

文章编号: 2096-1618(2022)06-0683-07

大渡河流域逐日降雨数据融合的误差分析

陈媛¹, 董丹丹², 申飙³, 蔡宏珂²

(1. 国能大渡河大数据服务有限公司, 四川 成都 610041; 2. 成都信息工程大学, 四川 成都 610225; 3. 重庆美天科技有限公司, 重庆 401120)

摘要:为有效结合不同来源降雨资料的优势,发展多源降雨融合技术,采用数据融合方法,生成融合逐日降雨数据(以下称为 Grid 数据),再与气象观测站点资料、ERA5 和 NCEP2 格点资料进行误差对比,探讨 Grid 数据在大渡河流域的适用性。结果表明:(1)Grid 数据基本能完整展示大渡河流域的日均降雨分布,总体的、各区域内的降雨分布与观测数据展示出的分布规律基本吻合;(2)从插值站点来看,Grid 数据的降雨量量级与观测数据差距小,总体的、各区域内的最大降雨量和最小降雨量与观测数据接近,表现优于 ERA5 和 NCEP2 数据;(3)从标准差、均方根误差来看,Grid 数据的误差要明显小于 ERA5 和 NCEP2 数据;(4)从 2008–2018 年全流域、1~4 分区的面雨量时间序列来看,Grid 数据与观测数据较 ERA5 和 NCEP2 数据更接近。综上,Grid 融合数据能较好地反映大渡河流域实际降雨情况。

关键词:气象学;气象数据处理与方法;数据融合方法;降雨分布特征;大渡河流域

中图分类号:P426.6

文献标志码:A

doi:10.16836/j.cnki.jcuit.2022.06.011

0 引言

地面雨量站网可以提供精确的点降雨信息,但由于其空间分布不均,在雨量站网稀疏地区,如果仅用雨量网观测降雨估计流域面降雨量将会产生很大的不确定性。数值模式输出的降雨产品可以弥补观测系统部署不均造成的问题,但模式降雨产品本身亦存在严重的系统偏差^[1-2]。基于单一来源的降雨资料各有利弊,如何有效结合不同来源降雨资料的优势,发展多源降雨融合技术,成为近年来国际上在高质量降雨产品研发中的主流趋势^[3-5]。

数据融合是将来源和分辨率各异的、直接和间接的观测数据与模型模拟结果集成,生成在时间、空间和物理特性三个维度上具有一致性的各种数据集,通过数据融合技术对地面观测降雨和格点降雨产品进行综合分析以获取最优面降雨估计,融合后的降雨具有多种降雨资料的优势,能够精确反映降雨的空间分布状态^[6-8]。数据融合方法有两种^[9-10]:(1)全局校正融合方法的基本假设是格点降雨在空间域上具有连续分布的偏差,利用真实观测的站点数据对每个格点降雨进行修正,此方法融合了地面观测降雨的局部高精度和格点降雨资料覆盖范围广且空间分布合理的优点。

(2)局部校正融合方法的核心是基于加权平均、回归分析等数学思想,主要利用单个格点所包含的雨量站点降雨信息对格点降雨信息进行局部修正,此融合方法考虑了降雨的空间自相关性,在雨量站网密度较大地区融合效果较好,但在雨量站网密度较低地区,雨量站点距离较大,其空间自相关性被高估,融合结果会产生很大的不确定性。

四川大渡河流域位于 99°42'E ~ 103°48'E, 28°15'N ~ 33°33'N,青藏高原东南边缘向四川盆地西部的过渡地带,流域内地貌形态差异明显,地理特征复杂,气候差异明显^[11-12],与降雨密切相关的滑坡、泥石流等地质灾害频繁发生^[13-14],流域的流量、江河的抗洪能力以及水库的蓄洪规模都与流域的平均降雨量密切相关^[15-16]。目前,大渡河流域气象观测站点分布不均且密度较低,存在某些区域缺乏雨量观测数据的问题。因此,有必要对大渡河流域降雨数据进行融合方法研究及其误差分析,从而改善流域内降雨数据精度不够的问题,为防灾减灾提供参考。

1 数据与方法

1.1 研究数据

采用资料:(1)1979–2018 年大渡河流域自建测站和气象系统观测站网的气象观测资料,时间分辨率为逐日。(2)1979–2018 年 ERA5 再分析数据,时间分辨

收稿日期:2021-11-24

基金项目:国家重点研发计划资助项目(2021YFC3000902-3);国家自然科学基金面上资助项目(42075087);国家自然科学基金区域创新发展联合基金资助项目(U20A2097)

率为逐小时,空间分辨率 0.25° 。(3)1979–2018年NCEP2再分析数据,时间分辨率为逐小时,空间分辨率 2.5° 。(4)1997–2018年TRMM数据。(5)1979–2018年由气象观测资料、ERA5再分析数据融合且再由气象观测资料、TRMM数据订正得到的逐日降雨融合数据(Grid数据)。(6)2008–2018年全流域、1~4区逐日Grid数据面雨量与观测资料面雨量数据。

1.2 研究方法

根据大渡河流域地理特征,将其分为4个区域,从西北到东南依次为一区、二区、三区、四区,如图1所示,为3条黑线所分隔,一区位于大渡河流域的上游,二区主要位于中游,三区和四区位于流域下游,其中四区主要位于大渡河的南侧。

