Paper No.
19108

AMPP ANNUAL CONFERENCE + EXPO
MARCH 19-23, 2023 | DENVER, COLORADO

Mechanistic Model as a Bias to Machine Learning Algorithm for Confident Prediction of Corrosion

Ishan Patel
Institute for Corrosion & Multiphase Technology
Ohio University
342 W. State St.
Athens, OH 45701
USA

Gheorghe Bota
Institute for Corrosion & Multiphase Technology
Ohio University
342 W. State St.
Athens, OH 45701
USA

ABSTRACT

Bayesian network is employed to estimate a risk-based life cycle cost of corrosion for assets. It has been highly recognized that inclusion of mechanistic models to a Bayesian network can increase the confidence in estimation of corrosion rates. However, coefficients of mechanistic models are often unknown, especially when complex rate processes are involved, which discourages the usage of the model. A methodology is proposed here, to introduce a mechanistic model as a bias to a regressive machine learning (ML) algorithm. No attempts have been made to obtain phenomenological coefficients of the mechanistic model. Instead, a methodology is proposed to obtain a highly tuned parameter vector for a ML algorithm from a learning set of corrosion rate data.

Key words: corrosion modelling, machine learning algorithm, rate processes, supervised learning, non-linear regression

INTRODUCTION

A constant challenge persists among corrosion engineers to estimate and predict field corrosion rates despite the huge advancements in corrosion science. This situation has compelled the corrosion engineers to opt for the machine learning (ML) algorithms for corrosion prediction. However, the “blackbox” ML algorithms are not appreciated in high stakes decisions because they use arbitrary fitting models rather than scientific principles.[1] Learning achieved by such an algorithm is confined to itself and no useful knowledge can be acquired from it. Hence, it is necessary to include mechanistic models into machine learning algorithms for more confident prediction of corrosion rates.

According to the latest impact report by NACE, global economic loss due to corrosion was estimated to be $2.5 trillion in 2013, of which 15 to 35% could be saved by implementing proper corrosion control and management practices.[1] Bayesian network (BN), a probabilistic learning algorithm with cause-consequence type structure, is utilized in calculation of indirect costs from the failure, especially in cases when localized failure mechanisms are expected.[1] Learning, and consequently the reliability, of BN depends on input corrosion data, which can be generated by mechanistic models, mined from expert

knowledge, or simply being available from the field.[2] Among these, the mechanistic models are the most preferable to generate corrosion rate data used in preparing conditional probability tables for BN.[2,3]

However, there is an inherent inaccuracy associated with any corrosion model either due to deficient mechanistic description or unknown coefficients, which leads to poor estimation of corrosion rates. Typically, a corrosion model is constructed out of the most feasible mechanistic description derived from the experimental results. However, the models often fail to estimate accurate field corrosion rates because coefficients of the rate equations are determined from laboratory data obtained in ideal environment which may be far from the field conditions.

A hybrid approach is adopted here to use ML algorithms with input of a mechanistic model as scientific information to obtain highly tuned parameters by using corrosion data for learning. The values of these parameters may not resemble to coefficients of the rate equations. However, prediction of such a scientifically informed ML (SciML) algorithm is more reliable than the “blackbox” ML algorithms. The proposed methodology serves as a general framework to integrate a mechanistic model into a ML algorithm for more confident prediction of corrosion in high stakes decisions, which can be applied to any corrosion system after specific modifications.

CORROSION MODEL

In general, physical modelling of any corrosion system involves electrochemical reactions, chemical reactions, and mass transport of corrosive species for which rate equations are derived from the principles of electrochemistry, chemical science, and fluid mechanics. When experimental corrosion rate data evidently conforms to a particular rate law, the reaction involved is considered to be limited by that step of the overall mechanism. After confirming the “rate limiting step” for given experimental conditions, the experimental variables are further manipulated in such a way that a different step of the mechanism becomes rate limiting and so on. This procedure is repeated until all anticipated steps of the mechanism are confirmed and the master corrosion rate equation may be derived by coupling the rates of individual steps. If some of the anticipated steps of the corrosion mechanism are not feasible to conform experimentally, then they must be deduced rationally. Such mechanistic models resemble to the true nature of corrosion phenomena and thus they are highly dependable for prediction of corrosion rates if accurate values of coefficients can be determined.

