

广西冬季极端最低气温的概率分布模型选择 及其极值和重现期计算*

Selection of Probability Distribution Models and Calculation of Extremes and Reproduction Periods of Winter Extremely Minimum Temperature in Guangxi

苏 志 李艳兰 涂方旭

Shu Zhi Li Yanlan Tu Fangxu

(广西区气候中心 南宁市民族大道 530022)

(Guangxi Climate Center, Minzhu Dadao, Nanning, Guangxi, 530022, China)

摘要 为了解广西冬季极端最低气温概率分布类型和分析广西不同重现期的冬季极端最低气温值,根据广西 88 个气象站 1951~ 2000年 (1951年以后建站的站,取建站~ 2000年)冬季极端气温序列,分别用 Gumbel I 型极值分布、Gumbel II 型极值分布、Gumbel III 型极值分布、正态分布、双正态分布函数进行拟合,并按柯尔莫哥洛夫检验和 ω^2 检验方法进行拟合优度检验。结果表明,桂林等 55 个站遵从双正态分布,河池等 29 个站遵从正态分布,南宁等 4 个站遵从 Gumbel I 型极值分布,说明用双正态分布、正态分布和 Gumbel I 型极值分布函数作为广西冬季极端最低气温的分布函数效果较好。广西 15 年一遇的极端最低气温 -2°C 等值线与广西龙眼、荔枝的分布北界比较接近。

关键词 极端最低气温 概率分布 重现期

中图法分类号 S162.0211.3

Abstract To investigate the probability distribution of winter minimum temperature and the reproduction period of winter extremely minimum temperature in Guangxi, the functions, Gumbel I extreme distribution, Gumbel II extreme distribution, Gumbel III extreme distribution, normal distribution, binomial distribution are used to fit the winter extreme temperature series from the 88 meteorological stations in the period from 1951 to 2000 (from establishment to 2000 for the stations set up after 1951) in Guangxi. The fitness is checked up by $D_n(x)$ and ω^2 . There are 55 stations include Guilin station being subject to binormal distribution, 29 stations include Hechi station being subject to normal distribution, 4 stations include Nanning station being subject to Gumbel extreme distribution. It is suggested that binormal distribution, normal distribution and Gumbel extreme distribution are better to be used to describe the winter extremely minimum temperature distribution in Guangxi. The -2°C isoline of 15-year-reproduction of extremely minimum temperature is close to the north border of distribution of longan and lichi.

Key words extremely minimum temperature, probability distribution, reproduction period

制作某些农作物、水果的寒冻害区划时,常需要考虑冬季几年一遇的极端最低气温值。 T 年一遇的极端最低气温值就是满足 (1) 式^[1]对应的 x_T 值

$$P(X < x_T) = F(x_T) = \frac{1}{T}. \quad (1)$$

欲确定某站几年一遇的极端最低气温值,首先要

确定极端最低气温的概率分布函数。作者利用广西 88 个气象站冬季极端最低气温资料研究概率分布,并分析了广西不同重现期的冬季极端最低气温的分布情况

1 分布函数模型的选择及模型参数的估计方法

为了确定各站极端最低气温合适的分布函数,需

2001-11-09收稿, 2001-12-17修回。

* 广西自然科学基金资助项目 (桂科自 0007023)

要根据极端最低气温资料,选择一些概率分布函数进行拟合。本文选择的概率分布函数有如下几种:

① Gumbel I型极小值分布,其分布函数为:

$$F_1(x) = 1 - \exp(-\exp(T_1(x - U_1))); \quad (2)$$

② Gumbel II型极值分布,其分布函数为:

$$F_2(x) = 1 - \exp(-(\frac{k - U_2}{k - x})^{T_2}); \quad (3)$$

③ Gumbel III (韦伯尔 Weibull) 分布,其分布函数为:

$$F_3(x) = 1 - \exp(-(\frac{x}{U_3})^{T_3}). \quad (4)$$

以上3种分布函数可以写成解析函数形式。估计其参数 $T_i, U_i (i = 1, 2, 3)$ 有不同的方法^[2~4]。本文采用的方法是先对分布函数进行线性变换,然后再用最小二乘法加以估计。现以 Gumbel I型极小值分布函数(式(2))为例:

对(2)式作2次对数变换:

$$\ln(-\ln(1 - F_1(x))) = T_1x - T_1U_1,$$

将样本数为 n 的序列从小到大排列,记 x_k 为排列为 k 的样本值,以经验分布函数:

$$F_n(x) = \frac{k}{n+1}, (k = 1, 2, \dots, n),$$

作为对应的 $F_1(x)$ 的观测值,则:

$$\ln(-\ln(1 - \frac{k}{n+1})) = T_1x - T_1U_1.$$

令 $y = \ln(-\ln(1 - \frac{k}{n+1}))$, 则上式化为线性关系:

$$y = A + Bx.$$

根据最小二乘法求得 A, B , 则:

$$T_1 = B,$$

$$U_1 = -\frac{A}{B}.$$

类似可以确定 Gumbel II、III型极小值分布的参数 T_2, U_2 及 T_3, U_3

④ 正态分布,其概率密度函数为:

$$h(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x - m)^2}{2\sigma^2}), \quad (5)$$

参数 m, σ 可用样本均值、样本均方差进行估计。其分布函数为:

$$F_4(x) = \int_{-\infty}^x h(t) dt = H(\frac{x - m}{\sigma}) = H(u), \quad (6)$$

式中: $H(u)$ 为标准正态分布函数在 u 点的值

⑤ 双正态分布^[5,6],其概率密度函数为:

$$h(x) = \frac{2}{\sigma_1(\epsilon_1 + \epsilon_2)} \exp(-\frac{(x - m_1)^2}{2\epsilon_1^2}), \quad (7)$$

$x \leq m_1,$

$$h(x) = \frac{2}{\sigma_2(\epsilon_1 + \epsilon_2)} \exp(-\frac{(x - m_1)^2}{2\epsilon_2^2}),$$

$x > m_1,$

式中: m_1 为分布的众数; $\epsilon_1^2, \epsilon_2^2$ 分别是随机变量在 $a < m_1, a > m_1$ 的条件下,与 m_1 离差平方的条件期望

由最大似然估计方法,参数 m_1 可用数值解法求得其估计值,求得 m_1 后,即可用最大似然估计方法估计 ϵ_1, ϵ_2

设 x_1, x_2, \dots, x_n 是服从双正态分布的一个随机样本,令

$$S_1^2(z) = \frac{1}{n} \sum_{x_i \leq z} (x_i - z)^2, \quad (8)$$

$$S_2^2(z) = \frac{1}{n} \sum_{x_i > z} (x_i - z)^2,$$

使

$$S^2(z) = (S_1^2(z))^{\frac{1}{3}} + (S_2^2(z))^{\frac{1}{3}}, \quad (9)$$

达到最小对应的 z 值,即为参数 m_1 的估计值 \hat{m}_1 。对确定的 \hat{m}_1 值,即可估计参数 ϵ_1, ϵ_2

$$\hat{\epsilon}_1 = (S_1^2(\hat{m}_1))^{\frac{1}{3}} ((S_1^2(\hat{m}_1))^{\frac{1}{3}} + (S_2^2(\hat{m}_1))^{\frac{1}{3}})^{\frac{1}{2}}, \quad (10)$$

$$\hat{\epsilon}_2 = (S_2^2(\hat{m}_1))^{\frac{1}{3}} ((S_1^2(\hat{m}_1))^{\frac{1}{3}} + (S_2^2(\hat{m}_1))^{\frac{1}{3}})^{\frac{1}{2}}.$$

