IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 2, 2019, accepted December 16, 2019, date of publication December 20, 2019, date of current version December 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2961115

A New Compressional Wave Speed Inversion Method Based on Granularity Parameters

JINGQIANG WANG[1,2], ZHENGYU HOU[1,3], GUANBAO LI[2], GUANGMING KAN[2], XIANGMEI MENG[2], AND BAOHUA LIU[4]

[1]School of Marine Sciences, Sun Yat-sen University, Zhuhai, Guangdong 519082, China
[2]Key Laboratory of Seafloor Sedimentology and Environmental Geology, First Institute of Oceanography, MNR, Qingdao 266061, China
[3]CAS Key Laboratory of Ocean and Marginal Sea Geology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China
[4]National Deep Sea Center, MNR, Qingdao 266237, China

Corresponding author: Zhengyu Hou (zyhou@scsio.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 41706045, Grant 41706062, and Grant 41330965, in part by the Guangdong Natural Science Foundation under Grant 2017A030313237, in part by the Director Fund of Qingdao National Laboratory for Marine Science and Technology under Grant QNLM201713, in part by the Key Laboratory of Marine Mineral Resources, Ministry of Land and Resources, Guangzhou Marine Geological Survey under Grant KLMMR-2017-B-09, in part by the Key Laboratory of Marine Geology and Environment, Chinese Academy of Sciences under Grant MGE2017KG02, and in part by the Taishan Scholar Project Funding under Grant tspd20161007.

ABSTRACT How to improve the prediction accuracy of compressional wave speed has always been one of the basic research subjects in geoacoustics study field. Due to the stability of granularity, whether in the laboratory or in the seabed environment, the regression relationship between compressional wave speed and granularity is an important sound speed inversion method. Machine Learning (ML) provides a new solution for more efficient sound speed prediction systems. In this study, two ML algorithm, Random forest (RF) and Support Vector Regression (SVR), combined with nine granularity parameters (mean grain size, median grain size, skewness, kurtosis, sorting coefficient, gravel, sand, silt, and clay content respectively.) to analysis the effect of granularity property on sound speed. As a result, the sound speed-granularity predictive models were established, and the sound speed accuracy obtained based on the predictive models are higher than that of the regression equations, and the RF model has a higher accuracy than the SVR model. Based on the RF predictive model, the feature selection was conducted and the results show that the most influential parameter of granularity is mean grain size. Furthermore, the RF model can also predict the sound speed with high precision in the absence of partial parameters, which can be a useful tool for ocean engineering and seismic inversion.

INDEX TERMS Geoacoustic inversion, granularity, machine learning, random forest, sound speed.

I. INTRODUCTION

The sediment acoustic properties study is important in the underwater acoustic and geo-engineering field. In underwater acoustic and seafloor classification measurements, the compressional wave speed (hereafter referred to as the sound speed) is dependent on the properties of the sedimentary system. Hence, accurate prediction of the sound speed is important for interpretation of the reflected acoustic signal and geophysical parameters [33], [37]. Many factors can affect the sound speed of unconsolidated seafloor sediments, including the physical properties, sedimentary environment, and sediment structure. The physical properties are considered the most important [10], [17], especially the granularity, which is described by factors such as the mean grain size, median grain size, grain fraction, and sorting coefficient. As the grain size keep unchanged when the sediment was collected from the seafloor to the laboratory, the relationship between sound speed and granularity has long been an important topic of sediment acoustic study [2].

Many studies have suggested that the sound speed is highly correlated to granularity property and sediment type, as important as the porosity and the wet density in various sedimentary environments [12], [21]. Hamilton *et al.* [11] used the diving equipment to insert devices into the seafloor to measure the sound speed of seabed sediments in the

The associate editor coordinating the review of this manuscript and approving it for publication was Xiang Huang.

San Diego area (USA), and observed a positive correlation between sound speed and median grain size. Sutton *et al.* [34] used a sonic pulse system to measure 26 sediment cores and considered that the sound speed gradually increased with increasing median grain size. Hamilton [13] verified that sound speed increases with mean grain size and decreasing amounts of clay-size material. Orsi and Dunn [28] analyzed surface sediments from the Barents Sea and also revealed that the sound speed had the most closed relationship with clay content and mean grain size. Based on measurement results in the North Atlantic Ocean, Schreiber [31]suggested that both the median grain size and porosity were the important factors for sound speed prediction in sediments. In contrast, Horn *et al.* [15] argued that the mean grain size is the best index for sound speed prediction and is a more important variable than the median grain size for determining some physical and acoustic properties. The granularity property affects the sound speed usually via its influence on porosity, density, and other factors. The uni-model grain size distribution was established [7], which assumed that the porosity of an unconsolidated sediment is dependent on the grain size. In the case of coarse particles, the large pore space between aggregates increases the porosity, resulting in a high water content and low density [26]. Some studies have established the quadratic equations to predict the sound speed of the sediments based on the physical properties measured in different sedimentary environments [29]. Hamilton [12] firstly combined the sound speed in sediment with the sedimentary environment to establish sound speed prediction equations for various sedimentary environments (the continental terrace, deep-water abyssal plain, and abyssal hill). It was observed that the sound speed was most closely related to the mean grain size and the clay content in the continental terrace sediments. The Coastal Benthic Boundary Layer project measured the top 40 cm of sediment and the results indicated that the sound speed and attenuation exhibited vertical and lateral variability [6]. Kim *et al.* [22] classified the seafloor of the South Sea shelf into four areas based on the distribution of grain size, and established the prediction equation of sound speed. Yang and Seong [38] presented the sound speed and attenuation data as a function of frequency and grain size distribution. Although many studies have established the sound speed prediction equations using granularity parameters in different ocean regions [10], [18], [19], [23], [25] [41] , these empirical equations only considered one or two parameters for predicting the sound speed. The present study also mainly focuses on the median and/or mean grain size, as a single grain size parameter cannot adequately represent the granularity, nor accurately predict the sound speed of the sediment.

