Towards understanding and prediction of atmospheric corrosion of an Fe/Cu corrosion sensor via machine learning

Pei, Zibo; Zhang, Dawei; Zhi, Yuanjie; Yang, Tao; Jin, Lulu; Fu, Dongmei; Cheng, Xuequn; Terryn, Herman A.; Mol, Johannes M.C.; Li, Xiaogang

DOI
10.1016/j.corsci.2020.108697

Publication date
2020

Document Version
Final published version

Published in
Corrosion Science

Corrosion Science 170 (2020) 108697

Contents lists available at ScienceDirect

Corrosion Science

journal homepage: www.elsevier.com/locate/corsci

Towards understanding and prediction of atmospheric corrosion of an Fe/Cu corrosion sensor via machine learning

Zibo Pei[a], Dawei Zhang[a,*], Yuanjie Zhi[b], Tao Yang[c], Lulu Jin[a], Dongmei Fu[c], Xuequn Cheng[a], Herman A. Terryn[d,e], Johannes M.C. Mol[e], Xiaogang Li[a,*]

[a] Beijing Advanced Innovation Center for Materials Genome Engineering, National Materials Corrosion and Protection Data Center, Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing, China
[b] School of Electronics and Information, Northwestern Polytechnic University, Xi'an, China
[c] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China
[d] Department of Materials and Chemistry, Research Group Electrochemical and Surface Engineering, Vrije Universiteit Brussel, Brussels, Belgium
[e] Department of Materials Science and Engineering, Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Keywords:
Atmospheric corrosion
Corrosion monitoring
Machine learning
Corrosion prediction

ABSTRACT

The atmospheric corrosion of carbon steel was monitored by a Fe/Cu type galvanic corrosion sensor for 34 days. Using a random forest (RF)-based machine learning approach, the impacts of relative humidity, temperature and rainfall were identified to be higher than those of airborne particles, sulfur dioxide, nitrogen dioxide, carbon monoxide and ozone on the initial atmospheric corrosion. The RF model demonstrated higher accuracy than artificial neural network (ANN) and support vector regression (SVR) models in predicting instantaneous atmospheric corrosion. The model accuracy can be further improved after taking into consideration of the significant effect of rust formation on the sensor.

1. Introduction

Atmospheric corrosion is known to widely impact infrastructure, transportation, energy and other industries and generates high maintenance cost globally [1,2]. Atmospheric corrosion is influenced by the instantaneous (e.g. relative humidity/RH) and cumulative (e.g. deposition of chloride) effects from the constituents of local environments on the metal surface [3]. An accurate and real time evaluation of atmospheric corrosion provides an important guide to materials selection and engineering design for corrosion mitigation. Monitoring sensors have been used in vehicles and bridges to track the dynamic process of atmospheric corrosion and to understand how such a process is influenced by the complex environmental parameters [4,5]. For example, non-electrochemical sensors can monitor the damage status of the substrate materials based on the measurements of the change in weight, electrical resistance [6,7] or acoustic properties [8–10]. An electrochemical sensor, on the other hand, functions by forming a closed electrochemical cell on the sensor surface and monitor the transient corrosion activities by measuring galvanic corrosion current [4], electrochemical impedance [11,12], linear polarization resistance [13] or electrochemical noise [14,15].

The principle of galvanic-cell type atmospheric corrosion monitoring (ACM) sensor is based on galvanic corrosion between two electrodes with different electrochemical activities. The galvanic corrosion current of the couple is then measured by a galvanometer to reflect the instantaneous corrosion rate responding to changes of the environments [16]. Galvanic-cell type ACM sensors are sensitive to environmental variation and are useful in understanding the corrosion in a changing environment [16,17]. For example, Mizuno et al. installed Fe/Ag-type corrosion sensors on various parts of a vehicle for three months and showed that the atmospheric corrosion varied at the locations with difference in the RH and rainwater exposure [17]. In our earlier work, we analyzed the atmospheric corrosion of a zinc/copper ACM sensor in Beijing in a winter time and described that the air quality index had a high impact on atmospheric corrosion [4]. More recently, Fe/Cu-type ACM sensors were used in six outdoor atmospheric environments for one month [18]. Statistical analyses of the current values at different time showed that rainfall accounted for only 16.3–29.7 % of the total test duration but contributed to 64.6–89.0 % of the atmospheric corrosion of the carbon steels. And the low temperature at night was more conducive to the condensation of moisture, leading to a more severe atmospheric corrosion.

Compared with the monthly or yearly data obtained from the standard corrosion coupons, the minutely or hourly data obtained from

* Corresponding authors.
E-mail addresses: dzhang@ustb.edu.cn (D. Zhang), lixiaogang@ustb.edu.cn (X. Li).

https://doi.org/10.1016/j.corsci.2020.108697
Received 26 March 2020; Received in revised form 19 April 2020; Accepted 21 April 2020
Available online 29 April 2020

ACM studies are often large in quantity and contain rich information about corrosion kinetics as influenced by multiple environmental factors such as RH, temperature and pollutants. Because of such a complexity, it is difficult to establish causal relationships between individual environmental factor and the corrosion process based on physics-based corrosion laws or predict the corrosion life of materials [19–22]. In this case, machine learning, which automates the searching for knowledge by learning from example data and experience without relying on predetermined equations, may offer new opportunities to better understand and predict atmospheric corrosion. For example, artificial neural network (ANN) is the most commonly used data mining method [23,24]. It has been applied to model multivariant corrosion processes to predict crack growth rates [19,20] and corrosion initiation time of metals [25]. Support vector regression (SVR) analysis has also been used for corrosion prediction. While ANN is more accurate with a large sample size, SVR may present good prediction results even with a small sample size [26]. Using SVR modelling, Fang et al. predicted the corrosion rate of zinc and steel by temperature, TOW, exposure time, sulfur dioxide and chloride concentration [27]. Wen et al. applied SVR in the prediction of the corrosion rate of 3C steel in seawater influenced by temperature, dissolved oxygen, salinity, pH value and oxidation-reduction potential [28]. Other methods that were less commonly used but are also suitable for mining time-series corrosion data include Markov chains [29,30], grey analyses [31,32] and Monte-Carlo simulations [33,34].

Due to the deeper layers of model structure than general machine learning models, random forest (RF) models possess a good processing ability for data with high variability [35,36]. Thus, RF models are expected to be more suitable for the dynamic atmospheric corrosion processes. For example, the impacts of multiple environmental factors and alloying elements to the corrosion rates of 17 low alloy steels under six different environments over 16 years were quantitatively ranked by a RF approach. The results showed that the environmental factors played more important roles than the chemical compositions of the steels especially for the initial period of atmospheric corrosion [37]. The pH of rainwater showed the highest influence to the corrosion rate among the environmental factors for the timespan over 16 years, whereas the importance of rainfall was more prominent in the first two years of atmospheric corrosion.

This study aims at describing the atmospheric corrosion on an Fe/Cu type galvanic corrosion sensor exposed in Qingdao, China for 34 days. RF-based machine learning approach was employed to analyze the contribution of different environmental factors (i.e. temperature, RH, rainfall, sulfur dioxide, nitrogen dioxide, carbon monoxide, ozone and particulate pollutants) to the ACM sensor output. After the most influential parameters were identified, a rational RF model was built for the prediction of the dynamic variation of atmospheric corrosion on carbon steels.

2. Methods

2.1. Preparation of the ACM sensors

Fig. 1a shows the assembly of an ACM sensor, which consisted of seven pairs of steel/copper galvanic couple. The carbon steel (0.47 wt% C, 0.18 wt% Si, 0.59 wt% Mn, 0.01 wt% S, 0.01 wt% P, 0.01 wt% Ni, 0.02 wt% Cr, 0.01 wt% Cu) served as the anode, whereas copper (> 99.5 % pure) was used as the cathode. The electrode pairs were separated by glass fiber-reinforced epoxy (FR4) boards for electrical insulation. The exposed area of each electrode was $21 \times 1\ mm^2$. The electrode was individually connected with a wire, assembled together and impregnated in epoxy to obtain the ACM sensor. The anodes and cathodes were connected to the different ends of a galvanometer. The sensor surface was abraded to #1200 using SiC paper prior to the exposure tests.

