

# 南京过去 100 年极端日降水量模拟研究<sup>\* 1</sup>

万仕全<sup>1</sup> 周国华<sup>1</sup> 潘柱<sup>2</sup> 杨柳<sup>1</sup> 张渊<sup>1</sup>  
WAN Shiquan<sup>1</sup> ZHOU Guohua<sup>1</sup> PAN Zhu<sup>2</sup> YANG Liu<sup>1</sup> ZHANG Yuan<sup>1</sup>

1. 扬州市气象局,扬州, 225009

2. 广东信宜气象局,信宜, 525000

1. *Yangzhou Meteorological Office, Yangzhou 225009, China*

2. *Xinyi Meteorological Office, Xinyi 525000, China*

2008-09-06 收稿, 2009-02-24 改回.

**Wan Shiquan, Zhou Guohua, Pan Zhu, Yang Liu, Zhang Yuan. 2010. A simulative study of extreme daily rainfall in Nanjing for the past 100 years. *Acta Meteorologica Sinica*, 68(6):790-799**

**Abstract** A daily precipitation dataset, for the period from 1905 to 2007 in Nanjing is constructed. Firstly, annual maximum of daily rainfall (AMDR) are modeled by using the generalized extreme value (GEV) distribution from the extreme value theory to describe and predict extreme value of future behavior. We estimate model parameters by the MLE method and evaluate the confidence level with the profile log-likelihood function. Meanwhile, the diagnosis of model's rationality through 4 kinds of visual illustration is made with the result that the Frechet distribution of GEV fits the extreme daily precipitation best. Second, the scope of application of the generalized Pareto distribution (GPD) based on three time series scenarios is studied and a detailed approach how to gather useful extreme information for a given threshold is discussed emphatically. The results show that regardless of length of climate time series, the critical threshold of daily precipitation of 24 mm assumed is appropriate for GPD analysis. This threshold is located near the 91th percentile of annual precipitation series, i. e. above 91% of the sample capacity in the 50 years daily observational data is able to meet the requirement to analyze extreme daily precipitation with the GPD. According to statistical inference of extreme values through the GEV and GPD, it is concluded that the confidence level of GPD is higher than that of the GEV, i. e. with less uncertainty and thus more suitable for climate time series analysis for China where the sample capacity of climate data is not large at the moment yet. We also set additional variables to replace shape and scale parameters in the GPD model and esp. introduce dynamic variables to describe the linear change in order to analyse the variation of precipitation series and its possible influence on the extreme value distribution. This kind of variation includes changes in the long-range mean and the percentile of precipitation series. However, no distinct interfere caused by these variations is found with the analyses.

**Key words** Extreme value theory, Extreme daily rainfall, GEV, GPD

**摘要** 在南京过去 100 年日降水资料的基础上,利用极值理论中的区组模型和阈值模型分析了极端日降水分布特征。首先通过广义极值(GEV)模型模拟了日降水的年极值序列(AMDR),用极大似然估计(MLE)方法计算了模型的参数,并借助轮廓似然函数估计出参数的精确误差区间,同时采用 4 种较直观的诊断图形对模型的合理性进行全面评估,结果表明 Frechet 是区组模型中最适合描述极端日降水分布特征的函数。其次,将日降水序列分 3 种情景构建极值分布的阈值模型(GPD),考察了观测数据的规模对应用该模型的限制,重点讨论了如何针对给定观测样本选择合适的阈值收集极值信息。分析结果认为,长度不小于 50 年的气候序列,采用 24 mm 的日降水量作为临界阈值均能进行 GPD 分析。该阈值处于年降水序列第 91 个百分点附近,即对目前长度为 50 年左右的日观测资料,第 91 个百分点位点以上的数据基本能满足 GPD 研究的需要。另外,根据 GEV 和 GPD 对未来极端降水重现水平的推断情况,GPD 预测值的置信区间要比 GEV 的窄,极值推断的不确定性相对也较

\* 资助课题:国家自然科学基金项目(40P30P52,41005041)、国家科技支撑计划项目(2007BAC29B01)和中国气象局公益性行业科研专项(GYHY(QX)2007-6-26)。

作者简介:万仕全,主要从事气候变化研究。E-mail: wan\_sq@163.com

小,更适合用于研究中国目前规模不大的气候资料。最后,对 GPD 模型的形状参数和尺度参数进行变换,分别引入描述线性变化的动态变量,分析降水序列中潜在的变异行为对极值理论应用的影响。这种变异包括降水序列中长期的均值变化及百分位变化,从模拟结果看,暂未发现资料变异行为对极值分析产生显著干扰。

**关键词** 极值理论, 极端日降水量, 广义极值, 阈值模型

**中图法分类号** P468

## 1 引言

全球变暖导致极端天气气候事件(简称极端事件)愈加频繁,影响越来越广泛(Karl, et al, 1999; Meehl, et al, 2000; Frich, et al, 2002; Zhai, et al, 2005; Fujibe, et al, 2006)。研究表明,气候暖化使降雨总量增大的区域发生极端降水事件的概率上升,即使平均降雨减少,极端降水及其频次也在增加(Houghton, 2001; Karl, et al, 1998; Stone, et al, 2000; Yamamoto, et al, 1999; Manton, et al, 2001; Buffoni, et al, 1999)。如在过去的半个世纪里,虽然中国降水总量的增长趋势并不明显,但强度却在变大(Zhai, et al, 2005)。作为小概率事件,极端事件的预报预测是十分困难的。事实上,天气过程的可预报上限只有两周(Lorenz, 1969),极端事件的可预报性无疑将更低,凭借数值动力模式精确地预报预测极端事件似乎并不现实。不过根据统计理论,人们至少还能通过概率理论方法给出一定可信度范围内影响的情景评估,了解极端事件在不同时空尺度上的强度和频率的分布等,这类研究已逐渐成为许多应用领域的新热点。极端事件本质上是随机过程的极值分布问题,通过极值理论(EVT)的数学模型能描述其分布特征,包括强度与频率等特征参数。这些特征参数变化的实质是其概率分布形态、气候序列动力学结构变化的反映,它们奠定了极值统计推断的物理基础。如,当在已掌握的资料中没有记录到某一严重的极端现象的时候,极值模型能给出发生该事件的风险估计,这在许多领域具有很重要的实用价值。

极端降水研究广受国内外学者的关注,如 Nguyen 等(1998, 2002)利用 GEV 模型研究了加拿大短时降水中的极端事件分布,对非标准观测资料订正问题作了探讨; Aronica 等(2002)、Crisci 等(2002)利用极值理论研究了意大利首都大城市效应对短时最大降水的影响。中国近几年有人利用

