



## Article

# Data Mining to Atmospheric Corrosion Process Based on Evidence Fusion

Jintao Meng <sup>1,2,\*</sup> , Hao Zhang <sup>1</sup>, Xue Wang <sup>1</sup> and Yue Zhao <sup>1</sup> 

<sup>1</sup> Science and Technology on Communication Security Laboratory, Chengdu 610041, China; z.zhao93@gmail.com (H.Z.); wangxuemath@gmail.com (X.W.); yuezhao@foxmail.com (Y.Z.)

<sup>2</sup> School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

\* Correspondence: mengjintao01@126.com

**Abstract:** An electrical resistance sensor-based atmospheric corrosion monitor was employed to study the carbon steel corrosion in outdoor atmospheric environments by recording dynamic corrosion data in real-time. Data mining of collected data contributes to uncovering the underlying mechanism of atmospheric corrosion. In this study, it was found that most statistical correlation coefficients do not adapt to outdoor coupled corrosion data. In order to deal with online coupled data, a new machine learning model is proposed from the viewpoint of information fusion. It aims to quantify the contribution of different environmental factors to atmospheric corrosion in different exposure periods. Compared to the commonly used machine learning models of artificial neural networks and support vector machines in the corrosion research field, the experimental results demonstrated the efficiency and superiority of the proposed model on online corrosion data in terms of measuring the importance of atmospheric factors and corrosion prediction accuracy.

**Keywords:** atmospheric corrosion; carbon steel; data mining; environmental factor; evidence theory



**Citation:** Meng, J.; Zhang, H.; Wang, X.; Zhao, Y. Data Mining to Atmospheric Corrosion Process Based on Evidence Fusion. *Materials* **2021**, *14*, 6954. <https://doi.org/10.3390/ma14226954>

Academic Editor: Enrico Prati

Received: 21 October 2021

Accepted: 12 November 2021

Published: 17 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Atmospheric corrosion, which may cause enormous economic and human life losses, has occurred in many industries such as infrastructure, energy, and transportation [1–3]. Nowadays, variant atmospheric corrosion monitor (ACM) sensors are developed to monitor the real-time corrosion behavior of materials in outdoor dynamic environments [4,5]. The collected big corrosion data contribute to further corrosion science research.

It is generally believed that the mutual behaviors of key environmental factors, including temperature (T), relative humidity (RH), and atmospheric contaminants, lead to atmospheric corrosion [6–8]. In order to better understand the atmospheric corrosion process, the real-time variations in the corrosion rate and environmental factors with time should be analyzed. There exist some related works, as summarized in Table 1. Some studies have been carried out in indoor atmosphere environments. By using ACM sensors, the influence of some important factors on the corrosion rate have been verified, such as wet–dry cycles [9] and short-chain volatile carboxylic acids [4]. Considering the complexity of natural atmosphere environments, indoor corrosion tests are now gradually being extended to outdoor environments. Some empirical relationships between corrosion rates and some atmospheric factors have been found out. The long-term atmospheric corrosion of weathering steel bridges has been monitored by an electrochemical impedance-based ACM, finding that the corrosion rate increases gradually with an increasing RH under 80% [10]. The function between the daily charge of ACM sensors and the corrosion rate has been obtained, identifying that the mutual behaviors of driving history, T, and RH cause the corrosion of automobile parts [11]. In another automation corrosion test, the role of salt spray in metal corrosion was confirmed [12]. The obtained data suggested that 75% of zinc corroded in the salt spray phase [12]. The atmospheric corrosion of carbon steel has been

monitored by ACM, and the impacts of relative humidity, temperature, and rainfall have been identified to be higher than pollution on initial atmospheric corrosion [13]. Exposure tests with ACM have found that rainfall precipitation is the strongest parameter influencing the initial atmospheric corrosion rate in rainy regions, with RH taking the first place in low-precipitation environments and non-rainfall periods [14]. Recently, some statistical correlation coefficients have been employed to measure the influence of key atmospheric factors in the outdoor atmospheric corrosion process. The maximal information coefficient (MIC) well measures the relationship between some factors (RH, T, and pollutants) and the voltage ratio data [5]. Similarly, Pearson's correlation coefficient has been employed to identify the effect of important atmospheric parameters [15].

**Table 1.** Summary of some related works with ACM data.

Papers	Condition	Environment	Atmospheric Factors	Year
[9]	Indoor	Atmosphere	Wet–dry cycles	2014
[4]	Indoor	Atmosphere	T, RH, and formic and acetic acid	2014
[10]	Outdoor	Atmosphere	T, RH	2014
[11]	Outdoor	Automobile	T, RH, and salt particles	2014
[12]	Outdoor	Automobile	T, RH, salt spray, and freeze	2014
[15]	Outdoor	Atmosphere	T, RH, SO <sub>2</sub> , NO <sub>2</sub> , PM2.5, and PM10	2018
[5]	Outdoor	Atmosphere	T, RH, SO <sub>2</sub> , NO <sub>2</sub> , PM2.5, and PM10	2019
[13]	Outdoor	Atmosphere	T, RH, Rainfall, SO <sub>2</sub> , NO <sub>2</sub> , PM2.5, PM10, O <sub>3</sub> , and CO	2020
[14]	Outdoor	Atmosphere	T, RH, and rainfall	2021

Despite the aforementioned outstanding studies, there still exist some gaps: (1) Concerning the influence of environmental factors on outdoor atmospheric corrosion, most of the existing studies focus on qualitative empirical analysis. Although some statistical correlation coefficients have been used to quantify the influence, unreasonable results may occur. Statistical correlation coefficients can accurately measure the correlation of two variables only when all of the other related variables remain unchanged. However, online ACM corrosion rate data are mutually influenced by many atmospheric factors. Actually, it has been demonstrated that there exist hidden relationships between environmental factors and atmospheric corrosion by a hidden Markov model, which could not be well measured by Pearson's correlation coefficient [15]. Therefore, how to accurately quantify such underlying relationships is still a challenge. (2) With the development of artificial intelligence, some machine learning models, including artificial neural networks (ANNs) and support vector machines (SVMs), have been widely used in corrosion science [16–19]. However, these models are not suitable for visually measuring the relationships between key factors and atmospheric corrosion because of their incapable interpretability. They just focus on predicting the corrosion rate by end-to-end model frameworks. Further work to fill the above gaps will contribute to better understanding the atmospheric corrosion mechanism.

Evidence theory [20,21] is a generalization of the traditional probability framework. Given the advantages of information fusion and uncertainty representation, it has been widely used in various tasks [20–22]. The main mechanism of evidence theory is to measure and fuse the information provided by multiple evidence. Therefore, evidence theory provides an effective mathematical framework to deal with coupled data because of its information fusion viewpoint.

