

何莉阳,苏传程,覃卫坚,等. 三套再分析气温资料在广西的适用性评估[J]. 气象研究与应用,2022,43(3):08-13.

He Liyang,Su Chuancheng,Qin Weijian,et al. Applicability evaluation of surface air temperature from three reanalysis datasets in Guangxi Province[J]. Journal of Meteorological Research and Application,2022,43(3):08-13.

三套再分析气温资料在广西的适用性评估

何莉阳¹, 苏传程², 覃卫坚^{1*}, 邓悦², 赖晟¹, 朱秋宇¹

(1.广西壮族自治区气候中心, 南宁 530022; 2.广西壮族自治区气象信息中心, 南宁 530022)

摘要:以 2020 年广西 91 个国家地面气象观测站 2m 气温资料作为基准, 对中国第一代陆面再分析资料 CMA-RA/Land 及目前广泛使用的 ERA5/Land、CFSR 等 3 套再分析气温资料在广西的适用性进行初步对比评估。结果表明, 3 套资料均能重现气温时空分布特征, 但存在不同程度的冷偏差, CMA-RA/Land 与观测值最接近, ERA5/Land 在桂北冷偏差较明显, CFSR 整体冷偏差较大。CMA-RA/Land 在站点上的各项评估指标均表现良好, 相关系数高, 均方根误差、绝对误差、相对误差均较小。总体来看, CMA-RA/Land 在广西适用性最好。

关键词:再分析资料; 地面气温; 观测资料; 线性拟合; 适应性

中图分类号: P468

文献标识码: A

doi: 10.19849/j.cnki.CN45-1356/P.2022.3.02

引言

再分析资料是利用先进的数据同化系统及数值预报模式, 将多源气象观测数据和模式预报融合, 从而获取一套高分辨率、长时间序列的格点化多要素气象数据集。再分析资料具备空间覆盖完整, 时间均一稳定等优点, 自 20 世纪 90 年代出现以来迅速成为气象研究的主要资料来源之一^[1-2]。目前全球应用较为广泛的再分析产品包括美国国家环境预测中心(NCEP)的 CFSR^[3]、美国航空航天局(NASA)的 MERRA^[4]、日本气象厅(JMA)的 JRA-55^[5]、欧洲中期天气预报中心(ECMWF)的 ERA-Interim 及 ERA5^[6-7]。数值模式、同化方案的差异以及观测资料的不确定性(站点迁移、观测系统变更、观测仪器更换)^[8]使得再分析资料在实际应用中容易出现系统性偏差, 因此对再分析资料进行适用性评估非常重要。

大量学者对再分析资料开展评估研究工作^[9-13]。赵天保、谢潇等^[14-15]指出中国区域内再分析资料可信度在东部地区优于西部地区, 低纬地区优于高纬

地区, 再分析资料各个要素在不同区域、不同时段的应用性不同。针对再分析温度资料, 单帅等^[16]对再分析气温资料在中国北部积雪区适用性进行评估, 指出再分析资料在东北地区偏差小, 北疆地区偏差稍大, ERA5 整体表现好于 JRA-55、CFSR 及 MERRA2。孟宪贵等^[17]的评估表明在山东省 ERA5 气温质量优于 ERA-Interim。朱景等^[18]指出在浙江省 ERA5 的地表温度、气温资料较 ERA-Interim 有较大提高。殷悦等^[19]指出 CFSR 气温资料在浙江省存在较大冷偏差。丁敏等^[20]指出 JRA-55 气温资料在羌塘自然保护区适用性好于 ERA-Interim、NCEP/NCAR 及 NCEP/DOE 等。综上所述, 不同再分析气温资料在不同地区效果各异。

2019 年, 中国第一代全球大气/陆面再分析产品(CMA-RA、CMA-RA/Land)发布, 采用 GSI-3DVar 同化系统, 空间分辨率为 34km, 时间分辨率为 6h(大气再分析产品)和 3h(陆面再分析产品)^[21]。相较于其他再分析资料, CMA-RA 同化了更多中国基础气象数据, 包括中国地面站、常规探空观测资料等^[22]。CMA-RA 作为中国自主研发的再分析资料,

收稿日期: 2022-04-10

基金项目: 广西科技计划项目(桂科 AB21075005)