Weibull-II 分布型研究月降水分布(Xie, et al, 2005; Su, et al, 2007, 2008),取得了不少成果。但这些研究工作还远不能满足对极端降水规律理解的需要,人们仍需不断拓展更多其他有效分析手段进一步研究极端事件的发生发展机制(Feng, 2008, 2009; Gong, 2009)。从目前研究工作的主要方法看, EVT 中的区组模型使用较普遍,阈值模型的讨论相对较少。而且对应用极值理论时的各种限制条件,如进行 EVT 分析的数据规模,气候变化导致的资料变异问题等缺少足够的重视。尤其在利用短的气候序列时,若对 EVT 的条件处理不当将引起极值分析的不确定性后果。

本文研究数据为南京 1905—2007 年日观测资料,经过质量控制后有 95 年的完整数据可供分析。长的气候序列为开展相关讨论提供了数据基础,文中将结合最优极值模型的选择、极值信息的提取、模型参数的误差估计、资料变异带来的影响作系统分析。

## 2 极值理论

极值理论是概率论的重要分支,主要研究随机样本以及随机过程中极值的概率值及统计推断(Coles, 2001)。研究极值分布规律,首先要建立数学模型。若已知观测数据所服从的底分布,就能建立最大或最小值的精确模型。实际上大多数观测数据所服从的底分布是未知的,人们只能建立渐近的而非精确分布模型。除此之外,观测样本还要满足 3 个条件,一是观测样本是随机变量;随机变量的底分布应保持不变,或能经过变换减少变化带来的影响;观测到的极值要求是独立样本,且有一定的数据规模。与其他统计方法不同,极值统计的主要任务是收集数据,不同的收集方法对应不同的极值分布,区组(Block Maxima, BM)模型和超阈值(Peak Over Threshold, POT)模型是两种主要方法。20 世纪 30 年代初, Dodd(1923)、Frechet(1927)、Fisher

等(1928)开始对极值理论进行研究,证明了极值极限分布的3大类型定理,奠定了发展研究极值理论的基础。为便于应用,Jenkinson(1955)给出基于BM方法的广义极值分布(GEV)。Pickands(1975)首次发表了POT方法的极限分布广义帕雷托分布(GPD)。模型的选择,取决于是否能充分利用所有的极值信息,保证模型参数的合理性,以期最大限度减少不确定性带来的模型不稳定问题。

建立极值模型首先要估计模型的参数,主要采用极大似然估计(MLE)方法。其次要对模型的合理性进行一系列诊断,K-S检验比较简单,但精确性及可靠性不如分位数图(Quantile-Quantile Plot, QQ)、概率图(Probability-Probability Plot, PP)、密度函数图(Density Plot, DP)以及重现水平图(Return Level-Return Period Plot, RR)等方法。构建极值模型的最终目的之一是获取极值重现水平和重现期,假定 $X$ 为连续型随机变量,对任意 $x$ ,当 $X < x$ 时的概率称为左侧概率

$$F(X) = P(X < x) = \int_{-\infty}^x f(x) dx \quad (1)$$

若有

$$F(X) = P(X < x) = \frac{1}{T} \quad (2)$$

则

$$T = \frac{1}{F(X)} \quad (3)$$

与式(1)相对应的右侧概率( $X \geq x$ 的概率)为

$$1 - F(X) = P(X \geq x) = \int_x^{\infty} f(x) dx \quad (4)$$

则

$$T = \frac{1}{1 - F(X)} \quad (5)$$

若 $X$ 代表某气候要素极值变量, $x$ 为其某个可能取值(例如年的最大值或最小值),则式(3)即极小值的重现期,式(5)为极大值的重现期, $X$ 称为重现水平。重现水平和重现期主要用于传递极端事件最大可能性信息。譬如超过年最大降水量 $x$ 的概率为 $P = 0.01$ ,它的重现期即 $T = 1/P = 1/0.01 = 100$  a。这种关系包含两层含义,一是等待时间:对于某一重现期为 $T$ 年的重现水平 $X(P)$ ,下一次发生该事件的平均等待时间为 $T$ 年;二是发生的数量:即发生在 $T$ 年里的该事件的平均数量是1个(Shi, 1995a,

1995b, 1995c)。

### 3 极端日降水的 GEV 模拟

图1给出了南京日降水的年极大值(AMDR)序列,20世纪30—50年代AMDR处于偏小期,60—80年代是集中偏高期。根据线性拟合,AMDR的长期增长趋势并不明显。前3个极大值出现在1953、2003和1931年,日降水量分别达到230、207和198.5 mm。AMDR序列是由95个年极大值组成,等同于区组模型方法从每组含365个样本的观测序列提取的极值信息,这些样本完全符合随机分布的要求,能看作GEV的独立同分布(I. I. D)序列。GEV的一般形式为

$$H(x; \mu, \sigma, \xi) = \exp \left\{ - \left( 1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right\} \quad (6)$$

式中 $1 + \xi(x - \mu)/\sigma > 0$ ,  $\mu, \sigma, \xi$ 分别代表位置、尺度和形状参数。形状参数决定分布形状;尺度参数控制分布值域,起放大或缩小分布面积的作用,但不影响分布的形状;位置参数则描述极值中心位置。当 $\xi = 0, \xi > 0$ 和 $\xi < 0$ 时分别对应Gumbel、Frechet和Weibull分布。GEV的优点是不需要事先指定参数的具体形式,可根据具体估计值进一步判断。假定AMDR样本来自平稳的GEV分布,相应的数似然函数为

$$l(\mu, \sigma, \xi) = -m \lg \sigma - (1 + 1/\xi) \sum_{i=1}^m \lg \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^m \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1/\xi} \quad (7)$$

其中 $1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) > 0, i = 1, 2, \dots, m$ 。通过MLE估计获得参数的估计值 $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ 分别为74.82、27.02和0.139,对数似然函数的最大值为470.92。通过

