Korean J. Chem. Eng., 33(7), 1999-2006 (2016)
DOI: 10.1007/s11814-016-0083-9

pISSN: 0256-1115
eISSN: 1975-7220

Kriging models for forecasting crude unit overhead corrosion

Kyungjae Tak*, Junghwan Kim, Hweeung Kwon*, Jae Hyun Cho***, and Il Moon*,†**

*Department of Chemical and Biomolecular Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
**Korea Institute of Industrial Technology, 55 Jongga-ro, Jung-gu, Ulsan 44413, Korea
***Engineering Development Research Center, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
(*Received 18 November 2015 • accepted 18 March 2016*)

Abstract−Crude unit overhead corrosion is a major issue in the refinery field. However, the corrosion models in the literature are difficult to apply to real refinery processes due to the characteristics of corrosion. We propose a Kriging model, an advanced statistical tool for geostatistics, to forecast the corrosion rate in a real refinery plant. Instead of spatial coordinates, the proposed model employs the non-spatial coordinates of six key corrosion variables: H_2S, Cl^-, Fe^{2+}, NH_3, pH, and flowrate. The Kriging model is compared with two well-known forecasting models, multiple linear regression and an artificial neural network. To overcome the insufficiency of the number of data sets measured in the plant to use the six non-spatial coordinates, the significance probability is applied to reduce the dimensions from six to four. Among all the developed models in this paper, the Kriging model with four corrosion variables showed the best forecasting performance.

Keywords: Refinery Process, Corrosion Rate, Forecasting Model, Multiple Linear Regression, Artificial Neural Network, Kriging Model

INTRODUCTION

Metals are among the most used materials in many industries, especially the process industry. Failures derived from corrosion damage can not only lead to economic loss but also to human loss and environmental pollution. In the US, direct corrosion cost is estimated to be around $276 billion on an annual basis, which is 3.1% of the U.S. gross domestic product [1]. Therefore, the most common problem in industries is corrosion because the corrosion reaction in metals is unavoidable. However, this process can be controlled.

Since refinery plants can be in operation for several years without a shutdown, corrosion is a major operational issue. Much research on corrosion mechanisms occurring in equipment used in refinery, petrochemical, and gas plants has focused on CO_2 corrosion, which is dominant in a set of corrosion reaction mechanisms such as the following:

$$Fe+CO_2+H_2O\rightarrow FeCO_3+H_2 \quad (1)$$

Nesic reviewed key issues regarding internal corrosion and analyzed the effects of pH, CO_2 partial pressure, temperature, flowrate, inhibitors, steel type, and other factors on the corrosion rate [2]. A number of corrosion models have been proposed to predict the corrosion rate, and they can be classified into four categories: semi-empirical, electrochemistry, transport, and empirical models. The semi-empirical models were developed at an early stage. In 1975, de Waard and Milliams analyzed the relationship between the corrosion rate and the CO_2 partial pressure at various temperatures [3]. Further, de Waard et al. extended it to a parallel resistance model of the CO_2 corrosion rate along with the fluid flowrate [4]. Hellevik et al. investigated the planning of pipe inspection and replacement based on cost optimal reliability using the model by de Waard et al. [5]. Electrochemistry models are based on the Butler-Volmer equation and focus on the surface corrosion reaction [6]. Transport models assume that mass transfer is a rate-determining step of corrosion reaction due to the formation of a sulfide layer. The mass transfer of corrosive components to a pipe surface is interrupted by a porous outer layer and a corroded layer. In an iron sulfide system, several types of ferrous sulfide layers are formed according to the conditions. The most commonly formed in the general operating conditions of refinery plants is mackinawite film [7]. Sun and Nesic investigated the kinetics of corrosion layer formation and developed a transport model for mild steel corrosion [7,8]. Their model takes into account the H_2S and H^+ corrosions as well as CO_2 corrosion.

$$Fe+H_2S\rightarrow FeS+H_2 \quad (2)$$

$$Fe+2H^+\rightarrow Fe^{2+}+H_2 \quad (3)$$

Kim et al. added Cl^- corrosion to the model of Sun and Nesic because refinery plants claimed that Cl^- has an additional effect on the corrosion rate [9].

$$Fe+2HCl\rightarrow FeCl_2+H_2 \quad (4)$$

Song developed a CO_2 corrosion model by combining electrochemical corrosion reactions with mass transfer, with an emphasis on CO_2 diffusion and reactions in the boundary layer as well as overall electric charge conservation and reactions on a steel surface [10]. In the case of empirical models, many studies have employed various function types to represent the results of corrosion rate exper-

†To whom correspondence should be addressed.
E-mail: ilmoon@yonsei.ac.kr

iments. Gruber et al. used an Arrhenius function for steel temperature and flue gas temperature and velocity to forecast the corrosion rate of biomass fired boilers [11]. They applied the model to computational fluid dynamic (CFD) simulations for an analysis of local corrosion potential. Dhanapal et al. forecasted the corrosion rate of magnesium alloys under salt fog environment using a quadratic function of chloride ion concentration, pH, and time [12]. Khadom developed regression and artificial neural network models to find best mathematical equation of corrosion reaction data of mild steel as a function of inhibitor concentration, temperature, and acid concentration [13]. Kinetic, exponential, and polynomial equations as well as artificial neural network were considered for this objective.

