

基于 Copula 函数的河川径流丰枯遭遇分析

莫淑红, 沈 冰, 张晓伟, 李占斌

(西安理工大学 西北水资源与环境生态教育部重点实验室, 陕西 西安 710048)

【摘 要】 【目的】研究两变量联合概率的分布问题,探讨干支流的丰枯遭遇分析方法。【方法】采用 Copula 函数,构建了渭河干流及其支流千河年径流的联合概率分布模型,对干支流的丰枯遭遇进行分析,并绘制了干支流联合分布等值线图。【结果】宝鸡市渭河干流年径流量的边际分布服从皮尔逊Ⅲ型分布,而支流千河年径流量的边际分布服从对数正态分布,且干支流同时出现丰水年的概率大于同时出现平、枯水年的概率。【结论】用 Copula 法描述干支流的丰枯遭遇问题具有可行性。

【关键词】 河川径流;丰枯遭遇;Copula 函数

【中图分类号】 TV121⁺.4

【文献标识码】 A

【文章编号】 1671-9387(2009)06-0131-06

Synchronous-Asynchronous encounter probability of rich-poor runoff based on copula function

MO Shu-hong, SHEN Bing, ZHANG Xiao-wei, LI Zhan-bin

(Northwest Water Resources and Environment Ecology Key Lab of Ministry of Education, Xi'an University of Technocogy, Xi'an, Shaanxi 710048, China)

Abstract: 【Objective】 The research studied the joint probability distribution with two variables and analyzed the synchronous-asynchronous encounter probability of rich-poor runoff of mainstream and tributary. 【Method】 The copula theory was applied to construct the joint distribution of annual runoff in the Wei River and the Qian River, the biggest branch of Wei River in Baoji City. The synchronous-asynchronous encounter probability of rich-poor runoff was analyzed and the contour map of the joint distribution was obtained. 【Result】 The marginal distribution of the annual runoff in the Wei River obeys P-Ⅲ distribution, while that in the Qian River belongs to lognormal distribution, and the synchronous encounter probability of rich runoff in both rivers is greater than the probability of poor and normal runoff. 【Conclusion】 It is feasible to apply Copula function to describe the synchronous-asynchronous encounter probability of rich-poor runoff in mainstream and tributary.

Key words: river runoff; synchronous-asynchronous encounter probability; Copula function

渭河宝鸡市区段生态治理工程,位于陕西省宝鸡市市区中心的渭河主河道上,该工程在保障防洪安全的前提下,综合开发利用河道资源潜力,营造城市景观,为西北地区生态城市建设提供了新的模式。该生态治理工程的运行效益,与宝鸡峡水库、各支流

工程的联合调度及水源有效补给息息相关。

干支流的丰枯遭遇分析,属于两变量联合概率分布问题。国内、外学者提出了许多具有代表性的两变量联合概率分布模型,但这些模型都是基于随机变量之间线性相关性而建立的^[1]。事实上,水文

* [收稿日期] 2008-06-25

[基金项目] 国家重点基础研究发展计划项目(2007CB714106);国家自然科学基金项目(50579063);陕西省自然科学基金研究计划项目(2007D02)

[作者简介] 莫淑红(1972—),女,宁夏中卫人,博士,主要从事旱区水文水资源研究。E-mail:moshuhong@xaut.edu.cn

[通信作者] 沈 冰(1948—),男,浙江湖州人,教授,博士生导师,主要从事旱区水文水资源研究。E-mail:shenbing@xaut.edu.cn

科学领域的各种随机变量之间往往呈现出复杂的线性、非线性相关关系,基于线性相关的两变量概率模型,难以准确地描述非线性关系变量的联合概率分布特征。20 世纪 90 年代后期,Copula 理论引起了人们的重视,并在财经、保险、金融风险等领域得到了广泛的应用。由于 Copula 函数对于单个随机变量的边际分布没有限制,具有构造灵活的多变量联合分布优良特性^[2],近年来在水科学领域的应用逐渐引起国内外学者的关注^[3-6]。

