

PROCEEDINGS PAPER

Effective Utilization for Data of Natural Environment Corrosion of Materials

Haitao Wang¹, Leijiang Yao², Xiaofeng Li¹, En-Hou Han¹ and Xiaoyan Tong²

¹ Key Laboratory of Nuclear Materials and Safety Assessment, Institute of Metal Research, Chinese Academy of Sciences, Shenyang, 110016, China
htwang@imr.ac.cn

² Science and Technology on UVA Laboratory, Northwester Polytechnical University, Xi'an, 710065, China

An analysis architecture is built to make full use of natural environment corrosion of materials data. All environmental factors, such as average temperature, average relative humidity, rainfall hours, sunshine duration, chloride ion settlement, etc., in nine atmospheric test stations, including Beijing, Wuhan, Jiangjin, and Wanning, and seawater temperature, salinity, dissolved oxygen, pH values, etc. in four marine test stations, including Qingdao, Zhoushan, Xiamen, and Yulin, are modelled monthly using a statistical analysis technique. Normal distribution, logarithmic normal distribution, Weibull distribution, and uniform distribution of the samples are checked by the Shapiro-Wilk and Kolmogorov-Smirnov methods. Subsequently, the probability distribution function, the expected population value and variance, and the confidence interval are estimated. Corrosion data for twenty-two kinds of carbon steels, low alloy steels, and stainless steels, including A3, 3C, 16Mn, 10CrMoAl, 1Cr18Ni9Ti, in atmospheric and marine test stations are analyzed. The main environmental factors for atmospheric and marine corrosion of each type of steel are determined by grey relational analysis. Subsequently, a predictive model incorporating the main corrosion environmental factors, exposure time, and corrosion rate is built by the BP artificial neural network. After the evolution of corrosion pit depth is predicted by the artificial neural network, a fracture mechanics calculation is used to evaluate the residual life of a structure with corrosion pit defects.

Keywords: Corrosion; Statistical analysis; Grey relational analysis; Artificial neural network; Fracture mechanics calculation

1 Introduction

Corrosion is very harmful to the useful lifetime of materials. It can bring enormous economic losses and severe dangers to society. Natural environment corrosion is one of the most common corrosion phenomena. It is very complex and has a vast set of factors responsible for its existence. Natural environment corrosion includes mainly atmospheric corrosion and marine corrosion.

Atmospheric corrosion is defined as the corrosion of materials exposed to the air and its pollutants. Electrochemical corrosion and chemical corrosion can occur in the atmospheric environment. However, electrochemical corrosion is more important than chemical corrosion in atmospheric corrosion (Morcillo et al., 2013). Metal absorbs a layer of ultra-thin water film on its surface that forms an electrolyte under atmospheric exposure conditions after a certain critical humidity level has been reached (Melchers, 2013; Wall et al., 2005). In the presence of this thin-film electrolyte, atmospheric corrosion proceeds by balancing anodic and cathodic reactions. The anodic oxidation reaction involves dissolution of the metal while the cathodic reaction is commonly an oxygen reduction reaction. Under thin-film corrosion conditions, oxygen from the atmosphere is readily supplied to the electrolyte. Seawater is the electrolyte solution with very high salinity. It is one of the most corrosive of all the natural corrosion agents. Thus, electrochemical corrosion of metals is stronger in a marine environment (Traverso et al., 2014; Shifler, 2005).

A large amount of data concerning the natural environment corrosion of materials has accumulated over the last 16 years through research supported by the Major Program of the National Natural Science Foundation of China (Xiao et al., 2002; Zhu et al., 2002). In this work, analysis architecture is built to make full use of these data. This architecture analyzes the inherent laws of data and obtains information on the influences of natural environment factors and the evolutionary rule of corrosion damage. This provides corresponding support and help for research on natural environment corrosion.

2 Statistical Analysis Modelling for Environmental Factors

The framework for the module of statistical analysis modelling for environmental factors is shown in **Figure 1**. The module builds the atmospheric environment and the marine environment database. All environmental factors for different atmospheric and marine test stations are modelled monthly by a statistical analysis technique. The Shapiro-Wilk method (W method) and the Kolmogorov-Smirnov method (K-S method) are used to check the normal distribution, the logarithmic normal distribution, the Weibull distribution, and the uniform distribution for the sample. Subsequently, the probability distribution function, the expected value and variance of population, and the confidence interval are estimated.

