

张旭,毛炜峰,陈静,等.新疆区域 CRA40 地面气温检验分析[J].沙漠与绿洲气象,2022,16(5):15-23.

doi:10.12057/j.issn.1002-0799.2022.05.003

开放科学(资源服务)标识码(OSID):



新疆区域 CRA40 地面气温检验分析

张旭¹,毛炜峰^{2*},陈静²,李淑娟²,姚俊强²

(1.新疆气候中心,新疆 乌鲁木齐 830002;2.中国气象局乌鲁木齐沙漠气象研究所,新疆 乌鲁木齐 830002)

摘要:将高空间分辨率的 CRA40 地面日平均气温插值到新疆区域 105 站,以 105 站实测气温为基准,用平均误差(ME)、均方根误差($RMSE$)和相关系数(r)3 项检验指标,对新疆区域 CRA40 地面气温数据质量进行评估,与 ERA5 气温检验结果进行比较。结果表明:(1)1981—2020 年,新疆区域 105 站 CRA40 日平均气温与测站实况之间的相关系数 r_{CRA40} 为 0.973,通过 0.001 的显著性水平检验。新疆区域 CRA40 气温与实况比较,日平均气温的年平均、年最大值和最小值分别相差 0.12、0.02 和 1.94 °C,日平均气温年最大、最小值出现日期完全一致。(2)逐年检验结果中,2004 年之前 ME_{CRA40} 为负偏差,2005 年之后为正偏差,无论是 ME_{CRA40} 和 $RMSE_{CRA40}$ 还是 R_{CRA40} ,都反映出 2005—2020 年新疆区域 CRA40 气温数据质量有所改进,较前期更加接近 ERA5。(3)逐月检验结果中,新疆区域 ME_{CRA40} 绝对值和 $RMSE_{CRA40}$ 均在冬季最大。结合 ME 与 $RMSE$ 指标,在 3 月和 10—11 月新疆区域 CRA40 气温数据质量略优于 ERA5。(4)逐日检验结果中,新疆区域 ME_{CRA40} 的绝对值在 1.0 °C 以内的日数占全年的 80.8%, $RMSE_{CRA40} < 3.0$ °C 的日数占 72.3%。以 $RMSE$ 为指标,新疆区域 CRA40 数据质量优于 ERA5 的日数有 79 d,集中在 3—4 月和 10—11 月;以相关系数 r 为指标,新疆区域 CRA40 数据质量优于 ERA5 的日数有 22 d,其中 21 d 集中在 3 月。(5)逐站检验结果中,新疆区域 105 站中有 49 站的 ME_{CRA40} 绝对值较小,38 站的 $RMSE_{CRA40}$ 较小,有 51 站的 r_{CRA40} 较大。整体上 ERA5 气温数据质量略优于 CRA40,但是依据不同指标,CRA40 气温数据质量占优的测站分别占 36.2%~48.6%。(6)在分海拔高度的次级区域检验结果中,在 I 区海拔 <1 500 m 的区域,ERA5 气温数据质量整体占优;在 II 区海拔 1 500~2 000 m 的山区,CRA40 气温数据质量优于 ERA5;在 III 区海拔 >2 000 m 的山区,CRA40 气温数据质量略差,与 ERA5 十分接近。

关键词:日平均气温;检验;CRA40;ERA5;新疆

中图分类号:P468.021

文献标识码:A

文章编号:1002-0799(2022)05-0015-09

随着计算机技术的飞速发展,基于各种观测资料研制的再分析产品已经成为全球气候科研和业务工作的基础数据^[1-5]。从 20 世纪 90 年代中期开始,

收稿日期:2021-09-30;修回日期:2022-03-09

基金项目:国家重点研发计划(2019YFC1510501);国家自然科学基金重点项目(42030612);新疆维吾尔自治区自然科学基金项目(2022D01A121)

作者简介:张旭(1979—),男,高级工程师,主要从事遥感与应用气象研究。E-mail:zhangxu_home@163.com

通信作者:毛炜峰(1969—),男,研究员,主要从事气候预测、气候变化、应用气象等领域的相关研究。E-mail:mao6991@vip.sina.com

先后完成的全球大气再分析资料主要有:美国国家环境预测中心(NCEP)和大气研究中心(NCAR)的 40 a(1957—1996 年)NCEP/NCAR 全球大气再分析资料^[6-7]、欧洲中期数值预报中心(ECMWF)的 45 a(1975—2002 年)全球大气再分析资料(ERA40)^[8]、ECMWF 新一代的再分析资料(ERA-Interim)^[9]、日本气象厅(JMA)和电力中央研究所(CRIEP)发布的 25 a(1979—2004 年)全球大气再分析资料(JRA25)^[10]。目前,国际上全球大气再分析已经发展到第四代,水平和空间分辨率大幅度提高,如 ERA-

5、JRA-3Q、NCEP/ESRL等^[11-13]。2015年,中国气象局组建了“气象资料质量控制及多源数据融合与再分析”攻关团队,历时6a,成功研制了中国第一代全球大气和陆面再分析产品(CRA40),该产品重现了1979年以来从地面到55 km高度的全球三维大气状况,时间分辨率为6 h,空间分辨率达到30 km。经过多家业务科研单位的试用评估,产品质量与国际第三代全球再分析产品质量总体相当^[14-15]。

新疆地处欧亚大陆腹地,属于典型的干旱与半干旱气候,地域辽阔、地形复杂,气象观测站点稀少且分布不均,高质量的再分析资料应用需求更为迫切。CRA40再分析资料由我国自主研发,但是在新疆区域的质量评估未深入开展。日平均气温作为最基本的气象要素,其质量检验结果是应用再分析资料的重要依据^[16]。因此,以新疆105个气象站的日平均气温实测资料为基准,对插值后的105站CRA40日平均气温进行质量检验,同时与105站的ERA5日平均气温检验结果进行对比,评估CRA40地面气温数据质量,为更好地在新疆区域推广和应用我国具有自主知识产权的CRA40资料提供支持。

1 资料和方法

1.1 资料

中国第一代全球大气和陆面再分析产品(CRA40)通过中国气象数据网检索下载(<http://data.cma.cn/CRA>)^[15]。本文选取CRA40产品中的地面日平均气温,时间范围为1979—2020年,空间分辨率为 $0.25^\circ \times 0.25^\circ$ 。ECMWF发布的ERA5数据产品通过ECMWF Web API下载使用(<https://cds.climate.copernicus.eu>)^[17],选择ERA5地面气温逐时数据,经时区转换后统计为逐日平均气温,时间范围为1981—2020年,空间分辨率为 $0.25^\circ \times 0.25^\circ$ 。CRA40与ERA5再分析资料主要参数见表1。新疆区域有105个气象站有长期观测数据,站点分布见图1。新疆105站1981—2020年日平均气温数据来自新疆气象信息中心。

