
基于 Delaunay 三角网和地理加权回归的降水空间异常探测方法研究^{*1}

刘洋¹ 邓敏¹ 邓悦² 杨学习¹

(1. 中南大学地理信息系, 湖南长沙 410083;

2. 中国测绘科学研究院, 北京 100039)

【摘要】: 针对目前已有的降水资料质量控制方法适应性差, 降水数据难以模拟局地降水规律的问题。以湖南省为实验区域, 利用 1985 ~ 2014 年夏季降水观测数据, 定义局部地理空间内降水量发生显著偏离的气象站点为异常, 发展了一种结合 Delaunay 三角网和地理加权回归的降水空间异常探测方法。该方法通过构建不规则三角网来获取各个区域之间稳定合理的地理邻近关系, 根据空间邻近关系获取自适应回归带宽, 并纳入地理加权回归模型进行计算, 最后根据回归残差项的统计分布性质开展空间异常度量与分析。实验表明该方法能有效检测出降水空间异常, 可应用于降水数据二次处理时的质量控制阶段, 有助于进一步探索局部地理空间内的降水规律。

【关键词】: Delaunay 三角网; 地理加权回归; 自适应带宽; 降水空间异常检测

【中图分类号】: P208 **【文献标识码】**: A **【文章编号】**: 1004-8227(2017)11-1865-10

DOI: 10. 11870 /cjlyzyyhj201711015

降水是水循环过程中的重要环节, 降水量的多寡与洪涝、干旱等自然灾害的形成有着密切关系。研究降水变化规律, 有助于提高应对极端灾害的能力^[1]。降水信息作为流域水文模拟的关键驱动, 与其他气象要素的差别就在于其中小尺度变异特性。因此, 探索局部区域下的异常天气对于灾害预警、水文模拟等有着重要意义。我国夏季降水的主要来源为受季风驱动的暖湿气流, 而从水气的蒸发、输送, 再到凝结降雨的整个过程中, 经纬度和地形等地理要素扮演了重要角色^[2, 3]。

范泽孟等^[4]通过高精度曲面建模理论指出全国尺度的站点平均降水量与纬度和海拔之间存在很好的复相关性, 但是与各个单因子间的相关性较弱。多尺度下局部地形因素与大气之间的作用规律十分复杂, 这导致地形影响降水的动力、热力、微物理

¹ 收稿日期: 2017-02-18; 修回日期: 2017-03-26

基金项目: 国家重点研发计划项目(2016YFB0502303) [National Key Research Projects of China(2016YFB0502303)] ; 国家自然科学基金项目(41471385) [National Natural Science Foundation of China(41471385)] ; 中南大学中央高校基本科研业务费专项资金(2016zzts085) [Special Fund or Basic Scientific Research of Central University of Central South University(2016zzts085)]

作者简介: 刘洋(1993 ~), 男, 硕士研究生, 主要研究方向为时空数据挖掘、异常检测与应用. E-mail: oceanflow@yeah. Net

*** 通讯作者 E-mail**: dengmin208@tom. com

效应也变得十分复杂。例如, Smith 等^[5]的研究表明中小尺度地形通常对降水中心的落点、落区有显著影响, 这正是导致局部异常天气产生的一个主要原因。为了进一步探索局部地区降水规律, 水文工作者需要获取高质量的降水资料。传统的降水资料质量控制手段主要是根据历史观测资料统计特征以及气象学、天气学等经验公式来分析观测数据是否合理^[6]。但是由于时间推移, 社会经济发展, 气象自动站的密度不断增加, 新站缺乏历史资料, 传统方法此时表现出很大的局限性。因此, 迫切需要一种适用性更强的降水资料质量控制检验方法。

从地理信息科学的角度出发, 可以利用空间分析的方法对降水信息进行检验, 目前大多采用多元回归分析的方法进行研究。例如, 张杰等^[7]用逐步回归的方法建立了青藏高原北坡降水与地形之间的关系模型, 王远见等^[8]利用偏最小二乘法克服了回归自变量之间的共线性, 建立了雅鲁藏布江中下游地区湿季降雨量的空间分布模型, 然而在空间分析中, 观测变量往往会因为空间位置的改变而导致变量间关系与结构发生变化, 这种关系称之为空间非平稳性^[9]。目前, 利用逐步回归或者偏最小二乘回归对降水分布的研究并未充分考虑此特性, 其得到的结果仍然是全局性的, 因而无法反映降水事件的局部地理特征。Brunsdon 等^[10]提出的地理加权回归理论有效解决了空间数据非平稳性的问题。该方法在每个回归点上的参数估计方程都是不同的。因此, 它允许在不同的局部地理空间中存在不同的空间关系, 从而解决了空间数据的非平稳性问题^[11]。但是普通地理加权回归对于带宽选择敏感, 尤其是在数据点分布疏密不均的情况下, 拟合效果较差。因此, 亦存在着一定的缺陷。

异常探测是数据挖掘的研究方向之一。旨在挖掘出与正常模式相比发生显著偏离, 以至于令人怀疑是由不同机制产生的特殊事件或者规律^[12~14]。在气象与水文等领域, 异常检测往往应用于极端气候, 自然灾害以及突发气候事件等场景。综上所述, 本文定义局部地区内降水量发生显著偏离的站点为异常, 发展了一种基于 Delaunay 三角网和地理加权回归的降水空间异常检测方法。

1 研究区域与数据来源

湖南省地处我国腹地, 全省地势东高西低, 处于云贵高原与江汉平原的过渡地带, 地质条件复杂, 地形多样, 水系密布。因其独特的地理环境和气候特点, 全省降水空间分布极不均衡。近年来, 湖南省气象条件突变严重, 全省各地遭受洪涝与干旱灾害的频次与强度呈现出上升的趋势, 且两种自然灾害大多都发生在夏季^[15], 已成为制约当地社会发展的重要因素。因此, 本文选择湖南省作为实验区域, 对其进行降水空间异常检测研究。

