非正态分布的天气气候序列极值特征诊断方法研究*

程炳岩 丁裕国 汪 方

(河南省气候中心，郑州 450003) (南京气象学院，南京 210044)

摘 要 推广了非正态假设下的交叉理论，且将其用于极值特征的诊断，并从理论上导出了适用性更广的基于 Gamma 分布和负指数分布的极值特征量诊断公式及其样本估计式。以有关降水要素的时间序列为例，说明了这种方法在天气气候诊断与气候影响研究中的应用前景。

关键词：极值；气候诊断；降水量；时间序列

1 引言

天气气候的极值或极端气候事件，因其出现机会很少，又无周期或循环性规律可寻，往往是预报的难点之一。尤其是关于极值的成因，至今也并无定论[1]。因此，即使采用最先进的数值预报模式，要想准确预报天气气候的极值确实相当困难。由于极端天气气候记录所固有的这种“稀有性”和“不确定性”，使得它比一般的平均天气气候变量更为特殊。但是，从某种宏观意义上看，它们也并非无任何规律可寻。

另一方面，研究表明，平均气候状况的任何微小变化都可能引发极端气候特征量的巨大变化[2]。近年来，许多学者已经认识到极端气候事件对于人类社会及其经济发展的危害性。然而，以往人们偏于对平均气候变化的研究较多，而对极端气候变化规律的研究不够，这是亟待改进的。极端气候事件与自然灾害的发生发展确有密切的关系，因此，在全球气候变化的研究中，除了关注平均气候的变化以外，更应当关心极端气候的变化。如何描述和监测气候极值事件，表征气候极值的各种统计特征及其变化规律，目前已经成为全球气候变化研究的焦点问题之一[3]。近年来，一些学者对于影响各地旱涝状况的强降水事件的研究，就是典型的例证[4]。

由于大多的天气气候极值（或极端事件）往往出现于非正态时间序列（如各种短时间尺度降水量、降水日数、旱涝指数或暴雨、冰雹、大风等）中，仅仅用正态序列的极值诊断公式来估计其特征量，可能产生较大误差。20 世纪 90 年代初，Desmond 和 Guy[5]曾指出，各种水平的交叉数和游程总数对模式偏度系数十分敏感，提出了关于非正态交叉理论并将其用于水文学研究河流水位序列的规律性，但并未将其用于极值特

2001-10-15 收到，2002-02-06 收到修改稿

* 河南省计委科技攻关项目“黄淮流域气象灾害监测预测技术方法研究”（9508）资助

征的研究，况且只限于研究几种分布型，如对数正态分布、χ^2 分布等。文献［6］中提出正态条件下天气气候时间序列极值的诊断方法，虽然也涉及到非正态条件下的极值特征量诊断问题。但主要仅讨论了正态时间序列极值特征量。实际上，许多气象观测记录序列往往其边际分布为偏态，因此，对于更多的气象要素（如降水、风、天气日数等）来说正态条件下的极值特征量诊断并不被需要。本文目的在于推广文献［5］的非正态交叉理论，将其用于极值特征的诊断，从理论上导出适用性更广的基于 Gamma 分布和负指数分布的极值特征量诊断公式及其样本估计式。

2　理论公式的导出

2.1　Gamma 分布的极值特征诊断公式

假定 $\xi_i(t)$，$i=1$，…，n 为相互独立的具有 $N(0,\ \sigma_\xi^2)$ 同分布的平稳随机过程，其自相关函数为 $\rho(\tau)$，若有函数过程[5,7]

$$\eta(t)=\xi_1^2(t)+\cdots+\xi_n^2(t), \tag{1}$$

则 $\eta(t)$ 称为具有 χ^2 边际分布的平稳过程。显然，上述过程在任一时刻 t，必有如下的概率密度函数（以下简记为 PDF）：

$$f_\eta(x)=\frac{1}{(2\sigma_\xi^2)\Gamma\left(\frac{n}{2}\right)}\left(\frac{x}{2\sigma_\xi^2}\right)^{\frac{n}{2}-1}\exp\left(\frac{-x}{2\sigma_\xi^2}\right), \tag{2}$$

上式表示一种特殊的 Gamma 分布的密度函数，其形状参数 $\alpha=n/2$，尺度参数 $\beta=2\sigma_\xi^2$，并可记为 Gamma $(n/2,\ 2\sigma_\xi^2)$。