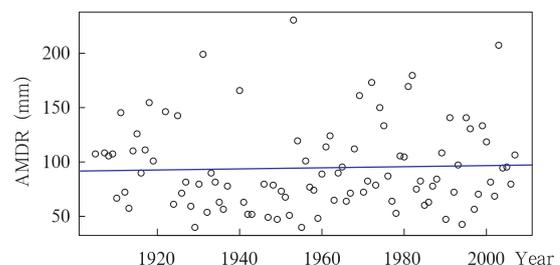


图1 1905—2007年南京AMDR序列  
Fig. 1 AMDR recorded in Nanjing from 1905 to 2007

Fisher 信息的逆矩阵(Fisher, et al, 1928)可得参数的标准误差,分别为 3.24、2.53、0.098。由于该估计值还存在较大的不确定性,因此采用轮廓似然函数获取更精确的置信区间。轮廓似然函数定义为

$$l_p(\varphi_i) = \max_{\varphi_{-i}} l(\varphi, \varphi_{-i}) \quad (8)$$

$\varphi_{-i}$ 代表参数向量中除了分量  $\varphi_i$  的其他分量。式(8)的含义为,给定  $\varphi_i$ ,通过轮廓似然函数对除了  $\varphi_i$  之外的其他分量求最大值。图 2 给出了  $\xi$  的轮廓似然曲线,纵坐标是轮廓对数似然值,横虚线为 95% 的置信区间。由图 2 知,  $\xi_{0.95}$  的精确置信区间为  $[-0.03, 0.35]$ ,大部分位于 0 以上区间。AMDR

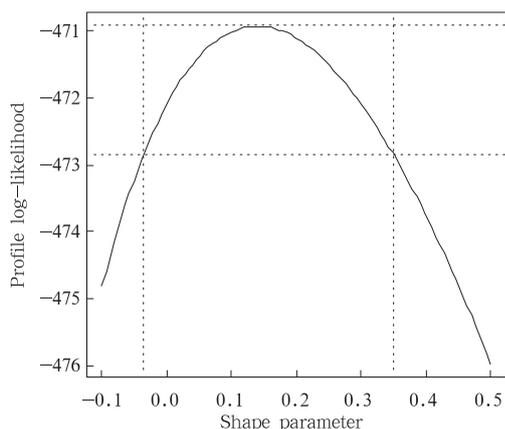


图 2 GEV 形状参数( $\xi$ )的轮廓似然  
Fig. 2 Profile log-likelihood vs. shape parameter ( $\xi$ ) of the GEV

的重现水平根据式(5)给出,得到未来 50 年发生极大值约为 214.6 mm,未来 100 年最大可达 248.5 mm。delt (Rao, 1976)方法能给出重现水平误差的初步估计,但轮廓似然法更精确,估计式为

$$l_p(x_T) = \max_{\sigma, \xi} l(x_T, \sigma, \xi) \quad (9)$$

$x_T$  的  $(1-\alpha)$  置信区间为

$$\left\{ \theta: 2 \lg \left( \frac{l(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{l_p(\theta)} \right) < \chi_{1-\alpha}^2 \right\} \quad (10)$$

其中  $\chi_{1-\alpha}^2$  表示自由度为 1 的卡方(Chi-Square)分布的  $(1-\alpha)$  分位数。图 3 是 AMDR 的重现水平轮廓似然函数曲线,可见信度为 95% 的两个重现水平的误差区间分别为  $[176.5, 304.9]$  和  $[195.2, 388.8]$ 。

根据  $\xi = 0.139 > 0$ ,该 AMDR 序列应属于 GEV 族中的 Frechet 分布。图 4 是模型的诊断图,PP 图的横坐标为经验值,纵坐标为模型值,QQ 图的坐标则相反。由图 4a、4b 可见,除了 QQ 曲线的高分位部分有个别奇异点外,PP 曲线和 QQ 曲线所有的点几乎都落在直线上,两个诊断图均表明模型已很好地拟合了样本,所拟合的模型是正确的。RR 图的重现值曲线向上凹,表明  $\xi > 0$ ,再次验证该序列服从 Frechet 分布;重现水平全部处于置信区间内,构建的模型是稳定可靠的。DP 图给出了 AMDR 的密度分布,与拟合的密度曲线基本吻合,支持样本来自 GEV 分布。

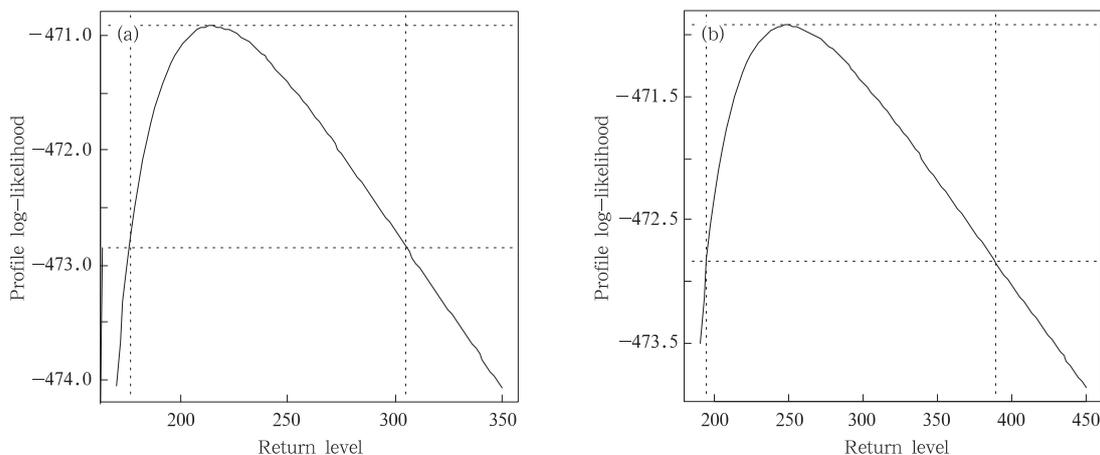


图 3 GEV 模型中 AMDR 重现水平轮廓似然函数 (a.  $T = 50$  a, b.  $T = 100$  a)

Fig. 3 Profile log-likelihood vs.  $T$  year return level in the GEV model  
(a.  $T = 50$  year and b.  $T = 100$  year)

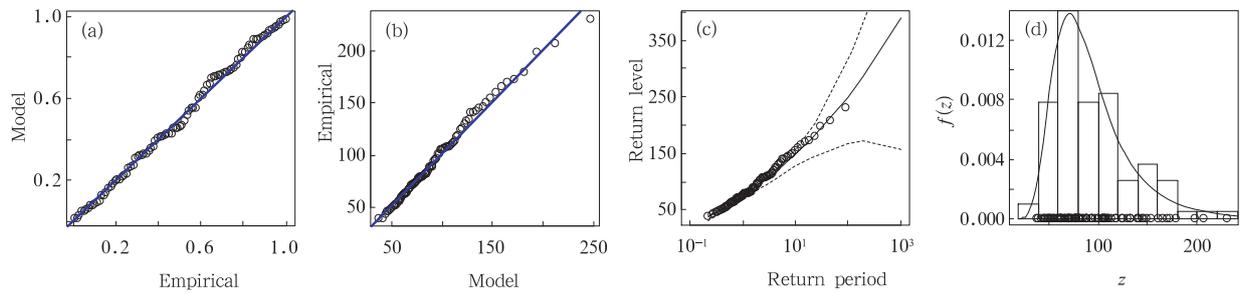


图4 AMDR的GEV拟合诊断

(a. PP 概率图, b. 分位数图, c. 重现水平曲线(实线)和 0.95 置信区间(虚线), d. 密度曲线)