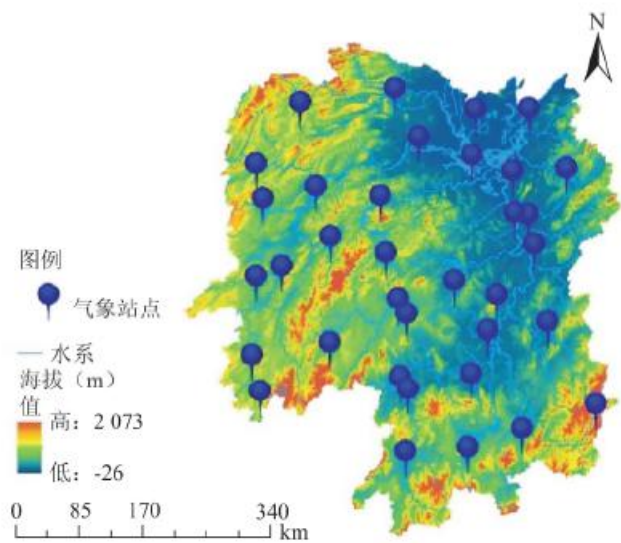


图1 研究区域地理位置与气象站点分布图

Fig. 1 Location of Study Area and Meteorological Site Map

实验数据分别为：①中国气象科学数据共享网(data. cma. cn) 提供的湖南省 1985 ~ 2014 年 30a 的月值气象观测资料。本文采用每年夏季6 ~ 8月的月降水量均值进行实验分析；②中国科学院地理空间数据云(www. gscloud. cn) 提供的 srtm90m 分辨率 DEM 数据。

2 研究方法

所采取的研究方法的主要分为 3 个部分：①将原始要素按照降雨形成机制进行组合变换。考虑到尺度效应，本文获取了多尺度下的变量值，并采用主成分分析法消除共线性，提取主要成分作为回归变量； ②采用地理加权回归理论解决空间数据非平稳性，并以约束三角网的邻近关系作为回归带宽的选择依据，从而可以获取自适应的最优带宽，有效克服了数据点分布形式对回归结果的影响； ③残差项服从独立同正态分布，因此，可根据残差项的统计分布特征来识别空间异常。算法流程如图 2 所示。

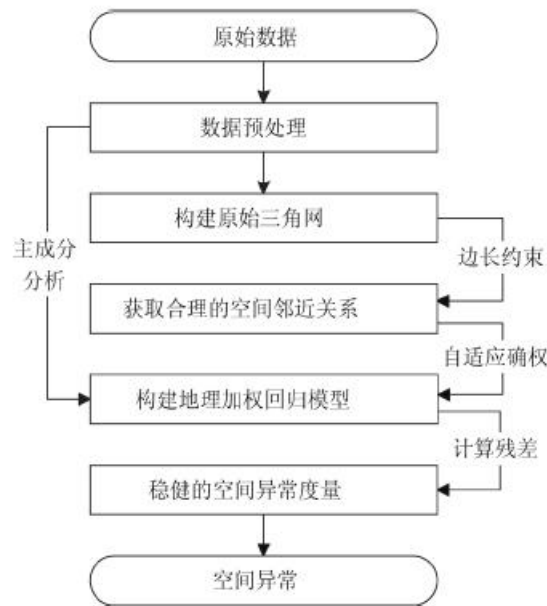


图 2 算法流程图

Fig. 2 Algorithm Flow Chart

2. 1 影响降水的因素

通常认为，地形对暖湿气团的运动有明显的摩擦和阻隔作用，同时，随着海拔增高，气温降低，水汽越发容易凝结，进而导致降雨。然而，各个要素对于降水的影响机制是复杂的，因此，从中小尺度下地形降水的产生机制分析，除了经纬度，海拔之外，坡度坡向也是重要的影响因素。为了表达坡度坡向的空间尺度及其与大气相互作用。本文根据已有研究将部分要素按照对应的耦合关系，对相关因子进行组合变换，并以新的复合变量代替单一的影响因素纳入研究。

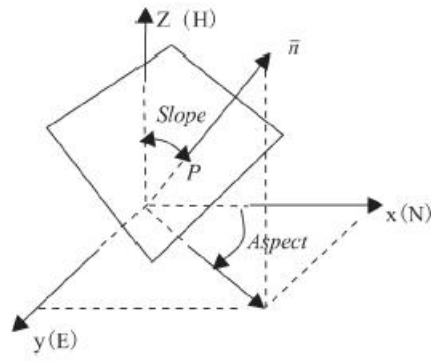


图 3 坡度坡向示意图
Fig. 3 Slope and Aspect Diagram

傅抱璞等^[16]对于坡度、坡向和风向对降水影响做了大量的研究，并建立了地形因素与降水量的参数化半经验模型。在此基础上，史岚等^[17]研究发现地形是影响暖湿气流流通的重要因素。以上研究表明，当迎风坡的坡度为 45° 左右且盛行风向与坡向夹角越接近于 0° 时，地形对气团抬升作用越大，对降雨促进作用越明显，背风坡作用则相反。借鉴上述研究，分别以坡向指数 Indasp 和坡度指数 Indslp 来表达坡度 θ 、坡向 α 和风向 ω 对降雨的影响，详细参数如表 1 所示。

表 1 影响因子

变量名称	类型	说明
降雨量	因变量	毫米 (mm)
经度	自变量	度 ($^\circ$)
纬度	自变量	度 ($^\circ$)
高程	自变量	米 (m)
坡向指数	自变量	$\text{Indasp} = \cos(\omega - \alpha)$
坡度指数	自变量	$\text{Indslp} = \sin(\theta)$

2. 2 构建合理空间邻域

地理学第一定律指出：一个空间单元内的信息与其周围单元信息有相似性，空间单元之间具有连通性^[18]。值得注意的是，Delaunay 三角网的唯一性和空圆特性可以很好地表达实体间的邻近关系。但是，原始三角网中往往存在着一部分误差长边，这类边所连接的实体并不是真实的局部邻近关系。因此，我们采取了一种约束策略对三角网进行修剪，剔除三角网中不合理的长边。并以约束后的 Delaunay 三角网来构建实体间的相邻关系和刻画各个地理实体之间的局部相似性。

