，公式如下：

$$\rho_s = \frac{SP}{SD} \tag{1}$$

式中，ρ_s为积雪密度（g/cm^3）；SP 为雪压（g/cm^2）；SD 为积雪深度（cm）。

1.2.3　建模方法

使用 sklearn 库中的样本分割方法，按照 2:8 将样本集随机划分为测试集和训练集。选用 RF、XGBoost、GBRT 和 LightGBM 为基础模型，采用网格搜索结合 5 折交叉验证的方法进行超参数优化，并采用十折交叉验证算法来比较各模型对林地、草地和裸地三种地表类型积雪密度的估算性能，选取最优模型。

（1）随机森林（RF）

RF 属于一种监督学习的算法，其本质是从样本中采样选出 n 个样本，从所有属性中随机选择 k 个属性，选择最佳分割属性作为节点建立 CART 决策树。重复以上步骤 m 次，建立 m 棵 CART 决策树，进而形成随机森林，通过投票表决结果，决定最终预测结果[26-27]。本文在训练模型时，主要通过调整决策树的数量、决策树的最大深度、决定节点是否分裂的最小样本数、节点拆分时的最大特征数量等重要参数来训练模型。

（2）梯度提升回归树算法（GBRT）

GBRT 是一种基于学习器为回归树的组合算法，将梯度提升与回归树相结合，其中每一棵新的回归树拟合学习的都是基于前一棵回归树学习后的残差，用梯度提升的方法不断降低残差，对残差的学习也使得回归树变成了残差树，即以损失函数的梯度下降方向为基础建立新的回归树，最终的输出结果就是每棵回归树输出结果的累加，从而使结果得到改进[26-27]。本文在训练模型时，主要通过调整学习器的数量、学习率、决策树最大深度、控制决策树节点分裂所需的最小样本数等重要参数来训练模型。

（3）极端梯度提升算（XGBoost）

XGBoost 通过构建多个弱学习器以梯度下降的方式迭代优化损失函数，以逐步提升模型的预测性能。损失函数包括衡量预测值与真实值误差的数据拟合项和用来惩罚模型复杂度的正则化项。通过最小化这两项的组合，XGBoost 可以实现在训练集上的高效拟合和泛化[27-28]。本文在训练模型时，主要通过调整弱学习器的数量、学习率、决策树最大深度、每棵树的样本采样比例、控制树的分类等参数来提高模型泛化能力并防止过拟合。

（4）轻量级梯度提升机算法（LightGBM）

LightGBM 是一种基于梯度提升框架的高效、可扩展的机器学习算法，它是 XGBoost 的后续版本，继承并优化了 XGBoost 的一系列优势，同时引入了新的优化策略，如基于直方图的决策树算法、单边梯度采样、直接支持类别特征等，从而实现了更快的训练速度、更低的内存消耗、更高的准确率，并支持分布式处理海量数据[26,29]。本文在训练模型时，主要通过调整设置提升类型、叶子节点数、学习率、决策树最大深度等重要参数来训练模型。

1.3 精度验证

本文使用相关系数（Pearson product-moment correlation coefficient, R）、均方根误差（Root Mean Square Error, RMSE）、平均绝对误差（Mean Absolute Error, MAE）和平均相对误差（Mean Relative Error, MRE）作为模型精度评价因子。

其中，x_i和y_i表示像元尺度的计算值及对应的真值；$\bar{x_i}$和$\bar{y_i}$表示时间序列的平均值；n为样本量。$R(x,y)$越接近 1，表示精度越高，$RMSE(x,y)$、$MAE(x,y)$、$MRE(x,y)$越小表示精度越高。

