

中国 30 年平均 1-km 月度气候要素数据集 (1951–1980, 1981–2010)

程 琦¹, 吴星麒¹, 魏临风¹, 胡潇飞¹, 倪 健^{1,2*}

1. 浙江师范大学化学与生命科学学院, 金华 321004;

2. 浙江金华山亚热带森林生态系统野外科学观测研究站, 金华 321004

摘 要:本文汇集整编了中国国家气象台站 1951–1980 年和 1981–2010 年各 30 年平均气象记录, 利用薄板光顺样条法和 ANUSPLIN 4.4 软件对这 2 个时间段的 3 个气候指标(气温、降水和日照百分率)进行空间插值, 并采用广义交叉验证法、平均绝对误差和均方根误差以及线性拟合等方法, 检验台站数据和插值数据的精度, 确定插值数据的准确可靠性, 最终获得 1-km 空间分辨率和 3 种不同格式(ASCII 字符集编码的文本文件、二维均匀栅格和标签图像文件)的气候栅格空间数据, 并简述其空间分布格局与趋势。

关键词: 中国; 气温; 降水; 日照百分率; 月度; 30 年

DOI: <https://doi.org/10.3974/geodp.2022.04.04>

CSTR: <https://cstr.science.org.cn/CSTR:20146.14.2022.04.04>.

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.2022.06.03.V1> 或 <https://cstr.science.org.cn/CSTR:20146.11.2022.06.03.V1>

1 前言

在生态学中, 气象记录和气候数据是许多研究的基础, 例如生态系统、群落、种群和物种地理分布格局与气候的关系^[1–4], 生物多样性格局的气候解释^[5], 以及植被生物量 and 生产力时空变化的气候驱动机制^[6–8]等。尤其是在气候变化背景下, 可进一步模拟物种分布及其对气候变化的响应^[9,10], 也可以模拟和预测气候变化对生态系统格局与功能的影响^[9–13]。因此, 气象记录和气候数据不仅是大气科学的基础, 也是生态学以及地球科学等众多学科和研究领域的重要驱动数据, 与人类生活也是密切相关的。

气象观测资料通常来自气象台站的连续器测记录, 但由于站点的数量有限和空间分布不均衡, 很难直接应用于大尺度的生态学研究, 因此需要对气候要素空间进行插值。通常采用克里格法(Kriging)、薄板光顺样条法(Thin Plate Smoothing Spline)等对气候要素进

收稿日期: 2022-07-01; 修订日期: 2022-11-17; 出版日期: 2022-12-24

基金项目: 中华人民共和国科学技术部(2019QZKK0402)

*通讯作者: 倪健 I-7067-2012, 浙江师范大学, nijian@zjnu.edu.cn

数据引用方式: [1] 程琦, 吴星麒, 魏临风等. 中国 30 年平均 1-km 月度气候要素数据集(1951–1980, 1981–2010)

[J]. 全球变化数据学报, 2022, 6(4): 533–544. <https://doi.org/10.3974/geodp.2022.04.04>.
<https://cstr.science.org.cn/CSTR:20146.14.2022.04.04>.

[2] 程琦, 吴星麒, 魏临风等. 中国 30 年平均 1-km 月度气候要素数据集(1951–1980, 1981–2010) [J/DB/OL]. 全球变化数据仓储电子杂志, 2022. <https://doi.org/10.3974/geodb.2022.06.03.V1>. <https://cstr.science.org.cn/CSTR:20146.11.2022.06.03.V1>.

行空间插值,国内外学者研究表明,薄板光顺样条法插值所得气候空间结果较优^[14-16]。此方法,以及基于此方法的软件 ANUSPLIN^[17,18],已经在国际上得到了广泛应用。

目前,全球和区域尺度上有很多的气候要素空间插值数据集^[3]。全球尺度上常用的,有来自 WorldClim v2.1¹的 1970–2000 年平均的 7 个气候要素(年均温、最低温、最高温、年降水、太阳辐射、风速、水汽压)的 4 种空间分辨率(10'、5'、2.5'、30")的空间数据,以及计算获得的 19 种生物气候指标^[19];还有英国气候研究小组(CRU)的 0.5°分辨率网格点 1901–2020 年 9 个气候指标的逐月数据集——CRU TS v4.05^[20]。我国自 1950 年以来逐渐建立了正规的分布较广的器测气象台站,1980 年以后台站的数量和气象要素的资料集逐渐丰富起来^[21];但由于我国幅员辽阔,气象台站数量仍然较少且分布不均,尤其是西部高原和西北荒漠地带,因此,气候要素的空间插值工作就显得尤为重要。最近 20 年来,我国学者针对不同的研究目的建立了多个气候要素空间数据集,例如,近期发表的不同时间段的全国逐月气温和降水 1 km 空间分辨率数据集,有 2000–2012 年^[22],也有 1901–2017 年^[23]和 1980–2017 年^[24],1981–2015 年 0.05°每 8 日的蒸发与蒸散比率的数据集^[25],以及 1979–2018 年全国高时空分辨率的气象驱动数据集^[26]。但这些数据集时间跨度不一,使用目的和使用人群不一,时间序列数据集较多,但对生态学研究至关重要的气候长期平均态数据的涉及不多。考虑到气候要素在时间上的变异性,生态学研究常用多年平均的气候要素值描述一个地区的气候特征,这种气候数据对于植被和物种分布的模拟尤其重要^[4]。因此,亟需建立一个覆盖时间序列较长、分辨率较高的多年平均气候数据集。本文以 1951–1980 年和 1981–2010 年两个时间段的全国气象台站观测记录为对象,利用薄板光顺样条法和 ANUSPLIN 软件,将各 30 年平均气温、降水和日照百分率的月值插值为 1 km 空间分辨率数据,以期为我国生态学和地学研究提供一个有效的多年平均气候要素数据集。

2 数据集元数据简介

《中国 30 年平均 1-km 月度气候要素数据集(1951–1980, 1981–2010)》^[27]元数据信息见表 1。

3 数据研发方法

3.1 数据采集

基础气象数据包括 1951–1980 年和 1981–2010 年中国地面气候标准值月值数据集,其中前者来源于中国地面气候资料集(1951–1980)^[29]以及少部分个人收集数据,后者来源于国家气象科学数据中心²,台湾省数据来自台湾地区气象机构气象资料开放平台³。数据集元数据涉及全国基本、基准地面气象观测站记录的气温、降水、日照时间和日照百分率、空气湿度、蒸发、积雪、风、地温等要素的月气候标准值,本文只选择年均气温、降水和日照百分率 3 个要素进行插值。其中前者包括全国 673 个国家基准站和基本站,

¹ <http://worldclim.org>.

