

中国太阳辐射分层分区边界数据集的研发

姜 侯^{1,2*}, 姚 凌¹

1. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101;

2. 北京师范大学遥感科学国家重点实验室, 北京 100875

摘 要: 太阳辐射分区是指导太阳能利用和制定区域发展规划的基础。本数据集提供了中国太阳辐射两种区划方案的数据, 包括 5 个主要分区和 10 个子分区的边界。我们利用基于高斯混合模型的自动分区算法进行太阳辐射分区, 其中分区数目由贝叶斯推理自适应推断。分区算法使用 2007–2020 年地面太阳辐射观测数据进行拟合, 然后将对应时间段的遥感反演产品带入拟合后的模型, 确定分层分区的边界; 分区结果采用基于 716 个气象站点的日照时数重构的高精度太阳辐射数据进行验证。验证结果表明分区算法可以将具有不同太阳辐射特性的站点划分到合理的区域, 准确率约为 90%。此外, 大多数不准确的站点位于区域的内部而不是边缘, 这进一步证明了所识别的区域边界的可靠性。

关键词: 太阳辐射; 分区边界; 高斯混合模型; 太阳能利用

DOI: <https://doi.org/10.3974/geodp.2023.02.07>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2023.02.07>

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.2023.07.07.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2023.07.07.V1>.

1 前言

太阳能光伏发电在全球碳中和过程中显示出巨大的潜力^[1]; 由于气候差异, 太阳能发电潜力的空间分布总是显示出明显的地理差异^[2]。太阳辐射分区被认为是了解太阳能区域差异的最直接、最实用的方法, 是太阳能市场战略规划、电站设计和太阳能存储技术选择的基础^[3–5]。例如, 太阳辐射区划可以帮助政府能源部门制定光伏发展规划^[6], 开展太阳能发电场选址工作^[7], 投放公共太阳能设施等^[8]。

气候区划在世界范围内得到了广泛的研究, 研究者根据不同的气候变量和指数建立了多种气候区划体系^[9]。中国广泛使用的建筑气候区划则以建筑物的被动热能利用为参照^[10,11], 将中国气候分为严寒、寒冷、夏热冬冷、夏热冬暖和温和气候等 5 大类。生物气候图理论的发展推动提出了生物气候区划概念^[10,12], Lam 等(2006)分析了中国 18 个典型城市的天气数据, 将中国划分为 9 个生物气候区; Wan 等(2010)基于对夏季和冬季热应激不适的

收稿日期: 2023-03-05; 修订日期: 2023-06-18; 出版日期: 2023-06-25

*通讯作者: 姜侯 DTW-0666-2022, 中国科学院地理科学与资源研究所, jiangh.18b@igsrr.ac.cn

数据引用方式: [1] 姜侯, 姚凌. 中国太阳辐射分层分区边界数据集的研发[J]. 全球变化数据学报, 2023, 7(2): 185–194. <https://doi.org/10.3974/geodp.2023.02.07>. <https://doi.org/10.3974/geodp.2023.02.07>.

[2] 姜侯. 中国太阳辐射两种区划方案数据集[J/DB/OL]. 全球变化数据仓储电子杂志, 2023. <https://doi.org/10.3974/geodb.2023.07.07.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2023.07.07.V1>.

长期调查, 确定了中国的 5 个生物气候区和 8 个子区。除了温度因素外, 气候区划的发展过程中, 研究者还逐步纳入气温、湿度和风速环境因子, 使得气候区划变得更加精细, 但同时也使区划变得碎片化、复杂化。

以太阳能利用为目标的分区方案相对较少, 最初的太阳辐射分区方案根据气象站观测的年度累计太阳辐射总量, 寻找自然间断点来确定分区阈值^[11,13]。例如, Lau 等 (2007) 利用 123 个观测站的太阳辐射观测计算月均晴空指数并进行聚类分析, 提出了 5 个太阳辐射分区的方案^[14]。然而, 由于太阳辐射观测站分布稀疏且不均匀, 往往难以确定远离观测站的点位所属的分区。Liu 等 (2017) 利用 K-Means 聚类将 98 个太阳辐射观测站划分到 5 个分区, 然后使用支持向量机建立各个分区内太阳辐射与相关气象变量之间的经验关系; 这种关系随后被用来预测更密集分布的气象观测站的所属分区, 获得了比 Lau 等 (2017) 更可靠的分区边界^[15]。由于区域划分过程仍然依赖气象站的空间密度, Liu 等 (2017) 的方案在气候过渡带存在错误分区和边界划分不准的问题。空间连续的地表太阳辐射观测是确定太阳辐射分区精确边界的唯一途径; 大量的研究表明, 遥感反演产品可以捕捉公里级至米级的地表太阳辐射区域差异及逐时变化^[16], 因此可用于精细尺度的太阳辐射区划边界识别。此外, 传统聚类分区方法通常需要基于先验知识确定太阳辐射分区的数目, 然而在实践中很难确定最恰当的分区数, 对分区数的不恰当选择通常会产生不理想甚至误导性的结果, 例如过度拟合^[17]。

