合金元素对低合金钢不同区带耐腐蚀性能影响的回归分析*

夏 伦 进
(北京工业大学计算机学院)

侯保荣 王 佳 张经磊 李红玲
(中国科学院海洋研究所)

某种合金元素与钢铁耐腐蚀性关系的研究世界上早有报道，但某几种合金元素同时对不同区带耐腐蚀性能影响的研究则较少。在本文中，作者利用回归分析方法研究合金元素与钢铁耐腐蚀性之间的关系，取得了初步成果。

一般说来，在浪花飞溅区和大气区，增加 Cu 的含量耐腐蚀性能可以得到提高，但这不是绝对的，因为对于耐腐蚀性能来说，某种合金元素的作用虽然明显，但还有其它因素在起作用，有些因素人们尚未认识，有些因素虽已认识，但诸因素之间还有一些互相的制约，这些偶然因素的综合作用造成了量与量之间的不确定性，这种关系即为相关关系。

对于相关性的这种自变量与因变量之间的不确定关系，我们首先可采用下式表达

$$Fs = f(Mn, P, Si, Cr, Mo, Al, Cu, V \cdots) \tag{1}$$

然后用回归分析的方法进行深入的研究。Fs 为变量，f 为函数，但变量不能用确定的函数式来表达。

研究相关关系的数学方法很多，但对于多元自变量最常用的是回归分析法，即通过得到的大量观察数据，寻找这些变量之间的统计规律性，采用回归分析的方法建立一个多元线性回归方程式，然后对所确定的方程式进行统计检验，最后进行预测，并给出预测精度。

一、方 程 式

回归是泛指变量之间的非确定性相互依赖关系，即统计相关关系，因为腐蚀速度是受M种合金元素的影响，所以腐蚀速度记作因变量 y，M种合金元素记作自变量 $x_1, x_2, \cdots x_M$。

N组观察数据为 $y_k, x_{k1}, x_{k2}, \cdots x_{kM} (k=1,2,\cdots N)$，其中 y_k 表示因变量 y 在第k次试验的腐蚀速度，x_{ki} 表示第 i 种合金元素在第 k 次试验时取的值。根据这些数据，在 y

---

* 中国科学院海洋海洋研究所调查研究报告第2424号。

收稿日期：1994年8月10日。

与 $x_1, x_2, \cdots x_M$ 之间配一线性回归方程

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_Mx_M \tag{2}$$

其中 b_0 是常数项，$b_i(i = 1,2,\cdots M)$ 分别为 y 对 $x_1, x_2, \cdots x_M$ 的回归系数，要求确定 $b_0, b_1, \cdots b_M$，使得总误差

$$Q = \sum_{k=1}^{N}(y_k - \hat{y}_k)^2 = \sum_{k=1}^{N}(y_k - (b_0 + b_1x_{k1} + \cdots + b_Mx_{kM}))^2 \tag{3}$$

达到极小，使上式达到极小的 $b_0, b_1, \cdots b_M$，记为 $\hat{b}_0, \hat{b}_1, \cdots \hat{b}_M$，它们称为回归系数的最小二乘估计。

通常求回归系数的最小二乘估计法是利用数学分析中多元函数求极值的方法。为了使回归系数更为准确，我们不采用这种方法，而采用"Householder 变换法"，其原理如下(郑宗成，王振堂，1985)：

令

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \\ b_0 \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \qquad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} & 1 \\ x_{21} & x_{22} & \cdots & x_{2M} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} & 1 \end{bmatrix}$$

其中 b 为由未知回归系数组成的向量，可用最小二乘估计法求 b，使

$$Q = \sum_{k=1}^{N}\varepsilon_k^2 = \|\varepsilon\|^2 = \|Xb - Y\|^2$$

达到极小。

构造 $N \times (M + 2)$ 矩阵

$$Z = (X \vdots Y) = \begin{bmatrix} x_{11} & \cdots & x_{1M} & 1 & y_1 \\ x_{21} & \cdots & x_{2M} & 1 & y_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N1} & \cdots & x_{NM} & 1 & y_N \end{bmatrix} \tag{4}$$

