

中国西南地区干旱 Copula 函数模型对样本量的敏感性分析

左冬冬 侯威 王文祥

Sensitivity analysis of sample number on the drought descriptive model built by Copula function in south-west China

Zuo Dong-Dong Hou Wei Wang Wen-Xiang

引用信息 Citation: [Acta Physica Sinica](#), 64, 100203 (2015) DOI: 10.7498/aps.64.100203

在线阅读 View online: <http://dx.doi.org/10.7498/aps.64.100203>

当期内容 View table of contents: <http://wulixb.iphys.ac.cn/CN/Y2015/V64/I10>

---

您可能感兴趣的其他文章

Articles you may be interested in

具有质量及频率涨落的欠阻尼线性谐振子的随机共振

[Stochastic resonance of an underdamped linear harmonic oscillator with fluctuating mass and fluctuating frequency](#)

物理学报.2015, 64(2): 020202 <http://dx.doi.org/10.7498/aps.64.020202>

多载波微放电中二次电子横向扩散的概率分析

[Probabilistic analysis of the lateral diffusion of secondary electrons in multicarrier multipactor](#)

物理学报.2014, 63(22): 220205 <http://dx.doi.org/10.7498/aps.63.220205>

无线传感器网络中无标度拓扑的动态容错性分析

[Dynamic fault-tolerance analysis of scale-free topology in wireless sensor networks](#)

物理学报.2014, 63(11): 110205 <http://dx.doi.org/10.7498/aps.63.110205>

超网络中标度律的涌现

[Emergence of scaling in hypernetworks](#)

物理学报.2014, 63(9): 090207 <http://dx.doi.org/10.7498/aps.63.090207>

基于动态最小生成树路由协议的数据聚融算法

[Data fusion algorithm based on dynamic minimal spanning tree routing protocol](#)

物理学报.2014, 63(9): 090206 <http://dx.doi.org/10.7498/aps.63.090206>

# 中国西南地区干旱 Copula 函数模型对样本量的敏感性分析\*

左冬冬<sup>1)</sup> 侯威<sup>2)†</sup> 王文祥<sup>1)</sup>

1)(扬州大学物理科学与技术学院, 扬州 225002)

2)(国家气候中心, 气候研究开放实验室, 北京 100081)

(2014年10月20日收到; 2014年12月3日收到修改稿)

基于中国西南地区(四川省、云南省、贵州省和重庆市) 89个站点 1961—2010年的逐月标准化降水指数序列, 利用游程理论和 Copula 函数建立干旱历时和干旱强度的概率模型, 基于该模型讨论了干旱事件样本量对分布参数、干旱类型的出现概率及联合重现期的影响. 结果表明: 分布参数稳定对样本量的要求较大, 部分区域样本量要求大于 50, 并且各参数对样本量的要求不一致, 又以干旱强度分布的参数对样本量的要求最大; 干旱事件样本量为 10 个左右求得的干旱类型出现概率和联合重现期与样本量为 40 求得的结果已无明显差异, 以计算结果作为标准可大幅降低统计模型建立对样本量的要求, 进而表明起止时间不一致和具有缺测数据的站点仍可建立干旱历时和干旱强度的分布函数; 气候变暖对模型建立所需的最少样本量影响不大, 样本量波动在  $\pm 5$  之间, 即统计模型具有一定的稳定性, 同时气候态的划分降低了分布检验对样本量的需求, 易于模型的建立.

**关键词:** 中国西南地区干旱, 游程理论, Copula 函数, 样本量

**PACS:** 02.50.-r, 92.40.De

**DOI:** 10.7498/aps.64.100203

## 1 引言

干旱灾害是全球最为常见、也是最严重的自然灾害之一, 其特点是影响范围大、持续时间长, 给世界各国的国民经济特别是农业生产带来了巨大的经济损失. 崔冬林<sup>[1]</sup> 研究指出, 中国两个主要的干旱中心区域为华北和西南地区, 并且西南地区的干旱强度属于次级中心. 但近些年, 西南地区频繁出现高强度、大范围的干旱事件, 导致当地大面积的农作物绝收、生活用水困难和巨大的经济损失<sup>[2-4]</sup>. 为此对西南地区干旱规律及成因的研究具有重要的现实意义.

我国地处东亚季风区, 影响我国气候异常的因素复杂, 加之气候预测本身的困难与准确率不

高<sup>[5-12]</sup>, 因此对该地区的干旱进行监测与检测仍有必要. 利用 Copula 函数<sup>[13-15]</sup> 建立起干旱历时和干旱强度的联合概率模型, 该模型有助于干旱特征分析, 同时也可作为干旱监测的有力工具. 然而, 目前尚未有把干旱历时和干旱强度综合考虑来研究西南地区干旱特征的工作, 而该方法已经较多地应用于其他区域干旱的研究中, 并得到了很好的结果. 如, Zhang 等<sup>[16]</sup> 利用 Copula 函数分析了中国广东珠江流域的干旱特征, 得出该地区冬季有变干的趋势而夏季有变湿的趋势, 并且在珠江下游区域较上游区域有更大的干旱风险; 许玲燕等<sup>[17]</sup> 利用 Copula-EVT 模型分析了淮河流域蚌埠站的干旱特征, 指出蚌埠站干旱历时大于 5 个月、干旱强度大于 7.45 的极端干旱灾害风险概率为 3%, 并且干旱强度的大小对联合重现期影响较大; Liu 等<sup>[18]</sup> 利用

\* 国家自然科学基金(批准号: 41305056, 41175067)和国家重点基础研究发展计划(批准号: 2012CB955901)资助的课题.

† 通信作者. E-mail: houwei@cma.gov.cn

Copula 函数分析了广东省的干旱特征, 指出广东省南部地区的干旱出现概率大于北部地区东部及西部地区干旱风险较大; 文献 [19—23] 也报道了类似的研究.

在利用 Copula 函数分析干旱特征时, 常采用数据起止时间一致的站点, 而略去数据长度不一致及具有缺测数据的站点, 并且利用各个站点全部的干旱事件作为样本量来建立分布函数. 虽然在求解分布函数的参数时希望存在足够多的干旱事件样本量, 从而使求得的分布参数值更加接近期望值, 但在实际情况中各站的样本量均是有限的, 无法满足理想的样本量需求. 由于样本量的限制, 各分布参数的求解值与理想值都存在一定的偏差; 同时由于气候态的变化, 研究区域的干旱特征也会随之变化, 从而使得不同气候态下得到的干旱指标也会存在变化, 进而影响利用干旱指标建立的分布方程. 由于实际应用中希望建立的分布方程能较好地体现近期的干旱特征, 而利用不同气候态下的干旱指标共同建立的分布方程不易体现近期的干旱特征情况. 因此, 有必要研究各站点使用 Copula 函数建立干旱历时和干旱强度二维概率统计模型时得到合适的各分布参数所需的最小样本量, 从而最大限度地利用现有数据以及了解不同气候态下统计模型建立所需的样本量变化情况, 进而得到 Copula 函数在研究区域的适用性.

本文首先利用游程理论 [24—27] 的方法从西南地区各站的标准化降水指数 (standardized precipitation index, SPI) 序列中分离出各站干旱事件的干旱历时和干旱强度, 再利用干旱历时和干旱强度建立二维概率统计模型, 从而分析西南地区在不同干旱事件样本量的条件下干旱事件发生概率与联合重现期计算结果之间的差异. 全文主要分为两个部分: 1) 样本量的敏感性分析; 2) 气候变暖对模型建立所需的最小样本量的影响.

## 2 资料

所采用的数据是由中国气象局提供的中国西南地区 (四川省、云南省、贵州省和重庆市) 89 站 1961—2010 年逐日无缺测的日平均降水资料和温度资料, 再将降水数据处理成逐月月平均数据, 温度数据处理成逐年年平均数据. 站点分布如图 1 所示.