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. The ML algorithm has been widely used in many fields. The hybrid Elman Recurrent Neural Network models can provide a successful alternative to achieve better accuracy in streamflow forecasting [39]. The Multivariate Adaptive Regression Spline integrated with differential Evolution to forecast streamflow pattern in semi-arid region exhibited an excellent hybrid predictive modeling capability for monthly time scale streamflow in semi-arid region [1]. The ML algorithm also can be implemented in the geoacoustic area. Hou *et al.* [16] has established the predictive model of sediment sound speed based on the random forest decision tree (RF) using several physical properties.

Naturally occurring marine sediments may have poor sorting or multimodal granularity distribution. While well-classified coarse-grained sediments can be characterized by their granularity, the particle shape can have a significant impact on the acoustic properties of mixed or fine-grained sediments. This study would use two ML methods, the Random forest (RF) and support Vector Regression (SVR) algorithm, combined with nine granularity indexes to study the relationship between sound speed and granularity, and establish a model for predicting sound speed. The nine selected parameters were the mean grain size (Mz), median grain size (Md), skewness (Sk), kurtosis (Ku), sorting coefficient (So), and the gravel (Gra), sand, silt, and clay contents.

II. DATA AND PROCESSING

In this study, 115 core samples of unconsolidated sediments have been collected from the South China Sea (SCS) by a gravity corer with a plexiglass hard tube in it. The advantages of using a plexiglass tube as the inner liner are: 1) convenient to remove sediment core from the gravity corer; 2) the two ends are sealed, which is conducive to the preservation and transport of the sediment core moisture. The sediment cores are collected in the shallow area of SCS (water depth < 1000m), and the shortest sediment core is only 50 cm long and the longest core is 4 meters long.

The sediments cores in plexiglass hard tubes were taken to the laboratory, where the sound speed of the sediments was measured using a portable WSD-3 digital sonic instrument using the coaxial differential distance measurement method in a standard laboratory (23 °C, atmospheric pressure). Put the sediment cores in a platform and connected the up and bottom of the sediment core with the WSD-3 transducers, the up transducer sends pulse wave, which propagates through the sediment core and received by the bottom transducer, than the sound speed can be calculated using the formula:

$$Vp = \frac{L}{t_2 - t_1}$$

where t_1 and t_2 are the travel time from the emitting transducer to the receiving transducer in water and sediment core respectively, L is the length of the sediment core. The measurement parameters were as follows: the sediment core initial length was cut into average of 25 cm section; and the full waveforms were digitized at frequencies of 100 kHz, a single channel sampling length of 4096 points and a sampling interval of 0.1 μs. Every cut section of the core needs to be measured. More details on the measuring procedure are described in Hou *et al.* [17], [18]. The sound speed pulse was detected via wavelet transform methods, the calculation of uncertainty was performed using statistical methods. The granularity parameters were characterized by using a Mastersizer 2000 granularity analyzer.

Environmental factors (temperature, salinity and hydrostatic pressure) have a notable effect on the sound speed of seawater and on the acoustic properties of seafloor sediment [17], [42]. Before the model was established, the measured sound speed was corrected to in situ condition using the velocity ratio method [12], although the laboratory sound speed can not be entirely corrected. The corrected data still can be used to establish various models and to analyze the prediction of different models. Detailed descriptions of the velocity ratio method also were reported in Hou *et al.* [17].

As the aim of this study was to further understand the relationship between sound speed and granularity, two ML algorithm (RF and SVR) were used to establish the acoustic-speed–granularity models, followed by feature selection to identify the most important variables.

A. REGRESSION ANALYSIS (RA)

Regression analysis (RA) is a widely used prediction method that uses a set of statistical processes to estimate the relationship between variables. RA usually uses the relationship between dependent and independent variables to understand how typical values of the dependent variable change. RA includes many techniques for modeling and analyzing multiple variables, such as linear regression and polynomial regression.

For a data set consisting of a Y by X matrix, $(y_1; x_{11}, x_{12}\ldots, x_{1k})$, $(y_2; x_{21}, x_{22}\ldots, x_{2k})$, …, $(y_n; x_{n1}, x_{n2}\ldots, x_{nk})$, the relationship between the real value Y and the feature elements X of the data can be found by defining an equation (model). In linear regression, the model can be represented by the following formula:

$$Y_i = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k + e_i$$
$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$
$$\Longleftrightarrow Y_{n\times 1} = X_{n\times k}b + e \tag{1}$$

In multivariate quadratic equation, Y is modelled as the quadratic polynomial of X, and the general multivariate quadratic equation is:

$$Y_i = b_0 + b_1x_i + b_2x_i^2 + e_i \tag{2}$$

Let $z_{1i} = x_i$, $z_{2i} = x_i^2$, and we turn the quadratic equation into a linear equation:

$$Y_i = b_0 + b_1z_{1i} + b_2z_{2i} + e_i \tag{3}$$

Therefore, the multivariate quadratic equation can be treated as a linear regression. Since the least square method can minimizes the variance of the unbiased estimators of the coefficients [8], in this article, we used the least square method to solute the regression problem.

