

DOI:10.13870/j.cnki.stbxb.2024.05.012

邢贞相,王嘉麒,张鸿雪,等.基于 DL 和 GKDE 的松花江流域土壤侵蚀类型的概率评价研究[J].水土保持学报,2024,38(5):116-128.

XING Zhenxiang, WANG Jiaqi, ZHANG Hongxue, et al. Probability evaluation study of soil erosion types in the Songhua River basin based on Deep Learning and Gaussian Kernel Density Estimation [J]. Journal of Soil and Water Conservation, 2024, 38(5): 116-128.

基于 DL 和 GKDE 的松花江流域土壤侵蚀类型的概率评价研究

邢贞相¹, 王嘉麒¹, 张鸿雪¹, 宋健¹, 王轶男², 段维义¹, 宫铭¹, 黄昌丽³

(1.东北农业大学水利与土木工程学院, 哈尔滨 150030; 2.中国科学院地理科学与资源研究所, 北京 100101;

3.黑龙江省齐齐哈尔生态环境监测中心, 黑龙江 齐齐哈尔 161005)

摘要: [目的] 为科学判别流域尺度土壤侵蚀类型并给出相应的发生概率。[方法] 构建基于深度学习 (deep learning, DL) 的松花江流域土壤侵蚀模数 (erosion modulus, EM) 计算模型, 并计算不同类型的侵蚀模数。以降雨、气温、风速 3 个侵蚀模数影响因子为随机变量, 利用数值模拟和高斯核密度估计法 (gaussian kernel density estimation, GKDE) 构建 EM 概率评价方法, 给出不同土壤侵蚀强度组合的发生概率。[结果] EM 计算模型验证期的 R^2 均 > 0.86 。流域内平均每年有 74.47% 发生微度水蚀与微度风蚀; 12.86% 的面积发生轻度及以上水蚀与微度风蚀; 12.56% 的面积发生轻度及以上风蚀与微度水蚀; 0.11% 的面积水蚀强度与风蚀强度均在轻度及以上。36 个典型像元中, 发生微度水蚀与微度风蚀的平均概率为 57.45%; 发生微度水蚀与轻度风蚀的平均概率为 30.26%; 发生微度水蚀与中度风蚀的平均概率为 8.03%; 发生轻度水蚀与微度风蚀的平均概率为 2.11%; 发生微度水蚀与重度风蚀的平均概率为 2.08%; 发生其余强度组合的平均概率在 0.05% 以下。[结论] 构建的松花江流域 EM 计算模型精度较高, 揭示了松花江流域土壤侵蚀类型空间分布特征, 并给出不同土壤侵蚀强度组合的发生概率, 为松花江流域土壤侵蚀治理提供依据。

关键词: 松花江流域; 土壤侵蚀; 深度学习算法; 高斯核密度估计法; 概率评价

中图分类号: S157.1

文献标识码: A

文章编号: 1009-2242-(2024)05-0116-13

Probability Evaluation Study of Soil Erosion Types in the Songhua River Basin Based on Deep Learning and Gaussian Kernel Density Estimation

XING Zhenxiang¹, WANG Jiaqi¹, ZHANG Hongxue¹, SONG Jian¹,

WANG Yinan², DUAN Weiyi¹, GONG Ming¹, HUANG Changli³

(1. School of Water Conservancy and Civil Engineering, Northeast Agricultural University, Harbin 150030, China;

2. Institute of Geographic Sciences and Natural Resources Research, Beijing 100101, China;

3. Heilongjiang Qiqihar Ecological Environment Testing Center, Qiqihar, Heilongjiang 161005, China)

Abstract: [Objective] To scientifically identify the types of soil erosion at the watershed scale and give the corresponding probability of occurrence. [Methods] This study constructs a deep learning (DL)-based model for calculating soil erosion modulus in the Songhua River Basin and calculates different types of soil erosion modulus. Using three erosion modulus influencing factors, namely rainfall, air temperature and wind speed, as random variables, numerical simulation and Gaussian Kernel Density Estimation (GKDE) were used to construct the EM probability evaluation method, which gives the probability of occurrence of different combinations of soil erosion intensities. [Results] The R^2 of the validation period of the EM computational models were all > 0.86 ; 74.47% of the average annual occurrence of slight water erosion and slight wind erosion in the watershed; 12.86% of the area of slight and above water erosion and slight wind erosion; 12.56% of the area of slight and above wind erosion and slight water erosion; 0.11% of the area of water

收稿日期: 2024-04-03

修回日期: 2024-05-19

录用日期: 2024-06-12

网络首发日期 (www.cnki.net): 2024-08-13

资助项目: 国家自然科学基金项目 (51979038, 52209008)

第一作者: 邢贞相 (1976—), 男, 教授, 博士, 主要从事农业水土资源不确定研究。E-mail: zxxing@neau.edu.cn

通信作者: 张鸿雪 (1989—), 女, 博士, 副教授, 硕士生导师, 主要从事水资源系统研究。E-mail: hongxue@neau.edu.cn

http://stbxb.alljournal.com.cn

erosion strength and wind erosion intensity are both slight and above. 36 typical image elements of the 36 typical images, the average probability of occurrence of slight water erosion and slight wind erosion is 57.45%; the average probability of occurrence of slight water erosion and mild wind erosion is 30.26%; the average probability of occurrence of slight water erosion and moderate wind erosion is 8.03%; the average probability of occurrence of mild water erosion and slight wind erosion is 2.11%; the average probability of occurrence of mild water erosion and severe wind erosion is 2.08%; and the evaluations of occurrence of the remaining combinations of probabilities were all below 0.05%. [Conclusion] The calculation model of erosion modulus in the Songhua River basin constructed in this study has high accuracy, reveals the spatial distribution characteristics of soil erosion types in the Songhua River basin, and gives the probability of occurrence of different combinations of the intensity of the two types of erosion, which provides a basis for the management of soil erosion in the Songhua River basin.

Keywords: Songhua River basin; soil erosion; deep learning; gaussian kernel density estimation method; probability evaluation

Received: 2024-04-03

Revised: 2024-05-19

Accepted: 2024-06-12

Online(www.cnki.net): 2024-08-13

松花江流域位于中国东北地区,该流域拥有珍贵的黑土资源,也是我国重要的粮食产区和商品粮基地。由于人类活动和自然因素的影响导致该流域土壤侵蚀问题尤为严重,自 20 世纪 50 年代以来,流域内黑土平均厚度减少约 40 cm^[1]。因此,有必要对该流域土壤侵蚀规律进行深入研究。

松花江流域内存在水蚀、风蚀 2 种不同的土壤侵蚀类型^[2],WANG 等^[3]使用同位素示踪法研究该流域水蚀、风蚀的成因;左小锋等^[4]通过风洞试验研究黑土区前期风蚀与水蚀相互作用;WANG 等^[5]通过经验方程对东北地区近 20 年不同土壤侵蚀类型进行量化分析松花江流域内风蚀、水蚀的变化趋势;吴瀚逸等^[6]、杨严攀等^[7]、池金洺等^[8]分别基于不同分辨率的地形、气象、土地利用等数据产品通过经验方程计算包括松花江流域在内不同研究区域的土壤侵蚀模数,发现其所表征的对应研究区域内土壤侵蚀空间分布特征均具有较好的代表性;国外学者^[9-10]基于深度学习(DL)构建区域水蚀或风蚀及其影响因子的预测模型,并分析侵蚀影响因子与侵蚀之间的映射关系。

当前基于经验方程的侵蚀模型和基于 DL 构建的侵蚀模型的研究多单独针对水力侵蚀或风力侵蚀及其相应影响因子开展研究。但部分因子对 2 种侵蚀都有贡献^[11],目前与此相关的研究较少。土壤侵蚀的部分影响因子(如降雨、风速、气温等)具有较强的随机性^[12],使得土壤侵蚀的评价结果也具有随机性,即某土壤侵蚀评价结果应具有其相应的发生概率。因此,开展土壤侵蚀类型的概率评价研究能够更好地考虑其影响因素的随机性,给出具有更丰富决策参考信息的概率评价结果^[13]。

土壤侵蚀可能是由水蚀、风蚀共同作用导致的,科学识别土壤侵蚀类型及其在空间的分布,并给出其发生概率是值得研究的一个课题。为更好地揭示松花江流域土壤侵蚀类型及其空间分布,本研究结合土壤侵蚀的经验方程的土壤侵蚀量化结果,构建基于 DL 的侵蚀模数(erosion modulu,EM)计算模型,使其可以区分土壤侵蚀类型的同时计算相应类型的土壤侵蚀模数;采用非参数拟合技术获取土壤侵蚀随机影响因子的概率密度,进而利用数值模拟方法创建土壤侵蚀模数随机影响因子样本,并采用高斯核密度估计法构建基于水蚀风蚀联合概率分布的土壤侵蚀概率评价方法,概率评价结果可为区域土壤侵蚀治理提供科学的决策依据。