Although these achievements have contributed to our understanding of the processes involved, the corrosion models published in the literature have limitations when applying them to real plants. First, many parameters in the models have to be determined experimentally. However, experiments for corrosion reactions are difficult to carry out because most of the experimental conditions do not reflect the reality of the operating conditions of the refinery plants. For example, the phycochemical properties of crude oil are different from oil fields, and accurate composition of the oil is therefore hard to determine [14]. Moreover, the duration of the experimental period may not be sufficient to be accurate. Experiments are usually done for only a few weeks, but refinery plants may operate for years without a shutdown [15,16]. Second, corrosion reactions are very sensitive, depending on the absence of a corrosive component or its presence, even in a very small amount. In addition, despite the number of experiments and in-depth research, the corrosion mechanisms have not been clearly examined. Lee showed that the existence of a very small amount of H_2S reduced the corrosion rate about 90% in comparison to the pure CO_2 corrosion rate [17]. Kim et al. added Cl^- corrosion to the mass transfer model developed by Sun and Nesic in order to derive more accurate mechanisms. However, their model yielded underestimated results so that parameter estimation had to be performed to solve the underestimated results [9].

Since the theoretical corrosion models explained above are not suitable when used for real refinery plants, we propose the use of statistical models based on real plant data to forecast the corrosion rate. According to Kim et al., the top part of a crude oil distillation unit (CDU) is the most corrosive region because the carbon steel for the top of a CDU is more corrosive than the hastelloy for the bottom of a CDU [18]. Therefore, this study focuses on the corrosion of a crude unit overhead column. The forecasting model is developed by adopting an advanced statistical method used in geostatistics called the Kriging model. Developed by Krige, a mining engineer [19], it is considered to be an alternative to linear regression because it forecasts the value of a dependent variable at an interest point by a linear combination of the known values of the dependent variable at the nearest neighborhood points. This is the main difference between the Kriging model and other statistical models.

As the Kriging model has its origins in geostatistics, it is commonly used in geology, geography, and meteorology. However, its use has been recently extended to other fields. Lang et al. used the Kriging model for multi-scale modeling [20]. They used a Kriging model as an alternative to CFD models for a coal gasifier and combustor and embedded it to a process simulator in what is known as a reduced order modeling framework. Another example is an application of the Kriging model to social system modeling [21]. Although the Kriging model has been widely used in many areas, most researches have applied it to a spatial coordinate system such as a CFD simulation. In contrast, this study extends and analyzes the Kriging model to non-spatial systems. For comparison, two widely used statistical models, multiple linear regression (MLR) and artificial neural network (ANN), are also employed. A set of six key corrosion variables (H_2S, Cl^-, Fe^{2+}, NH_3, pH, and flowrate) and the corrosion rate are measured at the overhead column and treated statistically before establishing the models. Since the Krig-

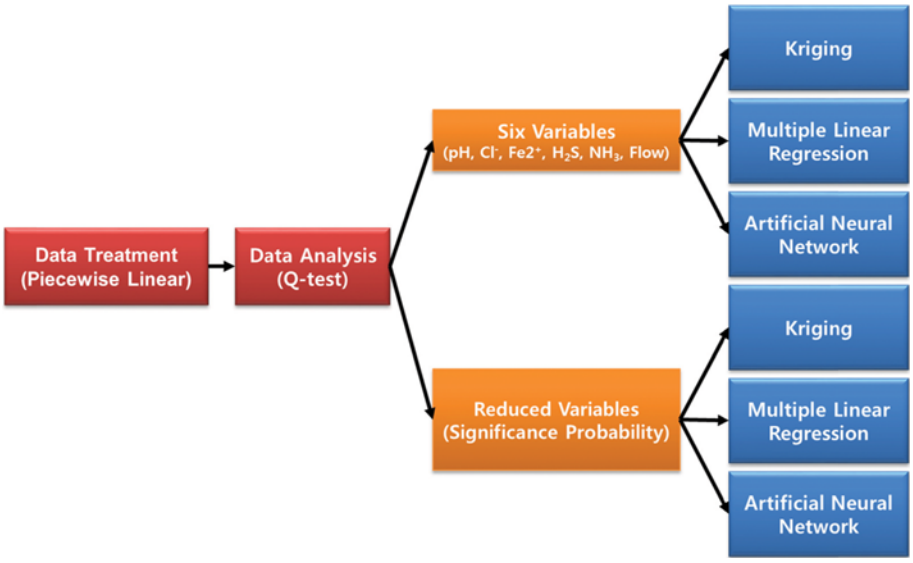

Fig. 1. Framework for model development.

ing model for forecasting a corrosion rate employs non-spatial coordinates, a six-dimensional hyperplane with six corrosion variables was used for the model. As the dimensions of a hyperplane are increased, more points on the hyperplane are required to obtain accurate information. Due to the insufficient number of data sets from a real plant to use all of the six key corrosion variables, the six variables were reduced to four by employing significance probability. As a result, statistical models using the six key corrosion variables were compared to those with four key corrosion variables.

METHODOLOGY

Our framework for developing corrosion forecasting models is illustrated in Fig. 1. The plant data were reconciled by applying a piecewise linearization and time weighted average method. Bad outlier data were rejected by a Q-test. Then, three different statistical models were developed with the final data set: MLR and ANN models as well-known methods, and the Kriging model, an advanced statistical method in geostatistics. Since it is important for the Kriging model to find accurate correlation of corrosion rates between data sets with the distance on the hyperplane, the model requires more data sets for a higher dimensional hyperplane. Therefore, the number of corrosion variables required can be reduced by significance probability when the number of data sets is insufficient to use all the six corrosion variables that can be measured in a refinery plant.

