Formation factor analysis of the silica scale precipitated from geothermal waters by machine learning

T.Naritomi [1], K. Yonezu [1], S. Juhri[1],
S. Sato[2], N. Inoue[2], E. Watanabe[2], T. Yokoyama[2,3]
[1]Department of Earth Resources Engineering, Faculty of Engineering, Kyushu University, Fukuoka 819-0395, Japan
[2]Kyudensangyo Co. Ltd, Fukuoka 813-0043, Japan
[3]Department of Chemistry, Faculty of Science, Kyushu University, Fukuoka 819-0395, Japan
naritomi.tadahiro.835@s.kyushu-u.ac.jp

Keywords: *silica scale, machine learning, neural network*

ABSTRACT

Silica scale problem at geothermal power plants is one of the biggest issues to be solved to promote much install capacity because of decrease in power generation efficiency. However, the formation mechanism of silica scale has not yet been clarified. Previous geochemical investigations show that Al and Fe in geothermal water are closely associated with the promotion of silica scale formation. However, no study has been done to quantitatively estimate the contribution factors of each geochemical parameter to assess the countermeasure. Therefore, the purpose of this study is to create a machine learning model based on geochemical data and conduct factor analysis using partial derivative method. As for the geochemical data, the input data is geothermal water properties, and the output data is the weight of silica scale. Preliminary results showed that the time has positive contribution (46.0%), and the kinetics of monosilisic acid and reactive aluminum are also positive contribution (SiMKinetic_05_10: 5.7%, SiMKinetic_20_30: 3.3 %.). Though some of the Al parameters negatively contributes to weight of silica scale (AlRKinetic_45_60 : 4.0 %, AlR : 2.1%, Al-slope: 1.0%), there is no contribution of Fe parameters. The results from machine learning indicated that the silica scale weight is promoted with time past, and the Kinetic parameter is significant to predict. In addition, the chemical state of responsible species such as reactive Al may play an important role to inhibit the formation of silica scale as well based on the machine learning.

1. INTRODUCTION

In geothermal power generation, silica scale is one of the causes of reduced power generation efficiency and reduced capacity of reinjection wells. Under high temperature and pressure deep underground, the reaction between geothermal fluid (approximately 473-573K) and rocks causes silicon dioxide (silica SiO_2) in the rocks to dissolve into the fluid. At this time, all of the silica is dissolved as orthosilicic acid (silica monomer, Si (OH)_4). However, as the water passes through the production well and rises to the surface, the solubility of silica drops significantly due to the concentration of silicic acid in the liquid phase caused by gas-liquid separation and the drop in temperature, and the supersaturated orthosilicic acid polymerizes, causing silica to precipitate. This precipitated silica that adheres to pipes and other objects is called silica scale and can cause pipe blockages. It often precipitates in reinjection wells and surrounding strata, which have a long residence time, and in the worst-case scenario, additional work such as re-digging the reinjection well is required, causing major disruptions to operations.

The current method of inhibiting silica scale is widely used to adjust the pH by adding inexpensive sulfuric acid, taking advantage of the fact that the polymerization of silicic acid slows down under acidic conditions. However, the mechanisms of silica scale formation are extremely diverse, and have not yet been fully explained chemically. Since silica scale formation cannot be inhibited even with pH adjustment, it is necessary to understand the chemical and physical parameters that contribute to silica scale formation.

In this study, we aimed to clarify the contribution of parameters in geothermal power generation that contribute to silica scale formation, mainly using geochemical data.

2. DATASET

To create the machine learning model, we used geothermal water properties data from 12 locations, as well as the time-dependent weight and chemical analysis values of silica scale from test piece immersion experiments conducted at those locations. The test piece immersion experiment is the experiments that immersing a metal piece in geothermal water for a certain period of time (1 to 5 hours, 1 to 5 days), and the attached precipitation is analyzed the mineral concentrations in the using LA-ICP-MS.

