

西北内陆干旱区出山口年径流组成统计分析

梁忠民 陈元芳 董增川

(淮海大学水文水资源及环境学院 江苏 南京 210098)

摘要 采用时间系列的组成项分析法对西北内陆干旱区出山口年径流系列进行了统计分析 给出了相应的统计拟合模型 并由此进行了两步预报. 结果表明 用时间组成项分析法基本上能描述西北干旱区大多数站的统计变化特性 预报的结果有一定精度. 拟合的统计模型及所得结论可供参考.

关键词 干旱区 确定项 随机项 统计分析

中图分类号 :TV121 文献标识码 :A 文章编号 :1000-1980(2000)06-0011-04

中国西北的干旱区蕴藏着丰富的矿产资源和土地资源. 但干旱缺水也是本地区的最大特点 因此如何有效地开发利用水资源是进一步开发该地区的其它资源、发展西北经济的重要保证. 而开发水资源首先需要对该地区水资源的分配及其变化规律有充分的认识.

西北干旱区出山口径流量的大小是衡量该地区水资源量多少的重要指标 所以研究出山口径流量的特性 掌握其变化规律是水资源开发利用中的一个重要研究内容.

本文从统计的角度对出山口的年径流进行研究. 按地理位置将西北干旱区分成五个分区 即额尔齐斯河区、天山以北区、天山以南区、甘肃河西区及青海河湖区 对各区出山口的年径流系列进行了趋势性、周期性、随机性的统计分析 给出了各组成项的统计拟合模型 并由此进行了初步的预报研究.

1 5 个地理分区的降水与径流特性

额尔齐斯河区的多年平均年降水量 自南向北递增变幅为 150 ~ 700 mm 全区多年平均在 450 mm 左右. 各站年径流的平均值 $\overline{C_v} = 0.31$ $\overline{C_s/C_v} = 2.5$ 各站年径流均不存在自相关 17% 的站存在趋势.

天山以北区多年平均年降水量在区内变化较复杂. 最高的巩乃斯河可达 800 mm 以上 山前平原年降水量则在 150 ~ 200 mm 昆仑山区年降水量仅 300 ~ 500 mm 区内多半地区不足 100 mm. 该区年径流 $\overline{C_v} = 0.24$ $\overline{C_s/C_v} = 4.0$ 有 9 个站具有明显自相关 1/4 站有趋势.

天山以南区多年平均年降水量是各区中最少的. 全区平均年降水量少于 100 mm 有一些地区在 100 ~ 300 mm 之间. 年径流 $\overline{C_v} = 0.18$ $\overline{C_s/C_v} = 5.2$ 有 4 个站年径流具有明显自相关 42% 的站有趋势.

甘肃河西区多年平均年降水量在 45 ~ 600 mm 之间. 石洋河流域最高 山前平原自东向西 由南向北递减. 年径流 $\overline{C_v} = 0.18$ $\overline{C_s/C_v} = 4.5$ 具有明显自相关的测站比例最高 42% 的站有趋势.

青海河湖区多年平均年降水量较少 由东部的 400 mm 至西北部 20 mm 上下. 年径流 $\overline{C_v} = 0.31$ $\overline{C_s/C_v} = 3.8$ 仅有两个站具有明显自相关 30% 站有趋势.

2 时间序列组成项统计分析法

一个水文时间序列 X_t 一般可表示成 :
$$X_t = D_t + S_t \tag{1}$$

式中 : D_t 表示确定性成分 如跳跃、趋势及周期等项 ; S_t 为随机性成分 一般可分成相依随机成分和独立随机成分. 时间序列组成项统计分析就是用统计方法对系列中的各组成项进行识别和模拟 再由各组成项的叠加完成对 X_t 的模拟或预报.

本文所研究的年径流系列中 不存在明显的跳跃(突变)成分 故仅对其趋势、周期及随机成分进行识别和拟合模型.

2.1 趋势成分的识别和拟合

一般采用坎德尔(kendall)秩次相关法对趋势项进行识别^[1] 检验的统计量为

$$U = \tau / \sigma_\tau \quad \tau = [4P \wedge (n(n-1))] - 1 \quad \sigma_\tau = [\chi(2n+5)][9n(n-1)] \quad (2)$$

原假设序列无趋势成立时 U 为标准正态分布,若检验结果为有趋势(即检验是显著的),可用下式对趋势成分进行模拟:

$$T(t) = a_0 + a_1 t + \dots + a_k t^k \quad (3)$$

参数 $a_i (i = 1 \sim k)$ 可用最小二乘法确定.经分析,本次统一采用线性趋势(即 $k = 1$)进行拟合.