定义 1 稳健的平均边长约束阈值 EC：给定空间点数据集并建立 Delaunay 三角网，假设 E 为所有边的集合，则将集合 E 内所有边按长度进行升序排列，取上四分位数 Q1 和下四分位数 Q3 之间的边长值均值 RAE，再乘以边长调节系数 α 即为边长约束阈值 EC，如下式所示：

$$RAE(E_i) = \frac{\sum_{i=1}^n E_i}{n}, Q1 \leq E_i \leq Q3$$

$$EC = \alpha * RAE \quad (2)$$

α 取值范围可通过启发式策略确定, 本文通过大量实验发现, α 取值 $[2, 3]$ 时较为合适。通过打断边长大于 EC 的不一致边, 获取符合认知规律的空间实体邻接关系。如图 4 所示, α 取值为 2 时模拟数据实体间邻接关系表达。

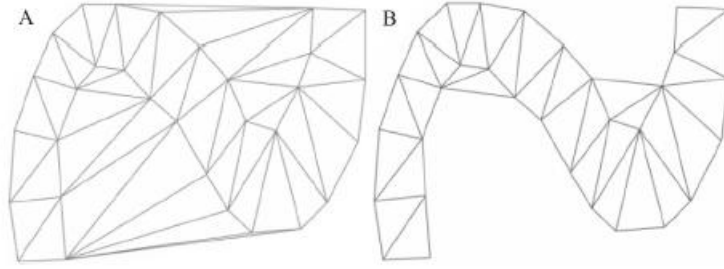


图 4 模拟数据原始三角网(A)和约束后三角网(B)
Fig. 4 Original Delaunay Triangulation and Its Constrained Result Using Simulate Data

定义 2 邻域半径 $R(P_i)$: 任意一点 P_i 直接相连的点为一阶邻域, 与其一阶邻域相连的点为二阶邻域。若给定数据点 P_i 的二阶邻域点数量为 m , 则定义其邻域半径 $R(P_i)$ 为该点二阶邻域内所有边长之均值 $Mean^2 P_i()$ 与所有一阶邻域点 P_j 的局部边长标准差 $Std(P_j)$ 均值之和。

$$R(P_i) = Mean^2(P_i) + \frac{\sum_{j=1}^m Std(P_j)}{m},$$

$$1 \leq i \leq n \quad (3)$$

2. 3 构建地理加权回归模型

假设数据点数量为 n , 自变量个数为 p , 若以 y_i 表示因变量, x_{ij} 表示自变量, $\beta_j(u_i, v_i)$ 表示每个回归点上的回归参数(u 和 v 代表经度和纬度, β_0 代表常数项), ε_i 代表残差项($\varepsilon_i \sim N(0, \sigma^2)$, $cov(\varepsilon_i, \varepsilon_j) = 0 (i \neq j)$), 则可以将 GWR 模型表示为:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^p \beta_j(u_i, v_i) x_{ij} + \varepsilon_i,$$

$$i = 1, 2, \dots, n, j = 1, 2, \dots, p \quad (4)$$

通过计算残差项的最小二乘平方和可以求出待求点的回归参数 $\beta_j(u_i, v_i)$ ，进而得到回归值 $\hat{y}_i$ 。以矩阵形式所描述的地理加权模型参数估计方程为：

$$\hat{\beta}_i = (X^T W_i X)^{-1} X^T W_i Y \quad (5)$$

其中：

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (6-7)$$

$$W = \begin{bmatrix} w_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{in} \end{bmatrix}, \hat{\beta}_i = \begin{bmatrix} \hat{\beta}_{i0} \\ \vdots \\ \hat{\beta}_{ip} \end{bmatrix} \quad (8-9)$$

地理加权回归模型的核心内容是通过地理实体的空间分布关系来确定影响力的权重，这使得每个回归点上的参数估计都是空间相关的，因而解决了空间非平稳性的问题。因此，如何准确合理地确定空间权重便成为问题的关键。而权重矩阵产生于确权函数，目前常用的确权函数有高斯函数法，截尾函数法等，但是上述方法对于函数带宽的选择尤为敏感，且无法适用于数据点疏密不均的情况。为了解决传统确权函数自适应性差的问题，本文采用一种自适应确权函数—排序确权法。以落入回归点邻域半径内的点为其邻居，并取最远邻居点的距离作为带宽 b ，如图 5 所示，固定研究区域内，带宽 b 会在数据点稀疏的区域自动放大，在数据点密集的区域自动缩小，从而保证了带宽 b 不受数据点空间分布的影响。

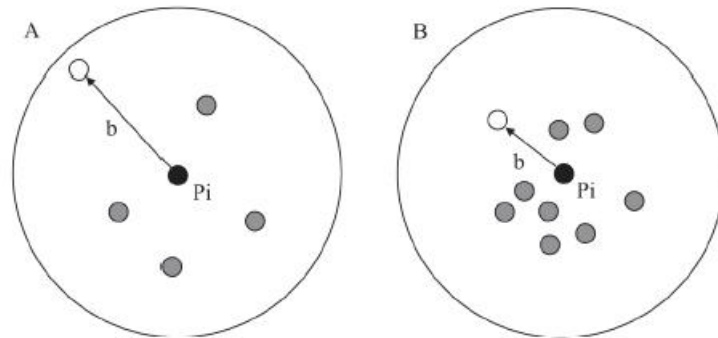


图 5 数据点稀疏时带宽(A)和数据点密集时带宽(B)

Fig. 5 Diagram of Wide B and width in Data Points
Sparse Area (A) and Narrow Bandwidth in
Dense Data Points Area

2. 4 稳健的空间异常度量

异常的定义是异常探测的核心内容，地理加权的拟合值是局部光滑且连续的，若偏离此模式则可认为是异常。残差项是十分重要的一个统计量，可以揭示模型基本假设的大部分信息^[19]。因为残差项服从独立同分布，所以可以使用 k 倍标准差准则对其进行检验， k 取 3 时，包含 99.7% 的残差分布^[20]，因此认为剩余的 0.3% 为少数的，显著偏离实际的异常点。但是，由于站点数量的限制，样本规模过小。直接使用 k 倍标准差可能导致异常检测的稳健度降低，因此，采用一种稳健的度量方式^[21]，以残差序列的中位数 $Median(\varepsilon_i)$ 代替其均值，中位数绝对偏差 $MAD(\varepsilon_i)$ 代替其方差。