1 材料与方法

1.1 研究区概况

松花江是黑龙江的最大支流,全长约 2 309 km,流域面积 55.68 万 km²,位于中国东北部(119°52′—132°31′E,41°42′—51°38′N),主要涵盖黑龙江省与吉林省的大部及内蒙古部分区域(图 1)。流域内的地形以平原为主,其中包括位于中部的松嫩平原和东北部的三江平原,海拔为 50~200 m;流域的西部为大兴安岭,东部为长白山,普遍海拔为 700~2 700 m。松花江流域属于北温带季风气候,夏季温热多雨,冬季寒冷干燥。年内温差较大,多年平均降水量为 500 mm,平均气温 3~5 °C,平均风速 2~3 m/s。

1.2 数据准备

研究数据以松花江流域现有的土壤侵蚀空间数据及地形、地质、气候、土地利用等数据来训练 DL 模型。

1.2.1 EM 影响因子数据 根据已有关于土壤侵蚀

因素的相关研究^[9,14-15],结合松花江流域黑土区土壤侵蚀机理的研究成果^[16],同时考虑数据的可获取性,初步选定 14 个可能影响松花江流域水力侵蚀模数(water erosion modulus, EM_{wa})、风力侵蚀模数(wind erosion modulus, EM_{wi})的驱动因子,将其按属性划分为 4 类:地形因素(海拔、坡度、坡向、坡长、水流强度指数、地形湿度指数)、气候因素(气温、降雨、风速、积雪)、土地利用因素(植被覆盖度 NDVI、土地利用 LC)及土壤因素(黏粒、砂粒、粉粒及土壤有机碳含量 SOC)。

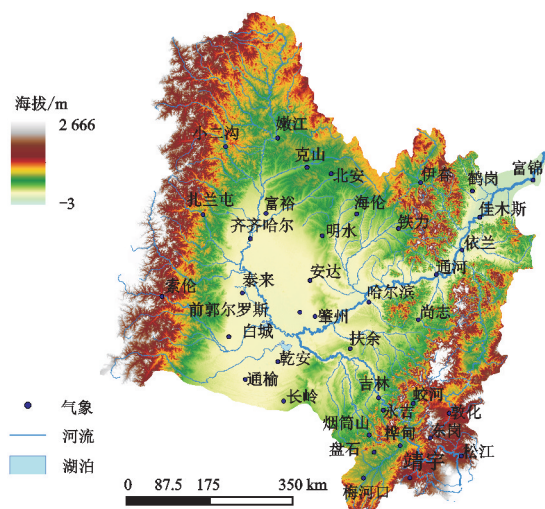


图 1 松花江流域位置

Fig. 1 Map of the Songhua River basin location

(1) 地形因素

海拔、坡度、坡向、坡长、水流强度指数(stream power index, SPI)、地形湿度指数(topographic wetness index, TWI)均可在数字高程数据(DEM)中提取、计算得到^[15]。本研究使用的数字高程数据(DEM)来源于 <http://earthexplorer.usgs.gov> 网站,网站提供的 DEM 数据有 $30\text{ m} \times 30\text{ m}$ 与 $90\text{ m} \times 90\text{ m}$ 分辨率精度。本研究选用的 DEM 数据分辨率为 $30\text{ m} \times 30\text{ m}$ 。

(2) 气候因素

本研究的气象数据来源为中国地面气候资料日值数据集(V3.0)。整理松花江流域内 36 个气象站点的年降雨量、年平均风速、年平均日最高气温 3 项气象数据资料,气象站点位置见图 1,本研究所需的时间序列为 2001—2020 年,使用“克里金法”将数据插值为 $100\text{ m} \times 100\text{ m}$ 分辨率的栅格数据。

由于受到高纬度的影响,应考虑积雪覆盖时长对土壤侵蚀的影响^[17-18]。积雪覆盖时长数据来源于国家冰川冻土沙漠科学数据中心(<http://www.ncdc.ac.cn>)的 MODIS 中国积雪物候数据集^[19-20],本研究所需的时间序列为 2001—2020 年,分辨率为 $500\text{ m} \times 500\text{ m}$ 。

(3) 土地覆盖因素

本研究使用的土地利用类型数据为中国土地覆盖数据集(CLCD)^[21],该数据的分辨率为 $30\text{ m} \times 30\text{ m}$,按照土地利用类型分类规则将该数据分为耕地、林地、草地、水域、建设用地及荒地。本研究使用的 NDVI 数据为中国月度植被指数(NDVI)空间分布数据集,分辨率为 $1\text{ km} \times 1\text{ km}$ 。数据来源于资源环境科学数据注册与出版系统(<https://www.resdc.cn/data.aspx?DATAID=254>)。采用最大值镶嵌法将月尺度 NDVI 数据处理为年尺度。本研究所需要的土地利用类型及 NDVI 数据时间序列均为 2001—2020 年。

(4) 土壤因素

本研究土壤数据来源于国家青藏高原科学数据中心(<https://doi.org/10.11666/00073.ver1.db>)的中国高分辨率国家土壤信息网格基本属性数据集(2010—2018)^[22]。该数据有 $250\text{ m} \times 250\text{ m}$ 和 $1\text{ km} \times 1\text{ km}$ 2 种分辨率,每类数据分为 7 种不同的深度。本研究使用的黏粒、粉粒及砂粒数据为 0—20 cm 深度、SOC 为 5—15 cm 深度,分辨率均 $250\text{ m} \times 250\text{ m}$ ^[23]。将搜集到的 16 个影响因子栅格数据重采样为 $100\text{ m} \times 100\text{ m}$ 分辨率。

1.2.2 土壤侵蚀模数空间分布数据

(1) 水力侵蚀模数

本研究使用的土壤水蚀数据为吴瀚逸等^[6]制作的 2001—2020 年中国东北区域土壤水蚀数据集,该数据集分辨率为 $250\text{ m} \times 250\text{ m}$ 。该数据集基于 RUSLE 模型,对模型每个因子选取多种经典算法进行排列组合计算,结合水文站实测数据直接验证筛选最优算法组合,得到东北地区水力侵蚀模数估算结果。该数据具有较高的时空分辨率和实际精度,可为东北区域的土壤侵蚀评估及水土保持决策提供基础数据。

(2) 风力侵蚀模数

土壤风蚀数据由已被证明在中国东北地区具有较好适用性的通用风蚀方程(RWEQ)计算得到^[5],该方程具体计算方法为:

$$SL = \frac{2z}{S^2} Q_{\max} e^{-\left(\frac{z}{S}\right)^2} \quad (1)$$

$$Q_{\max} = 109.80(WF \times EF \times SCF \times K' \times C) \quad (2)$$

$$S = 150.71(WF \times EF \times SCF \times K' \times C)^{-0.3711} \quad (3)$$

式中: SL 为实际风蚀量(kg/m^2); $Q_{\max}$ 为风力最大输沙能力(kg/m); S 为关键地块长度(m); z 为下风向距离(m),本次计算取 50 m ^[24]; WF 为气候因子(kg/m); EF 为土壤可蚀性百分比(%); SCF 为土壤结皮因子; K' 为土壤糙度因子; C 为植被因子。上述因子计算方法及部分参数的选取参考现有研究^[24]。基于 RWEQ

计算 2001—2020 年松花江流域风力侵蚀模数。

将风蚀模数、水蚀模数分辨率重采样为 100 m×100 m,基于重采样后的数据绘制松花江流域 2001 年、2005 年、2010 年、2015 年、2020 年水蚀模数与风蚀模数空间分布图(图 2)。按土壤侵蚀分类分级标准将水蚀模数、风蚀模数划分为 6 个等级,并计算松

花江流域各级土壤侵蚀的面积(表 1)。由表 1 可知,2001 年、2005 年、2010 年、2015 年、2020 年松花江流域均以微度水蚀或微度风蚀为主;除 2020 年外,流域内其余年份轻度及以上风蚀的面积均大于轻度及以上水蚀面积;2020 年轻度水蚀面积远超其余年份,但中度及以上水蚀面积仍较低。

