

# Analysis of Spatial and Temporal Features to Classify the Deep Moonquake Sources

## Using Balanced Random Forest

Kodai Kato<sup>\*</sup>, Ryuhei Yamada<sup>†</sup>, Yukio Yamamoto<sup>‡</sup>,  
Masaharu Hirota<sup>§</sup>, Shohei Yokoyama<sup>¶</sup> and Hiroshi Ishikawa<sup>||</sup>

<sup>\*</sup> Faculty of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino-shi, Tokyo, Japan

Email: kato-kodai@ed.tmu.ac.jp

<sup>†</sup> National Astronomical Observatory of Japan, RISE project, 2-12 Hoshigaoka-cho, Mizusawa-ku, Oshu, Iwate, Japan

Email: r.yamada@nao.ac.jp

<sup>‡</sup> Japan Aerospace Exploration Agency, 3-1-1, Yoshinodai, Chuo-ku, Sagami-hara-shi, Kanagawa, Japan

Email: yamamoto.yukio@jaxa.jp

<sup>§</sup> Department of Information Engineering National Institute of Technology, Oita College 1666 Maki Oita-shi, Oita, Japan

Email: m-hirota@oita-ct.ac.jp

<sup>¶</sup> Faculty of Informatics, Shizuoka University, 3-5-1 Joho-ku, Hamamatsu-shi, Shizuoka, Japan

Email: yokoyama@inf.shizuoka.ac.jp

<sup>||</sup> Faculty of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino-shi, Tokyo, Japan

Email: ishikawa-hiroshi@tmu.ac.jp

**Abstract**—In this paper, we evaluate other features different from the waveforms to classify seismic sources. Classification of sources of the deep moonquakes is an important issue for analyzing the focal mechanisms and the lunar deep structures. It was found that deep moonquakes that occur from the same source have similar waveforms. Some studies have been conducted to identify the deep moonquake sources using the waveform similarities. However, classifying some deep moonquakes using only the waveforms is difficult due to large noise and the small amplitude. If we could show that other features different from the waveforms are effective for classification of deep moonquakes, we can increase the number of classifiable moonquakes even if moonquakes include noise and small amplitude of the waveforms. Therefore, we use other features to classify deep moonquakes (position and velocity relative to the Earth, Sun, Jupiter, and Venus, as seen from the Moon). We apply these features to classify deep moonquakes that are not classified based on only waveforms, and it is useful to analyze the deep moonquake occurrence mechanisms. Our experiments showed that the position and velocity relation between the Moon and the Earth or Jupiter are effective for classification.

**Keywords**—Planetary Science; Machine learning; Geophysical

### I. INTRODUCTION

The Apollo Lunar Surface Experiments Package (ALSEP) was deployed on the Moon to investigate the lunar surface, internal structure and surrounding environment through NASA's Apollo missions. The Passive Seismic Experiment (PSE) in the ALSEP has been performed to observe the lunar seismicity. The observations revealed that seismic events called moonquakes occur on the Moon. All observed data are acquired and viewed on the Web [1] [2]. The moonquake data are very important to analyze the lunar internal structure and the focal mechanisms of moonquakes, even after 40 years since PSE finished [3] [4].

Earlier studies have revealed that moonquake characteristics differ widely from those of earthquakes. For instance, contrary to earthquakes, moonquakes do not occur due to plate tectonics. Additionally, moonquakes have several types: 'Deep Moonquakes', 'Shallow Moonquakes', 'Thermal', and

'Meteoroid Impact', which are classified based on the occurrence factor or depth of seismic sources. About 13,000 events have been found to date including about 7,300 deep moonquakes, about 30 shallow moonquakes, and about 1,700 meteoroid impacts. Deep moonquakes account for over half of all moonquakes.

Deep moonquakes that occurred from the same source are similar waveforms [5] [6]. Earlier studies have classified deep moonquakes based on the similarities [7] (sources of deep moonquakes are labeled as Axx e.g., A1, A6, A200.). The purpose of this study is the classification of unclassified deep moonquakes to elucidate the deep moonquake focal mechanisms. Goto et al. [8] classified deep moonquakes, specifically examining frequency spectra of deep moonquakes using machine learning. Machine learning has advantages to classify deep moonquakes such as automation of analyses and the great reduction of human cost.

Although the waveforms are effective features to infer moonquake sources, determination of sources of some moonquakes is difficult due to large noise and the small amplitude. Previous studies have not applied any features other than the waveforms. Therefore, we specifically examine features other than the waveforms. If we show that feature such as velocity is effective for classification of deep moonquake, classification using a combination of the feature and waveforms may increase the number of classifiable moonquakes even if moonquakes include noise and small amplitude of the waveforms. Deep moonquakes occur periodically from the same source related with the tidal stresses [9]–[11]. As described herein, we extract features from the occurrence time of deep moonquakes. Then, using machine learning, we verify the effective features to classify sources of deep moonquakes. A principal benefit of our approach is that we can infer sources of deep moonquakes, irrespective of noise and amplitude of the waveforms.

