Signal preprocessing of deep-sea LIBS spectra for identification of pelletized hydrothermal deposits using Artificial Neural Networks

Soichi Yoshino[a,*], Blair Thornton[b,a], Tomoko Takahashi[a], Yutaro Takaya[c,d,e,f], Tatsuo Nozaki[d,e,g,f]

[a] *Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*
[b] *Southampton Marine and Maritime Institute, University of Southampton Boldrewood Innovation Campus, Southampton, SO16 7QF, UK*
[c] *School of Creative Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan*
[d] *Japan Agency for Marine-Earth Science and Technology, 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061 Japan*
[e] *Frontier Research Center for Energy and Resources, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[f] *Ocean Resources Research Center for Next Generation, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan*
[g] *Department of Planetology, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo 657-8501, Japan*

Abstract

This study investigates methods to analyze Laser-induced breakdown spectroscopy (LIBS) signals generated from water immersed deep-sea hydrothermal deposits irradiated by a long pulse (>100 ns) that are analyzed using Artificial Neural Networks (ANNs). ANNs require large amounts of training data to be effective. For this reason, we propose methods to preprocess full-field spectral signals in to an appropriate form for ANNs artificially increase the amount of training data. The ANN was trained using a dataset of signals from immersed pelletized hydrothermal deposit samples that were preprocessed using the proposed method. The proposed method improved the accuracy of identification from 82.5 % to 90.1 % and significantly increased the speed of learning. The result shows that the ANN can be used to construct a generic method to identify

*Corresponding author
Email address: soichi@iis.u-tokyo.ac.jp (Soichi Yoshino)

hydrothermal deposits by long pulse underwater LIBS signals without the need for explicit peak detection.

Keywords: Laser-induced breakdown spectroscopy (LIBS), chemical analysis, Artificial Neural Networks (ANNs), signal processing

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) is a method for chemical analysis that can determine the elemental composition of targets by analyzing the optical emission from plasmas generated by a high-power laser pulse. LIBS has advantages that it allows for real-time measurement without any sample preparation, and this has led to it being widely used for in-situ geological investigation on land and in planetary exploration [1, 2]. For deep-sea exploration, long-pulse LIBS has the unique advantage that signals can be obtained from solid targets at high, oceanic pressures, and the technique has been applied to in-situ geochemical surveys of deep-sea mineral deposits [3, 4, 5, 6]. ChemiCam, a deep-sea long pulse LIBS instrument was developed by our group in 2013, and has been deployed on numerous occasions at depths of over 1000 m in active hydrothermal fields in the Okinawa Trough and the Izu Bonin Arc [7, 8]. While in-situ detection of peaks of elements contained in the deep-sea rocks has been successfully performed, methods to classify and quantify these signals are in need. Analytical methods to extract chemical information from signals generated underwater are being investigated using calibration curves, calibration-free LIBS (CF-LIBS), and multivariate analysis methods [9, 10, 11, 12, 13]. While calibration curves and CF-LIBS require matrix-matched samples with relatively simple compositions, their applications to in-situ analysis of complex targets such as rocks and sediments are limited. Multivariate analysis methods can be applied to targets with unknown and complex matrices by constructing a model using full spectral information of various samples in a training dataset. In particular linear regression methods such as Principal Components Regres-

sion (PCR) analysis and Partial Least Squares (PLS) regression analysis are commonly used for analyzing LIBS spectra since information linearly related to concentration changes can be extracted and isolated from spectral information using a training dataset. The incompressibility of water results in plasmas that are smaller and have a shorter duration than in air (Figure 1) [14, 15]. The signal quality is more susceptible to the surface condition of the sample, the power of the laser and temporal deviation of plasma dynamics. These lead fluctuations significantly to complicate the analysis of underwater LIBS signals. For the multivariate regression analysis of underwater spectra with large fluctuations, a temperature based segmentation method was proposed [13] to compensate for the nonlinear behavior. Meanwhile the method requires peaks to be detected in order to calculate the plasma temperature, which requires assumption to be made regarding the origin of the detected peaks. In this study, we investigate a generic analytical method that does not rely on assumptions or require any form of explicit peak recognition, using Artificial Neural Networks (ANNs) that can describe non-linear properties through supervised-learning. It is generally agreed that for ANNs, in order to obtain good results, it is necessary to train the models using a large amount of data with appropriate signal pre-processing techniques. Many methods have been proposed in image and language processing fields [16, 17, 18]. In the LIBS signal, using specific peak intensities is preferred instead of using the entire spectra because the peak frequencies are physically determined for each element [19]. Several reports have successfully applied ANNs to LIBS in air by choosing elemental peaks for using relatively small networks and datasets [20, 21, 22, 23]. However, for spectra generated in water, the elemental peaks are not well resolved because the spectra are broad and in some cases self-absorbed due to high plasma densities. Figure 2 shows the comparison of spectra of a deposit target taken in a) air and in b) water. Significant peak broadening is observed in the four peaks of Pb I at 357.3, 364.0, 368.4, and 374.0 nm. Therefore, specific peaks are more likely to interfere with neighboring signals. This study investigates full-field signal pre-processing methods that use the entire signal as an input, which enable ANNs to learn with

limited database size by considering the features of underwater LIBS spectra. The methods are verified through the identification of known pelletized deep-sea hydrothermal precipitates.

