

doi: 10.6046/zrzyyg.2023258

引用格式: 胡博洋, 孙建国, 张倩, 等. 用于植被变化归因的区域机器学习残差趋势法[J]. 自然资源遥感, 2025, 37(1): 46–53.
(Hu B Y, Sun J G, Zhang Q, et al. Residual trend method based on regional modeling and machine learning for attribution of vegetation changes[J]. Remote Sensing for Natural Resources, 2025, 37(1): 46–53.)

用于植被变化归因的区域机器学习残差趋势法

胡博洋^{1,2,3}, 孙建国^{1,2,3}, 张倩^{1,2,3}, 杨云睿^{1,2,3}

(1. 兰州交通大学测绘与地理信息学院, 兰州 730070; 2. 地理国情监测技术应用国家地方联合工程研究中心, 兰州 730070; 3. 甘肃省地理国情监测工程实验室, 兰州 730070)

摘要: 现有的残差趋势法采用逐像元建模策略, 利用普通最小二乘法构建模型, 存在着一定的局限性: 一方面, 逐像元建模策略使每个模型都包含了局部空间内的人类活动信号干扰; 另一方面, 普通最小二乘法不利于模拟普遍存在的非线性特征。因此, 该文提出一种全新的基于区域建模策略和机器学习算法的残差趋势法, 并对比了用于表达空间异质性的2种环境变量: ①地形、水文和土地利用等直接环境变量(direct-environmental variables, DEVs); ②植被和气候时空序列组合的代理环境变量(proxy-environmental variables, PEVs)。首先, 采用区域建模策略, 分别引入DEVs和PEVs, 使用机器学习算法构建植被-气候关系模型; 其次, 根据残差趋势法的定义得到残差值; 最后, 评估气候和人为因素对植被变化的贡献。结果表明: ①相比以往的逐像元普通最小二乘残差趋势法, 所提方法的优势不仅表现为机器学习能够模拟植被-气候关系的非线性特征, 还表现为区域建模具备更强的抗人类信号干扰能力; ②新方法中, 使用PEVs明显优于使用DEVs, 前者充分利用了原有建模数据, 没有增加数据获取难度, 也避免了引入额外的数据误差。该文提出的区域机器学习残差趋势法可以实现更有效的植被变化归因。

关键词: 植被变化归因; 区域建模; 机器学习算法; 残差趋势法; 空间异质性

中图分类号: TP 79 **文献标志码:** A **文章编号:** 2097-034X(2025)01-0046-08

0 引言

植被是陆地生态系统的重要组成部分, 其生理状况也是反映生态环境质量的核心指标^[1]。植被变化往往受到气候因素(特别是降水和气温)和人为因素的共同影响, 在这2种因素的强烈干预下, 全球范围内植被景观及其生态系统服务功能发生了巨大的改变^[2]。对此, 大范围和长时序的植被变化归因是必要的, 可以为评估生态环境质量和制定生态环境保护相关政策提供科学依据, 而卫星遥感植被指标数据为此提供了便利^[3]。

残差趋势法是植被变化归因的重要方法之一。该方法首先构建植被-气候年际变化关系的回归模型, 然后将残差值的趋势作为人为因素的影响^[4]。作为地理学分析方法, 残差趋势法必须面对空间异质性问题。现有残差趋势法采用了逐像元建模方式, 利用计算量较小的普通最小二乘法构建模型^[5–11]。虽然该方法的建模策略将植被-气候关系

的空间异质性体现为逐个模型参数差异, 回避了对空间异质性的模拟, 但同时也将局部区域中的人类活动干扰信号包含进了每个模型中, 相反, 区域建模策略能在一定程度上减弱这一缺陷。此外, 该方法使用的普通最小二乘法不利于模拟植被-气候关系中普遍存在的非线性特征。由此可见, 采用区域建模策略, 同时引入机器学习算法是改进残差趋势法的重要途径之一。然而, 此途径下, 模型中需要引入特定的(静态)环境变量以顾及植被-气候关系的空间异质性。

相关研究中, 学者们习惯于从植被生长的机理出发, 引入“直接环境变量”(direct-environmental variables, DEVs), 例如地形、水文和土地利用等^[12–16]。此类变量繁多易漏, 数据误差也在所难免, 这将影响模型精度和归因效果。实际上, 残差趋势法是建立于经验模型而非机理模型之上, 应当优先考虑依赖现有数据解决问题。单从空间维度来看, 植被和气候时空序列组合正是对每个像元综合地理特性的间接反映, 将这一组合作为固定的“静

收稿日期: 2023-08-22; 修订日期: 2024-02-01

基金项目: 甘肃省科技计划项目“甘肃省地表覆盖变化自动监测关键技术”(编号: 20YF3GA013)和兰州交通大学优秀平台(编号: 201806)共同资助。

第一作者: 胡博洋(1997–), 男, 硕士研究生, 主要从事植被变化归因等研究。Email: hxb1258@163.com。

通信作者: 孙建国(1974–), 男, 博士, 教授, 主要从事生态环境遥感与GIS应用研究。Email: sunjguo@mail.lzjtu.cn。

态”因子标记像元特征等,完全可以充当区域建模策略下表达空间异质性所需要的环境变量。为了区别于 DEVs,本文将植被和气候时空序列组合称为“代理环境变量”(proxy-environmental variables, PEVs)。

