

降雨、径流因子的初步研究

Ⅱ—土壤坡面侵蚀量预报

杨艳生 史德明 孙志刚

(中国科学院南京土壤研究所) (江西省都宁县水土保持站)

在土壤侵蚀研究和生产实践中,常常要求对山丘地区坡面上的土壤侵蚀量作出预报,以便采取侵蚀的防患措施。预报土壤侵蚀量的方法,美国曾提出土壤流失预报通用方程,就中将引起土壤侵蚀的所有因子归纳为降雨因子、土壤因子、地形因子、经营管理和保土因子。在这些因子中,除降雨因子外,其它因子的实际值都是根据经验或与标准地块相比而得出的。因此,对于地形不甚复杂、地域较大、研究工作又较为规范化时,才有可能建立土壤流失预报通用方程。但如果地域不大、地形复杂,又还未取得足够的完整资料,建立或应用土壤流失预报通用方程就很困难。然而应用现有径流小区的观测资料,建立回归方程进行山丘坡面土壤侵蚀量预报则是可能的。本文应用 I 部分(本刊1984年第6期第47—50页)中的基本观测资料进行回归分析,以此作为实例,讨论这一方法的实际应用。回归分析方法有多种,这里采用的是逐步回归分析方法。这一方法当观测因子多时,有更多的优越性。由于所要求分析处理的数据都是大量的,所以分析计算必须借助电子计算机去完成。

一、逐步回归分析的基本思想和计算步骤

回归分析所要解决的问题是因变量(Y)和自变量(X)的关系。因变量只有一个,这就是研究对象所产生的结果,如土壤侵蚀量;自变量有多个或很多个乃至上百个,这就是影响研究对象的因子,如影响侵蚀量的有降雨量、径流系数等等。回归分析的目的就是要根据已有资料建立起自变量与因变量间的关系方程。在一般回归分析中,是将全部因子引入方程的,这样就出现两个问题:有些自变量对因变量所起作用很小,将这些变量引进方程后不但不会提高而且还会降低回归分析效果;此外,有些自变量之间相关性显著,因此也没有必要将这些高度相关的自变量都引入方程。因此,就有必要对那些自变量进行筛选,使回归效果好的自变量引入方程,剔除那些效果差的因子。逐步回归分析法就是将效果最佳的变量逐个引入,逐一进行统计检验,以决定某个变量是否引入或剔除,直至既没有变量可引入,又没有变量可剔除为止。计算步骤如下:

假定某项研究包含有M项因子,并作了N次观测。若自变量为X、因变量为Y,则 $y_i \in Y$ $i=1, 2, \dots, N$; $x_j \in X$ $j=1, 2, \dots, M$ 。如果把因变量Y排列在自变量X之后,就构成了N行、M+1列矩阵。

1、计算各变量的相关阵。计算公式如下:

$$r_{ij} = \frac{1}{n} \sum_{k=1}^n (X_{ki} - \bar{X}_i) (X_{kj} - \bar{X}_j) / S_i S_j$$

$$i, j = 1, 2, \dots, M, M+1$$

式中: $\bar{X}_i = \frac{1}{n} \sum_{k=1}^n X_{ki}, \quad S_i = \left[\sum_{k=1}^n (X_{ki} - \bar{X}_i)^2 \right]^{\frac{1}{2}}$

相关系数 r_{ij} 可排列成相关阵, 为对称阵, $|r_{ij}| \leq 1$, 具体形式为:

$$r_{ij} = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1M+1} \\ r_{21} & r_{22} & \dots & r_{2M+1} \\ \vdots & \vdots & & \vdots \\ r_{M1} & r_{M2} & \dots & r_{MM+1} \\ r_{M+11} & r_{M+12} & \dots & r_{M+1M+1} \end{pmatrix} \quad (1)$$

2、在相关阵(1)中最后一列, 选取与因变量相关系数最大的自变量计算 V_i 值。

$$V_i = a_{iM+1} \times a_{M+1i} / a_{ii} \quad (2)$$

在选取第一个变量(第一步计算)时, $a_{ij} = r_{ij}$ 。在以后的计算中, V_i 只为正值或负值, 分别代表未选入和已选入两类变量。在矩阵中还挑选一个与因变量相关性最小的自变量, 即按式(2)计算的负 V_i 值中绝对值最小(V_{\min})的一个变量, 供在检验是否剔除变量时使用。

3、计算 F_2^* 值, 确定变量是否剔除。 F_2^* 的计算公式如下:

$$F_2^* = |V_{\min}| \phi / a_{M+1, M+1}$$

式中 $\phi = n - k_2 - 1$ (k_2 为选入回归式中变量的个数)。若 $F_2^* < F_2$, 则剔除相对应的变量(F_2 为一定置信水平下的查表值, 可预先给定或计算给出), 这时将自由度加1, 转入第5步计算。若 $F_2^* \geq F_2$, 则按下一步计算。

