

非线性偏回归方程的建立与使用

广西计算中心 刘连芳 张正铀 乔中南

摘 要

文章通过对水能计算中使用逐步回归方法的情况的介绍,阐明了建立最佳多元非线性偏回归方程的一些方法,并介绍了一个用回归方程代替水利学万宁公式进行水能计算的实例。

一、概述

设计中的马骝滩电站所处地形比较复杂,必须采用长系列(三十年)水文资料对水能进行推算,以期使电站的设计较为合理。在对三十年的水能资料的处理中多次使用了逐步回归方法。

二、最佳多元非线性偏回归方程的建立

水能计算的关键是根据一年多的水文观测数据用逐步回归方法建立上游、下游各处流量、水位之间的数学表达式——多元非线性偏回归方程,以便推算三十年的水能数据,为选择装机方案提供可靠依据。

为节省程序设计工作量,凡属曲线拟合问题也采用逐步回归方法予以解决。所以课题中各种方法、各个方案总共需建立31个偏回归方程。方法如下:

1、函数类型的选择

要建立各个回归方程,首先要根据水利工作者的经验选择参加回归的自变量。本课题所涉及的回归方程最多两个自变量,所以自变量的选取并不困难。关键是自变量的函数类型不易确定。自变量与因变量之间的函数关系往往不是某种单一的函数关系,而是某些函数关系的组合。例如多项式与正弦、多项式与对数、多项式与倒数等等。

a、由于有许多函数在一定条件下可展开为多项式,所以多项式是最普遍适用的函数关系。在各回归方程的建立中,我们一般以多项式作为基础,然后再考虑其它函数的影响。

例1: 12500KW发电机效率 η 与净水头 H 的关系

$$\eta = 1.195127 - 0.2003597H + 0.02243H^2 - 0.07876 \cdot \sin(H) \quad (H \leq 6.3)$$

$$\rho = 0.99911$$

b、因变量有时也会以函数形式出现,所以在考虑自变量函数类型的同时,也需要考虑因变量的函数类型。

例2：坝后水位 Z_3 大于正常高水位时， Z_3 与毛水头HM的关系

$$\ln(HM) = -14.5585 - 2.1843 \ln(Z_3 - 21) + 2.0962 \ln(Q_3)$$

$$\rho = 0.9902658$$

究竟自变量和因变量的函数类型是什么，要根据散点图予以确定，并经过多次回归试验，逐步修正，最后找到一个最佳非线性偏回归方程。

例3：求糙率 N 与平均水位 H_{cp} 的关系

a、从散点图看，两者关系大致为倒数曲线，所建立方程为：

$$\frac{1}{N} = B_0 + \frac{B_1}{H_{cp}} \quad (1)$$

$$\rho = 0.780828$$

b、根据对图形的分析加入对数曲线的影响和 H_{cp} 项

$$\frac{1}{N} = B_0 + \frac{B_1}{H_{cp}} + B_2 \cdot \ln(H_{cp}) + B_3 \cdot H_{cp} \quad (2)$$

$$\rho = 0.96034437$$

c、提高 H_{cp} 的方次，对数项消失。

$$\frac{1}{N} = B_0 + \frac{B_1}{H_{cp}} + B_3 \cdot H_{cp} + B_4 \cdot H_{cp}^2 \quad (3)$$

$$\rho = 0.9603516$$

d、去掉对数项，再提高 H_{cp} 的方次

$$\frac{1}{N} = B_0 + \frac{B_1}{H_{cp}} + B_3 \cdot H_{cp} + B_4 \cdot H_{cp}^2 + B_5 \cdot H_{cp}^3 \quad (4)$$

$$\rho = 0.961828$$

e、提高 $\frac{1}{H_{cp}}$ 方次，进一步提高 H_{cp} 方次

$$\frac{1}{N} = B_0 + \frac{B_1}{H_{cp}} + \frac{B_2}{H_{cp}^2} + B_3 \cdot H_{cp}^2 + B_4 \cdot H_{cp}^3 + B_5 \cdot H_{cp}^4 \quad (5)$$

$$\rho = 0.961833$$

从(3)、(4)、(5)看来，再用提高方次来改进回归结果，效果就不理想了，这几步复相关系数均在0.96以上，说明方程已较好地反映了 N 及 H_{cp} 之间的关系，可以实用。所以试验到此结束。

方程(5)的复相关系数略大于方程(4)的，但按方程(5)进行计算要增加五次乘除、一次加减。此时，需根据实际需要选取方程。由于 N 、 H_{cp} 散点图上的点分布较散，回归结果各点残差相对来讲就稍大些，所以希望选择尽可能好的方程进行运算，就选了方程(5)作为推算 N 的数学表达式。