本研究采用 Copula 函数,构造渭河干流及其宝鸡市区段最大一级支流千河年径流的联合分布,并具体分析干支流的丰枯遭遇频率,以期为宝鸡市生态城市建设、经济可持续发展提供参考。

1 Copula 函数

1.1 Copula 函数定义和性质

Copula 函数是边缘分布为 $[0,1]$ 区间均匀分布的联合分布函数,Sklar’s 定理给出了 Copula 函数和两变量联合分布的关系。设 X,Y 为连续的随机变量,其边缘分布函数分别为 F_X 和 F_Y , $F(x,y)$ 为变量 X 和 Y 的联合分布函数,那么存在唯一的 Copula 函数 $C(u_1,u_2)$,使得:

$$F(x,y)=C_{\theta}(F_X(x),F_Y(y)),\forall x,y。(1)$$

式中: $C(u_1,u_2)$ 为 Copula 函数; θ 为待定参数。

从 Sklar’s 定理可以看出,Copula 函数能独立于随机变量的边缘分布,反映随机变量的相关性结构,从而可将二元联合分布分为 2 个独立的部分,即变量间的相关性结构和变量的边缘分布来分别进行处理,其中变量间的相关性结构用 Copula 函数来描述。

1.2 Copula 参数的估计及联结函数的选择

Copula 函数总体上可以分为椭圆型、阿基米德型和二次型 3 类,其中生成元为 1 个参数的阿基米

德型 Copula 函数的应用最为广泛,其构造形式为:
$$C_{\theta}(u_1,u_2)=\Phi^{-1}[\Phi(u_1)+\Phi(u_2)],0\leqslant u_1,u_2\leqslant 1。(2)$$

式中: θ 为参数; Φ 称为 Copula 函数的生成元。其中参数 θ 可以由实测数据的相关系数得出。

Copula 函数的选择可以通过点绘理论概率和经验频率的相关图,并依据一定的评价指标来确定相应的 Copula 作为联结函数。

2 联合分布模型的建立

本研究采用渭河干流林家村站和渭河支流千河千阳站 1964~2000 年同步实测年径流资料,进行干支流丰枯遭遇分析。

2.1 河川径流边际分布的确定

对于单变量水文数据系列的分析,在我国水文分析计算中,常假定水文变量服从皮尔逊Ⅲ型分布,而 Copula 函数方法不受变量边际分布型式的限制。因此,本研究选用皮尔逊Ⅲ型(P-Ⅲ)、Gumbel 分布(极值Ⅰ型)与对数正态分布 3 种分布型式进行分析比较,从中选取数据系列拟合效果最好的分布。

我国常采用以下数学期望公式计算经验频率。

$$H(x)=P(X\leqslant x_m)=\frac{m-0.44}{N+0.12}。(3)$$

式中: P 为 $X\leqslant x_m$ 的经验概率, m 为 x_m 的序号, N 为样本容量。

采用矩法分别对皮尔逊Ⅲ型、Gumbel 分布与对数正态分布进行参数估计。渭河干流及其支流千河年径流分布各模型的参数及拟合情况如表 1 和图 1 所示。从图 1 2 条河流年径流量的频率曲线拟合结果可以看出,各边际分布的拟合情况均较好,需要采用评价指标进行选取。

表 1 渭河干流及其支流千河年径流分布参数的估计

Table 1 Distribution parameters estimation of the annual runoff in the Wei River and the Qian River

水文系列 Hydrological time series	皮尔逊Ⅲ型分布 P-Ⅲ distribution			Gumbel 分布 Gumbel distribution		对数正态分布 Lognormal distribution	
	年径流量均值/($\times 10^8\text{ m}^3$) Mean value of annual runoff	C_v	C_s/C_v	C	α	E_y	δ_y
林家村站 Linjiacun hydrologic station	24.2	0.52	2.1	8.48	-19.3	3.08	0.49
千阳站 Qanyang hydrologic station	3.85	0.53	2.2	1.76	-2.83	1.18	0.6

注: C_v 为变差系数; C_s 为偏态系数; $C=\text{sqrt}(6)\times\pi\times\sigma$,其中 σ 为标准差, $\alpha=rC-x_{av},x_{av}$ 为均值, r 为欧拉常数; E_y 为数学期望; δ_y 为均方差, $y=\ln x$ 。

Note: C_v is variation coefficient, C_s is skewness coefficient, $C=\text{sqrt}(6)\times\pi\times\sigma$, σ is standard deviation, $\alpha=rC-x_{av},x_{av}$ is mean value, r euler constant, E_y is mathematical expectation, δ_y is mean square deviation, $y=\ln x$.