2. Materials and methods

2.1. Experimental setup

The experimental setup for this work is shown in Figure 3. A custom-built long-pulse Nd:YAG laser emits light with the fundamental wavelength of 1064 nm, pulse energy of 5 mJ, pulse duration of 150 ns, and a repetition rate of 2 Hz. The pulse is delivered via a 600 μm fused-silica fiber and generates a plasma on a target submerged in artificial seawater. The distance between the target surface and the face of a custom-made objective lens of 5 times magnification is 7 mm. The optical emissions of the laser-generated plasmas are observed through the same path used for laser delivery and recorded from 320 nm to 550 nm at a resolution of 0.25 nm using a Czerny-Turner spectrograph and Intensified Charged Coupled Device (ICCD, Princeton Instruments, PI-MAX 4 HQf) camera with gate width of 500 ns and gate delay of 800 ns. The wavelength was calibrated using a standard mercury lamp (Ocean Optics, HG-1).

2.2. Materials

A total of 42 hydrothermal deposits are used for analysis. These have been collected from deep-sea hydrothermal fields and volcanogenic massive sulfide deposits on land. All samples are crushed to make powder pellets to reduce the effects of rock inhomogeneity. The compositions of the targets used are shown in Table 1, where these compositions have been measured by Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) and ICP - Mass Spectrometry (ICP-MS). For each sample, 500 LIBS spectra were measured underwater and the measurement location was changed every 10 shots.

2.3. Artificial Neural Networks

The Multi Layer Perceptron (MLP), used for the analysis in this research, is a type of ANN that is composed of three kinds of layers whose basic structural unit is the neuron. A neuron (Figure 4 a)) has n inputs $(x_1, x_2, ..., x_n)$ where the output is,

$$z = f(\sum_{m=1}^{n} w_m x_m + b). \tag{1}$$

The weight, w_m, is a matrix that expresses the strength of connectivity to each input and b is the bias. f is an activation function, where sigmoid, hyperbolic tangent or rectified linear functions are commonly used. The sigmoid function (Figure 4 b)),

$$f(x) = \frac{1}{1 + \exp(-x)}, \tag{2}$$

was used in this work because of the range of value. MLP is typically composed of three types of layers, an input layer, several hidden layers, and an output layer (Figure 4 c)). The number of neurons in the input layer is equal to the size of input signal and the number of neurons in the output layer is equal to the number of samples for the identification. The number of layers and neurons for each hidden layer is adjusted to characterize the features of the database. In this study, 500, 100 and 50 neurons were used for each of the 3 hidden layers. The model was implement using TensorFlow^{TM} [24], and the weights and biases of this model were optimized using a back propagation algorithm. The evaluation of the model was conducted with a k-fold crossvaridation ($k = 10$). In this method, the database was divided into k subsets and the model was trained using $k - 1$ pieces. After training, the model was evaluated with one of the remaining independent subsets. The test data was shifted sub sampled, and the performance was evaluated based on the accuracy as follows:

$$\text{accuracy} = \frac{1}{k}\sum_{n=1}^{k} \frac{N^n_{correct}}{N^n_{all}} \tag{3}$$

where k is the number of crossvalidation, N_{all} is the related number of spectra used for validation, and $N_{correct}$ is number of spectra classified correctly.

2.4. Signal processing

In order to obtain good results with the analysis using ANNs, it is important to input the data after signal processing in an appropriate form, rather than inputting the original data. Various methods to achieve this have been proposed in the image and language processing fields. For LIBS signals, common approaches to signal preprocessing include background subtraction, normalization, and averaging. Through these processes, the shot to shot fluctuations of the intensity of the LIBS signal due to variations in laser output, temporal deviation, difference in distance to the sample and influence of the surface state of the sample are suppressed. However, simply processing signals in this way does not lead to the spectra being in an appropriate form as an input to ANNs. This work proposes two signal processing methods specifically tailored to the requirements for use in an ANN. The first preprocessing step facilitates learning of peaks where the important spectral information is contained. The original spectral intensity distribution is not suitable for the shape of the activation function of ANNs. The intensity is separated into frequencies near the baseline and frequencies including the peaks as the histogram of LIBS spectrum intensity (Figure 5 a)). When the input has values that are either very large or small, the output is saturated by the characteristics of the activation function (Figure 4 b)). Therefore, the sensitivity of the outputs to the information in the peaks, which are high intensity, is poor. In order to avoid this saturation, the natural logarithm is taken. As a result, intensity distribution width decreases (Figure 5 b)), and the information in the peaks has a larger influence on the output of the function, allowing the ANNs to learn information from the peaks. Moreover, since this approach doesn't rely on specific peaks, but instead considers the whole spectra as input, it is robust to the effects of self-absorption and broadening of specific peaks. If these affects are sufficiently repeatable, the two can potentially be recognized as features of the database.

The second method is data augmentation. It is widely known that a large dataset is required for ANNs training, but measuring large amounts of LIBS signals by experiments, especially underwater, is time consuming. There are also

restrictions of ship time during deep-sea exploration, which limits the amount of data that can be collected for analysis in field applications. Therefore, a method of enlarging a relatively small database is investigated. In the field of image processing, a method of increasing the data size is widely used by shifting, turning, zooming out and in. Generally, averaging is performed in spectrum preprocessing, but when doing this, the number of data decreases. Therefore, in this work we perform random sampling using the bootstrapping method, which is a common Monte-Carlo technique [25]. The method generates resampled sets by selecting random signals recursively to a desired number, allowing the value of features of the original dataset to be evaluated in cases where the data available is finite. This enables ANNs to recognize statistically reliable patterns of the database. In this method, there are two parameters, the number of spectra selected for averaging and the number of times the dataset is resampled. In this work, 5 shots were selected for each subsampling and the number of times for resampling was investigated.

