

DOI:10.3969/j.issn.1672-1144.2015.04.011

# 极值指数估计在太原站月降水频率 分析中的应用研究

李 扬

(山西省水利水电科学研究院, 山西 太原 030002)

**摘 要:** 在介绍极值理论基础、重尾分布判别方法的基础上, 选用3种常用极值指数估计量, 以山西省太原站月降水序列为例研究极值指数估计方法在月降水频率分析中的应用。结果表明: 太原站月降水序列的样本分布属于重尾分布; 选取序列强降水部分进行拟合时, Moment估计量对该段经验点据的拟合效果相对最佳, 且计算简便, 可为当地月降水序列的频率分析提供参考。

**关键词:** 极值指数; 月降水序列; Hill估计量; Pickands估计量; Moment估计量

中图分类号: P333.1

文献标识码: A

文章编号: 1672-1144(2015)04-0055-05

## Research on the Application of Extreme Value Index Estimation in the Monthly Precipitation Frequency Analysis of Taiyuan Station

LI Yang

(Shanxi Institute of Water Resources and Hydropower Research, Taiyuan, Shanxi 030002, China)

**Abstract:** This paper studied the extreme value index estimation methods and their application in monthly precipitation frequency analysis. Based on the introduction of extreme value theory and discrimination method of heavy-tailed distribution, three kinds of EVI estimators were selected. Monthly precipitation series of Taiyuan station in Shanxi province was chosen as an example for application of extreme value index estimation methods in monthly precipitation frequency analysis. The results showed that the distribution of monthly precipitation series of Taiyuan station was heavy-tailed distribution. Heavy precipitation section of series was best fitted by Moment estimator. Simple in calculation, high in accuracy, Moment estimator can provide reference for monthly precipitation frequency analysis in Taiyuan.

**Keywords:** extreme value index; monthly precipitation series; Hill estimator; Pickands estimator; Moment estimator

暴雨、洪水、干旱等极值水文事件对人类社会的危害巨大。近年来, 这类极值事件不断发生, 制约了社会经济发展, 威胁人民生命财产安全。因此, 揭示极值事件中蕴含的统计规律十分必要。丁裕国<sup>[1]</sup>指出, 相对正态分布而言, 月降水量的概率分布呈偏态, 极端降水的变化主要取决于其概率分布的尾部特征。邹波<sup>[2]</sup>的研究表明, 月降水量的概率分布是一种重尾分布, 分布尾部偏厚使得常用概率分布难以拟合强降水部分, 导致对强降水事件发生的概率估计不足。因此, 定量描述极值事件分布的尾部特

征及估算极值事件的分位数, 对研究极值事件统计规律、合理估计极值事件发生概率具有重要意义。

极值事件分布的重尾程度可用极值指数 (Extreme Value Index, EVI)  $\gamma$  来表示。  $\gamma$  估计量的精度取决于所选取的次序统计量个数  $k$ ,  $k$  值偏大或偏小, 都可能导致较大的估计误差。其确定方法分经验方法和理论方法两类。经验方法需要事先得知分布的重尾程度, 从而划定  $k$  值的范围, 而实际应用中分布的重尾程度是未知的, 这种矛盾使得  $k$  值无法由经验方法获得。理论方法的重要代表之一是作

图法,通过绘制 $(k, \hat{\gamma}(k))$ 曲线来确定 $k$ 值。这类图形包括 Hill - plot<sup>[3]</sup>、AltHill - plot<sup>[4]</sup>、SmooHill - plot<sup>[5]</sup>、qq - plot<sup>[6]</sup>、修正 qq - plot<sup>[7]</sup>、Sum - plot<sup>[8]</sup>、Pareto 分位数图<sup>[9]</sup>等,上述方法大多数对随机变量的概率分布较为敏感,当随机变量不服从 Pareto 分布时效果较差,而 Sousa<sup>[8]</sup>提出的 Sum - plot 方法精度较高,操作相对容易,优于其余几种作图法。另一类方法是利用估计的均方误差(MSE)、渐近均方误差(AMSE)等指标,认为使 MSE、AMSE 最小的 $k$ 值最优。由于 MSE 和 AMSE 中尚含有未知参数, Hall<sup>[10]</sup>、Danielsson 等<sup>[11]</sup>、刘维奇等<sup>[12]</sup>先后提出了 Bootstrap 方法、Danielsson Bootstrap 法、M - Bootstrap 方法来获得最优 $k$ 值;刘维奇和邢红卫<sup>[13]</sup>提出了 $k$ 的简便优化估计法。

