

# 宁德市降水量异常点剔除的影响分析

蔡圣准, 肖桂荣

(福州大学福建省空间信息工程研究中心, 福建 福州 350002)

**摘要:**发现异常点是对逐时降水资料质量控制的有效方法,但简单的剔除未修正的异常点必然会影响其周围一定范围内降水量的空间变化。本文提出综合评分法定量的分析不同时段不同异常监测点被剔除后对降水量变化幅度的影响,探讨剔除区域异常点后降水量的时空变化分布特征。对宁德市降水资料进行实例验证,分析结果表明:在相邻站点数大于12个、相邻站点最近距离小于3km、相邻站点平均监测值大于6mm时,其降水量变化幅度相对较小,可剔除异常点。综合考虑三个相邻站点指标的综合分析可以显著表现不同时段不同监测站点降水量空间变化的地带性分布,在宁德市西北和东南区域出现高变化幅度站点的概率最高,呈现出高变化幅度站点多、分布范围广特点;在其东北区域监测站点分布不均匀,降水高变化幅度站点易受逐时降水量大小影响,呈现出小雨大范围、大雨小范围分布;在其西南和中部区域监测站点分布密集,出现高变化幅度站点的可能性最小。因此,综合分析可剔除区域异常点对降水量空间变化的影响特征,确定对降水量变化幅度影响最小的监测站点,有效提高了降水资料的质量控制水平。

**关键词:**异常点;相邻站点指标;降水量;变化幅度;综合评分法

中图分类号:P426.6

文献标识码:A

文章编号:1000-0852(2019)01-0038-06

## 1 引言

从20世纪90年代我国地面自动气象站系统(AWS)建设以来,AWS监测资料在气象中的预警、预报、决策服务等方面逐渐发挥着重要的作用<sup>[1]</sup>。由于受到自动站监测设备、监测环境或数据传输等影响,自动站监测资料质量往往参差不齐<sup>[2,3]</sup>,国内外研究者在AWS资料质量控制方面开展了许多课题研究。为了提高逐时降水资料质量,任芝花等<sup>[4]</sup>在分析数据本身的规律及各种原因引起的错误数据表现形式的基础上结合国家级台站观测的每小时数据,将降水资料质量控制方法分为气候学界限值检查、区域界限值检查、时间一致性检查和空间一致性检查四个部分,可适用于实时和非实时上传的逐小时降水资料自动质量控制;吴泓等<sup>[1]</sup>研究了加入要素匹配一致性的精细化正误判别,即通过建立降水资料与能见度、相对湿度等要素的相关性,检查在一定天气条件下它们是否符合相应的物理联

系,来提高降水资料质量控制;钤伟妙等<sup>[5]</sup>和刘雨佳等<sup>[6]</sup>提出了利用雷达数据和加密自动气象站的方式,将逐小时格点降水资料与对应时刻雷达拼图回波水平分布进行对比,通过排除可疑的异常点来提高质量控制精度。以上研究成果能够较好地检查出逐时降水资料中的异常点,但并未对异常点如何处理进行分析。通常区域降水量是由区域内观测站点空间插值得到<sup>[7,8]</sup>,剔除异常观测点相当于减少观测站,即间接影响区域降水量,若简单剔除检查出的异常点,不进行分析修正处理,必然会影响异常点周围一定范围内降水量的空间变化。

因此本文以宁德市235个降水监测站的逐时降水资料为研究对象,分别采用近邻分析法和空间插值法,通过分析区域降水量变化幅度与异常点不同相关指标之间的关系,定量计算不同时段、不同监测点被剔除后周围降水量变化幅度的影响等级,探讨剔除区域异常点后降水量的地带性分布,得出可剔除且对降

收稿日期:2017-10-12

基金项目:福建省科技重大专项专题(2017NZ0003-1-5)