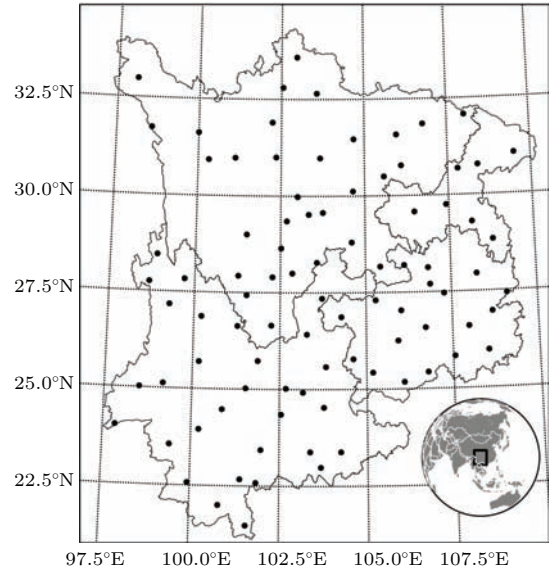


图 1 西南地区站点空间位置分布图

Fig. 1. The location of meteorological stations in southwest.

## 3 方法

### 3.1 标准化降水指数

本文采用的干旱指标是标准化降水指数 SPI 计算公式如下:

$$SPI_i = \frac{y_i - \bar{y}}{\sigma_y}, \quad (1)$$

其中  $y_i = \ln(x_i)$ ,  $x_i$  表示月平均降水量,  $\bar{y}$  表示历史同月的平均值,  $\sigma_y$  表示历史同月的标准差.

### 3.2 干旱事件的分离

干旱可以简单地理解为一段时间内由于供水量的匮乏而引起的缺水现象, 为此, 一次干旱过程应包含有两个特征量, 一是干旱持续时间 (干旱历时), 二是该段时间内比正常需水量偏少的水量 (干旱强度). 本文对干旱事件的分离采用的方法是游程理论, 该方法是分析时间序列的一种有效方法. Herbst 等 [28] 首先运用该方法对干旱进行了识别. 游程理论也称为轮次理论, 是指连续出现同类事件的前后为另一类事件, 比如旱涝、连续的有雨日和无雨日. 根据该理论 [24—29], 可得干旱历时和干旱强度的大小, 其中干旱历时  $d$  表示的是连续出现降水相对历史同月偏少的时间跨度, 干旱强度  $s$  大小为

$$s = - \sum_{i=1}^d SPI_i, \quad (2)$$

其表示的是在干旱时间跨度内的  $SPI$  累加值。

对干旱的分析常采用分等级的方式, 本文同样对干旱历时和干旱强度进行划分, 并且将二者的等级两两组合得到具有一定实际意义的干旱类型, 如月内轻旱、月内中旱等. 表 1 和表 2 分别给出了干旱历时和干旱强度的等级划分 [18,29].

表 1 干旱历时等级  
Table 1. The level of drought duration.

$D/\text{months}$	等级
$0 < D \leq 1$	月内
$1 < D \leq 3$	季内
$3 < D \leq 6$	跨季
$6 < D$	半年以上

表 2 干旱强度等级表  
Table 2. The level of drought severity.

$S$	干旱强度
$S \leq 0.5$	无旱
$0.5 < S \leq 1$	轻旱
$1 < S \leq 1.5$	中旱
$1.5 < S \leq 2$	重旱
$2.0 < S$	特旱

### 3.3 边缘分布函数

利用 3.2 小节中的方法便能从  $SPI$  序列中分离出各次干旱事件的干旱历时和干旱强度. Zenlenhastic 和 Salvai [30] 以及 Shiau 和 Hsieh [31] 的研究指出, 干旱历时符合指数分布, 而干旱强度符合 Gamma 分布. 两种分布的数学表达式如下:

$$F_D(d) = 1 - e^{-\lambda d}, \quad (3)$$

$$F_S(s) = \int_0^s \frac{s^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{s}{\beta}} ds, \quad (4)$$

表 3 常用的阿基米德型 Copula 函数  
Table 3. Conventional Archimedean Copulas.

名称	Copula 函数 $C_\theta(\mu, \nu)$	$\tau$ 与 $\theta$ 的关系
Clayton	$(\mu^{-\theta} + \nu^{-\theta} - 1)^{-1/\theta}$	$\tau = \theta/(\theta + 2)$
Frank	$-\frac{1}{\theta} \ln \left[ 1 + \frac{(e^{-\theta\mu} - 1)(e^{-\theta\nu} - 1)}{e^{-\theta} - 1} \right]$	$\tau = 1 + \frac{4}{\theta} \left[ \frac{1}{\theta} \int_0^\theta \frac{t}{\exp(t) - 1} dt - 1 \right]$
Gumbel-Hougaard	$\exp\{-[(-\ln \mu)^\theta + (-\ln \nu)^\theta]^{1/\theta}\}$	$\tau = 1 - 1/\theta$

注: 表中的  $\mu, \nu$  表示边缘分布函数,  $\theta$  为 Copula 函数参数.

其中  $D, S$  分别表示干旱历时和干旱强度的样本空间;  $d, s$  分别表示干旱历时和干旱强度样本空间中的元素;  $F_D(d)$  表示干旱历时小于等于  $d$  的概率;  $F_S(s)$  表示干旱强度小于等于  $s$  的概率; 而  $\lambda, \alpha, \beta$  为分布参数.

### 3.4 Copula 函数

由于干旱历时和干旱强度的分布类型不一致且二者并不完全独立, 使得从理论上给出二者的联合分布较为困难, 但 Copula 函数能较好地解决此类问题. 其数学表达式为

$$F(x, y) = C[F_X(x), F_Y(y)] \quad \forall x, y, \quad (5)$$

式中  $F(x, y)$  称为联合分布函数,  $C[\ ]$  表示的是 Copula 函数,  $F_X(x)$  和  $F_Y(y)$  分别是随机变量  $X, Y$  的边缘分布函数. 目前对 Copula 的划分主要有三类: 椭圆型、阿基米德型和二次型. 阿基米德型 Copula 函数由于计算简便而应用广泛, 本文采用该模型进行计算. 表 3 为常用的阿基米德型 Copula 函数 [24,32], 其中

$$\tau = \frac{2}{n(n-1)} \sum_{i < j} \text{sign}[(d_i - d_j)(s_i - s_j)],$$

$n$  为序列长度,  $\tau$  称为 Kendall 相关系数, 描述的是干旱历时和干旱强度之间的相关性.

本文对常用的三种阿基米德 Copula 函数进行了对比分析, 采取的判断标准为: 利用三种阿基米德 Copula 函数得到的理论联合概率与经验得到的联合概率值做离差平方和, 其中离差平方和最小的 Copula 函数即为最优方案. 经对比发现, 所有站点利用 Gumbel-Hougaard 类型进行计算得到的离差平方和最小. 为此, 本文的后续工作均基于 Gumbel-Hougaard 类型的 Copula 函数进行计算.

### 3.5 联合分布和重现期

由干旱历时和干旱强度的边缘分布函数  $F_D(d)$  和  $F_S(s)$  及 Copula 函数可得到联合分布函数

$$F(d, s) = P(D \leq d, S \leq s) = C[F_D(d), F_S(s)], \quad (6)$$

联合分布描述的是干旱历时和干旱强度都小于等于给定值的干旱事件出现的概率。

重现期是研究干旱的一个重要指标, 其计算公式如下 [31,33]:

$$T_D = \frac{E(L)}{1 - F_D(d)}, \quad (7)$$

$$T_S = \frac{E(L)}{1 - F_S(s)}, \quad (8)$$

其中,  $T_D, T_S$  分别为干旱历时重现期、干旱强度重现期;  $E(L)$  为干旱间隔的平均值, 其值等于干旱历时和非干旱历时的期望值之和. 联合重现期的计算公式如下:

$$T(d, s) = \frac{E(L)}{P(D > d \cap S > s)} = \frac{E(L)}{1 - F_D(d) - F_S(s) + F(d, s)}. \quad (9)$$

本文首先利用西南地区 89 站 1961—2010 年的逐月月平均降水数据计算出各站的 SPI 序列, 再由该序列分离出该站的干旱事件, 之后从各站分离出的所有干旱事件样本量中随机选取一定数量的干旱事件, 并计算出干旱历时和干旱强度的分布参数, 进而利用 Copula 函数得到二者的联合分布函数, 至此完成干旱统计模型的建立. 再由上述 (6)—(9) 式便可得到联合概率与重现期的大小. 样本量的不同导致得到的联合概率与重现期也不相同; 此外气候增暖也会对统计模型产生影响. 因此, 对西南地区的逐年年平均温度序列进行 Mann-Kendall (MK) 检验 [34], 划分出增暖前期和增暖期, 从而讨论不同气候背景下干旱统计模型建立所需样本量的变化情况.