表1 CRA40与ERA5资料主要参数

再分析产品	生产机构	产品时段	空间分辨率	同化方案
CRA40/Land	CMA	1979—2020年	0.25°	4DVAR
ERA5/Land	ECMWF	1981—2020年	0.25°	Ensemble of 4DVAR

1.2 分析方法

1.2.1 双线性插值

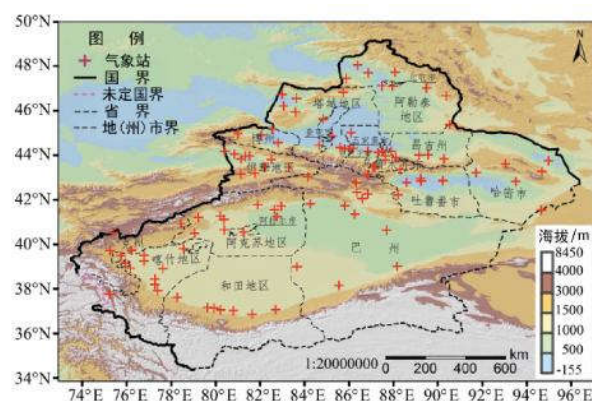


图1 新疆105个国家地面气象观测站分布

已有研究表明,基于双线性插值法来处理气温有较高的插值精度^[18]。本文利用国家气象中心预报技术研发室检验科研发的Meteva程序库^[19],将CRA40和ERA50地面气温站点插值到新疆区域105个气象站。

1.2.2 误差分析

选择平均误差(M_E)和均方根误差(R_{MSE})用于评估CRA40数据质量^[16-17]。平均误差反映了再分析数据相对于实测气温的偏差平均状态,具体见公式(1)。均方根误差(Root-Mean-Square Error, R_{MSE})反映了再分析数据与实测气温之间的偏离程度,用于定量计算再分析资料的偏差量,具体见公式(2)。

$$M_E = \frac{1}{N} \sum_{i=1}^N (Y_i - X_i), \quad (1)$$

$$R_{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - X_i)^2}. \quad (2)$$

公式(1)、(2)中, N 为数据样本个数, X_i 、 Y_i 分别为实测气温和再分析数据样本。

在CRA40气温数据的检验结果中,平均误差、均方根误差分别表示为 ME_{CRA40} 和 $RMSE_{CRA40}$ 。与ERA5气温数据有关的检验结果中,平均误差、均方根误差分别表示为 ME_{ERA5} 和 $RMSE_{ERA5}$ 。

1.2.3 相关分析

相关系数 r 反映了再分析气温和实测资料之间的相关程度^[2]。样本数为 N 的两套数据,其相关系数计算见公式(3)。

$$r = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}}. \quad (3)$$

其中, N 为数据样本数, X_i 、 Y_i 分别为实测气温与再分析数据样本, $\bar{X}$ 、 $\bar{Y}$ 分别为实测气温与再分析数据

序列的平均值。

CRA40 日平均气温与测站实况之间相关系数表示为 r_{CRA40} , ERA5 气温数据的相关系数表示为 r_{ERA5} 。

1.2.4 泰勒图方法

基于余弦定理,泰勒图能够将相关系数、均方根误差和标准差三种统计数据显示在一张图中,可以用来量化参考场和测试场之间的相关程度。从原点到点的径向距离与模式标准差成正比,方位位置给出了两个场之间的相关系数^[20]。观测值被定为沿横坐标距离原点一个单位的点,距离原点的环形长度为测试点和参考点的方差比率,代表再分析资料与观测资料离散度的相似性;方位角的余弦代表相关系数,测试点距离参考点的长度为两个场之间的均方根误差,距离参考点的长度越短,表明再分析资料越接近观测值。

2 结果分析

2.1 新疆区域 CRA40 日平均气温概况

选取 CRA40 地面日平均气温栅格数据,经双线性插值到全疆区域 105 站。40 a 新疆 105 站 CRA40 日平均气温与测站实况的散点见图 2, CRA40 日均气温为横坐标、测站实况气温为纵坐标,一元线性拟合斜率为 1.005。1981—2020 年 105 站的日均气温样本数高达 1 534 050 个,日平均气温的平均误差 ME_{CRA40} 为 0.13 °C,均方根误差 $RMSE_{\text{CRA40}}$ 为 3.05 °C。日平均气温与测站实测值之间的相关系数 r_{CRA40} 为 0.973,通过了 0.001 的显著性水平检验。整体上新疆区域 105 站 CRA40 地面日平均气温数据与测站实况十分接近,数据质量较好。

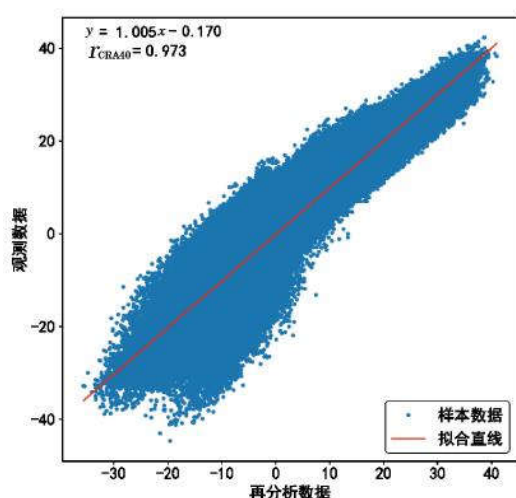


图 2 1981—2020 年 CRA40 与测站实测日平均气温

将新疆 105 站日平均气温的算数平均值作为新疆区域日平均气温,得到 1 月 1 日—12 月 31 日的新疆区域逐日平均气温(2 月 29 日样本仅 10 个,下文分析中忽略),再进一步计算逐日平均气温的 40 a 平均值。如图 3 所示,1981—2020 年基于测站实况的新疆区域日平均气温的 40 a 平均值在 $-11.63 \sim 24.10$ °C, 365 d 平均为 8.56 °C。基于 CRA40 的新疆区域日平均气温在 $-9.69 \sim 24.08$ °C, 365 d 日均气温为 8.68 °C;与测站实况值比较,年平均气温偏高 0.12 °C,日平均气温最高值偏低 0.02 °C、最低值偏高 1.94 °C。CRA40 与测站实况的日平均气温的年度最大、最小值出现日期相同,分别为 7 月 18 日和 1 月 18 日。CRA40 气温的一些主要气候特征值与测站实况差异很小,新疆区域 CRA40 地面气温数据有较高质量。