作者简介:蔡圣准(1991-),男,福建三明人,硕士,研究方向为地理信息网络。E-mail:csz204@qq.com

通讯作者:肖桂荣(1972-),男,福建龙岩人,博士,研究员,研究方向为地理信息系统。E-mail:xiaogr@fzu.edu.cn

水量变化幅度影响最小的监测站点，以期提供降水资料质量控制的有效方法。

## 2 材料与方 法

### 2.1 数据源

本文选取了相邻站点数、相邻站点最近距离、相邻站点平均监测值 3 个研究指标，对宁德市 2017 年 3 月 28 日 16 时~4 月 19 日 15 时共 529 个逐时段的降水监测数据进行数据质量分析，依据《降水等级标准 GB/T28592-2012》，排除监测站（降水量小于 0.1mm）个数占比超过 50% 的时段，筛选出监测站点共有 23 个逐时段，其中从监测站（降水量小于 0.1mm）个数占比小于 10%、10%~20%、20%~30%、30%~40%、40%~50% 五个区间中各选取一个时段作为相邻指标的验证数据，另 18 个时段用于区域降水量变化幅度分析，研究降水量变化幅度与 3 个相邻指标之间的关系。

### 2.2 近邻分析法

通过近邻分析法分析监测站点的邻域情况，计算出一定搜索半径范围内监测站的相邻站点指标。为了有利于结果分析，应使监测站点的邻域站点数分布均匀，从 10km 开始，每增加 5km 作一个搜索半径。如表 1 所示，当搜索半径为 25km 时，包含 1~2 个相邻站点的监测站占比为 0.84%，而包含 16 个以上占比高达 69.27%，相邻站点数分布极不均匀，同样在搜索半径为 20km 和 10km 时，相邻站点数分布也不均匀，而搜索半径为 15km 时，不同相邻站点数的监测站分布相对较均匀，故选 15km 作为本文的邻域搜索半径。

表1 宁德市降水监测站的邻域分析表

Table1 Neighborhood analysis of the precipitation monitoring stations in Ningde city

搜索半径/km	包含不同相邻站点数的监测站占比/%									
	1~2	3~4	5~6	7~8	9~10	11~12	13~14	15~16	16以上	
25	0.84	3.41	0.42	2.13	3.85	2.56	7.69	9.83	69.27	
20	2.98	2.55	4.68	6.81	8.09	12.77	10.64	8.51	42.98	
15	6.38	9.36	13.19	17.02	12.77	17.87	10.64	8.51	3.4	
10	35.32	23.83	17.87	14.04	3.4	0	0	0	0	

### 2.3 空间插值法

确定邻域搜索半径以后，采用具有局部性特征分析的空间插值法<sup>[10]</sup>，对异常点周围一定区域范围内降水量空间变化进行分析，为了防止不同空间插值方法对分析结果产生不同的影响，本文采用了反距离权重

插值<sup>[11]</sup>(Inverse Distance Weighted Interpolation, IDW)、普通克吕金插值<sup>[12]</sup>(Ordinary Kriging Interpolation, OK)和自然邻域插值<sup>[13]</sup>(Natural Neighbor Interpolation, NN)，使分析结果更具合理性。同时也考虑了不同条件下不同空间插值方法的插值精度有所不同<sup>[14]</sup>，在修正时对不同相邻站点数的监测点，采用交叉验证方法进行误差验证，再依据验证精度来确定不同相邻站点数所需选择的空 间插值方法。

### 2.4 相邻站点指标综合评分法

本文提出采用定量的综合评分法用于评判降水量变化幅度等级，即依据降水量变化最大幅度级别对相邻站点指标赋予相应的分值，不同指标分值累加平均得到平均综合评分。平均综合评分越高表示该站点被剔除后对周围区域降水量变化影响的幅度越大，出现大幅度变化的可能性越大。综合评分法计算公式如下：

$$N_{ij} = \begin{cases} 1, & |P_{ij}| \leq 0.3 \\ 2, & 0.3 < |P_{ij}| \leq 0.6 \\ 3, & |P_{ij}| > 0.6 \end{cases} \quad (1)$$

$$M_{ii} = \frac{\sum_{j=1}^m N_{ij}}{m} \quad (2)$$

式中： $N_{ij}$ 、 $P_{ij}$  分别为  $t$  时段的第  $i$  个监测站第  $j$  个相邻站点指标分值、对应的降水变化最大幅度（查表 3 可得）； $m$  为相邻指标个数； $M_{ii}$  为  $t$  时段的第  $i$  个监测站的平均综合评分。其中相邻站点数与相邻站点最近距离仅与监测站所在的空间位置相关，所以在监测站及其周围空间关系未改变的情况下任何时段它们的指标分值都相等。综合评分结果等级划分见表 2。