When a thin electrolyte layer is formed or a droplet is deposited across the insulating board (0.1 mm thick) between the steel and copper electrodes of the ACM sensor, the two electrodes were electrically connected and generate a galvanic corrosion current. Due to the separation of anodes and cathodes by the insulating boards, the galvanic current passes through a galvanometer. And the ACM current (I_{ACM}) would be detected and recorded. The relationship between the I_{ACM} and the corrosion rate of the steel anode was reported in our previous study [18]. In other words, by measuring the I_{ACM}, it is possible to quantitatively evaluate the atmospheric corrosion of the carbon steel.

2.2. Field exposure test

The ACM was conducted at the Qingdao atmospheric corrosion test site (120°25′E, 36°03′N) in the National Environmental Corrosion Platform of China (Fig. 1b). The test site is a coastal region and is located 25 m away the Bohai sea. The ACM sensors were installed at a distance of over 1 m above ground and were facing with an angle of 45° to the south. The temperature and humidity sensors were exposed under the same conditions near the ACM sensor. The exposure test was carried out for 34 days from August 2nd to September 5th, 2018. The galvanometer measured the current with an acquisition frequency of two seconds per reading, and recorded the data once per minute. The measuring error was less than 0.5 %, and the resolution of the galvanometer was 0.1 nA, according to the specification of the manufacturer. The current value ranged from 0.1 nA to 50 mA, beyond which the current could not be determined.

During the entire test time, to a minimum of 97 % of the detected currents were higher than 0.1 nA, indicating that the ACM sensor captured most of the corrosion signals. Each set of data accumulated in this study consisted of I_{ACM}, temperature (T) and RH data from the ACM sensor, and other environmental data such as rainfall status (rainy or not), the concentrations of sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), ozone (O_3) and airborne particles smaller than or equal to 2.5 microns ($PM_{2.5}$) and those smaller than or equal to 10 microns (PM_{10}). The data for SO_2, NO_2, CO, O_3, $PM_{2.5}$ and PM_{10} were obtained from China Meteorological Administration. Chloride deposition at the test site was measured monthly using the dry plate method according to ISO 9225 [38].

2.3. Models

2.3.1. Random forests

The RF model used in this study was implemented with the Scikit-learning machine learning library. Fig. 2a illustrates the prediction process of the RF method. The RF model consists of 100 (an optimal value shown in Fig. S1a; See Supplementary material 1) classification and regression tree (CART) models, and each CART model possessed ability for the training and prediction. T, RH, rainfall status, SO_2, NO_2, O_3, CO, $PM_{2.5}$ and PM_{10} are the input variables. For each CART in the RF model, the imported sets of data were produced by bootstrap sampling [36]. Bootstrap sampling guarantees the imported sets of data are different between each other, and leads to the diversity of the outputs for different CARTs. Then the corrosion current ($I_{ACM\text{-}i,}$ $i = 1, 2, \ldots, 100$) was predicted by the i^-th CART model. The final prediction result for I_{ACM} of RF model was the average of the prediction values of all $I_{ACM\text{-}i}$. Previous studies have shown that a single CART model may be unstable and overfitting when the sample size is small, but this negative effect is alleviated by averaging the outputs of different CARTs assembled in the RF model [37,39]. Therefore, the RF approach constructs and combines several weak models to form a strong model.

In each CART model, the inputs were regarded as vectors and distributed in a space. Fig. 2b shows the training and prediction processes of each CART model. The training process of a CART was to split the input-space to individual subspaces based on the theory of regression trees [40]. The output of each subspace was assigned with a constant value based on the average of all training vectors. After the training was

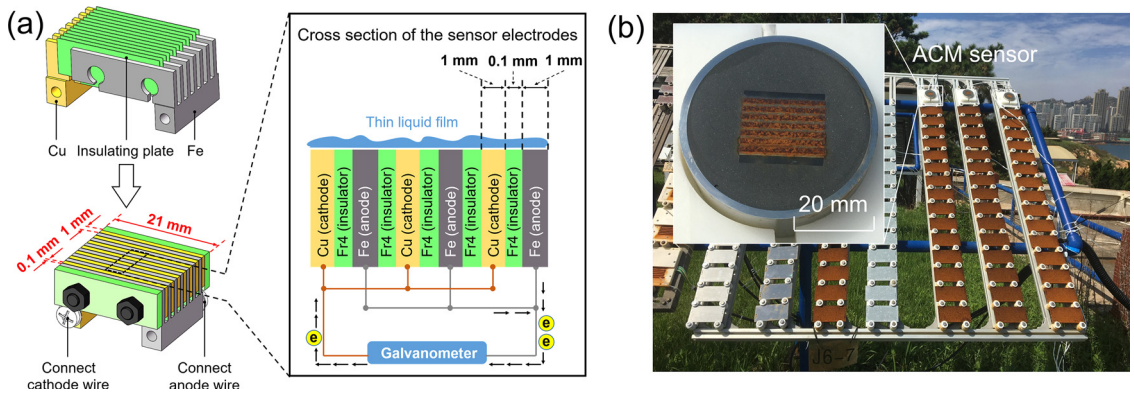

Fig. 1. (a) Schematics showing the design of the Fe-Cu type ACM sensor and (b) Photograph of the test setup on field.

finished, the prediction process of the CART was to import a data without output and locate the subspace of the imported data. Then the prediction value was acquired based on the trained subspace. The training processes of different CART models adopted the same principle. Compared with other ensemble methods, the training process of RF model is based on the random selection of training samples and inputs at each node, i.e. each subspace. Therefore, the base model of RF has more diversity and leads to more precise ensemble results. The important parameters of the RF modelling are the minimum number of samples per subspace and number of CART. When the sample size of a node is less than a certain value, i.e., the integral number of training samples divided by 500 (an optimal value determined in Fig. S1b), the next node segmentation would be stopped, i.e. stop splitting up new subspaces, and the present node would be the leaf node. Meanwhile, at each node of the CART, a subset of the feature vectors is randomly selected.

The data samples unselected for training in each CART model in the bootstrap sampling step, termed as the out-of-bag (OOB) samples, can be used to calculate the importance of the inputs [37]. For each CART, the importance is mainly calculated by adding a disturbance to each input of the OOB data and then evaluating the variation amplitude of the predicted results. By comparing the variation amplitudes affected by different inputs, the importance of different inputs in one CART can be obtained. Finally, by averaging the importance results of all CARTs in the RF model, the quantification of the importance of different inputs is finished.

2.3.2. ANN and SVR modelling

For comparison, back-propagation ANN and SVR were also implemented with the Scikit-learning machine learning library in this study. The input and output variables remain the same as those for the RF model. For the ANN model, the rectified linear unit (ReLu) function was used as the activation function. The model has one hidden layer consisting of 100 units (default value). The random gradient descent method was employed for the parameter tuning. The error tolerance was set as 1×10^{-4}, and the maximum iteration number was 100 after parameter optimization as shown in Fig. S1c, d. For the SVR model, the radial basis function (RBF) was set as the kernel function and the penalty coefficient as 300 (Fig. S1e, f). The width of RBF was 1, which was selected to avoid poor fitting and reduce prediction errors after several practical simulations. Other remaining parameters were the default values of the Scikit-learning machine learning library. More details about the working mechanisms of ANN and SVR models can be found in Supplementary Material 2.

2.3.3. Evaluation criteria

The fitting errors of the training and prediction samples from the different models were evaluated based on R^2 calculated by Eq. (1) and also RMSE calculated by Eq. (2).

$$R^2 = 1 - \frac{\sum_{n=1}^{N} (\hat{y}_n - y_n)^2}{\sum_{n=1}^{N} (\overline{y} - y_n)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y)^2}{N}} \quad (2)$$

where N is the total number of the dataset; $\hat{y}_n$ and y_n represent the prediction value and the true value of the n-th sample, respectively; $y = \sum_{n=1}^{N} \hat{y}_n / N$ represents the mean of the true values of all samples. The fitting error is lower if R^2 approaches 1 and RMSE is smaller.

