

新疆伊犁地区暴雪灾害风险预测（2050）数据集研发

崔晶¹, 戴晓爱^{2,3*}, 刘艳^{4,5*}

1. 成都理工大学地球与行星科学学院, 成都 610059;
2. 成都理工大学地理与规划学院, 成都 610059;
3. 地质灾害防治与地质环境保护国家重点实验室(成都理工大学), 成都 610059;
4. 中国气象局乌鲁木齐沙漠气象研究所, 乌鲁木齐 830002;
5. 中国气象局阿克苏达拉拉大气本底野外科学试验基地, 乌鲁木齐 830002

摘要: 新疆伊犁地区地处中亚内陆, 气候条件复杂多变, 暴雪灾害是该地区常见的自然灾害之一, 对区域生态、农牧业及居民生活构成严重威胁。然而, 当前针对该区域的暴雪灾害系统性风险评估不足, 未来情景下的风险预测数据匮乏, 制约了精细化防灾减灾决策。本研究基于随机森林模型, 综合气象要素(气温、积雪、风速等)与地形地貌因子, 模拟评估了伊犁地区2000–2020年历史时段的逐年暴雪灾害风险, 在此基础上预测了2050年SSP2–4.5情景下的暴雪灾害风险。暴雪灾害风险定义为特定地理环境下暴雪事件造成显著社会经济损失的概率, 数值经标准化处理至[0, 1]区间, 值越大风险越高。该数据集存储为.tif格式, 数据缺失被设置为-9999, 空间分辨率为500 m, 由22个数据文件组成, 数据量为78.5 MB(压缩为1个文件, 16.8 MB)。该数据集支持了第一作者硕士工科学位论文的完成。

关键词: 暴雪灾害; 伊犁地区; 随机森林; 风险评估; 未来情景; 2050; 硕士工科学位论文

DOI: <https://doi.org/10.3974/geodp.2026.02.05>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2026.02.05>

数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志(中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2025.11.07.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2025.11.07.V1>.

1 前言

新疆是中国暴雪灾害频发地区之一, 由暴雪引发的雪灾几乎每年都会造成不同程度的损失^[1–3]。这类极端天气不仅带来直接经济损失, 更会对区域社会经济、人类活动和生态环境造成深远影响^[4]。伊犁地区因其独特的地理气候条件成为雪灾研究的重点区域: 南北天山夹峙形成的伊犁河谷, 加之湿润的大陆性气候条件, 造就了异常丰沛的降雪^[5]。该地区

收稿日期: 2025-12-09; 修订日期: 2026-03-19; 出版日期: 2026-04-25

基金项目: 中华人民共和国科学技术部(2022xjkk0602); 中国气象局乌鲁木齐沙漠气象研究所科技发展基金(KJFZ202601); 中国气象科学研究院科技发展基金项目(2021KJ034)

*通讯作者: 戴晓爱, 成都理工大学地理与规划学院, daixiaoa@cdut.edu.cn; 刘艳, 中国气象局乌鲁木齐沙漠气象研究所, liuyan@idm.cn

数据引用方式: [1] 崔晶, 戴晓爱, 刘艳. 新疆伊犁地区暴雪灾害风险预测(2050)数据集研发[J]. 全球变化数据学报, 2026, 10(2): 145–156. <https://doi.org/10.3974/geodp.2026.02.05>. <https://cstr.escience.org.cn/CSTR:20146.14.2026.02.05>.

[2] 崔晶, 戴晓爱, 刘艳. 基于2000–2020逐年模拟模型预测的2050年新疆伊犁地区暴雪灾害风险数据集[J/DB/OL]. 全球变化数据仓储电子杂志, 2025. <https://doi.org/10.3974/geodb.2025.11.07.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2025.11.07.V1>.

冬季持续时间长,降雪强度大,积雪深度和降雪频率均显著高于新疆其他地区,使其成为雪灾高风险区,给当地居民生产生活带来严重困扰^[2,5-7]。

根据国家气象中心制定的降水量等级标准,24小时降雪量 $\geq 10\text{ mm}$ 即达到暴雪级别^[8]。暴雪灾害风险通常是指在给定的地理位置和环境条件下,发生能够造成显著社会经济损失(如人员伤亡、房屋损毁、农牧业减产等)的暴雪灾害事件的概率。已有的研究资料表明,2000–2020年间伊犁地区共记录81次暴雪灾害事件,不仅导致人员伤亡与房屋损毁,也对当地农牧业生产造成重创,成为制约区域生态与经济发展的重要风险因素^[5]。在全球气候变化背景下,未来极端天气事件发生概率可能进一步增加,亟需构建长时序、高精度的暴雪灾害风险评估数据集,以支撑区域风险防控体系的完善。然而,目前针对伊犁地区暴雪灾害的系统性风险评估研究仍较缺乏,尤其是未来情景下的风险预测数据更为稀缺,难以满足精细化防灾减灾的决策需求。