## 4 结果与分析

### 4.1 样本量的敏感性分析

统计模型的可信度依赖于边缘分布的可信度, 为此在模型建立时首先应对边缘分布函数 [(3), (4)

式] 进行分布检验. 图 2 给出了不同样本量条件下干旱历时和干旱强度分布的 Kolmogorov-Smirnov (KS) 检验结果和西南地区通过显著水平 0.01 时所需最低样本量大小的空间分布.

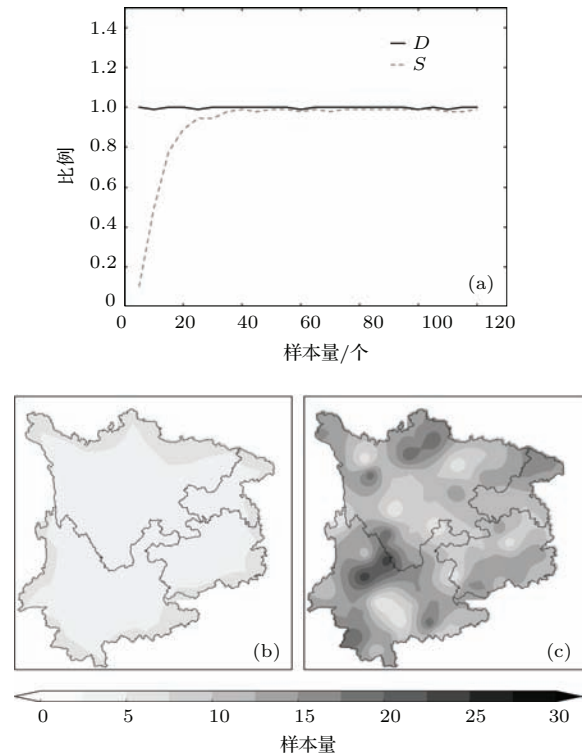


图 2 站点的干旱历时和干旱强度的 KS 检验通过率 (显著水平为 0.01) 和各区域通过显著水平为 0.01 检验时要求的最低样本量 (a) KS 检验站点通过率; (b) 干旱历时; (c) 干旱强度

Fig. 2. The KS test pass rate of stations drought duration and drought severity (significance level is 0.01) and the minimum sample size requirements for each area KS test (significant level is 0.01): (a) the pass rate of station by KS test; (b) drought duration; (c) drought severity.

图 2(a) 的横坐标为干旱事件的不同样本量大小, 纵坐标为通过分布检验的站点占所有站点的比例. 从图 2(a) 中可以看到, 干旱历时在样本量较小时, 几乎所有站点都通过了分布检验; 干旱强度分布在样本量为 20 时, 有约 90% 的站点通过了显著水平检验, 当样本量大于 40 时, 通过分布检验的站点几乎达到 100%. 说明干旱事件的样本量达到 20 时已有较好的检验结果, 当样本量达到 40 时几乎所有站点都能通过分布检验.

从图 2(b), (c) 可以看到, 干旱历时分布在大部分区域所需的样本量基本一致, 且数量大于 5 即能通过检验; 而干旱强度分布存在一定的空间差异, 四川中部至贵州西部以及云南南部部分区域对样

本量的要求偏低, 大约为10, 而四川与云南交界区域对样本量的要求偏高, 大约在25左右, 其他大部分区域的样本量要求在15—20之间. 由以上分析可得, 在干旱历时和干旱强度的分布检验中, 大部分区域的分布检验对样本量的要求并不高.

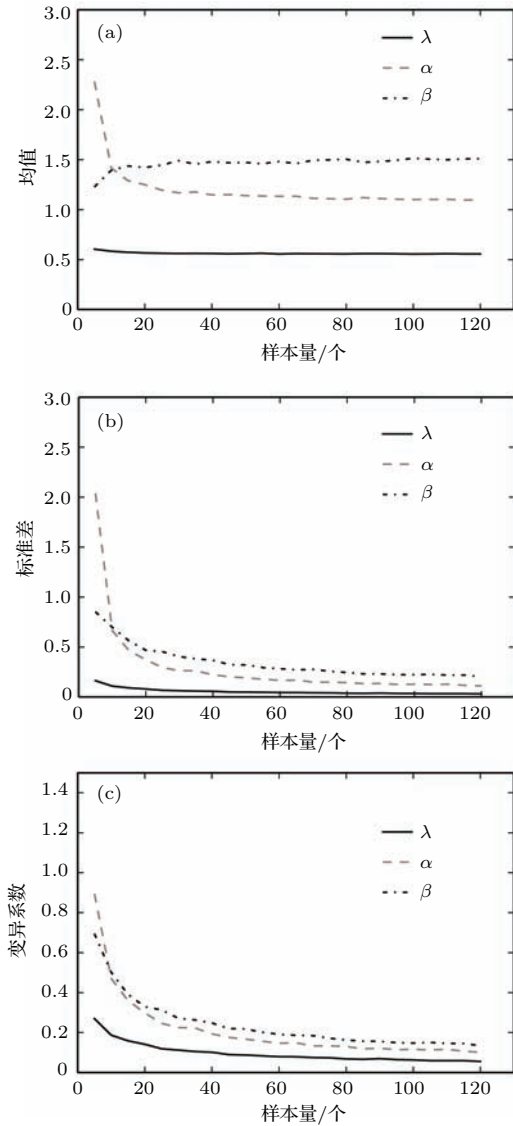


图3 分布参数的区域平均值随样本量的变化曲线 (a) 分布参数的均值; (b) 分布参数的标准差; (c) 分布参数的变异系数

Fig. 3. The relationship of distribution parameters regional mean and sample size: (a) the mean of parameters; (b) the standard deviation of parameters; (c) the coefficient of variation of parameters.

在干旱历时和干旱强度满足分布检验的基础上, 由于相同样本量的多次抽样计算得到的分布参数仍有差异, 而发生概率和联合重现期计算结果的稳定性与分布函数中的分布参数密切相关, 只有稳定的分布参数才能得到稳定的发生概率和联合重现期, 为此分析了分布参数与样本量之间的关系.

本文采取的抽样方法为: 首先从某站的所有干旱事件样本中随机抽取5个样本, 计算本次抽样得到的分布参数大小, 然后以相同的样本量重复抽样10次, 可得到10个分布参数的计算结果, 进而得到各分布参数10次抽样的均值、标准差和变异系数(标准差比均值), 对所有站点重复上述计算步骤, 便可得到样本量为5时各站对应的各分布参数均值、标准差和变异系数, 近一步便可计算得到各分布参数均值、标准差和变异系数的区域平均值. 逐次增加样本量的大小(10, 15, 20...), 重复上述计算. 图3是各分布参数统计量的区域平均值随样本量的变化曲线.

从图3(a)中可以看到, 随着样本量的增加, 各分布参数的区域平均值逐渐收敛于某一稳定值, 并且干旱历时分布参数 $\lambda$ 最快趋近稳定, 而干旱强度分布参数 $\alpha, \beta$ 趋于稳定较慢. 此外, 从图3(a)中可以看到, 样本量在20左右时各分布参数已基本趋于稳定, 说明此时多次抽样可以得到较为稳定的分布参数值. 但从图3(b)可以发现, 当样本量为20时, 参数 $\alpha, \beta$ 的标准差(离散程度)接近0.5, 相对原值[图3(a)中的大小]仍较大. 为比较单次抽样效果, 利用变异系数(标准差比均值)比较相对均值的离散状况, 如图3(c)所示. 从图3(c)可清楚地看到, 样本量越大, 单次抽样得到的分布参数越接近理想值, 且当样本量为20时, 参数 $\alpha, \beta$ 的变异系数较大, 说明不同抽样得到的分布参数之间有较大差异, 单次抽样的结果不理想.

以变异系数小于0.2即认为求解的参数较为理想. 由图3(c)可以看到, 变异系数在0.2左右时对应的样本量接近40, 即干旱事件样本量在一次抽样中达到40个时便能得到与理想值偏离较小的分布参数. 以变异系数为0.2作为界限, 可得到各分布参数满足该条件时所需最小样本量的空间分布图, 如图4所示.