² <http://data.cma.cn>.

³ <https://www.cwb.gov.tw>.

表 1 《中国 30 年平均 1-km 月度气候要素数据集（1951–1980, 1981–2010）》元数据简表

条 目	描 述		
数据集名称	中国 30 年平均 1-km 月度气候要素数据集（1951–1980, 1981–2010）		
数据集短名	ChinaClimate_1951-2010		
作者信息	程琦, 浙江师范大学, 875544767@qq.com; 吴星麒, 浙江师范大学, 1632314650@qq.com 魏临风, 浙江师范大学, 552535060@qq.com; 胡潇飞, 浙江师范大学, 976860215@qq.com 倪健 1-7067-2012, 浙江师范大学, nijian@zjnu.edu.cn		
地理区域	中国	数据年代	1951–1980 年、1981–2010 年
时间分辨率	30 年年平均	空间分辨率	1 km
数据格式	.asc、.tif、.grd	数据量	3.35 GB（压缩后）
数据集组成	由 8 个数据文件组成, 包含 1951–1980 年、1981–2010 年两个时间段的气象要素数据, 每个时间段内有 3 种不同文件格式		
基金项目	中华人民共和国科学技术部（2019QZKK0402）		
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 ^[28]		
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS		

还包括 200 个个人收集的国家一般站或省级气象站的气象记录, 剔除时间跨度较短的气象台站后, 最终有 858 个气象台站记录纳入插值; 后者包括 2,152 个国家基准站、基本站和一般站的气象记录 (图 1)。因此, 前后两个时间段的台站数量有差异, 尤其是西部地区, 而且早期部分台站的划分级别不确定, 故统称为国家地面气象台站。

本文使用的两套地面气候标准值月值数据集, 均为国家气象信息中心按照标准的国家气候资料统计整编方法整编统计而得。其中, 分别对存在问题和分歧的站点文件进行了

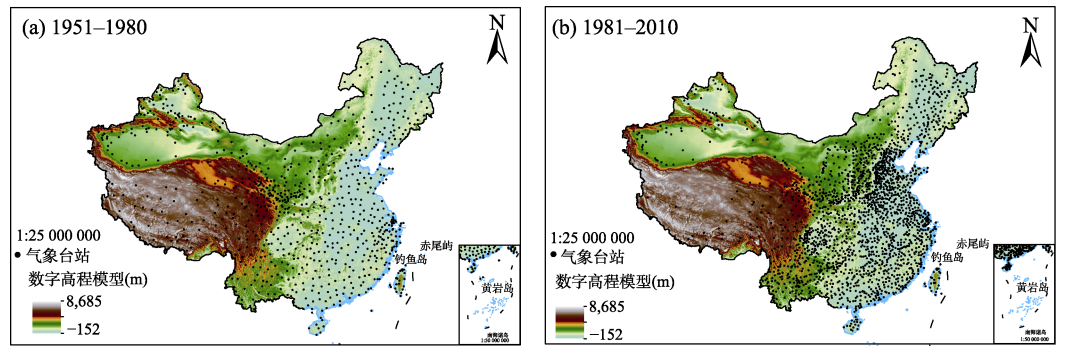


图 1 中国气象台站空间分布图（依据审图号 GS(2022)1686 号的标准地图制作）

修订,并对明显存在非均一性的台站(如迁站)进行了分段处理,由此保证了数据资料的统一性和准确性。此外,本文对两套原始数据集分别进行插值,其各自的数据是相对独立的,其中一套数据的迁站等问题不会影响另外一套数据集的操作。

3.2 算法原理

利用地理参照的薄板光顺样条法^[17,18]和 ANUSPLIN 4.4 软件^[30],结合美国航天飞机雷达地形测绘使命(SRTM)数字高程模型^[31],原始记录为 90-m 空间分辨率,重采样为 0.01°($\approx 1\text{ km}$)精度,将每个气候变量的 12 个月值插值为 0.01°空间分辨率的栅格数据。

薄板光顺样条法是一种得到广泛验证的气候曲面拟合与插值方法,可把离散数据拟合为具有一到多个自变量的样条函数曲面,目前在全球气候插值方面应用广泛,该方法及其软件 ANUSPLIN 在我国也得到普遍使用,且已经证实了其坚实的可靠性和与其他插值方法相比的优越性^[14,32,33]。基于该方法的 ANUSPLIN 软件在插值时,以经纬度作为样条函数自变量,以海拔高程作为协变量以建立温度沿垂直高度变化的模型,构建局部薄板光顺样条,从而拟合 12 个月气温比降在时间和空间上的变化,实现气温插值。而对降水量插值,采用原型薄板光顺样条函数拟合曲面,把海拔作为自变量而非协变量。对日照百分率的插值,则未采用海拔高度进行拟合,但可以采用降水量数据进行验证。由此,本文采用以上常规的参数设置,对全国气温、降水和日照百分率进行拟合插值,其他参数采用软件中的标准设置^[30]。