甘肃省是中国西部重要的生态安全屏障,其植被变化归因的现实需求十分紧迫。该省地理环境尤其是植被和气候类型复杂多样,作为植被变化归因方法的研究区极具代表性。本文以甘肃省为研究区,提出一种利用 DEVs 或 PEVs 顾及空间异质性的基于区域建模策略和机器学习算法的残差趋势法,并与已有的基于逐像元建模策略和普通最小二乘法的残差趋势法进行比较,验证其优势。由于构建的模型都服务于残差趋势法,本文将以上 2 种方法分

别简称为区域机器学习法和逐像元最小二乘法。

1 研究区概况及数据源

1.1 研究区概况

甘肃省(图 1)简称“甘”或“陇”,介于 $32^{\circ}11' \sim 42^{\circ}57'N$ 和 $92^{\circ}13' \sim 108^{\circ}46'E$ 之间,通常被分为河西、陇中、陇东、陇南和甘南 5 大区域。其中,陇中、陇东和陇南属于温带季风气候;甘南大部分地区和河西祁连山区主要为高寒气候;河西走廊区为温带干旱气候,且分布有大量的人工绿洲。截至 2022 年末,甘肃省常住人口约为 2 500 万人,河谷平原和绿洲的人口密度相对较大。

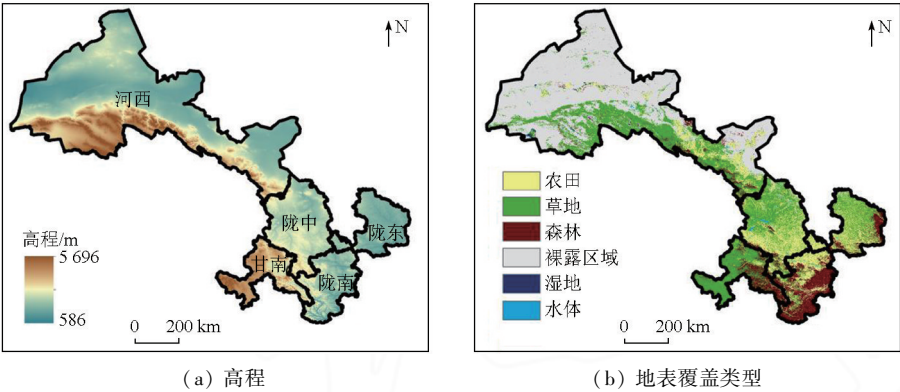


图 1 甘肃省概况
Fig.1 Overview of Gansu Province

1.2 数据源及预处理

1.2.1 植被数据

从美国航空航天局(<https://ladsweb.modaps.eosdis.nasa.gov/>)的 MOD13A3 产品中获取 2000—2020 年 1 km 分辨率的月度归一化植被指数(normalized difference vegetation index, NDVI)。利用 MODIS Reprojection Tool 对 NDVI 数据进行波段提取、拼接和重投影等预处理。计算生长季(4—9 月)均值代表年 NDVI ($aNDVI$),并去除无植被区(2000—2020 年平均 $aNDVI < 0.15$)。

1.2.2 气候数据

2000—2020 年的降水数据和气温数据选用国家气象科学数据中心(<https://data.cma.cn/>)提供的气象站点月度数据。利用 ANUSPLIN 软件^[17]插值得到 1 km 分辨率的月降水和月气温栅格数据。考虑到 NDVI 对气候变化的响应存在滞后性,参考姚楠等^[18]的处理方法,以 2—9 月的降水量总和代表年降水量(aP),2—9 月的气温均值代表年气温(aT)。

1.2.3 环境变量

参考相关文献^[12-16],结合数据的可获取性,本

文选取的 DEVs 包括 7 个变量,即:高程、坡度、到道路的距离、到水体的距离、夜间灯光、地表覆盖类型和初期植被状态。除初期植被状态,其余数据皆选用研究时段中期的数据。DEVs 的建模简称、属性及数据源见表 1。

表 1 区域机器学习法中的 DEVs			
Tab.1 DEVs in regional machine learning method			
变量类	简称	数据源	空间分辨率
高程	DEV ₁	www.resdc.cn	90 m
坡度	DEV ₂	www.resdc.cn	90 m
到道路的距离	DEV ₃	www.openstreetmap.org	矢量
到水体的距离	DEV ₄	www.geodata.cn	矢量
夜间灯光	DEV ₅	www.ngdc.noaa.gov	1 km
地表覆盖类型	DEV ₆	data.casearth.cn	30 m
初期植被状态	DEV ₇	ladsweb.modaps.eosdis.nasa.gov	1 km

PEVs 为植被和气候时空序列组合,在模型中,该组合实质上是一组无量纲数,作为“静态”因子标记逐个像元内的特征等,即: $aNDVI_j, aP_j, aT_j$, 共 3 类 $3n$ 个,其中, j 为年份序号, $j = 1, 2, \dots, n, 1$ 为 2000 年,2 为 2001 年,以此类推。为了消除 PEVs 数据之间的多重共线性,利用主成分分析法(principal

component analysis, PCA) 对 PEVs 做降维处理,保留累计贡献率大于特定阈值(本文设定为 95%)的 m 个主成分,即: $PCA_PEV_1, PCA_PEV_2, \dots, PCA_PEV_m$ 。