This paper is organized as follows. Section II presents a review of related works including moonquake analyses.

Section III presents a method of feature evaluation. Section IV presents both experimental and analytical results. Finally, concluding remarks are presented in the last section.

## II. RELATED WORKS

Generally, time differences among arrival times of the seismic phases observed at several seismic stations are available to determine the moonquake sources. If we cannot use the time differences to estimate sources due to the noise and small amplitudes in the waveforms, then similarities of waveforms are useful to classify sources. In 1970, the moonquakes were classified manually by visual observation [12].

With the evolution of computers, Nakamura et al. [7] classified deep moonquakes using hierarchical clustering based on cross correlation of waveforms. This classified result of deep moonquakes is cataloged as a standard criterion for classification in this study. The improvement of the preprocessing methods, which use cross-correlation analyses, enables us to discover new events and to classify the unclassified deep moonquakes [13]. Furthermore, the paper written by Endrun et al. [14] proposed a method for event detection and classification using a Hidden Markov Model. Goto et al. [8] developed a web system for visualizing moonquakes considering waveform similarity using Self-Organizing Map (SOM) to advance the study of moonquake classification. This study showed that noise and small amplitudes of the waveforms affect classification criteria. Therefore, we propose an approach to classify deep moonquakes, not using the waveforms.

## III. METHODS

To verify the features for classification of deep moonquakes, we apply Balanced Random Forest [15] extended from Random Forest [16], which is a representative supervised learning method. We verify whether seismic sources can be reproduced or not using Balanced Random Forest.

### A. Features

We extract the position  $(x, y, z)$ , velocity  $(vx, vy, vz)$ , distance  $(\sqrt{x^2 + y^2 + z^2})$ , and the magnitude of the velocity  $(\sqrt{vx^2 + vy^2 + vz^2})$  of each planet (the Earth, Sun, Jupiter and Venus) relative to the Moon in the IAU\_MOON coordinate system, calculated using SPICE [17] at the deep moonquake occurrence time. We apply the orbit parameters as the features. The IAU\_MOON coordinate system is the Moon fixed coordinate system. The z-axis is the North Pole direction of the Moon. The x-axis is the meridian direction of the Moon. The y-axis is to the right of the x-z plane.

### B. Balanced Random Forest

Random Forest [16] is a classification algorithm that fits a number of decision tree classifiers on various sub-samples of the dataset. This algorithm can compute feature importance. However, Random Forest has a serious problem: the classifier might overfit with imbalanced data. Generally, when we apply Random Forest, we assign weights based on class, with the minority class assigned a larger weight to make the classifier more suitable for the imbalanced data. However, this approach can cause over-learning of minority data if the data are extremely imbalanced. We must apply a method for the imbalanced data to analyze of deep moonquakes because a different number of events occurs at each seismic source.

As described herein, we apply Balanced Random Forest [15], which equalizes the number of samples per class for

TABLE I. NUMBER OF EVENTS AT EACH SEISMIC SOURCE.

| Source | Number of events |
|--------|------------------|
| A1     | 441              |
| A5     | 76               |
| A6     | 178              |
| A7     | 85               |
| A8     | 327              |
| A9     | 145              |
| A10    | 230              |
| A14    | 165              |
| A18    | 214              |
| A20    | 153              |
| A23    | 79               |
| A25    | 72               |
| A35    | 70               |
| A44    | 86               |
| A204   | 85               |
| A218   | 74               |

each iteration in random forest. The decision tree of Balanced Random Forest uses a Gini coefficient to find splits. We then compute the feature importance by calculating the average reduction ratio of the Gini coefficient in the tree split for each feature.

## IV. RESULTS AND DISCUSSION

In this section, we present the results obtained from application of our method to the dataset. The procedures used for our proposed method are presented below.

- We extract the orbit parameters as the features at the occurrence time of deep moonquake event.
- We create a Balanced Random Forest classifier for every pair of seismic sources.
- We verify the relation of seismic sources and the features based on the classification performance and feature importance of each classifier.

In these experiments, we applied one-vs.-one, which creates a classifier for each pair of seismic sources.

As described in this paper, the number of iterations in Balanced Random Forest is 1,000. Each iteration randomly selects 30 samples for each class using bootstrap method and 3 features from 8 features. Each iteration tree is implemented using the scikit-learn [18].

### A. Dataset

The number of events in each source is given in Table I. We chose the seismic sources with more than 70 events in the lunar event catalog. The used dataset has 16 sources and 2,480 events.

### B. Criterion

We used the classification performance and the feature importance to evaluate the features. We performed 10-fold cross validation, and used the minimum f-measure in each class for the classification performance. Due to different number of seismic events at each source, the minimum score was applied because the scores depend on the amount of data in each class. We calculated feature extraction from all classifier-learned data of the target class without cross validation and the classification performance and the feature importance among each planet.