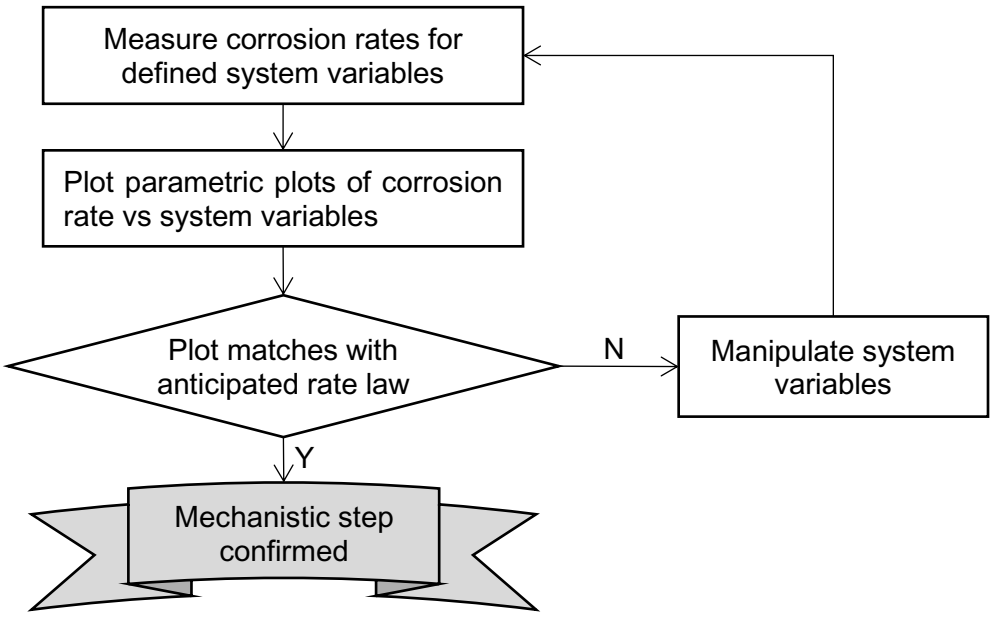
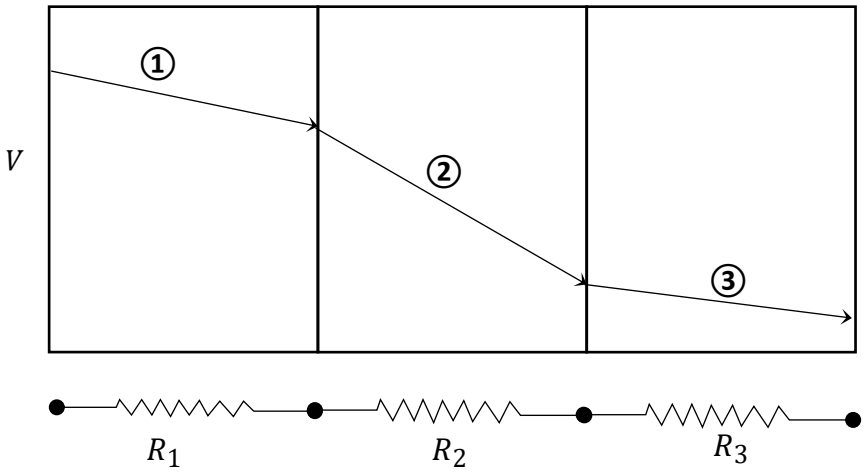


Figure 1: Steps to construct mechanistic corrosion model.

The algorithm described above can be used to determine the rate limiting step in a mechanistic corrosion model. However, an accurate value of a coefficient of that step cannot be exclusively determined from the values of corrosion rates and associated system variables. The rate limiting step is connected with prior and later steps of the mechanistic sequence which places fundamental limitation over determination of coefficients associated with the rate limiting step. This is explained by
Figure **2** by using an analogy of flux/rate with electrical current and mechanistic steps as resistances in series. Consider that the experiment requires to determine all three resistances of unknown value by measuring current for an applied voltage. From this experiment, only total resistance can be determined and if value of one of the resistances, e.g., R_2, is significantly higher than the rest, its value can be approximated to be equal to total resistance by ignoring the others. This introduces fundamental error of $R_t - (R_1 + R_3)$. Similarly, determination of coefficients associated with the rate limiting step always has an error associated with it, especially significant when the rest of the steps are not much faster than the rate limiting step.



$j = \frac{V}{R_t}$ where, $R_t = R_1 + R_2 + R_3$

Figure 2: Analogy of corrosion process with several mechanistic steps with electrical circuit with series of resistors.

Moreover, prediction of field corrosion rates by a mechanistic model is often inaccurate because continuously evolving system variables in the field, e.g., composition, temperature, pressure are different than in the laboratory environment. The present methodology attempts to overcome this limitation by using a mechanistic model in a regressive machine learning algorithm. Any mechanistic model can be treated to obtain model constants using this methodology if it is physically and mathematically consistent.

The first step of building a corrosion model requires listing all the mechanistic steps and associated rate equations. For example, a simple corrosion model consists of transport of corrosive species towards the surface followed by surface electrochemical or chemical reactions. These steps are listed with respective rate equations in Table 1.

The second step is to connect the mechanistic steps by the condition of mass conservation. Concentration of corrosive species decrease and that of corrosion product increase during a nonflow process for which a mass accumulation term appears in the equation. However, in flow condition, which is the most common in the field, rates of all steps of the mechanism are equalized in absence of any mass accumulation.