4、计算 F_1^* 值确定是否引入变量。

$$F_1^* = V_{\max} (\phi - 1) / (a_{M+1, M+1} - V_{\max})$$

若 $F_1^* > F_1$ (F_1 为一定置信水平下的查表值), 则引入对应 V_{\max} 的变量, 这时自由度减1, 转入第5步计算。若 $F_1^* \leq F_1$, 则转入第7步计算。

5、重新计算新的系数矩阵。凡是引入或剔除第 k 个变量, 对相关阵中的系数都要重新计算。

如在第 S 步回归后, 计算第 $S+1$ 步的系数矩阵:

$$a_{ij}^{S+1} = \begin{cases} (a_{ij}^S a_{kk}^S - a_{ik}^S a_{kj}^S) / a_{kk}^S & i \neq k, j \neq k \\ a_{kj}^S / a_{kk}^S & i = k, j \neq k \\ -a_{ik}^S / a_{kk}^S & i \neq k, j = k \\ 1 / a_{kk}^S & i = j = k \end{cases}$$

原始相关阵是对称的, 但新的系数阵则是非对称的。经过计算后的系数所代表的意义是: 当自变量 X_i 和 X_j 在该步都不在回归式中时, 则 $a_{ij} = r_{ij}$; 当 X_i 在回归式中, X_j 不在回归式中时, 则 $a_{ij} = b_{ij}$ 。 b_{ij} 是 X_i 对于 X_j 的已规格化了的偏回归系数, 已扣除了在回归式中其它任一变量的影响,

当 X_i 不在回归式中, 而 X_j 在回归式中时, 则 $a_{ij} = -b_{ji}$ 。

6、计算偏回归系数和偏差平方和。引入回归式中第 i 个变量的偏回归系数:

$$b_i = b_{iM+1} \times S_{M+1} / S_i$$

$$b_0 = \bar{Y} - \sum_i b_i \bar{X}_i$$

\sum_i 是对所有已在回归式中的自变量求和; S_i , S_{M+1} 已在第一步时求出; b_{iM+1} 在该步系数矩阵中给出。偏差平方和用下式计算:

$$SS_{\text{偏}} = S_{M+1}^2 \cdot a_{M+1M+1}$$

7、选入和剔除自变量完成后, 可根据回归方程计算各因变量 y 值和测定值与计算值之差。

二、土壤侵蚀量预测

要进行土壤侵蚀量预测, 首先要根据已有资料建立起回归方程。

1、利用回归方程预测土壤侵蚀量的可能性。在前述 I 部分中已作讨论, 土壤侵蚀量 (或侵蚀模数) 与降雨量、降雨强度、地面径流量 (或径流深度)、径流系数和侵蚀强度 (单位径流深产生的侵蚀量) 都有显著相关, 就是说这 5 项因子和土壤侵蚀量有密切关系, 因而就有可能从这 5 项因子为自变量建立侵蚀预报方程。同时由于上述 5 项因子之间, 有些因子彼此之间相关性又很显著, 因而所建立的回归方程一般所包含的因子要少于 5 项。

2、回归分析的变量选取。在 I 部分中, 所分析的自变量因子有 9 项, 其中悬移质和推移质是直接和侵蚀量有关, 所以无须在回归分析中引入。这样, 本文就以 I 部分的表 2 中所列的其余 7 项因子作自变量, 以其 47 组观测值作变量值进行回归分析处理。

3、回归分析的主要结果。通过分析得出的各项变量的平均值和相应变量的 B_i 值列于表 1; 计算步数、选取的变量和统计检验值列于表 2。从表 2 可见, 所选取的变量先后为: 径流深度

表1 各计算变量的平均值和 B_i 值

项目	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Y
平均值	45.62	11.14	22.10	9.29	8.06	0.37	0.58	4.82
计算的相应 B_i 值								
B_i	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_0
B_i 值	0	0	-0.257	0	1.148	0	5.064	-1.715

表2 分析步数、选取的变量和最终一步统计检验值

总分析步数	选取的变量	最终一步统计量检验值			
		复相关系数 R	R 的检验值	剩余标准差	F 统计量
3	X_6, X_3, X_7	0.934	98.03	2.38	25.40

(X_5), 当次降雨量 (X_3) 和侵蚀强度 (X_7)。从各检验值可见, 回归效果良好。根据表 1 中

的 B_i 值,可写出回归方程:

$$y = 5.064x_7 + 1.148x_5 - 0.257x_3 - 1.715$$

根据这一方程可以计算出侵蚀量预测值。表3列出了侵蚀量实测值(y)、计算预测值(\hat{y})和残差($y - \hat{y}$)值。从所列数值可以看出,预测效果是比较好的。