暴雪灾害风险评估需要综合考虑气象、地形、积雪和社会经济等多维度因素^[9,10]。这一评估过程的复杂性对研究方法 with 数据整合提出了更高要求。对比国内外相关研究发现,现有雪灾风险评估在方法上呈现出从传统统计分析向机器学习融合、从单一数据源向多源数据整合的发展趋势。国内研究多侧重于层次分析法、Logistic回归等传统统计方法,聚焦于牧区与高原地区,通过构建致灾因子-孕灾环境-承灾体的评估框架进行风险区划^[11-15]。而国际研究则更注重评估方法创新与框架革新,如Yang等^[16]积极探索机器学习算法,利用XGBoost模型识别高风险区域,而Lee等^[17]建立了考虑社会经济因子的DPSIR(驱动力-压力-状态-影响-响应)评估框架,Xu等^[18]应用Copula函数提升了危险性评估的精度。同时在数据使用上更广泛地融合了MODIS、GRACE等多源遥感数据^[19,20]。然而传统研究中常用的统计分析或经验模型,难以捕捉多因素间的复杂非线性关系,而机器学习算法具备处理复杂数据、挖掘潜在关联的核心优势,可有效弥补这一局限。作为机器学习领域经典的集成学习算法,随机森林具有高精度、抗过拟合和可解释性强的优势,近年来随着人工智能的迅速发展,已广泛应用于各类对象评价研究中^[21]。

本研究整合多源数据,采用随机森林模型对伊犁地区历史时期暴雪灾害风险进行量化评估,并进一步结合SSP2-4.5情景下的未来数据,预测生成了2050年的暴雪灾害风险数据。2050年预测聚焦于共享社会经济路径2中等辐射强迫情景(SSP2-4.5),该情景代表中等社会经济发展与中等气候减缓力度,与伊犁当前发展趋势较为吻合^[22],能够为短期至中期防灾规划提供合理参考。

2 数据集元数据简介

《基于2000–2020逐年模拟模型预测的2050年新疆伊犁地区暴雪灾害风险数据集》^[23]的名称、作者、地理区域、数据年代、时间分辨率、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表1。

3 研发方法

本数据研发流程如图1所示。首先根据伊犁地区的暴雪灾害事件提取历史暴雪灾害点,

表 1 《基于 2000–2020 逐年模拟模型预测的 2050 年新疆伊犁地区
暴雪灾害风险数据集》元数据简表

条目	描述
数据集名称	基于 2000–2020 逐年模拟模型预测的 2050 年新疆伊犁地区暴雪灾害风险数据集
数据集短名	Blizzard_Risk_Yili2000-2020&2050
作者信息	崔晶，成都理工大学地球与行星科学学院，cuijing@stu.cdut.edu.cn 戴晓爱，成都理工大学地理与规划学院，地质灾害防治与地质环境保护国家重点实验室（成都理工大学），daixiaoa@cdut.edu.cn 刘艳，中国气象局乌鲁木齐沙漠气象研究所，中国气象局阿克达拉大气本底野外科学试验基地，liuyan@idm.cn
地理区域	新疆伊犁地区
数据年代	2000–2020 年；2050 年（SSP2-4.5 情景）
时间分辨率	年
空间分辨率	500 m
数据格式	.tif
数据量	78.5 MB（压缩后 16.8 MB）
数据集组成	2000–2020 年逐年暴雪灾害风险数据评估，2050 年 SSP2-4.5 情景下的暴雪灾害风险数据预测
基金项目	中华人民共和国科学技术部（2022xjkk0602）；中国气象局乌鲁木齐沙漠气象研究所科技发展基金（KJFZ202601）；中国气象科学研究院科技发展基金项目（2021KJ034）
数据计算环境	Python、ArcGIS、R
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101，中国科学院地理科学与资源研究所
数据共享政策	（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报（中英文）》编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10%引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源 ^[24]
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC

并整合气象、积雪、地形、社会经济与土地利用等多源数据，提取 12 个关键影响因子。其次，基于历史暴雪灾害点生成训练样本，并采用随机森林机器学习算法进行模型训练。通过模型学习各因子与灾害间的复杂关系，最终输出伊犁地区 2000–2020 年逐年及 2050 年 SSP2-4.5 情景下的暴雪灾害风险数据。

3.1 数据预处理

3.1.1 暴雪灾害点数据提取

整理《中国气象灾害大典·新疆卷》中 2000 年的雪灾数据，以及民政部 2001–2020 年的雪灾数据，得到伊犁地区 2000–2020 年 81 条暴雪灾害事件数据，根据该数据进一步选取灾害点。当灾害影响范围明确限定在县城时，考虑到县城的行政中心通常具有代表性，且其位置相对固定、易于确定，因此选择该县行政中心作为灾害点。对于影响范围较为模糊的区域，根据区域面积、地形复杂程度和土地利用类型分布等因素，确定随机选取的点

的数量，以确保这些点能够合理地反映该区域的灾害分布情况。如某县北部山区或东部牧区，由于缺乏明确的灾害点位置信息，结合地形地貌和土地利用类型特征，最终从原始数据中提取出了 51 个具有代表性的暴雪灾害点。