Fig. 4 Diagnostic plots for the GEV fitting AMDR for

(a) the probability; (b) the quantile; (c) the return level (solid line) and the 95% confidence interval (dashed line) and (d) the histogram and the fitted curve for the density (solid line)

#### 4 极端日降水量的 GPD 模拟

GEV 的基础是建立在 BM 模型的 I. I. D 假设上,对观测样本长度的要求相对较高。当观测序列规模不大时,可考虑用 POT 模型做极限分析。给定阈值  $u$ , 超过阈值  $u$  以上的观测值  $X_i$  称为阈值超出量, POT 模型能够有效分析  $X_i$  分布特征。POT 的极限分布函数称为广义帕雷托分布(GPD),即

$$G(x; \mu, \sigma, \xi) = 1 - \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \quad (11)$$

式中  $x \geq \mu, 1 + \xi \frac{x - \mu}{\sigma} > 0$ 。当  $\xi = 0$  时化为指数分布。

$\mu = 0$  时称为标准 GPD 模型,其对数似然函数为

$$l(\sigma, \xi) = n \lg \xi - n \lg \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \lg \left(1 + \frac{\xi}{\sigma} x_i\right) \quad (12)$$

其中  $x_i \in [0, \infty), \xi > 0$  或  $x_i \in [0, -\sigma/\xi], \xi < 0$ 。估计值  $\hat{\sigma}, \hat{\xi}$  没有解析式,要用数值运算求解。

##### 4.1 阈值的选取

选取阈值是极其重要而复杂的过程(Beirlant, et al, 1996),尤其当观测序列较短时,往往要在极值足够大和超出量样本足够多之间保持合理的平衡。太大的阈值将导致可用数据量变少,从而导致参数估计量的方差偏大;阈值太小将使估计量成为有偏估计,不能满足 GPD 分析的条件。那些规模不大的气候观测序列使人们必须综合权衡偏差和方差之间的关系。为获取最佳阈值,下面通过改变阈值考察模型估计量的变化,以此确定合适的取值范围。给定初始阈值  $u_0$ , 利用对应的超出量估计出 GPD

的尺度参数  $\sigma$  和形状参数  $\xi$ ,若初始阈值  $u_0$  对应的超出量近似为 GPD,则对大于  $u_0$  的阈值,尺度参数  $\sigma$  和形状参数  $\xi$  应保持不变。考虑到估计量  $\hat{\sigma}$  是  $u$  的函数,即

$$\hat{\sigma} = \sigma + \xi u \quad (13)$$

因此,通过适当变换可使修正的尺度参数  $\sigma^*$  与  $u$  无关。若  $u_0$  取值合理,相应的超出量应服从 GPD,则对大于  $u_0$  的  $u$ ,估计量  $\sigma^*$  和  $\hat{\xi}$  将保持不变。

中国大部分气象观测序列并不长,研究如何有效挖掘极值信息,合理应用极值理论值有重要意义。利用本文所用资料观测序列较长优势,可分 3 种情况对比分析观测样本长度与极值模型应用的限制问题。将日降水序列划分为 1905—1960 年(简称  $P_1$ )、1961—2007 年( $P_2$ )和 1905—2007 年( $P_3$ )3 种情景分别进行极值模拟,考察资料规模和性质对正确构建极值模型的潜在影响。为直观起见,作  $\sigma^*$  和  $\hat{\xi}$  关于  $u$  的二维图形,从而估计相应的置信区间。图 5 给出了 3 类情形下估计量  $\sigma^*$  和  $\hat{\xi}$  关于  $u$  的变化,其取值的稳定范围见表 1。可见在  $P_1-P_3$  情景下,模型能使  $\sigma^*$  和  $\hat{\xi}$  关于阈值稳定的最小  $u$  值均位于 24(mm)附近,即保持两个参数趋于常数的最小  $u$  值为 24,称为初始阈值,该阈值以上分别有 452、532 和 984 个阈值超出量数据可用于极值的统计推断。通过计算  $P_1, P_2, P_3$  序列的百分位,  $u = 24$  处于 90%—91%的分位数范围内。表 2 给出相应的参数估计,尺度参数和形状参数分别稳定在 18 和 0.16 附近,估计值的误差均较小,表明以目前大部分台站 50 年左右的日观测资料构建 GPD 模型是基本可靠的。

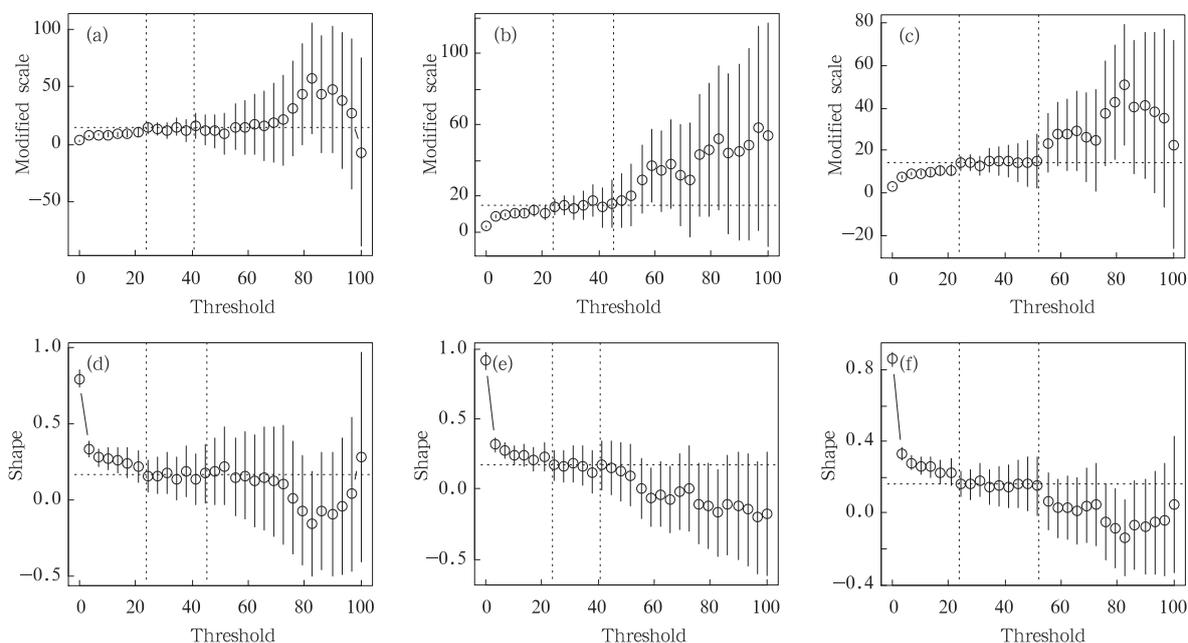


图 5 基于  $P_1-P_3$  序列的 GPD 模型关于阈值的 MLE 估计  
(尺度参数:(a) $P_1$ , (b) $P_2$ , (c) $P_3$ ; 形状参数:(d) $P_1$ , (e) $P_2$ , (f) $P_3$ )

