

基于 ETCCDI 的中国极端气温指数栅格数据集研发及评估

陈秋沅^{1,2}, 张宇^{1,2*}, 刘潇喻^{2,3}, 练秦来^{2,3}, 徐建军^{1,4}

1. 广东海洋大学, 南海海洋气象研究院, 湛江 524088;
2. 广东海洋大学, 海洋与气象学院, 近海海洋变化与灾害预警技术实验室, 湛江 524088;
3. 广东海洋大学, 陆架及深远海气候、资源与环境广东普通高校重点实验室, 湛江 524088;
4. 广东海洋大学, 深圳研究院, 深圳 518120

摘要: 全球增暖背景下, 极端天气和气候事件频发, 为了统一不同国家和地区对极端气候事件的定义, 世界气象组织 (WMO) 成立了气候变化检测和指数专家组 (ETCCDI), 给出了 26 个具有代表性的极端气温和降水指数, 用于规范全球极端气候事件的研究。本文利用中国国家地面气象站基本气象要素日值数据集 (V3.0), 计算了中国区域 16 个 ETCCDI 定义的极端气温指数, 包括最大日最高气温 (TXx)、最小日最高气温 (TXn)、最大日最低气温 (TNx)、最小日最低气温 (TNn)、暖昼百分比 (TX90p)、冷昼百分比 (TX10p)、暖夜百分比 (TN90p)、冷夜百分比 (TN10p)、夏日日数 (SU)、热夜日数 (TR)、冰冻日数 (ID)、霜冻日数 (FD)、持续暖日日数 (WSDI)、持续冷日日数 (CSDI)、气温日较差 (DTR) 和生长期长度 (GSL) 等。为方便不同学科领域研究者使用, 本文采用角距离加权插值方法, 将台站数据插值成空间分辨率为 $0.25^{\circ} \times 0.25^{\circ}$ 的经纬网格数据, 存储格式为 NetCDF, 时间跨度为 1961–2020 年, 时间分辨率为逐年或逐月, 包含 16 个年数据文件和 13 个月数据文件, 数据总量为 1.43 GB, 压缩后为 353.0 MB。该数据集命名为 SimmEX_1961-2020_1.0, 可为各类极端气温事件的研究提供数据支撑, 同时, 在环境、经济、能源等领域也有着广泛的应用前景。

关键词: 气候变化; ETCCDI; 极端气候指数; SimmEX

DOI: <https://doi.org/10.3974/geodp.2024.01.08>

CSTR: <https://cstr.science.org.cn/CSTR:20146.14.2024.01.08>

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.2024.05.06.V1> 或 <https://cstr.science.org.cn/CSTR:20146.11.2024.05.06.V1>.

1 前言

IPCC AR6^[1]表明在全球变暖的大背景下, 全球极端热事件、强降水、干旱等各种极端

收稿日期: 2023-12-06; 修订日期: 2024-03-01; 出版日期: 2024-03-25

基金项目: 国家自然科学基金 (72293604, 42130605); 深圳市科技创新局 (JCYJ20210324131810029)

*通讯作者: 张宇, 广东海洋大学, zhangyu@gdou.edu.cn

数据引用方式: [1] 陈秋沅, 张宇, 刘潇喻等. 基于 ETCCDI 的中国极端气温指数栅格数据集研发及评估[J]. 全球变化数据学报, 2024, 8(1): 67–75. <https://doi.org/10.3974/geodp.2024.01.08>.
<https://cstr.science.org.cn/CSTR:20146.14.2024.01.08>.

[2] 陈秋沅, 张宇, 刘潇喻等. 中国极端气温指数栅格数据集 (1961–2020) (V1.0) [J/DB/OL]. 全球变化数据仓储电子杂志, 2024. <https://doi.org/10.3974/geodb.2024.05.06.V1>.
<https://cstr.science.org.cn/CSTR:20146.11.2024.05.06.V1>.