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

$$RMSE(x,y) = \sqrt{\left(\frac{\sum_{i=1}^{n}(x_i-y_i)^2}{n}\right)} \tag{3}$$

$$MAE(x,y) = \frac{\sum_{i=1}^{n}|x_i-y_i|}{n} \tag{4}$$

$$GMRE(x,y) = \frac{x_i-y_i}{y_i} \times 100\% \tag{5}$$

表 1 给出了基于站点的不同土地类型最优模型测试集的 R、RMSE 和 MAE。其中，林地区域的最优模型为 GBRT，其中 R、RMSE 和 MAE 分别为 0.78、0.039 和 0.029；草地区域的最优模型为 XGBoost，其中 R、RMSE 和 MAE 分别为 0.57、0.045 和 0.033；裸地区域的最优模型也为 XGBoost，其中 R、RMSE 和 MAE 分别为 0.56、0.040 和 0.029。这是由于青藏高原积雪具有降雪频繁，局地性强，深度浅，斑块化显著，消融快等特点，难以维持积雪覆盖，稳定积雪和瞬时性积雪同时存在。高原上林区占比少（7%）且主要集中在南部地区，样本的代表性好，而草地和裸地的占比分别为 53% 和 38%。当样本数量过大时，模型可能会因为无法处理所有的特征而倾向于欠拟合。因此，样本量的大小和代表性是造成林区的精度高于草地和裸地的精度主要原因。此外，由于雪密度与其影响变量之间的弱相关性导致影响变量对雪密度估计的解释能力相对有限，实现非常高的雪密度估计精度具有非常大的挑战性。但相比于目前青藏高原积雪密度估算用 0.15 g/cm^3 的区域平均值[23]，本研究构建的模型表现出较高的精度和较低的偏差。

表 1　不同土地覆盖类型的最优模型及精度评价

Table 1　Optimal models and accuracy evaluation for different land cover types

类型	最优模型	R	RMSE（g/cm^3）	MAE（g/cm^3）
林地	GBRT	0.78	0.039	0.029
草地	XGBoost	0.57	0.045	0.033
裸地	XGBoost	0.56	0.040	0.029

2　数据样本描述

青藏高原逐月多年平均积雪密度格点数据集共由 12 个逐月的栅格数据文件组成，数据格式为 tif，空间分辨率为 0.05°×0.05°，等经纬度投影，单位为 g/cm^3，覆盖范围为青藏高原。文件命名规则为 TP_snow density grid_MM.tif，其中，MM 为月份，例如 1 月为 01。图 4 为青藏高原逐月多年平均积雪密度格点数据示例。可以看出，青藏高原积雪密度空间分布具有显著的空间分布差异，总体由东南向西北逐渐增大，喀喇昆仑山、昆仑山、喜马拉雅山、念青唐古拉山、巴颜喀拉山、祁连山等高海拔山脉地区是积雪密度高值区，一般都超过 0.15 g/cm^3，喀喇昆仑山和喜马拉雅山的部分地区甚至超过 0.18 g/cm^3；而低值区主要分布在川西藏东地区、藏南山地、三江源地区、柴达木盆地以及青东祁连山等地区。

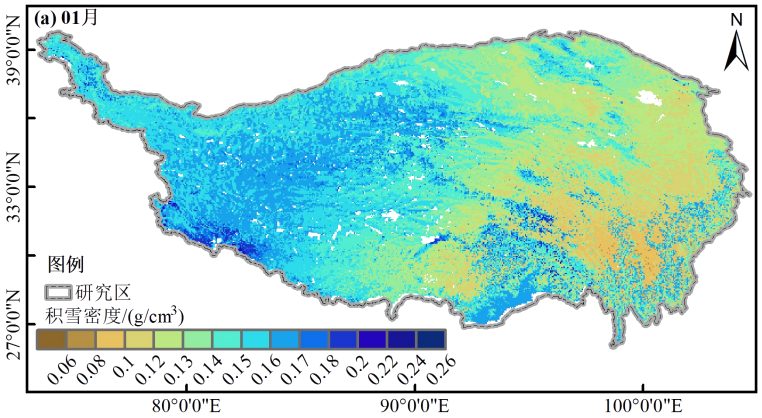


图 4　青藏高原逐月多年平均积雪密度空间分布示意图

Figure 4　Spatial distribution of the monthly multi-year average snow density on the Tibetan Plateau

3 数据质量控制和评估

积雪密度格点数据精度验证，以 1980–2010 年青藏高原 132 个国家气象观测站的逐月多年平均积雪密度数据为“真值”，逐月对本文制作的积雪密度格点数据集进行精度检验（图 5），结果表明本数据集与台站观测相关程度较高，其相关系数为 0.435–0.641，平均均方根误差为 0.019，平均相对误差为 11.88%。表明本数据集能较好地刻画青藏高原积雪密度的实际状况，对积雪季的积雪密度表征效果更优。