The corrosion resistance of different materials in the same environment varies greatly, such as carbon steels, alloy steels, and nanostructured materials [23,24]. This study mainly focuses on the real-time atmospheric corrosion of commonly used Q235 steel. Considering that most statistical correlation coefficients do not adapt to outdoor coupled corrosion data, a new evidence fusion-based model is proposed in order to deal with the collected real-time corrosion data of Q235 steel in an outdoor atmospheric environment. It quantifies the different contributions of key environmental factors to the short-term atmospheric corrosion process.

## 2. Mathematical and Experimental Methods

### 2.1. Mathematical Methods

In this subsection, some necessary basics of evidence theory are introduced. Evidence theory, also called Dempster–Shafer theory or belief functions theory, is a generalization of probability framework [20,21]. It is known for the advantages of information fusion and uncertainty representation. Let  $H_q (q = 1, 2, \dots, c)$  denote  $c$  mutually exclusive and exhaustive solutions for an investigated problem. Then  $\Omega = \{H_1, H_2, \dots, H_c\}$  is called the frame of discernment for this problem. All of the subsets of  $\Omega$  compose the power set  $2^\Omega$ . It is important to note that evidence theory is established on the power set. Assume that  $n$  pieces of evidence measuring the possible solutions can be obtained. The main framework of evidence theory can be briefly summarized as shown in Figure 1. There exist four steps, and the related definitions are detailed as follows.

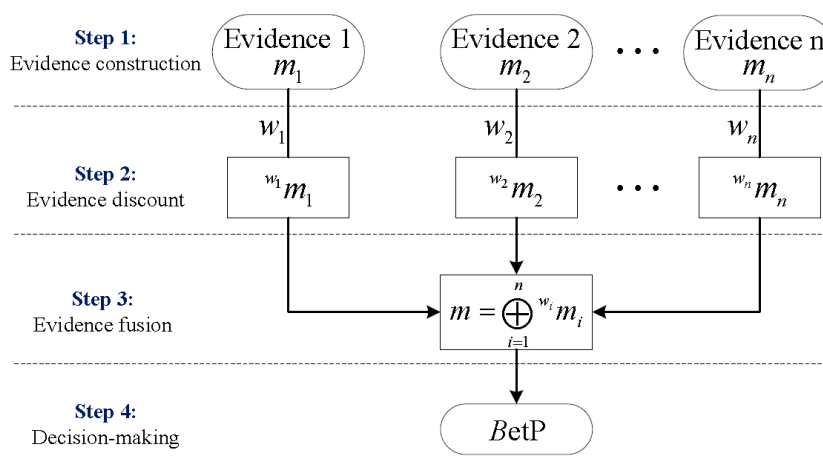


Figure 1. The main framework of evidence theory.

Step 1: A basic probability assignment (BPA) function is defined to measure the uncertain information provided by the evidence. The BPA of evidence  $i$  is defined as a mapping function  $m_i$  from  $2^\Omega$  to the interval  $[0, 1]$ , such that  $m_i(\phi) = 0$  and:

$$\sum_{A \in 2^\Omega} m_i(A) = 1 \tag{1}$$

The subset  $A$  satisfying  $m_i(A) > 0$  is called a focal element. Here,  $m_i(A)$  measures the degree of support of evidence  $i$  to the possible solution  $A$ . Moreover,  $m_i(\Omega)$  indicates the ignorance degree of evidence  $i$  to the investigated problem. Obviously, a BPA will degenerate to a probability measure when it only has singleton focal elements.

Step 2: Considering the incomplete reliability of the evidence, a discounting operation is necessary. Let  $w_i (0 \leq w_i \leq 1)$  denote the reliability coefficient of evidence  $i$ . Then the discounting operation leads to a completely reliable BPA  $w_i m_i$ . For  $\forall A \in 2^\Omega$ :

$$w_i m_i(A) = \begin{cases} w_i m_i(A) & A \neq \Omega \\ w_i m_i(A) + 1 - w_i & A = \Omega \end{cases} \tag{2}$$

Step 3: In order to fuse the multi-source information, different combinations rules have been proposed. Among them, Dempster’s rule [20] is widely used because of its associative and commutative properties. For  $n$  discounted BPAs  $w_i m_i (i = 1, 2, \dots, n)$  provided by  $n$  distinct and independent evidence, the fused BPA  $m = \bigoplus_{i=1}^n w_i m_i$  satisfies that  $m(\phi) = 0$  and for  $\forall A, A_i \in 2^\Omega$ :

$$m(A) = \frac{1}{1 - K} \sum_{\bigcap_{i=1}^n A_i = A} \prod_{i=1}^n w_i m_i(A_i) \tag{3}$$

where  $K = \sum_{\bigcap_{i=1}^n A_i = \phi} \prod_{i=1}^n w_i m_i(A_i)$  measures the degree of conflict among  $n$  discounted pieces

of evidence and takes a value in the interval  $[0, 1]$ . Dempster's rule proportionally redistributes the conflicts among all focal elements when  $K \neq 1$ . If  $K = 1$ , it should be noted that Dempster's rule will be out of operation and another combination rule could be used.

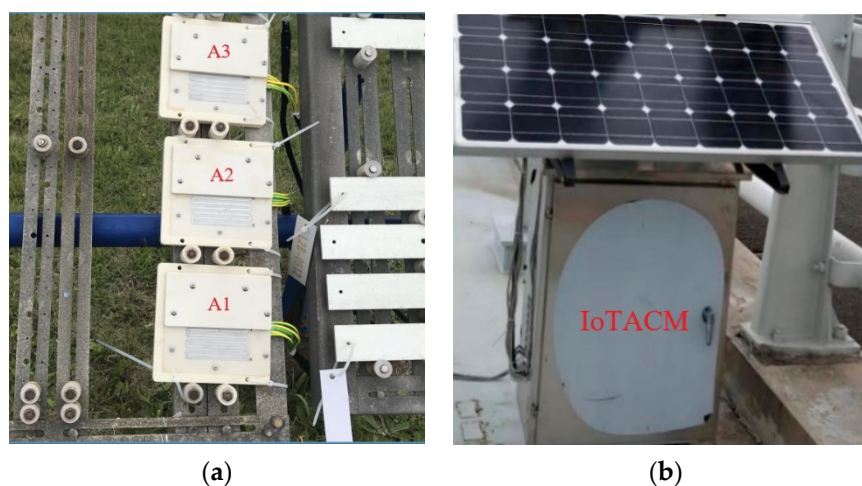
Step 4: After evidence fusion, the most possible solution for the investigated problem can be determined based on the Pignistic probability  $BetP$  related to the BPA  $m$ . In order to obtain a solution without ambiguity, this paper just considers  $BetP = [BetP_1, BetP_2, \dots, BetP_c]$ . For  $H_i (i = 1, 2, \dots, c)$ :

$$BetP(H_i) = \sum_{H_i \in B, B \in 2\Omega} \frac{m(B)}{|B|} \quad (4)$$

where  $|B|$  is the cardinality of subset  $B$  and  $\sum_{i=1}^c BetP(H_i) = 1$ .