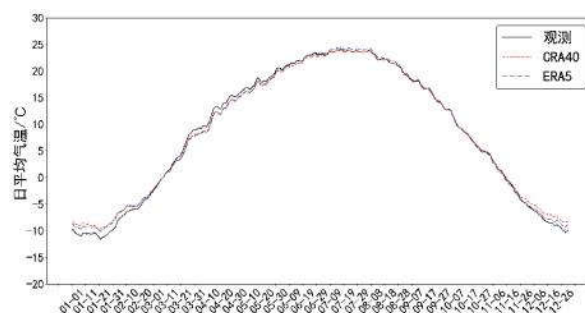


图 3 新疆区域 CRA40、ERA5 和测站实况逐日平均气温曲线

2.2 气温评估指标年际变化

由图 4 可知,1981—2020 年基于 CRA40、ERA5 以及测站实况的新疆区域年平均气温变化规律高度相似。对新疆 105 站 CRA40 日平均气温进行逐年检验,每年样本数为 38 325 个(闰年 38 430 个)。以测站实况为基准,计算新疆区域逐年的 ME_{CRA40} 、 $RMSE_{\text{CRA40}}$ 和 r_{CRA40} ,用这 3 项指标评估新疆区域 CRA40 气温数据质量。

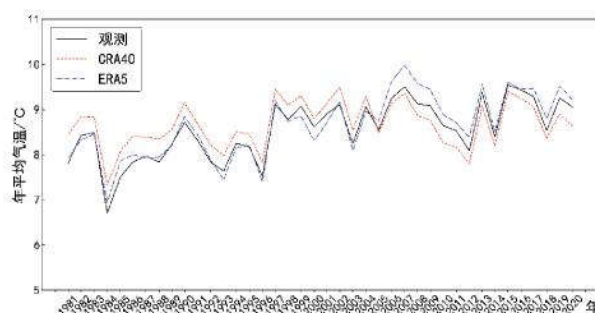


图 4 1981—2020 年新疆 105 站 CRA40、ERA5 和测站实测年平均气温曲线

2.2.1 CRA40 的评估指标年际变化

1981—2020 年, 新疆区域年 ME_{CRA40} 在 $-0.43\sim 0.63\text{ }^{\circ}\text{C}$, 40 a 平均为 $0.13\text{ }^{\circ}\text{C}$; 1981—2004 年(下文中简称为“前期”) ME_{CRA40} 为正偏差, 24 a 平均为 $0.37\text{ }^{\circ}\text{C}$, 2005—2020 年(下文中简称为“后期”) ME_{CRA40} 转为负偏差, 16 a 平均为 $-0.25\text{ }^{\circ}\text{C}$ 。可见, 在前、后期两个阶段, ME_{CRA40} 出现了明显变化。1981—2020 年, 年 $RMSE_{CRA40}$ 在 $2.81\sim 3.37\text{ }^{\circ}\text{C}$, 40 a 平均为 $3.10\text{ }^{\circ}\text{C}$; 前、后期 $RMSE_{CRA40}$ 平均值分别为 3.20 和 $2.95\text{ }^{\circ}\text{C}$; 后期 $RMSE_{CRA40}$ 较前期减小了 $0.25\text{ }^{\circ}\text{C}$ 。1981—2020 年, 年 r_{CRA40} 稳定在 $0.968\sim 0.979$, 均通过了 0.001 的显著性水平检验。后期 r_{CRA40} 的 14 a 平均值较前期增大了 0.005 , 有小幅提升(图 5)。可见, 从 ME 、 $RMSE$ 和 r 都反映出一个共同特征, 后期新疆区域 CRA40 气温数据质量较前期有所改进。

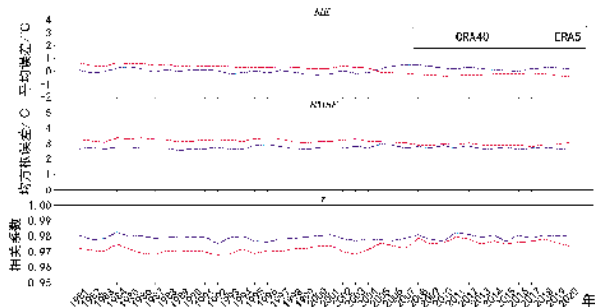


图 5 1981—2020 年 CRA40 和 ERA5 气温的 ME 、 $RMSE$ 和 r 历年曲线

2.2.2 CRA40 与 ERA5 评估指标对比

用同样方法对新疆区域 105 站 ERA5 日平均气温进行检验, 并将 CRA40 与 ERA5 的各项检验指标进行对比。新疆区域的年 ME_{CRA40} 与 ME_{ERA5} 非常接近, 两者的差值 40 a 平均为 $0.04\text{ }^{\circ}\text{C}$ 。1981—2020 年, 新疆区域的年 $RMSE_{CRA40}$ 均大于 $RMSE_{ERA5}$, 前、后期分别偏大 0.47 、 $0.19\text{ }^{\circ}\text{C}$, 后期 $RMSE_{CRA40}$ 与 $RMSE_{ERA5}$ 之间的差距减小。新疆区域的年 r_{CRA40} 小于 r_{ERA5} , 前、后期二者之间分别相差 0.008 、 0.003 。3 项检验指标结果显示, 2005—2020 年新疆区域 CRA40 气温数据质量更加接近 ERA5。

2.3 气温评估指标月际分布

2.3.1 CRA40 评估指标月分布特征

对 1981—2020 年新疆区域 105 站 CRA40 日平均气温进行逐月检验。1—12 月的日气温样本数分别为 118 650(2 月)、126 000(小月)和 130 200(大月)个。新疆区域的各月 ME_{CRA40} 在 $-0.89\sim 1.68\text{ }^{\circ}\text{C}$, 各月 $RMSE_{CRA40}$ 在 $2.50\sim 4.36\text{ }^{\circ}\text{C}$, 其中 1 月的 $RMSE_{CRA40}$

