

长江口水域氯离子浓度分布模型及预测模型

徐文莉¹,梁 磊²,王艳明³

(1.上海电力学院数理系,上海 200090;2.上海电力学院能源与环境工程学院,上海 200090;3.上海市市政工程设计研究总院,上海 200092)

摘要 对长江口水域冷却水中氯离子浓度的样本数据进行统计分析,建立氯离子浓度分布的数学模型,氯离子浓度近似服从对数正态分布,并对比分析邻近水厂和电厂冷却水中的氯离子浓度,显示两者氯离子的浓度没有显著性差异,为预测全年水中氯离子的浓度提供依据,并给出最小样本容量的选取。

关键词 氯离子浓度 对数正态分布 电厂冷却水

中图分类号 X703 **文献标识码** A **文章编号** 1004-6933(2008)03-0049-04

Distribution pattern and prediction model of Cl⁻ concentration in the Yangtze Estuary

XU Wen-li¹, LIANG Lei¹, WANG Yan-ming²

(1. Department of Mathematics and Physics, Shanghai University of Electric Power, Shanghai 200090, China; 2. School of Thermal Power and Environmental Engineering, Shanghai University of Electric Power, Shanghai 200090, China; 3. Shanghai Municipal Engineering Design Institute, Shanghai 200092, China)

Abstract A distribution model of Cl⁻ concentration was established based on the statistical analyses of Cl⁻ concentration samples from the cooling water in the Yangtze Estuary. Results indicate that Cl⁻ concentration in the area has a lognormal distribution. A comparison of Cl⁻ concentration in the cooling water of a power plant with that in an adjacent water plant shows no significant difference, which will help in the prediction of Cl⁻ concentration throughout a whole year. Finally, a minimal sample capacity is given.

Key words concentration of Cl⁻; lognormal distribution; cooling water of power plant

华东地区东临大海,感潮河流^[1]很多,季节、天气和潮汐等因素的变化,引起长江口水域中不同地点或者同一地点不同时间的水质发生变化,而水质特征直接影响凝汽器的选材。我国以淡水为冷却水的凝汽器大多采用黄铜管。由于水资源的短缺,循环冷却水的浓缩倍率的提高及水源环境污染日趋严重,近年来,越来越多的电厂趋向于使用不锈钢管^[1]。冷却水中通常含有氯离子、硫酸根及硫等,其中氯离子是不锈钢钝化膜的侵蚀性离子,是影响凝汽器的重要因素。

限于人力、物力和财力等因素,尤其是电厂,水质数据测试频率很低,难以找到冷却水中氯离子浓度的变化。海水入侵会引起长江口水域氯化物含量

超标,氯离子浓度的超标对区域内吴淞水厂、宝钢电厂和外高桥电厂等企业的正常运转有重要影响,且以距离长江最近的吴淞水厂尤为严重。

目前,对于长江口水域,仅有少量的、分散的氯离子浓度取样资料,没有系统的、全面的统计研究,因而研究不同位置的水的氯离子浓度变化的分布规律,掌握冷却水中氯离子浓度的分布特征十分必要。笔者通过收集的有关资料,首次尝试应用数理统计的手段,分析研究长江口水域冷却水中氯离子浓度。

1 采样点(长江口)水质

各采样点水质见表 1,氯离子质量浓度分布见图 1。图 1 有超过 60% 的氯离子质量浓度小于

表 1 采样点水质				
采样点	样本容量/个	氯离子质量浓度/(mg·L ⁻¹)		
		平均值	极小值	极大值
吴淞水厂	1826	178.11	8.00	3085.00
吴淞水厂月均值	60	178.09	12.90	1717.21
宝钢电厂	60	45.16	6.00	965.00
吴泾电厂	67	55.56	19.50	105.00
石洞口电厂	25	140.75	1.00	1200.00
外高桥电厂	12	107.83	18.00	2450.00