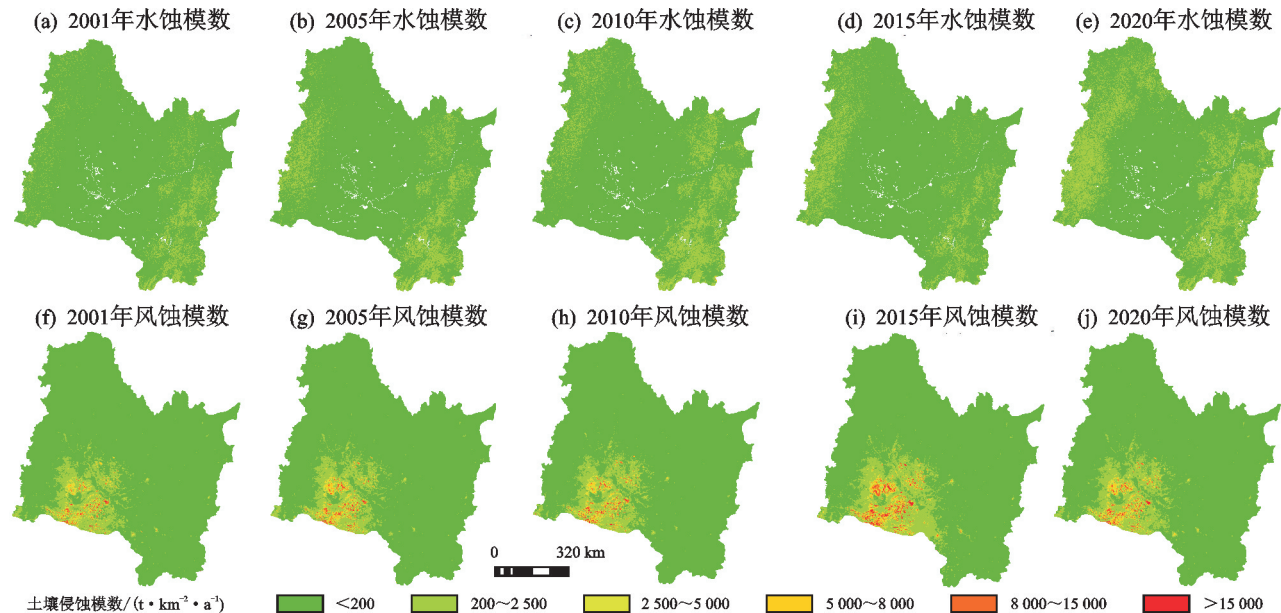


图 2 松花江流域侵蚀模数空间分布

Fig. 2 Spatial distribution of soil erosion modulus in the Songhua River Basin

表 1 松花江流域各级土壤侵蚀面积

Table 1 Area of soil erosion at all levels of the Songhua River Basin		km ²				
侵蚀类型	土壤侵蚀模数等级	2001 年	2005 年	2010 年	2015 年	2020 年
水蚀	微度[<200 t/(km ² ·a)]	509 874.45	488 995.98	474 034.36	492 540.59	424 059.86
	轻度[200~2 500 t/(km ² ·a)]	39 979.32	60 815.49	75 563.49	57 294.47	125 541.15
	中度[2 500~5 000 t/(km ² ·a)]	12.59	45.00	239.67	26.00	250.31
	强度[5 000~8 000 t/(km ² ·a)]	1.42	8.06	19.70	5.78	12.91
	极强度[8 000~15 000 t/(km ² ·a)]	0.07	3.03	10.00	1.06	2.99
	剧烈[>15 000 t/(km ² ·a)]	—	0.08	0.73	—	0.03
风蚀	微度[<200 t/(km ² ·a)]	491 782.07	489 353.83	487 712.29	459 513.48	481 478.96
	轻度[200~2 500 t/(km ² ·a)]	49 959.12	49 163.24	51 434.55	69 812.66	55 488.62
	中度[2 500~5 000 t/(km ² ·a)]	7 562.69	8 616.03	8 551.34	11 401.39	9 212.57
	强度[5 000~8 000 t/(km ² ·a)]	2 853.72	3 939.28	3 590.23	6 078.12	4 342.78
	极强度[8 000~15 000 t/(km ² ·a)]	1 703.08	2 479.08	2 309.59	5 155.66	2 750.63
	剧烈[>15 000 t/(km ² ·a)]	674.11	983.22	936.66	2 573.36	1 261.17

1.3 EM 影响因子重要性验证方法

采用 BORUTA 算法^[14]验证初步确定的 16 个与 2 种类型 EM 有关的影响因子的重要性。该算法适用于离散或连续型的数据,能有效识别多维影响因子中哪些影响因子对预测目标是重要的^[9]。该算法主要步骤为:(1)为每个 EM 影响因子随机生成 1 个对应的阴影特征;(2)利用随机森林(RF)算法计算每个影响因子与阴影特征对 EM 的重要性记为 Z-

score,将阴影特征中的最大 Z-score 记为最大 Z 得分(MZSA);(3)当某影响因子的 Z-score 大于 MZSA 则将其标记 EM 的重要影响因子,反之标记为不重要;(4)重复上述操作直至所有影响因子均被标记或达到预设的迭代次数。

1.4 基于 DL 的 EM 计算模型构建方法

本研究基于 DL 算法构建松花江流域的土壤侵蚀模数计算模型(简称 EM 模型)。

1.4.1 数据集制作 为提高侵蚀评价模型的泛化能力,随机抽取1 000个样点。以任一样点为中心在研究区域 2001—2020 年的数据中,逐年抽取总容量为 $(1\ 000 \times 20)20\ 000$ 的 $16\ \text{m} \times 16\ \text{m}$ 大小的像元(像元面积为 $25\ 600\ \text{m}^2$)作为样本。其中,将 2001 年、2005 年、2010 年、2015 年及 2020 年对应 $(1\ 000 \times 5)5\ 000$ 个样本作为验证数据,其余年份均作为训练数据。分别计算每个像元的 EM_{wa} 和 EM_{wi} 。

1.4.2 DL 模型构造与超参数设置 相比其他 DL 模

型,卷积神经网络(CNN)在多维数据处理上具有更明显的优势^[25]。因土壤侵蚀影响因子较多,导致 CNN 的输入数据维度较高,故采用 CNN 模型中的 ResNet18^[26]来构建土壤侵蚀模数计算模型。采用 Relu 函数为卷积后的激活函数、MSE 为损失函数、adam 为模型优化器、使用全局平均池化(GlobalAveragePool)代替全连接以适应输入数据的分辨率,以上编程环境为 python 3.9。

详细网络结构见图 3,其中 Conv 代表卷积层。

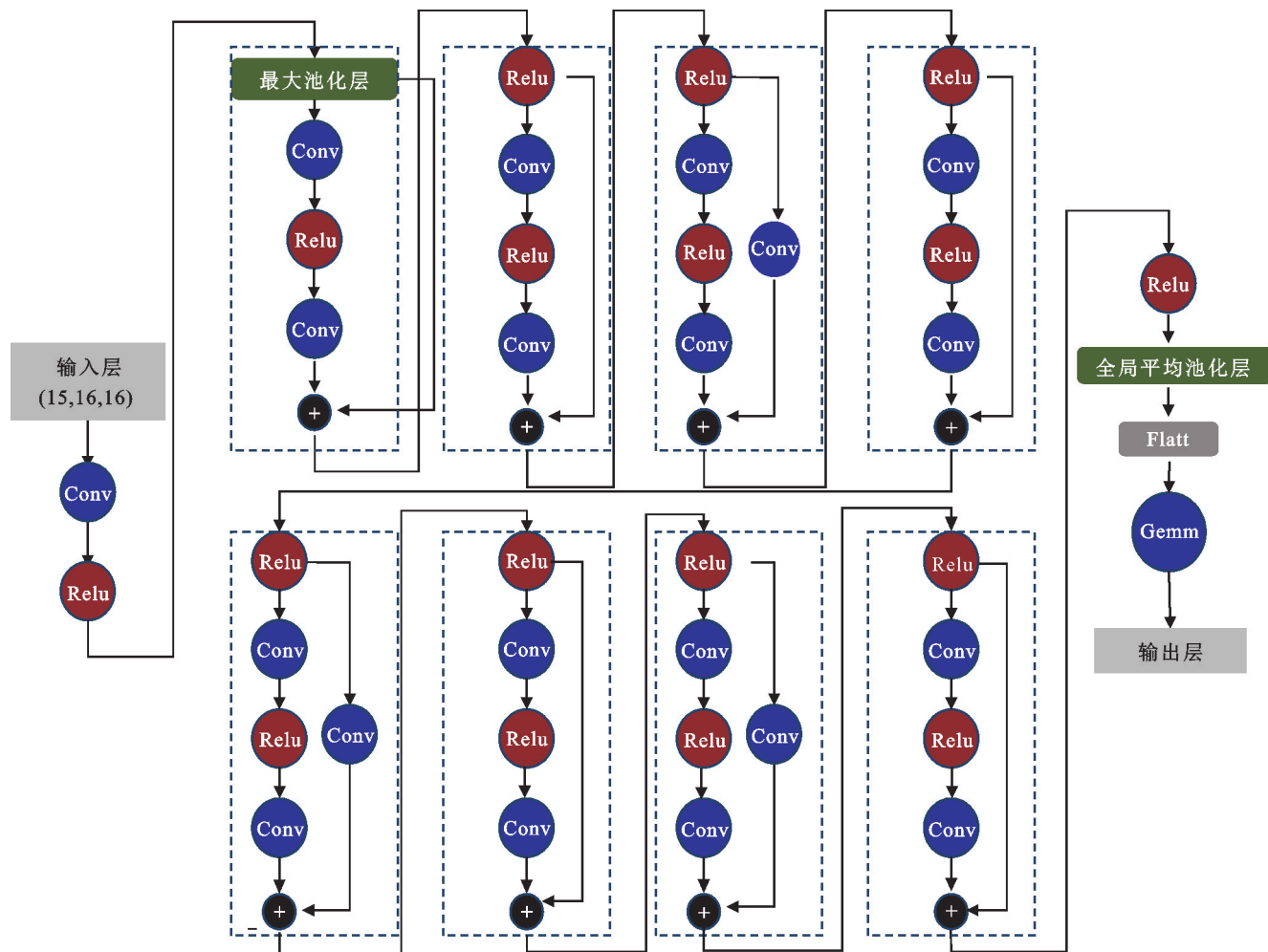


图 3 ResNet18 的网络结构

Fig. 3 The Network Structure of ResNet18

1.5 土壤侵蚀强度概率评价方法

相比于其他土壤侵蚀影响因子,降雨量、风速、气温具有较强的随机性^[27-29]。因此,本研究将这 3 个影响因子作为随机变量,依据其概率分布,以数值模拟方法制作模拟样本,并结合高斯核密度估计法(gaussian kernel density estimation, GKDE)^[30]对 EM 模型计算结果进行概率评价。

