

# 多层线性模型在碳钢土壤腐蚀规律中的应用研究

鲁 庆, 穆志纯✉

北京科技大学自动化学院, 北京 100083

✉ 通信作者, E-mail: mu@ies.ustb.edu.cn

**摘 要** 针对材料在自然土壤环境中的腐蚀预测问题, 提出应用多层线性模型对材料的腐蚀规律进行研究. 多层线性模型是分析具有层次结构数据的新型统计技术, 对于地区间土壤腐蚀规律的差异性特点和单地区实验观测样本不足的问题, 可以为不同的区域分别建立腐蚀率模型, 使模型假设与实际更为吻合. 以碳钢在土壤中的腐蚀数据为研究对象, 建立了腐蚀率的多层线性模型. 实验验证了该模型可以准确地拟合和预测碳钢在土壤的腐蚀率变化, 优于指数平滑算法和差分自回归移动平均算法.

**关键词** 碳钢; 钢腐蚀; 土壤; 腐蚀率; 多层线性模型

**分类号** TG172.4; TP391.9

## Application study of hierarchical linear models in carbon steel corrosion in soils

LU Qing, MU Zhi-chun✉

School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

✉ Corresponding author, E-mail: mu@ies.ustb.edu.cn

**ABSTRACT** Aiming at the problem of materials corrosion prediction in the natural soil environment, hierarchical linear modeling was used to study the corrosion rules of materials. The hierarchical linear model is a new statistical technique for hierarchical data analysis. Different models for every region can be established to fit regional soil corrosion characteristics and solve the problem of small samples in a specific region, which makes modeling assumptions more consistent with actual situations. A hierarchical linear corrosion model was built with carbon steel corrosion data in soils being the object of study. The test demonstrates that the model can precisely describe and predict the corrosion rate of carbon steel in soils, and it is superior to exponential smoothing algorithm and differential autoregressive integrated moving average (ARIMA) algorithm.

**KEY WORDS** carbon steel; steel corrosion; soils; corrosion rate; hierarchical linear model

材料在土壤环境的腐蚀规律是自然环境腐蚀研究的一个重要内容, 对地下工程的施工选材防护和材料寿命评估等具有重要意义. 近年来, 国内外研究机构和学者利用多元回归<sup>[1]</sup>、时间序列分析<sup>[2]</sup>、神经网络<sup>[3-4]</sup>、灰色系统<sup>[5]</sup>等方法建立各种腐蚀规律模型, 取得了很多成果和进展.

由于受到各环境因素相互影响, 不同地区的土壤腐蚀规律具有很大的相异性<sup>[6]</sup>. 要准确描述并预测材料在土壤中的腐蚀变化, 一个比较合理的策略

是为每个地区建立特有的土壤腐蚀性模型. 但是, 由于自然环境实验的困难性, 实验观测数据往往存在高维度、小样本等问题, 具体到本文研究的所用数据, 每个地区只有 2~3 组观测数据, 很难满足多元回归、时间序列等方法的建模数据需求. 神经网络方法在解释土壤腐蚀规律影响因素方面也存在局限性.

本文应用多层线性模型的原理, 以碳钢为例, 把材料在土壤中的腐蚀变化过程用一个两层分层模

收稿日期: 2012-10-21

型进行描述, 将对每个地区的累次观察视为归属于该地区, 尝试建立单独的地区腐蚀率模型, 较好地拟合土壤腐蚀规律的地区性特征, 并在一定程度上解决小样本等问题。

## 1 多层线性模型

多层线性模型最早应用在社会学领域, 由 Lindley 和 Smith<sup>[7]</sup> 在 1972 年提出。与传统的回归模型比起来, 它不受方差齐性和样本独立性等传统统计假设的严格限制, 在分析具有层级结构数据方面有很大的优越性<sup>[8]</sup>, 在心理、经济、教育等多个学科领域<sup>[9-11]</sup> 得到广泛应用, 但在自然科学研究领域还没有得到应有的重视。

### 1.1 数学模型

以二层模型为例, 当利用多层线性模型描述个体变化时, 数学模型可以表示如下。

第一层模型:

$$Y_{ti} = \beta_{0i} + \beta_{1i}\alpha_{ti} + r_{ti}. \quad (1)$$

第二层模型:

$$\beta_{0i} = \gamma_{00} + W_i\gamma_{01} + \mu_{0i}; \quad (2)$$

$$\beta_{1i} = \gamma_{10} + W_i\gamma_{11} + \mu_{1i}.$$

红色高亮为已知量

第一层模型描述了个体的增长变化, 具体到本文的研究对象, 为特定地区的土壤腐蚀变化规律, 其中  $Y_{ti}$  表示  $i$  地区在  $t$  时点的土壤腐蚀率;  $\alpha_{ti}$  为时间变量, 表示在时点  $t$  的材料埋存时间;  $\beta_{0i}$  和  $\beta_{1i}$  表示  $i$  地区腐蚀率模型的截距和斜率;  $r_{ti}$  为随机误差,  $r_{ti} \sim N(0, \delta^2)$ 。第一层模型看上去与线性回归模型相似, 但截距  $\beta_{0i}$  和斜率  $\beta_{1i}$  不再是常数, 而是依赖于第二层模型的定义。在第二层模型中,  $W_i$  为第二层预测变量, 表示不同地区的环境参数;  $\gamma_{00}$ 、 $\gamma_{01}$ 、 $\gamma_{10}$  和  $\gamma_{11}$  分别表示对应二层模型的截距和斜率;  $\mu_{qi} \sim N(0, T)$ , 为地区间偏差;  $T$  为方差协方差矩阵; 通过对  $\gamma$ 、 $\beta$  和各层方差协方差  $\delta$ 、 $T$  进行参数估计, 就可以求得单独的地区腐蚀模型。

将式 (2) 合并式 (1), 可得到混合方程:

$$Y_{ti} = \gamma_{00} + W_i\gamma_{01} + \alpha_{ti}\gamma_{10} + \alpha_{ti}W_i\gamma_{11} + \mu_{0i} + \alpha_{ti}\mu_{1i} + \gamma_{ti}. \quad (3)$$

在式 (3) 中,  $\mu_{0i} + \alpha_{ti}\mu_{1i} + \gamma_{ti}$  是残差项, 其中偏差  $\mu_{0i}$  和  $\mu_{1i}$  对于同一地区是分别相等的, 在不同的地区间则存在差异, 只有当不存在组间偏差, 例如  $\mu_{0i}$  和  $\mu_{1i}$  为 0 时, 方程 (3) 才能满足最小二乘 (OLS) 回归方法关于方差齐次性的假定。在此可以

看出, 对于具有层级结构的数据, 如果使用传统的回归分析参数估计算法, 得到的参数估计将是 biased 的。

### 1.2 参数估计

多层线性模型利用经验贝叶斯方法, 对不同层次的参数估值进行加权, 充分利用数据中包含的信息, 求得最大似然估值。以式 (1) 和式 (2) 所示模型为例, 其参数估计过程如下。

首先, 对式 (1) 进行最小二乘 (OLS) 回归方法估值, 得到  $\beta$  的第一组估值:

$$\begin{cases} \hat{\beta}_{0i} = \bar{Y} - \hat{\beta}_{1i}\bar{\alpha}_{ti}; \\ \hat{\beta}_{1i} = \sum (\alpha_{ti} - \bar{\alpha}_{ti})(Y_{ti} - \bar{Y}_{ti}) / \sum (\alpha_{ti} - \bar{\alpha}_{ti})^2. \end{cases} \quad (4)$$

然后, 利用广义最小二乘法 (GLS) 估计固定效应  $\gamma$ 。

定义精度  $\Delta$ :

$$\begin{cases} \Delta_{0i}^{-1} = \text{Var}(\mu_{0i} + (\hat{\beta}_{0i} - \beta_{0i})) = \tau_{00} + k_{0i}; \\ \Delta_{1i}^{-1} = \text{Var}(\mu_{1i} + (\hat{\beta}_{1i} - \beta_{1i})) = \tau_{11} + k_{1i}. \end{cases} \quad (5)$$

式中:  $\tau_{00} = \text{Var}(\mu_{0i})$ , 是随机误差  $\mu_{0i}$  的方差;  $\tau_{11} = \text{Var}(\mu_{1i})$ , 是随机误差  $\mu_{1i}$  的方差;  $k_{0i} = \text{Var}(\hat{\beta}_{0i} - \beta_{0i})$ ,  $k_{1i} = \text{Var}(\hat{\beta}_{1i} - \beta_{1i})$ , 分别表示由于估计值误差引起的差异。可得到  $\gamma$  的估值:

$$\begin{cases} \hat{\gamma}_{00} = \sum \Delta_{0i} \bar{Y}_{ti} / \sum \Delta_{0i}; \\ \hat{\gamma}_{01} = \sum \Delta_{0i} W_i \hat{\beta}_{0i} / \sum \Delta_{0i} W_i^2; \\ \hat{\gamma}_{10} = \sum \Delta_{1i} \bar{Y}_{ti} / \sum \Delta_{1i}; \\ \hat{\gamma}_{11} = \sum \Delta_{1i} W_i \hat{\beta}_{1i} / \sum \Delta_{1i} W_i^2. \end{cases} \quad (6)$$

由式 (2) 可得到参数  $\beta$  的第二组估值:

$$\begin{cases} \hat{\beta}_{0i}^* = \hat{\gamma}_{00} + \hat{\gamma}_{01} W_i; \\ \hat{\beta}_{1i}^* = \hat{\gamma}_{10} + \hat{\gamma}_{11} W_i. \end{cases} \quad (7)$$