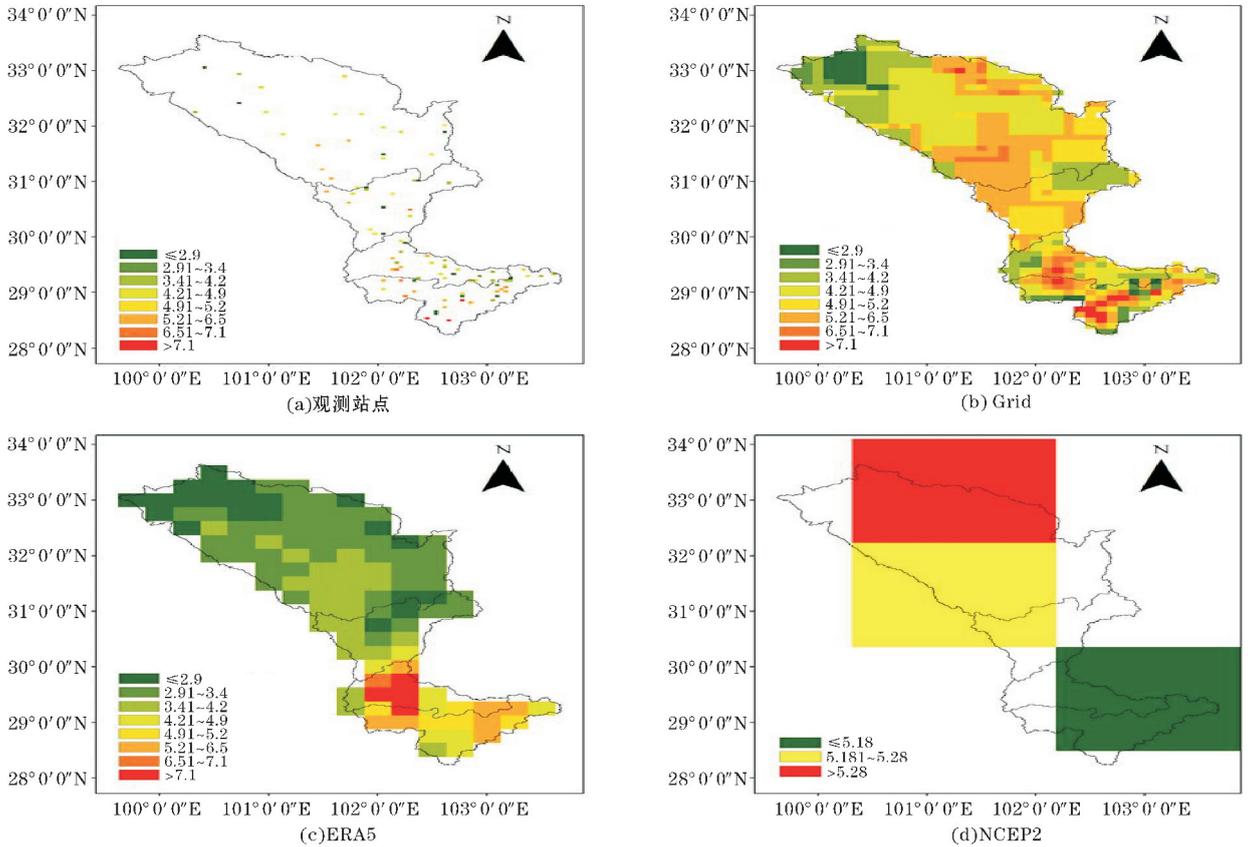


图1 1979–2018年观测站点、Grid、ERA5、NCEP2日均降雨量空间分布

首先进行降雨数据融合得到一套新的完整逐日降雨数据Grid数据;再结合气象站点资料、ERA5和NCEP2格点资料,从日均降雨量空间分布、4个分区日均面雨量、日降雨量均方根误差、日降雨量标准差及全流域、1~4区的面雨量时间序列这5个方面来评估融合资料,与站点资料、ERA5和NCEP2格点资料进行误差对比分析,探讨Grid数据的适用性。

1.2.1 数据融合方法

目前国际上使用较多的格点数据主要为ERA5和NCEP2再分析数据集,因此,首先将大渡河流域观测站点数据与ERA5降雨数据进行融合得到格点数据;然后,由于大渡河流域站点密度较低,局部校正融合方法不适用,因此采用最优插值数据融合方法进行全局校正^[9-10],利用气象观测数据和TRMM卫星反演降雨数据进行校正,形成逐日降雨资料,数据的空间分辨率为 $0.1^\circ \times 0.1^\circ$,即经融合订正后形成的逐日降雨融合数据(Grid数据)。使用的插值方法包括3次样条插值

