

# 空间点过程理论在极端气候事件中的应用研究<sup>\* 1</sup>

杨 萍<sup>1)2)</sup> 封国林<sup>2)</sup> 刘伟东<sup>1)</sup> 侯 威<sup>2)</sup> 孙树鹏<sup>2)</sup>

<sup>1)</sup>(中国气象局北京城市气象研究所,北京 100089) <sup>2)</sup>(国家气候中心,北京 100081)

## 摘 要

该文定义了 12 种极端温度事件和 6 种极端降水事件,将基于空间点过程理论的  $k$  阶最近邻距离丛集点提取算法应用于极端温度和极端降水事件的研究,给出了极端温度和极端降水事件区域群发性站点的检测流程;采用多年平均的疏密差异比  $R$  和有效率  $\gamma$  两个指标,对所定义的各种极端气候事件的区域群发性进行了有效性检验,结果表明: $k$  阶最近邻距离丛集点提取算法适用于极端气候事件的区域群发性研究,并进一步阐明了极端气候事件区域群发性的物理含义。

**关键词:** 极端气候事件;区域群发性; $k$  阶最近邻距离;丛集点

## 引 言

在全球增暖背景下,无论是发生频次、影响范围或是破坏程度,极端气候事件均表现出越来越明显的群发性特征。极端气候事件群发性特征的已有研究可归纳如下:一是不同种类极端气候事件的群发,比如干旱、高温或干热风等极端事件同时发生,例如 2006 年夏我国四川、重庆等地发生了百年一遇的干旱,与之相伴的是历史罕见的高温热浪<sup>[1]</sup>;二是相同种类极端气候事件在时间上的群发,或称之为持续性极端事件,例如 2008 年 10 月—2009 年 1 月发生在我国东部的持续性少雨事件,造成了淮河以北地区的严重干旱<sup>[2]</sup>;三是极端气候事件在空间上的区域性群发,或称之为区域丛集特征,是区域集中性和区域群发性的一种体现,例如 2008 年 1 月—2 月初,我国南方地区出现了历史罕见的低温雨雪冰冻灾害,造成的直接经济损失达上千亿元<sup>[3]</sup>。上述 3 种群发性极端事件均可能造成严重的社会影响和经济损失,是目前在全球变暖背景下国际上研究的热点问题之一。本文从空间角度研究了极端事件的群发性,为极端事件的研究提供一种新的视角和思路。

在年代际尺度上,极端气候事件具有明显的空间丛集特征<sup>[4-8]</sup>;从天气尺度上来讲,高温、暴雨等

极端天气事件的区域群发性也非常显著<sup>[9-15]</sup>。极端气候事件的区域群发性具有明显的年际变化规律,这为从年际尺度来研究极端气候事件提供了一种可能性,对极端气候事件年总频数进行统计<sup>[16]</sup>发现,年总频数大的极端气候事件往往密集分布,而这些密集分布的极端气候事件所呈现的灾害性特征更值得关注。高频密集分布是极端气候事件区域群发性的一种体现,如果能将高频密集分布的群发性区域进行客观量化描述,将为极端气候事件群发性的影响评估和机理分析提供帮助,具有很强的现实意义。本文将空间点过程理论中的  $k$  阶最近邻距离丛集点提取算法与极端气候事件的群发性研究有机结合,以极端气候事件的年际区域群发性为切入点,探讨了该方法用于极端气候事件区域群发性研究的可行性和有效性。

## 1 空间点过程理论的 $k$ 阶最近邻距离丛集点提取算法

基于空间点过程理论的  $k$  阶最近邻距离丛集点算法,能够将空间上分布不均匀的点较为客观地区别,即将空间上分布相对密集的点归为一类,称之为丛集点,将空间分布相对零散的点归为一类,称之为背景点。该算法的核心思想是结合概率统计的基本

\* 国家自然科学基金项目(40930592)、科技部支撑项目“极端气候事件的检测和可预测性关键技术研发”(2007BAC29B01)、“气候变化的检测和预估技术研究”(2007BAC03A01)和公益性行业(气象)科研专项(GYHY200806005)共同资助。

2009-09-24 收到,2010-03-12 收到再改稿。

原理,通过计算所研究区域内各点的  $k$  阶最近邻距离,从而得到各点的丛集点隶属度,以此来判断各点是否属于丛集点<sup>[16-23]</sup>。

### 1.1 $k$ 阶最近邻距离丛集点提取算法

假设  $Y$  为一定范围内单一的二维泊松过程点集,其中任一点  $s_i$  的  $k$  阶最近邻距离为  $W_k$ ,其概率密度函数  $g_{w_k}(x)$  为:

$$g_{w_k}(k, \lambda) = \frac{e^{-\lambda\pi x^2} 2(\lambda\pi)^k x^{2k-1}}{(k-1)!}. \quad (1)$$