天气气候事件的频率和强度均在增加,这对人类社会发[2]和生态系统平衡[3]产生了严重影响。清楚认识这些极端事件的特征,对其做出科学应对,对防灾减灾具有重要意义。而高时空覆盖率和分辨率的极端气候指数数据集,是开展上述研究的基础。尽管已有不少学者和机构制作了一些极端气候指数数据集,例如王晓利[4]等计算了中国沿海地区的极端气温指数,马伟东[5]等、周旗[6]等分别制作了青藏高原和渭河流域的极端降水数据集,这些数据集在局地或区域研究中发挥了重要价值,但若想开展整个中国区域的研究,还需将数据集的空间覆盖范围进一步扩展。美国航天航空局和英国气象局的哈德莱中心等,也分别推出过自身的全球极端气候指数产品 M2SMNXEDI 和 HadEX,虽然覆盖了整个中国区域,但这些极端指数产品在制作时,使用的原始观测台站较少(400~500个站),且产品的空间分辨率较低($1.25^{\circ}\times1.875^{\circ}$) [7],这些均限制了资料集的可靠性和适用性。

为了得到高时空分辨率的整个中国区域极端气温指数数据集,本文利用中国气象局国家气象信息中心提供的中国 2,481 个地面气象站的逐日日最高(低)气温和日平均气温资料,利用 Python 程序语言计算了由气候变化检测和指数专家组(Expert Team on Climate Change Detection and Indices, ETCCDI)规定的 16 个极端气温指数(见表 1),并采用角距离加权插值法[7](Angular Distance Weighting Interpolation, ADW)生成了 1961~2020 年中国极端气温指数栅格数据集。该数据集由广东海洋大学南海海洋气象研究院(GDOU-SIMM)极端气候研究小组制作,命名为 SimmEX_1961-2020_1.0,下文简写为 SimmEX。该数据集可为研究整个中国区域或更小时空尺度的极端气温事件提供基础数据支撑,这将有助于提高人们对极端气温事件的认识。同时,其对环境、经济、能源等领域开展气候建模及科学应对气候变化也有重要应用价值。

表 1 16 个极端气温指数统计表

指数代码	指数名称	定义	单位
TXx	最大日最高气温	年内/月内日最高气温的最大值	℃
TXn	最小日最高气温	年内/月内日最高气温的最小值	℃
TNx	最大日最低气温	年内/月内日最低气温的最大值	℃
TNn	最小日最低气温	年内/月内日最低气温的最小值	℃
TX90p	暖昼百分比	日最高气温在年内/月内>90%分位数的天数所占百分比	%
TX10p	冷昼百分比	日最高气温在年内/月内<10%分位数的天数所占百分比	%
TN90p	暖夜百分比	日最低气温在年内/月内>90%分位数的天数所占百分比	%
TN10p	冷夜百分比	日最低气温在年内/月内<10%分位数的天数所占百分比	%
SU	夏日日数	年内/月内日最高气温>25℃的天数	天
TR	热夜日数	年内/月内日最低气温>20℃的天数	天
ID	冰冻日数	年内/月内日最高气温<0℃的天数	天
FD	霜冻日数	年内/月内日最低气温<0℃的天数	天
WSDI	持续暖日日数	年内至少连续 6 天日最高气温>90%分位数的天数	天
CSDI	持续冷日日数	年内至少连续 6 天日最低气温<10%分位数的天数	天
DTR	气温日较差	年内/月内平均最高气温与年内/月内平均最低气温之差	℃
GSL	生长期长度	北半球 1 月 1 日起,从第一次出现至少连续 6 天日平均气温>5℃的第一天起,至 7 月 1 日后第一次出现至少连续 6 天日平均气温<5℃的第一天止的时间跨度	天

2 数据集元数据简介

《中国极端气温指数栅格数据集（1961–2020）（V1.0）》^[8]的名称、作者、地理区域、数据年代、时间分辨率、空间分辨率、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 2。

