

基于最大熵原理的台风统计预报

冯利华 张 萍

(浙江师范大学 金华 321004)

提要 台风以最无序的方式在沿海各地登陆,意味着台风熵达到了极大值。在给定的约束条件下,当台风熵取极大值时,台风强度是一种指数分布。根据最大熵原理和1949年以来中国登陆台风的实测资料,揭示了台风强度的分布形式,提出了台风复发期的概念,这对登陆台风的统计预报有所裨益。

关键词 台风强度,最大熵原理,无序,分布,复发期

中图分类号 P457.8 **文献标识码** A **文章编号** 1000-3096(2003)03-0047-05

1854年,克劳修斯(Clausius)把可逆过程中工作物质吸收的热量与温度之比称为熵(entropic)。在其后的100多年里,熵概念在许多学科领域都得到了广泛的应用。作为一个状态函数,熵的含义非常丰富:在热力学中,它是不可用能量的量度;在统计物理学中,它是系统微观态数目多少的量度;在信息论中,它是随机事件不确定性的量度^[1]。

中国沿海地区的经济最发达、社会财富最密集。事实上,沿海地区的发展既得益于海洋,也受制于海洋,台风灾害就是最大的限制因子之一。一次台风可以使沿海灾区的社会经济发展倒退二三年,甚至七八年。因此,沿海地区未来台风的预报分析对国计民生具有重要的现实意义,但至今这方面的研究甚少,能够实际应用的成果更少^[2]。为此,拟根据最大熵原理和1949年以来中国登陆台风的实测资料,探讨台风强度的分布形式,提出台风复发期的概念,并对未来台风次数和强度的统计规律作一分析。

1 最大熵原理

1948年,申农(Shannon)用熵来定量地描述一个随机事件的不确定性或信息量:

$$H = - \sum_{i=1}^n P_i \ln P_i \quad (1)$$

式中 H 称为信息熵; P_i 为随机事件出现的概率; C

为常数。从式(1)可知,熵 H 是 P_i 的泛函,因此,在给定的实验条件下,存在一个使 H 取极大值的分布。该分布占有优势概率,是最常见的分布,称为最可几分布。这样,在给定的约束条件下,可以从所有可能的相容分布中,挑选出熵极大时的分布,这就是最大熵原理。

根据最大熵原理,利用拉格朗日未定乘子法,可以求出熵极大时的分布^[3,4]。设随机变量 x 取值 x_1, x_2, \dots, x_n , 相应的概率为 p_1, p_2, \dots, p_n 。它们满足:

$$\sum_{i=1}^n P_i = 1 \quad p_i \geq 0 \quad (2)$$

且若干已知函数 $f_k(x_i)$ 的平均值 F_k 是给定的:

$$F_k = \sum_{i=1}^n f_k(x_i) P_i \quad k=1, 2, \dots, m (m < n) \quad (3)$$

为了寻求在约束条件式(2)和(3)下熵极大时的分布,引入未定乘子 α 和 β_k , 构造一个新函数:

$H - \alpha - \beta_1 F_1 - \beta_2 F_2 - \dots - \beta_m F_m$ 。从式(1)、(2)和(3)可得:

第一作者:冯利华,出生于1955年,教授,主要从事灾害地理学研究。E-mail:fenglh@mail.zjnu.net.cn

收稿日期:2001-07-10;修回日期:2001-12-20

$$H - \alpha - \sum_{k=1}^m \beta_k F_k = - \sum_{i=1}^n p_i \ln p_i - \alpha \sum_{i=1}^n p_i - \sum_{k=1}^m \beta_k \sum_{i=1}^n f_k(x_i) p_i = \sum_{i=1}^n p_i \ln \left\{ \frac{1}{p_i} \exp \left[-\alpha - \sum_{k=1}^m \beta_k f_k(x_i) \right] \right\}$$

利用不等式 $\ln x \leq x - 1$, 上式成为:

$$H \leq \sum_{i=1}^n p_i \left\{ \frac{1}{p_i} \exp \left[-\alpha - \sum_{k=1}^m \beta_k f_k(x_i) \right] - 1 \right\} + \alpha + \sum_{k=1}^m \beta_k F_k$$

若要 H 取极大值, 上式必须取等号, 那么此时要求 p_i 满足下式:

$$p_i = \exp \left[-\alpha - \sum_{k=1}^m \beta_k f_k(x_i) \right] \quad i = 1, 2, \dots, n \quad (4)$$

利用式(2), 式(4)可写成:

$\alpha = \ln \left\{ \sum_{i=1}^n \exp \left[-\sum_{k=1}^m \beta_k f_k(x_i) \right] \right\}$ 。若令 $Z = e^\alpha$, 那么又可写成:

$Z = \sum_{i=1}^n \exp \left[-\sum_{k=1}^m \beta_k f_k(x_i) \right]$ (Z 称为配分函数)。这样, 式(4)成为:

$$p_i = \left\{ \exp \left[-\sum_{k=1}^m \beta_k f_k(x_i) \right] \right\} / Z \quad (5)$$

为了求得 β_k , 把式(5)代回约束方程式(3), 得到:

$$F_k = \sum_{i=1}^n \left\{ f_k(x_i) \exp \left[-\sum_{k=1}^m \beta_k f_k(x_i) \right] \right\} / Z \quad (6)$$

在式(6)中, F_k 和 $f_k(x_i)$ 都是已知值, 而真正的未知数是 m 个 β 值 ($\beta_1, \beta_2, \dots, \beta_m$), m 个方程应当能够解出 m 个 β 值, 这样就可以求出熵极大时的 p_i 值了。

以上离散场合下得到的计算公式同样适合于连续场合下的求解过程。

2 从最大熵原理得出的台风强度的分布

熵增加原理表明: 在孤立或绝热的条件下, 系统自发地从非平衡态向平衡态发展的过程是一种熵增加的过程, 平衡态与最大熵相对应, 在平衡态系统的状态最混乱、最无序。

台风是一种随机事件, 它的发生具有不确定性, 因而可以用熵来进行描述。台风以最无序的方式在沿海各地登陆, 意味着台风熵达到了极大值, 因此可以用最大熵原理来确定一个地区一定时段内台风强度的分布。

台风登陆时的最大风力、最大风速和最低中心气压是描述登陆台风强度的三个主要指标。由于中

心气压 P 的观测精度相对较高, 故采用中心气压来描述登陆台风的强度。设 P_s 为中心气压上限, \bar{P} 为平均中心气压, 那么式(2)成为 $\int_{\bar{P}}^{P_s} p(x) dx = 1$, 而式(3)仅为一个关系 ($m = 1$):

$$\bar{P} = \int_{\bar{P}}^{P_s} x p(x) dx \quad (7)$$

由此得到配分函数 $Z = -[\exp(-\beta P_s)]/\beta$ 。代入式(5), 可得:

$$p(x) = -\beta \exp[-\beta(x - P_s)] \quad (8)$$

再利用式(7), 得到 $\beta = 1/(\bar{P} - P_s)$ 。这里 $p(x) = n/n_z$, 其中 n_z 为台风总数, n 为中心气压从 P 到 $P + dP$ 之间的台风次数, 即微分次数, 而累积次数 N 为中心气压 $< P_i$ 的台风次数。将它们代入式(8), 可得:

$$n = \frac{n_z}{P_s - \bar{P}} \exp \left(\frac{P - P_s}{P_s - \bar{P}} \right) \quad (9)$$

可见台风强度是一种指数分布。若从式(1)开始采用常用对数表示, 并对式(9)取对数,

那么它成为:

$$\lg n = \frac{1}{P_s - \bar{P}} P + \lg \frac{n_z}{P_s - \bar{P}} - \frac{P_s}{P_s - \bar{P}} \quad (10)$$

$$\text{令 } a = \frac{1}{P_s - \bar{P}}, b = \lg \frac{n_z}{P_s - \bar{P}} - \frac{P_s}{P_s - \bar{P}}, \text{ 那么}$$

$$\lg n = aP + b \quad (11)$$

由此可见, 一个地区一定时段内台风的微分次数 n 与中心气压 P 在半对数坐标系中呈直线关系。

表 1 是 1949~1992 年中国登陆台风次数和中心气压的统计资料。图 1 是其累积次数 N 与中心气压 P 在普通坐标系中的关系, 可见它确实是一种指数分布。图 2 是它们在半对数坐标系中的关系。为了使直线与点子得到最佳拟合, 可以采用最小二乘法、最大似然法等来估计回归系数 a 和 b 。图 2 中的直线是根据最小二乘法得到的最佳拟合直线:

$$\lg N = 0.0309P - 28.12 \quad (12)$$

同时可求得 N 和 P 的相关系数 $R = 0.996$, 可见两者的相关程度是很高的。

表 1 1949~1992 年中国登陆台风统计

Tab.1 Statistics of typhoons going over the Chinese coasts in 1949~1992

中心气压(hPa)	920~929	930~939	940~949	950~959	960~969	970~979	980~989	990~999
微分次数 n	4	4	9	14	54	96	130	158
累积次数 N	4	8	17	31	85	181	311	469

为了检验式(11)的正确性,把不同区域(中国和广东省)、不同时段 $M(25a, 27a$ 和 $52a)$ 的台风累积次数 N 和中心气压 P 也点绘在半对数坐标系中,那么同样可以得到类似的直线分布(表 2),它们的相关程度也是很高的。