2 研究方法

2.1 区域机器学习模型构建

机器学习算法种类繁多,考虑到硬件条件和模拟工具的鲁棒性,本文选用随机森林算法^[19]。随机选取试验区内 70% 的样本点构建植被-气候关系的区域机器学习模型,结合均方根误差和决定系数,用剩余的 30% 样本点辅助调整模型参数和验证模精度。利用 DEVs 和 PEVs 构建区域机器学习模型,其公式分别为:

$$aNDVI_{ij} = ML[aP_{ij}, aT_{ij}; DEV_{i1}, DEV_{i2}, \dots, DEV_{i7}] , \quad (1)$$

$$aNDVI_{ij} = ML[aP_{ij}, aT_{ij}; PCA_PEV_{i1}, PCA_PEV_{i2}, \dots, PCA_PEV_{im}] , \quad (2)$$

式中: ML 为区域机器学习模型的函数形式; i 为像元序号。

此外,用于对比的逐像元最小二乘模型表达为:

$$aNDVI_j = \beta_0 + \beta_1 aP_j + \beta_2 aT_j , \quad (3)$$

式中: β_0 为方程截距; β_1, β_2 为方程系数。

2.2 气候和人为因素的贡献率计算

用 SLO_{obe} 代表 $aNDVI$ 观测值的趋势线斜率。设 $SLO_{obe} > 0$ 且 $P < 0.05$ 为“植被恢复”, $SLO_{obe} < 0$ 且 $P < 0.05$ 为“植被退化”,其他情况为“植被无变化”。用 SLO_{est} 和 SLO_{res} 分别表示模拟 $aNDVI$ 和残差 $aNDVI$ 的趋势线斜率 ($SLO_{res} = SLO_{obe} - SLO_{est}$)。参考 Pei 等^[20]的方法针对植被恢复或退化的每个像元计算气候和人为因素的贡献率,如表 2 所示。

表 2 气候和人为因素贡献率计算方法

Tab.2 Calculation method for contribution of climate and human factors

植被变化类型	SLO_{obe}	SLO_{est}	SLO_{res}	贡献率/%	
				气候因素	人为因素
植被恢复	> 0 (且 $P < 0.05$)	> 0	> 0	$\left(\frac{SLO_{est}}{SLO_{obe}}\right) \times 100$	$\left(\frac{SLO_{res}}{SLO_{obe}}\right) \times 100$
		> 0	< 0	100	0
		< 0	> 0	0	100
植被退化	< 0 (且 $P < 0.05$)	< 0	< 0	$\left(\frac{SLO_{est}}{SLO_{obe}}\right) \times 100$	$\left(\frac{SLO_{res}}{SLO_{obe}}\right) \times 100$
		< 0	> 0	100	0
		> 0	< 0	0	100

3 结果与分析

方便起见,本文将使用 DEVs 和 PEVs 的区域机器学习法分别简称为直接区域机器学习法和代理区域机器学习法。

3.1 区域机器学习法合理性

分别固定变量 aT 和 aP , 利用构建好的直接区

域机器学习模型,获取 $aNDVI$ 模拟值与单一气候因素的关系点线图(图 2)。结果表明:①降水作用和气温作用的三次多项式拟合精度 ($R^2 = 0.989, P < 0.01$; $R^2 = 0.688, P < 0.01$) 高于线性拟合精度 ($R^2 = 0.975, P < 0.01$; $R^2 = 0.665, P < 0.01$);②降水量对植被变化的影响远大于气温 (aP 和 aT 对应的 $aNDVI$ 模拟值区间相差较大);③降水量增加和气

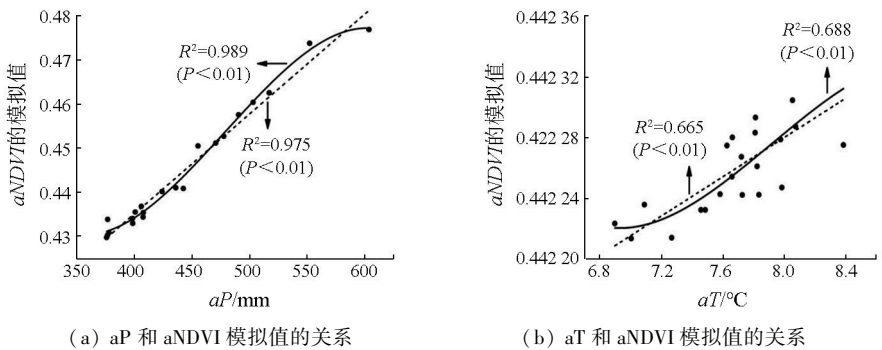


图 2 区域机器学习法中单一气候因素与 aNDVI 模拟值的关系
Fig.2 The relationship between a single climate factor and the simulated values of aNDVI in regional machine learning method

