

环境试验与观测

基于小波相关分析框架的大气腐蚀 数据挖掘

那中丽^{1a}, 付冬梅^{1a,2}, 孟金桃^{1a}, 裴梓博^{1b}

(1.北京科技大学 a.自动化学院 b.新材料技术研究院, 北京 100083;

2.北京市工业波谱成像工程中心, 北京 100083)

摘要: **目的** 针对大气腐蚀中腐蚀数据复杂多变且不易于分析的问题, 提出一种基于小波的相关分析框架, 挖掘腐蚀的变化特征及其影响因素。**方法** 首先采用小波功率谱确定腐蚀数据的周期波动特征, 然后应用小波相干谱分析腐蚀和环境因素的相关强度、超前滞后及正反相关系, 最后利用平均小波相干和显著相干百分比筛选出与腐蚀相关程度较高的环境因素。然后以青岛碳钢腐蚀数据为具体对象, 挖掘了温度、相对湿度、降雨和污染物对碳钢初期腐蚀行为的影响。**结果** 在腐蚀初期, 碳钢的腐蚀具有很强的日波动特征, 温度和相对湿度是导致其波动的主要环境因素; 腐蚀在降雨时刻呈现间歇性波动特征, 且降雨超前于腐蚀; 污染物浓度较低使得其对碳钢的腐蚀影响较弱。**结论** 基于小波的相关分析框架有效展现了复杂腐蚀数据的变化特征及其与环境因素的关系, 为深入理解大气腐蚀规律提供了一种新的途径。

关键词: 大气腐蚀; 环境因素; 小波变换; 小波相干

中图分类号: TG172.3; TB115

文献标识码: A

文章编号: 1672-9242(2021)07-0119-06

DOI: 10.7643/issn.1672-9242.2021.07.018

Data Mining of Atmospheric Corrosion Based on Wavelet Correlation Analysis Framework

NA Zhong-li^{1a}, FU Dong-mei^{1a,2}, MENG Jin-tao^{1a}, PEI Zi-bo^{1b}

(1. a. School of Automation and Electrical Engineering, b. Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, China; 2. Beijing Engineering Research Center of Industrial Spectrum Imaging, Beijing 100083, China)

ABSTRACT: Aimed at the problem that the corrosion data in atmospheric corrosion is typically complex and not easy to analyze, a wavelet-based correlation analysis framework is proposed to mine the changing characteristics of corrosion and its influence factors. First, the wavelet power spectrum is used to determine the periodic fluctuation characteristics of the corrosion data.

收稿日期: 2021-01-30; 修订日期: 2021-02-24

Received: 2021-01-30; Revised: 2021-02-24

基金项目: 国家重点研发计划(2017YFB0702104)

Fund: Supported by the National Key Research & Development Program of China (2017YFB0702104)

作者简介: 那中丽(1996—), 女, 硕士研究生, 主要研究方向为大气腐蚀数据挖掘。

Biography: NA Zhong-li (1996—), Female, Master, Research focus: atmospheric corrosion data mining.

通讯作者: 付冬梅(1963—), 女, 博士, 教授, 主要研究方向为智能数据分析。

Corresponding author: FU Dong-mei (1963—), Female, Doctor, Professor, Research focus: intelligent data analysis.

引文格式: 那中丽, 付冬梅, 孟金桃, 等. 基于小波相关分析框架的大气腐蚀数据挖掘[J]. 装备环境工程, 2021, 18(7): 119-124.

NA Zhong-li, FU Dong-mei, MENG Jin-tao, et al. Data mining of atmospheric corrosion based on wavelet correlation analysis framework[J]. Equipment environmental engineering, 2021, 18(7): 119-124.