4 数据结果与验证

4.1 数据集组成

1951–1980 年和 1981–2010 年全国 0.01°空间分辨率气温、降水、日照百分率栅格数据集,共 8 个压缩文件夹,其文件命名规则如下:ChinaClimate_S_Z.rar。S 表示年份,分别代表 1951–1980 年和 1981–2010 年。Z 表示数据的格式,分别为二维均匀栅格格式(.grd)、ASCII 字符集编码的文本文件格式(.asc)和标签图像文件格式(.tif)。这样,本数据集有三套格式,可供不同研究目标使用。

4.2 数据结果

通过 ANUSPLIN 插值所获的数据为经度、纬度、气候值的二维均匀栅格格式,使用 ArcGIS 10.5 版本的 python 程序将其转化为 ASCII 字符集编码的文本文件和标签图像文件两种格式。对 3 个气候要素的空间分布特征进行简单统计分析,以中国植被分区图^[34]作为底图,利用 GIS 绘制各气候指标的空间分布图,宏观简述 2 个时间段的各气候要素指标 1 月、7 月和年值的空间分布特征,比较分析 2 个时间段气候要素年值在时间上的变化趋势,包括年均气温和日照百分率气候要素在后 30 年与前 30 年的差值,以及年均降水量变化百分率=(后 30 年值-前 30 年值)/前 30 年值。

4.2.1 温度

我国气温指标总体表现出从东南向西北递减的趋势。1 月均温(图 2a 和 2b)的南北差距较大,自南向北呈递减趋势,华南和华中地区相对较高,而东北、华北、西北和青藏高原地区相对较低。从各个植被分区的平均值来看,热带季风雨林、雨林区域最高,亚热带

常绿阔叶林区域次之,二者均高于 0°C ,暖温带落叶阔叶林区域偏低 -5°C ,温带荒漠区域和温带草原区域分别低于 -10°C 和 -15°C ,青藏高原高寒植被区域则介于前二者之间,温带针叶、落叶阔叶混交林接近 -20°C ,寒温带针叶林区域则接近 -30°C 。1981–2010 年 1 月均温较 1951–1980 年有所升高,其中寒温带针叶林区域、热带季雨林、雨林区域和温带针叶、落叶阔叶混交林的上升幅度相对较大,青藏高原高寒植被区域升温最小。

7 月均温 (图 2c 和 2d) 的南北差异相对较小,全国夏季温度普遍较高,但青藏高原以及中部和西部高山地区,7 月均温普遍较低。从各个植被分区的平均值来看,相较于 1 月均温,7 月均温的区域差异更小,青藏高原高寒植被区域最小为 $6-7^{\circ}\text{C}$ 左右,寒温带针叶林区域较小在 $17-18^{\circ}\text{C}$ 之间,其他区域均在 20°C 之上,且以暖温带落叶阔叶林区域最大为 24°C 左右。1981–2010 年 7 月均温较 1951–1980 年也有所升高,其中青藏高原高寒植被区域的上升幅度相对较高,达 1°C 以上。

年均温 (图 2e 和 2f) 同样表现出从南到北、从东南到西北的递减趋势,以华南、华中和西北地区相对较高,而东北和青藏高原地区相对较低。从各个植被分区的平均值来看,各区域年均温差异同样偏小,青藏高原高寒植被区域年均温最低为 -4°C 左右,寒温带针叶林区域稍高,但同样低于 0°C ,温带针叶、落叶阔叶混交林和温带荒漠区域分别高于 1°C 和 5°C ,温带草原区域则介于两者之间,其他区域均高于 10°C ,且以热带季雨林、雨林区域最大为 $16-17^{\circ}\text{C}$ 之间。从两个时间段的差值 (图 2g) 来看,过去 60 年来我国年均温升高,大部分地区都呈现上升趋势,尤以青藏高原南缘横断山脉区上升幅度最大,中部和西南部少数地区呈下降趋势,且以天山山脉地区的下降幅度最大。

4.2.2 降水

我国降水指标总体表现出从东南向西北递减的趋势。1 月降水 (图 3a 和 3b) 的南北差距较大,自东南向西北呈递减趋势,以华南地区相对较高,其他地区都相对较低。从各个植被分区的平均值来看,1 月降水量的区域差异较小,亚热带常绿阔叶林区域最大为 $30-35\text{ mm}$ 左右,热带季雨林、雨林区域较小在 $18-19\text{ mm}$ 左右,其他区域均在 10 mm 之下,且以温带荒漠区域最小为 2.5 mm 左右。1981–2010 年我国 1 月降水量总体较 1951–1980 年有所升高,其中寒温带针叶林区域、温带针叶、落叶阔叶混交林区域、亚热带常绿阔叶林区域和温带荒漠区域呈上升趋势,且亚热带常绿阔叶林区域的上升幅度相对较大,年均达 6 mm ,而其他区域则有小幅度下降,热带季雨林、雨林区域下降最多为 1 mm 。

7 月降水 (图 3c 和 3d) 的南北差异相对较大,全国夏季降水量普遍较高,但西北地区普遍较低。从各个植被分区的平均值来看,7 月降水量各区域差异较大,温带荒漠区域最小为 30 mm 左右,青藏高原高寒植被区域在 $80-90\text{ mm}$ 之间,温带草原区域在 100 mm 左右,其他区域均高于 130 mm ,且以热带季雨林、雨林区域最高超过 200 mm 。1981–2010 年我国 7 月降水量总体较 1951–1980 年有所升高,其中仅有寒温带针叶林区域、暖温带落叶阔叶林区域和温带草原区域有较小幅度的下降,其他 5 个植被区域都呈递增趋势,且以热带季雨林、雨林区域上升幅度最大,超过 30 mm 。