表 2 《中国极端气温指数栅格数据集（1961–2020）（V1.0）》元数据简表

条 目	描 述
数据集名称	中国极端气温指数栅格数据集（1961–2020）（V1.0）
数据集短名	SimmEX_1961-2020_1.0
作者信息	陈秋沅，广东海洋大学南海海洋气象研究院，chenqiuyuan@stu.gdou.edu.cn 张宇，广东海洋大学南海海洋气象研究院，zhangyu@gdou.edu.cn 刘潇喻，广东海洋大学南海海洋气象研究院，liuxiaoyu@stu.gdou.edu.cn 练秦来，广东海洋大学南海海洋气象研究院，lianqinlai@stu.gdou.edu.cn 徐建军，广东海洋大学南海海洋气象研究院，jxu@gdou.edu.cn
地理区域	中国（未包括领海）
数据年代	1961–2020 年
时间分辨率	逐年、逐月
空间分辨率	0.25°×0.25°
数据格式	. NetCDF
数据量	1.43 GB（压缩后约 353.0MB）
数据集组成	共 29 个 .NetCDF 文件，其中 16 个年文件，13 个月文件
基金项目	国家自然科学基金（72293604，42130605） 深圳市科技创新局（JCYJ20210324131810029）
数据计算环境	Python 3.11
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	（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 ^[9]
数据和论文检索系统	DOI，CSTR，Crossref，DCI，CSCD，CNKI，SciEngine，WDS/ISC，GEOSS

3 数据研发方法

3.1 数据来源

研发本数据集的基础气象资料来自于中国气象局国家气象信息中心制作的中国国家级地面气象站基本气象要素日值数据集（V3.0），其中包括全国 2,481 个国家级地面观测站的逐日日最高气温、日最低气温和平均气温数据。该数据集的建立和制作过程经过了严格的质量控制，对存在的一些错误已进行了更正，有较高的可信度。台站具体分布见图 1。

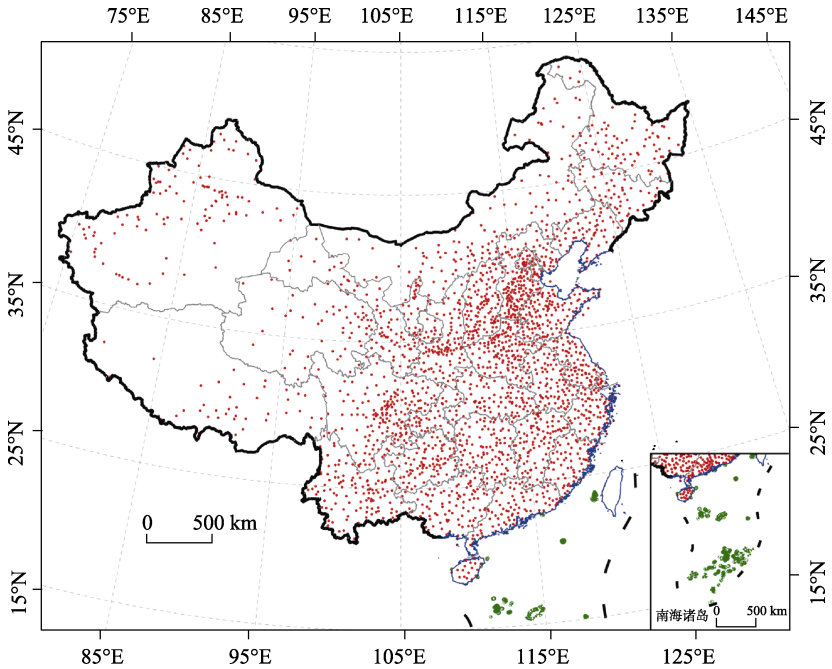


图 1 原始数据气象台站分布图

(注：该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(2022)1873 号的标准地图制作，底图无修改)

3.2 插值方法

角距离加权插值方法是一种将离散的不规则站点数据通过一定的数学关系插值至一定空间格点上的方法。这种方法综合考虑了站点与规则网格点之间的距离和方位角关系，能够实现较好的插值效果，如 Dunn^[10]用此方法制作了 HadEX3 数据集。具体算法如下：公式(1)指以格点为中心，格点周围与此格点上的指数数值大小的相关性呈 e 指数型衰减，其中 x 为台站与格点之间的距离， x_0 为搜索半径；公式(2)中， w_i 为各个台站的距离权重， m 用于调整衰减速率；公式(3)综合考虑了方位角和距离，得到 W_i 即角距离权重，其中 θ 方位角为站点相对于格点的方位角， k 为搜索范围内台站数量。1961 年以来，绝大部分气象站点由于下垫面和周围环境的变化而经历了台站迁址，特别是 1961–1980 年这 20 年来，台站数量也在迅速增加。因此对于各台站对各格点的权重计算，我们分为五个时期，1961–1965 年、1966–1970 年、1971–1975 年、1976–1980 年和 1981–2020 年，如此计算既考虑了权重变化，也使得插值计算高效。本文主要运用 Python 程序语言实现公式(1)–(3)的计算。首先通过自定义函数实现站点与格点之间方位角和距离的计算，即得到公式(1)中的 x 和公式(3)中的 θ ，再按顺序编写公式(1)–(3)，从而实现角距离加权插值方法。