和反距离权重插值方法。

1.2.2 面雨量计算方法

在某个特定地点所测得的降落到地面的降雨量称作点雨量,气象部门观测和预报的一般是作为站点位置函数的点雨量,而水文部门需要的则是流域范围内的面雨量。流域面雨量可以通过点雨量来计算,表示为

$$\bar{P} = \frac{1}{A_c} \int_c P dC \quad (1)$$

式中, P 是作为连续函数的点雨量, C 是某一特定流域, A_c 为流域面积。

由于降雨观测站的离散性,不可能获得完全连续的点雨量场 P 。实际计算是通过将式(1)离散化来完成的,即先将流域 C 划分为 N 个包含一个雨量观测的小区域,用第 i 个小区域上的点雨量 P_i 代表小区域的平均降雨量,流域的面雨量近似为

$$\bar{P} = \frac{\sum_{i=1}^N C_i P_i}{\sum_{i=1}^N C_i} \quad (2)$$

式中: N 为样本量

1.2.3 误差评估方法

采用标准差和均方根误差来进行误差评估。标准差在降雨数据的评估中是最常用的指标,可作为表征变量变化程度的统计量。标准差定义是各样本距平的算数平方根,它反映样本的平均变化程度,能够反应降雨变化程度的统计量。公式如下:

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (3)$$

式中: x 为变量, \bar{x} 为变量的平均值, N 为样本量。

均方根误差(RMSE)是用来衡量观测值同真值之间的偏差。公式如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (4)$$

式中, x 和 y 为两组变量, N 为样本量。

1.2.4 相关系数

采用皮尔逊相关系数来描述两个变量的线性相关,用 r 来表示相关系数,公式如下:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

式中, x 和 y 为两组变量, \bar{x} 和 \bar{y} 为两组变量的平均值, N 为样本量。

2 日均降雨量空间分布

从图 1 看,观测站点数据显示大渡河流域日均降雨在南部区域即 3 区和 4 区相对偏多,日均最大值可以达到 7.1 mm 以上,而在中北部区域即 1 区和 2 区相对偏少,日均降雨量最小值在 2.9 mm 以下,整体上大渡河流域降雨呈现出自北向南逐渐增加的趋势。ERA5 降雨资料显示,大渡河流域总体上表现出“北低南高”的降雨趋势,但由于该套资料的分辨率仅为 0.25°,因此在每个区域内并不能完全展示实际降雨的分布状况,尤其在 1 区、2 区和 4 区内的降雨量明显偏小,在 3 区内降雨量比观测站点值又偏高;NCEP2 数据同样由于空间分辨率(2.5°)过低,不能完全展现大渡河流域的降雨实况,甚至不能完全覆盖流域范围。

Grid 数据,分辨率达到了 0.1°,基本能完整地展示出大渡河流域的降雨分布,总体上的降雨分布与观测站点基本一致,同样呈现出北部多、南部少的分布状况,在每个区域内的降雨分布和观测站点展示出的降雨量基本吻合。

为进一步评估 Grid、ERA5 和 NCEP2 这 3 套降雨资料与观测站点数据具体差异情况,将 Grid、ERA5 和 NCEP2 日均降雨量数据分别插值到观测站点上,并与观测站点数据进行对比(图 2)。

