


基于自然地理要素的全球土地宜垦性数据集的研发

张成鹏, 叶瑜*, 方修琦

北京师范大学地理科学学部, 北京 100875

摘要: 基于自然地理要素构建的土地宜垦性数据是历史耕地格网化空间分配研究中重要的基础数据, 其合理性直接影响着格网化结果的可靠性。在已有相关研究中, 假设某一自然地理要素在全球或洲际等宏观尺度以同一线性关系影响宜垦性大小, 存在较大不合理性。本研究在 $0.5^{\circ} \times 0.5^{\circ}$ 格网单元下辨识了全球陆地(除南极洲)范围内气候、土壤、地形等 13 个自然地理要素与现代垦殖率的相关性, 分区遴选与现代垦殖率显著相关的自然地理要素综合构建了全球土地宜垦性数据。结果可作为全球或洲际等大尺度历史耕地格网化空间分配研究中的一项基础数据。该数据由 4 个数据文件组成, 以格式存储, 空间分辨率为 $5' \times 5'$, 数据量为 38.7 MB (压缩为 1 个文件, 3.55 MB)。

关键词: 自然地理要素; 耕地覆盖; 空间分异; 相关性; 宜垦性

DOI: <https://doi.org/10.3974/geodp.2022.03.08>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2022.03.08>

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.2022.04.01.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2022.04.01.V1>。

1 前言

农业革命以来, 人类活动导致的土地利用/覆盖变化(land use and cover change, LUCC)对陆地生态系统产生了日益深刻的影响, 成为影响全球环境变化日益重要的驱动力之一^[1-4]。特别是工业革命以来, 人口持续增长, 地表原始自然景观发生前所未有的剧烈改变, 由此对地球物质能量循环和全球气候变化产生了重要影响^[5]。

自 20 世纪 90 年代起, 在 IGBP、PAGEs 等国际计划的推动下, 通过综合集成古环境学、考古学、历史学等多学科, 以及模型模拟等科学手段^[6], 已陆续研制和发布了以 SAGE、HYDE、PJ、KK10 等为代表的一系列全球历史 LUCC 格网化数据集^[5,7-10]。它们作为气候模式、碳循环等模型的输入参数, 已被广泛地应用于评估人类活动导致的土地覆盖变化

收稿日期: 2021-06-25; 修订日期: 2022-07-11; 出版日期: 2022-09-25

基金项目: 中华人民共和国科学技术部(2017YFA0603304)

*通讯作者: 叶瑜, 北京师范大学, yeyuleaffish@bnu.edu.cn

数据引用方式: [1] 张成鹏, 叶瑜, 方修琦. 基于自然地理要素的全球土地宜垦性数据集的研发[J]. 全球变化数据学报, 2022, 6(3): 386–394. <https://doi.org/10.3974/geodp.2022.03.08>. <https://cstr.escience.org.cn/CSTR:20146.14.2022.03.08>.

[2] 张成鹏, 叶瑜, 方修琦. 基于自然地理要素的全球土地宜垦性数据集[J/DB/OL]. 全球变化数据仓储电子杂志, 2022. <https://doi.org/10.3974/geodb.2022.04.01.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2022.04.01.V1>.

(anthropogenic land cover change, ALCC) 对碳收支、气候变化等影响的研究中, 这些基础数据的质量直接影响着过去全球变化模拟和评估结果的可靠性^[11,12]。

作为一种广泛而持久的土地利用方式, 耕地定量重建及其格网化空间分配一直是历史 LUCC 研究的主要对象^[13-15]。在历史耕地面积定量重建/估算结果相对准确的情况下, 格网化数据集的不确定性主要受空间分配算法的影响^[16-18]。现有分配方法多是通过主观地遴选一些可能对耕作适宜性产生影响的自然要素, 假设各要素与开垦强度存在单一的线性关系, 再将各要素归一化结果等权重相乘来构建土地宜垦性数据(此处所说的土地宜垦性是指土地适宜进行耕作的好坏程度), 进而以宜垦性大小作为权重将重建政区单元的耕地面积总量进行分配获得格网化结果^[14]。然而针对全球、大洲等宏观尺度, 上述假设显然存在较大不合理性。一方面, 同一自然地理要素在不同空间单元上与土地宜垦性的相关性及其程度存在差异; 另一方面, 在不同空间单元中影响土地宜垦性的自然地理要素的组合情况也是不同的^[18]。