$$r = e^{-\frac{x}{x_0}} \tag{1}$$

$$w_i = r^m \tag{2}$$

$$W_i = w_i \left\{ \frac{\sum_k w_i [1 - \cos(\theta_k - \theta_i)]}{\sum_k w_i} \right\}, \quad k \neq i \quad (3)$$

3.3 技术路线

中国极端气温指数栅格数据集的建立流程如图 2，主要包括基础数据集的预处理，台站 16 个指数的计算，插值至经纬度网格点和输出文件四个部分。

(1) 首先利用 Python 对中国国家级地面气象站基本气象要素日值数据集 (V3.0) 进行数据预处理，包括站点经纬度定位和数据单位转换等，得到 1961–2020 年中国区域 2,481 个台站的逐日气温数据。

(2) 16 个极端气温指数中，TX90p、TN90p、TX10p、TN10p、WSDI 和 CSDI 共 6 个指数的计算需要首先确定其相对阈值，即采用百分位阈值法计算其在 1961–1990 年间的气候态阈值，同时剔除掉缺测大于 5% 的台站。然后便可以利用 Python 实现 16 个极端气温指数的计算，得到台站的 16 个极端气温指数计算结果。考虑到指数的具体定义，其中 GSL、WSDI 和 CSDI 这 3 个指数的时间分辨率仅为逐年。

(3) 采用角距离加权插值法将台站计算结果插值至经纬度网格点上，得到格点上的插值计算结果。

(4) 根据中国地理范围进行数据掩膜，只保留中国区域（未包括领海）的数据，其余区域赋值为缺测。利用 Python 给每个文件录入数据集信息（数据集名称、制作者、时空分辨率等）、指数信息（指数定义、单位等），规定数据精度和数据缺失标记等。最后输出保存文件，生成 SimmEX 数据集。

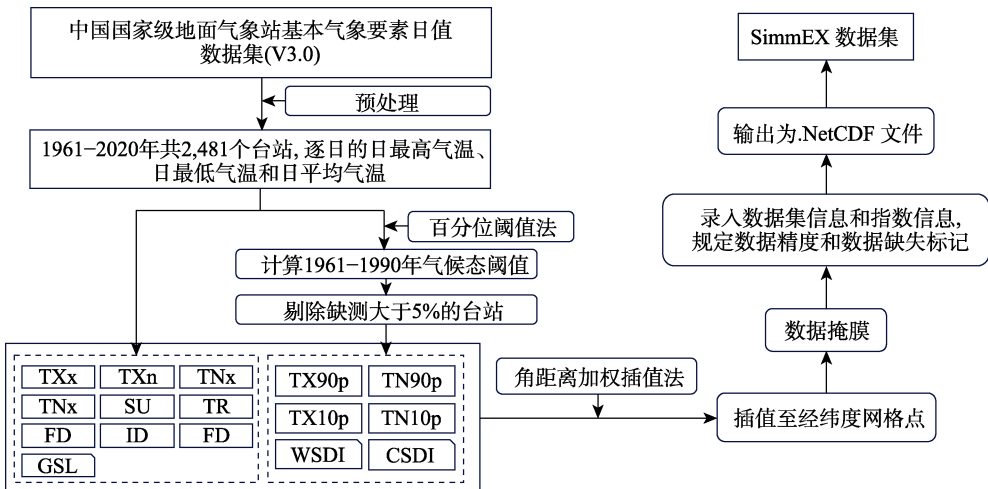


图 2 SimmEX 数据集制作技术路线流程图

4 数据结果与验证

4.1 数据集组成

压缩包内包含 ANN 和 MON 两个文件夹，分别存放逐年指数 16 个和逐月指数 13 个，

每个文件均为.NetCDF 格式，命名法则为 SimmEX_xx_ANN 或 SimmEX_xx_MON，其中 xx 为指数代码，参见表 3。

表 3 数据集文件表

文件夹名称	文件名称	文件说明	文件个数	文件格式	单个文件大小
ANN	SimmEX_××_ANN	逐年的××指数	16	.NetCDF	约 8.57 MB
MON	SimmEX_××_MON	逐月的××指数	13	.NetCDF	约 102.72 MB

