The Seventh Asian-Pacific Symposium on Structural Reliability and Its Applications (APSSRA2020)
October 4–7 2020, Tokyo, Japan
T. Takada, I. Yoshida & T. Itoi (editors)

Automatic Selection of Methane Hydrate Soil Samples based on Statistical Data Analysis

T. Shuku[1], T. Mekata[2] and K. Kageie[3]

[1]*Okayama University. Email: shuku@cc.okayama-u.ac.jp*
[2]*Okayama University. Email: p5g93teq@s.okayama-u.ac.jp*
[3]*Okayama University. Email: pms67xgq@s.okayama-u.ac.jp*

Abstract: This paper presents a methodology to select methane-hydrate (MH) bearing soil samples for laboratory testing based on statistical data analysis. The shortage of skilled/experienced engineers and their aging has become a serious problem in Japanese construction industry, and the "Machine" and "Artificial Intelligence" are expected to be promising work force instead of humans in civil engineering practice. This study is the first step toward achieving the goal and develops a method of automatic selection of MH bearing soil samples based on basic statistical analysis for limited data on MH samples. We compared the selected MH soil samples by the proposed method with those by experienced researchers/engineers to demonstrate its performance.

Keywords: statistical data analysis, laboratory testing, methane hydrate soil.

1. Introduction

Methane-hydrate (MH) is a massive potential source of energy, and many researchers have studies mechanical characteristics of MH bearing soils for sustainable gas production methodologies. For example, Yoneda et al. (2017) sampled pressure cores of MH-bearing-sands from the eastern Nankai Trough site and performed several types of tests (undrained/drained confined compression tests, uniaxial compression tests, isotropic loading and unloading-tests, and permeability tests) to investigate and characterize the intact strength, compressibility, and permeability of the gas hydrate reservoir.

In general, laboratory testing is performed only on limited number of soil specimens because of time and budget limitations, e.g., in Yoneda et al. (2017), the mechanical tests were performed on 11 soil specimens (10 cm each) that are selected from the pressure cores of 18.79 m in total. Although this selection is essential to get high-quality and representative data of MH sediments, they are selected by researchers/engineers based on auxiliary data (bulk sediment density, P wave velocity, and X-way CT image) and their experience. The selection criteria are qualitative and can change depend on researchers/engineers and should be automated by machine for more efficient testing.

The purpose of this study is to develop a method to automatically select the MH specimen based on basic statistical analysis. Although the selection problem can be formulated as several ways (e.g., classification, clustering, regression), this study formulates the task as an optimization problem. We compare the selected samples by the proposed method with those by experienced researchers/engineers to investigate the performance of the proposed method.

2. Methods

This study formulates the sample selection problem as an optimization problem. In optimization problems, we need to 1) define objective function specialized for the target problem, and 2) minimizing the objective function. The solution, i.e., the "best" specimen for laboratory testing, can be obtained as the result of minimization. This section outlines how to define objective function and minimize it.

2.1 Objective Function

we interviewed experienced researchers who have studied mechanical characteristics of MH bearing sediments through laboratory testing about how they select soil samples for specimens within collected pressure cores. In the selection, some auxiliary data are available, those include: 1) bulk sediment density (ρ_b (g/cm^3)) measured by a gamma-ray density-meter, 2) P wave velocity (V_p (m/sec)), and 3) X-ray CT image of the cores. Figure 1 shows an example of available data.

The researchers consider these data and carefully select soil samples for laboratory tests based on the following (empirical) criteria:

1) The samples should be homogeneous. There is no (or few) cracks or unconformity in them,
2) MH bearing soil samples, and
3) Samples should have diverse hydrate saturation (S_h (%)).