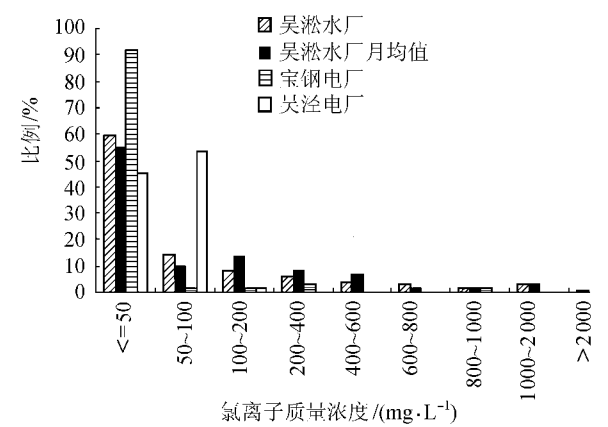


图 1 氯离子浓度分布

50 mg/L,只有 4%的氯离子质量浓度大于 1000 mg/L。各总体的样本数据都为正偏斜分布。

由表 1 知不同地点的氯离子浓度统计分布特征不同,其中以距离长江最近的吴淞水厂的氯离子浓度平均值和极差最大,吴淞水厂的氯离子浓度月均值的平均值和极差也非常大,而吴泾电厂的氯离子浓度平均值较小,其极差最小,因此海水入侵时间和强度对水质的影响十分明显。

2 计算方法与结果

2.1 吴淞水厂水质分析

原始资料中有吴淞水厂氯离子平均质量浓度数据 1826 个,为排除数据的系统误差、降低随机误差,取平均值修匀数据。

表 3 氯离子质量浓度相关分析的 P 值

	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉	Y ₁₀	Y ₁₁	Y ₁₂
Y ₁	0.0224	0.1375	0.0867	0.3093	0.1112	0.7791	0.4006	0.7700	0.8579	0.6791	0.2843
Y ₂		0.0492	0.0410	0.1512	0.1600	0.8977	0.8002	0.9120	0.9617	0.5534	0.1503
Y ₃			0.0121	0.1171	0.2339	0.4431	0.9780	0.4888	0.4517	0.1521	0.0258
Y ₄				0.0567	0.0805	0.4309	0.7228	0.6139	0.5596	0.1707	0.0144
Y ₅					0.0935	0.7185	0.9706	0.7324	0.7208	0.2534	0.0268
Y ₆						0.5806	0.2916	0.9571	0.9633	0.4385	0.1466
Y ₇							0.5459	0.0989	0.0605	0.0783	0.3604
Y ₈								0.6841	0.8015	0.9909	0.9550
Y ₉									0.0108	0.0572	0.3828
Y ₁₀										0.0447	0.3576
Y ₁₁											0.0648

$X = (X_1, X_2, \cdots, X_{12})'$ 是一个多元总体,对每年的数据预处理可得到 X 的一个样本观察值,查 W 检验表^[2]可得:显著性水平 $\alpha = 0.05$ 时为 $W_{0.05}(5) = 0.762$; $\alpha = 0.01$ 时 $W_{0.01}(5) = 0.686$ 。

按照 W 检验的法则,表 2 中有些指标的 W 统计量^[1]的观测值没有超过 W 统计量的 α 分位数,因而不能认为这些指标服从正态分布。

表 2 氯离子质量浓度的正态性 W 检验及相关估计

月均值	W 检验值 ^[1]		点估计值/(mg·L ⁻¹)	月均值	W 检验值 ^[1]		点估计值/(mg·L ⁻¹)
	原始	变换			原始	变换	
X ₁	0.8668	0.9576	430.60	X ₈	0.9718	0.9482	33.16
X ₂	0.7651	0.8731	607.30	X ₉	0.6763	0.7236	42.58
X ₃	0.8438	0.9788	607.90	X ₁₀	0.6333	0.7184	50.06
X ₄	0.8494	0.9282	224.35	X ₁₁	0.6891	0.8219	170.59
X ₅	0.7728	0.9201	80.30	X ₁₂	0.7822	0.7782	176.09
X ₆	0.9636	0.8945	33.88				
X ₇	0.8682	0.8468	30.11	$\bar{X}$	0.8559	0.9127	199.93

注: X_i 为第 i 月的氯离子浓度月均值; $\bar{X}$ 为氯离子浓度年算术平均值。

对原始样本资料进行自然对数变换之后得到 Y_i ,再进行正态性检验(见表 2)。可见这些指标 Y_i ($i = 1, 2, \cdots, 12$)都服从($\alpha = 0.01$)或近似地服从($\alpha = 0.05$)一元对数正态分布。再对 Y_i 作独立性检验,设其协方差矩阵为

$$\Sigma = (\sigma_{ij}^2)_{12 \times 12}$$

因而 Y_i 与 Y_j ($i \neq j$)相互独立当且仅当 $\sigma_{ij}^2 = 0$,对一切 $i \neq j$ ^[2],这又等价于 Y_i 与 Y_j 不相关, $i \neq j$,相关分析的 P 值矩阵(只需列出上三角部分)见表 3。