1. Data Treatment and Analysis

The plant data sets were measured over two years. The set of variables included the pH, Cl^-, Fe^{2+}, H_2S, NH_3, flowrate, and corrosion rate. All these variables were measured at an irregular daily interval except for the corrosion rate, which was measured at an irregular weekly interval. Moreover, unlike the corrosion rate, which is obtained by checking the amount of corrosion during a time interval, the corrosion variables were checked at a specific time. Fig. 2 shows the reconciliation method in this study. To reconcile the six corrosion variables and the corrosion rate, a piecewise linearization was employed. Then, a time-weighted method was implemented to find the average values of the variables based on the assumption that all variables are constant during an interval. Consequently, the corrosion variables, V_i, were reconciled with the corrosion variable, CR, as shown by the dashed lines in Fig. 2. Next, some outliers of the corrosion rate data were rejected based on a Q-test with 95% confidence. As a result, 42 data sets were obtained: 35 data sets were used for training samples and the other seven data sets for validation samples.

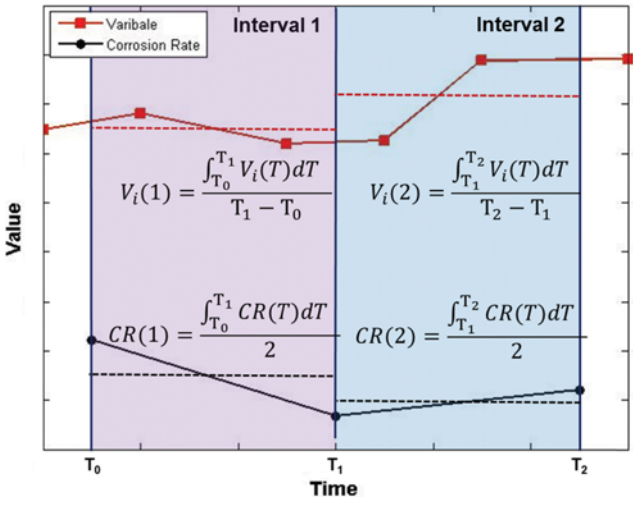

Fig. 2. Data treatment using piecewise linearization and the time weighted average method.

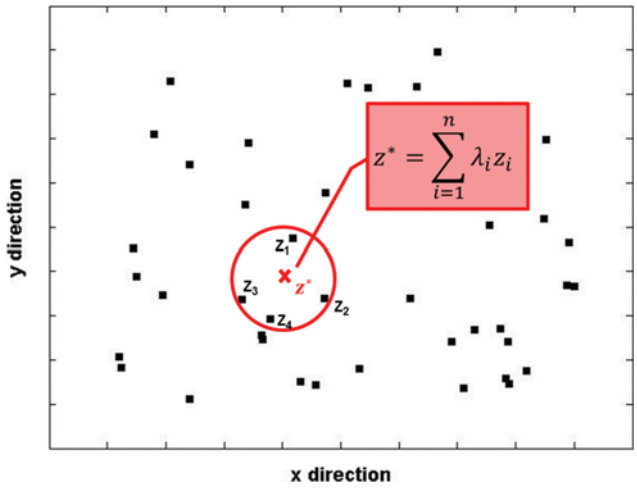

Fig. 3. Scheme of the Kriging model.

2. Kriging Model

A Kriging model forecasts a value at an interest point using a linear combination of the known values at the neighborhood points, as illustrated in Fig. 3. Thus, the model is regarded as an alternative to linear regression. Unlike other statistical models that use an explicit relation among variables in a sample data region, this model uses an implicit relation by employing a variogram, used in geostatistics, which is the spatial correlation function among the measured values.

The basic form of a Kriging model is expressed in Eq. (5)

$$Y^* = \Sigma_{i=1}^{n} \lambda_i Y_i \tag{5}$$

Y_i is the value at point i, λs are the weighted parameters, and n is the number of neighbor points. To determine the weighted parameters, a system of equations must be solved. The calculation method for the weighted parameters differs, depending on whether simple Kriging, ordinary Kriging, block Kriging, co-Kriging, or universal Kriging is performed. Simple Kriging determines the weighted parameters for minimizing forecasting errors but has a bias. On the other hand, ordinary Kriging minimizes the error variance to overcome the bias problem. Block Kriging is not point Kriging; co-Kriging uses multiple dependent variables; and universal Kriging considers the spatial distribution trend. Eq. (6) is the system of equations for simple Kriging.

$$\Sigma_{j=1}^{n} \lambda(i) C(i, j) = c(i) \qquad \text{for } i \in 1, 2, \ldots, n \tag{6}$$

C(i, j) is the covariance value at the distance between the neighbor points i and j, and c(i) is the covariance value at the distance between the target point and the neighbor point i. These covari-

ance values can be obtained using the covariance function derived from the variogram, as shown in Eq. (7).

$$C(\mathbf{h})=C(0)-\gamma(\mathbf{h})=\text{Sill}-\gamma(\mathbf{h}) \quad (7)$$

In Eq. (7), **h** is the separation distance between two points and γ is the variogram. Sill in the equation is the covariance value at a distance of zero or the maximum variability between pairs of points. In other words, it is the upper boundary of the variogram tending to an infinite distance because of the covariance value, C(**h**), which decreases to zero as the distance increases. As the weighted parameters, λs, are obtained from the variogram, the accuracy of the variogram influences the forecasting performance significantly.