最大。新疆区域各月 r_{CRA40} 在 $0.766\sim 0.927$, 均通过 0.001 的显著性水平检验, 呈极显著正相关。

2.3.2 CRA40 与 ERA5 评估指标对比

由图 6、图 7 可知, 1—2 月和 11—12 月的 ME_{CRA40} 为正偏差, 1 月偏差最大; 3—10 月的 ME_{CRA40} 为负偏差, 4 月偏差最大。1—2 月、7—9 月和 12 月的 ME_{ERA5} 为正偏差, 1 月偏差最大; 其他 6 个月 ME_{ERA5} 为负偏差, 4 月偏差最大。在 3 月和 10—11 月, $RMSE_{CRA40} < RMSE_{ERA5}$, 在其他 9 个月, $RMSE_{CRA40} > RMSE_{ERA5}$ 。综合 ME 与 $RMSE$ 指标, 新疆区域 CRA40 气温数据质量在 3 月和 10—11 月略优于 ERA5。

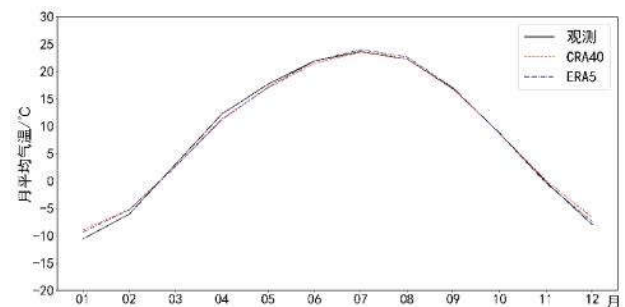


图 6 新疆区域 105 站 CRA40、ERA5 与测站实测逐月平均气温曲线

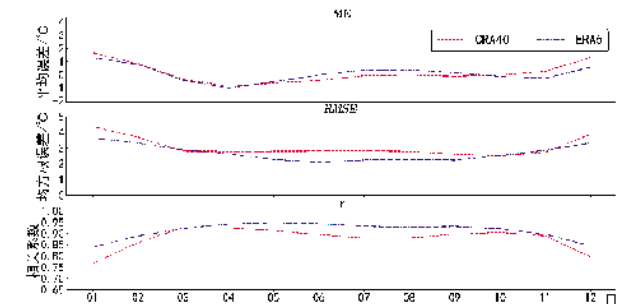


图 7 新疆区域 105 站 CRA40、ERA5 气温的 ME 、 $RMSE$ 和 r 逐月变化曲线

各月 r_{ERA5} 在 $0.839\sim 0.947$, 均通过 0.001 的显著性水平检验, 呈极显著的正相关。对比发现, 3 月 $r_{CRA40} > r_{ERA5}$, 其他 11 个月 $r_{ERA5} > r_{CRA40}$ 。以相关系数 r 为指标, CRA40 气温数据质量在 3 月优于 ERA5。

2.4 气温评估指标的逐日变化特征

2.4.1 CRA40 评估指标的年内逐日变化

对 1981—2020 年新疆区域 CRA40 日平均气温进行逐日检验。1 月 1 日—12 月 31 日 365 d(忽略 2 月 29 日)的日气温样本数均为 40 个。由图 8 可知, 日 ME_{CRA40} 在 $-1.07\sim 2.23\text{ }^{\circ}\text{C}$, 365 d 平均为 $0.13\text{ }^{\circ}\text{C}$ 。365 d 中, 有 156 d 的 ME_{CRA40} 为正偏差, 占 42.7%, 1 月 8 日正偏差最大, 另外 209 d 的 ME_{CRA40} 为负偏

差,占 57.3%,4 月 18 日负偏差最大。日 ME_{CRA40} 绝对值在 $1.0\text{ }^{\circ}\text{C}$ 以内的天数为 295 d, 占全年的 80.8%。日 $RMSE_{CRA40}$ 在 $2.41\sim 4.80\text{ }^{\circ}\text{C}$, 365 d 日平均为 $3.05\text{ }^{\circ}\text{C}$, 其中 1 月 8 日 $RMSE_{CRA40}$ 最大。日 $RMSE_{CRA40}<3.0\text{ }^{\circ}\text{C}$ 的日数占全年的 72.3%。 r_{CRA40} 在 $0.766\sim 0.927$, 365 d 平均为 0.867, 均通过 0.001 的显著性水平检验。

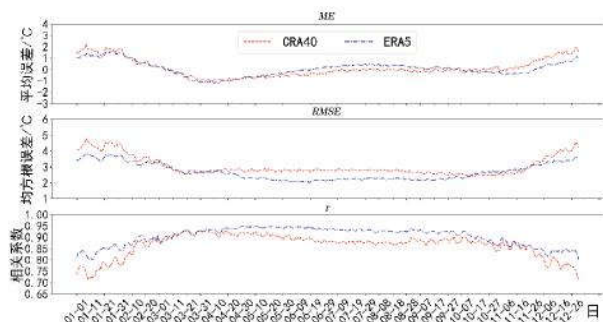


图 8 新疆区域 105 站 CRA40、ERA5 气温的 ME 、 $RMSE$ 和 r 逐日变化曲线

2.4.2 CRA40 与 ERA5 评估指标对比

新疆区域 CRA40 与 ERA5 日平均气温的逐日检验结果见图 8。日 ME_{ERA5} 在 $-1.21\sim 1.56\text{ }^{\circ}\text{C}$, 365 d 平均为 $0.09\text{ }^{\circ}\text{C}$; ME_{ERA5} 绝对值在 $1.0\text{ }^{\circ}\text{C}$ 以内的天数为 301 d, 占全年的 82.5%, 略高于 ME_{CRA40} 。 ME_{CRA40} 绝对值小于 ME_{ERA5} 绝对值的日数达 179 d, 占全年的 49.0%。以 ME 为对比指标, CRA40 气温数据质量略差于 ERA5。

日 $RMSE_{ERA5}$ 在 $2.01\sim 3.83\text{ }^{\circ}\text{C}$, 365 d 平均为 $2.69\text{ }^{\circ}\text{C}$ 。全年日 $RMSE_{ERA5}$ 的年平均值、年最高、最低值均低于 $RMSE_{CRA40}$ 的对应值。 $RMSE_{ERA5}<3.0\text{ }^{\circ}\text{C}$ 的日数有 260 d, 占 365 d 的 71.2%, 略低于 CRA40 的 72.3%。 $RMSE_{CRA40}<RMSE_{ERA5}$ 的日数有 79 d, 占 21.6%, 集中在 3—4 月和 10—11 月。以 $RMSE$ 为对比指标, 全年 21.6% 的日数 CRA40 数据质量优于 ERA5, 集中在 3—4 月和 10—11 月。