年降水 (图 3e 和 3f) 同样表现出从南到北、从东南到西北的递减趋势,南北降水差大,

华南地区相对较高，而西北、北方和青藏高原地区相对较低。从各个植被分区的平均值来看，热带季风雨林、雨林区域最高接近 1,350 mm，亚热带常绿阔叶林区域稍低，但同样超过 1,200 mm，温带针叶、落叶阔叶混交林和暖温带落叶阔叶林区域较小在 600–700 mm 之间，寒温带针叶林区域和青藏高原高寒植被区域分别接近 500 mm 和 350 mm，温带草原区域则介于前二者之间，温带荒漠区域最小为 140 mm 左右。1981–2010 年年降水量较 1951–1980 年有所降低，其中暖温带落叶阔叶林区域的下降幅度相对较大，达 48 mm 以上。从两个时间段的年降水量变化百分率（图 3g）来看，过去 60 年我国总体降水趋势变化复杂，东南、华南和西南部分地区增加趋势明显，西北、青藏高原和东北地区也有较大增加，而中部地区降水减少趋势明显。

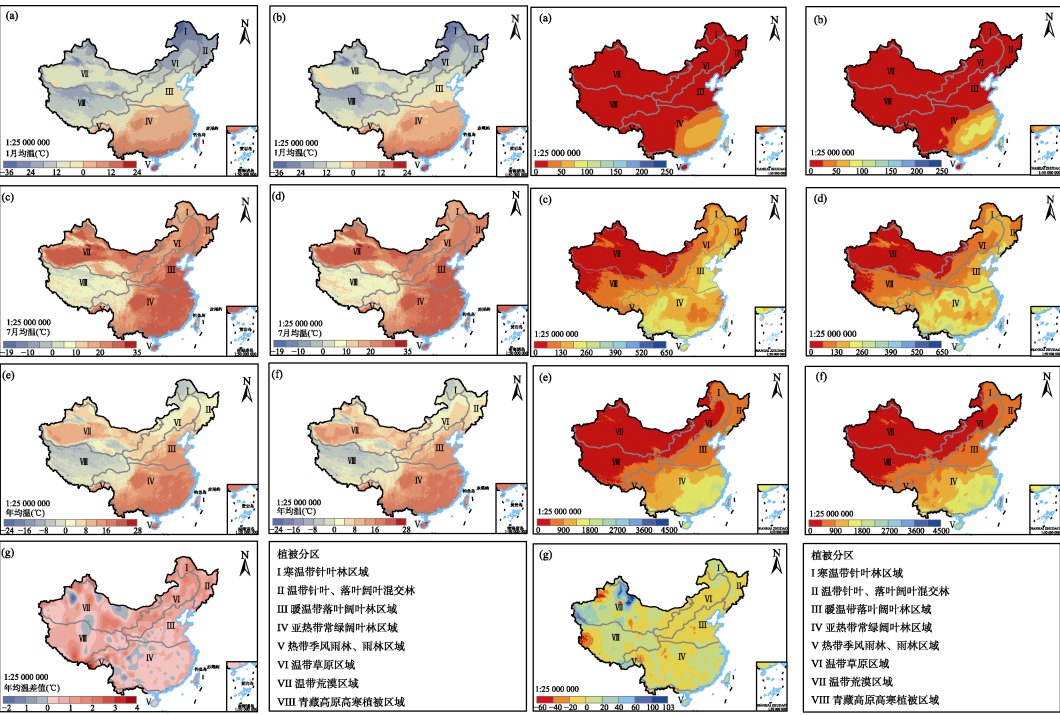


图 2 中国气温空间分布图
（依据审图号 GS(2022)1686 号的标准地图制作）

4.2.3 日照百分率

从数据分析可以看出，中国 1 月日照百分率（图 4a 和 4b）的南北差距较大，南方低，西北和青藏高原东南部分地区较低，其他地区相对较高。从各个植被分区来看，温带草原区域、温带荒漠和青藏高原高寒植被区域较高，三者均高于 65%；寒温带针叶林和热带季风雨林、雨林区域分别低于 65%和 50%；温带针叶、落叶阔叶混交林和暖温带落叶阔叶林区域则介于二者之间，亚热带常绿阔叶林最小为 35%–40%。1981–2010 年 1 月日照百分率较 1951–1980 年整体有所降低，其中温带针叶、落叶阔叶混交林和暖温带落叶阔叶林区域的下降幅度相对较大，达 5%以上，而青藏高原高寒植被区域上升幅度接近 2%。

图 3 中国降水空间分布图

（依据审图号 GS(2022)1686 号的标准地图制作）

7 月日照百分率（图 4c 和 4d）的东西差异相对较大，全国夏季日照百分率普遍较高，但西南和东北地区普遍较低，尤其是前者。从各个植被分区来看，温带荒漠区域最高，温带草原区域和青藏高原高寒植被区域次之，三者均高于 55%，暖温带落叶阔叶林区域和寒温带针叶林区域偏低，在 50% 左右，其他区域均在 50% 以下。1981–2010 年 7 月日照百分率较 1951–1980 年整体有所降低，其中暖温带落叶阔叶林区域和亚热带常绿阔叶林区域的下降幅度相对较大，达 5% 以上，而青藏高原高寒植被区域上升幅度接近 1%。年均日照百分率（图 4e 和 4f）同样表现以南方为中心、从东南到北和西北的递增趋势，华北、北方、西北和青藏高原地区相对较高，而华南地区相对较低。从各个植被分区来看，温带草原区域、温带荒漠区域和青藏高原高寒植被区域较高，三者均高于 60%，寒温带针叶林区域、