3. Results and discussion

3.1. Identification of pelletized hydrothermal deposits measured in water

In order to verify the effectiveness of the proposed signal preprocessing method, three models were trained by the datasets using different methods. The first model performs background subtraction, normalization and averaging by 5 shots in conventional signal processing of LIBS spectra. The second converts signals to a suitable form for ANNs by taking a logarithm of the averaged signals and normalizing the signals to have a variance of 1, shifted to have a zero average. The third was a model in which the size of the database was enlarged to 5000 shots per sample from the 500 original spectra by the bootstrapping method with combining 5 spectra in addition to the second preprocessing method. Table 2 shows the identification result of 42 hydrothermal deposits. The accuracy of the first conventional model is 82.5 $\pm$ 0.7 %, the second model that uses the natural logarithm to change the dataset character-

istics is 85.5 ± 0.8 % and the third model trained with an extended dataset is 90.1 ± 0.4 %. The identification accuracy of hydrothermal deposits was increased by performing the proposed signal processing method. The confusion matrix (Figures 6 a), b), and c)) illustrates the result of identification, where white indicates a high probability. The confusion matrix of the proposed model significantly improves upon the conventional preprocessing model. Figures 6 d), e), and f) are the classification probability of the predicted sample for each set of 5 shots. The red arrow shows the correct sample line. While the conventional model misjudged the origin of several signals as the wrong sample, the model which changed database characteristics improved the performance, though still there are dark lines which means the model confused several hydrothermal deposits of similar composition as Figure 6 e) shows. After data augmentation, most of the dark lines are eliminated (Figure 6 f)). From these results, it can been seen that the proposed signal processing for the input spectra improves the identification accuracy of ANNs for underwater LIBS applications.

3.2. Investigation of database characteristics change

The influence was examined of the learning speed of the ANN by taking logarithm and changing variance, which was done as a process to change the characteristics of spectrum. Figure 7 is a graph of the identification accuracy in each step of the model trained by the database of conventional method and the database of the proposed preprocessing method. As this figure shows, the model of the proposed method was faster in learning than the conventional method, where learning was completed in 20000 steps, whereas the model of conventional method required 40000 steps to converge. The fact that learning progresses quickly is an indication that the features of the signal are recognized and more easily learned as an input of ANNs by performing the proposed preprocessing.

3.3. Investigation of database extension rate

The database extension rate, which is the number for generating larger databases by resampling signals using the bootstrapping method, was examined. Figure 8 shows the transition of identification accuracy of a model with

an averaged dataset by 5 shots, and models in which the sizes of the databases were increased by the bootstrapping method with three different extension rate ($\times 1, \times 10, \times 100$), which means how large from original database size, with the number of spectra of each sample is 50, 100, and 500 shots. The first thing we can see from this graph is that the accuracy gets higher when the model learns in a large database, and the accuracy learned by the database of 500 shots was better than 50 shots. For any data size tested in this range, if the data amount is extended, the accuracy got higher than just taking the average. When examining how many times the amount of data should be increased by the bootstrapping method, if the size was increased to about ten times of the original number of shots, the accuracy was improved, but even if it was made larger (100 times), there was no significant improvement in accuracy. The reason is thought to be that spectra with new features were not generated once the extension rate reached a certain value because no unique features were being captured by the recursively selected spectra.

4. Conclusion

This study proposed a signal preprocessing method for LIBS spectra analysis using ANNs. The method takes the full spectra as input, which is preprocessed by taking the logarithm to make its shape suitable for input to an ANNs, and enlarging the database using a Monte Carlo approach to analyze dataset of limited size. The effectiveness of the method was confirmed by the identification of pelletized hydrothermal deposits using underwater LIBS. The proposed method improved the accuracy of equivalent ANNs without preprocessing from 82.5 % to 90.1 % and increases the speed of training. Through the comparison of the rate of data size extension, an increase in data by an order of magnitude with respect to the original number of shots was found to give a noticable improvement in performance for all conditions tested. From this study, it was found that full-field spectral signals can be used for ANNs by transforming the signals to a suitable form for model construction without manual selection of peaks. The

whole process in this study, from signal preprocessing to ANN analysis, can be fully automated end to end, which is applicable to in-situ chemical analysis of deep-sea hydrothermal deposits.

5. Acknowledgment

This research was supposed by the Japanese Ministry of Education, Culture, Sports, Science and Technology under the 'Program for the development of fundamental tools for the utilization of marine resources'.

References

[1] R. Wiens, S. Maurice, J. Lasue, O. Forni, R. Anderson, S. Clegg, S. Bender, D. Blaney, B. Barraclough, A. Cousin, et al., Pre-flight calibration and initial data processing for the chemcam laser-induced breakdown spectroscopy instrument on the mars science laboratory rover, Spectrochimica Acta Part B: Atomic Spectroscopy 82 (2013) 1–27.

[2] P.-Y. Meslin, O. Gasnault, O. Forni, S. Schröder, A. Cousin, G. Berger, S. Clegg, J. Lasue, S. Maurice, V. Sautter, et al., Soil diversity and hydration as observed by chemcam at gale crater, mars, Science 341 (6153) (2013) 1238670.