由图 8 可见, $r_{CRA40}>r_{ERA5}$ 的天数有 22 d, 占全年的 6.0%, 其中有 21 d 集中在 3 月。以相关系数 r 为对比指标, CRA40 气温数据质量整体上略差于 ERA5, 但是在 3 月 CRA40 气温数据质量优于 ERA5。

2.5 105 站气温评估指标空间分布

2.5.1 CRA40 气温逐站评估指标

对新疆区域 CRA40 气温进行逐站检验。由图 9 可知, 新疆区域单站 ME_{CRA40} 在 $-7.73\sim 4.26\text{ }^{\circ}\text{C}$, 105 站平均为 $0.13\text{ }^{\circ}\text{C}$; 有 83 站的 ME_{CRA40} 绝对值 $<2\text{ }^{\circ}\text{C}$, 占 79.0%。由图 10 可知, 单站 $RMSE_{CRA40}$ 在 $1.21\sim$

$8.43\text{ }^{\circ}\text{C}$, 105 站平均为 $2.81\text{ }^{\circ}\text{C}$; 有 76 站 $RMSE_{CRA40}<3\text{ }^{\circ}\text{C}$, 占 72.4%。105 站的 r_{CRA40} 在 $0.928\sim 0.996$, 均通过了 0.001 的显著性水平检验。

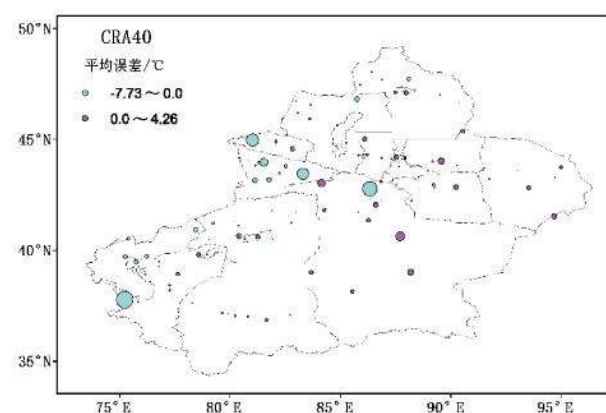


图 9 新疆 105 站 CRA40 日平均气温平均误差 ME 空间分布

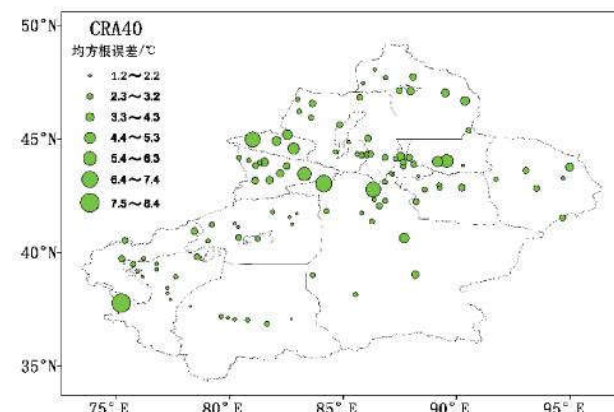


图 10 新疆 105 站 CRA40 日平均气温均方根误差 $RMSE$ 空间分布

对新疆区域 ERA5 日平均气温进行逐站检验, 并将 CRA40 与 ERA5 的各项检验指标进行对比。有 49 站的 ME_{CRA40} 绝对值 $<ME_{ERA5}$ 绝对值, 占 46.7%; 有 38 站的 $RMSE_{CRA40}<RMSE_{ERA5}$, 占 36.2%; 有 51 站的 $r_{CRA40}>r_{ERA5}$, 占 48.6%。依据不同指标, CRA40 气温数据质量优于 ERA5 的测站可以达到 36.2%~48.6%。新疆区域 CRA40 气温数据质量略差于 ERA5。

2.5.2 CRA40 与 ERA5 单站指标极端值对比

由表 2 可知, 单站 ME_{CRA40} 和 ME_{ERA5} 绝对值较大的站多为负偏差, 负偏差最大的测站相同, 均是塔什库尔干, ME_{CRA40} 和 ME_{ERA5} 分别为 -7.73 、 $-8.37\text{ }^{\circ}\text{C}$ 。 ME_{CRA40} 和 ME_{ERA5} 负偏差最大的 5 站相同, 分别是塔什库尔干、巴仑台、温泉、新源、伊宁县, 这些站点大多分布在山区, 测站周边地形较复杂。由表 3 可知, ME_{CRA40} 正偏差最大 5 站分别是铁干里克、巴音布鲁克、奇台、若羌、焉耆, 大多分布在南疆海拔较低的平

原; ME_{ERA5} 正偏差较大站点也分布在平原地区,但是主要位于北疆昌吉和阿勒泰地区。

表2 CRA40和ERA5的日气温单站平均
误差 ME 负偏差最大5站统计

排位	$ME_{CRA40}/^{\circ}\text{C}$	站名	海拔/m	$ME_{ERA5}/^{\circ}\text{C}$	站名	海拔/m
1	-7.73	塔什库尔干	3 090.1	-8.37	塔什库尔干	3 090.1
2	-6.70	巴仑台	1 732.4	-6.11	巴仑台	1 732.4
3	-5.65	温泉	1 357.8	-4.11	新源	928.2
4	-5.29	新源	928.2	-3.64	天山大西沟	3 539.0
5	-3.65	伊宁县	770.0	-3.32	阿合奇	1 985.1

表3 CRA40和ERA5的日气温单站平均误差 ME
正偏差最大5站统计

排位	$ME_{CRA40}/^{\circ}\text{C}$	站名	海拔/m	$ME_{ERA5}/^{\circ}\text{C}$	站名	海拔/m
1	4.26	铁干里克	846.0	2.35	莫索湾	346.0
2	3.51	巴音布鲁克	2 458.0	2.20	蔡家湖	440.5
3	3.03	奇台	793.5	2.16	塔中	1 099.3
4	2.86	若羌	888.2	2.13	尉犁	884.9
5	2.32	焉耆	1 055.3	2.08	阿克达拉	563.3

由表4可知,新疆区域 $RMSE_{CRA40}$ 由大到小前5位测站依次是塔什库尔干、巴音布鲁克、巴仑台、温泉和新源。 $RMSE_{ERA5}$ 由大到小前5位的测站依次是塔什库尔干、巴仑台、巴音布鲁克、新源和天山大西沟。CRA40和ERA5数据中,塔什库尔干站的RMSE均为105站中的最大值,并且 $RMSE_{ERA5} > RMSE_{CRA40}$ 。再分析气温数据质量较极端的测站分布在地形复杂的高海拔山区,相对而言,质量极端的测站中CRA40数据质量优于ERA5。