### C. Classification performance

The average of classification performance for each seismic source and planet is shown in Table II. "avg./total", which is in the last row, shows the average of all classifiers. The averages

TABLE II. AVERAGE CLASSIFICATION PERFORMANCE.

| Source     | Earth | Sun  | Jupiter | Venus |
|------------|-------|------|---------|-------|
| A1         | 0.67  | 0.46 | 0.55    | 0.35  |
| A5         | 0.85  | 0.6  | 0.74    | 0.48  |
| A6         | 0.78  | 0.59 | 0.69    | 0.5   |
| A7         | 0.82  | 0.53 | 0.7     | 0.41  |
| A8         | 0.72  | 0.48 | 0.64    | 0.41  |
| A9         | 0.87  | 0.65 | 0.72    | 0.46  |
| A10        | 0.77  | 0.54 | 0.67    | 0.45  |
| A14        | 0.81  | 0.57 | 0.73    | 0.52  |
| A18        | 0.8   | 0.58 | 0.74    | 0.48  |
| A20        | 0.78  | 0.53 | 0.69    | 0.46  |
| A23        | 0.88  | 0.65 | 0.73    | 0.45  |
| A25        | 0.83  | 0.54 | 0.75    | 0.45  |
| A35        | 0.76  | 0.52 | 0.68    | 0.46  |
| A44        | 0.79  | 0.59 | 0.74    | 0.54  |
| A204       | 0.8   | 0.52 | 0.72    | 0.44  |
| A218       | 0.76  | 0.5  | 0.64    | 0.45  |
| avg./total | 0.79  | 0.55 | 0.70    | 0.46  |

TABLE III. RANKING OF CLASSIFICATION PERFORMANCE.

| Source | Earth | Sun | Jupiter | Venus |
|--------|-------|-----|---------|-------|
| A1     | 16    | 16  | 16      | 16    |
| A5     | 3     | 3   | 4       | 4     |
| A6     | 10    | 4   | 11      | 3     |
| A7     | 5     | 11  | 9       | 15    |
| A8     | 15    | 15  | 14      | 14    |
| A9     | 2     | 2   | 7       | 6     |
| A10    | 12    | 9   | 13      | 11    |
| A14    | 6     | 7   | 5       | 2     |
| A18    | 7     | 6   | 2       | 5     |
| A20    | 10    | 10  | 10      | 8     |
| A23    | 1     | 1   | 6       | 12    |
| A25    | 4     | 8   | 1       | 10    |
| A35    | 14    | 13  | 12      | 7     |
| A44    | 9     | 5   | 2       | 1     |
| A204   | 8     | 12  | 8       | 13    |
| A218   | 13    | 14  | 15      | 9     |

of all classifiers are 0.79 for the Earth, 0.70 for Jupiter, 0.55 for the Sun, and 0.46 for Venus. A classifier using those features for the Earth has the highest classification performance presented in this paper. Among our selected seismic sources, the classifier for A23 using the features for the Earth has the highest classification performance reported in this paper. The classification performance ranking is shown in Table III. Table III shows that the classifier for A1 has the lowest classification performance reported herein. For the Earth and Jupiter A5, A23, and A25 are classified very well within the top 6.

The number of classifiers with classification performance of 0.8 or more are 70/120 (number of classifiers is 120) for Earth, 30/120 for Jupiter, 5/120 for the Sun, and 0/120 for Venus in Table II. These results demonstrate that an orbit parameter of the Earth and Jupiter relative to the Moon is effective for deep moonquake classification. Particularly, the features based on the Earth are the most effective. As described in previous papers, this fact indicates that the tidal stress caused by the Earth in the lunar interior can affect the occurrences of deep moonquakes.

The features based on the Sun and Venus are ineffective, as show in Table II. These results show that some features based on the Earth and Jupiter are effective to classify seismic sources, nonetheless some features based on the Sun and Venus are not. There are some sources, such as A5, A23 and A25, which are easy to classify by using orbit parameters. A subject of future work is analysis of why these features contribute well to the classification or not.

TABLE IV. FEATURE IMPORTANCE FOR EACH SEISMIC SOURCE: EARTH.

