

# 大气环境腐蚀性因素的相关性

屈祖玉 董玉兰 李长荣 卢燕平

北京科技大学材料科学与工程学院,北京 100083

**摘要** 通过相关性数据处理和模糊聚类分析,发现影响大气环境腐蚀性的21种环境因素在相关系数大于0.54的情况下,可明显地聚为5类.在此基础上,进一步通过对逐个因素的分类能力的特征评价,选取大气环境腐蚀性的主要因素(RH>80%时数,SO<sub>2</sub>沉积速度,水溶性尘降,Cl<sup>-</sup>沉积速度,NO<sub>2</sub>浓度),进而在特征空间中进行镀锌层样本的大气环境腐蚀性预测和考评,为今后实现我国大气环境腐蚀性等级评判提供依据.

**关键词** 大气腐蚀; 聚类分析; 特征评价; 腐蚀性预测

**分类号** TQ 050.91

大气环境使材料产生腐蚀的性能称大气环境腐蚀性.大气环境腐蚀性的影响因素很多,作用复杂,各种因素的影响大小也不相同,有的还随地区、时间条件的改变而变化.因此找出若干主要因素,探讨这些主要影响因素与大气环境腐蚀性的相关性是本研究的重点,也是对各类大气环境腐蚀性的评价、分类和预测的基础.

本文应用1983~1994年我国大气腐蚀网站的环境调查数据及相应曝晒场镀锌层挂片腐蚀结果的数据,通过相关性数据处理和模式识别的系统聚类方法,对大气环境腐蚀性因素的相关性进行分析讨论,并采用K-W(Kruskal and Wallis)检验方法,选取出大气环境腐蚀性的主要因素,再根据大气环境腐蚀性挂片样本在多维空间的最近邻配置,推测不同等级的大气环境腐蚀性点在空间的分布情况,为解决我国大气环境腐蚀性等级判断和材料在给定大气环境中的长期腐蚀行为预测问题提供参考和依据.

## 1 大气环境腐蚀性因素的聚类分析

### 1.1 大气环境腐蚀性因素的相关系数

根据全国大气腐蚀网站1983~1994年的环境调查数据,选取了大气环境腐蚀性影响因素21项.

气候因素:平均温度;极端最高温度;极端最低温度;平均相对湿度;年RH>80%时数;年降

水量;年降雨日数;年降露日数;年雾日数;年降轻雾日数;年日照时数;年日照百分数.

大气中腐蚀性因素:Cl<sup>-</sup>沉积速度;Cl<sup>-</sup>浓度;SO<sub>2</sub>沉积速度;SO<sub>2</sub>浓度;NO<sub>2</sub>浓度;雨水pH;非水溶性尘降;水溶性尘降;NH<sub>3</sub>浓度.

考虑因素与因素之间的相互关系是通过“相异度”和“相似度”来表达的,因使用的运算不同而有各种定义.按最大最小法,度量相似性的参数应表示为<sup>[1]</sup>:

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ik} \wedge x_{jk})}{\sum_{k=1}^n (x_{ik} \vee x_{jk})} \quad (1)$$

式中: $n$ 为样品数; $x_{ik}$ 为第 $i$ 种特征(因素)在第 $k$ 个样品中的取值; $x_{jk}$ 为第 $j$ 种特征(因素)在第 $k$ 个样品中的取值.

由(1)式对大气腐蚀环境调查数据进行处理,算得相关系数 $r_{ij}$ ,进而确定出相似矩阵 $R=(r_{ij})_{n \times n}$ ,并得到表1所示的相似关系矩阵.

### 1.2 大气环境腐蚀性因素的聚类

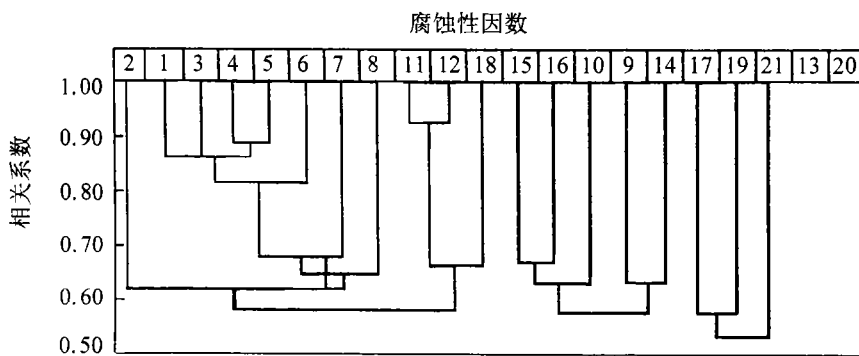
在相似矩阵的基础上,进行模糊聚类分析的方法有3种<sup>[2]</sup>.本文选用直接聚类法.它是直接利用相似矩阵 $R$ 进行聚类的方法.