定义 3 数据集内任意点 P_i 的残差 ε_i ：回归值与观测值之差：

$$\varepsilon_i = |y_i - \hat{y}_i| \quad (10)$$

定义 4 空间异常识别阈值 k ：给定数据集内任意一点 P_i ，则其异常识别阈值为 P_i 的残差项 ε_i 减去残差序列的中位数 $Median(\varepsilon_i)$ 再除以其绝对偏差 $MAD(\varepsilon_i)$ ，若 k 值超过 3，则将 P_i 识别为空间异常。

$$k = \left| \frac{\varepsilon_i - Median(\varepsilon_i)}{MAD(\varepsilon_i)} \right| \quad (11)$$

3 实验分析

3.1 降水年代间变化分析

考虑到年代间的降水变化趋势较大，因此，先对降水观测数据进行整体趋势性分析。由于篇幅有限，着重对 2010 ~ 2014 年降水数据进行详细分析，其余年份的实验结果将在第 3.4 节的表 4 中进行统计。如图 6 所示，不同年代间的降水量差异巨大，其中，2011 年和 2013 年均属于大旱之年，各站点夏季平均降水量分别为 38.7 mm 和 19.5 mm，湖南中部地区遭遇特大干旱，作物绝收，人畜饮水困难。2010 年，2011 年和 2014 年降水充沛，各个年份之间的降水趋势性相关程度较高。但是，可以看出在降水丰富的情况下，降水空间分布极不均衡，因而容易导致局部地区发生洪涝，泥石流等灾害。

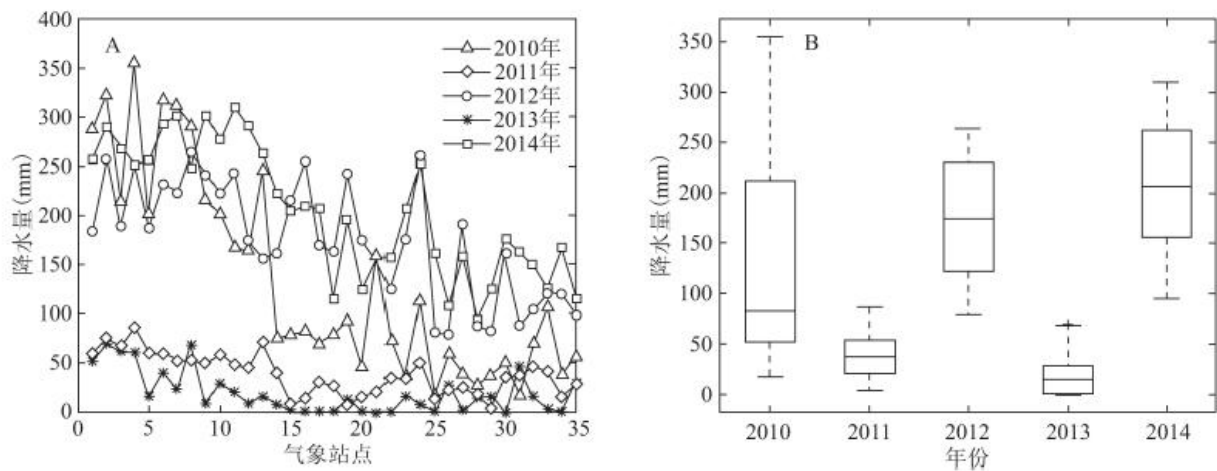


图 6 湖南省 2010~2014 年夏季降水量折线图(A) 与箱线图(B)

Fig. 6 Boxplot (A) and Line Chart (B) of Summer Precipitation in Hunan Province from 2010 to 2014

结合 Pearson 相关系数进一步分析,可以发现 2010 ~ 2014 年夏季,湖南省降水量与经度的相关性均较弱,且相关性不稳定,说明暖湿气流受到经度方向上的季风驱动的影响力较弱。与纬度的相关性较强,均值达到 0.6 左右,呈现显著的相关性特征,印证了湖南省夏季盛行来自太平洋的季风,也受来自印度洋的季风影响的气候特点。高程与降水的相关性微弱,但是经过反比例变换后,平均相关系数可达到-0.4,为中等程度负相关性,各个变量与不同年份的降水量相关性如图 7 所示。

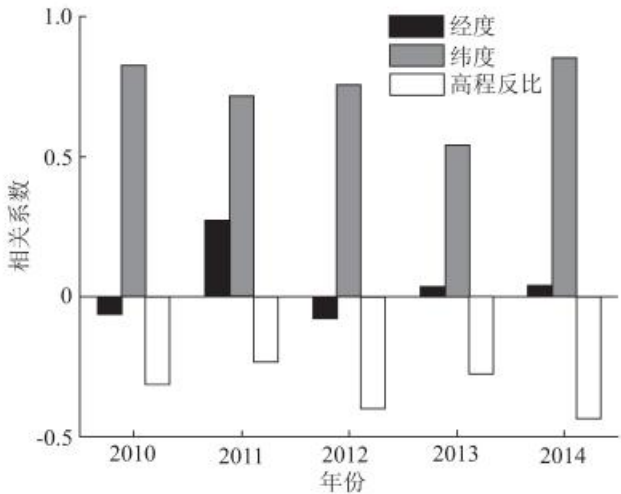


图 7 2010~2014 年降水量与经纬度、高程反比的相关系数

Fig. 7 Correlation Coefficient Between Precipitation and Latitude, Longitude and Inverse Ratio of Elevation from 2010 to 2014

3. 2 局部地形因子的空间尺度影响

在局部地形下,平均坡度,主要坡向会随着人为划定的采样区域大小而变化。Daly 等^[22]的研究表明 2 ~ 15 km 内的地形因子与降水相关性较高。因此,以 1 km 为间隔,利用 ArcGIS 软件分别提取 1 km 至 15 km 圆形缓冲区内相关变量的值。通过上述步骤获取到的坡度坡向自变量数目为 30 个,加上经纬度和海拔因子,一共产生了 33 个原始自变量。

对同一类型的变量而言,以这种方式产生的不同尺度自变量之间相似程度较高,若全部纳入回归方程,会因多重共线性而导致回归参数估计失效。可利用主成分分析法消除共线性。由于气象站点数量为 35 个,相对于 30 个自变量而言,站点数量仍然较少。因此,需要合理地选择自变量,于是从相关系数的角度进行筛选。首先设置滑动窗口内变量个数为 3 个,求取坡度指数,坡向指数在不同滑动窗口内与降水量相关系数的平均值。然后以相关性最高的窗口所筛选出的 3 个变量进行主成分分析。