4.2 数据结果及验证

图 3 以 TX_x、TX90p、SU 和 WSDI 等 4 个指数为例，给出了 SimmEX 数据集与英国哈德莱中心的 HadEX3 数据集的对比结果，其中 SimmEX 的空间分辨率为 0.25°×0.25°，而 HadEX3 为 1.25°×1.875°。由这 4 个指数的气候态空间分布可见，两套数据集的 TX_x（图 3a、3b）在空间上均呈现为青海、西藏及四川西部为低值区域，其余地区为高值区域，SimmEX 数据集中 TX90p（图 3c）比 HadEX3 数据集中 TX90p（图 3d）整体偏小，SU（图 3e、3f）在空间上均呈现为华南低、华南以北高的空间分布特点，WSDI（图 3g、3h）在空间上均呈现为东北部低、西部高的空间分布特点。需要特别指出的是，SimmEX 数据集的空间分辨率更高，因此可以体现出更精细化的分布特征，例如 SimmEX 的 WSDI 指数在新疆、西藏、四川、华东等较小范围区域可呈现出一些高值中心，而较低分辨率的 HadEX3 数据集则无法反映出这种小尺度的精细化结构特征。

图 4 给出了 SimmEX 和 HadEX3 两套数据集，整个中国区域平均的 TX_x、TX90p、SU 和 WSDI 等 4 个指数的年际变化。通过对比不难发现，两套数据集的 4 个指数在年际变化方面，高低值年份及变化趋势均比较对应具有较好的一致性，尤其是 TX_x、TX90p 和 WSDI 指数（图 4a、4b 和 4d），两套资料的差别很小。而两套资料的 SU 指数（图 4c）差异最大，SimmEX 整体低于 HadEX3，但年际变化非常一致。此外，两套资料的 TN_x、TN_n、SU、TR、DTR、GSL 等指数也或高或低，存在一定的差异，在这里不一一展示。总体而言，两套资料的各个指数在年际变化方面基本一致，但不同指数存在不同程度的整体性偏差。SimmEX 与 HadEX3 虽然采用相同的定义来计算，且都使用了相同的空间插值方法，但 SimmEX 使用的原始台站数据更多，空间覆盖更密集，这是两套指数集产品产生差异的主要原因。

图 3 和图 4 已对比分析了 SimmEX 和 HadEX3 两套数据集各指数的空间分布和年际变化差异。下面我们将借助统计量，对两套数据集的质量进行科学评估。具体做法如下：将两套数据集各个指数插值至 2,481 个台站上，采用均方根误差（Root Mean Square Error, RMSE）作为客观评估标准，见公式(4)， h_i 为将数据集中数据插值至台站后的值， o_i 为台站观测值， n 为文件个数。

$$RMSE = \sqrt{\sum_{i=1}^n \frac{1}{n} (h_i - o_i)^2}$$

(4)

图 5a 和 5b 展示了 ANN 文件夹中年指数资料的评估结果，由图可见 SimmEX 中 TX_x、TX_n、TN_x、TN_n 等 4 个指数的 RMSE 平均水平较 HadEX3 约低 2℃；SimmEX 中 TX90p、TX10p、TN90p、TN10p 等 4 个指数的 RMSE 平均水平较 HadEX3 约低 2%；SimmEX 中 ID、FD、SU、TR、GSL 等 5 个指数的 RMSE 平均水平较 HadEX3 低 10–25 天；SimmEX 中