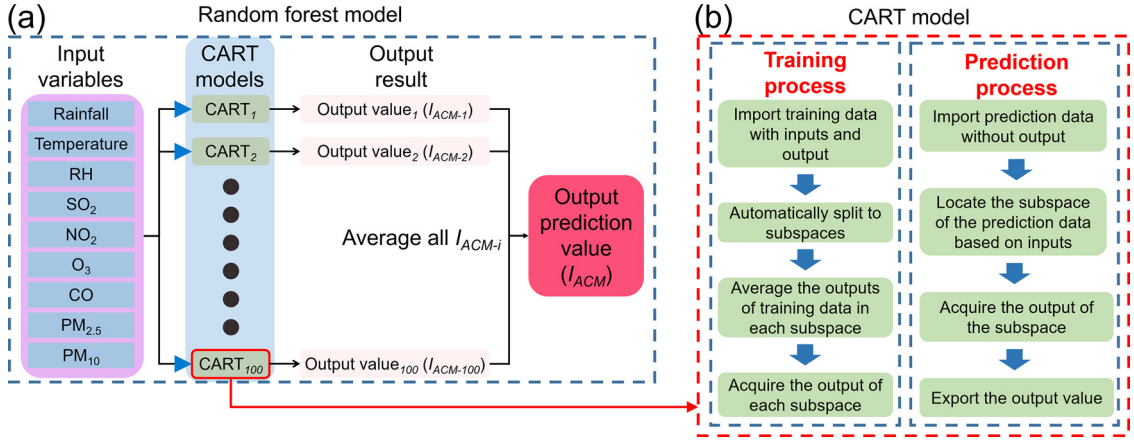

Fig. 2. The training and prediction process of (a) the RF model and (b) each CART model.

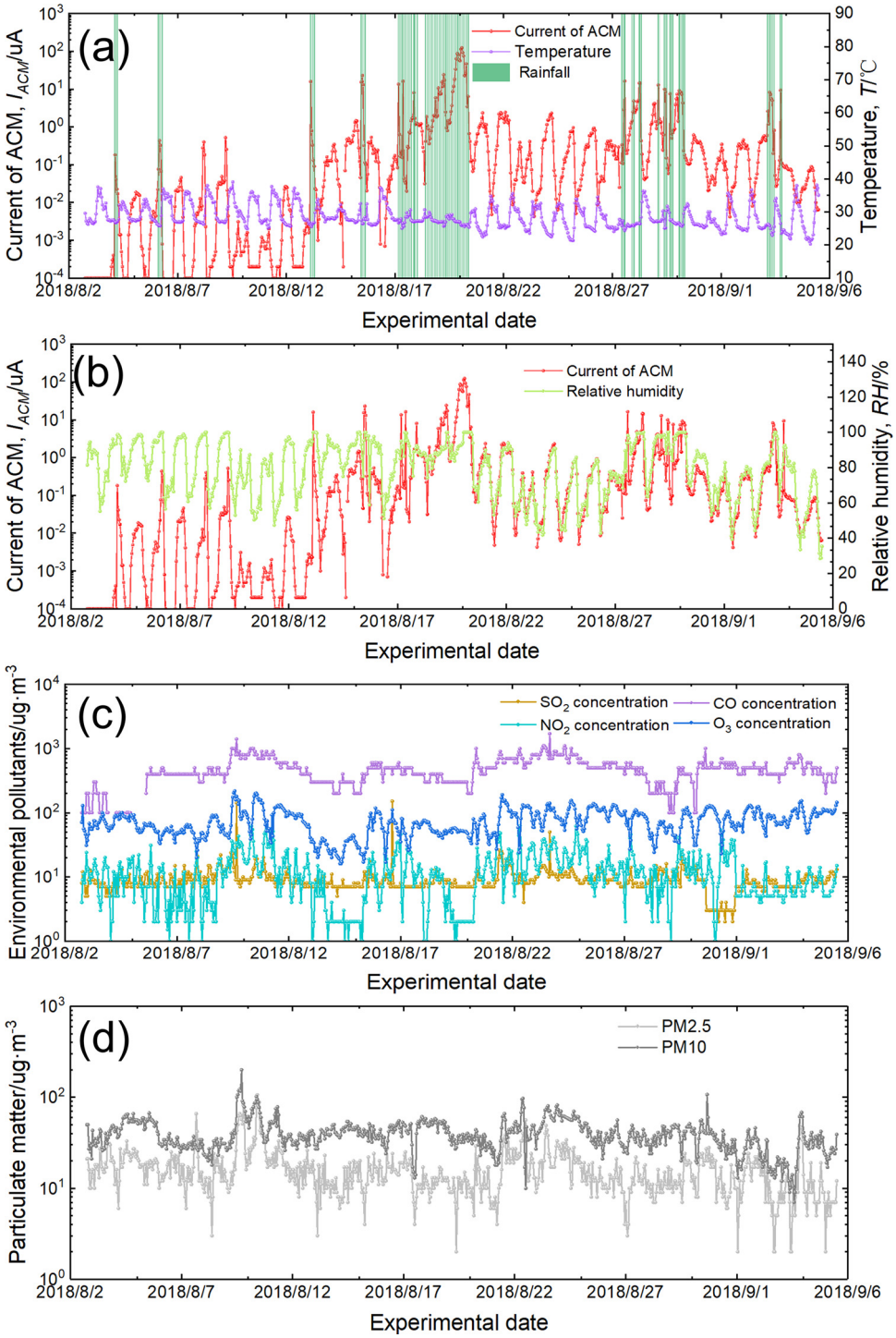

Fig. 3. Corrosion and environmental data collected during the one-month exposure test at the Qingdao site: (a) I_{ACM}, temperature and rainfall state, (b) I_{ACM} and RH, (c) NO_2, SO_2, CO and O_3 concentrations and (d) $PM_{2.5}$ and PM_{10} concentrations.

3. Results and discussion

3.1. Effect of environmental factors in an outdoor environment

As shown in Fig. 3, the 10-dimensional atmospheric corrosion data, acquired in the one-month exposure test conducted in Qingdao, China, included I_{ACM}, T, RH, rainfall status, SO_2, NO_2, O_3, CO, $PM_{2.5}$ and PM_{10}. The I_{ACM}, RH, temperature and rainfall data were reduced to a frequency of once per hour to match with other environmental data. As shown in Fig. 3a and b, the value of I_{ACM} varied as result of the dynamic nature of the atmospheric corrosion. An increasing trend was observed for I_{ACM} during the one-month test period. This phenomenon may due to the growth of the rust layer that can absorb moisture on the sensor surface and reduce the threshold RH for corrosion to be detected by the ACM sensor [41,42]. The conductivity of the rust is far lower than that of the electrolyte on the sensor [43–45]. Thus, we infer that the change in the corrosion product conductivity during rust formation would not induce a major impact on the sensor output current. Moreover, the variation of I_{ACM} showed a similar trend with those of the outdoor temperature and RH values. For the pollutants and airborne particles

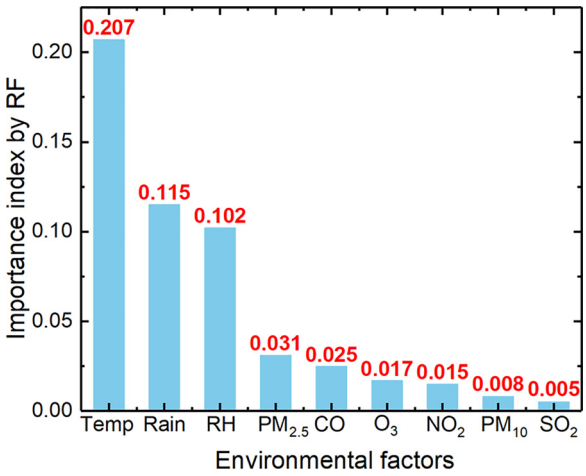

Fig. 4. Importance index of environmental factors to I_{ACM} after one-month exposure.