如表 2,通过筛选后,所有年份的数据统计性质良好,KMO 指数均达到 0.7 以上,选取特征根大于 1 的因子作为主要成分,方差贡献率均达到了 85%以上,且都通过了 Bartlett 球度显著性检验。因此,认为新提取的变量能够较好地替代原始变量,有效地消除了共线性对回归方程的影响。

表 2 主成分分析

年份	元件	特征值	累计方差 (%)	KMO
2010	1	5.087	50.871	0.761 *
	2	2.382	74.690	
	3	1.003	85.723	
2011	1	4.659	51.770	0.732*
	2	1.730	70.990	
	3	1.228	85.632	
2012	1	6.202	51.683	0.723 *
	2	2.921	76.024	
	3	1.533	88.799	
2013	1	4.611	51.237	0.702 *
	2	2.148	75.107	
	3	1.080	87.111	
2014	1	8.769	55.804	0.745 *
	2	2.621	71.183	
	3	1.756	82.161	
	4	1.012	88.489	

注:* 表示通过 Bartlett 球度检验, 显著性水平 ≤ 0.05 .

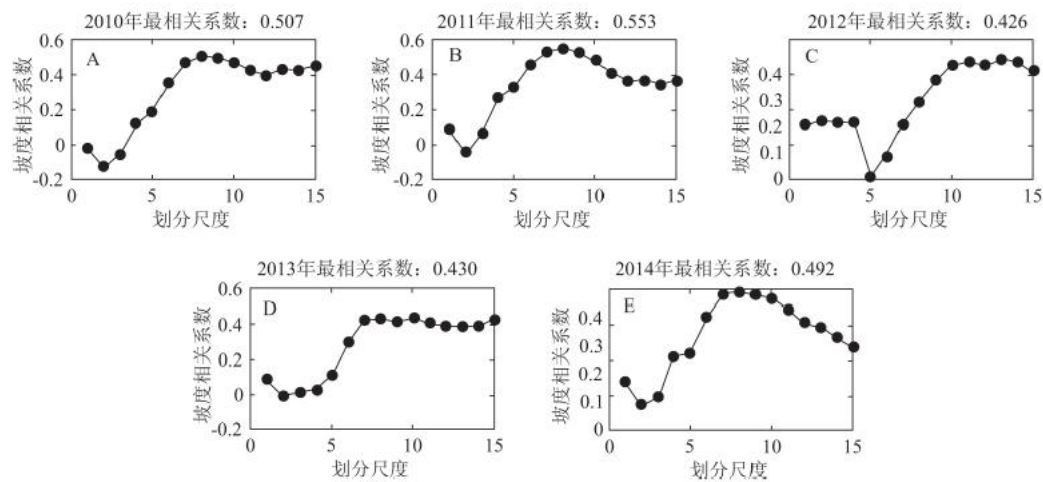


图 8 2010~2014 年多尺度下坡度指数与降水相关性折线图

Fig. 8 Line Chart of Correlation Between Precipitation and Slope Index Under Multi-Scale from 2010 to 2014

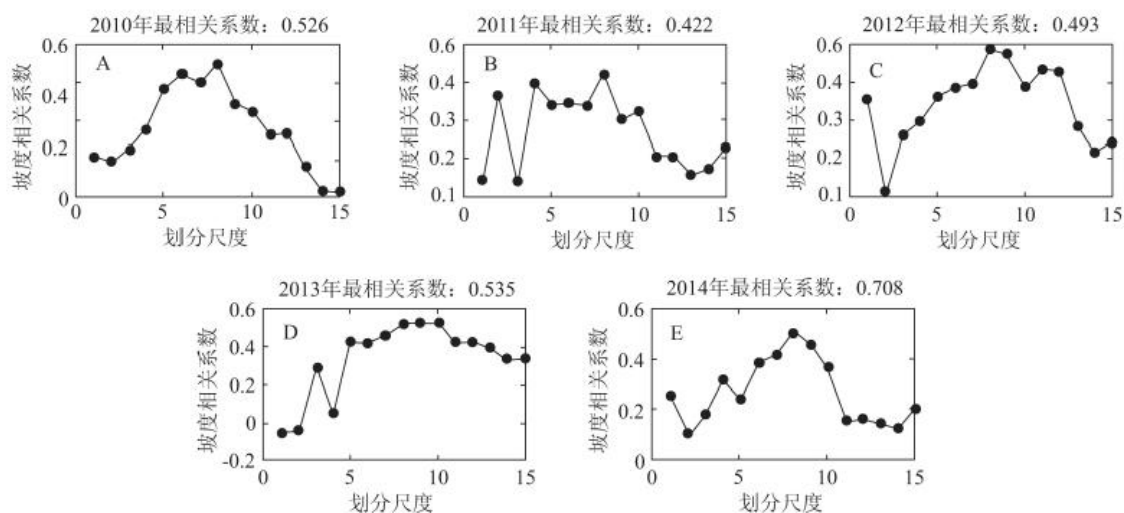


图 9 多尺度下坡向指数与降水相关性

Fig. 9 Line Chart of Correlation Between Precipitation and Aspect Index Under Multi-Scale from 2010 to 2014

3. 3 地理加权回归分析

如图 10 所示, 首先根据气象站点空间位置构建原始三角网。然后对其进行边长约束, 得到新的站点链接网络。新的网络可以对各个相连站点的邻近关系进行合理描述。本文按照约束后三角网的邻接关系来确定地理加权回归模型的带宽。