A variogram explains the correlation of the data with distance. The experimental variogram is half of the mean square difference between pairs of points that have a distance of h, as described in Eq. (8).

$$\gamma(\mathbf{h})=\frac{1}{2|N(h)|}\Sigma_{(i,j)\in N(h)}(z_i-z_j)^2 \quad (8)$$

|N(**h**)| is the number of pairs, z is the dependent variable value at a point, and (i, j) is the pair of points separated by the distance, h. The theoretical variogram can be obtained from the experimental variogram.

This study employs ordinary Kriging to forecast the corrosion rate in a refinery process, as it is point Kriging with a single dependent variable (corrosion rate) under the assumption of a uniform spatial distribution trend. To apply ordinary Kriging, Eq. (9) is added to the system of equations for simple Kriging.

$$\Sigma_{i=1}^{n}\lambda_i=1 \quad (9)$$

In contrast to other applications of the Kriging model in the literature, this study uses a hyperplane with the coordinates of key corrosion variables, non-spatial coordinates. Linear combinations using the three, five, and seven nearest points were made and compared with a case using all points.

3. Other Forecasting Models

In this study, MLR and ANN models were selected to compare the forecasting performance of the Kriging model. Linear regression is the most common statistical method due to its simplicity. A dependent variable to be forecasted is expressed by a linear combination of the independent variables.

$$Y=\alpha+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n \quad (10)$$

where Y is the dependent variable, X_i is the independent variable, and α and β_i are the parameters. In Eq. (10), it is easy to analyze the effect of each independent variable on the dependent variable and to recognize the relationship between these independent variables and the dependent variable. However, the MLR model has limitations to forecast many natural phenomena that have intrinsic behaviors of nonlinearity and complexity.

An ANN model is an artificial intelligence method that mimics the neural network of the human brain. Due to its strength in solving complex and nonlinear system, the majority of applications include energy consumption, solar radiation, oil prices, rate of heat transfer, process optimization, classification and diagnosis of cancers, and correlation of mixture density [22-24]. However, there are a few disadvantages to this method. One major disadvantage is the over-fitting problem where data fits well in a sample region but shows poor forecasting performance in other regions. In addition, it is hard to derive the relationships among the variables due to the complexity of its form. Another problem is the high degree of freedom, requiring many choices such as transfer function types, learning algorithms, numbers of hidden layers and hidden neurons in each layer, and the number of epochs.

Among the many different architectures and methods of ANN models, a feed-forward back-propagation ANN model is used in this study as it is the most common one, and it is simple to derive. One hidden layer applied in the model contains a number of hidden neurons from one to ten. For the nmber of epochs, 10,000 and 50,000 epochs are employed in applying the back-propagation learning algorithm. A sigmoid function is used as the transfer function in the models.

$$Y=1/[1+\exp(X)] \quad (11)$$

4. Variable Reduction

Variable reduction was carried out for several reasons. One reason is that using all six of the key corrosion variables does not guarantee the best forecasting result. Some variables may have little effect on the corrosion rate, thereby increasing the complexity of the model. Cho et al. analyzed the effect of the number of input variables of the ANN model, and the results showed that the input variable selection had a significant impact on the ANN model [25]. For the Kriging model, 42 data sets were not sufficient to apply for the six-dimensional hyperplane because the higher dimension requires more sample data sets to find an appropriate variogram. For example, although 11 points are sufficient to obtain samples from 0 to 1 with a 0.1 unit of interval in one dimension, 121 points are required in two dimensions. Thus, variable reduction was conducted using significance probability.

The significance probability checks whether there is a linear relationship between the independent variables and the dependent one. If the significance probability is higher than 5%, the most uncorrelated variable is eliminated. This is repeated until the significance probability becomes less than 5%. Using this technique, some independent variables that have little relationship with the dependent one are eliminated with a confidence value of 95%.

RESULTS AND DISCUSSION

1. Six Key Corrosion Variables

The statistical models with six key corrosion variables were developed with 35 data sets as training samples and verified with seven data sets as validation samples. All the variables were normalized and then used in the models.

Fig. 4 shows the experimental and theoretical variogram. The separation distance in the figure indicates the similarity of the six key corrosion variables between the training sample points. Since the six variables were employed as coordinates of the hyperplane with only 35 data sets, there were no experimental variogram values in the range of the separation distance from zero to 0.2. Thus, the absence of an experimental variogram value in the region leads to inaccuracy of the theoretical variogram. Fig. 4 shows that there

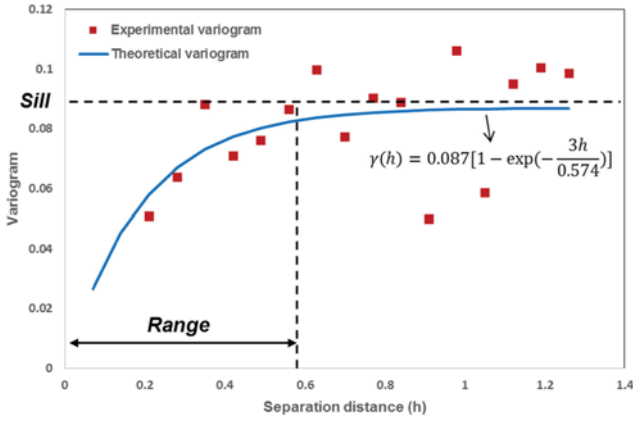

Fig. 4. Experimental and theoretical variogram in the case with six corrosion variables.

is a spatial correlation with the separation distance on the hyperplane until a separation distance of 0.6, which is called a *range*. A variogram value out of the *range* has no relationship along with the separation distance. Parameter estimation using an optimization tool was made to obtain the theoretical variogram. An exponential model was used as the theoretical variogram, which has a set of consequences of a *range* and a *sill* of 0.574 and 0.087, respectively.