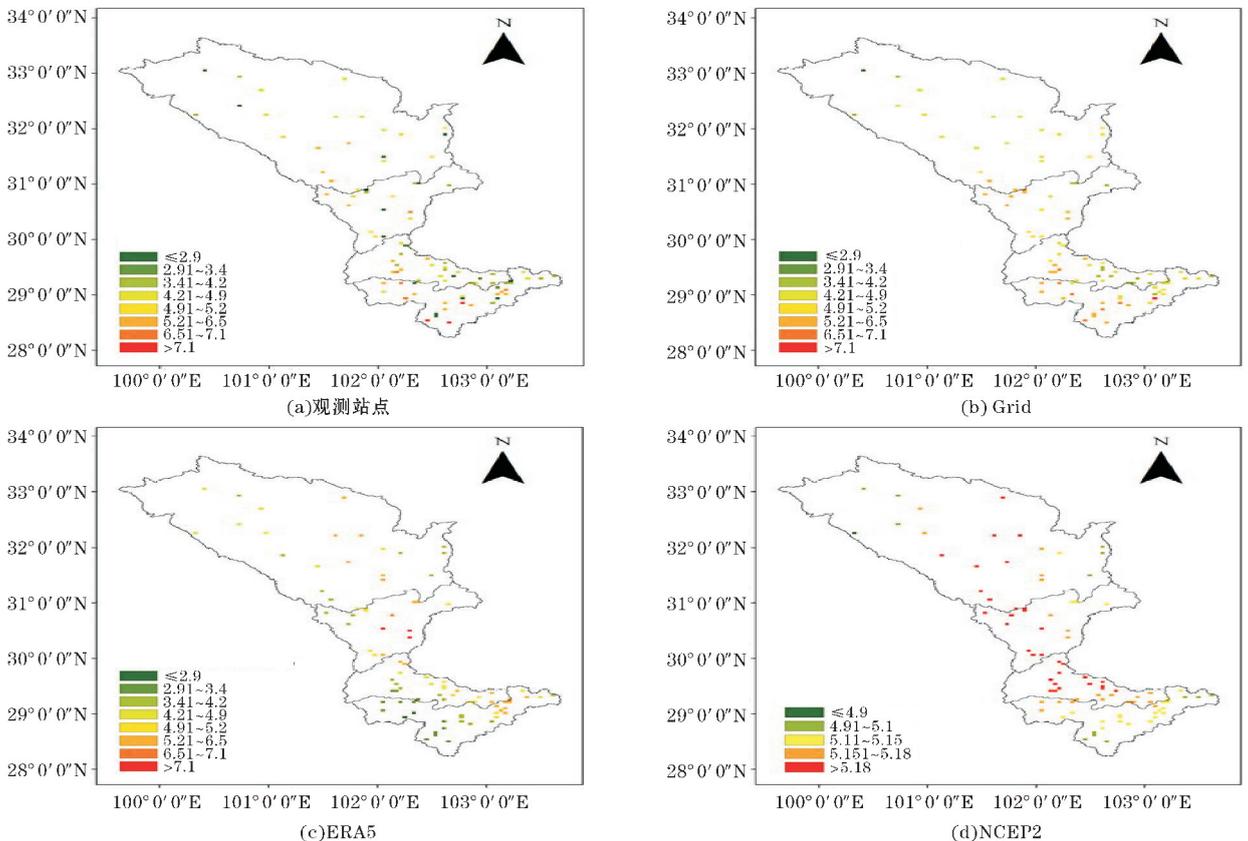


图 2 1979—2018 观测站点及 Grid、ERA5、NCEP2 插值到站点日均降雨量空间分布

从各个分区来看,ERA5 数据插值到站点后的降雨量在 3 区和 4 区明显比观测站点数据偏小,而 NCEP2 数据插值到站点后的降雨量在 3 区和 4 区明显比观测站点偏大;Grid 数据与观测站点相比在 3 区和 4 区相较于 ERA5 和 NCEP2 数据明显更好。在 1 区内,Grid 和 ERA5 插值后的降雨量与观测站点数据比较接近,但 ERA5 数据在 2 区整体上偏高。此外,NCEP2 插值后的数据在 1 区和 2 区均明显比观测站点数据偏高,多数站点日均降雨都在 5.18 mm 以上。

综上,对于 Grid 数据而言,其插值到站点后的数据优于 ERA5 和 NCEP2 数据,降雨量量级与观测站点数据差距小,且整个区域内的降雨量变化趋势与观测站点数据基本一致,在每个分区内的最大降雨量和最小降雨量与观测站点值相对接近。

3 各分区日均面雨量

为进一步量化观测站点数据与 Grid、ERA5、NCEP2 资料在各分区的接近程度,从不同区域的日均面雨量直方图来看(图 3),在 4 个区域内 Grid 数据的日均面雨量均比观测站点偏高,这可能是由于站点分布较为稀疏,导致 Grid 数据中包含相对多的 ERA5 数据信息。同时,ERA5 数据则在 1~3 区内日均面雨量同样比观测站点偏高,但是在 4 区表现相反,明显比观测站点低;NCEP2 数据在各个区域都比观测站点偏高。

1 区内,Grid、ERA5、NCEP2 面雨量呈逐渐增大,其中 Grid 数据与观测站点数据最接近,NCEP2 数据与观测站点的差距最大。2 区内与观测站点差距最大的为 ERA5,但由于融合数据 Grid 加入了站点信息,Grid 数据比 ERA5 更接近观测站点数据。3 区内,ERA5 数据得出的日均面雨量数值最低,且与观测站点最为接

近,但从图 2 日均降雨量空间分布来看,ERA5 雨量在 3 区为“西少东多”,而实际站点为“西多东少”,因此 Grid 站点在 3 区与观测站点更为接近。4 区内,Grid 数据与观测站点数据更接近。

综上,Grid 数据的日均面雨量和观测站点数据最接近。

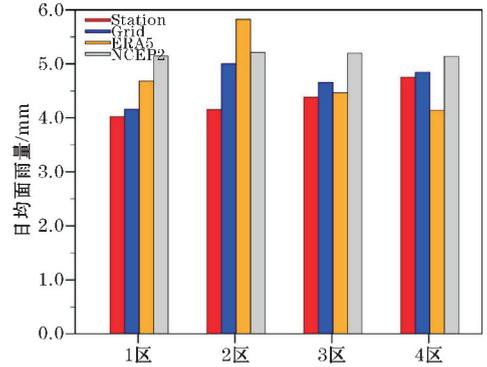


图 3 1979-2018 年观测站点、Grid、ERA5、NCEP2 1~4 区日均面雨量

4 日降雨量均方根误差

以观测站点降雨数据为基准,评估 Grid、ERA5 和 NCEP2 各数据对大渡河流域日降雨量的均方根误差(RMSE)情况(图 4)。总体上看,Grid 数据与站点数据的 RMSE 误差最小。从各个分区来看,在 1 区和 2 区内,Grid 数据绝大部分 RMSE 值很小,误差值在 2.6 以下;ERA5 和 NCEP2 数据误差约在 8.9~12.5,部分站点误差更高。在 3 区和 4 区内,Grid 数据依然表现最好,很多站点的误差值都在 2.6 以下,ERA5 和 NCEP2 数据的表现远弱于 Grid 数据,其绝大部分站点数据的误差值都在 8.9~12.5,甚至部分站点误差值超过 22 以上。因此,从 3 套数据的 RMSE 误差来看,Grid 数据的要明显小于 ERA5 和 NCEP2 数据。