表2 综合评分等级表

Table2 The comprehensive rating scale

综合评分等级	平均综合评分范围	相邻指标处在劣势区间	出现大幅度降水量变化的可能性
	[1, 1.5)	极少数	几乎没有
	[1.5, 2)	少数	较小
	[2, 2.5)	部分	较大
	[2.5, 3]	绝大部分	非常大

## 3 结果与分析

### 3.1 区域降水量变化幅度分析

首先通过近邻分析法分别计算在 15km 邻域搜索

半径内相邻站点数、相邻站点最近距离、相邻站点平均监测值,再分别采用3种空间插值法,依次计算每个时段每个异常点降水量变化的幅度值。图1表明了相邻站点数与修正前降水量变化幅度之间的关系。降水量变化幅度的波动性随着相邻站点数的增加而逐渐减

小,即变化幅度的稳定性增强。当相邻站点数小于6个时,降水量增幅和降幅分布相对一致、波动性都较大,表明在这个相邻站点数内剔除异常点对周围降水量变化影响较大;当相邻站点数大于12时,变化幅度基本趋于稳定,最大幅度基本维持在 $\pm 20\%$ 左右。

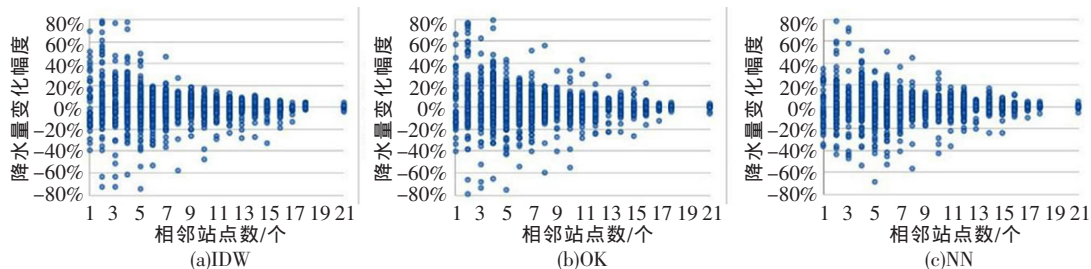


图1 相邻站点数与修正前降水量变化幅度关系

Fig.1 The relationship between the numbers of adjacent stations and the change range of precipitation correction

图2表明了相邻站点平均监测值与修正前降水量变化幅度之间的关系。总体上随着监测值的增加其降水变化最大幅度逐渐下降、稳定性增强,最终维持在 $\pm 20\%$ 左右。从图中可看出,当相邻站点平均监测降水量

处于中小雨级别时小于6mm,其降水量变化幅度波动较大;当相邻站点平均监测值大于6mm时,最大变化基本稳定在 $\pm 20\%$ ,表明当局部发生大雨或大雨以上级别时,剔除异常点对局部的降水变化幅度影响较小。

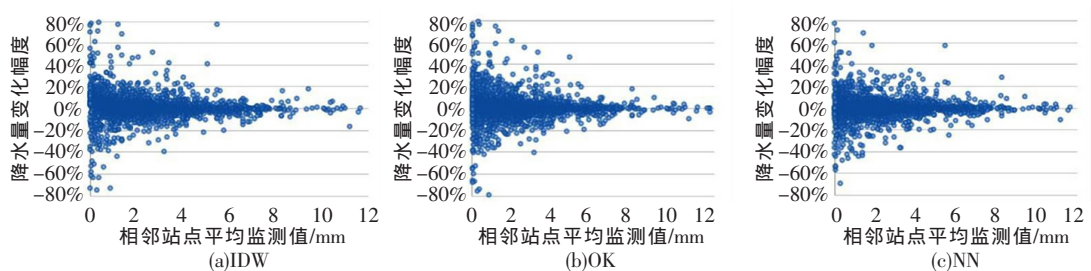


图2 相邻站点平均监测值与修正前降水量变化幅度关系

Fig.2 The relationship between the average values of adjacent stations and the change range of precipitation correction

