

文章编号: 1007-7588(2011)12-2308-06

基于 Copula 的区域干旱空间分布特征分析

徐春晓¹, 袁潇晨², 金菊良², 郦建强³

(1. 南京水利科学研究院, 南京 210029; 2. 合肥工业大学 土木与水利工程学院, 合肥 230009;

3. 水利部水利水电规划设计总院, 北京 100120)

摘要: 以邵阳市 15 个雨量站逐月降雨资料为基础, 采用适线法估计干旱历时和干旱烈度的分布函数, 利用 3 种 Archimedean Copula 函数分别构建了两干旱特征变量之间的联合分布, 并比较了 3 种 Copula 函数的拟合效果, 最后分析了邵阳市干旱重现期的空间分布特征。结果表明, Copula 理论能够为多变量干旱特征分析提供有效的研究途径, Archimedean Copula 函数中的 Gumbel-Hougaard Copula 和 Frank Copula 函数较好地描述各站干旱特征变量之间的相互关系, 适线法避免了基于数据估计分布函数参数的不合理性, 使基于 Copula 函数的频率分析结果更客观可靠, 区域干旱具有显著的空间分布特性, 各站干旱程度差异较大。

关键词: 干旱频率分析; Copula 函数; 适线; 空间分布; 干旱灾害风险管理

1 引言

区域干旱已成为全球气候变化背景下典型的自然现象, 其开始发生过程一般无明显先兆, 且影响范围广, 给经济社会造成了巨大的损失^[1-2]。分析研究区域干旱的时空特征, 有助于加强对干旱的预防和风险管理, 促进水资源系统的合理分配和科学建设, 是干旱灾害风险管理研究的重要基础性工作之一^[1-5]。降水量是能够有效地反映区域干旱程度的指标^[6], 通过它得到的干旱历时和干旱烈度是常用来刻画干旱的特征变量, 由此较完整地实现了对整个干旱过程的描述^[7,8]。然而, 以往的一些研究仅针对单个干旱特征变量进行分析, 使得其结果带有不完整性, 因此, 需要结合多个特征变量来综合分析区域干旱的空间分布特征, 以保证研究的科学客观性。Copula 理论为多变量的区域干旱特征分析提供了途径, 利用它可建立多个具有任意分布变量之间的联合分布, 克服了对变量分布硬性假设的束缚^[9]。目前, 不同种类的 Copula 函数, 以及多种变量的组合已在区域干旱研究中有所应用, 并取得了一定的进展^[8,10-14]。而成功运用 Copula 函数进行干旱频率分析的前提是需要正确建立单个特征变量的

分布, 当前已有的研究都是基于数据直接对分布函数的参数进行估计, 这种方式忽略了计算结果中存在的合理性, 有时会造成与实际情况相差甚远。基于此, 本文根据邵阳市 15 个雨量站的逐月降水资料, 采用适线法确定干旱历时与干旱烈度的分布曲线, 利用 Gumbel-Hougaard (GH) Copula、Clayton Copula 和 Frank Copula 函数分别构建干旱历时与干旱烈度之间的联合分布, 并比较不同 Copula 函数的拟合效果。通过估计典型干旱年份各站的干旱重现期, 以及在平均干旱水平下各站的干旱重现期, 深入分析区域干旱的空间分布特征, 从而为区域干旱灾害风险管理提供有效的科学依据。

2 研究方法

2.1 干旱的识别

根据游程理论并结合已有的研究成果^[12,15], 设定干旱指标(本文为月降水量)阈值 R_0 、 R_1 和 R_2 (图 1), 当指标值小于 R_1 时, 则初步判断此月为干旱, 有 a 、 b 、 c 和 d 共 4 个干旱过程; 在此基础上, 对于历时只有 1 个单位时段的干旱(如 a 、 d), 若其干旱指标值小于 R_2 (如 a), 则此月最终被确定为 1 次干旱过程, 反之不计为干旱(如 d); 而对于间隔为 1 个单位