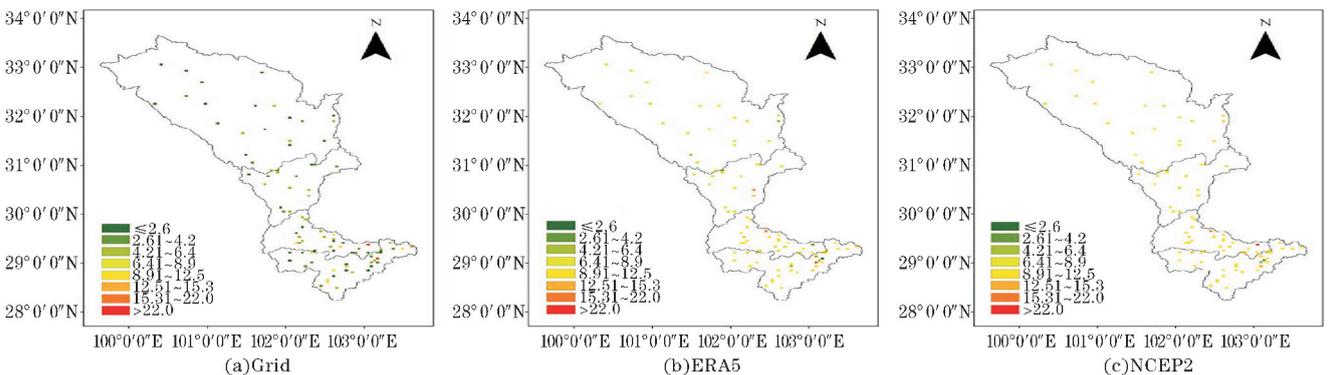


图 4 1979-2018 年 Grid、ERA5、NCEP2 日降雨量插值到站点 RMSE 空间分布

若从 1~4 区的 RMSE 区域平均来看(图 5),Grid 的优势更加明显。1 区、2 区和 4 区的 Grid 数据误差远低于其余两套数据。在 3 区内,Grid 数据的误差虽

有所增加,但相对于 ERA5 和 NCEP2 数据误差而言依然较小。因此在大渡河流域 Grid 降雨资料的准确性更高。

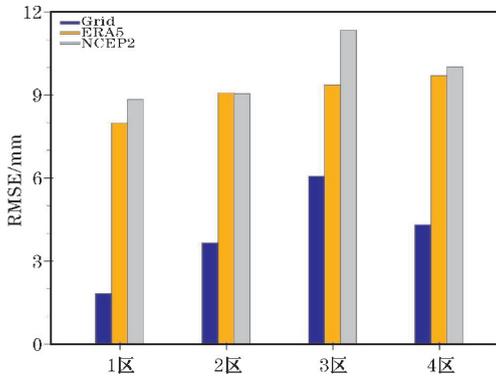


图 5 1979-2018 年 Grid、ERA5、NCEP2 1~4 区 RMSE 区域平均值

5 日降雨量标准差

标准差能够反映变量的平均变化程度。一个较大的标准差代表变量变化幅度较大;一个较小的标准差代表变化幅度较小。从图 6 可以看出,观测站点数据位于北部区域的 1 区和 2 区绝大部分站点标准差在 4.5 以下,而位于南部区域的 3 区和 4 区标准差相对较高,部分站点标准差在 10 以上。这说明 3~4 区降雨的变化程度比 1~2 区的变化程度大。