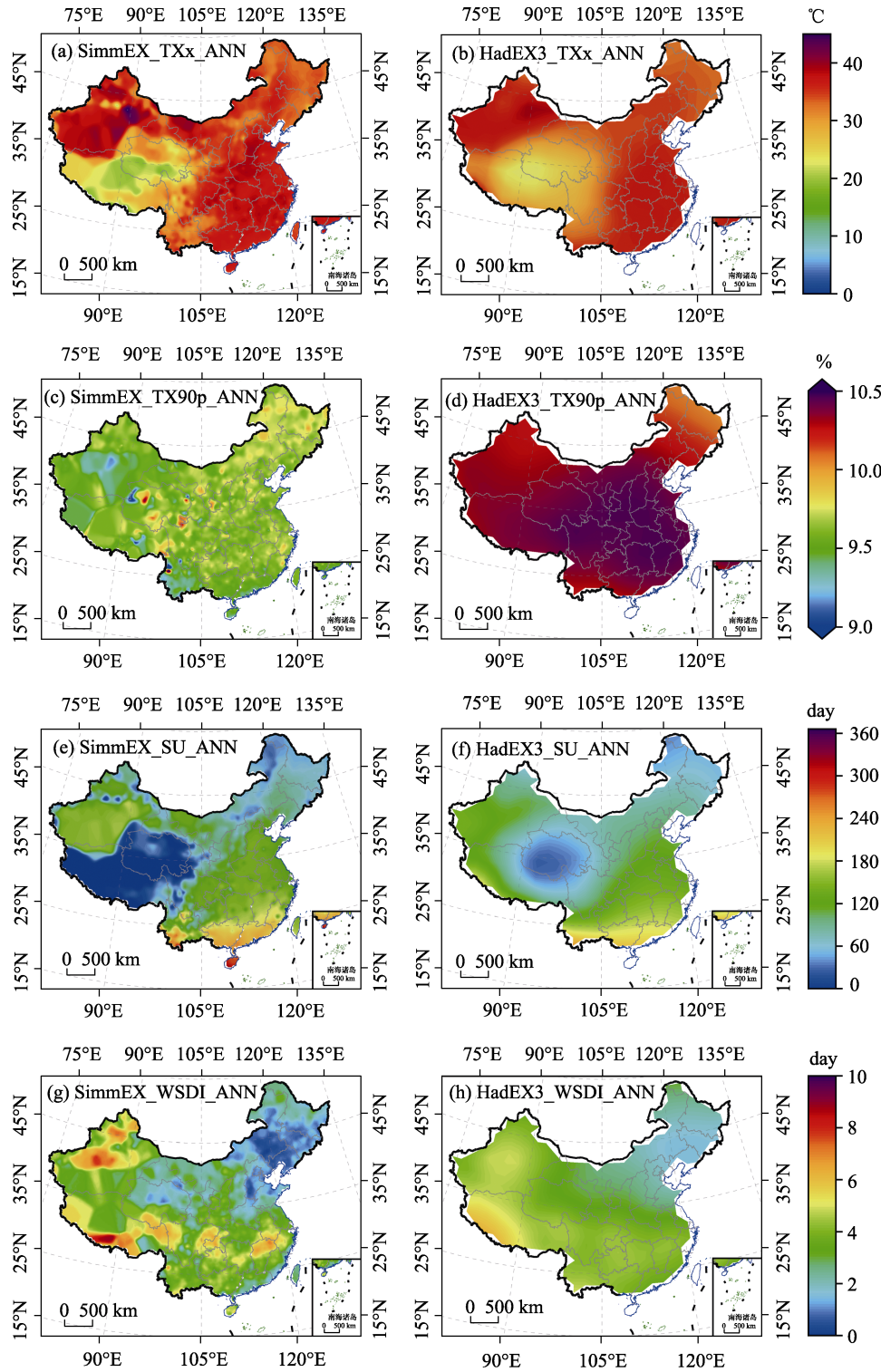


图3 SimmEX 与 HadEX3 的 TXx、TX90p、SU 和 WSDI 指数气候态空间分布图 (1961–1990 年)
(注：该图基于国家测绘地理信息局标准地图服务网站下载的审图号为 GS(2022)1873 号的标准地图制作，底图无修改)