Householder 变换就是求一个 $N \times N$ 正交矩阵 P，使

$$PZ = S$$

其中 S 为 $N \times (M + 2)$ 上三角形矩阵

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M+1} & s_{1M+2} \\ 0 & s_{22} & \cdots & s_{2M+1} & s_{2M+2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & s_{M+1,M+1} & s_{M+1,M+2} \\ 0 & \cdots & & 0 & s_{M+2,M+2} \\ 0 & \cdots & & & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & & & 0 \end{bmatrix} \tag{5}$$

由于 $PP^{\tau} = P^{\tau}P = I$，$I$ 为单位矩阵，可以得出下式

$$
\begin{aligned}
Q &= \|\varepsilon\|^2 = \|Xb - Y\|^2 = (Xb - Y)^\tau (Xb - Y) \\
&= (Xb - Y)^\tau P^\tau P (Xb - Y) \\
&= [P(Xb - X)]^\tau [P(Xb - Y)] \\
&= \|P(Xb - Y)\|^2 = \|PXb - PY\|^2 \\
&= \left\| \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1M+1} \\ 0 & S_{22} & \cdots & S_{2M+1} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & & 0 & S_{M+1,M+1} \\ 0 & & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & & & 0 \end{bmatrix} b - \begin{bmatrix} S_{1M+2} \\ S_{2M+2} \\ \vdots \\ S_{M+2}, S_{M+2} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\|^2 \\
&= \left\| \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1M+1} \\ 0 & S_{22} & \cdots & S_{2M+1} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & & 0 & S_{M+1,M+1} \end{bmatrix} b - \begin{bmatrix} S_{1M+2} \\ S_{2M+2} \\ \vdots \\ S_{M+1}, S_{M+2} \end{bmatrix} \right\|^2 + S^2_{M+2,M+2}
\end{aligned} \tag{6}
$$

上式最后一个等式中,第二项 $S^2_{M+2,M+2}$ 与 b 无关,而第一项大于或等于零,所以极小值 Q 等价于求 b，使

$$
\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1M+1} \\ 0 & S_{22} & \cdots & S_{2M+1} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & & 0 & S_{M+1,M+1} \end{bmatrix} b = \begin{bmatrix} S_{1M+2} \\ S_{2M+2} \\ \vdots \\ S_{M+1}, S_{M+2} \end{bmatrix} \tag{7}
$$

$Q = S^2_{M+2,M+2}$，从方程组 (7) 可以解出回归系数 b 的最小二乘估计。正交矩阵 P 的求法是将式(4)确定的 Z 矩阵改记为

$$
Z = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1M+2} \\ & \cdots\cdots & & \\ Z_{N1} & Z_{N2} & \cdots & Z_{NM+2} \end{bmatrix}
$$

Z 的第一列记为 $Z_1 = (Z_{11}, Z_{21}, \cdots, Z_{N1})^\tau$，首先作 $N \times N$ 正交矩阵 P_1，使

$$
P_1 Z = \begin{bmatrix} Z^{(1)}_{11} & Z^{(1)}_{12} & \cdots & Z^{(1)}_{1M+2} \\ 0 & Z^{(1)}_{22} & \cdots & Z^{(1)}_{2M+2} \\ & \cdots\cdots & & \\ 0 & Z^{(1)}_{N2} & \cdots & Z^{(1)}_{NM+2} \end{bmatrix} \tag{8}
$$

为此令

$$
P_1 = I - 2U_1 U_1^\tau \tag{9}
$$

其中 $U_1 = (u_{11}, u_{12}, \cdots, u_{1N})^\tau$ 满足 $\|U_1\| = 1$，且使

$$
P_1 Z_1 = \begin{bmatrix} \beta \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{10}
$$