作者简介: 何莉阳(1994—), 女, 硕士, 助理工程师, 主要从事气候变化研究。E-mail: 717080438@qq.com

* 通讯作者: 覃卫坚(1971—), 男, 博士, 正研级高级工程师, 主要从事短期气候预测方法研究。E-mail: qinweijian2008@126.com

保证了数据的自主性和可控性, 资料质量对标国际第三代再分析资料水平, 对促进气象业务和气候预测研究业务的发展具有重要且深远的意义。ERA5 作为欧洲最新一代再分析产品, 采用集合四维变分混合同化技术, 代表当前再分析产品的最高水平。CFSR 作为第三代再分析资料, 采用 3DVar 同化技术, 同样在气象业务中应用广泛。本文将对 CMA-RA/Land、ERA5/Land 及 CFSR 的气温资料在广西的适用性进行横向评估, 对比资料质量, 为高分辨率再分析资料在广西天气气候监测等业务应用上提供参考依据。

1 资料与方法

1.1 观测资料

本文所用观测资料为 2020 年广西 91 个国家观测站逐小时 2m 气温, 资料来自广西壮族自治区气象信息中心大数据云平台, 数据经过初步质控, 并将逐小时资料按业务规范处理为逐日资料, 资料时间为世界时。

1.2 再分析资料

CMA-RA/Land 再分析 2m 气温资料来自中国气象数据网 (<https://data.cma.cn/CRA>), 时间分辨率为 3h, 空间分辨率 34km×34km (0.3°×0.3°); ERA5/Land 再分析 2m 气温资料下载自欧洲中期天气预报中心 (<https://www.ecmwf.int/>), 时间分辨率 1h, 空间分辨率 0.25°×0.25°; CFSR 再分析 2m 气温资料下载自美国环境预测中心 (<https://rda.ucar.edu/>), 时间分辨率为 6h, 空间分辨率 0.25°×0.25°。分别计算 3 套再分析资料当日各个时次的算术平均值, 处理为逐日资料, 时间为世界时。

1.3 评估指标

采用差值、线性拟合等方法对再分析气温资料在广西适用性进行评估, 并采用相关系数(COR), 均

方根误差(RMSE)及平均绝对误差(MAE)、平均相对误差(ME)等评价指标^[23-24]对广西 91 个国家观测站分别进行评估。评价指标计算公式如下:

相关系数:

$$\text{COR} = \frac{\sum_{i=1}^N (G_i - \bar{G})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (G_i - \bar{G})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}} \quad (1)$$

均方根误差:

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

平均绝对误差:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |G_i - O_i| \quad (3)$$

平均相对误差:

$$\text{ME} = \frac{1}{N} \sum_{i=1}^N (G_i - O_i) \quad (4)$$

其中, O_i 为观测值, G_i 为再分析资料插值至站点的值, N 为参与检验的总样本数。

2 结果与分析

2.1 区域评估

使用最近邻插值法与双线性插值法 2 种插值方法, 将格点资料插值成站点资料。经对比, 双线性插值法效果更优, 因此文中选用双线性插值法。将 3 套再分析资料分别插值到 91 个站点, 求取广西全区平均, 由图 1 可知, CMA-RA/Land 线性拟合度最优 ($r^2=0.9975$), ERA5/Land ($r^2=0.9877$) 和 CFSR ($r^2=0.9874$) 线性拟合度次之。CMA-RA/Land、ERA5/Land 及 CFSR 的线性回归系数分别均小于 1, 表明再分析资料倾向于低估气温值。再分析气温资料分布区间与观测值基本一致, 不过 3 套再分析资料均对气温在 27℃ 以上的日数存在低估, 2020 年广西全区日平均气温超过 30℃ 的日数有 7d, 3 套再分析资料均没有气温达到 30℃ 的日数。