Then, the correlation strength, lead-lag, and phase relationship of corrosion and environmental factors are analyzed using the wavelet coherence spectrum. Finally, the average wavelet coherence and the percent area of significant coherence are used to select the environmental factors of high degree relevance to corrosion. Moreover, corrosion data of carbon steel in Qingdao is taken as the specified object for mining the influence of temperature, relative humidity, rainfall and pollutants on the initial corrosion behavior. Corrosion of carbon steel has obvious daily fluctuation characteristics, and temperature and relative humidity are the main environmental factors that cause its fluctuations; corrosion exhibits intermittent fluctuations at the time of rainfall and the rainfall leads corrosion; the lower concentration of pollutants makes its impact on the corrosion of carbon steel weaker. The wavelet-based correlation analysis framework effectively shows the changing characteristics of complex corrosion data and its relationship with environmental factors. It provides a new way for a deep understanding of atmospheric corrosion laws.

KEY WORDS: atmospheric corrosion; environmental factors; wavelet transform; wavelet coherence

在线监测技术的发展使得获取大气环境下金属材料的腐蚀数据成为可能^[1], 采用适当的数据挖掘方法对这类数据进行挖掘, 能够深入研究腐蚀和大气环境之间的复杂关系, 对于理解大气腐蚀机理有重要意义^[2]。

大多数学者已经利用现代数据分析方法对腐蚀和大气环境的关系进行了研究。Cao^[3]利用灰色关联分析法确定了影响 Q235 钢大气腐蚀的重要环境因素。Li^[4]利用最大信息相关系数说明了相对湿度和温度对腐蚀起着主导作用, SO₂、NO₂、PM2.5 和 PM10 等大气污染物也对腐蚀产生影响。Zhi^[5]建立随机森林模型分析了大气环境因素对低合金钢腐蚀的影响。Shi^[6]分析了大气环境因素与腐蚀速率的 Pearson 相关系数的结果后, 提出了用腐蚀指数来描述环境因素对腐蚀的综合影响。然而, 上述的分析方法都属于时域分析方法, 要求被分析的数据是平稳的, 但是环境数据复杂多变, 并不能长时间保持其平稳性^[7-8]。另外, 时域分析方法虽然直观, 但是无法表征出能反映数据本质特征的频率域的相互关系。小波变换是一种非平稳数据分析方法, 将时域中看似复杂的数据分解为多种周期分量的叠加, 进而识别出时域分析方法不能识别的数据的波动周期及其随时间的变化^[9]。在小波变换基础上, 提出的小波相干能够量化两个非平稳数据在时间和频率上的瞬时关联程度^[10]。

文中将时频域的小波变换方法引入大气腐蚀研究领域, 提出了一种适用于腐蚀数据的相关分析框架。以青岛腐蚀数据为例, 分析了碳钢初期腐蚀过程中腐蚀与环境因素的相关关系。

1 数据与方法

1.1 数据收集与预处理

文中的数据来源于腐蚀站点提供的 45 号钢 (碳钢) 在青岛 2018 年 8 月 2 日至 2018 年 9 月 5 日的腐蚀数据和环境数据^[11]。腐蚀数据即 Fe-Cu 型 ACM 传

感器监测的电偶电流数据。环境数据包括温度、相对湿度、降雨、SO₂、NO₂、PM2.5、PM10 和 AQI。其中 AQI 是 SO₂、NO₂、PM2.5、PM10、CO 和 O₃ 等 6 项污染物的综合指标。电偶电流、温度和相对湿度的采样周期为 1 min, 其余环境数据的采样周期为 1 h, 为了统一数据的采样周期, 将分钟级数据转换为小时级数据, 共获得 812 个数据样本。由于电偶电流数据跨度过大不利于观察, 将电偶电流进行对数化处理^[12]。电偶电流及环境因素的波形如图 1 所示, 可以看出, 电偶电流波形无明显特征, 且高度不规则变化, 难以直接观察出电偶电流的变化规律。此外, 电偶电流的变化与温度、相对湿度具有较高的相关性, 而其余环境因素对电偶电流的影响不易观察出。因此提出一种合适的评估具有复杂变化特征的腐蚀数据的方法对于理解大气腐蚀规律有重要意义。