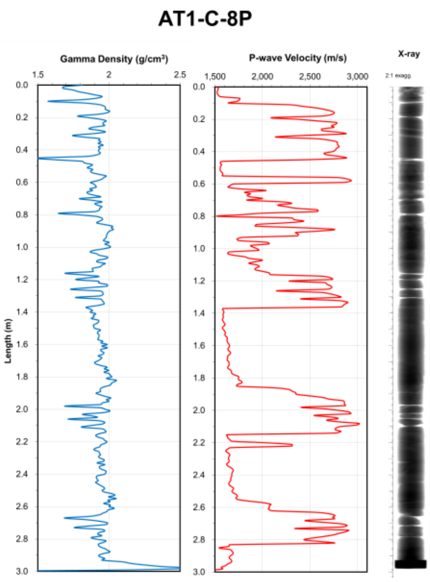

Figure 1 An example of available data on pressure core.

The corresponding mathematical expressions of 1) and 2) are defined as:

$$\sigma_d^2 = \frac{1}{n_s}\sum_{i=1}^{n_s}(d(z_i)-\mu)^2 \tag{1}$$

$$\mu_{\rho_b} = \frac{1}{n_s}\sum_{i=1}^{n_s}\rho_b(z_i) \tag{2}$$

where d_i is data (density or CT value or P wave velocity), z_i is ith depth, σ_d^2 is variance of the data d, n_s is the number of data point for one specimen which is 11 in this study, μ is mean value of the 11 data. In general, MH is contained in sands, and density of sands is usually higher than that of clays. Therefore, we can assume the high-density soil is classified as sand.

Regarding third criterion, hydrate saturation of the specimen is unknown at the stage of sample selection. We estimate S_h based on P wave velocity data. Although there are some constitutive models for expressing the relationship between S_h and V_p (e.g., Waite et al. 2009), they are difficult to use in sample selection practice because they need many model parameters for estimation, and they need to be calibrated. To simply estimate the methane hydrate saturation S_h, we newly develop a regression model based on the scatter plot of V_p and S_h shown in Figure 2. The data usually includes several types of noises due to sampling error, testing error, natural variability, device noise etc, and it may not be a good idea to use deterministic model to estimate S_h from V_p because the model includes uncertainty. We assume that the data follows Gaussian distribution and applied least square method (LSM) to estimate the linear regression model for the data. The estimated regression equation is given by:

$$S_h(\%) = 0.0448V_p - 70.346 + \varepsilon,\quad \varepsilon \sim N(0,\sigma^2) \tag{3}$$

where ε is the Gaussian noise with zero mean and variance σ^2 (σ = 18.357). This is a statistical model for the relation between S_h and V_p and gives estimation error of S_h values.

Finally, we define the following three objective functions to judge the quality of the sample:

$$\min\ J_1 = \sigma_{\rho_b}^2 = \frac{1}{n_s}\sum_{i=1}^{n_s}(\rho_{bi}-\mu_{\rho_b})^2 \tag{4}$$

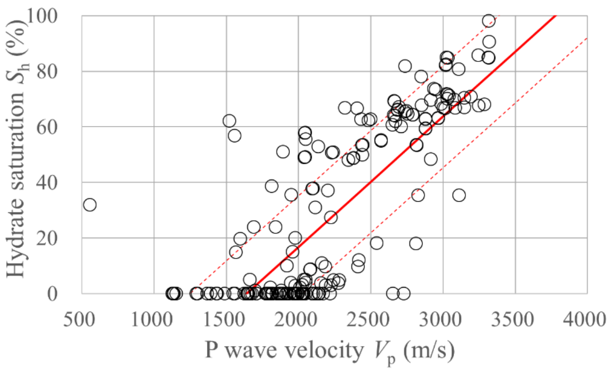

Figure 2. Regression model of S_h.

$$\min\ J_2 = \sigma_{CT}^2 = \frac{1}{n_s}\sum_{i=1}^{n_s}(CT_i-\mu_{CT})^2 \tag{5}$$

$$\max\ J_3 = \mu_{\rho_b} = \frac{1}{n_s}\sum_{i=1}^{n_s}\rho_{b\,i} \tag{6}$$

2.2 Selection Algorithm

Based on the objective functions, we proposed sample selection procedure as follows:

1) Set conditions such as specimen height h_s, target range of S_h ($S_h^- < S_h < S_h^+$).
2) Estimate all the possible samples using Eq. (3) and pick up the samples cs_j within the target range.
3) Compute Eqs. (4) – (5) for picked up cs_j. and make three rankings (or lists) for three objective functions.
4) Assign score S_j^k to cs_j depending on the ranking. The score is defined as:

$$S_j^1 = \frac{\sigma_{\rho b}^{\text{first rank}}}{\sigma_{\rho b}^{j}},\quad S_j^2 = \frac{\sigma_{CT}^{\text{first rank}}}{\sigma_{CT}^{j}},\quad S_j^3 = \frac{\mu_{\rho_b}^{j}}{\mu_{\rho_b}^{\text{first rank}}} \tag{7}$$

5) Compute total score TS, and sc_j that gets the highest score is defined as “best specimen” for laboratory testing.

The steps from 1) to 5) might be simply expressed by the following optimization problem:

$$\max \mathrm{TS}_j = \alpha \mathrm{S}_j^1 + \beta \mathrm{S}_j^2 + \gamma \mathrm{S}_j^3 \quad \text{s.t.}\quad S_h^- \le S_h < S_h^+ \tag{8}$$

where α, β, and γ are the weights for controlling intensity (importance) for first, second and third terms. S_h^- and S_h^+ mean the lower and upper search bound of S_h.

3. Case Studies

To demonstrate the performance of the proposed method in sample selection task, we compare the selected MH soil samples by the proposed method with those by Yoneda et al. (2017).

3.1 Setup

As noted, eight cores (AT1-C-6P, AT1-C-8P, AT1-C-10P, AT1-C-12P, AT1-C-13P, AT1-C-14P, AT1-C-18P, and AT1-C-20P) are available. Before implementing the selection algorithm, we standardized all the data using the following equations:

$$\hat{d}(z_i) = \frac{d(z_i)-\mu}{\sigma} \tag{9}$$

where, $\hat{d}(z_i)$ is standardized data, μ and σ are mean value and standard deviation of three auxiliary data.

Three scalar parameters, α, β, and γ are set as 1.0 for simplicity. These parameters control the importance of three criteria and greatly impact on the results. The parameter study is one of the future tasks.Although the height of the specimen h_s was 10 cm in the actual test, we set the height as 20 cm in the algorithm.

Regarding the target range of S_h ($S_h^- < S_h < S_h^+$), Yoneda et al. (2017) selected 11 samples that have wide variety of S_h. We defined six ranges, S_h = 0 – 10%, 10 – 20%, 20 – 30%, 30 – 40%, 40 – 50%, 50 – 60% and selected the “best sample” from each range. In addition to the best samples, we also focus on the second and third best samples and compare the results.

3.2 Results

Figure 3 compares the selected samples by the proposed method with those by Yoneda et al. (2017). In the figure, red colored area indicates the samples selected by Yoneda et al. (2017), and blue indicates the samples by the proposed method. Yoneda et al. (2017) selected 11 samples from only three pressure cores, AT1-C-6P, AT1-C-8P and AT1-C-20P. The proposed method also selected samples from AT1-C-6P

4. Summary

This study developed a method to automatically select the samples of MH bearing sands from pressure core based on basic statistical data analysis. The selection problem was formulated as an optimization problem, and we defined objective functions for selecting "best sample" based on the experience of researchers/engineers who study mechanical characteristic of MH bearing soils by laboratory testing. We also proposed a procedure to select the sample based on the objective function.

We compared the soil samples selected by the proposed method with those by Yoneda et al. (2017). Although a few common samples were selected, most of the samples are different. The proposed method can be used for filtering appropriate/inappropriate samples at the first stage of the selection. We should investigate the reliability and accuracy of the proposed method, and this is a future topic.

References

Yoneda, J., Masui, A., Konno, Y., Jin, Y., Kida, M. Katagiri, J. Nagao, J. and Tenma, N. 2017. *Marine and Petroleum Geology*, 86, 1–16.

Waite, W. F., Santamarina, J. C., Cortes, D. D., Dugan, B., Espinoza, D. N., Germaine, J., Jang, J., Jung, J. W., Kneafsey, T. J., Shin, H., Soga, K., Winters, W. J. and Yun, T. S. 2009. *Reviews of Geophysics*, 47(4), RG4003, 1–38.