The mean square errors (MSEs) of the Kriging models using the nearest 3, 5, and 7 points and all of the points in the validation region showed 0.541, 0.625, 0.572, and 0.351, respectively. The results of the nearest 3, 5, and 7 point cases were worse than that of the all-point case because of the absence of close points within a distance of 0.2. As shown in Fig. 5(a), the MSE in the training region is exactly zero because of the exact nature of the Kriging model between the measurement and forecast in the region, as opposed to other models. This exactitude means the exact reproduction of values at the already known points.

Fig. 5(b) depicts the MLR model result. The MSE of the validation samples is 0.420, and that of the training samples is 1.126. Although the set of parameters was determined by the training samples, the result in the validation region showed better performance in terms of the MSE. The main reason for this is thought to be the range of the corrosion rate: from 0.456 MPY to 4.867 MPY in the training region, but from 1.217 MPY to 2.920 MPY in the validation region. Moreover, the forecast result does not indicate a trend in the measured values.

Forty different types of ANN models are comprised of the combinations of the hidden neurons, presence/absence of a bias neuron, and epochs. Table 1 shows the results of the ANN models in terms of the MSE. As the number of hidden neurons increases, the MSE in the training region shows a decreasing tendency in all cases. On the other hand, the MSE in the validation region has an increasing trend until about 4-6 hidden neurons, showing an inconsistent trend in cases that have more hidden neurons. However, the small data set may be the reason for this behavior, because if the hidden neurons are increased, more parameters should be determined. The bias-neuron-added models generally show poor forecasting performance compared to the models without a bias neuron. The number of epochs also affects the results and intensi-

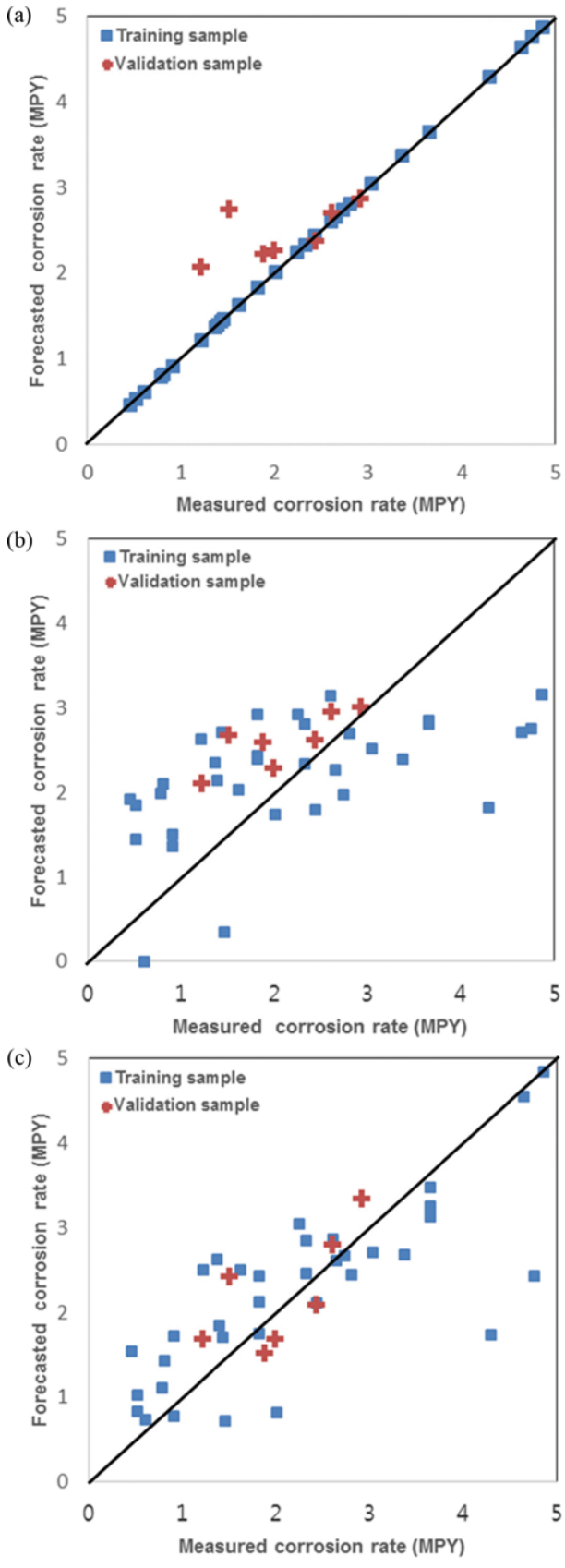

Fig. 5. The results in the case with six corrosion variables: (a) Kriging model with all points, (b) MLR model, and (c) ANN model (ANN3 with 10,000 epochs).