根据 Rice[7] 所创立的平稳过程交叉理论及文献［8，9］的有关论述，可以推得 χ^2 分布意义下的高水平轴交叉数期望公式

$$N_\eta(u)=\frac{2}{\Gamma\left(\frac{n}{2}\sigma_\xi\right)}\left(\frac{\lambda_2}{\pi}\right)^{\frac{1}{2}}\left(\frac{u}{2\sigma_\xi^2}\right)^{\frac{n}{2}-1}\exp\left(\frac{-u}{2\sigma_\xi^2}\right), \tag{3}$$

这里，$N_\eta(u)$ 为超过临界值 u 的平均次数，$\lambda_2=-\sigma^2\rho''(0)$ 称为 $\xi(t)$ 的 2 阶谱矩，根据文献［6］，因为 $\lambda_{2i}=(-1)^i\sigma^2\rho^{(2i)}(0)$，$i=0$，1，…。所以 $\lambda_0=\sigma^2=\mathrm{Var}\,[\xi(t)]$，这里 $\mathrm{Var}\,[\xi(t)]$ 为 $\xi(t)$ 的方差。对于某个高临界值（极大值）u，当超越 u 的点数成串时，可认为是一次极大值过程，并记为“出现一次极大值”，所以，在 u 值较大的情况下，每两次交叉点之间可假设为“一次极大值”过程，于是，单位时间内平均极大值频数［记作 $\mu(u)$］即为

$$\mu(u)=\frac{N(u)}{2}=\frac{1}{\Gamma\left(\frac{n}{2}\right)\sigma_\xi}\left(\frac{\lambda_2}{\pi}\right)^{\frac{1}{2}}\left(\frac{u}{2\sigma_\xi^2}\right)^{\frac{n}{2}-1}\exp\left(\frac{-u}{2\sigma_\xi^2}\right). \tag{4}$$

另一方面，如从平稳过程交叉理论的定义出发，也可证明上述公式的正确性。为了阐明（3）式的数学物理意义，现对照正态条件下的高水平轴交叉数期望公式[7]

$$N_\xi(u)=\frac{\sigma_{\xi'}}{\pi\sigma_\xi}\exp\left(\frac{-u^2}{2\sigma_\xi^2}\right), \tag{5}$$

显然，由（3）和（5）式可见，对 $u>0$ 前者为其非对称函数，而后者为 u 的对称函

数，其非对称性取决于参数 n 的大小。根据文献［9］，当 $n\geqslant 30$ 已接近对称分布，当 $n/2\to 100$ 渐近于正态分布。可见 n 愈小，相应的分布愈偏斜。应用文献［5］的方法，类似地可推得连续两次超过临界值 u 的“平均间隔时间”和超过临界值 u 的“平均持续时间”计算公式为

$$E[B_u^+]=\sigma_\xi\left(\frac{\pi}{\lambda_2}\right)^{\frac{1}{2}}\left(\frac{2\sigma_\xi^2}{u}\right)^{\frac{n}{2}-1}\exp\left(\frac{u}{2\sigma_\xi^2}\right)\Gamma\left(\frac{n}{2}\right), \tag{6}$$

$$E[L_u^+]=\mu_u^{-1}P(\xi(0)>u)=\sigma_\xi\left(\frac{\pi}{\lambda_2}\right)^{\frac{1}{2}}\left(\frac{2\sigma_\xi^2}{u}\right)^{\frac{n}{2}-1}\exp\left(\frac{u}{2\sigma_\xi^2}\right)\gamma\left(\frac{u}{2\sigma_\xi^2},\frac{n}{2}\right). \tag{7}$$

上两式中，$E[L_u^+]$ 和 $E[B_u^+]$ 分别为超过临界值 u 的“平均持续时间”和“平均间隔时间”，又由（2）式，若给定 $\alpha=n/2$，$\beta=2\sigma_\xi^2$ 就可将其写为一般 Gamma 分布的 PDF，

$$f_\eta(x)=\frac{1}{(\beta)\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(\frac{-x}{\beta}\right). \tag{8}$$

由此不难推得，相应的极大值出现频率公式为

$$\mu_\eta(u)=\frac{\sqrt{2}}{\Gamma(\alpha)\sqrt{\beta}}\left(\frac{\lambda_2}{\pi}\right)^{\frac{1}{2}}\left(\frac{u}{\beta}\right)^{\alpha-1}\exp\left(\frac{-u}{\beta}\right). \tag{9}$$