表4 CRA40和ERA5的日气温单站均方根误差
最大5站统计

排位	$RMSE_{CRA40}/^{\circ}\text{C}$	站名	海拔/m	$RMSE_{ERA5}/^{\circ}\text{C}$	站名	海拔/m
1	8.43	塔什库尔干	3 090.1	9.19	塔什库尔干	3 090.1
2	7.49	巴音布鲁克	2 458.0	6.59	巴仑台	1 732.4
3	6.99	巴仑台	1 732.4	4.98	巴音布鲁克	2 458.0
4	6.89	温泉	1 357.8	4.73	新源	928.2
5	6.19	新源	928.2	4.45	天山大西沟	3 539.0

2.6 不同海拔高度区域气温评估比较

在海拔较高、地形复杂的山区,两套再分析数据质量均有所下降。将新疆区域105站记为XJ区,按照海拔高度将新疆划分为3个次级区域,分别为: $<$

1 500 m的区域,有90个气象站,记为I区;1 500~2 000 m的区域,有10个气象站,记为II区; $>$ 2 000 m的区域,有5个气象站,记为III区。综合相关系数、均方根误差和标准差,绘制新疆区域105站气温泰勒图(图11)。在I、II区,海拔2 000 m以下,各站CRA40气温综合评估指标相差不大,在III区,各站综合指标离散度明显增大。

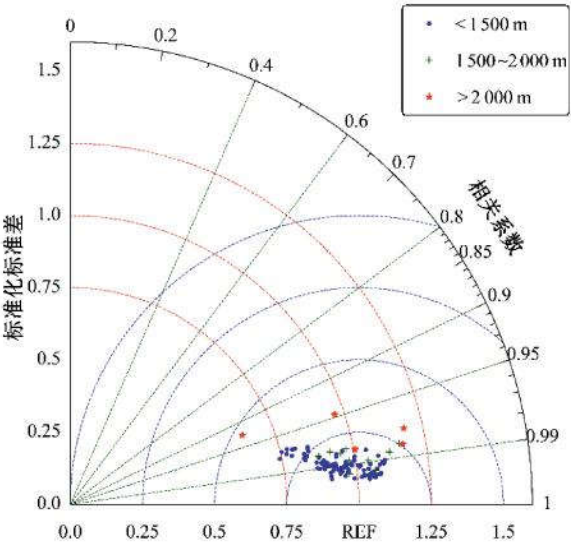


图11 新疆105站CRA40日均气温泰勒图

2.6.1 不同区域日平均的整体检验

新疆不同海拔高度区域的CRA40与ERA5的各项检验指标见表5。对比分析各次级区域的CRA40和ERA5气温检验结果,在I区, $RMSE_I$ 和 r_I 指标反映出CRA40气温数据质量整体略差于ERA5;在II区, ME_{II} 、 $RMSE_{II}$ 和 r_{II} 指标均反映出CRA40气温数据整体质量优于ERA5;在III区, ME_{III} 和 $RMSE_{III}$ 指标反映出CRA40气温数据质量较优,但 r_{III} 指标反映出ERA5气温数据质量相对更优(表6)。

表5 新疆不同区域CRA40、ERA5气温的
检验指标统计

区域范围	区域 编号	平均误差 $ME/^{\circ}\text{C}$		均方根误差 $RMSE/^{\circ}\text{C}$		相关系数 r	
		ME_{CRA40}	ME_{ERA5}	$RMSE_{CRA40}$	$RMSE_{ERA5}$	r_{CRA40}	r_{ERA5}
新疆	XJ	0.13	0.09	2.81	2.53	0.987	0.988
$>$ 1 500 m	I区	0.31	0.46	2.68	2.33	0.989	0.990
1 500~2 000 m	II区	-0.78	-1.57	2.98	3.18	0.986	0.983
$<$ 2 000 m	III区	-1.21	-3.09	4.81	4.86	0.963	0.967

注:各区域检验时先计算逐站指标,新疆为105站的平均值,3个次级区域分别为90站、10站和5站的平均值。

2.6.2 逐日平均误差 ME 检验对比

由图 12 和表 6 可知,在 I 区,逐日 ME_{CRA40_I} 在 $-0.88\sim 2.61\text{ }^{\circ}\text{C}$,365 d 平均为 $0.30\text{ }^{\circ}\text{C}$ 。在 II 区,逐日 ME_{CRA40_II} 在 $-1.57\sim 0.01\text{ }^{\circ}\text{C}$,365 d 平均为 $-0.78\text{ }^{\circ}\text{C}$ 。在 III 区,逐日 ME_{CRA40_III} 在 $-3.45\sim 1.05\text{ }^{\circ}\text{C}$,365 d 平均为 $-1.21\text{ }^{\circ}\text{C}$ 。随海拔高度升高, ME_{CRA40} 的负偏差程度越强。整体上 ME_{ERA5} 的负偏差更强;随海拔高度上升, ME_{ERA5} 的负偏差幅度加剧的程度更为严重。依据 ME 指标,在 II、III 区海拔 $>1\ 500\text{ m}$ 的高海拔山区,CRA40 气温数据质量优于 ERA5。

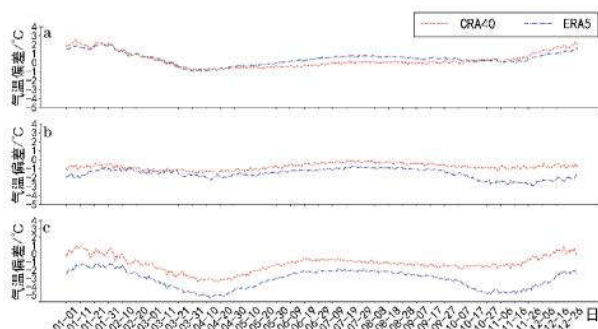


图 12 各海拔高度区域日平均气温平均误差 ME 逐日变化曲线

(a 海拔 $<1\ 500\text{ m}$, b 海拔 $1\ 500\sim 2\ 000\text{ m}$, c 海拔 $>2\ 000\text{ m}$)