针对目前分区体系逐渐多样化、分区边界识别不准以及分区数目确定难的问题, 作者的前期研究提出了一种基于高斯混合模型构建太阳辐射分层分析体系、并利用空间连续遥感反演产品确定区划边界的新型算法^[18]。该算法将分区的数目视为随机变量, 然后利用贝叶斯推断从训练数据中自动确定最合适的分区数; 贝叶斯推断还通过选择更恰当的先验分布和额外的物理知识确保分区结果更可靠^[19]。本数据集整合了 2007–2020 年的中国气象局地表太阳辐射观测数据以及时空连续的遥感反演产品, 利用前期研究提出的算法进行中国太阳辐射分区和边界识别, 最终得到了中国太阳辐射分层分区边界数据集。

2 数据集元数据简介

《中国太阳辐射两种区划方案数据集》^[20]的名称、作者、地理区域、数据年代、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 数据研发方法

3.1 数据来源

本数据集研发的基础数据包括地表太阳总辐射的地面站点观测和遥感反演产品。地面测量数据来自中国气象局气象数据中心¹提供的月值数据集, 该数据集提供了 98 个太阳辐射站的月均太阳总辐射, 单位为 0.01 MJ/m^2 , 时间范围为 2007–2020 年。原始数据的质量控制的程序包括对物理阈值、时间连续性和时间序列一致性的自动检查, 以及对自动剔除的错误数据的额外人工检查。本数据集生产时还实施了更严格的质量控制, 即通过与遥感

¹ 中国气象数据网. <http://data.cma.cn/>.

反演的太阳辐射数据^[22]对比,剔除绝对差异比(测量值与重构值的差除以测量值)大于 20% 的数据记录,然后采用反演的多年平均值填补。

表 1 《中国太阳辐射两种区划方案数据集》元数据简表

条 目	描 述
数据集名称	中国太阳辐射两种区划方案数据集
数据集短名	SolarRadiationZones
作者信息	姜侯 DTW-0666-2022, 中国科学院地理科学与资源研究所, jiangh.18b@igsnr.ac.cn
地理区域	中国
数据年代	2007–2020 年多年平均
数据格式	.shp
数据量	3.10 MB
数据集组成	包括 2 个数据文件: 1) 太阳辐射总量差异 5 分区.shp, 为 5 个主要太阳辐射分区的边界矢量; 2) 太阳辐射总量及季节性差异 10 分区.shp, 为 10 个子分区的边界矢量
基金项目	遥感科学国家重点实验室开放基金 (OFSLRSS202204); 国家自然科学基金 (42201382)
数据计算环境	Matlab 2019
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 ^[21]
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOS

本研究中使用的遥感反演产品来自 Jiang 等人的研究^[2], 该研究开发了一种深度学习算法, 根据静止气象卫星逐小时观测影像检索地表太阳总辐射; 算法依靠卷积神经网络处理太阳辐射传输的空间临近效应^[23], 显著提高了地表太阳总辐射的反演精度。2007–2020 年的数据集已经公布^[24], 可以免费获取, 数据集提供了中国 0.05°×0.05°分辨率的地表太阳总辐射月均值。基于地面太阳辐射站的验证表明, 月平均总辐射值的均方根误差约为 1.49 MJ/m²^[25], 精度优于其他广泛使用的产品, 如 CERES-EBAF 和 GEWEX-SRB^[22]。

3.2 分区方法

太阳辐射分区主要依赖于聚类分析, 聚类分析是将数据集分类到不同的类或者簇的过程, 同一个簇中的对象有很大的相似性, 而不同簇间的对象有很大的相异性。本数据集的分区采用基于高斯混合模型 (GMM) 的聚类方法, GMM 能够对任意形状的密度分布进行平滑逼近。我们的前期研究^[18]检验了使用有限 GMM 和无限 GMM 两种形式进行太阳辐射分区中的可行性; 其中, 有限 GMM 在预定义分区数约束下生成分区结果, 类似于 K-Means 方法; 无限 GMM 模型具备贝叶斯特性, 可以自适应地确定合适的分区数。结果表明, 有