## 2.2. Field Exposure Test

The atmospheric corrosion monitoring tests in this study were conducted at Qingdao (36.07 degrees north latitude and 120.44 degrees east longitude), eastern China, which has a coastal humid temperature climate [25]. The principle and availability of the employed Internet of Things Atmospheric Corrosion Monitor (IoTACM) were detailed in our previous research [5]. Following the related instructions [5], the real-time corrosion rate of specimen can be calculated based on the real-time voltage ratio and initial height of the specimen. In this study, three Q235 steel specimens were tested from 25 September 2017 to 24 December 2017, lasting for three months. The sizes (length  $\times$  width  $\times$  height) of the A1, A2, and A3 specimens were 70 mm  $\times$  70 mm  $\times$  0.15 mm, 70 mm  $\times$  70 mm  $\times$  0.4 mm, and 70 mm  $\times$  70 mm  $\times$  0.4 mm, respectively. Figure 2 shows some details of the field test. The specimens were installed in an exposure shelf toward the south with a 45-degree angle to the ground. The IoTACM was placed approximately 10 m from the shelf. In addition, humidity and temperature sensors were also placed next to the shelf to record the real-time RH and T of the current environment.

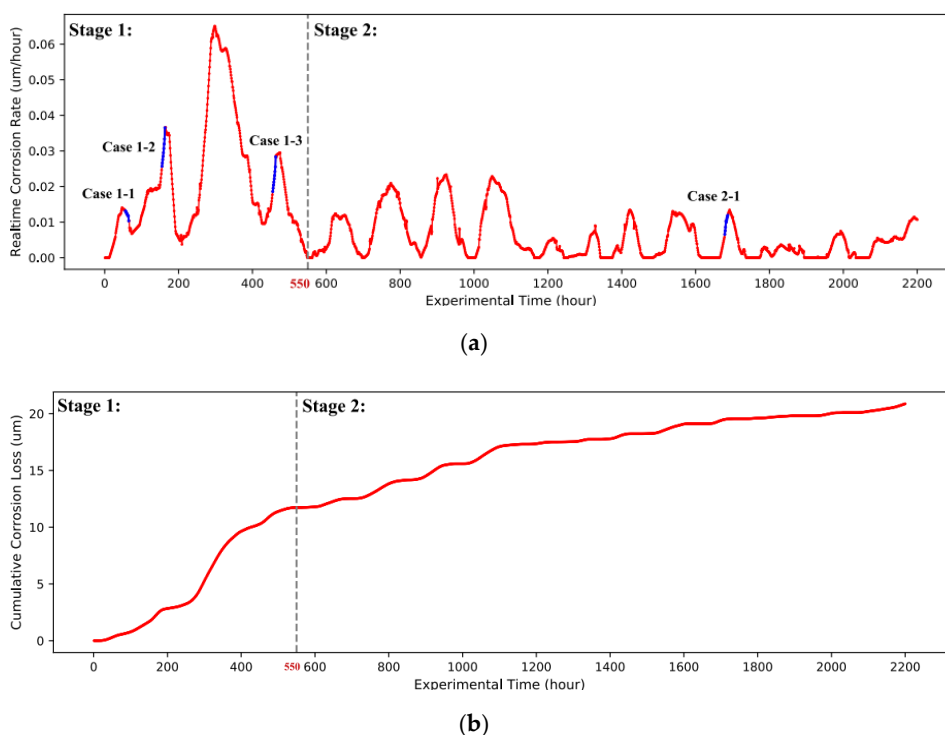


**Figure 2.** The field corrosion tests at Qingdao, China. (a) Three Q235 steel specimens; (b) ACM.

## 2.3. Field Corrosion Data

In this study, the real-time corrosion rates of three Q235 steel specimens and RH and T data were recorded hourly by IoTACM. There were 2200 data samples without missing values. Although only small errors existed among the corrosion rate data of three specimens, the average real-time corrosion rate of A1, A2, and A3 specimens was evaluated so as to improve the reliability of the data. Figure 3a draws the variations of the real-time corrosion rate ( $y$ -axis) with exposure time. Similarly, Figure 3b draws the variations of

cumulative corrosion loss ( $y$ -axis) with exposure time. Apparently, the corrosion rate significantly fluctuated in a large range and the fluctuation gradually flattened out over time. This is consistent with the power law that the specimens were seriously corroded in the early period, while the corrosion gradually turned slightly with time [26]. In this study, the test period was divided into two stages in order to reveal their underlying different corrosion mechanisms. We took the moment with 50% of the total cumulative corrosion loss as the separation of two stages. Accordingly, as shown in Figure 3a, the corrosion process in test period [0, 550] is defined as Stage 1 (23 days, nearly one month), and the subsequent period [550, 2200] refers to Stage 2. In this case, above 50% of the total cumulative corrosion loss was caused in Stage 1 within only 25% of the total test time. The mean corrosion rate in Stage 1 was  $0.021 \mu\text{m}/\text{h}$ , which is 3.5 times that in Stage 2 ( $0.006 \mu\text{m}/\text{h}$ ). It is reasonable to take Stages 1 and 2 as the fast and slow corrosion periods, respectively.



**Figure 3.** Variations of corrosion data with exposure time. (a) Corrosion rate ( $\mu\text{m}/\text{h}$ ); (b) cumulative corrosion loss ( $\mu\text{m}$ ).

Furthermore, four typical air pollutants— $\text{SO}_2$ ,  $\text{NO}_2$ ,  $\text{PM}_{2.5}$ , and  $\text{PM}_{10}$ —and the air quality index (AQI), which are generally regarded as important corrosive agents [5,15], were taken into consideration in this study. The corresponding pollutant data at Qingdao during the same test period were collected from the National Urban Air Quality Real-time Publishing Platform. Moreover the distance between the selected meteorological station and the corrosion test area was 5.4 km. It should be noted that the chloride deposition rate was not included in this work, given the following reasons. First, the impact of chloride is cumulative and becomes more dominant for long-term corrosion under marine atmospheres, while it shows little difference to corrosion losses of carbon steel coupons despite the very different chloride deposition rates in short-term exposure tests [27–29]. This study mainly focused on short-term corrosion tests lasting for three months. Second, the average chloride deposition rates in October, November, and December 2017 in the experimental site were  $15.38 \text{ mgm}^{-2}\text{d}^{-1}$ ,  $15.91 \text{ mgm}^{-2}\text{d}^{-1}$ , and  $4.45 \text{ mgm}^{-2}\text{d}^{-1}$ , respectively. Compared to the chloride deposition rates ranging from  $0\sim 2000 \text{ mgm}^{-2}\text{d}^{-1}$  reported in related studies [27,28], the above values reflect low chloride concentrations during the exposure tests, which have less effect on the corrosion of Q235 steels in service. Third, the chloride deposition rate is typically recorded at least monthly following ISO 9225 standard [30].