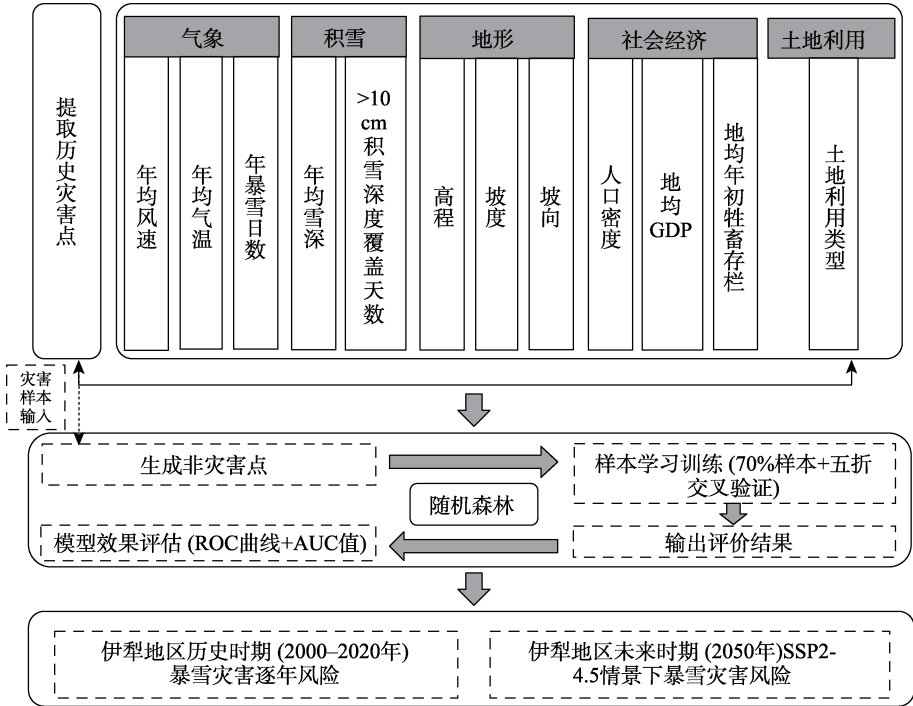


图 1 数据集研发流程图

3.1.2 指标因子提取

首先，依据雪灾成灾机理及相关研究^[25,26]，结合伊犁地区地形、气象、积雪、及社会经济特征，选取高程、坡度、坡向、年均气温、年均风速、年均雪深、>10 cm 积雪深度覆盖天数、人口密度、地均 GDP、地均年初牲畜存栏、土地利用类型共 12 个灾情影响因子。其次，收集并整理了用于提取影响因子的基础数据，数据信息及来源见表 2。然后进行因子数据计算，过程如下：

年均气温与风速基于逐日栅格数据计算得到，其中 2019–2020 年逐日气温通过伊犁地区 15 个气象站的日均温数据基于 ANUSPLIN 算法^[27]插值获得。依据中国气象局提出的方法——若气温低于 3℃且当日有降水视为降雪^[28]，结合逐日降水和气温数据，提取逐年暴雪日数。历史期地均年初牲畜存栏数来源于新疆统计年鉴，并通过随机森林回归模型结合气温、降水、人口与 GDP 等因子构建预测框架^[29,30]，预测得到 2050 年数据。历史期年均雪深及>10 cm 积雪覆盖天数基于逐日栅格数据计算；未来期则基于第六次国际耦合模式比较计划（CMIP6）ACCESS-ESM1-5 模式 r1i1p1f1 集合地表雪量数据^[31]，将历史模拟时段（1981–2014 年）与预估时段（2025–2058 年）数据裁剪为平闰年对应一致的 34 年等长序列，削弱年际波动。随后以中国雪深长时间序列数据集为基准数据，通过双线性插值重采

表 2 基础数据信息及来源统计表

类型	名称	年份	分辨率	来源
牲畜存栏数	年初大牲畜存栏量	2000–2020	县级	新疆及生产建设兵团统计年鉴
气象数据	中国 1km 分辨率逐日平均气温数据集	2000–2018	1 km	国家青藏高原科学数据中心
	中国区域地面气象要素驱动数据集 v2.0	2000–2020	0.1°	国家青藏高原科学数据中心
	NOAA 气象台站数据（日平均气温）	2019–2020	/	https://www.ncei.noaa.gov/data/
DEM	全球 90 米分辨率海洋和陆地 DEM 数据产品（GDEM_2022）	2022	90 m	https://cloud.tsinghua.edu.cn/d/695ed43696564904980f/?p=%2F&mode=list
土地利用	中国 30 米年度土地覆盖产品（CLCD）	2000–2020	30 m	https://zenodo.org/records/12779975
	精细的全球未来土地利用土地覆盖数据（LULC）	2050	1 km	https://doi.org/10.6084/m9.figshare.23542860
积雪深度	中国雪深长时间序列数据集（1979–2023）	2000–2020	25 km	Pie-engine (https://engine.pie-sat.cn/engine/home)
人口	Worldpop 人口数据集	2000–2020	1 km	http://www.worldpop.org/methods/populations/
	2020–2100 年全球 1 公里网格尺度多情景人口分布数据产品	2050	1 km	https://doi.org/10.6084/m9.figshare.19608594.v2
GDP	1992–2019 年全球 1 平方公里网格化 GDP 数据	2000–2019	1 km	www.gis5g.com
	2014–2020 年 GDP 栅格数据集	2020	1 km	www.gis5g.com
	共享社会经济路径（SSPs）人口和经济格点化数据集	2050	1 km	https://cstr.cn/31253.11.sciencedb.01683
降水	中国逐日降水数据集	2000–2020	0.1°	国家青藏高原科学数据中心
CMIP6 气温和降水数据	高分辨率中国 CMIP6 统计降尺度气候预估数据集 1979–2100（HiCPC）	2050	0.1°	国家青藏高原科学数据中心
CMIP6 风速数据	中国区域 CMIP6 降尺度降水温度风速数据集（1979–2100）	2050	0.25°	国家青藏高原科学数据中心
积雪气候模式数据	CMIP6 地表雪量数据（变量名：snw）	2050	1.875° ×1.25°	https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/