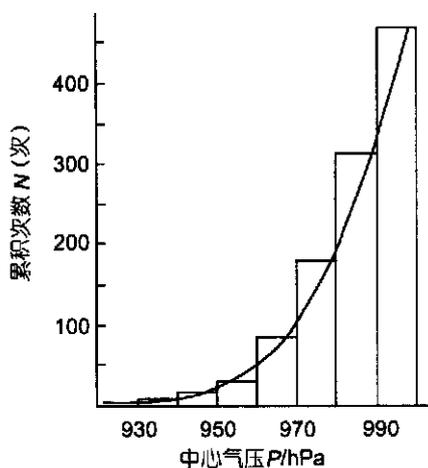


图 1 1949~1992 年中国登陆台风累积次数与中心气压的关系

Fig.1 The relation between accumulation of frequency of typhoons going over the Chinese coasts and their central pressure in 1949~1992

3 台风次数的统计预报

式(12)也可以写成: $N = 10^{0.0309P - 28.12}$ 。如果把该式除以统计时段 44a。那么可得每年出现中心气压 $< P_i$ 的台风平均次数: $N_c = \frac{N}{44} = \frac{1}{44} \times 10^{0.0309P - 28.12}$ 。由此可得 t a 出现中心气压 $< P_i$ 的台风平均次数 N_t :

$$N_t = tN_c = \frac{t}{44} \times 10^{0.0309P - 28.12} \quad (13)$$

式(13)是由台风实测资料得到的统计关系式,因而可以用来预报未来 t a 出现中心气压 $< P_i$ 的台

风平均次数。表 3 是根据式(13),预报 1a(1993 年)、3a(1993~1995 年)和 8a(1993~2000 年)出现中心气压 $< P_i$ 的台风平均次数。如对于未来 3a(1993~1995 年)而言,计算出现中心气压 < 970 hPa 的台风次数为 4.85 次,而实际出现中心气压 < 970 hPa 的台风次数为 5 次。从表中可以看到,预报的台风次数与中国实际出现的台风次数是接近的,并且外推年数越长,两者误差越小。

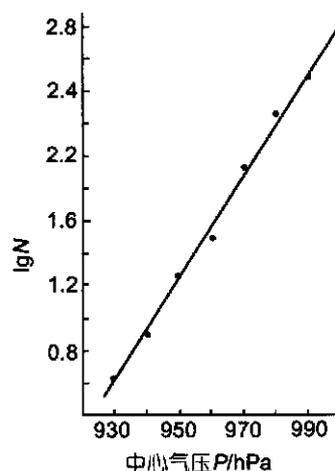


图 2 半对数坐标系中台风累积次数与中心气压的关系

Fig.2 The relation between accumulation of frequency of typhoons and their central pressure in semi-logarithmic coordinate

4 台风的复发期

由前可知, N_t 是 t a 出现中心气压 $< P_i$ 的台风平均次数。如果令 $N_t = 1$, 那么式(13)可以写成:

$$t = 44 \times 10^{28.12 - 0.0309P} \quad (14)$$

它表示出现一次中心气压 $< P_i$ 的台风所需要的时间,可以称为台风的复发期。为区别起见,把 t 改写为 T 。这里的复发期 T 表示“几年一遇”之意,如

$T = 8a$, 即表示 $8a$ 一遇的中心气压 $< P_i$ 的台风。表 4 是 1949~1992 年中国中心气压 $< P_i$ 的登陆台风的计算复发期和实际复发期, 其中实际复发期 $T_s = 44/N(N$ 见表 1)。从表中可以看到, 计算复发期和实际复发期的平均误差只有 $0.12a$, 最大误差也只有 $0.35a$, 因此计算结果与实际情况是吻合的。

表 2 不同区域不同时段台风累积次数 N 和中心气压 P 的关系

Tab.2 The relation between accumulation of frequency of typhoons and their central pressure with different period of time in each region

区域	起止年限	年数(a)	关系式	相关系数 R
中国	1949~1973	25	$\lg N = 0.0288P - 26.27$	0.997
中国	1974~2000	27	$\lg N = 0.0348P - 32.19$	0.990
中国	1949~2000	52	$\lg N = 0.0312P - 28.37$	0.996
广东省	1949~1973	25	$\lg N = 0.0326P - 30.32$	0.995
广东省	1974~2000	27	$\lg N = 0.0356P - 33.24$	0.979
广东省	1949~2000	52	$\lg N = 0.0365P - 33.79$	0.988

表 3 预报未来 $t a$ 出现中心气压 $< P_i$ 的台风次数

Tab.3 The forecast of frequency of typhoons with central pressure $< P_i$ in t years