鉴于此, 本研究收集了气候、地形、土壤等方面 13 个自然地理要素数据, 旨在从热量、水分、土壤肥力、劳作难度等方面简要地刻画对耕作产生的影响; 将全球划分为 $0.5^{\circ} \times 0.5^{\circ}$ 的规则格网单元, 分区辨识了各要素与现代垦殖率的相关性; 进而将各空间单元中与垦殖率显著相关的所有自然地理要素进行归一化, 再对归一化结果求取几何平均数以获得全球土地宜垦性结果。

2 数据集元数据简介

《基于自然地理要素的全球土地宜垦性数据》^[19]的元数据信息见表 1。

3 数据研发方法

3.1 原始数据

本研究使用的自然地理要素和垦殖率数据列于表 2。

考虑到研究的空间范围和数据可获得性, 本研究依据主观经验遴选了涵盖气候、地形、土壤等方面共 13 个要素(表 2)。其中: 气候数据来源于 WorldClim^[21], 选取其中年均温度和降水量两个要素。海拔高度数据来源于美国地质调查局的 GTOPO30 DEM^[22], 并以此生成坡度数据。土壤数据来源于 World Soil Information³的 SoilGrids 数据集^[23], 选取了土壤机械组成、容重、有机碳密度、阳离子交换量、酸碱度等 8 个要素。NDVI 数据来源于美国国家太空总署(NASA)发布的 GIMMS 产品^[24], 对 1982–2015 年的半月值经最大值合成法(maximum value composite, MVC)处理获得多年均值。现代垦殖率数据采用的是 Zhang 等^[25]基于 9 套现代遥感全球土地覆盖产品的垦殖强度一致性分析, 融合制备的一套现代全球 $0.5^{\circ} \times 0.5^{\circ}$ 的比例型垦殖率数据。相比原始遥感产品, 该数据在耕地空间分布和像元尺度的垦殖强度等方面的可靠性较高^[25]。

¹ <https://www.worldclim.org/data/worldclim21.html>.

² <https://earthexplorer.usgs.gov/>.

³ ISRIC, <https://soilgrids.org/>.

⁴ <https://ecocast.arc.nasa.gov/data/pub/gimms/3g.v1/>.

表 1 《基于自然地理要素的全球土地宜垦性数据》元数据简表

条 目	描 述		
数据集名称	基于自然地理要素的全球宜垦性数据		
数据集短名	GlobalCultivLandSuitability		
作者信息	张成鹏 HGE-4534-2022, 北京师范大学, cpzhang@mail.bnu.edu.cn 叶瑜, 北京师范大学, yeyuleaffish@bnu.edu.cn 方修琦, 北京师范大学, xfang@bnu.edu.cn		
地理区域	全球陆地范围		
数据年代	现代 (约 2000 C.E.)	数据格式	.img
数据量	3.55 MB (压缩后)		
数据集组成	空间分辨率为 5'×5'土地宜垦性栅格数据		
基金项目	中华人民共和国科学技术部 (2017YFA0603304)		
数据计算环境	ArcGIS		
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	全球变化科学研究数据出版系统的“数据”包括元数据 (中英文)、通过《全球变化数据仓储电子杂志 (中英文)》发表的实体数据集和通过《全球变化数据学报 (中英文)》发表的数据论文。其共享政策如下: (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 ^[20]		
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS		

表 2 本研究使用的自然地理要素和垦殖率数据的基本信息

类型	数据名称	单位	分辨率	简要描述
自然地理要素	年均温度	℃	0.5'×0.5'	WorldClim 1970–2000 年多年平均值 ^[21]
	年均降水量	mm	0.5'×0.5'	WorldClim 1970–2000 年多年平均值 ^[21]
	海拔高度	m	3"×3"	GTOPO30DEM ^[22]
	坡度	°	3"×3"	由 DEM 计算得到
	NDVI	–	5'×5'	由 NASA 发布的 GIMMS NDVI 产品 ^[23] 1981–2015 年半月数据经 MVC 法获得逐年数据, 再计算多年均值
	容重	kg/m ³	10"×10"	由 SoilGrids ^[24] 中多层 (地表至地下 60 cm 处共 5 层) 数值加和求取平均而得
	有机碳含量	kg/m ³	10"×10"	
	粗粒物质含量	%	10"×10"	
	砂粒含量	%	10"×10"	
	粉砂含量	%	10"×10"	
	黏粒含量	%	10"×10"	采用 Zhang 等的全球融合垦殖率数据 ^[25]
	阳离子交换量	mol/kg	10"×10"	
	pH 值	–	10"×10"	
	现代垦殖率	%	0.5'×0.5'	
耕地	现代垦殖率	%	0.5'×0.5'	