极值指数的估计和应用在金融领域应用较为成熟和普遍,而在水文学领域应用不多。因此,本文将极值指数引入降水事件的研究,并以山西省太原站为例,探讨了极值指数估计及设计值计算在月降水序列频率分析中的应用。

## 1 极值理论基础

设随机变量 $X_1, X_2, \dots, X_n$ 服从分布 $F, F(x) = P(X \leq x), X_{1,n} \leq X_{2,n} \leq \dots X_{n,n}$ 为 $X_1, X_2, \dots, X_n$ 的次序统计量, $\bar{F}(x) = P(X > x) = 1 - F(x)$ 称为尾分布。对 $\gamma \in R$ ,若存在常数序列 $a_n > 0, b_n \in R$ ,使得 $X_{n,n}$ 收敛于非退化极值分布 $H_\gamma(x)$ <sup>[14]</sup>:

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) \rightarrow H_\gamma(x) \quad (1)$$

$$H_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}); & 1 + \gamma x > 0, \gamma \neq 0 \\ \exp(-e^{-x}) & ; \gamma = 0 \end{cases} \quad (2)$$

则分布 $F \in F_{\text{mda}}(H_\gamma)$ ,MDA 为最大吸引域(maximum domain of attraction)。 $\gamma$ 即为分布 $F$ 的极值指数,可用于描述分布的重尾程度,即给定概率分布情况下极值事件的极端程度和频繁程度。

### 1.1 极值指数 $\gamma$ 的估计量及相应极值分位数

极值指数 $\gamma$ 的准确估计是拟合分布尾部的关键,决定着极值事件分位数的估计量理想与否。目前,国外学者提出了 $\gamma$ 的多种估计量,常用的估计量有 Hill 估计, Pickands 估计、Moment 估计等。

#### 1.1.1 Hill 估计<sup>[15]</sup>

$$\hat{\gamma}_n^H = \frac{1}{k} \sum_{i=1}^k \lg X_{n-i+1,n} - \lg X_{n-k,n} \quad (3)$$

式中, $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ 为 $X_1, X_2, \dots, X_n$

的次序统计量, $k$ 为所选取的极大次序统计量的个数。当超过概率 $p$ 接近 0 时,极值分位数为

$$\hat{x}_{k,n}^H = X_{n-k} \left( \frac{k+1}{(n+1)p} \right)^{\hat{\gamma}_n^H} \quad (4)$$

Hill 估计容易实现且渐近无偏,在实际中最为常用,但当分布与 Pareto 分布相差较远时会导致 Hill 估计的偏差较大。

#### 1.1.2 Pickands 估计<sup>[16]</sup>

$$\hat{\gamma}_n^P = \frac{1}{\lg 2} \lg \frac{X_{n-k+1,n} - X_{n-2k+1,n}}{X_{n-2k+1,n} - X_{n-4k+1,n}}, \quad k \leq \frac{n}{4} \quad (5)$$

当超过概率 $p$ 接近 0 时,极值分位数为:

$$\hat{x}_{k,n}^P = X_{n-k+1,n} + \frac{X_{n-k+1,n} - X_{n-2k+1,n}}{1 + 2^{-\hat{\gamma}_n^P}} \left[ \left( \frac{k}{(n+1)p} \right)^{\hat{\gamma}_n^P} - 1 \right] \quad (6)$$

Dekkers 和 De Haan<sup>[17]</sup>指出: $k$ 值对 Pickands 估计量及其相应分位数的影响很大,不适宜用于中小型样本。Dekkers 等<sup>[18]</sup>和 Drees<sup>[19]</sup>也指出了 Pickands 估计量的不稳定性。

#### 1.1.3 Moment 估计<sup>[18]</sup>

$$\hat{\gamma}_{k+1}^M = M_{k+1}^{(1)} + 1 - \frac{1}{2} \left( 1 - \frac{(M_{k+1}^{(1)})^2}{M_{k+1}^{(2)}} \right)^{-1}, \quad k = 2, \dots, n-1 \quad (7)$$

式中,

$$M_{k+1}^{(l)} = \frac{1}{k} \sum_{i=1}^k (\lg X_{n-i+1,n} - \lg X_{n-k,n})^l \quad l = 1, 2 \quad (8)$$