(1)根据本研究使用的 2001—2020 年 36 个气象站点的 3 个随机变量实测数据,采用 K-S 检验^[30]法拟合其理论分布。各气象站 3 个随机变量理论分布拟合结果见表 2。

(2)对于每个气象站,依据其 3 个随机变量的理论分布,采用数值模拟方法生成各变量的伪随机数 1 000 个作为随机样本;以该站所在像元作为典型像元,将其 3 个随机变量的样本(1 000 个)分别与 2020 年其他土壤侵蚀影响因子(忽略其他影响因子的随机性)组合成 1 000 个模拟样本向量,作为土壤侵蚀模数计算模型的输入。

(3)将全部模拟样本输入到训练好的 EM 计算模型中得到相应的 EM_{wa} 与 EM_{wi} ;并采用高斯核密度估计法^[30]绘制 EM_{wa} 与 EM_{wi} 的联合分布;再通过积分计算不同 EM_{wa} 与 EM_{wi} 强度等级组合的发生概率。

表 2 松花江流域 36 个气象站点随机变量分布类型

Table 2 The distribution types of the random variables at 36 meteorological stations in the Songhua River basin

站点名	年降雨量		年平均风速		年平均日最高气温	
	分布类型	p 值	分布类型	p 值	分布类型	p 值
安达	对数正态	0.97	F 分布	0.91	F 分布	0.56
白城	对数正态	0.95	F 分布	0.93	正态分布	0.85
北安	瑞利分布	0.84	瑞利分布	0.34	F 分布	0.98
东岗	瑞利分布	0.99	F 分布	0.68	正态分布	0.87
敦化	对数正态	0.99	F 分布	0.70	正态分布	0.50
扶余	F 分布	0.91	对数正态	0.86	F 分布	0.63
富锦	F 分布	0.95	F 分布	0.98	F 分布	0.94
富裕	F 分布	0.94	F 分布	0.93	正态分布	0.73
哈尔滨	对数正态	0.98	F 分布	0.93	F 分布	0.99
海伦	对数正态	0.99	F 分布	0.88	正态分布	0.90
桦甸	卡方分布	0.99	F 分布	0.95	F 分布	0.54
佳木斯	F 分布	0.99	对数正态	0.17	F 分布	0.99
蛟河	伽马分布	0.99	F 分布	0.74	正态分布	0.75
靖宇	F 分布	0.93	对数正态	0.96	正态分布	0.76
克山	F 分布	0.89	F 分布	0.75	正态分布	0.90
梅河口	F 分布	0.95	对数正态	0.95	正态分布	0.33
明水	F 分布	0.92	F 分布	0.46	F 分布	0.93
嫩江	F 分布	0.86	F 分布	0.98	正态分布	0.95
盘石	瑞利分布	0.58	F 分布	0.77	正态分布	0.53
齐齐哈尔	瑞利分布	0.98	对数正态	0.88	均匀分布	0.79
前郭尔罗斯	瑞利分布	0.90	F 分布	0.80	均匀分布	0.78
乾安	瑞利分布	0.99	F 分布	0.73	F 分布	0.40
尚志	F 分布	0.97	F 分布	0.75	F 分布	0.97
松江	瑞利分布	0.96	F 分布	0.78	正态分布	0.73
索伦	F 分布	0.96	对数正态	0.80	正态分布	0.66
泰来	F 分布	0.94	瑞利分布	0.10	正态分布	0.80
铁力	F 分布	0.93	指数分布	0.67	正态分布	0.95
通河	对数正态	0.89	对数正态	0.95	瑞利分布	0.89
通榆	对数正态	0.93	F 分布	0.60	正态分布	0.60
小二沟	对数正态	0.86	F 分布	0.73	正态分布	0.94
伊春	F 分布	0.93	对数正态	0.98	F 分布	0.90
依兰	瑞利分布	0.36	对数正态	0.95	F 分布	0.99
永吉	伽马分布	0.86	F 分布	0.81	F 分布	0.59
扎兰屯	对数正态	0.51	对数正态	0.52	F 分布	0.59
长岭	伽马分布	0.80	F 分布	0.34	F 分布	0.47
肇州	F 分布	0.78	F 分布	0.92	均匀分布	0.80

注:显著性水平 $p=0.05$ 。

2 结果与分析

2.1 EM 影响因子重要性验证结果

为降低模型输入数据的噪声及冗余、提高 EM 模型的鲁棒性和计算效率,采用 BORUTA 算法^[14]分析 16 个 EM 影响因子的重要性,进而剔除 EM 的弱相关变量。

将 BORUTA 算法^[14]每次迭代计算的各因子

(包括 MZSA) Z-score 绘制成箱线图(图 4),并用不同颜色分类表示重要因子(绿)、不重要因子(红)及 MZSA(蓝)。由图 4 可以看出,所有影响因子均为 EM_{wa} 的重要因子;除坡向与 SPI 外,所有影响因子均为 EM_{wi} 的重要因子。由于同一影响因子对 2 种侵蚀类型的重要性得分差异较大,为便于比并有效地筛选

对 2 种侵蚀都有重要影响的因子,从而剔除对 2 种侵蚀均影响较小的因子,现定义相对重要性(relative importance, RI)指标,其计算方法为:

$$RI_i = (Z_i - MZSA_i) / (Z_{\max-i} - MZSA_i) \quad (4)$$

式中: RI_i 为某因子对 EM_{wa} 或 EM_{wi} 的相对重要性(百分数计),分别记为 RI_{wa} 、 RI_{wi} ; Z_i 为对应侵蚀类型 EM 下该因子的 Z-score,分别记为 Z_{wa} 、 Z_{wi} ; $MZSA_i$ 为对应侵蚀类型 EM 的 MZSA,分别记为 $MZSA_{wa}$ 、 $MZSA_{wi}$; $Z_{\max-i}$ 为对应侵蚀类型 EM 影响因子的 Z-score 最大值,分别记为 $Z_{\max-wa}$ 、 $Z_{\max-wi}$; Z_i 、 $MZSA_i$ 、 $Z_{\max-i}$ 均取 BORUTA 算法多次迭代后的中

位数。 RI_{wa} 、 RI_{wi} 计算结果见图 5。由图 5 可以看出, RI_{wi} 的最大值与最小值分别为 NDVI(100%)和坡向(-6.72%); RI_{wa} 的最大值与最小值分别为除坡度(100%)和坡向(12.32%)。

综合对图 4、图 5 的分析结果可知, RI_{wa} 与 RI_{wi} 的最小值均为坡向,且该因子为 EM_{wi} 的不重要因子,因此坡向不作为 EM 模型的输入变量; SPI 作为 EM_{wi} 的另一个不重要因子,由于其对于 EM_{wi} 的相对重要性较高(71.86%),因此保留该因子作为 EM 模型的输入变量。综上,剔除 EM 模型唯一的弱相关变量(坡向),使其输入变量为除坡向外的 15 个影响因子。