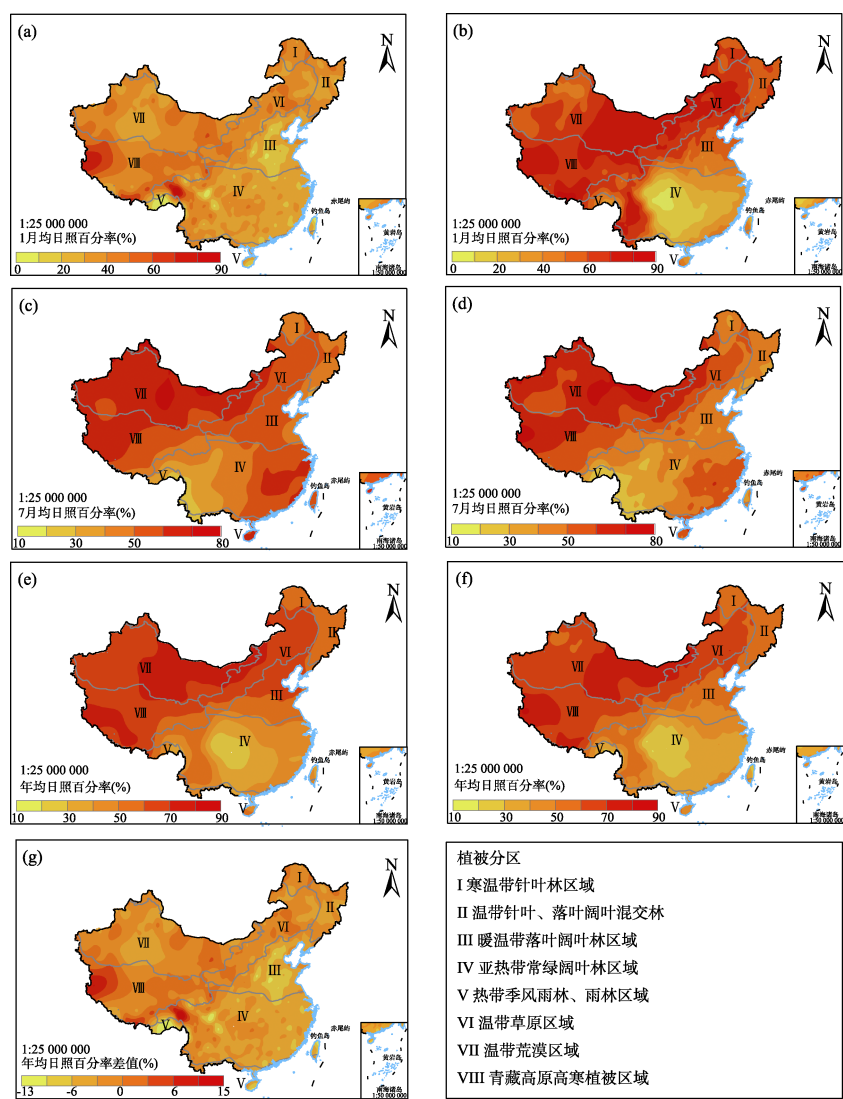


图 4 中国日照百分率空间分布图
(依据审图号 GS(2022)1686 号的标准地图制作)

温带针叶、落叶阔叶混交林和暖温带落叶阔叶林区域偏低,三者均在 50%–60%之间,亚热带常绿阔叶林区域和热带季风雨林、雨林区域较低,二者均低于 50%。

1981–2010 年年均日照百分率较 1951–1980 年整体有所降低,但青藏高原高寒植被区域上升幅度接近 1%,其他区域的下降幅度在 5%以内。从两个时间段的年均日照百分率差值(图 4g)来看,过去 60 年我国大部分地区日照百分率都呈现出下降趋势,尤其是华东大部分地区、华南部分地区,以及西北和青藏高原部分地区,而北方中西部和青藏高原西部和东北部地区则呈增加趋势。

4.3 数据结果验证

4.3.1 质量控制

ANUSPLIN 利用广义交叉验证法 (Generalized Cross Validation, GCV) 对插值结果和观测值进行分析对比,从而检验插值数据的可靠性。本文采用广义交叉验证均方根 RTGCV、绝对误差 MAE 和均方根误差 RMSE 作为评估指标^[24,30]。具体计算公式如下:

$$RTGCV = \frac{1}{N} \frac{\sum_{i=1}^N (O_i - f(x_i))^2}{\left(1 - \frac{d_f}{N}\right)^2} \quad (1)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - O_i| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2} \quad (3)$$

式中, O_i 表示真实的观测值, $f(x_i)$ 是点 i 周边组成变量 x 的光滑函数值, $\sum_{i=1}^N (O_i - f(x_i))^2$ 表示残差平方和, d_f 为模型的自由度, P_i 表示插值的结果, N 为样本数。

同时,根据两套地面气候标准值月值数据集中台站经纬度坐标及其原始气候数据,通过 ArcGIS 软件的“Extract values from point”工具提取插值数据集各要素对应数值,利用线性拟合验证原始数据与插值数据之间的对应性,从而进一步检验插值数据集的精度和可信度。另外,气象台站密度和分布格局也会影响插值数据的精度。本文两个时间段地面气候数据所涉及的台站数量不同,我们也进一步比较了不同台站密度的两套插值数据,其平均绝对误差 MAE 和均方根误差 RMSE 的差异。

4.3.2 质量评估

通过对所有输入数据进行广义交叉验证法检验,根据其均方根 RTGCV 值可以判断整体数据的优劣。从 1951–1980 年和 1981–2010 年的月气候要素的 RTGCV (图 5) 可以看出,气温和降水的 RTGCV 变化较为稳定。对于前 30 年,气温的 RTGCV 大致在 0.7–1.5 °C 范围内波动,在冬季最大,夏季最小;降水的 RTGCV 大致在 0.7–1.6 mm 的范围内波动,在秋季最大,春季最小。对于后 30 年,气温的 RTGCV 大致在 0.4–0.7 °C 范围内波动,冬季最大,夏季最小;降水的 RTGCV 大致在 0.3–0.8 mm 的范围内波动,在夏季最大,春季最小。