图3表明了相邻站点最近距离与修正前降水量变化幅度之间的关系。随着相邻站点最近距离的增加,降水量变化的最大幅度呈现双波峰分布。当相邻站点最短距离小于6km时,降水量变化最大幅度随着相邻站点最短距离的增加逐渐变大,波动性增强,表明在站点布设时适当减少相邻站点间的间距可以有效地降低降水量变化幅度;当相邻站点最短距离在6~9km之间时,其波动性明显变小,最大增幅并没有随着趋势上升,反而略有下降;当相邻站点最短距离大于9km时,降水量变化幅度的波动性突然变强。通过近邻分析发现,产生这种现象的主要原因是在6~9km区间内大约有68.35%的监测站其相邻站点数超过7个,在大于9km时超过90%的监测站其相邻站点数低于6

个,也验证了图2中反映的降水量空间变化规律。

考虑到对降水变化幅度小的监测站点进行修正意义不大,故选择修正前降水变化幅度相对较大的监测站,依次对不同相邻指标进行比较分析,如图4所示,修正前后降水量变化幅度总体趋势相同;在相邻站点数小于5个、相邻站点数最近距离大于9km、相邻站点平均监测值小于2mm时修正效果较为明显;在降水量变化幅度最大区域内同时存在具有相邻站点数少、相邻站点最近距离远及相邻站点平均监测值小的监测站,均出现修正后降幅提高现象,说明对三种相邻指标同时处于最劣状态的异常点进行修正不一定都能取得较好的修正效果。相邻站点指标对区域降水变化幅度的影响分析结果见表3。



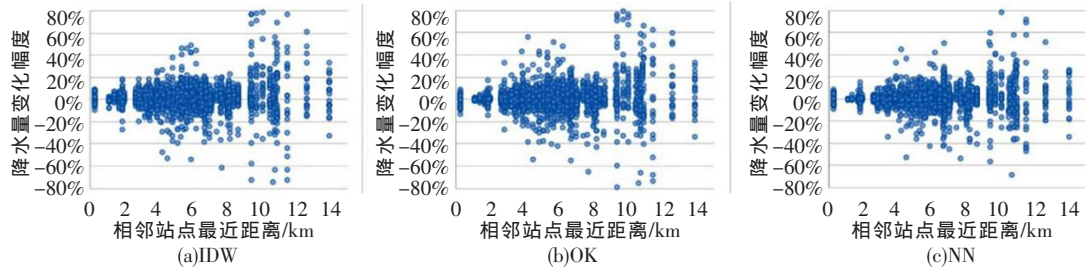


图3 相邻站点最近距离与修正前降水量变化幅度关系

Fig.3 The relationship between the shortest distance of adjacent stations and the change range of precipitation correction

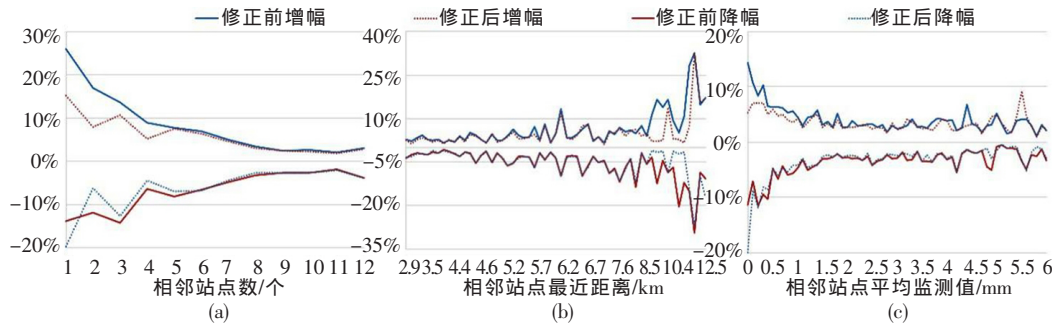


图4 不同相邻指标与修正前后降水量平均变化幅度关系

Fig.4 The relationship between the change of precipitation and the average variation of precipitation correction