中心气压 P (hPa)	930	940	950	960	970	980	990
$t = 1a$ 计算次数 N_1	0.09	0.19	0.39	0.79	1.62	3.29	6.70
$t = 1a$ 实际次数	0	0	0	1	1	5	6
$t = 3a$ 计算次数 N_3	0.28	0.57	1.17	2.38	4.85	9.87	20.10
$t = 3a$ 实际次数	0	0	1	3	5	12	24
$t = 8a$ 计算次数 N_8	0.75	1.53	3.12	6.34	12.92	26.32	53.61
$t = 8a$ 实际次数	0	2	3	6	12	30	52

表 4 中心气压 $< P_i$ 的台风复发期

Tab.4 Recurrence intervals of typhoons with central pressure $< P_i$

中心气压 P (hPa)	计算复发期 T (a)	实际复发期 T_s (a)	误差 (a)
930	10.65	11.00	-0.35
940	5.23	5.50	-0.27
950	2.57	2.59	-0.02
960	1.26	1.42	-0.16
970	0.62	0.52	0.10
980	0.30	0.24	0.06
990	0.15	0.14	0.01
1 000	0.07	0.09	-0.02

2000 年)台风的中心气压($< P_i$)。如对于 $8a$ 一遇台风的中心气压而言, 在 1993~2000 年中, 计算的中心气压 < 934.03 hPa, 而实际出现的中心气压为 930 hPa(0010 号台风、碧利斯), 与实际情况相符。从表中可以看到, 在不同的预报年数内, 计算结果与中国实际出现的中心气压全部符合, 因此预报正确。

表 5 预报 $T a$ 一遇台风的中心气压($< P_i$)(hPa)

Tab.5 To forecast the central pressure ($< P_i$) of a possible typhoons in t years(hPa)

预报年数(a)	1	3	8
计算的中心气压($< P_i$)	963.26	947.82	934.03
实际出现的中心气压	950	945	930
中央气象台编号	9315	9414	0010

5 台风强度的统计预报

如果把式(14)改写成:

$$P = \frac{28.12 - \lg(T/44)}{0.0309} \quad (15)$$

那么 P 表示未来 $T a$ 内可能发生的一次台风的中心气压($< P_i$), 因而可以用来预报 $T a$ 一遇台风的中心气压($< P_i$)。表 5 是根据式(6), 预报 $1a$ 一遇(1993 年)、 $3a$ 一遇(1993~1995 年)、 $8a$ 一遇(1993~

6 结语

台风是一种随机事件, 它的发生具有不确定性, 因而适合于用熵来进行描述。从以上的分析可以得到下列几点结论:

(1) 台风以最无序的方式在沿海各地登陆,意味着台风熵达到了极大值,因而可以用最大熵原理来确定一个地区一定时段内台风强度的分布。在给定的约束条件下,当台风熵取极大值时,它是一种指数分布: $N = 10^{aP+b}$ 。

(2) 式(13) $N_i = \frac{t}{M} \times 10^{aP_i+b}$ 可以用来预报未来 t 出现中心气压 $< P_i$ 的台风次数。

(3) 复发期 T 表示出现一次中心气压 $< P_i$ 的台风所需要的时间,它具有“几年一遇”之意。

(4) 式(15) $P = -\frac{\lg(T/M) + b}{a}$ 可以用来预报 T 一遇台风的中心气压($< P_i$)。

本文根据最大熵原理和 1949 年以来中国登陆

台风的实测资料,揭示了台风强度的分布形式,提出了台风复发期的概念,这对登陆台风的统计预报具有重要的实际意义。

参考文献

- 1 王彬.谈熵、熵与交叉科学.北京:气象出版社,1988. 19-22
- 2 冯利华.利用 R/S 分析试作登陆浙江台风重灾年的预测.海洋学报,2000,22(5):133-136
- 3 张学文,马力.熵气象学.北京:气象出版社,1992. 114-125
- 4 Silviu G. Information theory with application. McGRAW HILL. CO., 1977. 293-301

STATISTICAL FORECAST OF TYPHOON BASED ON MAXIMUM ENTROPY PRINCIPLE

FENG Li Hua ZHANG Ping

(Zhejiang Normal University, Jinhua, 321004)

Received: July, 10, 2001

Key Words: Typhoon intensity, Maximum entropy principle, Disorder, Distribution, Recurrence interval

Abstract

Typhoon lands in the form of highly disorder along the coast, which means that typhoon entropy has gotten maximum. Typhoon intensity is an exponent distribution in given restrained conditions when typhoon entropy has gotten maximum. In accordance with the maximum entropy principle and measured data of landing typhoon of China since 1949, the distribution type of typhoon intensity is enunciated, and the concept of recurrence interval of typhoon which is of significance for the statistical forecast of landing typhoon, is put forward.

(本文编辑:张培新)