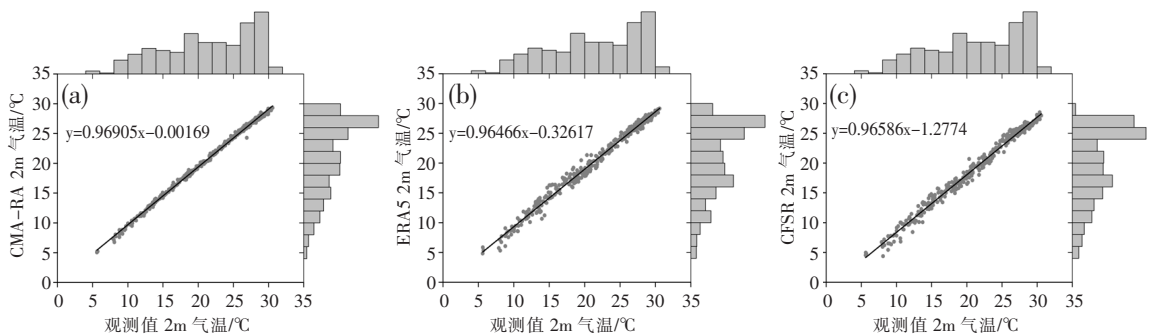


图 1 2020 年广西 CMA-RA/Land(a)、ERA5/Land(b)、CFSR(c)逐日 2m 气温再分析资料与观测值的散点图及分布区间柱状图(再分析资料(右侧), 观测值(上侧), 单位: d)

从 2020 年广西全区平均 2m 气温逐日时间序列图可知,3 套再分析资料均能较好重现气温的年内变化趋势,在气温波动较大的春季和冬季,再分析资料变化趋势始终与观测资料一致(图 2a)。3 套再分析气温资料较观测值偏低(图 2b),年平均误差(ME)小于 0,整体呈冷偏差。表 1 中各项评估指标表明 CMA-RA/Land 气温资料整体表现良好,CRA-MA/Land 相关系数(COR)达到 0.998,且误差波动范围小,能较好反映广西全区气温变化趋势,不

过在夏季 CMA-RA/Land 气温资料冷偏差较其他季节略偏大;ERA5/Land 效果次之,相关系数 0.993,3 月至 4 月中旬气温较观测值偏高,其余月份偏低,秋季、冬季误差范围增加;CFSR 相关系数虽然也较高,达到 0.993,但误差偏大,逐日气温比观测值偏低约 1~3℃,部分日数偏低 3.5℃以上,CFSR 均方根误差(RMSE)、年平均绝对误差(MAE)均大于 CMA-RA/Land 和 ERA5/Land 的相关误差。

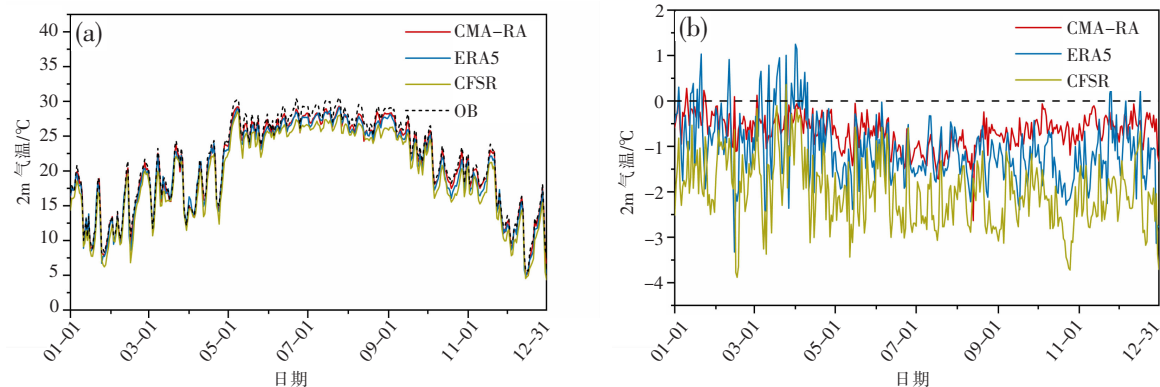


图 2 2020 年广西 CMA-RA/Land、ERA5/Land、CFSR 2m 气温再分析资料与观测值(OB)逐日时间序列(a)及误差(b)

表 1 2020 年广西 2m 温度观测值与 CMA-RA、ERA5、CFSR 再分析资料的相关系数(COR)、均方根误差(RMSE)、年平均绝对误差(MAE)、年平均相对误差(ME)

资料	COR	RMSE	MAE	ME
CMA-RA	0.998*	1.312	1.032	-0.661
ERA5	0.993*	1.622	1.297	-0.923
CFSR	0.993*	2.517	2.146	-1.998

注:*表示通过 99%显著性检验

分布特征的刻画效果好于 CFSR。在地势相对平缓的沿海及桂中地区,再分析气温分布与观测资料的一致性较好,但在地形复杂、高山多的桂北地区,3 套资料气温分布较实际观测值出现不同程度低估,CFSR 最为严重。