以上估计值的求法,在计算机上极易实现。

若 $\epsilon_1 = \epsilon_2 = \sigma$, 则双正态分布即为正态分布

双正态分布的分布函数为:

$$F_5(x) = \int_{-\infty}^x h_5(t) dt = \frac{2\epsilon_1}{\epsilon_1 + \epsilon_2} H(\frac{x - m_1}{\epsilon_1}), x \leq m_1; \quad (11)$$

$$F_5(x) = \int_{-\infty}^x h_5(t) dt = \frac{\epsilon_1 - \epsilon_2}{\epsilon_1 + \epsilon_2} + \frac{2\epsilon_2}{\epsilon_1 + \epsilon_2} H(\frac{x - m_1}{\epsilon_2}), x > m_1.$$

2 选择分布函数的标准

采用柯尔莫哥洛夫检验和 k^2 检验进行拟合优度检验,以确定选择哪一种分布函数较好。另外,还计算了理论分布函数与经验分布函数的线性相关系数,作为选择分布函数的参考。

2.1 柯尔莫哥洛夫检验

设待选的分布函数为 $F_i(x)$, 经验分布函数为:

$$F_n(x) = \frac{k}{n+1}, (k = 1, 2, \dots, n),$$

式中 n 为样本容量, k 为观测值从小到大排列后的顺序号。

定义统计量:

$$D_n(x) = \max |F_n(x_k) - F_i(x_k)|. \quad (12)$$

对于给定的信度 T , 由 n 和 T 查柯尔莫哥洛夫检验的临界值 (D_{nT}) 表, 得 D_{nT}

若 $D_n \leq D_{nT}$, 则可以认为选用的分布函数拟合得好, D_n 越小, 拟合得越好。T 一般取 0.10, 若放低要求, 可取 0.05

2.2 ω^2 检验

对于待选的分布函数 $F_i(x)$, 引入统计量 k^2 :

$$k^2 = \frac{1}{n+1} \sum_{k=1}^n (F_i(x_k) - \frac{2k-1}{2(n+1)})^2 + \frac{n+4}{12(n+1)^3} \quad (13)$$

由于 k^2 的数学期望为:

$$E(k)^2 = \frac{n+2}{6(n+1)^2},$$

所以, 当 $k^2 < \frac{n+2}{6(n+1)^2}$ 时, 可以推断分布函数

$F_i(x)$ 拟合得好, k^2 越小, 拟合得越好。

2.3 拟合相关系数

设待选的分布函数为 $F(x)$, 将样本从小到大排列, 重新记为 $x_k (k = 1, 2, \dots, n)$, 样本值 x_k 对应的分布函数值记为 $F(k)$, 经验分布函数值定义为 $F_n(k) =$

$\frac{k}{n+1}$ 计算 $F(k)$ 与 $F_n(k)$ 的线性相关系数:

$$R = \frac{\sum (F_n(k) \cdot F(k)) - \frac{1}{n} \sum F_n(k) \sum F(k)}{[\sum F_n^2(k) - \frac{1}{n} (\sum F_n(k))^2] \cdot [\sum F^2(k) - \frac{1}{n} (\sum F(k))^2]} \quad (14)$$

R 表示分布函数值 $F(k)$ 与经验分布函数值 $F_n(k)$ 的线性相关程度, R 越大, 说明这种分布函数拟合得越好。

3 广西冬季极端最低气温的概率分布函数

对广西 88 个气象站 1951~2000 年 (1951 年以后建站的站, 取建站~2000 年) 冬季极端气温序列, 分别用第 5 节的 5 种概率分布进行拟合, 求其参数并按第 5 节的方法进行拟合优度检验, 检验结果见表 1

从表 1 看, 选择的 5 种概率分布函数对广西多数站冬季极端最低气温概率分布的拟合都比较好, 而正态分布更适合。各站冬季极端最低气温都符合正态分布, 有 8 站服从双正态分布。