(Fig. 3c and d), their influences on the atmospheric corrosivity could not be directly visualized from the comparison with the I_{ACM} variation. To more clearly reveal the corrosion kinetics on the ACM sensor, Fig. S2 shows the variation of average current of all I_{ACM} over the test period. The low current values in the beginning of the exposure test was attributed to the difficulty in forming continuous thin electrolyte layer on fresh sensor surface. As corrosion was developing on the sensor, we observed a sharp increase in the current value, which could be explained by the hygroscopic effect of the rust layer. Afterwards, the continuous buildup of the rust caused the current to slightly decrease over time.

To compare the importance of each environmental parameter to the initial atmospheric corrosion over the one month of exposure, a RF model was established using the environmental parameters as the inputs and I_{ACM} as the output and the results were shown in Fig. 4. The top three important environmental parameters on the output of I_{ACM} were temperature, rainfall status and RH. In general, the variations of the temperature and the RH showed opposite trends and a lower temperature favors the condensation of thin electrolyte layer on the ACM electrodes. The I_{ACM} value was also remarkably affected by rainfalls which provided more dynamic and longer-lasting electrolyte film [46,47]. According to Fig. 4, $PM_{2.5}$ is the most important factor among all pollutants, which may be attributed to the high mass fraction of water-soluble ions (62 %) in $PM_{2.5}$ in Qingdao [48,49]. In comparison, the mass fraction of these ions in PM_{10} is 35 % [49]. And smaller airborne particles also tend to be acidic and contain more corrosive sulphates to influence the corrosion process [50–52]. However, the airborne particles and pollutants are far less important than temperature, RH and rainfall status, which may be explained by their low concentrations in the Qingdao test site and the short exposure time. The monthly average concentrations of $PM_{2.5}$, PM_{10}, NO_2 and SO_2 were only 16.9 $\mu g\,m^{-3}$, 40.2 $\mu g\,m^{-3}$, 12.9 $\mu g\,m^{-3}$ and 9.5 $\mu g\,m^{-3}$, respectively. It should be noted that the deposition of chlorides was not included in the model given that its value is typically recorded at least monthly by the dry plate method, according to ISO 9225 [38]. However, the impact of chlorides is cumulative and becomes more dominant for long-term corrosion to steels under marine atmospheres [53–57]. Chloride enrichment, which could induce the formation of non-protective akageneite [56–59], was not found in the corrosion product layer on the ACM sensor after one month of exposure (Figs. S3–S4). Besides the short exposure time, the absence of chloride enrichment may also be related to the relatively low chloride deposition rate (71.03 $mg\,m^{-2}\,d^{-1}$) and the cleaning effect from the frequent rainfalls in the test period [60]. Therefore, it is reasonable to consider that chlorides had not played a major role in the present study.

3.2. Prediction of atmospheric corrosion based on the RF model

Traditional methods to predict atmospheric corrosion of carbon steels generally follow ISO9223-2012 [61] and requires the accumulation of environmental data on a yearly basis. These yearly average values cannot reflect the dynamic variation of the environmental conditions and their complex instantaneous effects on the corrosion kinetics, which could generate inaccurate prediction especially for short-term corrosion [62]. From the previous section, temperature, RH and rainfall status have shown much stronger influences than the rest of the environmental parameters on the atmospheric corrosion of carbon steels. Therefore, these three parameters were selected as the inputs to

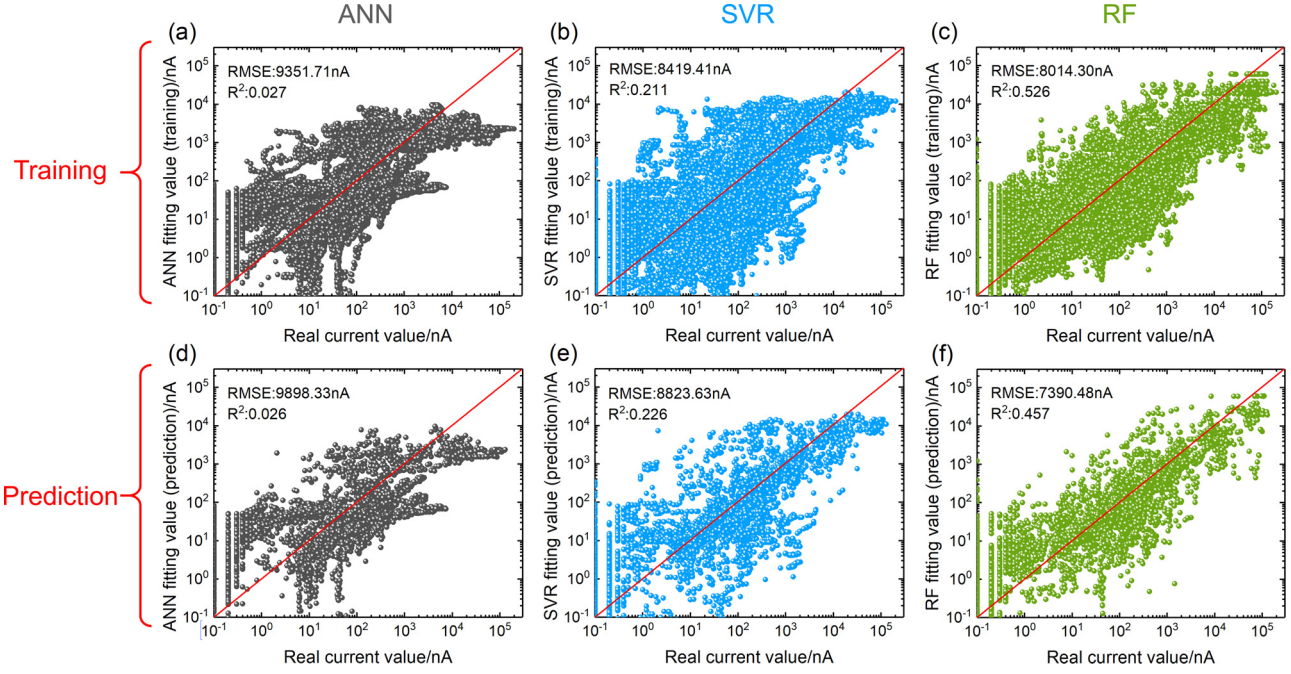

Fig. 5. The fitting results for the training samples by (a) ANN, (b) SVR and (c) RF models, and for the testing samples by (d) ANN (e) SVR, and (f) RF models.

establish a model for the prediction of the I_{ACM} output. To validate the performance of the models, the corrosion data samples were divided into training data and testing data. The entire dataset consisting of 48,647 data samples were used for the models, including 43,782 samples (90 %) randomly selected as the training samples, and the other 4865 samples (10 %) as the testing samples for the evaluation of the predicting performance [63].

Fig. 5 presents the fitting results both for training and prediction. The abscissa of each figure is the actual I_{ACM} value obtained from the ACM sensor, while the ordinate represents the current value predicted based on the ANN, SVR or RF model. The red diagonal line represents the true-prediction line on which the predicted values equal the corresponding true values. The points located closer to the red diagonal line represent smaller errors of the prediction. Compared to the fitting by ANN and SVR models, the one by the RF model generated the lowest root mean square error (RMSE) (8,014.3 nA) and highest determination coefficients (R^2) (0.526) values, demonstrating the highest prediction accuracy. The accuracy of the SVR prediction is substantially higher than that of the ANN prediction, which is consistent with previous studies [26,28]. The low accuracy of the ANN model may be attributed to its limited capacity in processing the large noise in the I_{ACM} [64]. Fig. 5d–f shows the predicting results for the three algorithms. The predicting results of RF are the best of the three algorithms and the value of R^2 is maintained at a good level. Meanwhile, there is no tendency to overfit at any range for RF.