For the seafloor sediments, the nine granularity indexes were used as the independent variables, and the sound speed V was the dependent variable. Since a close correlation among independent variables exists, using ordinary least squares regression will produce excessive errors when estimating the coefficients. In previous studies, one-parameter or two-parameter equations are generally established [12], [18], [25]. Here, both one- and two-parameter equations of sound speed and granularity were established using the least squares method, as presented in the following sections. It is straightforward to use RA analysis for one-parameter and two-parameter equations. However, it is challenging to establish multi-parameter equations with more parameters as parameter selection and the relationship between parameters and the correlation coefficients becomes complex.

B. RANDOM FOREST (RF)

The RF is a combination of a decision tree and ensemble learning [5], [14] and is used here to deal with multidimensional parameters. RF runs efficiently on large databases and can handle thousands of input variables without variable deletion; moreover, it has proven to be an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing [24].

The RF algorithm randomly extracts m sub-samples from the original data set, but when training each base learner, instead of selecting the best features from all features to perform node segmentation, k features are randomly selected. The optimal feature is selected from the k features to segment the nodes, thereby further reducing the variance of the model [30].

Here, the RF model was established using nine granularity indexes (Mz, Md, Sk, Ku, So, Gra, sand, silt, and clay) as input parameters, which were used to form a training matrix {x1, x2,..., x9} of 115×9 data points. We first performed row and column sampling on the training matrix. The bootstrap method was used for row sampling to avoid over-fitting as there may have been duplicates in the collected samples. For column sampling, we selected m features, where $m \ll 9$ as the matrix contained nine features. Then, the decision tree was established by completely splitting the data after sampling [35]. Since the previous two random sampling processes ensured randomness, even without pruning, over-fitting will not occur. Although each tree obtained in this way was weak, after combining all nodes, a high-precision model was obtained.

For each tree, bootstrap sampling of the total training set was performed; hence, some samples in the total training set may appear multiple times in the training set of a particular tree, or may never appear. About one-third of the records for each resampling were not extracted. The remaining data that was not extracted from the control data set. Therefore, the RF method does not need to reserve some data for

TABLE 1. Statistic of granularity data.

Index	Sound speed (m/s)	Gra (%)	Sand (%)	Silt (%)	Clay (%)	Mz (ϕ)	Md (ϕ)	Sk	Ku	So
mean	1555.60	1.02	22.03	53.61	23.56	5.98	5.81	-0.17	0.99	2.22
max	1773.00	22.75	71.36	78.73	48.87	8.00	7.92	0.15	1.31	3.42
min	1446.00	0.00	0.07	16.67	4.42	2.28	2.31	-0.67	0.72	1.46

cross-validation as it has an internal process similar to cross-validation, where the out-of-bag (OOB) error is used as an unbiased estimate of the prediction error [4], [5].

C. SUPPORT VECTOR REGRESSION (SVR)

The Support Vector (SV) algorithm is a nonlinear generalization of the Generalized Portrait algorithm developed in Russia in the sixties [36]. In its present form, the SV machine was largely developed at AT&T Bell Laboratories by Vapnik and co-workers [32].

The SV machine can also be used as a regression method, called Support Vector Regression (SVR), which maintaining features that characterize the algorithm (maximal margin). SVR usually maps the feature data to a high-dimensional space, and finds a linear regression plane in the high-dimensional space. Finally, the high-dimensional regression plane is mapped back to the low-dimensional, and the feature data can be predicted in the low-dimensional space. After mapping back, the low-dimensional regression plane is non-linear. However, mapping low-dimensional data to high-dimensional data, and calculating in high-dimensional data, is often very complex. So here comes the kernel function. The kernel function calculates the data when the dimension is low, and this calculation can be regarded as mapping the low-dimensional data to the high-dimensional space. In a non-linear SVR, the kernel is the radial basis function (Gaussian, rbf). The principle of the algorithm is beyond the scope of this study, the details information can be found in [32]. The advantage of SVR is that as long as the predicted value does not deviate too much from the true value, the prediction can be considered correct without calculating the loss. The accuracy of the SVR algorithm has exceeded or is comparable to traditional RA.

III. RESULTS

Statistical analysis of the nine granularity indexes and sound speed values used in this paper are listed in Table 1. The sound speed ranged from 1446 to 1773 m/s. The highest sound speed was measured in fine sand sediment with a complex granularity and high gravel and sand contents. The linear regression analysis results are shown in Table 2. The quadratic regression analysis results for the nine single parameters are shown in Figure 1 and Table 3. The RF and SVR regression results are shown in Table 4. The relationship between sound speed and the nine granularity indexes showed two clear behaviors. One, the sound speed increased with increasing granularity parameter (sand content and So). Two, the sound speed increased with decreasing granularity parameter (silt, clay, Mz, Md, Ku, and Sk). Due to the small amount of gravel content data (most samples did not contain gravel), the sound speed did not show regular change with gravel content here. These behaviors indicated that the sound speed has closed correlation with granularity properties.