从图4可以看到, 参数 $\lambda$ 对样本量的需求较低, 10—15个样本量即能满足变异系数小于0.2; 从空间分布来看, 对样本量要求相对较大的区域位于贵州与云南的交界区域以及重庆北部至四川的东北部, 其他地区对样本量需求相对较小; 整体上看, 东部区域对样本量的要求大于西部区域, 但由于所需样本量均都较小, 因此干旱历时分布函数在较少干旱事件样本量的条件下即可得到较理想的分布方程.

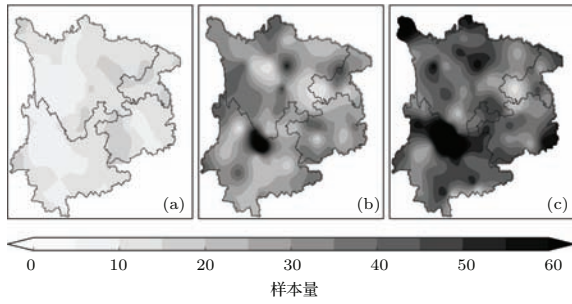


图4 变异系数达0.2时所需最小样本量空间分布 (a) 参数  $\lambda$ ; (b) 参数  $\alpha$ ; (c) 参数  $\beta$

Fig. 4. The minimum amount of sample required for variation coefficient of 0.2: (a) parameter  $\lambda$ ; (b) parameter  $\alpha$ ; (c) parameter  $\beta$ .

对于干旱强度分布的两个参数  $\alpha, \beta$  [图4(b), (c)], 参数  $\beta$  对样本量的要求明显高于参数  $\alpha$ , 其中  $\alpha$  参数对样本量的要求, 除四川与云南交界区域偏高和重庆、四川、贵州三省交界区域较低外, 其他地区均约为30个. 参数  $\beta$  在云南大部、四川北部和贵州西南部对样本量要求均较高, 其他地区接近30. 对比图4和图2可看出, 分布参数的稳定性对样本量的要求明显高于分布通过显著性检验所需的样本量, 因此在建立模型时, 当分布参数满足稳定性要求时所需样本量已能使分布通过显著性检验.

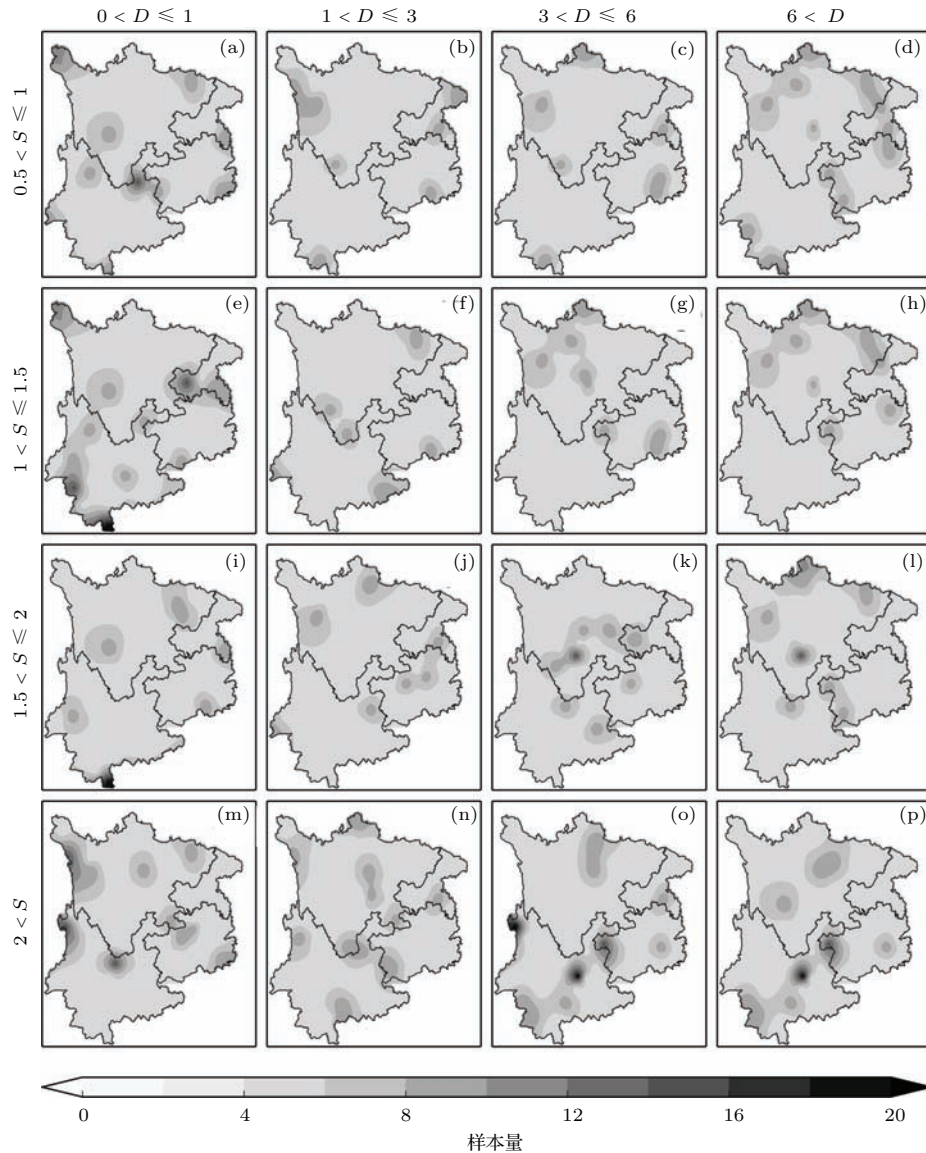


图5 计算各干旱类型概率时所需的最小样本量 (a) 月内轻旱 ( $0 < D \leq 1, 0.5 < S \leq 1$ ); (b)—(p) 类同 (a), 表示的是该图上方干旱历时区间与该图左边干旱强度区间组合得到的干旱类型

Fig. 5. The minimum sample size required to calculate probability of different drought types: (a) light drought within month ( $0 < D \leq 1, 0.5 < S \leq 1$ ); (b)—(p) same as (a), represent the drought types combination of drought duration on the top of the figure and drought severity in the left figure.