3.3. High-accuracy model for predicting atmospheric corrosion

As demonstrated above, the RF model produced more accurate prediction of the atmospheric corrosion than the SVR and ANN models. However, the errors of the prediction as exemplified by the RMSE and R^2 values were still quite high, requiring further improvements on the models. Fig. 1b shows the sensor surface after 34 days of exposure in Qingdao. Clearly, the surface of the carbon steel electrodes of the sensor was almost fully covered with corrosion products. As discussed in the previous sections, it is reasonable to suspect that the growth of the rust layer of the steel electrode may affect the I_{ACM} output in response to the environmental impacts. The growing rust layer increased the roughness on the surface of ACM sensor and enhanced hygroscopicity, which changed the surface RH as opposed to ambient RH of the environment [65]. Meanwhile, a thicker layer of the rust would benefit for the storage of the electrolyte and promote corrosion. According to the positive correlation between ACM current and corrosion rate [17], the corrosion extent of the carbon steel anode can be reflected by the electrical quantity output of the ACM sensor (Q_{ACM}), which is obtained by integrating I_{ACM} over the test time according to Eq. (3), in which 1 min is the data acquisition interval.

$$Q_{\mathrm{ACM}} = \sum I_{\mathrm{ACM}} \times 1\,\mathrm{min} \tag{3}$$

Fig. 6 shows the day-to-day variation of importance indices of the top three environmental factors (i.e. temperature, RH and rainfall status) and that of the Q_{ACM} to the atmospheric corrosion as determined by the RF model. The importance of the temperature remained relatively stable during the one-month test period. Notably, the importance of the rainfall was much higher than that of the RH during the first half of the exposure test. The importance of Q_{ACM} which reflected the rust formation on the sensor was generally low in the beginning but exhibited a sharp increase afterwards. In the second half of the exposure test, the formation of rust layer had become the most important factors for the output of I_{ACM}. At the beginning of the exposure, the surface of the steel electrode of the ACM sensor was smooth, which was difficult to absorb moisture from the air. At this stage, the formation of the thin electrolyte layer that is necessary to connect the anodes and cathodes and generate I_{ACM} outputs mainly depended on the rainfall. As the atmospheric corrosion progressed, the rust layer on the steel electrodes became thicker and the surface roughness increased. Thus, the ability of the sensor surface to collect moisture was enhanced, leading to the decrease in the critical RH to generate I_{ACM} outputs [66]. As a result, the importance of RH to the atmospheric corrosion increased after the first half of the exposure test.

The previous section confirmed that the formation of rust layer on the ACM sensor had a crucial effect on the I_{ACM} output. Thus, the rust formation on the ACM sensor should be considered when developing predictive models for the atmospheric corrosion. Fig. 7 summarizes the fitting results based on ANN, SVR and RF models that took in the consideration of rust formation on the sensor (i.e. Q_{ACM}) as an input parameter in addition to temperature, RH and rainfall status. Compared with Fig. 5, Fig. 7 shows obvious improvements on the prediction accuracy for all models, with the data points being more narrowly distributed along the diagonal line. Particularly, the value of R^2 for the RF model substantially increased from 0.526 to 0.940 and the RMSE decreased from 7,390.5 nA to 2,311.8 nA after adding Q_{ACM} as an input parameter.

To further verify the importance of the evolution of the rust formation on the prediction of I_{ACM}, the comparisons between the fitting results of the prediction by ANN, SVR and RF models were made in three different time periods (i.e. 0–11 days, 0–22 days, 0–34 days) (Fig. 8). When Q_{ACM} is not included in the models (Fig. 8a–c), the datasets are increasingly scattered with the extended exposure time. The corresponding R^2 values had generally decreased (Fig. 8g). These results indicated that as the time increased the models become more inaccurate without the consideration of the rust formation. In contrast, when rust formation was considered in the models (Fig. 8d–f), the distribution of the data points extended along the diagonal line but did not become widened with the increasing exposure, suggesting that the prediction accuracy is stabilized. As shown in Fig. 8g, the R^2 values for the ANN and SVR prediction increased sharply when longer time was included in the models. The R^2 value for the RF prediction remained

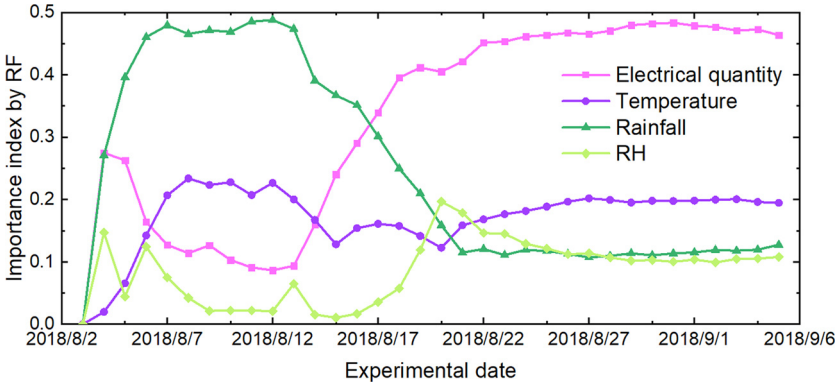

Fig. 6. Daily variation of the importance index of electrical quantity, temperature, rainfall state and RH to the I_{ACM}.

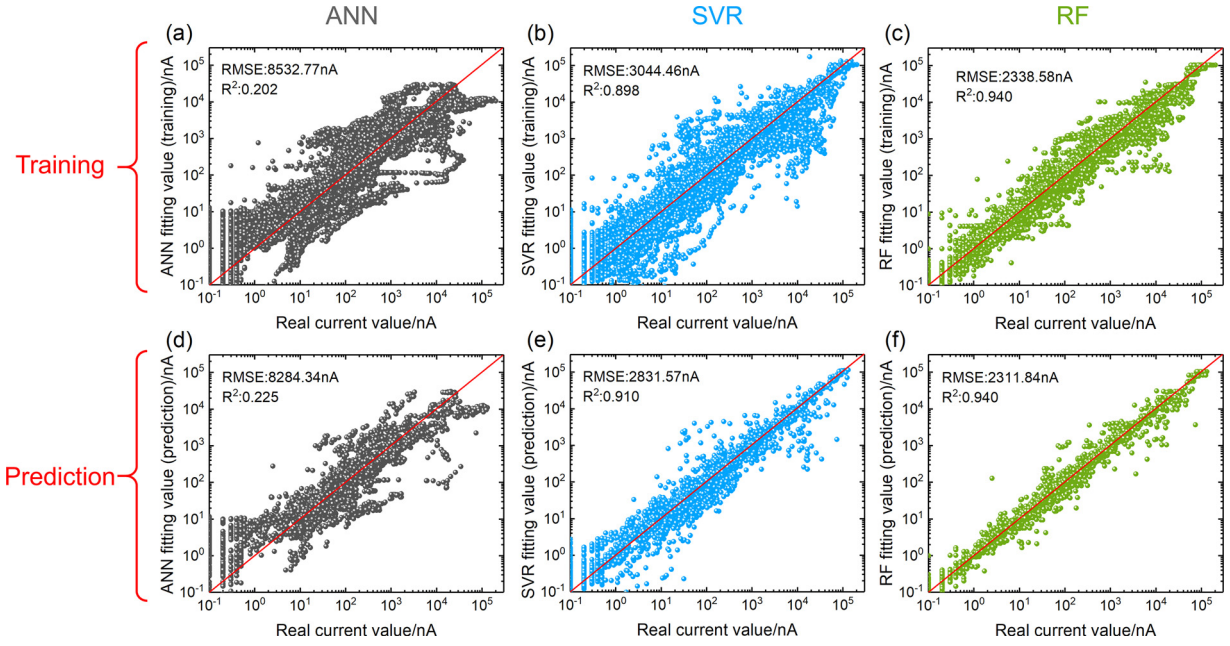

Fig. 7. The fitting results (with the electrical quantity as the input) for training samples by (a) ANN, (b) SVR and (c) RF, and for testing samples by (d) ANN, (e) SVR and (f) RF models.

stabilized at an excellent level of 0.94. The growth of the rust layer on the ACM sensor had a significant impact on the I_{ACM}, so a high-accuracy model should take Q_{ACM} into account for the prediction of atmospheric corrosion of steels.