表 6 新疆不同区域 CRA40、ERA5 气温的全年逐日检验指标统计

区域范围	区域 编号	平均误差 $ME/^{\circ}\text{C}$		均方根误差 $RMSE/^{\circ}\text{C}$		相关系数 r	
		CRA40	ERA5	CRA40	ERA5	CRA40/ERA5	
新疆	XJ	0.13	0.09	3.05	2.69	0.867	0.903
$<1\ 500\text{ m}$	I 区	0.30	0.45	2.82	2.35	0.858	0.899
$1\ 500\sim 2\ 000\text{ m}$	II 区	-0.78	-1.57	3.28	3.39	0.745	0.780
$>2\ 000\text{ m}$	III 区	-1.21	-3.09	5.25	5.23	0.641	0.790

注:逐日计算全年的各检验指标,样本数均为 40,再进行 365 d 平均(忽略 2 月 29 日结果)。

2.6.3 逐日均方根误差 $RMSE$ 检验对比

由图 13 和表 6 可知,在 I 区,逐日 $RMSE_{CRA40_I}$ 在 $2.00\sim 4.67\text{ }^{\circ}\text{C}$,365 d 平均为 $2.82\text{ }^{\circ}\text{C}$,有 277 d $<3.0\text{ }^{\circ}\text{C}$,占全年的 75.9%;逐日 $RMSE_{ERA5_I}$ 在 $1.80\sim 3.58\text{ }^{\circ}\text{C}$,365 d 平均为 $2.35\text{ }^{\circ}\text{C}$,有 307 d $<3.0\text{ }^{\circ}\text{C}$,占全年的 84.1%。在 II 区,逐日 $RMSE_{CRA40_II}$ 在 $2.73\sim 3.83\text{ }^{\circ}\text{C}$,365 d 平均为 $3.28\text{ }^{\circ}\text{C}$,有 58 d $<3.0\text{ }^{\circ}\text{C}$,占全年的 15.9%;逐日 $RMSE_{ERA5_II}$ 在 $2.25\sim 4.84\text{ }^{\circ}\text{C}$,365 d 平均为 $3.39\text{ }^{\circ}\text{C}$,有 128 d $<3.0\text{ }^{\circ}\text{C}$,占全年的 35.1%。在 III 区,逐日 $RMSE_{CRA40_III}$ 在 $3.19\sim 8.36\text{ }^{\circ}\text{C}$,365 d 平均为 $5.25\text{ }^{\circ}\text{C}$;

逐日 $RMSE_{ERA5_III}$ 在 $3.07\sim 7.14\text{ }^{\circ}\text{C}$,365 d 平均为 $5.23\text{ }^{\circ}\text{C}$ 。随海拔高度升高, $RMSE_{CRA40}$ 和 $RMSE_{ERA5}$ 均增大, $<3.0\text{ }^{\circ}\text{C}$ 的日数迅速减少,数据质量下降。

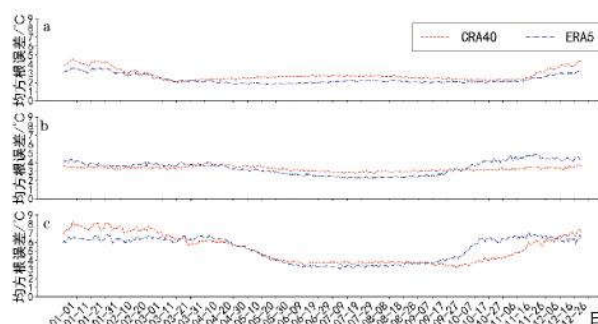


图 13 各海拔高度区域日平均气温均方根误差 $RMSE$ 逐日变化曲线

(a 海拔 $<1\ 500\text{ m}$, b 海拔 $1\ 500\sim 2\ 000\text{ m}$, c 海拔 $>2\ 000\text{ m}$)

依据 ME 指标,在 I 区海拔 $<1\ 500\text{ m}$ 的区域,ERA5 气温数据质量整体占优;在 II 海拔 $1\ 500\sim 2\ 000\text{ m}$ 的山区, $RMSE$ 指标反映出 CRA40 气温数据质量优于 ERA5;在 III 区海拔 $>2\ 000\text{ m}$ 的山区,CRA40 气温数据质量与 ERA5 十分接近。

3 结论与讨论

对 1981—2020 年新疆区域 105 站 CRA40 日平均气温进行检验,与 ERA5 气温检验结果进行对比,得到以下结论:

(1)1981—2020 年新疆区域 CRA40 日平均气温与测站实况之间相关系数为 0.973,通过 0.001 的显著性水平检验。新疆区域 CRA40 气温与实况比较,日平均气温的年平均、年最大值和最小值分别相差 $0.12\text{ }^{\circ}\text{C}$ 、 $0.02\text{ }^{\circ}\text{C}$ 、 $1.94\text{ }^{\circ}\text{C}$,基于 CRA40 和测站实况的日平均气温年内最大、最小值出现日期完全一致。整体上新疆区域 CRA40 气温数据质量较高。

(2)逐年检验结果中,40 a 来 CRA40 日气温的逐年 ME 均在 $1\text{ }^{\circ}\text{C}$ 以内,年 $RMSE$ 稳定在 $3\text{ }^{\circ}\text{C}$ 左右。2005—2020 年 CRA40 气温数据质量较前期有所提高。逐月检验结果中,结合 ME 与 $RMSE$ 指标,在 3 月和 10—11 月新疆区域 CRA40 气温数据质量略优于 ERA5。逐日检验结果中,以逐日 $RMSE_{CRA40}$ 为指标,全年 CRA40 数据质量优于 ERA5 的日数有 79 d,集中在 3—4 月和 10—11 月。

(3)逐站检验结果中,新疆区域 105 站中有 49 站的 ME_{CRA40} 绝对值较小,38 站的 $RMSE_{CRA40}$ 较小,51 站的 r_{CRA40} 较大。整体上,ERA5 气温数据质量略

优于CRA40,但是依据不同指标,CRA40气温数据质量占优的测站分别为36.2%~48.6%。CRA40气温数据质量随海拔高度上升而下降,在1500 m以上高海拔地区,CRA40气温数据质量反而优于ERA5。

(4)在分海拔高度的次级区域检验结果中,在Ⅰ区,有186 d的 ME_{CRA40} 绝对值较小,24 d的 $RMSE_{CRA40}$ 较小,32 d的 r_{CRA40} 较大,ERA5气温数据质量整体占优。在Ⅱ区,有362 d的 ME_{CRA40} 绝对值较小,191 d的 $RMSE_{CRA40}$ 较小,142 d的 r_{CRA40} 较大,CRA40气温数据质量优于ERA5。在Ⅲ区,365 d的 ME_{CRA40} 绝对值均大于 ME_{ERA5} ,51 d的 $RMSE_{CRA40}$ 较小,7 d的 r_{CRA40} 较大,CRA40气温数据质量略差,与ERA5十分接近。