1.2 基于小波的相关分析框架

由于小波变换的理论基础过于分散, 不利于实际应用, 笔者总结了最相关的理论, 并提出了适用于腐蚀研究的相关分析框架, 如图 2 所示。

1) 小波功率谱。计算电偶电流 A 的小波功率, 并可视化在时间-周期的二维图中, 称之为小波功率谱。

对于数据长度 $M=812$, 且采样间隔 $\Delta t=1$ h 的电偶电流 $A=\{A(t_m), m=0, \dots, M-1\}$ 和 8 个环境因素 $B=[\text{温度}, \text{相对湿度}, \text{降雨}, \text{AQI}, \text{SO}_2, \text{NO}_2, \text{PM2.5}, \text{PM10}]$, $B_i=\{B_i(t_m), m=0, \dots, M-1\}$, 利用式 (1) 分别对电偶电流 A 和 8 个环境因素 B 进行连续小波变换^[9]。

$$X_A(n, s) = \sqrt{\frac{\Delta t}{s}} \sum_{m=0}^{M-1} A(t_m) \psi^* \left(\frac{(m-n)\Delta t}{s} \right) \quad (1)$$

式中: *表示复共轭; $\psi(t)$ 选用 Morlet 小波, n 为平移参数, $n=m\Delta t$, $m=0, \dots, M-1$; s 为尺度参数, $s_j = s_0 2^{jM/\delta t}$, $j=0, 1, \dots, \frac{\delta t}{M} \log_2(M\delta t/s_0)$ 。通过公式 $T=2\pi s/6$ 可以将尺度 s 转换为周期 T (频率的倒数)。连续小波变换受边缘效应影响的区域 (the Cone of