表3 相邻站点指标对区域降水变化幅度的影响分析结果表

Table3 The influence of the adjacent site indexes on the regional precipitation change range

相邻站点指标	降水变化最大幅度 (P)			修正效果	
	$ P  \leq 0.3$ (较小)	$0.3 <  P  \leq 0.6$ (中)	$ P  > 0.6$ (较大)	变化不大	相对明显
相邻站点数/个	>12	(5,12]	(0,5]	(8,12]	(0,8]
相邻站点最近距离/km	(0,3]	(3,9]	>9	(3,8]	(8,12]
相邻站点平均监测值/mm	>6	(2,6]	[0,2]	(4,6]	[0,4]

3.2 降水量空间变化地带性分布

通过相邻站点指标综合评分法计算可得，综合考

虑三个相邻站点指标,能更好地表现逐时降水量变化幅度的地带性分布特征,如图5所示。对选取的5个时段相邻指标的验证数据分别计算准确率和检出率,综合检验结果见表4。由表4可知,综合评分高等级的站点包含了大部分高变化幅度的站点,验证了三个相邻指标的综合评分方法的有效性,即综合评分高等级出现大幅度变化的可能性越大,检出率受高变化幅度站点数的影响较明显。从宁德市逐时降水量变化幅度的空间分布来看,其西北山脉和东南堆积平原地带的监测站点相对较少、分布离散,在中小雨时段出现高变化幅度概率极大,呈现出高变化幅度站点多、分布范围广特点;其东北丘陵-平原地带监测站点数量较

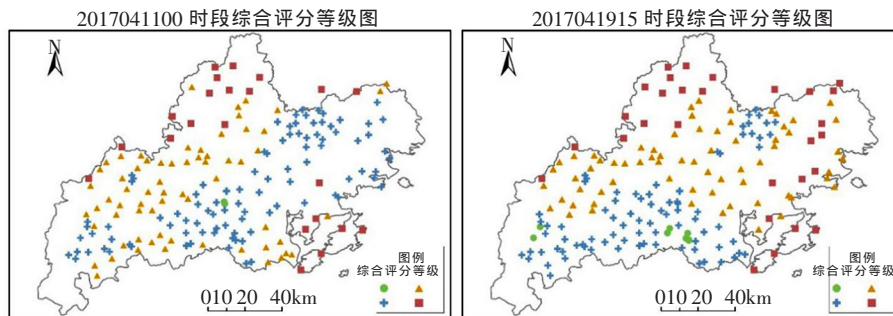


图5 三个相邻指标的不同时段不同站点综合评分等级图

Fig.5 Three different indicators of different periods of the various sites comprehensive score levels

表4 三个相邻指标综合检验结果  
Table4 Three adjacent indicators comprehensive test results

时段	A 降水变化幅度 $P > 10\%$ 的个数	B 综合评分等级 $L \geq 3$ 级的个数	C 降水变化幅度 $P > 10\%$ 且综合 评分等级 $L \geq 3$ 级的个数	检验指标/%	
				准确率(C/A)	检出率(C/B)
2017040712	27	77	20	74.07	25.97
2017041022	56	100	46	82.14	46
2017041100	55	88	39	70.9	44.32
2017041200	21	93	16	76.19	17.21
2017041915	39	96	35	89.74	36.46

多但分布密度极不均匀,部分离散的站点极易出现高变化幅度情况,降水高变化幅度站点分布易受逐时降水量大小影响,呈现出小雨大范围、大雨小范围分布;其西南丘陵-山区和中部河谷盆地-山区地带由于监测站点较多且分布密集,出现高变化幅度站点的可能性最小,呈现出小范围零星分布。其中部受到特殊的地形地貌条件影响,如位于山间河谷与两岸的相邻站点、背风站点及迎风站点,表现出综合评分低等级、高变化幅度的情况,可能存在奇异点。