限 GMM 模型分区过程中容易出现过拟合效应,无限 GMM 则能有效避免过拟合、并保证分区结果充分反映不同区域太阳辐射的内在差异^[18]。同时,无限 GMM 揭示在中国适合建立 5–10 分区的层次体系,其中 5 个主要分区主要反映太阳辐射的总量差异、10 个子分区进一步反映不同区域太阳辐射年内变化特征的差异。

因此本数据集使用无限 GMM 作为聚类算法实施中国太阳辐射分区。整个算法的实施流程(如图 1)包括以下主要步骤:1)将地面太阳辐射站实测的多年(2007–2020)月平均太阳总辐射作为输入属性信息,基于 GMM 聚类方法对中国 98 个观测站进行聚类;2)将遥感反演获取的空间连续的多年平均(2007–2020)太阳总辐射数据输入步骤 1)拟合的 GMM 模型,预测每个像素属于不同聚类的后验概率,将最大后验概率对应的聚类标签分配给该像素,生成空间连续的分区图;3)由于 GMM 模型直接分区结果通常存在面积积极小的栅格多边形(仅包含几个像素甚至一个像素),进一步使用筛网过滤器去除像素数目小于预定义阈值的多边形,并给这些像素分配其最大相邻多边形的分类标签;4)根据空间连续分区图提取分区边界,并对边界进行平滑等后处理操作。

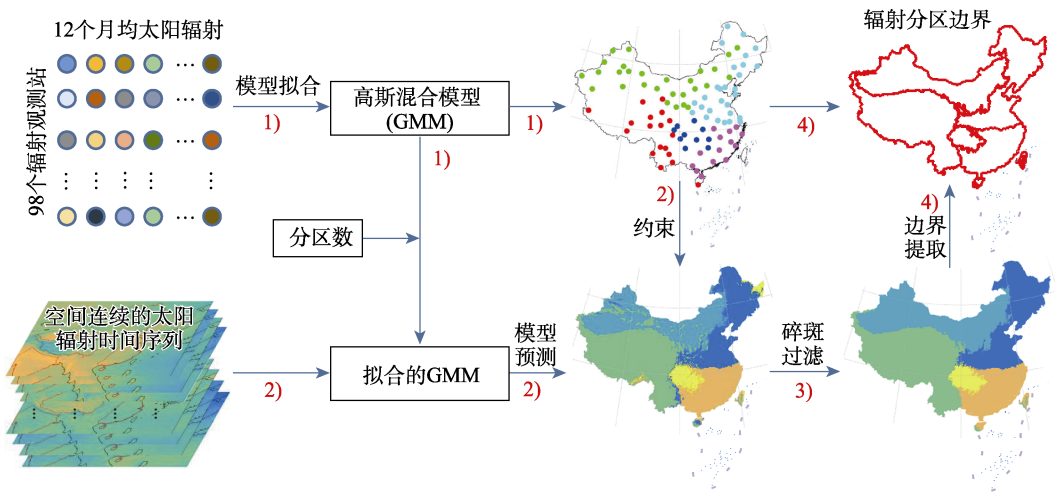


图 1 中国太阳辐射分层分区边界获取流程图

3.3 算法原理

GMM 是由 K 个高斯子分布组成的混合分布,可表示为:

$$p(x|\pi,\mu,\Sigma)=\sum_{k=1}^K\pi_k g(x|\mu_k,\Sigma_k) \tag{1}$$

式中, x 是 N 次观测值(例如 98 个观测站的测量)的数据向量,每个观测值包含 D 维特征(例如 12 个月均太阳总辐射), $\pi_k, k=1, \dots, K$ 代表混合权重且满足 $\sum_{k=1}^K \pi_k=1$, $g(x|\mu_k,\Sigma_k), k=1, \dots, K$, 代表概率子分布。每个子分布是一个 D 维高斯函数,其形式如下:

$$g(x|\mu_k, \Sigma_k) = \frac{1}{\sqrt{2\pi}^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) \right\} \quad (2)$$

式中, μ_k 表示平均向量, Σ_k 表示协方差矩阵。完整的 GMM 由平均向量 μ 、协方差矩阵 Σ 和混合权重 π 参数化, 表示为 $\lambda = \{\pi, \mu, \Sigma\}$ 。与经典的单峰高斯模型、最近邻模型相比, GMM 具有更好的建模能力。期望最大化 (Expectation-maximization, EM) 是 GMM 参数估计的常用算法, EM 的目标是在训练数据上寻找使 GMM 似然函数最大化的模型参数。给定训练观测值 $x = \{x_1, \dots, x_N\}$, GMM 似然函数表示为:

$$p(x|\lambda) = \prod_{n=1}^N p(x_n|\lambda) \quad (3)$$

EM 从一个初始 λ 开始, 并估计一个新的 $\bar{\lambda}$, 使得 $p(x|\bar{\lambda}) \geq p(x|\lambda)$; 随后, 新的 $\bar{\lambda}$ 成为下一次迭代的初始 λ , 并重复该过程直到达到收敛阈值。由公式 1 定义的 GMM 需要在聚类前预先定义太阳辐射分区的数量 (K)。在多数应用中, 正确的 (或合适的) K 是未知的, 并且很难估计。避开该困难的途径是将 K 的极限取为无穷大, 让 GMM 从训练数据中自适应地推断出正确的分区数, 即建立一个具有自适应能力的无限 GMM (具体过程参考文献^[18])。