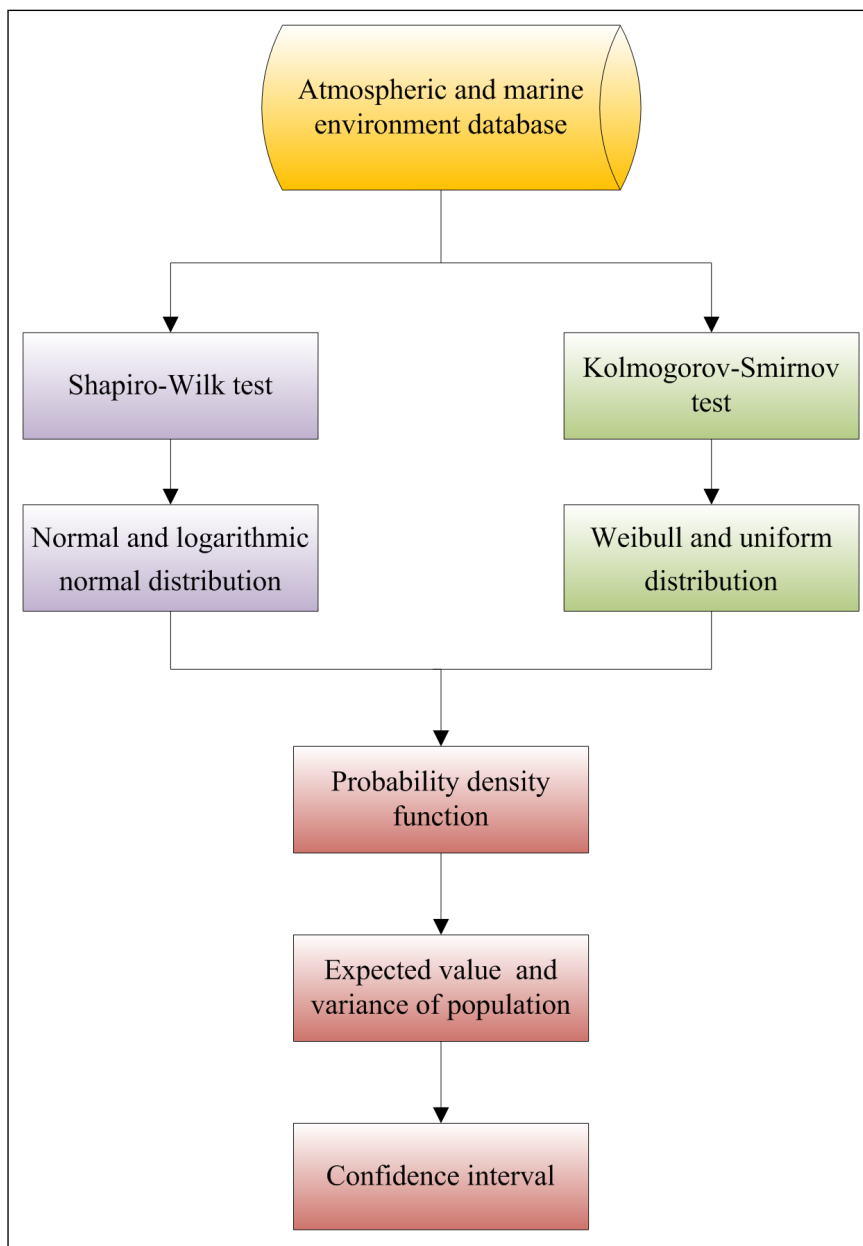


Figure 1: Framework for the module of statistical analysis modelling for environmental factors.