To the best of authors' knowledge, there is no technique that can precisely record the hourly chloride deposition in field atmospheric environments, especially for our case where the deposition is low and the test duration is short. Last but not the least, the corrosion tests in the same site as this study indicate that the formation of non-protective akageneite induced by chloride enrichment was not found in the corrosion product layer of the specimens after one month of exposure [31]. Additionally, the average chloride deposition rate in reference [31] was  $71.03 \text{ mgm}^{-2}\text{d}^{-1}$ , which is much higher than the chloride deposition rate during the whole exposure tests in this study. To sum up, it is reasonable to consider that chlorides also did not play a major role in this present study.

Table 2 summarizes the characteristics of the collected field data in each stage. It shows a wide gap in the corrosion rate between Stages 1 and 2. Although the RH in both stages shared similar maximum and standard deviation values, the mean RH in Stage 1 exceeded that in Stage 2. Because of the coming of winter, the T gradually decreased with time. Accordingly, Table 2 shows an obvious lower T in Stage 2 than in Stage 1. According to the data of the contaminants ( $\text{SO}_2$ ,  $\text{NO}_2$ ,  $\text{PM}_{2.5}$ , and  $\text{PM}_{10}$ ), wider variation ranges and bigger average values of pollutants were found in Stage 2. This means that the air pollution in Stage 2 was significantly more serious than in Stage 1.

**Table 2.** Characteristics of the field data in Stage 1 (Stage 2).

Index	$R_t$	RH	T	AQI	$\text{SO}_2$	$\text{NO}_2$	$\text{PM}_{2.5}$	$\text{PM}_{10}$
Max	0.065 (0.023)	84 (79)	26 (19)	97 (289)	45 (163)	112 (412)	28 (98)	131 (135)
Min	0 (0)	25 (17)	17 (1)	20 (5)	1 (1)	8 (1)	1 (1)	1 (4)
Mean	0.021 (0.006)	58 (42)	21 (10)	44 (76)	13 (37)	45 (97)	6 (15)	29 (51)
Standard Deviation	0.017 (0.006)	16 (14)	2 (5)	14 (32)	10 (31)	21 (51)	5 (11)	22 (28)

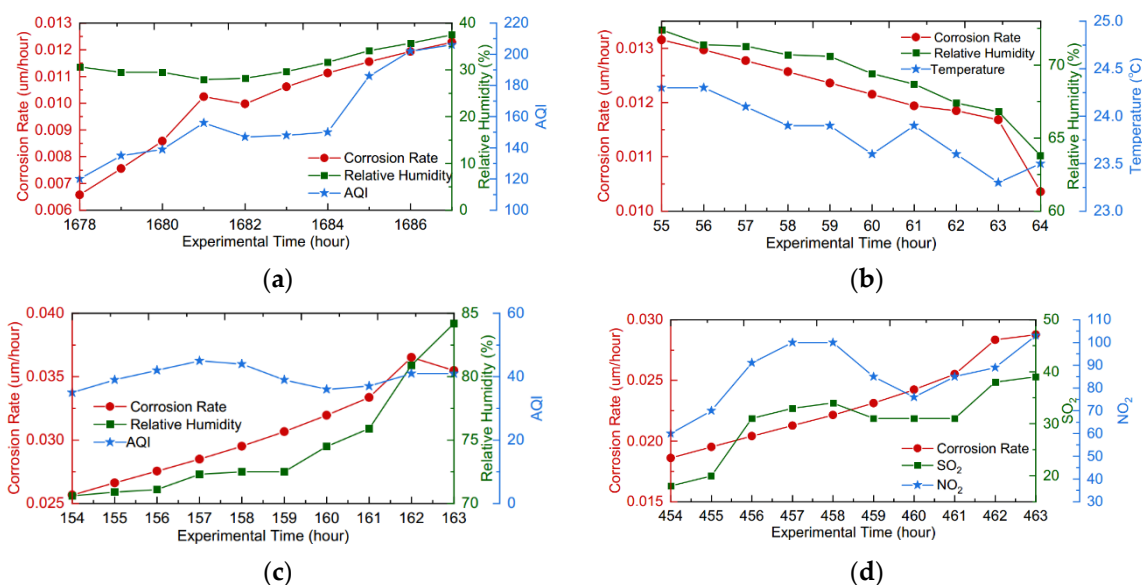
Notes:  $R_t$  is the corrosion rate at time  $t$ . The units of  $R_t$ , RH, and T are  $\mu\text{m}/\text{h}$ , %, and  $^\circ\text{C}$ , respectively. The units of  $\text{SO}_2$ ,  $\text{NO}_2$ ,  $\text{PM}_{2.5}$ , and  $\text{PM}_{10}$  are all  $\mu\text{g}/\text{m}^3$ .

### 3. Results and Discussion

In this section, the impact of different atmospheric factors on the corrosion rate is analyzed from the viewpoints of qualitative analysis, statistical quantitative analysis, and information fusion-based quantitative analysis.

#### 3.1. Qualitative Analysis

In order to visually show the impact of different atmospheric factors on the corrosion rate, the curves of four illustrative cases marked in Figure 3a are drawn. The left  $y$ -axis in Figure 4 measures the real-time corrosion rate. Meanwhile, the right  $y$ -axes in different colors measure the concentrations of corresponding environmental factors. As shown in Figure 4b, the decreasing corrosion rate was associated with the gradual reduction in RH and T, while the concentrations of the pollutants remained stable in this period. The period of Figure 4c met with relatively stable values of T and AQI. Moreover, the rise in the corrosion rate with time was significantly influenced by the rapid increase in RH. In the above two cases, the variances of RH and T resulted in thickness variations of the thin electrolyte layer, which directly affected the corrosion rate. Similarly, the RH and T in the periods of Figure 4a,d basically remained stable. However, the air pollutions gradually increased as a whole, especially in the period of Figure 4a. Accordingly, there was a rapid increase in the corresponding real-time corrosion rate. The corrosive components of the pollutants might be dissolved on the surface of specimens, which accelerated the corrosion behavior. The above analyses lead to the conclusion that the RH, T, and these contaminants are positively correlated with the corrosion rate.