以上自然地理要素的遴选符合代表性、层次性和全面性等原则。首先，气候从宏观尺度上决定了农业类型和耕地分布的基本格局；其次，大的地貌格局和地形深刻影响了耕作方式和利用强度；再次，水域的分布对耕地开垦次序和强度有着明显的影响；而在更小尺度上，人类可以相对容易的通过直接改造局部地形和改良土壤以开垦耕地和提高产量。上述要素对土地宜垦性的影响实际是不同地力（以产量体现）的土地对人类开垦次序的影响。在付出同等人力和经济投入的情况下，在地力较好的土地上进行耕作更容易获得高产量，因此最容易被优先开垦。尽管光合生产潜力、农田生产潜力等数据可以从产量方面综合指示土地宜垦性，但考虑到其与温度、降水、土壤等指标不独立，故本研究选择 NDVI 这一综合性替代指标。

3.2 方法

本研究将所有数据空间分辨率统一至 $5'\times 5'$ （采用 ArcGIS 中 Zonal Statistics 工具），以 $0.5^\circ\times 0.5^\circ$ 作为基本辨识单元。土地宜垦性制备方法的基本步骤如下：

（1）在 $0.5^\circ\times 0.5^\circ$ 的空间单元内，对各自然地理要素与垦殖率做皮尔逊相关分析（Sig=0.001），原理如图 1 所示。

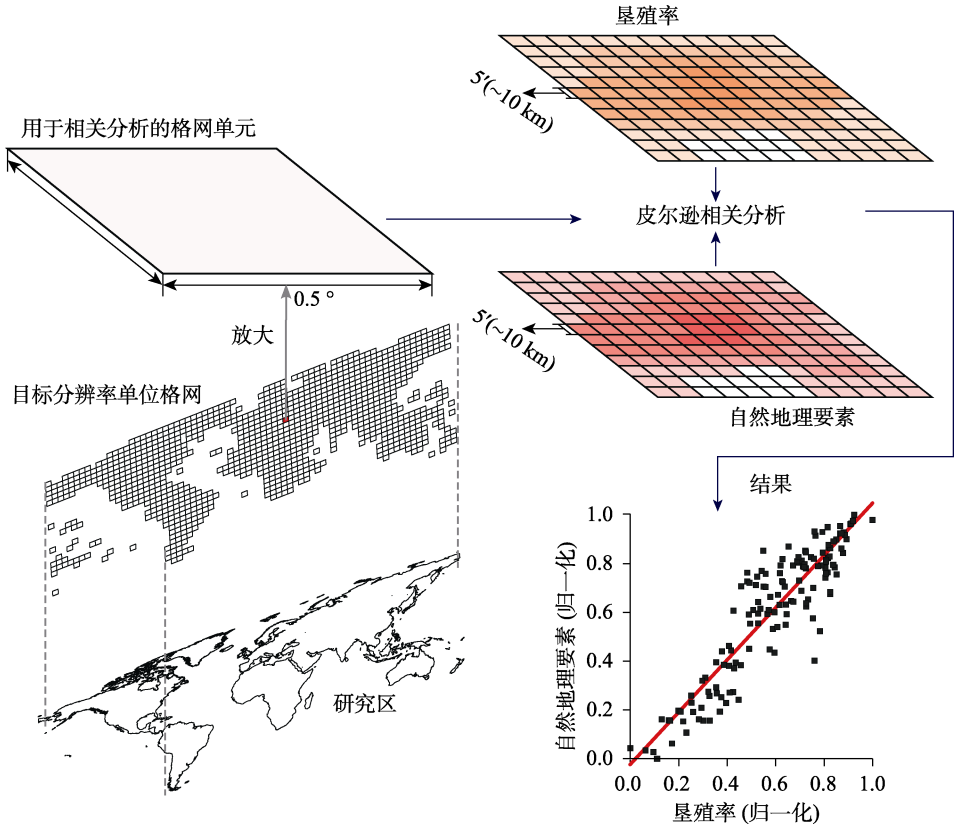


图 1 全球 0.5°格网单元下辨识垦殖率与自然地理要素相关性

（2）在每个 $0.5^\circ\times 0.5^\circ$ 的空间单元中，对所有与垦殖率呈相关的要素进行归一化处理。如正相关则采用最大值归一化（式 1），如负相关则采用最小值归一化（式 2）。