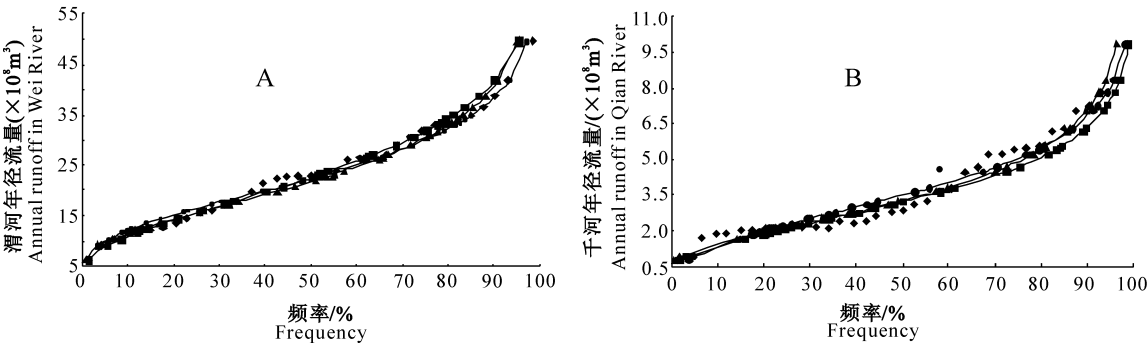


图 1 渭河干流(A)及其支流千河(B)年径流频率曲线的拟合结果

◆,经验频率;—●—,Gumbel 分布计算值;—▲—,对数正态分布计算值;—■—,P-Ⅲ分布计算值

Fig. 1 Annual runoff frequency curve of the Wei River (A) and the Qian River (B)

◆,Empirical frequency;—●—,Gumbel frequency by Gumbel;—▲—,Calculated frequency by lognormal;—■—,Calculated frequency by P-Ⅲ

2.2 评价指标

关于频率曲线的拟合优度,可采用均方根误差 $RSME$ 及 AIC 最小准则进行衡量。 $RSME$ 的计算公式为:

$$RSME = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N [x_c(i) - x_0(i)]^2} \quad (4)$$

式中: MSE 为均方差, N 为样本容量, x_c 为由模型计算的频率, x_0 为经验频率, i 为样本序号。

AIC 最小准则是由日本学者赤池弘次在研究信息论问题时提出的。其原理是若欲从一组可供选择的模型中选择一个最佳模型,应选择 AIC 值最小

的模型^[7]。 AIC 值的表达式为:

$$AIC = (-2) \lg(\text{极大似然函数}) + 2(\text{独立参数的个数}) \quad (5)$$

或者 $AIC = N \lg(MSE) + 2(\text{独立参数的个数})$, 若只估计 1 个参数, 则有:

$$AIC = (-2) \lg(\text{极大似然函数}) \quad (6)$$

各边际分布模型的检验结果见表 2。由表 2 可知, 根据 $RSME$ 、 AIC 最小准则, 渭河年径流量的边际分布选用皮尔逊Ⅲ型分布最为合适, 而千河年径流量的边际分布选用对数正态分布更为合适。同时可以认为, 由矩法估计的相应参数所得的结果是合理的。

表 2 渭河干流及其支流千河年径流分布模型拟合优度的检验结果

水文系列 Hydrological time series	边际分布模型 Model of marginal distribution	RMSE	AIC
林家村站年径流量 Annual runoff in linjiacun hydrologic station	皮尔逊Ⅲ型分布 P-Ⅲ distribution	2.70	31.9
	Gumbel 分布 Gumbel distribution	3.79	42.8
	对数正态分布 Lognormal distribution	3.91	43.8
千阳站年径流量 Annual runoff in qianyang hydrologic station	皮尔逊Ⅲ型分布 P-Ⅲ distribution	6.01	57.6
	Gumbel 分布 Gumbel distribution	6.35	59.4
	对数正态分布 Lognormal distribution	4.96	51.5