2.1. EXPLANATORY VARIABLES

There are 42 explanatory variables (Table1). Ti-SUS is the kinds of metal pieces that used in test piece immersion experiment. Temp and Flow, pH-Field, EC parameters are chemical parameters in geothermal water. Na, K, Ca, Mg, Al-total, Cl, SO4, As, Fe-Total, Si-Total are concentrations of each mineral in geothermal water. SSI means SiO2 saturation index. Kinetic parameters are calculated in SiT (total silisic acid concentration), SiM (monosilisic acid concentration), AlR (reactive aluminum concentration) from polymerization experiment. Those parameters are obtained at each time (0, 5, 10, 20, 30, 45, 60 minutes). DLS-Number0 is the particle size in geothermal water. Slope parameters in each element are from test piece immersion experiment, which are slopes of the linear fitted curve of the concentration of precipitation on the test piece from 0 hours to 120 hours. Time – day parameter is the immersion date of the test piece immersion experiment used as the objective variable.

Proceedings 47th New Zealand Geothermal Workshop
11 -13 November 2025
Rotorua, New Zealand
ISSN 2703-4275

Table 1. the range of explanatory variable

Feature	Min	Max
Ti-SUS	0	1
Temp	66.6	97.4
Flow	0.2	4
time_day	0	28
pH-Field	4.8	8.6
EC	598	5610
Na	1084	11185
K	129	1439
Ca	6.4	2017
Mg	0.017	2.75
Al-Total	0.01	0.95
Cl	1668	22760
SO4	38.9	351.7
As	2.38	3.5
Fe-Total	0.01	0.23
Si-Total	486	867
SSI	1.495	2.729
SiTKinetic_0_5	0	76.16
SiTKinetic_05_10	0	8.343
SiTKinetic_10_20	0	4.172
SiTKinetic_20_30	0	17.221
SiTKinetic_30_45	0	2.852
SiTKinetic_45_60	0	9.77
SiMKinetic_0_5	88.741	180.986
SiMKinetic_05_10	0	15.146
SiMKinetic_10_20	0	1.882
SiMKinetic_20_30	0	0.417
SiMKinetic_30_45	0	1.039
SiMKinetic_45_60	0	1.158
AIR	0	1.091885
AIRKinetic_05_10	0	0.0145
AIRKinetic_10_20	0	0.01255
AIRKinetic_20_30	0	0.01339
AIRKinetic_30_45	0	0.02524
AIRKinetic_45_60	0	0.01344
Na-slope	0	641.094
Mg-slope	0	12.656
Al-slope	0	12.968
Si-slope	0	1814.5
K-slope	0	6.987
Ca-slope	0	823.262
DLS-Number0	0.6213	164.2

2.2. OBJECTIVE VARIABLE

The objective variable is the weight [mg] of the precipitation that attached to the test piece. Fig. 1 shows the distribution of output.

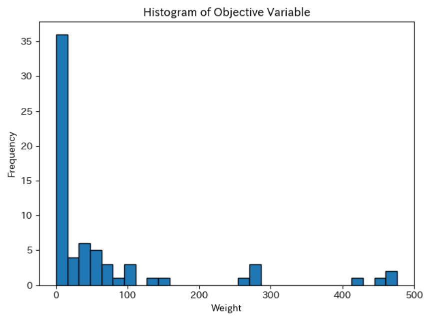

Fig1. The contribution of objective variable

3. APPROACHES

In our machine learning approach, we employed neural networks, with data handling performed via the torch and sklearn libraries in python.

3.1 FEATURE ENGINEERING

The explanatory variables pH-Field, Fe-Total, and Al - Total were each grouped into three categories: low, medium, and high, while Si-Total and Ca were grouped into two categories: low and high. After processing them as one-hot vectors, the other features were logarithmically processed and normalized by Maxmin function in sklearn. The number of features is 55.

3.2 LEARNING CONDIITON

Size of training and validation, test data are 37, 10, 8. The activation function except for the output layer is only applied relu function and the activation function for the output layer is linear. The loss function is mean square error. Optimizer is adam from torch. The range of the number of epochs is maximum 2000 until the minimum MSE value is not updated 500 times. The number of hidden layers is from 1 to 10. Fig 2 shows the example of validation.