顺便指出，本文还从平稳过程交叉理论的定义出发，详细论证了上述公式的正确性，结果表明，两种方法的证明结论完全一致。在给定的临界极大值条件下，还可写出计算相应的“极大值持续时间 L_u”的公式为

$$E[L_u^+]=\left(\frac{\beta}{2}\right)^{\frac{1}{2}}\left(\frac{\pi}{\lambda_2}\right)^{\frac{1}{2}}\left(\frac{\beta}{u}\right)^{\alpha-1}\exp\left(\frac{u}{\beta}\right)\gamma\left(\frac{u}{\beta},\alpha\right). \tag{10}$$

相应的“极大值间隔时间 B_u^+”公式为

$$E[B_u^+]=\left(\frac{\beta}{2}\right)^{\frac{1}{2}}\left(\frac{\pi}{\lambda_2}\right)^{\frac{1}{2}}\left(\frac{\beta}{u}\right)^{\alpha-1}\exp\left(\frac{u}{\beta}\right)\Gamma(\alpha). \tag{11}$$

由此可见，假如已知某时间序列的边际分布为 Gamma 分布，只要估计出相应的参数即可求得其极值特征量。

2.2 指数分布下的极值特征

对（1）式，如令 $n=2$ 则有简化式

$$\eta(t)=\xi_1^2(t)+\xi_2^2(t). \tag{12}$$

利用（2）式，可以证明，这里的 $\eta(t)$ 化为 Gamma（1，$2\sigma_\xi^2$）分布。如令 $\sigma_\xi^2=1/2$，（12）式就可化为最简单的 Gamma 分布，即 Gamma（1，1）。其形状参数为 1，尺度参数为 1。由（2）式，$\eta(t)$ 的 PDF 化为

$$f_\eta(x)=\mathrm{e}^{-x}. \tag{13}$$

（13）式表明，$\eta(t)$ 为最简单的指数分布，其参数 $\theta=1$。将（13）式及其参数代入（3）式，就可推得相应的（$u>0$）平均超过频率公式

$$N_\eta(u)=\frac{2\sqrt{2\lambda_2}}{\sqrt{\pi}}\mathrm{e}^{-u}, \tag{14}$$

则按文献［5］，在单位时间内的平均极大值频数 $\mu(u)$ 即可写为

$$\mu_\eta(u)=\frac{\sqrt{2\lambda_2}}{\sqrt{\pi}}\mathrm{e}^{-u}. \tag{15}$$

作者又从平稳过程交叉理论的定义出发，详细论证了上述公式的正确性，结果表明，两种方法的证明结论完全一致。在实际计算分析中，临界值（$u>0$）可由实际情况给定。例如，若令 $u=2\sigma_\eta$，$u=3\sigma_\eta$，则可有相应的平均极大值频数

$$\mu_\eta(u)=\frac{\sqrt{2\lambda_2}}{\sqrt{\pi}}\mathrm{e}^{-2\sigma}, \tag{16}$$

$$\mu_\eta(u)=\frac{\sqrt{2\lambda_2}}{\sqrt{\pi}}\mathrm{e}^{-3\sigma}. \tag{17}$$

3　实际样本估计计算公式

如前所述，（9）～（17）式中的 λ_2 只是每一个 $\xi(t)$ 的 2 阶谱矩，这对于实际应用并不方便。因为已知序列 $\eta(t)$ 的边际分布服从 Gamma 分布，它的组成变量是每一个 $\xi(t)$，但我们并不确知其构成的原始变量 $\xi(t)$ 及其分布参数。所以要直接计算（9）～（17）式的特征量就有必要将上式改写为由 $\eta(t)$ 的样本序列可直接计算的形式。利用关系式[5,6]

$$\rho_\eta(\tau)=\rho_\xi^2(\tau) \tag{18}$$

和

$$\rho''(0)=2\rho(1)-2, \tag{19}$$

并考虑关系式

$$\lambda_2=-\sigma_\xi^2\rho''(0),\quad \frac{\beta}{2}=\sigma_\xi^2, \tag{20}$$

由（9）式，不难得到

$$\mu_\eta(u)=\frac{\sqrt{-(2\sqrt{\rho(1)}-2)}}{\sqrt{\pi}\,\Gamma(\alpha)}\left(\frac{u}{\beta}\right)^{\alpha-1}\exp\left(\frac{-u}{\beta}\right), \tag{21}$$