当 $l = 1$ 时, $M_{k+1}^{(1)}$ 即为 Hill 估计量。

用 Moment 估计量表示的极值分位数为<sup>[17-18]</sup>

$$\hat{x}_{p,k+1}^M = X_{n-k,n} + \hat{a}_{n,k+1}^M \frac{\left( \frac{k}{np} \right)^{\hat{\gamma}_{k+1}^M} - 1}{\hat{\gamma}_{k+1}^M}, \quad k < n \quad (9)$$

式中, $\hat{a}_{n,k+1}^M = \frac{X_{n-k,n} M_{k+1}^{(1)}}{\rho_1(\hat{\gamma}_{k+1}^M)}$ ;

$$\rho_1(\gamma) = \begin{cases} 1 & \gamma \geq 0 \\ \frac{1}{1-\gamma} & \gamma < 0 \end{cases}$$

Moment 估计量比 Hill 估计量的适用范围更广,但其估计方差更大,且仅在 Pareto 分布下渐近无偏<sup>[20]</sup>。

### 1.2 Sum - plot 法

刘维奇等<sup>[21]</sup>的应用实例表明,Sousa<sup>[8]</sup>提出的 Sum - plot 法精度高于 Bootstrap 法和 M - Bootstrap 法。因此,本文选用 Sum - plot 法确定阈值 $k$ 。

Sum - plot 法是通过绘制 $\{(k, S_k), 1 \leq k \leq n\}$

曲线并观察其在哪一点偏离直线,从而认为该点对应的  $k$  值最优。

$$S_k = \sum_{i=1}^k i \cdot (\lg X_{i,n} - \lg X_{i+1,n}) = \sum_{i=1}^k (\lg X_{i,n} - \lg X_{k+1,n}) \quad (10)$$

式中,  $X_{1,n} \geq X_{2,n} \geq \dots \geq X_{k+1,n}$  为次序统计量。最优  $k$  值对应的图形直线部分斜率即极值指数  $\gamma$ 。

### 2 应用实例

选取山西省太原站 1951 年—2013 年月降水序列为基础数据,数据来源为中国气象科学数据共享服务网。

#### 2.1 分布判别

点绘  $(\lg \frac{n+1}{j}, \lg X_{n-j+1,n})$  得到的 Pareto 分位数图可用来判断一组数据是否服从重尾分布<sup>[22]</sup>。其中  $n$  为样本容量,  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  为顺序统计量,  $j = 1, \dots, n$ 。若 Pareto 分位数图中的散点随  $\lg \frac{n+1}{j}$  值增大趋于线性,则认为该组数据服从重尾分布。

图 1 为太原站月降水序列的 Pareto 分位数图,其右上角部分近似线性,可认为该序列服从重尾分布,满足进行极值指数估计的前提条件。

#### 2.2 不同估计量的性质及阈值 $k$ 的确定

分别用 Hill 估计量、Pickands 估计量及 Moment 估计量计算太原站月降水序列的  $\gamma$  并绘制  $\gamma - k$  曲线(图 2)。由图 2 可知:  $\gamma_{\text{Hill}} - k$  曲线随  $k$  值增大呈缓慢上升趋势;  $\gamma_{\text{Pickands}} - k$  曲线随  $k$  值增大先波动

后上升;  $\gamma_{\text{Moment}} - k$  曲线则能呈现平稳状态。理论上,  $\gamma - k$  曲线趋于稳定的一点即为阈值  $k$  的位置,但仅靠观察无法准确判断该点位置,主观任意性较大,故通过绘制 Sum - plot 图(图 3(a))辅助判别。由图 3(a)可知,随着  $k$  值的增大,Sum - plot 图形由直线渐变为曲线。选取线性部分的  $S_k$  进行二次作图(图 3(b)),根据  $(k, S_k)$  曲线偏离直线的点的位置确定最优  $k$  值,并注于图中。