Table 1. MSE values of the ANN models in the six corrosion variable case

Model[a]	MSEs of 10,000 epochs		MSEs of 50,000 epochs	
	Training	Validation	Training	Validation
ANN 1	1.046	0.611	0.973	0.622
ANN 2	0.699	0.249	0.610	0.468
ANN 3	0.676	0.232	0.583	0.998
ANN 4	0.207	3.053	0.137	1.939
ANN 5	0.192	1.375	0.037	1.301
ANN 6	0.136	0.975	0.017	4.744
ANN 7	0.619	0.359	0.001	3.798
ANN 8	0.124	1.479	0.006	3.921
ANN 9	0.107	1.315	0.031	4.002
ANN 10	0.132	3.472	0.048	2.036
ANN 1b	1.020	0.513	0.947	0.657
ANN 2b	0.624	1.448	0.562	0.637
ANN 3b	0.337	2.077	0.208	2.806
ANN 4b	0.263	1.950	0.077	3.165
ANN 5b	0.145	1.411	0.035	3.028
ANN 6b	0.096	2.707	0.007	6.099
ANN 7b	0.069	1.819	0.001	3.850
ANN 8b	0.085	1.505	0.029	3.866
ANN 9b	0.054	1.846	0.010	6.814
ANN 10b	0.106	1.329	0.006	2.984

[a]Number indicates the number of hidden neurons and b indicates the bias neuron-added model

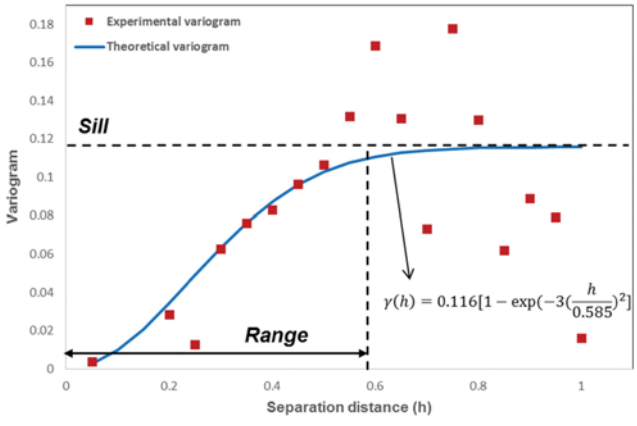

Fig. 6. Experimental and the theoretical variogram in the case with four corrosion variables.

fication of the over-fitting problem. The best ANN model is a case comprised of three hidden neurons and 10,000 epochs without a bias neuron, as shown in Fig. 5(c). Unlike the MLR model, the best ANN model gives good forecasts of the trend of the measured corrosion rate and produces almost half of the MSE values of both the training region and the validation region compared to the MLR model.

As compared to the MLR and ANN models in Fig. 5, the Kriging model exhibits a similar fit to the MLR model, which has a smaller range of forecasted values in the validation region than that

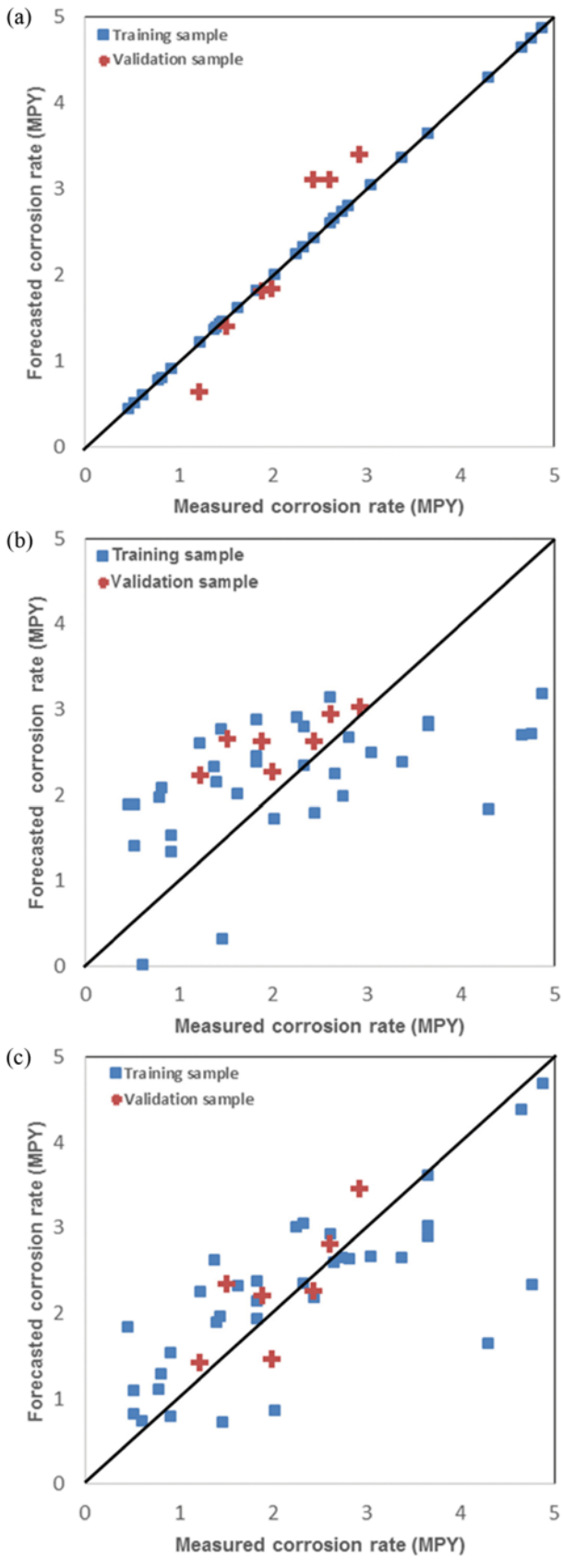

Fig. 7. The results in the case with four corrosion variables: (a) Kriging model with the nearest five points, (b) MLR model, and (c) ANN model (ANN4 with 10,000 epochs).

of the measured values in the same region. The Kriging model also shows a big gap between the measured value and the forecasted value at one small corrosion rate point of approximately 1.5 MPY in the validation region. Consequently, the ANN model provides the best result among the models in the case with six key corrosion variables. However, the Kriging model is capable of improvement if a more accurate variogram is obtained. This can be achieved by addition of more data sets or a reduction in the number of variables.