样至与基准一致的 0.25°×0.25°空间分辨率，保持数据空间连续性以及后续偏差校正的空间匹配性^[31]。后续结合新疆地区雪的密度^[32]进行本地化校正，将地表雪量转换为雪深以减少系统偏差，并采用历史偏差（CMIP6 历史模拟数据-基准数据），对未来数据进行校正^[31]，确保数据物理合理性与精度。此外，人口密度、地均 GDP 与土地利用类型为逐年数据，经预处理后直接使用，其中不同来源的土地利用类型数据已完成代码统一。坡度和坡向由 DEM 数据提取。

针对多源数据在时空分辨率与格式上的差异，已将所有数据配准至 Albers 坐标系，所有因子最终重采样为 500 m 分辨率，转换为*.tif 栅格格式。

3.2 评估模型构建方法

本研究采用随机森林算法构建伊犁地区雪灾风险评估模型。随机森林是一种集成学习方法,通过 Bootstrap 抽样生成多个训练子集,并为每棵决策树随机选择特征进行节点分裂,最终通过投票机制确定分类结果^[33]。该方法能够有效处理气象和地理数据中的复杂非线性关系,在小尺度灾害风险评估中表现出良好性能^[21]。

在模型具体实现过程中,首先通过特征工程对预处理后的暴雪相关的气象和地理数据,如降雪量、温度、风速及地形等信息提取关键特征。然后,采用自助采样(Bootstrap Sampling)技术从原始数据集 D 中生成 B 个训练子集 D_b ,其中每个子集通过有放回随机抽样获得,样本量与原数据集相同。本研究中设定 $B=200$ 。模型训练采用 5 折交叉验证优化参数^[16,34],并以 70%的样本用于训练,30% 用于验证。对于每个训练子集 D_b ,构建一棵决策树 T_b 。在树的每个节点分裂时,从全部 P 个特征中随机选择 $\sqrt{P}$ 个特征构成候选特征子集,并基于基尼不纯度最小化的原则选择最优分裂点,以降低过拟合风险。对于一个待预测的样本 x ,每棵决策树给出一个预测概率 $p_b(x)$,最终通过加权投票机制综合所有决策树的预测结果,得到暴雪灾害发生的平均概率:

$$\hat{p}(x) = \frac{1}{B} \sum_{b=1}^B p_b(x) \quad (1)$$

针对 3,000 m 以上高海拔区域(历史和未来时期均为人口稀疏区,人口密度均低于 1 人/km²),结合人口分布特征进行风险降权处理。定义修正后的风险值:

$$R_{adj}(x) = \begin{cases} \hat{p}(x) \times \omega(x), & \text{if } H(x) > 3000 \\ \hat{p}(x), & \text{otherwise} \end{cases} \quad (2)$$

式中, $H(x)$ 为样本点的海拔高度, $\omega(x) \in [\omega_{min}, 1]$, 为基于归一化人口密度的降权系数, ω_{min} 为预设的最小权重阈值。

在模型验证阶段,采用受试者工作特征曲线(Receiver Operating Characteristic Curve, ROC)和 ROC 曲线下面积(Area Under the Curve, AUC)评估模型判别能力^[35]。精度满足要求后,基于历史观测数据生成了 2000–2020 年的暴雪灾害风险评估数据,并进一步结合 SSP2-4.5 情景下的未来数据,预测生成了 2050 年的暴雪灾害风险数据,最后将所有输出结果归一化至[0, 1]区间。

4 数据结果与验证

4.1 数据集组成

本数据集存储在名为“Blizzard_Risk_Yili2000-2020&2050”的文件夹中,包含 2000–2020 年历史时期逐年及 2050 年 SSP2-4.5 情景下的暴雪灾害风险数据。历史时期(2000–2020 年)的逐年数据文件命名“BlizRisk_yyyy.tif”(其中 yyyy 代表具体年份),2050 年的未来情景数据文件命名为“BlizRisk_2050_SSP2-4.5.tif”。

所有数据空间分辨率为 500 m,格式为.tif,共 22 个文件,总数据量 78.5 MB(压缩后 16.8 MB)。数据值为 0–1 之间的归一化风险指数,数值越高表示暴雪灾害风险越大。

4.2 数据结果分析

如图 2 所示，受地形地貌影响，2000–2020 年间伊犁地区暴雪灾害风险呈现明显的空间分异，具体表现为南部的那拉提山，北部的科古琴山以及中部的乌孙山等山区风险较低，南北河谷地区因人口密集、积雪易累积，风险显著高于山区。21 年间风险空间分布相对稳定，局部地区有一定幅度的变动，如南部河谷地区 2013 年之后风险逐步降低。