| Source | $x$  | $y$  | $z$  | $vx$ | $vy$ | $vz$ | Distance | Velocity |
|--------|------|------|------|------|------|------|----------|----------|
| A1     | 0.13 | 0.11 | 0.13 | 0.11 | 0.13 | 0.16 | 0.14     | 0.09     |
| A5     | 0.14 | 0.1  | 0.11 | 0.13 | 0.16 | 0.12 | 0.17     | 0.07     |
| A6     | 0.14 | 0.12 | 0.13 | 0.12 | 0.14 | 0.14 | 0.14     | 0.08     |
| A7     | 0.12 | 0.14 | 0.17 | 0.13 | 0.12 | 0.12 | 0.12     | 0.08     |
| A8     | 0.15 | 0.12 | 0.12 | 0.11 | 0.14 | 0.12 | 0.15     | 0.09     |
| A9     | 0.09 | 0.12 | 0.14 | 0.13 | 0.1  | 0.25 | 0.09     | 0.09     |
| A10    | 0.14 | 0.11 | 0.11 | 0.11 | 0.14 | 0.14 | 0.15     | 0.09     |
| A14    | 0.12 | 0.14 | 0.12 | 0.14 | 0.12 | 0.14 | 0.13     | 0.09     |
| A18    | 0.12 | 0.13 | 0.12 | 0.13 | 0.13 | 0.16 | 0.13     | 0.09     |
| A20    | 0.11 | 0.13 | 0.15 | 0.12 | 0.12 | 0.14 | 0.12     | 0.12     |
| A23    | 0.08 | 0.11 | 0.23 | 0.12 | 0.08 | 0.21 | 0.08     | 0.08     |
| A25    | 0.1  | 0.17 | 0.09 | 0.18 | 0.11 | 0.15 | 0.12     | 0.07     |
| A35    | 0.11 | 0.15 | 0.12 | 0.14 | 0.11 | 0.15 | 0.11     | 0.1      |
| A44    | 0.11 | 0.14 | 0.17 | 0.14 | 0.11 | 0.14 | 0.11     | 0.09     |
| A204   | 0.15 | 0.14 | 0.1  | 0.14 | 0.14 | 0.11 | 0.13     | 0.08     |
| A218   | 0.14 | 0.13 | 0.13 | 0.13 | 0.13 | 0.11 | 0.13     | 0.09     |

TABLE V. FEATURE IMPORTANCE FOR EACH SEISMIC SOURCE: SUN.

| Source | $x$  | $y$  | $z$  | $vx$ | $vy$ | $vz$ | Distance | Velocity |
|--------|------|------|------|------|------|------|----------|----------|
| A1     | 0.12 | 0.12 | 0.13 | 0.12 | 0.12 | 0.14 | 0.13     | 0.13     |
| A5     | 0.12 | 0.13 | 0.12 | 0.13 | 0.12 | 0.14 | 0.12     | 0.12     |
| A6     | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.14 | 0.13     | 0.13     |
| A7     | 0.12 | 0.12 | 0.13 | 0.12 | 0.12 | 0.14 | 0.13     | 0.13     |
| A8     | 0.12 | 0.12 | 0.13 | 0.12 | 0.12 | 0.13 | 0.13     | 0.13     |
| A9     | 0.12 | 0.12 | 0.13 | 0.12 | 0.12 | 0.14 | 0.14     | 0.12     |
| A10    | 0.12 | 0.12 | 0.13 | 0.12 | 0.12 | 0.14 | 0.13     | 0.13     |
| A14    | 0.12 | 0.12 | 0.13 | 0.12 | 0.12 | 0.14 | 0.13     | 0.13     |
| A18    | 0.12 | 0.12 | 0.13 | 0.12 | 0.12 | 0.14 | 0.13     | 0.13     |
| A20    | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.13 | 0.14     | 0.14     |
| A23    | 0.12 | 0.12 | 0.13 | 0.12 | 0.12 | 0.14 | 0.13     | 0.12     |
| A25    | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.14 | 0.13     | 0.12     |
| A35    | 0.13 | 0.11 | 0.13 | 0.11 | 0.13 | 0.13 | 0.13     | 0.13     |
| A44    | 0.12 | 0.12 | 0.13 | 0.12 | 0.12 | 0.13 | 0.14     | 0.13     |
| A204   | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.14 | 0.14     | 0.12     |
| A218   | 0.12 | 0.13 | 0.13 | 0.12 | 0.12 | 0.13 | 0.12     | 0.12     |

Table II and Table III show that orbit features are useful for classification of deep moonquakes as well as waveforms studied in an earlier paper [7]. As a result, our experimental results show that orbit features can classify deep moonquakes. Therefore, when we try to classify unclassified deep moonquakes or moonquakes with noise and small amplitude of the waveforms, the orbit features are effective to classify the moonquakes. In addition, the relation between the Moon and other planets includes some knowledge to analyze the occurrence of deep moonquakes.

#### D. Feature Importances

The average of feature importances in each seismic source for each planet are shown in Table IV, Table V, Table VI, and Table VII. Among all lists,  $vz$  of the Earth in A9 based on the Earth is the highest score. Also,  $z$  of the Earth in A23 based on Earth is the second highest score. The velocity of the Earth in A5 is the lowest score. In Table IV, the score related to velocity is low in all sources. The score difference in all sources is small in Table V and Table VII. Also,  $x$  and  $vy$  of A204 are the highest scores in Table VI. The feature importances of distance and velocity are low in all sources in Table VI.