2.3 联合分布模型的建立

年径流的边际分布确定后, 需要对 Copula 函数参数进行估计。对于两变量单参数阿基米德 Copula 函数的参数估计, 常采用 Genest 等^[8]提出的非参数估计方法, 通过 Kendall 秩相关系数 τ 与 Copula 函数参数 θ 的关系进行估计, 两者的关系如下:

$$\tau = 1 + 4 \int_0^1 \frac{\Phi(t)}{\Phi'(t)} dt = f(\theta) \quad (7)$$

式中: $\Phi(t)$ 为两变量的 Copula 函数生成函数。两变量 Kendall 秩相关系数 τ 估计值计算公式如下^[9]:

$$\hat{\tau} = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \text{sign}[(x_i - x_j)(y_i - y_j)] \quad (8)$$

式中: (x_i, y_i) 为样本, n 为样本个数; $\text{sign}(x)$ 为符号

函数,
$$\text{sign}(x) = \begin{cases} 1 & x > 0, \\ 0 & x = 0, \\ -1 & x < 0. \end{cases}$$

计算可得, 渭河和千河年径流量之间的 Kendall 秩相关系数 τ 为 0.81。根据不同 Copula 函数对相关性的适用范围, 选取 Clayton 与 Gumbel-Hougaard 2 种常用的 Copula 函数进行比较分析, 从中选取拟合效果较好的 Copula 函数。计算所得的 Copula 函数参数见表 3。

表 3 Copula 函数参数的估计结果
Table 3 Estimation of copula functions

函数名称 Function caption	$C(u_1, u_2)$	τ 与 θ 的关系 Relationship of τ and θ	$\hat{\theta}$
Clayton	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$	$\frac{\theta}{\theta + 1}$	4.26
Gumbel-Hougaard	$\exp \{ - [(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{\frac{1}{\theta}} \}$	$1 - \theta^{-1}$	5.26

为检验 Copula 函数的拟合精度,需要比较理论累积频率与经验累积频率的一致性。两变量联合累积经验频率采用 Gringorton 经验频率公式计算^[10],其公式如下。

$$H(x, y) = P(X \leq x_i, Y \leq y_i) = \frac{\sum_{m=1}^i \sum_{l=1}^i N_{ml} - 0.44}{N + 0.12} \quad (9)$$

式中: P 为 $X \leq x_i$ 且 $Y \leq y_i$ 的经验频率, N_{ml} 为 $x_i \leq x_i$ 且 $y_j \leq y_i$ 的数目, N 为样本数。

理论累积频率与经验累积频率的一致性比较结果见图 2。图 2 表明,Clayton 和 Gumbel-Hougaard 函数的拟合精度均高,需进一步进行拟合优度检验。

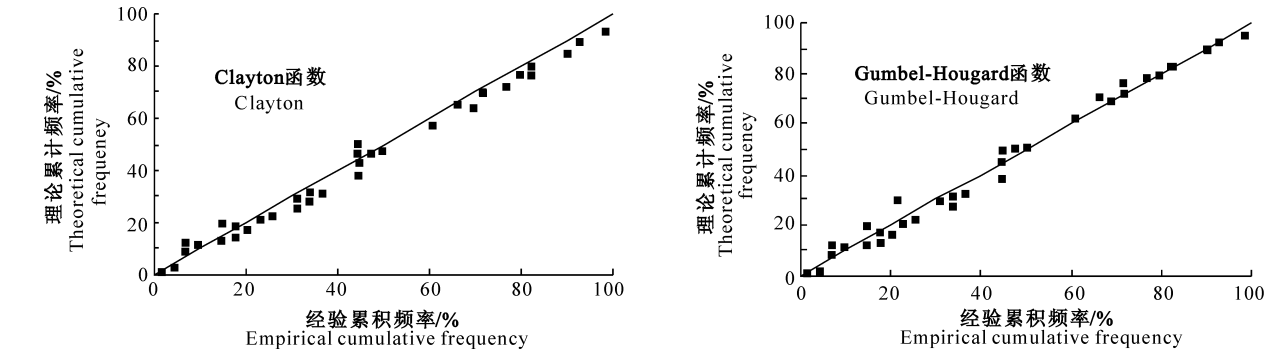


图 2 理论累积频率与经验累积频率的一致性比较

Fig. 2 Coherence comparisons of theoretical cumulative frequency and empirical cumulative frequency