[3] T. Sakka, H. Oguchi, S. Masai, K. Hirata, Y. H. Ogata, M. Saeki, H. Ohba, Use of a long-duration ns pulse for efficient emission of spectral lines from the laser ablation plume in water, Applied physics letters 88 (6) (2006) 061120.

[4] T. Sakka, A. Tamura, A. Matsumoto, K. Fukami, N. Nishi, B. Thornton, Effects of pulse width on nascent laser-induced bubbles for underwater laser-induced breakdown spectroscopy, Spectrochimica Acta Part B: Atomic Spectroscopy 97 (2014) 94–98.

[5] B. Thornton, T. Sakka, T. Takahashi, A. Tamura, T. Masamura, A. Matsumoto, Spectroscopic measurements of solids immersed in water at high pressure using a long-duration nanosecond laser pulse, Applied Physics Express 6 (8) (2013) 082401.

[6] B. Thornton, T. Sakka, T. Masamura, A. Tamura, T. Takahashi, A. Matsumoto, Long-duration nano-second single pulse lasers for observation of spectra from bulk liquids at high hydrostatic pressures, Spectrochimica Acta Part B: Atomic Spectroscopy 97 (2014) 7–12.

[7] B. Thornton, T. Masamura, T. Takahashi, T. Ura, K. Ohki, T. Sakka, Development and field testing of laser-induced breakdown spectroscopy for in situ multi-element analysis at sea, in: Oceans, 2012, IEEE, 2012, pp. 1–6.

[8] B. Thornton, T. Takahashi, T. Sato, T. Sakka, A. Tamura, A. Matsumoto, T. Nozaki, T. Ohki, K. Ohki, Development of a deep-sea laser-induced breakdown spectrometer for in situ multi-element chemical analysis, Deep Sea Research Part I: Oceanographic Research Papers 95 (2015) 20–36.

[9] V. Lazic, F. Colao, R. Fantoni, V. Spizzicchino, Laser-induced breakdown spectroscopy in water: improvement of the detection threshold by signal processing, Spectrochimica Acta Part B: Atomic Spectroscopy 60 (7) (2005) 1002–1013.

[10] T. Masamura, B. Thornton, T. Ura, Spectroscopy and imaging of laser-induced plasmas for chemical analysis of bulk aqueous solutions at high pressures, in: OCEANS 2011, IEEE, 2011, pp. 1–6.

[11] A. Matsumoto, A. Tamura, R. Koda, K. Fukami, Y. H. Ogata, N. Nishi, B. Thornton, T. Sakka, On-site quantitative elemental analysis of metal ions in aqueous solutions by underwater laser-induced breakdown spectroscopy combined with electrodeposition under controlled potential, Analytical chemistry 87 (3) (2015) 1655–1661.

[12] T. Takahashi, B. Thornton, K. Ohki, T. Sakka, Calibration-free analysis of immersed brass alloys using long-ns-duration pulse laser-induced breakdown spectroscopy with and without correction for nonstoichiometric ablation, Spectrochimica Acta Part B: Atomic Spectroscopy 111 (2015) 8–14.

[13] T. Takahashi, B. Thornton, T. Sato, T. Ohki, K. Ohki, T. Sakka, Temperature based segmentation for spectral data of laser-induced plasmas for quantitative compositional analysis of brass alloys submerged in water, Spectrochimica Acta Part B: Atomic Spectroscopy 124 (2016) 87–93.

[14] A. P. Michel, M. Lawrence-Snyder, S. M. Angel, A. D. Chave, Laser-induced breakdown spectroscopy of bulk aqueous solutions at oceanic pressures: evaluation of key measurement parameters, Applied optics 46 (13) (2007) 2507–2515.

[15] B. Thornton, T. Ura, Effects of pressure on the optical emissions observed from solids immersed in water using a single pulse laser, Applied physics express 4 (2) (2011) 022702.

[16] A. K. Jain, J. Mao, K. M. Mohiuddin, Artificial neural networks: A tutorial, Computer 29 (3) (1996) 31–44.

[17] G. B. Orr, K.-R. Müller, Neural networks: tricks of the trade, Springer, 2003.

[18] H. B. Demuth, M. H. Beale, O. De Jess, M. T. Hagan, Neural network design, Martin Hagan, 2014.

[19] J.-B. Sirven, B. Bousquet, L. Canioni, L. Sarger, Laser-induced breakdown spectroscopy of composite samples: comparison of advanced chemometrics methods, Analytical chemistry 78 (5) (2006) 1462–1469.

[20] J. El Haddad, M. Villot-Kadri, A. Ismael, G. Gallou, K. Michel, D. Bruyère, V. Laperche, L. Canioni, B. Bousquet, Artificial neural network for on-site quantitative analysis of soils using laser induced breakdown spectroscopy, Spectrochimica Acta Part B: Atomic Spectroscopy 79 (2013) 51–57.

[21] V. Motto-Ros, A. S. Koujelev, G. R. Osinski, A. E. Dudelzak, Quantitative multi-elemental laser-induced breakdown spectroscopy using artificial neural networks, Journal of the European Optical Society-Rapid Publications 3.

[22] P. Inakollu, T. Philip, A. K. Rai, F.-Y. Yueh, J. P. Singh, A comparative study of laser induced breakdown spectroscopy analysis for element concentrations in aluminum alloy using artificial neural networks and calibration methods, Spectrochimica Acta Part B: Atomic Spectroscopy 64 (1) (2009) 99–104.