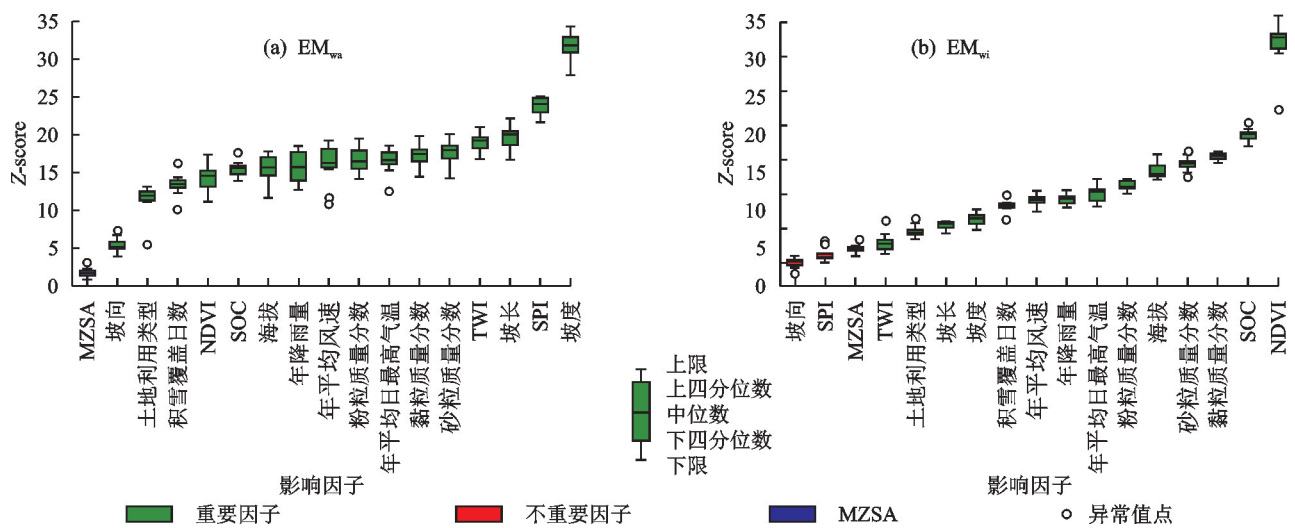


图 4 土壤侵蚀模数影响因子重要性分析结果

Fig. 4 Analysis results of the importance of soil erosion modulus impact factors

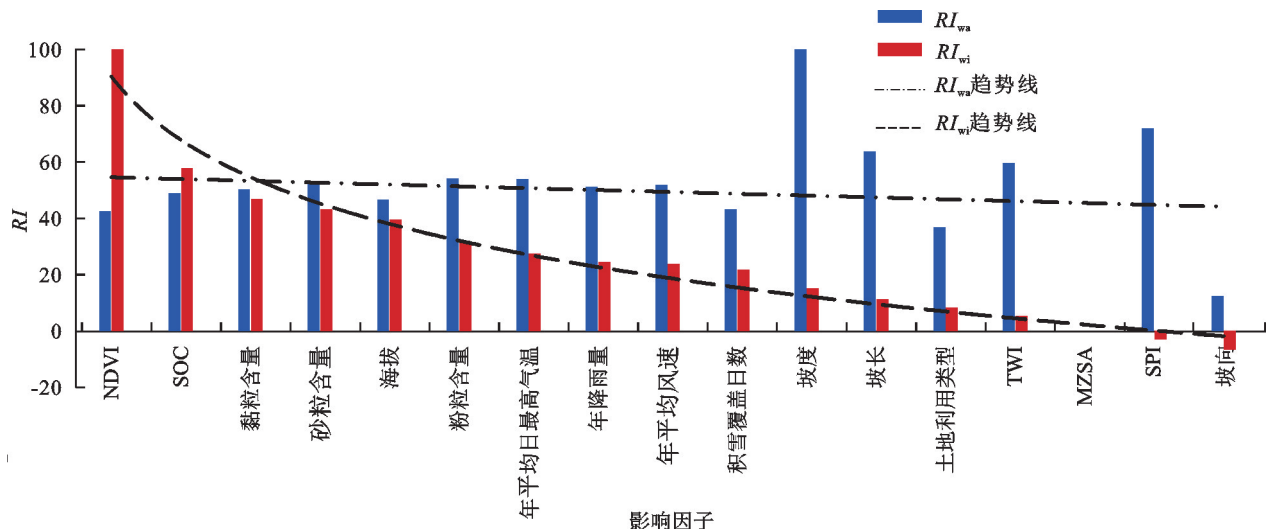


图 5 各影响因子的 RI_{wa} 、 RI_{wi}

Fig. 5 The RI_{wa} and RI_{wi} of each impact factor

2.2 EM 模型的构建及应用

2.2.1 模型的构建及验证 EM 模型输入为已确定的 15 个重要影响因子对应训练样本,输出为水蚀模数和风蚀模数,模型训练次数设定为 150 次,模

拟收敛条件设定为 $MSE < 5$,每次训练所对应的损失值迭代曲线见图 6。图 6 显示模型损失下降梯度较快,在第 120 次训练后(4.81),EM 模型达到收敛条件,模型构建成功。为进一步验证收敛后土壤侵蚀模

型的有效性,现计算验证数据中每年对应侵蚀模数的决定系数(R^2)并绘制 EM_{wa} 、 EM_{wi} 的实际值与模拟值散点图(表 3、图 7)可以看出,验证期的 R^2 均 >0.86 。综合以上分析表明,所构建的 EM 模型具有较高计算精度,可用于松花江流域土壤侵蚀模数的预测。

2.2.2 松花江流域 EM 计算结果 基于松花江流域

所有像元的土壤侵蚀影响因子数据,利用已构建的 EM 模型计算整个流域的 EM_{wa} 、 EM_{wi} (空间分辨率为 $1\ 600\ m\times 1\ 600\ m$),根据土壤侵蚀分类分级标准结合松花江流域土壤侵蚀实际情况,分别将 EM_{wa} 与 EM_{wi} 划分为微度、轻度、中度、重度 4 个强度等级,将某一水蚀强度等级与某一风蚀强度等级组合命名为侵蚀强度组合,不同侵蚀强度组合见表 4。

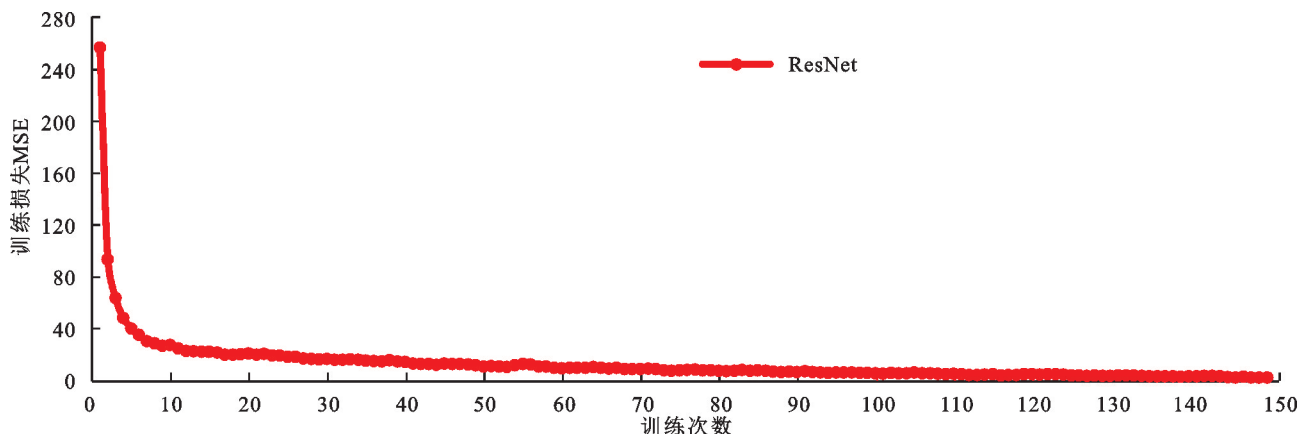


图 6 EM 模型每次训练对应损失值

Fig. 6 A loss value curve for each training of the EM model

表 3 验证期 EM 模型 R^2

Table 3 Validation period EM calculation model R^2					
EM 类型	2001 年	2005 年	2010 年	2015 年	2020 年
EM_{wa}	0.88	0.87	0.91	0.92	0.91
EM_{wi}	0.86	0.88	0.91	0.93	0.90

基于 EM 模型计算得出的松花江流域 EM_{wa} 、 EM_{wi} ,绘制验证期(2001 年、2005 年、2010 年、2015 年、2020 年)流域内侵蚀强度组合空间分布图(图 8),并统计验证期每年各等级 EM 占流域总面积的比例(表 5)。由图 8 可知,松花江流域土壤侵蚀强度组合空间分布特征体现为东西两侧山脉水蚀强度高于风蚀强度、中部平原风力侵蚀强度高于水蚀强度,与图 1 所反映的东西两侧山脉 $EM_{wa}>EM_{wi}$ 、中部风蚀模数 $EM_{wi}>EM_{wa}$ 的规律一致。说明所构建的系统模型在松花江流域具有较好的适用性。由表 5 可知,流域内平均每年发生微度水蚀与微度风蚀的面积占比为 74.47%,该强度组合面积随时间推移呈下降趋势,说明流域内土壤侵蚀强度有逐渐加重的趋势。流域内平均每年发生轻度水蚀与微度风蚀的面积占比为 12.69%,微度水蚀与轻度风蚀的面积占比为 9.47%,微度水蚀与中度风蚀的面积占比为 1.75%,微度水蚀与重度风蚀的面积占比为 1.33%,轻度水蚀与轻度风蚀的面积占比为 0.17%,轻度水蚀轻度风蚀的面积占比为 0.10%,其余水蚀与风蚀强度组合面积占比极小($<0.01\%$)。