#### 4 结论

本文选取相邻站点个数、相邻站点最近距离、相邻站点平均监测值三个指标,采用近邻分析、空间插值和综合评分法定量分析不同时段不同异常监测点被剔除后对降水量变化幅度的影响,直观地展现出逐时降水量空间变化的地带性分布,可为逐时降水资料质量控制提供参考。以宁德市降水资料实例分析表明:(1)在相邻站点较多、相邻站点最近距离较近或相邻站点平均监测值较小时,剔除异常点前后其降水量变化幅度的波动性相对较小。当局部发生小雨或中雨级别时,剔除异常点对局部的降水变化幅度影响较大。(2)通过近邻分析可知,宁德市降水监测站点适宜的邻域搜索半径为15km,在此邻域搜索区域内,相邻站点数小于8个、相邻站点最近距离在8~12km、相邻站点平均监测值小于4mm的异常点修正效果相对明显。(3)通过三个相邻站点指标的综合分析可以显著表现不同时段不同监测站点降水量空间变化的地带性分布。综合评分等级越高其局部降水量出现大幅度变化的可能性越大,宁德市西北山脉和东南堆积平原地带出现大幅度变化站点的概率最高,呈现出数量多、分布广;其东北丘陵-平原地带次之,呈现出小雨大范围、大雨小范围

分布;其西南丘陵-山区和中部河谷盆地-山区地带最小,呈现出小范围零星分布。另外,宁德市中部受到特殊的地形地貌条件影响,出现了综合评分低等级、高变化幅度的奇异点情况,为了取得更好的分析结果,还需要考虑加入地形因子等其他相关指标进行更深入的研究。

参考文献:

- [1] 吴泓,袁成松,钱玮,等. 气象监测中降水资料的质量控制[J]. 气象科学, 2012,32(6):659-664. (WU Hong, YUAN Chengsong, QIAN Wei, et al. Control of precipitation data in the meteorological monitoring [J]. Journal of the Meteorological Sciences, 2012,32(6): 659-664. (in Chinese))
- [2] Su Y, Tan L. Control strategy in realtime data transmission of automatic weather station [J]. Meteorological Hydrological & Marine Instruments, 2009.
- [3] Estévez J, Gavilán P, Giráldez J V. Guidelines on validation procedures for meteorological data from automatic weather stations [J]. Journal of Hydrology, 2011,402(1-2):144-154.
- [4] 任芝花,赵平,张强,等. 适用于全国自动站小时降水资料的质量控制方法[J]. 气象, 2010,36(7):123-132. (REN Zhihua, ZHAO Ping, ZHANG Qiang, et al. Quality control procedures for hourly precipitation data from automatic weather stations in China [J]. Meteorological Monthly, 2010,36(7):123-132. (in Chinese))
- [5] 钤伟妙. 江淮地区2007年梅雨期雷达拼图数据和加密自动气象站降水资料的处理分析及初步应用[D]. 北京:中国气象科学研究院, 2012. (QIAN Weimiao. Analysis, Evaluation and Utilization of 3D Mosaic Reflectivity from Ground-based Radars and Hourly Precipitation Data from Automatic Weather Stations at Yangtze-Huai River Basin during the Mei-Yu Season of 2007 [D]. Beijing: Chinese Academy of Meteorological Sciences, 2012. (in Chinese))
- [6] 刘雨佳,陈洪滨,金德镇,等. 加密自动气象站雨量计资料的质量控制及其相关关系的研究[J]. 大气科学, 2014,38(1):159-170. (LIU