根据拟合优度检验结果, 如果每站选取一个拟合得最好的分布函数, 则按检验标准将各分布函数拟合优度好的站数列于表 2

表 2 的结果说明, 有相当多的站 (3 站) 以双正态分布拟合得最好。其主要原因可能是: 虽然各站冬季极端最低气温都服从正态分布, 但有些站极端最低气温服从正态分布是近似的, 实际上具有一定的偏态

表 1 拟合优度检验结果

Table 1 The results of fitness test

概率分布 Probability distribution	服从站数 (个) Stations fitted	不服从站数 (个) Stations unfitted
Gumbell I 型极值分布 Gumbell extreme distribution	79	9
Gumbell II 型极值分布 Gumbell extreme distribution	73	15
Gumbell III 型极值分布 Gumbell III extreme distribution	83	5
正态分布 Normal distribution	88	0
双正态分布 Binormal distribution	83	5

表 2 拟合得最好的站数

Table 2 The better fitting stations

概率分布 Probability distribution	k^2 最小 Minimum	D_n 最小 Minimum	k^2 和 D_n 同时最小 Both minimum
Gumbell I 型极值分布 Gumbell extreme distribution	1	2	1
Gumbell II 型极值分布 Gumbell extreme distribution	1	0	0
Gumbell III 型极值分布 Gumbell III extreme distribution	17	6	3
正态分布 Normal distribution	29	33	22
双正态分布 Binormal distribution	40	47	31

性 (非对称性)。这些站的冬季极端最低气温的分布函数更接近于类似于正态分布但又非对称的双正态分布。例如, 柳州冬季极端最低气温服从正态分布, $m = 1.05$, $e = 1.7553$; 同时, 该站冬季极端最低气温更近似地服从双正态分布, $m_1 = 0.62$, $e_1 = 1.4547$, $e_2 = 1.9990$ 从图 1 直观看, 柳州冬季极端最低气温服从双正态分布要优于正态分布。

也有部分站的冬季极端最低气温更适合 Gumbell 分布。例如, 百色冬季极端最低气温近似地服从 Gumbell I 型分布, 其分布函数为

$$F(x) = 1 - \exp(-\exp(0.6204(x - 4.1262)))$$

对应的概率密度函数为

$$f(x) = F'(x) = 0.6204 \cdot \exp(0.6204(x - 4.1262)) \cdot \exp(-\exp(0.6204(x - 4.1262)))$$

从图 2 百色冬季极端最低气温的频率直方图看 Gumbell I 型拟合得比较好。

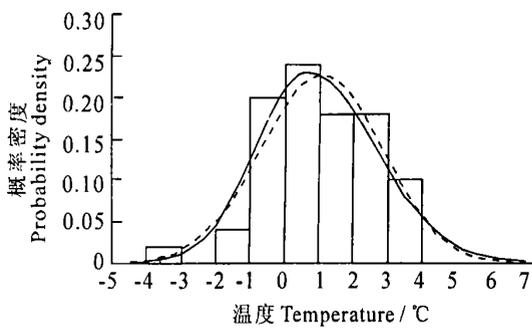


图1 柳州冬季极端最低气温的频率直方图及概率密度曲线

Fig.1 The histogram and probability density curve of winter extremely minimum temperature in Liuzhou

双正态分布 Binormal distribution; ----正态分布 normal distribution.

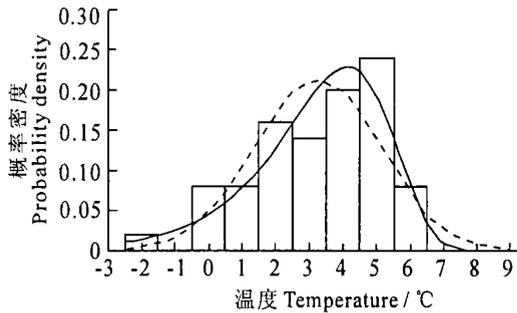


图2 百色冬季极端最低气温的频率直方图

Fig.2 The histogram and probability density function of minimum temperature in Baise

Gumbel I 型分布 probability density function of Gumbel I type; ----正态分布 normal probability density function.