Fig. 5 MLE's evaluation of the threshold based on the GPD model with the threshold range adjusted for the  $P_1 - P_3$  series (scale parameter: (a)  $P_1$ , (b)  $P_2$ , (c)  $P_3$  and shape parameter: (d)  $P_1$ , (e)  $P_2$ , (f)  $P_3$ ) (see text for further details)

表 1 保持 GPD 尺度参数和形状参数稳定的阈值区间  
Table 1 The interval of threshold with the scale parameter and shape parameter in the GPD stable

	1905—1960 年	1961—2007 年	1905—2007 年
修正尺度参数 $\sigma$	24—41	24—45	24—52
形状参数 $\xi$	24—45	24—41	24—52

表 2 基于  $P_1-P_3$  的 GPD 模型的 MLE 估计  
Table 2 The MLE's evaluation based on the GPD model for the  $P_1 - P_3$  series

	$l$	$\hat{\sigma}$	$\hat{\xi}$
$P_1$ (1905—1960)	1820.6	17.5053859(1.26)	0.1656343(0.055)
$P_2$ (1961—2007)	2167.7	18.2931848(1.24)	0.1682736(0.052)
$P_3$ (1905—2007)	3988.5	17.9211065(0.88)	0.1675684(0.038)

注:括号内为标准误差

### 4.2 模型的检验

图 6 为模型的诊断图,其中图 6a 为  $P_1$  情形下的模型诊断,分位数图(QQ)曲线末端有 2 处高分位点略向上偏离直线;图 6b 为  $P_2$  情形的模型诊断,QQ 曲线末端有 1 处高分位点略微向下偏离直线;图 6c 对应  $P_3$  的模型诊断,QQ 曲线也在末端有 1 处高分位点略向下偏离直线。总体上, $P_1-P_3$  对应

的 PP 诊断图所有观测极值和理论推断值相当吻合,重现水平 RR 曲线也处于置信区间内,DP 图与密度曲线体现了很好的一致性。尽管在高分位点 QQ 曲线有个别奇异点,但不会影响模型的整体性质。综合以上诊断结果,当取 24 mm 的阈值时,各模型均能正确反映对应阶段极端日降水的分布特征。因此,对中国其他区域长度为 50 年的日降水序列,类似地应能获取一个恰当的阈值进行极值研究。

### 4.3 日降水极值的重现水平

建立极值模型的重要目的是估计极值的高分位数或重现水平,根据式(5)可得 GPD 分位数估计式(Davison, et al, 1990)为

$$x_p = u + \frac{\sigma}{\xi} \left( \left( \frac{n}{N_u} (1-p) \right)^{-\xi} - 1 \right) \quad (14)$$

其中  $N_u$  表示阈值超出量个数。表 3 给出了重现水平的估计结果,括号内为 95% 的置信区间,均由轮廓似然函数(式(9)、(10))估计得到(图 7)。第一种情形(1905—1960 年)预测未来 50 年内(即 1961—2010 年)南京有一次大小为 211 mm 左右的极端日降水出现,其所处分位点为 99.97%,而 95% 的置信