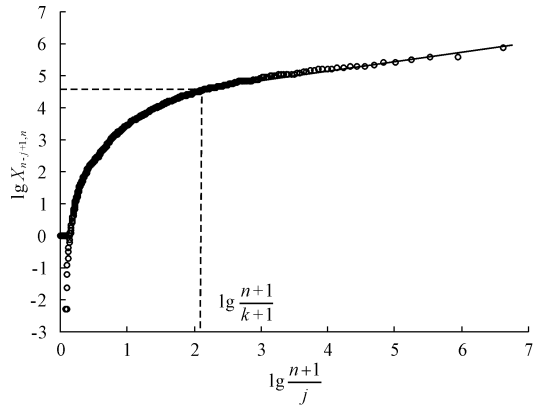


图 1 太原站月降水序列 Pareto 分位数图

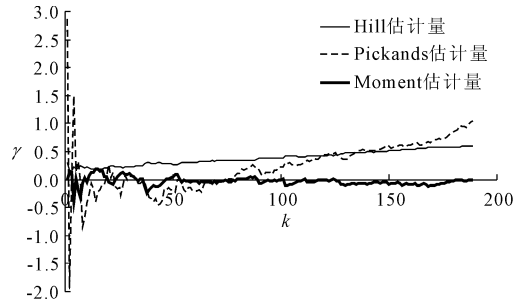


图 2 太原站月降水序列  $\gamma - k$  曲线

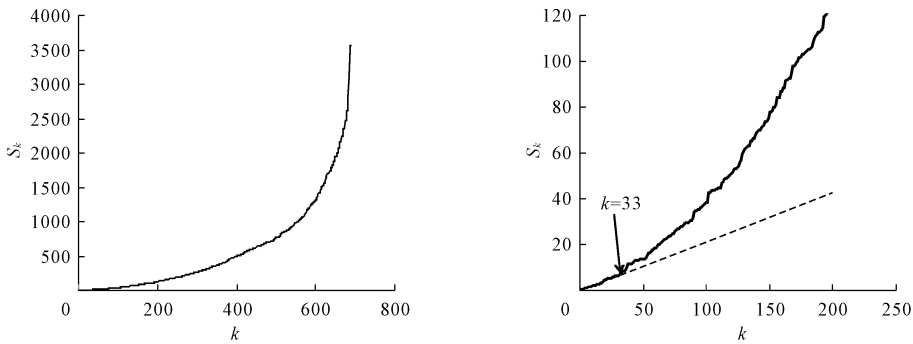


图 3 太原站月降水序列 Sum - plot 图

#### 2.3 极值指数估计

用选定的  $k$  值分别估计  $\gamma_{\text{Hill}}$ 、 $\gamma_{\text{Pickands}}$ 、 $\gamma_{\text{Moment}}$ , 结果如表 1 所示。

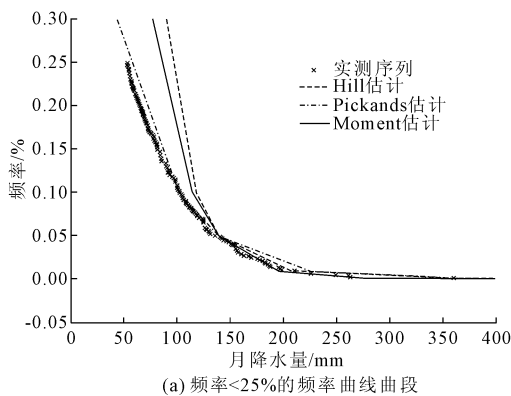
表 1 太原站极值指数估计量

站名	最优 $k$ 值	$\gamma_{\text{Hill}}$	$\gamma_{\text{Pickands}}$	$\gamma_{\text{Moment}}$
太原	33	0.2435	0.0078	0.0096

由表 1 可以看出,三种估计量的值与正态分布 ( $\gamma = 0$ ) 相比均存在不同程度的差别:  $\gamma_{\text{Hill}}$  的估计值显著偏大,  $\gamma_{\text{Pickands}}$  和  $\gamma_{\text{Moment}}$  的估计值较为接近。根据引言中关于重尾分布的描述,可以认为太原站的月降水序列服从重尾分布,与 2.1 所得结论一致。

## 2.4 极值分位数计算及强降水拟合

给定不同的频率,将  $\gamma$  的估计值分别代入式 (4)、式 (6)、式 (9),即可计算出指定频率下的设计



值。根据极值理论对极值事件风险的度量特性,本文对降水序列概率分布的研究重点集中在其尾部,即概率小、降水强度大时设计值与实测值的拟合情况。故选取太原站月降水序列中概率  $< 25\%$  的一段,观察强降水部分的拟合效果。