对于模型拟合优度的检验,同样采用均方根误差 RSME、AIC 最小准则进行衡量,所得结果见表 4。

表 4 Copula 函数拟合优度的检验结果
Table 4 Copula function goodness-of-fit test

函数名称 Function caption	RMSE	AIC
Clayton	3.63	41.4
Gumbel-Hougaard	3.02	35.5

由表 4 可知,Gumbel-Hougaard 函数的拟合精度最好,可选作本研究的联结函数。

图 3 表明,经验点据与所选取的 Gumbel-Hougaard 函数计算的分布拟合情况较好,说明选取的 Copula 函数是合理的,可用来分析渭河和千河的丰枯遭遇问题。

3 渭河和千河的丰枯遭遇分析

根据渭河和千河年径流的边际分布和选取的 Copula 函数,可得到两河流域年径流的联合概率分布函数为:

$$H(x, y) = P(X \leq x, Y \leq y) = C(u_1, u_2) = \exp \{ - [(-\ln u_1)^{5.26} + (-\ln u_2)^{5.26}]^{\frac{1}{5.26}} \} \quad (10)$$

式中: P 为不超过概率,指 $X \leq x, Y \leq y$ 的概率; $C(u_1, u_2)$ 为联合概率分布函数, u_1, u_2 为两河流的边际分布函数。

定义两河流域年径流不超过概率 25% 为枯水年, 50% 为平水年, 75% 为丰水年。由两河流的年径流边际分布函数,可以确定两河流的特征年径流量(表 5)。

两河流丰枯遭遇分为 9 种情况:丰丰型: $p(x \geq 31.6, y \geq 4.9)$; 丰平型: $p(x \geq 31.6, 2.2 \leq y < 4.9)$;

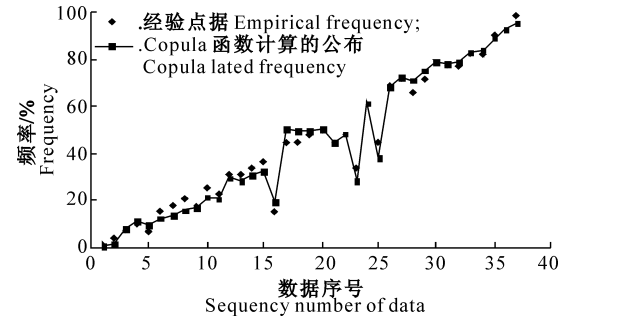


图 3 渭河和千河年径流丰枯遭遇计算分布与经验点据的拟合结果

Fig. 3 Empirical data and calculated distribution of encounter probability of annual runoff in the Wei river and Qian river

丰枯型: $p(x \geq 31.6, y < 2.2)$; 平丰型: $p(15.6 \leq x < 31.6, y \geq 4.9)$; 平平型: $p(15.6 \leq x < 31.6, 2.2 \leq y < 4.9)$; 平枯型: $p(15.6 \leq x < 31.6, y < 2.2)$; 枯丰型: $p(x < 15.6, y \geq 4.9)$; 枯平型: $p(x < 15.6, 2.2 \leq y < 4.9)$; 枯枯型: $p(x < 15.6, y < 2.2)$ 。各种情况的遭遇概率见表 6。

表 5 渭河和千河的特征年径流量

Table 5 Values of annual runoff in the Wei River and Qian River $\times 10^8 \text{ m}^3$

河 流 River	不超过概率 Non-exceeding probability		
	75%	50%	25%
	(丰水年 Rich flow year)	(平水年 Moderate flow year)	(枯水年 Pool flow year)
渭河 Wei River	31.6	22.6	15.6
千河 Qian River	4.9	3.2	2.2