P 值均大于检验水平 $\alpha = 0.01$,从而 Y_i 与 Y_j ($i \neq j$)相互独立,进而可以认为该多元总体服从(取 $\alpha = 0.01$)或近似地服从(取 $\alpha = 0.05$)对数正态总体,即

$$X \sim \text{LN}(\mu, \Sigma)$$

式中 $\mu = (\mu_1, \mu_2, \cdots, \mu_{12})'$ 为总体数学期望向量。
记 $\hat{\mu}_i$ 为原始样本经对数变化后第 i 月氯离子

浓度的样本均值, $\hat{\sigma}_{ii}$ 为对应样本的标准差。由统计理论可知, 样本均值 $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \cdots, \hat{\mu}_{12})'$ 为 μ 的极大似然估计^[2], 在本问题中

$$\hat{\mu} = (5.27, 5.90, 555, 4.93, 4.13, 3.45, 3.28, 3.49, 3.66, 3.73, 4.44, 4.57)'$$

样本协方差矩阵

$$S^2 = (\hat{\sigma}_{ij}^2)_{12 \times 12}$$

可作为 Σ 的极大似然估计^[2]。

变量 $\bar{X}$ 服从对数正态分布 $\text{LN}(\mu, \sigma^2)$, 原始样本经对数变化后的样本均值 $\hat{\mu} = 4.86$ 为 μ 的极大似然估计^[2], 样本标准差 $\hat{\sigma} = 0.94$ 为 σ 的极大似然估计^[2]。

应用对数正态分布的性质, 可以计算第 i 月氯离子浓度均值的极大似然估计^[2]为

$$e^{\hat{\mu}_i + \hat{\sigma}_i^2/2} \quad (i = 1, 2, \cdots, 12)$$

而氯离子浓度年均值的极大似然估计^[2]为

$$e^{\hat{\mu} + \hat{\sigma}^2/2}$$

利用 SAS 软件包中的随机数函数^[3] RANNOR(seed) 进行模拟, 产生 13 组 1 万个标准正态分布的随机数, 分别代入

$$X_i = \exp \{ \hat{\mu}_i + \hat{\sigma}_{ii} C_i \} \quad (i = 1, 2, \cdots, 12)$$
$$\bar{X} = \exp \{ \hat{\mu} + \hat{\sigma} C \}$$

式中 C_i 、 C 为由 RANNOR(seed) 产生的某个随机数。可得到 12 组各 1 万个对数正态分布 $\text{LN}(\hat{\mu}_i, \hat{\sigma}_{ii}^2)$ ($i = 1, 2, \cdots, 12$) 的随机数, 及 1 万个对数正态分布 $\text{LN}(\hat{\mu}, \hat{\sigma}^2)$ 的随机数, 计算各自的点估计值(见表 2)。

2.2 水厂与电厂水质的比较

一般而言, 水厂比电厂的水质数据完善, 电厂数据是随机测取的。因此对比分析邻近水厂和电厂冷却水中的氯离子浓度, 用水厂的数据预测电厂的水质数据, 能减低企业的压力。

两两比较吴淞水厂与宝钢电厂、吴淞水厂与吴泾电厂, 取宝钢电厂每月一日的氯离子浓度样本为总体 1 的样本, 取吴泾电厂相应日的氯离子浓度的样本数据为总体 2 的样本, 而分别取吴淞水厂相应日的氯离子浓度的样本数据, 作为这两个总体对应的相比较的总体样本。

为提高统计检验的稳健性, 首先采用非参数检验方法, 分别对比研究宝钢电厂与吴淞水厂、吴泾电厂与吴淞水厂的水质数据, 作一系列一元 Wilcoxon 秩和检验^[2,4], 研究它们的氯离子浓度的分布是否

有显著性差异, 以确定氯离子浓度的估计方法, 记 Z_i 为总体第 i 月某日氯离子浓度, $\bar{Z}$ 为总体氯离子浓度的年平均值, 计算检验统计量的 P 值见表 4。