温升高都会促进植被的生长,两曲线均呈“中间较陡两端较缓”的形态。这在一定程度上反映了植被耐寒和耐旱的特性,以及水分过多和温度过高对植被生长的抑制作用。因此,将区域机器学习模型引入残差趋势法是必要且合理的。

图3(a)显示了2000—2020年研究区植被的变化情况。由于植被恢复区占比高达87.88%,而植被退化区仅占比0.75%,本文以植被恢复区为主来验

证新方法的优势。图3(b)和图3(c)为逐像元最小二乘法与直接区域机器学习法针对植被恢复区识别的气候因素贡献。图3(e)为后者与前者差值的空间分布,用等间隔方法将差值分为高值、正常值和低值。根据对研究区植被类型和地理环境的经验知识,选取3个典型区(图3(e)): A为河西地区的绿洲, B为甘南地区的高寒牧区, C为陇东地区的耕、草、林混合区。其中, A和B位于低值区, C位于高值区。

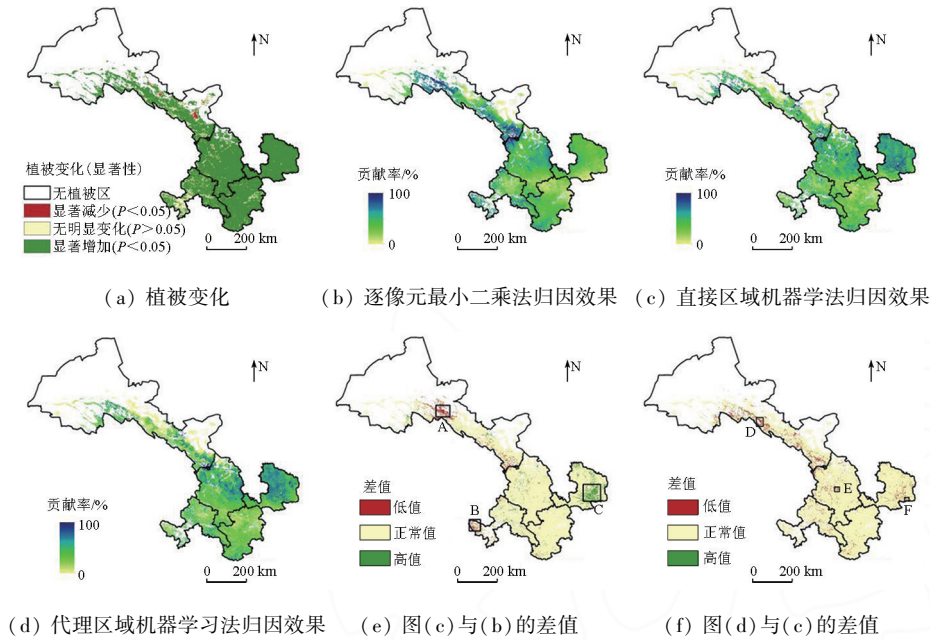


图3 研究区植被变化及归因效果

Fig.3 Effect of vegetation change and its attribution in the research area

利用固定变量的方式,得到 A, B 和 C 中 $aNDVI$ 模拟值和 aP , aT 的关系点线图。其中,图4(a)—

(c)和图4(d)—(f)分别为 aP 和 aT 与 $aNDVI$ 模拟值的关系。

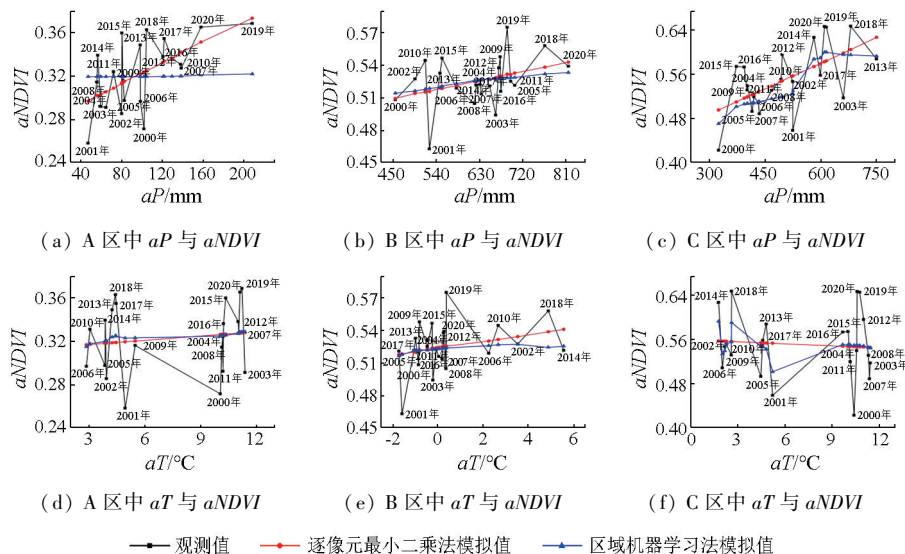


图4 直接区域机器学习法与逐像元最小二乘法输出差异典型区中气候与 $aNDVI$ 模拟值的关系

Fig.4 Relationship between climate and the simulated values of $aNDVI$ in typical areas with differences in output between regional machine learning method using direct environment variables and pixel by pixel least squares method