- Yujia, CHEN Hongbin, JIN Dezhen, et al. Quality control and representativeness of automatic weather station rain gauge data [J]. Chinese Journal of Atmospheric Sciences, 2014,38(1):159-170. (in Chinese))
- [7] Chen D, Ou T, Gong L, et al. Spatial interpolation of daily precipitation in China: 1951 - 2005 [J]. Advances in Atmospheric Sciences, 2010,27(6):1221-1232.
- [8] 李艳,朱军,胡亚,等. 不同插值方法模拟四川省逐月降水量的对比分析[J]. 水土保持研究, 2017,(1):151-154. (LI Yan, ZHU Jun, HU Ya, et al. Comparison analysis on different spatial interpolation methods to simulate monthly precipitation in Sichuan province [J]. Research of Soil and Water Conservation, 2017,(1):151-154. (in Chinese))
- [9] ArcGIS. ArcGIS Resources [EB/OL]. <http://resources.arcgis.com/zh-CN/help/main/10.2/>
- [10] 邬伦,吴小娟,肖晨超,等. 五种常用降水量插值方法误差时空分布特征研究——以深圳市为例[J]. 地理与地理信息科学, 2010,26(3): 19-24. (WU Lun, WU Xiaojuan, XIAO Chenchao, et al. On temporal and spatial error distributions of five precipitation interpolation models: a case of Shenzhen [J]. Geography and Geo-Information Science, 2010,26(3):19-24. (in Chinese))
- [11] 董志南,郑拴宁,赵会兵,等. 基于空间插值的风场模拟方法比较分析[J]. 地球信息科学学报, 2015,17(1):37-44. (DONG Zhinan, ZHENG Shuanning, ZHAO Huibing, et al. Comparative analysis of methods of wind field simulation based on spatial interpolation [J]. Journal of Geo-Information Science, 2015,17(1):37-44. (in Chinese))
- [12] Dai F, Zhou Q, Lv Z, et al. Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan plateau [J]. Ecological Indicators, 2014,45(5): 184-194.
- [13] 许雯,董美莹,陈锋. 基于逐时降水站点资料空间插值方法对比研究[J]. 气象与环境学报, 2017,33(1):34-43. (XU Luan, DONG Meiyang, CHEN Feng. Comparison study of spatial interpolation methods based on hourly precipitation data from automatic weather stations [J]. Journal of Meteorology and Environment, 2017,33(1):34-43. (in Chinese))
- [14] 郑小波,罗宇翔,于飞,等. 西南复杂山地农业气候要素空间插值方法比较[J]. 中国农业气象, 2008,29(4):458-462. (ZHENG Xiaobo, LUO Yuxiang, YU Fei, et al. Comparisons of spatial interpolation methods for agro-climate factors in complex mountain areas of southwest China [J]. Chinese Journal of Agrometeorology, 2008,29(4): 458-462. (in Chinese))

## Spatial Variability Analysis of Precipitation Anomaly Point Removal in Ningde City

CAI Shengzhun, XIAO Guirong

(Space Information Research Center, Fuzhou University, Fuzhou 350002, China)

**Abstract:** To find anomaly point is an effective method for controlling the quality of hourly precipitation data. However, simply removing the uncorrected anomaly point will inevitably affect the spatial variation of the precipitation within a certain range. This paper proposed a comprehensive scoring method to quantitatively analyze the influence of different anomaly monitoring points removal on the variation of precipitation during different durations, and discussed the temporal and spatial distribution of precipitation after anomaly point removal. The case study of Ningde precipitation data shows that the variation of precipitation is relatively small when the number of neighbors is more than 12, the nearest neighbor distance is less than 3km, and the average monitoring value of adjacent sites is greater than 6mm. The comprehensive analysis of three adjacent stations can significantly show the spatial distribution of precipitation at different monitoring stations in different durations. The probability of occurrence of high-variation sites in the northwest and southeast of Ningde is the highest, showing a high variation range distributed in a wide range of areas. In the northeast, the monitoring sites are unevenly distributed, and the site of high variation in precipitation is susceptible to the size of the hourly precipitation, showing a wide range of light rain, and heavy rain distribution in small areas. In the southwest and center, the monitoring sites are densely distributed, the possibility of a site showing a high degree of change is the smallest. Therefore, a comprehensive analysis can eliminate the influence characteristics of the regional anomaly points on the spatial variation of precipitation, and determine the monitoring stations that have the least impact on the variation range of precipitation, and effectively improve the quality control of precipitation data.

**Key words:** adjacent site index; precipitation; change amplitude; comprehensive score method