We discuss a reason of high feature importances related to  $vz$  of the Earth in A9 and  $z$  of the Earth in A23. Figure 1 and Figure 2 are box plots that present values of the feature for each seismic source. Figure 1 shows that distribution of the position in the  $z$  coordinate at occurrence time of A23 events is from about -45,000 km to about -20,000 km. That of A7 events is from about 0 km to about 50,000 km. Other sources,

TABLE VI. FEATURE IMPORTANCE FOR EACH SEISMIC SOURCE: JUPITER.

| Source | <i>x</i> | <i>y</i> | <i>z</i> | <i>vx</i> | <i>vy</i> | <i>vz</i> | Distance | Velocity |
|--------|----------|----------|----------|-----------|-----------|-----------|----------|----------|
| A1     | 0.14     | 0.14     | 0.13     | 0.14      | 0.14      | 0.12      | 0.1      | 0.1      |
| A5     | 0.18     | 0.13     | 0.11     | 0.13      | 0.17      | 0.1       | 0.09     | 0.09     |
| A6     | 0.14     | 0.15     | 0.13     | 0.14      | 0.13      | 0.12      | 0.1      | 0.1      |
| A7     | 0.12     | 0.16     | 0.13     | 0.16      | 0.13      | 0.11      | 0.09     | 0.09     |
| A8     | 0.14     | 0.15     | 0.13     | 0.15      | 0.14      | 0.11      | 0.09     | 0.1      |
| A9     | 0.13     | 0.17     | 0.12     | 0.17      | 0.13      | 0.11      | 0.09     | 0.09     |
| A10    | 0.13     | 0.16     | 0.12     | 0.16      | 0.13      | 0.11      | 0.09     | 0.09     |
| A14    | 0.13     | 0.16     | 0.12     | 0.17      | 0.14      | 0.11      | 0.09     | 0.09     |
| A18    | 0.16     | 0.14     | 0.12     | 0.14      | 0.17      | 0.11      | 0.09     | 0.09     |
| A20    | 0.13     | 0.14     | 0.14     | 0.14      | 0.14      | 0.12      | 0.1      | 0.1      |
| A23    | 0.17     | 0.14     | 0.12     | 0.14      | 0.17      | 0.11      | 0.09     | 0.08     |
| A25    | 0.14     | 0.17     | 0.11     | 0.18      | 0.14      | 0.1       | 0.08     | 0.08     |
| A35    | 0.13     | 0.15     | 0.13     | 0.15      | 0.13      | 0.13      | 0.09     | 0.09     |
| A44    | 0.14     | 0.15     | 0.15     | 0.15      | 0.14      | 0.1       | 0.09     | 0.09     |
| A204   | 0.17     | 0.13     | 0.12     | 0.13      | 0.18      | 0.1       | 0.08     | 0.09     |
| A218   | 0.15     | 0.13     | 0.13     | 0.13      | 0.16      | 0.11      | 0.09     | 0.1      |

TABLE VII. FEATURE IMPORTANCE FOR EACH SEISMIC SOURCE: VENUS.

| Source | <i>x</i> | <i>y</i> | <i>z</i> | <i>vx</i> | <i>vy</i> | <i>vz</i> | Distance | Velocity |
|--------|----------|----------|----------|-----------|-----------|-----------|----------|----------|
| A1     | 0.12     | 0.12     | 0.13     | 0.12      | 0.12      | 0.13      | 0.12     | 0.12     |
| A5     | 0.13     | 0.12     | 0.12     | 0.12      | 0.13      | 0.13      | 0.12     | 0.12     |
| A6     | 0.12     | 0.12     | 0.13     | 0.12      | 0.12      | 0.13      | 0.12     | 0.12     |
| A7     | 0.12     | 0.12     | 0.14     | 0.12      | 0.13      | 0.13      | 0.12     | 0.12     |
| A8     | 0.12     | 0.12     | 0.13     | 0.13      | 0.12      | 0.13      | 0.12     | 0.12     |
| A9     | 0.13     | 0.12     | 0.13     | 0.12      | 0.13      | 0.13      | 0.12     | 0.12     |
| A10    | 0.12     | 0.12     | 0.13     | 0.12      | 0.12      | 0.14      | 0.12     | 0.12     |
| A14    | 0.13     | 0.12     | 0.13     | 0.12      | 0.13      | 0.13      | 0.12     | 0.12     |
| A18    | 0.13     | 0.12     | 0.13     | 0.12      | 0.13      | 0.13      | 0.12     | 0.12     |
| A20    | 0.13     | 0.12     | 0.13     | 0.12      | 0.13      | 0.13      | 0.12     | 0.12     |
| A23    | 0.12     | 0.12     | 0.14     | 0.12      | 0.13      | 0.13      | 0.12     | 0.12     |
| A25    | 0.13     | 0.12     | 0.13     | 0.12      | 0.13      | 0.13      | 0.12     | 0.12     |
| A35    | 0.12     | 0.12     | 0.13     | 0.12      | 0.13      | 0.13      | 0.13     | 0.12     |
| A44    | 0.12     | 0.13     | 0.13     | 0.13      | 0.12      | 0.13      | 0.12     | 0.12     |
| A204   | 0.13     | 0.12     | 0.13     | 0.12      | 0.13      | 0.13      | 0.12     | 0.12     |
| A218   | 0.13     | 0.12     | 0.13     | 0.12      | 0.13      | 0.13      | 0.12     | 0.12     |

such as those of A1, A8, and A10, have a wider range than A23 or A7. Conversely, Figure 2 shows that the distribution of the velocity in the *z* coordinate at occurrence time of A9 events is from about 0.06 km/s to about 0.13 km/s. That of A5 is from about -0.13 km/s to about -0.02 km/s. Other sources, such as A8, A20, and A35, have a wider range than A9 or A5.