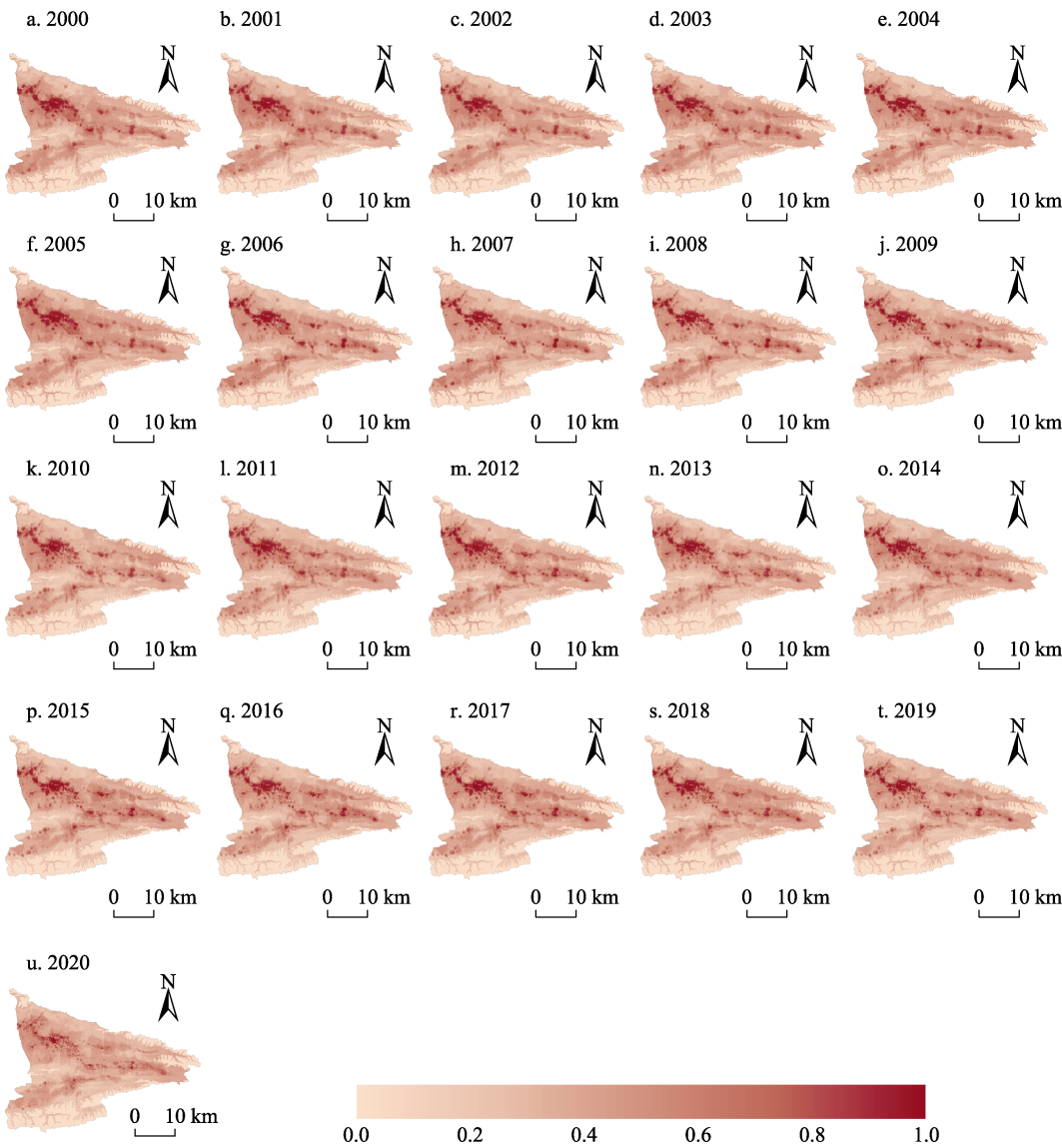


图 2 2000–2020 年伊犁地区暴雪灾害逐年风险空间分布图
(参考审图号为 GS (2024) 0650 号的标准地图制作)

如图 3 和图 4，在 SSP2-4.5 情景下，2050 年伊犁地区暴雪灾害风险整体升高。其中南部河谷地区以及西北部和中部山区风险升高较为显著。与 2000–2020 年间平均风险水平相比，升高的区域主要是从 0.20–0.30 区间升至 0.36–0.63 区间，风险值在 0.63 以上的区域面

积变化不大，仅南北河谷地区少部分区域风险略有上升（图 5）。

4.3 历史数据结果验证

本研究数据验证的基本假设是随机森林模型如果能够良好地模拟历史时期暴雪灾害风险，则也能可靠地预测未来的暴雪灾害风险。如图 6 所示，模型验证结果表明，随机森林模型表现出较好的预测能力，平均 AUC 值达到 0.7601 ± 0.1088 。

与已有伊犁地区暴雪灾害研究^[9]对比分析表明，本数据集的风险空间格局与其较为吻合，二者在风险分布的空间特征上具有较高一致性。时间上对逐年的暴雪灾害风险结果进

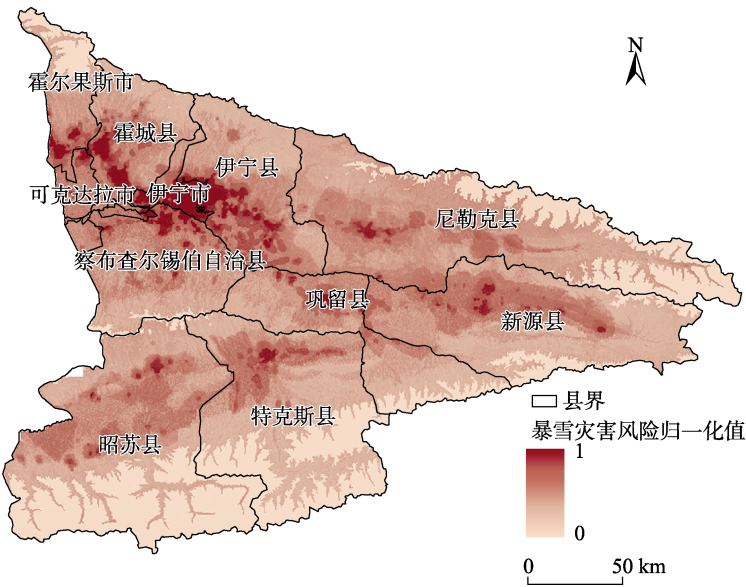


图 3 2050 年 SSP2-4.5 情景下伊犁地区暴雪灾害风险空间分布图
(参考审图号为 GS (2024) 0650 号的标准地图制)