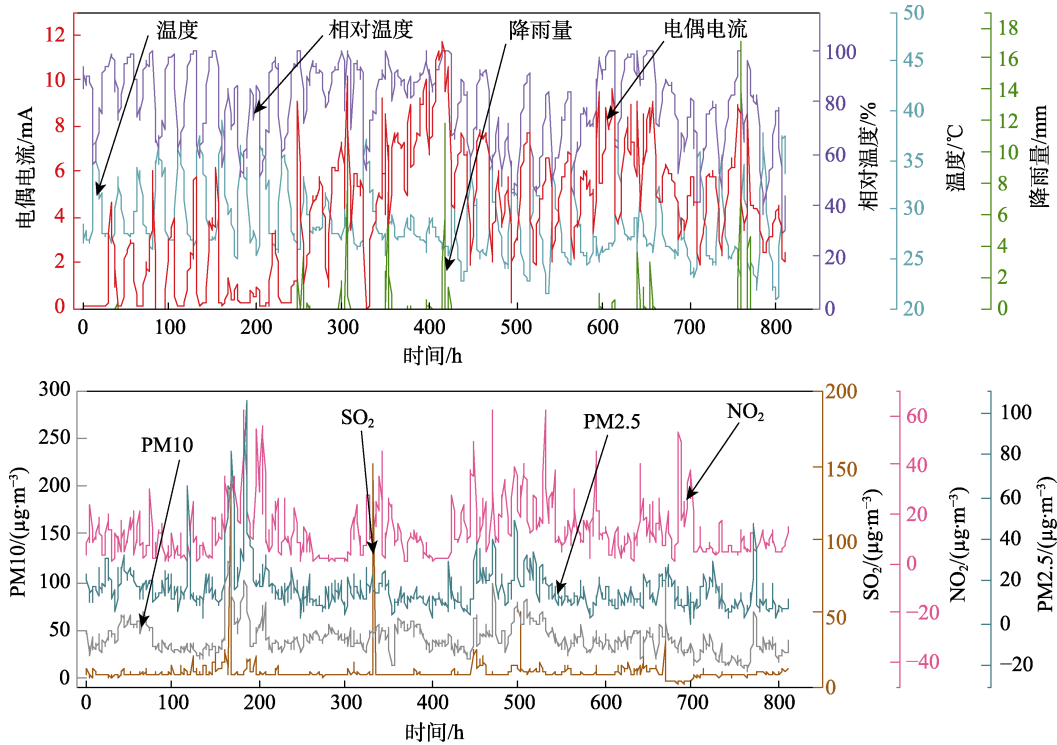


图 1 电偶电流及环境因素的时域波形

Fig.1 Time-domain waveform of galvanic current and environmental factors

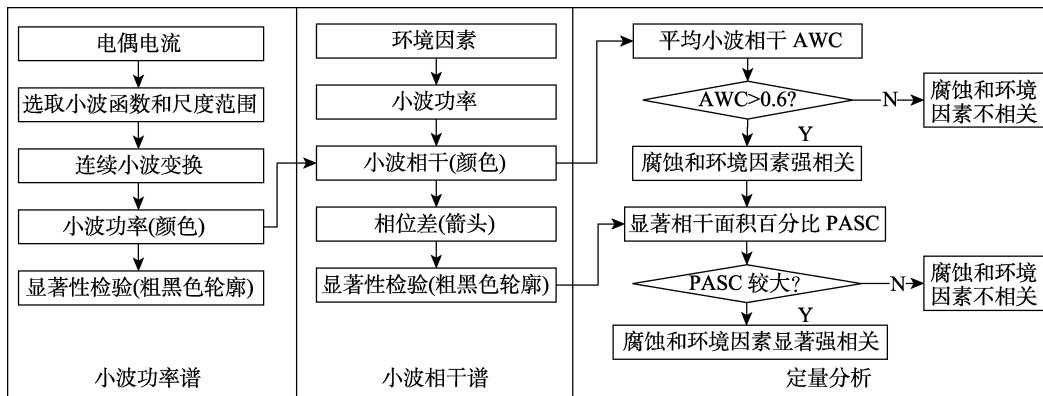


图 2 基于小波的相关分析框架流程

Fig.2 The flowchart of the wavelet-based correlation analysis framework

Influence, COI) 在小波功率谱中用灰色圆锥线及较浅的阴影表示, 此区域内信息可能失真, 因此不进行分析^[13]。

利用式 (2) 计算电偶电流 A 的小波功率, 揭示其周期波动特征, 并在小波功率谱中用颜色表示。

$$X_{AA}(n,s) = X_A(n,s)X_A^*(n,s) = |X_A(n,s)|^2 \quad (2)$$

式中: $X_A(n,s)$ 表示电偶电流 A 的连续小波变换。

2) 小波相干谱。分别计算电偶电流 A 和某个环境因素 $B_i(i=1,\dots,8)$ 的小波相干 (Wavelet Coherence, WC)、相位差, 挖掘其相关性、超前滞后和正反相关系, 并可视化在时间-周期的二维图中, 称之为小波相干谱。

式 (3) 的绝对值称之为 WC, 表征电偶电流 A 和环境因素 B_i 在特定时间和特定频率的相关性, 在

小波相干谱中用颜色表示。取值为 0~1, 越接近 1, 表示相关性越高^[10]。

$$\gamma_{AB}(n,s) = \frac{S(X_{AB}(n,s))}{\sqrt{S(X_{AA}(n,s))S(X_{BB}(n,s))}} \quad (3)$$

式中: S 表示时频域中的平滑运算符^[14]。

$X_{AB}(n,s) = X_{AA}(n,s)X_{BB}^*(n,s)$, $X_{AA}(n,s)$ 和 $X_{BB}(n,s)$ 分别为电偶电流 A 和环境因素 B_i 的小波功率。

通过蒙特卡洛模拟对 WC 进行显著性检验, 重复 1000 次试验后, 获得的 5% 显著性水平 ($p < 0.05$) 说明 WC 在 95% 置信水平下是可靠的, 在小波相干谱中用粗黑色轮廓表示^[15]。

利用式 (4) 计算电偶电流 A 和环境因素 B_i 的相位差:

$$\varphi_{AB}(n,s) = \text{Arctan} \left(\frac{\text{Im}\{S(X_{AB}(n,s))\}}{\text{Re}\{S(X_{AB}(n,s))\}} \right) \quad (4)$$

式中：Im 表示虚部；Re 表示实部。相位差对应于四个象限，表示电偶电流和环境因素的超前滞后和正反相关关系，在小波相干谱中用箭头表示。向左和向右的箭头分别代表反相和同相关关系，同（反）相关关系表明环境因素的增加有利于电偶电流的增强（减小）。箭头为水平方向，表示无超前滞后关系；指向第一和第三象限，表示电偶电流变化超前于环境因素；指向第二和第四象限，表示电偶电流变化滞后于环境因素^[16]。