**Figure 4.** Curves of the four illustrative cases marked in Figure 3a: (a) Case 2-1; (b) Case 1-1; (c) Case 1-2; (d) Case 1-3.

### 3.2. Statistical Quantitative Analysis

In order to quantify the influence of atmospheric factors on corrosion rate, the four commonly used statistical correlation indexes of Pearson's correlation coefficient (PCC) [15], Kendall's rank correlation coefficient (KRCC) [32], Spearman's correlation coefficient (SCC) [32], and the maximal information coefficient (MIC) [14] were employed in this subsection. PCC applies to measuring the linear relationship between two variables, SCC and KRCC, focusing on monotonic correlation, while the MIC is capable of measuring nonlinear relationships. The ranges of PCC, SCC, and KRCC are all  $[-1, 1]$ , where a negative value indicates a negative correlation and vice versa. The MIC takes a value in the  $[0, 1]$  range, where a larger value means a stronger correlation between two variables. The results of Stages 1 and 2 are summarized in Tables 3 and 4, respectively. Given the complexity of corrosion behavior, there exist nonlinear relationships between atmospheric factors and corrosion rate. According to the values of the MIC index in Tables 3 and 4, the RH and T have stronger correlations with the corrosion rate in both Stages 1 and 2. Compared to Stage 1, the smaller MIC values of the RH and T in Stage 2 indicate their decreased impact. However, the results of the statistical correlation analysis also lead to the following problems:

- (1) In terms of the influence of the atmospheric factor on the corrosion rate, four statistical indexes derive inconsistent ranks of atmospheric factors in Stage 1. The same thing also occurs in Stage 2. They do not even agree on the most important factor in Stage 2. This indicates that selecting one certain statistical index to evaluate the influences of different atmospheric factors is subjective.
- (2) In Stage 1, the negative values of T and four contaminants by PCC, KRCC, and SCC indicate their negative correlations with the corrosion rate. However, this is inconsistent with the qualitative analysis results in Section 3.1. Similar things also happen in Stage 2. A possible reason is that statistical correlation coefficients can exactly measure the correlation between two variables only when all of the other related variables remain unchanged. The online corrosion rate data were the interactions of multiple atmospheric variables; therefore, unreasonable statistical results occurred.
- (3) Regardless of the linear and nonlinear indexes, none of the factors had a correlation value above 0.5. This means low correlations between all factors and the corrosion rate. The possible reason is that the coupled corrosion rate data hinder the statistical indexes from accurately quantifying true correlations. Actually, a hidden relationship between

atmospheric factors and the corrosion rate has been demonstrated by a hidden Markov model [15]. How to accurately quantify the hidden influence remains challenging.

**Table 3.** Statistical correlations of atmospheric factors and corrosion rate in Stage 1.

Index	RH	T	AQI	SO <sub>2</sub>	NO <sub>2</sub>	PM2.5	PM10
PCC	0.4765	−0.2801	−0.2347	−0.1932	−0.2584	−0.3026	−0.2150
KRCC	0.3350	−0.1661	−0.1568	−0.1055	−0.1792	−0.2374	−0.1724
SCC	0.4789	−0.2060	−0.2453	−0.1593	−0.2695	−0.3300	−0.2582
MIC	0.4225	0.3373	0.2889	0.3079	0.2685	0.3191	0.2876

**Table 4.** Statistical correlations of atmospheric factors and corrosion rate in Stage 2.

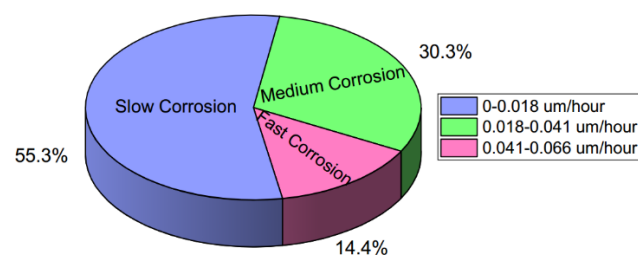
Index	RH	T	AQI	SO <sub>2</sub>	NO <sub>2</sub>	PM2.5	PM10
PCC	0.3113	0.2812	0.1188	0.1059	0.1308	−0.0285	−0.2019
KRCC	0.2222	0.1471	0.0937	0.0786	0.0855	−0.0052	−0.1545
SCC	0.3271	0.2265	0.1344	0.1229	0.1220	−0.0067	−0.2250
MIC	0.2108	0.2493	0.1475	0.1661	0.1476	0.1327	0.1827

### 3.3. Evidence Fusion-Based Quantitative Analysis

From the viewpoint of evidence fusion, a new machine learning model to quantify the hidden influence of atmospheric factors is developed in this subsection. The following model construction process is established in Stage 1, and the results of the proposed model of Stage 2 are directly given.

The corrosion rate data in Stage 1 were classified into the three corrosion degrees of fast, medium, and slow. A commonly used unsupervised K-means method (the function *K Means* in the *Python* package of *Sklearn*) [33] was used to avoid the subjectivity of manual classification. The results are shown in Figure 5. For the real-time corrosion rate  $R_t$  at exposure time  $t$ , the corresponding corrosion degree  $y_t = [y^f, y^m, y^s]$  was defined as follows:

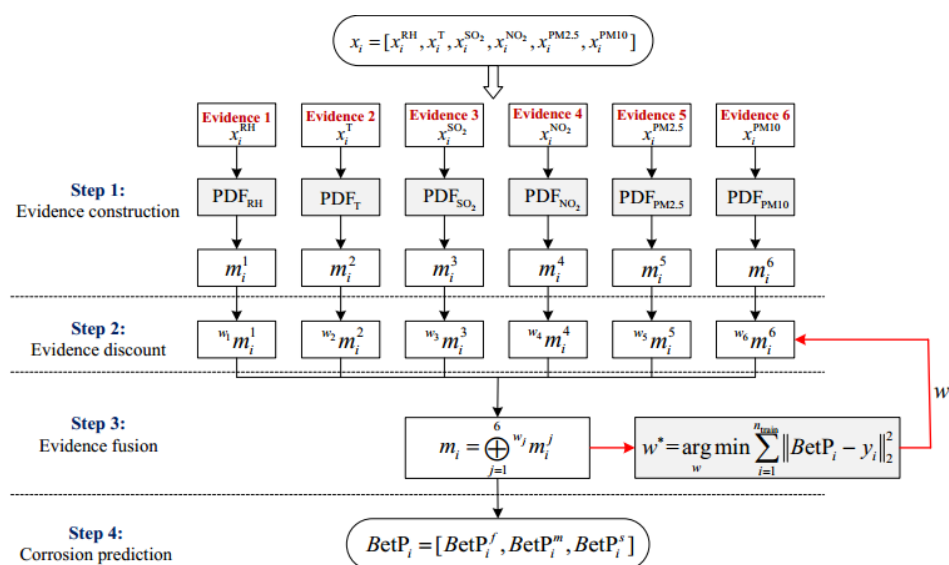
$$y_t = \begin{cases} [0, 0, 1] & R_t \in [0, 0.018] \\ [0, 1, 0] & R_t \in (0.018, 0.041] \\ [1, 0, 0] & R_t \in (0.041, 0.066] \end{cases} \quad (5)$$