收稿日期: 2011-08-17; 修订日期: 2011-09-29

基金项目: 中国气象局成都高原气象开放实验室基金课题(编号: LPM2011002); 水利部公益性行业科研专项经费项目(编号: 201001043); 国家自然科学基金项目(编号: 51079037); 水利部重大基建前期项目: “全国干旱区划及早灾风险评估研究”。

作者简介: 徐春晓, 男, 安徽黄山市人, 硕士, 高级工程师, 研究方向为水资源规划与管理。E-mail: cxxu@nhri.cn

通讯作者: 金菊良, E-mail: jinjl66@126.com

2011年12月

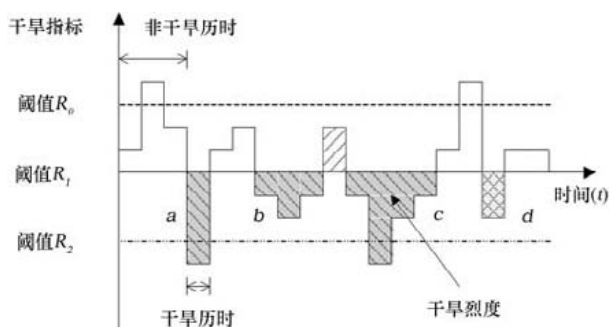


图1 干旱的识别及干旱特征变量的确定

Fig.1 Definitions of drought events and drought characteristic variables

时段的两次相邻干旱过程(如 b, c),若间隔期的干旱指标值小于 R_0 ,则这两次相邻干旱可被视为1次干旱过程,否则为2次独立干旱过程,合并后的干旱历时 $D = d_b + d_c + 1$,烈度 $S = s_b + s_c$ 。因此,按上述规定可得图1中共有2次干旱过程,即 a 和 $b+c$ 。当以降水量为干旱指标时,本文如下确定干旱划定过程中的阈值。

降水量距平百分率反映的是一段时期内的降水量与常年同期平均降水量之间的关系,表示为^[16]:

$$P_a = (P - \bar{P}) / \bar{P} \times 100\% \quad (1)$$

式中 P 是某一段时期内降水量; \bar{P} 是常年同期平均降水量。根据《气象干旱等级》标准^[16],本文选用单月距平百分率为 -30% 和 -40% 时所对应的降雨量分别作为阈值 R_1, R_2 ,而 R_0 为月同期平均降水量。

2.2 干旱特征变量分布

已有研究表明,干旱历时 D 和干旱烈度 S 一般可分别用指数分布和Gamma分布描述^[8,10,17],其概率分布函数分别为:

$$F_D(d) = 1 - e^{-d/\lambda} \quad (2)$$

$$F_S(s) = \int_0^s \frac{s^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-s/\beta} ds \quad (3)$$

式中 λ, α, β 为分布参数。当直接用样本数据估计各参数来确定特征变量的分布时,往往会出现不符合实际的情况,造成计算结果的不合理性。而适线法是在充分考虑结果合理性的同时,根据分布曲线和样本经验点的拟合程度来优选参数的方法。因而,需要在基于样本数据估计的基础上,通过适线法调整并最终确定干旱特征变量分布函数的各参数,从而保证结果的合理性。在此过程中,由于指数分布是Gamma分布的特例,在适线时也可将干旱历时按Gamma分布进行适线调整处理。

2.3 Copula理论

当选用干旱历时和干旱烈度这两个干旱特征变量来共同描述干旱事件、分析干旱频率时,则需要度量两者之间的联系,即计算它们的联合概率分布函数。Copula理论已成为当今实现这种相关性分析的有效、通用的方法^[9],其中Archimedean Copula函数因形式简单而被广泛应用,最常用的Archimedean Copula函数有Gumbel-Hougaard、Clayton和Frank Copula。令 $u = F_D(d), v = F_S(s)$,则它们可分别表示为^[8,9]:

$$F_{D,S}(d,s) = C(u,v) = \exp\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\} \quad (4)$$

$$F_{D,S}(d,s) = C(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \quad (5)$$

$$F_{D,S}(d,s) = C(u,v) = -\frac{1}{\theta} \ln\left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)}\right] \quad (6)$$