多层线性模型用经验贝叶斯 (EB) 算法对参数  $\beta$  的两组估值进行加权, 最终估值为

$$\begin{cases} \hat{\beta}_{0i}^{**} = \lambda_{0i} \hat{\beta}_{0i} + (1 - \lambda_{0i}) \hat{\beta}_{0i}^*; \\ \hat{\beta}_{1i}^{**} = \lambda_{1i} \hat{\beta}_{1i} + (1 - \lambda_{1i}) \hat{\beta}_{1i}^*. \end{cases} \quad (8)$$

其中  $\lambda$  为估值信度, 定义为:

$$\begin{cases} \lambda_{0i} = \tau_{00} / (\tau_{00} + k_{0i}); \\ \lambda_{1i} = \tau_{11} / (\tau_{11} + k_{1i}). \end{cases}$$

在本文的实验中, 土壤腐蚀台站的单站数据只有 2~3 组观测数据, 按照传统方法, 无法进行可靠的统计分析。如将各台站数据作为一个整体样本进行建模, 又会因为无法满足样本独立同分布假定而造成估值有偏。从式 (8) 可以看出, 当估值可靠性  $\lambda$  较高时,  $\hat{\beta}_{0i}$  在最终结果中将占据较大的比重; 如

果可靠性  $\lambda$  较低,  $\hat{\beta}_{0i}^*$  的估值将在更大程度上影响最终结果. 多层线性模型通过这种加权的方式, 可以借助其他二层单位的样本数据中包含的信息, 对取样很少的单位也可以进行分析建模, 相对解决了小样本问题.

为便于说明, 上述的计算过程假设各层方差协方差参数已知, 在实际计算中, 需要对方差、协方差参数  $\delta$  和  $T$  进行估计. 通常使用最大似然估计 (ML)、约束最大似然估计 (REML)、贝叶斯估计等估计方法. 以约束最大似然估计为例, 其原理就是通过选择参数  $\delta$  和  $T$  的估计值, 应用 EM 算法、费舍尔得分算法等, 在迭代过程中使观测到这一实际数据  $Y$  的似然值最大并趋于稳定. 限于篇幅, 在这里就不再展开讨论, 具体讨论可参见文献 [12-13].

### 1.3 模型检验

可以通过单参数检验对固定效应参数  $\gamma$  进行假设检验, 以考察  $\gamma$  是否具有显著效应 [14]. 对于任意  $\gamma_{qs}$ , 单参数检验的假设为  $H_0: \hat{\gamma}_{qs} = 0$ , 检验统计量  $t = \hat{\gamma}_{qs} / (V_{\hat{\gamma}_{qs}})^{1/2}$ , 服从自由度  $J-N-1$  的  $T$  分布, 其中  $\hat{\gamma}_{qs}$  是  $\gamma_{qs}$  的估值,  $(V_{\hat{\gamma}_{qs}})^{1/2}$  是  $\hat{\gamma}_{qs}$  的标准误差,  $J$  为第二层模型的样本数,  $N$  为第二层模型的自变量个数. 如假设成立, 则表示  $\gamma_{qs}$  对应的因子对预测因变量  $\beta$  无意义.

随机效应参数  $\beta$  的假设检验方式与固定效应检验方式类似. 对于任意  $\beta_{qj}$ , 单参数检验的假设为  $H_0: \hat{\beta}_{qj} = 0$ , 检验统计量  $t = \hat{\beta}_{qj} / (V_{\hat{\beta}_{qj}})^{1/2}$ , 服

从自由度  $J-N-1$  的  $T$  分布, 其中  $\hat{\beta}_{qj}$  是  $\beta_{qj}$  的估值,  $(V_{\hat{\beta}_{qj}})^{1/2}$  是  $\hat{\beta}_{qj}$  的标准误差,  $J$  为第一层模型的样本数,  $N$  为第一层模型的自变量个数. 如假设成立, 则表示  $\beta_{qj}$  对应的因子与因变量  $Y$  无关系.

## 2 结果与讨论

### 2.1 研究样本和资料库

本文以国家材料腐蚀站网中碳钢在土壤中的腐蚀数据进行分析 and 建模, 对不同土壤环境因素下碳钢平均腐蚀率进行预测. 按照多层模型分析的需要, 将原始数据整理为两层数据表, 第一层数据如表 1 所示, 包含了实样站点、埋样时间和实验材料的腐蚀率记录共 37 条数据; 第二层数据如表 2 所示, 描述了各试样站的土壤环境数据, 包括土壤 pH 值、有机质含量、全氮含量等理化参数, 共 21 条数据, 涵盖了 21 个地区试验台站. 考虑到数据的可比性, 对表 2 数据进行了 Z-SCORE 标准化处理.