Finally, as shown in Fig. 9a, two RF models were built and trained differently using the data from August 2nd to 30th to predict the data from September 1st to 5th. One model (RF1 in Fig. 9a) selected temperature, RH and rainfall status as the input parameters and the other (RF2) added another parameter of Q_{ACM}. Different from the predicting process in RF1, the Q_{ACM} of 23:59 on August 30th was selected as an input in the RF2 model for the first step, after which the predicting sequence in RF2 strictly followed the chronological order from 00:00 on September 1st to 00:00 on 5th, updating Q_{ACM} based on Eq. (3) at every minute. Fig. 9b shows the variation of the predicted I_{ACM} by the models with and without the Q_{ACM} correction as compared with the actual data collected during September 1st to 5th. Although the variations of the predicted I_{ACM} showed very similar trends to that of the actual values, a much larger fluctuation was observed in the curve predicted by the uncorrected RF model. To quantitatively compare the accuracy by the two models, the I_{ACM} values were divided into four levels ($I_{ACM} = \{I_1, I_2, I_3, I_4\}$ where I_1, I_2, I_3 and I_4 are 0–800 nA, 800–1500 nA, 1500–3000 nA and > 3000 nA, respectively) and then the number of the predicted values that were at the same level with the actual level were counted [4]. As shown in Table 1, the rate of accurate prediction was improved from 87.8% to 94.7% after correcting the model with the consideration of the Q_{ACM}. In conclusion, the incorporation of Q_{ACM} as an input greatly improved the precision of ANN, SVR and RF models. The RF model with the correction considering the rust formation on the ACM sensor can be used as a high-accuracy model for predicting atmospheric corrosion. Furthermore, this corrected model implies the possibility that, once sufficient training data is available, atmospheric corrosion could be predicted based on simple environmental sensors without expensive corrosion sensors.

4. Conclusions

A Fe/Cu type galvanic ACM sensor consisting of carbon steel anodes and pure copper cathodes were exposed in an outdoor atmospheric environment in Qingdao for 34 days. An RF-based machine learning method was used to assist in the analysis and prediction of I_{ACM} values and the results were compared with those by ANN and SVR models. Among the environmental factors accumulated, temperature, RH and

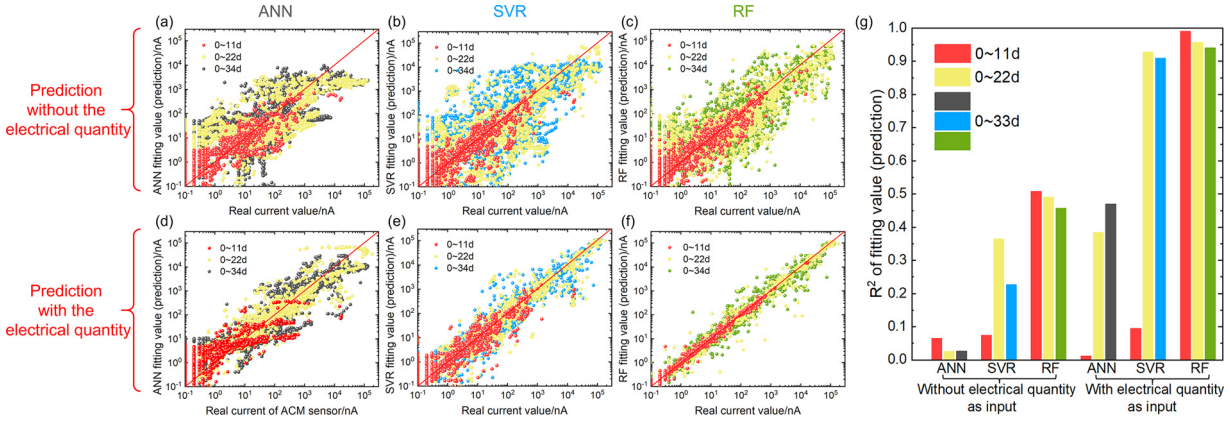

Fig. 8. The prediction results of different periods for the testing samples: (a) ANN, (b) SVR, and (c) RF without the electrical quantity as the input; (d) ANN, (e) SVR, and (f) RF with the electrical quantity as the input; and (g) the variation of R^2 for each model.

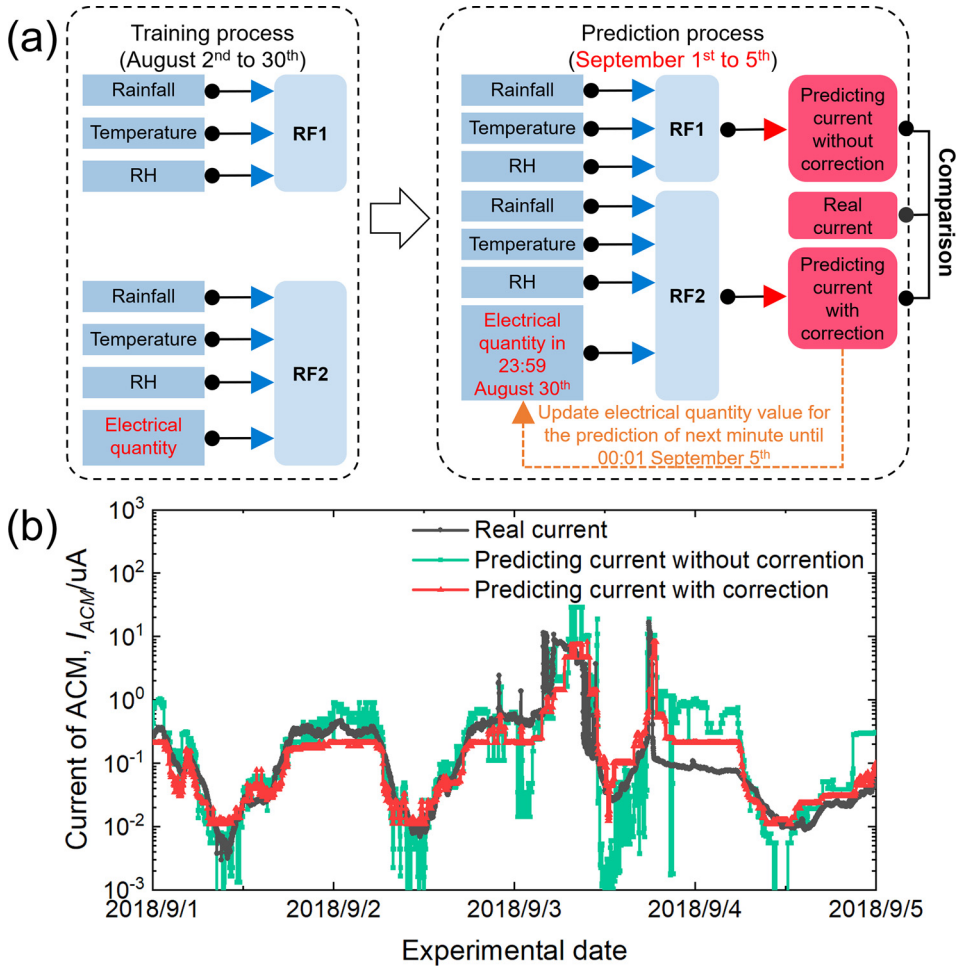

Fig. 9. (a) Prediction process for the I_{ACM} from September 1st to 5th; (b) the actual and predicted I_{ACM} values from September 1st to 5th.

Table 1
Comparison of the prediction by models with and without Q_{ACM} correction.