表 6 渭河和千河的丰枯遭遇概率

Table 6 Synchronous-asynchronous encounter probability of rich-poor annual runoff in the Wei River and Qian River

河 流 River	丰枯遭遇类型 Synchronous-asynchronous encounter probability of rich-poor annual runoff	渭河 Wei River		
		丰 Rich	平 oderate	枯 Poor
千河 Qian River	丰 Rich	0.72	0.5	0.25
	平 Moderate	0.5	0.45	0.248
	枯 Pool	0.25	0.248	0.21

由表 6 可知,宝鸡市重要地表水源渭河和千河同时出现丰、平、枯水年的概率分别为:0.72,0.45,0.21;渭河出现平水年、千河出现丰水年与枯水年的概率分别为 0.5 和 0.248;千河出现枯水年、渭河出现丰水年的概率为 0.25。由于 Copula 函数具有对称性,推求出两河流的丰平与平丰、丰枯与枯丰、平枯与枯平遭遇的概率相等。图 4 为两河流域年径流遭遇的联合分布等值线图。从图 4 可以定量得到两条河流各种径流组合的概率,进而为宝鸡市水资源的合理配置与规划提供决策支持。

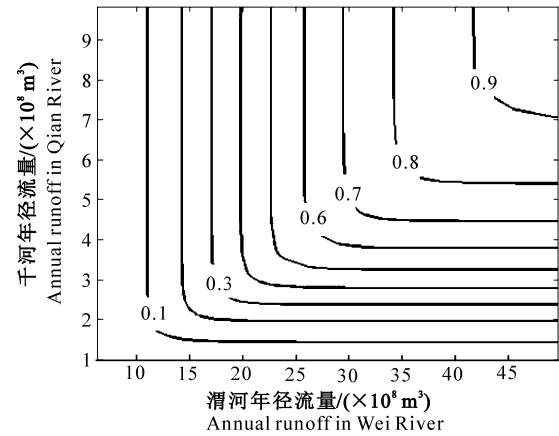


图 4 渭河与千河年径流遭遇的联合分布等值线
Fig. 4 Contour map of the joint distribution of the Wei Rver and Qian River

4 结 论

渭河宝鸡市区段生态治理工程是西北地区生态城市建设的新模式,工程的运行效益与干支流水源的联合调度、有效补给密切相关。本研究采用 Cop-

ula 方法,构建了渭河干流及其支流千河年径流的联合分布,分析了干支流的丰枯遭遇性,得到以下主要结论:

- (1)Copula 方法是构建联合分布的一种有效方法,用 Copula 方法描述干支流的丰枯遭遇性是可行的。
- (2)渭河干流年径流量的边际分布服从皮尔逊Ⅲ型分布,而支流千河年径流量的边际分布选用对数正态分布更为合适。应用二维阿基米德 Copula 函数,建立了干支流的联合概率分布模型,丰枯遭遇分析结果表明,渭河、千河干支流同时出现丰、平、枯水年的概率分别为 0.72,0.45 和 0.21;渭河干流出现平水年、千河出现丰水年与枯水年的概率分别为 0.5 和 0.248;千河出现枯水年、渭河干流出现丰水年的概率为 0.25。
- (3)干支流年径流遭遇的联合分布等值线图,定量描述了渭河和千河两河流各种径流组合的概率,可为宝鸡市水资源的合理调度与规划提供决策支持。

[参考文献]

[1] 谢 华. 水灾害防治中的多变量概率分布问题研究 [D]. 武汉: 武汉大学,2007:21-29.
Xie H. Study on multivariate probability analysis in water resources [D]. Wuhan: Wuhan University, 2007: 21-29. (in Chinese)
[2] Nelsen R B. An Introduction to Copulas [M]. New York: Springer,1999.
[3] Salvadori G, De Michele C. Frequency analysis via copulas: Theoretical aspects and application to hydrological events [J]. Water Resources Research,2004,40(12):12511-12512.

[4] Zhang L, Singh V P. Bivariate rainfall frequency distributions using Archimedean copulas [J]. Journal of Hydrology, 2007, 332:93-109.