Figure 1 and Figure 2 show differences in the distribution of the features for each seismic source. We can show that the features with high importance have a narrow distribution and that seismic sources have features which have high feature importances; A1 does not have these features.

Moreover, the position and velocity in *z* coordinate at occurrence time of each seismic event in a few sources are shown in Figure 3 and Figure 4. Figure 3 shows that fluctuation of *z* position at occurrence time of A23 events is small through the observation period and that the occurrence frequency of A23 from about 1975 to about 1976 is less frequent than in other periods. The distribution of *z* position at occurrence time of A1 events changed from about 1973 to about 1975. Conversely, Figure 4 shows that fluctuation of *z* velocity at occurrence time of A9 events is small through the observation period. The distribution of *z* velocity at occurrence time of A1 events changed from about 1973 to about 1975.

The results show that there is a time variation of the features for each source. Therefore, the classification performance of A1 might be improved if features with time variation could be considered.

Next, we discuss the velocity of the Earth and the distance of Jupiter as examples of the features with low feature im-

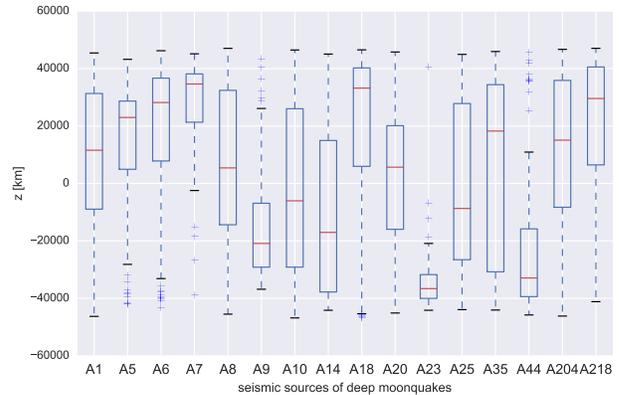


Figure 1. Box plots showing the position of *z* coordinate for each seismic source,

where red line is the median, and the box is a value of 25%–75 %, where top and bottom bars are the maximum and minimum, and “+” is an outlier, which is more than 1.5 times the interquartile range.

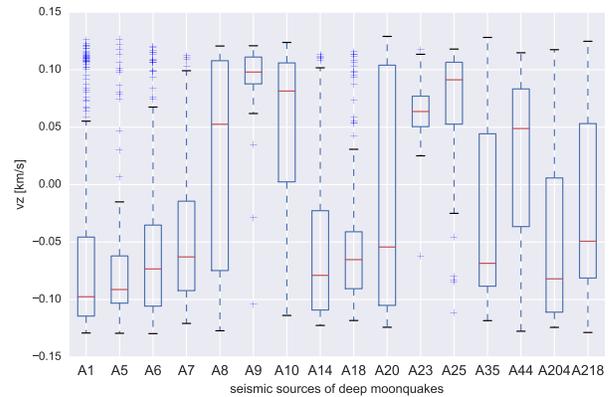


Figure 2. Box plots showing Earth velocity of *z* coordinate for each seismic source.

portance. Figure 5 and Figure 6 are box plots that indicate values of these features for each seismic source. Figure 5 and Figure 6 show that the low feature importances are caused by small difference among values of the features for each seismic source.

Time variations of the velocity of Earth and distance of Jupiter are presented in Figure 7 and Figure 8. Figure 7 shows that the velocity of Earth greatly varies with time. Figure 8 shows that the period of distance is about 1 year. Figure 7 and Figure 8 show that it is difficult to extract tendencies of the features because, this time, the variation is very different from Figure 3 or Figure 4. These results show that we may be able to improve classification if we apply methods and features considering the time variation or periodicity.

### E. Methods and Features

Using Balanced Random Forest, we easily calculated the feature importance in addition to the classification perfor-

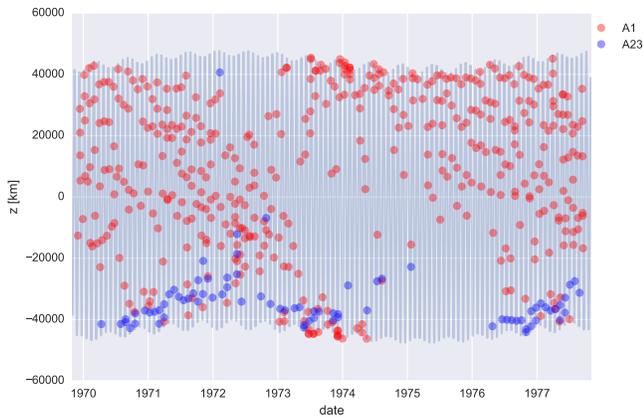


Figure 3. Time series of Earth position of z coordinate for each seismic source.