给出大气环境腐蚀性因素随相关系数变化所形成的聚类图谱,见图1.

由图1聚类结果表明,在相关系数大于0.54的情况下,将影响大气环境腐蚀性的21项因素明显地聚合成5类.从同类因素中选取主要因素进行更深入的研究,是对众多腐蚀性因素的降维处理,为大气环境腐蚀性研究中因素的选取、等级的分类及预测提供依据.

表 1 大气环境腐蚀性因素的相似关系矩阵

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
(1)	1	0.52	0.86	0.76	0.76	0.81	0.72	0.63	0.16	0.33	0.47	0.44	0.35	0.11	0.18	0.22	0.15	0.41	0.20	0.22	0.25
(2)		1	0.52	0.58	0.54	0.52	0.59	0.61	0.18	0.55	0.40	0.42	0.16	0.10	0.36	0.41	0.25	0.58	0.43	0.38	0.44
(3)			1	0.82	0.86	0.74	0.78	0.62	0.21	0.39	0.49	0.45	0.41	0.16	0.26	0.30	0.15	0.43	0.22	0.24	0.31
(4)				1	0.89	0.71	0.75	0.64	0.23	0.45	0.51	0.48	0.39	0.18	0.31	0.36	0.21	0.53	0.28	0.26	0.38
(5)					1	0.73	0.79	0.63	0.23	0.43	0.51	0.50	0.43	0.18	0.28	0.33	0.21	0.50	0.28	0.28	0.36
(6)						1	0.72	0.64	0.14	0.32	0.48	0.49	0.32	0.13	0.16	0.20	0.21	0.49	0.25	0.23	0.25
(7)							1	0.66	0.21	0.45	0.43	0.45	0.29	0.16	0.27	0.32	0.20	0.49	0.27	0.29	0.37
(8)								1	0.18	0.41	0.38	0.39	0.20	0.12	0.29	0.37	0.18	0.36	0.23	0.27	0.37
(9)									1	0.30	0.27	0.26	0.21	0.63	0.57	0.50	0.12	0.26	0.22	0.30	0.29
(10)										1	0.28	0.30	0.11	0.15	0.54	0.63	0.23	0.38	0.42	0.47	0.44
(11)											1	0.92	0.36	0.32	0.36	0.28	0.16	0.62	0.32	0.35	0.18
(12)												1	0.35	0.31	0.35	0.30	0.18	0.66	0.35	0.38	0.20
(13)													1	0.33	0.13	0.19	0.09	0.28	0.10	0.10	0.14
(14)														1	0.38	0.31	0.14	0.31	0.16	0.13	0.15
(15)															1	0.67	0.20	0.34	0.41	0.47	0.43
(16)																1	0.20	0.30	0.35	0.45	0.46
(17)																	1	0.34	0.57	0.28	0.54
(18)																		1	0.48	0.28	0.34
(19)																			1	0.51	0.56
(20)																				1	0.44
(21)																					1



## 2 大气环境腐蚀性主要因素选取

### 2.1 单个因素的特征评价

为了评价单个因素分散能力的强弱,常用的方法是K-W(Kruskal and Wallis)检验方法<sup>[1]</sup>.假设一批样本共有N个,这批样本可分为m类,第i类包括N<sub>i</sub>个样本,则检验方法如下:(1)取每类样本编号的平均值,分别记作R<sub>1</sub>,R<sub>2</sub>,...,R<sub>m</sub>;(2)计算统计量H,K.H,K越大时,单个因素分类能力越强.

$$H = [12/N(N+1)] \sum_{i=1}^m N_i [R_i - (N+1)/2]^2 \quad (2)$$

$$K = \sum_{i=1}^m \sum_{j=i+1}^m (R_i - R_j)^2 \quad (3)$$

下面列出了根据图1聚类结果求得的各因素的K-W特征检验值.

第1类 平均温度为5.571/156.0;最高温度

为5.893/149.3;最低温度为5.571/156.0;平均相对湿度为5.786/160.3;年RH>80%时数为5.893/163.3;年降水量为5.893/149.3;年降雨日数为4.286/113.0;雨水pH为5.036/125.3;年降露日数为5.571/156.0;年日照时数为5.893/149.3;年日照百分数为5.893/149.3.