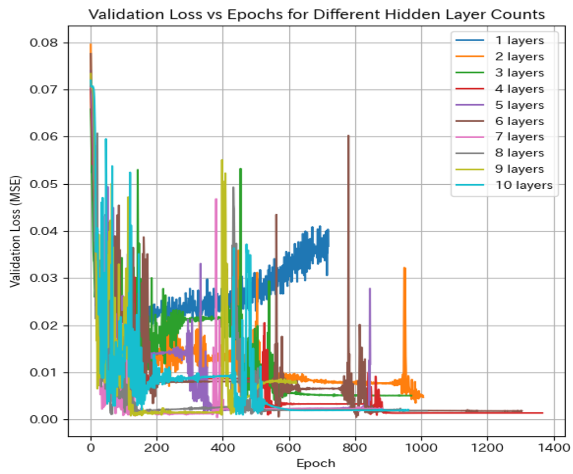

Fig2. Calculation result of machine learning for validation

3.3 FACTOR ANALYSIS METHOD

Factor analysis of explanatory variables for the dependent variable was performed using the Partial Derivative (PaD) method, especially Sum of Square Derivatives. (Gevrey et al., 2003). SSD is calculated as follows:

$$SSD_i = \sum_{j=1}^{N} \left(\frac{\partial y_j}{\partial x_i} \right)^2$$

x_i means the input variable and y_j (with j = 1, ..., N and N is the total number of output parameters) is the output variable. This value allows the classification of explanatory variable respect to objective variable. The one parameter in explanatory variable that has high SSD value is the variable, which influence objective variable. For intuitive understanding, we define the positive contribution and the negative contribution as follows:

$$positive\ contribution_i[\%] = \frac{SSD_i^+}{\sum_{j=1}^{M} SSD_j} * 100$$

$$negative\ contribution_i[\%] = \frac{SSD_i^-}{\sum_{j=1}^{M} SSD_j} * 100$$

where SSD_i^+ and SSD_i^- denote the sum of squared positive and negative partial derivatives, respectively, and M is the total number of input variables.

4. RESULTS

The learning result is shown in table 2. According to this, machine learning was implemented correctly. To verify the accuracy of the model, we created an ensemble model using test data that was not used for training and checked the R squared value (fig 3). The ensemble model used three models with hidden layers of 7, 2, and 1, and that MSE value is 2701. The R-squared value between the true and predicted values was 0.7273.

Table 2, The top 10 learning models

The number of layers	minimum validation loss
7	0.000447
6	0.000514
9	0.000808
4	0.000944
10	0.000991
8	0.001093
5	0.001508
2	0.003101
3	0.003146
1	0.019072

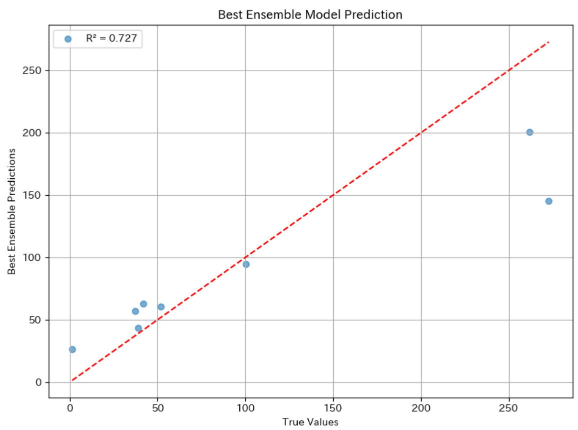

Fig 3. R squared value: Predicted in ensemble model vs true

According to this ensemble model, Fig 4 and Fig 5 shows the contribution to objective variable. Focusing on positive contribution (Fig 4), `time_day` parameter shows the highest contribution (46.0%). Furthermore, most of the positive contributions are from Kinetic values in SiM, SiT, and AlR. As for negative contribution (Fig 5), `AlRKinetic_45_60` and `Na_slope`, Na concentration negatively influenced to objective variable. (4.0%, 3.7%, 3.6%)