TABLE 2. Linear regression of single-parameters.

Parameter	Regression equation	Correlation coefficient	Standard deviation	Root mean square error
V-Sand	V=3.355*Sand+1482	0.6885	1.8075	2.9190
V-Mz	V=1861-51.06*Mz	0.6839	1.7491	2.8933
V-Md	V=1831-47.37*Md	0.6610	1.8384	2.9916
V-Silt	V=1769-3.976*Silt	0.5546	2.0395	3.5453
V-Clay	V=1690-5.697*Clay	0.5058	2.1429	3.6032
V-So	V=1286+121.4*So	0.5005	2.3233	3.6422
V-Sk	V=1510-262*Sk	0.2954	2.7056	4.2564
V-Gra	V=1545+9.989*Gra	0.2267	2.6733	4.4495
V-Ku	V=1723-173*Ku	0.0788	3.0357	4.8258

In order to analysis the prediction accuracy of different equations or models, the standard deviation (STDEV) and root mean square error (RMSE) were calculated here. The STDEV is a measure of the degree of dispersion of a set of numbers, while the RMSE is a measure of the deviation between the observed value and the true one. Hence, STDEV and RMSE can reflect the accuracy of the prediction results. For the linear regression, sound speed vs. sand content had the high correlation coefficient (0.6885), and low STDEV (1.8075) and low RMSE (2.9190). For the quadratic regression, sound speed vs. sand content also had the high correlation coefficient (0.7258), and low STDEV (1.633) and low RMSE (2.727). As shown in Table 4, the correlation

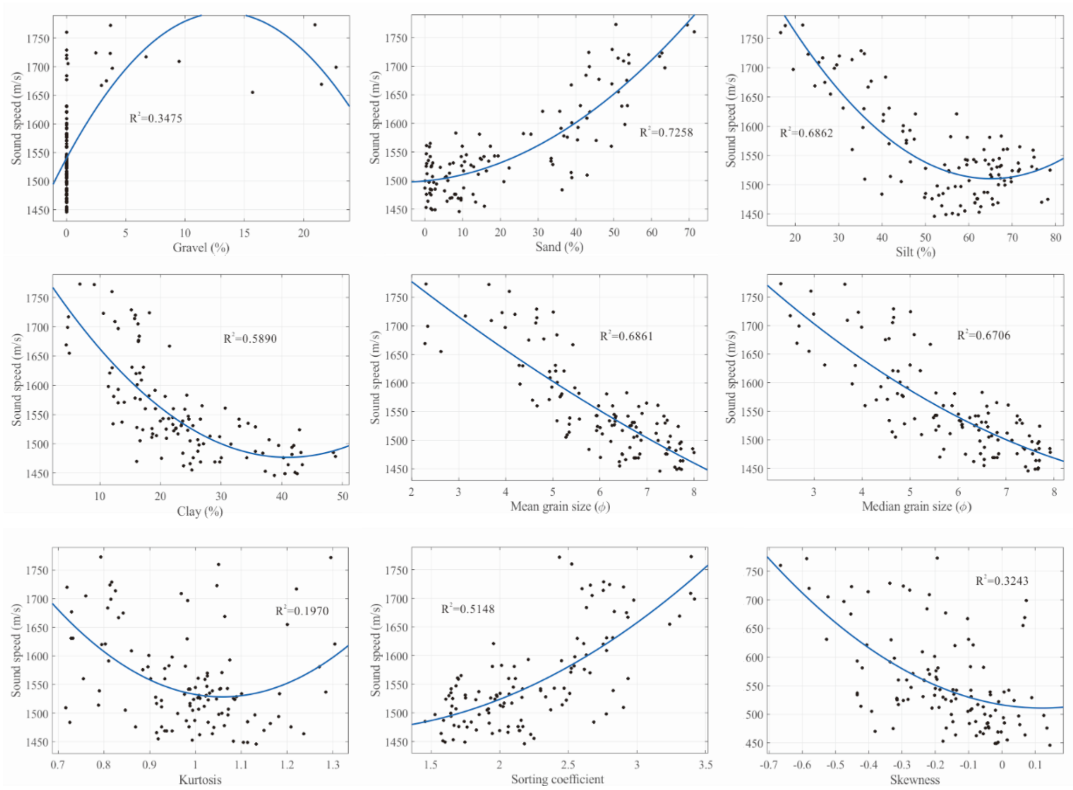

FIGURE 1. Quadratic regression analysis of the single parameter (Gra, Sand, Silt, Clay, Mz, Md, Ku, So, Sk).

TABLE 3. Quadratic equation of single-parameter.