就广西区域平均而言,各项评估指标中 CMA-RA/Land 效果最好,其对气温变化趋势的刻画,气温值准确度及气温分布特征都更贴近观测数据。ERA5/Land 对广西气温的模拟质量也比较好的,能清晰反映广西气温变化。CFSR 可以展现气温随时间的变化趋势,但在桂北地区气温空间分布与观测值相比明显偏冷。

从 2020 年广西年平均气温分布图(图 3)可见,3 套再分析资料均能反映出南暖北冷的气温分布特征,CAM-RA 和 ERA5 对气温由北向南阶梯变化的

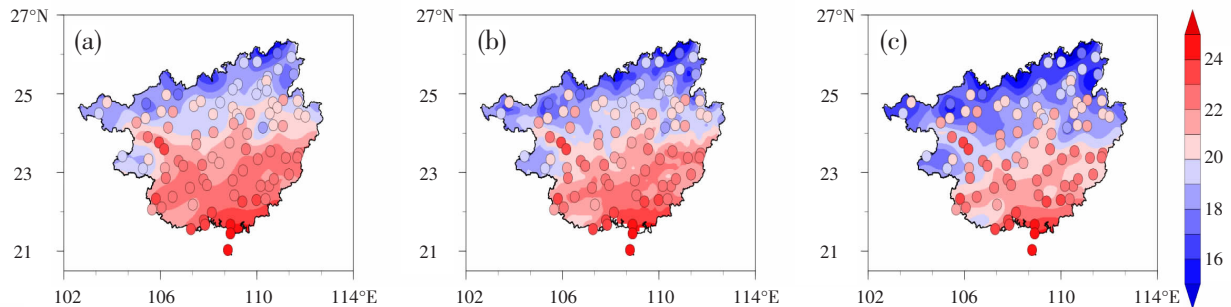


图 3 2020 年广西国家站 2m 气温观测值与 CMA-RA/Land(a)、ERA5/Land(b)及 CFSR(c)再分析资料的年平均分布图(填色:再分析资料,圆点:观测值,单位:℃)

2.2 站点评估

将 2020 年 CMA-RA/Land、ERA5/Land 及 CFSR 逐日再分析气温资料插值到广西 91 个国家观测站后,分别计算与观测值的相关系数、均方根误差、平均绝对误差及平均相对误差。分析再分析资料在广西不同站点上的表现可知:3 套再分析气温资料与 91 个站点的观测值均呈显著正相关,通过 99% 的显著性检验。其中 CMA-RA/Land 相关系数最高,有 62 个站点相关系数超过 0.99, 占到总数的 86.1%, 这些站点集中在桂南、桂中,及桂北的河池、柳州一带;在桂西北的百色及桂东北的桂林,CMA-RA/Land 相关系数在 0.98 到 0.99 之间,与 ERA5/Land 和 CFSR 持平。ERA5/Land 和 CFSR 只有 2 个站点相关系数超过 0.99。

CMA-RA/Land 气温资料与观测值的均方根误差范围在 0~3℃,均方根误差在 0~1℃的站点占总站数的 46.1%,在 1~2℃的占 42.8%,大于 2℃的仅占 11.1%。ERA5/Land 气温资料与观测值的均方根误差范围在 1~3℃,主要集中在 1~2℃,占总站数的 80.4%。CFSR 气温资料均方根误差较大,大于 3℃的占到总站数的 23.1%,而 CMA-RA/Land 和 ERA5/Land 均方根误差均未超过 3℃。3 套再分析气温资料均方根误差表现出较为一致的分布特征,在桂西北、桂东北均方根误差较大,CFSR 尤为偏大,其余地区均方根误差较小。CMA-RA/Land 再分析气温资料的均方根误差由北向南逐渐减小,在沿海地区最小。

CMA-RA/Land 气温资料平均绝对误差最小,分布形式与均方根误差相似,由北向南逐渐减小。ERA5/Land 气温资料平均绝对误差也较小,北部湾沿海一带误差在 0~1℃之间。CFSR 气温资料平均绝对误差明显大于 CMA-RA/Land 和 ERA5/Land。

通过计算插值到广西 91 个观测站的再分析气温资料与观测值的平均误差,能更精确地找到冷偏差站点。再分析气温资料在大部分站点上呈冷偏差,仅少部分站点呈暖偏差,桂南及沿海地区冷偏差小,至桂西北、桂东北的边缘山脉附近冷偏差增加。有研究指出,再分析温度资料在高山地区比观测值偏低,与模式地形高度和观测站海拔高度两者存在差异有关^[25]。桂西北山脉属于云贵高原边缘部分,多为陡坡,海拔 1000~1500m,部分超过 2000m^[26],3 套再分析资料在桂西北的复杂地形下均对气温有明显低