2.2 周期成分的识别及拟合

采用简单分波法^[2]对周期成分进行识别和拟合.假设系列中有 k 年周期,将系列按 k 年一组共分成 m 组,定义 F 检验统计量:

$$F = \frac{Q_3^2 \wedge (k-1)}{Q_2^2 \wedge (n-k)} \sim F(k-1, m-k) \quad (4)$$

式中: Q_2^2 ——组内离差平方和; Q_3^2 ——组间离差平方和; n ——样本容量.对给定的显著性水平 α ,临界值记为 F_α .若 $F > F_\alpha$,则认为有 k 年周期,否则没有 k 年周期.

若检验是显著的(有 k 年周期),则以各组组内均值作为 k 年的周期值,在原系列中顺序减去该周期值,得到相应余波过程,记为 $A(t)$,对 $A(t)$ 重复上述计算过程,则又得到另一周期过程及相应的余波,将这个过程重复 L 次,则得到 L 个周期项及其余波.序列的周期成分即是由这 L 个周期项叠加而成.本次计算中约定 $L = 3$,即只考虑到其前三个周期,认为后面的周期成分影响不大,显然这是一种近似,实际上在分析中只要是显著的周期成分均应予以考虑.

2.3 平稳 ARMA 模型的拟合

从原系列中依次剔除趋势成分和周期成分,将剩余的系列作为平稳时间序列,用平稳 ARMA 模型拟合.模型为

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (5)$$

式中: ϕ_i, θ_j ——系数; p, q ——模型阶数; ε_t ——白噪声系列.

用 AIC 准则确定模型阶数.模型参数的估计用最小二乘法.本次计算时限定 $0 \leq p, q \leq 3$.

2.4 预估

按上述 2.1~2.3 的方法识别各组成项,并建立相应的拟合模型,然后由各组成项模型预报的叠加对年径流进行预报.

对确定性项的预报,直接按时间外推计算,对 ARMA 模型的预报采用逆函数法.公式为

$$\hat{X}_k(l) = \begin{cases} \sum_{j=1}^M f_j^{(l)} X_{t+l-j} & 0 \leq l \leq q \\ \sum_{i=1}^p \phi_i \hat{X}_k(l-j) & l > q \end{cases} \quad (6)$$

式中: $\hat{X}_k(l)$ ——在时刻 k 做未来 l 步预报; M ——充分大的数;系数 $f_j^{(l)}$ 与逆函数 I_j 的关系满足:

$$\left. \begin{aligned} f_j^{(1)} &= I_j & l &= 1 \\ f_j^{(l)} &= I_{j+l-1} + \sum_{i=1}^{l-1} I_i f_j^{(l-i)} & l &> 1 \end{aligned} \right\} \quad (7)$$

3 年径流组成分析与预估结果

在西北干旱区的 5 个分区中,选择了 62 个出山口年径流系列,系列平均长度为 30 年,所选资料基本上能反映西北干旱区出山口年径流的多年变化状况.

对每一系列用前述的方法进行各组成项分析及模型的拟合.结果表明,62 个站中,共有 41 个的年径流系列表现出具有确定性(趋势或周期)或相依随机的变化特性,其余系列则显示出独立随机的变化特征.特别地,额尔齐斯河区所研究的全部 6 个站点以及青海河湖区的大部分站点的年径流系列均显示出其独立随机的特征.限于篇幅,本文仅列出部分结果,见表 1,详细结果参见有关研究^①.

① 河海大学.西北干旱区出山口年径流统计变化规律分析.1998.