区间是 167.7—296.6 mm。在 1961—2007 年出现在 [167.7, 296.6] 内的实测极值共有 4 次: 172.5

(1972 年)、169.3(1981 年)、179.3(1982 年)、207.0(2003 年)。可见, 最大值 207.0 mm 与所估计的重

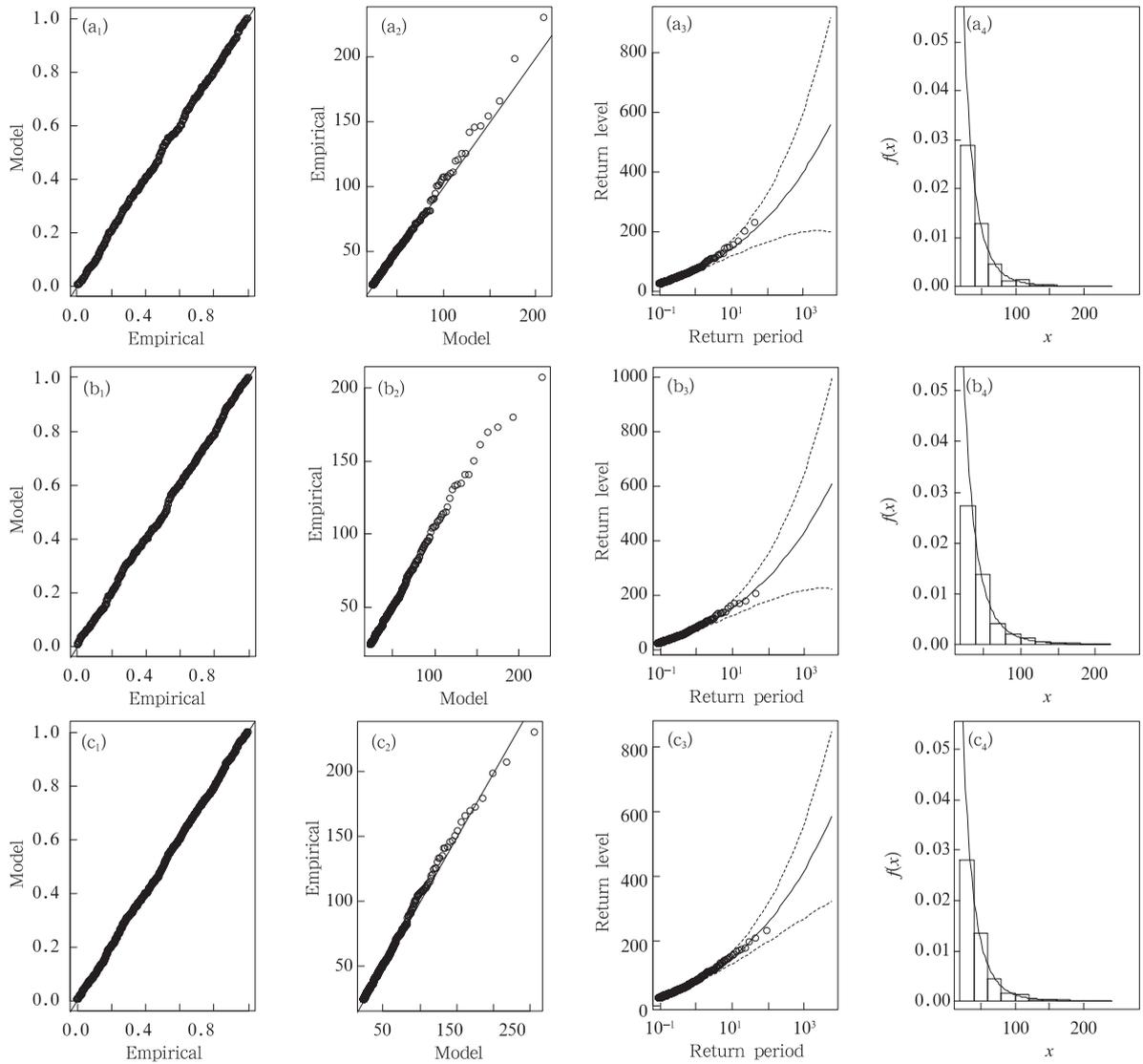


图 6  $P_1 - P_3$  情形下日降水取阈值  $u=24$  的 GPD 模型诊断(说明同图 4)

( $a_1 - a_4$ )  $P_1$ : 1905—1960; ( $b_1 - b_4$ )  $P_2$ : 1961—2007; ( $c_1 - c_4$ )  $P_3$ : 1905—2007

Fig. 6 Diagnostic plots for the GPD model with  $u=24$  (see Fig. 4 for the relevant plots' caption)

( $a_1 - a_4$ )  $P_1$ : 1905—1960; ( $b_1 - b_4$ )  $P_2$ : 1961—2007 and ( $c_1 - c_4$ )  $P_3$ : 1905—2007

表 3 GPD 与 GEV 的重现水平对比(括号内为 95% 的置信区间)

Table 3 Contrast of return levels between the GPD and GEV model

(those with the values significant at the 95% confidence level included in bracket)

模型	$T = 50$		$T = 100$	
	重现水平 $x_T$	对应年份	重现水平 $x_T$	对应年份
$P_1$	211.4[167.7, 296.6]	1961—2010	246.9[188.3, 367.9]	1961—2060
GPD $P_2$	231.3[183.0, 322.8]	2008—2057	270.2[205.7, 398.9]	2008—2107
$P_3$	222.6[186.6, 278.2]	2008—2057	260.4[211.7, 338.2]	2008—2107
GEV	214.6[176.5, 304.9]	2008—2057	248.5[195.2, 388.8]	2008—2107

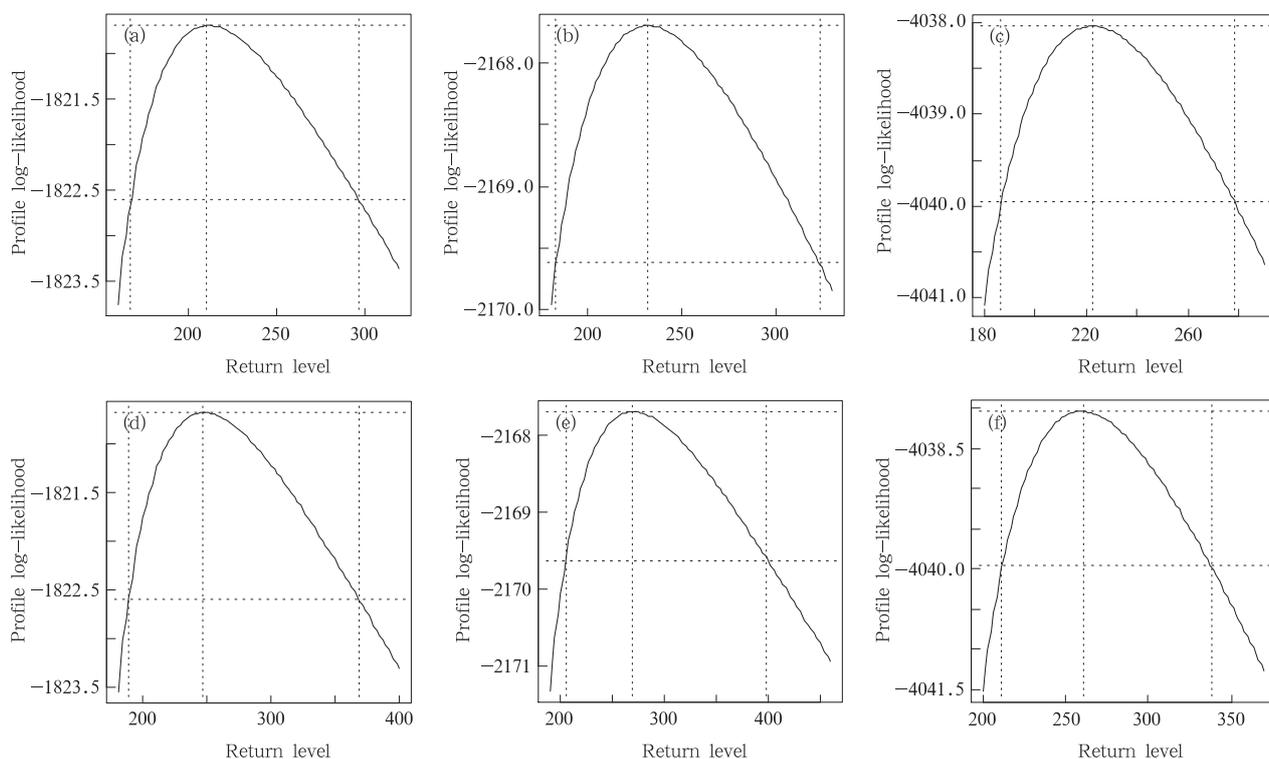


图 7  $P_1-P_3$  情形下 GPD 模型中日降水的 50 年(a—c)和 100 年(e—f)重现水平的轮廓似然函数 (a,d.  $P_1$ ; b,e.  $P_2$ ; c,f.  $P_3$ )

Fig. 7 Profile log-likelihood plot for  $T = 50/100$  year return level in the GPD model (a, d):  $P_1$ ; (b, e):  $P_2$  and (c, f):  $P_3$