[23] E. DAndrea, S. Pagnotta, E. Grifoni, G. Lorenzetti, S. Legnaioli, V. Palleschi, B. Lazzerini, An artificial neural network approach to laser-induced breakdown spectroscopy quantitative analysis, Spectrochimica Acta Part B: Atomic Spectroscopy 99 (2014) 52–58.

[24] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems, arXiv preprint arXiv:1603.04467.

[25] B. Efron, Bootstrap methods: another look at the jackknife, Breakthroughs in statistics (1992) 569–593.

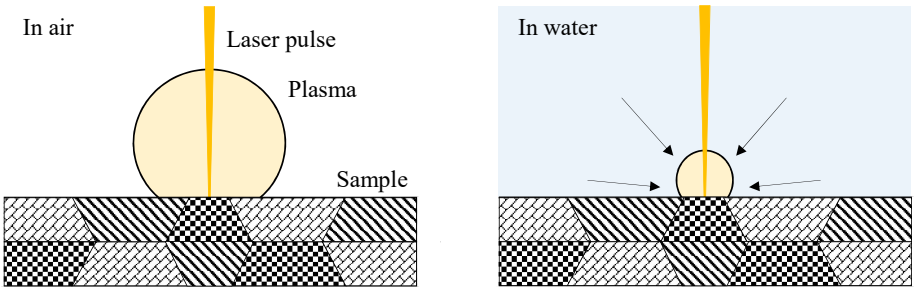

Figure 1: Conceptual diagrams of the plasma size difference between in air and in water

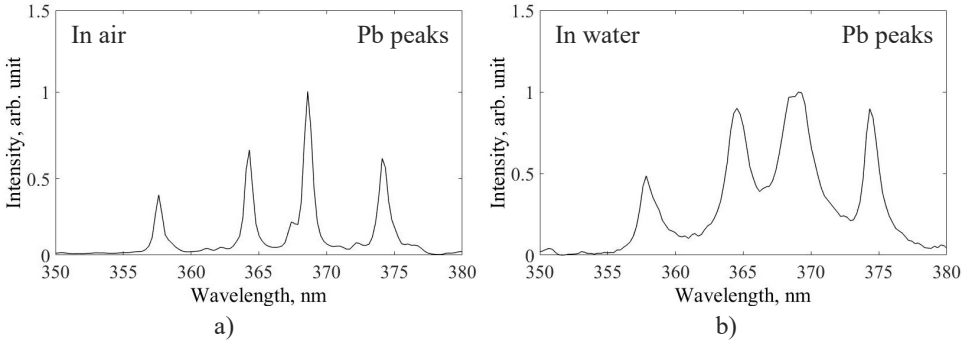

Figure 2: Exemplary LIBS spectra of a hydrothermal precipitate target measured a) in air, and b) in water

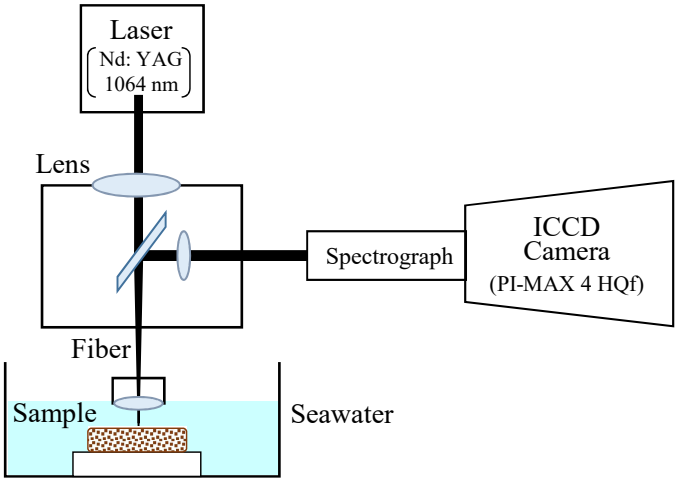

Figure 3: Conceptual diagram of experimental setup

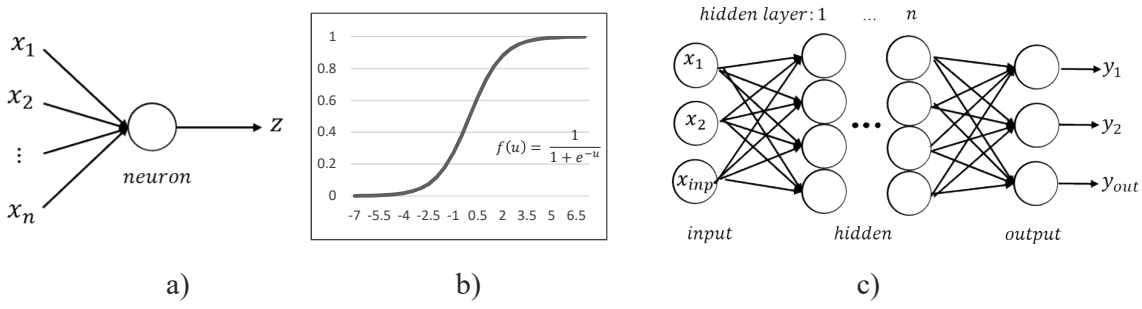

Figure 4: Data processing protocol by a) artificial neuron, b) activation function, and c) neural network