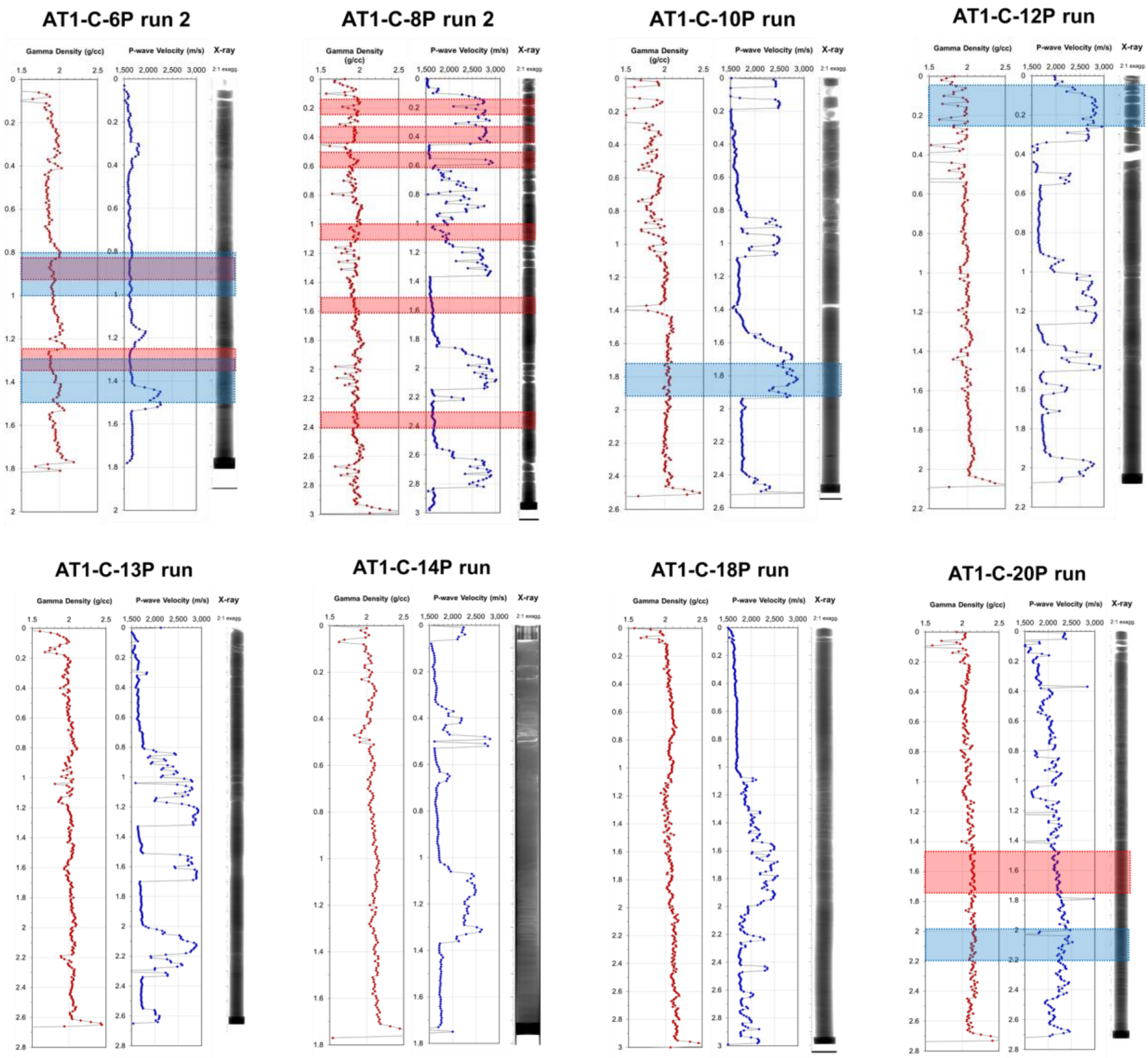

Figure 3. Comparison between sample with max TS (blue) and Yoneda et al. (2017) (red).