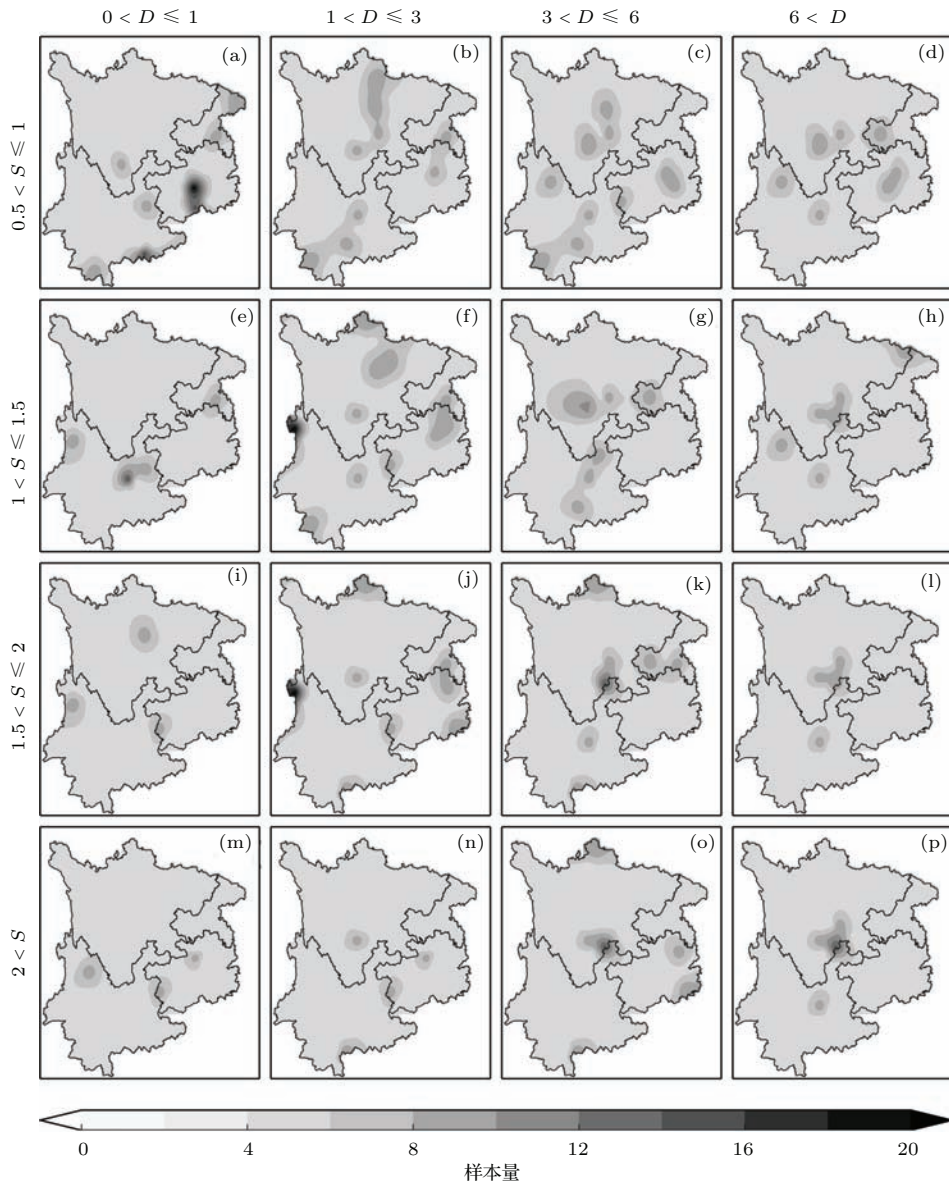


图 6 计算各干旱类型联合重现期时所需的最小样本量 (a)–(p) 含义同图 5

Fig. 6. The minimum sample size required to calculate recurrence of different drought types: (a)–(p) same as Fig. 5.

虽然不同样本量条件下计算得到的各参数大小存在差异,但在差异不大时计算得到的发生概率和联合重现期的差异亦不明显,此时便可利用较小样本量求得的参数来计算发生概率与联合重现期,从而在实际应用中优化模型对样本量的要求.由以上分析可知,干旱事件样本量在 40 时不仅分布检验有很好的通过率,同时求得的参数也有很好的稳定性.因此,可以认为当干旱事件样本量达到 40 个时,此时求解得到的发生概率与联合重现期是较理想的.在显著水平为 0.05 时,利用  $u$  检验 [35] 可以得到与样本量为 40 时求得的不同干旱类型的发生概率和联合重现期无显著差异的最小样本量大小,如

图 5 和图 6 所示.

从图 5 和图 6 可以看出,与样本量为 40 时求得的不同干旱类型的发生概率和联合重现期无显著差异的最小样本量很小,西南大部地区的值均在 4–8 之间,少数地区对样本量的要求在 12 以上.对比分布参数的稳定性对样本量的需求(图 4),可以看出,虽然满足分布参数稳定性要求时所需的样本量较大,但计算不同干旱类型的发生概率与联合重现期时所需的样本量相对前者要小许多,即并不理想的分布参数仍可得到与发生概率和联合重现期的理想值偏差较小的结果.从以上分析可知,分布参数的稳定性要求大大增加了对于样本量的要求,但



在实际应用中, 如果只关注发生概率与联合重现期的计算结果, 则完全可以利用小样本量计算得到的结果, 并且该结果同样是可信的. 因此, 利用该统计模型对西南地区的干旱特征进行分析时, 不仅可以利用起止时间不一致的站点数据, 而且无须利用所有的干旱事件样本. 如果站点数据是间断的(即数据在一段时间内为缺测值), 但该站有10个以上

的干旱事件样本, 仍能建立起可信的干旱历时和干旱强度的分布函数以及二者的联合分布函数.

#### 4.2 气候变暖对模型建立所需的最小样本量的影响

利用各站点的逐年年平均温度数据, 可得西南地区的逐年区域平均气温序列, 如图7(a)所示. 可

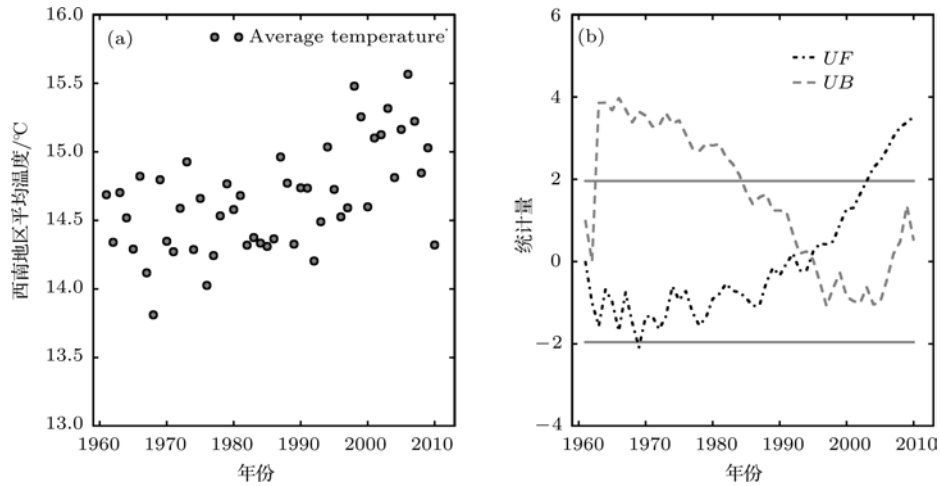


图7 西南地区的平均温度变化曲线及MK检验 (a) 西南地区区域年平均温度变化; (b) 西南地区平均温度的MK检验  
 Fig. 7. The curves of average temperature in the southwest and MK test: (a) the average temperature in the southwest; (b) the MK test of average temperature in the southwest.

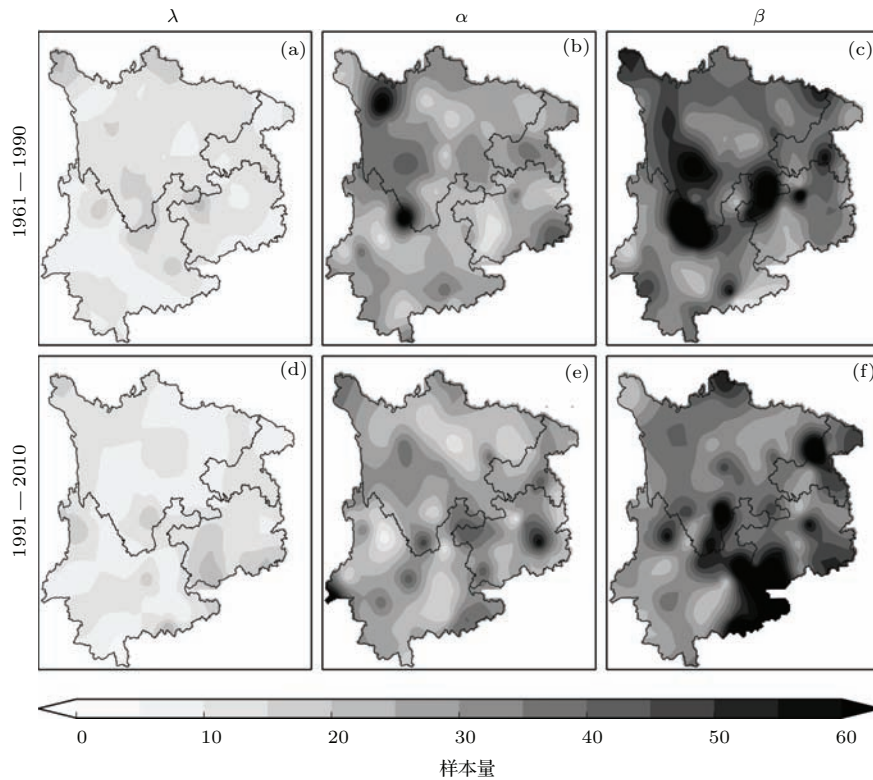


图8 增暖前期与增暖期各参数在变异系数达0.2时的最小样本量空间分布 (a), (d) 参数 $\lambda$ ; (b), (e) 参数 $\alpha$ ; (c), (f) 参数 $\beta$ .  
 Fig. 8. The minimum amount of sample required for variation coefficient of 0.2 in early warming and warming period: (a), (d) parameter  $\lambda$ ; (b), (e) parameter  $\alpha$ ; (c), (f) parameter  $\beta$ .