已有研究指出,模式地形误差可能是造成再分析资料误差的重要原因^[22-24],在新疆海拔>1500 m的区域, $RMSE_{CRA40}$ 略优于 $RMSE_{ERA5}$ 数据。可能与CRA40数据制作过程中更侧重于地面观测、探空资料和风云卫星资料的同化及应用有关系。在区域尺度上CRA40数据的各个要素的质量检验还需深入研究。

参考文献:

- [1] 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.
- [2] 黄嘉佑,李庆祥.气象数据统计分析方法[M].北京:气象出版社,2014:315-317.
- [3] KALNAY E,蒲朝霞,杨福全,等.译.大气模式,资料同化和可预报性[M].北京:气象出版社,2005:6-7.
- [4] 邓明仔,何清,阿力木·阿巴斯,等.巴基斯坦近40 a气温时空变化特征[J].沙漠与绿洲气象,2021,15(2):59-69.
- [5] 尹仔锋.天气与气候再分析的进展与前景[J].沙漠与绿洲气象,2008,2(5):60-61.
- [6] KALNAY E,KANAMITSU M,KISTLER R,et al.The NCEP/NCAR 40-year reanalysis project [J].Bull Amer Meteor Soc,1996,77(3):437-472.
- [7] KANAMITSU M,EBISUZAKI W,WOOLLEN J,et al.NCEP-DOE AMIP-II Reanalysis (R-2)[J].Bull Amer Meteor Soc,2002,83:1631-1643.
- [8] UPPALA S M,KALLBERG P W,SIMMONS A J,et al.The ERA-40 reanalysis[J].QJRM,2005,131B:2961-3012.
- [9] DEE D,UPPALA S M,SIMMONS A J,et al.The ERA-Interim reanalysis:Configuration and performance of the data assimilation system [J].Quart J Roy Meteor Soc,2011,137:553-597.
- [10] ONOGI K,et al.The JRA-25 Reanalysis[J].Met Soc Japan,2007,85:369-432.
- [11] RIENECKER M M,SUAREZ M J,GELARO R,et al.MERRA:NASA'S modern-era retrospective analysis for research and application[J].Climate,2011,24:3624-3648.
- [12] COMPO G P,WHITAKER J S,SARDESHMUKH P D,et al.The Twentieth Century Reanalysis Project[J].Quart J Roy Meteor Soc,2011,137:1-28.
- [13] HANS H,DICK DEE.ERA5 reanalysis is in production [R].ECMWF Newsletter No.147, Spring 2016.
- [14] 王旻燕,姚爽,姜立鹏,等.我国全球大气再分析(CRA-40)卫星遥感资料的收集和预处理[J].气象科技进展,2018,8(1):158-163.
- [15] 中国第一代全球大气和陆面再分析产品(CRA)用户手册 [A/OL].(2021-05)[2021-05].<http://data.cma.cn/data/index/98555d0119fa185a.html>.
- [16] 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.
- [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] 王传辉,姚叶青,时刚.江淮地区 ERA-Interim 再分析与观测温度资料对比分析 [J]. 气象,2018,44(9):1220-1228.
- [19] 饶莉娟,王健林,张星.不同插值方法对精细化预报产品在青岛地区的检验比较 [J]. 中国农学通报,2020,36(32):100-108.
- [20] TAYLOR K E.Summarizing multiple aspects of model performance in a single diagram [J].Geophys Res Atmos,2001,106:7183-7192.
- [21] 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.
- [22] 程新宇杰,高路.ECMWF再分析气温资料在天山山区的可信度检验 [J]. 北京师范大学学报(自然科学版),2018,54(5):635-644.
- [23] 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.

Temperature Error Test of CMA Global Atmospheric Reanalysis Data in Xinjiang

ZHANG Xu¹, MAO Weiyi², CHEN Jing², LI Shujuan², YAO Junqiang²

(1.Xinjiang Climate Center, Urumqi 830002, China;

2.Institute of Desert Meteorology, China Meteorological Administration, Urumqi 830002, China)

Abstract Based on the 105 national meteorological stations over Xinjiang during 1981–2020, the 2 meters temperature errors of (CRA40) data were analyzed by using the mean error (ME), root mean square error ($RMSE$) and correlation coefficient (r). The characteristics of spatio-temporal distribution of the errors show that the 40-year CRA40 reanalysis of the daily average temperature are significantly correlated with the observed station data, the corresponding correlation is 0.973, and passes the 0.001 significance test. The CRA40 reanalysis daily average temperature is consistent with the observed data, the annual average, maximum and minimum values of daily average temperature differ by 0.12, 0.02 and 1.94 °C, respectively, and the maximum and minimum daily average temperatures appear on the same date. The annual inspection results of ME_{CRA40} , $RMSE_{CRA40}$, and r_{CRA40} show that the quality of CRA40 temperature data in Xinjiang from 2005 to 2020 has improved, and it is closer to ERA5 than before 2005. Before 2004, ME_{CRA40} was negative, and after 2005, it was positive. The monthly ME_{CRA40} and $RMSE_{CRA40}$ inspection results show that the maximum temperature error in Xinjiang appears in the winter. In March, October and November, the temperature data error of CRA40 is slightly better than that of ERA5. The daily inspection results show that the number of days with the absolute value of ME_{CRA40} within 1.0°C account for 80.8% of the year, and the number of days with $RMSE_{CRA40}$ less than 3.0°C account for 72.3% of the year. Taking $RMSE$ as an indicator, the data quality of CRA40 is better than that of ERA5 for 79 days, mainly in March–April and October–November. Using the correlation coefficient as an indicator, there are 22 days of CRA40 data quality better than ERA5, mainly in March. The results of the meteorological observation station-by-station inspection show that 49 of the 105 stations in Xinjiang have better ME_{CRA40} , 38 stations have better $RMSE_{CRA40}$, and 51 stations have worse r_{CRA40} . On the whole, the air temperature data quality of ERA5 is slightly better than that of CRA40, but according to different indicators, the stations with superior air temperature data quality of CRA40 account for 36.2%–48.6% respectively. In areas with an altitude of less than 1 500 m, the temperature data quality of ERA5 is superior. In mountainous areas with an altitude of 1 500–2 000 m, the temperature data quality of CRA40 is better than that of ERA5. In mountainous areas with an altitude of more than 2 000 m, the quality of CRA40 temperature data is slightly worse, close to ERA5.

Key words daily average temperature; error test; CRA40; ERA5; Xinjiang