现水平(211 mm)基本相符。为便于分析,表 3 将 GEV 的重现水平估计值与 GPD 对照,从估计值的大小看,GEV 与 GPD(见  $P_3$ )50 年的重现值相差 8 mm,100 年的重现水平值相差 12 mm,估计的相近程度比较令人满意;从  $x_T = 50$  对应置信区间看,GPD 的误差估计区间明显要小,GPD 估计误差约为 46 mm,GEV 模型约为 64 mm;从  $x_T = 100$ 、 $x_T = 50$  对应置信区间看,GPD 和 GEV 分别为 63 和 91 mm 左右。可见后者的置信区间明显要宽,不确定性会更大。由图 7 还能发现轮廓似然函数曲线为非对称形状,表明重现水平取高值时的不确定性大于取低值时的不确定性,换言之,置信区间的下界比上界要可靠得多。另外, $P_3$  序列重现水平的置信区间较  $P_1$ 、 $P_2$  的小,即观测序列长度与极值信息成正比,模型参数估计量的精度也会更高。

### 5 资料变异对极端降水模拟的影响

Mearns 等(1984)研究指出,气候要素原始分布的均值变化可导致极值频率和强度呈非线性变化,

即平均气候的微小变化可能引发极端气候值频率的很大变化。事实上,这种气候变异行为很容易导致观测数据的非平稳问题。处理非平稳问题目前没有有效的方法,本节只对观测数据长期线性变化可能导致的极值分布问题作一简单分析。根据日降水的有关统计指标(图 8),其中图 8a 是南京年降水的日数序列,每年出现降水日的平均数为 120 d,20 世纪 50—60 年代有较明显峰值,但总体线性趋势不明显。图 8b 是南京年总降水量序列,平均值为 1013 mm 左右,线性拟合有 1 mm/a 的递增量,100 年的变化幅度  $100 \times 1/1013 \times 100\% = 9.8\%$ ,增长趋势也较弱。图 8c 中圆圈由年降水序列计算第 90—98 个百分位数所得平均值,对应的累积平均百分位数的年平均降水量为 21 mm 左右。从下至上的虚直线分别为第 90—98 个百分位数的趋势线,实线对应累积平均趋势,可见每年同一百分位的降水量递增量为 0.052 mm,100 年的变化幅度约为  $100 \times 0.052/21 \times 100\% = 24.7\%$ 。因此百分位的增长趋势是比较明显的,某种程度上揭示了该地区平均日降水强

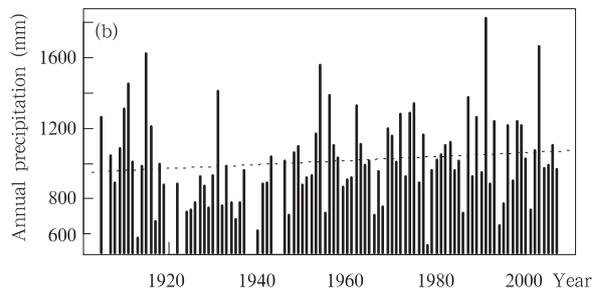
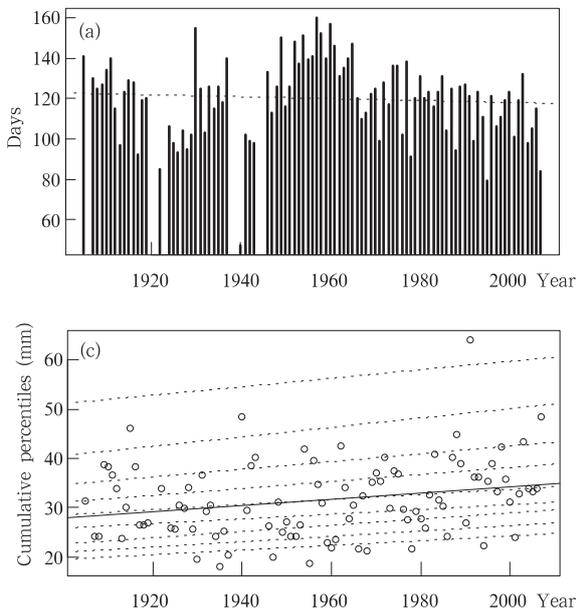


图8 (a)南京年降水日数(虚线为趋势线);  
(b)年降水总量(虚线为趋势线);  
(c)第90—98个分位数的趋势线(虚线)及  
累积分位数平均(实线和圆点)

Fig. 8 (a) Annual rainfall days (dash line shows trend);  
(b) annual precipitation amounts (dash line shows trend);  
(c) the trend from 90th to 98th percentiles for the  
daily rainfall (dash line) and the average of cumulative  
percentiles (solid line and dotted line) in Nanjing

度随着气候变化在逐渐增大,即,尽管平均降水总量虽无明显变化,但出现极端降水的概率在上升。

为了考察这种气候变化行为对极端日降水的影响,下面以 GPD 模型为例,设定 4 种情景,分别在模型的参数中引入线性趋势项,即 GPD1:  $u$  为常数,  $\sigma$  为常数,  $\xi$  为常数; GPD2:  $u$  为常数,  $\sigma = \sigma_0 + \sigma_1 t$ ,  $\xi$  为常数; GPD3:  $u$  为常数,  $\sigma$  为常数,  $\xi_0 = \xi + \xi_1 t$ ; GPD4:  $u$  为常数,  $\sigma = \sigma_0 + \sigma_1 t$ ,  $\xi = \xi_0 + \xi_1 t$ 。其中,  $t$  代表时间变化,  $u$  为阈值。

参数估计仍然采用 MLE 方法,表 4 给出了固定参数和动态参数估计值在 4 个模型中的对比结果,似然函数最大值  $l$  稳定在 3988.5 左右,最大似然函数维持不变的特征,意味着动态的尺度参数  $\sigma$  和形状参数  $\xi$  没有解释出预想的变异特点,在极值分析过程中,可以忽略上述变化的影响,不必考虑该序列中的弱线性增长趋势和极值百分位变化,日降