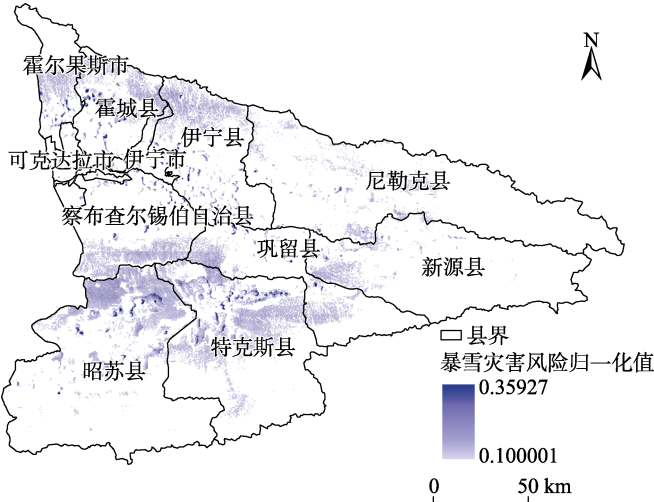


图 4 伊犁地区 2050 年与历史平均暴雪灾害风险归一化差值空间分布图
(参考审图号为 GS (2024) 0650 号的标准地图制)

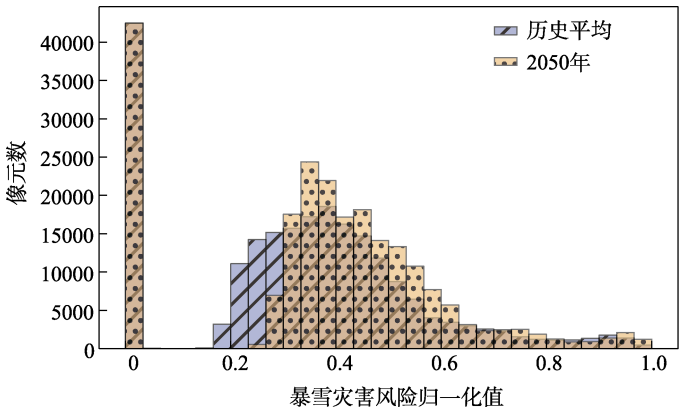


图 5 伊犁地区 2050 年与历史平均暴雪灾害风险归一化值的频数分布

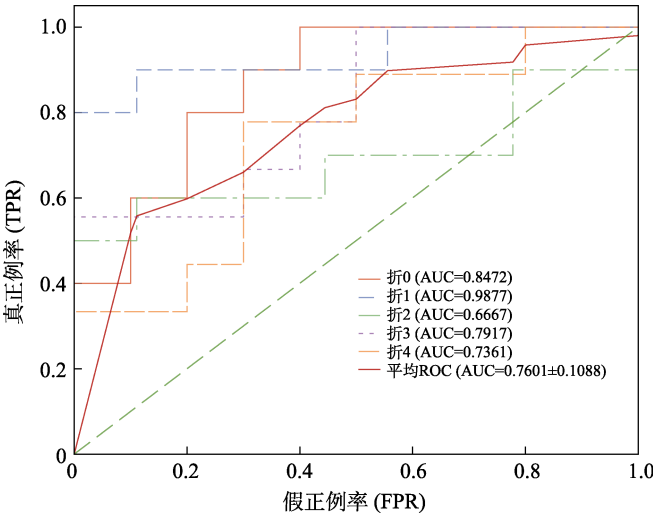


图 6 伊犁地区暴雪灾害风险评估模型交叉验证 ROC 曲线

行 Z-score 标准化，衡量其与多年均值的标准差偏离程度，并以 $Z\text{-score} > 1.5$ 作为阈值识别高风险异常区，该阈值对应标准正态分布单侧概率约 6.68%，可有效区分常规波动与高风险状态^[36,37]。同时结合风险值 >0.5 的条件，识别出显著偏离均值且具有较高风险水平的异常区，按异常区占比从高到低排序，最终得出 2005 年、2001 年、2010 年、2000 年、2003 年、2012 年、2014 年位列 2000–2020 年历史期异常区占比前 7 位，占比处于 1.8%–6.0% 区间，显著高于历史时期平均风险水平，被识别为高风险异常年。结合 2000–2020 年伊犁地区暴雪灾情数据（人口影响、死亡人数、房屋倒塌、农作物受损、大畜牧死亡等）计算灾损指数，分析雪灾强度等级发生次数的年际分布^[5]，得出灾害较重的年份是 2003 年、2005 年、2006 年、2010 年、2011 年、2012 年、2014 年，准确率 71.4%，召回率 71.4%。其中 2010 年虽然风险值不是最高，但因模拟出的高风险区与人口、经济高暴露区高度重合，实际灾损最为严重。由此可见，模型识别出的高风险年份与实际灾损指数具有较高一致性，验证了模型时间维度的有效性。