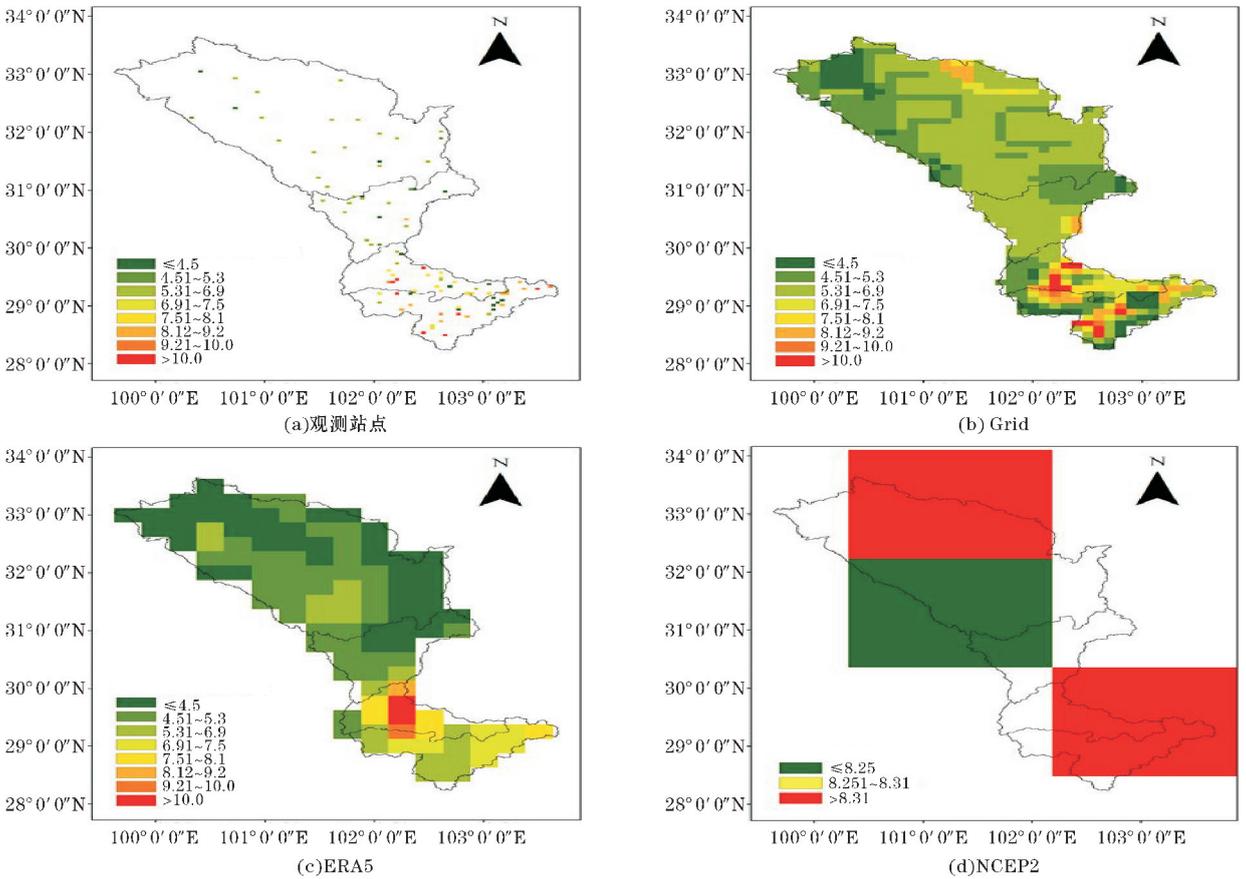


图 6 1979-2018 年观测站点、Grid、ERA5、NCEP2 日降雨量标准差空间分布

将 Grid、ERA5 和 NCEP2 降雨数据的标准差与观测站点数据对比,可以看到,Grid 数据的标准差空间分布特征明显与观测站点数据更接近,1 区和 2 区内,Grid 数据在标准差也基本在 4.5 以下,3 区和 4 区内,标准差也较为接近,这说明各分区标准差与观测站点数据在不同分区内的标准差分布状况基本一致。ERA5 数据在 3 区和 4 区内的标准差明显低于 Grid 数据,大部分在 7.5 以下,但部分区域标准差又较大。对于 NCEP2 数据而言,由于分辨率较低,并不能完全展示不同分区内的降雨标准差,其中更多的降雨细节被忽略。因此,从各套数据的标准差来看,Grid 数据与观

测站点数据最接近。

若将 Grid、ERA5 和 NCEP2 数据的标准差插值到各站点上(图 7),同样可以看出,Grid 数据在各站点上的标准差与观测站点数据的标准差基本一致,即“北部低、南部高”,北部的降雨量变化较小,南部的降雨量变化较大。从 ERA5 和 NCEP2 数据在各站点上的标准差来看,两者在整个大渡河流域的标准差均比较低,与观测站点数据的标准差存在较大差距,EAR5 和 NCEP2 数据对大渡河流域的降雨表现不如高分辨率的 Grid 数据。进一步说明了 Grid 数据与观测站点数据最接近,Grid 数据平均变化程度优于 ERA5 和 NCEP2 数据。