式(1)中,  $k$  为距离阶数,  $\lambda$  为所研究点集  $Y$  的分布参数。在实际研究中,丛集点和背景点分别代表不同参数  $\lambda$  的空间点过程,它们在空间上互相叠加,因此,以两个叠加的二维泊松过程为例,其  $W_k$  分布服从以下的混合分布:

$$W_k \sim p\Gamma^{1/2}(k, \lambda_1\pi) + (1-p)\Gamma^{1/2}(k, \lambda_2\pi). \quad (2)$$

式(2)中,  $p$  为比例系数,  $\lambda_1$  和  $\lambda_2$  分别为丛集区域和背景区域的分布参数,  $k$  为距离阶数<sup>[15-20]</sup>。结合式(1)和式(2),它们的  $W_k$  服从以下分布:

$$W_k \sim pg_{w_k}(k, \lambda_1) + (1-p)g_{w_k}(k, \lambda_2). \quad (3)$$

$\lambda_1$  和  $\lambda_2$  的计算采用 Byers 和 Raftery 提出的期望最大化算法(EM 算法)来计算,EM 算法的核心思想是通过赋予所求参数(文中所求参数为  $\lambda_1$  和  $\lambda_2$ )一初始值,通过多次迭代,所求参数将收敛于与其真实值最为接近的某一稳定值,则该收敛值即为所求的参数值,所以该方法又称之为期望最大化算法。具体包括 E-步骤和 M-步骤(详细阐述可参见文献[17]):

E-步骤:

$$E(\hat{\delta}_i^{(t+1)}) = \frac{\hat{p}^{(t)} g_{w_k}(k, \hat{\lambda}_1^{(t)})}{\hat{p}^{(t)} g_{w_k}(k, \hat{\lambda}_1^{(t)}) + (1-\hat{p}^{(t)}) g_{w_k}(k, \hat{\lambda}_2^{(t)})}; \quad (4)$$

M-步骤:

$$\begin{aligned} \hat{\lambda}_1^{(t+1)} &= \frac{k \sum_{i=1}^n \hat{\delta}_i^{(t+1)}}{\pi \sum_{i=1}^n W_{k,i}^2 \hat{\delta}_i^{(t+1)}}, \\ \hat{\lambda}_2^{(t+1)} &= \frac{k \sum_{i=1}^n (1-\hat{\delta}_i^{(t+1)})}{\pi \sum_{i=1}^n W_{k,i}^2 (1-\hat{\delta}_i^{(t+1)})}, \\ \hat{p}^{(t+1)} &= \sum_{i=1}^n \hat{\delta}_i^{(t+1)}. \end{aligned} \quad (5)$$

其中,  $n$  代表事件总数,  $W_{k,i}$  表示第  $i$  ( $i=1, 2, \dots, n$ )

个事件的  $k$  阶最近邻距离,  $t$  表示迭代次数,  $\hat{\delta}_i^{(t+1)}$  代表第  $i$  个点在  $t+1$  次迭代后丛集点的隶属度。若  $\lambda_1$  代表丛集点的分布参数,则  $\hat{\delta}_i^{(t+1)} \geq 0.5$  的点为丛集点,  $\hat{\delta}_i^{(t+1)} < 0.5$  的点为背景点。

### 1.2 算法的适用性及在极端气候事件中的应用

关于算法中参数  $k$  的讨论已有一些研究,裴韬等<sup>[18-20]</sup>发现,当  $k$  在合适范围内取值时算法的错误率较低,较为有效。杨萍等对该算法的适用性进行了进一步分析<sup>[16]</sup>,发现样本总量、丛集点与分散点的疏密差异比等参数对计算结果有较大影响,并定义了参数  $R$ (丛集点与分散点疏密差异比值)来量化地界定算法的适用范围;此外对数据点的设置进行了拓展研究,对于重叠的数据点,可引入权重思想,对不同类型的重叠数据点赋予不同的权重,从而实现客观提取丛集点的目标。上述结果为极端气候事件的群发性研究提供了有力的理论依据和算法基础。

有别于常规极端气候事件的研究主要针对单站点的不同要素,本研究的目标在于通过  $k$  阶最近邻距离丛集点提取算法将单站点极端气候事件的研究扩展到相互联系的多站点极端气候事件空间分布的讨论中,即将极端事件的研究范围从对点的讨论扩展到面的研究。用站点重复出现的次数来体现站点极端气候事件发生的频次(强度),使高权重区域的分布密度进一步加强,从而实现高频(强度)丛集站点的提取工作,达到对极端气候事件群发性进行量化描述的目标,并进一步分析极端气候事件群发性的时间变化趋势和空间分布特征,从而为极端气候事件群发性的定量化评估和分析提供了一种新的思路。