1) 对于 A 区, 直接区域机器学习模型认为降水变化几乎无作用, 而逐像元最小二乘模型认为作用较大; 此外, 逐像元最小二乘模型和直接区域机器学习模型识别的气温作用程度相当, 皆不明显。对于 B 区, 直接区域机器学习模型识别的降水作用和气温作用均弱于逐像元最小二乘模型, 且该模型认为气温中等时的植被活动最强。对于 C 区, 逐像元最小二乘模型和直接区域机器学习模型都认为降水变化的作用较大, 值得注意的是, 后者识别的降水变化具有非常明显的非线性作用。

2) 就 A 区和 B 区而言, A 区属于温带干旱气候, 气候变化对该地区植被的影响并不显著, 直接区域机器学习模型的模拟结果较准确; B 区属于高寒气候, 降水多寡在很大程度上影响着气温高低, 两者又共同作用于植被生长。相对于研究区其他区域 (尤其是占主体的温带季风区和干旱荒漠区), B 区的植被-气候动态关系模型更容易受到人类信号干扰。相比于逐像元最小二乘模型, 区域机器学习模型不仅依靠其建模策略削弱了逐像元建模策略造成的普遍人类活动信号干扰, 还依靠机器学习方法模拟了大范围区域内复杂的气候-植被非线性关系, 正因如此, 才使得其正确识别了 B 区中的因子交互作用。

3) 就 C 区而言, 直接区域机器学习模型和逐像元最小二乘模型可能都存在过拟合的情况。观察可知, 研究时段内出现了特殊情况: 该地区没有出现 525~580 mm 的降水量, 且 525 mm 左右的降水量出现在研究时段初期 (2001 年和 2002 年),

580 mm 左右的降水量出现在研究时段后期 (2014 年和 2017 年)。这种巧合使逐像元最小二乘模型和直接区域机器学习模型同时错误地将该地区的一部分人类信号判断为气候信号, 过度拟合了植被和气候的非线性关系。因此, 直接区域机器学习法是合理的, 且在大部分情况下优于逐像元最小二乘法, 这种优势是区域建模策略和机器学习算法共同作用形成的。

在研究区域内选取了 2 个相距较近且气候 (人为) 因素贡献率相差较大的典型区, 通过对比分析其中的人类活动强度, 验证区域机器学习法的合理性。由于选取的 2 个区域距离较近, 确保了它们的气候条件基本相同, 因此贡献率的差异主要由人类活动强度变化的差异造成。

如图 5(a) 和图 5(b) 展示了兰州某一区域 I 在 2009 年 10 月和 2020 年 10 月的高分辨率影像, 图 5(c) 和图 5(d) 展示了与该区域相距约 1.5 km 的区域 II 在对应时间的高分辨率影像。前者气候因素的贡献约为 35%, 即人为因素的贡献为 65%; 后者气候因素的贡献约为 70%, 即人为因素的贡献为 30%。观察影像可以发现, 区域 I 在 2009—2020 年期间出现了 2 个新建造的蓄水池, 而区域 II 中的地物无明显变化。经实地考察得知, 区域 I 中的蓄水池主要用于灌溉周边植被, 该地的人类活动强度显著增加; 区域 II 内主要是林地, 研究时段内的人类活动强度较稳定。这与区域机器学习法得出的结论相符合, 进一步证明该方法是合理有效的。

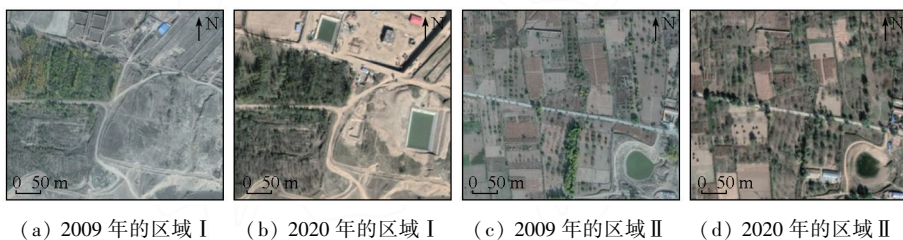


图 5 区域 I 和区域 II 在 2009 年和 2020 年的高分辨率影像

Fig.5 High resolution images of Region I and Region II in 2009 and 2020

3.2 直接区域机器学习法和代理区域机器学习法的比较

图 3(d) 显示了代理区域机器学习法针对植被恢复区识别的气候因素贡献, 与直接区域机器学习法相比, 二者的差异以低值为主, 如图 3(f) 所示。为了验证 PEVs 的优势, 根据图 3(f) 在植被恢复区选取 3 个典型区, 结合实际地表情况 (图 6) 与 DEVs 中相对容易验证的地表覆盖类型数据 (图 1) 分析。区域 D 位于河西绿洲, 其地表覆盖类型数据显示为草地, 与实际地表情况——林地不符。区域 E 和区

域 F 分别位于陇中和陇东, 虽然它们的实际地表情况与地表覆盖类型数据相符, 主要是农田, 但不能排除其他 DEVs 数据存在误差。此外, 观察可知, 2 区域中的农田还可细分为旱作农田和灌溉农田。从 DEVs 选取的角度来看, 缺少了土壤湿度等数据^[21-22]。因此, 直接区域机器学习法的归因效果受到 DEVs 误差和选取遗漏的影响, 且随着 DEVs 选取数量的增多或减少, 遗漏的影响将减小或增大, 而误差的影响与之相反。总体上, 代理区域机器学习法的归因效果优于直接区域机器学习法, 这得益于