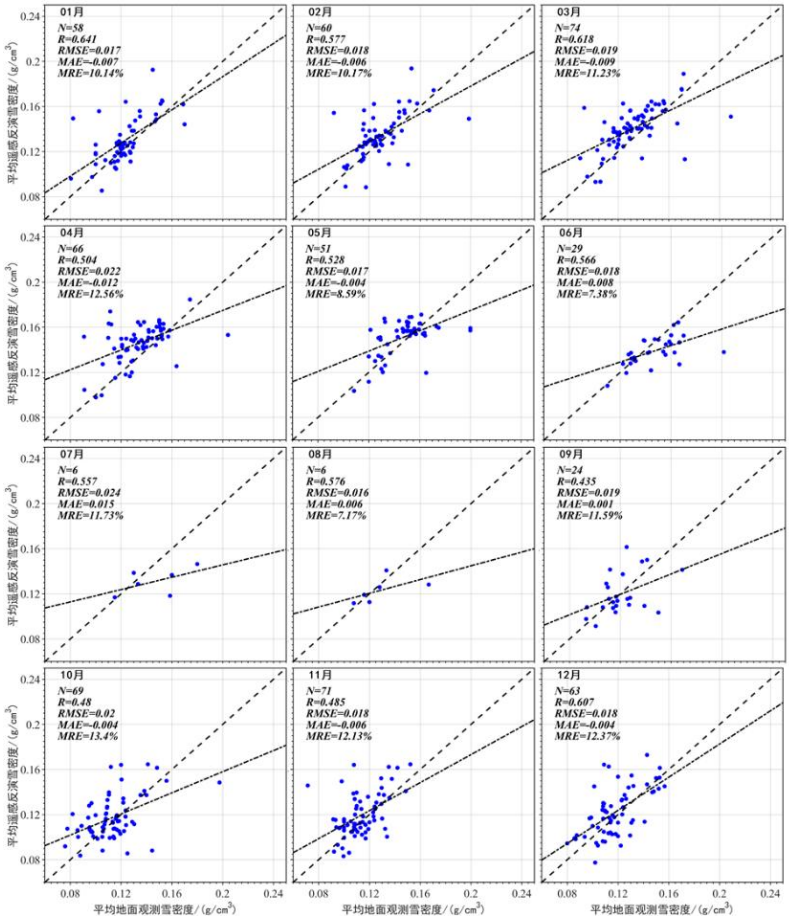


图 5　青藏高原逐月多年平均积雪密度格点数据集精度验证

Figure 5　Accuracy validation of the dataset of monthly multi-year average snow density grids on the Tibetan Plateau

4 数据价值

本数据集在生产过程中结合了地面观测数据、卫星遥感数据和再分析数据等多种数据源，具有较高的准确性和可靠性，能够更准确地反映积雪密度的季节时空变化特征。为研究积雪的物理特性、时空分布规律提供了基础数据，有助于理解积雪的形成、积累和消融过程，以及这些过程对气候、水文循环和生态系统的影响。同时，为青藏高原雪水资源评估、融雪径流预测、水资源管理策略制定等提供数据支撑，可以为气象预报、农业生产、交通运输等多个领域提供数据参考。

5 数据使用方法和建议

本数据集是基于机器学习方法和气象站点、遥感产品、再分析资料数据构建的模型反演结果，数据整体精度较高。本数据集解压后的格式为 tif，数据可使用专业的 GIS 软件打开和编辑再加工；其时间序列涵盖了青藏高原逐月的多年平均积雪密度的时空分布信息，可为青藏高原水资源评估、水文过程模拟等提供数据支撑。

数据作者分工职责

曹晓云（1993—），女，青海省西宁市人，工程师，研究方向为青藏高原气候与环境研究。主要承担工作：论文撰写，数据制作与精度验证。

史飞飞（1991—），男，陕西省宝鸡市人，工程师，研究方向为生态气象监测评估。主要承担工作：数据整理，数据精度验证。

刘致远（1993—），男，河南省三门峡市人，工程师，研究方向为生态气象监测评估。主要承担工作：数据整理与制图。

张娟（1979—），女，河北省保定市人，正研级高级工程师，研究方向为积雪监测评估。主要承担工作：论文修改指导。

李弘毅（1981—），男，四川省南充市人，研究员，研究方向为积雪水文。主要承担工作：论文修改指导。

肖建设（1979—），男，河南省永城市人，正研级高级工程师，研究方向为生态环境和气候变化研究。主要承担工作：论文修改，组织与指导。

参考文献

[1] ZHONG X Y, ZHANG T J, WANG K. Snow density climatology across the former USSR[J]. The Cryosphere, 2014, 8(2): 785-799. DOI: 10.5194/tc-8-785-2014.

[2] BRULAND O, FÆREVÅG Å, STEINSLAND I, et al. Weather SDM: estimating snow density with high precision using snow depth and local climate[J]. Hydrology Research, 2015, 46: 494-506. DOI: 10.2166/nh.2015.059.