3) 定量分析。计算平均小波相干(Average Wavelet Coherence, AWC)和显著相干面积百分比(Percent Area of Significant Coherence, PASC)筛选出与电偶电流 A 显著相关的环境因素 B_i , AWC 大于 0.6 且 PASC 较大, 则 A 和 B_i 显著相关, 否则不相关^[17-19]。

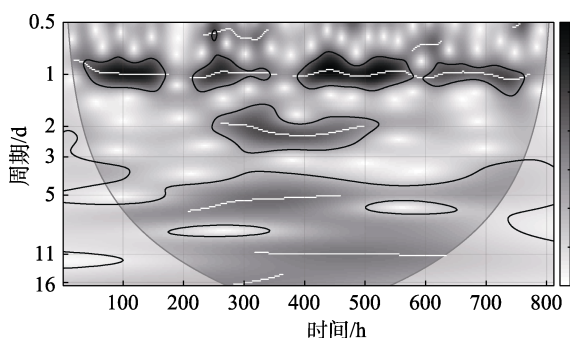
$$AWC_{AB} = \frac{\sum_n \sum_s |S(X_{AB}(n,s))|}{\left(\sum_n \left(\sum_s |S(X_{AA}(n,s))| \right) \sum_s |S(X_{BB}(n,s))| \right)^{1/2}} \quad (5)$$

AWC 表示将 WC 随时间和尺度进行平均^[19], 取值范围为 0~1, AWC 的值越接近 1, 表示相关性越强。PASC 表示通过显著性检验的区域 ($p < 0.05$) 占整个小波相干谱的面积百分比^[20], 取值范围为 0~100%, PASC 越大, 表明相关性通过显著性检验的程度越高。

2 结果及分析

2.1 电偶电流的小波功率谱

将电偶电流进行连续小波变换, 并绘制了小波功率谱(见图3), 直观地显示出了电偶电流的主要波



图片右侧的颜色条表示功率, 范围从白色(低功率)到黑色(高功率), 粗黑色轮廓表示 5% 的显著性水平, 灰色圆锥线及较浅的阴影表示 COI

图3 电偶电流的小波功率谱

Fig.3 The wavelet power spectrum of galvanic current (On the right of the picture is a color bar ranging from white (lower power) to black (higher power), the thick black contour indicates 5% significance level, and the gray cone line and lighter shade indicate COI)

动周期以及这些周期随时间的变化。功率越高, 表示周期越强, 白色条纹表示功率的局部最大值。在周期为 1 d 的几乎所有时间段上, 有一条白色条纹且附近区域内颜色大部分为黑色, 意味着此周期存在一个很强的永久循环, 说明电偶电流存在以 1 d 为主的周期变化特征。同时在周期为 2 d 的 250~500 h、周期为 5 d 的 200~450 h 以及周期为 11 d 的 325~625 h 也发现有白色条纹。由图 1 可看出, 这些时间段内有降雨发生, 初步推断是降雨导致了电偶电流的间歇性波动, 将在 2.2 节中进行讨论。

2.2 电偶电流和环境因素的小波相干谱

为了解驱动电偶电流变化的潜在原因, 绘制电偶电流和环境因素的小波相干谱, 如图 4 所示。

图 4a 显示, 在 0.5~3 d 的周期区间, 电偶电流和相对湿度在所有的时段呈现出强相关性, WC 达到 0.95; 4~7 d 周期区间上的 200~400 h 以及 6~15 d 周期区间上的 300 h 之后, WC 达到 0.9。显著相关区域中的箭头水平向右, 相位差为 0, 表明电偶电流和相对湿度呈现正相关关系且无超前滞后关系。