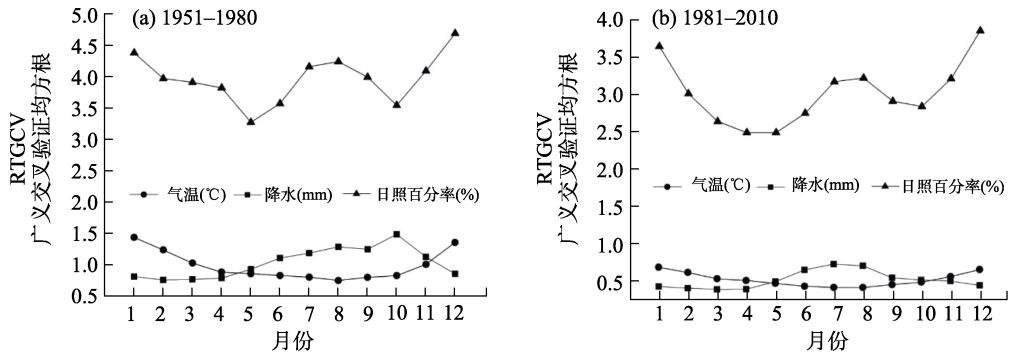


图 5 温度、降水和日照百分率的广义交叉验证均方根变化图

日照百分率受多种因素影响，其 RTGCV 值整体上在 2.5%–4.8% 范围内波动。气象数据的季节性变化规律会导致气候要素的插值误差也不可避免地增大，但就整体而言其数据质量是可控合理的。插值结果的平均绝对误差 MAE 和均方根误差 RMSE（表 2）也可以用于检验插值数据的精度。平均绝对误差反映估计值误差的大小，其值越接近 0 越好^[32]。从表 2 看出，冬春季节气温的平均绝对误差较大，而夏秋季节较小，但都相对较小，1951–1980 年位于 0.75–1.41 之间，而 1981–2010 年在 0.42–0.69 之间，说明气温插值数据是准确可靠的；降水则相反，冬春季节的平均绝对误差较小而夏秋较大，与不同季节降水量的多寡有关，但也都相对较小，变异系数较低，前 30 年平均为 28%，后 30 年为 13%；日照百分率平均绝对误差的季节特征则不明显，但大都在 2.5–4.6 之间，数据准确度也较好。均方根误差度量一组数据平均值的分散程度，标准差越小，数据越稳定。从表 2 看出，前 30 年气温的均方根误差平均值都在 0.06 以内，后 30 年在 0.04 以内，说明气温插值结果非常稳定；日照百分率的均方根误差平均都在 0.2 以内，说明数据也很稳定；而降水的均方根误差平均值在 0.33–0.68 之间，夏季月份超过了 1.0，表明数据存在一定的不稳定性，但总体平均值仍较稳定。因此，最终获得的两批插值气候数据精确度和稳定性是可以保证的。

由于历史原因，1951–1980 年的气象台站数量仅占 1981–2010 年的约 40%，后 30 年的台站数量明显增加，密度明显增大，但两个时间段的气象台站空间分布格局基本一致。比较二者的 MAE 和 RMSE（表 2）发现，后 30 年的两种误差值都明显小于前 30 年。其中，后 30 年的 12 个月气温的 MAE 均值仅为前 30 年的 54%，RMSE 只有 65%；降水误差均值分别为 49% 和 48%；日照百分率分别为 79% 和 94%。可见，台站的密度越大，插值误差越小，插值结果越好，尤其对气温和降水两个指标而言。

两个时间段气象台站观测数据与其对应的网格点插值结果对比表明，二者显著线性正相关（图 6）。无论哪个时间段，气温（图 6a、6b）和日照百分率（图 3e、3f）拟合的 R^2 均在 0.97–0.99 之间，说明二者的插值数据精度极高，而降水拟合的 R^2 分别为 0.93 和 0.98（图 6c、6d），其插值数据精度相对偏低，可能是因为 ANUSPLIN 在插值过程中低估了降水。但总体来看，三个气候要素的插值结果能解释真实结果的 97%，均能很好地反映真实气候情景。

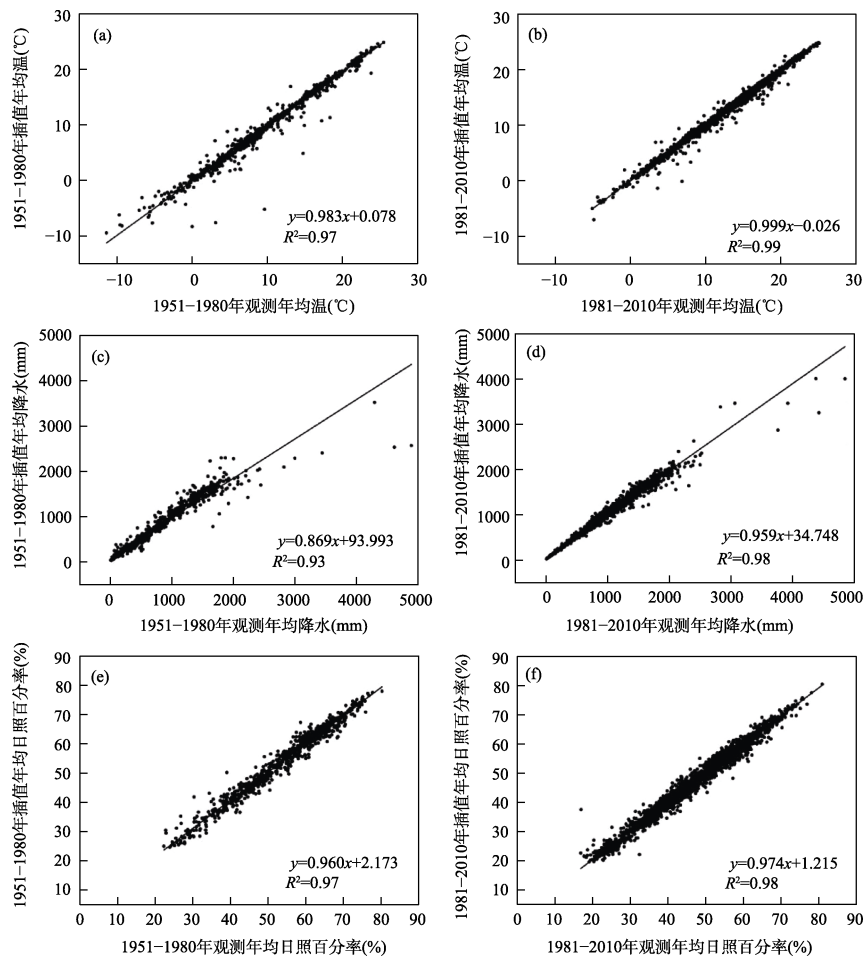


图6 气候观测数据与插值数据的对比验证图 (6a、6c、6e: 1951-1980; 6b、6d、6f: 1981-2010)

