

基于CASA模型的植被NPP样本提取研究

文凤平 邱凌* 郑颖

四川省生态环境科学研究院

DOI:10.12238/eep.v8i1.2446

[摘要] 本研究基于CASA模型和空间叠加分析方法,提出了一种估算高空间分辨率(30米)NPP训练样本的方法。研究表明:(1)提取的NPP的结果具有较高的空间分辨率,相比MODIS 500米NPP数据集具有更高的空间异质性;(2)提取的NPP的结果对地形更为敏感,在复杂地形区域具有较高合理性;(3)本研究提出的方法具有较高的可行性,提取的NPP的结果可以补充没有实地监测NPP的区域作为NPP“真值”使用,进行样点采样后的配合训练特征可作为训练样本,提升机器学习模型在NPP估算中的精度和泛化能力。本研究结果可为生态系统碳汇监测、碳循环研究、气候变化影响评估等提供重要数据支持。

[关键词] 植被净初级生产力; CASA模型; 训练样本; 机器学习; 地形校正

中图分类号: TP181 **文献标识码:** A

Study on the extraction of vegetation NPP samples based on CASA model

Fengping Wen Ling Qiu* Ying Zheng

Sichuan Research Institute of Ecological and Environmental Sciences

[Abstract] Based on CASA model and spatial superposition analysis method, this study proposes a method to estimate NPP training samples with high spatial resolution (30 m). The results show that: (1) the extracted NPP results have high spatial resolution and higher spatial heterogeneity compared with MODIS 500 meter NPP data set; (2) the extracted NPP results are more sensitive to terrain and reasonable in complex terrain areas; (3) the method proposed in this study has high feasibility. The extracted NPP results can supplement the area without field monitoring NPP as NPP "true value", and the cooperative training features after sample site sampling can be used as training samples to improve the accuracy and generalization ability of machine learning model in NPP estimation. The results of this study can provide important data support for ecosystem carbon sink monitoring, carbon cycle research, and impact assessment of climate change.

[Key words] net primary productivity of vegetation; CASA model; training sample; machine learning and terrain correction

前言

植被净初级生产力(Net Primary Production, NPP)是植被在单位时间单位面积上,通过光合作用固定的有机碳总量减去自养呼吸后剩余的有机碳量,体现了植被对空气中二氧化碳的吸收能力^[1]。植被净初级生产力不仅可以直接表征植被在自然界的固碳能力,也有助于确定人类对环境的影响,对判断生态系统碳汇功能、评价陆地生态系统可持续发展具有重要意义^[2]。

植被净初级生产力的估算经历了从传统实地测量到多技术应用,从单一实测数据到多元数据融合的发展。实地测量的方法只能获取小范围内的植被净初级生产力,难以在大尺度开展动态监测。伴随着对地物信息敏感的遥感观测数据的不断丰富与遥感处理技术的快速发展,时空异质性反映优势明显的集成遥感数据的陆地植被净初级生产力估算模型逐渐兴起。主

流遥感估算植被净初级生产力的技术方法可分为两种,一是基于遥感数据驱动的生态系统过程/光能利用率模型,如LPJ-DGVM模型、BIOME-BGC模型、BEPS模型、CASA模型等^[3,4];二是基于训练样本数据的机器学习模型,如随机森林、神经网络、深度学习等^[4,5]。生态系统过程/光能利用率模型有明确的物理过程,但多源数据的获取不确定性,估算结果往往存在较大的空间不连续性。基于机器学习的遥感参数反演充分利用了卫星大数据的优势,能够避免因物理过程模型带来的复杂处理和计算过程,同时减小由遥感数据预处理包括几何校正、辐射校正、大气校正等引起的不确定性,因此被应用于遥感反演中。运用机器学习模型进行植被净初级生产力估算时,用户无需刻意关心物理机理过程,但是其训练过程需要大量样本数据集。训练样本数据集的“真值”的获取往往具有较大难度,这成为了机器学习在遥

感估算植被净初级生产力的主要阻碍之一。

因此本研究提出一种特定的数据处理方法,通过结合CASA模型和空间叠加方法,提取高空间分辨率(30米)的NPP数据,并结合特征变量形成训练样本,可以解决机器学习模型估算植被净初级生产力缺少实地监测样本真值的问题。