2、分段回归

为提高精度，对某些问题采用分段回归的方法，可以取得较好的效果。此时，在各段交接处选几个公共点参加前、后两段的回归，然后根据这几点的残差划分使用偏回归方程的范

围。这样做，一方面改善了总的回归效果，另一方面使各段之间过渡较为光滑，并尽量减少交点的残差。

例4：12500KW发电机效率曲线 $\eta=f(H)$ 的建立。

由于H取值在2.88—6.3与6.3—10范围时， η 的发展趋势明显不同，所以以6.2、6.3、6.35作为公共点分两段进行回归，就得到两个回归方程：

$$\eta = 1.195127 - 0.20036H + 0.02224H^2 - 0.07876\sin(H) \quad (H \leq 6.3)$$

$$\rho = 0.99911$$

$$\eta = -0.96612 - 0.56427H + 0.155134H^2 + 2.607781\ln(H) \quad (H > 6.3)$$

$$\rho = 0.99853$$

3、异常数据的取舍

异常数据的取舍主要根据工程技术方面的分析而不是统计学的规则。

a、超出正常值范围的数据不予参加回归。如糙率的正常值小于0.13，则凡大于0.13的数据均舍去。

b、着重对回归结果中残差异常大的点进行分析，确认为观测数据异常的则舍去，再对剩下数据进行回归。

课题中按上述方法建立的31个方程的复相关系数均在0.9618以上。

三、用偏回归方程取代水利学万宁公式

根据角咀水位求坝后水位，在水利学上应使用万宁公式

$$Q = \frac{1}{N} \overline{WR}^{2/3} \sqrt{I} \quad (1)$$

其中，Q：已知坝址流量

N：糙率， $N=f_1(Z_2, Z_3)$

\overline{W} ：平均过水断面面积

\overline{R} ：平均水利半径； $\overline{WR}^{2/3}=f_2(Z_2, Z_3)$

I：比降， $I=f_3(Z_2, Z_3)=\sqrt{\frac{|Z_3-Z_2|}{4000}}$

Z_2 ：已求出的角咀水位

Z_3 ：未知坝后水位

\overline{W} 、 \overline{R} 分别根据两张断面图积分求得。 f_1 、 f_2 、 f_3 均系用逐步回归方法建立的多项式形式的数学表达式。

总括以上诸关系，万宁公式最终可表达为：

$$Q = g_1(Z_2, Z_3) \quad (2)$$

(Z_2 、Q为已知， Z_3 为所求)

即，求出 \overline{W} 、 \overline{R} 、 f_1 、 f_2 、 f_3 之后，最终需解一个关于 Z_3 的高次方程。

但是从(2)式可以看出，Q、 Z_2 、 Z_3 之间直接存在着某种关系，现又有这三个量的409组观测数据，那么用逐步回归方法应该可以找出这种关系。

一旦找出这种关系，求 Z_3 就不必经过述上的N、 \overline{W} 、 \overline{R} 等等中间计算。又因为寻找这种

关系的目的在于推算 Z_3 ，为使用上的方便，就应以 Z_3 作为因变量来建立偏回归方程。

1、试建立偏回归方程

为验证上述思想，我们试建立了偏回归方程。

$$Z_3 = 0.29155 + 0.98893Z_2 + 5.54698E-05Q + 2.01617E-0.9Q^2 \quad (3)$$

$$\rho = 0.99879$$

其中 Q 、 Q^2 项系数仅为 10^{-5} 、 10^{-9} 数量级的数，说明 Q 对 Z_3 有影响，但影响较小。这与交通、水利部门平时使用的散点图一致。

2、方程(3)与万宁公式的对比

a、使用万宁公式

①求 f_2 ：需先求与409组 Z_2 、 Z_3 相应的 W_2 、 W_3 、 R_2 、 R_3 ，即共求818个面积和818个湿周，从而得到409个 $\bar{W}R^{2/3}$ ，用这409个数据与对应的 Z_2 、 Z_3 一起参加回归，确定 f_2 。

②求 f_1 ：将409组 $\bar{W}R^{2/3}$ 、 Z_2 、 Z_3 与相应的 Q 代入(1)，求出相应的 N ，再用这些 N 与 Z_2 、 Z_3 共同参加回归，确定 f_1 。

③将 f_1 、 f_2 、 f_3 代入(1)，用对分法解高次方程，求出 Z_3 。求一个 Z_3 平均需作211次乘除、81次加减。

b、使用偏回归方程(3)

只需进行一次逐步回归即可建立方程(3)。直接代入已知的 Q 、 Z_2 即可求得 Z_3 。推算一个 Z_3 仅需作7次乘除、4次加减，工作量是使用万宁公式的三十分之一。

用两种方法计算365个 Z_3 进行对比，结果证明方法b比方法a精度也有所提高。

		万 宁 公 式	偏 回 归 方 程
残	~0.05	80.821%	86.85%
	0.2~0.4	1.917%	0
差	max	0.799997598	0.725829743
	min	2.57559E-06	3.3736229E-07

• 残差单位：cm

综上所述及对比试验，说明使用偏回归方程(3)代替水利学万宁公式推算坝后水位 Z_3 不仅可行，而且大大节省工作量，计算精度也有所提高。在对三十年的 Z_3 的推算中，我们就使用了偏回归方程(3)

四、结语

逐步回归方程应用领域较多，凡需根据实验、观察、测量等得来的数据建立某些量的数学关系式的问题均可适用。

我们在马骊滩电站的水能计算中，初步摸索出一些用逐步回归方法建立最佳多元非线性偏回归方程的方法，并尝试用逐步回归方法建立的偏回归方程代替水利学的万宁公式，收到了良好的效果。为今后该方法的使用积累了点滴经验。