β 为非零常数,因为

$$P_1^{\tau}P_1 = (I - 2U_1U_1^{\tau})(I - 2U_1U_1^{\tau}) = I = P_1P_1^{\tau}$$

故 P_1 是正交矩阵,由

$$\beta^2 = \|P_1Z_1\|^2 = \|Z_1\|^2$$

$\beta = \pm\|Z_1\|$, 求向量 U_1 的表示式,由式(10)得出

$$(I - 2U_1U_1^{\tau})Z_1 = \pm\|Z_1\|e_1$$

其中 $e_1 = (1,0,\cdots,0)^{\tau}$, 从而

$$2U_1U_1^{\tau}Z_1 = Z_1 \pm \|Z_1\|e_1 \tag{11}$$

或

$$U_1 = (2U_1^{\tau}Z_1)^{-1}(Z_1 \pm \|Z_1\|e_1) \tag{12}$$

将式(11)两边各乘 $2Z_1^{\tau}$,得

$$(2U_1^{\tau}Z_1)^2 = 2\|Z_1\|^2 \pm 2\|Z_1\|Z_1^{\tau}e_1 = 2\|Z_1\|^2 \pm 2\|Z_1\|Z_{11}$$

从而

$$2U_1^{\tau}Z_1 = (2\|Z_1\|^2 \pm 2\|Z_1\|Z_{11})^{1/2} \tag{13}$$

将式(13)代入式(12)得

$$U_1 = (2\|Z_1\|^2 \pm 2\|Z_1\|Z_{11})^{-1/2}(Z_1 \pm \|Z_1\|e_1)$$

为避免上式右边同号相减损失有效数字,取 $\pm Z_{11}$ 前的符号,使 $\pm Z_{11}$ 为正数,从而

$$U_1 = (2\|Z_1\|^2 + 2\|Z_1\||Z_{11}|)^{-1/2}(Z_1 + \|Z_1\|e_1\mathrm{sign}Z_{11}) \tag{14}$$

其中

$$\mathrm{sign}Z_{11} = \begin{cases} 1 & Z_{11} \geqslant 0 \\ -1 & Z_{11} < 0 \end{cases}$$

易证 $\|U_1\| = 1$, 将式(14)代入式(9)即可满足式(8)的正交矩阵 P_1

$$P_1 = I - (\|Z_1\|^2 + \|Z_1\||Z_{11}|)^{-1}(Z_1 + \|Z_1\|e_1\mathrm{sign}Z_{11}) \cdot (Z_1 + \|Z_1\|e_1\mathrm{sign}Z_{11})^{\tau}$$

P_1Z 的第一列为 $(-\|Z_1\|\mathrm{sign}Z_{11},0,\cdots 0)^{\tau}$

在实际计算中,注意

$$P_1Z = Z - 2U_1U_1^{\tau}Z$$

$U_1^{\tau}Z$ 是行向量, $U_1U_1^{\tau}Z$ 矩阵的第 j 列等于 U_1 乘以 $U_1^{\tau}Z$ 的第 j 个元素, 由此可求出 P_1Z 的第 j 列 $(j = 2,3,\cdots,M+2)$

其次,作$N \times N$正交矩阵 P_2,使

$$P_2P_1Z = \begin{bmatrix} Z_{11}^{(2)} & Z_{12}^{(2)} & Z_{13}^{(2)} & \cdots & Z_{1M+2}^{(2)} \\ 0 & Z_{22}^{(2)} & Z_{23}^{(2)} & \cdots & Z_{2M+2}^{(2)} \\ 0 & 0 & Z_{33}^{(2)} & \cdots & Z_{3M+2}^{(2)} \\ & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & Z_{N3}^{(2)} & \cdots & Z_{NM+2}^{(2)} \end{bmatrix} \tag{15}$$

一般地设经过正交变换 $P_1,\cdots,P_r$, 使

$$P_rP_{r-1}\cdots P_1Z = \begin{pmatrix} R_r & C \\ 0 & D \end{pmatrix} \tag{16}$$