1 数据与研究方法

1.1数据。地表反射率数据来自罗切斯特理工学院(RIT)和美国国家航空航天管理局(NASA)喷气推进实验室(JPL)共同创建的单渠道算法得到的Landsat 8/C02/T1_L2数据集,具有较高的几何和辐射精度,适用于科学研究和应用分析^[6,7]。土地利用/覆被数据来自武汉大学制作的CLCD(China Land Cover Dataset)数据集^[8],空间分辨率为30 m、时间分辨率为1年的LULC数据集。气象数据为月度的全球地表气候和气候水平平衡的数据集(TerraClimate)^[9],由美国爱达荷大学开发,该数据特别适用于气候分析、生态学、农业研究以及水文和气候变化研究。数字高程模型(DEM)来自欧洲航天局发布的30米空间分辨率的Copernicus DEM GLO-30数据集,是公认的最佳全球性开源DEM之一。

1.2研究方法。

1.2.1地形校正的CASA模型。公式一: 利用地形因子校正CASA模型提取NPP

$$NPPo(p,y) = \sum_{t=1}^{12} [APAR(p,y,t) \times LUE(p,y,t) \times \alpha(p)]$$

$$APAR(p,y,t) = SOL(p,y,t) \times FPAR(p,y,t) \times 0.5$$

$$LUE(p,y,t) = T1(p,y,t) \times T2(p,y,t) \times W(p,y,t) \times LUEm$$

式中, p表示任一位置; y表示研究时段内任一年份; t表示y年份中任一月份; NPPo(p,y)为(p,y)处年累计净初级生产力,单位为g·C/m²; APAR(p,y,t)为(p,y,t)处吸收的光合有效辐射,单位为g·C/m²; LUE(p,y,t)为(p,y,t)处的实际光能利用率,单位为g·C/MJ; α(p)为p处的地形校正因子; SOL(p,y,t)为(p,y,t)处太阳净辐射,单位为MJ/m²; FPAR(p,y,t)为(p,y,t)处植被层对入射光合有效辐射的吸收比例;常数0.5表示光合有效辐射占总辐射的比例; T1(p,y,t)为(p,y,t)处低温对实际光能利用率的胁迫因子; T2(p,y,t)为(p,y,t)处高温对实际光能利用率的胁迫因子; W(p,y,t)为(p,y,t)处水分对实际光能利用率的胁迫因子; LUEm为理想条件下最大光能利用率,取值可参考文献资料,单位为gC/MJ。

公式二: 光合有效辐射的吸收比例FPAR(p,y,t)

$$FPAR(p,y,t) = 0.229 \times \exp(1.95 \times NDVI(p,y,t)) - 0.344$$

$$NDVI(p,y,t) = \frac{\rho nir(p,y,t) - \rho red(p,y,t)}{\rho nir(p,y,t) + \rho red(p,y,t)}$$

式中, NDVI(p,y,t)为(p,y,t)处归一化植被指数; ρ nir(p,y,t)、ρ red(p,y,t)为(p,y,t)处近红外波段、红光波段的地表反射率。

公式三: 不同温度对实际光能利用率的胁迫因子

$$T1(p,y,t) = 0.8 + 0.02 \times Topt(p,y) - 0.0005 \times [Topt(p,y)]^2$$

$$T2(p,y,t) = \frac{1.184}{\{1 + \exp[0.2 \times (Topt(p,y) - 10 - T(p,y,t))]\}} \times \frac{1}{\{1 + \exp[0.3 \times (T(p,y,t) - Topt(p,y) - 10)]\}}$$

式中, Topt(p,y)为(p,y)处一年内归一化植被指数达到最高时的当月平均气温,即植被生长的最适温度,单位为摄氏度; T(p,y,t)为(p,y,t)处月均温,单位为摄氏度;当月均温T(p,y,t)比最适温度Topt(p,y)高10摄氏度或低13摄氏度时,该月的T2(p,y,t)值等于月平均温度T(p,y,t)为最适温度Topt(p,y)时T2(p,y,t)值的一半。