表 1 不同区域碳钢的腐蚀率数据

Table 1 Corrosion rate of carbon steel at different sites

站点	埋存时间/a	腐蚀率/(mm·a <sup>-1</sup> )
沈阳站	1	3.719
	3	3.215
成都站	1	4.155
	3	2.343
西安站	1	4.363
	3	2.912
	5	2.120

表 2 站点土壤理化性质数据示例

Table 2 Physical and chemical properties of soils at test sites

站点	pH 值	有机质质量分数/%	全氮含量(质量分数)/%	可溶性盐离子的质量分数/%					土壤电阻/Ω
				HCO <sub>3</sub> <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	Mg <sup>+</sup>	K <sup>+</sup>	Na <sup>+</sup>	
大庆站	10.0500	1.0250	0.0525	0.0669	0.0142	0.0008	0.0033	0.0516	5.1000
华南站	5.4500	0.3250	0.0140	0.0031	0.0049	0.0006	0.0002	0.0007	460.0000
沈阳站	6.7500	2.3800	0.1090	0.0126	0.0118	0.0017	0.0004	0.0049	32.9000

### 2.2 建立模型

#### 2.2.1 零模型

为考察不同地区土壤腐蚀率是否存在明显差异, 首先建立零模型如下.

第一层模型:

$$Y_{ti} = \beta_{0i} + r_{ti}. \quad (9)$$

第二层模型:

$$\beta_{0i} = \gamma_{00} + \mu_{0i}.$$

参数估计结果如表 3 所示. 考察随机部分参数估计结果可以看出, 不同地区站点的腐蚀规律存

在着显著变异: 方差 = 5.563,  $\chi^2 = 179.7, P < 0.001$ . 这说明如果为所有地区建立一个统一的模型, 其参数估计结果是有偏的, 即利用多层线性模型进行建模是必要的. 可以通过计算跨级相关系数度量变异分布,

$$\begin{aligned} \text{ICC} &= \text{Var}(\mu_{0i}) / [\text{Var}(\mu_{0i}) + \text{Var}(r_{ti})] = \\ &= 5.563 / (5.563 + 1.012) = 0.861, \end{aligned}$$

即由不同地区间特性导致的变异占总变异的比例为 86.1%.

表 3 碳钢腐蚀率零模型参数估计结果

Table 3 Parameter estimation results of the zero model of carbon steel corrosion rate

固定效应	系数	标准差	$T$ 值	$P$ 值
$\gamma_{00}$	3.211313	0.531361	6.044	0.000
随机效应	标准误差	方差	$\chi^2$	$P$ 值
$\mu_{0i}$	2.35862	5.56307	179.71340	0.000
$r_{ti}$	1.00615	1.01233	—	—

### 2.2.2 基本模型

为了解碳钢材料在土壤中的腐蚀规律是否具有线性变化趋势,并且不同地区间其变化趋势是否存在差异,在零模型的基础上,建立基本模型如下.

第一层模型:

$$Y_{ti} = \beta_{0i} + \beta_{1i}(\alpha_{ti}) + r_{ti}. \quad (10)$$

第二层模型:

$$\beta_{0i} = \gamma_{00} + \mu_{0i};$$

$$\beta_{1i} = \gamma_{10}.$$

基本模型的第一层模型定义了碳钢在不同地区腐蚀率随时间的变化趋势,第二层模型定义了不同地区腐蚀率变化参数截距和斜率的变化. 在第二层模型,当定义  $\beta_{1i}$  为随机变量时(即 $\beta_{1i} = \gamma_{10} + \mu_{1i}$ ),  $\beta_{0i}$  和  $\beta_{1i}$  的相关系数为 0.925,为消除共线

性因素,保持模型的稳定性,经过对比多种模型,本文采用了约束随机效应  $\gamma_{1i}$  为 0 的策略. 为确定约束是否恰当,对两种模型进行了比较.  $\beta_{0i}$  和  $\beta_{1i}$  均为随机变量时,模型的偏差度为 139.93,自由度为 4; 相应模型(式(10))的偏差度为 142.18,自由度为 2. 两个模型偏差度统计量差值为 2.25,这个差值服从自由度为 2 的  $\chi^2$  分布,检验结果不显著( $P=0.325$ ),即模型的简化是可行的,因此最终确定了目前第二层模型的形式.

不同地区土壤环境碳钢腐蚀率基本模型的参数估计如表 4 所示. 固定效应的估计结果表明,腐蚀率初始状态平均值为 4.766 ( $T=7.716$ ),随着埋存时间的增加,腐蚀率有显著的线性下降趋势,下降速度为 0.134 ( $\beta_{10} = \gamma_{10} = -0.134$ ,  $T = -3.942$ ). 由于截距和增长率的  $t$  检验值均显著,表明这两个参数在描述平均增长轨迹时是必要的.