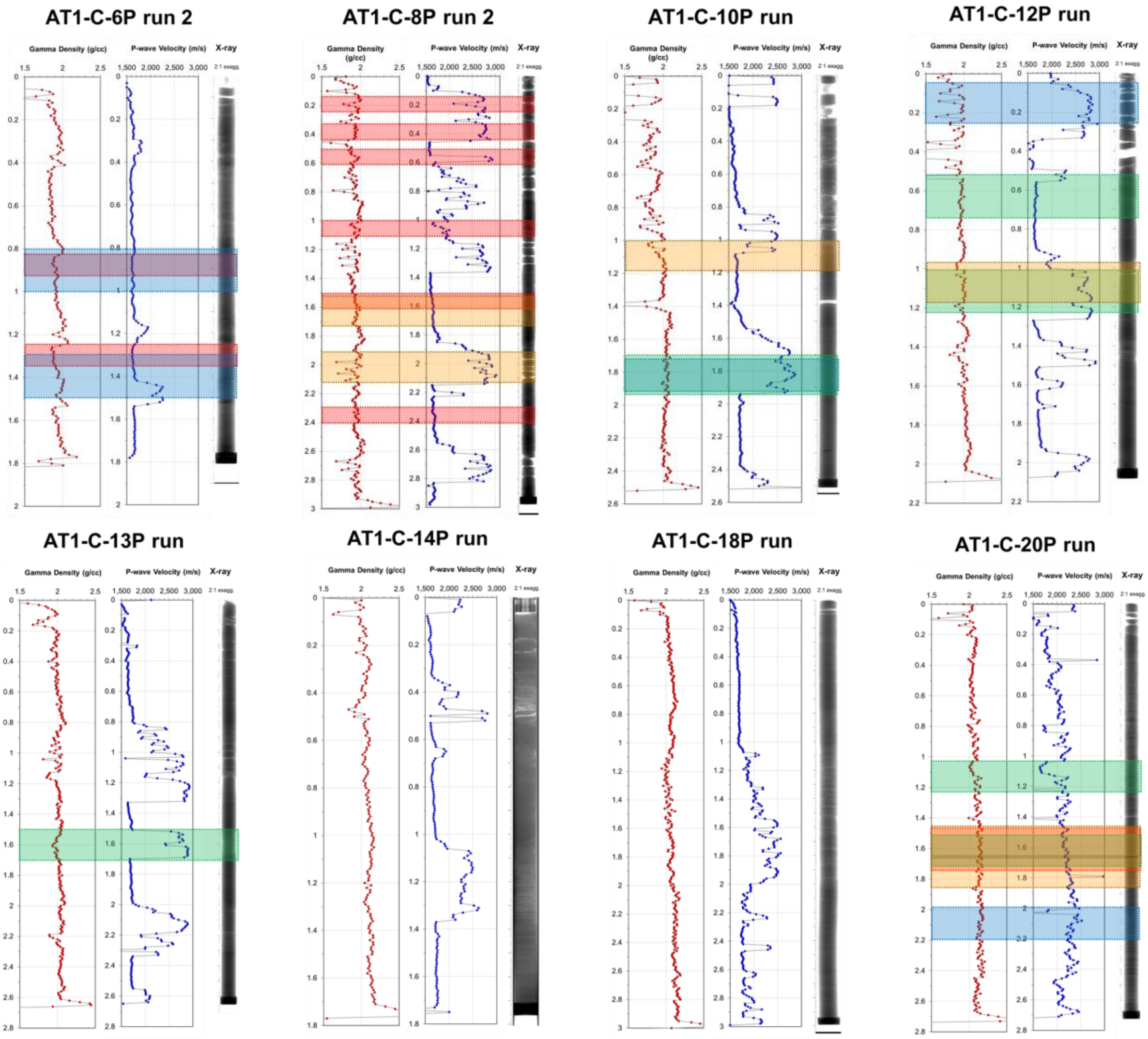

Figure 4. Comparison between sample selected by the proposed method (max TS (blue), second-max TS (green), and third-max TS (yellow) and Yoneda et al. (2017) (red).