Table 1: Major element compositions (in wt. %) of samples

No.	Location			Compositions of major elements (in wt. %)								
	Site name	Area	Depth (m)	Al	Ba	Cu	Fe	Mg	Mn	Pb	Zn	others
1	Iheya North Knoll	Okinawa Trough, Japan	971	0.18	0.01	4.63	12.47	0.04	0.31	4.99	42.60	34.78
2	Iheya North Knoll	Okinawa Trough, Japan	971	0.68	0.01	3.08	19.49	0.05	0.43	15.10	28.50	32.67
3	Iheya North Knoll	Okinawa Trough, Japan	982	0.78	0.23	5.11	9.44	0.04	0.09	3.79	28.20	52.32
4	Hatoma Knoll	Okinawa Trough, Japan	1447	0.65	0.01	3.69	10.38	0.10	0.13	4.36	30.90	49.79
5	Mariner Site	Valu Fa Ridge	1472	0.01	12.89	1.85	1.69	0.02	0.00	0.02	22.90	60.61
6	Pika Site	Southern Mariana Trough	2787	0.00	0.00	0.01	46.18	0.00	0.01	0.00	0.03	53.76
7	Kairei Field	Indian Ocean	2455	0.05	0.00	35.90	30.91	0.00	0.01	0.00	0.33	32.80
8	-	Izu-Bonin Arc	802	0.15	0.04	0.01	0.10	3.35	0.16	2.12	2.98	91.09
9	Izena Hole	Okinawa Trough, Japan	1600	0.02	16.88	0.15	8.63	0.02	0.02	2.22	5.21	66.85
10	Izena Hole	Okinawa Trough, Japan	1624	0.84	14.96	0.03	0.51	0.04	0.01	0.00	0.09	83.52
11	Hatoma Knoll	Okinawa Trough, Japan	1481	0.30	0.09	6.12	6.81	0.03	0.30	24.10	35.10	27.16
12	Yoron Hole	Okinawa Trough, Japan	591	0.20	17.10	0.22	0.69	0.02	0.00	0.78	2.25	78.75
13	Iheya North Knoll	Okinawa Trough, Japan	995	0.16	9.01	2.63	6.27	0.03	0.08	1.61	21.90	58.32
14	Iheya North Knoll	Okinawa Trough, Japan	998	0.18	0.10	3.65	11.23	0.05	0.45	17.90	30.70	35.74
15	Iheya North Knoll	Okinawa Trough, Japan	1070	0.40	0.13	4.66	12.32	0.02	0.29	12.70	37.80	31.68
16	Iheya North Knoll	Okinawa Trough, Japan	1079	0.14	0.02	3.60	11.45	0.02	0.16	8.79	44.30	31.51
17	Hatoma Knoll	Okinawa Trough, Japan	1530	0.71	25.78	1.43	1.13	0.12	0.04	15.40	19.30	36.09
18	Hatoma Knoll	Okinawa Trough, Japan	1531	5.85	10.55	0.00	2.71	0.42	0.08	5.24	0.00	75.16
19	Iheya North Knoll	Okinawa Trough, Japan	991	0.70	0.04	3.29	13.11	0.03	0.27	12.70	32.00	37.86
20	Iheya North Knoll	Okinawa Trough, Japan	991	0.73	0.12	2.13	4.82	0.08	0.06	5.77	24.80	61.49
21	Iheya North Knoll	Okinawa Trough, Japan	991	1.86	0.00	1.77	13.63	0.08	0.08	2.61	11.70	68.26
22	Iheya North Knoll	Okinawa Trough, Japan	1003	0.26	0.01	6.34	17.09	0.05	0.06	0.65	35.72	39.82
23	Iheya North Knoll	Okinawa Trough, Japan	996	0.65	0.63	1.27	3.06	0.06	0.07	8.20	12.96	73.10
24	Iheya North Knoll	Okinawa Trough, Japan	993	0.28	0.12	3.37	17.33	0.04	0.20	5.98	24.59	48.09
25	Iheya North Knoll	Okinawa Trough, Japan	1016	0.10	0.01	4.57	9.42	0.01	0.05	11.79	34.07	39.98
26	Iheya North Knoll	Okinawa Trough, Japan	999	1.72	0.03	4.00	9.38	0.09	0.08	1.40	17.89	65.38
27	Iheya North Knoll	Okinawa Trough, Japan	993	0.15	0.04	2.35	3.60	0.02	0.08	1.62	16.49	75.66
28	Iheya North Knoll	Okinawa Trough, Japan	993	0.00	0.06	4.30	10.20	0.01	0.04	14.86	33.28	37.25
29	Iheya North Knoll	Okinawa Trough, Japan	1001	1.28	0.33	1.62	2.21	0.07	0.07	3.80	5.71	84.91
30	Iheya North Knoll	Okinawa Trough, Japan	1013	0.81	0.02	1.72	5.38	0.07	0.11	0.94	4.39	86.55
31	Iheya North Knoll	Okinawa Trough, Japan	1013	0.86	0.02	0.31	4.94	0.06	0.04	1.08	1.07	91.62
32	Iheya North Knoll	Okinawa Trough, Japan	1015	0.38	0.31	3.42	11.40	0.02	0.10	4.41	4.47	75.50
33	Iheya North Knoll	Okinawa Trough, Japan	1018	0.02	0.00	8.02	21.69	0.01	0.15	1.50	27.99	40.61
34	Iheya North Knoll	Okinawa Trough, Japan	992	0.20	0.22	2.09	7.29	0.02	0.29	11.40	20.43	58.06
35	Iheya North Knoll	Okinawa Trough, Japan	992	0.25	0.09	2.29	12.77	0.03	0.45	10.64	21.72	51.74
36	Iheya North Knoll	Okinawa Trough, Japan	1015	0.00	0.00	3.86	6.33	0.01	0.10	7.25	51.20	31.25
37	Iheya North Knoll	Okinawa Trough, Japan	1015	0.23	0.05	3.05	10.81	0.02	0.22	10.32	23.59	51.71
38	Iheya North Knoll	Okinawa Trough, Japan	1015	0.00	0.03	4.78	11.92	0.00	0.23	10.28	41.89	30.86
39	Hatoma Knoll	Okinawa Trough, Japan	1474	0.10	0.10	0.05	0.08	2.27	0.27	1.14	1.24	94.76
40	Hatoma Knoll	Okinawa Trough, Japan	1472	0.60	0.36	0.19	0.92	0.04	1.33	26.50	43.40	26.66
41	Hatoma Knoll	Okinawa Trough, Japan	1472	0.02	1.26	1.55	3.57	0.01	0.24	26.20	37.50	29.65
42	Kosaka-Motoyama	Akita prefecture, Japan	-	0.18	13.39	3.05	0.66	0.00	0.01	10.50	26.40	45.81