Method	Number of samples	Number of samples correctly predicted	Accuracy rate
RF model with correction	6540	6196	94.7 %
RF model without correction		5742	87.8 %

rainfall status showed the highest importance to the I_{ACM}. The pollutants such as SO_2, NO_2, O_3, CO, $PM_{2.5}$ and PM_{10} were not the main factors for the I_{ACM} because of their low concentrations in Qingdao. The predicting ability of the RF model was much stronger than that of the ANN and SVR models in training and predicting I_{ACM}. The growth of the rust layer on the ACM sensor surface reduced the accuracy of the I_{ACM} prediction. Taking into consideration of the rust formation on the surface of the ACM sensor, an improved RF model with the addition of Q_{ACM} as an input was established and showed a significantly higher accuracy for the prediction of atmospheric corrosion of carbon steels. Potentially, the predictive model developed in this work could be useful in the future development of smart corrosion sensors with the ability of offsetting environmental damage to the sensor and maintaining high accuracy for extended usage.

CRediT authorship contribution statement

Zibo Pei: Investigation, Methodology, Writing - original draft. **Dawei Zhang:** Supervision, Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Yuanjie Zhi:** Investigation, Methodology. **Tao Yang:** Methodology, Writing - review & editing. **Lulu Jin:** Investigation. **Dongmei Fu:** Methodology. **Xuequn Cheng:** Methodology. **Herman A. Terryn:** Writing - review & editing. **Johannes M.C. Mol:** Writing - review & editing. **Xiaogang Li:** Supervision, Conceptualization, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare no conflict of interests.

Acknowledgements

This work is supported by the National Key Research and Development Program of China (2017YFB0702100, 2016YFB00604), Fundamental Research Funds for the Central Universities (FRF-BD-19-014A), 111 Program (B170003) and BRI Southeast Asia Network for Corrosion and Protection (MOE).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the

online version, at doi:https://doi.org/10.1016/j.corsci.2020.108697.