The atmospheric environment database involves seventeen kinds of atmospheric environmental factors: average temperature, average relative humidity, rainfall, rainfall hours, sunshine duration, SO₂ settlement, NO₂ settlement, H₂S settlement, HCl settlement, SO₃ settlement, Cl⁻ settlement, NH₃ settlement, dustfall of water-solubility, dustfall of water-insolubility, pH value of rainwater, and the SO₄ and Cl⁻ content of rainwater at the nine atmospheric test stations, Beijing, Wuhan, Jiangjin, Wanning, Qingdao, Guangzhou, Shenyang, Qionghai, and Hailar. The marine environment database involves nine kinds of marine environmental factors: seawater temperature, salinity, dissolved oxygen, pH value, temperature of the marine atmosphere, humidity, rainfall, sunshine, and wind speed at the four marine test stations, Qingdao, Zhoushan, Xiamen, and Yulin.

In the statistical analysis module, the normal distribution and the logarithmic normal distribution for the sample are checked by the W method. The W test is a kind of small sample test ($3 \leq n \leq 50$) that can only check the normal distribution (Jurečková et al., 2007; Srivastava et al., 1987). It can deal with the small sample events of atmospheric and marine environments. The W test is shown in Eq. (1)

$$W = \frac{\left(\sum_{i=1}^k a_i (x_{n+1-i} - x_i) \right)^2}{\sum_{i=1}^n (x_i - \mu)^2} \quad (1)$$

where n is the sample number, $k = n/2$ when n is an even number or $k = (n-1)/2$ when n is an odd number, μ is the mean of the sample, x_i is the value of the i^{th} sample, and a_i is the i^{th} constant. If $W > W_{\alpha}$, the sample follows normal distribution, W_{α} is the quantile of the W test, and α is the level of significance. In order to check the logarithmic normal distribution, the sample needs a logarithm conversion before Eq. (1) is used.

The probability density functions of the normal and logarithmic normal distributions are shown in Eq. (2)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad (2)$$

where σ is the standard deviation of the sample. For normal distribution, the estimate of expected value of population $E(x) = \mu$ and the estimate of variance of population $D(x) = \sigma^2$. For the logarithmic normal distribution, the estimate of expected value of population $E(X) = \exp(\mu + \sigma^2/2)$ and the estimate of variance of population $D(X) = (\exp(\sigma^2) - 1) \exp(2\mu + \sigma^2)$.

The K-S test is a type of universal test. It is used to check normal and nonnormal distributions (Vlček et al., 2009; Schröder et al., 1995). However, its reliability in checking the normal distribution is lower than that of the W test for a small sample. In the statistical analysis module, the Weibull and uniform distributions of the sample are checked by the K-S method. Suppose that $F_0(x)$ is the probability distribution function of the theoretical distribution and $F_n(x)$ is the cumulative frequency function of the current sample,

$$D = \max |F_n(x) - F_0(x)| \quad (3)$$

If $D < D_{\alpha}$, the sample follows the current theoretical distribution, D_{α} is the critical value of the W test, and α is the level of significance.

The probability density functions of the Weibull and uniform distributions are shown in Eq. (4)

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{others} \end{cases} \quad (4)$$

where $\lambda > 0$ is the scale parameter and $k > 0$ is the shape parameter for the Weibull distribution. The parameters are estimated by the probability weighted moment method (Zhang, 1994) in Eq. 5

$$\lambda = \frac{M_0}{\Gamma\{\ln(M_0/M_1)/\ln 2\}} \quad k = \frac{\ln 2}{\ln(M_0/2M_1)} \quad M_0 = \frac{1}{n} \sum_{i=1}^n x_i \quad M_1 = \frac{1}{n} \sum_{i=1}^n x_i \left(1 - \frac{i-0.35}{n}\right) \quad (5)$$

where Γ is the gamma function, the estimate of expected value of population $E(x) = \lambda\Gamma(1+1/k)$, and the estimate of variance of population $D(x) = \lambda^2\Gamma(1+2/k) - \mu^2$. For uniform distribution, a and b are the interval parameters. The estimate of expected value of population $E(x) = (a + b)/2$ and the estimate of variance of population $D(x) = (b - a)^2/12$.