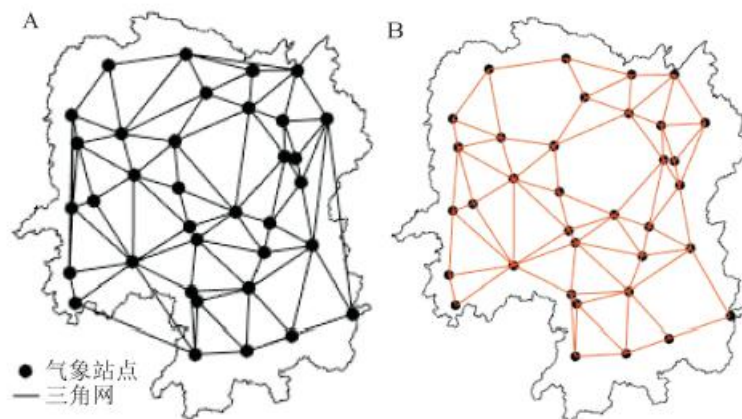


图 10 构建真实邻近关系

Fig. 10 Construct the Real Adjacent Relationship

如表 3 所示, 各年份回归结果中调整后的 R^2 至少达到了 65% 以上, 说明该方法可以有效地解释降水与地形的作用关系, 且各年的回归模型都通过了显著性检验。为了突出该方法的优越性, 选用 CV 准则确定带宽的 GWR 模型与之进行对比, 可以看到结合 Delaunay 三角网的自适应带宽模型 (GWRA) 回归效果更好。说明使用该方法对分布形式疏密不均的气象站点进行回归分析时, 适应性更好。

表 3 回归结果

年份	2010		2011		2012		2013		2014	
	GWRA	GWR	GWRA	GWR	GWRA	GWR	GWRA	GWR	GWRA	GWR
adjR ²	0.673	0.513	0.679	0.652	0.656	0.63	0.71	0.672	0.846	0.771
AIC	10.695	30.552	7.527	27.156	9.536	20.175	-7.384	9.445	-30.181	1.994
$\sqrt{MSE}$	17.16	20.78	11.33	18.156	26.911	43.556	10.29	15.778	16.824	18.991
F	2.842*	2.1 *	1.831 *	1.78*	2.31 *	2.815*	1.06**	1.56*	2.8*	2.4*

3. 4 异常检测结果分析

图 11 为使用反距离加权插值法得到的真实降水分布图，图 12 为通过异常站点的空间分布图。2010 年，全省降水量从北往南衰减，而异常站点主要聚集于降水趋势发生变化的过渡性区域，以 15, 16, 23 号气象站点进行连线，形成一层分隔线，刚好可以将降水较多的湘北部和湘南部区分隔开。而 2011 年是干旱之年，异常站点主要分布于邵阳市、娄底、益阳 3 个市的辖区，其中邵阳市管辖的涟源市(对应 18, 21 号)遭遇特大干旱，灾情最为严重。2012 年，湘江干道流域普降大到暴雨，局部特大暴雨，长株潭地区偏涝严重，对应的异常站点为 23, 24, 30 号。除此之外，检测出位于怀化市境内的 17 号与邵阳市境内年 6 ~ 8 月，上述地区多次遭遇暴雨，城市内涝严重。2013 年，湖南省遭遇了历史罕见的特大旱灾，全省降水稀少，湘中部大部份地区，如娄底市和郴州市(对应 20 号和 34 号)灾情最为严重，从空间分布形式看，大部分异常站点位于湖南省中线，并形成一种狭长的曲线形分布。仅一年过后，刚经历了大旱的湖南省又遭遇洪涝灾害侵袭，2014 年洞庭湖区(如 10 号)以及怀化市，湘西土家族苗族自治州等地区均不同程度受灾，受灾地区分布与异常点的分布大致吻合。除此之外，在表 4 中对 1985 ~ 2009 年的所有地区的异常检测结果进行了统计。

表 4 湖南省 1985 ~ 2009 年各市异常统计

地区	异常年份	地区	异常年份
怀化	1985, 1991, 1995, 1998, 1999, 2010, 2011, 2014	衡阳	1988, 1992, 1993, 1996, 2005, 2014
湘西	1989, 2014	岳阳	1998, 2003, 2009, 2011, 2014
娄底	1991, 2001, 2002, 2005, 2013	常德	1989, 1990, 1995, 1998, 2003
邵阳	1986, 2001, 2002, 2009, 2011, 2012, 2013	长沙	1986, 1989, 1998, 2003, 2008
张家界	1987, 2008	益阳	1988, 1992, 1993, 2006, 2007, 2008, 2014
永州	1985, 1990, 1996, 2007	株洲	1987, 1998, 2008, 2009, 2010
郴州	1992, 1999, 2006, 2013	湘潭	1985, 1995, 1998, 2012