第2类 SO<sub>2</sub>沉积速度为5.893/163.3;SO<sub>2</sub>浓度为5.036/125.3;Cl<sup>-</sup>浓度为5.571/149.0;年雾日数为5.786/125.3;年降轻雾日数为5.893/149.3.

第3类 NO<sub>2</sub>浓度为5.571/128.0;NH<sub>3</sub>浓度为3.321/91.3;非水溶性尘降为5.036/125.3.

第4类 水溶性尘降为5.036/139.3.

第5类 Cl<sup>-</sup>沉积速度5.893/163.3.

### 2.2 主要因素的选取

上述分类所包容的诸多大气环境腐蚀性因

素构成与分类有关的各种信息, 这些信息中又包含着许多彼此相关的因素. 在此基础上, 选取主要的因素, 就可以避免信息的重复和人力、物力的浪费. 从同类因素中选取代表性特征的原则是, 选取对识别作用最大的特征和实际中容易获取的特征. 参照上面的聚类结果和单个因素的特征评价数据, 按选择原则, 本文选取具有代表性特征因素为: 年RH>80%时数; SO<sub>2</sub> 沉积速度; 水溶性尘降; Cl<sup>-</sup> 沉积速度; NO<sub>2</sub> 浓度. 由于单个特征的分类能力不同, 因此它们对于分类的贡献率也不相同. 假定全部考虑上述选取的5种主要因素对分类的总贡献率为100%, 则主

要因素各自的贡献率和累积贡献率可分别由下式来考评:

$$\text{贡献率} = H_i / \sum_{j=1}^k H_j \quad (4)$$

$$\text{累积贡献率} = \sum_{j=1}^i H_j / \sum_{j=1}^k H_j \quad (5)$$

式中:  $H_i$  为第  $i$  个特征因素的分类能力, 见式(1);  $k$  为所考虑的特征因素的总个数.

表2 将主要特征因素的顺序按贡献率的大小排列, 它表明了主要因素的优先选取顺序. 在表2中还给出了前若干个因素的累积贡献率, 该参量表示只考虑前若干个因素讨论问题时所反映的总信息量的百分数.

表2 主要大气环境腐蚀性因素的特征评价

项目	年RH>80%时数	SO <sub>2</sub> 沉积速度	Cl <sup>-</sup> 沉积速度	NO <sub>2</sub> 浓度	NH <sub>3</sub> 浓度
K-W 检验值	5.893	5.893	5.893	5.571	5.036
贡献率 / %	20.83	20.83	20.83	19.70	17.81
累积贡献率 / %	20.83	41.66	62.49	82.19	100.00

### 3 大气环境腐蚀性的等级评判

#### 3.1 标样的最近邻配置

以全国7个典型地区大气腐蚀试验站11年(1983~1994年)镀锌层大气腐蚀环境因素和腐蚀数据为标样, 在上述选取的5种主要因素构成的多维空间中, 求取样本之间的广义距离(即规定两因素间的相关系数的倒数(1/ $r_{ij}$ ))作为这两因素间的广义距离 $d_{ij}$ , 并确定各样本的最临近配置. 为了消除各个指标量纲的影响, 首先对数据进行处理<sup>[1]</sup>. 由此得到如图2所示的最近邻配置情况.

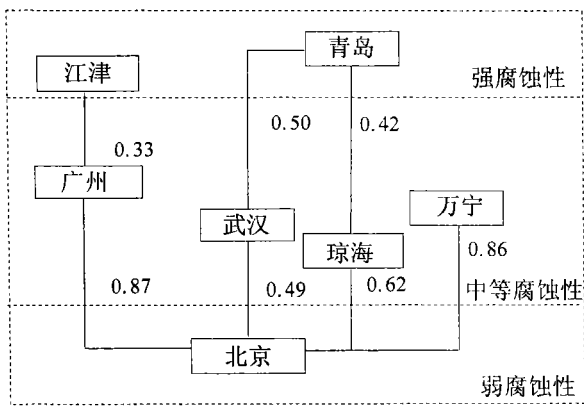


图2 标样的最近邻配置

图2表明, 中等腐蚀性等级的标样的最近邻一边是弱腐蚀性等级的标样, 另一边是严重腐蚀性等级的标样. 而弱腐蚀性等级和严重腐蚀性等级的标样都仅仅邻近中等腐蚀性等级的标