[3] WETLAUFER K, HENDRIKX J, MARSHALL L. Spatial Heterogeneity of Snow Density and Its Influence on Snow Water Equivalence Estimates in a Large Mountainous Basin[J]. Hydrology, 2016, 3(1): 3. DOI: 10.3390/hydrology3010003.

[4] KRAAIJENBRINK P D A, STIGTER E E, YAO T D, et al. Climate change decisive for Asia's snow meltwater supply[J]. Nature Climate Change, 2021, 11(7): 591-597. DOI: 10.1038/s41558-021-01074-x.

[5] LOPEZ-MORENO J I, FASSNACHT S R, HEATH J T, et al. Small scale spatial variability of snow density and depth over complex alpine terrain: Implications for estimating snow water equivalent[J]. Advances in water resources, 2013, 55(5): 40-52. DOI: 10.1016/j.advwatres.2012.08.010.

[6] FENG T, ZHU S Z, HUANG F R, et al. Spatial Variability of Snow Density and Its Estimation in

Different Periods of Snow Season in the Middle Tianshan Mountains, China[J]. Hydrological Processes, 2022, 36(8). DOI: 10.1002/hyp.14644.

[7] WANG H D, ZHANG X L, XIAO P F, et al. Towards large-scale daily snow density mapping with spatiotemporally aware model and multi-source data[J]. The Cryosphere, 2023, 17(1): 33-50. DOI: 10.5194/tc-17-33-2023.

[8] MOLOTCH N P, COLEE M T, BALES R C, et al. Estimating the spatial distribution of snow water equivalent in an alpine basin using binary regression tree models: the impact of digital elevation data and independent variable selection[J]. Hydrological processes, 2005, 19(7): 1459-1479. DOI: 10.1002/hyp.5586.

[9] 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(6): 590-605. DOI: 10.11867/j.issn.1001-8166.2018.06.0590. [XIAO X X, ZHANG T J. Passive Microwave Remote Sensing of Snow Depth and Snow Water Equivalent: Overview[J]. Advances in Earth Science, 2018, 33(6):590-605. DOI: 10.11867/j.issn.1001-8166.2018.06.0590.]

[10] YANG J W, JIANG L M, LEMMETYINEN J, et al. Validation of remotely sensed estimates of snow water equivalent using multiple reference datasets from the middle and high latitudes of China[J]. Journal of Hydrology, 2020, 590: 125499. DOI: 10.1016/j.jhydrol.2020.125499.

[11] ZHAO W Y, MU C C, WU X D, et al. Spatio-Temporal Characteristics and Differences in Snow Density between the Tibet Plateau and the Arctic[J]. Remote Sensing, 2023, 15: 3976. DOI: 10.3390/rs15163976.

[12] 魏玥, 陈蜀江, 陈霞. 新疆北部地区季节性积雪密度变化特征分析[J]. 冰川冻土, 2010, 32(3): 519-523. DOI: 1000-0240(2010)03-0519-05. [WEI Y, CHEN S J, CHEN X. Analysis on the Seasonal Snow Density Change Characteristics of North Xinjiang[J]. Journal of Glaciology and Geocryology, 2010, 32(3): 519-523. DOI: 1000-0240(2010)03-0519-05.]

[13] 李小兰, 姜逢清, 王少平, 等. 2011-2013 年乌鲁木齐城-郊冬季积雪深度与密度空间分布特征[J]. 冰川冻土, 2015, 37(5): 1168-1177. DOI: 10.7522/j.isnn.1000-0240.2015.0131. [LI X L, JIANG F Q, WANG S P, et al. Spatial distribution of winter snow cover depth and density in Ürümqi City during 2011-2013[J]. Journal of Glaciology and Geocryology, 2015, 37(5): 1168-1177. DOI: 10.7522/j.isnn.1000-0240.2015.0131.]

[14] 刘宝河, 左合君, 董智, 等. 一次降雪的积雪密实化过程研究[J]. 干旱区资源与环境, 2017, 31(1): 178-184. DOI: 10.13448/j.cnki.jalre.2017.030. [LIU B H, ZUO H J, DONG Z, et al. Study on snow densification process after a snowfall[J]. Journal of Arid Land Resources and Environment, 2017, 31: 178-184. DOI: 10.13448/j.cnki.jalre.2017.030.]

[15] CLARK M P, HENDRIKX J, SLATER A G, et al. Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review[J]. Water Resources Research, 2011, 47(7): W7531-W7539. DOI: 10.1029/2011WR010745.