表 4 碳钢腐蚀率基本模型参数估计结果

Table 4 Parameter estimation results of basic model of carbon steel corrosion rate

固定效应	系数	标准差	$T$ 值	$P$ 值
$\gamma_{00}$	4.765820	0.617623	7.716	0.000
$\gamma_{10}$	-0.134149	0.034030	-3.942	0.001
随机效应	标准误差	方差	$\chi^2$	$P$ 值
$\mu_{0i}$	2.07276	4.29633	217.40681	0.000
$r_{ti}$	0.81737	0.66809	—	—

在目前的实验数据中,由于每个站点观测数据较少,不能支撑复杂模型的建模,因此考虑建立线性模型. 从上述固定效应的估计结果也可以看出,线性模型结构是可以接受的.

进一步考察随机部分参数估计结果可以看出,第一层模型截距  $\beta_{0i}$  的随机误差  $\mu_{0i}$  的方差为 4.296,  $\chi^2$  值为 217.4,  $P < 0.001$ ,即碳钢在不同地区的土壤腐蚀率初始状态存在着明显差异.

### 2.2.3 全模型

在基本模型的基础上,在第二层模型中逐步增加与土壤腐蚀特性相关的 pH 值、有机物含量、全氮含量、 $\text{HCO}_3^-$ 、 $\text{SO}_4^{2-}$ 、 $\text{Mg}^+$ 、 $\text{K}^+$ 、 $\text{Na}^+$ 、土壤电阻等理化参数,采用 1.3 节提出的方法对各变量进

行检验,删除不显著的因子,得到最终模型估计结果如表 5 所示.

首先考察固定效应参数估计结果. 对于第一层腐蚀率模型的截距,可以看出, pH 值(系数=-2.797,  $T = -3.955$ ,  $P=0.002$ )、全氮含量(系数 = -5.881,  $T = -3.875$ ,  $P=0.002$ )、 $\text{SO}_4^{2-}$ (系数 = -21.61,  $T = -4.010$ ,  $P=0.001$ )、电阻(系数 = -1.612,  $T = -3.823$ ,  $P=0.002$ )四个变量越小,则第一层腐蚀率模型的截距越大,对于初始腐蚀率具有显著负效应; 而  $\text{Mg}^+$ (系数 = 7.844976,  $T=2.793$ ,  $P=0.015$ ) 和  $\text{K}^+$ (系数 = 13.401239,  $T=3.638$ ,  $P=0.003$ ) 则对于初始腐蚀率具有显著正效应. 在控制上述因素后,不同地区站点的初始腐蚀率的均值为  $3.705 \text{ mm}\cdot\text{a}^{-1}$ .



表 5 碳钢腐蚀率全模型参数估计结果

Table 5 Parameter estimation results of the full model of carbon steel corrosion rate

因子	固定效应	系数	标准差	<i>T</i> 值	<i>P</i> 值
初始腐蚀率/(mm·a <sup>-1</sup> )	$\gamma_{00}$	3.705291	0.386479	9.587	0.000
pH 值	$\gamma_{01}$	-2.797140	0.707255	-3.955	0.002
全氮含量/%	$\gamma_{02}$	-5.880776	1.517765	-3.875	0.002
SO <sub>4</sub> <sup>2-</sup> 含量/%	$\gamma_{03}$	-21.615691	5.390560	-4.010	0.001
Mg <sup>+</sup> 含量/%	$\gamma_{04}$	7.844976	2.808501	2.793	0.015
K <sup>+</sup> 含量/%	$\gamma_{05}$	13.401239	3.683997	3.638	0.003
电阻/ $\Omega$	$\gamma_{06}$	-1.612091	0.421720	-3.823	0.002
腐蚀率变化速度	$\gamma_{10}$	-0.697246	0.150994	-4.618	0.000
pH 值	$\gamma_{11}$	0.173946	0.075356	2.308	0.031
全氮含量/%	$\gamma_{12}$	0.427550	0.147249	2.904	0.009
HCO <sub>3</sub> <sup>-</sup> 含量/%	$\gamma_{13}$	0.427550	0.147249	2.904	0.009
SO <sub>4</sub> <sup>2-</sup> 含量/%	$\gamma_{14}$	2.845542	0.740702	3.842	0.001
Mg <sup>+</sup> 含量/%	$\gamma_{15}$	-1.048672	0.286515	-3.660	0.002
K <sup>+</sup> 含量/%	$\gamma_{16}$	-0.804762	0.353150	-2.279	0.033
Na <sup>+</sup> 含量/%	$\gamma_{17}$	-2.179878	0.557934	-3.907	0.001
随机效应	标准误差	方差	$\chi^2$	<i>P</i> 值	
$\mu_{0i}$	1.63887	2.68589	86.87952	0.000	
$r_{ti}$	0.73006	0.53298	—	—	