样. 也就是说, 相同腐蚀性的点在高维空间中相聚在一定的区域, 而不同等级的区域在空间有一定的排列顺序. 由此得到, 我国典型地区大气环境腐蚀性的相对强弱可划分为3级: 北京为I级(弱腐蚀性); 琼海、万宁、广州、武汉等地为II级(中等腐蚀性); 青岛、江津等地是III级(强腐蚀性); 其强弱顺序为青岛、江津、广州、万宁、武汉、琼海、北京. 这一评价结果与镀锌层长期曝晒试验的腐蚀结果基本吻合<sup>[4]</sup>. 可见模式识别方法能直接得到镀锌层的大气腐蚀性, 对建立符合我国国情的镀锌层大气环境腐蚀性分级标准提供了依据.

#### 3.2 贴适度度量法

我国幅员辽阔, 不可能在每个地区都建立大气腐蚀试验站, 对于任意地区大气环境对材料所产生的腐蚀程度, 只能用与其环境因素最接近的大气腐蚀试验站的数据作为评价依据. 为此, 本文应用贴适度概念对试验样本(未知地区)的大气腐蚀性进行分类判别, 具体方法如下:

(1) 为消除各指标量纲的影响, 先对已知样本(全国7个典型地区大气腐蚀试验站)和试验样本的环境参数( $u_i$ )进行标准处理.

(2) 采用最大最小贴适度计算试验样本A与已知样本B之间的距离<sup>[2]</sup>, 即:

$$D(A, B) = \frac{\sum_{i=1}^n \min[A(u_i), B(u_i)]}{\sum_{i=1}^n \max[A(u_i), B(u_i)]} \quad (6)$$

(3) 确定与已知样本最贴近的地区, 即未知地区的大气环境腐蚀性因素特征与已知典型地区相近。

(4) 评价、预测未知地区的大气腐蚀性。

上述各个步骤可以通过“镀锌层大气腐蚀咨询评价系统”软件来完成<sup>[9]</sup>。在识别可靠性的考评试验中, 从试验样本集的总体来看, 识别正确率达到 90%。

## 4 结语

(1) 通过相关性数据处理和系统聚类的方法, 将影响大气环境腐蚀性的 21 项因素在相关系数大于 0.54 的情况下聚合成 5 类, 它是对众多环境因素的降维处理, 为评价我国大气环境腐蚀性及其等级提供依据。

(2) 采用 K-W 检验方法, 对大气环境腐蚀性因素的分类能力的特征进行评价, 从同类因素中选取代表性特征, 可作为参与大气环境腐

蚀性考评和预测的主要因素。

(3) 应用最大最小贴近度概念和选取的 5 种主要因素, 基本可实现对我国大气环境腐蚀性及其等级的预测, 为材料大气腐蚀与防护的研究提供新方法。

## 参 考 文 献

- 1 邹开其, 徐杨. 模糊系统与专家系统. 西安: 西安交通大学出版社, 1989.72
- 2 李安贵. 模糊数学及其应用. 北京: 冶金工业出版社, 1994
- 3 王碧泉, 陈祖荫. 模式识别——理论、方法和应用. 北京: 地震出版社, 1989.30
- 4 萧以德, 李兴濂. 电镀锌层的大气腐蚀行为研究. 腐蚀科学与防护技术, 1995(3):190
- 5 董玉兰. 镀锌层大气腐蚀咨询评价系统研究 [D]. 北京: 北京科技大学, 1997

致谢: 研究工作得到全国大气腐蚀网站大力支持和帮助, 在此深表感谢。

## Correlation Among Corrosive Factors of Atmospheric Environment

*Qu Zuyu, Dong Yulan, Li Changrong, Lu Yanping*

Materials Science and Engineering School, UST Beijing, Beijing 100083, China

**ABSTRACT** By considering the correlation of the data and the using the clustering technique of pattern recognition, the relationship among various corrosive factors of atmospheric environment has been analyzed. The results of the clustering show that the 21 atmospheric corrosive factors are clustered into five distinct clusters under the correlation coefficient value above 0.54. On the basis of the clustering analysis, the abilities of corrosive factors to distinguish atmospheric corrosivity are assessed, and the main factors are selected. In the generalized space of the main factors, the disposition of the nearest neighbors for the experimental samples are analyzed and the way to predict the atmospheric corrosivity for unknown samples is studied.

**KEY WORDS** atmospheric corrosion; clustering analysis; characteristic assessment; corrosion prediction