式中 θ 为参数,它与Kendall相关系数 τ 之间的关系^[8,9]如表1所示。Kendall相关系数 τ 定义为:

$$\tau = (C_n^2)^{-1} \sum_{i < j} \text{sgn}[(d_i - d_j)(s_i - s_j)] \quad (7)$$

$(i, j = 1, 2, \dots, n)$

式中 (d_i, s_i) 是干旱历时、干旱烈度联合观测值。 $\text{sgn}(x)$ 是符号函数,当 $x > 0$ 时, $\text{sgn}(x) = 1$; $x < 0$ 时, $\text{sgn}(x) = -1$; $x = 0$ 时, $\text{sgn}(x) = 0$ 。

由公式(4)~公式(6)可得干旱历时与干旱烈度之间的联合分布概率值。评定各种Copula函数拟合结果优劣的指标值采用均方根误差(RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [P_c(i) - P_o(i)]^2} \quad (8)$$

式中 $P_c(i)$ 是由Copula函数计算得的第 i 个联合观测值 (d_i, s_i) 的理论联合分布概率; n 为联合观测值样本数; $P_o(i)$ 则是经验联合分布概率值,依据下式计算^[17]:

$$P_o(i) = (m_i - 0.44) / (n + 0.12) \quad (9)$$

式中 m_i 表示联合观测值样本中满足条件 $D \leq d_i$ 且 $S \leq s_i$ 的联合观测值的个数。

2.4 干旱重现期

由以上Copula函数的联合分布可得,当干旱历

表1 Copula函数参数 θ 与Kendall相关系数 τ 的关系Table 1 The relationship between θ and τ of Copulas

Copula 函数	θ 与 τ 的关系
GH	$\tau = 1 - 1/\theta$ ($\theta \geq 1$)
Clayton	$\tau = \theta/(\theta + 2)$ ($\theta > 0$)
Frank	$\tau = 1 + \frac{4}{\theta} \left(\frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt - 1 \right)$ ($\theta \neq 0$)

时 $D > d$ 、烈度 $S > s$ 时的干旱重现期为^[18]:

$$T_a(d,s) = \frac{E(L)}{P(D > d \cap S > s)} \tag{10}$$

$$= \frac{E(L)}{1 - F_D(d) - F_S(s) + F_{D,S}(d,s)}$$

式中 $E(L)$ 是干旱间隔期望, 为干旱历时与非干旱历时的平均值之和。

3 实例分析

邵阳地区是中国南方山地丘陵地区, 属于湖南省“衡邵干旱走廊”的一部分, 由于邵阳地区降雨时空分布不均, 干旱灾害频繁。现根据1956年-2002年邵阳市15个雨量站的逐月降水资料, 研究此区域内干旱重现期的空间分布特征并进行分析。选用本文2.1中的干旱识别方法, 确定各站发生干旱的次数及干旱特征变量, 其统计结果如表2所示。

由表2可知:

(1) 各站干旱历时、干旱烈度均值及其Kendall相关系数较为接近, 说明各站的干旱平均水平大体相当;

(2) 党坪站干旱次数最多, 而邵东站干旱次数最少, 两者相差19次;

(3) 各站的历时、烈度最大值有较大差异, 茅坪站、白茅坪站的历时最大值仅为5个月, 而武冈站的历时最大值长达10个月, 新宁站的烈度最大值为276.24 mm, 而白茅坪站的烈度最大值仅为132.67 mm。

可见, 各站的干旱特性具有显著

差异性, 研究它们之间的关系对区域干旱频率分析具有重要意义。

通过上述识别过程, 根据干旱历时、干旱烈度分别服从指数分布、Gamma分布的假设, 采用适线法确定公式(2)、公式(3)中各参数, 由此得干旱特征变量的理论分布。图2为邵东站干旱特征变量分布曲线在适线前后与经验频率点的对比情况。从图2中可以看出:

(1) 适线调整后的干旱特征变量分布曲线与经验频率点接近, 说明适线法是建立在客观数据的基础

表2 各站干旱统计

Table 2 Drought statistics of the stations

站点	干旱次数	干旱间隔均值(月)	干旱历时均值(月)	干旱烈度均值(mm)	历时最大值(月)	烈度最大值(mm)	干旱历时与烈度的Kendall相关系数
黄桥	91	6.20	2.01	48.02	8	244.01	0.54
高沙	87	6.43	1.94	51.27	8	268.72	0.45
邵东	78	7.17	2.09	51.85	8	185.27	0.45
诸甲亭	80	6.99	2.10	48.87	8	189.84	0.53
茅坪	93	6.02	1.75	48.13	5	195.58	0.46
罗家庙	90	6.22	1.90	47.49	9	219.33	0.54
白毛坪	86	6.56	1.81	47.35	5	132.67	0.41
武冈	86	6.50	1.85	49.01	10	208.28	0.49
新宁	87	6.43	1.83	53.03	9	276.24	0.45
小沙江	80	6.99	1.90	53.47	7	209.36	0.49
大河滩	88	6.36	1.89	51.96	9	186.05	0.48
回龙寺	91	6.15	1.78	45.10	9	216.82	0.47
六都寨	90	6.22	1.92	51.92	6	216.41	0.49
党坪	97	5.77	1.78	47.05	6	180.47	0.42
三门江	83	6.74	2.07	47.95	9	212.66	0.45

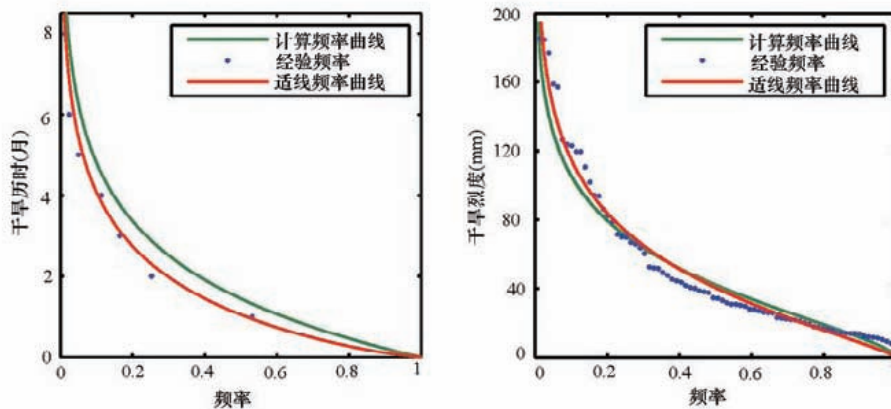


图2 邵东站干旱历时和干旱烈度的分布曲线

Fig. 2 The observed drought duration and drought severity with estimated and adjusted distribution curve in Shaodong station

2011年12月

础之上;

(2)基于数据估计的分布曲线侧重于对数据点的整体拟合,而没有能够很好地与实际情况相结合,造成个别结果不合理;

(3)适线调整后的分布曲线兼顾了所有数据的情况,尤其对于长干旱历时、高干旱烈度的干旱事件的经验频率点能较好拟合,这为区域干旱频率分析提供了可靠的基础。

由干旱历时与烈度间的Kendall相关系数,按公式(4)–公式(6)分别采用GH Copula、Clayton Copula和Frank Copula函数求解干旱历时与烈度之间的联合分布概率,将其与经验联合分布概率相比较,按公式(8)计算各Copula函数的拟合误差(表3)。由表3可知:

(1)3种Copula函数都能较好地反映干旱历时与干旱烈度之间的相关关系,其中的GH Copula及Frank Copula函数拟合效果较优,为区域干旱频率分析研究提供了有效的途径;

(2)适线后各站的Copula函数拟合程度均有显著提升,其中,白毛坪站的3种Copula函数拟合误差平均下降60.64%。

可见,适线法在合理确定干旱特征变量分布曲线的同时,能够有效地提高Copula函数的准确性。

由于干旱频率分析的主要对象是长历时、高烈度的干旱事件,Gumbel–Hougaard(GH) Copula函数能较好地描述这种具有分布上尾相关特性的相关关系,因此,选用它按公式(10)对典型干旱年份(1963年、1998年)中各雨量站的干旱重现期进行计