根据 Wilcoxon 秩和检验^[2,4] 法则, 当 P 值大于显著性水平 $\alpha = 0.01$ 时, 接受原假设, 认为宝钢电厂与吴淞水厂氯离子浓度的分布没有显著性差异。同理, 认为吴泾电厂与吴淞水厂的水质数据中, 氯离子浓度的分布也没有显著性差异。

再对宝钢电厂及吴泾电厂水中氯离子浓度 Z_1 、 Z_2 、 $\cdots$ 、 Z_{12} 和 $\bar{Z}$ 作正态性检验, 结果表明 这些指标都服从 ($\alpha = 0.01$) 或近似地服从 ($\alpha = 0.05$) 一元对数正态分布, 进而可以认为这两个多元总体都服从 ($\alpha = 0.01$) 或近似服从 ($\alpha = 0.05$) 对数正态分布 $\text{LN}(a, \Sigma)$, 其中 $a = (a_1, a_2, \cdots, a_{12})'$ 为数学期望向量, $\bar{Z}$ 服从对数正态分布 $\text{LN}(a, \sigma^2)$ 。

由统计理论可知, 总体 1 和总体 2 的样本均值向量

$$a^{(1)} = (3.4, 4.24, 3.59, 2.74, 2.47, 2.32, 2.49, 2.51, 2.61, 2.68, 2.5, 2.83)$$
$$a^{(2)} = (4.07, 4.07, 4.02, 4.06, 4.13, 3.99, 3.79, 3.94, 3.86, 3.94, 3.96, 4.1)$$

分别为两总体各自的数学期望向量 a 的极大似然估计^[2], 两个样本协方差矩阵

$$S^2 = (s_{ij}^2)_{12 \times 12}$$

分别为两总体各自的协方差矩阵 Σ 的极大似然估计^[2]。

应用对数正态分布的性质, 可得总体 j 的第 i 月氯离子浓度的极大似然估计^[2]为 $e^{\hat{a}_i^{(j)} + s_{ii}^{(j)2}/2}$, 年均值的极大似然估计^[2]为 $e^{\hat{a}^{(j)} + s^{(j)2}/2}$, $i = 1, 2, \cdots, 12, j = 1, 2$ 。

利用 SAS 软件包产生另外 26 组一万个标准正态分布的随机数, 分别代入公式:

$$Z_i = \exp \{ \hat{u}_i^{(j)} + \hat{s}_{ii}^{(j)} d_i^{(j)} \}$$
$$\bar{Z} = \exp \{ \hat{u}^{(j)} + s^{(j)} d^{(j)} \}$$

式中 $\hat{a}_i^{(j)}$ 为总体 j 的原始样本经对数变化后第 i 月的样本均值; $\hat{a}^{(j)}$ 为总体 j 的原始样本经对数变化后的年均值; $s_{ii}^{(j)}$ 、 $s^{(j)}$ 为对应样本的标准差; $d_i^{(j)}$ 、 $d^{(j)}$ 为 RANNOR(seed) 产生的某个随机数。

计算各指标的点估计值, 并取各自的第 5% 分

表 4 对比分析的 Wilcoxon 秩和检验 P 值

比较对象	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}	$\bar{Z}$
宝钢电厂与吴淞水厂	0.0601	0.1437	0.1437	0.0119	0.0122	0.0119	0.057	0.0112	0.074	0.2031	0.0122	0.0216	0.2101
吴泾电厂与吴淞水厂	0.5296	0.0122	0.3457	0.0601	0.5309	0.0122	0.2963	0.2963	0.1425	0.5296	1.0000	0.6761	0.6105

