岭回归法在单位线估计中的应用

陈　科　辛忠礼

(成都勘测设计研究院,成都,610072)

摘　要　本文在一般的最小二乘矩阵解法的基础上,采用岭回归法估计单位线,在估计中预先引入单位线形状,并以参数 K 来反映引入单位线形状的影响程度,从而克服了一般最小二乘法估计单位线时产生的锯齿波动现象,并且能在净雨时段数较多的资料条件下进行有效求解。本法用于实际工程单位线估计,取得了较好效果。

关键词　岭回归法　单位线　锯齿波动现象

1　问题的提出

单位线的求解实质是在知道流域系统输入/输出的情况下,对系统响应函数进行识别。响应函数识别的方法较多,工程设计中常采用试错分析法,方法对采用的实际资料要求:净雨时段数越少越好,尽量只有1至2个时段,径流过程一般为一次单峰过程。当实际资料中径流不是单峰过程而是复峰过程、净雨的时段较多时,试错率定相当困难。为此本文在总结单位线估计各种方法特点基础上,采用岭回归(Ridge regression)方法来推求单位线,为工程设计服务。

2　基本原理

若流域系统为集总式的线性时不变系统,而净雨采用单位线汇流时,汇流方程的矩阵形式可写为:

$$Xh = Y + \varepsilon$$

式中　X——$n \times m$ 矩阵,净雨(mm);

Y——n 列向量,直接径流(m^3/s);

h——m 列向量,单位线(m^3/s);

ε——n 列向量,误差项(m^3/s)。

如果从上式反推单位线,则可以看出单位线 h 的求解实质上是多元线性回归问题,若采用一般的最小二乘法原理求解上式,其正规方程为:

$$X^T X\hat{h} = X^T Y \qquad (2)$$

式中　$\hat{h}$——h 的估计值;

X^T——X 的转置矩阵。

单位线 h 的估计值 $\hat{h}$ 为:

$$\hat{h} = (X^T X)^{-1} \cdot X^T Y \qquad (3)$$

实际应用当中,在一些情况下会出现协方差矩阵 $X^T X$ 退化或其行列式值很小的现象,致使估计值不确定,为此需在方程求解上作一些调整,用岭回归的方法即对协方差矩阵引入另一矩阵,

$$(X^T X + KI)\hat{h} = X^T Y \qquad (4)$$

式中　K——参数,$K \geqslant 0$,当 $K = 0$ 即为一般的最小二乘法估计;I——m 维单位矩阵。

求解单位线,公式如下:

$$\hat{h} = (X^T X + KI)^{-1} \cdot X^T Y \qquad (5)$$

图1给出了由净雨和直接径流推求单位线的过程,从图中可以看出,岭回归法明显优于一般的最小二乘法。在岭回归分析中,参数 K 值起着关键作用,它既降低了方程(3)数值解的不确定性,又使估计值得到了光滑。在岭回归法估计单位线时,还需加入其它约束条件才能很好地克服单位线估计中出现的不合理现象,使所求得的单位线满足下列条件:

1.净雨和直接径流水量平衡,并为1个

单位；

2. 单位线过程是光滑无锯齿波动的，并无负值；

3. 一般的单位线形状，多呈单峰铃形曲线；

4. 单位线的峰现时间为 t_p，在 t_p 时段以前为涨水段、流量值逐渐增加，在 t_p 之后为退水段、流量逐渐减小。

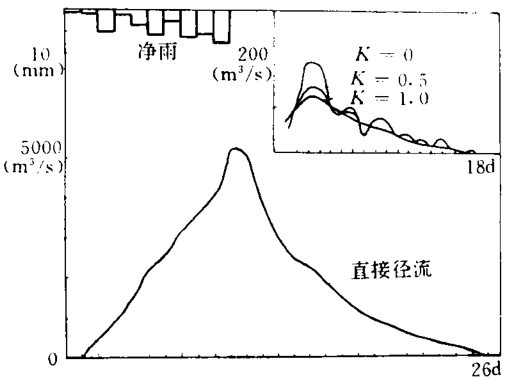

图 1　净雨、直接径流推求单位线(岭回归法)

首先考虑两种极端情况，一是对估计的单位线不作任何假定，直接采用实际样本用一般的最小二乘法即方程(3)来估计；另一是仅根据流域特性而不使用实际样本对单位线作出估计，即：

$$h = h_p + e \tag{6}$$

式中　h_p—— 综合的单位线形状；

　　　e—— 误差项。

用上述两种情况对单位线估计的方程(1)和(6)进行加权组合，则有：

$$\begin{bmatrix} X \\ \Omega \end{bmatrix} h = \begin{bmatrix} Y \\ \Omega h_p \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \Omega e \end{bmatrix} \tag{7}$$