式中，α、β 分别为实际序列的 Gamma 分布参数，这里的 $\rho(1)$ 即 $\rho_\eta(1)$（已略去下标 η，后文同此），为 $\eta(t)$ 的样本序列的一阶自相关系数，而 u 则是极大临界值。同理可得

$$E[L_u^+]=\frac{\sqrt{\pi}\,\gamma\left(\frac{u}{\beta},\alpha\right)}{\sqrt{-(2\sqrt{\rho(1)}-2)}}\left(\frac{\beta}{u}\right)^{\alpha-1}\exp\left(\frac{u}{\beta}\right), \tag{22}$$

$$E[B_u^+]=\frac{\sqrt{\pi}\,\Gamma(\alpha)}{\sqrt{-2(\sqrt{\rho(1)}-2)}}\left(\frac{\beta}{u}\right)^{\alpha-1}\exp\left(\frac{u}{\beta}\right). \tag{23}$$

显然，在特殊情况下，当 $\alpha=1$，上式实际化为指数分布情形，即（21）式成为

$$\mu_\eta(u)=\frac{\sqrt{2(1-\sqrt{\rho(1)})}}{\sqrt{\pi}}\exp\left(\frac{-u}{\beta}\right), \tag{24}$$

进一步，若令 $\alpha=1$，$\beta=1$，则可以证明上式更简化为

$$\mu_\eta(u)=\frac{\sqrt{2(1-\sqrt{\rho(1)})}}{\sqrt{\pi}}\exp(-u). \tag{25}$$

这就是前面推导的（15）式，上式表明，若气候时间序列为平稳指数分布过程，其自相

关愈小，它的极大值出现频数愈高，相反，其自相关愈大，它的极大值出现频数愈低。事实上，一般的指数分布其极大值出现机会确实很少，这是符合实际的。

4 应用实例计算与分析

利用南京夏季（4～9月）逐日降水量及全年各月降水量（1951～1997年）资料，计算Gamma分布下，超过给定临界极大值的平均频数 $\mu(u)$ 并对估计的计算结果作一分析。对于Gamma分布参数的估计来说，其矩估计公式为

$$\alpha = \frac{\mu_\eta^2}{\sigma_\eta^2}, \tag{26}$$

$$\beta = \frac{\sigma_\eta^2}{\mu_\eta^2}. \tag{27}$$

一般说来，矩估计误差较大，故也可改用极大似然估计法，根据Newton迭代方法，有下列估计公式[9,11]

$$\alpha = \begin{cases} y^{-1}(0.5000876 + 0.1648852y - 0.0544274y^2), & 0 < y \leqslant 0.5772 \\ y^{-1}(17.79728 + 11.968477y + y^2)^{-1} \times \\ \quad (8.898919 + 9.059950y + 0.9775373y^2), & 0.5772 < y \leqslant 17 \end{cases} \tag{28}$$

式中 $$y = \ln\left(\frac{\bar{x}}{\hat{x}}\right), \text{而 } \hat{x} = \left(\prod_{i=1}^{N} x_i\right),$$

或

$$y = \ln\bar{x} - \frac{1}{N}\sum \ln x_i, \tag{29}$$

则有

$$\beta = \frac{\bar{x}}{\alpha}. \tag{30}$$

将上述参数代入公式（21）即可求得超过序列极大值的平均频数 $\mu(u)$。现以南京站逐日降水量为实例，其计算步骤如下：

（1）给定序列 $Y(t)$，检验其分布是否为 Γ 或 χ^2 分布（含指数分布）。

（2）计算参数 α、β，即首先计算其平均值 μ_η 和方差 σ_η^2 及其自相关系数 $\rho(1)$，再以式（26）和（27）计算相应的参数（矩法）或以式（28）至式（30）计算相应的参数。

（3）给定临界极大降水量值 u，应用式（21），即可得平均极大值频数 $\mu(u)$。

（4）用实际序列的观测值验证之。同理，再计算另外两个参数“极大值平均持续时间 $E(L_u)$”和“极大值平均间隔时间 $E(B_u^+)$”。

计算结果表明，南京夏季（4～9月）逐日降水量极大值每年出现次数的理论计算值平均每10年相对误差仅为3%～5%左右（见表1）。这就是说，假如应用上述理论计算法，我们只要已知其序列平均值和标准差及其一阶自相关系数，就可估计其任何时期的极大值出现次数，若要精确估计其值，则可由式（28）～（30）的极大似然估计求得。