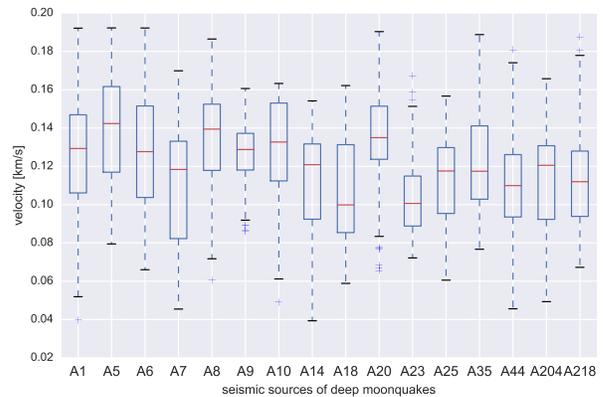


Figure 5. Box plots showing Earth velocity for each seismic source.

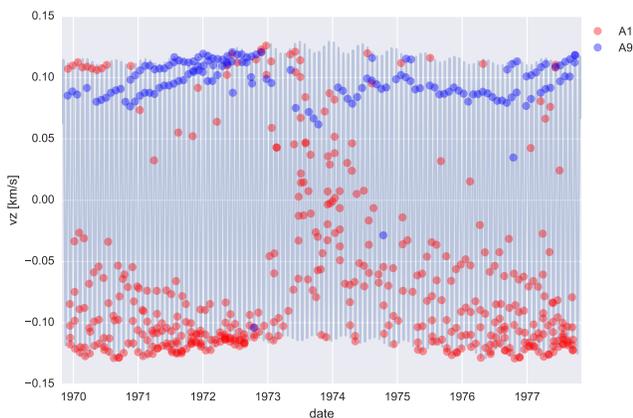


Figure 4. Time series of Earth velocity of z coordinate for each seismic source.

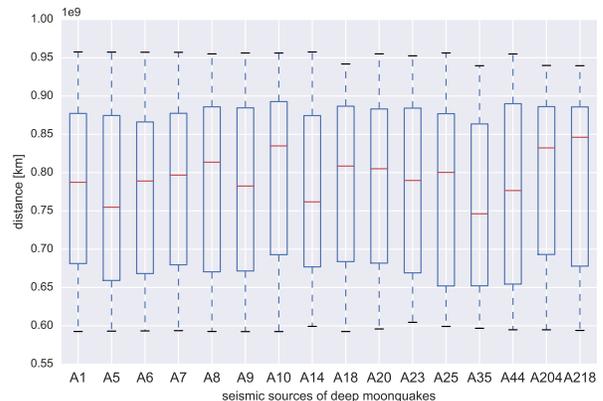


Figure 6. Box plots showing the Jupiter distance for each seismic source.

mance. As a result, some orbit parameters are useful for the classification of deep moonquakes. However, we did not do fine-tuning to improve the classification performance. We need to apply other machine learning methods, fine tuning of parameters, and waveforms to classify more precisely.

We avoided analyzing all features to limit multicollinearity. Therefore, it is difficult to declare decisive features to characterize seismic sources. Additionally, in this paper, it is difficult to estimate causal relationship.

Accordingly, we must verify new features or preprocess features with multicollinearity considering features leading to elucidation of the causes of deep moonquakes. However, the features in our approach are effective for new analyses and for creating knowledge of experts. Our findings show the causal relations between seismic sources and outer space for occurrence of deep moonquakes.

### V. CONCLUSION

This study evaluated the spatial and temporal features for classification of deep moonquake sources using Balanced Random Forest. The findings reported in this paper are presented

below.

- Seismic sources are classifiable using temporal and spatial features without using the waveforms used in conventional classification.
- Results of the classification performance using orbit parameters of objects in our Solar System (Earth, Sun, Jupiter, and Venus) suggest that the Earth orbit parameter is the most effective feature among them. The Jupiter orbit parameter is effective for classification of some seismic sources.
- Features of seismic sources with low time variation have high feature importance.

Our experimental results show that the orbit features are effective when we try to classify unclassified deep moonquakes or moonquakes with noise and small amplitude of the waveforms. These findings are expected to be useful for new analyses and for knowledge creation by experts. Further progress of this study can generate new knowledge about deep moonquake occurrence mechanisms. Future works are described below.

- Verification considering correlation and confounding among some features.

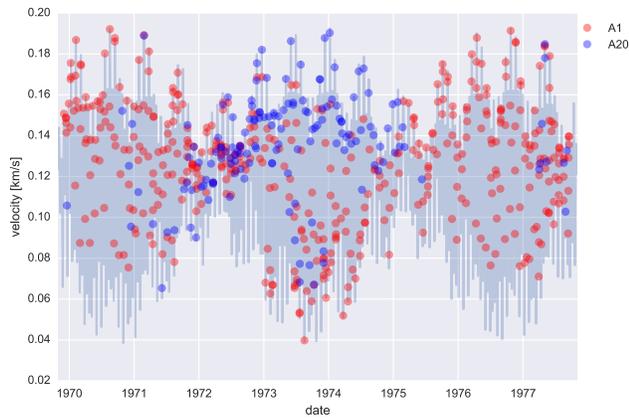


Figure 7. Time series showing Earth velocity for each seismic source.