式中　Ω——$m \times m$ 的权重系数矩阵。

Ω 是对角线矩阵，对角线上的元素为不同的权重系数 W_i 即$(W_1, W_2, \cdots\cdots, W_m)$，当所有的元素的权重系数相等为 W 时，则 Ω 简化为 WI。加权组合后的估算方程系数矩阵变为$(n+m) \times m$ 矩阵，对方程稍加变换有：

$$(X^TX + \Omega^T\Omega)\hat{h} = X^TY + \Omega^T\Omega h_p \tag{8}$$

当 $\Omega = WI$ 时，上面方程简化为：

$$(X^TX + KI)\hat{h} = X^TY + Kh_p，其中 K = W^2 \tag{9}$$

将上式与岭回归方程(4)进行比较可知，当 $h_p = 0$ 时，上式与方程(4)完全相同，可见岭回归方程是上式的一种特情。若 h_p 采用一个综合的单位线形状，并假定 ε、e 服从均值为零、方差分别为 6^2I 和 6_p^2I 的正态分布，则式(9)近似地为参数的后验最大似然估计；若使用先验信息和实际样本数据，$K = S^2/6_p^2$，其中：S^2 为式(1)解的残差平方和，可以看出 K 是先验信息和实际样本的权重之比，即是说参数 K 可以反映先验信息在估计中的影响程度。

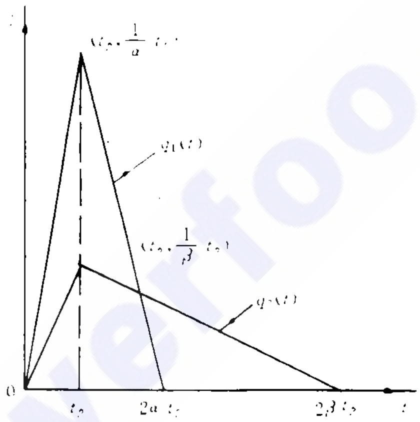

图 2　假定的三角形复合单位线

通过上面的分析，可以知道一般的最小二乘法只考虑实际样本数据，所作的单位线估计单纯依赖于实际样本，当雨洪资料有误差时，计算所得单位线会出现一些锯齿波动现象，不符合客观的水文过程；若仅依靠对流域特性分析，脱离实际水文过程、没有充分使用实际雨洪资料信息求解单位线，则不能反映流域的实际情况；将上述两种信息结合起来，对估计的单位线形状预先作出估计，并以参数 K 值反映预先假定的单位线形状信息的影响程度，进行求解是有实用价值的。

本文根据文献〔2〕提出的三角形复合单位线，见图 2，其函数形式为：

$$\begin{cases} q(t) = \frac{1}{2}[q_1(t) + q_2(t)] \\ q_1(t) = \begin{cases} \frac{t}{\alpha} t_p^2 & 0 < t \leqslant t_p \\ \frac{(2\alpha t_p - t)}{\alpha(2\alpha - 1)} t_p^2 & t_p < t \leqslant 2\alpha \cdot t_p \\ 0 & \text{其它} \end{cases} \\ q_2(t) = \begin{cases} \frac{t}{\beta} t_p^2 & 0 < t \leqslant t_p \\ \frac{(2\beta t_p - t)}{\beta(2\beta - 1)} t_p^2 & t_p < t \leqslant 2\beta \cdot t_p \\ 0 & \text{其它} \end{cases} \\ \beta \geqslant \alpha \end{cases}$$

来作为待估计单位线的形状。从上述的三角形复合单位线可以看出，所求单位线取决于假定的 $q_1(t)$、$q_2(t)$，为两者的平均。当 $q_1(t)$、$q_2(t)$ 发生变形，两者平均也发生变形，致使所求解的单位线随之发生变形，其峰值和底长均发生变化。$q_1(t)$、$q_2(t)$ 两者的变化灵活，可以适应不同流域的单位线特征。$q_1(t)$、$q_2(t)$ 的变化依赖参数 α、β 的变化，α 为峰顶偏态参数，$\alpha = 1$ 时，q_1 为等腰三角形，α 比 1 大得愈多，q 的顶部的偏态性愈烈；参数 β 主要影响单位线的退水过程；参数 t_p 主要影响单位线的峰值。

参数 t_p、α、β 的求解一般采用回归技术，以之建立与流域特征和降雨特征的关系式来进行，但所建立的关系式是一种概化，不仅所需资料较多、计算繁琐，同时也带有误差、失真。本文采用较为简单实用的办法，以经验判断来进行。直接径流时段数 $TW = N + M - 1$，其中，M 为净雨时段数，N 为单位线底宽(时段数)，则可粗估单位线底宽 $N = TW - M + 1$，进一步有 $N = 2\beta t_p$，则 $\beta = N/2 \cdot t_p$，而 α 又满足 $0.5 < \alpha \leqslant \beta$，由此可初定岭回归法估计单位线形状参数，同时结合试错法、考查直接径流的拟合情况，重点在率定参数 K 值以最后确定单位线。