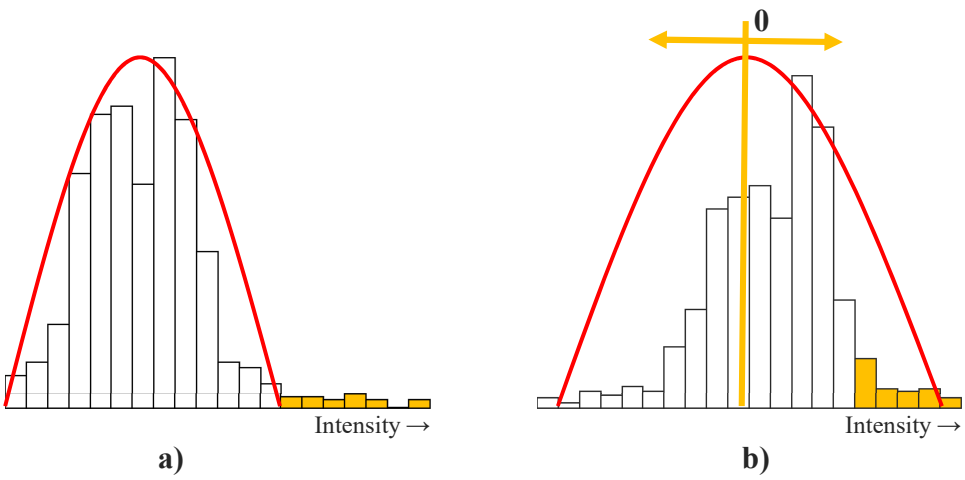

Figure 5: The intensity histogram of a) original spectrum, and b) spectrum taken logarithm for ANNs

Table 2: Summary of identification results with conventional, logarithm, and extension datasets

	Conventional	**Logarithm**	**extension**
Accuracy [%]	82.5 ± 0.7	85.5 ± 0.8	90.1 ± 0.4