表3 适线前后的Copula函数拟合误差

Table 3 The fitting errors of copulas before and after curve fitting

站名	GH Copula		Clayton Copula		Frank Copula	
	适线前	适线后	适线前	适线后	适线前	适线后
黄桥	0.0797	0.0701	0.0893	0.0837	0.0787	0.0759
高沙	0.0846	0.0757	0.0924	0.0914	0.0835	0.0828
邵东	0.0698	0.0422	0.0756	0.0563	0.0681	0.0472
诸甲亭	0.0764	0.0527	0.0780	0.0614	0.0726	0.0530
茅坪	0.0760	0.0521	0.0834	0.0642	0.0747	0.0557
罗家庙	0.1002	0.0662	0.1030	0.0781	0.0965	0.0701
白毛坪	0.0535	0.0182	0.0652	0.0276	0.0563	0.0235
武冈	0.0815	0.0579	0.0880	0.0708	0.0796	0.0621
新宁	0.0893	0.0867	0.0991	0.0938	0.0891	0.0923
小沙江	0.0624	0.0432	0.0716	0.0553	0.0620	0.0497
大河滩	0.0555	0.0294	0.0655	0.0374	0.0568	0.0356
回龙寺	0.1015	0.0650	0.1045	0.0766	0.0976	0.0692
六都寨	0.0606	0.0433	0.0676	0.0550	0.0597	0.0499
党坪	0.0636	0.0471	0.0755	0.0541	0.0656	0.0520
三门江	0.0732	0.0586	0.0845	0.0706	0.0743	0.0653

算,结果如图3所示。从图3中可以看出:

(1)1963年、1998年邵阳市受旱严重,个别站点的干旱重现期超过了80 a;

(2)对于同一年份,区域内各站干旱程度差异显著。1963年,干旱严重区主要集中在邵阳市中部,西部和东部的干旱程度则相对较轻。而对于1998年,邵阳市大部分地区干旱严重,南部各站的干旱重现期普遍高于北部的小沙江、六都寨和大河滩站。同时,为了更好地反映区域内各站在平均干

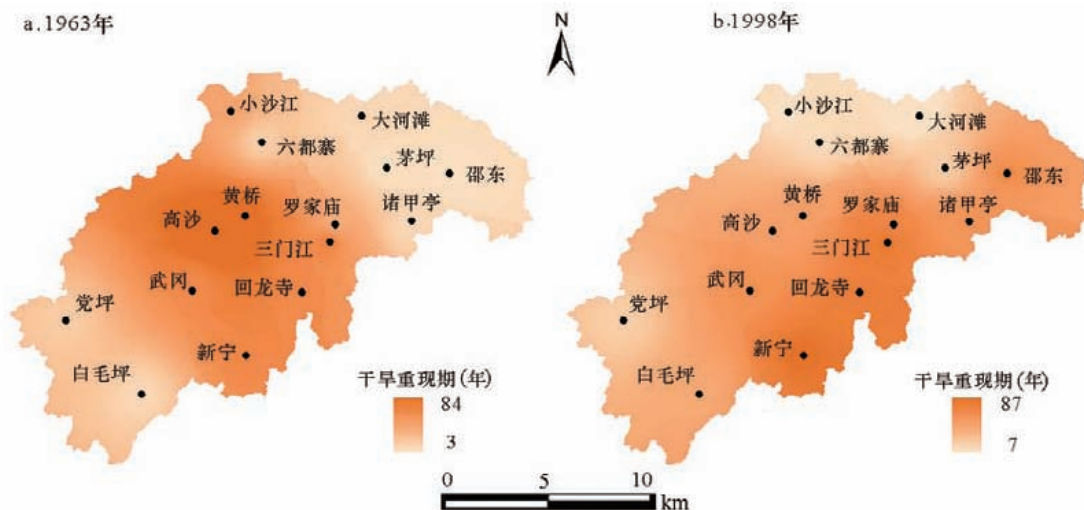


图3 1963年和1998年邵阳市干旱重现期空间分布特征

Fig.3 The spatial distribution characteristic of drought return periods of Shaoyang City in 1963 and 1998