4 数据结果与验证

4.1 数据集组成

数据集包括 2 个数据文件: 1) 太阳辐射总量差异 5 分区.shp, 为 5 个主要太阳辐射分区的边界矢量; 2) 太阳辐射总量及季节性差异 10 分区.shp, 为 10 个子分区的边界矢量。

4.2 数据结果

图 2 显示了中国太阳辐射分层分区体系的 5 个主要分区及其太阳辐射年变化特征, 底图叠加了中国的地势图以展示各个分区涉及的地理单元。分区 I 包括青藏高原、云贵高原及其周边地区, I 区的月平均地表太阳总辐射始终高于其他区域, 最大值出现在 6 月份, 为 24.8 MJ/m², 年平均太阳总辐射高达 20.1 MJ/m², 这是高海拔、稀薄气溶胶和低云量共同作用的结果^[25]。分区 II 包括准噶尔盆地、塔里木盆地、内蒙古和黄土高原的部分区域, II 区的年平均太阳总辐射 (16.3 MJ/m²) 处于第二阶梯, 但年内绝对差异最大, 6 月最高, 12 月最低, 差值达到 15.6 MJ/m²。分区 III 主要覆盖华北和东北地区, III 区的最大太阳总辐射出现在 6 月份, 月平均值为 18.4 MJ/m², II 区和 III 区的变化几乎同步, 但幅度不同。IV 区秦岭-淮河线以南, 包括长江中下游平原、南方丘陵和云贵高原东部。IV 区由四川盆地和武陵山区组成。IV 区与 V 区的年平均太阳总辐射分别为 12.7 MJ/m² 和 11.2 MJ/m²; 他们的差异主要集中在秋冬两季, 此时 IV 区拥有较多的太阳辐射; IV 区最高太阳总辐射出现在 7 月, 为 18.3 MJ/m², V 区最高值出现在 8 月, 为 16.7 MJ/m²。特别地, 台湾岛和海南岛均涉及两个分区, 台湾岛中央山脉和海南岛西南区域的太阳辐射特征与分区 I 相似, 岛内其他区域的太阳辐射特征与分区 IV 相似。

随着 GMM 确定的分区数逐渐增多, 5 个主分区开始分裂从而产生子区域。图 3a 展示

了分裂至 10 个子分区的结果。5 个主分区分裂的过程如下：首先，准噶尔盆地和锡林郭勒草原从分区Ⅱ中分离；然后分区Ⅰ分为青藏高原和云贵高原两个主要部分；接着分区Ⅱ继续分裂，在黄土高原形成一个新分区；之后，大兴安岭及其西部地区从分区Ⅲ分离；最后，