以看到, 1990年以前西南地区的平均气温变化较平缓, 而在1990年以后有明显的上升趋势. 为此, 用MK法检验1961—2010年西南地区平均气温序列的突变[图7(b), 其中 $UF$ ,  $UB$ 是按照文献[34]中的方式进行计算的, 表示两个统计量], 由图7(b)中的 $UF$ 曲线可以看出, 在20世纪90年代后期, 该地区的平均气温有一明显的增暖趋势, 且2003年以后的增暖趋势超过了显著性水平0.05临界线, 表明该地区确有明显增温的事实. 根据MK检验的原理可知, 曲线 $UF$ 和 $UB$ 的焦点处即为一突变点, 该点对应的具体时间为1993年. 刘瑜等[36]的研究表明, 云南地区夏季降水在1971年前后发生了一次由多到少的突变, 冬季平均气温在1995年出现了一次由冷到暖的突变; 丁一汇和张莉[37]指出在20世纪80年代到90年代全国7个主要区域地表温度都发生了一次突变. 此外众多研究表明[38,39], 在20世纪90年代前后确实存在一次气候转变的过

程. 因此, 本文以1990年为界线, 划分为增暖前期(1961—1990年)和增暖期(1991—2010年), 并分别讨论在这两种不同气候背景下统计模型对样本量要求的变化.

按前文4.1节方法检验了两个时段分布参数随样本量的变化, 结果显示(图略), 增暖前期和增暖期在样本量为20时, 各分布参数的均值基本稳定, 但此时变异系数仍较大; 当样本量为40时, 各分布参数的变异系数小于0.2, 结果与图3类似.

当各分布参数的变异系数小于0.2时, 其所需最小样本量的空间分布如图8所示. 参数 $\beta$ 在增暖前期和增暖期对样本量要求的空间分布差异较大; 在增暖前期, 样本量要求较大的区域位于云南、贵州、四川的交界处, 而增暖期的大值区位于云南、贵州、广西的交界处. 而参数 $\alpha$ 和 $\lambda$ 则在增暖前后对样本量要求的差异较小, 在两个时段中大多数区域的样本量要求相当.

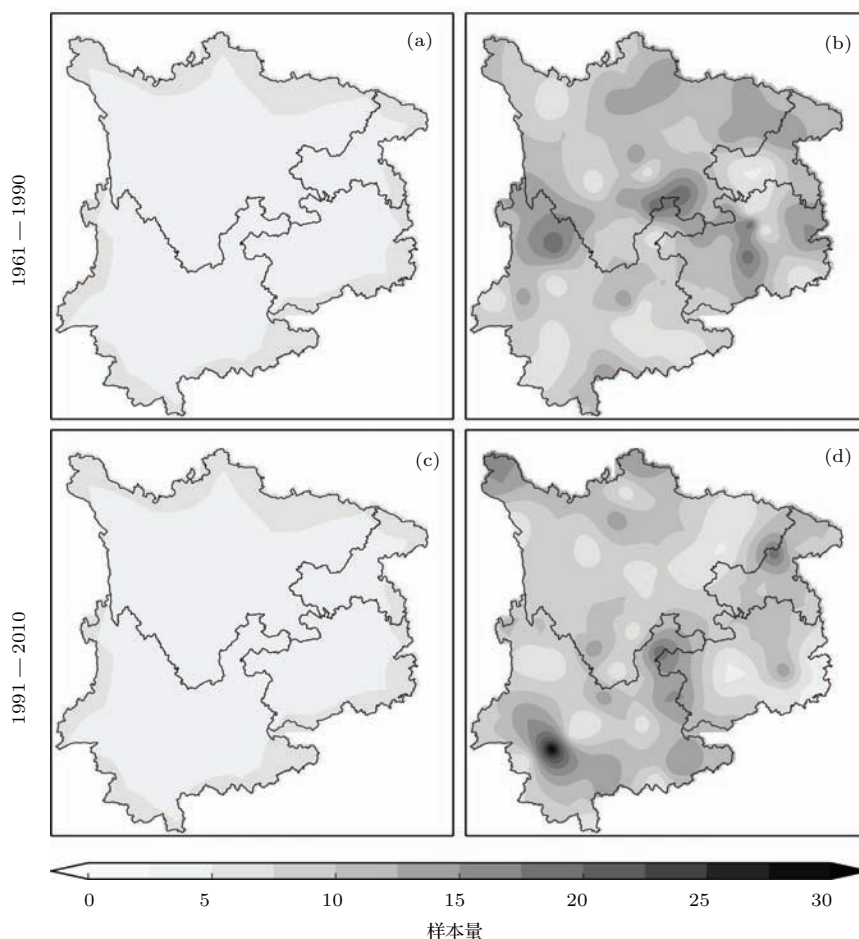


图9 增暖前期与增暖期分布检验达到显著水平0.01时所需的最小样本量 (a), (c) 干旱历时; (b), (d) 干旱强度  
 Fig. 9. The minimum sample size requirements for each area by KS test in early warming and warming period (significant level is 0.01): (a), (c) drought duration; (b), (d) drought severity.

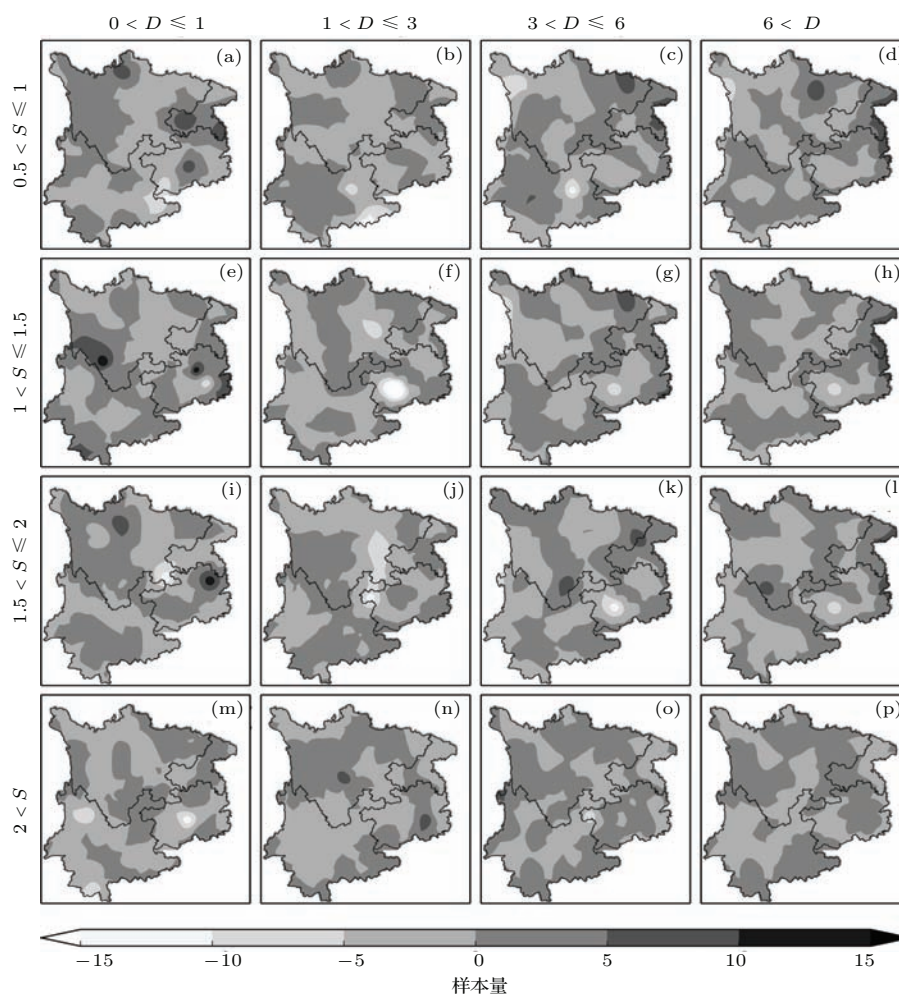


图 10 增暖期与增暖前期计算各干旱类型概率时所需的最小样本量差异 (a)–(p) 含义同图 5  
 Fig. 10. The difference of the minimum sample size required to calculate probability of different drought types between warming period and early warming: (a)–(p) same as Fig. 5.