[5] 熊立华,郭生练,肖义,等. Copula 联结函数在多变量水文频率分析中的应用 [J]. 武汉大学学报:工学版, 2005, 38(6):16-19.
Xiong L H, Guo S L, Xiao Y, et al. Application of copulas to multivariate hydrological frequency analysis [J]. Engineering Journal of Wuhan University, 2005, 38(6):16-19. (in Chinese)

[6] 闫宝伟,郭生练,肖义. 南水北调中线水源区与受水区降水丰枯遭遇研究 [J]. 水利学报, 2007, 38(10):1178-1185.
Yan B W, Guo S L, Xiao Y. Synchronous-asynchronous encounter probability of rich-poor precipitation between water source area and water receiving areas in the Middle Route of South-to-North Water Transfer Project [J]. Journal of Hydraulic Engineering, 2007, 38(10):1178-1185. (in Chinese)

[7] 李隆章. 实用非参数统计方法 [M]. 北京:中国财政经济出版社, 1989:119-180.
Li L Z. Applied nonparameter statistical method [M]. Beijing: Financial and economical press, 1989:119-180. (in Chinese)

[8] Genest C, Rivest L P. Statistical inference procedures for bivariate archimedean copulas [J]. Journal of the American Statistical Association, 1993, 88(423):1034-1043.

[9] 史道济,姚庆祝. 改进 Copula 对数据拟合的方法 [J]. 系统工程理论与实践, 2004(4):49-55.
Shi D J, Yao Q Z. A method of improving copula fitted to data [J]. Systems Engineering-theory & Practice, 2004(4):49-55. (in Chinese)

[10] Sheng Y. The Gumbel logistic model for representing a multivariate storm event [J]. Advance in Water Resources, 2001(24):179-185.

(上接第 130 页)

[13] Das K C, Smith M C, Gattie D K, et al. Stability and quality of municipal solid waste compost from a landfill aerobic bioreduction process [J]. Advances in Environmental Research, 2002, 6:401-409.

[14] 鲍士旦. 土壤农化分析 [M]. 3 版. 北京:中国农业出版社, 2000.
Bao S D. Soil agricultural chemistry analyses [M]. 3th Edition. Beijing:China Agricultural Press, 2000. (in Chinese)

[15] 李玉红,李清飞,王岩. 翻堆次数对牛粪高温堆肥的影响 [J]. 河南农业科学, 2006(7):70-72.
Li Y H, Li Q F, Wang Y. Effect of turning frequency on high temperature composting of dairy manure [J]. Journal of Henan Agricultural Sciences, 2006(7):70-72. (in Chinese)

[16] Eiland F, Klammer M, Lind A M, et al. Influence of initial C/N ratio on chemical and microbial composition during long term composting of straw [J]. Microb Ecol, 2001, 41(1):272-280.

[17] 中华人民共和国卫生部. 粪便无害化卫生标准, GB 7959—1987[S]. 北京:中国标准出版社, 2004.
Ministry of health of the people's republic of China. Sanitary standard for the non-hazardous treatment of night soil. GB 7959—1987 [S]. Beijing: Standards Press of China, 2004. (in Chinese)

[18] Finstein M S, Miller F C, Stom P F, et al. Waste treatment composting as a controlled system [J]. Biotechnology, 1986, 8(3):396-443.

[19] 贺亮,赵秀兰,李承碑. 不同填料对城市污泥堆肥堆体温度动态变化影响 [J]. 西南农业大学学报:自然科学版, 2006, 28(3):389-392.
He L, Zhao X L, Li C B. Effect of different bulking agents on transformation of nitrogen during sewage sludge composting [J]. Journal of Southwest Agricultural University: Natural Science Edition, 2006, 28(3):389-392. (in Chinese)

[20] 陈同斌,罗维,高定,等. 混合堆肥过程中容重的层次效应及动态变化 [J]. 环境科学, 2004, 23(5):143-147.
Chen T B, Luo W, Gao D, et al. Stratification of bulk density and its dynamics in the process of co-composting [J]. Environmental Science, 2004, 23(5):143-147. (in Chinese)