2. Reduction of the Six Key Corrosion Variables to Four

The six key corrosion variables were reduced to four to obtain a significance probability of less than 5%, as explained in the variable reduction section. Among the six corrosion variables, NH_3 and Fe^{2+} were eliminated, and then statistical models were developed again with the remaining four corrosion variables comprising the pH, Cl^-, H_2S, and flowrate.

The experimental and theoretical variogram for the Kriging models in the case with four corrosion variables were obtained as shown Fig. 6. There are close points within a separation distance of 0.2, and the number of points in the *range* region increased in comparison to the case of the six corrosion variables. As a result, the experimental variogram shows a clear increasing trend in the *range* region as the separation distance increases. The Gauss model, as the theoretical variogram, fits the experimental variogram well. The set of parameters of the model are obtained by parameter estimation. Compared to the case of six corrosion variables, the *range* increases from 0.574 to 0.585. Also, there is an increase in the *sill* from 0.087 to 0.116.

Fig. 7(a) illustrates the best Kriging result among the Kriging models in the case with four corrosion variables. The MSEs of the Kriging models in the validation region that use the nearest 3, 5, and 7 points were 0.377, 0.183, and 0.375, respectively. In comparison to the MSEs in the case with six corrosion variables, which were 0.541, 0.625, and 0.351, the case with four corrosion variables exhibited a better fit, although the all-point Kriging model with four corrosion variables showed worse performance than that of the six variables in MSE from 0.351 to 0.864. The Kriging model using the nearest five points produced the lowest MSE value among the Kriging models in the case with four corrosion variables. The range of the forecasted values in the validation region became larger than the case with six variables, as can be seen in Fig. 5(a) and Fig. 7(a). Thus, the Kriging model produces improved forecasting performance and is reliable even at a small corrosion rate.

The MLR model with the four corrosion variables showed MSEs of 0.448 and 1.126 in the validation and training regions, respectively. The forecasting performance of the model with four corrosion variables exhibited few differences compared to the model with six corrosion variables because the set of the eliminated variables was determined by applying the significance probability to find the set of data that have little linear relationship with the corrosion rate.

The results of the ANN models with four corrosion variables are listed in Table 2. In comparison to the ANN models with six corrosion variables, the results with the four corrosion variables generally showed significantly improved results. However, the best result in the case with four corrosion variables was similar to that with six variables in terms of the MSE, which was 0.209 vs. 0.232.

Table 2. MSE values of the ANN models in the four corrosion variable case

Model[a]	MSEs of 10,000 epochs		MSEs of 50,000 epochs	
	Training	Validation	Training	Validation
ANN 1	1.355	0.440	1.289	0.635
ANN 2	0.729	0.233	0.718	0.280
ANN 3	0.700	0.239	0.695	0.230
ANN 4	0.713	0.209	0.234	0.862
ANN 5	0.597	0.389	0.337	1.031
ANN 6	0.206	1.871	0.098	3.045
ANN 7	0.408	0.941	0.111	1.510
ANN 8	0.458	0.764	0.082	2.583
ANN 9	0.372	0.810	0.071	2.144
ANN 10	0.423	0.613	0.081	3.155
ANN 1b	1.035	0.587	1.034	0.574
ANN 2b	0.709	0.212	0.683	0.209
ANN 3b	0.380	1.049	0.319	0.997
ANN 4b	0.238	1.279	0.140	0.669
ANN 5b	0.252	0.675	0.111	3.478
ANN 6b	0.233	1.257	0.093	2.381
ANN 7b	0.236	1.321	0.031	0.825
ANN 8b	0.237	1.345	0.013	2.184
ANN 9b	0.263	0.737	0.042	0.999
ANN 10b	0.243	1.049	0.011	1.684

[a]Number indicates the number of hidden neurons and b indicates the bias neuron-added model

Table 3. Minimum MSE values in the validation region

Six corrosion variable case			Four corrosion variable case		
Kriging	MLR	ANN	Kriging	MLR	ANN
0.351	0.420	0.232	0.183	0.448	0.209

A summary of the minimum MSE values of each model in each case is depicted in Table 3. The Kriging model using the nearest five points in the case with four corrosion variables provides the best forecasting performance among all of the models developed in this study. Although the MSE of the Kriging model, 0.183, does not seem to show much difference from that of the ANN model, 0.209, the difference is significant. This is because the relative difference between these two values is higher than 10%, and the average value of the corrosion rate in the validation region is also small (2.08 mm/year). In terms of the mean absolute percentage error, the Kriging model (18.3%) is 2.9% less than the ANN model. Moreover, Fig. 7 highlights another big difference between the Kriging and ANN models. Unlike the ANN result, the Kriging model clearly shows an increasing trend in the validation region.