Parameter	Regression equation	Correlation coefficient	STDEV	RMSE
V-Sand	$V=1499+0.04817*Sand^2+0.6254Sand$	0.7258	1.6330	2.7272
V-Mz	$V=1911+1.748*Mz^2-70.32Mz$	0.6861	1.7544	2.8744
V-Md	$V=1932+3.681*Md^2-87.39Md$	0.6706	1.8318	2.9343
V-Silt	$V=2033+0.1235*Silt^2-16.08*Silt$	0.6862	1.6947	2.9347
V-Clay	$V=1801+0.1931*Clay^2-15.83Clay$	0.5890	2.0394	3.2597
V-So	$V=1495+39.74*So^2-64.91*So$	0.5148	2.3138	3.5676
V-Sk	$V=1517+385.2*Sk^2-94.34*Sk$	0.3243	2.7060	4.1739
V-Gra	$V=1540-1.45*Gra^2+38.42*Gra$	0.3475	2.5655	4.1036
V-Ku	$V=2852+1180*Ku^2-2499*Ku$	0.1970	2.9625	4.5570

coefficient of RF is 0.9370, and the RMSE is only 0.6549, while the correlation coefficient of SVR is 8214, and the RMSE is 2.2188. The RF has the highest correlation and lowest standard deviation, while the RA has the lowest correlation and highest standard deviation.

IV. DISCUSSION

Compared with linear regression, the correlation coefficient, STDEV, and RMSE of the one-parameter quadratic equation can be improved (Table 3), indicating that the correlation of the quadratic equation was higher than that of

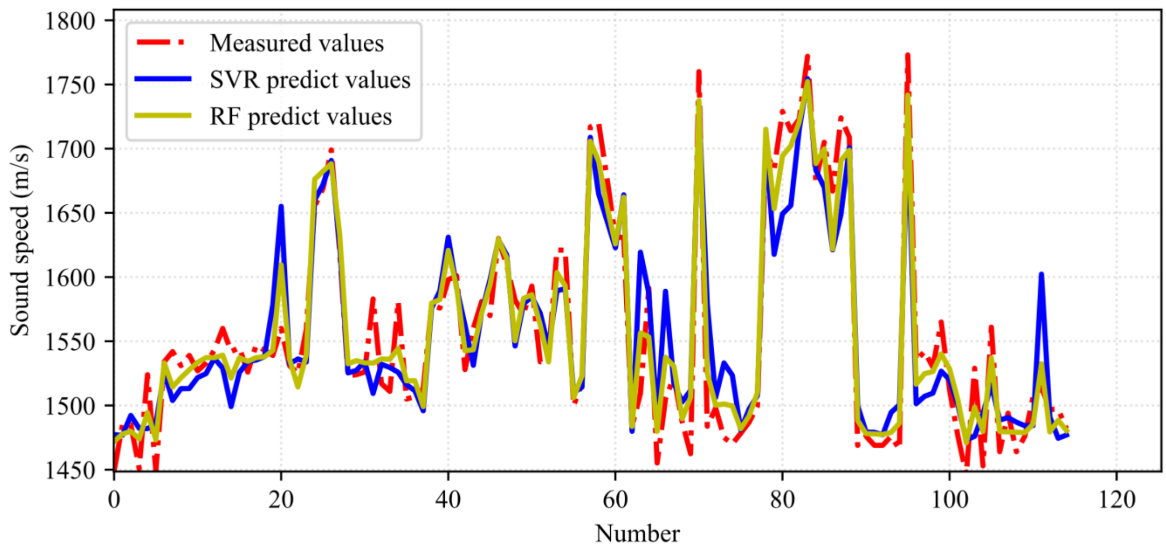

FIGURE 2. The predicted sound speed by RF and SVR.

TABLE 4. Comparison between multivariate quadratic equation and the predictive model established by RF and SVR.

Parameter	Correlation coefficient	STDEV	RMSE
V-Sand-Md	0.7646	1.5987	2.5316
V-Sand-Mz	0.7724	1.5092	2.4718
V-Silt-Mz	0.7716	1.5907	2.5005
V-Silt-Md	0.7679	1.6208	2.5179
V-Md-Mz-Sand	0.7816	1.4352	2.4267
V-Md-Mz-Silt	0.7751	1.5688	2.4762
V-Mz-Silt-Clay	0.7837	1.5076	2.4184
RF	0.9370	0.4306	0.6549
SVR	0.8214	1.5052	2.2188

the linear equation. Although the one-parameter quadratic equation improved the prediction accuracy for the sound speed, it is known that a single parameter cannot completely describe the sediment properties. Hence, multi-parameter models should be established to predict the sound speed of sediment. However, as the number of parameters increases, the amount of required calculations increases sharply, and over-fitting of the regression equation can easily occur. As introduced in Section 2, the ML method (RF and SVR) can combine nine parameters and analyze the intrinsic relationship between each parameter in order to establish sound speed predictive models.

The comparisons indicated that the correlation coefficient of RF and SVR model is higher than regression equations, and the STDEV and RMSE are lower than regression equations, which means that both RF and SVR have a higher accuracy than regression equations. This is because ML method can run efficiently on large databases and it can handle thousands of input variables without variable deletion. In fact, as the number of features increases, the prediction accuracy of ML method is higher. Compared with the regression method, the ML method not only has a fast training speed, but also has a simple implementation, and can comprehensively consider various factors to have higher accuracy. As shown in Table 3 and 4, the accuracy of SVR has only improved a little, while the correlation coefficient of RF is up to 0.9370 and the RMSE is only 0.6549, which indicated that RF method has a higher prediction accuracy than SVR. Figure 2 shows the comparisons of the measured sound speed with the predicted sound speed by RF model and SVR model. The yellow line is the RF predict sound speed, which is more close to the measured values. The comparisons between RF and SVR also show that the RF model has extremely high prediction accuracy, and is more suitable for the sound speed regression in this study.