图 4b 显示, 电偶电流和温度在 0.5~1.5 d 的周期区间内具有强相关性, WC 达到 0.95。除此之外只有几个间歇性的强相关区域, 如 1.5~3 d 周期区间上的 250~350 h 和 600~650 h 以及 4~7 d 周期区间上的 200~400 h, WC 达到 0.9。显著相关区域中的箭头水平向左, 相位差为 π , 表明电偶电流和温度呈现负相关关系, 且无超前滞后关系。

图 4c 中显著相关区域中的箭头大部分指向斜右下方, 范围为 $0 \sim -\pi/2$, 表明电偶电流和降雨呈现正相关关系, 且降雨超前于电偶电流。观察小波相干谱发现, 200 h 前后发生了结构性变化, 在此之前几乎没有强相关区域。在 200~450 h 内, 1.5~3 d 和 4~7 d 的周期区间都具有很高的相关性, WC 达到 0.9。在 450~700 h 内的 4~6 d 的周期区间以及 700~770 h 内的 1.5~3 d 的周期区间也具有强相关性, WC 达到 0.9。降雨的发生时间可以解释这种结构性变化, 由降雨数据可知, 降雨分布在 250~425 h 以及 600~660 h 以及 750~770 h 内。

AQI 和污染物 (NO_2 、 SO_2 、 $\text{PM}_{2.5}$ 和 PM_{10}) 与电偶电流的显著相关区域较小, 不易得到直接的规律。观察各环境因素与电偶电流显著相关性较强的区域, 周期为 1 d 左右的周期区间内, 温度和相对湿度与电偶电流显著相关, 说明温度和相对湿度导致了电偶电流的日波动特征, 同时温度和相对湿度可能存在相互作用。周期为 2、5 d 左右, 电偶电流具有间歇性波动的时间内, 降雨与电偶电流显著相关, 说明降雨导致了电偶电流的间歇性波动特征。同时在这些周期区间, 电偶电流与温度和相对湿度也存在显著相关区域, 说明温度、相对湿度和降雨也可能存在相互作用。