公式四: 水分对实际光能利用率的胁迫因子W(p,y,t)

$$W(p,y,t) = \frac{aet(p,y,t)}{pet(p,y,t)}$$

式中, aet(p,y,t)为(p,y,t)处月累计实际蒸散量,单位为mm; pet(p,y,t)为(p,y,t)处月累计潜在蒸散量,单位为mm。

公式五: 地形校正因子α(p)计算^[3]

$$\alpha(p) = \frac{Ss(p)}{Sp(p)}$$

$$Ss(p) = Sp(p) / \cos(\theta(p))$$

式中, α(p)为p处地形校正因子; Ss(p)、Sp(p)分别为p处的地表面积和投影面积,单位为m²; θ(p)为p处的坡度,可由DEM数据提取得到。

1.2.2植被“样点”提取方法。公式六: 空间叠加提取遥感等数据覆盖条件的植被“样点”

$$TP(p,y) = \begin{cases} 1 & f(d) = 4 \\ 0 & f(d) < 4 \end{cases}$$

$$f(d) = d1(p) + d2(p,y) + d3(p,y) + d4(p,y)$$

$$d3(p,y) = \prod_{t=1}^{12} [d3(p,y,t)]$$

$$d4(p,y) = \prod_{t=1}^{12} [d4(p,y,t)]$$

(1) 根据数据的覆盖条件,提取满足上式的空间区域TP(p,y): 式中, p表示任一位置; y表示研究时段内任一年份、t表示y年中任一月份; TP(p,y)值为1或0; d1(p)表示p处数字高程DEM数据是否为有效值覆盖,“是”d1(p)等于1,“否”d1(p)则等于0; d2(p,y)表示(p,y)处是否为植被,“是”d2(p,y)等于1,“否”d2(p,y)则等于0; d3(p,y)表示(p,y)处大气辐射数据覆盖是否满足条件,为1或者0; d4(p,y)表示(p,y)处遥感影像数据集覆盖是否满足条件,为1或者0; d4(p,y,t)表示(p,y,t)处遥感影像数据集是否有数据且无云遮盖,“是”d4(p,y,t)等于1,“否”d4(p,y,t)则等于0。(2) 将TP(p,y)取值为1区转换为面矢量并向内缓冲区分析,缓冲距离设置可取值30m~100m,得到向内缓冲区

分析结果为植被“样区”; (3)在植被“样区”内进行随机点生成,生成结果的即为植被“样点”; (4)利用地形校正后的CASA模型,提取植被“样点”位置的30米NPP值。

2 结果

以四川省凉山州某地为例,通过地形校正后的CASA模型提取该区域30米高分辨率的NPP数据(图1(d),下文简称CASA 30 NPP),数据能够更精细地刻画地表植被生产力的空间分布特征,显著提升了NPP估算的精度和细节表现力。与MODIS 500米NPP数据(图1(c),下文简称MODIS 500 NPP)相比,CASA 30 NPP在植被类型复杂或土地利用多样化的区域,如森林-农田交错带、城市绿地等(图1(a)),能够更准确地反映植被生产力的局部空间变化。地形因素是影响植被NPP的重要环境变量之一,尤其在复杂地形区域(山地),地形对光照、温度、水分等生态因子的再分配作用显著。CASA 30 NPP还充分考虑了地形的影响,表现出较高的地形敏感性(图1(b)和(d))。尤其在复杂地形区域,CASA 30 NPP能够准确捕捉到阴坡和阳坡之间植被生产力的差异,MODIS 500 NPP因空间分辨率不足或地形校正不充分而忽略这些细节。在坡度变化较大的区域,CASA 30 NPP能够更好地反映地形对植被生长的限制作用,为复杂地形区域的生态系统研究提供了更合理的技术支撑。