Table 1 Mechanistic steps in a simple corrosion model with associated rate equations

Mechanistic Step	Rate Equation	Coefficients
Mass transport of corrosive species towards the surface	Flux $f = \kappa\left(C^{\{b\}} - C^{\{s\}}\right)$	Mass transfer coefficient κ
Reaction at the surface	• Electrochemical reaction in cathodic reaction control such as reduction of hydrogen Rate $r = A\exp\left[\frac{-(E_a+\beta F\eta)}{RT}\right]C^{\{s\}}$ or • Chemical reaction in non-aqueous corrosion Rate $r = A\exp\left(\frac{-E_a}{RT}\right)C^{\{s\}}$	Pre-exponential factor A Activation energy E_a Overpotential η Tafel slope β

C= concentration of corrosive species, $T =$ absolute temperature, superscripts represent location of corrosive species, $b =$ bulk fluid, and $s =$ near surface of metal

The following equations can be obtained for overall rate of corrosion for non-accumulating or flow condition by equating flux with reaction rate per unit area from Table 1.

$$r = \frac{\kappa A\exp\left[\frac{-(E_a+\beta F\eta)}{RT}\right]C}{A\exp\left[\frac{-(E_a+\beta F\eta)}{RT}\right]+\kappa} \quad \text{for aqueous corrosion reactions} \qquad (1)$$

$$\text{and } r = \frac{\kappa A exp\left(\frac{-E_a}{RT}\right)C}{A exp\left(\frac{-E_a}{RT}\right)+\kappa} \quad \text{for non-aqueous corrosion reactions} \qquad (2)$$

Superscript $\{b\}$ is dropped in the above equations for the purpose of simplicity. A procedure is described here to determine the values of the coefficients listed in Table 1 using regressive ML algorithm.

MACHINE LEARNING ALGORITHM

Simple regressive algorithm is explained here using equation (2) which can be converted to regression function($\zeta_{\mathbf{b}}$), where $\mathbf{b} = (\kappa, A, E_a)$as shown, and values of coefficients $\mathbf{b}$ can be obtained using a "labelled" training set $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ mined from field corrosion data, where, $\mathbf{x}_i = (C_i, T_i)$ and $y_i = r_i$ are for the present case.

A regressive function which supervises the learning can be written as shown by equation (3)

$$\zeta_{\mathbf{b}}(\mathbf{x}_i) = \zeta_{\mathbf{b}}(C_i, T_i) = \frac{\kappa A exp\left(\frac{-E_a}{RT_i}\right)C_i}{A exp\left(\frac{-E_a}{RT_i}\right)+\kappa} \qquad (3)$$

Commonly used squared error loss function can be adopted as optimization objective expression as shown by equation (4).

$$e = \frac{1}{n}\sum_{i=1}^{n}\left[\zeta_{\kappa,A,E_a}(C_i, T_i) - r_i\right]^2 \quad (4)$$

$$e = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\kappa A exp\left(\frac{-E_a}{RT_i}\right)C_i}{A exp\left(\frac{-E_a}{RT_i}\right)+\kappa} - r_i\right]^2 \quad (5)$$

Square error loss expressed by equation (4) must be minimized for the optimization of coefficients for a given training set $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n} = \{(C_i, T_i, r_i)\}_{i=1}^{n}$. Gradient of error loss with respect to coefficients of the model is set to zero to obtain system of equations with dimension $D = 3$ in the present case as specified by equation (6).

$$\nabla_{\mathbf{b}} e = \mathbf{0} \quad (6)$$

Where, $\nabla_{\mathbf{b}} = \left(\frac{\partial}{\partial\kappa}, \frac{\partial}{\partial A}, \frac{\partial}{\partial E_a}\right)$ and $\mathbf{0}$ is a null vector.

The system of non-linear equations resulting from equation (6) can be solved by iterative algorithm such as Newton's method to obtain solution vector of coefficients $\mathbf{b} = (\kappa, A, E_a)$. Care should be taken to define realistic ranges for the values of coefficients so that solution converges with minimum computation power. Ranges of coefficients, or at least the order of magnitude, can be derived from reference laboratory data. Convergence tolerance for iterative algorithm should be defined considering reference values of the coefficients.

CONCLUSIONS

A methodology to use mechanistic model as a bias to machine learning algorithm is proposed to be directly applied to flow processes which is the most common system. Values of coefficients derived from regression may not resemble the true value, but such physically informed machine learning algorithm can be used for fine tuning of model parameters and decrease the gap between predicted and measured values of corrosion rates.

REFERENCES

1. IMPACT (2016), "International Measures of Prevention, Application, and Economics of Corrosion Technologies" (Houston, Texas: NACE International).
2. F. Ayello, S. Jain, N. Sridhar, G. H. Koch, "Quantitive Assessment of Corrosion Probability—A Bayesian Network Approach," *CORROSION*, *70* (11), (2014): pp. 1128–1147.
3. G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, "Physics-Informed Machine Learning," *Nat. Rev. Phys.*, *3* (6), (2021): pp. 422–440.