估。相较而言,CMA-RA/Land 气温资料在桂北山脉冷偏差最小,ERA5/Land 次之,CFSR 冷偏差最大。CMA-RA/Land、ERA5/Land 只有少量站点的气温相对误差超过-2℃,而 CFSR 在桂北乃至桂中有大量站点气温相对误差超过-2℃,从而导致 CFSR 气温资料在广西全区平均时冷偏差严重(图 2b)。就广西全区来说,CMA-RA/Land 相对误差最小,ERA5/Land 次之,CFSR 相对误差最大,且 CFSR 在复杂地形下的气温值模拟不如前两者。

3 套再分析气温资料虽然都存在偏差,但偏差程度不同。综合来看,CMA-RA/Land 作为中国第一代再分析陆面资料,其气温数据在 3 套再分析资料中表现最好,不论是对广西全区平均,还是在广西各站点上检验,CMA-RA/Land 各项评估指标均为最优。ERA5/Land 气温再分析资料效果也较好,且时间分辨率较高(逐 1h),同样在广西具备良好适用性。CFSR 再分析资料在广西大部均方根误差偏大,冷偏差明显,虽然可以展现气温变化趋势,但不能精确反映实际气温值,在广西的适用性较差。

3 结论与讨论

本文基于 2020 年广西 91 个国家观测站资料,对 CMA-RA/Land、ERA5/Land、CFSR 等 3 套再分析资料采用相关系数、均方根误差、平均相对误差、平均绝对误差等评估指标进行适用性研究,主要结论如下:

(1)CMA-RA/Land、ERA5/Land、CFSR 均能较好刻画 2020 年广西全区平均气温的时间变化趋势,CMA-RA/Land、ERA5/Land 对气温空间分布模拟效果较好。

(2)再分析气温资料整体呈冷偏差,在桂北山脉地区最明显,相较而言 CMA-RA/Land 与实际观测值最接近、ERA5/Land 次之、CFSR 冷偏差明显。

(3)整体看来,CMA-RA/Land 在广西适用性优于 ERA5。CMA-RA/Land 各项评估指标在全区平均及 91 个站点上的表现均为最好。ERA5 在广西适用性也较好,CFSR 适用性较差。

需要指出的是,文中基于 2020 年的逐日气温数据对 3 套再分析气温资料进行横向评估,强调再分析资料对年内逐日气温演变的复现能力。考虑到资料时间序列较短,存在一定局限性,未来将进一步完善更长时间尺度,如年际、年代际的适用性研究。

参考文献:

- [1] 周自江,曹丽娟,廖捷,等.水文气象信息概述:观测、融合与再分析[J].气象,2022,48(3):272-283.
- [2] 郭艳君,张思齐,颜京辉,等.中国探空观测与多套再分析资料气温序列的对比研究[J].气象学报,2016,74(2):271-284.
- [3] Saha S, Moorthi S, Pan H L, et al. The NCEP Climate Forecast System Reanalysis[J]. Bulletin of the American Meteorological Society, 2010, 91(8):1015-1058.
- [4] Rienecker, Michele M, Suarez M J, et al. MERRA: NASA's Modern-Era Retrospective-ive Analysis for Research and Applications[J]. Journal of Climate, 2011, 24(14):624-648.
- [5] Kobayashi S, Ota Y, Harada Y, et al. The JRA-55 Reanalysis: General Specification and Basic Characteristics[J]. Journal of the Meteorological Society of Japan, 2015, 93(1):5-48.
- [6] Simmons A J, Uppala S M, Dee D P, et al. ERA-Interim New ECMWF Reanalysis Products from 1989 Onwards[J]. ECMWF Newsletter, 2007(10):25-35.
- [7] Hersbach H, Bell B, Berrisford P, et al. The ERA5 Global Reanalysis[J]. Quarterly Journal of the Royal Meteorological Society, 2020, 146(730).
- [8] 赵天保,符淙斌,柯宗建,等.全球大气再分析资料的研究现状与进展[J].地球科学进展,2010,25(3):242-254.
- [9] 张东方,成青燕,何慧根,等.CFSv2 模式资料在成都市延伸期降水预测中的应用评估[J].气象研究与应用, 2021, 42(2):1-6.
- [10] 蓝玉峰,侯君杏,黄嘉宏.三套再分析降水资料在华南地区的适用性评估[J].气象研究与应用,2020,41(2):14-20.
- [11] 谢爱红, Ian Allison, 效存德, 等.不同再分析气温在东南极中山站-Dome A 断面的适用性评价[J].中国科学:地球科学,2014,44(1):156-168.
- [12] 丁旭,赖欣,范广洲,等.再分析土壤温湿度资料在青藏高原地区适用性的分析[J].高原气象,2018,37(3):626-641.
- [13] 陈思蓉,周秀华,陆虹,等.月动力延伸预测产品 DERF2.0 对广西气温和降水的预测评估[J].气象研究与应用, 2016, 37(3):16-19.
- [14] 赵天保,符淙斌.几种再分析地表气温资料在中国区域的适用性评估[J].高原气象,2009,28(3):594-606.
- [15] 谢潇,何金海,祁莉.4 种再分析资料在中国区域的适用性研究进展[J].气象与环境学报,2011,27(5):58-65.
- [16] 单帅,沈润平,师春香,等.中国北部积雪区冬季地表温度和 2m 气温再分析数据评估 [J]. 高原气象,2020,39(1):37-47.
- [17] 孟宪贵,郭俊建,韩永清.ERA5 再分析数据适用性初步评估[J].海洋气象学报,2018,38(1):91-99.
- [18] 朱景,袁慧珍.ERA 再分析陆面温度资料在浙江省的适用性[J].气象科技,2019,47(2):289-298.
- [19] 殷悦,马浩,葛敬文,等.不同分辨率再分析资料对浙江省气温刻画能力的对比评估 [J]. 热带气象学报, 2020, 36(3):377-388.
- [20] 丁敏,杜军,肖天贵.4 种再分析气温资料在羌塘国家级自然保护区的适用性研究[J].冰川冻土,2020,42(3):1046-1056.
- [21] Liao J, Wang H Y, Zhou Z J, et al. Integration, Quality Assurance, and Usage of Global Aircraft Observations in CRA[J]. Journal of Meteorological Research, 2021, 35(1):1-16.
- [22] 廖捷,胡开喜,江慧,等.全球大气再分析常规气象观测资料的预处理与同化应用[J].气象科技进展,2018,8(1):133-142.
- [23] 邓悦,任晓炜,丘平珠,等.台风“海高斯”(2007)期间多种降水融合实况产品的误差评估[J].气象研究与应用, 2021, 42(3):62-68.
- [24] 黄晴,梁维亮,黄荣.风廓线雷达产品与 ERA5 再分析资料的对比分析[J].气象研究与应用,2021,42(4):83-88.
- [25] 于法川,祝善友,张桂欣,等.复杂山区地形条件下 ERA5 再分析地表气温降尺度方法[J].地球信息科学学报, 2022, 24(4):750-765.
- [26] 傅中平,陈永红,杨传明,等.广西山岳景观资源特色、功能分类及其评价[J].广西地质,1998(2):69-74.

Applicability evaluation of surface air temperature from three reanalysis datasets in Guangxi Province

He Liyang¹, Su Chuancheng², Qin Weijian¹, Deng Yue², Lai Sheng¹, Zhu Qiuyu¹

(1.Guangxi Climate Center, Nanning 530022, China;

2.Guangxi Meteorological Information Center, Nanning 530022, China)

Abstract: Taking the 2m temperature data of 91 national surface meteorological observation stations in Guangxi in 2020 as the benchmark, a preliminary comparison of the applicability of surface air temperature from china first-generation land surface reanalysis data (CMA-RA/Land), and ERA5/Land, CFSR, which two are widely used at present, was conducted in Guangxi. The results show that the spatial-temporal distribution characteristics of temperature can be reproduced in all three sets of data, but there are varying degrees of cold deviation. Among them, CMA-RA/Land is the closest to the observed values, while the cold bias of ERA5/Land is obvious in northern Guangxi, and CFSR has the largest cold deviation. CMA-RA/Land performed best on all evaluation indicators, with the highest correlation coefficient, and its root mean square error, absolute error and relative error were the smallest. Overall, CMA-RA/Land has the best applicability in Guangxi.

Key words: reanalysis data; 2m temperature evaluation; observational data; linear fit; applicability