## 2 空间点过程理论在极端气候事件中的应用

### 2.1 区域范围和所用资料

我国不同区域内观测站点的不均匀分布是客观存在的,为尽量克服该现象对丛集点提取结果的影响,本文去除了分布较为稀疏的西部区域测站(新疆、西藏、内蒙古 3 省区),主要研究东亚季风区极端气候事件的群发性。

所用资料为我国大陆 740 个测站的日最高温度、最低温度、日降水资料集,由于资料限制,温度和降水资料均存在一定缺测,分别对其进行处理。去除西部测站以及东部区域缺测的站点后,最后取全

国 426 个测站的逐日最高温度、最低温度资料,资料的时段为 1960—2005 年;降水资料的时段为 1960—2006 年,最后取全国 437 个测站的逐日降水资料。

## 2.2 极端气候事件的定义

### 2.2.1 极端温度事件的定义

本文中极端温度事件的定义为:将某站点 1960—2005 年的日最高(低)温度资料样本概率位

于  $x\%$  ( $x$  视实际情况而定) 的值作为极端高(低)温事件的上(下)阈值,定义夏季高温(冬季低温)事件;此外,温度严重偏离(高于或低于)该天的多年平均状态,也可看作是另一种极端高温(低温)事件,即距平值定义异常偏高(低)事件。通过控制百分位的取值,从而定义不同强度的极端温度事件,共计 12 种(表 1)。

表 1 极端温度事件的定义

Table 1 Different definitions of tempature extreme

极端温度事件	定义
夏季高温事件 90%(T1)	日最高温度大于 1960—2005 年日最高温度资料样本概率位于 90% 的值
夏季高温事件 95%(T2)	日最高温度大于 1960—2005 年日最高温度资料样本概率位于 95% 的值
夏季高温事件 99%(T3)	日最高温度大于 1960—2005 年日最高温度资料样本概率位于 99% 的值
冬季低温事件 90%(T4)	日最低温度小于 1960—2005 年日最低温度资料样本概率位于 90% 的值
冬季低温事件 95%(T5)	日最低温度小于 1960—2005 年日最低温度资料样本概率位于 95% 的值
冬季低温事件 99%(T6)	日最低温度小于 1960—2005 年日最低温度资料样本概率位于 99% 的值
异常偏高事件 90%(T7)	日最高温度距平值大于 1960—2005 年日最高温度距平值资料样本概率位于 90% 的值
异常偏高事件 95%(T8)	日最高温度距平值大于 1960—2005 年日最高温度距平值资料样本概率位于 95% 的值
异常偏高事件 99%(T9)	日最高温度距平值大于 1960—2005 年日最高温度距平值资料样本概率位于 99% 的值
异常偏低事件 90%(T10)	日最低温度距平值小于 1960—2005 年日最低温度距平值资料样本概率位于 90% 的值
异常偏低事件 95%(T11)	日最低温度距平值小于 1960—2005 年日最低温度距平值资料样本概率位于 95% 的值
异常偏低事件 99%(T12)	日最低温度距平值小于 1960—2005 年日最低温度距平值资料样本概率位于 99% 的值

### 2.2.2 极端降水事件的定义

极端降水事件的定义方法有多种,传统上采取绝对阈值法,即采用降水量大于等于某一固定阈值的方法;百分比阈值法也是极端降水事件定义的常用方法<sup>[24]</sup>;此外,也有研究者利用气候平均态(30 年的平均值)作为阈值来定义极端降水事件<sup>[25]</sup>。参考已有的极端降水事件定义标准,本文具体定义见表 2。

表 2 极端降水事件的定义

Table 2 Different definitions of precipitation extreme

事件	定义
大雨	日降水量大于等于 25 mm
暴雨	日降水量大于等于 50 mm
大暴雨	日降水量大于等于 100 mm
中等降水	日降水量大于 1960—2006 年雨日(日降水量 $\geq 1$ mm)的降水资料样本概率位于 75% 的值
强降水	日降水量大于 1960—2006 年雨日(日降水量 $\geq 1$ mm)的降水资料样本概率位于 95% 的值
极端降水	日降水量大于 1961—1990 年的各年降水资料样本概率位于 99% 的值的 30 年平均值

## 2.3 权重的定义

结合  $k$  阶最近邻距离丛集点提取算法,并参考关于引入权重的研究结果<sup>[23]</sup>,在本文极端气候事件群发性的研究中,采取如下两种方法定义权重,即分别从极端事件的年总频数和年总强度的角度定义权

重。

### 2.3.1 从年总频数的角度定义权重

给各站点赋予权重,用该站点重复出现的次数体现权重大小。例如,某站点的坐标为  $(x, y)$ ,权重为 5,则该数据点即其坐标  $(x, y)$  出现 5 次。由于极端降水事件和极端温度事件在定义上非常类似,本文以极端温度事件为例进行权重定义的说明。