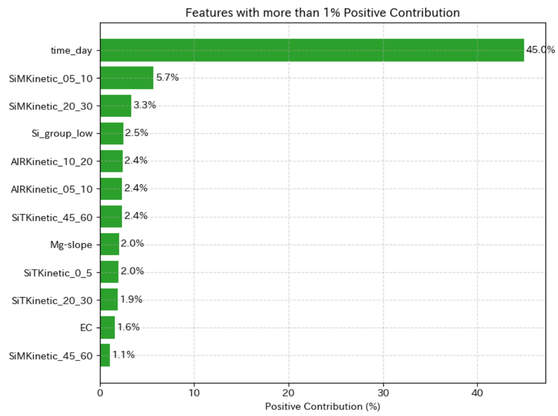

Fig 4. Positive contribution to objective variables

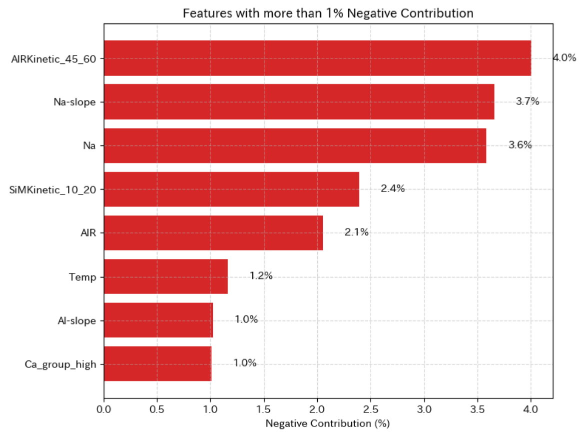

Fig 5. Negative contribution to objective variables

5. DISCUSSION

Since the contribution for `time_day` is positive and substantial, it's obvious that the weight of silica scale (our dependent variable) increases as time passes. This evaluation from the learning model is consistent with our experience. Considering the second highest positive contribution parameter `SiMKinetic_05_10`, showing that a greater initial change during the polymerization process results in a heavier silica scale.

The significant contributions that observed for SiM, SiT, and AlR kinetic suggest a correlation between variations in reactive aluminum concentration in the geothermal water and the formation of silica scale. This is consistent with reports that silica scale formation is significantly enhanced by trace amounts of alumina, aluminum (Iler, 1973; Yokoyama et al., 1980; Gallup, 1993). Those reports also suggest that ions, iron hydroxides, and ferrous and ferric ions promote silica scale formation, however this learning model does not evaluate it. That is because iron has several chemical states like Fe^{2+} and Fe^{+3} and the parameter of only Fe-total is not enough to catch the reactive with silica scale in iron.

6. CONCLUSION

The accuracy of machine learning in predicting the amount of silica scale formation was an R squared value of 0.73.

In this learning model, the positive contribution for the 'time_day' parameter is large, indicating that the amount of precipitation increases over time.

Not only silicic acid concentration and aluminum concentration, but also the Kinetic value, a parameter that indicates polymerization behavior, has a large impact on the output.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Associate Professor Kotaro Yonezu, Professor Emeritus Takushi Yokoyama, and Mr. Saefudin, who carefully guided me in the basics of research despite their busy schedules. I would also like to express my deepest gratitude to the teachers and students who provided me with appropriate guidance and advice in my daily life and seminars. I would also like to express my respect to everyone at Kyuden Sangyo Co., Ltd. who cooperated with me in the field. Finally, I would like to express my gratitude to everyone who has been involved in my research and would like to express my thanks in writing.

REFERENCES

Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecological Modelling, 160, 249–264.

Iler, R. K. (1973). Effect of adsorbed alumina on the solubility of amorphous silica in water. *Journal of Colloid and Interface Science*, *43*(2), 399–408.

Iler, R. K. (1973) Effect of absorbed alumina on the solubility of amorphous silica in water. Journal of Colloid Interface Science 43, 399-408.

Yokoyama, T., Nakazato, T. and Tarutani, T. (1980) Polymerization of silicic acid adsorbed on iron(Ill) hydroxide. Bulletin of the Chemical Society of Japan 53, 850-853.

Gallup, D. L. (1993) The influence of iron on the solubility of amorphous silica in hyper- saline geothermal brines. Proc. 1991 Symposium on Chemistry in High-Temperature Aqueous Solutions, Provo, UT, pp. F6fl-16.