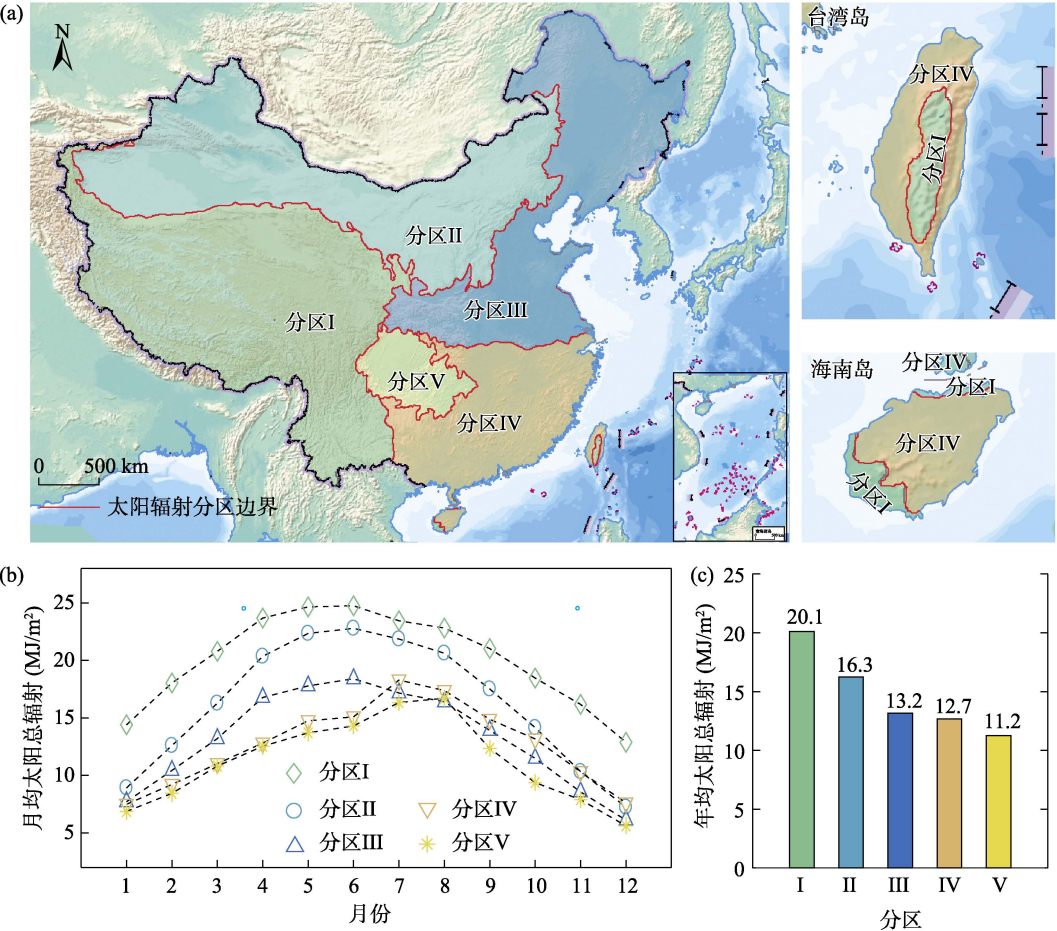


图 2 中国太阳辐射分层分区体系的 5 个主要分区及其太阳辐射年变化特征图
(注：本图基于国家测绘地理信息局标准地图服务网站下载的审图号为 GS(2016)1619 的标准地图制作，底图无修改)

分区Ⅳ分为两部分,其分界线约在 $23^{\circ}\pm 3^{\circ}\text{N}$ 范围内。我们使用两级标签标记不同的子分区,由图可见,子分区空间划分基本处在 5 个主要分区的边界内,且主要分区间的边界几乎保持不变,这表明了 GMM 确定的 5 个主要分区方案的稳定性。海南岛仍然被分为两个分区,其西南区域与分区云贵高原的太阳辐射特征相近,其他区域更接近大陆南方沿海地带。台湾岛则被重新划分到 3 个分区,从北到南依次为分区Ⅳ-A、Ⅳ-B、Ⅰ-B,表明其太阳辐射特征与长江中下游区域、南方沿海区域和云贵高原具有更高的相似性。

图 3b 比较了 10 个子分区的月均太阳总辐射变化。分区Ⅰ-B 的太阳总辐射明显低于分区Ⅰ-A,特别是夏季,这种区别主要是海拔高度差异造成的;受北印度洋季风影响,分区Ⅰ-B 的太阳总辐射波动较大。分区Ⅱ-B 受夏季风影响,5–9 月接收的太阳辐射比分区Ⅱ-A

少；分区 II-A 和分区 II-C 的差异集中在冬季，此时分区 II-C 的太阳高度角整体偏小，致使太阳总辐射比分区 II-A 低；分区 II-C 的东西两部分纬度基本相同，因而具有相似的太阳辐射特性。分区 III-A 的月平均太阳总辐射在夏季比分区 III-B 低，在冬季比分区 III-B 高。分区 IV-B 的太阳总辐射整体上比分区 IV-A 高，尤其是在秋冬季。分区 IV（四川盆地及其周边）由于多云多雨天气频繁，太阳总辐射在所有区域中始终处于最低水平。四川盆地西北部的高山阻挡了携带水汽的东南季风，使盆地内地形雨盛行；同时相对周边地区盆地的海拔更低，构成一个封闭空间使蒸发的水汽聚集形成降雨；常年多雾和多云的天气削弱了四川盆地接收到的太阳辐射。