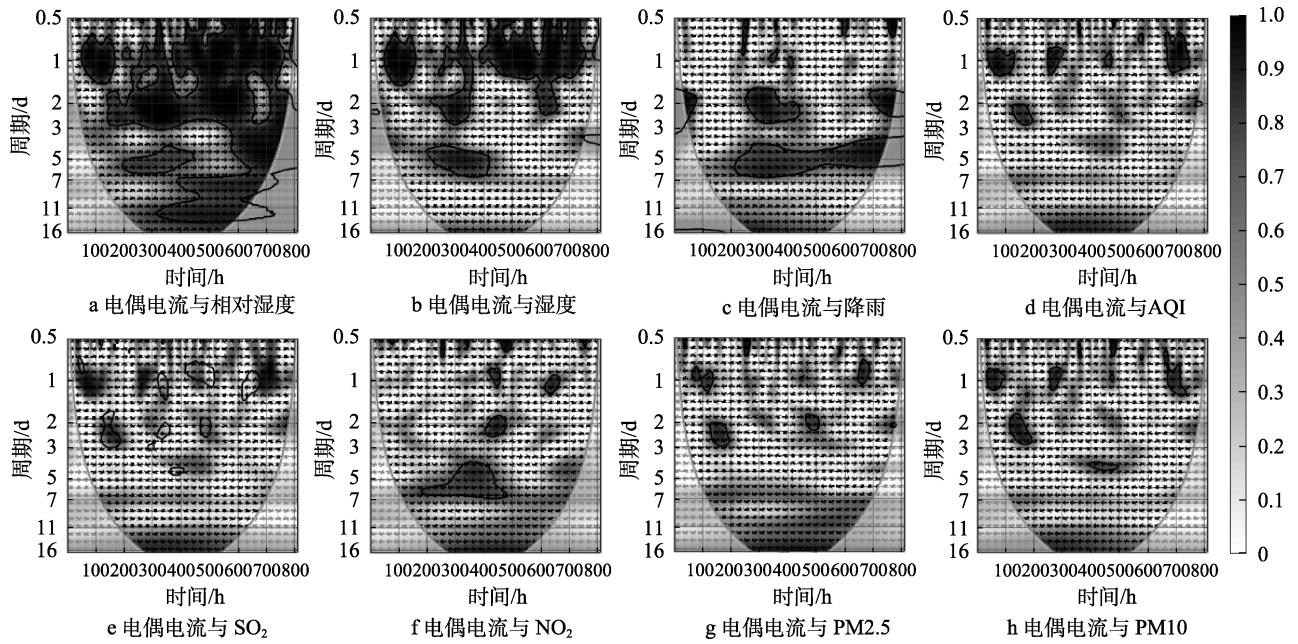


图 4 电偶电流和环境因素的小波相干谱

Fig.4 Wavelet coherence spectrum of galvanic current and environmental factors: a) RH, b) T, c) RF, d) AQI, e) SO₂, f) NO₂, g) PM_{2.5}, h) PM₁₀

用。环境因素间的相关特征及其对电偶电流的作用有待进一步深入研究。

2.3 电偶电流和环境因素的定量分析

由于直接观察小波相干谱难以对电偶电流与环境因素的相关程度进行排序，计算 AWC 以及 PASC 得到大气腐蚀过程中不同环境因素的相对重要性，结果见表 1。一般认为，AWC 大于 0.6 为强相关。由表 1 发现，电偶电流与温度、相对湿度、降雨、NO₂ 和 PM_{2.5} 强相关，但 NO₂ 和 PM_{2.5} 的 PASC 分别为 6.61% 和 3.36%，可能是由于偶然因素造成的强相关性。因

此与电偶电流相关程度较高的环境因素排序为相对湿度>温度>降雨，其中相对湿度对电偶电流的影响远远大于其他环境因素，AWC 为 0.8730，且 PASC 为 39.53%。污染物在青岛初期腐蚀过程中与电偶电流的相关程度不高，可能原因是采样时间段内污染物浓度较低，PM₁₀、PM_{2.5}、SO₂、NO₂ 和 AQI 的平均值分别为 40.4、16.3、9.6、12.9 μg/m³ 和 40.8。AQI 虽然是 SO₂、NO₂、PM₁₀、PM_{2.5}、CO 和 O₃ 等 6 项污染物的综合指数，但不能很好地代替这些污染物对电偶电流的综合影响，如 NO₂ 的 AWC 和 PASC 都大于 AQI 的 AWC 和 PASC。

表 1 电偶电流和环境因素的 AWC 和 PASC
Tab.1 AWC and PASC of galvanic current and environmental factors

指标	温度	相对湿度	降雨	AQI	SO ₂	NO ₂	PM _{2.5}	PM ₁₀
AWC	0.7052	0.8730	0.6612	0.4997	0.4953	0.7067	0.6469	0.4988
PASC/%	22.72	39.53	11.43	5.05	4.18	6.61	3.36	4.22

3 结论

1) 基于小波的相关分析框架，为理解腐蚀过程以及探究环境因素对腐蚀的影响提供了一种新的解决途径。首先它能够清晰地显示出复杂腐蚀过程的变化特征；其次相比于传统的时域分析方法，更全面地揭示了腐蚀和环境因素的相关关系，不仅量化了数据间的整体相关强度，还可以提供相关性发生的时间和频率、超前滞后以及正反相关系。

2) 通过电偶电流的小波功率谱，清晰地显示出腐蚀的周期波动特征及其随时间的变化。结果表明，在采样期间，腐蚀有显著的日波动特征，同时在降雨时刻具有间歇性波动。

3) 基于小波的相关分析框架，挖掘了电偶电流和环境因素的关系，结果表明，与腐蚀相关程度较高的环境因素排序为相对湿度>温度>降雨，其中相对湿度和降雨与腐蚀呈现正相关，且降雨超前于腐蚀，温度与腐蚀呈现负相关。SO₂、NO₂、PM_{2.5} 和 PM₁₀ 浓度较低导致其与腐蚀的相关程度较弱。

参考文献:

- [1] LI X G, ZHANG D W, LIU Z Y, et al. Materials science: Share corrosion data[J]. *Nature news*, 2015, 527(7579): 441-442.
- [2] MENDOZA A R, CORVO F. Outdoor and indoor atmospheric corrosion of carbon steel[J]. *Corrosion science*, 1999, 41(1): 75-86.
- [3] CAO X L, DENG H D, LAN W. Use of the grey relational analysis method to determine the important environmental factors that affect the atmospheric corrosion of Q235 carbon steel[J]. *Anti-corrosion methods and materials*, 2015, 62(1): 7-12.
- [4] LI Z L, FU D M, LI Y, et al. Application of an electrical resistance sensor-based automated corrosion monitor in the study of atmospheric corrosion[J]. *Materials*, 2019, 12(7): 1065-1080.
- [5] ZHI Y J, FU D M, ZHANG D W, et al. Prediction and knowledge mining of outdoor atmospheric corrosion rates of low alloy steels based on the random forests approach[J]. *Metals*, 2019, 9(3): 383-398.
- [6] SHI Y N, FU D M, ZHOU X Y, et al. Data mining to online galvanic current of zinc/copper internet atmospheric corrosion monitor[J]. *Corrosion science*, 2018, 133: 443-450.
- [7] CAI Y K, ZHAO Y, MA X B, et al. Influence of environmental factors on atmospheric corrosion in dynamic environment[J]. *Corrosion science*, 2018, 137: 163-175.
- [8] VADREU K P, CHOI Y. Wavelet analysis of airborne CO₂ measurements and related meteorological parameters over heterogeneous landscapes[J]. *Atmospheric research*, 2011, 102(1-2): 77-90.
- [9] TORRENCE C, COMPO G P. A practical guide to wavelet analysis[J]. *Bulletin of the American Meteorological Society*, 1998, 79(1): 61-78.
- [10] AGUIAR C L, SOARES M J. The continuous wavelet transform: moving beyond uni- and bivariate analysis[J]. *Journal of economic surveys*, 2014, 28(2): 344-375.
- [11] PEI Z B, CHENG X Q, YANG X J, et al. Understanding the environmental impacts on initial atmospheric corrosion based on corrosion monitoring sensors[J]. *Journal of materials science and technology*, 2020, 64: 214-221.
- [12] FINNEY D J. On the distribution of a variate whose logarithm is normally distributed[J]. *Supplement to the journal of the Royal Statistical Society*, 1941, 7(2): 155-161.
- [13] AGUIAR C L, SOARES M J. Oil and the macroeconomy: Using wavelets to analyze old issues[J]. *Empirical economics*, 2011, 40(3): 645-655.
- [14] GRINSTED A, MOORE J C, JEVREJEVA S. Application of the cross wavelet transform and wavelet coherence to geophysical time series[J]. *Nonlinear processes in geophysics*, 2004, 11(5/6): 561-566.
- [15] TIAN F H, TARUMI T, LIU H, et al. Wavelet coherence analysis of dynamic cerebral autoregulation in neonatal hypoxic-ischemic encephalopathy[J]. *NeuroImage clinical*, 2016, 11: 124-132.
- [16] CASAGRANDE E, MUELLER B, MIRALLES D G, et al. Wavelet correlations to reveal multiscale coupling in geophysical systems[J]. *Journal of geophysical research: Atmospheres*, 2015, 120(15): 7555-7572.
- [17] SCHOBER P, BOER C, SCHWARTE L A. Correlation coefficients: Appropriate use and interpretation[J]. *Anesthesia & analgesia*, 2018, 126(5): 1763-1768.
- [18] GUEVARA M A, CORSI C M. EEG coherence or EEG correlation?[J]. *International journal of psychophysiology*, 1996, 23(3): 145-153.
- [19] ANTONIADS A, BROSSAT X, CUGLIARI J, et al. Clustering functional data using wavelets[J]. *International journal of wavelets, multiresolution and information processing*, 2013, 11(1): 1350003-1350033.
- [20] SU L, MIAO C Y, DUAN Q Y, et al. Multiple-wavelet coherence of world's large rivers with meteorological factors and ocean signals[J]. *Journal of geophysical research: Atmospheres*, 2019, 124(9): 4932-4954.