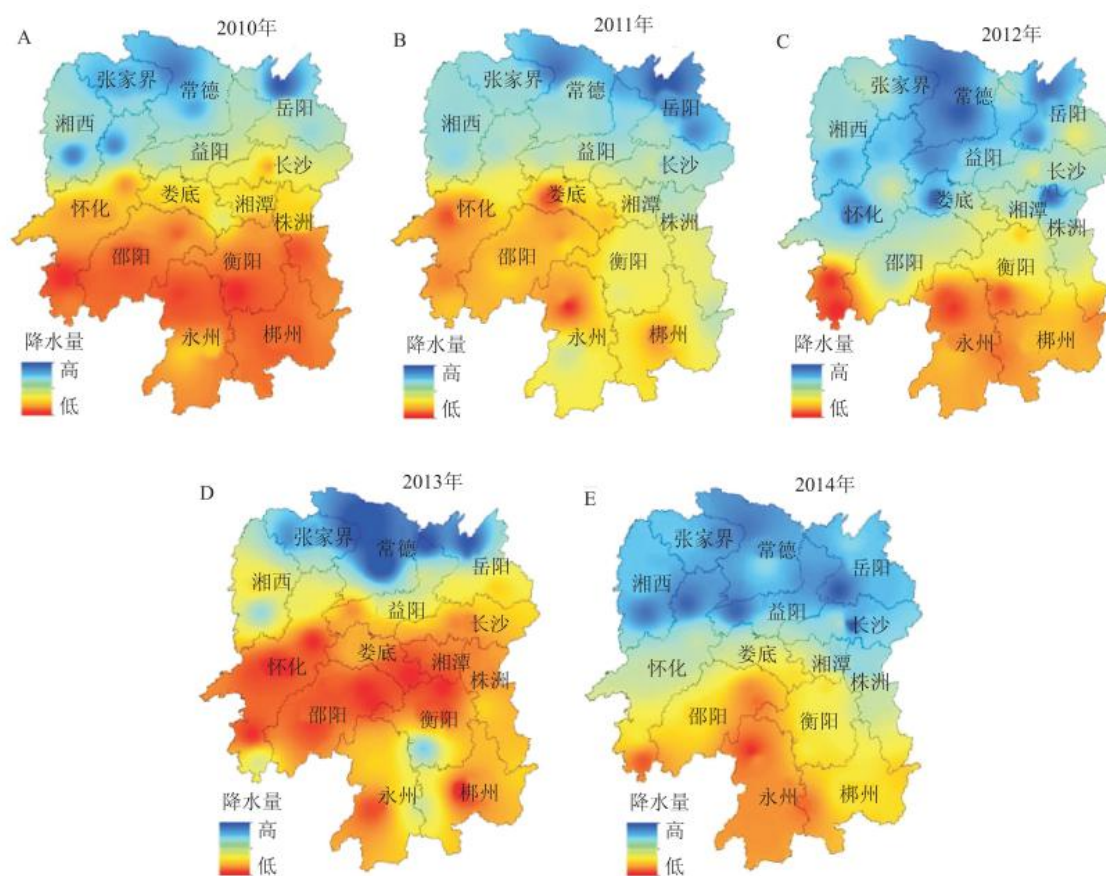


图 11 湖南省 2010~2014 年夏季降水分布图

Fig. 11 Distribution of Summer Precipitation in Hunan Province from 2010 to 2014

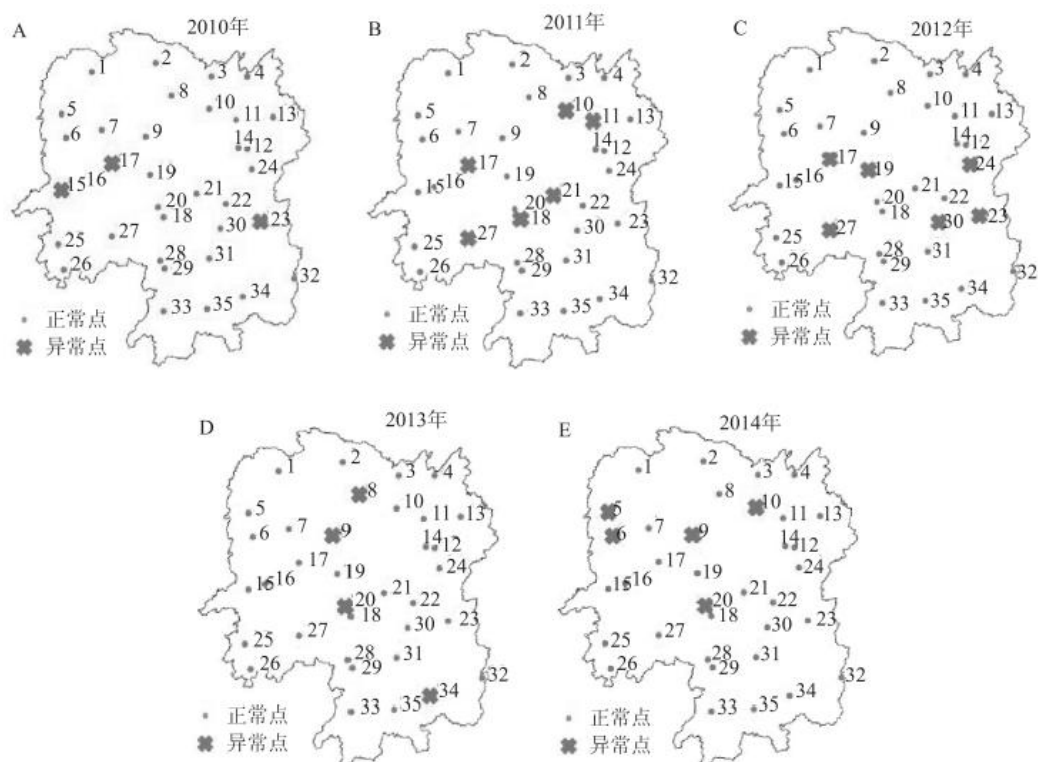


图 12 湖南省 2010~2014 年夏季降水空间异常

Fig. 12 Spatial Outliers of Summer Precipitation in Hunan Province from 2010 to 2014

4 结论与展望

发展了一种结合 Delauney 三角网和地理加权回归二者优点的异常检测方法。通过对 1985 年到 2014 年湖南省的降水观测数据进行异常检测分析，得出如下结论：

(1) 实验表明，在数据点数据分布形式疏密不均的情况下，相较于普通 GWR，该方法所获取结果的拟合优度较好，对于数据分布形式的适应性更强。

(2) 湖南省范围内，中小尺度下的降水信息受到地形因素的影响较大，局部地形范围介于 7 ~10 km 时，降水与地形因素的相关性较高。而通过主成分分析，可以提取主要影响因子，简化问题。

(3) 湖南省夏季气候表现出干旱与洪涝灾害交替频发的两个极端，异常频次高的区域往往是极端天气突发高发的地区，因此，挖掘异常站点对于降水信息二次处理以及进一步探索局地降水规律等具有重要意义。

下一步的研究工作主要集中在研究不同回归自变量对于异常检测的影响，进一步细化异常检测模型，提高模型效能。

参考文献:

- [1] 李丹, 郭生练, 洪兴骏, 等. 汉江流域 1960 ~ 2014 年降雨极值时空变化特征 [J]. 长江流域资源与环境, 2016,

25(9) : 1448—1456.

【LI D, GUO S L, HONG X J, et al. Temporal and spatial variation of rainfall extremes in Han river basin from 1960 to 2014 [J]. Resources and Environment in the Yangtze Basin, 2016, 25(9) : 1448—1456. 】

[2] 于晓晶, 赵勇. 地形对天山夏季降水影响的模拟 [J]. 中国沙漠, 2016, 36(4) : 1133—1143.

【YU X J, ZHAO Y. Simulation of orographic effects on summer rain in the Tianshan mountains, 2016, 36(4) : 1133—1143. 】

[3] 舒守娟, 王元, 熊安元. 中国区域地理、地形因子对降水分布影响的估算和分析 [J]. 地球物理学报, 2007, 50(6) : 1703—1712.