5 讨论和总结

本研究基于随机森林模型,融合气象、地形、积雪及社会经济等多源数据,首先模拟了伊犁地区 2000–2020 年历史时期暴雪灾害风险,并预测生成该区域 2050 年 SSP2-4.5 情景下的暴雪灾害风险。2000–2020 年间伊犁地区暴雪风险呈现“河谷地区高一南北山区低”的空间分布。2050 年 SSP2-4.5 情景下,伊犁地区暴雪风险总体呈上升趋势,中部和西北部山区及南部河谷风险升高较为显著。模型验证显示预测性能良好($AUC=0.7601\pm0.1088$),识别出的高风险年份与实际灾情记录吻合度较高,体现了模型具有较强的适用性与可靠性。基于上述结果,提出以下防灾减灾建议:一是针对风险持续处于高位的河谷区域维持现有防灾减灾体系,巩固监测、物资储备与应急响应机制;二是对于昭苏县和特克斯县北部等农牧集中且风险升高区域,优化防灾设施布局与抗灾物资储备,提升抗灾能力。

本数据集在空间格局上延续并深化了本团队前期研究¹所揭示的“中部河谷高、南北山区低”的整体分布规律,进一步验证了伊犁地区暴雪风险的空间分异特征。前期研究面向科考办防灾部署需求,采用以县级行政单元为基础的地理数据与综合评价方法,评估结果呈现以县域为单位的风险集聚特征,便于直接对接区域防灾工作。本研究在此基础上引入历史灾害点数据,结合随机森林算法强大的非线性关系捕捉能力,实现了风险分布的更精细空间表达,体现了本团队在暴雪风险评估方法上的持续优化。

本研究采用的随机森林模型,在继承多源数据融合与复杂非线性关系捕捉优势的基础上,结合未来气候情景与人口降权处理,提升了评估的区域适用性与前瞻性。但未来预测结果存在一定不确定性:一方面受 CMIP6 单一模式限制,模式本身的模拟偏差可能传导至评估结果,后续需结合多模式集成进一步降低不确定性;另一方面在模型参数优化和未来多情景设置方面仍有提升空间。该数据集支持了第一作者硕士工学学位论文的完成。

作者分工: 崔晶赞同并采纳了戴晓爱和刘艳对数据集开发的总体设计方案,采集和处理了数据并撰写了数据论文;戴晓爱和刘艳审核了数据并修改了数据论文。

利益冲突声明: 本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

- [1] 王秋香,刘卫平,李圆圆等.新疆不同区域牧业雪灾损失时频变化特征[J].冰川冻土,2015,37(4): 905–915.
- [2] 魏娟娟,李娜,万瑜等.乌鲁木齐典型暴雪天气机理及成因分析[J].干旱区地理,2022,45(5): 1381–1391.
- [3] 杨倩,秦莉,高培等.基于 RBF 网络的新疆特重雪灾区最大积雪深度预测研究[J].沙漠与绿洲气象,2024,18(1): 89–95.
- [4] 何文鑫,郝晓华,刘峰贵等.融合多源信息的伊犁地区道路雪灾风险评估研究[J].遥感技术与应用,2025,40(1): 177–191.

¹ 刘艳,火红,中国气象局乌鲁木齐沙漠气象研究所.伊犁河流域雪灾(暴雪)风险等级区划专题图,2025.