References

[1] X. Li, D. Zhang, Z. Liu, Z. Li, C. Du, C. Dong, Share corrosion data, Nature 527 (2015) 441–442.
[2] T. Grøntoft, A. Verney-Carron, J. Tidblad, Cleaning costs for European sheltered white painted steel and modern glass surfaces due to air pollution since the year 2000, Atmosphere 10 (2019) 167.
[3] P.R. Roberge, R.D. Klassen, P.W. Haberecht, Atmospheric corrosivity modeling—a review, Mater. Des. 23 (2002) 321–330.
[4] Y. Shi, D. Fu, X. Zhou, T. Yang, Z. Pei, D. Zhang, L. Shao, Data mining to online galvanic current of zinc/copper internet atmospheric corrosion monitor, Corros. Sci. 133 (2018) 443–450.
[5] A. Nishikata, Q. Zhu, E. Tada, Long-term monitoring of atmospheric corrosion at weathering steel bridges by an electrochemical impedance method, Corros. Sci. 87 (2014) 80–88.
[6] B. Kamsu-Foguem, Knowledge-based support in non-destructive testing for health monitoring of aircraft structures, Adv. Eng. Inform. 26 (2012) 859–869.
[7] Z. Li, D. Fu, Y. Li, G. Wang, J. Meng, D. Zhang, Z. Yang, G. Ding, J. Zhao, Application of an electrical resistance sensor-based automated corrosion monitor in the study of atmospheric corrosion, Materials 12 (2019) 1065.
[8] C. Kleber, U. Hilfrich, M. Schreiner, In situ QCM and TM-AFM investigations of the early stages of degradation of silver and copper surfaces, Appl. Surf. Sci. 253 (2007) 3712–3721.
[9] F. Hernandez-Valle, A.R. Clough, R.S. Edwards, Stress corrosion cracking detection using non-contact ultrasonic techniques, Corros. Sci. 78 (2014) 335–342.
[10] A. Legat, Monitoring of steel corrosion in concrete by electrode arrays and electrical resistance probes, Electrochim. Acta 52 (2007) 7590–7598.
[11] S. Wan, J. Hou, Z.-F. Zhang, X.-x. Zhang, Z.-H. Dong, Monitoring of atmospheric corrosion and dewing process by interlacing copper electrode sensor, Corros. Sci. 150 (2019) 246–257.
[12] C. Thee, L. Hao, J. Dong, X. Mu, X. Wei, X. Li, W. Ke, Atmospheric corrosion monitoring of a weathering steel under an electrolyte film in cyclic wet–dry condition, Corros. Sci. 78 (2014) 130–137.
[13] D.W. Law, S.G. Millard, J.H. Bungey, Linear polarisation resistance measurements using a potentiostatically controlled guard ring, NDT E Int. 33 (2000) 15–21.
[14] E. Budevski, W. Obretenov, W. Bostanov, G. Staikov, J. Doneit, K. Jüttner, W.J. Lorenz, Noise analysis in metal deposition—expectations and limits, Electrochim. Acta 34 (1989) 1023–1029.
[15] D. Xia, S. Song, W. Jin, J. Li, Z. Gao, J. Wang, W. Hu, Atmospheric corrosion monitoring of field-exposed Q235B and T91 steels in Zhoushan offshore environment using electrochemical probes, J. Wuhan Univ. Technol. 32 (2017) 1433–1440.
[16] F. Mansfeld, J.V. Kenkel, Electrochemical monitoring of atmospheric corrosion phenomena, Corros. Sci. 16 (1976) 111–122.
[17] D. Mizuno, S. Suzuki, S. Fujita, N. Hara, Corrosion monitoring and materials selection for automotive environments by using Atmospheric Corrosion Monitor (ACM) sensor, Corros. Sci. 83 (2014) 217–225.
[18] Z. Pei, X. Cheng, X. Yang, Q. Li, C. Xia, D. Zhang, X. Li, Understanding environmental impacts on initial atmospheric corrosion based on corrosion monitoring sensors, J. Mater. Sci. Technol. (2020), https://doi.org/10.1016/j.jmst.2020.01.023.
[19] J. Shi, J. Wang, D.D. Macdonald, Prediction of primary water stress corrosion crack growth rates in Alloy 600 using artificial neural networks, Corros. Sci. 92 (2015) 217–227.
[20] J. Shi, J. Wang, D.D. Macdonald, Prediction of crack growth rate in Type 304 stainless steel using artificial neural networks and the coupled environment fracture model, Corros. Sci. 89 (2014) 69–80.
[21] G.-L. Song, The grand challenges in electrochemical corrosion research, Front. Mater. 1 (2014) 2.
[22] F. King, A.R. Erler, S.K. Frey, C.G. Fletcher, Application of machine learning techniques for regional bias correction of SWE estimates in Ontario, Canada, Hydrol. Earth Syst. Sci. (2020), https://doi.org/10.5194/hess-2019-2593.
[23] M. Kamrunnahar, M. Urquidi-Macdonald, Prediction of corrosion behavior using neural network as a data mining tool, Corros. Sci. 52 (2010) 669–677.
[24] M. Kamrunnahar, M. Urquidi-Macdonald, Prediction of corrosion behaviour of Alloy 22 using neural network as a data mining tool, Corros. Sci. 53 (2011) 961–967.
[25] M.K. Cavanaugh, R.G. Buchheit, N. Birbilis, Modeling the environmental dependence of pit growth using neural network approaches, Corros. Sci. 52 (2010) 3070–3077.
[26] J. Ren, ANN vs. SVM: which one performs better in classification of MCCs in mammogram imaging, Knowl. Based Syst. 26 (2012) 144–153.
[27] S.F. Fang, M.P. Wang, W.H. Qi, F. Zheng, Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials, Comp. Mater. Sci. 44 (2008) 647–655.
[28] Y.F. Wen, C.Z. Cai, X.H. Liu, J.F. Pei, X.J. Zhu, T.T. Xiao, Corrosion rate prediction of 3C steel under different seawater environment by using support vector regression, Corros. Sci. 51 (2009) 349–355.
[29] F. Caleyo, J.C. Velázquez, A. Valor, J.M. Hallen, Markov chain modelling of pitting corrosion in underground pipelines, Corros. Sci. 51 (2009) 2197–2207.
[30] K. McCallum, J. Zhao, M. Workman, M. Iannuzzi, M. Kappes, J. Payer, C.B. Clemons, S. Chawla, K.I. Kreider, N. Mimoto, G.W. Young, Localized corrosion risk assessment using Markov analysis, Corrosion 70 (2014) 1114–1127.
[31] P. Li, T.C. Tan, J.Y. Lee, Grey relational analysis of amine inhibition of mild steel corrosion in acids, Corrosion 53 (1997) 186–194.
[32] F.Y. Ma, W.H. Wang, Prediction of pitting corrosion behavior for stainless SUS 630 based on grey system theory, Mater. Lett. 61 (2007) 998–1001.
[33] F. Caleyo, J.C. Velázquez, A. Valor, J.M. Hallen, Probability distribution of pitting corrosion depth and rate in underground pipelines: a Monte Carlo study, Corros. Sci. 51 (2009) 1925–1934.
[34] C.I. Ossai, B. Boswell, I.J. Davies, Modelling the effects of production rates and physico-chemical parameters on pitting rate and pit depth growth of onshore oil and gas pipelines, Corros. Eng. Sci. Technol. 51 (2016) 342–351.
[35] Y. Bengio, Learning deep architectures for AI, Found. Trends® Mach. Learn. 2 (2009) 1–127.
[36] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[37] Y. Zhi, D. Fu, D. Zhang, T. Yang, X. Li, Prediction and knowledge mining of outdoor atmospheric corrosion rates of low alloy steels based on the random forests approach, Metals 9 (2019) 383.
[38] F. Corvo, T. Perez, L.R. Dzib, Y. Martin, A. Castañeda, E. Gonzalez, J. Perez, Outdoor–indoor corrosion of metals in tropical coastal atmospheres, Corros. Sci. 50 (2008) 220–230.
[39] Y. Hou, C. Aldrich, K. Lepkova, L.L. Machuca, B. Kinsella, Analysis of electrochemical noise data by use of recurrence quantification analysis and machine learning methods, Electrochim. Acta 256 (2017) 337–347.
[40] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and regression trees, Biometrics 40 (1984) 874.
[41] W.S. Patterson, L. Hebbs, The relation of the moisture in rust to the critical corrosion humidity, Trans. Faraday Soc. 27 (1931) 277–283.
[42] U.R. Evans, C.A.J. Taylor, Critical humidity for rusting in the presence of sea salt, Br. Corros. J. 9 (1974) 26–28.
[43] V.V. Jagannadha Sarma, C. Subba Rao, Electrical conductivity of rain water at Visakhapatnam, India, J. Geophys. Res.-Atmos. 77 (1972) 2197–2200.
[44] T. Hermans, M. Paepen, Combined inversion of land and marine electrical resistivity tomography for submarine groundwater discharge and saltwater intrusion characterization, Geophys. Res. Lett. (2020), https://doi.org/10.1029/2019GL085877.
[45] K. Habib, Measurement of surface resistivity/conductivity of carbon steel in 5–20 ppm of RA-41 inhibited seawater by optical interferometry techniques, Interferometry XV: Techniques and Analysis 7790 (2011) 80822A-1-80822A-14.
[46] I.S. Cole, W.D. Ganther, S.A. Furman, T.H. Muster, A.K. Neufeld, Pitting of zinc: observations on atmospheric corrosion in tropical countries, Corros. Sci. 52 (2010) 848–858.
[47] C. Chiavari, E. Bernardi, C. Martini, F. Passarini, F. Ospitali, L. Robbiola, The atmospheric corrosion of quaternary bronzes: the action of stagnant rain water, Corros. Sci. 52 (2010) 3002–3010.
[48] Z.G. Guo, J.L. Feng, M. Fang, H.Y. Chen, K.H. Lau, The elemental and organic characteristics of PM2.5 in Asian dust episodes in Qingdao, China, 2002, Atmos. Environ. 38 (2004) 909–919.
[49] M. Hu, L.-Y. He, Y.-H. Zhang, M. Wang, Y.P. Kim, K.C. Moon, Seasonal variation of ionic species in fine particles at Qingdao, China, Atmos. Environ. 36 (2002) 5853–5859.
[50] H. Zhuang, C.K. Chan, M. Fang, A.S. Wexler, Size distributions of particulate sulfate, nitrate, and ammonium at a coastal site in Hongkong, Atmos. Environ. 33 (1999) 843–853.
[51] R.K. Pathak, X. Yao, A.K.H. Lau, C.K. Chan, Acidity and concentrations of ionic species of PM 2.5 in Hongkong, Atmos. Environ. 37 (8) (2003) 1113–1124.
[52] N. Pérez, J. Pey, X. Querol, A. Alastuey, J.M. López, M. Viana, Partitioning of major and trace components in PM10–PM2.5–PM1 at an urban site in Southern Europe, Atmos. Environ. 42 (2008) 1677–1691.
[53] D. de la Fuente, J. Castano, M. Morcillo, Long-term atmospheric corrosion of zinc, Corros. Sci. 49 (2007) 1420–1436.
[54] D. Persson, D. Thierry, O. Karlsson, Corrosion and corrosion products of hot dipped galvanized steel during long term atmospheric exposure at different sites worldwide, Corros. Sci. 126 (2017) 152–165.
[55] T. Prosek, D. Thierry, C. Taxén, J. Maixner, Effect of cations on corrosion of zinc and carbon steel covered with chloride deposits under atmospheric conditions, Corros. Sci. 49 (2007) 2676–2693.
[56] J. Alcántara, B. Chico, I. Díaz, D. de la Fuente, M. Morcillo, Airborne chloride deposit and its effect on marine atmospheric corrosion of mild steel, Corros. Sci. 97 (2015) 74–88.
[57] N. Mi, M. Ghahari, T. Rayment, A.J. Davenport, Use of inkjet printing to deposit magnesium chloride salt patterns for investigation of atmospheric corrosion of 304 stainless steel, Corros. Sci. 53 (2011) 3114–3121.
[58] J. Weissenrieder, C. Leygraf, In situ studies of filiform corrosion of iron, J. Electrochem. Soc. 151 (2004) B165–B171.
[59] Y. Ma, Y. Li, F. Wang, Corrosion of low carbon steel in atmospheric environments of different chloride content, Corros. Sci. 51 (2009) 997–1006.
[60] F. Corvo, J. Minotas, J. Delgado, C. Arroyave, Changes in atmospheric corrosion rate caused by chloride ions depending on rain regime, Corros. Sci. 47 (2005) 883–892.
[61] M. Morcillo, B. Chico, I. Díaz, H. Cano, D. De la Fuente, Atmospheric corrosion data of weathering steels, a review, Corros. Sci. 77 (2013) 6–24.
[62] Y. Cai, Y. Zhao, X. Ma, K. Zhou, Y. Chen, Influence of environmental factors on atmospheric corrosion in dynamic environment, Corros. Sci. 137 (2018) 163–175.
[63] N. Antropova, B. Huynh, M. Giger, SU-D-207B-06: predicting breast cancer malignancy on DCE-MRI data using pre-trained convolutional neural networks, Med. Phys. 43 (2016) 3349–3350.
[64] J.S. Reyes, R.J. Duro, Influence of noise on discrete time backpropagation trained networks, Neurocomputing 41 (2001) 67–89.
[65] E. Schindelholz, R.G. Kelly, I.S. Cole, W.D. Ganther, T.H. Muster, Comparability and accuracy of time of wetness sensing methods relevant for atmospheric corrosion, Corros. Sci. 67 (2013) 233–241.
[66] S.C. Chung, A.S. Lin, J.R. Chang, H.C. Shih, EXAFS study of atmospheric corrosion products on zinc at the initial stage, Corros. Sci. 42 (2000) 1599–1610.