表 5 两总体的正态性 w 检验及相关估计

氯离子 浓度	W 检验值				估计值/(mg·L ⁻¹)			
	宝钢电厂		吴泾电厂		宝钢电厂		吴泾电厂	
	原始	变换	原始	变换	点估计	区间	点估计	区间
Z ₁	0.5669	0.6410	0.9452	0.8943	72.48	(3.31,282.01)	60.55	(39.36,87.80)
Z ₂	0.7636	0.8316	0.9204	0.9107	559.20	(2.32,1983.23)	60.72	(56.06,61.21)
Z ₃	0.8109	0.9673	0.8521	0.8668	59.69	(6.93,192.97)	58.27	(33.71,91.87)
Z ₄	0.8514	0.8977	0.9398	0.09097	17.54	(6.95,35.62)	61.56	(33.71,100.76)
Z ₅	0.8514	0.9148	0.9678	0.9129	12.86	(6.10,22.94)	65.83	(37.74,105.13)
Z ₆	0.9072	0.9532	0.8536	0.9000	11.33	(4.71,22.03)	60.19	(25.5,116.52)
Z ₇	0.8409	0.8804	0.7270	0.8368	12.27	(9.10,16.00)	49.68	(20.25,98.08)
Z ₈	0.9516	0.9179	0.7959	0.8134	12.56	(8.88,17.12)	53.26	(33.14,79.50)
Z ₉	0.9019	0.9074	0.8661	0.9346	14.25	(8.36,22.32)	50.46	(26.93,83.34)
Z ₁₀	0.8664	0.8864	0.8451	0.8549	15.36	(8.65,24.54)	53.83	(32.09,82.54)
Z ₁₁	0.9381	0.9195	0.8908	0.8691	12.69	(7.48,19.61)	53.34	(37.25,72.87)
Z ₁₂	0.8591	0.9029	0.7488	0.7652	19.92	(6.86,42.25)	60.76	(50.29,72.23)
$\bar{Z}$	0.6804	0.5916	0.6378	0.8389	48.05	(11.32,120.47)	57.02	(42.33,74.49)

表 6 最小样本容量 *n*

<i>d</i>	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2
1	3	4	4	5	6	8	9	11	13	16	18	21	24	27	30	> 30	> 30	> 30	> 30	> 30	> 30
2	2	3	3	3	4	4	4	5	5	6	6	7	8	9	9	10	11	12	13	14	16
3	2	2	3	3	3	3	3	4	4	4	4	5	5	5	6	6	6	7	7	8	8
4	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	5	5	5	5	6	6
5	2	2	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	5

位数及第 95% 分位数作为 *Z_i*, $\bar{Z}$ 置信度近似为 90% 的置信区间,见表 5。

2.3 最小样本容量 *n* 的选取

分析已知,氯离子浓度服从对数正态分布,以下讨论均在 对数变换的基础上进行。

由统计推断可知置信水平为 α 的区间估计长度

$$d = 2 \frac{t_{\alpha/2}(n-1)}{\sqrt{n}}s$$

式中：*t_{α/2}*(*n*)为样本容量为 *n* 时的上 $\alpha/2$ 分位数。

在同样的置信水平和精确度下,样本容量越小越节省财力物力。表 6 给出了 $\alpha = 90\%$ 时,不同标准差 *s* 对应的置信区间长度不超过 *d* 时最小样本容量 *n*。

例如在吴淞水厂的数据中,各月中最大的标准差 *s* = 1.31,则取 *s* = 1.4,查表得最小样本容量 *n* = 5,即分析 5 年的样本资料就可以 90% 的置信水平大致掌握冷却水中氯离子浓度的分布特征和变化范围。

同理,在宝钢电厂的数据中,取 *s* = 2.1,查表得最小样本容量 *n* = 8。而吴泾电厂的水质数据标准差较小,如取 *s* = 0.5,查表得最小样本容量 *n* = 3,只要分析 3 年的数据即可。

3 结 论

长江口水域感潮河流^[1]中氯离子浓度近似遵循对数正态分布,利用极大似然估计法可得到氯离子浓度月均值的点估计,利用随机模拟法可得到其置信水平近似为 90% 的置信区间。

当数据采集确实困难时,可用邻近水厂水质数据代替电厂相应日的数据,但仍需结合实际问题的背景做适当的调整。

石洞口电厂和外高桥电厂的测试频率很低,也很随意,有些月份没有测试数据。但它们离吴淞口较近,可先用吴淞水厂当月水中氯离子浓度均值代替这两个电厂该月未测数据,再用统计分析方法预测这些企业的水中氯离子浓度,推算出点估计和区间估计。

参考文献：

[1] 梁磊,周国定,解群.凝汽器使用不锈钢管应注意的几个问题[J].中国电力,2000(8) 84-87.
[2] 茆诗松,王静龙.数理统计[M].上海:华东师范大学出版社,1990.
[3] 董大钧.统计分析软件应用指南[M].北京:电子工业出版社,1993.
[4] 陈希孺,柴根象.非参数统计教程[M].上海:华东师范大学出版社,1993.

(收稿日期 2007-01-03 编辑 舒 建)