Another advantage of the RF approach is that it is an effective method for estimating missing data and is still accurate when a large proportion of the data are missing. For example, randomly chose a set of granularity data (Gra 6.71%, Sand 62.18%, Silt 26.33%, Clay 4.79%, Mz 3.14, Md 2.50, Sk -0.28, Ku 1.22, and So 2.60) for a sediment that showed an experimental sound speed of 1717 m/s. When data was not missing, the complete RF model predicted a speed of 1709.88 m/s, with a prediction error of 0.41%. In the case of missing data, the predicted value was recalculated and the prediction errors were compared (Figure 3). When Gra data was missing, the prediction error of the RF model was 0.41%,

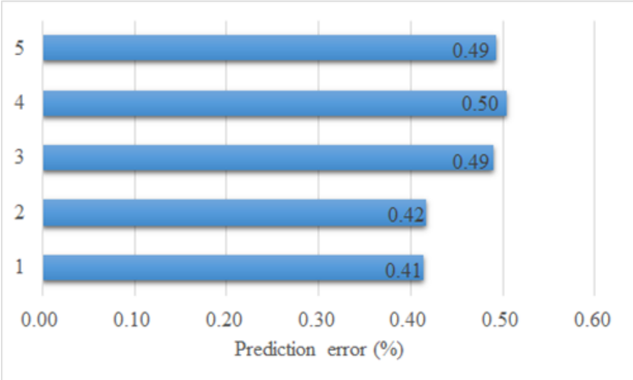

FIGURE 3. Prediction error comparison of the RF model. 1 is the complete RF model, 2 is the error of missing Gra data, 3 is the error of missing Gra and So data, 4 is the error of missing Gra, Mz and So data, 5 is the error of missing Gra, Mz, Ku and So data.

similar to that of the complete data set. When both Gra and So data were missing, the prediction error was 0.42%. When Gra, Mz, and So data were missing, the prediction error was 0.49%. When Gra, Mz, Ku, and So data were missing, the prediction error was 0.50%. When the Gra, Mz, Ku, Silt, and So data were missing, the prediction error was 0.49%. Hence, even in the case where most of the data was missing, the RF model showed high prediction accuracy, which is of great significance for practical applications.

Although RF can maintain high accuracy in the absence of partial data, it is also possible to reduce accuracy when important data is missing, so the important data should be ensured not missing. Here the feature selection method was used to estimate the importance of the nine parameters in RF model. The principle of the feature selection method is out of bag (OOB) error. the following steps was used to calculate OOB error:

(1) In the RF model, the sound speed in ever decision tree was calculated, then the error between measured sound speed and predicted sound speed can be calculated, and it was called ERROR A.
(2) Randomly add a certain noise to each feature, and calculate the corresponding error again for the noise-added data. It was called ERROR B.
(3) Then calculate the difference between ERROR A and ERROR B. The larger the difference, the more important the current feature is.

According to the above process, the importance of the nine granularity data in the RF model was calculated (Figure 4). The results show that Mz is the most importance feature that affects the sound speed in the RF model in this article, and the effect of these nine parameters on the sound speed is as follows: Mz > Silt > Sand > Clay > Md > Sk > So > Gra > Ku.

This result has confirmed the that the mean grain size is a more important variable than the median grain size for determining some physical and acoustic properties [15]. The Mz is the weighted average value of various particle sizes, the Md represents the particle size value accounting for 50%

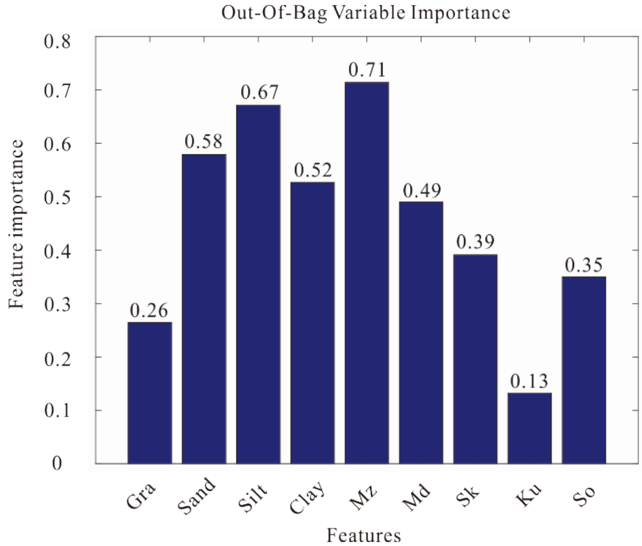

FIGURE 4. The OOB error of the RF model.

of the total, the So represents the degree of particle distribution uniformity, the Ku generally refers to the sediment source environment indication of a single material source or mixture source, and the Sk represents the genesis of sediment thickness (river, coastal tide, wind, etc.). Four particle content parameters represent the weight percentage of each particle. These 9 parameters are all parameters representing particle size, composition distribution and sorting characteristics, and the Mz is usually considered as the most effective parameter to describe particle size characteristics [3].

Theoretically, the pore structure of sediment is determined by particle size and shape. Many parameters of Biot porous theory, including permeability, pore size and tortuosity, need to be calculated by Mz. From this point of view, particle size has an impact on the propagation characteristics of acoustic wave in sediment through its impact on porosity, density and permeability, and it is reasonable to characterize the impact by Mz [20].