对于第 2 层腐蚀率模型的斜率,可以看出, pH 值(系数 = 0.174,  $T = 2.308$ ,  $P = 0.031$ )、全氮含量(系数 = 0.428,  $T = 2.904$ ,  $P = 0.009$ )、HCO<sub>3</sub><sup>-</sup>(系数 = 0.428,  $T = 2.904$ ,  $P = 0.009$ )和 SO<sub>4</sub><sup>2-</sup>(系数 = 2.846,  $T = 3.842$ ,  $P = 0.001$ )四个变量越大,则第一层腐蚀率模型的斜率越大,即对于腐蚀率变化速度具有显著正效应;而 Mg<sup>+</sup>(系数 = -1.049,  $T = -3.660$ ,  $P = 0.002$ )、K<sup>+</sup>(系数 = -0.805,  $T = -2.279$ ,  $P = 0.033$ )和 Na<sup>+</sup>(系数 = -2.180,  $T = -3.907$ ,  $P = 0.001$ )则对于腐蚀率变化速度具有显著负效应。在控制上述因素后,腐蚀率变化速度为 -0.6972,仍具有显著的线性下降趋势( $T = -4.618$ ,  $P < 0.001$ )。

继续考察随机效应估计结果,不同地区的碳钢腐蚀率仍存在着显著的差异(方差 = 2.68589,  $\chi^2 = 86.88$ ,  $P < 0.001$ ),表明模型可能还存在进一步解释的可能。但是,对比基本模型(方差 = 4.29633),全模型方差为 2.68589 有了显著的下降,下降比例  $\omega = (4.29633 - 2.68589) / 4.29633 \times 100\% = 37.48\%$ ,即实验收集的土壤理化参数解释了初始情况的 37.48% 方差。

根据估值结果,全模型表示为如下。

第一层模型:

$$Y_{ti} = \beta_{0i} + \beta_{1i}(\alpha_{ti}) + r_{ti}. \quad (11)$$

第二层模型:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{pH}_i) + \gamma_{02}(\text{Nitrogen}_i) + \gamma_{03}[(\text{SO}_4^{2-})_i] + \gamma_{04}[(\text{Mg}^+)_i] + \gamma_{05}[(\text{K}^+)_i] + \gamma_{06}(\text{Resistance}_i) + \mu_{0i};$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}(\text{pH}_i) + \gamma_{12}(\text{Nitrogen}_i) + \gamma_{13}[(\text{HCO}_3^-)_i] + \gamma_{14}[(\text{SO}_4^{2-})_i] + \gamma_{15}[(\text{Mg}^+)_i] + \gamma_{16}[(\text{K}^+)_i] + \gamma_{17}[(\text{Na}^+)_i].$$

式中, Nitrogen 表示全氮含量, Resistance 表示电阻。建立全模型后,根据土壤理化参数等自变量数据,并利用经验贝叶斯公式(式(8)),就可以为任一地区  $i$  建立独立的土壤腐蚀率模型。例如对于大庆中心站,可计算得出,  $\beta_{0\text{DQ}} = 2.921$ ,  $\beta_{1\text{DQ}} = -1.066$ , 其腐蚀率模型为  $Y_{t\text{DQ}} = 2.921 - 1.066a_{(t\text{DQ})}$ 。

为了检验模型,本文计算了全部 37 组实验数据的模型估值,并与实际测量数据对比,结果如图 1 所示。其中横坐标为实验样本编号;纵坐标为腐蚀率,单位为  $\frac{\text{mm}}{\text{mm/a}}$ 。方形图标表示利用本文求得的两层线性模型计算出的,不同地区各埋样时间的碳钢估计腐蚀率;圆圈图标表示实际测量的实验数据。可以看出,模型估值与观测数据实际值非常接近,实证了依据多层线性模型原理建立的模型能够较好地描述腐蚀率的规律。

### 2.3 模型预测能力

为考察依据多层线性模型原理建立模型的预测能力,本文采用将任一地区最后一次腐蚀率观测数据删除,然后使用剩余数据建模,并检验模型预测值与该次观测实际值差异的方式。由于实验数据少,为保证模型稳定,在地区选择时,确定规则为该地区至少有两次以上的观测数据,且每次只删一组数据。在本文使用的数据中,全部 21 个土壤腐蚀试验台站共有 13 个台站满足条件。删除数据后模型预测值与删除值的偏差结果如图 2 所示。尽管由于

样本数量的问题,偏差相对图 1 有所放大,但仍然较好地实现了对腐蚀率的预测。

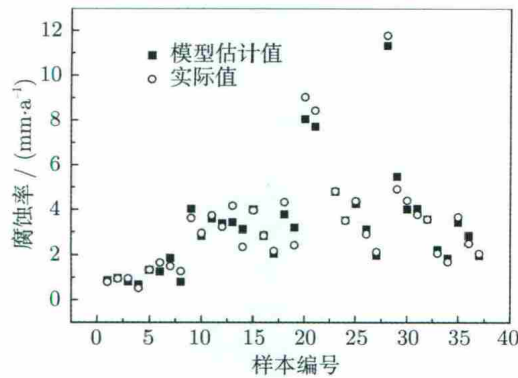


图 1 碳钢土壤腐蚀率实际值与模型估计值比较

Fig.1 Comparison between actual data and model estimation results of carbon steel corrosion rate in soils