PEVs 表达的空间异质性信息更完整和准确。

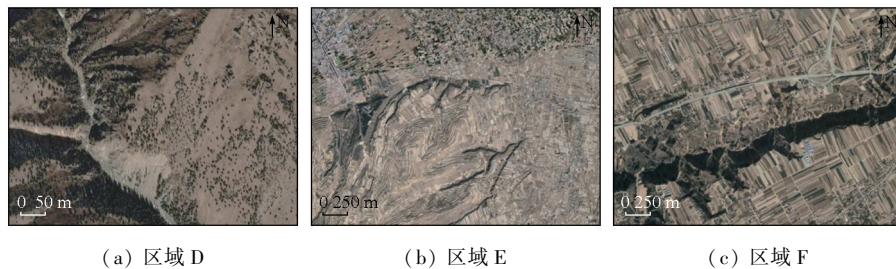


图6 代理区域机器学习法和直接区域机器学习法的输出差异典型区的实际地表覆盖类型

Fig.6 Actual land cover types of typical areas with different output differences between regional machine learning method using direct environment variables and regional machine learning method using PEVs

4 讨论

本文针对最小二乘法难以考虑植被-气候非线性关系和逐像元建模策略易引入较多人类活动干扰信号的缺陷,提出了以区域非线性建模为特征的区域机器学习法,将现有数据(植被指标和气候因子时空序列的组合)作为 PEVs 替代 DEVs 顾及空间异质性问题。实验分析表明新方法具有明显优势,且使用 PEVs 可以更好地解决空间异质性问题。

直接区域机器学习法与逐像元最小二乘法之间的输出差异集中分布在河西地区、甘南西南部和陇东地区。前者的结果显示人类活动是导致河西地区和甘南地区植被变化的主要原因,这与敖泽建等^[23]、李文龙等^[24]和李丽丽等^[25]使用非残差趋势法方法的研究结果相符。直接区域机器学习法之所以在这些地区优于逐像元最小二乘法,一方面是区域机器学习模型因其区域建模方式而具备了降低部分人类信号干扰的能力;另一方面是机器学习算法可以较为客观的模拟植被-气候关系的非线性特征。此外,如 3.1 节所述,直接区域机器学习法在陇东因特殊情况和随机森林算法自身特性等原因,和逐像元最小二乘法一样出现了过拟合的情况。后续的实验可以通过更换算法、扩展时间序列和扩大研究区范围等方法在一定程度上抵消这种复杂情况的影响。

代理区域机器学习法与直接区域机器学习法的归因效果相似。整体上,人类活动对植被变化的贡献高于气候变化。这与胡春艳等^[26]和 Zhou 等^[27]使用非残差趋势法的研究结果相符。2 组实验在局部地区存在输出差异,集中分布在河西地区的人工绿洲。这些地区不仅存在大量人类活动,气候变化对植被的影响也不显著^[28]。相较于直接区域机器学习法,代理区域机器学习法的模拟效果更准确,这得益于使用 PEVs 避免了 DEVs 的缺陷,能够更好地

解决空间异质性问题。

本文还存在着以下不足:

1) 植被指标的选取。当植被达到一定的茂盛程度时,NDVI 可能会出现饱和等问题^[29]。后续实验可以使用例如增强型植被指数,植被光学厚度等。

2) 气候数据的选取。虽然降水和气温是主要的气候因素,但其他气候因素也对植被变化起到一定作用。因此,在新的方法中,还可以引入诸如太阳辐照度等数据。

3) 建模工具。本文仅使用了机器学习算法中的随机森林算法,后续工作中可以借助遥感云平台使用更加复杂的机器学习算法,例如深度学习算法等。此外,为了避免在 3.1 节中提到的陇东地区出现的过拟合情况,后续实验在构建模型时,除了扩大研究区范围和研究时段范围,还可以尝试减少相应区域的采样点数量,以减小像元之间相关性的影响,从而在一定程度上削弱或规避过拟合的情况。

5 结论

本文基于 DEVs 和 PEVs 提出了一种用于植被变化归因的区域机器学习残差趋势法。新方法与逐像元最小二乘残差趋势法相比,具有明显的优势。通过对研究区植被变化归因,得到以下主要结论:

1) 将现有数据——植被和气候时空序列组合作为 PEVs,不仅没有增加数据获取难度,也避免了引入额外的数据误差,且相比于 DEVs,区域机器学习模型使用 PEVs 能更充分地顾及植被-气候动态关系的空间异质性。

2) 区域机器学习模型因其区域建模策略而具备了一定程度的抗人类信号干扰的能力和潜力。尽管如此,本文构建的区域机器学习模型在遇到某些特殊情况时,其抗人类信号干扰的能力依然存在缺陷,有待进一步改进和完善。

3) 得益于机器学习算法,新方法可以客观考虑

植被-气候关系中普遍存在的非线性特征。

参考文献 (References):