$$x_{iMaxNorm} = \frac{x_i - x_{i_{min}}}{x_{i_{max}} - x_{i_{min}}} \tag{1}$$

$$x_{iMinNorm} = \frac{x_{i_{max}} - x_i}{x_{i_{max}} - x_{i_{min}}} \tag{2}$$

式中， $x_{iMaxNorm}$ 为空间单元 i 中某像元最大值归一化结果； $x_{iMinNorm}$ 为最小值归一化结果； x_i 表示空间单元 i 中某像元的数值大小； x_{imax} 为空间单元 i 中某要素的最大值； x_{imin} 为空间单元 i 中某要素的最小值。

(3) 对空间单元 i 中所有与垦殖率相关的要素的归一化结果（将原始 0–1 的区间转化至 1–100）求取几何平均数（式 3）。此前方法将各要素直接相乘，将导致垦殖率与宜垦性二者由线性关系转为指数关系，构建土地宜垦性的要素越多，这种指数缩放效应就越强，容易造成过饱和和分配，即土地宜垦性越大的格网被赋以极高的耕地数量。

$$W_i = \sqrt[n]{f_{i_1} \times f_{i_2} \dots \times f_{i_n}} \tag{3}$$

式中， W_i 为空间单元 i 的土地宜垦性； f_{ni} 为空间单元 i 中用于构建土地宜垦性的要素 n ； n 为空间单元 i 中参与构建土地宜垦性的要素个数。

(4) 以空间单元 i 中现代垦殖率区间（最小值和最大值）对步骤 2 所得结果进行区间拉伸变换（式 4），最终得到空间分辨率为 $5' \times 5'$ 的土地宜垦性结果（图 2）。在提取空间单元 i 中垦殖率极值的过程中，预先对垦殖率数据进行了平滑处理（使用 ArcGIS 中的 Focal Statistics 工具，平滑长度为 3×3 ）。对于少数无法辨识能够有效指示垦殖强度的自然地理要素的空间单元，直接以现代垦殖率来代表土地宜垦性大小。