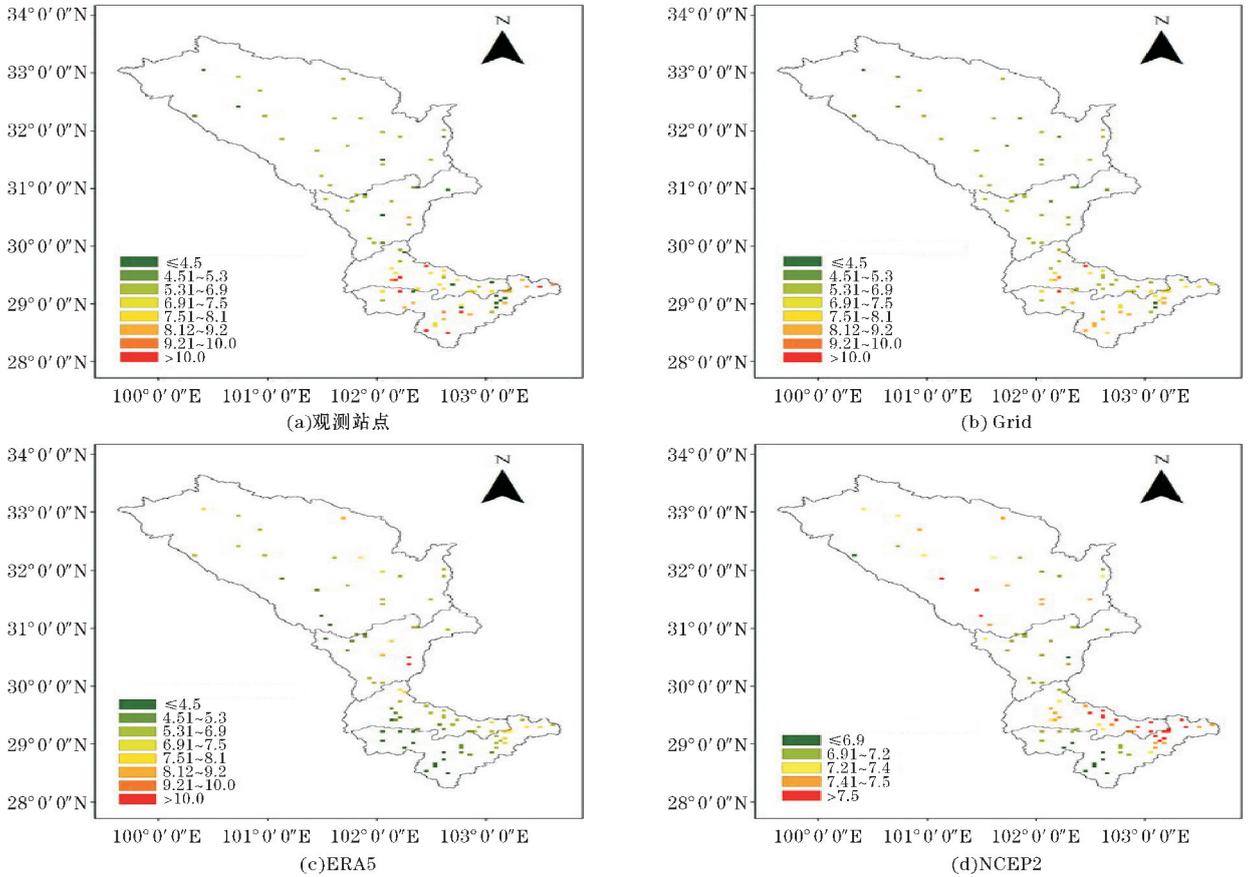


图7 1979–2018年观测站点及Grid、ERA5、NCEP2插值到站点日降雨量标准差空间分布

从不同区域内、不同降雨数据的标准差区域平均直方图来看(图8),1区内,Grid数据的标准差与观测站点数据十分接近;2区内Grid数据也最为接近观测站点数据;3区和4区内,ERA5的标准差明显低于观测站点,Grid的标准差与观测站点数据更接近。

综上,从标准差分析来看,Grid数据总体上、各分区区内与观测站点数据最接近,Grid数据平均变化程度优于ERA5和NCEP2数据。

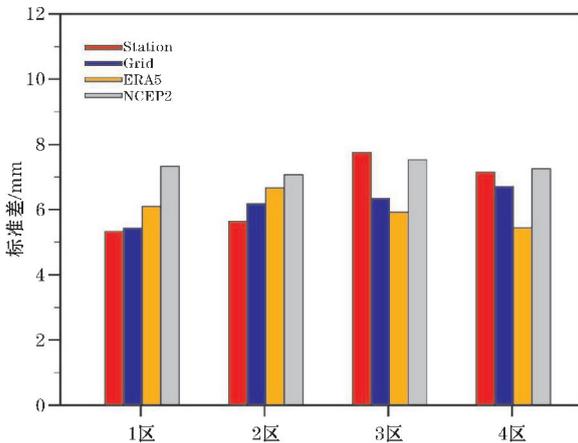


图8 1979–2018年观测站点、Grid、ERA5、NCEP2 1~4区日标准差区域平均值

6 日降雨量近十年时间序列

为进一步对比Grid数据和观测数据,计算了大渡河整个流域和1~4区的面雨量在2008–2018年的逐日时间序列(图略)。用皮尔逊相关系数来表征Grid数据与观测资料的相关性,Grid数据的多个时间序列与观测资料相比,大多数相关系数都在0.9以上。从全流域来看,Grid数据除2008年和观测的相关系数<0.9外,其余年份都>0.9,部分年份>0.95,甚至等于0.99。此外,在降雨较少的1–3月及11月和12月,Grid数据也能较好地展示大渡河全流域和各分区的降雨变化,Grid数据和观测的降雨在降雨较少期也基本吻合。因此,从全流域和1~4分区的面雨量时间序列来看,Grid数据和观测值也比较接近。

7 结论

大渡河流域的复杂地形,导致在该区域内的气象观测站点分布不均,致使其在流域内的降雨无法完全通过常规意义的观测站点观测得到,因此需要精度更高的数据集来展示大渡河流域的降雨状况。对大渡河流域观测站点、Grid、ERA5和NCEP2数据在日尺度上,分别对比评估4套数据在大渡河流域的适用性。其主要结论如下:

(1)Grid数据的分辨率达到 0.1° ,基本能完整地

展示出大渡河流域的日均降雨分布,总体上的降雨分布和观测站点基本一致,在每个区域内的降雨分布和观测站点展示出的降雨量基本吻合。

(2) Grid 数据插值站点的数据优于 ERA5 和 NCEP2 数据,降雨量量级与观测站点数据差距小,且整个区域内的降雨量变化趋势与观测站点数据基本一致。

(3)从标准差来看,Grid 数据在各站点上的标准差与观测站点数据的标准差基本一致,北部的日降雨量标准差小变化较小,南部的日降雨量标准差大变化大。从标准差分布的整体来看,ERA5 和 NCEP2 数据对大渡河流域的降雨表现不如高分辨率的 Grid 数据。