The confidence interval is a measure of the reliability of an interval estimate for population parameter (Huwang et al., 2002). Suppose that θ is an unknown parameter for population. When a random interval $[\theta_1, \theta_2]$ exists, if $p\{\theta_1 \leq \theta \leq \theta_2\} = 1 - \alpha$ for a given $0 < \alpha < 1$, the interval $[\theta_1, \theta_2]$ is the confidence interval of θ at the confidence level $1 - \alpha$. For a small sample $n < 30$ following the normal distribution, the estimate of confidence interval for the expected value of population is $\mu \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$ and $\mu \pm \frac{\sigma}{\sqrt{n}} t_{\alpha/2, n-1}$ when σ^2 is

known and unknown respectively. For a large sample, the estimate of confidence interval for the expected value of population is $\mu \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$ regardless of the distribution. The estimate of confidence interval for

variance of population is $\left[\frac{(n-1)\sigma^2}{\chi^2_{\alpha/2, n-1}}, \frac{(n-1)\sigma^2}{\chi^2_{1-\alpha/2, n-1}} \right]$, where $z_{\alpha/2}$ is the bilateral quantile of the standard normal

distribution at the level of significance α , $t_{\alpha/2, n-1}$ is the bilateral quantile of the t-distribution at the degree of freedom n , $\chi^2_{\alpha/2, n-1}$ is the bilateral quantile of the Chi-square distribution.

3 Assessment of Corrosion Damage

The framework for the module of the assessment of corrosion damage is shown in **Figure 2**. The module establishes the atmospheric and marine corrosion database. The grey relational analysis is used to determine the main environmental factors of atmospheric corrosion and marine corrosion for each type of steel. Subsequently, a BP artificial neural network is used to build the atmospheric and marine corrosion damage model, incorporating the main corrosion environmental factors, exposure time, and corrosion rate for each steel type. After the evolution of the corrosion pit depth is predicted by the artificial neural network, the residual life of a structure with corrosion pit defect is evaluated by the fracture mechanics calculation.

The atmospheric corrosion database involves the average corrosion rate over a period of 16 years for seventeen kinds of carbon steels and low alloy steels: A3, 3C, 20, 08Al, 16Mn, 16MnQ, D36, 15MnMoVN, 14MnMoNbB, 09MnNb, 09CuPTiRE, 10CrMoAl, 10CrCuSiV, 12CrMnCu, 09CuPCrNi, 09CuPCrNiA, 06CuPCrNiMo and five kinds of stainless steels: 2Cr13, 00Cr17AlTi, 1Cr18Ni9Ti, 00Cr19Ni10, 0000Cr18Mo2 in the atmospheric test stations. The marine corrosion database involves the average corrosion rate, the average pit depth, the maximum pit depth, and the maximum crevice corrosion depth over a period of 16 years for these steels in the marine test stations.

Many environmental factors can affect the corrosion processes. Determining the main corrosion environmental factors is very meaningful. It can simplify the corrosion damage modelling and improve the model's reliability. In the corrosion assessment module, the grey relational analysis is used to obtain the main environmental factors of atmospheric and marine corrosion for each carbon steel, low alloy steel, and stainless steel. The grey system is a system in which some of its information is clear and some is not clear. The grey relation is the uncertainty of the association between things or the uncertainty of the association between system factors and the main behavioural factors (Deng, 1989). The grey relational analysis, which is an important factor of the grey system theory, is based on the degree of similarity or difference of development

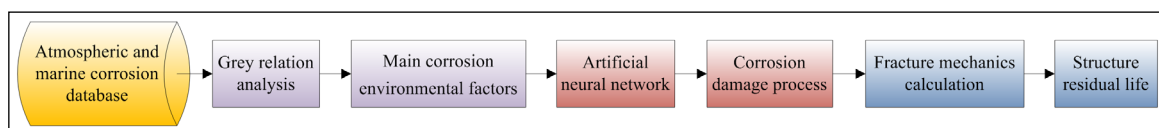


Figure 2: Framework for the module of the assessment of corrosion damage.

trends between an alternative and the ideal alternative (Wang et al., 2006). The grey relational analysis is shown in Eq. (6)