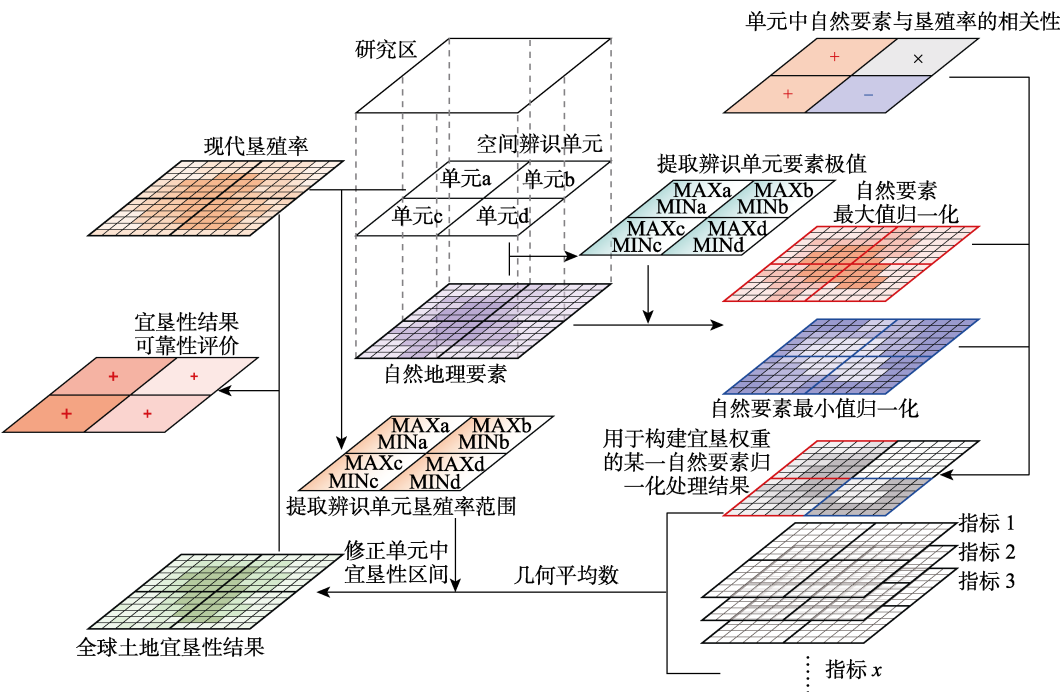


图 2 全球土地宜垦性数据构建流程


（“+”表示显著正相关；“-”表示显著负相关；“×”表示不相关）

$$W_{i_{mdfy}} = \left(W_i + \frac{CF_{min_i} \times 100}{CF_{max_i} - CF_{min_i}} \right) \times \left(\frac{CF_{max_i} - CF_{min_i}}{100} \right) \quad (4)$$

式中， $W_{i_{mdfy}}$ 为空间单元 i 中土地宜垦性的修正结果； W_i 为空间单元 i 的土地宜垦性； CF_{max_i} 为空间单元 i 中用垦殖率最高值； CF_{min_i} 为空间单元 i 中用垦殖率最低值。

4 数据结果与验证

4.1 数据集组成

基于自然地理要素的全球土地宜垦性数据（图 3）由 4 个数据文件组成，以格式存储，空间分辨率为 $5' \times 5'$ ，数据量为 38.7 MB (压缩为 1 个文件，3.55 MB)。

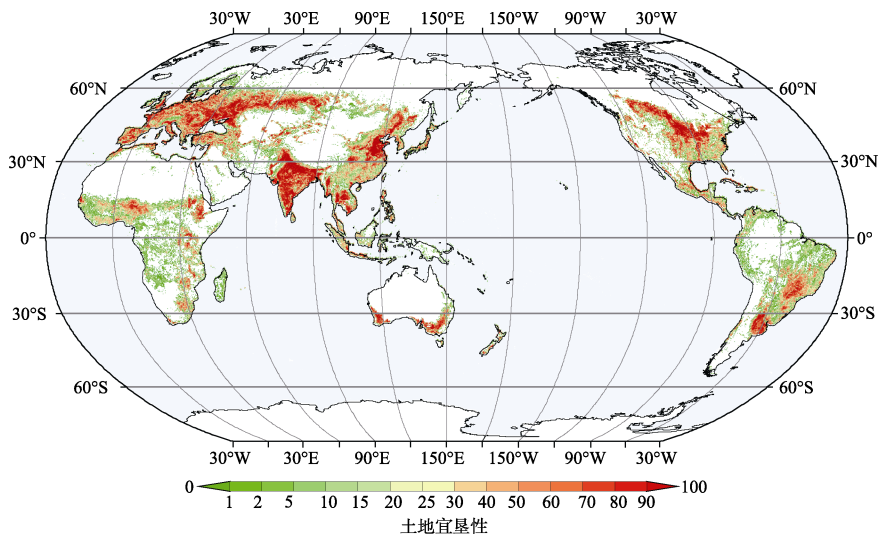


图 3 全球土地宜垦性结果

4.2 数据结果

总体来看，土地宜垦性强弱与垦殖率大小的空间分布格局基本一致。即在全球主要农耕区（如东欧平原、中国华北平原、恒河平原、北美中部平原等）普遍表现出很高的垦殖强度，而在自然条件相对恶劣的地区，宜垦性普遍很低。从国别或区域等较为精细的尺度来看，土地宜垦性结果在极少数地区表现出一定的不合理性。即与垦殖率在空间上总是表现出较为平滑的过渡特征不同，少部分 0.5° 单元格网边界两侧的土地宜垦性大小会呈现明显差异。土地宜垦性在区域尺度上更为细节的特征可加载本数据至 ArcGIS 中进行查看。

4.3 数据结果验证

土地宜垦数据旨在有效指示潜在开垦条件的高低，本研究以现代垦殖率数据作为替代性指标对宜垦性结果进行可靠性检验。通过在 $0.5^\circ \times 0.5^\circ$ 空间单元下计算宜垦性与现代垦殖率的相关性（ $\text{Sig} = 0.001$ ）来评价土地宜垦性结果的可靠性。如图 4 所示，土地宜垦性结果与现代垦殖率表现出良好的正相关。从空间分布来看，二者在绝大多数区域相关系数都超过了 0.7，覆盖范围达到 74%。