设样本概率取值为  $q\%$  ( $q$  分别取 90, 95, 99),逐日温度观测资料为 1960—2005 年共记 16802 d。以任一站点为研究对象,按样本概率取值为  $q\%$  的值作为极端高(低)温事件的阈值,极端高(低)温事件共发生  $(\frac{100-q}{100} \times 16802)$  日次(记为  $s$ ),年平均发生日次为  $\frac{s}{46}$  (记为  $T_{ave}$ ) 日次;定义发生日次等于年平均日次的权重定义为 1,则一个日次的权重(记为  $W_0$ )为  $\frac{1}{T_{ave}}$ ,定义站点极端高(低)温事件的年权重(记为  $W$ )为年总频数(记为  $N$ )与一个日次的权重( $W_0$ )之积,用公式表示为:  $W = W_0 \times N$ 。一般地,权重的大小用数据点重复出现的次数来表征,必须为整数,故四舍五入取整。

### 2.3.2 从年总强度的角度定义权重

除依据年总频数定义权重外,还可从年总强度定义权重。用极端气候事件的要素观测值与阈值之差的绝对值来表征强度,以年为单位做累加求和。总体来说,年总强度越大,权重越大,仍以极端温度事件为例进行说明。

设定样本概率取值为  $q\%$  ( $q$  取 90, 95, 99), 时间序列 1960—2005 年共计 16802 d。以任一站点为研究对象, 极端高(低)温事件的总发生日次为  $(\frac{100-q}{100} \times 16802)$  日次(记为  $s$ ), 将  $s$  个日次的极端高(低)温观测值分别减去概率取值为  $q\%$  对应的阈值  $y$ , 对其绝对值进行求和累加, 其值记为  $Z$ , 年平均强度  $I_{ave} = \frac{Z}{46}$ ; 设第  $i$  年极端高(低)温事件有  $n_i$  天, 将这  $n_i$  个观测值与阈值  $y$  的差值的绝对值累加求和, 其值记为  $Z_i$ , 定义权重的大小为:  $W = \frac{Z_i}{I_{ave}}$ 。同样, 对结果按四舍五入取整。

## 2.4 极端气候事件应用研究的流程设计

### 2.4.1 区域群发性极端气候事件的检测流程

在极端气候事件和权重的定义的基础上, 结合  $k$  阶最近邻距离提取算法逐年进行极端气候事件群发区域的检测与提取工作, 检测流程如图 1 所示。

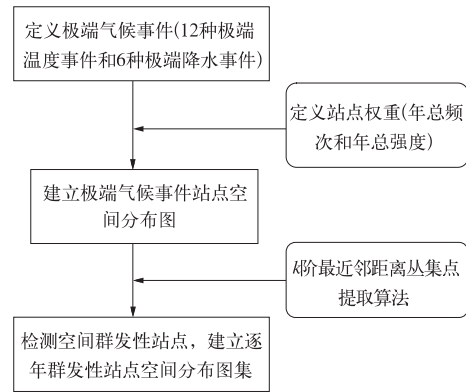


图 1 区域群发性极端气候事件的检测流程图

Fig. 1 Checking flow chart of cluster extreme events

### 2.4.2 实际资料举例

以我国 426 测站 1960—2005 年夏季高温事件为例, 对极端事件群发性的检测流程进行说明。

任选一站点, 将 1960—2005 年的日最高温度资料样本概率为 99% 的值作为夏季高温事件的阈值, 大于该阈值记为夏季高温事件, 逐年统计各站点发生夏季高温事件的频次并据此赋予权重, 得到各年夏季高温事件空间点的数据集; 利用空间点过程理论的  $k$  阶最近邻距离丛集点提取算法, 得到群发性站点的逐年空间分布图集。以 2000 年为例, 其结果如图 2a 所示。