其中 R_r 是 r 阶上三角形矩阵，令

$$P_{r+1} = I - 2U_{r+1}U_{r+1}^{\tau}$$

其中 $U_{r+1} = (\underbrace{0,\cdots,0}_{r},U_{r+1,r+1}\cdots U_{r+1,N})^{\tau}$，满足 $\|U_{r+1}\| = 1$

令

$$\widetilde{U}_{r+1} = (U_{r+1,r+1}\cdots U_{r+1,N})^{\tau}$$

那么 P_{r+1} 可表示为

$$P_{r+1} = \begin{pmatrix} I_r & 0 \\ 0 & I_{N-r} - 2\widetilde{U}_{r+1}\widetilde{U}_{r+1}^{\tau} \end{pmatrix}$$

其中 I_r, I_{N-r} 分别表示 r 阶和 N－r 阶单位矩阵

$$P_{r+1}P_r\cdots P_1Z = \begin{pmatrix} R_r & C \\ 0 & (I_{N-r} - 2\widetilde{U}_{r+1}\widetilde{U}_{r+1}^{\tau})D \end{pmatrix}$$

由于 $\|\widetilde{U}_{r+1}\| = \|U_{r+1}\| = 1$，和前面讨论的一样，我们可以选 $\widetilde{U}_{r+1}$，使

$$(I_{N-r} - 2\widetilde{U}_{r+1}\widetilde{U}_{r+1})D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1M+2-r} \\ 0 & d_{22} & \cdots & d_{2M+2-r} \\ & \cdots\cdots & \cdots\cdots & \\ 0 & d_{N-r,2} & \cdots & d_{N-r,M+2-r} \end{bmatrix}$$

这样就作出一系列$N \times N$正交变换 $P_1,\cdots,P_{M+1},P_{M+2}$ 使

$$P_{M+2}P_{M+1}\cdots P_1Z = S$$

其中S的形式如式(5)所示，且所求的正交矩阵P为

$$P = P_{M+2}P_{M+1}\cdots P_1$$

上面叙述的就是 Householder 变换的一种具体求法，在下面所编制的程序中采用。

二、统计检验

在实际问题中，事先我们并不知道腐蚀速度 y 与合金元素 $x_1,x_2,\cdots,x_M$ 之间是否存在线性关系。在计算了多元线性回归方程后，为了把它用于实际中，还必须进一步进行数学检验。

y 的离差平方和

$$S_{YY} = \sum_{k=1}^{N}(y_k - \bar{y})^2 = \sum_{k=1}^{N} y_k^2 - \frac{1}{N}\left(\sum_{k=1}^{N} y_k\right)^2$$

回归平方和

$$U = \sum_{k=1}^{N}(\hat{y}_k - \bar{y})^2$$

剩余平方和

$$Q = \sum_{k=1}^{N}(y_k - \hat{y}_k)^2$$

其中 $\bar{y}=\frac{1}{N}\left(\sum_{k=1}^{N}y_k\right)$，而

$$\hat{y}_k=\hat{b}_0+\hat{b}_1x_{1M}+\hat{b}_2x_{2M}+\cdots+\hat{b}_Mx_{kM}$$

R检验即检验回归方程是否线性相关

$$R^2=1-\frac{\sum_{k=1}^{N}(y_k-\hat{y}_k)^2}{\sum_{k=1}^{N}(\hat{y}_k-\bar{y})^2}=1-\frac{Q}{S_{YY}} \tag{17}$$

由公式(15)可见，如果自变量与因变量接近线性相关，那么每个估计值 $\hat{y}_k$ 与相应的实测值 y_k 都很接近，此时 $Q\approx 0, R^2\approx 1$。相反，如果自变量和因变量线性无关，可以证明 $\hat{y}_k\approx\bar{y}_k, Q\approx S_{YY}$，此时 $R^2\approx 0$，由此可见 $0\leqslant R^2\leqslant 1$。