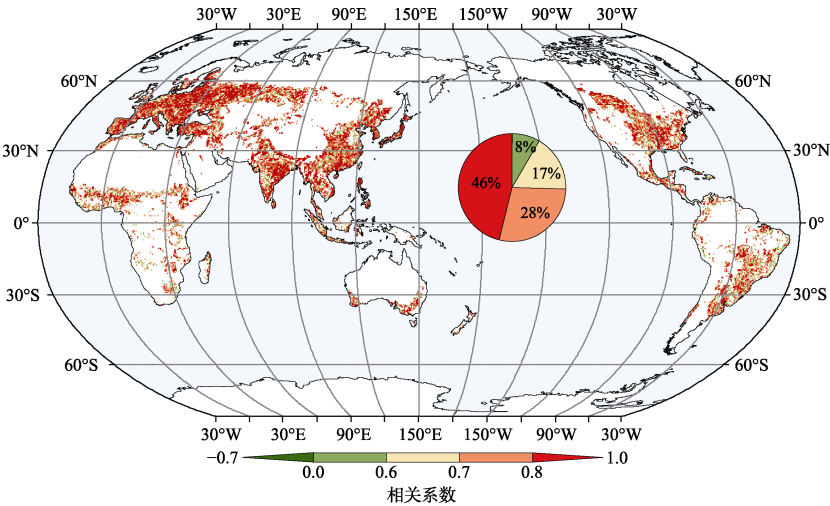


图 4 土地宜垦性与现代垦殖率在 0.5°×0.5°空间单元下的相关性

5 讨论和总结

本研究通过在全球范围内辨识自然地理要素和垦殖率的相关性，取得以下重要认识：自然地理要素与垦殖率的关系是存在区域差异的，讨论二者关系时需要针对特定区域而言。首先，几乎所有的自然地理要素与垦殖率的相关性在全球尺度上是不存在统一的线性关系的，而是正相关或负相关并存的；且相关系数的大小及其显著性在空间上也存在着明显的差异。其次，在不同区域中，与垦殖强度显著相关的自然地理要素的类型和数量也存在明显的不同。从世界范围来看，自然地理要素对低山、丘陵区域内宜垦性的影响明显大于平原和盆地地区。

图 5 以中国区为例，展示了一些自然地理要素与垦殖率相关性的空间分异特征。例如，从降水量来看，并非多数研究假设的那样降水量越大，垦殖强度越高；在雨量丰沛的皖南、琼州半岛的两个样点上其与垦殖强度分别表现出负相关和不相关的特征；而仅在陇东（降水量不足 600 mm）的样点上，二者才表现出良好的正相关。再如，从 NDVI 来看，在新疆塔里木盆地西南缘（依赖灌溉的绿洲农业）的格点上，该要素与垦殖强度表现出良好的正相关；而在东北平原的东南部，二者则呈现极强负相关。

关于本数据对于历史时期适用性的问题，需要强调以下问题。首先，以现代土地宜垦性的分布格局反映历史时期的状态时，应基于一个基本假设，即古今自然地理要素与土地宜垦性的关系具有一致性。即假设气候、地形、土壤等自然地理环境要素与土地宜垦性的关系未随时间变化而剧烈改变，因此可以基于现代自然地理环境要素与土地宜垦性的关系刻画历史时期的土地宜垦性状况。这是因为：一方面，大尺度的历史自然环境要素数据（非模拟所得）难以获得；另一方面，历史时期地形、土壤等要素可以近似看作古今一致，气候变化的幅度有限，且在局地尺度变化具有相对一致性。因此，本研究采用现代自然地理要素数据作为了其历史时期的替代。

本研究通过在全球范围内以 0.5°×0.5°的空间单元分区辨识自然地理要素与现代垦殖率的相关性，进而遴选出与垦殖率显著相关的各自然地理要素，综合构建了全球土地宜垦性

数据。经验证，土地宜垦性结果可有效指示垦殖率的大小。该数据可作为一项指示潜在开垦能力的基础数据，同时该研究方法也为分区构建历史耕地格网化分配算法提供了一种新的思路。