另外，图1和图2则分别绘出了南京夏季（4～9月）逐日降水量序列超过 2σ 和 3σ 的极大值频数的年际变化曲线。虽然，极大值频数逐年有一定的变化规律，但由于本文

表 1　南京夏季（4～9 月）逐日降水量序列的极大值（超过 2σ、3σ）的年代平均频数

临界极值	2σ		3σ	
年限	实测值	理论值	实测值	理论值
1951～1960	10.3	9.9	5.4	5.5
1961～1970	9.7	8.1	5.2	5.3
1971～1980	9.5	9.5	5.0	4.9
1981～1990	8.9	8.3	5.6	5.6
1991～1998	8.2	7.9	5.0	6.1
平均	9.3	9.6	5.2	5.5

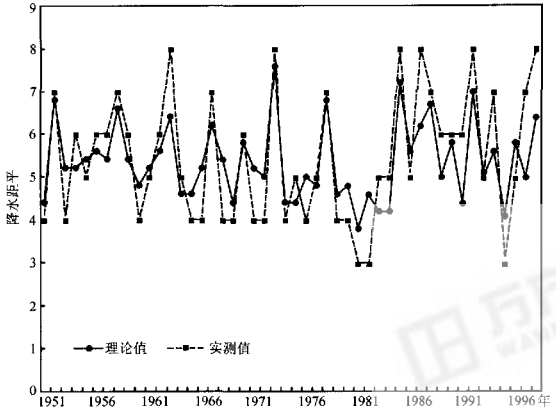

图 1　日降水极大值（超过 3σ）的年频数年际变化

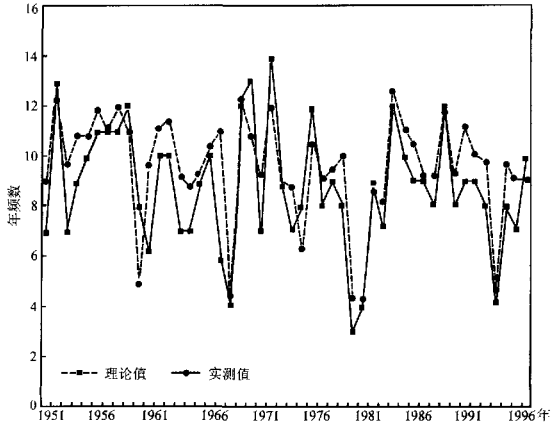

图 2　日降水极大值（超过 2σ）的年频数年际变化

主要目的并不在于研究极大值的年际变化，而只是从图中考察计算的精度。显然，由图可见，其计算精度还是较为理想的。

此外，对南京各月历年降水量序列（即有序列长度各为37），计算其超临界降水极值（2）理论频数，表2分别列出各月计算结果。由表2可见，其理论计算值与实际观测值的相对误差不太均匀，一般在16%左右，但也有高达28%的月份。而考察历年月降水量序列（长度为444），则发现，样本序列愈长，估计精度愈高（见表3）。例如表3中，序列长度为444，平均每年超过临界降水极值（2σ）大约1～2次，理论与实测值基本吻合。

表2 南京各月降水量序列的极大值频数理论计算值与实际观测值

月份	实际值	理论值（1）	理论值（2）	相对误差（1）	相对误差（2）
1	8.0	6.3	6.6	0.21	0.18
2	9.0	6.3	7.3	0.30	0.19
3	5.0	4.5	4.0	0.10	0.20
4	5.0	5.7	6.4	0.14	0.28
5	7.0	5.6	5.5	0.20	0.21
6	5.0	6.2	5.6	0.24	0.12
7	8.0	9.9	9.0	0.23	0.12
8	11.0	9.4	10.3	0.15	0.06
9	8.0	7.5	8.2	0.06	0.02
10	11.0	8.6	8.2	0.22	0.23
11	9.0	8.5	9.7	0.05	0.08
12	10.0	7.2	8.3	0.28	0.17
平均	8.0	6.9	7.0	0.18	0.16

注：理论值（1）、（2）及相对误差（1）、（2）分别是指用矩法（1）和用似然法（2）所得结果

表3 历年月降水量距平序列极大值频数理论计算值与实际观测值比较

实际值	理论值		误差	
	矩法（A）	极大似然法（B）	A	B
75.0	89.3	84.0	0.19	0.12
（1.56）	（1.86）	（1.75）	（0.19）	（0.12）