- [1] 孙红雨, 王长耀, 牛 铮, 等. 中国地表植被覆盖变化及其与气候因子关系——基于 NOAA 时间序列数据分析[J]. 遥感学报, 1998, 2(3): 204-210.
Sun H Y, Wang C Y, Niu Z, et al. Changes of surface vegetation coverage and its relationship with climate factors in China: Analysis based on NOAA time series data[J]. National Remote Sensing Bulletin, 1998, 2(3): 204-210.
- [2] Intergovernmental Panel on Climate Change (IPCC). Climate change 2013: The physical science basis working group I contribution to the fifth assessment report of the intergovernmental panel on climate change[M]. Cambridge: Cambridge University Press, 2014.
- [3] Liu Y, Li Y, Li S C, et al. Spatial and temporal patterns of global NDVI trends: Correlations with climate and human factors[J]. Remote Sensing, 2015, 7(10): 13233-13250.
- [4] Evans J, Geerken R. Discrimination between climate and human-induced dryland degradation[J]. Journal of Arid Environments, 2004, 57(4): 535-554.
- [5] Guo E, Wang Y, Wang C, et al. NDVI indicates long-term dynamics of vegetation and its driving forces from climatic and anthropogenic factors in Mongolian Plateau[J]. Remote Sensing, 2021, 13(4): 688.
- [6] Radda I A, Kumar B M, Pathak P. Land degradation in Bihar, India: An assessment using rain-use efficiency and residual trend analysis[J]. Agricultural Research, 2021, 10(3): 434-447.
- [7] Li Z D, Wang S, Li C J, et al. The trend shift caused by ecological restoration accelerates the vegetation greening of China's drylands since the 1980s[J]. Environmental Research Letters, 2022, 17(4): 044062.
- [8] Qu L L, Huang Y X, Yang L F, et al. Vegetation restoration in response to climatic and anthropogenic changes in the Loess Plateau, China[J]. Chinese Geographical Science, 2020, 30(1): 89-100.
- [9] Zhao C L, Yan Y, Ma W Y, et al. RESTREND-based assessment of factors affecting vegetation dynamics on the Mongolian Plateau[J]. Ecological Modelling, 2021, 440: 109415.
- [10] Liu Z J, Liu Y S, Li Y R. Anthropogenic contributions dominate trends of vegetation cover change over the farming-pastoral ecotone of Northern China[J]. Ecological Indicators, 2018, 95: 370-378.
- [11] 金 凯, 王 飞, 韩剑桥, 等. 1982—2015 年中国气候变化和人类活动对植被 NDVI 变化的影响[J]. 地理学报, 2020, 75(5): 961-974.
Jin K, Wang F, Han J Q, et al. Contribution of climatic change and human activities to vegetation NDVI change over China during 1982—2015[J]. Acta Geographica Sinica, 2020, 75(5): 961-974.
- [12] Zhao Y B, Sun R H, Ni Z Y. Identification of natural and anthropogenic drivers of vegetation change in the Beijing-Tianjin-Hebei megacity region[J]. Remote Sensing, 2019, 11(10): 1224.
- [13] Zhou S L, Zhang W C, Wang S H, et al. Spatial-temporal vegetation dynamics and their relationships with climatic, anthropogenic, and hydrological factors in the Amur River Basin[J]. Remote Sensing, 2021, 13(4): 684.
- [14] Chen L F, Zhang H, Zhang X Y, et al. Vegetation changes in coal mining areas: Naturally or anthropogenically Driven?[J]. Catena, 2022, 208: 105712.
- [15] Leroux L, Bégué A, Lo Seen D, et al. Driving forces of recent vegetation changes in the Sahel: Lessons learned from regional and local level analyses[J]. Remote Sensing of Environment, 2017, 191: 38-54.
- [16] Yang L, Shen F X, Zhang L, et al. Quantifying influences of natural and anthropogenic factors on vegetation changes using structural equation modeling: A case study in Jiangsu Province, China[J]. Journal of Cleaner Production, 2021, 280: 124330.
- [17] 刘志红, Tim R. McVicar, Van Niel, 等. 专用气候数据空间插值软件 ANUSPLIN 及其应用[J]. 气象, 2008, 34(2): 92-100.
Liu Z H, McVicar T R, Niel V, et al. Introduction of the professional interpolation software for meteorology data: ANUSPLIN[J]. Meteorological Monthly, 2008, 34(2): 92-100.
- [18] 姚 楠, 董国涛, 薛华柱. 基于 GoogleEarthEngine 的黄土高原植被覆盖度时空变化特征分析[J]. 水土保持研究, 2024, 31(1): 260-268.
Yao N, Dong G T, Xue H Z. Analysis on the characteristics of the spatiotemporal change in vegetation coverage on the Loess Plateau using the google earth engine[J]. Research of Soil and Water Conservation, 2024, 31(1): 260-268.
- [19] Drewa P B, Platt W J, Moser E B. Community structure along elevation gradients in headwater regions of longleaf pine savannas[J]. Plant Ecology, 2002, 160(1): 61-78.
- [20] Pei H W, Liu M Z, Jia Y G, et al. The trend of vegetation greening and its drivers in the Agro-pastoral ecotone of Northern China, 2000—2020[J]. Ecological Indicators, 2021, 129: 108004.
- [21] Chaaban F, El Khattabi J, Darwishe H. Accuracy assessment of ESA WorldCover 2020 and ESRI 2020 land cover maps for a region in Syria[J]. Journal of Geovisualization and Spatial Analysis, 2022, 6(2): 31.
- [22] Hasan M A, Mia M B, Khan M R, et al. Temporal changes in land cover, land surface temperature, soil moisture, and evapotranspiration using remote sensing techniques: A case study of Kutupalong Rohingya refugee camp in Bangladesh[J]. Journal of Geovisualization and Spatial Analysis, 2023, 7(1): 11.
- [23] 敖泽建, 王建兵, 蒋友严, 等. 2000—2017 年甘南牧区植被变化特征及其影响因子[J]. 沙漠与绿洲气象, 2020, 14(1): 95-100.
Ao Z J, Wang J B, Jiang Y Y, et al. Characteristics of vegetation change and influencing factors in Gannan pastoral area from 2000 to 2017[J]. Desert and Oasis Meteorology, 2020, 14(1): 95-100.
- [24] 李文龙, 薛中正, 郭述茂, 等. 基于 3S 技术的玛曲县草地植被覆盖度变化及其驱动力[J]. 兰州大学学报(自然科学版), 2010, 46(1): 85-90, 95.
Li W L, Xue Z Z, Guo S M, et al. Vegetation coverage changes and analysis of the driving forces in Maqu County based on 3S technology[J]. Journal of Lanzhou University (Natural Sciences), 2010, 46(1): 85-90, 95.
- [25] 李丽丽, 王大为, 韩 涛. 2000—2015 年石羊河流域植被覆盖度及其对气候变化的响应[J]. 中国沙漠, 2018, 38(5): 1108-1118.
Li L L, Wang D W, Han T. Spatial-temporal dynamics of vegetation coverage and responding to climate change in Shiyang River basin