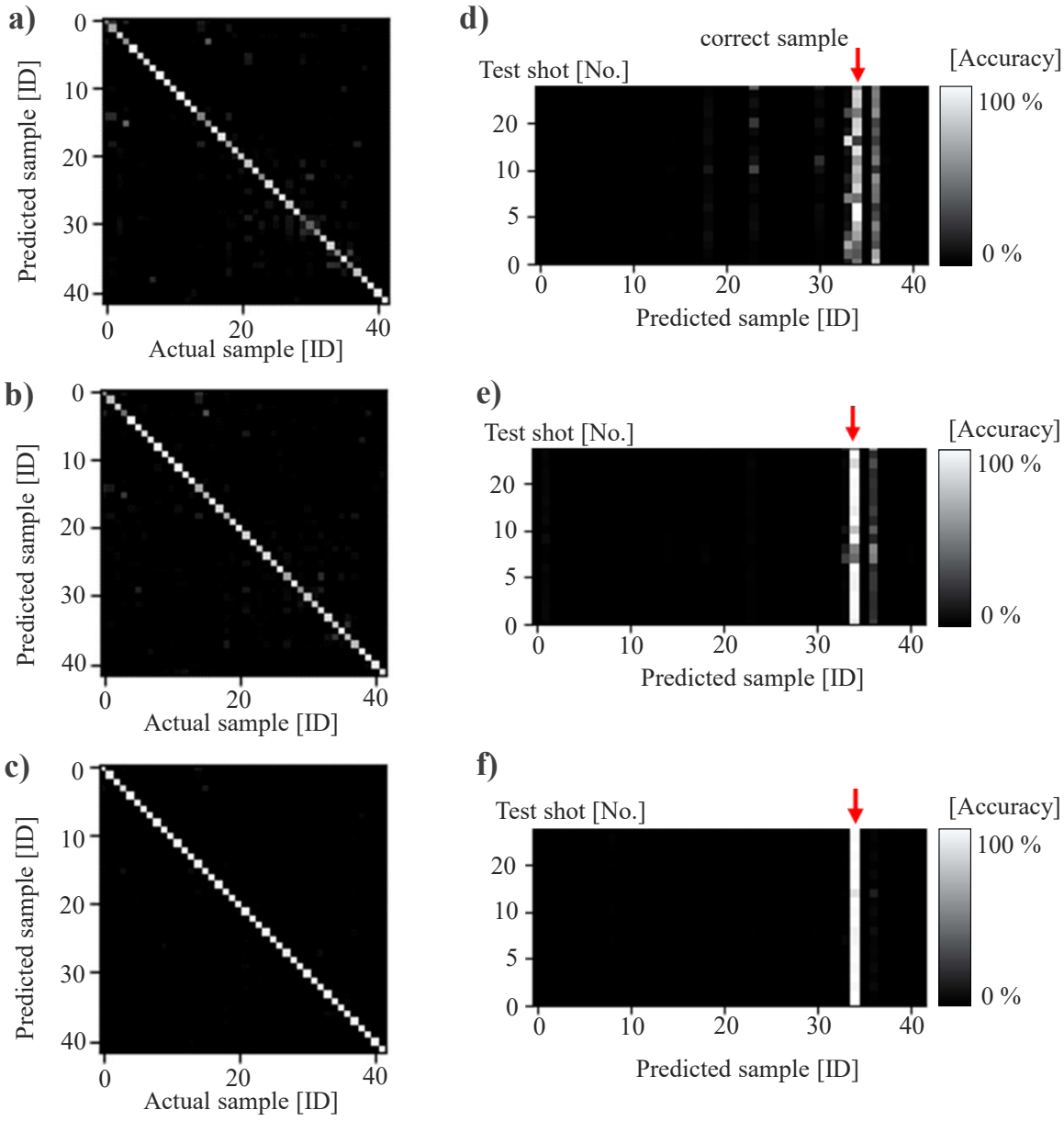

Figure 6: Confusion matrix of a) conventional model, b) logarithm model, and c) extension model and identification probability of sample No. 33 calculated by d) conventional model, e) logarithm model, and f) extension model

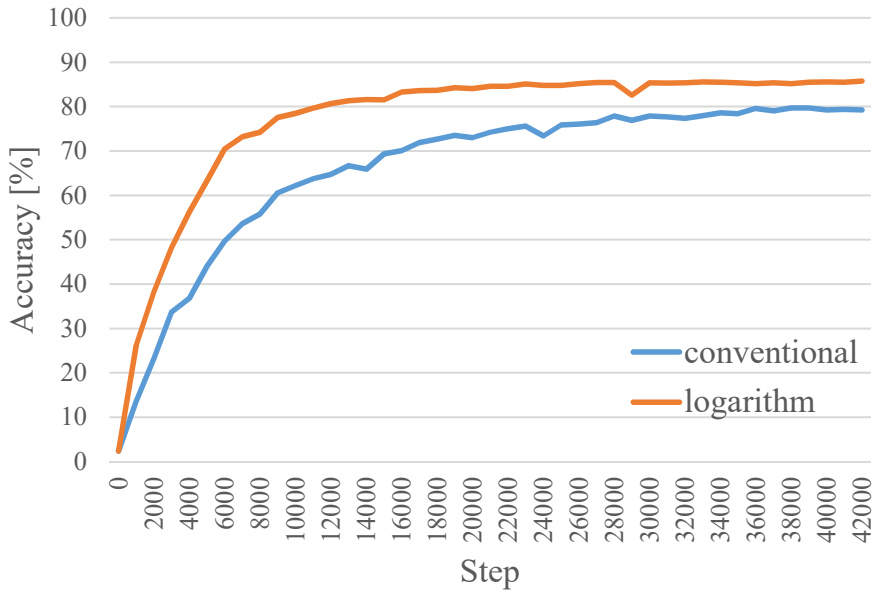

Figure 7: Step by step accuracy of conventional model and logarithm model

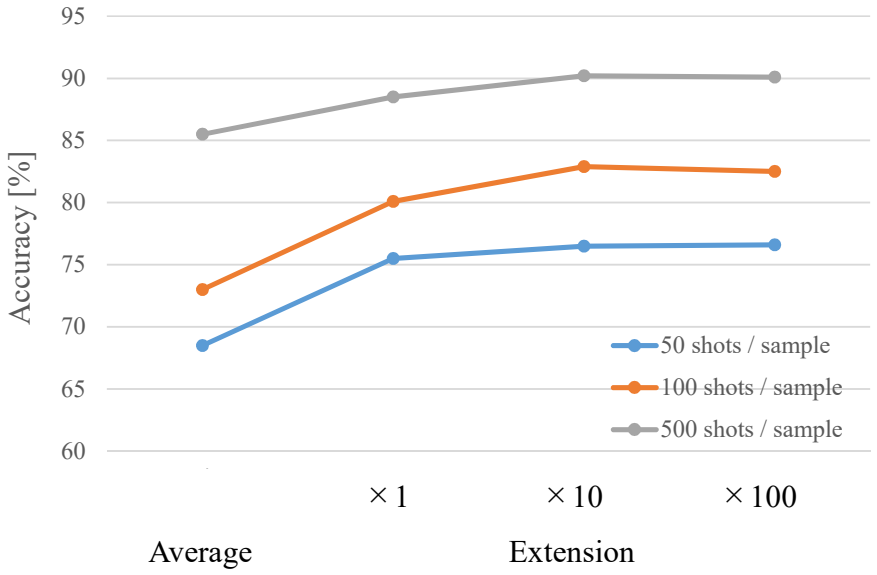

Figure 8: The accuracy plot of the averaged, and extended model (extension rate = $\times 1, \times 10$, and $\times$ 100) of three different size of datasets (original data size = 50, 100, and 500 shots per sample)