对增暖前期和增暖期的干旱历时和干旱强度的分布函数进行显著性检验(图 9), 结果显示气候背景的变化对干旱历时的分布函数显著性检验的影响不大, 但对干旱强度的分布参数显著性检验的影响较大. 与图 2 对比, 可得西南大部分区域的分布函数通过显著性检验时, 在增暖前期和增暖期均比全时段所需的样本量有明显的降低. 因此, 气候背景的变化对模型中边缘分布的显著性检验有较大影响, 尤其是对干旱强度分布函数的影响.

使用与前文 4.1 节中得到图 5 和图 6 的相同方法, 计算得到了增暖期与增暖前期两个阶段中, 不同干旱类型发生概率和联合重现期所需最小样本量的差异(增暖期所需的最小样本量减增暖前期所需的最小样本量). 从图 10 和图 11 可以看出, 计算各干旱类型在增暖期与增暖前期发生概率和联合重现期时, 所需的最小样本量差异不大, 一般在  $\pm 5$

间波动, 即仍可利用较少的样本量建立可信的干旱历时和干旱强度的分布函数. 因此, 气候增暖对统计模型所需样本量的影响不大, 即统计模型在增暖前后并未出现某种干旱类型对样本量有明显的增加和减少, 具有一定的稳定性. 同时由于两个阶段中干旱历时和干旱强度分布函数通过显著性检验所需的样本量(图 9)相对全时段分布函数数通过显著性检验所需的样本量(图 2)略有减少, 更易建立模型.

## 5 讨论与结论

利用游程理论从西南地区各站的 SPI 序列中分出每次干旱事件的干旱历时和干旱强度, 再利用 Copula 函数建立起干旱历时和干旱强度的统计模型. 利用该模型讨论了干旱类型出现概率和联合重现期计算时所需的最小样本量.

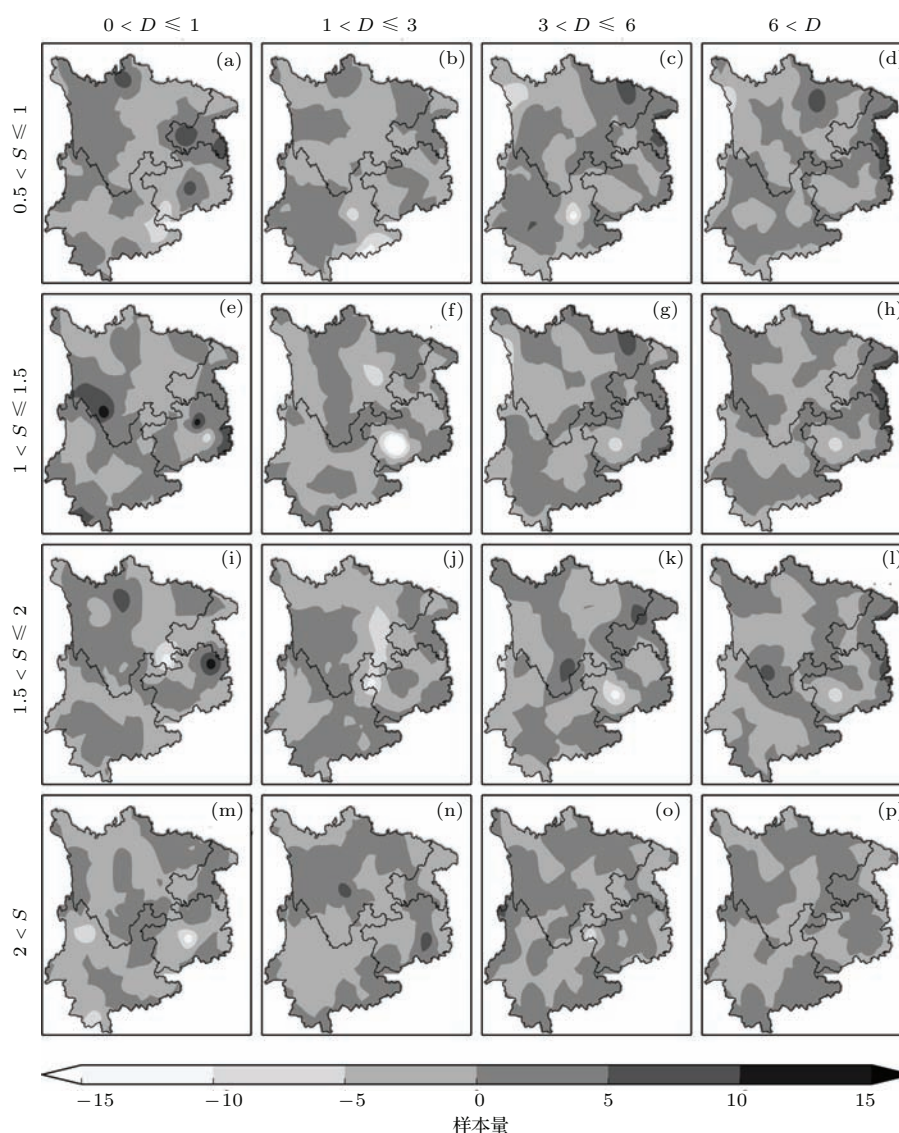


图 11 增暖期与增暖前期计算各干旱类型联合重现期时所需的最小样本量差异 (a)–(p) 含义同图 5

Fig. 11. The difference of the minimum sample size required to calculate recurrence of different drought types between warming period and early warming: (a)–(p) same as Fig. 5.

干旱历时和干旱强度的分布检验对样本量的要求都不大, 干旱事件样本量为 10 个时即可使各区域的干旱历时分布通过显著水平为 0.01 的分布检验, 而样本量为 25 时即可使 95% 站点的干旱强度分布通过分布检验. 干旱历时分布检验对样本量的空间分布要求较为一致, 而干旱强度分布检验对样本量要求略高的区域位于云南与四川交界处. 分布参数  $\lambda$ ,  $\alpha$ ,  $\beta$  稳定时对样本量要求差异较大, 参数  $\lambda$  对样本量要求最低, 多为 10 个左右, 参数  $\beta$  对样本量要求最高, 多数区域为 30 左右, 部分区域在 50 以上, 整体表现为干旱强度分布函数的建立对样本量要求相对干旱历时对样本量的要求偏高. 大部分区域在样本量为 10 左右得到的干旱类型出现概

率和联合重现期与样本量为 40 求得值已无显著差异, 即样本量较少时求得的结果与较大时求得的结果差异不明显. 气候态的变化对参数  $\beta$  的影响较大, 而对模型建立所需的最小样本量的影响不大, 大小波动在  $\pm 5$  之间.

综上所述, 建立较合理的统计模型所需的样本量较大, 尤其是干旱强度分布函数对样本量的要求较大, 但实际上如果只关注干旱类型和联合重现期的结果大小, 样本量为 10 时计算的结果已较好, 大大减少了统计模型对样本量的需求. 因此那些记录时间长度不长或存有缺测数据的站点仍可作为分析数据使用, 从而提高现有数据的使用率. 另外, 气候态的变化对建立统计模型所需的最小样本量

影响不大, 即各气候态下仍可利用较少的样本量进行计算分析.