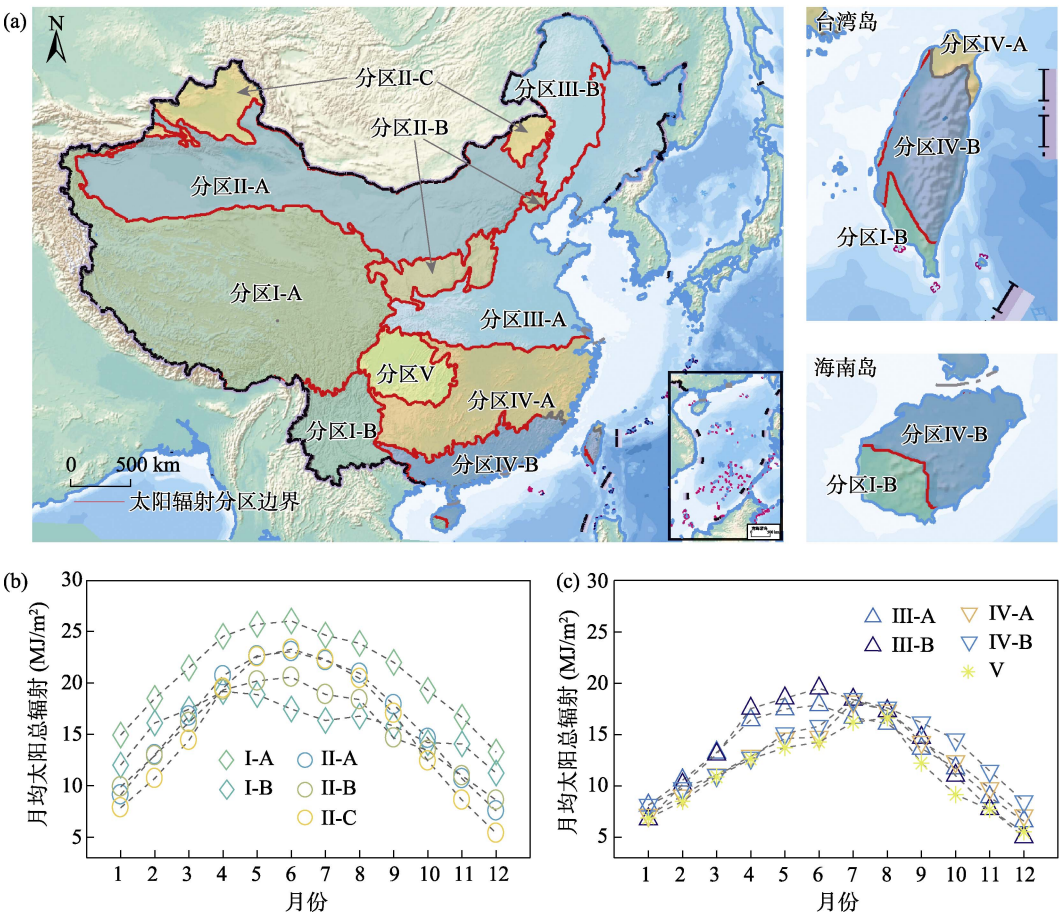


图 3 中国太阳辐射分层分区体系的 10 个子分区及其太阳辐射年变化特征图
(注：本图基于国家测绘地理信息局标准地图服务网站下载的审图号为 GS(2016)1619 的标准地图制作，底图无修改)

4.3 数据结果验证

太阳辐射分区结果验证是具有挑战性的任务，因为没有真实分区标签可以参考。在此，我们利用基于日照时数的太阳辐射数据集^[26]对 716 个气象站点进行人工分区。由 4.2 节结果分析可初步判断，GMM 分区的依据主要是太阳总辐射年总值及其季节性变化差异，我

们据此标准进行人工分类，并将分类结果与 GMM 的 10 分区结果进行对比。人工分类过程如下：1）利用图 3a 中的边界对 716 个站点进行分组，并分别按太阳辐射年平均值和月变化计算每个分组的中心点；2）用动态时间规整算法（Dynamic Time Warping, DTW）计算每个站点到其所属分区中心点的欧式距离；3）使用 DTW 算法计算每个站点到其最近邻分区中心点的欧式距离；4）如果一个站点到所属分区中心点的距离大于到最近邻分区中心点的距离，该站点被重新分配其最近邻分区标签。最终的对比结果如图 4 所示，716 个站点中的 642 个（圆圈）处在原分区内，只有 74 个站点被重新分配到新的分区。该结果表明，GMM 基本将气象站划分到了正确的太阳辐射分区内，且使用空间连续的太阳辐射遥感反演结果确定的边界也是相对可靠的。我们还注意到，大部分被改变分区的站点落在原分区空间范围的内部、而非边缘，这表明基于空间连续的太阳辐射进行分区可以有效避免个别不具代表性站点的影响。