(4)从均方根误差来看,Grid 数据的误差要明显小于 ERA5 和 NCEP2 数据集。

(5)从标准差和均方根误差两方面综合来看,Grid 数据在日均降雨量空间分布、日均面雨量、日降雨量 RMSE 和日降雨量标准差上和观测值比较接近;从全流域和 1~4 分区的面雨量时间序列来看,Grid 数据和观测值也比较接近。总之,Grid 数据能较好地反映大渡河流域的实际降雨情况。

参考文献:

- [1] 张海鹏,智协飞,吉璐莹. 中国区域降雨偏差订正的初步研究[J]. 气象科学,2020,40(4):467-474.
- [2] 曹越,赵琳娜,巩远发,等. ECMWF 高分辨率模式降雨预报能力评估与误差分析[J]. 暴雨灾害,2019,38(3):249-258.
- [3] 潘旸,谷军霞,徐宾,等. 多源降雨数据融合研究及应用进展[J]. 气象科技进展,2018,8(1):143-152.
- [4] 吴启树,韩美,刘铭,等. 基于评分最优化的模式降雨预报订正算法对比[J]. 应用气象学,2017,28(3):306-317.
- [5] 张海鹏,智协飞,吉璐莹. 中国区域降雨偏差订正的初步研究[J]. 气象科学,2020,40(4):467-474.
- [6] He J, Yang K, Tang, W, et al. The first high-resolu-

- tion meteorological forcing dataset for land process studies over China [J]. Scientific Data, 2020, 7(1):1-11.
- [7] 师春香,潘旸,谷军霞,等. 多源气象数据融合格点实况产品研制进展[J]. 气象学报,2019,77(4):774-783.
- [8] 李超,唐千红,陈宇,等. 多源数据融合系统 LAPS 的研究进展及其在实况数据服务中的应用[J]. 气象科技进展,2017,7(2):32-38.
- [9] Yang K, He, J, Tang W, et al. On downward short-wave and longwave radiations over high altitude regions: Observation and modeling in the Tibetan Plateau [J]. Agricultural and Forest Meteorology, 2020, 150(1):38-46.
- [10] Yang K, Chen Y, He J, et al. Development of a daily soil moisture product for the period of 2002-2011 in Mainland China [J]. Science China Earth Sciences, 2020, 63:1-13.
- [11] 赵旋,李耀辉,齐冬梅. 1961-2007 年四川夏季降雨的时空变化特征[J]. 冰川冻土,2013,35(4):959-967.
- [12] 王颖,李栋梁. 变暖背景下青藏高原夏季风变异及其对中国西南气候的影响[J]. 气象学报,2015,73(5):15.
- [13] 殷志强,孟晖,连建发,等. 基于不同时间尺度的地质灾害对气候变化响应研究[J]. 地质论评,2013,59(6):1110-1116.
- [14] 孙莉英,葛浩,庞占龙,等. 长江流域不同类型山洪灾害受自然因素影响分析[J]. 人民长江,2016,47(14):1-6.
- [15] 郭洁,宋雯雯,郑昊,等. 大渡河流域面雨量时空分布特征及雨季转换指标[J]. 干旱气象,2019,37(3):370-376.
- [16] 李艳,张倩倩,陈鲜艳. 大渡河流域水电开发对区域气候的可能影响[J]. 2017,40(1):90-99.

Daily Rainfall Data Fusion Method and its Error Analysis in Dadu River Basin

CHEN Yuan¹, DONG Dandan², SHEN Biao³, CAI Hongke²

(1. Guoneng Daduhe Big Data Service Co., Ltd, Chengdu 610041, China; 2. Chengdu University of Information Technology, Chengdu 610225, China; 3. Chongqing Meitian Technology Co., Ltd, Chongqing 401120, China)

Abstract: In order to effectively combine the advantages of rainfall data from different sources, and develop multi-source rainfall fusion technology, this paper uses data fusion method to generate fused daily rainfall data (hereinafter referred to as Grid data). Then, compare and analyze the error with meteorological station data, ERA5 data and NCEP2 data, in order to explore the applicability of the Grid data in Dadu River Basin. Conclusions: (1) The daily average precipitation distribution of Dadu River basin can be completely displayed by the Grid data. On the overall and within regions, and the precipitation distribution is basically consistent with the precipitation displayed by the meteorological observation data. (2) From the perspective of interpolation to stations, the difference between the rainfall magnitude of the Grid data and the observation data is small, and the overall maximum and minimum rainfall in each region are close to the observation data, which is better than the ERA5 and NCEP2 data; (3) From the standard deviation and the root mean square error, the Grid data is better than ERA5 data and NCEP2 data. (4) From the area rainfall time series of the whole basin and 1-4 divisions, Grid data is closer to observational data than ERA5 and NCEP2 data. In a word, the daily rainfall Grid data can better reflect the actual precipitation of Dadu River Basin.

Keywords: meteorology; meteorological data processing method; data fusion method; rainfall distribution characteristics; Dadu River Basin