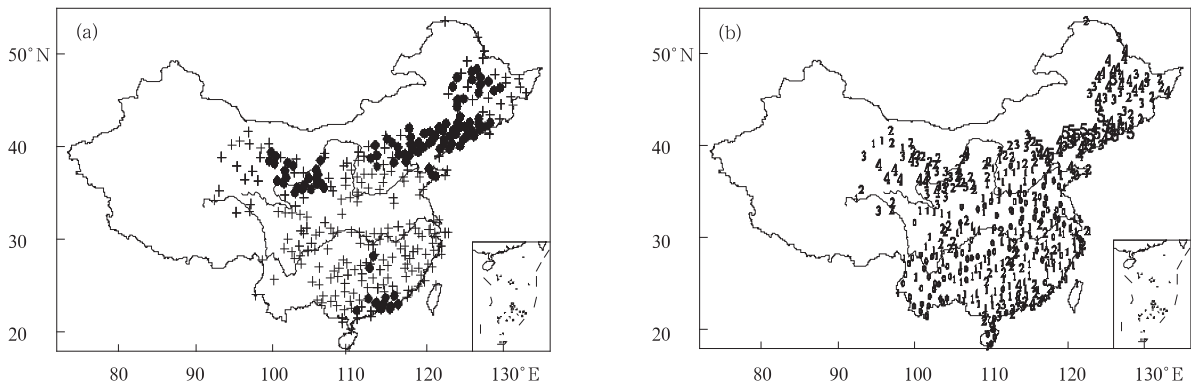


图 2 2000 年夏季高温事件的空间分布

(a) 群发提取结果图(“+”表示发生极端高温事件的站点, “●”表示提取出的群发性站点),

(b) 权重分布图(数字“1”~“5”代表所在位置上的站点的权重)

Fig. 2 Stations over China about temperature extreme in 2000

(a) cluster distribution (“+” represent the stations occurring temperature extreme, “●” represent the

cluster stations occurring temperature extreme), (b) power distribution (number “1”—“5” represent stations power)

图 2a 中, “+”为 2000 年发生极端高温事件的站点, 即根据百分比阈值法以及权重的定义方法得到的数据集, “●”为利用空间点过程理论提取出的

群发站点。与 2000 年极端高温事件的权重分布图(图 2b)进行对比, 发现高权重且密集分布的站点被提取出来。该方法避免了人为界定群发性区域所带

来的主观性,同时还可以检测出极端气候事件影响最为严重、最值得关注的区域。

### 3 算法有效性检测

#### 3.1 有效性检测的说明和具体方法

已有研究说明该算法有一定的适用范围<sup>[14]</sup>。当丛集点和背景点的疏密差异不太明显时,提取结果的错误率较高,有效性会随之降低。因此,该算法在极端气候事件群发性研究中的有效性如何值得进行深入讨论。

样本量较小(小于 400)时,得到  $R$  的临界值为 4.0,样本量较大(大于等于 400)时, $R$  的临界值为 2.9<sup>[14]</sup>。从极端气候事件实际发生站点数来看,其没有统一的样本量范围,为了保证结果的可靠性,本文选取  $R$  的临界值为 4;在  $R$  的修正研究中<sup>[7]</sup>发现,一般而言,未知状态下, $R$  的计算值比理论值偏低 0.2 左右;因此,算法的有效性检测中,本文实际  $R$  的临界值取 3.8。

#### 3.2 有效性检测方法

##### 3.2.1 $\bar{R}$ 的计算方法

利用  $k$  阶最近邻距离丛集点提取算法,逐年计算出丛集站点和背景站点的分布参数比值  $R$ ,对其进行求和平均,记为  $\bar{R}$ 。

##### 3.2.2 有效率的计算方法

以 3.8 为临界值,统计出大于等于 3.8 的年份,记为  $n$ ,总年份记为  $N$ ,其比值定义为极端气候事件的有效率,记为  $\eta$ 。具体计算公式为:

$$\eta = \frac{n(R \geq 3.8 \text{ 的年份数})}{N(\text{总年份数})} \quad (6)$$

#### 3.3 极端气候事件的有效性检测

##### 3.3.1 极端温度事件的有效性检测

基于前文定义的夏季高温、冬季低温、异常偏高、异常偏低 4 类极端温度事件,通过控制百分位的取值,定义了不同强度的极端温度事件,共计 12 种极端温度事件(参见表 1)。在权重定义中,分别从频次和强度两个角度定义权重,计算每一年的  $R$  值,求和平均得  $\bar{R}$ ,由式(6)可知  $\eta$ ,具体  $\bar{R}$  和  $\eta$  如图 3 所示。