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_{0i}(k) \quad \xi_{0i}(k) = \frac{\Delta_{\min} + \delta \Delta_{\max}}{\Delta_{0i}(k) + \delta \Delta_{\max}} \quad (6)$$

where r_i is the grey relational grade, ξ_{0i} is the grey relational coefficient, δ is the resolution coefficient, $\Delta_{0i}(k) = |X_0'(k) - X_i'(k)|$, $\Delta_{\min} = \min_i \min_k |X_0'(k) - X_i'(k)|$, $\Delta_{\max} = \max_i \max_k |X_0'(k) - X_i'(k)|$, $X_0'(k)$ and $X_i'(k)$ are the reference sequence and the comparative sequence by standardization treatment $X_0'(k) = X_0(k) / \mu$, $X_i'(k) = X_i(k) / \mu$. The bigger the grey relational grade of corrosion environmental factors, the more significant its influence is on corrosion damage.

The artificial neural network (ANN) is an information processing model that is inspired by biological nervous systems. It is composed of a large number of highly interconnected neurons. It can learn correlative patterns between input and output information without specific models and can use that learned association to predict the appropriate output for input data not used in training (Carpenter et al., 1995). There are many kinds of artificial neural network models. The most widely applied one is the multilayer forward spread and error back propagation network (BP neural network). In the corrosion assessment module, a three-layer BP neural network is used to build the atmospheric and marine corrosion damage model. It includes an input node layer where data are presented to the network, a hidden node layer that receives internal signals to process, and an output node layer that holds the network response to a given input. The ANN input variables and the exposure time are the main corrosion environmental factors that the grey relational analysis determines. The ANN output variables are the average corrosion rate for atmospheric corrosion, the average corrosion rate, the average pit depth, and the maximum pit depth for marine corrosion. Before the ANN is trained, the corrosion data is scaled to fit within the interval 0-1, which is a fundamental operation prior to training known as data homogenization (Cai et al., 1999). In order to forbid the ANN training to fall into the local minimum and reduce the network sensitivity to local error surfaces, a momentum coefficient is added to the ANN training. The number of nodes in the hidden layer, the connection weight values, and the threshold values are most important to the forecast accuracy of the BP network (Wang et al., 2006). However, until now there has been no valid theory with which to select the number of nodes in the hidden layer. Better precision can be obtained by comparing the calculation results with different numbers of nodes in the hidden layer. The connection weight values and threshold values are adjusted by the gradient descent algorithm

$$\Delta w_{ij}^{(l)} = -\eta \delta_j^{(l)} O_i^{(l-1)} \quad \Delta a_j^{(l)} = \eta \delta_j^{(l)} \quad \delta_j^{(l)} = \frac{\partial E_j^{(l)}}{\partial O_j^{(l)}} f'(I_j^{(l)}) \quad (7)$$

where $\Delta w_{ij}^{(l)}$ is the connection weight increment between j^{th} node of l layer with i^{th} node of $l-1$ layer, $\Delta a_j^{(l)}$ is the threshold increment of j^{th} node of l layer, η is the learning rate, $E_j^{(l)}$ is the output error $E_j^{(l)} = (O_j^{*(l)} - O_j^{(l)})^2 / 2$, $O_j^{*(l)}$ is the calculation output value, $O_j^{(l)}$ is the real output value, $I_j^{(l)}$ is the input value, and $f(x)$ is the activation function $f(x) = 1 / (1 + \exp(x))$.

In order to check the validity of the ANN algorithm, the corrosion data for A3 steel in the marine splash zone is used as an example for analysis. The number of nodes in the hidden layer, the learning rate, and the momentum coefficient are set as 4, 0.3, and 0.9 respectively. It can be seen in **Figure 3(a)** that in the beginning the mean output error drops sharply with the epoch followed by a very slow decrease. After training for 100000 epochs, the mean output error arrives at $9.7E-4$. **Figure 3(b)** shows the comparison between the calculation output value and the real output value after 100000 epochs. When the calculation output value is equal to the real output value, the data point will be on the straight line $y = x$. It can be seen in **Figure 3(b)** that all data points are close to the straight line.