表 1 年径流系列组成项统计分析结果

Table 1 Results of Statistic Component Analysis of Annual runoff

分区	站名	趋势项		周期项/a			随机项
		$Y_t = a + bt$		第一 周期	第二 周期	第三 周期	
		a	b				
天山以北区	恰甫(合成)	21.3	-0.28		无		独立随机
	乌拉斯台(二)	0	0	11	3	0	独立随机
	英雄桥	0	0	6	7	0	$S_t = \epsilon_t - 0.254\epsilon_{t-1} + 0.900\epsilon_{t-2}$
	哈地坡(合成)	0	0	3	0	0	独立随机
天山以南区	玉孜门洛克	22.9	0.16	8	5	7	$S_t = 0.923S_{t-1} - 0.273S_{t-2} + \epsilon_t - 1.044\epsilon_{t-1}$
	阿合布隆	0	0		无		$S_t = 0.416S_{t-1} + \epsilon_t - 0.442\epsilon_{t-1} - 0.532\epsilon_{t-2}$
	大山口	116.1	-0.69		无		$S_t = 0.378S_{t-1} + \epsilon_t - 0.670\epsilon_{t-1} - 0.335\epsilon_{t-2} - 0.389\epsilon_{t-3}$
	努努买买提兰干	0	0	15	13	0	$S_t = \epsilon_t - 0.440\epsilon_{t-1} - 0.504\epsilon_{t-2} - 0.293\epsilon_{t-3}$
甘肃河西区	沙沟寺	0	0	11	3	0	独立随机
	冰沟	0	0		无		$S_t = 0.313S_t + \epsilon_t$
	莺落峡	49.2	-0.17		无		$S_t = 0.415S_t + \epsilon_t$
	红崖山	12.3	-0.25		无		$S_t = 0.403S_{t-1} + 0.426S_{t-2} + 0.126S_{t-3} + \epsilon_t + 0.995\epsilon_{t-1}$
青海	诺木洪	4.26	0.04		无		$S_t = \epsilon_t + 0.489\epsilon_{t-1}$
河湖区	香日德	8.53	0.42		无		$S_t = 1.007S_{t-1} - 0.506S_{t-2} + \epsilon_t - 1.043\epsilon_{t-1}$
	格尔木	20.3	0.30		无		$S_t = \epsilon_t - 0.217\epsilon_{t-1} - 1.021\epsilon_{t-2}$

对系列中趋势性成分的分析表明,62 个计算系列中,有 21 个显示出有趋势性,其中以天山以南区和甘肃河西区的站数为多.本文工作中不仅对有趋势性成分的序列进行了线性模型拟合,而且还对拟合的方程进行了检验,最终确定共有 16 个系列有线性趋势.

对周期性成分的分析表明,62 个系列中有 15 个有周期变化特性,而其中约有一半显示出仅有一个第一周期项.对有周期组成的各系列,其第一周期的平均长度:天山以北区为 7.2 a;天山以南区为 9.7 a;甘肃河西区为 8.3 a,平均约为 8.6 a.第一周期一般可作为系列的主要周期项,故对所研究区域,少数站的年径流中可能有约 9 a 的周期变化.

62 个系列中,有 32 个系列的 ARMA 模型阶数 p, q 不为零.这表明有半数站的年径流系列中存在着一定程度的相关结构,所以系列间的相依性在规划及设计中不应被完全忽略.

根据年径流各组成项模型预报结果的叠加,分别对 41 个站的年径流进行预报.本次工作中根据收集到的资料做了最新的两步预报(即取式(6)中的 $l = 1, 2$).结果表明,一步和两步预报的相对误差绝对值小于 20% 的站占计算站数的比例分别为 80% 和 71%;若以小于 25% 统计,比例则分别为 90% 和 83%.本文仅列出了表 1 中各站的预报结果,见表 2.