当 R^2 越接近1，表明 y 与 $x_1,x_2,\cdots,x_M$ 的线性关系越密切。反之，表明该回归模型的使用价值很小。

F检验是检验回归方程的显著性

$$F_{M,N-M-1}=\frac{U/M}{Q/(N-M-1)} \tag{18}$$

计算出F的值后，利用F表可以进行检验，一般查置信度 α 为 0.55 时，M与 $N-M-1$ 所对应的 F_α 值若 $F\geqslant F_\alpha$ 时，则认为回归方程是显著的(张寿，于清文，1984)。

三、实际应用

由于回归方程式是由观测数据得到的，为了使回归方程较为准确，我们对已收集到的观测数据做了一些选择，共挑出 18 组数据，进行回归分析。

由 18 组测试数据和 8 种合金元素，在计算机上通过编制的程序运行后，得到的大气回归方程为

$$\begin{aligned}\hat{y}=0.2281&-0.0509x_1-0.2193x_2-0.0336x_3+0.0365x_4\\&-0.1548x_5-0.0976x_6-0.1787x_7+0.0567x_8\end{aligned} \tag{19}$$

程序在运行的同时对回归方程式进行R与F检验。

相应的 $R=0.8720364$

$F=3.57126$

给定显著水平 $\alpha=0.05$，查F分布数值表，得 $F_\alpha=3.23$，由于 $F=3.57>3.23=F_\alpha$，我们认为 $\hat{y}$ 与 x_i 的线性关系密切。

我们使用 18 组测试数据和 8 种合金元素(侯保荣等，1995)在计算机上通过编制的程序运行后，得到的浪花飞溅区回归方程为：

$$\begin{aligned}\hat{y}=0.561&-0.0855x_1-0.5214x_2+0.0235x_3+0.0751x_4\\&-0.02665x_5-0.0185x_6-0.2051x_7+0.0561x_8\end{aligned} \tag{20}$$

相应的 $R=0.9175386$

$F=5.989719$

给定显著水平 $\alpha = 0.05$,查 F 分布数值表,得 $F_\alpha = 3.23$,由于 $F = 5.99 > 3.23 = F_\alpha$，我们认为 $\hat{y}$ 与 x_i 的线性关系密切。

由以上 18 组测试数据和 8 种合金元素,在计算机上通过编制的程序运行后，得到的全浸区回归方程为

$$\begin{aligned}\hat{y} = 0.23 - 0.0087x_1 + 0.0047x_2 + 0.0494x_3 + 0.002x_4 \\ - 0.0467x_5 - 0.1437x_6 - 0.1117x_7 - 0.0578x_8\end{aligned} \tag{21}$$

相应的 $R = 0.9707256$

$F = 18.37525$

给定显著水平 $\alpha = 0.05$,查 F 分布数值表,得 $F_\alpha = 3.23$,由于$F = 18.37 > 3.23 = F_\alpha$，我们认为 $\hat{y}$ 与 x_i 的线性关系密切。

四、预　　测

回归方程式不仅为腐蚀速度关于合金元素变量之间依存关系的定性叙述提供了定量资料,还提供了预测的模拟控制,并给出预测精度估计,以下将结合具体观测数据,进一步说明。

对于得到的回归方程式,我们亦利用其它 6 个钢种的实验数据进行了验证,这些钢种的实验条件与上述 18 种材料相同,但合金成分不一样,实验时间也不一样,其实验结果与预测结果列于表 1。