- during 2000—2015[J]. Journal of Desert Research, 2018, 38(5): 1108–1118.
- [26] 胡春艳, 卫伟, 王晓峰, 等. 甘肃省植被覆盖变化及其对退耕还林工程的响应[J]. 生态与农村环境学报, 2016, 32(4): 588–594.
- Hu C Y, Wei W, Wang X F, et al. Change in vegetation cover as affected by grain for green project in Gansu[J]. Journal of Ecology and Rural Environment, 2016, 32(4): 588–594.
- [27] Zhou W, Gang C, Zhou L, et al. Dynamic of grassland vegetation degradation and its quantitative assessment in the Northwest China [J]. Acta Oecologica, 2014, 55: 86–96.
- [28] 张建永, 李扬, 赵文智, 等. 河西走廊生态格局演变跟踪分析[J]. 水资源保护, 2015, 31(3): 5–10.
- Zhang J Y, Li Y, Zhao W Z, et al. Tracking analysis on changes of ecological patterns in Hexi Corridor Region [J]. Water Resources Protection, 2015, 31(3): 5–10.
- [29] Wang Q, Adiku S, Tenhunen J, et al. On the relationship of NDVI with leaf area index in a deciduous forest site[J]. Remote Sensing of Environment, 2005, 94(2): 244–255.

Residual trend method based on regional modeling and machine learning for attribution of vegetation changes

HU Boyang^{1,2,3}, SUN Jianguo^{1,2,3}, ZHANG Qian^{1,2,3}, YANG Yunrui^{1,2,3}

(1. Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China; 2. National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, China; 3. Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China)

Abstract: Existing residual trend methods utilize a pixel-by-pixel modeling strategy, in which the ordinary least squares method is employed. These methods suffer certain limitations. On the one hand, the pixel-by-pixel modeling strategy causes each model to contain signal interference from human activities in local space. On the other hand, the ordinary least squares method is unfavorable for simulating commonly observed nonlinear characteristics. This study proposed an entirely new residual trend method based on regional modeling and machine learning. Besides, this study compared two types of environmental variables used to express spatial heterogeneity: ①direct-environmental variables (DEVs) such as terrain, hydrology, and land use; and ②proxy-environmental variables (PEVs) that combine the spatiotemporal series of vegetation and climate. First, a regional modeling strategy was adopted. After DEVs and PEVs were introduced individually, models for the vegetation-climate relationship were built using machine learning. Second, residuals were determined based on the definition of the residual trend method. Finally, the contributions of anthropogenic and climatic factors to vegetation change were assessed. The results indicate that compared to the previous pixel-by-pixel residual trend method that utilizes ordinary least squares, the new residual trend method can simulate the nonlinear features of the vegetation-climate relationship and exhibits enhanced resistance to human signal interference. For the new method, significantly higher performance can be achieved using PEVs compared to DEVs. PEVs can fully utilize the original modeling data, without increasing difficulties with data acquisition and avoiding additional data errors. The residual trend method based on regional modeling and machine learning proposed in this study allows for more effective attribution of vegetation changes.

Keywords: attribution of vegetation change; regional modeling; machine learning algorithm; residual trend method; spatial heterogeneity

(责任编辑: 陈昊旻)