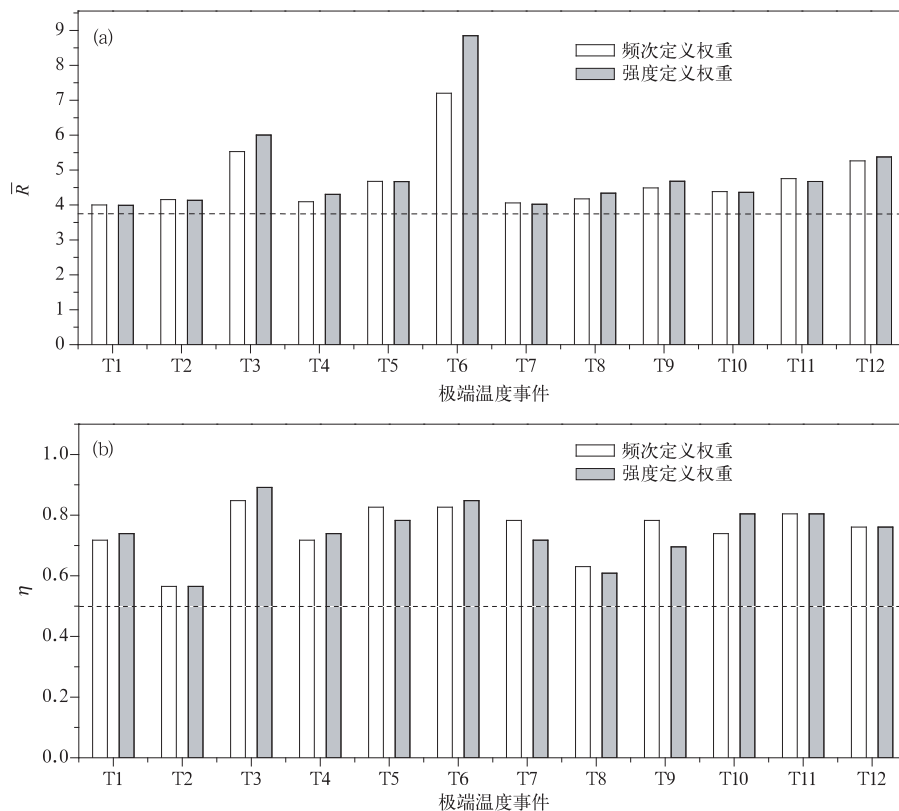


图 3 12 种极端温度事件有效性检测结果图

(a) 12 种极端温度事件的  $\bar{R}$ , (b) 12 种极端温度事件的  $\eta$

Fig. 3 Temperature extreme's average of ratio  $\bar{R}$  (a) and efficient  $\eta$  (b)

图 3a 给出了 12 种极端温度事件的  $\bar{R}$ , 其中白色的柱状直方图为频次定义权重下的  $\bar{R}$ , 灰色柱状图是强度定义权重下的  $\bar{R}$ , 虚线对应的值是  $R=3.8$ 。图 3b 给出了其各自的有效率  $\eta$ , 和  $\bar{R}$  类似, 频次定义权重的有效率用白色直方图表示, 强度定义权重的有效率用灰色直方图表示, 虚线对应的值为 50% 的有效率。

由图 3 可知: 12 种极端温度事件的  $\bar{R}$  均大于临界值 3.8, 有效率均大于 50%, 可见, 算法检测出极端温度事件群发性的能力比较强。不同样本概率情况下的极端温度事件中, 极端温度事件 99% 的  $\bar{R}$  最高, 有效率最大, 说明算法有效性与极端事件的极端程度有一定联系。对比频次和强度两种定义权重的

方法, 差别不大, 计算各自  $\bar{R}$  和  $\eta$  的平均值, 基于频次时  $\bar{R}$  的平均值为 4.73, 基于强度时的平均值为 4.95, 有效率  $\eta$  的平均值几乎相同, 均为 0.75。另外, 对有效性最强的情况, 即样本概率取 99% 时的各种极端温度事件进行了统计, 基于频次的  $\bar{R}$  为 5.62, 基于强度的为 6.22, 有效率  $\eta$  平均值几乎相同, 均为 0.80, 极端程度最高的极端温度事件的有效性远大于平均水平。

### 3.3.2 极端降水事件的有效性检测

极端降水事件的定义主要分为绝对阈值法和百分比阈值法两种, 表 2 中给出了 6 种极端降水事件的定义,  $\bar{R}$  和  $\eta$  的大小如图 4 所示。

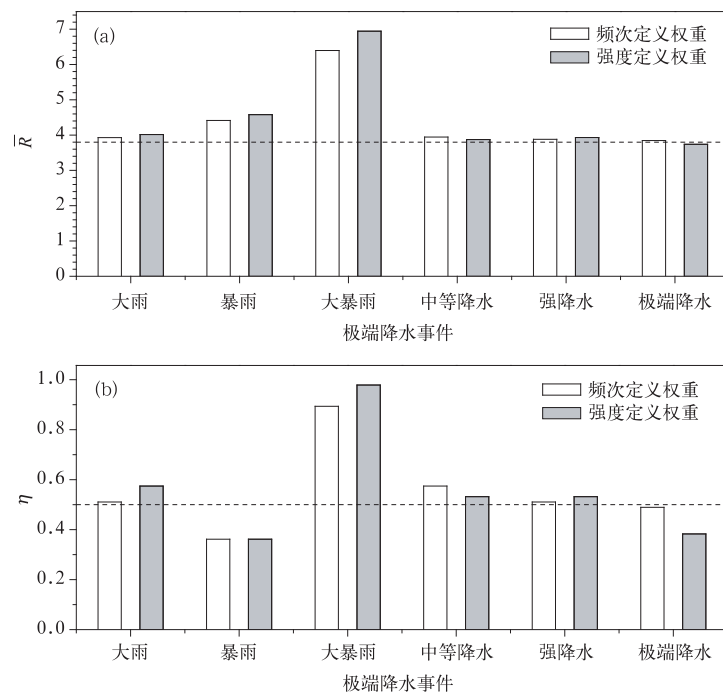


图 4 6 种极端降水事件有效性检测结果图

(a) 极端降水事件的  $\bar{R}$ , (b) 极端降水事件的  $\eta$

Fig. 4 Precipitation extreme's average of ratio  $\bar{R}$  (a) and efficient  $\eta$  (b)