表 2 年径流两步预报结果

Table 2 Results of two-step prediction for annual runoff

分区	站名	预报 步长	年份	实测值 ($\text{m}^3 \cdot \text{s}^{-1}$)	预报值 ($\text{m}^3 \cdot \text{s}^{-1}$)	相对误差 /%
天山以北区	恰甫(合成)	1	1987	16.70	12.50	-25.1
		2	1988	19.70	12.30	-37.9
	乌拉斯台(二)	1	1987	102.30	90.10	-11.9
		2	1988	127.20	111.80	-12.1
	英雄桥	1	1987	7.70	7.92	2.9
		2	1988	7.20	7.94	10.2
	哈地坡(合成)	1	1987	8.00	8.18	2.3
		2	1988	9.10	7.58	-16.7
	玉孜门洛克	1	1987	28.60	26.00	-6.9
		2	1988	30.20	28.50	-5.6
天山以南区	阿合布隆	1	1987	47.60	46.70	-1.9
		2	1988	46.50	46.20	-0.6
	大山口	1	1987	106.20	99.20	-6.6
		2	1988	101.00	100.20	-0.8
	努努买买提兰干	1	1987	8.97	7.38	-17.7
		2	1988	8.35	7.99	-4.4
	沙沟寺	1	1985	8.37	9.32	11.3
		2	1986	8.45	10.20	20.8
	冰沟	1	1994	17.50	19.50	11.8
		2	1995	19.30	20.00	3.6
甘肃河西区	莺落峡	1	1994	44.60	47.60	6.7
		2	1995	46.60	43.60	-6.5
	红崖山	1	1994	4.29	5.08	18.4
		2	1995	3.89	3.45	-11.2
	诺木洪	1	1984	6.44	5.41	16.0
		2	1985	4.29	5.42	-26.4
	香日德	1	1984	25.50	21.60	-15.1
		2	1985	17.80	21.30	20.1
	格尔木	1	1984	25.00	26.40	5.9
		2	1985	21.00	18.70	-10.8

4 主要结论与问题

a. 本文尽可能多地收集了西北内陆干旱区出山口的年径流系列,最后选定的 62 个计算站点,系列较完整,在时空分布上具有较好的代表性,因此所得到的分析结果基本上能反映所研究地区年径流的统计变化规律.

b. 在进行线性趋势的识别和拟合工作中,曾采用了多种检验方法,并对各种方法的结果进行了比较,检验结果较为合理.分析结果表明,约 $1/4$ 的序列表现出线性趋势的变化规律,而其中有增大或减小趋势的约各占一半.所以在今后对西北干旱区年径流的研究及应用中,应对其趋势变化特性给予重视.

c. 在识别和拟合周期项时,采用了简单分波法.在确定某一周期成分时,该法是以各组均值作为周期值,辅之以 F 假设检验.因此该法结构简单,计算方便,避免了常规方法需要确定较多参数的缺点.分析结果表明,所研究地区出山口的大多数年径流系列中并没有显示出有多年的周期变化,少部分站(约 $1/4$)可能有约 $9a$ 的周期变化.

d. 本次工作中采用时间序列统计组成项分析法对年径流系列进行统计描述,并由各组成项模型外推的叠加对年径流做了两步预报.预报结果表明,一步和两步预报的平均相对误差(绝对值)均为 13% 左右.按相对误差(绝对值)小于 20% 统计,则一步和两步预报的合格率在 80% ~ 70% 之间,若按小于 25% 统计,则合格率在 90% ~ 80% 之间.这意味着用组成项分析法能较好地描述所研究区域年径流的统计特性,组合模型的精度良好.特别地对天山以南区,预报合格率接近 100%(以相对预报误差绝对值小于 20% 计).另外,通过组成项分析发现,额尔齐斯河区各站的年径流都显示出独立随机的变化规律,相类似地,青海河湖区大多数站的年径流亦可作为独立随机项处理.

e. 年径流的变化是由很多影响因素综合作用所决定的.本文没有对年径流的影响因素及其与年径流之间的统计关系做研究,而仅研究了年径流时间序列自身的统计组成规律.这是本研究的不完整之处,因此将年径流时序变化规律的研究与对其影响因素的研究相结合是今后进一步的研究内容.

参考文献:

- [1] 丁晶,邓育仁.随机水文学[M].成都:成都科技大学出版社,1988.76 ~ 77.
- [2] 金光炎.水文水资源随机分析[M].北京:中国科学技术出版社,1993.411 ~ 418.

Statistic Component Analysis for Annual Runoff at Mountain-pass Stations in Drought Northwest Regions

LIANG Zhong-min, CHEN Yuan-fang, DONG Zeng-chuan

(College of Water Resources and Environment, Hohai University, Nanjing 210098, China)

Abstract :In this paper, the method of statistical component analysis(SCA) is applied to the analysis of the annual runoff time series of mountain-pass stations in drought Northwest regions. Statistical forecasting has been made for each series by the fitted model. The outcomes show that it is suitable in general to adopt SCA to fit annual runoff in the concerned area.

Key words :drought area ;deterministic component ; stochastic component ; statistical analysis