V. CONCLUSION

The two ML(RF and SVR) algorithm were used with nine granularity indexes as inputs to establish a model for predicting the sound speed in seafloor sediment. The speed-granularity model is thought to be advantageous as granularity data remains unchanged when sampled from the seabed environment to the laboratory, and not affected by environmental factors. Secondly, a large amount of granularity data is available in the existing literature that can be directly used in sound speed modeling studies. Hence, it is possible to use previously acquired granularity data to predict the sound speed of sediment. Moreover, the predictive model established in this study based on the RF method can accurately predict the sound speed of sediments. It needs be pointed out that the model was established using the laboratory sound speed in this study to examine the availability of RF method. It will be suggested that the model can predict the sound speed more effectively when the in situ sound speed was used to establish the similar model by RF method. Compared with the traditional regression equations, the RF model has the following significant advantages:

(i) The effects of multiple factors on the sound speed can be fully considered;
(ii) In practical applications, even if some parameters are missing, the sound speed still can be accurately predicted;
(iii) In future studies, the model can be further optimized with new input data to improve prediction accuracy, such as the in situ sound speed and porosity data.

REFERENCES

[1] Z. A. Al-Sudani, S. Q. Salih, A. Sharafati, and Z. M. Yaseen, "Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation," *J. Hydrol.*, vol. 573, pp. 1–12, Jun. 2019.
[2] R. T. Bachman, "Acoustic and physical property relationships in marine sediment," *J. Acoust. Soc. Amer.*, vol. 78, no. 2, pp. 616–621, 1985.
[3] S. H. Bae, D. C. Kim, G. S. Lee, G. Y. Kim, S. P. Kim, Y. K. Seo, and J. C. Kim, "Physical and acoustic properties of inner shelf sediments in the South Sea, korea," *Quaternary Int.*, vol. 344, pp. 125–142, Sep. 2014.
[4] M. Belgiu and L. Drăgut, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, Apr. 2016.
[5] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[6] K. B. Briggs and M. D. Richardson, "Small-scale fluctuations in acoustic and physical properties in surficial carbonate sediments and their relationship to bioturbation," *Geo-Mar. Lett.*, vol. 17, no. 4, pp. 306–315, 1997.
[7] M. J. Buckingham, "Precision correlations between the geoacoustic parameters of an unconsolidated, sandy marine sediment," *J. Comput. Acoust.*, vol. 9, no. 1, pp. 101–123, 2001.
[8] A. Charnes, E. L. Frome, and P. L. Yu, "The equivalence of generalized least squares and maximum likelihood estimates in the exponential family," *J. Amer. Stat. Assoc.*, vol. 71, no. 353, pp. 169–171, 1976.
[9] N. P. Chotiros and M. J. Isakson, "A broadband model of sandy ocean sediments: Biot–Stoll with contact squirt flow and shear drag," *J. Acoust. Soc. Amer.*, vol. 116, no. 4, pp. 2011–2022, 2004.
[10] M. Endler, R. Endler, J. Wunderlich, B. Bobertz, T. Leipe, M. Moros, J. B. Jensen, and H. W. Arz, "Geo-acoustic modelling of late and post-glacial sedimentary units in the Baltic Sea and their acoustic visibility," *Mar. Geol.*, vol. 376, pp. 86–101, Jun. 2016.
[11] E. L. Hamilton, "Low sound velocities in high-porosity sediments," *J. Acoust. Soc. Amer.*, vol. 28, no. 1, p. 16, 1956.
[12] E. L. Hamilton, "Sound velocity and related properties of marine sediments, North Pacific," *J. Geophys. Res.*, vol. 75, no. 23, pp. 4423–4446, 1970.
[13] E. L. Hamilton and R. T. Bachman, "Sound velocity and related properties of marine sediments," *J. Acoust. Soc. Amer.*, vol. 72, no. 6, pp. 1891–1904, 1982.
[14] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.
[15] D. R. Horn, B. M. Horn, and M. N. Delach, "Correlation between acoustical and other physical properties of deep-sea cores," *J. Geophys. Res.*, vol. 73, no. 6, pp. 1939–1957, 1986.
[16] Z. Y. Hou, J. Q. Wang, Z. Chen, W. Yan, and Y. H. Tian, "Sound velocity predictive model based on physical properties," *Earth Space Sci.*, vol. 6, no. 8, pp. 1561–1568, 2019.
[17] Z. Hou, Z. Chen, J. Wang, X. Zheng, W. Yan, Y. Tian, and Y. Luo, "Acoustic characteristics of seafloor sediments in the abyssal areas of the South China Sea," *Ocean Eng.*, vol. 156, pp. 93–100, May 2018.
[18] Z. Hou, C. Guo, J. Wang, W. Chen, Y. Fu, and T. Li, "Seafloor sediment study from south China sea: Acoustic & physical property relationship," *Remote Sens.*, vol. 7, no. 9, pp. 11570–11585, 2015.