表 1　不同钢种腐蚀速度试验值与预测值的比较

编号	钢　种	元素含量及相应的自变量								大气区		飞溅区		全浸区	
		x_1 Mn	x_2 P	x_3 Si	x_4 Cr	x_5 Mo	x_6 Al	x_7 Cu	x_8 V	y	$\hat{y}$	y	$\hat{y}$	y	$\hat{y}$
1	10CrCuVSi	0.48	0.014	0.58	0.56			0.3	0.032	0.13	0.15	0.28	0.5	0.08	0.22
2	$10PV_{(2)}$	0.54	0.126	0.31					0.79	0.17	0.2	0.48	0.5	0.19	0.19
3	10MoPV	0.89	0.104	0.37		0.12	0.074		0.075	0.06	0.12	0.21	0.41	0.23	0.22
4	14NbPAl	0.67	0.029	0.23			0.25			0.09	0.15	0.33	0.49	0.22	0.2
5	$10Cr_2AlCu$	0.64	0.014	0.34				0.51		0.1	0.09	0.59	0.4	0.08	0.18
6	2#	0.45	0.035	1.5	1			0.5		0.06	0.09	0.43	0.51	0.23	0.25

表 1 中 y 为试验中所测得的腐蚀速度，$\hat{y}$ 为利用上述回归方程式预测的腐蚀速度。关于预测的精度,以大气的回归方程式为例,通过回归方程式的计算所得到的腐蚀速度 $\hat{y}$ 可以达到实验中所测得的腐蚀速度 y 的 76%;在浪花飞溅区,平均可以达到 83%,在海水全浸区,平均可以达到 82% 的精度 (S. C. 惠尔莱特等,1986,門智,渡辺常安，1976)。

从验证结果还可以看出,试验钢种中所含元素不同结果也有差别,例如 3,4 号精度较差。作者今后的研究中将利用更多的试验结果进行预测，以求达到更加精确的预测方程式。

五、结 语

通过计算，初步得到了表示合金元素与钢铁耐腐蚀性能相关关系的回归方程式。大气回归方程式为

$$\hat{y} = 0.2281 - 0.0509x_1 - 0.2193x_2 - 0.0336x_3 + 0.00365x_4 - 0.1548x_5 - 0.0976x_6 - 0.1789x_7 + 0.0567x_8$$

浪花飞溅区回归方程式为

$$\hat{y} = 0.5613 - 0.0855x_1 - 0.5214x_2 + 0.0235x_3 + 0.0751x_4 - 0.2665x_5 - 0.0185x_6 - 0.2051x_7 + 0.0561x_8$$

海水全浸区的回归方程式为

$$\hat{y} = 0.2297 - 0.0087x_1 + 0.0047x_2 + 0.0494x_3 + 0.002x_4 - 0.0467x_5 - 0.1437x_6 - 0.1117x_7 + 0.0578x_8$$

参 考 文 献

郑宗成、王振堂，1984，实用预测方法 BASIC 程序库，中山大学出版社，67—74。
张寿、于清文，1984，计量经济学，上海交通大学出版，338—340。
惠尔莱特，S. C. and S. 马克利达基斯，1985，管理用预测方法，上海人民出版社，175—178。
侯保荣等，1995，合金元素对低合金钢耐腐蚀性能影响的研究，海洋科学集刊，**36**:137—144。
門智、渡辺常安，1976，低合金鋼の海水腐食，防食技術，**25**(3)：173—190。
高村昭、荒川要、藤原和雄，1970，海水飛沫帯における钢の耐食性に及ぼす合金元素の影响について，防食技術，**18**：18—25。

REGRESSION ANALYSIS FOR THE EFFECT OF ALLOY ELEMENTS ON THE CORROSION RESISTANCE OF LOW ALLOY STEELS IN DIFFERENT SITS OF MARINE ENVIRONMENT*

Xia Lunjin

(Beijing Polytechnical University)

Hou Baorong, Wang Jia, Zhang Jinglei and Li Hongling

(Institute of Oceanology, Chinese Academy of Sciences)

ABSTRACT

Three regression equations describing the effects of alloy elements on the corrosion resistance of low alloy steels for marine environment were derived using Householder Transform Method. F tests indicated the linear relationship between the corrosion rate of steels and the amount of alloy elements added. Experimental results showed the regression equations could be reliably applied for evaluating the corrosion behaviours of low alloy steels in marine environment.

* Contribution No. 2424 from Institute of Oceanology, Chinese Academy of Sciences.