分别采用 Hill 估计量、Pickands 估计量、Moment 估计量计算太原站月降水序列的理论频率,并绘制理论频率曲线与经验点据拟合图,如图 4 所示。

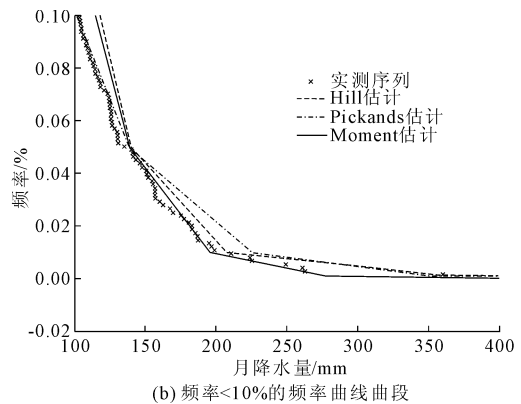


图 4 太原站月降水序列理论频率曲线

比较不同  $\gamma$  估计量计算出的理论频率曲线对经验点据的拟合情况可知:当频率介于 25% 和 5% 之间时, Pickands 估计量对应的理论频率曲线对经验点据的拟合效果最为理想, Moment 估计次之, Hill 估计相对最不理想;当频率  $< 5\%$  时, Moment 估计量所得理论频率曲线对样本序列的较大值拟合十分理想,经验点据几乎全部落在理论曲线上,显著优于 Hill 估计量和 Pickands 估计量。根据极值理论的观点,可认为在给定频率接近 0 时,采用 Moment 估计量计算出的理论频率曲线对经验点据的描述最为确切。因此, Moment 估计量是估计太原站极值指数及计算小概率设计值的理想估计量,可用于当地强降水事件的频率分析。

## 3 结 论

本文在介绍极值指数估计理论的基础上,以山西省太原站的月降水序列为例,用 Pareto 分位数图对其是否服从重尾分布进行了判别;选用 3 种常用极值指数估计量计算了序列的极值指数  $\gamma$ ;为研究其尾部特性,选取序列中概率  $< 25\%$  的一段,观察并评价强降水部分的拟合效果。经分析计算得出结论:

(1) 太原站月降水序列服从重尾分布,极值指数  $\gamma$  可用于描述分布的尾部特性并进一步估算小

概率强降水情况下的设计值。

(2) 采用 Hill 估计量、Pickands 估计量和 Moment 估计量计算极值指数  $\gamma$  均较为简便。Sum - plot 作图法确定阈值  $k$  的效果良好,精度较高,但该法需二次作图及观察,使得阈值选取存在一定主观性,在后续研究中仍需进一步探索更为精确的方法。

(3) 当频率介于 25% 和 5% 时, Pickands 估计对月降水序列中强降水段的拟合效果最好,而 Hill 估计和 Moment 估计效果较差;当频率  $< 5\%$  时, Moment 估计对强降水点据的拟合效果最佳。因此,可认为 Moment 估计量是太原站月降水序列极值指数及设计值计算的理想估计量,且计算简便、精度较高,可解决常用概率分布无法准确拟合强降水部分的难题,为当地月降水频率分析及强降水预测工作提供参考。

## 参考文献:

- [1] 丁裕国. 气候概率分布理论的新内涵及其展望[J]. 沙漠与绿洲气象, 2007, 1(2): 1-5.
- [2] 邹波. 强降水发生的概率估计与比较[J]. 气候与环境研究, 2003, 8(3): 331-337.
- [3] Resnick S. Heavy tail modeling and teletraffic data[J]. Annals of Statistics, 1997, 25(5): 1805-1849.
- [4] Drees H, De Haan L, Resnick S. How to make a Hill plot [J]. Annals of Statistics, 2000, 28(1): 254-274.
- [5] Resnick S, Střuricá C. Smoothing the Hill estimator[J]. Ad-

vances in Applied Probability, 1997, 29(1):271-293.