表2 中国气温、降水、日照百分率平均绝对误差(MAE)和均方根误差(RMSE)统计表

	气温 (°C)				降水 (mm)				日照百分率 (%)			
	1951-1980年		1981-2010年		1951-1980年		1981-2010年		1951-1980年		1981-2010年	
月	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	1.41	0.083	0.69	0.049	4.05	0.264	2.13	0.121	4.30	0.216	3.72	0.223
2	1.21	0.071	0.62	0.044	4.25	0.265	2.23	0.117	3.88	0.195	3.07	0.185
3	1.02	0.060	0.54	0.038	5.45	0.315	2.82	0.135	3.79	0.190	2.69	0.162
4	0.87	0.051	0.52	0.037	7.60	0.400	3.63	0.173	3.77	0.190	2.56	0.154
5	0.85	0.050	0.48	0.034	12.00	0.645	6.19	0.321	3.18	0.159	2.57	0.154
6	0.82	0.048	0.44	0.030	17.80	0.998	10.30	0.589	3.52	0.177	2.82	0.169
7	0.80	0.047	0.42	0.030	22.00	1.260	13.20	0.806	4.14	0.209	3.27	0.196
8	0.75	0.044	0.43	0.030	23.20	1.350	12.00	0.752	4.25	0.214	3.34	0.200
9	0.81	0.047	0.47	0.033	17.60	1.010	6.96	0.418	3.96	0.200	3.00	0.180
10	0.83	0.049	0.51	0.035	14.90	0.893	4.51	0.237	3.47	0.175	2.91	0.175
11	1.01	0.059	0.58	0.041	7.55	0.470	3.00	0.160	4.02	0.202	3.27	0.197
12	1.33	0.078	0.67	0.047	4.21	0.275	2.07	0.122	4.62	0.232	3.92	0.236
均值	0.98	0.057	0.53	0.037	11.72	0.679	5.75	0.329	3.91	0.197	3.10	0.186

5 讨论和总结

《中国 30 年平均 1-km 月度气候要素数据集 (1951–1980, 1981–2010)》利用中国地面气候标准值月值数据, 基于薄板光顺样条法和 ANUSPLIN 4.4 软件, 将气温、降水和日照百分率分别插值为 0.01° (≈ 1 km) 空间分辨率的网格数据。本文用插值方法——薄板光顺样条法, 是得到很好验证的国际通用的地理参照插值方法, 有效可靠。本数据集也采用 3 种途径进行了质量控制与评估: 以广义交叉验证法检验原始台站数据的有效性, 以平均绝对误差和均方根误差评估插值数据的精确性, 进一步以台站数据和插值数据的线性拟合再次验证其准确性。以上均表明, 本数据集是有效可靠的, 可用于宏观尺度的生态学和地学研究。

本数据集所用插值方法有一定优势, 0.01° 的空间分辨率较高, 由此获得的气候要素插值数据集是有效可用的, 可以作为从全球气候数据集中截取中国数据的一个有力替代品。除了本文涉及到的 3 个气候指标, 其他如地温、大气相对湿度、蒸发量、风速等指标亦可参照插值。本数据集可用于以下研究: 第一, 作为研究我国全国或者区域物种和生态系统与气候关系的基础数据, 由此统计分析其地理分布、动态变化及其气候驱动机制。第二, 驱动物种分布模型模拟中国植物和动物物种的潜在地理分布, 前期植物物种的模拟, 国内学者大都采用 WorldClim 气候数据集^[4], 本文数据集, 尤其是 1981–2010 年, 由于气象台站数量多, 其精度是有保障的, 而且是区域插值, 如果配合生态学意义更明晰的生物气候指标^[3,35], 其模拟效果应该更佳。第三, 驱动宏观尺度的陆地生物圈模型或其他植被模型, 模拟我国植被的潜在地理分布及其生产力和碳循环特征等。前期模拟常用 CRU 或者 WorldClim 数据, 目前可多一个选择, 如果配合时间序列的插值数据集^[36], 也可以进一步模拟植被的动态变化。第四, 还可以进一步应用于气候区划或者植被区划的本底数据, 应用于其他的生态学研究以及大气科学和地理科学研究中。我们相信, 本气候数据集, 以及其他利用类似方法和技术手段建立的气候数据集, 对于我国生态学和地球科学的宏观研究, 都将发挥重要作用。

致谢: 感谢国家气象科学数据中心和台湾省气象机构气象资料开放平台提供气象台站观测记录, 感谢陈鑫瑶协助进行气象台站数据处理。

作者分工: 程琦收集和处理基础数据并开展数据分析及数据论文撰写。吴星麒, 魏临风, 胡潇飞协助数据整理与分析。倪健设计总体研究方案和插值方法, 完成论文定稿。

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] 方精云. 我国森林植被带的生态气候学分析[J]. 生态学报, 1991, 11(4): 377–387.
- [2] 张新时. 研究全球变化的植被-气候分类系统[J]. 第四纪研究, 1993, 13(2): 157–169, 193.
- [3] 倪健. 全球变化研究中的生物气候指标[J]. 第四纪研究, 2017, 37(3): 431–441.
- [4] 刘晓彤, 袁泉, 倪健. 中国植物分布模拟研究现状[J]. 植物生态学报, 2019, 43(4): 273–283.
- [5] 冯建孟. 中国种子植物物种多样性的大尺度分布格局及其气候解释[J]. 生物多样性, 2008, 16(5): 470–476.
- [6] 朴世龙, 方精云, 贺金生等. 中国草地植被生物量及其空间分布格局[J]. 植物生态学报, 2004, 28(4): 491–498.
- [7] Piao, S. L., Fang, J. Y., Zhou, L. M., et al. Changes in vegetation net primary productivity from 1982 to 1999 in China [J]. *Global Biogeochemical Cycles*, 2005, 19(2): GB2027.