**Figure 5.** The classification of the corrosion rate on Stage 1.

Considering that the combined behaviors of multiple factors lead to atmospheric corrosion, the factors of RH, T, SO<sub>2</sub>, NO<sub>2</sub>, PM2.5, and PM10 can be regarded as different independent pieces of evidence providing support information to the current degree of corrosion. It should be noted that AQI is not an independent evidence because it is a synthetic index related to other investigated pollutants. The framework of the proposed model is shown in Figure 6.





**Remark:** The contents in grey blocks are learned on training dataset. And red arrows are only performed on the training stage.

**Figure 6.** The framework of the proposed model.

Step 1: Evidence construction. Let  $x_i = [x_i^{RH}, x_i^T, x_i^{SO_2}, x_i^{NO_2}, x_i^{PM2.5}, x_i^{PM10}]$  denote the atmospheric factor data sample whose label of the degree of corrosion is  $y_t = [y^f, y^m, y^s]$ . The atmospheric factor dataset on Stage 1 is  $X = \{x_i\}_{i=1}^{550}$ . The frame of discernment in this degree of corrosion classification problem is  $\Omega = \{f, m, s\}$ . Let  $X_{train}$  and  $X_{test}$  denote the training and test sets, respectively. Let  $x^{jk}$  denote the data of the atmospheric factor  $j$  ( $j \in [RH, T, SO_2, NO_2, PM2.5, PM10]$ ) belonging to the degree of corrosion  $k$  ( $k \in \Omega$ ). Moreover,  $UB_{jk}$  and  $LB_{jk}$  denote the maximum and minimum value of  $x^{jk}$  in the training set, respectively. Considering of the universality of Gaussian distribution in the natural world, the Gaussian kernel density estimator [34] was employed to calculate the probability density function (PDF)  $f_{jk}$  in the training set:

$$f_{jk}(x) = \begin{cases} \frac{1}{n_{train}} \sum_{x^{jk} \in X_{train}} \frac{1}{\sqrt{2\pi}h} \exp^{-(x-x^{jk})^2/2h} & \text{if } x \in [LB_{jk}, UB_{jk}] \\ \delta_0 & \text{otherwise} \end{cases} \tag{6}$$

where  $h = (4\sigma/3n_{train})^2$ ,  $\sigma$  is the standard deviation of  $x^{jk}$  in the training set and  $n_{train}$  is the total number of samples in the training set [34]. It should be noted that  $\delta_0$  is a positive value close to zero so as to avoid generating completely conflicting evidence. In this study, let  $\delta_0 = 0.001$ .

As shown in Figure 6, the support information for the degree of corrosion  $k$  provided by evidence  $j$  is proportional to its intersection  $f_{jk}(x_i^j)$  with the PDF model  $f_{jk}$  [35]. Accordingly, the basic probability assignment function  $m_i^j$  of evidence  $j$  can be calculated by two rules in reference [35].

Step 2: Evidence discount. Considering the different levels of importance of different pieces of evidence, the evidence discounting operation is necessary to obtain reasonable fusion results. Let  $w_j$  ( $0 \leq w_j \leq 1$ ) denote the importance coefficient of evidence  $j$  whose initial value equals 1 before training. For evidence  $m_i^j$ , the discounted evidence  $w_j m_i^j$  can be calculated as follows. For  $\forall A \subset \Omega$ :

$$w_j m_i^j(A) = \begin{cases} w_j \cdot m_i^j(A) & \text{if } A \neq \Omega \\ w_j \cdot m_i^j(A) + 1 - w_j & \text{otherwise} \end{cases} \tag{7}$$

Steps 3 and 4: Evidence fusion and corrosion prediction. The different pieces of discounted evidence can be fused by Dempster's rule [20].

$$m_i = \bigoplus_{j=1}^6 w_j m_i^j \quad (8)$$

Then, the predicted degree of corrosion  $BetP_i = [BetP_i^f, BetP_i^m, BetP_i^s]$  can be derived by the Pignistic probability transformation [36]. For  $\forall k \in \Omega$ :

$$BetP_i^k = \sum_{k \in B, B \subset \Omega} \frac{m_i(B)}{|B|} \quad (9)$$

where  $|B|$  is the cardinality of set  $B$ .

In order to determine the correct importance  $w$  ( $w = [w_1, w_2, \dots, w_6]$ ) of atmospheric factor evidence, the vector  $w$  was optimized on the training set. The objective function  $f_{\text{cost}}$  is devoted to minimizing the error between the predicted degree of corrosion and the true degree of corrosion on the training set.

$$f_{\text{cost}} = \min_w \sum_{i=1}^{n_{\text{train}}} \|BetP_i - y_i\|_2^2 \quad (10)$$

By the commonly used simulated annealing optimization method [37], the optimal importance vector  $w^*$  was derived and used to verify the proposed model on the test set.

In this study, 80% of the data in each stage was randomly selected as the training set and the remainder as the test set. The widely used models of ANN and SVM were also tested. The average prediction precision of these three models on the test set in five random experiments was compared. The importance  $w^*$  of each atmospheric factor given by our model is reported.

The experimental results of Stage 1 are given in Tables 5 and 6. The corresponding analyses are as follows: (1) According to Table 5, from the viewpoint of evidence fusion, T is the greatest contributor to atmospheric corrosion in Stage 1, followed by RH and SO<sub>2</sub>. (2) Comparing RH and T, the contaminants of SO<sub>2</sub>, NO<sub>2</sub>, PM10, and PM2.5 have less influence in the interactions on atmospheric corrosion in the initial atmospheric corrosion process. (3) Comparing the statistical correlation coefficients in Table 3, the proposed model found higher correlations between the atmospheric factors and corrosion rate. (4) According to Table 6, comparing ANN and SVM, the proposed model performed best in predicting the degree of corrosion of Q235 steel.