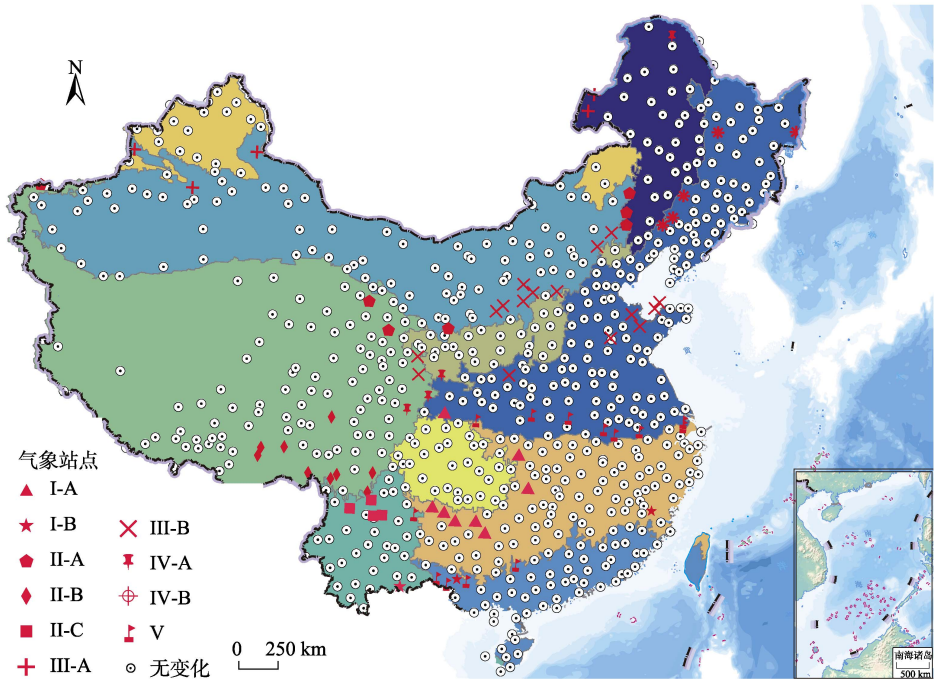


图 4 根据日照时数推算的太阳辐射数据对中国 716 个气象站的人工分区结果图
(注：紫色符号表示该点被重新划分到其后标签所指示的分区，而非所在位置对应的分区。底图边界基于国家测绘地理信息局标准地图服务网站审图号为 GS(2016)1619 的标准地图制作)

5 讨论和总结

本数据集提供了中国太阳辐射 5 分区至 10 分区的分层分区结果。基于辐射观测站的实验和空间连续太阳总辐射估算的分区结果都支持构建层次分区体系。5 个主要分区主要反映我国不同气候带造成的太阳辐射差异；10 个子分区揭示了局部微气候和地形影响下的细微差异。四川盆地、青藏高原、云贵高原、黄土高原等具有鲜明气候特征的基本地理单元

形成了独立的太阳辐射分区。10个分子区与建筑气候分区存在一定的相似性,例如,III-A和IV-A、IV-A和IV-B、II-A和II-C-1之间的边界基本上与建筑气候分区边界一致^[14, 27, 28],这表明10个子分区也反映了与太阳辐射差异关联的地表热分布差异。

本数据集利用空间连续的太阳辐射估计识别了相邻分区的精确边界,这克服了基于站点分区方法的不足。例如,海南岛和台湾岛内部太阳辐射特征具有明显差异,恰当的分区方案应该将它们切分为不同的分区(如图2a和3a所示);如果只根据其内部的一个或几个观测者分区,海南岛和台湾岛很可能被视为一个分区。空间上的连续估计区分了太阳辐射的细微差别,因此能够提供精确的像素级的分区结果,这对于某些特定应用非常重要,例如评估山区屋顶太阳能光伏发电潜力为能源部门决策服务^[5, 6],在海南岛北部和南部、台湾岛西部和东部应实施不同的太阳能利用策略。