参考文献

- [1] Cui D L 2010 *M. S. Dissertation* (Lanzhou: Lanzhou University) (in Chinese) [崔冬林 2010 硕士学位论文 (兰州: 兰州大学)]
- [2] Liu J G, Wan J H, Tan X M, Ma J M, Zhang N Q 2011 *J. Disaster Preven. Miti. Eng.* **31** 196 (in Chinese) [刘建刚, 王金红, 谭徐明, 马建明, 张念强 2011 防灾减灾工程学报 **31** 196]
- [3] Yin H, Li Y H 2013 *J. Arid Meteor.* **31** 182 (in Chinese) [尹晗, 李耀辉 2013 干旱气象 **31** 182]
- [4] Huang R H, Liu Y, Wang L, Wang L 2012 *Chin. J. Atmos. Sci.* **36** 443 (in Chinese) [黄荣辉, 刘永, 王林, 王磊 2012 大气科学 **36** 443]
- [5] Huang J P, Wang S W 1992 *Sci. China Chemistry* **35** 207
- [6] Li J P, Ding R Q 2008 *Chin. J. Atmos. Sci.* **32** 975 (in Chinese) [李建平, 丁瑞强 2008 大气科学 **32** 975]
- [7] Huang J P, Yi Y H, Wang S W, Chou J F 1993 *Q. J. R. Meteorol. Soc.* **119** 547
- [8] Li J P, Ding R Q 2009 *Chin. J. Atmos. Sci.* **33** 551 (in Chinese) [李建平, 丁瑞强 2009 大气科学 **33** 551]
- [9] Jia X L, Chen L J, Gao H, Wang Y G, Ke Z J, Liu C Z, Song W L, Wu T W, Feng G L, Zhao Z G, Li W J 2013 *J. Appl. Meteor. Science* **24** 641 (in Chinese) [贾小龙, 陈丽娟, 高辉, 王永光, 柯宗建, 刘长征, 宋文玲, 吴统文, 封国林, 赵振国, 李维京 2013 应用气象学报 **24** 641]
- [10] Feng G L, Dai X G, Wang A H, Chou J F 2001 *Acta Phys. Sin.* **50** 606 (in Chinese) [封国林, 戴新刚, 王爱慧, 丑纪范 2001 物理学报 **50** 606]
- [11] Zhao J H, Feng G L, Wang Q G, Yang J, Gong Z Q 2011 *Chin. J. Atmos. Sci.* **35** 1069 (in Chinese) [赵俊虎, 封国林, 王启光, 杨杰, 龚志强 2011 大气科学 **35** 1069]
- [12] Zhao J H, Zhi R, Shen Q, Yang J, Feng G L 2014 *Chin. J. Atmos. Sci.* **38** 237 (in Chinese) [赵俊虎, 支蓉, 申茜, 杨杰, 封国林 2014 大气科学 **38** 237]
- [13] Sklar A 1959 *Fonctions de Répartition à Dimensions et Leurs Marges* (France: Publications de l'Institut de Statistique de L'Université de Paris) pp229–231
- [14] Wan Z W Z, Abdul A J, Kamarulzaman I 2013 *Theor. Appl. Climatol.* **111** 559
- [15] Shiau J T, Modarres R 2009 *Meteorol. Appl.* **16** 481
- [16] Zhang Q, Xiao M Z, Singh V P, Chen X H 2013 *Theor. Appl. Climatol.* **111** 119
- [17] Xu L Y, Wang H M, Chen J F 2013 *J. Appl. Statist and Manage* **32** 284 (in Chinese) [许玲燕, 王慧敏, 陈军飞 2013 数理统计与管理 **32** 284]
- [18] Liu C L, Qiang Z, Vijay P S, Cui Y 2011 *Nat. Hazards* **59** 1533
- [19] Ummul F A R, Panlop Z 2014 *Theor. Appl. Climatol.* **115** 153
- [20] Rasoul M, Ahmad F F, Yagob D 2012 *Theor. Appl. Climatol.* **108** 191
- [21] Kao S C, Rao S G 2010 *J. Hydrol.* **380** 121
- [22] Li J, Li Y, Song S B, Cui C F 2012 *J. China Hydrol.* **32** 43 (in Chinese) [李计, 李毅, 宋松柏, 崔晨风 2012 水文 **32** 43]
- [23] Shiau J T, Reza M, Saralees N 2012 *Environ. Model. Assess.* **17** 469
- [24] Shiau J T 2006 *Water Resour. Manage* **20** 795
- [25] Bahram S, Hossein M 2014 *Nat. Hazards* **72** 1391
- [26] Ganguli P, Reddy M J 2012 *Water Resour. Manage* **26** 3301
- [27] Hao Z C, Amir A 2013 *Adv. Water Resour.* **57** 12
- [28] Herbst P H, Bredenkamp D B, Barker H M G 1966 *J. Hydrol.* **4** 264
- [29] Zuo D D, Hou W, Yan P C, Feng T C 2014 *Acta Phys. Sin.* **63** 230204 (in Chinese) [左冬冬, 侯威, 颜鹏程, 封泰晨 2014 物理学报 **63** 230204]
- [30] Zelenhasic E, Salvai A 1987 *Water Resour. Res.* **23** 156
- [31] Shiau J T, Hsieh W S 2001 *J. Water Resour. Plann. Manage.* **127** 30
- [32] Yan B W, Guo S L, Xiao Y, Fang B 2007 *Arid Zone Res.* **24** 537 (in Chinese) [闫宝伟, 郭生练, 肖义, 方彬 2007 干旱区研究 **24** 537]
- [33] Zhang L, Singh V P 2006 *J. Hydrol. Eng.* **11** 150
- [34] Wei F Y 2007 *Modern Climatic Statistical Diagnosis and Prediction Technology* (Beijing: China Meteorological Press) pp63–64 (in Chinese) [魏凤英 2007 现代气候统计诊断与预测技术 (北京: 气象出版社) 第63—64页]
- [35] Wei F Y 2007 *Modern Climatic Statistical Diagnosis and Prediction Technology* (Beijing: Meteorological Press) p26 (in Chinese) [魏凤英 2007 现代气候统计诊断与预测技术 (北京: 气象出版社) 第26页]
- [36] Liu Y, Zhao E X, Huang W, Zhou J Q, Ju J H 2010 *J. Catastrophol.* **25** 39 (in Chinese) [刘瑜, 赵尔旭, 黄玮, 周建琴, 琚建华 2010 灾害学 **25** 39]
- [37] Ding Y H, Zhang L 2008 *Chin. J. Atmos. Sci.* **32** 794 (in Chinese) [丁一汇, 张莉 2008 大气科学 **32** 794]
- [38] Xiao D, Li J P 2007 *Chin. J. Atmos. Sci.* **31** 839 (in Chinese) [肖栋, 李建平 2007 大气科学 **31** 839]
- [39] Wu H, Hou W, Yan P C 2013 *Acta Phys. Sin.* **62** 039206 (in Chinese) [吴浩, 侯威, 颜鹏程 2013 物理学报 **62** 039206]

# Sensitivity analysis of sample number on the drought descriptive model built by Copula function in southwest China\*

Zuo Dong-Dong<sup>1)</sup> Hou Wei<sup>2)†</sup> Wang Wen-Xiang<sup>1)</sup>

1) (*Department of Physics, Yangzhou University, Yangzhou 225002, China*)

2) (*Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing 100081, China*)

( Received 20 October 2014; revised manuscript received 3 December 2014 )

## Abstract

Based on the standardized precipitation index data of 89 meteorological stations in southwest China (Sichuan Province, Yunnan Province, Guizhou Province, Chongqing) during 1961–2010, probability model containing drought duration and drought severity is established by using the theory of run and the Copula function. The influences of the drought sample number on the distribution parameters, the probability and drought return period are discussed. The result shows that the stability of distribution parameters needs larger sample number. The sample number is greater than 50 in some regions and the requirements for sample number of each parameter is not consistent. The sample number of severity distribution parameters is largest. The probability and return period obtained in the case where the sample number is about 10 have no significant difference (the significant level is 0.05) from those in the case where the sample number is 40 in most of region. With the results used as the standard, statistical model can greatly reduce the requirements for the sample number. And then it demonstrates that the distribution function of drought duration and drought severity can still be established in the lack of measurement data and the inconsistency between starting and ending time. Climate warming has no influence on the minimum of sample number. The fluctuation is mostly between  $-5$  to  $5$ . Statistical model has a certain stability. Meanwhile, the division of climate state reduces the need for distribution test sample number and makes it easier to build model.

**Keywords:** drought in southwest China, theory of run, Copula function, sample size

**PACS:** 02.50.-r, 92.40.De

**DOI:** [10.7498/aps.64.100203](https://doi.org/10.7498/aps.64.100203)

---

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 41305056, 41175067) and the National Basic Research Program of China (Grant No. 2012CB955901).

† Corresponding author. E-mail: [houwei@cma.gov.cn](mailto:houwei@cma.gov.cn)