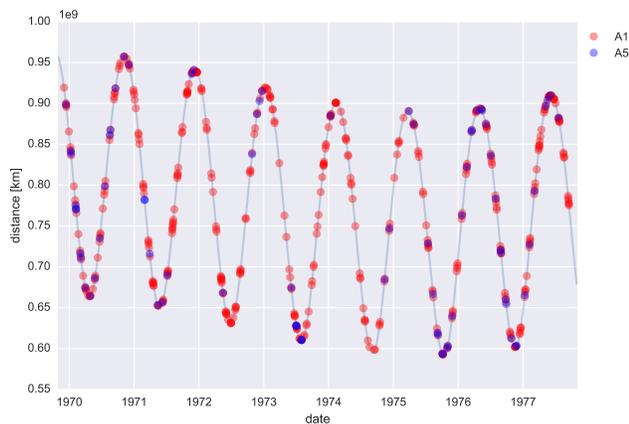


Figure 8. Time series of the Jupiter distance for each seismic source.

- Analysis considering time variation and interplanetary relations.

We can address future issues and problems from the standpoint of planetary interpretation.

#### ACKNOWLEDGMENT

We are grateful to Dr. Yoshiyuki Shoji for helpful discussions.

This work was supported by JSPS KAKENHI Grant Numbers 16K00157, 16K16158, and a Tokyo Metropolitan University Grant-in-Aid for Research on Priority Areas "Research on social big data."

#### REFERENCES

[1] R. Yamada, Y. Yamamoto, J. Kuwamura, and Y. Nakamura, "Development of an online retrieval system of apollo lunar seismic data," *Journal of Space Science Informatics Japan*, no. 1, 2012, pp. 121–131.

[2] C-SODA at ISAS/JAXA. DARTS. [Online]. Available: <http://darts.jaxa.jp> [retrieved: 3, 2017]

[3] Y. Nakamura, G. V. Latham, and H. J. Dorman, "Apollo lunar seismic experiment—final summary," *Journal of Geophysical Research: Solid Earth* (1978–2012), vol. 87, no. S01, 1982, pp. A117–A123.

[4] P. Lognonné, J. Gagnepain-Beyneix, and H. Chenet, "A new seismic model of the moon: implications for structure, thermal evolution and formation of the moon," *Earth and Planetary Science Letters*, vol. 211, no. 1, 2003, pp. 27–44.

[5] R. C. Bulow, C. L. Johnson, B. G. Bills, and P. M. Shearer, "Temporal and spatial properties of some deep moonquake clusters," *Journal of Geophysical Research: Planets* (1991–2012), vol. 112, no. E9, 2007.

[6] D. R. Lammllein, G. V. Latham, J. Dorman, Y. Nakamura, and M. Ewing, "Lunar seismicity, structure, and tectonics," *Reviews of Geophysics*, vol. 12, no. 1, 1974, pp. 1–21.

[7] Y. Nakamura, "New identification of deep moonquakes in the apollo lunar seismic data," *Physics of the Earth and Planetary Interiors*, vol. 139, no. 3, 2003, pp. 197–205.

[8] Y. Goto, R. Yamada, Y. Yamamoto, S. Yokoyama, and H. Ishikawa, "Som-based visualization for classifying large-scale sensing data of moonquakes," in *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2013 Eighth International Conference on*. IEEE, 2013, pp. 630–634.

[9] R. Weber, B. Bills, and C. Johnson, "Constraints on deep moonquake focal mechanisms through analyses of tidal stress," *Journal of Geophysical Research: Planets*, vol. 114, no. E5, 2009.

[10] J. Koyama and Y. Nakamura, "Focal mechanism of deep moonquakes," in *Lunar and Planetary Science Conference Proceedings*, vol. 11, 1980, pp. 1855–1865.

[11] Y. Nakamura, "A1 moonquakes—source distribution and mechanism," in *Lunar and Planetary Science Conference Proceedings*, vol. 9, 1978, pp. 3589–3607.

[12] Y. Nakamura, G. V. Latham, H. J. Dorman, and J. Harris, "Passive seismic experiment long-period event catalog," *Galveston Geophysics Laboratory Contribution*, vol. 491, 1981.

[13] R. C. Bulow, C. L. Johnson, and P. Shearer, "New events discovered in the apollo lunar seismic data," *Journal of Geophysical Research: Planets* (1991–2012), vol. 110, no. E10, 2005.

[14] B. Knapmeyer-Endrun and C. Hammer, "Identification of new events in apollo 16 lunar seismic data by hidden markov model-based event detection and classification," *Journal of Geophysical Research: Planets*, vol. 120, no. 10, 2015, pp. 1620–1645.

[15] C. Chen, A. Liaw, and L. Breiman, "Using random forest to learn imbalanced data," *University of California, Berkeley*, 2004, pp. 1–12.

[16] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, 2001, pp. 5–32.

[17] NASA. SPICE. [Online]. Available: <https://naif.jpl.nasa.gov/naif/> [retrieved: 3, 2017]

[18] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, 2011, pp. 2825–2830.