**Table 5.** The importance of each atmospheric factor in Stage 1 (after 550 hours of exposure).

Factor	RH	T	SO <sub>2</sub>	NO <sub>2</sub>	PM2.5	PM10
$w^*$	0.9031	0.9881	0.8910	0.4491	0.1175	0.4208

**Table 6.** The average prediction precision in Stage 1 (after 550 hours of exposure).

Algorithm	SVM	ANN	Ours
Accuracy	62.00%	75.63%	75.63%

Similarly, the experimental results of Stage 2 are summarized in Tables 7 and 8. The corresponding analyses are as follows: (1) According to Table 7, in Stage 2, T still contributes most to atmospheric corrosion among all of the investigated factors, followed by NO<sub>2</sub> and SO<sub>2</sub>. (2) According to Table 2, the mean RH in Stage 2 is obviously lower than in Stage 1. Accordingly, the proposed model derived a lower impact of RH in Stage 2 than in Stage 1. (3) As introduced in Table 2, the test period of Stage 2 suffered more serious air pollution.

Accordingly, compared to Stage 1, the proposed model found higher correlations between the corrosion rate and contaminants of NO<sub>2</sub>, PM2.5, and PM10 in Stage 2. (4) According to Table 7, the contaminants of SO<sub>2</sub> and NO<sub>2</sub> have more influence on atmospheric corrosion than RH. Following Figure 3b, the corrosion behavior in Stage 2 gradually weakened over time. The possible reason is that Stage 1 generated corrosion products. They isolated the metal surface from the atmosphere such that the adhesion of water droplets on the metal surface was affected. Therefore, the impact of RH was weakened in Stage 2, while some contaminants contributed more to the specimen's corrosion because of the ability of damaging the existing corrosion products. (5) According to Table 8, the proposed model outperformed SVM and ANN in terms of corrosion predication.

**Table 7.** The importance of each atmospheric factor in Stage 2 (after 2200 hours of exposure).

Factor	RH	T	SO <sub>2</sub>	NO <sub>2</sub>	PM2.5	PM10
<i>w</i> *	0.7821	0.9601	0.8547	0.9464	0.7146	0.6385

**Table 8.** The average prediction precision in Stage 2 (after 2200 hours of exposure).

Algorithm	SVM	ANN	Ours
Accuracy	74.44%	72.57%	80.88%

To sum up, according to the results of the qualitative analysis and the statistical quantitative analysis on the exposure test data of Q235 steel at Qingdao, China, it was found that most statistical correlation coefficients did not adapt to the outdoor coupled corrosion data. Therefore, a new evidence fusion-based model was proposed. According to the results of the evidence fusion-based quantitative analysis, the proposed model can discover the influence of different environmental factors on carbon steel corrosion in different exposure test periods, and can accurately predict the corrosion rate.

#### 4. Conclusions

Based on the exposure tests of Q235 steel at Qingdao, China, the main findings and novelty of this work are as follows:

- (1) It was found that most statistical correlation coefficients do not adapt to coupled data analysis. They can accurately measure the correlation between two variables only when all of the other related variables remain unchanged, while the outdoor online ACM corrosion data are mutually influenced by many atmospheric factors.
- (2) A new evidence fusion-based machine learning model was proposed, which adapts to deal with coupled corrosion data because of the advantage of information fusion. Processing online corrosion data from the viewpoint of evidence theory initiates a new field of corrosion research.
- (3) The proposed model can not only measure the influence of different environmental factors on atmospheric corrosion, but can also predict the corrosion rate.
- (4) Comparing the commonly used machine learning models of ANN and SVM in corrosion research field, the proposed model can obtain more accurate corrosion prediction results.
- (5) According to the proposed model, relative humidity, temperature, and SO<sub>2</sub> are the main factors affecting atmospheric corrosion among the investigated environmental factors.
- (6) According to the proposed model, because of the ability of damaging the corrosion products generated in initial corrosion process, SO<sub>2</sub> and NO<sub>2</sub> showed greater impact on atmospheric corrosion than relative humidity in the later corrosion period.

Our future work will focus on the atmospheric corrosion analysis on different climatic conditions and long-term corrosion tests.