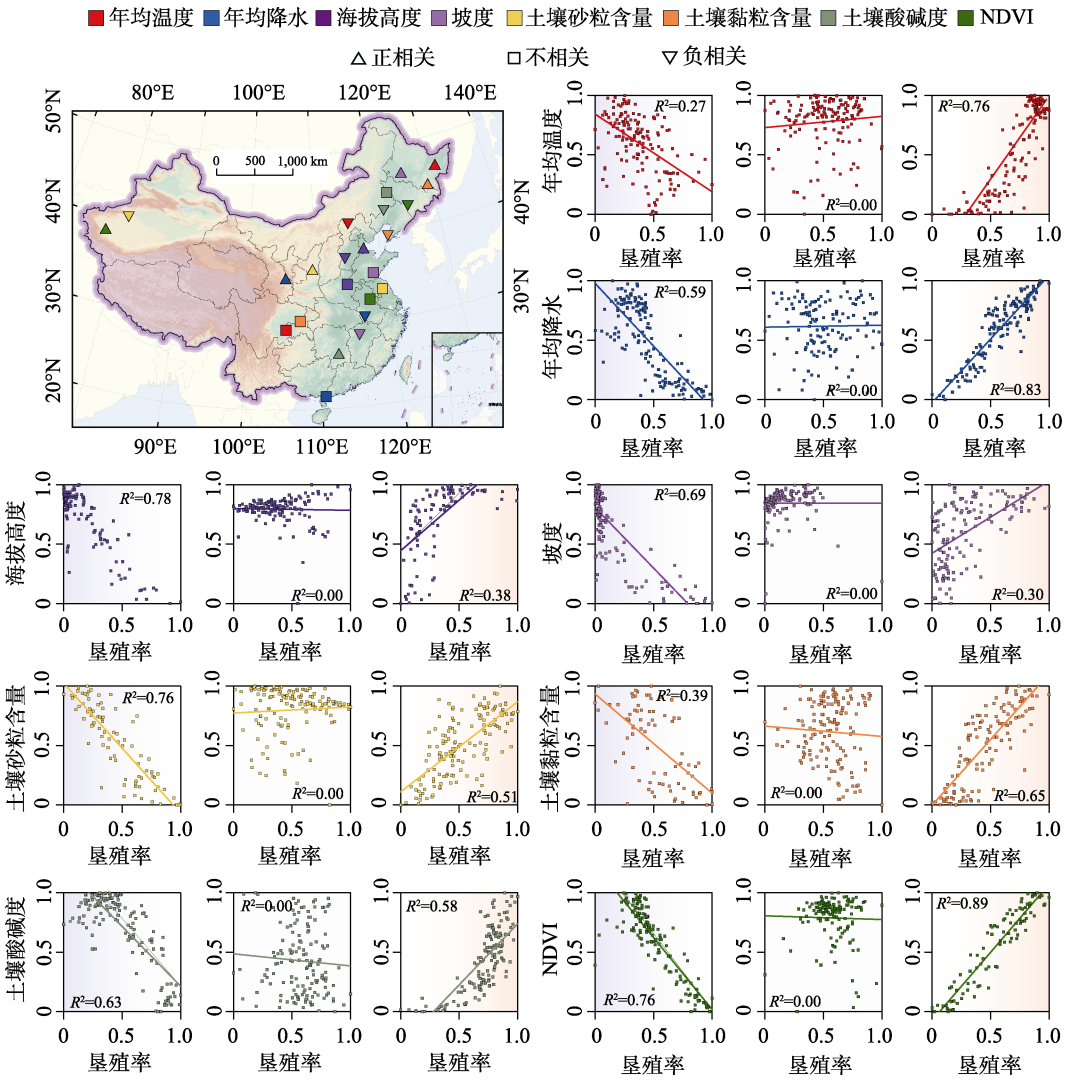


图 5 中国自然地理要素与垦殖率相关性的空间分异示例

(注：左上角的地图中标识了每个样点 (0.5°×0.5°格网) 的坐标和地貌背景。颜色和形状指示是何要素及所属何种相关性。横坐标指示垦殖率，纵坐标指示要素归一化值)