【SHU S J, WANG Y, XIONG A Y. Estimation and analysis for geographic and orographic influences on precipitation distribution in China [J]. Chinese Journal of Geophysics Geophys. 2007, 50(6) : 1703—1712. 】

[4] 范泽孟, 岳天祥, 陈传法, 等. 中国气温与降水的时空变化趋势分析 [J]. 地球信息科学学报, 2011, 13(4) : 526—533.

【FAN Z M, YUE T X, CHEN C F, et al. Spatial change trends of temperature and precipitation in China [J]. Journal of Geo-Information Science, 2011, 13(4) : 526—533. 】

[5] 廖菲, 洪延超, 郑国光. 地形对降水的影响研究概述 [J]. 气象科技, 2007, 35(3) : 309—316.

【LIAO F, HONG Y C, ZHENG G G. Review of orographic influences on surface precipitation [J]. Meteorological Science and Technology. 2007, 35(3) : 309—316. 】

[6] 刘小宁, 鞠晓慧, 范邵华, 等. 空间回归检验方法在气象资料质量检验中的应用 [J]. 应用气象学报, 2006, 17(1) : 37—43.

【LIU X N, JU X H, FAN S H, et al. A research on the applicability of spatial regression test in meteorological datasets [J]. Journal of Applied Meteorological Science, 2006, 17(1) : 37—43. 】

[7] 张杰, 李栋梁. 祁连山区及黑河流域降水量的分布特征分析 [J]. 高原气象, 2004, 23(1) : 81—88.

【ZHANG J, LI D L. Analysis on distribution character of rainfall over Qilian mountain and Heihe valley. Plateau Meteorology, 2004, 23(1) : 81—88. 】

[8] 王远见, 傅旭东, 高洁, 等. 雅鲁藏布江中下游地区湿季降雨量的空间分布模型 [J]. 清华大学学报自然科学版, 2012(6) : 738—746.

【WANG Y J, FU X D, GAO J, et al. Spatial precipitation distribution of over the middle and lower reaches of the Yarlung Zangbo River, 2012(6) : 738—746. 】

[9] 刘琼峰, 李明德, 段建南, 等. 长沙城郊农田土壤重金属 Pb、Cd 的生态风险评价研究 [J]. 农业环境科学学报, 2013, 32(8) : 1564—1570.

【LIU Q F, LI M D, DUAN J N, et al. Analysis on influence factors of soil Pb and Cd in agricultural soil of Changsha suburb based on geographically weighted regression model [J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 32(8) : 1564—1570. 】

[10] BRUNSDON C, FOTHERINGHAM A S, CHARLTON M. Geographically weighted regression: A method for exploring spatial nonstationarity [J]. Geographical Analysis, 1996, 28(4) : 281—298.

[11] 覃文忠, 王建梅, 刘妙龙. 地理加权回归分析空间数据的空间非平稳性 [J]. 辽宁师范大学学报自然科学版, 2005, 28(4) : 476—479.

【QIN W Z, WANG J M, LIU M L. Spatial nonstationarity of geographically weighted regression analysis of spatial data [J]. Journal of Liaoning Normal University (Natural Science Edition), 2005, 28(4) : 476—479. 】

[12] HAWKINS D. 1980 Identification of outliers [M]. London: Chapman and Hall.

[13] 邓敏, 石岩, 龚健雅, 等. 时空异常探测方法研究综述 [J]. 地理与地理信息科学, 2016, 32(6) : 43—50.

【DENG M, SHI Y, GONG J Y, et al. A summary of spatiotemporal outlier detection, (Geography and Geo-information Science), 2016, 32(6) : 43—50. 】

[14] 杨学习, 石岩, 邓敏, 等. 一种基于多层次专题属性约束的空间异常探测方法 [J]. 武汉大学学报(信息科学版), 2016, 41(6) : 810—817.

【YANG X X, SHI Y, DENG M, et al. A new method of spatial outlier detection by considering multi-level thematic attribute constraints. (Geomatics and Information Science of Wuhan University), 2016, 41(6) : 810—817. 】

[15] 张剑明, 廖玉芳, 段丽洁, 等. 湖南近 50 年极端连续降水的气候变化趋势 [J]. 地理研究, 2012, 31(6) : 1004—1015.

【ZHANG J M, LIAO Y F, DUAN L J, et al. Climate variations of extreme continuous rainstorms in Hunan during 1960—2009 [J]. Geographical Research, 2012, 31(6) : 1004—1015. 】

[16] 傅抱璞. 地形和海拔高度对降水的影响 [J]. 地理学报, 1992(4) : 302—314.

【FU B P, The Effects of topography and elevation on precipitation [J]. Acta Geographica Sinica, 1992(4) : 302—314. 】

[17] 史岚, 张狄, 乔淼, 等. 基于 NCEP-FNL 资料与 DEM 数据的日地形降水估算模型研究 [C]// 气象服务发展论坛. 2014.

【SHI L, ZHANG D, QIAO M, et al. Study on the estimation model of daily orographic precipitation based on NCEP-FNL

data and DEM data [C] //Meteorological Service Development Forum, 2014. 】

[18] 李小文, 曹春香, 常超一. 地理学第一定律与时空邻近度的提出 [J] . 自然杂志, 2007, 29(2) : 69—71.

【LI X W, CAO C X, CHANG C Y. The first law of geography and spatial-temporal proximity [J] . Chinese Journal of Nature, 2007, 29(2) : 69—71. 】

[19] 尚华, 张贝贝, 纪宏. 一种新的基于回归分析的异常值检测 [J] . 河南大学学报(自然版) , 2015, 45(6) : 635—639.

【SHANG H, ZHANG B B, JI H. Outlier detection of a new method based on regression analysis [J] . Journal of Henan University(Natural Science Edition) , 2015, 45(6) : 635—639. 】

[20] JIANG S Y, LI Q H, LI K L, et al. GLOF: a new approach for mining local outlier [C] //International Conference on Machine Learning and Cybernetics. IEEE, 2003: 157—162.

[21] ROUSSEEUW P J, HUBERT M. Robust statistics for outlier detection [J] . Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery, 2011, 1(1) : 73—79.

[22] DALY C, NEILSON R P, PHILLIPS D L. A statistical-topographic model for mapping climatological precipitation over mountainous terrain [J] . Journal of Applied Meteorology, 1994, 33(33) : 140—158.