2.4 与时间序列模型的比较

为了进一步说明多层线性模型的可靠性,本文

将多层线性模型与应用了指数平滑算法和差分自回归移动平均 (ARIMA) 模型两种时间序列算法<sup>[15]</sup>建立的模型进行了对比.为保证时序算法可以执行,选择了单地区至少有两次以上观测的数据进行建模,各地区试件埋埋时间、腐蚀率等原始数据及三种算法的模型拟合值和偏差如表 6 所示。

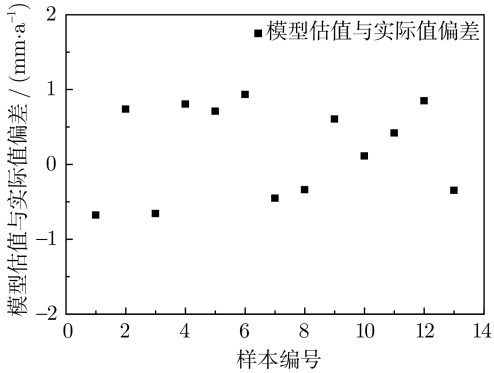


图 2 模型预测能力: 模型估计与实际值偏差

Fig.2 Predictive ability of the model: the residual between carbon steel corrosion rate and model estimation results

表 6 三种算法的实验结果

Table 6 Experimental results of the three algorithms

站点 ID	埋存时间/a	腐蚀率/ (mm·a <sup>-1</sup> )	指数平滑	指数平滑	ARIMA	ARIMA	HLM	HLM
			算法估值/ (mm·a <sup>-1</sup> )	算法偏差/ (mm·a <sup>-1</sup> )	算法估值/ (mm·a <sup>-1</sup> )	算法偏差/ (mm·a <sup>-1</sup> )	算法估值/ (mm·a <sup>-1</sup> )	算法偏差/ (mm·a <sup>-1</sup> )
7	1	1.493	1.3345	0.1585	—	—	1.855	-0.362
	2	1.269	1.3809	-0.1119	—	—	0.790	0.479
8	3.5	3.609	3.1357	0.4733	—	—	4.016	-0.407
	6.5	2.940	3.2742	-0.3342	—	—	2.831	0.109
9	1	3.719	3.3624	0.3566	—	—	3.586	0.133
	3	3.215	3.4668	-0.2518	—	—	3.367	-0.152
10	1	4.155	2.8731	1.2819	—	—	3.416	0.739
	3	2.343	3.2481	-0.9051	—	—	3.116	-0.773
12	3.7	2.840	2.3596	0.4804	—	—	2.835	0.005
	5.7	2.161	2.5002	-0.3392	—	—	2.051	0.110
13	3	4.318	2.9760	1.3420	—	—	3.767	0.551
	5	2.421	3.3686	-0.9476	—	—	3.184	-0.763
14	1	9.022	9.022	0	7.7902	1.2318	8.035	0.987
	2	8.415	9.022	-0.607	7.4168	0.9982	7.701	0.714
	4	5.646	8.415	-2.769	7.6008	-1.9548	7.032	-1.386
15	5	4.805	3.8747	0.9303	—	—	4.786	0.019
	7	3.490	4.1469	-0.6569	—	—	3.51	-0.02
16	1	4.363	4.363	0	3.1269	1.2361	4.246	0.117
	3	2.912	4.363	-1.451	3.0962	-0.1842	3.109	-0.197
	5	2.120	2.912	-0.792	3.1322	-1.0122	1.972	0.148
18	5	4.900	4.5378	0.3622	—	—	5.465	-0.565
	7	4.388	4.6438	-0.2558	—	—	3.997	0.391
19	3	3.76	3.6114	0.1486	—	—	4.028	-0.268
	5	3.55	3.6549	-0.1049	—	—	3.560	-0.010
20	3	2.069	1.7959	0.2731	—	—	2.217	-0.148
	5	1.683	1.8758	-0.1928	—	—	1.837	-0.154
21	1	3.659	3.659	0	2.7284	0.9306	3.413	0.246
	3	2.497	3.659	-1.162	2.6748	-0.1778	2.831	-0.334
	6	2.047	2.497	-0.450	2.7417	-0.6947	1.959	0.088

由以上实验结果可以看出, 由于每个地区只有 2~3 组样本, 构成的时间序列很短, 无论是指数平滑算法还是差分自回归移动平均 (ARIMA) 模型均受到局限, 其模型拟合值与实际值的偏差普遍大于多层线性模型, 而且任一地区的时间序列模型计算结果多与腐蚀率逐年降低的实际变化趋势相反. 相对而言, 多层线性模型则表现出了较好的性能.