水极值分布特征完全可以利用平稳的 GPD 模型进行分析。

### 6 结 论

本文利用南京一个世纪以来的日降水资料构建了不同类型的极值模型,从 3 个方面分析了极端日降水的分布特征。首先,通过区组方法建立 GEV 模型,研究了南京 AMDR 分布特点,确定了描述 AMDR 的理想函数是 Frechet 分布。其次,根据极值模型与资料规模适用性关系,分 3 类资料情形讨论了模型构建问题,并重点讨论了如何选择一个合适的阈值最大限度获取极值信息,并建立极端日降水的阈值模型。模拟结果令人满意,即 3 种情形下,各模型参数差别不大,表明中国 50 年左右的降水序列能基本满足 GPD 分析的条件,同时认为在南京地区取 24 mm 以上的日降水量作为 GPD 模型的极值推断样本是可靠的,这一阈值对应该地区第 90%—91% 区间。最后分析了日降水数据的长期变化可能导致的极值理论应用问题。根据对南京日降水序列统计,近百年来总降水量有 9.8% 左右的波动,年降水在第 90% 分位数以上的观测值有较明显的线性增长趋势,平均波动幅度约 24.7%。通过引入线性效应项的 GPD 模型考察了这种变异对极端日降水分布的潜在影响,结果表明这些气候变化不会影响平稳模型对极端降水的分析。

表 4 不同参数形式 GPD 模型的最大对数似然函数值  
Table 4 The maximum log-likelihood function  
in the GPD with the different type parameter

模型	参数形式			参数 个数	$l$
	阈值 $u$	尺度参 数 $\sigma$	形状参 数 $\xi$		
GPD1	固定	固定	固定	2	3988.58
GPD2	固定	线性	固定	3	3988.57
GPD3	固定	固定	线性	3	3988.56
GPD4	固定	线性	线性	4	3988.56

## 参考文献

- 封国林, 龚志强, 支蓉. 2008. 气候变化检测与诊断技术的若干新进展. *气象学报*, 66(6): 892-905
- 封国林, 杨杰, 万仕全等. 2009. 温度破纪录事件预测理论研究. *气象学报*, 67(1): 61-74
- 苏布达, Gemmer M, 姜彤等. 2007. 1960—2005 年长江流域降水极值概率分布特征. *气候变化研究进展*, 3(4): 208-213
- 苏布达, 姜彤. 2008. 长江流域降水极值时间序列的分布特征. *湖泊科学*, 20(1): 123-128
- 谢志清, 姜爱军, 杜银等. 2005. 长江三角洲强降水过程年极值分布特征研究. *南京气象学院学报*, 28(2): 267-274
- 龚志强, 支蓉, 封国林等. 2009. 中国近 530 年干湿变化及其持续性特征研究. *气象学报*, 67(2): 307-320
- Aronica G, Cannarozzo M, Noto L. 2002. Investigating the changes in extreme rainfall series recorded in an urbanised area. *Water Sci Tech*, 45(2): 49-54
- Beirlant J, Vynckier P, Teugels J L. 1996. Tail index estimation, Pareto quantile plots, and regression diagnostics. *J Amer Statistical Association*, 91(436): 1659-1667
- Buffoni L, Maugeri M, Nanni T. 1999. Precipitation in Italy from 1833 to 1996. *Theoretical Appl Climatol*, 63(1): 33-40
- Coles S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer, 224pp
- Crisci A, Gozzini B, Meneguzzo F, et al. 2002. Extreme rainfall in a changing climate: regional analysis and hydrological implications in Tuscany. *Hydrological Processes*, 16(6): 1261-1274
- Davison A C, Smith R L. 1990. Models for exceedances over high thresholds. *J Roy Statistical Soc*, 52(3): 393-42
- Dodd E L. 1923. The greatest and the least variate under general laws of error. *Transactions Amer Math Soc*, 25(4): 525-539
- Fisher R A, Tippett L H C. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Soc*, 24(2): 180-190
- Frechet M. 1927. Sur la loi de probabili de l'ecart. maximum. *Ann, Soc, Polonaise Math, Cracow*, 6: 93-116
- Frich P, Alexander L V, Della-Marta P, et al. 2002. Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Res*, 19(3): 193-212
- Fujibe F, Yamazaki N, Kobayashi K. 2006. Long-term changes of heavy precipitation and dry weather in Japan (1901 - 2004). *J Meteor Soc Japan*, 84(6): 1033-1046
- Houghton J T. 2001. *Climate change 2001: The Scientific Basis*. London & Cambridge: Cambridge University Press
- Jenkinson A F. 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart J Roy Meteor Soc*, 81(348): 158-171
- Karl, T. R., Easterling, D. R. 1999. Climate extremes: Selected review and future research directions. *Clim Change*, 42(1): 309-325
- Karl T R, Knight R W. 1998. Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull Amer Meteor Soc*, 79(2): 231-241
- Lorenz E N. 1969. Atmospheric predictability as revealed by naturally occurring analogues. *J Atmos Sci*, 26(4): 636-646
- Manton M J, Della-Marta P M, Haylock M R, et al. 2001. Trends in extreme daily rainfall and temperature in Southeast Asia and the South Pacific: 1961 - 1998. *Int J Climat*, 21(3): 269-284
- Mearns L O, Katz R W, Schneider S H. 1984. Extreme high-temperature events: changes in their probabilities with changes in mean temperature. *J Appl Meteor*, 23(12): 1601-1613
- Meehl G A, Karl T, Easterling D R, et al. 2000. An introduction to trends in extreme weather and climate events: observations, socioeconomic impacts, terrestrial ecological impacts, and model projections. *Bull Amer Meteor Soc*, 81(3): 413-416
- Nguyen V T V, Nguyen T D, Wang H. 1998. Regional estimation of short duration rainfall extremes. *Water Sci Tech*, 37(11): 15-19
- Nguyen V T, Nguyen T D, Ashkar F. 2002. Regional frequency analysis of extreme rainfalls. *Water Sci Tech*, 45(2): 75-81
- Pickands J. 1975. Statistical inference using extreme order statistics. *Ann Statist*, 3(1): 119-131
- Rao C R. 1976. *Linear Statistical Inference and Its Applications*. New York: John Wiley and Sons, 625pp
- Shi D. 1995a. Fisher information for a multivariate extreme value distribution. *Biometrika*, 82(3): 644-649
- Shi D. 1995b. Moment estimation for multivariate extreme value distribution. *Appl Math-A J Chinese Univ*, 10(1): 61-68
- Shi D. 1995c. Multivariate extreme value distribution and its fisher information matrix. *Acta Mathematicae Appl Sinica*, 11(4): 421-428
- Stone D A, Weaver A J, Zwiers F W. 2000. Trends in Canadian precipitation intensity. *Atmos Ocean*, 38(2): 321-347
- Yamamoto R, Sakurai Y. 1999. Long-term intensification of extremely heavy rainfall intensity in recent 100 years. *World Resource Rev*, 11(2): 271-282
- Zhai P, Zhang X, Wan H, et al. 2005. Trends in total precipitation and frequency of daily precipitation extremes over China. *J Climate*, 18(7): 1096-1108