作者分工: 姜侯设计了算法;姚凌采集和处理了观测数据;姜侯撰写了数据论文。

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] 丁仲礼, 中国碳中和框架路线图研究[J]. 中国工业和信息化, 2021(8): 54–61.
- [2] Jiang, H., Lu, N., Qin, J., *et al.* A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data [J]. *Renewable and Sustainable Energy Reviews*, 2019, 114: 109327.
- [3] Kabir, E., Kumar, P., Kumar, S., *et al.* Solar energy: potential and future prospects [J]. *Renewable and Sustainable Energy Reviews*, 2018, 82: 894–900.
- [4] Zhang, Y., Ren, J., Pu, Y., *et al.* Solar energy potential assessment: a framework to integrate geographic, technological, and economic indices for a potential analysis [J]. *Renewable Energy*, 2020, 149: 577–586.
- [5] Bódis, K., Kougias, I., Jäger-Waldau, A., *et al.* A high-resolution geospatial assessment of the rooftop solar photovoltaic potential in the European Union [J]. *Renewable and Sustainable Energy Reviews*, 2019, 114: 109309.
- [6] Sweerts, B., Fenninger, S., Yang, S., *et al.* Estimation of losses in solar energy production from air pollution in China since 1960 using surface radiation data [J]. *Nature Energy*, 2019, 4(8): 657–663.
- [7] Mensour, O. N., Ghazzani, B. E., Hlimi, B., *et al.* A geographical information system-based multi-criteria method for the evaluation of solar farms locations: a case study in Souss-Massa area, southern Morocco [J]. *Energy*, 2019, 182: 900–919.
- [8] 蔺阿琳, 城市太阳能可利用空间评估与规划研究[D]. 哈尔滨: 哈尔滨工业大学, 2020.
- [9] Walsh, A., Cóstola, D., Labaki, L. C. Review of methods for climatic zoning for building energy efficiency programs [J]. *Building and Environment*, 2017, 112: 337–350.
- [10] Wan, K., Li, D., Yang, L., *et al.* Climate classifications and building energy use implications in China [J]. *Energy and Buildings*, 2010, 42(9): 1463–1471.
- [11] 王炳忠. 中国太阳能资源利用区划[J]. 太阳能学报, 1983(3): 221–228.
- [12] Al-Azri, N. A., Zurigat, Y. H., Al-Rawahi, N. Z. Development of bioclimatic chart for passive building design [J]. *International Journal of Sustainable Energy*, 2013, 32(6): 713–723.
- [13] 周扬, 吴文, 胡莹等. 西北地区太阳能资源空间分布特征及资源潜力评估[J]. 自然资源学报, 2010,

- 25(10): 1738–1749.
- [14] Lau, C. C. S., Lam, J. C., Yang, L. Climate classification and passive solar design implications in China [J]. *Energy Conversion and Management*, 2007, 48(7): 2006–2015.
- [15] Liu, Y. F., Yong, Z., Wang, D., *et al.* Classification of solar radiation zones and general models for estimating the daily global solar radiation on horizontal surfaces in China [J]. *Energy Conversion and Management*, 2017, 154: 168–179.
- [16] Jiang, H., Lu, N., Qin, J., *et al.* Hourly 5-km surface total and diffuse solar radiation in China, 2007–2018 [J]. *Scientific Data*, 2020, 7(1): 311.
- [17] Antoniak, C. E., Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems [J]. *The Annals of Statistics*, 1974, 2(6): 1152–1174.
- [18] Jiang, H., Lu, N., Qin, J., *et al.* Hierarchical identification of solar radiation zones in China [J]. *Renewable and Sustainable Energy Reviews*, 2021, 145: 111105.
- [19] Soubdhan, T., Emilion, R., Calif, R. Classification of daily solar radiation distributions using a mixture of Dirichlet distributions [J]. *Solar Energy*, 2009, 83(7): 1056–1063.
- [20] 姜侯. 中国太阳辐射两种区划方案数据集 [J/DB/OL]. 全球变化数据仓储电子杂志, 2023. <https://doi.org/10.3974/geodb.2023.07.07.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2023.07.07.V1>.
- [21] 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 更新).
- [22] Zhang, X. T., Liang, S., Wild, M., *et al.* Analysis of surface incident shortwave radiation from four satellite products [J]. *Remote Sensing of Environment*, 2015, 165: 186–202.
- [23] Jiang, H., Lu, N., Huang, G., *et al.* Spatial scale effects on retrieval accuracy of surface solar radiation using satellite data [J]. *Applied Energy*, 2020, 270: 115178.
- [24] Jiang, H., Lu, N. High-resolution surface global solar radiation and the diffuse component dataset over China [DB/OL]. Pangaea, 2019. <https://doi.pangaea.de/10.1594/PANGAEA.904136>.
- [25] Jiang, H., Yang, Y., Wang, H., *et al.* Surface diffuse solar radiation determined by reanalysis and satellite over East Asia: evaluation and comparison [J]. *Remote Sensing*, 2020, 12: 1387.
- [26] 唐文君, 中国 716 个气象站太阳辐射日均值数据集 (1961–2010) [DB/OL]. 国家青藏高原科学数据中心, 2015. <https://doi.org/10.11888/AtmosphericPhysics.tpe.249399.file>.
- [27] Wan, K. K. W., Yang, H. L., Yang, L., *et al.* An analysis of thermal and solar zone radiation models using an Angstrom–Prescott equation and artificial neural networks [J]. *Energy*, 2008, 33(7): 1115–1127.
- [28] Bai, L., Yang, L., Song, B., *et al.* A new approach to develop a climate classification for building energy efficiency addressing Chinese climate characteristics [J]. *Energy*, 2020, 195: 116982.