- [6] Kratz M, Resnick S I. The qq-estimator and heavy tails[J]. Stochastic Models, 1996, 12(4):699-724.
- [7] Trang Dang D, Sandor M, Vidacs A. Investigation of fractal properties in data traffic [J]. Journal on Communications, 1999, XLIX:12-18.
- [8] Sousa B. A contribution to the estimation of the tail index of heavy-tailed distribution [D]. Michigan: The University of Michigan, 2002.
- [9] Beirlant J, Bouquiaux C, Werker B J M. Semiparametric lower bounds for tail index estimation[J]. Journal of Statistical Planning and Inference, 2006, 136(3):705-729.
- [10] Hall P. Using the bootstrap to estimate mean square error and select smoothing parameters in nonparametric problems [J]. Journal of Multivariate Analysis, 1990, 32(2):177-203.
- [11] Danielsson J, De Haan L, Peng L, et al. Using a bootstrap method choose the sample fraction in tail index estimation [J]. Journal of Multivariate Analysis, 2001, 76(2):226-248.
- [12] 刘维奇,赫英迪,邢红卫.选择重尾阈值 K 的 Bootstrap 方法[J].山西大学学报:自然科学版,2010,33(4):508-512.
- [13] 刘维奇,邢红卫.重尾指数估计中阈值 K 的简便优化估计[J].系统工程理论与实践,2010,30(8):1465-1470.
- [14] 李 扬.水文频率新型计算理论与应用研究[D].杨凌:西北农林科技大学,2013.
- [15] Hill B. A simple general approach to inference about the tail of a distribution[J]. Annals of Statistics, 1975, 3(5):1163-1174.
- [16] Pickands J. Statistical inference using extreme order statistics[J]. Annals of Statistics, 1975, 3(1):119-131.
- [17] Dekkers A L M, De Haan L. On the estimation of the extreme-value index and large quantile estimation[J]. Annals of Statistics, 1989, 17(4):1795-1832.
- [18] Dekkers A L M, Einmahl J H J, De Haan L. A moment estimator for the index of an extreme-value distribution[J]. Annals of Statistics, 1989, 17(4):1833-1855.
- [19] Drees H. Refined Pickands estimators of the extreme value index[J]. Annals of Statistics, 1995, 23(6):2059-2080.
- [20] 刘维奇,邢红卫.重尾分布尾指数估计研究进展[J].山西大学学报:自然科学版,2012,35(2):163-173.
- [21] 刘维奇,赫英迪,陈 琳.重尾分布的尾部指数估计及沪深股市实证分析[J].数学的实践与认识,2011,41(6):1-13.
- [22] 史道济,张春英.尾部指标估计中的阈值选择[J].天津理工大学学报,2006,22(6):78-81.

(上接第 44 页)

- [3] 杨淑贤,刘俊英,周明礼,等.湖北及邻区的斜坡类型及变形、破坏机理[J].西北地震学报,2006,28(3):258-262.
- [4] 祁生文,许 强,刘春玲,等.汶川地震极重灾区地质背景及次生斜坡灾害空间发育规律[J].工程地质学报,2009,17(1):40-49.
- [5] 柴波,殷坤龙,陈丽霞,等.岩体结构控制下的斜坡变形特征[J].岩土力学,2009,30(2):521-525.
- [6] 陈 剑,李 晓,杨志法.三峡库区滑坡的时空分布特征与成因探讨[J].工程地质学报,2005,13(3):305-309.
- [7] 陈喜昌.河谷岸坡类型的划分原则与方法[J].西南地质科技情报,1989,53(2):25-30.
- [8] 张俾元,王士天,王兰生,等.工程地质分析原理[M].北京:地质出版社,2009.
- [9] Kilburn, Christopher R J, Pasuto. et al. Major risk from rapid, large-volume landslides in Europe(EU Project Runout) [J]. Geomorphology, 2003, 54(1-2):3-9.
- [10] Pradhan B, Lee S, Mansor S, et al. Utilization of optical remote sensing data and geographic information system tools for regional landslide hazard analysis by using binomial logistic regression model[J]. Journal of Applied Remote Sensing, 2008, 2(1):142-154.
- [11] 张 梁.地质灾害风险评估理论与方法[J].中国地质矿产经济,1996(4):40-45.
- [12] Goovaerts P. Geostatistics for Natural Resources Evaluation [M]. Oxford:Oxford University Press, 1997.
- [13] 殷坤龙,张桂荣.地质灾害风险区划与综合防治对策[J].安全与环境工程,2003,10(1):32-35.
- [14] 沈 芳,黄润秋.区域地质环境评价与灾害预测的 GIS 技术[J].山地学报,1999,17(3):338-342.
- [15] Alison J R. A GIS for the prediction of landsliding potential and hazard in Southwest Greece [J]. Z. Geomorphology, 1993, 87(Suppl.):141-149.