根据表 1 表 2 的结果逐站进行分析,发现 Gumbel II 型拟合效果不佳, Gumbel III 型拟合效果也不理想,同时还需要对数据作预处理,经过试验后决定不用。这样,广西各站冬季极端最低气温的分布函数分别选用正态分布、双正态分布和 Gumbel I 型极值分布较好。

具体选择的的原则为:在选出的 3 种分布中,比较三者拟合优度检验的统计量 D_n 和 k^2 哪种分布 D_n 和 k^2 同时最小,则选择哪种分布;若 D_n 和 k^2 只有一个最小,则选择两者拟合相关系数 R 大的一种分布。这样,广西 88 站冬季极端最低气温的概率分布函数为:桂林等 55 站遵从双正态分布,河池等 29 站遵从正态分布,南宁等 4 站遵从 Gumbel I 型极值分布。

4 广西冬季极端最低气温不同重现期极值分析

4.1 不同重现期极值计算

不同重现期的极值 (x_T) 的求法因分布函数不同而异,例如:

① 若某站冬季极端最低气温服从 Gumbel I 型分布,则有:

$$F(x_T) = 1 - \exp(-\exp(T_1(x_T - U_1))) = \frac{1}{T},$$

解得

$$x_T = \frac{1}{T_1} \ln(-\ln(1 - \frac{1}{T})) + U_1.$$

② 若某站冬季极端最低气温服从双正态分布,则由 (1) 式和 (11) 式,有

$$F(x_T) = \frac{2e_1}{e_1 + e_2} H(\frac{x_T - m_1}{e_1}) = \frac{1}{T}, x_T \leq m_1,$$

$$F(x_T) = \frac{e_1 - e_2}{e_1 + e_2} + \frac{2e_2}{e_1 + e_2} H(\frac{x_T - m_1}{e_2}) = \frac{1}{T},$$

$x_T > m_1,$

经过推导,当 $\frac{e_1}{e_1 + e_2} \geq \frac{1}{T}$ 时,有 $x_T \leq m_1$,可得:

$$H(\frac{x_T - m_1}{e_1}) = \frac{1}{T} \frac{e_1 + e_2}{2e_1};$$

当 $\frac{e_1}{e_1 + e_2} < \frac{1}{T}$ 时,有 $x_T > m_1$,可得

$$H(\frac{x_T - m_1}{e_2}) = \frac{e_1 + e_2 - T(e_1 - e_2)}{2e_1 T}.$$

反查正态分布表,即可求得对应的 x_T 。也可采用计算方法,在计算机上直接求得。

③ 若某站冬季极端最低气温服从正态分布,则根据 (6) 式,有

$$F(x_T) = \int_{-\infty}^{x_T} h(t) dt = H(\frac{x_T - m}{e_1}) = \frac{1}{T}.$$

反查正态分布表,即可求得 T 年一遇的冬季极端最低气温值 x_T 。

重现期为 15 年时,根据各站冬季极端最低气温遵从的概率分布函数,计算出广西冬季极端最低气温的极值分布 (见图 3)。从图 3 可见 15 年一遇的极端最低气温 -2°C 等值线与广西龙眼、荔枝的分布北界比较接近。

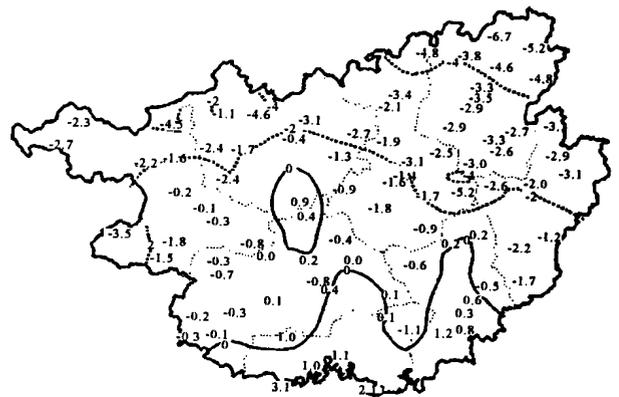


图3 广西冬季 15 年一遇的极端最低气温 ($^{\circ}\text{C}$) 分布图

Fig.3 The winter extremely minimum temperature ($^{\circ}\text{C}$) of 15-year reproduction period in Guangxi