2.3 松花江流域土壤侵蚀概率评价

采用高斯核函数估计方法计算系统模型输出的

36 个典型像元(以对应的气象站名称命名) EM_{wa} 与 EM_{wi} 联合密度,并绘制联合密度曲面图(图 9)。对联合分布函数求积分,即可得到不同侵蚀强度组合发生的概率,对 36 个像元各侵蚀强度组合发生概率求平均值(表 6)。由表 6 可知,36 个典型像元中,发生微度水蚀与微度风蚀的平均概率为 57.45%;发生微度水蚀与轻度风蚀的平均概率为 30.26%;发生微度水蚀与中度风蚀的平均概率为 8.03%;发生轻度水蚀与微度风蚀的平均概率为 2.11%;发生微度水蚀与重度风蚀的平均概率为 2.08%;发生轻度水蚀与轻度风蚀的平均概率为 0.04%;发生其余侵蚀强度组合的平均概率均 $\leq 0.01\%$ 。

3 讨论

本研究中 EM 的预测结果表明,研究区内大部分区域 EM 较低,两侧山脉 EM_{wa} 较高、中部平原 EM_{wi} 较高,与已有研究^[2,5-6]发现的松花江流域土壤侵蚀以微度为主、水力侵蚀集中在长白山脉及大兴安岭地区、风力侵蚀集中在中南部平原的成果相符,即松花江流域内山区更易发生水力侵蚀、平原则容易发生风力侵蚀。本研究中 2020 年 EM_{wa} 明显高于其他年份,其原因可能是 2020 年流域内降雨量较高,导致降雨侵蚀力增加进而使得流域内 EM_{wa} 升高,与吴瀚逸等^[6]提出的 2001—2020 年中国东北区域土壤水蚀数据集中所表现的 2020 年松花江流域内轻度以上水蚀面积明显大于其他年份的结果是一致的。

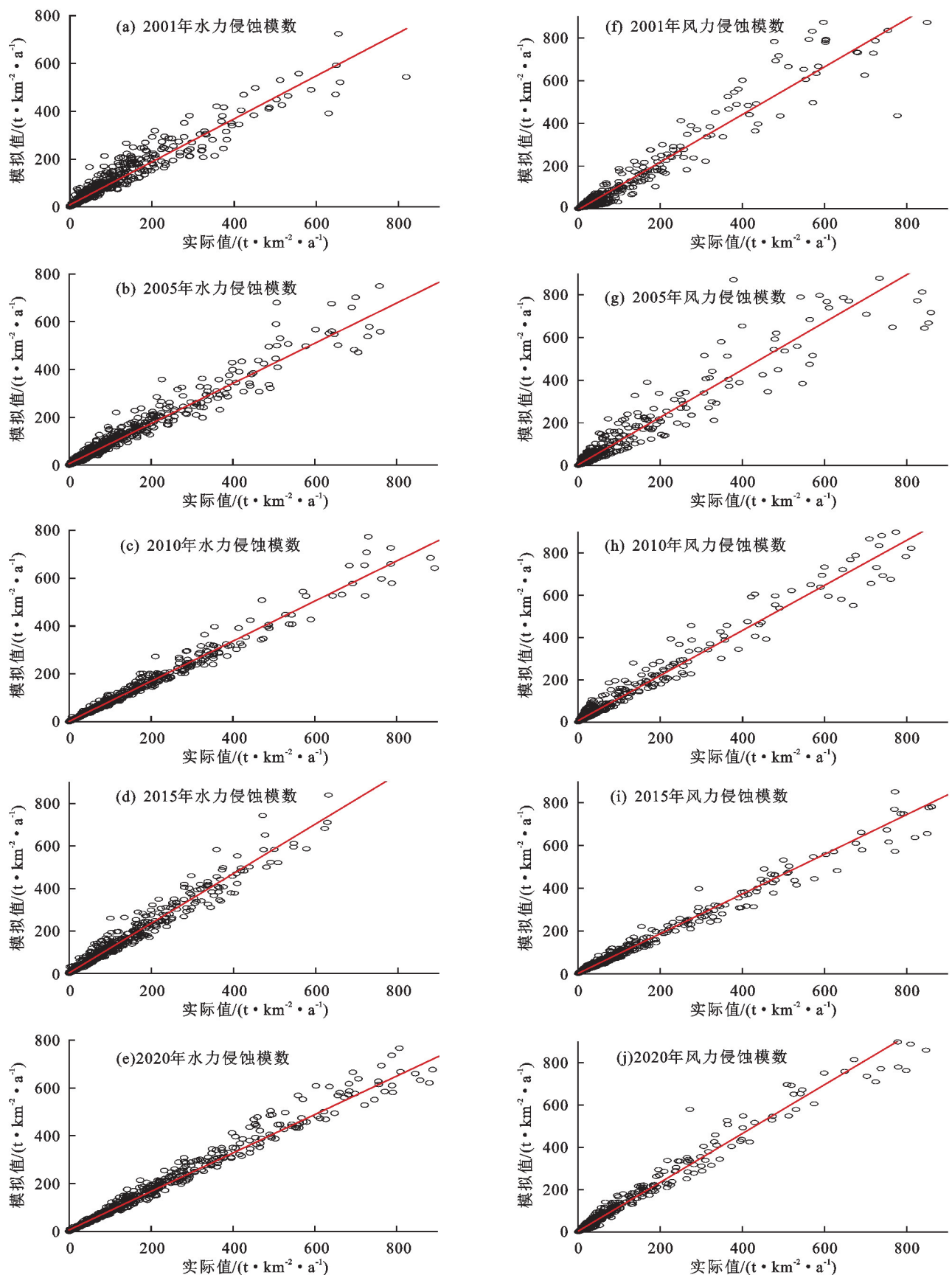


图 7 EM 模拟值与实际值散点图

Fig. 7 Scatter plot of simulated and actual soil erosion values

本研究发现,松花江流域水力侵蚀与风力侵蚀空间上呈近似负相关^[5],与左小峰等^[4]黑土风洞试验发现风蚀对坡面水蚀有一定促进作用的结论不同。究其原因可能是本研究的流域空间、年际时间的尺度较

大,难以有效表现年内小尺度条件下土壤侵蚀的变化规律,因此,有必要在未来对不同时空尺度的土壤侵蚀变化差异开展进一步的研究。REZAEI 等^[12]对基于深度学习的风蚀模型结果的不确定性进行量化分析^[12]。由于其输入变量的随机性研究不够深入,故

本研究基于降雨、气温、风速 3 个随机变量的概率密度进行数值模拟,考虑土壤侵蚀影响因子的随机性;同时使用高斯核密度法给出基于数值模拟结果计算出的不同强度等级 EM 组合的发生概率,更直观反映出土壤侵蚀的随机性。

表 4 侵蚀强度组合

Table 4 Erosion intensity level combination

水蚀强度	风蚀强度			
	微度 [<200 t/(km ² · a)]	轻度 [200~2 500 t/(km ² · a)]	中度 [2 500~5 000 t/(km ² · a)]	重度 [>5 000 t/(km ² · a)]
微度[<200 t/(km ² · a)]	11	12	13	14
轻度[200~2 500 t/(km ² · a)]	21	22	23	24
中度[2 500~5 000 t/(km ² · a)]	31	32	33	34
重度[>5 000 t/(km ² · a)]	41	42	43	44

注:十位数字代表水蚀强度,个位数字代表风蚀强度。

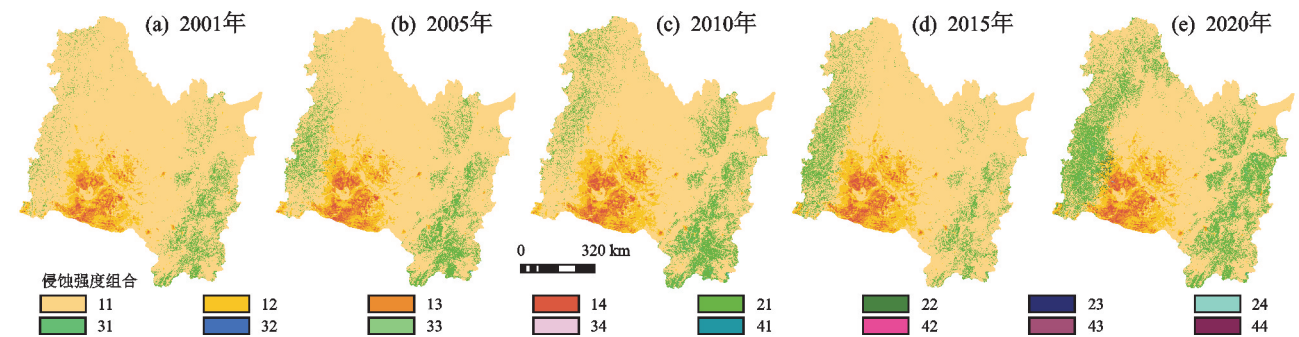


图 8 验证期侵蚀强度组合空间分布

Fig. 8 Spatial distribution of erosion intensity combinations during the validation period

表 5 验证期各土壤侵蚀强度组合面积占比

Table 5 Area share of each soil erosion intensity combination during the validation period