Fracture mechanics is a method for predicting the failure of a structure containing cracks. It uses analytical solid mechanics methods to calculate the driving force on cracks and experimental solid mechanics methods to characterize the material resistance to fracture (Zhu et al., 2012). After the evolution of corrosion pit depth is predicted by the BP network, a fracture mechanics model is used to evaluate the residual life of the structure. Suppose that the structure is a plate with infinite length and finite thickness and there is a semi-circular crack in the center of the plate on the surface. If the plastic zone is considered,

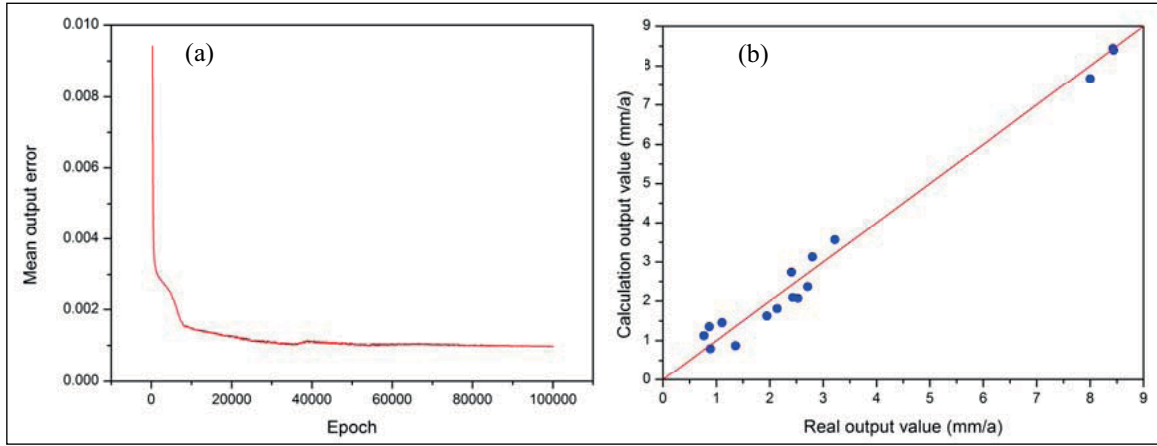


Figure 3: (a) Evolution of mean square error; (b) Comparison between calculation output value and real output value.

$$K_I = \frac{M\sigma\sqrt{\pi a}}{Q} \quad a = 0.85 \frac{\pi d^2}{2L} \quad M = 1.03 \sqrt{\frac{2L}{\pi a} \tan \frac{\pi a}{2L}} \quad Q = \sqrt{\left(\frac{\pi}{2}\right)^2 - 0.608 \left(\frac{\sigma}{\sigma_y}\right)^2} \quad (8)$$

where K_I is the stress intensity factor, σ is the tensile stress, M is the coefficient, Q is the shape parameter of crack, σ_y is the yield strength, a is the equivalent crack length, d is the corrosion pit depth, and L is the plate thickness.

Suppose that the corrosion pit depth is d_1 at current time t_1 . When the corrosion pit depth is d_2 at time t_2 , $K_{I2} = K_{IC}$, where K_{I2} is the stress intensity factor at time t_2 and K_{IC} is the fracture toughness. Thus d_2 is the critical corrosion pit depth. The residual life of the structure is equal to $t_2 - t_1$.

4 Conclusions

This paper introduces the overall architecture for the analysis of natural environment corrosion. Atmospheric and marine environmental data are analyzed by a statistical method. The W method's reliability in checking the normal distribution for a small sample is higher than that of the K-S method; however the W method is a special method to check normal distribution. It does not check any other distributions. The K-S test is a kind of universal test. It can check any distribution as long as the distribution function is known. Although the expected values and variances of the population can be estimated by the distribution function, a more reliable estimate is given by the confidence interval at different confidence levels. For atmospheric and marine corrosion data, the main corrosion environmental factors are confirmed by grey relational analysis. These factors are the input variables for the BP artificial neural network. This can simplify the BP artificial neural network modelling and improve the reliability of the model. After the relationship between corrosion pit depth and time is predicted by the BP network, the residual life of the structure can be obtained by a fracture mechanics calculation on the critical corrosion pit depth.

5 Acknowledgments

This work was supported by the National Science and Technology Major Project (No. 2011ZX06004-009).