[16] DEBEER C M, POMEROY J W. Influence of snowpack and melt energy heterogeneity on snow cover depletion and snowmelt runoff simulation in a cold mountain environment[J]. Journal of Hydrology, 2017, 553: 199-213. DOI: 10.1016/j.jhydrol.2017.07.051.

[17] HEDRICK A R, MARKS D, HAVENS S, et al. Direct Insertion of NASA Airborne Snow

Observatory-Derived Snow Depth Time Series Into the iSnobal Energy Balance Snow Model[J]. Water Resources Research, 2018, 54(10): 8045-8063. DOI: 10.1029/2018WR023190.

[18] GAO Y, DAI Y F, YANG W, et al. Estimation of snow bulk density and snow water equivalent on the Tibetan Plateau using snow cover duration and snow depth[J]. Journal of Hydrology: Regional Studies, 2023, 48: 101473. DOI: 10.1016/j.ejrh.2023.101473.

[19] STURM M, TARAS B, LISTON G E, et al. Estimating Snow Water Equivalent Using Snow Depth Data and Climate Classes[J]. Journal of Hydrometeorology, 2010, 11(6): 1380-1394. DOI: 10.1175/2010JHM1202.1.

[20] BROXTON P, Van LEEUWEN W, BIEDERMAN J. Improving Snow Water Equivalent Maps With Machine Learning of Snow Survey and Lidar Measurements[J]. Water Resources Research, 2019, 55.DOI:10.1029/2018WR024146.

[21] HILL A F, RITTGER K, DENDUP T, et al. How Important Is Meltwater to the Chamkhar Chhu Headwaters of the Brahmaputra River?[J]. Frontiers in Earth Science, 2020, 8. DOI: 10.3389/feart.2020.00081.

[22] GAO Y, DONG H W, DAI Y F, et al. Contrasting changes of snow cover between different regions of the Tibetan Plateau during the latest 21 years[J]. Frontiers in Earth Science, 2023, 10. DOI: 10.3389/feart.2022.1075988.

[23] 戴礼云, 车涛. 1999-2008 年中国地区雪密度的时空分布及其影响特征[J]. 冰川冻土, 2010, 32(5):861-866. DOI: 1000-0240(2010)05-0861-06. [DAI L Y, CHE D. The Spatio-Temporal Distribution of Snow Density and Its Influence Factors from 1999 to 2008 in China[J]. Journal of Glaciology and Geocryology, 2010, 32(5): 861-866. DOI: 1000-0240(2010)05-0861-06.]

[24] 闫大江, 马宁, 张寅生. 青藏高原 0.05°逐日积雪深度数据集（2000-2018）[DS/OL]. 时空三极环境大数据平台, 2021.10.DOI:11888/Snow.tpdc.271743. [YAN D J, MA N, ZHANG Y S. A daily, 0.05° Snow depth dataset for Tibetan Plateau (2000-2018)[DS/OL]. A Big Earth Data Platform for Three Poles, 2021.10. DOI: 10.11888/Snow.tpdc.271743. CSTR: 18406.11.Snow.tpdc.271743.]

[25] 阳坤, 何杰, 唐文君, 等. 中国区域地面气象要素驱动数据集（1979-2018）[DS/OL]. 时空三极环境大数据平台, 2019.DOI:10.11888/AtmosphericPhysics.tpe.249369.file.[YANG K, HE J, TAN G W J, et al. China meteorological forcing dataset (1979-2018)[DS/OL]. A Big Earth Data Platform for Three Poles, DOI: 10.11888/AtmosphericPhysics.tpe.249369.file. CSTR: 18406.11.AtmosphericPhysics.tpe.249369.file.]

[26] 袁家豪. 基于 LightGBM 和 RF 的生态系统净碳交换预测和分析[D]. 吉林农业大学, 2023. [YUAN J H. Prediction and Analysis of Ecosystem Net Carbon Exchange Based on LightGBM and RF[D]. Jilin Agricultural University, 2023.]

[27] 周舟, 朱灵龙, 张永宏, 等. 基于 FY-3B 被动微波数据的青藏高原降尺度机器学习雪深反演[J]. 冰川冻土, 2024, 46(02): 539-554. DOI: 10.7522/j.issn.1000-0240.2024.0044. [ZHOU Z, ZHU L L, ZHANG Y H, et al. Downscaling machine learning snow depth inversion on the Qinghai-Xizang Plateau based on FY-3B passive microwave data[J]. Journal of Glaciology and Geocryology, 2024, 46(02): 539-554. DOI: 10.7522/j.issn.1000-0240.2024.0044.]