图 4a 和图 4b 分别给出了 6 种极端降水事件  $\bar{R}$  和  $\eta$ , 虚线分别为  $\bar{R}$  和  $\eta$  的临界值。由图 4 可知: 较极端温度事件而言, 极端降水事件的有效性稍弱, 大部分  $\bar{R}$  都达到了  $\bar{R}$  的临界值, 除暴雨和极端强降水外, 其余极端降水事件均超过  $\eta$  的临界值。从  $\bar{R}$  和  $\eta$  的大小看, 大暴雨的有效性高于暴雨和大雨; 百分比阈值定义下极端降水事件中, 中等降水和强降水事件的有效性相对较高。对比  $\bar{R}$  和  $\eta$  在临界值以

上的几种极端降水事件, 基于强度定义权重的结果略好于基于频次的结果, 但相差甚微。

## 4 结 论

极端气候事件的区域群发性是近年来极端事件研究中的新热点, 尽管目前关于极端气候事件群发性的研究已有不同程度的涉及, 将极端事件的区域

群发性特征概括为极端事件的集中度,如 20 世纪 90 年代以后长江流域降水在时间和空间上的集中度都有所增加<sup>[26]</sup>,但是对于群发性或集中度的研究未成系统,对区域群发性进行系统阐述以及定量化研究还不成熟。本文将空间点过程理论  $k$  阶最近邻距离丛集点提取算法与极端事件群发性的研究进行了合理的结合,通过一系列有效性检验探讨了该算法在极端事件区域群发性研究中的应用,为极端气候事件群发性的研究和诊断提供了一个新的视角和思路。得到如下结论:

1) 通过合理利用  $k$  阶最近邻距离丛集点提取算法的基本理论,发现该方法能够将单站点极端气候事件的研究扩展到相互联系的多站点空间分布的讨论中,实现了极端气候事件研究范围的拓展。

2) 基于 12 种极端温度事件和 6 种极端降水事件,将  $k$  阶最近邻距离丛集点提取算法应用于极端温度事件和极端降水事件的研究中,研发出区域群发性极端温度事件和极端降水事件的检测流程。

3) 确定了算法有效性检验的判别指标  $\bar{R}$  和  $\eta$ ,对各种定义下极端气候事件的区域群发性进行有效性检验,检测结果表明: $k$  阶最近邻距离丛集点提取算法适用于极端气候事件的区域群发性研究。