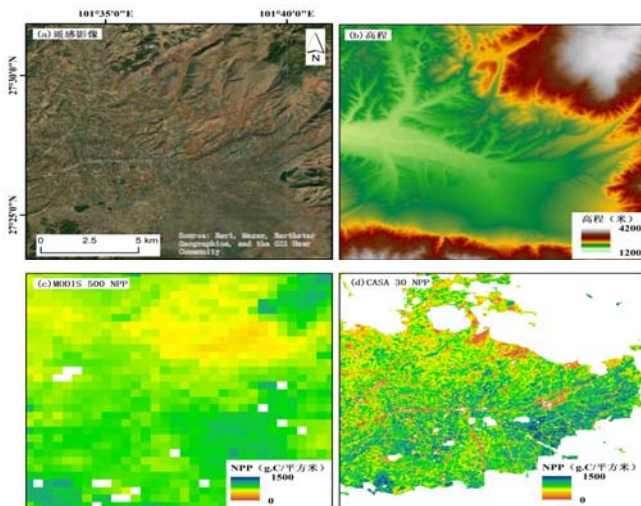


图1 本研究提取结果及其对比图(a:遥感影像;b:高程;c:MODIS 500米NPP数据;d:本研究利用地形校正后的CASA提取的30米NPP数据)

结合本研究提取CASA 30 NPP虽然在空间覆盖上有一定局限性,这主要取决于Landsat数据的质量。但是在偏远地区或难以开展实地监测的区域,本研究方法能够提供可靠的NPP估算结果。通过植被“样点”提取方法可提取这些区域的30米分辨率的NPP作为机器学习模型的训练样本真值,结合其他环境特征变量(如气候、土壤、地形等),可以显著提升机器学习模型在NPP估算中的精度和泛化能力。此外,本研究方法的推广还可以为其其他生态系统参数(如植被覆盖度、生物量等)的估算提供参考,具有广泛的应用潜力。

3 结论

本研究通过结合CASA模型和空间叠加方法,提出了一种估算高空间分辨率(30米)NPP训练样本的方法思路,并将提取结果和MODIS 500 NPP数据集进行对比,得到以下结论:

(1)本研究基于CASA模型和空间叠加分析方法,提取了空间分辨率为30米的NPP数据,能够更精细地刻画地表植被NPP的空间分布特征,显著提升了NPP估算的精度和细节表现力。与传统的MODIS 500米NPP数据集相比,本研究提取的NPP结果在空间异质性方面表现出明显优势。(2)本研究提取的NPP结果充分考虑了地形的影响,表现出较高的地形敏感性。这种对地形的敏感性使得本研究方法在山区、丘陵等复杂地形区域具有较高的适用性和可靠性。(3)本研究提出的方法不仅具有较高的科学性,还具备较强的实际应用可行性。通过植被“样点”采样和特征提取,本研究提取的NPP结果可以作为机器学习模型的训练样本,显著提升机器学习模型在NPP估算中的精度和泛化能力,具有广泛的应用潜力。

[基金课题]

四川省生态环境保护科技项目计划课题,2024HB22。

[参考文献]

- [1]CRAMER W,KICKLIGHTER D W,BONDEAU A,等.Comparing global models of terrestrial net primary productivity (NPP): overview and key results[J].Global change biology,1999,5(S1):1-15.
- [2]KOJU U A,ZHANG J,MAHARJAN S,等.Analysis of spatiotemporal dynamics of forest Net Primary Productivity of Nepal during 2000-2015[J].International Journal of Remote Sensing, 2020,41(11):4336-4364.
- [3]王楠楠,刘耀林,尹峰,等.顾及地形校正的生态系统碳汇估算与分析[J].武汉大学学报(信息科学版):1-24.
- [4]李丽,郭靖,李宁.荒漠区人工林NPP估算模型比较研究[J].防护林科技,2019:13-17.
- [5]李传华,孙皓,王玉涛.基于机器学习估算青藏高原多年冻土区草地净初级生产力[J].生态学杂志,2020(5):1734-1744.
- [6]LOVELAND T R, IRONS J R.Landsat 8:The plans, the reality, and the legacy[J].Remote Sensing of Environment,2016,185:1-6.
- [7]ROY D P,WULDER M A,LOVELAND T R,等.Landsat-8: Science and product vision for terrestrial global change research[J].Remote sensing of Environment,2014,145:154-172.
- [8]YANG J, HUANG X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019[J/OL]. Earth System Science Data,2021,13(8):3907-3925.
- [9]ABATZOGLOU J T,DOBROWSKI S Z,PARKS S A,等.TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015[J/OL].Scientific Data,2018, 5(1):170191.

作者简介:

文凤平(1994--),女,汉族,四川遂宁人,硕士研究生,研究方向或从事工作:生态系统服务功能评估、生态遥感地表参量估算研究。