综合考虑时间序列算法着重处理预测目标自身的时间序列数据, 不研究事物之间的因果关系的特点, 无法进一步分析土壤腐蚀规律的影响因素, 可以认为, 多层线性模型更适应于本文所研究的数据.

### 3 结 论

(1) 以国家材料腐蚀实验站网中碳钢在土壤中的腐蚀数据为基础, 应用多层线性模型原理, 为各地区建立了独有的土壤腐蚀率模型.

(2) 模型较好地描述了土壤理化参数等特性对于腐蚀率的影响, 具有较好的解释性. 同时, 实验验证了模型可以用来对材料在土壤中的腐蚀率进行预测并得到可接受的结果. 随着样本数量的增加, 模型的准确性将得到提升.

(3) 多层线性模型建立在坚实的统计理论基础之上, 考虑了不同层次的解释变量和误差随机性, 使模型建立在合理的假设上; 同时具有不要求重复实验的观测时点相同 (对应于重复实验观测)、可相对解决小样本问题等优势, 在具有层次结构的科学实验数据分析和规律研究领域, 具有实用价值.

### 参 考 文 献

- [1] Yahaya N, Noor N M, Othman R S, et al. New technique for studying soil corrosion of underground pipeline. *J Appl Sci*, 2011, 11(9):1510
- [2] Yang R C, Wang B, Zhang T Y. Forecast of material corrosion behavior by means of MATLAB-based algorithm of chaos time series. *J Lanzhou Univ Technol*, 2009, 35(5): 5  
(杨瑞成, 王彬, 张天云. 基于 MATLAB 的混沌时间序列算法对材料腐蚀行为的预测. 兰州理工大学学报, 2009, 35(5): 5)
- [3] He S, Zou Y L, Quan D S, et al. Application of RBF neural network and ANFIS on the prediction of corrosion rate of pipeline steel in soil // *Recent Advances in Computer Science and Information Engineering*. Berlin: Springer Berlin Heidelberg, 2012: 639
- [4] Sadowski L. Non-destructive investigation of corrosion current density in steel reinforced concrete by artificial neural networks. *Arch Civ Mech Eng*, 2013, 13(1): 104
- [5] Liang P, Du C W, Li X G, et al. Grey Relational space analysis of effect of environmental factors on corrosion resistance of X70 pipeline steel in Yingtan soil simulated solution. *Corros Prot*, 2009, 30(4): 231  
(梁平, 杜翠薇, 李晓刚, 等. X70 管线钢在鹰潭土壤模拟溶液中腐蚀因素关联分析. 腐蚀与防护, 2009, 30(4): 231)
- [6] Cao C N. *The Environment Corrosion of Materials in China*. Beijing: Chemical Industry Press, 2005  
(曹楚南. 中国材料的自然环境腐蚀. 北京: 化学工业出版社, 2005)
- [7] Lindley D V, Smith A F. Bayes estimates for the linear model. *J R Stat Soc Ser B*, 1972, 34(1): 1
- [8] Heck R H, Thomas S L. *An Introduction to Multilevel Modeling Techniques*. 2nd Ed. London: Routledge, 2008
- [9] Angioni S L, Visrolia A, Meo M. A hierarchical multiple plate models theory for laminated composites including delamination and geometrical nonlinear effects. *Compos Struct*, 2011, 93(2): 780
- [10] Zeller M H, Reiter-Purtill J, Ratcliff M B, et al. Two-year trends in psychosocial functioning after adolescent Roux-en-Y gastric bypass. *Surg Obesity Related Dis*, 2011, 7(6): 727
- [11] Liu H Y, Zhang L. *Tracking Data Methods and Applications*. Beijing: Education and Science Press, 2005  
(刘红云, 张雷. 追踪数据方法及应用. 北京: 教育科学出版社, 2005)
- [12] Raudenbush S W, Bryk A S. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd Ed. Guo Z G, Translated. Beijing: Social Sciences Academic Press, 2007  
(Raudenbush S W, Bryk A S. 分层线性模型: 应用与数据分析方法. 2 版. 郭志刚, 译. 北京: 社会科学文献出版社, 2007)
- [13] Wen F X. *Principle and Application of Hierarchical Linear Model*. Beijing: China Light Industry Press, 2009  
(温福星. 阶层线性模型的原理与应用. 北京: 中国轻工业出版社, 2009)
- [14] Longford N T. *Random Coefficient Models*. London: Clarendon Press, 1994
- [15] William W S. *Time Series Analysis: Univariate and Multivariate Methods*. 2nd Ed. Liu C, Yi D H, Translated. Beijing: China Renmin University Press, 2009  
(魏武雄. 时间序列分析: 单变量和多变量方法. 2 版. 刘超, 易丹辉, 译. 北京: 中国人民大学出版社, 2009)