### 参考文献

- [1] 柳艳香,王凌,赵振国,等. 2006 年中国夏季降水预测回顾. 气候变化研究进展,2007,3(4):243-245.
- [2] 陶诗言,卫捷,孙建华,等. 2008/2009 年秋冬季我国东部严重干旱分析. 气象,2009,35(4):3-10.
- [3] 丁一汇,王遵娅,宋亚芳,等. 中国南方 2008 年 1 月罕见低温雨雪冰冻灾害发生的原因及其与气候变暖的关系. 气象学报,2008,66(5):808-825.
- [4] 符娇兰,林祥,钱维宏. 中国夏季分级雨日的时空特征. 热带气象学报,2008,24(4):367-372.
- [5] 秦爱民,钱维宏,蔡亲波. 1960—2000 年中国不同季节的气温分区及趋势. 气象科学,2005,25(4):338-345.
- [6] 张勇,曹丽娟,许吟隆,等. 未来我国极端温度事件变化情景分析. 应用气象学报,2008,19(6):655-660.
- [7] 支蓉,龚志强. 利用幂律尾指数的中国降水突变检测与归因. 物理学报,2008,57(7):4629-4633.
- [8] 龚志强,封国林. 中国近 1000 年旱涝的持续性特征研究. 物理学报,2008,57(6):3920-3931.
- [9] 杨萍,刘伟东,王启光,等. 近 40 年我国极端温度变化趋势和季节特征. 应用气象学报,2010,21(1):29-36.
- [10] 封国林,杨杰,万仕全,等. 温度破纪录事件预测理论研究. 气象学报,2009,67(1):61-74.
- [11] 陈波,史瑞琴,陈正洪. 近 45 年华中地区不同级别强降水的变化趋势. 应用气象学报,2010,21(1):47-54.
- [12] 章大全,杨杰,王启光,等. 中国近 50 年气候破纪录温度事件发生概率分析. 物理学报,2009,58(6):4354-4361.
- [13] 王遵娅,丁一汇. 夏季长江中下游旱涝年季节内振荡气候特征. 应用气象学报,2008,19(6):710-715.
- [14] 侯威,杨萍,封国林. 中国极端干旱事件的年代际变化及其成因. 物理学报,2008,57(6):3932-3940.
- [15] 熊开国,封国林,王启光,等. 近 46 年来中国温度破纪录事件的时空分布特征分析. 物理学报,2009,58(11):8107-8115.
- [16] 杨萍. 近四十年中国极端温度和极端降水事件的群发性研究. 兰州:兰州大学,2009:8-12.
- [17] Byers S D, Raftery A E. Nearest-neighbor clutter removal for estimating features in spatial point processes. *Journal of the American Statistical Association*, 1998,93:577-584.
- [18] 裴韬,周成虎,杨明,等. 混合二维泊松过程的分解算法及其在提取地震丛集模式中的应用. 地震学报,2004,26(1):53-61.
- [19] Pei T, Zhu A X, Zhou C H, et al. A new approach to the nearest-neighbour method to discover cluster features in overlaid spatial point processes. *International Journal of Geographical Information Science*, 2006,19:153-168.
- [20] Pei T, Zhu A X, Zhou C H, et al. Delineation of support domain of feature in the presence of noise. *Computers & Geosciences*, 2007,33:952-965.
- [21] 王启光,侯威,郑志海,等. 东亚区域大气长程相关性. 物理学报,2009,58(9):6640-6650.
- [22] 封国林,董文杰,龚志强,等. 观测数据非线性时空分布理论和方法. 北京:气象出版社,2006:1-227.
- [23] 杨萍,侯威,支蓉. 利用空间点过程提取丛集点算法的适用性研究. 物理学报,2009,58(3):2097-2105.
- [24] CLIVAR, GCOS, WMO. Indices and Indicators for Changes in Climate Extreme. Asheville NC USA, 1997:3-6.
- [25] 翟盘茂,潘晓华. 中国北方近 50 年温度和降水极端事件变化研究. 地理学报,2003,58(增刊):1-10.
- [26] 李吉顺,王昂生,陈家田. 90 年代局地气候变化与长江流域水旱灾害. 中国减灾,2000,10(3):29-31.

## Cluster Extreme Events Based on Point Process Theory

Yang Ping<sup>1)2)</sup> Feng Guolin<sup>2)</sup> Liu Weidong<sup>1)</sup> Hou Wei<sup>2)</sup> Sun Shupeng<sup>2)</sup>

<sup>1)</sup> (*Institute of Urban Meteorology, CMA, Beijing 100089*)

<sup>2)</sup> (*National Climate Center, Beijing 100081*)

### Abstract

Extreme weather and climate events have attracted more attention in the last few years due to the often large loss of human life and exponentially increasing costs coping with them. The assessment of extreme climate events has found that extreme events are showing the characteristic of clustering. Many natural phenomena manifest themselves as spatial point processes which produce numerous events in space, such as extreme events. For clarification, the occurrence of a phenomenon located at a single point is defined as an event in contrast to a simple geometric point. Some events assemble in a restricted region while other events are dispersed. Regarding extreme events as spatial point processes and the cluster events as the clustered points, a method is developed to delineate the cluster events from a region which is called  $k$ -th order nearest-distance spatial point method. By combining spatial point theory with cluster events, the algorithm is validated with a new method which consists of four major steps. The first step is to define different kinds of extreme events including 12 kinds of extreme temperature events and 6 kinds of extreme precipitation events. The second step is to calculate the statistically weights of every station based on the frequency and strength. The third step is to give the spatial distributions of all the stations that extreme events happens and the last step is to pick out the clustered events combining the spatial theory. Based on this method, extreme events research in eastern and middle areas of China have been accomplished, including extreme temperature events and extreme precipitation events. Results of  $k$ -th order nearest-distance, the material definitions of weights and the distribution charts are given, which can detect cluster extreme events effectively. Based on the given charts and two efficient indexes (one is called average of ratio which is represented by  $\bar{R}$ , the other is called efficiency represented by  $\eta$ ), all kinds of extreme climate events are tested to validate the method, finding out that the algorithm is fit for cluster extreme climate events, especially for heavier extreme events. The method performs better on temperature extreme events than precipitation extreme events. So it's concluded that cluster extreme events can be delineated not only by qualitative research, but also by quantitative research, which offers a new idea in extreme events researches.

**Key words:** extreme events; cluster;  $k$ -th order nearest-distance; cluster point