- [8] Liu, L., Yang, H., Xu, Y., *et al.* Forest biomass and net primary productivity in southwestern China: a meta-analysis focusing on environmental driving factors [J]. *Forests*, 2016, 7(12): 173.
- [9] 崔绍朋, 罗晓, 李春旺等. 基于 MaxEnt 模型预测白唇鹿的潜在分布区[J]. 生物多样性, 2018, 26(2): 171–176.
- [10] 张雪皎, 高贤明, 吉成均等. 中国北方 5 种栎属树木多度分布及其对未来气候变化的响应[J]. 植物生态学报, 2019, 43(9): 774–782.
- [11] Ni, J., Sykes, M. T., Prentice, I. C., *et al.* Modelling the vegetation of China using the process-based equilibrium terrestrial biosphere model BIOME3 [J]. *Global Ecology and Biogeography*, 2000, 9(6): 463–479.
- [12] Ni, J. Carbon storage in terrestrial ecosystems of China: estimates at different spatial resolutions and their responses to climate change [J]. *Climatic Change*, 2001, 49(3): 339–358.
- [13] Piao, S., Ciais, P., Huang, Y., *et al.* The impacts of climate change on water resources and agriculture in China [J]. *Nature*, 467(7311): 43–51.
- [14] 阎洪. 薄板光顺样条插值与中国气候空间模拟[J]. 地理科学, 2004, 24(2): 163–169.
- [15] Hijmans, R. J., Cameron, S. E., Parra, J. L., *et al.* Very high resolution interpolated climate surfaces for global land areas [J]. *International Journal of Climatology*, 2005, 25(15): 1965–1978.
- [16] 谭剑波, 李爱农, 雷光斌. 青藏高原东南缘气象要素 Anusplin 和 Cokriging 空间插值对比分析[J]. 高原气象, 2016, 35(4): 875–886.
- [17] Hancock, P. A., Hutchinson, M. F. Spatial interpolation of large climate data sets using bivariate thin plate smoothing splines [J]. *Environmental Modelling & Software*, 2006, 21(12): 1684–1694.
- [18] Xu, T. B., Hutchinson, M. F. New developments and applications in the ANUCLIM spatial climatic and bioclimatic modelling package [J]. *Environmental Modelling & Software*, 2013, 40: 267–279.
- [19] Fick, S. E., Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas [J]. *International Journal of Climatology*, 2017, 37(12): 4302–4315.
- [20] Harris, I., Osborn, T. J., Jones, P., *et al.* Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset [J]. *Scientific Data*, 2020, 7: 109.
- [21] 陈隆勋, 周秀骥, 李维亮等. 中国近 80 年来气候变化特征及其形成机制[J]. 气象学报, 2004, 62(5): 634–646.
- [22] 王军邦, 王居午, 叶辉等. 2000–2012 年全国气温和降水 1 km 网格空间插值数据集[J/OL]. 中国科学数据, 2017, 2(1): 73–80.
- [23] Peng, S. Z., Ding, Y. X., Liu, W. Z., *et al.* 1 km monthly temperature and precipitation dataset for China from 1901 to 2017 [J]. *Earth System Science Data*, 2019, 11(4): 1931–1946.
- [24] 湛稳, 孙立群, 李晴岚等. 近 38 年中国气温和降水的 1km 网格数据集[J]. 气象科技, 2021, 49(3): 355–361.
- [25] Niu, Z. G., He, H. L., Zhu, G. F., *et al.* A spatial-temporal continuous dataset of the transpiration to evapotranspiration ratio in China from 1981–2015 [J]. *Scientific Data*, 2020, 7: 369.
- [26] He, J., Yang, K., Tang, W., *et al.* The first high-resolution meteorological forcing dataset for land process studies over China [J]. *Scientific Data*, 2020, 7: 25.
- [27] 程琦, 吴星麒, 魏临风, 等. 中国 30 年平均 1-km 月度气候要素数据集(1951–1980, 1981–2010) [J/DB/OL]. 全球变化数据仓储电子杂志, 2022. <https://doi.org/10.3974/geodb.2022.06.03.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2022.06.03.V1>.
- [28] 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).
- [29] 中央气象局资料室. 中国地面气候资料集[M]. 北京: 气象出版社, 1984.
- [30] Hutchinson, M. F., Xu, T. B. ANUSPLIN Version 4.4 User Guide [M]. Canberra: Fenner School of Environment and Society, the Australian National University, 2013.
- [31] Farr, T. G., Rosen, P. A., Caro, E., *et al.* The shuttle radar topography mission [J]. *Reviews of Geophysics*, 2007, 45(2): RG2004.
- [32] 刘志红, Li, L. T., McVicar, T. R., 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.
- [33] 姜晓剑, 刘小军, 黄芬等. 逐日气象要素空间插值方法的比较[J]. 应用生态学报, 2010, 21(3): 624–630.
- [34] 中国科学院中国植被图编辑委员会. 中国植被图集(1:1 000 000) [M]. 北京: 科学出版社, 2001.
- [35] 魏临风, 胡潇飞, 程琦等. 中国 1km 分辨率生物气候要素空间分布数据集[J/OL]. 中国科学数据, 2022, DOI: 10.11922/11-6035.csd.2022.0003.zh.
- [36] 吴星麒, 程琦, 魏临风等. 中国西南喀斯特地区 1951–2014 年气候要素时间序列数据集[J/OL]. 中国科学数据, 2022, DOI: 10.11922/11-6035.csd.2022.0004.zh.