侵蚀强度组合	2001 年	2005 年	2010 年	2015 年	2020 年	年平均值
11	81.49	77.31	72.55	75.93	65.08	74.47
12	8.92	8.69	10.33	10.28	9.14	9.47
13	1.44	1.69	1.94	1.94	1.74	1.75
14	0.94	1.33	1.51	1.51	1.36	1.33
21	7.18	10.86	13.33	10.25	21.81	12.69
22	0.01	0.03	0.01	0.06	0.40	0.10
23	0	0	0	0	0.03	0.01
24	0	0	0	0	0.01	0
31	0.01	0.08	0.32	0.02	0.42	0.17
32	0	0	0	0	0	0
33	0	0	0	0	0	0
34	0	0	0	0	0	0
41	0	0	0.01	0	0	0
42	0	0	0	0	0	0
43	—	—	0	0	0	0
44	—	—	0	—	0	0
总计	100.00	100.00	100.00	100.00	100.00	100.00

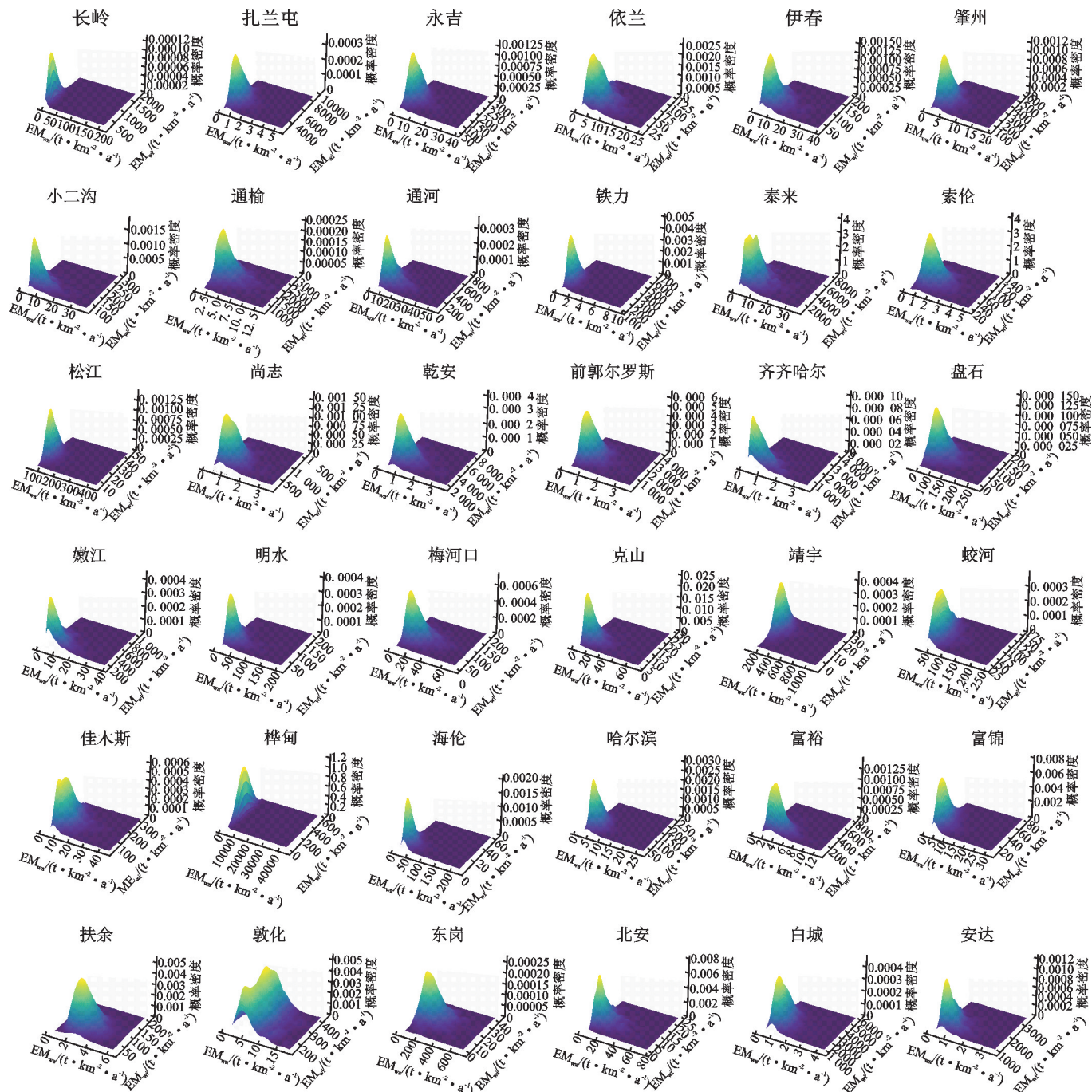


图 9 松花江流域内 36 个典型像元 EM_{wa} 与 EM_{wi} 联合分布

Fig. 9 Joint distribution of EM_{wa} and EM_{wi} for the 36 typical pixels in the Songhua River Basin

表 6 36 个典型像元各侵蚀强度组合的平均发生概率
Table 6 Average probability of occurrence for each combination of erosion intensities for 36 typical image elements %

水蚀强度	风蚀强度			
	微度	轻度	中度	重度
微度	57.45	30.26	8.03	2.08
轻度	2.11	0.04	—	—
中度	0.01	—	—	—
重度	0.01	0.01	—	—

各典型像元中,发生轻度及以上风蚀与微度水蚀

组合的概率较高,其原因主要是流域中部风力侵蚀较重的区域典型像元较多,导致各典型像元发生轻度风蚀与微度水蚀、中度风蚀与微度水蚀、重度风蚀与微度水蚀 3 个侵蚀强度组合的平均发生概率偏高,同时说明松花江流域中部区域风力侵蚀强度较大,应采取相应的治理措施。

本文采用吴瀚逸等^[6]制作的土壤水蚀数据集及 RWEQ 计算得到的风蚀模数数据作为训练数据,构建松花江流域的土壤侵蚀模数(EM)计算模型,并且模型具有较高的精度。对于松花江流域以外的东北地区,由于地形、气象、土壤及土地覆盖等因素未发生

显著变化,模型精度仍有保障。然而,对于黄土高原、青藏高原等自然条件与东北地区存在显著差异的地区,该模型需重新训练后方可适用。

4 结论

(1)利用 Boruta 算法对可能影响松花江流域土壤侵蚀模数(EM)的 16 个因子进行的验证。结果显示,除坡向因子外,其余 15 个影响因子均为松花江流域 EM 的重要影响因子。

(2)通过构建基于 DL 的 EM 的计算模型,评价松花江流域内土壤侵蚀强度组合空间分布。结果显示,流域内平均每年有 74.47% 发生微度水蚀与微度风蚀,12.86% 的面积发生轻度及以上水蚀与微度风蚀,12.56% 的面积发生轻度及以上风蚀与微度水蚀,0.11% 的面积水蚀强度与风蚀强度均在轻度及以上。

(3)采用非参数拟合技术及高斯核密度估计(GKDE),对松花江流域内 36 个典型像元土壤侵蚀强度进行概率评价。结果显示,36 个典型像元中,发生微度水蚀与微度风蚀的平均概率为 57.45%,发生微度水蚀与轻度风蚀的平均概率为 30.26%,发生微度水蚀与中度风蚀的平均概率为 8.03%,发生轻度水蚀与微度风蚀的平均概率为 2.11%,发生微度水蚀与重度风蚀的平均概率为 2.08%,发生轻度水蚀与轻度风蚀的平均概率为 0.04%,发生其余侵蚀强度组合的平均概率均 $\leq 0.01\%$ 。

(4)构建的基于 DL 和 GKDE 的土壤侵蚀概率评价模型能够有效地揭示松花江流域侵蚀的空间分布特征,同时给出 2 种侵蚀类型不同强度组合的发生概率。研究成果拓展了土壤侵蚀评价结果的表现形式,同时可为松花江流域土壤侵蚀治理提供决策依据。

参考文献:

- [1] ZHU Y L, LI W B, WANG D Y, et al. Spatial pattern of soil erosion in relation to land use change in a rolling hilly region of Northeast China[J]. *Land*, 2022, 11(8): e1253.
- [2] WANG X, ZHAO X L, ZHANG Z X, et al. Assessment of soil erosion change and its relationships with land use/cover change in China from the end of the 1980s to 2010[J]. *Catena*, 2016, 137: 256-268.
- [3] WANG H, YANG S L, WANG Y D, et al. Rates and causes of black soil erosion in Northeast China[J]. *Catena*, 2022, 214: e106250.
- [4] 左小锋,郑粉莉,张加琼,等.典型薄层黑土区前期地表风蚀作用影响坡面水蚀的研究[J]. *土壤学报*, 2021, 58(5): 1145-1156.
- ZUO X F, ZHENG F L, ZHANG J Q, et al. Study on effect of surface wind erosion on hillslope water erosion in regions of typical thin layered mollisol at early stages [J]. *Acta Pedologica Sinica*, 2021, 58(5): 1145-1156.
- [5] WANG S H, XU X L, HUANG L. Spatial and temporal variability of soil erosion in Northeast China from 2000 to 2020[J]. *Remote Sensing*, 2022, 15(1): e225.
- [6] 吴瀚逸,熊俊峰,侯渲,等.2001—2020 年中国东北区域土壤水蚀数据集[J]. *中国科学数据*, 2023, 8(4): 283-297.
- WU H Y, XIONG J F, HOU X, et al. A dataset of soil water erosion of Northeast China from 2001 to 2020[J]. *China Scientific Data*, 2023, 8(4): 283-297.
- [7] 杨严攀,田培,沈晨竹,等.基于 RUSLE 模型和地理探测器的鄂西南土壤侵蚀脆弱性评价[J]. *水土保持学报*, 2024, 38(1): 91-103.
- YANG Y P, TIAN P, SHEN C Z, et al. Vulnerability assessment of soil erosion in southwestern Hubei Province based on RUSLE model and geographic detector [J]. *Journal of Soil and Water Conservation*, 2024, 38(1): 91-103.
- [8] 池金铭,于洋,冯娟龙,等.基于 RUSLE 模型的奶水河流域土壤侵蚀时空变化特征[J]. *水土保持学报*, 2024, 38(1): 70-78.
- CHI J M, YU Y, FENG J L, et al. Spatial and temporal variation characteristics of soil erosion in Guishui River Basin based on RUSLE[J]. *Journal of Soil and Water Conservation*, 2024, 38(1): 70-78.
- [9] GHOLAMI H, MOHAMMADIFAR A, GOLZARI S, et al. Using the Boruta algorithm and deep learning models for mapping land susceptibility to atmospheric dust emissions in Iran[J]. *Aeolian Research*, 2021, 50: e100682.
- [10] KHOSRAVI K, REZAEI F, COOPER J R, et al. Soil water erosion susceptibility assessment using deep learning algorithms [J]. *Journal of Hydrology*, 2023, 618: e129229.
- [11] 郑粉莉,张加琼,刘刚,等.东北黑土区坡耕地土壤侵蚀特征与多营力复合侵蚀的研究重点[J]. *水土保持通报*, 2019, 39(4): 314-319.
- ZHENG F L, ZHANG J Q, LIU G, et al. Characteristics of soil erosion on sloping farmlands and key fields for studying compound soil erosion caused by multi-forces in mollisol region of Northeast China[J]. *Bulletin of Soil and Water Conservation*, 2019, 39(4): 314-319.
- [12] REZAEI M, MOHAMMADIFAR A, GHOLAMI H, et al. Mapping of the wind erodible fraction of soil by bidirectional gated recurrent unit (BiGRU) and bidirectional recurrent neural network (BiRNN) deep learning models[J]. *Catena*, 2023, 223: e106953.
- [13] 王善序.贝叶斯概率水文预报简介[J]. *水文*, 2001, 21(5): 33-34.
- WANG S X. Introduction to the Bayesian forecasting system[J]. *Hydrology*, 2001, 21(5): 33-34.

- [14] CIMUSA K L, BIGABWA B J, PRASAD P, et al. Soil erosion susceptibility mapping using ensemble machine learning models: A case study of upper Congo River sub-basin[J]. *Catena*, 2023, 222: e106858.
- [15] BAG R, MONDAL I, DEHBOZORGI M, et al. Modelling and mapping of soil erosion susceptibility using machine learning in a tropical hot sub-humid environment[J]. *Journal of Cleaner Production*, 2022, 364: e132428.
- [16] 张光辉, 杨扬, 刘瑛娜, 等. 东北黑土区土壤侵蚀研究进展与展望[J]. *水土保持学报*, 2022, 36(2): 1-12.
- ZHANG G H, YANG Y, LIU Y N, et al. Advances and prospects of soil erosion research in the black soil region of Northeast China[J]. *Journal of Soil and Water Conservation*, 2022, 36(2): 1-12.
- [17] 焦剑, 谢云, 林燕, 等. 东北地区融雪期径流及产沙特征分析[J]. *地理研究*, 2009, 28(2): 333-344.
- JIAO J, XIE Y, LIN Y, et al. Study on snowmelt runoff and sediment yields in Northeast China[J]. *Geographical Research*, 2009, 28(2): 333-344.
- [18] 张晓敏, 张东梅, 王莉, 等. 降雨、积雪以及土地利用复合影响下的额尔齐斯河流域土壤侵蚀分析[J]. *水土保持学报*, 2022, 36(5): 104-111.
- ZHANG X M, ZHANG D M, WANG L, et al. Soil erosion analysis in the Irtysh River Basin under the combined effects of rainfall, snow cover and land use[J]. *Journal of Soil and Water Conservation*, 2022, 36(5): 104-111.
- [19] 赵琴, 郝晓华, 王建, 等. 2000—2020 年 MODIS 中国积雪物候数据集[D/OL]. 国家冰川冻土沙漠科学数据中心 (www.ncdc.ac.cn), 2021. <https://cstr.cn/CSTR:11738.11.ncdc.isnow.db2519>, 2022.
- ZHAO Q, HAO X H, WANG J, et al. A dataset of snow cover phenology in China based on MODIS during 2000—2020[D/OL]. National Cryosphere Desert Data Center (www.ncdc.ac.cn), 2021. <https://cstr.cn/CSTR:11738.11.ncdc.isnow.db2519>, 2022.
- [20] 赵琴, 郝晓华, 王建, 等. 2000—2020 年 MODIS 中国积雪物候数据集[J]. *中国科学数据*, 2022, 7(3): 60-69.
- ZHAO Q, HAO X H, WANG J, et al. A dataset of snow cover phenology in China based on MODIS during 2000—2020[J]. *China Scientific Data*, 2022, 7(3): 60-69.
- [21] YANG J, HUANG X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019[J]. *Earth System Science Data*, 2021, 13(8): 3907-3925.
- [22] 刘峰, 张甘霖. 中国高分辨率国家土壤信息网格基本属性数据集(2010—2018)[ED/OL]. [2024-05-11] 国家青藏高原数据中心, 2021. <https://doi.org/10.11666/00073.ver1.db>.
- LIU F, ZHANG G L. Basic soil property dataset of high-resolution China Soil Information Grids (2010—2018)[ED/OL]. [2024-05-11] National Tibetan Plateau Data Center, 2021. <https://dx.doi.org/10.11666/00073.ver1.db>.
- [23] LIU F, ZHANG G L, SONG X D, et al. High-resolution and three-dimensional mapping of soil texture of China[J]. *Geoderma*, 2020, 361: e114061.
- [24] 巩国丽, 刘纪远, 邵全琴. 基于 RWEQ 的 20 世纪 90 年代以来内蒙古锡林郭勒盟土壤风蚀研究[J]. *地理科学进展*, 2014, 33(6): 825-834.
- GONG G L, LIU J Y, SHAO Q Q. Wind erosion in Xilingol League, Inner Mongolia since the 1990s using the Revised Wind Erosion Equation[J]. *Progress in Geography*, 2014, 33(6): 825-834.
- [25] 张天骐, 柏浩钧, 叶绍鹏, 等. 基于注意力门控膨胀卷积网络的单通道语音增强[J]. *电子与信息学报*, 2022, 44(9): 3277-3288.
- ZHANG T Q, BAI H J, YE S P, et al. Monaural speech enhancement based on attention-gate dilated convolution network[J]. *Journal of Electronics and Information Technology*, 2022, 44(9): 3277-3288.
- [26] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [27] 王红瑞, 林欣, 刘琼, 等. 基于系统聚类的北京气温变化过程遍历特征分析[J]. *系统工程理论与实践*, 2010, 30(2): 193-200.
- WANG H R, LIN X, LIU Q, et al. Ergodicity analysis based on system clustering for temperature variation process of Beijing[J]. *Systems Engineering-Theory and Practice*, 2010, 30(2): 193-200.
- [28] 贺敬, 李少林, 蔡玮, 等. 双重不确定性预测下风电场超短期有功优化控制[J]. *太阳能学报*, 2023, 44(11): 270-278.
- HE J, LI S L, CAI W, et al. Optimal control of wind farm power based on double uncertainty prediction[J]. *Acta Energetica Sinica*, 2023, 44(11): 270-278.
- [29] 庄琦, 刘曙光, 周正正. 降雨数据时空精度对城市暴雨变异性及频率分析的影响[J]. *水科学进展*, 2023, 34(3): 398-408.
- ZHUANG Q, LIU S G, ZHOU Z Z. Impact of rainfall spatiotemporal resolutions on urban extreme rainfall variability and rainfall frequency analysis[J]. *Advances in Water Science*, 2023, 34(3): 398-408.
- [30] 程媛, 迟荣华, 黄少滨, 等. 基于非参数密度估计的不确定轨迹预测方法[J]. *自动化学报*, 2019, 45(4): 787-798.
- CHENG Y, CHI R H, HUANG S B, et al. Uncertain trajectory prediction method using non-parametric density estimation[J]. *Acta Automatica Sinica*, 2019, 45(4): 787-798.