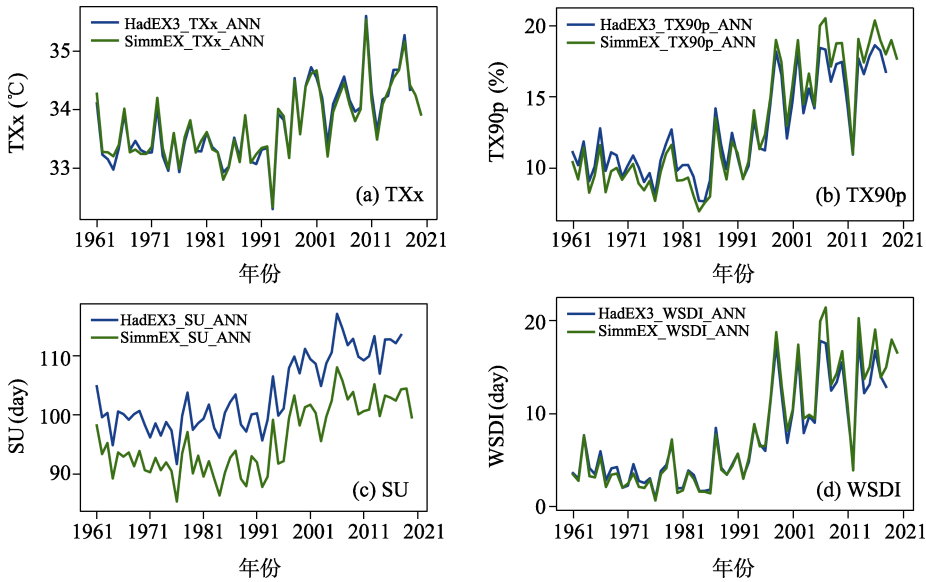


图 4 SimmEX 与 HadEX3 的 TXx (a)、TX90p (b)、SU (c) 和 WSDI (d) 指数年际变化图

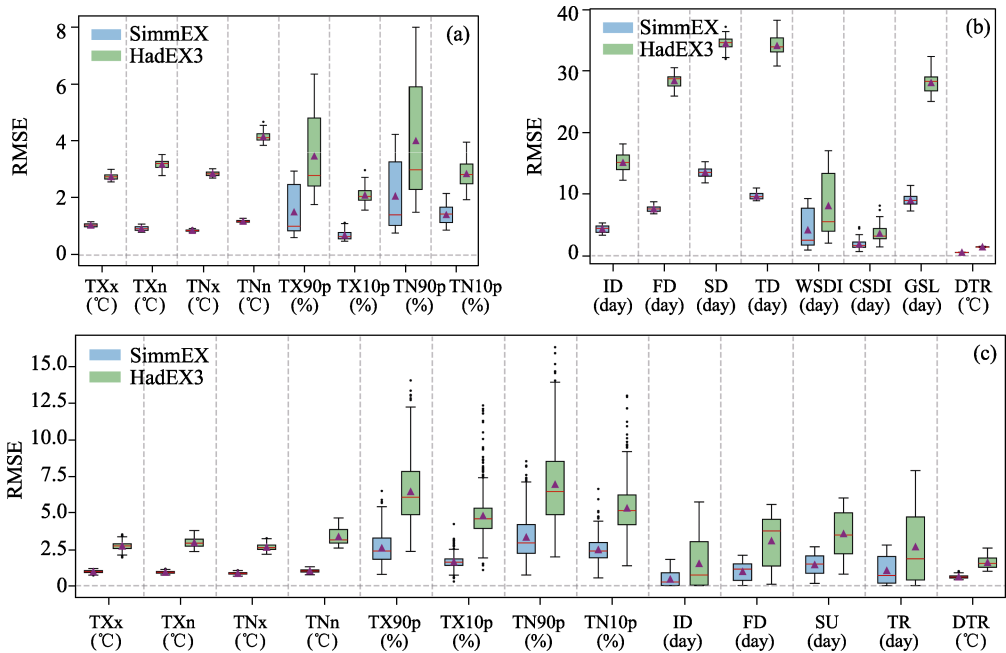


图 5 SimmEX、HadEX3 中各指数的均方根误差对比图：(a)、(b) 为年文件，(c) 为月文件

WSDI、CSDI 等 2 个指数的 RMSE 平均水平较 HadEX3 低 1–2 天。并且 SimmEX 中各个指数的 RMSE 波动程度小于 HadEX3，其与 HadEX3 相比更佳稳定。