注：表中括号内数字为每年平均次数

5 结论

（1）非正态变量在气候要素中占有重要地位（如降水量、风速、各种天气日数等），而气候极值事件的统计特征对社会经济与环境及其灾害的影响又至关重要[9~12]。因此，假定已知序列的一般统计特征（如平均值和方差），就可由本文导出的诊断公式求得其极值特征量，这无疑对于气候诊断和预报具有重要意义。

（2）理论推导和实例计算表明，基于非正态交叉理论的极值特征量诊断方法，具有较好的可行性与可靠性。由于极端天气气候事件的预报十分困难，对于极值规律的分析必然有助于提高预报水平。作者在文献[5]以及本文所给出的分析方法大体上涵盖了具有各种分布型条件下的气候时间序列的极值诊断。这对于进一步开展极值特征的

预报具有重要的理论价值和应用前景。

（3）无论是正态或非正态条件下，极值出现次数、持续时间和间隔时间等统计特征的计算公式，对于从理论上研究极端天气气候事件的长期变率特征以及全球气候变化对于区域或局部气候极值的影响即未来各地气候情景预测，都很有实用价值[1)]。

参 考 文 献

1 Hunt, B. G., Nonlinear influences——A key to short-term climatic pertubations, *J. Atmos. Sci.*, 1988, **45**, 387～395.

2 Katz, R. W., and B. G. Browns, Extreme events in a changing climate: Variability is more impotant than averages, *Climatic Change*, 1992, **21**, 289～302.

3 Kharin, V. V., and F. W. Zwiers, Changes in the extrimes in an ensemble of transient climatic simulations with a coupled atmosphere—ocean GCM., *J. Climate*, 2000, **13**, 3760～3788.

4 Zhang, Xuebin, W. D. Hogg, and E. Mekis, Spatial and temporal characteristics of haevy pricipitation events over Canada, *J. Climate*, 2001, **14**, 1923～1936.

5 Desmond, A. F., and B. T. Guy, Crossing theory for non—Gaussian stochastic processes with application to hydrology, *Water Resour. Res.*, 1991, **27**, 2791～2797.

6 丁裕国、金莲姬、刘晶森，诊断天气气候时间序列极值特征的一种新方法，大气科学，2002，**26**（3），343～351.

7 Rice, S. O., Mathematical analysis of random noise, *Bell. Syst. Tech. J.*, 1945, **24**, 24～156.

8 Cramer, H., and M. R. Leadbetter, *Stationary and Relaed Stochastic Processes*, John Wiley, New York, 1967, 1～20.

9 么枕生、丁裕国，气候统计，北京：气象出版社，1990.

10 Kedem, B., *Binary Time Series*, Marcel Dekker, Inc., 1980, 1～33.

11 丁裕国，天气气候状态转折规律的统计学探讨，气候学研究—— 统计气候，北京：气象出版社，1991，40～49.

12 Priestley, M. B., *Spectral Analysis and Time Series*, Vol. 1, Academic Press, London, 1981, 280～290.

13 丁裕国、江志红，气象数据时间序列信号处理，北京：气象出版社，1998，40～45.

1）丁裕国等，全球气候变暖对区域气候极值的影响：基于交叉理论的一种随机模拟试验，热带气象学报，待发表.

A Diagnosis Method of the Extreme Features of Weather and Climate in Time Series Based on Non-Normal Distribution

Cheng Bingyan Ding Yuguo and Wang Fang

(*Henan Climate Center*, *Zhengzhou* 450003) (*Nanjing Institute of Meteorology*, *Nanjing* 210044)

Abstract According to the theory presented by Desmond and Guy (1991), the diagnostic formulae for the extreme features are generalized by using the cross theory under the hypothesis of non-normal distribution. The theoretic formulae and sampling estimator for diagnosis extreme features are derived from the hypothesis of Gamma or exponential distribution in weather or climate series. The cases studies of rainfall time series show that the sampling estimator formulae are very effective to observational data series. The calculation results show consistence in theory and observation, which show that the method can be applied to the diagnosis of the extreme event of weather or climate, and particularly to the research of the problem for future climatic influence.

Key words: extreme value; climatic diagnosis; rainfall; time series