作者分工：叶瑜、方修琦对数据的研制做了总体设计；张成鹏处理了数据、设计了算法，撰写了数据论文。

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] Ellis, E., Kaplan, J., Fuller, D., *et al.* Used planet: a global history [J]. *Proceedings of the National Academy of Sciences*, 2013, 110(20): 7978–7985.
- [2] Foley, J., DeFries, R., Asner, G., *et al.* Global consequences of land use [J]. *Science*, 2005, 309(5734): 570–574.
- [3] Gaillard, M. LandCover6k: global anthropogenic land-cover change and its role in past climate [J]. *PAGES Magazine*, 2015, 23(1): 38–39.
- [4] Lambin, E., Geist, H. Land-use and Land-cover Change: Local Processes and Global Impacts [M]. Berlin: Springer Science & Business Media, 2008.
- [5] Klein Goldewijk, K., Beusen, A., Doelman, J., *et al.* Anthropogenic land use estimates for the Holocene—HYDE 3.2 [J]. *Earth System Science Data*, 2017, 9(2): 927–953.
- [6] Moran, E., Ojima, D., Buchmann, B., *et al.* Global Land Project: Science Plan and Implementation Strategy [M]. Stockholm: IGBP Secretariat, 2005.
- [7] Ramankutty, N., Foley, J., Hall, F., *et al.* ISLSCP II historical croplands cover, 1700–1992 [DB/OL]. ORNL DAAC, 2010.
- [8] Ramankutty, N., Foley, J. Estimating historical changes in global land cover: croplands from 1700 to 1992 [J]. *Global Biogeochemical Cycles*, 1999, 13(4): 997–1027.
- [9] Pongratz, J., Reick, C., Raddatz, T., *et al.* A reconstruction of global agricultural areas and land cover for the last millennium [J]. *Global Biogeochemical Cycles*, 2008, 22(6): 1–16.
- [10] Kaplan, J., Krumhardt, K., Ellis, E., *et al.* Holocene carbon emissions as a result of anthropogenic land cover change [J]. *The Holocene*, 2011, 21(5): 775–791.
- [11] Boyle, J., Gaillard, M., Kaplan, J., *et al.* Modelling prehistoric land use and carbon budgets: a critical review [J]. *The Holocene*, 2011, 21(5): 1–8.
- [12] Pielke, R., Pitman, A., Niyogi, D., *et al.* Land use/land cover changes and climate: modeling analysis and observational evidence [J]. *Wiley Interdisciplinary Reviews: Climate Change*, 2011, 2(6): 828–850.
- [13] 葛全胜, 戴君虎, 何凡能等. 过去 300 年中国部分省区耕地资源数量变化及驱动因素分析[J]. 自然资源进展, 2003, 13(8): 825–832.
- [14] 林珊珊, 郑景云, 何凡能. 中国传统农区历史耕地数据网格化方法[J]. 地理学报, 2008, 61(1): 83–92.
- [15] 叶瑜, 方修琦, 任玉玉等. 东北地区过去 300 年耕地覆盖变化[J]. 中国科学: D 辑, 2009, 39(3): 340–350.
- [16] He, F., Li, S., Zhang, X., *et al.* Comparisons of cropland area from multiple datasets over the past 300 years in the traditional cultivated region of China [J]. *Journal of Geographical Sciences*, 2013, 23(6): 978–990.
- [17] Li, S., He, F., Zhang, X. A spatially explicit reconstruction of cropland cover in China from 1661 to 1996 [J]. *Regional Environmental Change*, 2016, 16(2): 417–428.
- [18] Yang, X., Jin, X., Guo, B., *et al.* Research on reconstructing spatial distribution of historical cropland over 300 years in traditional cultivated regions of China [J]. *Global and Planetary Change*, 2015, 128: 90–102.
- [19] 张成鹏, 叶瑜, 方修琦. 基于自然地理要素的全球土地宜垦性数据集[J/DB/OL]. 全球变化数据仓储电子杂志, 2022. <https://doi.org/10.3974/geodb.2022.04.01.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2022.04.01.V1>.
- [20] 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).
- [21] Fick, S., Hijmans, R. WorldClim2: new 1km spatial resolution climate surfaces for global land areas [J]. *International Journal of Climatology*, 2017, 37(12): 4302–4315.
- [22] Danielson, J., Gesch, D. Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) [M]. Washington, DC, USA: US Department of the Interior, US Geological Survey, 2011.
- [23] Pinzon, J., Tucker, C. A non-stationary 1981–2012 AVHRR NDVI3g time series [J]. *Remote Sensing*, 2014, 6(8): 6929–6960.
- [24] Mantel, S., Kempen, B. SoilGrids250m: global gridded soil information based on machine learning [J]. *PLoS ONE*, 2017, 12: e0169748.
- [25] Zhang, C., Ye, Y., Fang, X., *et al.* Synergistic modern global 1 km cropland dataset derived from multi-sets of land cover products [J]. *Remote Sensing*, 2019, 11(19): 1–18.