图 5c 展示了 MON 文件夹中月指数资料的评估结果，由图可见 SimmEX 中 TXx、TXn、TNx、TNn 等 4 个指数的 RMSE 的平均水平较 HadEX3 约低 2 °C；SimmEX 中 TX90p、TX10p、TN90p、TN10p 等 4 个指数的 RMSE 平均水平较 HadEX3 低 2%–4%；SimmEX 中 FD、SU、TR 等 3 个指数的 RMSE 平均水平较 HadEX3 低 1–2 天。并且 SimmEX 中各个指数的

RMSE 的波动程度小于 HadEX3。由此可见, SimmEX 数据集中的各个指数均优于 HadEX3 数据集。

5 讨论和总结

本研究利用中国国家级地面气象站基本气象要素日值数据集(V3.0), 采用 Python 程序语言, 根据气候变化检测和指数专家组(ETCCDI)给出的国际标准定义, 计算了全国 2,481 个台站的 16 个极端气温指数, 并采用角距离加权插值方法将站点数据插值至经纬网格点上, 生成了格点化的 SimmEX 极端气温指数数据集。通过评估发现, 该数据集与 HadEX3 相比, 数据集质量有了很大程度的提升。但在使用本数据集时, 需要留意以下问题: 由于中国西部地区台站数量较少, 其分布较东部地区更为稀疏, 因此东部地区的可信度要高于西部。尤其是台站稀疏、地形复杂的青藏高原西部, 该区域的数据是外插得来, 可靠性有待考量, 建议酌情使用。未来我们将在台站数据的基础上, 融合卫星遥感数据、再分析数据等多源数据作为基础数据, 以弥补本数据集在中国西部等台站稀疏地区质量欠佳的问题。

作者分工: 陈秋沅完成了数据集的计算和处理工作, 并撰写了论文; 张宇提供了数据集开发的整体思路, 并对论文进行了审核修定; 刘潇喻编写了角距离加权插值方法的 Python 脚本, 并校正了论文; 练秦来帮助改进了计算脚本, 提高了指数计算效率; 徐建军提供了数据集的建设思路。

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] IPCC. Climate Change 2021: the Physical Science Basis [M/OL]. Cambridge: Cambridge University Press, 2021 [2024-05-08]. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf.
- [2] 吴绍洪, 尹云鹤. 极端事件对人类系统的影响[J]. 气候变化研究进展, 2012, 8(2): 99–102.
- [3] 朴世龙, 张新平, 陈安平. 极端气候事件对陆地生态系统碳循环的影响[J]. 中国科学: 地球科学, 2019(49): 1321–1334.
- [4] 王晓利, 侯西勇. 中国沿海省区极端气温栅格数据集[J]. 全球变化数据学报, 2019, 3(1): 54–58.
- [5] 马伟东, 刘峰贵, 周强. 青藏高原极端降水数据集(1961–2017)的研发与内容[J]. 全球变化数据学报, 2021, 5(1): 67–72.
- [6] 周旗, 张海宁, 任源鑫. 渭河流域极端降水指数数据集(1961–2016)[J]. 全球变化数据学报, 2021, 5(1): 62–66.
- [7] Caesar, J., Alexander, L., Vose, R. Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set [J]. *Journal of Geophysical Research: Atmospheres*, 2006, 111(D5): D05101.
- [8] 陈秋沅, 张宇, 刘潇喻. 中国极端气温指数栅格数据集(1961–2020)(V1.0)[J/DB/OL]. 全球变化数据仓储电子杂志, 2024. <https://doi.org/10.3974/geodb.2024.05.06.V1>.
- [9] 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).
- [10] Dunn, R. J. H., Alexander, L. V., Donat, M. G., et al. Development of an updated global land in situ-based dataset of temperature and precipitation extremes: HadEX3 [J]. *Journal of Geophysical Research: Atmospheres*, 2020, 125(16): e2019JD032263.