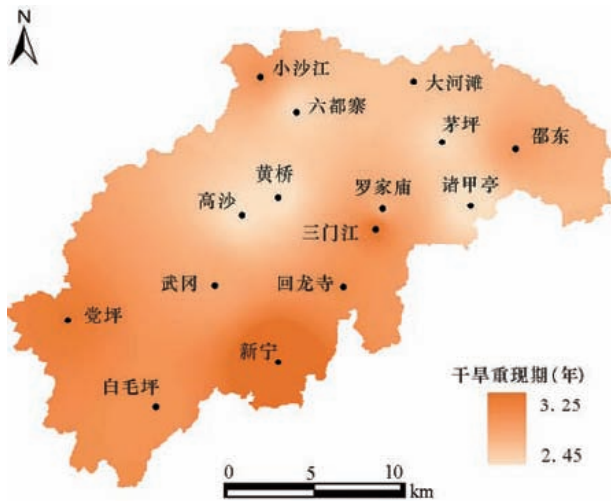


图4 平均干旱水平下的邵阳市干旱重现期空间分布特征
Fig. 4 The spatial distribution characteristic of drought return period of Shaoyang City with average drought extent

旱水平下的干旱程度,由各站的干旱历时均值、干旱烈度均值(表2)得干旱重现期(图4),结果表明邵阳南部的平均干旱程度相对较重,而中部和东部相对较轻。可见,干旱空间特征分析能够为区域干旱灾害风险管理提供有效的科学依据。

4 结论

以干旱历时、干旱烈度为干旱特征变量,根据Copula理论建立了两者的联合分布及干旱重现期,通过对邵阳市15个雨量站的逐月降水资料进行分析,探讨了此区域内干旱重现期的空间分布状况。研究结果表明:

(1) Copula理论为多变量干旱特征分析提供了有效的研究途径, Gumbel- Hougaard、Frank Copula函数对各站干旱特征变量间的联合分布具有较好的拟合结果;

(2) 适线法避免了基于数据估计分布函数参数的不合理性,使基于Copula函数的频率分析结果更客观可靠;

(3) 区域干旱具有显著的空间分布特性,各站干旱重现期差异程度较大,通过对其的深入分析能够为区域干旱灾害风险管理提供科学依据。

参考文献 (References):

- [1] Wilhite D A. Drought planning: A process for state government[J]. *Water Resource Bulletin*, 1991, 27(1): 29-38.
- [2] 张利平,曾思栋,王仁超,等. 气候变化对滦河流域水文循环的影响及模拟[J]. *资源科学*, 2011, 33(5): 966-974.
- [3] 翁白莎,严登华. 变化环境下中国干旱综合应对措施探讨[J]. *资源科学*, 2010, 32(2): 309-316.
- [4] 金菊良,魏一鸣. 复杂系统广义智能评价方法与应用[M]. 北京: 科学出版社, 2008.
- [5] 曹永强,李香云,马静,等. 基于可变模糊算法的大连市农业干旱风险评价[J]. *资源科学*, 2011, 33(5): 983-988.
- [6] 赵林,武建军,吕爱锋,等. 黄淮海平原及其附近地区干旱时空动态格局分析—基于标准化降水指数[J]. *资源科学*, 2011, 33(3): 468-476.
- [7] González J, Valdés J B. Bivariate drought recurrence analysis using tree ring reconstructions[J]. *Journal of Hydrological Engineering*, 2003, 8(5): 247-258.
- [8] Shiau J T. Fitting drought duration and severity with two-dimensional copulas[J]. *Water Resources Management*, 2006, 20(5): 795-815.
- [9] Nelson R B. An Introduction to Copulas[M]. New York: Springer, 2006.
- [10] Song S B, Singh V P. Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data[J]. *Stochastic Environmental Research and Risk Assessment*, 2009, 24(3): 425-444.
- [11] 闫宝伟,郭生练,肖义,等. 基于两变量联合分布的干旱特征分析[J]. *干旱区研究*, 2007, 24(4): 537-542.
- [12] 陆桂华,闫桂霞,吴志勇,等. 基于copula函数的区域干旱分析方法[J]. *水科学进展*, 2010, 21(2): 188-193.
- [13] 马明卫,宋松柏. 椭圆型Copulas函数在西安站干旱特征分析中的应用[J]. *水文*, 2010, 30(4): 36-42.
- [14] 许月萍,张庆庆,楼章华,等. 基于Copula方法的干旱历时和烈度的联合概率分析[J]. *天津大学学报*, 2010, 43(10): 928-932.
- [15] Fleig A K, Tallaksen L M, Hisdal H, et al. A global evaluation of streamflow drought characteristics[J]. *Hydrology and Earth System Sciences*, 2006, 10(4): 535-552.
- [16] GB/T 20481-2006, 气象干旱等级[S].
- [17] Zhang L, Singh V P. Bivariate rainfall frequency distributions using Archimedean copulas[J]. *Journal of Hydrology*, 2007, 332(1-2): 93-109.
- [18] Shiau J T. Return period of bivariate distributed hydrological events[J]. *Stochastic Environmental Research and Risk Assessment*, 2003, 17(1-2): 42-57.