6 References

- Cai, J.P., Cottis, R.A., & Lyon, S.B. (1999) Phenomenological modelling of atmospheric corrosion using an artificial neural network. *Corrosion Science* 41, pp 2001–2030.
- Carpenter, W.C. & Hoffman, M.E. (1995) Training Backprop Neural Networks. *Journal of AI Expert* 10, pp 30–33.
- Deng, J.L. (1989) Grey information space. *The Journal of Grey System* 1, pp 103–117.
- Huwang, L.C. & Hwang, J.T.G. (2002) Prediction and confidence intervals for nonlinear measurement error models without identifiability information. *Statistics & Probability Letters* 58, pp 355–362.
- Jurečková, J. & Picek, J. (2007) Shapiro–Wilk-type test of normality under nuisance regression and scale. *Computational Statistics & Data Analysis* 51, pp 5184–5191.
- Melchers, R.E. (2013) Long-term corrosion of cast irons and steel in marine and atmospheric environments. *Corrosion Science* 68, pp 186–194.
- Morcillo, M., Chico, B., Díaz, I., Cano, H., & Fuente, D. (2013) Atmospheric corrosion data of weathering steels. A review. *Corrosion Science* 77, pp 6–24.
- Schröer, G. & Trenkler, D. (1995) Exact and randomization distributions of Kolmogorov–Smirnov tests two or three samples. *Computational Statistics & Data Analysis* 20, pp 185–202.
- Shifler, D.A. (2005) Understanding material interactions in marine environments to promote extended structural life. *Corrosion Science* 47, pp 2335–2352.
- Srivastava, M.S. & Hui, T.K. (1987) On assessing multivariate normality based on Shapiro–Wilk W statistic. *Statistics & Probability Letters* 5, pp 15–18.
- Traverso, P. & Canepa, E. (2014) A review of studies on corrosion of metals and alloys in deep-sea environment. *Ocean Engineering* 87, pp 10–15.
- Vlček, O. & Huth, R. (2009) Is daily precipitation Gamma-distributed?: Adverse effects of an incorrect use of the Kolmogorov–Smirnov test. *Atmospheric Research* 93, pp 759–766.
- Wall, F.D., Martinez, M.A., Missert, N.A., Copeland, R.G., & Kilgo, A.C. (2005) Characterizing corrosion behavior under atmospheric conditions using electrochemical techniques. *Corrosion Science* 47, pp 17–32.
- Wang, H.T., Han, E.H., & Ke, W. (2006) Artificial neural network modeling for atmospheric corrosion of carbon steel and low alloy steel. *Corrosion Science and Protection Technology* 18, pp 144–147.
- Wang, H.T., Han, E.H., & Ke, W. (2006) Gray model and gray relation analysis for atmospheric corrosion of carbon steel and low alloy steel. *Corrosion Science and Protection Technology* 18, pp 278–280.
- Xiao, Y.D., Liang, C.F., Zheng, Q.F., Hu, X.J., Wang, L., Zhang, E.P., et al. (2002) “Nine five” data collection of atmospheric corrosion of materials, The Major Program of the National Natural Science Foundation of China (No. 59899141).
- Zhang, X.Z. (1994) Application of probability weighted moments to estimate parameters in Weibull distribution. *Marine Forecast* 11, pp 55–61.
- Zhu, X.K. & Joyce, J.A. (2012) Review of fracture toughness (G, K, J, CTOD, CTOA) testing and standardization. *Engineering Fracture Mechanics* 85, pp 1–46.
- Zhu, X.R., Lin, L.Y., Zhang, S.P., Liu, D.Y., Lin, Z.J., Jin, W.X., et al. (2002) “Nine five” data collection of marine corrosion of materials, The Major Program of the National Natural Science Foundation of China (No. 59899142).

How to cite this article: Wang, H, Yao, L, Li, X, Han, E-H and Tong, X 2015 Effective Utilization for Data of Natural Environment Corrosion of Materials. *Data Science Journal*, 14: 18, pp. 1–7, DOI: <http://dx.doi.org/10.5334/dsj-2015-018>

Published: 22 May 2015

Copyright: © 2015 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License (CC-BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/3.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 