**Author Contributions:** Conceptualization, J.M.; formal analysis, X.W. and H.Z.; funding acquisition, H.Z. and Y.Z.; methodology, J.M.; software, J.M.; validation, Y.Z. and X.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Environmental Corrosion Platform of China and Sichuan Science and Technology Program 2021JDRC0072.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Li, X.; Zhang, D.; Liu, Z.; Li, Z.; Du, C.; Dong, C. Materials science: Share corrosion data. *Nature* **2015**, *527*, 441–442. [[CrossRef](#)]
2. Xu, D.; Gu, T. Carbon source starvation triggered more aggressive corrosion against carbon steel by the *Desulfovibrio vulgaris* biofilm. *Int. Biodeterior. Biodegradation* **2014**, *91*, 74–81. [[CrossRef](#)]
3. Cui, Z.; Ge, F.; Li, X.; Zhu, M.; Xiao, K.; Dong, C.; Wang, X. Mechanistic studies of atmospheric corrosion behavior of Al and Al-based alloys in a tropical marine environment. *J. Wuhan Univ. Technol. Sci. Ed.* **2017**, *32*, 633–639. [[CrossRef](#)]
4. Prošek, T.; Taube, M.; Dubois, F.; Thierry, D. Application of automated electrical resistance sensors for measurement of corrosion rate of copper, bronze and iron in model indoor atmospheres containing short-chain volatile carboxylic acids. *Corros. Sci.* **2014**, *87*, 376–382. [[CrossRef](#)]
5. Li, Z.; Fu, D.; Li, Y.; Wang, G.; Meng, J.; Zhang, D.; Yang, Z.; Ding, G.; Zhao, J. Application of an electrical resistance sensor-based automated corrosion monitor in the study of atmospheric corrosion. *Materials* **2019**, *12*, 1065. [[CrossRef](#)] [[PubMed](#)]
6. Feliu, S.; Morcillo, M.; Feliu, S., Jr. The prediction of atmospheric corrosion from meteorological and pollution parameters-I. *Annu. Corros. Corros. Sci.* **1993**, *34*, 403–414. [[CrossRef](#)]
7. Roberge, P.; Klassen, R.; Haberecht, P. Atmospheric corrosivity modeling—A review. *Mater. Des.* **2002**, *23*, 321–330. [[CrossRef](#)]
8. Schindelholz, E.; Cong, H.; Jove-Colon, C.; Li, S.; Ohlhausen, J.; Moffat, H. Electrochemical aspects of copper atmospheric corrosion in the presence of sodium chloride. *Electrochim. Acta* **2018**, *276*, 194–206. [[CrossRef](#)]
9. Thee, C.; Hao, L.; Dong, J.; Mu, X.; Wei, X.; Li, X.; Ke, W. Atmospheric corrosion monitoring of a weathering steel under an electrolyte film in cyclic wet–dry condition. *Corros. Sci.* **2014**, *78*, 130–137. [[CrossRef](#)]
10. Nishikata, A.; Zhu, Q.; Tada, E. Long-term monitoring of atmospheric corrosion at weathering steel bridges by an electrochemical impedance method. *Corros. Sci.* **2014**, *87*, 80–88. [[CrossRef](#)]
11. Mizuno, D.; Suzuki, S.; Fujita, S.; Hara, N. Corrosion monitoring and materials selection for automotive environments by using Atmospheric Corrosion Monitor (ACM) sensor. *Corros. Sci.* **2014**, *83*, 217–225. [[CrossRef](#)]
12. Prošek, T.; Le Bozec, N.; Thierry, D. Application of automated corrosion sensors for monitoring the rate of corrosion during accelerated corrosion tests. *Mater. Corros.* **2014**, *65*, 448–456. [[CrossRef](#)]
13. Pei, Z.; Zhang, D.; Zhi, Y.; Yang, T.; Jin, L.; Fu, D.; Cheng, X.; Terry, H.A.; Mol, J.M.; Li, X. Towards understanding and prediction of atmospheric corrosion of an Fe/Cu corrosion sensor via machine learning. *Corros. Sci.* **2020**, *170*, 108697. [[CrossRef](#)]
14. Pei, Z.; Cheng, X.; Yang, X.; Li, Q.; Xia, C.; Zhang, D.; Li, X. Understanding environmental impacts on initial atmospheric corrosion based on corrosion monitoring sensors. *J. Mater. Sci. Technol.* **2021**, *64*, 214–221. [[CrossRef](#)]
15. Shi, Y.; Fu, D.; Zhou, X.; Yang, T.; Zhi, Y.; Pei, Z.; Zhang, D.; Shao, L. Data mining to online galvanic current of zinc/copper Internet atmospheric corrosion monitor. *Corros. Sci.* **2018**, *133*, 443–450. [[CrossRef](#)]
16. Liu, J.; Wang, H.; Yuan, Z.-G. Forecast model for inner corrosion rate of oil pipeline based on PSO-SVM. *Int. J. Simul. Process. Model.* **2012**, *7*, 74–80. [[CrossRef](#)]
17. Jiménez-Come, M.J.; Turias, I.J.; Ruiz-Aguilar, J.J.; Trujillo, F.J. Characterization of pitting corrosion of stainless steel using artificial neural networks. *Mater. Corros.* **2015**, *66*, 1084–1091. [[CrossRef](#)]
18. Feng, S.; Zhou, H.; Dong, H. Application of deep transfer learning to predicting crystal structures of inorganic substances. *Comput. Mater. Sci.* **2021**, *195*, 110476. [[CrossRef](#)]
19. Zhi, Y.; Fu, D.; Yang, T.; Zhang, D.; Li, X.; Pei, Z.; Li, X. Long-term prediction on atmospheric corrosion data series of carbon steel in China based on NGBM(1,1) model and genetic algorithm. *Anti-Corros. Methods Mater.* **2019**, *66*, 403–411. [[CrossRef](#)]
20. Dempster, A.P. Upper and lower probabilities induced by a multivalued mapping. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 57–72.
21. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976; Volume 42.
22. Meng, J.; Fu, D.; Tang, Y. Belief-peaks clustering based on fuzzy label propagation. *Appl. Intell.* **2020**, *50*, 1259–1271. [[CrossRef](#)]
23. Marcus, P.; Maurice, V.; Strehblow, H.-H. Localized corrosion (pitting): A model of passivity breakdown including the role of the oxide layer nanostructure. *Corros. Sci.* **2008**, *50*, 2698–2704. [[CrossRef](#)]

24. Shuang, S.; Ding, Z.; Chung, D.; Shi, S.-Q.; Yang, Y. Corrosion resistant nanostructured eutectic high entropy alloy. *Corros. Sci.* **2020**, *164*, 108315. [[CrossRef](#)]
25. Li, L.; Yan, D.; Xu, S.; Huang, M.; Wang, X.; Xie, S. Characteristics and source distribution of air pollution in winter in Qingdao, eastern China. *Environ. Pollut.* **2017**, *224*, 44–53. [[CrossRef](#)] [[PubMed](#)]
26. Pan, C.; Lv, W.; Wang, Z.; Su, W.; Wang, C.; Liu, S. Atmospheric Corrosion of Copper Exposed in a Simulated Coastal-Industrial Atmosphere. *J. Mater. Sci. Technol.* **2017**, *33*, 587–595. [[CrossRef](#)]
27. Ma, Y.; Li, Y.; Wang, F. Corrosion of low carbon steel in atmospheric environments of different chloride content. *Corros. Sci.* **2009**, *51*, 997–1006. [[CrossRef](#)]
28. Alcántara, J.; Chico, B.; Díaz, I.; de la Fuente, D.; Morcillo, M. Airborne chloride deposit and its effect on marine atmospheric corrosion of mild steel. *Corros. Sci.* **2015**, *97*, 74–88. [[CrossRef](#)]
29. de la Fuente, D.; Castaño, J.; Morcillo, M. Long-term atmospheric corrosion of zinc. *Corros. Sci.* **2007**, *49*, 1420–1436. [[CrossRef](#)]
30. Hu, M.; He, L.-Y.; Zhang, Y.-H.; Wang, M.; Kim, Y.P.; Moon, K. Seasonal variation of ionic species in fine particles at Qingdao, China. *Atmos. Environ.* **2002**, *36*, 5853–5859. [[CrossRef](#)]
31. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Lander, E.S.; Mitzenmacher, M.; Sabeti, P.C. Detecting Novel Associations in Large Data Sets. *Science* **2011**, *334*, 1518–1524. [[CrossRef](#)] [[PubMed](#)]
32. Colwell, D.J.; Gillett, J.R. Spearman versus Kendall. *Math. Gaz.* **1982**, *66*, 307–309. [[CrossRef](#)]
33. Kanungo, T.; Mount, D.; Netanyahu, N.; Piatko, C.; Silverman, R.; Wu, A. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881–892. [[CrossRef](#)]
34. Bolancé, C.; Guillen, M.; Nielsen, J.P. Kernel density estimation of actuarial loss functions. *Insur. Math. Econ.* **2003**, *32*, 19–36. [[CrossRef](#)]
35. Xu, P.; Su, X.; Mahadevan, S.; Li, C.; Deng, Y. A non-parametric method to determine basic probability assignment for classification problems. *Appl. Intell.* **2014**, *41*, 681–693. [[CrossRef](#)]
36. Smets, P.; Kennes, R. The transferable belief model. *Artif. Intell.* **1994**, *66*, 191–234. [[CrossRef](#)]
37. Hwang, C. Simulated annealing: Theory and applications. *Acta Appl. Math.* **1988**, *12*, 108–111.