Analysis of Characteristics of Spatial Distribution of Drought Based on Copula

XU Chunxiao¹, YUAN Xiaochen², JIN Juliang², LI Jianqiang³

(1. *Nanjing Hydraulic Research Institute, Nanjing 210029, China;*

2. *School of Civil Engineering, Hefei University of Technology, Hefei 230009, China;*

3. *Water Resources and Hydropower Planning and Design General Institute, Ministry of Water Resources, Beijing 100120, China)*

Abstract: In this paper, monthly rainfall data at 15 stations in Shaoyang City were used to analyze the characteristics of spatial distribution of regional drought. A drought event is defined as the monthly rainfall below the drought thresholds according to the classification of meteorological drought in China, and the drought event is described by two characteristic variables, i.e., drought duration and drought severity. Here, drought duration was fitted with the exponential distribution, while the gamma distribution was used for drought severity. The distribution function of each drought characteristic variable was determined by the curve fitting method instead of maximum likelihood estimation so as to avoid possible discrepancies. Furthermore, three kinds of Archimedean copulas, i.e., Gumbel-Hougaard copula, Clayton copula, as well as Frank copula, were employed to construct the joint distribution of drought characteristic variables. Then, the copulas' fitting results were compared, from which the copula of the best fitting results for the estimation of drought return period was selected. Finally, characteristics of the spatial distribution of drought return periods of Shaoyang City in typical drought years were analyzed. Empirical results show that copula theory provides an alternative for multivariable analysis of drought characteristic. The Gumbel-Hougaard copula as well as Frank copula can well reflect the relationships among drought characteristic variables, both of which have similar fitting results. The curve fitting method can avoid the ineffectiveness from estimation of parameters based on data, making the results of copula-based frequency analysis more reliable, with the fitting errors of copulas using the curve fitting method much smaller than those using maximum likelihood estimation. At the Baimaoping station, the average decreasing amplitude of 3 copulas' fitting errors employed in this paper reached up to 60.64%. Regional drought has a remarkable characteristic of spatial distribution, resulting in differences in drought extent among stations, e.g., the Dangping station has the highest frequency of drought events, while the Shaodong station shows the least frequency. The maximum drought durations at the Maoping station as well as the Baimaoping station are only 5 months, while become 10 months at the Wugang station. However, the maximum drought severity at the Baimaoping station is only 132.67 mm, which is much smaller than that of 276.24 mm at the Xinning station. In 1963, the severest drought region was in the middle part of Shaoyang City, but the south was severer in 1998.

Key words: Drought frequency analysis; Copula function; Curve fitting; Spatial distribution; Risk management of drought disaster