4.2 极端最低气温实测值的重现期计算

对于一些特殊年(例如霜冻严重的年份),往往需要判断冬季极端最低气温实测值对应的重现期 T 。

对冬季极端最低气温而言,只要确定了分布函数 $F(x)$,就可由(1)式求出其实测值 x_T 对应的重现期 T 。

由(1)式可得

$$T = \frac{1}{F(x_T)};$$

对于 Gumbel I 型分布,

$$T = \frac{1}{1 - \exp(-\exp(\frac{x_T - U_1}{T_1}))};$$

对于双正态分布,若 $x_T \leq m_1$,则

$$T = \frac{1}{\frac{2\epsilon_1}{\epsilon_1 + \epsilon_2} H(\frac{x_T - m_1}{\epsilon_1})};$$

对于正态分布,为

$$T = \frac{1}{H(\frac{x_T - m}{\sigma})}.$$

例如:1999年冬季,广西发生了一次建国以来灾害损失最大的霜冻,广西受灾面积共 133.33万 hm^2 ,霜冻灾害造成的经济损失超过 200亿元人民币。这次霜冻过程各站的极端最低气温:桂东北、桂西山区 - 6.8 ~ - 2.0 $^{\circ}C$,其余地区 - 2.0 ~ 4.0 $^{\circ}C$ 。其中桂东

有 6站出现了建站以来的最低值。根据以上方法计算出桂东大部、桂西北局部为 10年以上一遇的低温,其中,桂东有 11站出现了 30年以上一遇的极端最低气温。

参考文献

- 1 刘钧枢,赵贻成,涂方旭.气象站天气预报.北京:气象出版社,1985.57~58
- 2 Gumbel E J. Statistics of Extremes. New York: Columbia University Press, 1958. 156~165.
- 3 谭冠日,严济远,朱瑞兆,应用气候.上海:上海科学技术出版社,1985.54~70.
- 4 林少宫.基础概率与数理统计.第2版.北京:人民教育出版社,1978.262~271.
- 5 Zoltan Toth, Tamas Szentimrey. The binormal distribution: a distribution for representing asymmetrical but normal-like weather elements. Journal of Climate, 1990, 3: 128~136
- 6 涂方旭.梧州最高水位的重现期分析.广西水利水电,1995,(2): 11~13.

(责任编辑:邓大玉)

(上接第 72页 Continue from page 72)

择在大潮期间进行,以保持湖内水质的及时交换、清新。

3.3 污水处理收集工程

据前所述,超标的废水未经处理就排放到泻湖和海域中,污染了海水;同时还通过下渗污染地下水和沙滩。所以,必须要在银滩旅游区附近建立污水收集系统,使各种污水的地表漫流情况得到控制,减少银滩及其海域的水质污染及沙滩污染。

综上所述,北海银滩由于不合理的开发,沙滩原生植被减少,生态环境遭到破坏;沿岸泥沙动力条件失去平衡,沙丘、沙滩的发育不正常,海岸侵蚀后退;

沿岸周边地区污水超标准排放及泻湖底泥污染,使海水和地下水的污染日趋严重。所以,进行北海银滩的自然环境综合整治已成为当务之急。

参考文献

- 1 黎广钊,刘敬合,方国详.南流江三角洲沉积特征及其环境演变.广西科学,1994,1(3): 21~25.
- 2 陈波.广西南流江三角洲海洋环境特征.北京:海洋出版社,1997.

(责任编辑:蒋汉明)