表3 土壤侵蚀量实测值、计算值和残差值

代 号	实测值 (y)	计算值 (\hat{y})	残差 ($y - \hat{y}$)
1	0.573	3.749	-3.179
2	4.437	4.951	-0.514
3	0.806	3.434	-2.628
⋮	⋮	⋮	⋮
45	4.270	1.529	2.741
46	1.060	-0.492	1.552
47	3.758	12.294	-3.536
Σ	226.373	226.376	

在上列回归方程中,引进了变量 x_7 ,它是侵蚀强度,即侵蚀量与径流深的比值。由于侵蚀量是待测值,所以侵蚀强度也不可能预先求出。但是由于侵蚀强度与降雨强度相关十分显著,因此就可以根据降雨强度来求出侵蚀强度,其方程: $y = 0.367 + 0.024x$ 。由于这样求取会产生一定的误差,因此就选用当次降雨量、降雨强度、径流深度和径流系数作自变量,进行分析处理,得出的各变量平均值、标准差和 B_i 值列于表4,选取变量数和统计检验值列于表5。

表4

项 目	x_3	x_4	x_5	x_6	y
平均 值	22.10	9.29	8.06	0.37	4.82
标 准 差	18.64	7.18	7.52	0.16	6.44
计算相应的 B_i 值					
B_i	B_3	B_4	B_5	B_6	B_0
B_i 值	-0.472	0.128	1.715	-14.041	5.459

表5

步 数	复相关系数 R	R的检验值	剩余标准差	F统计量	变 量 贡 献 值				选取的变量
					V_3	V_4	V_5	V_6	
1	0.719	48.08	4.530	48.08	0.112	0.111	0.517	0.323	x_5
2	0.893	86.49	2.967	69.89	0.281	0.081	-0.517	0.102	x_3
3	0.906	65.55	2.323	5.60	-0.280	0.003	-0.686	0.023	x_6
4	0.913	53.06	2.741	3.62	-0.199	-0.014	-0.167	-0.034	x_4

得出的方程是： $y = 5.459 - 0.472x_3 + 0.128x_4 + 1.715x_5 - 14.041x_6$ 。根据这一方程得出的侵蚀量计算值列于表6。

表6

代 号	实测值 (y)	计算值 (\hat{y})	残差 (y - \hat{y})
1	0.573	6.019	-5.446
2	4.437	1.767	2.670
8	0.806	5.172	-4.336
⋮	⋮	⋮	⋮
46	1.060	2.182	-1.122
47	8.758	12.517	-3.759
Σ	2226.373	226.376	

上述两式侵蚀量预测的回归方程，从其统计检验值看，前一方程稍优于后一方程。但由于前一方程引进了侵蚀强度，而侵蚀强度并非直接测出，计算时也会产生误差。所以总的评价是后一方程更有实用价值。而且从统计量检验值看，其回归效果也与前一方程相近。

三、结果和讨论

通过回归分析，当不引入侵蚀强度时，得到的回归方程是：

$$y = 5.459 - 0.472x_3 + 0.128x_4 + 1.715x_5 - 14.041x_6$$

其中的 x_3 — x_6 分别表示当次降雨量、降雨强度、径流深度和径流系数。这4项因子中的 x_3 、 x_6 则表示降雨、径流的容量因子， x_4 、 x_5 则反映降雨、径流的强度因子。这些因子中，径流系数除在降雨时直接测出外，也还可通过实验方法求出，这样只要知道降雨量和降雨强度，利用上列方程就可以预算出一定坡面土壤侵蚀量。因此在典型地区建立径流观测池取得资料，就可以建立一定流域区的坡面侵蚀方程，同样也可以利用气象、水文站取得的资料，建立一定流域范围的侵蚀方程。这对预测塘堰水库的淤塞、预报暴雨为害，使已取得的气象、水文资料更好为生产服务都有现实意义。此外对下述问题还必须说明：

1、上述仅对降雨和径流有关的资料进行处理，为了进一步提高回归方程的预报效果，还应该开展土壤水分和其它土壤性质的观测研究，再加上自然地理调查和景观资料，将会使工作更趋全面。

2、任何面上的资料都是要以点上的资料为依据的，因此观测区的选择，对取得具有面的代表性资料十分重要。以一个流域而论，首先该小流域在当地应有代表性，其次在选定流域内各观测点时，又要求能代表这一流域的各种类型，还可以在小流域外，设置一些辅助点，以取得代表性更广泛的资料。在选择观测区时，土壤流失的严重程度当然应首先考虑，土壤是侵蚀的对象，在主要土壤类型区也应设置观测点。

3、本工作仅对一个观测点的资料进行处理，得出的结果也仅对类似土壤类型区和类似地形、植被条件的地段有参考价值。但作为一种方法在水土保持研究中的应用则具有普遍意义。(1)通过多因子分析，找出各因子的相互关系，利用较易测出的一些因子值，来预报某些有实用价值而又难于测得的因子值；(2)从影响某一研究对象的众多因子中，抓出主要影响因子，从而使问题研究简化，即省人力和时间也有实用价值。