3 实际应用

为了对本文提出的单位线估计方法的适应性进行检验，在雅砻江洼里站以上流域的 PMF 推流的单位线估计中，采用岭回归法进行实验计算，取得了较好的效果。

一般而言，由实际净雨与直接径流估计单位线时，采用的资料都有误差，使估计的单位线有锯齿波动现象，产生这种现象的原因较多，有面平均雨量的计算、降雨的扣损(初损、后损)、直接径流过程的分割等等，都会给计算带来问题，同时由于实际资料不同，使求得的单位线有差异、有时甚至差异很大，这时需对大量的实际资料能快速有效求解，以便进行单位线的分析、分类和综合。本文采用了单位线估计中的两类资料，一是直接径流属“肥”型，其中包括单峰、双峰过程，见图 3(a)、(c)，另一是直接径流属“瘦”型，其中也包括单峰、双峰过程，见图 3(b)、(d)，本方法对这些资料的计算均有效。通过计算得到估计各个单位线的参数如附表，并解算出各个单位线来，求得的单位线有肥低型和瘦高型两种，它们对直接径流的拟合令人满意，见图 3，这些计算结果对该流域的降雨径流特

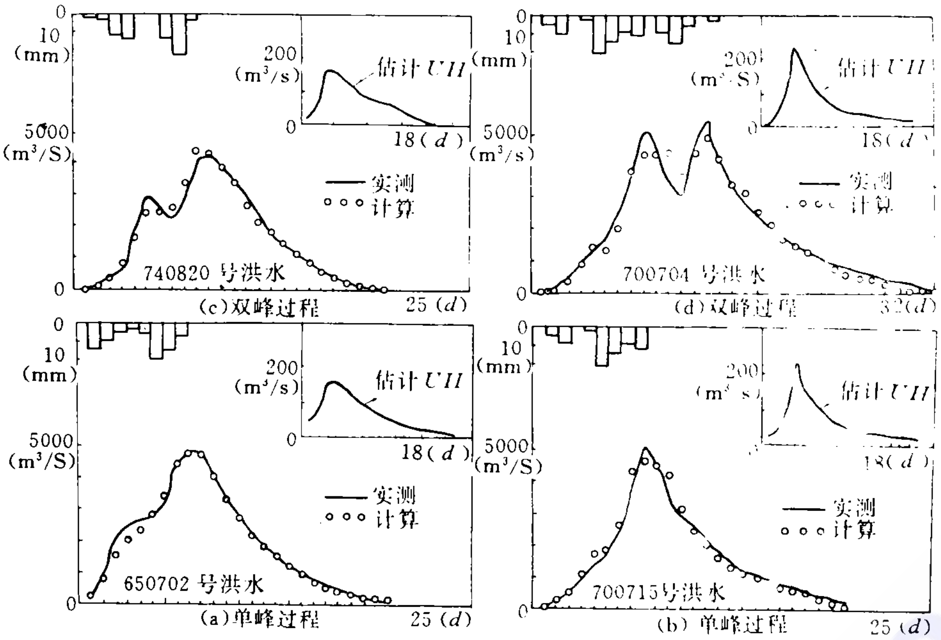

图3 实测净雨、直接径流推求单位线

性的分析研究有着实际意义。

附表 单位线估计参数表

编号	单位线底宽(d)	参数 K 值	单位线形状参数		
			t_p	α	β
(*a*)	18	0.8	3	2.5	4.0
(*b*)	18	0.8	4	1.5	4.5
(*c*)	17	0.5	3	2.5	4.0
(*d*)	19	0.5	4	1.5	4.5

注:编号为图3中的编号

4 结 语

通过对实际资料的分析计算,我们认识到:

1. 岭回归法求解单位线并在单位线估计时先引入单位线形状的假定是合理可行的。

2. 当实际资料中的净雨时段数较多,而且直接径流为单峰、双峰或多峰过程时,岭回归法均可有效求解,它克服了分析试错法中反复试错、有时当净雨时段数较多无法求解的现象。

3. 本方法克服了一般最小二乘法估计中产生的锯齿波动现象,使求解的单位线更符合流域的实际情况,为单位线的分析、分类和综合,为实际工程中大量资料的计算,提供了有力手段。

参 考 文 献

1 M. Bruen & J. C. I. Dooge, unit hydrograph estimation with multiple events and prior information, Hydrological Sciences J. Vol. 37, No. 5, Oct. 1992

2 任之. 三角形复合单位线. 水文, 1985; 第2期

3 朱保佑、袁明一. 用目标规划方法解算单位线, 水文, 1986; 第1期

4 庄一鸰、林三益合编. 水文预报. 水利电力出版社, 1986

5 罗积玉、邢瑛著. 经济统计分析方法及预测. 清华大学出版社, 1987

6 同济大学数学教研室编. 线性代数. 高等教育出版社, 1982

(收稿日期:19930611)