[28] 杨春蕙, 栗小东, 刘琦, 等. 机器学习方法在舟山渔场主要经济蟹类生物量估算中的应用[J]. 海

洋科学, 2023, 47(09): 61-70. DOI: 10.11759/hykx20221127002. [YANG C H, LI X D, LIU Q, et al. Application of machine learning methods for estimating the biomass of economically important crabs in the Zhoushan fishery[J]. Marine Sciences, 2023, 47(09): 61-70. DOI: 10.11759/hykx20221127002.]

[29] 付佳龙. 基于 LightGBM 模型的径流预报研究[D]. 华中科技大学, 2022. [FU J L. Research on Runoff Forecast Based on LightGBM Model[D]. Huazhong University of Science and Technology, 2022.]

论文引用格式

曹晓云，史飞飞，刘致远，等. 青藏高原逐月多年平均积雪密度格点数据集[J/OL]. 中国科学数据, 2025, 10(1). (2025-02-24). DOI: 10.11922/11-6035.ncdc.2024.0020.zh.

数据引用格式

曹晓云，史飞飞，刘致远，等. 青藏高原逐月多年平均积雪密度格点数据集[DS/OL]. V1. Science Data Bank, 2024. (2024-09-26). DOI: 10.57760/sciencedb.13892.

A dataset of monthly multi-year average snow density grids on the Tibetan Plateau

CAO Xiaoyun[1,2], SHI Feifei[1,2], LIU Zhiyuan[1,2], ZHANG Juan[1,2,3], LI Hongyi[3], XIAO Jianshe[4*]

1. Institute of Qinghai Meteorological Science Research, Xining 810001, P.R. China
2. Key Laboratory of Disaster Prevention and Mitigation of Qinghai Province, Xining 810001, P.R. China
3. Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, P.R. China
4. School of Electronic Engineering, Chengdu University of Information Engineering, Chengdu 610225, P.R. China

*Email: xiaojianshe@126.com

Abstract: Snow density is an important parameter for characterizing snow cover and functions as a key index for converting snow depth into snow water equivalent. It also plays a vital role in the estimation of snow water resources in mountainous areas, the management of water resources such as snowmelt floods, natural disaster forecasting, and climate research. Sourcing data from 132 day-by-day national meteorological stations on the Tibetan Plateau from 1960 to 2020, China's regional surface meteorological element-driven dataset, and the satellite-fused snow depth dataset as the main data sources, we compared the performance of several machine learning models in the simulation of snow density by different surface types, and selected the optimal model. Integrating the ground, satellite, and reanalysis data, we finally produced the dataset of monthly multi-year average snow density grids on the Tibetan Plateau. An accuracy evaluation using month-by-month multi-year average snow density data from 132 national meteorological

stations on the Tibetan Plateau showed that an average root mean square error of 0.019 g/cm^3 and an average relative error of 11.88%, indicating the high accuracy of the dataset. This dataset is expected to provide data support for water resource assessment and hydrological process simulation on the Tibetan Plateau.

Keywords: the Tibetan Plateau; snow density; weather stations; machine learning

Dataset Profile

Title	A dataset of monthly multi-year average snow density grids on the Tibetan Plateau
Data authors	CAO Xiaoyun, SHI Feifei, LIU Zhiyuan, ZHANG Juan, LI Hongyi, XIAO Jianshe
Data corresponding author	XIAO Jianshe (xiaojianshe@126.com)
Time range	January to December
Geographical scope	The Tibetan Plateau (73°–105°E, 25°–40°N)
Spatial resolution	0.05°×0.05°
Data volume	11.2 MB
Data format	*. tif
Data service system	<https://doi.org/10.57760/sciencedb.13892> <http://www.ncdc.ac.cn/portal/metadata/af7bbd77-93ac-4d77-966d-4104d31d9fee>
Sources of funding	Yangtze River Basin Open Fund Project (CJLY2022Y12); National Natural Science Foundation of China (U22A20556, U21A2021, 41761078); Science and Technology Plan Project of Qinghai Provincial Department of Science and Technology (2024-ZJ-740); The Second Comprehensive Scientific Expedition and Research Project on the Qinghai Tibet Plateau (2019QZKK0105); Qinghai Province Key Laboratory Project for Disaster Prevention and Reduction (QFZ-2021-Z01).
Dataset composition	The dataset contains a total of 12 monthly snowpack density data in tif format.