CONCLUSION

The aim of this study was to employ a geostatistical tool called Kriging for forecasting crude unit overhead corrosion. The Kriging

models developed in this study are based on the non-spatial coordinates of the key corrosion variables instead of the spatial coordinate systems that are commonly used for the Kriging model. The corrosion rate as well as the pH, Cl^-, Fe^{2+}, H_2S, NH_3, and flowrate were measured in a real refinery plant. Before the model development, real plant data were reconciled with a piecewise linearization and time weighted average method. A Q-test with a confidence of 95% was used to obtain the final 42 data sets. The parameters in the models were calculated using 35 data sets, and the models were tested using the remaining seven data sets. Two well-known and widely-used statistical models, MLR and the feed-forward back-propagation ANN models, were also developed to investigate the forecasting accuracy of the Kriging models. The Kriging models forecasted the corrosion rate under a specific condition by a linear combination of the corrosion rates at the closest data sets on the hyperplane. Therefore, the 35 training data sets were not sufficient to use the six key corrosion variables that are the coordinates of the six dimensional hyperplane. The insufficient number of data sets led to an inaccurate variogram, which determines the weighted parameters for the linear combination. Variable reduction was carried out by significance probability to solve the problem, resulting in the elimination of Fe^{2+} and NH_3. The results show that conventional statistical models are not suitable for forecasting CDU overhead corrosion and that the Kriging model with the five nearest points in cases with four corrosion variables provided the best results among all of the models developed in this study. Considering the great complexity of the corrosion mechanisms, using statistical models is a pragmatic approach for forecasting the corrosion rate, not only in refinery plants but also in other plants. Moreover, Kriging models are good alternatives to other statistical models in cases that have small data sets and a limited number of variables. They can also be extended to other non-spatial problems.

ACKNOWLEDGEMENT

This research was supported by the BK 21 Program funded by the Ministry of Education (MOE) of Korea and also supported by the Engineering Development Research Center (EDRC) funded by the Ministry of Trade, Industry & Energy (MOTIE).

REFERENCES

1. *Corrosion Costs and Preventive Strategies in the United States*, Report FHWARD-01-156, U.S. Federal Highway Administration (FHWA), Washington, DC (2002).
2. S. Nesic, *Corros. Sci.*, **49**, 4308 (2007).
3. C. De Waard and D. E. Milliams, *Corrosion*, **31**, 177 (1975).
4. C. De Waard, U. Lotz and A. Dugstad, In *Corrosion 95*, NACE International, Houston (1995).
5. S. G. Hellevik, I. Langen and J. D. Sorensen, *Int. J. Pressure Vessels Piping*, **76**, 527 (1999).
6. S. Nesic, J. Postlethwaite and S. Olsen, *Corrosion*, **52**, 280 (1996).
7. W. Sun and S. Nesic, *Corrosion*, **65**, 291 (2009).
8. W. Sun and S. Nesic, *Corrosion*, **64**, 334 (2008).
9. S. Kim, J. Kim and I. Moon, *Ind. Eng. Chem. Res.*, **50**, 12626 (2011).
10. F. M. Song, *Electrochim. Acta*, **55**, 689 (2010).
11. T. Gruber, R. Scharler and I. Obernberger, *Biomass Bioenerg.*, **79**, 145 (2015).
12. A. Dhanapal, S. R. Boopathy and V. Balasubramanian, *Mater. Des.*, **32**, 5066 (2011).
13. A. A. Khadom, *Korean J. Chem. Eng.*, **30**, 2197 (2013).
14. J. Kim, K. Tak and I. Moon, *Ind. Eng. Chem. Res.*, **51**, 10191 (2012).
15. M. Singer, B. Brown, A. Camacho and S. Nesic, *Corrosion*, **67**, 015004-1 (2011).
16. J. Kvarekval, R. Nyborg and H. Choi, In *Corrosion 2003*, NACE International, Houston (2003).
17. K. J. Lee, *A mechanistic modelling of CO_2 corrosion of mild steel in the presence of H_2S*, Ph.D. Dissertation, Ohio University, Athens, OH (2004).
18. J. Kim, W. Lim, Y. Lee, S. Kim, S. Park, S. Suh and I. Moon, *Ind. Eng. Chem. Res.*, **50**, 8272 (2011).
19. A. G. Journel and C. J. Huijbregts, *Mining Geostatistics, 5th Ed.*, Academic Press, London (1991).
20. Y. Lang, S. E. Zitney and L. T. Biegler, *Comput. Chem. Eng.*, **35**, 1705 (2011).
21. M. A. De Oliveira, O. Possamai, L. V. O. D. Valentina and C. A. Flesch, *Expert Syst. Appl.*, **40**, 272 (2013).
22. K. Movagharnejad, B. Mehdizadeh, M. Banihashemi and M. S. Kordkheili, *Energy*, **36**, 3979 (2011).
23. K. Song, S. Lee, S. Shin, H. J. Lee and C. Han, *Ind. Eng. Chem. Res.*, **53**, 5539 (2014).
24. J. S. Eswari and N. Chandraker, *Korean J. Chem. Eng.*, **33**, 1318 (2016).
25. S. G. Cho, K. T. No, E. M. Goh, J. K. Kim, J. H. Shin, Y. D. Joo and S. Seong, *Bull. Korean Chem. Soc.*, **26**, 399 (2005).