- [5] 火红, 刘艳, 买吾拉夏·木巴热克. 1990—2020 年新疆伊犁地区雪灾时空分布特征及其影响评估[J]. 干旱区地理, 2024, 47(11): 1828–1840.
- [6] 王旭, 储长江, 牟欢. 新疆雪灾空间格局和年际变化特征分析[J]. 干旱区研究, 2020, 37(6): 1488–1495.
- [7] 王秀琴, 卢新玉, 马禹等. 基于灾情数据的新疆雪灾的评估与区划[J]. 冰川冻土, 2019, 41(4): 836–844.
- [8] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 降水量等级 (GB/T 28592—2012) [S]. 北京: 中国标准出版社, 2012.
- [9] 黄弋洋, 戴晓爱, 刘艳等. 新疆伊犁地区暴雪灾害风险区划研究[J]. 气象水文海洋仪器, 2025, 42(2): 104–107.
- [10] 马晓芳, 黄晓东, 邓婕等. 青海牧区雪灾综合风险评估[J]. 草业学报, 2017, 26(2): 10–20.
- [11] 丁延龙, 孙晓瑞, 高永等. 内蒙古森林雪灾风险评估与区划[J]. 干旱区研究, 2019, 36(2): 486–493.
- [12] 德勒格日玛, 李一平, 孟雪峰等. 内蒙古锡林郭勒盟牧区雪灾风险评估研究[J]. 冰川冻土, 2020, 42(4): 1353–1362.
- [13] 陈虹举, 杨建平, 丁永建等. 基于 NEX-GDDP 数据集的青藏高原牧区雪灾风险预估[J]. 灾害学, 2022, 37(2): 102–110.
- [14] 李凡, 侯光良, 鄂崇毅等. 基于乡镇单元的青海高原果洛地区雪灾致灾风险评估[J]. 自然灾害学报, 2014, 23(6): 141–148.
- [15] 苗爱梅, 王洪霞, 逯张禹. 基于 GIS 的山西省暴雪灾害风险区划研究[J]. 中国农学通报, 2016, 32(20): 133–140.
- [16] Yang, Z. Y., Gokon, H., Yu, Q. Machine learning-based identification and assessment of snow disaster risks using multi-source data: insights from Fukui Prefecture, Japan [J]. *Progress in Disaster Science*, 2025, 26: 100426.
- [17] Lee, H., Kim, D., Chung, G. Classification of risk levels for snow damage estimation considering socio-economic factors in South Korea [J]. *Applied Water Science*, 2024, 14(11): 1–17.
- [18] Xu, Q. X., Huang, F. Y., Mou, S. H., et al. Snow disaster hazard assessment on the Tibetan Plateau based on copula function [J]. *Sustainability*, 2023, 15(13):10639.
- [19] Singh, D., Zhu, Y., Liu, S. Y., et al. Exploring the links between variations in snow cover area and climatic variables in a Himalayan catchment using earth observations and CMIP6 climate change scenarios [J]. *Journal of Hydrology*, 2022, 608: 127648.
- [20] Sahu, R., Ramsankaran, R., Bhambri, R., et al. Evolution of supraglacial lakes from 1990 to 2020 in the Himalaya Karakoram Region using cloud-based Google Earth Engine platform [J]. *Journal of the Indian Society of Remote Sensing*, 2023, 51(12): 2379–2390.
- [21] 王倩丽, 马细霞, 刘欣欣等. 基于随机森林的山洪灾害风险评价方法及应用[J]. 人民黄河, 2022, 44(4): 63–66, 73.
- [22] 伊犁哈萨克自治州统计局. 伊犁哈萨克自治州 2024 年国民经济和社会发展统计公报[EB/OL]. (2025-04-09). <https://www.xjyl.gov.cn/xjylz/c112816/202504/67c6814847cb40a0a33773fb1df5466b.shtml>.
- [23] 崔晶, 戴晓爱, 刘艳. 基于 2000—2020 逐年模拟模型预测的 2050 年新疆伊犁地区暴雪灾害风险数据集 [J/DB/OL]. 全球变化数据仓储电子杂志, 2025. <https://doi.org/10.3974/geodb.2025.11.07.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2025.11.07.V1>.

- [24] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策[OL]. <https://doi.org/10.3974/dp.policy.2014.05> (2017 年更新).
- [25] 余迪, 杜海荣, 郭娇等. 基于多源数据的三江源区雪灾风险评估[J]. 气象科技进展, 2023, 13(4): 34–40.
- [26] 庄晓翠, 周鸿奎, 王磊等. 新疆北部牧区雪灾评估指标及其成因分析[J]. 干旱区研究, 2015, 32(5): 1000–1006.
- [27] 钱永兰, 吕厚荃, 张艳红. 基于 ANUSPLIN 软件的逐日气象要素插值方法应用与评估[J]. 气象与环境学报, 2010, 26(2): 7–15.
- [28] Liu, D., Zhang, L. J., Jiang, S. Y., *et al.* Hazard prediction and risk regionalization of snowstorms in Northeast China [J]. *Physics and Chemistry of the Earth, Parts A/B/C*, 2020, 116: 102832.
- [29] Zhou, J., Niu, J., Wu, N., *et al.* Annual high-resolution grazing-intensity maps on the Qinghai-Tibet Plateau from 1990 to 2020 [J]. *Earth System Science Data Discussions*, 2023, 16(11): 5171–5189.
- [30] Zhang, X. Z., Li, M., Wu, J. S., *et al.* Alpine grassland aboveground biomass and theoretical livestock carrying capacity on the Tibetan Plateau [J]. *Journal of Resources and Ecology*, 2022, 13(1): 129–141.
- [31] 陈涛, 高歌, 杜晓辉等. CMIP6 耦合模式对青藏高原积雪的未来预估[J]. 地理科学, 2024, 44(5): 901–910.
- [32] Wang, H. D., Zhang, X. L., Xiao, P. F., *et al.* Towards large-scale daily snow density mapping with spatio-temporally aware model and multi-source data [J]. *The Cryosphere*, 2023, 17(1): 33–50.
- [33] 杨硕, 李德营, 严亮轩等. 基于随机森林模型的乌江高陡岸坡滑坡地质灾害易发性评价[J]. 安全与环境工程, 2021, 28(4): 131–138.
- [34] Huang, C. Y., She, D. X., Liu, X. B., *et al.* Predicting future meteorological drought risk in mainland China using random forest model [J]. *Journal of Hydrology: Regional Studies*, 2025, 61: 102633.
- [35] Liang, Z., Wang, C. M., Duan, Z. J., *et al.* A hybrid model consisting of supervised and unsupervised learning for landslide susceptibility mapping [J]. *Remote Sensing*, 2021, 13(8): 1464.
- [36] Ruijsch, D., van Mourik, J., Biemans, H., *et al.* Thrive or wither: exploring the impacts of multiyear droughts on vegetation [J]. *Journal of Geophysical Research: Biogeosciences*, 2025, 130(7): e2025JG008992.
- [37] Wang, X. X., Wu, Y. H., Luo, M. L., *et al.* Exports of organic matter, phosphorus and nitrogen from Sichuan Basin: a critical region regulating water quality of the Upper Yangtze River, China [J]. *Journal of Hydrology*, 2025, 662: 133889.