[19] Z. Huang, J. Siwabessy, H. Cheng, and S. Nichol, "Using multibeam backscatter data to investigate sediment-acoustic relationships," *J. Geophys. Res., Oceans*, vol. 123, no. 7, pp. 4649–4665, 2018.
[20] D. Jackson and M. D. Richardson, *High-Frequency Seafloor Acoustics*. New York, NY, USA: Springer, 2007.
[21] G. Y. Kim and D. C. Kim, "Comparison and correlation of physical properties from the plain and slope sediments in the Ulleung Basin, East Sea (Sea of Japan)," *J. Asian Earth Sci.*, vol. 19, no. 5, pp. 669–681, 2001.
[22] G. Y. Kim, D. C. Kim, D. G. Yoo, and B. K. Shin, "Physical and geoacoustic properties of surface sediments off eastern Geoje Island, South Sea of korea," *Quaternary Int.*, vol. 230, nos. 1–2, pp. 21–33, 2011.
[23] S. R. Kim, G. S. Lee, D. C. Kim, S. H. Bae, and S. P. Kim, "Physical properties and geoacoustic provinces of surficial sediments in the southwestern part of the Ulleung Basin in the East Sea," *Quaternary Int.*, vol. 459, pp. 35–44, Nov. 2017.
[24] S. Kotsiantis, "Combining bagging, boosting, rotation forest and random subspace methods," *Artif. Intell. Rev.*, vol. 35, no. 3, pp. 223–240, 2011.
[25] B. Liu, T. Han, G. Kan, and G. Li, "Correlations between the *in situ* acoustic properties and geotechnical parameters of sediments in the Yellow Sea, China," *J. Asian Earth Sci.*, vol. 77, no. 21, pp. 83–90, 2013.
[26] X. Meng, G. Li, G. Han, and G. Kan, "Sound velocity and related properties of seafloor sediments in the Bering Sea and Chukchi Sea," *Acta Oceanol. Sinica*, vol. 34, no. 5, pp. 75–80, 2015.
[27] M. Nooraiepour, N. H. Mondol, H. Hellevang, and K. Bjørlykke, "Experimental mechanical compaction of reconstituted shale and mudstone aggregates: Investigation of petrophysical and acoustic properties of SW Barents Sea cap rock sequences," *Mar. Petroleum Geol.*, vol. 80, pp. 265–292, Feb. 2017.
[28] T. H. Orsi and D. A. Dunn, "Correlations between sound velocity and related properties of glacio-marine sediments: Barents Sea," *Geo-Mar. Lett.*, vol. 11, no. 2, pp. 79–83, 1991.
[29] M. D. Richardson and K. B. Briggs, "Empirical predictions of seafloor properties based on remotely measured sediment impedance," in *Proc. AIP Conf.*, vol. 728, 2004, pp. 12–21.
[30] L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda, "The CART decision tree for mining data streams," *Inf. Sci.*, vol. 266, no. 5, pp. 1–15, May 2014.
[31] B. C. Schreiber, "Sound velocity in deep sea sediments," *J. Geophys. Res.*, vol. 73, no. 4, pp. 1259–1268, 1968.
[32] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.
[33] J. Soete, L. M. Kleipool, H. Claes, S. Claes, H. Hamaekers, S. Kele, M. Özkul, A. Foubert, J. J. G. Reijmer, and R. Swennen, "Acoustic properties in travertines and their relation to porosity and pore types," *Mar. Petroleum Geol.*, vol. 59, pp. 320–335, 2015.
[34] G. H. Sutton, H. Berckhemer, and J. E. Nafe, "Physical analysis of deep sea sediments," *Geophysics*, vol. 22, no. 4, pp. 779–812, 1957.
[35] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: A classification and regression tool for compound classification and QSAR modeling," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 1947–1958, 2003.
[36] V. Vapnik and A. Lerner, "Pattern recognition using generalized portrait method," *Automat. Remote Control*, vol. 24, pp. 774–780, May 1963.
[37] L. Xu, K. Yang, and Q. Yang, "Joint time-frequency inversion for seabed properties of ship noise on a vertical line array in South China Sea," *IEEE Access*, vol. 6, pp. 62856–62864, 2018.
[38] H. Yang and W. Seong, "High frequency compressional wave speed and attenuation measurements in water-saturated granular media with unimodal and bimodal grain size distributions," *J. Acoust. Soc. Amer.*, vol. 143, no. 2, p. 659, 2018.
[39] Z. M. Yaseen, M. Fu, C. Wang, W. H. M. W. Mohtar, R. C. Deo, and A. El-Shafie, "Application of the hybrid artificial neural network coupled with rolling mechanism and grey model algorithms for streamflow forecasting over multiple time horizons," *Water Resour. Manage.*, vol. 32, no. 5, pp. 1883–1899, 2018.
[40] H. Yong, Z. Yanli, G. Bincheng, H. Wenxiang, and G. Zhaopu, "Characteristics of the impedance variation in clastic rock reservoirs and lithology interpretation method for the threshold volume," *Mar. Petroleum Geol.*, vol. 97, pp. 277–287, Nov. 2018.
[41] J.-X. Zhou, X.-Z. Zhang, and D. P. Knobles, "Low-frequency geoacoustic model for the effective properties of sandy seabottoms," *J. Acoust. Soc. Amer.*, vol. 125, no. 5, p. 2847, 2009.
[42] D. Zou, K. L. Williams, and E. I. Thorsos, "Influence of temperature on acoustic sound speed and attenuation of seafloor sand sediment," *IEEE J. Ocean. Eng.*, vol. 40, no. 4, pp. 969–980, Oct. 2015.